SINGLE EVENT EFFECTS IN COMMERCIAL MICROPROCESSORS USING DYNAMIC CIRCUITRY

By

Xiaowei Zhu

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

May, 2002

Nashville, Tennessee

Approved by:

Prof. Lloyd Massengill

Prof. Bharat Bhuva

Prof. Daniel Fleetwood

Prof. Ronald Schrimpf

Prof. J. Fritz Barnes

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS iii

LIST OF TABLES v

LIST OF FIGURES vi

Chapter

I. INTRODUCTION 8

II. SOFT ERROR ISSUES IN MICROPROCESSORS 12

Single Event Basics 12
Dynamic Circuit Operation 15
Emerging SEE Issue with Technology Development and Current Research Status 18

III. THE ALPHA MICROPROCESSORS UNDER INVESTIGATION 24

General Information 24
Core Combinational Logic Design 27
Core Latch Design 28
Cache Design 32

IV. METHODOLOGY 35

Computational Method 35
Simulation Setup 39
Experimental Details 42

V. RESULTS AND DISCUSSION 46

Device Level – Q_{crit} and Charge Collection Efficiency 46
Circuit Level – Circuit Design Style 47
Frequency Dependence 57
Overall Chip-level SER 61

VI. CONCLUSION 65

ACKNOWLEDGMENTS

According to Webster, "acknowledgment" means "to express thanks for". To justifiably acknowledge those individuals who have helped me to reach this major accomplishment in my life, I would need another volume attached to my final thesis.

First I would like to thank my advisor, Dr. Lloyd W. Massengill for his guidance and for being such a good teacher. I am also indebted to Dr. Ron Schrimpf and Dr. Dan Fleetwood for their constant encouragement as well as constructive and insightful comments on this dissertation. I would also like to thank Dr. Bharat Bhuva and Dr. J. Fritz Barnes for serving on my committee, and Dr. Tim Holman for his unselfish support.

I would like to thank Compaq Computer Corporation, Alpha Development, for the opportunity to work with their SER team as an intern, and for providing me with the superb environment at Shrewsbury, Massachusetts. I am deeply indebted to Dr. Norbert Seifert, who not only contributed a great deal to this work, but also continuously encouraged and supported me during and after my stay at Compaq/ADG. I would like to thank my colleagues at Compaq/ADG. Special mention is due to Vinod Ambrose, Manjunath Bhat, Jim Chung, Heinrich Fisher, Ray Hokinson, Norm Leland, Phil Lin, David Moyer, Rob Mueller, Johann Nittmann, Llanda Richardson, Wendy Qin, Yuzhen Tang, Horst Wagner, and Bjorn Zetterlund, Tina Zou.

Deep gratitude is given to Dr. Claude Cirba for his patient support and guidance toward solving many problems encountered with the computer simulations and for his invaluable friendship. I would also like to acknowledge the help and friendship of Dr. Yuanfu Zhao and Dr. Sungchul Lee.

I am grateful to my fellow graduate students in the Vanderbilt Radiation Effects and Reliability Group, who were instrumental in providing such a constructive environment. Of special mention would be Scotter Ball, Hugh Barnaby, Younes Boulghassoul, Quping Hu, Daedra Lohr, Jason Lowe, Jim Felix, Yanfeng Li, Jinhong Liu, Jie Meng, Jayanth Shreedhara, Andrew Sternberg, Kevin Warren, and Hao Xiong. I would like to thank Connie Nichols and Pitsini Monkolkachit, very special friends of mine, who make me feel at home in America.

I would like to offer special thanks to my pingpong and badminton friends who made my graduate study life at Vanderbilt a beautiful memory. Of special mention would be Philippe

Adell, Kitti Attakitmongcol, Lei Guo, Yanfeng Li, Chris Lossin, Jingchun Ma, Yuansheng Yang, Xinhauai Ye, and Tong Zou.

Finally I want to thank my family, especially my mom and my brother, for their continuous support during all the years of my study. To reach this goal in my life would not have been possible without their unconditional love.

LIST OF TABLES

Table Page

1: SIA Roadmap of Microprocessor technology [1] 8

2: Comparison of dynamic logic and static CMOS logic........ 17

3: Probability of an error according to arrival time of the transient pulse 21

4: Microprocessor Features [14] 24

5: Critical charge of comparison (normalized unit) 50

6: Summary of simulation results at 20 MHz for nodes shown in Figure 15 59

7: Nomenclature used in Figure 37 [20]........ 62

LIST OF FIGURES

Figure Page

1: Microprocessor clock speed predicted by SIA road map of [1] 9

2: Single event occurs at PN junction of the semiconductor device, and is modeled as a transient current pulse at the circuit level 13

3: Soft error in memory circuits and combinational circuits 13

4: Concept of critical charge and window of vulnerability 14

5: The basic construction of a dynamic logic gate 15

6: Definition of setup and hold time of a latch 20

7: Schematic of a generic data path consisting of a inverter chain between two latches 22

8: Timing requirement for a voltage glitch occurrence at node S to be latched as a soft error 23

9: 21164 die photo[14] 25

10: 21264 die photo[14] 26

11: Dynamic inverter 28

12: 21164 Latch 29

13: 21264 Latch with no keeper 30

14: 21264 Latch with keepers 32

15: Circuit schematic of the SRAM sensing data-path 33

16: The circuit schematic of an SRAM cell 34

17: Probability density function and cumulative density function of collected charge for 5 MeV alpha particles[29] 37

18: SE transient current temporal profile for circuit-level simulations 39

19: Simulation setup of evaluating the SE vulnerability of the circuit in SPICE 41

20: Flow chart of the simulation method to find the critical charge 42

21: 21164 Alpha microprocessor with several memory arrays covered with epoxy (six gray rectangles in the middle of the chip) to mask the alpha particle 44

22: Measured SER of on-chip cache as a function of operating voltage for two different technologies. Circled values indicate the extrapolated values to the respective nominal supply voltages 46

23: Simulated critical charge of the SRAM cell for two different technologies 47

24: Experimental accelerated SER of core logic in two generations of Alpha microprocessors 49

25: Simulated critical charges are plotted for two different nodes of the 21164 core logic data-path in one clock cycle ... 51

26: The WOV decreases with increasing clock frequency, which results in a corresponding decrease in SER ... 52

27: Simulated critical charge of latch designs in 21264 Alpha microprocessor core logic ... 53

28: Computed failure rates are shown as a function of clock frequency for two types of latch designs ... 54

29: Simulated critical charge of various keeper designs ... 55

30: Computed failure rates are shown as a function of clock frequency for various keeper designs ... 56

31: Experimental accelerated SER as a function of clock frequency for the RETENTION pattern for 21164 Alpha microprocessor ... 58

32: Experimental accelerated SER as a function of clock frequency for the RETENTION pattern for 21264 Alpha microprocessor ... 58

33: Simulated critical charge of internal node of SRAM cell under reading/sensing condition at low frequency ... 60

34: Simulated critical charge of internal node of SRAM cell under reading/sensing condition at high frequency ... 60

35: Experimental accelerated SER of the TOAST pattern run on 21164 Alpha microprocessor is shown as a function of clock frequency with and without memory arrays covered ... 61

36: Experimental accelerated SER of the TOAST pattern run on 21264 Alpha microprocessor is shown as a function of clock frequency with and without memory arrays covered ... 62

37: The FIT rates due to α-particle strikes are plotted for several different Alpha microprocessor generations [21] ... 63

CHAPTER I

INTRODUCTION

The semiconductor industry has gone through an astounding revolution in the last few decades; the performance and integration density of microprocessors increase exponentially with time due to staggering technical advances in semiconductor fabrication process, circuit design and computer architecture. Table 1 shows the Semiconductor Industry Association roadmap [1] for microprocessor technology. As feature size continues to shrink, more transistors can be packed into the same chip area, therefore the transistor count per chip increases.

Table 1: SIA Roadmap of Microprocessor technology [1]

Product Year	1997	1999	2001	2003	2006	2009
Technology Generation	250nm	180nm	150nm	130nm	100nm	70nm
Microprocessor transistors/chip	11M	21M	40M	76M	200M	520M
Logic transistors/cm^2	3.7M	6.2M	10M	18M	39M	84M
NOR Cell Size (μm^2)	0.6	0.3	0.22	0.15	0.08	0.04
Nominal Supply Voltage (V)	1.8	1.5	1.2	1.0	0.9	0.6
Gate Delay Metric CV/I (ps)	16-17	12-13	10-12	9-10	7	4-5

Furthermore, the decreasing gate delay gained by the technology-scaling enables the on-chip clock speed to continue to increase. One way to improve the performance of a microprocessor is to increase the operating clock frequency. The demand for higher performance has resulted in clock speeds on the order of several GHz in modern high performance microprocessors, and this trend is forecast to continue in future generations, as is shown in Figure 1 [1].

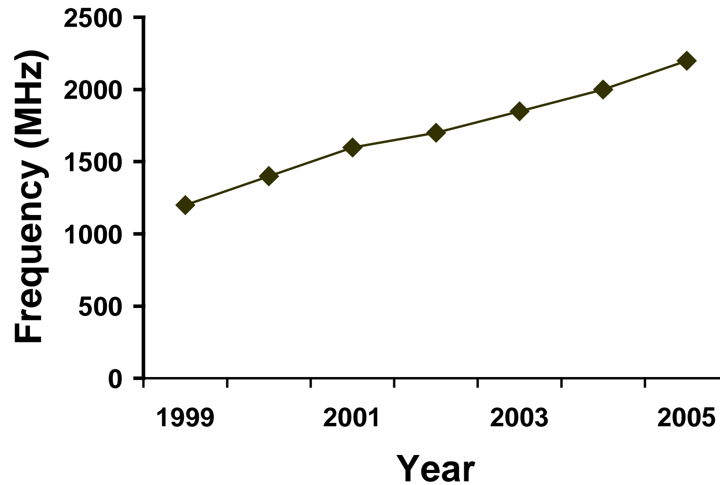


Figure 1: Microprocessor clock speed predicted by SIA road map of [1]

As feature sizes and operating voltages of electronic devices are continually reduced to satisfy the demand for higher density of integration and lower power consumption, the sensitivity of systems using these devices often increases. Radiation-induced charges can corrupt the information (presence or absence of pre-determined charge at circuit nodes) stored in a circuit. Because this type of error is nonpermanent, it is called soft error. This phenomenon is also called single event, since it is a localized effect caused by a single particle. The critical charge Q_c, often approximated as the product of the node capacitance C_{node} and the operating voltage V_{DD}, has been reduced to levels such that circuits are now vulnerable to low energy, ground based ions [2]. Radiation-induced reliability issues are no longer the sole concern of space circuit and device designers. Terrestrial cosmic rays and their interaction with electrical systems are attracting more and more attention. Single event effects (SEEs) are taking a prominent position in the mainstream integrated circuit industry. Many commercial manufacturers are coming to grips with the problem as a key reliability issue.

Advancement from one microprocessor generation to the next impacts reliability at the technology/device level, circuit level, and system level; therefore, the broad objective of this work is to study the impact of the Alpha[1] microprocessor maturation between two generations on soft error rates (SER) at three levels. At the device level, analysis is done on the error rate per information bit to weigh the relative impact of critical charge and charge collection efficiency.

[1] Compaq is registered in the U.S. Patent and Trademark Office. Alpha is a trademark of Compaq Computer Corporation.

Contrary to many researchers' beliefs that SER will increase in future technologies, the results obtained in this work indicate that on the device level the SER per on-chip memory cell decreases, rather than increases. At the circuit level, the core logic SER of these two generations is compared to analyze the impact of different circuit design techniques. At the system level, SER frequency trends are studied in major microprocessor components, such as the cache and core logic. For the 21164 Alpha core logic using dynamic transmission gate latch, unexpectedly the SER decreases with the clock speed. This trend is reported for the first time in this work and is well explained by the simulation analysis. The influence of architectural decisions such as error correction codes (ECC) on the chip-level SER is also explored. Through the comparison, the dominant SER contribution from the cache over the core logic found in this work justifies the extra cost in using error correction circuitry to protect the on chip cache. This result reinforces the previous observation by Howard et al.[3]. Both the experimental and computational methods are used in this work. Accelerated soft error testing was performed by direct exposure to Am241 foils while a program was run on the Alpha microprocessors with a Schlumberger ITS9000, IX tester. The SER was computed using the results obtained from the circuit level simulations.

The bulk of this work was done at Compaq Computer Corporation Alpha Development Group (ADG) at Shrewsbury, Massachusetts, during my internship in 2001. Some of the analysis described and conclusions drawn in this dissertation are a result of collaborative research efforts with the SER team at Compaq/ADG; however, they do not necessarily reflect the assessment methodology of Compaq/ADG.

This thesis has been organized as follows. Chapter 2 introduces the basic concepts related to the single-event upset (SEU) phenomenon and the basic operation of dynamic circuits, briefly reviews the technology development. Then the emerging soft error (SE) issues in microprocessors as technology advances are discussed, and the related work previously done in this area are presented. Chapter 3 discusses two generations of the Alpha microprocessor family under investigation in this work, briefly reviewing the circuit styles used in these microprocessors and the architecture of the microprocessor from the SE perspective. Chapter 4 explains the method and techniques used to study the problem, both experimentally and computationally. The testing setup and test patterns during the test are presented in detail. In addition, the simulation technique and the single particle hit modeling method used are described, and the SER computation equation is derived and the parameters used in the equation

and their justification are given. Chapter 5 presents experimental data and simulation analysis to explain the experimental observations. The results are organized to address SEE at the device, circuit and system levels. Chapter 6 summarizes all the key findings and concludes this work.

CHAPTER II

SOFT ERROR ISSUES IN MICROPROCESSORS

This chapter contains background information about single event effects in microprocessors. As technology evolves to deliver higher performance microprocessors, dynamic circuits are gaining increasing popularity thanks to its high speed and low area cost. First, the basic concepts and terms related to SEE are introduced. Then, the basic operation of dynamic circuitry is briefly reviewed. Finally, the emerging soft error issues associated with technology development are discussed, along with the related work previously done in this area.

Single Event Basics

Single events and soft errors in electrical systems caused by cosmic rays were first observed by Binder et al. [4] in the bipolar flip-flop circuits of a space-borne communication system in 1975. Soon after this observation in the space environment, May and Wood [5] experimentally observed soft errors in dynamic random access memory (DRAM) operating in the terrestrial environment. To evaluate the vulnerability of integrated circuits to radiation-induced soft errors or single event upsets (SEU) accurately, all sources of radiation must be accounted for. The primary sources of ionizing particles for terrestrial microelectronics include alpha particles from radioisotopic impurities in the packaging material and neutron reaction products induced by cosmic rays [6]. Although used interchangeably in the literature, in this work, single event upset means a corrupted electrical state at any storage node, while a soft error is counted only when the application run on the system misinterprets the corrupted state as valid information.

SEE occur when a single ionizing particle traverses semiconductor device material generating electron hole pairs along its track. Illustrated in Figure 2, A significant fraction of these carriers can be separated by the electric field in the depletion region and result in a transient current pulse at a circuit node [7]. If a single particle hits the memory element of the circuit and the induced current pulse magnitude and duration are sufficiently large, the feedback mechanism in the circuit memory element will cause a change in the logic state of the memory cell.

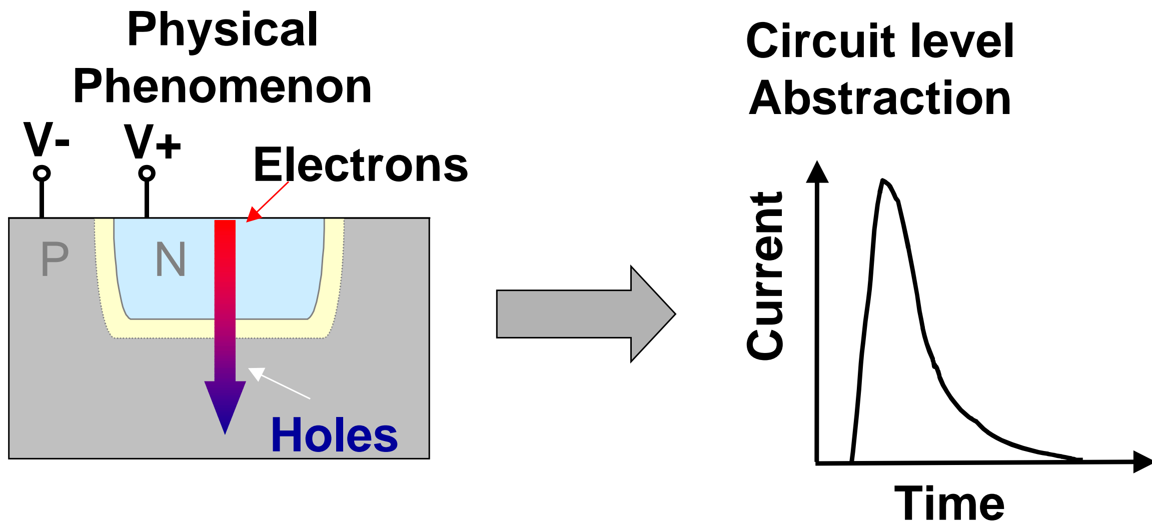


Figure 2: Single event occurs at PN junction of the semiconductor device, and is modeled as a transient current pulse at the circuit level

On the other hand, a single event occurring on combinational circuit nodes will lead to a transient voltage glitch. This voltage glitch can propagate through the logic gates and be latched into a memory element provided that the timing of the arrival and characteristics meet certain criteria. The criteria have been discussed by Baze [8] and Siehl et al. [9], and will be presented in detail later. These two error mechanisms in the circuits are depicted in Figure 3. Since the logic information loss at the circuit nodes due to single events is not permanent and can be corrected, the induced error is called a soft error (SE).

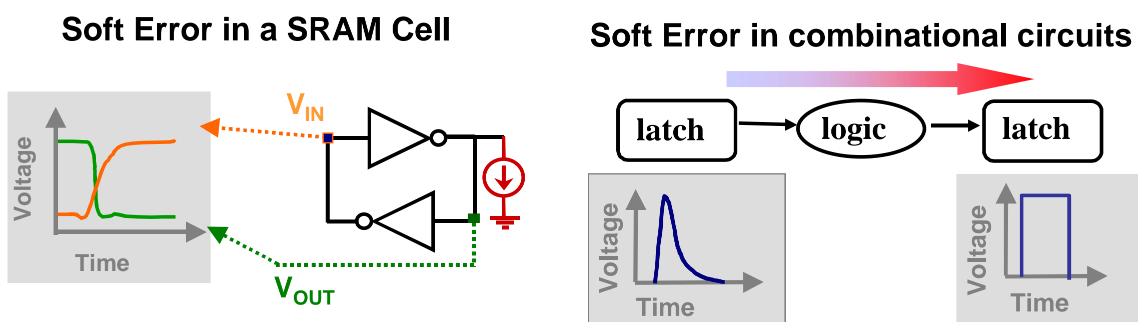


Figure 3: Soft error in memory circuits and combinational circuits

The vulnerability of a circuit is usually indicated by its critical charge (Q_{crit}), defined as the minimum amount of single-event-induced charge needed at a sensitive node to cause a circuit upset. A more detailed discussion about the concept of critical charge has been performed by Massengill [10] and Dodd and Sexton [11]. Because the SER sensitivity of a circuit may vary at

different clock phases, the critical charge of a circuit node is not necessarily constant, and may be a function of time. Another important quantity that characterizes the susceptibility of a memory cell to corruption of its logic state is the window of vulnerability (WOV), defined as the time during which a logic element is sensitive to SEU [12]. These basic concepts are best illustrated with an example shown in Figure 4. Q_{crit} for a certain circuit node can be a function of time. In the example, $Q_{crit}(t)$ is the lowest if the single event hit occurs on this node around a clock rising edge. If the collected charge (Q_{coll}) caused by a certain particle is higher than $Q_{crit}(t)$, a soft error would occur. Therefore, the period of time marked in dark color is the sensitive time window for this circuit node and this collected charge, i.e., the window of vulnerability. As indicated, the region in light-gray is "SE-safe" since the condition $Q_{coll} < Q_{crit}(t)$ holds. The larger the window of vulnerability and the lower Q_{crit}, the higher the soft-error rate (SER). Intuitively, the relation between window of vulnerability and clock cycle duration determines the frequency dependence of SER.

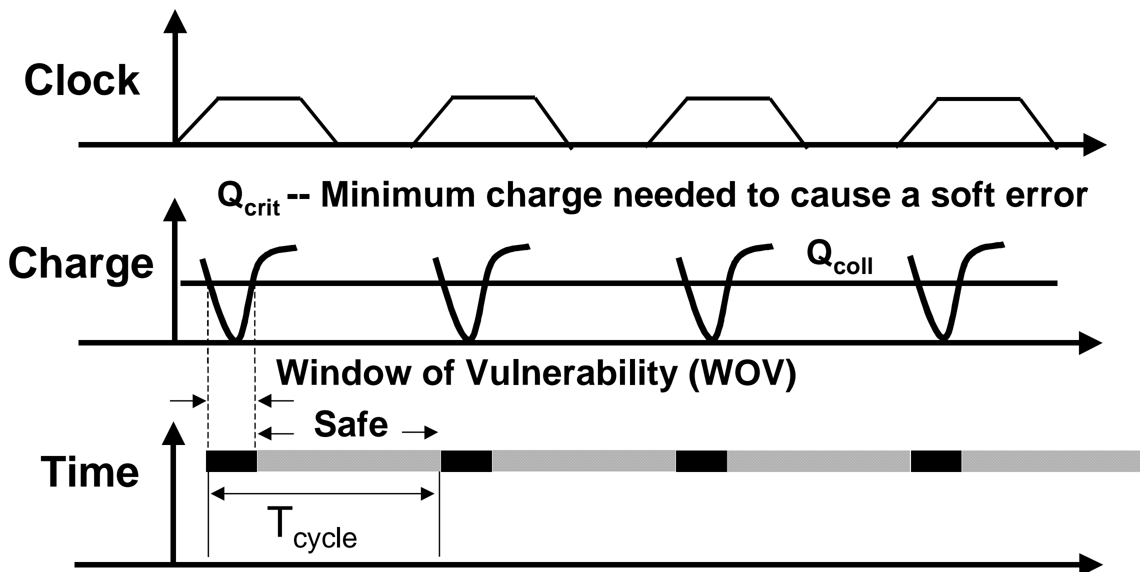


Figure 4: Concept of critical charge and window of vulnerability

Since the critical charge scales with shrinking device sizes and supply voltages, the deep submicron technologies are generally more sensitive than the old technologies. In order to reduce cost and improve performance, dynamic circuits are widely used. The SE vulnerability of dynamic circuits is closely related to its unique way of operation.

<u>Dynamic Circuit Operation</u>

Kang and Leblebici [13] provide more details on the design and issues in dynamic circuits. In this section, a brief review emphasizing issues related to SEU will be presented. Conventional static logic circuits for CMOS technologies use high and low voltages to indicate a logic state 1 or a logic state 0 at the output node. The charge and discharge of the output capacitance is generally carried out by p-channel and n-channel devices, respectively. With hole mobility being smaller than electron mobility, p-channel devices are often larger than n-channel devices. The presence of the large p-channel devices increases capacitive load at the output node of the previous gate resulting in higher power requirements and slower operating speeds.

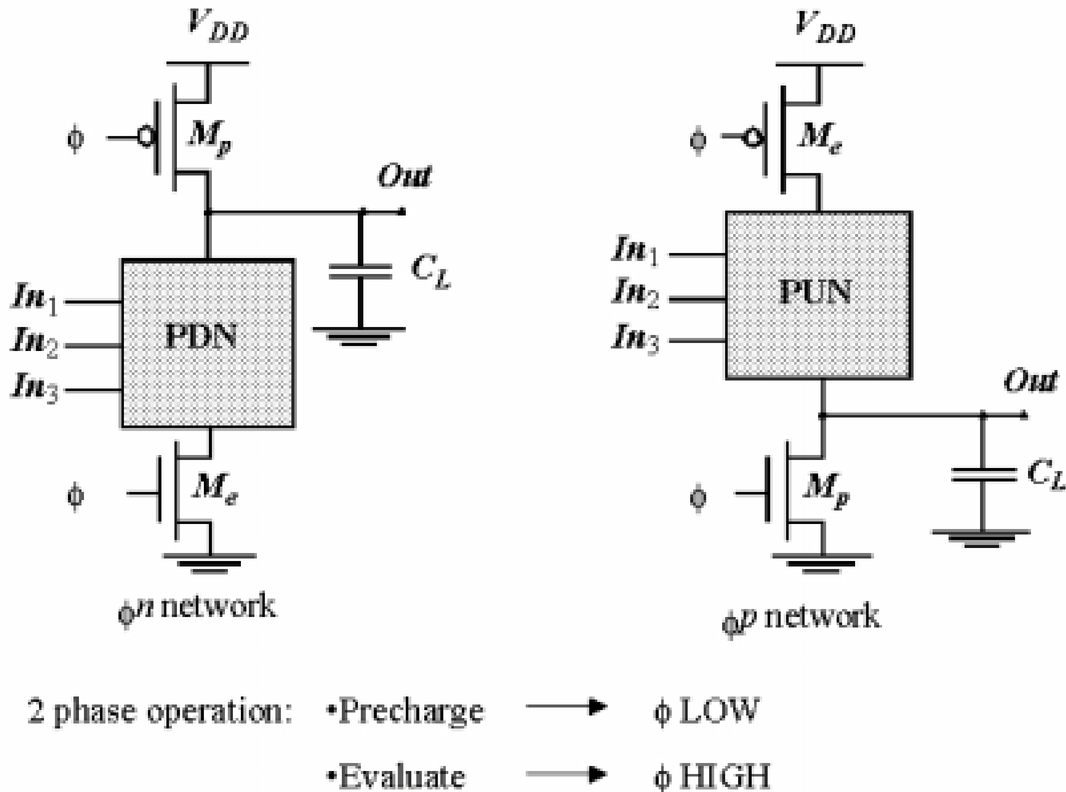


Figure 5: The basic construction of a dynamic logic gate

The basic construction of the dynamic logic circuits is plotted in Figure 5. For dynamic logic circuits, the logic gate operation is divided into two distinct phases of operation: pre-charge phase and evaluation phase. During the pre-charge phase, the output node of a logic gate is charged to a high logic voltage. During the evaluation phase, this output node voltage is either left at a logic state high or discharged to a logic state low. Since input evaluation only

necessitates a capacitance discharge, p-channel devices do not take part in logic function evaluation. Therefore, both the gate fan-in and fan-out capacitance are reduced. Moreover, in a dynamic logic circuit, a single p-type device with high current driving capability replaces the parallel combination of PMOS devices in its equivalents static counterpart. This change reduces the layout area, which results in lower inter-connect capacitance, further increasing the speed of the circuit. The single PMOS device always turns on in the pre-charge phase, and can quickly charge the output capacitance if the NMOS-network is off during evaluation. Furthermore, the switching point of the dynamic gate is set by the threshold voltage of the NMOS device. If the timing of the inputs is such that they are not asserted until after the pre-charge clock has been de-asserted, there is no crossover current during the output transition. Due to these reasons dynamic logic is usually much faster than its complementary equivalents in static logic [14].

The basic construction of an N-type dynamic logic gate and a P-type dynamic logic gate are shown in Figure 5. In an N-type dynamic logic gate, the pull-down network (PDN) is stacked with a clocked NMOS transistor. On the other hand, in a P-type dynamic logic gate, the pull-up network (PUN) is stacked with a clocked PMOS transistor. The pre-charge and evaluation phases of this circuit are controlled by the clock.

Pre-charge

In an N-dynamic gate, when the clock is low (CLK = 0), the output node *Out* is pre-charged to the supply voltage (V_{DD}). During this time, the evaluation NMOS transistor M_e is off. In a P-dynamic gate, when CLK = 1, the output node *Out* is pre-discharged to the ground (GND). During this time, the evaluation PMOS transistor M_e is off.

Evaluation

In an N-dynamic gate, when CLK = 1, the pre-charge transistor M_p is off, and the evaluation transistor M_e is turned on. The output node Out is conditionally discharged based on the PDN. In a P-dynamic gate, when CLK = 0, the pre-charge transistor M_p is off, and the evaluation transistor M_e is turned on. The output node Out is conditionally charged based on the PUN.

The logic state of output node *Out* depends on the amount of charge retained on this capacitive node. If the charge on this node is lost in the evaluation phase, there is no path to recharge the node.

A comparison of some of the important properties of dynamic logic and static CMOS logic gates is shown in Table 2.

Table 2: Comparison of dynamic logic and static CMOS logic

Circuit Type	Device count (a fan-in of N)	static power dissipation	Comments
static CMOS	2N	N	
Dynamic-logic	N+2	N	need clock

Dynamic logic occupies less chip area thanks to a reduced number of transistors per logic function when compared to static CMOS; as a result, it is generally faster due to smaller parasitic capacitors. In addition, dynamic gates do not have short circuit current, and all the current through the PDN or PUN goes into charging and discharging the load capacitance. No static current path exists between V_{DD} and GND in either phase; therefore, dynamic logic only consumes dynamic power. By contrast, there is a flow of current from V_{DD} to GND when the static gates are switching states, at which time both the PMOS and NMOS transistor are ON simultaneously.

During the evaluation phase, the PDN of a dynamic inverter starts to conduct when the input signal exceeds the threshold voltage (V_T) of the NMOS pull-down transistor. Therefore, it is reasonable to set the switching threshold (V_M) the gate to V_T. In addition, V_{IH}, the lowest acceptable high voltage representing logic “1”, and V_{IL}, the highest acceptable low voltage representing logic “0”, can be also approximated by V_T. Therefore, the noise margin for logic “0” is rather low ($=V_T$).

During the evaluation phase of dynamic logic circuit operation, the logic function of the output node relies on capacitive charge storage. As a result, the output node in this time period is very sensitive to single event upset (SEU). There can be no transition from '0' to '1', because the PMOS network is OFF during evaluation. Therefore, no regenerative mechanism exists during the evaluation phase.

For the above two reasons, the output node is extremely vulnerable to single events during the evaluation phase. This high degree of sensitivity to SE exhibited by dynamic circuitry make them unattractive for the space-based applications. However, cost efficiency and speed benefits of dynamic circuitry earn their popularity in the high-performance state-of-art microprocessors.

Emerging SEE Issue with Technology Development and Current Research Status

It is generally believed that the SE susceptibility increases sharply with device downscaling for both memory and logic circuits; thus the soft error rates have become an increasingly important reliability challenge. Single event effects (SEEs) are becoming a more important reliability issue in the commercial semiconductor industry. Intel not only started testing for alpha-induced soft errors in 0.18 μm logic designs, but also is anticipating the development of a strategy for hardening to terrestrial neutron SEEs [15]. IBM started the investigation of SER issue in 1978, and went through several years of research to verify its magnitude. Nowadays, SER study has become a significant factor in IBM's efforts toward improved product reliability [7]. In addition, as the demand for higher performance in space borne systems opens the door for smaller and faster commercial chips, it is beneficial to understand the SER trend with the technology scaling. In the discussion of the coming challenges in microarchitecture, Ronen et al. [16] claim the SER susceptibility increases by 43% per generation. Karnik et al. [17] predict from their study on the SER of the static latch built in 0.18 μm CMOS technology subjected to accelerated neutron beam that the neutron-induced SER is likely to stay constant for future technologies if the supply voltage scales only by 0.8x. However, a study on several SRAM technology generations conducted by Hareland et al. indicates that SER is reducing on a per bit basis for the future technologies. However, in these studies the trend in the susceptibilities of core logic and memory cells have only been investigated separately, and the question of the overall SER of a microprocessor (which is a combination of both contributions) has not been addressed. The goal of this study is to examine the SEE in the commercial state-of-the-art microprocessors under the operational condition, and investigate the impact of scaling at the device, circuit and system levels.

There are two means to improve the performance of the microprocessors, better electrical properties and higher density. The geometric device scaling helps the circuit speed by reducing

the nodal capacitance; at the same time, the supply voltage decreases in order to keep the electrical field reasonably small to avoid the deteriorative long term device characteristics. This trend results in a reduced critical charge. On the other hand, improvements in the fabrication process give the device designers the flexibility to tailor the device structure. Usually the charge collection efficiency is also reduced with the process scaling due to less diffusion area and higher doping level. It remains an open question as to which one of these two competing effects dominates at the device level.

In order to take full advantage of the process revolution and achieve the optimal performance, microprocessor design occurs in parallel with the manufacturing process development. This is the case for the development of the Alpha microprocessor [14]. The concurrent design means the concurrent change in both the process and circuit techniques. Obviously the influence of the circuit construction and configuration on the SER can not be ignored.

Combinational circuits are the basic building blocks of all digital systems. The same logic function can be implemented in various ways according to the performance or cost consideration for the system. The fact that combinational circuits contain non-arrayed logic block with different vulnerabilities to ion-induced perturbation complicates the analysis, since the regularity and vulnerability assumptions on which the analysis of memory circuits are based do not apply.

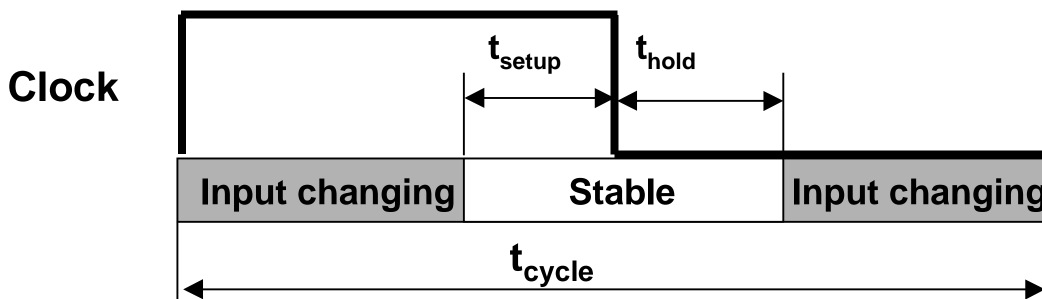
Digital ICs can be divided into two types of circuit blocks: latches (registers) and combinational blocks. Latches incorporate feedback paths to store data; combinational blocks achieve the logic function of the input data. The manifestation of a single ion strike at the circuit level is a transient voltage pulse. Due to the memory-less feature of combinational blocks, the "hit" nodes always return to their original voltage after a certain time. If the next stage is also a combinational block, the erroneous transient voltage, instead of being stored in a latch, may propagate to the next stage, until either latched into a register or attenuated.

In order for a transient pulse to be latched as an error in a register, three conditions have to be met, as proposed by Kaul et al. [18]. First, there must exist a sensitive path from the hit node to a register for this pulse to propagate. For example, consider a NAND gate. In NAND logic, logic '0' is the determining force, i.e., as long as one of the inputs of NAND logic is '0', the output will be logic '1' regardless of the state of other inputs. Therefore, a sensitive path exists for an input to the output only when all other inputs are '1' for NAND logic.

Second, the pulse must have sufficient amplitude and width when it arrives at a register if it does make its way there. The sensitive path from the 'hit' node to the register acts as a filter: it filters out the weak and narrow transient pulses, and allows the strong and wide transient pulses to pass. Intuitively, the pulse shape arriving at the register depends on the original pulse shape, the length of the sensitive path, and the temporal characteristics of the sensitive path.

Third, a pulse has to arrive at the right time to be latched into the register. As the circuits are usually synchronized by a clock signal, the SE transient pulse has to arrive in the 'sensitive time' or 'vulnerable window' to cause an error in the register. The term 'sensitive time' and 'vulnerable window' are interchangeable and are tightly related to the 'set-up and hold time' of the register.

Analysis of the first condition can be done in SPICE in a straightforward manner, while analysis of the second and third conditions is more involved. The general guidelines can be found the work on attenuation of single event induced pulses in CMOS combinational logic by Base and Buchner [8]. A transient voltage pulse caused by a single particle with amplitude less than the switching threshold voltage of the logic gate cannot propagate to the next stage. A pulse with sufficient amplitude has to last long enough to propagate through the gate. If the width of the pulse is greater than the logic transition time, the pulse will propagate through the gate without attenuation. If the width of the pulse is smaller than half the transition time, it will terminate. The pulses with width in the range between these two values will propagate with attenuation.



If changes on the input violate either setup or hold time, correct latch operation is not guarenteed

Figure 6: Definition of setup and hold time of a latch

The speed of combinational circuits and the distance from the hit node to the register are important parameters for the SE vulnerability analysis of the combinational circuit. The slower

the gate switches, the more stages an error signal has to propagate through to reach a register, and the lesser the possibility for the erroneous signal to lead to an upset.

Each synchronized input has a setup and hold time specification with respect to clock input. Illustrated in Figure 6, setup time is the amount of time the synchronous input has to be stable before the active edge of the clock signal. Hold time is defined [13] as the amount of time the synchronous input has to be stable after the arrival of the active edge of the clock signal.

In order for an SE induced transient voltage pulse to be latched into an edge-triggered register, it not only has to arrive at the right time (around the active clock edge), but also it has to be above the switching threshold voltage of the latch during the setup time and hold time of the latch. The probability that the transient pulse arrives at a time so as to cause an error is shown in Table 3, as proposed by Holland et al. [19].

Table 3: Probability of an error according to arrival time of the transient pulse

Temporal condition	Probability of an error
For transient pulse width <t_{clock}	$\frac{t_{set} + t_{pw} + t_{hold}}{t_{clock}}$
For transient pulse width >=t_{clock}	1.0

In Table 3, t_{pw} is the pulse width of the upset, and t_{set} and t_{hold} are the setup and hold time of the register, and t_{clock} is the period of clock signal for the register. These parameters are illustrated in Figure 6.

Analogous to the sensitive volume in a device, the time period from the beginning of the setup time to the end of the hold time of a register is called 'sensitive time' or 'vulnerable window', because only during this time the erroneous signal has a probability of being latched.

The above analyses suggest that frequency is an important parameter in the SE vulnerability analysis of the circuits, since it is impossible to compare the SE vulnerability of different circuit design styles without specifying their operating frequency. Therefore, the frequency dependence of SER is an issue on both the circuit level and system level.

In static SEU testing of memories such as SRAM's, a test pattern is written into the entire memory beforehand, and serves as a reference for determining the number of SEU's after irradiation with ions of predetermined fluence and linear energy transfer (LET). The upsets found from this kind of static measurement are not affected by clock frequency and thus are

called asynchronous [12] due to their relative insensitivity to the event's timing with respect to the microprocessor system clock [12]. On the other hand, the soft error rate (SER) of combinational logic is dependent on cycle time and consequently referred to as synchronous [18].

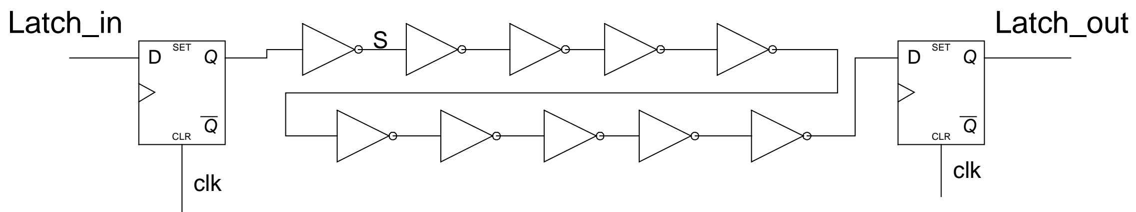


Figure 7: Schematic of a generic data path consisting of a inverter chain between two latches

For example, assuming a single particle hitting the circuit node S in the data path shown in Figure 7, the voltage glitch caused by the single event has to have sufficient amplitude and width in order to propagate through the inverter chain. In addition, in order for an error to be latched in the second latch, the voltage glitch has to arrive at a definite time prior to the clock edge and persist long enough to meet the setup and hold time requirement of the latch. Figure 8 illustrates the timing requirement for a voltage glitch at node S to be latched at the second latch as an error. Usually the single-event-induced transient voltage waveforms have a short rise time of about 50 ps and a fall time of several hundreds of ps. If an ion strikes node S during time period I, the transient voltage pulse seen at the input of the second latch will have decayed by the time of clock transition. On the other hand, if an ion strikes node S during time period II, the clock will have already transitioned when the transient voltage pulse arrives at the second latch after propagating through the inverter chain. In this example, the window of vulnerability of node S is defined with respect to the clock cycle. Therefore, we will expect SER arising from ion strikes at node S to be proportional to the clock speed of the circuit [12].

Buchner et al. have performed substantial work in studying the frequency dependency of logic circuits. They conclude that 1) the soft error rate is independent of the frequency in latch circuits when the setup and hold time is much less than clock period; 2) the soft error rate increases linearly with the operating clock frequency in the combinational circuits; 3) Soft errors in latch circuits dominate in present day technologies, but the errors in the combinational circuits will dominate in future technologies [12]. The purpose of the SER frequency dependence in this

work is two-folds. First, the dynamic latches used in the Alpha microprocessors are quite different from the static latch studied by Buchner et al. Second, as circuit operating frequency approaches GHz, the duration of the transient voltage pulse caused by a SE is comparable to the clock period. It will be advantageous to revisit these early conclusions.

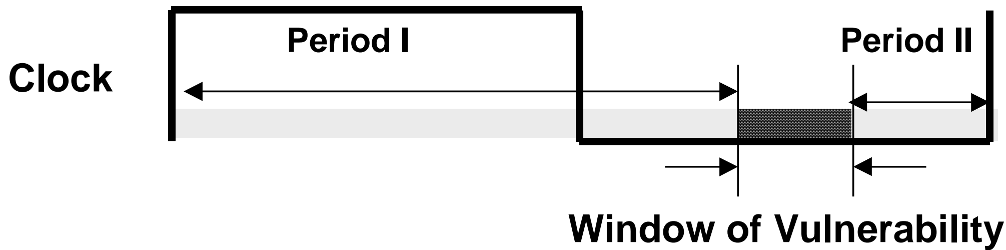


Figure 8: Timing requirement for a voltage glitch occurrence at node S to be latched as a soft error

The above analysis indicates that the overall SER trend with technology generations is complicated by the device and circuit level considerations. The final FIT (failure in time: 1 FIT equals 1 soft error in 10^9 hours of operation) rates of the system are further influenced by the architectural and packaging decisions, and depends strongly on the environment. ECC for the cache was introduced into the Alpha microprocessor in its third generation version, the 21264 chips. For both the 21164 and 21264 Alpha microprocessors, a lid coat or a wirebond attached chip capacitor (WACC) was placed in between the die and the package. The flip chip packaging technology is used in the 21264 Alpha microprocessors built in 0.18 mm CMOS technology. Flip chip microelectronic assembly is the direct electrical connection of face-down (hence, "flipped") electronic components onto substrates, circuit boards, or carriers, by means of conductive bumps on the chip bond pads. Since the flux of the alpha particle increases due to the lead present in the bumps, one would expect a negative impact from the flip chip on the overall chip level SER.

This chapter provides the background information about the SER issues related to the modern commercial microprocessors using dynamic circuits. In the next chapter, the specific dynamic design used in the Alpha microprocessors under the investigation will be discussed.

CHAPTER III

THE ALPHA MICROPROCESSORS UNDER INVESTIGATION

General Information

DIGITAL Equipment Corporation introduced the Alpha 21064 in 1992, the highest performance microprocessor in the industry at that time. Since then Alpha microprocessors have achieved industry-leading performance through process advancements, architectural improvements, and aggressive circuit design techniques [14][20]. The clock frequencies of Alpha microprocessors increased from 150 MHz to more than 1 GHz over four generations from the early 1990's to the present. In addition, Alpha microprocessor design has had reliability as a primary concern. The key parameters for the two generations of Alpha microprocessors under this investigation are listed in Table 4.

Table 4: Microprocessor Features [14]

	21164	21264
Transistor count (million)	9.3	15.2
Die size (mm^2)	18.1×16.5	16.7×18.8
Process Technology	0.35 μm	0.25 μm
Power supply (V)	2.5	2.0
Power dissipation (W)	50	72
Target Design frequency	300	600
Typical gate delay/cycle	14	12
On-chip cache	8KB L1-I 8KB L1-D 96KB L2	64KB L1-I 64KB L1-D

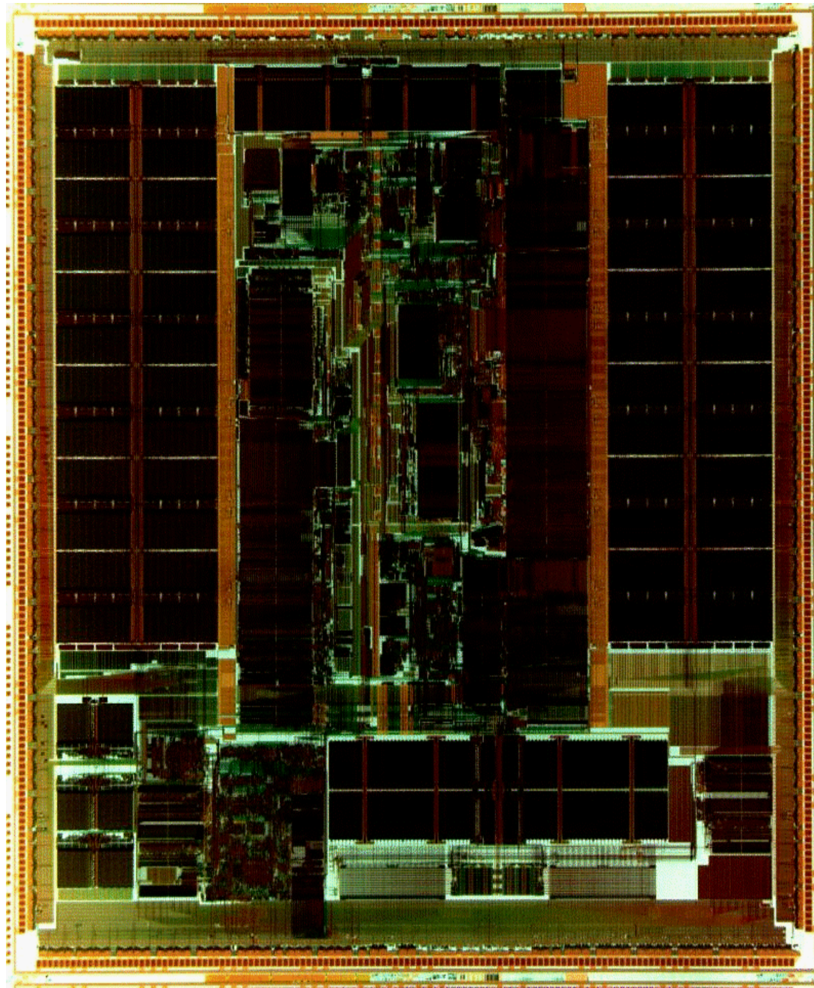


Figure 9: 21164 die photo[14]

The 21164, shown in Figure 9, is the second generation of Alpha architecture. It was designed to operate at 300 MHz in a 0.5 μm n-well CMOS process using a 3.3 V supply, and it dissipates 50 W. The number of gate delays per cycle was reduced from 16 in the first generation to 14 to provide an additional 10% reduction in cycle time beyond process scaling. The die is roughly 3.0 cm^2 and contains a total of 9.3 million transistors. The non-cache transistor count is tripled from the previous generation design to 2.5 million. Although originally designed to operate at 300 MHz, migration of this design to a 0.35 μm process has allowed the operating frequency to be increased to 600 MHz.

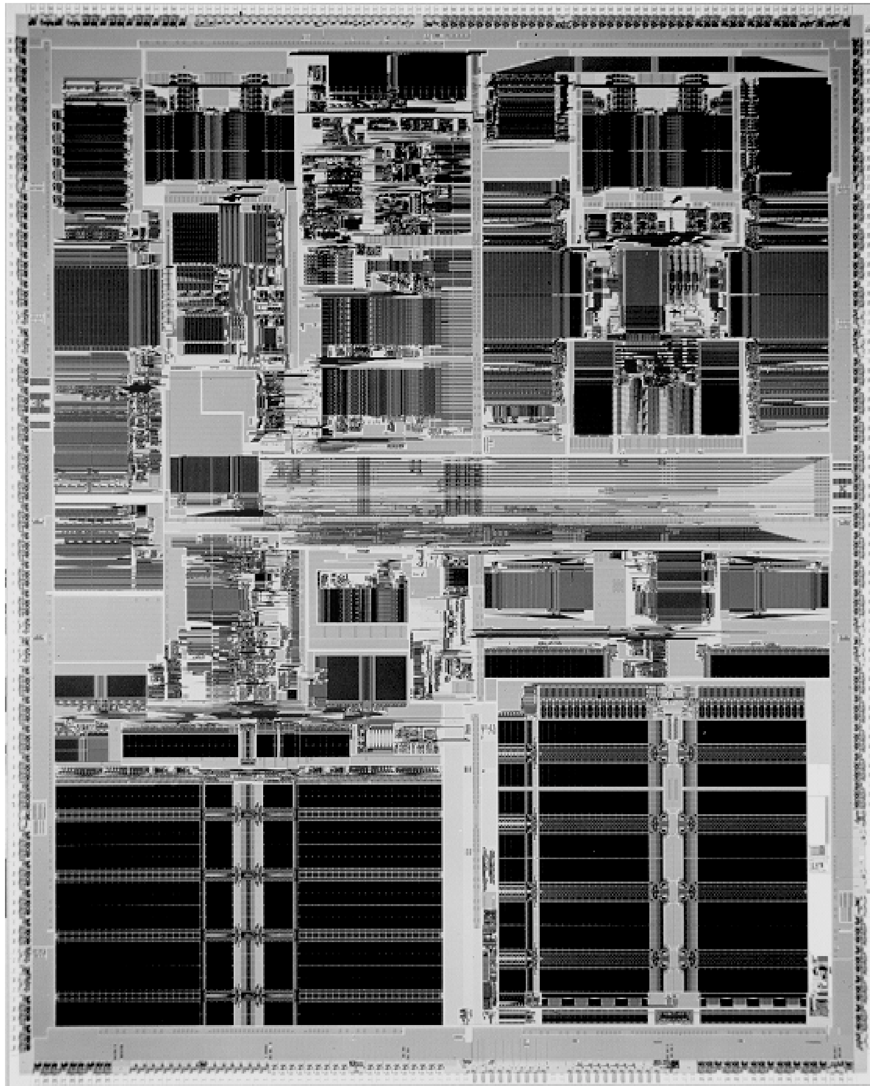


Figure 10: 21264 die photo[14]

The 21264, shown in Figure 10, is the third-generation Alpha microprocessor. It is fabricated in a 0.25 μm n-well CMOS process, and is targeted to operate at 600 MHz. The number of gate delays per cycle has been further reduced to 12, again providing an additional 10% reduction in cycle time relative to the previous design. A nominal supply voltage of 2.2 V is used to limit power dissipation to an estimated 72 W, but the design and process can operate reliably up to 2.5 V. The die is 3.1 cm^2, and contains 15.2 million transistors. The non-cache transistor count is more than double that of the 21164.

As process technology advances, better electrical properties and higher device densities, gained from the process scaling, improve the performance of the microprocessors via two means. First, the speed of the circuit is directly related to the time to charge and discharge the nodal

capacitance, which was reduced by shrinking transistor size and interconnects. Second, geometric scaling enables more circuitry to occupy the same die area, thus offering the designer flexibility to enhance the performance.

In order to achieve higher performance than offered by the above-mentioned process scaling in each generation, full-custom circuit design methodologies have been used universally by the Alpha microprocessor design teams. Fewer than 10% of the circuits were designed using automatic synthesis approaches. Therefore, a wide range of circuit styles including both the conventional complementary static CMOS logic and dynamic logic can be found in these microprocessors [14]. As a result, it is important to identify the circuit type that dominates the SE response of the core logic.

Core Combinational Logic Design

In the core logic of the Alpha microprocessors a wide range of circuit styles have been implemented including conventional complementary CMOS logic, and single- and dual-rail dynamic logic [21]. Dynamic circuits are one of the most commonly used circuit styles due to the advantages discussed earlier; however, dynamic circuits are very sensitive to noise, and require very careful design and extensive verification to ensure functionality.

Figure 11 shows a simple data path consisting of an inverter chain between two latches. The data path itself was not a part of the Alpha microprocessor design, but the circuit elements such as inverters and latches are extracted from the actual implementation to represent the general circuit structure in the core logic. All inverters in the chain are static except one. Controlled by the clock signal CLK, transistors P1 and N1 form a dynamic inverter between the nodes D_{in} and D_{out}. This conditional dynamic inverter is commonly used in both generations of the Alpha microprocessor.

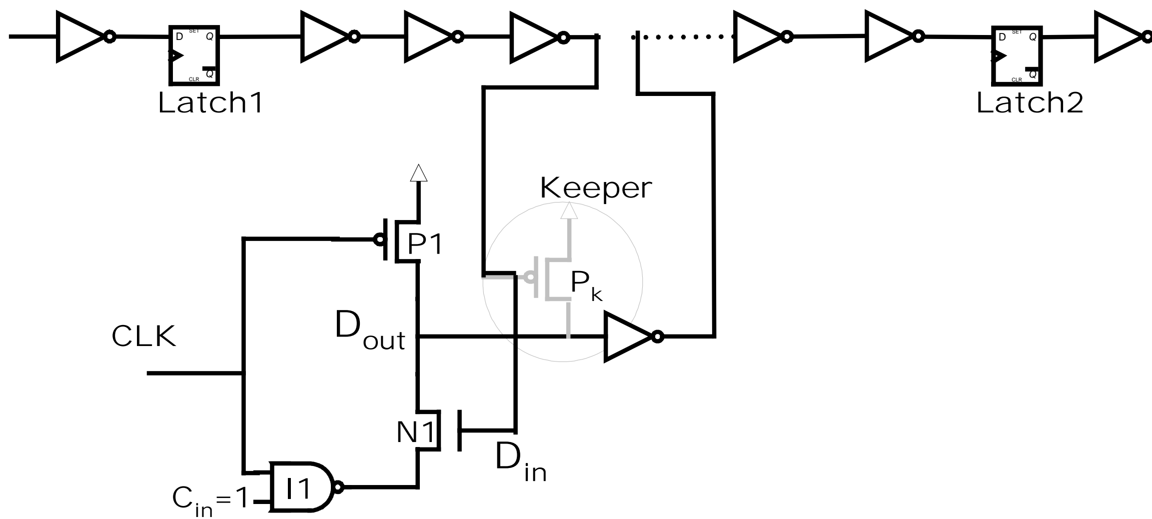


Figure 11: Dynamic inverter

At clock low phase, the output node of the dynamic inverter D_{out} is pre-charged to the logic state high by P1. The evaluation of this dynamic inverter, however, is activated only when the logic value of C_{in} is "1" at clock high phase. Under this condition, the output of the NAND gate I1 supplies a virtual ground to the NMOS device N1.

If D_{in} is low during the evaluation phase, both P1 and N1 are off, and D_{out} remains pre-charged to logic high value. However, D_{out} is, in this condition, very vulnerable to noise, because it can easily lose its logic high value from charge leakage. A weak pull-up device P_k, also known as a keeper device, drawn in gray in Figure 11, provides a feedback mechanism from D_{out} to D_{in}, efficiently preventing charge leakage. Whenever the voltage on D_{out} drops, P_k is turned on, and restores the charge on D_{out}.

Core Latch Design

Latches not only constitute an important implementation issue in the microprocessor circuit design strategy, but also have direct impact on the SER contribution from core logic, since all the single-events originated from combinational circuits have to be latched into a memory element to be observed. Because latch design is closely related to clocking strategy and has great influence on performance, high speed has been the primary goal of the latch design in each generation of the Alpha microprocessor. The high clock frequencies and small number of gate delays available per cycle on Alpha implementations have made low-latency latch design essential.

The Alpha 21164 uses the dynamic CMOS transmission gate latch family shown in Figure 12. This is also a level-sensitive latch that is transparent when CLK is high. When CLK goes low, the data are capacitively stored on node OUT and blocked by the pass gate. Because the charge retained on the node capacitance leaks away with time, this latch also requires a minimum operating frequency. The capability of embedding logic functions directly in the latch helps reduce the number of gate delays per cycle. For this transmission gate latch, the first inverter can simply be converted into a logic gate to incorporate logic. This latch requires true and complementary clock signals, one of which is generated locally. Because CLK is not in the input stack, the resulting low fan-in and clock load is an important performance advantage for the transmission gate latch. In addition, the delay and edge rate of the local clock signal are under good control because the inverter that generates the local clock for each latch type is custom designed. However, this inverter delays the clocking of the latch by one gate delay after the global clock transitions. Therefore, when the preceding latch opens with the global clock transition and when the succeeding latch remains closed due to that one gate delay, the input of the preceding latch may race through. In order to minimize the possibility of data race-through, the utilization of these latches comes with one design constraint: at least one minimum logic delay element is required between all latches [14].

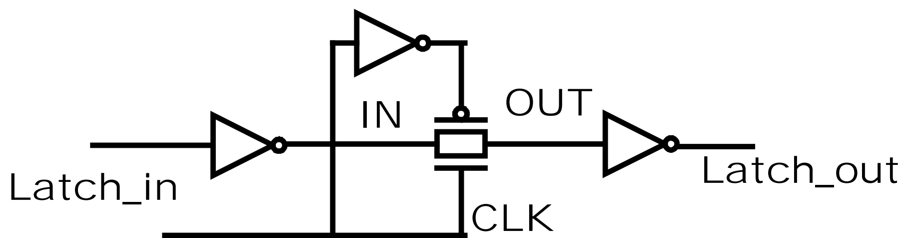


Figure 12: 21164 Latch

Variations of the edge-triggered latch (or flip-flop) are used as the primary storage element of the third generation Alpha microprocessor; one version is shown in Figure 13. This CMOS differential sense amplifier latch consists of two stages: a sense amplifier and a cascode header. It is static for pre-charged inputs, or can be made static by the addition of weak pull-down devices within the latch. When the nodes D_H and D_L are pre-charged to a high state during the low clock phase, the NMOS push-pull pairs in the cascode header are off. Therefore, the sense amplifier and cascode header are de-coupled at this phase, and the state of the latch is maintained by the cross-coupled inverters. The evaluation phase starts on the rising edge of the

clock cycle. As CLK rises, the input data are sampled at IN_H and IN_L. As a result, the NFET stack gated by the higher input voltage discharges. Based on the input values IN_H and IN_L, either D_H or D_L will transition low. Therefore, one of the NMOS push-pull pairs is activated and new content is written into the cascode header.

Figure 13: 21264 Latch with no keeper

Once the cross-coupled NFETs and PFETs lock the latch into the evaluated state, further changes on the inputs will not affect the latch content. Therefore, the hold time of this type of

latch depends on the speed at which the amplifier responds to a voltage difference at its inputs, and can be arbitrarily reduced by increasing the driving ability of the NMOS device N6. However, reversion of the input values at this moment will leave D_H and D_L no path to ground. There are two ways to solve this problem. The first one is to use the pre-charged inputs; since any change on the inputs will be preceded by the pre-charged high values, no reversion of the inputs will occur. The second solution is to add two weak pull-down devices, i.e., keepers, to the latch at D_H and D_L, as shown in Figure 14. Because the keepers provide paths from D_H and D_L to ground, at no time are D_H and D_L floating.

The amplifier-based latch exhibits excellent properties of low latency and arbitrarily small hold time. However, it suffers from relatively high power dissipation due to its complementary dual-rail operation. Regardless of the previous state of the flip-flop, one or the other of the complementary nodes of its first two stages will always switch twice per cycle. Moreover, the amplifier-based latch needs both the data and its complements as inputs. Consequently, it is difficult to incorporate logic within the flip-flop to reduce its latching overhead [22].

Figure 14: 21264 Latch with keepers

Cache Design

A large portion of the silicon area of the microprocessor is occupied by regular memory arrays. More than half of the transistors in today's high-performance microprocessors are devoted to cache memories, and this ratio is expected to increase in the future [23]. Therefore, the SER sensitivity of the cache directly impacts the overall chip-level SER.

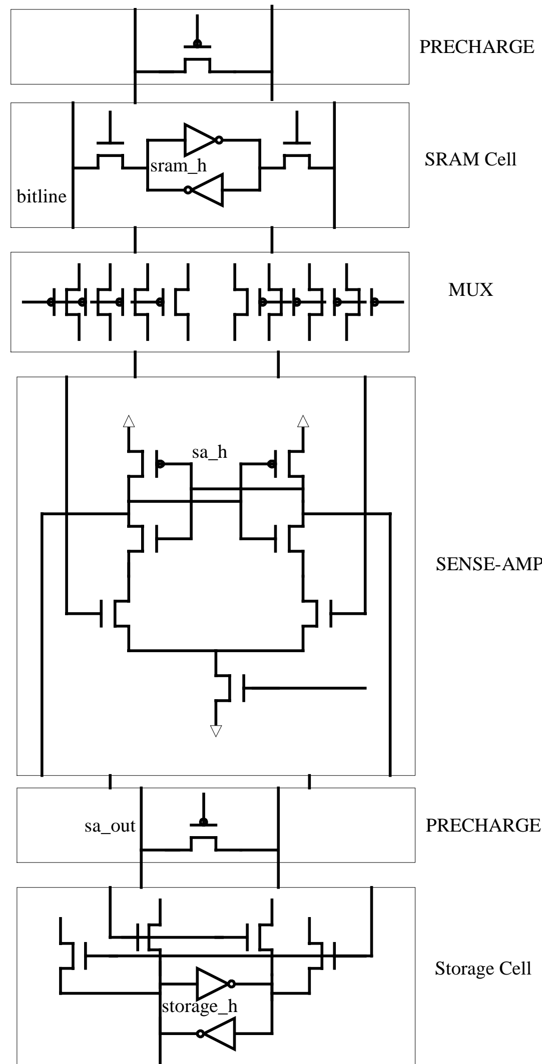


Figure 15: Circuit schematic of the SRAM sensing data-path

The Alpha 21164 includes an 8-kB instruction cache and an 8-kB data cache. A second-level 96-kB cache was present to reduce the overhead of accessing the memory. In the next generation, the Alpha 21264, the size of the instruction and data caches was increased from 8 to 64 kB, thus eliminating the need for the second level cache [14].

Figure 15 illustrates the cache read path extracted from the real design. It includes all the key components in the cache design, SRAM cell, pre-charge and charge balance circuitry, sense amplifier and storage cell. The function and operation of these components will be discussed briefly.

The cache memory is organized in units of words where each word is 64 bits wide. One word at a time is selected for reading or writing. The SRAM cell in Figure 15 represents a bit of the selected word. The selection is indicated by the assertion of word line signal WL, which turns on the pass transistors M5 and M6, as shown in Figure 16. After M5 and M6 are on, the information stored in the SRAM cell will transfer to two bit lines BL and $\overline{BL}$. Since the bit lines are shared by many cells in one column, they are pre-charged before the assertion of the word line signal. The voltage swing on the bit lines is substantially lower than the supply voltage; therefore, the function of the sense amplifier is to restore the full rail-to-rail amplitude. After amplification, the information goes to a storage cell for further use. The control signals on this cache read path, such as enabling the word line and the sense amplifier, are generated internally by the peripheral self-timing control circuitry. Therefore, no external timing signals are needed.

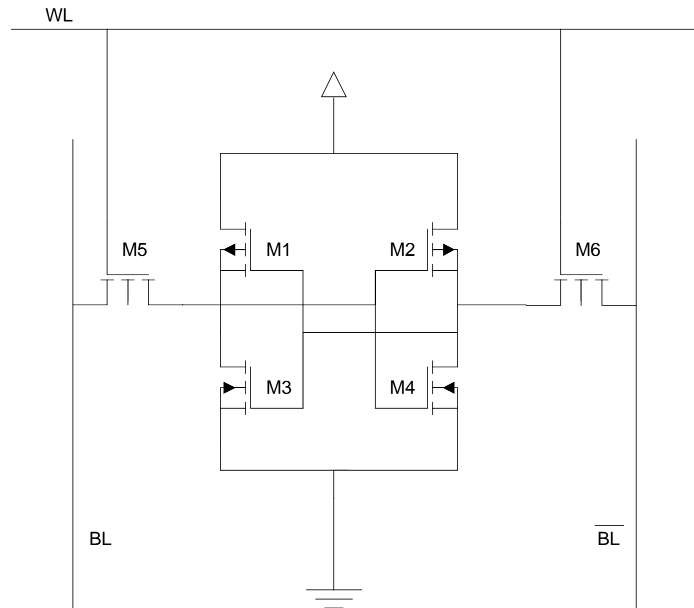


Figure 16: The circuit schematic of an SRAM cell

CHAPTER IV

METHODOLOGY

Computational Method

The goal of the computational modeling is to predict the soft error rates accurately at the chip level by simulating the interaction of a radiation environment with the microelectronic integrated circuits. Since single events are random phenomena, there is a certain probability prob(Q_I) associated with the collection of the charge Q_I by the circuit node per single particle hit. For memory cells under static SEU testing, the cell component of the soft error rate is given by [24][25]

$$SER_{SRAM} = \Phi_{\alpha} A_{SRAM} \int_{Q_{crit}}^{\infty} prob(Q_I)\, dQ_I \qquad (4\text{-}1)$$

where Φ_{α} is the particle flux, A_{SRAM} is the sensitive area of the SRAM cell, and Q_{crit} is the critical charge of the SRAM cell. The lower boundary for the integration is Q_{crit} since errors can only be observed when Q_I is larger than Q_{crit}.

The critical charge Q_{crit} of an SRAM cell is related to the supply voltage V_{DD}, and the following relationship has been proposed by one group [26]

$$Q_{crit} = C_{node} V_{DD} + I_{DP} T_F \qquad (4\text{-}2)$$

where C_{node} is the total capacitance of the node, I_{DP} is the maximum drain conduction current of the PMOS transistor, T_F is the flipping time and defines the irreversibility point after which the feedback mechanism of the SRAM will take over to continue the cell flipping process.

The SER of a SRAM cell can be obtained as a function of the critical charge by measuring the SER at various supply voltages. In an alpha-particle environment, the SER of SRAM cells is found empirically to be exponentially dependent on the supply voltage [18][27]. In fact, the second term $I_{DP}T_F$ in Equation (4-2) is found to be negligible, and the critical charge is, in fact, linearly proportional to the supply voltage [26]. Therefore, we have

$$SER_{\mathrm{SRAM}}(Q_{crit}) = Be^{-CQ_{crit}} \qquad (4\text{-}3)$$

where B and C are deduced from fitting to experimental SER vs. V_{DD} data.

Equating the right side of Equations (4-1) and (4-3) and carrying out the derivative on both sides, we have

$$BCe^{-CQ_{crit}} = \Phi_\alpha A_{SRAM}\, prob(Q_{crit}) \tag{4-4}$$

Rearranging the terms, we obtain

$$prob(Q_{crit}) = \frac{BCe^{-CQ_{crit}}}{\Phi_\alpha A_{SRAM}} \tag{4-5}$$

At the SRAM cell circuit node the probability to collect charge Q_{coll} per alpha particle, therefore, can be expressed as

$$prob(Q_{coll}) = \frac{BCe^{-CQ_{coll}}}{\Phi_\alpha A_{SRAM}} \tag{4-6}$$

The probability of collecting a certain amount of charge is an exponential function of the collected charge. Therefore, the chance to collect a smaller amount of charge is much larger. In Ref. [27][28] it is shown that this exponential relation will no longer hold at very low Q_{crit}/V_{DD} values. Instead, the SER saturates in this range. One possible explanation is as follows. Illustrated in Figure 17, the probability density function of the collected charge of 5 MeV alpha particles peaks at 6 fC [28]. The SER as a function of Q_{crit} is proportional to the integral of this probability density function, i.e., the cumulative density function. The values of collected charge with the highest probability due to low energy alpha particles are rather low compared to the Q_{crit} of the circuits studied in this work. Therefore, no deviation from the exponential dependence was observed down to the smallest V_{DD} values investigated in our study. The saturation effect observed in Ref. [27] [28] occurs when Q_{crit} of the circuit is less than the Q_{peak} indicated in Figure 17. Further, the exponential dependence of the SER as a function of supply voltage is an empirical observation and only valid for alpha particles; it is not true in the case of neutrons. Because of the higher charge deposition rates associated with high-energy neutron events, the neutron induced SER is less voltage dependent [28].

Since the charge collection process is fairly independent of the cell type, the charge collection probability function derived from 6T SRAM SER data should be applicable to all other cells/data-paths for a given process [18]. It has to be noted that the charge collection varies slightly with the geometrical shape of the collection area [24]. The 6T SRAM cell is found to have the highest soft error susceptibility among all storage cells on the chip; therefore, by using the charge collection probability density function derived from experimental data of 6T SRAM and applying it to all circuit nodes on the same chip, one overestimates the charge collection probability for those circuit nodes other than those in the 6T SRAM cells.

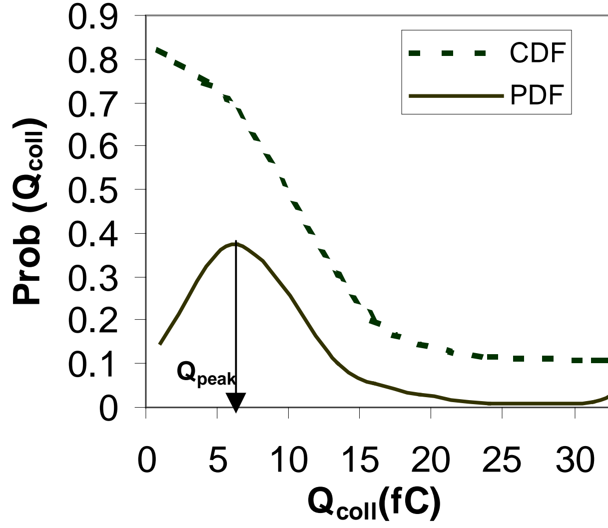


Figure 17: Probability density function and cumulative density function of collected charge for 5 MeV alpha particles[28]

Defining the observed SER as the average over all upsets for all collected charges Q_i, at all injection times t_{inj}, and for all nodes n, the SER of any circuit can be expressed as

$$SER = \phi_\alpha \sum_{n}^{nodes} A_n \sum_{i}^{Q} \sum_{j=t_{inj}}^{T_{cycle}} upset_{n,i,j}\, prob(n, Q_i, t_{inj}) \tag{4-7}$$

where ϕ_α denotes the α-particle flux, A_n the node area, T_{cycle} the cycle time, $prob(n,Q_i,t_{inj})$ the probability that charge Q_i is collected when the single hit occurs at node *n* at the time t_{inj} per alpha particle, and $upset_{n,i,j}$ equals 1 if and only if the circuit is upset by a SE hit at node *n* with collected charge Q_i at time t_{inj} [29].

Since the node at which the single hit occurs, when the single hit occurs, and the amount of charge collected by this single hit are independent of each other, the probability is the product of these three terms

$$prob(n, Q_i, t_{inj}) = prob(n)\, prob(Q_i)\, prob(t_{inj}) \tag{4-8}$$

It is assumed the alpha particle flux is uniformly distributed in space and time, therefore,

$$prob(n) = \frac{A_n}{A_{total}} \tag{4-9}$$

$$prob(t_{inj}) = \frac{1}{T_{cycle}} \Delta t \qquad (4\text{-}10)$$

The error rate is then given by

$$SER = \Phi_{\alpha} \sum_{n}^{nodes} A_n [\sum_{i}^{Q} prob(Q_{i,n})(\sum_{j=t_{inj}}^{T_{cycle}} \frac{1}{T_{cycle}} upset_{n,i,j} \Delta t) \Delta Q] \qquad (4\text{-}11)$$

Defining

$$\Delta \tau_{n,i} = \sum_{j=t_{inj}}^{T_{cycle}} upset_{j,i,n} \Delta t \qquad (4\text{-}12)$$

Equation (4-11) becomes

$$SER = \frac{\Phi_{\alpha}}{T_{cycle}} \sum_{n}^{nodes} A_n \sum_{i}^{Q} \Delta \tau_{n,i} \; prob(Q_{i,n}) \Delta Q \qquad (4\text{-}15)$$

where $\Delta\tau$ equals the so-called “window of vulnerability” (WOV) [12] [29]. It denotes the time during which a circuit is sensitive to SEU.

The overall chip-level soft error rate for a microprocessor is the sum of the contribution from the cache and the core logic, and can be vastly simplified and expressed as

$$SER_{chip} = \delta^{cache} SER_{cache} + \delta^{core} SER_{core} \qquad (4\text{-}16)$$

where δ^{cache} and δ^{core} are the derating factors of the cache and core contributions. When the cache is protected, either by hardware means such as shielding from the alpha particle source, or through software means, for example, by activating the ECC protection, the derating factor δ^{cache} becomes zero. Therefore, the SER measured under this condition is purely the contribution from the core logic. The chip-level SER depends on the program that is run on the microprocessor. The derating factors denote the fraction of soft errors that will actually be noticed by the system. When a test pattern/real program is run, the instruction cache and data cache are partially filled with instruction and data, respectively. The upsets occur in the unused portion of the cache will not be observed, so δ^{cache} is not one in the case when the cache is not protected. As for the core logic, if an upset of a logic state in the floating point unit occurs, it might not lead to a noticeable error when the application compares two integer numbers at the time of the upset. Even if the upset occurs at circuit nodes that are being used by the program, the erroneous information may still be blocked by the succeeding logic during the following clock cycle [30]. The derating factors depend on the architecture of the chip as well as on the

application, and can be determined from VHDL-level simulations, as done in [30]. However, in this work the derating factor of the core logic is assumed to be one for simplicity. Therefore, the SER computed in this work accounts for errors that might be undetected by the user, thus is overestimated.

Simulation Setup

The interaction of the single particles with the IC can be modeled at many levels. In this work, the random bombardment of the circuit by an alpha particle is modeled as a transient current pulse at the circuit level. Plotted in Figure 18, in all spice-level simulations particle strikes were modeled with a double exponential, time-dependent transient current waveform to account for funneling and diffusion charge collection. It can be expressed as [10]

$$I(t) = I_0(1-e^{-\frac{t}{t_{rise}}}); \qquad (t < t_{peak})$$
$$I(t) = I_0(1-e^{-\frac{t_{peak}}{t_{rise}}})e^{-\frac{t-t_{peak}}{t_{fall}}}; \qquad (t > t_{peak}) \tag{4-17}$$

where t_{rise} is the rise time of the pulse, which is on the order of tens of ps, t_{fall} is the fall time, and is on the order of a few hundreds ps; t_{peak} is the time when the current pulse reach its peak value I_0. These parameters are illustrated in Figure 18.

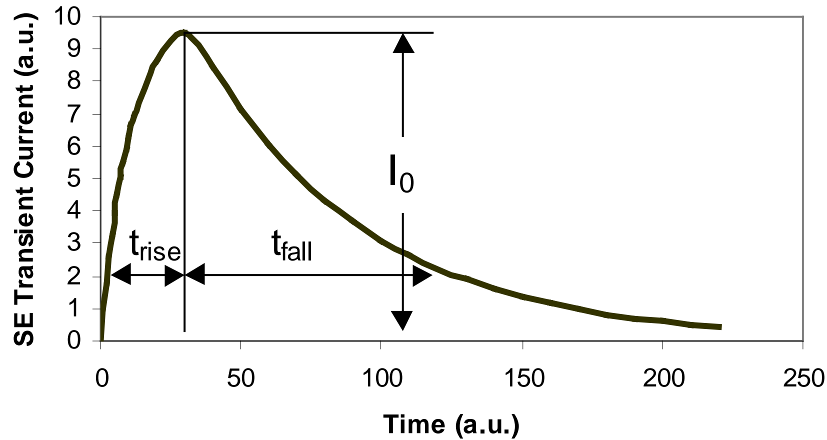


Figure 18: SE transient current temporal profile for circuit-level simulations

The total deposited charge at the circuit node by this current pulse is

$$Q_I = I_0[t_{peak} + t_{fall} - t_{rise} - (t_{fall} - t_{rise})e^{-\frac{t_{peak}}{t_{rise}}}] \quad (4\text{-}18)$$

Typical parameters are a rise time on the order of several ps and a fall time on the order of tens to hundreds of ps [10]. t_{rise} and t_{fall} used in the simulations are process dependent parameters and are obtained from foundry data. Since these are Compaq/Digital proprietary information, the actual number is not given here [31]. In both technologies studied in this work t_{fall} is much larger than t_{rise}, therefore, the charge delivered by the current pulse Q_I can be approximated by $I_0 \times t_{fall}$.

DEC SPICE, the Compaq proprietary circuit simulator, was used for all simulations. The model equations and parameters were developed and calibrated to the experiments at Compaq/Digital for the corresponding Digital processes (CMOS6 and/or CMOS728). Since these models are Compaq/Digital proprietary information, they are not discussed in this dissertation.

Due to its large size, evaluating the SER vulnerability of each circuit node in the cache memory using SPICE simulation is not feasible. However, because of the regularity of the cache memory, modeling of SER contributed from the cache can be simplified. The read path illustrated in Figure 15 is used to study the SE sensitivity of the cache memory. Since it includes all the key components in the cache design, the relative SE vulnerability and its frequency dependence can be derived from the simulation study. In contrast, modeling of the core logic part of the microprocessor is not an easy task due to the wide circuit varieties. The first step is to identify the most sensitive circuit type; it is also important have a count of the number of this type of circuit present in the chip in order to have an estimate of the SER of the core logic. This is done by using the simple data path shown in Figure 7 to find the most sensitive circuit elements in the core logic part of the microprocessor. To represent the worst case scenario, the inverter chain in Figure 7 consists of inverters with minimum device size. For each generation of the Alpha microprocessor design, there is a standard latch library from which variation of the latches and several sizes of the same latch design are available at the designers' choice. The latch in Figure 7 was replaced by the latch shown in Figure 12, Figure 13 and Figure 14 to represent the 21164 and the 21264 Alpha microprocessor core logic circuitry, respectively. All latches from the standard library are evaluated for their SE vulnerability.

The reason that the inverter with minimum device size represents the most sensitive combinational circuit element lies in equation (4-15). There are three important factors that determine the contribution of a certain circuit node to the SER. The first one is the diffusion area of

the node. The second one is the window of vulnerability, which depends on the circuit configuration. Since we assume a uniform distribution of the alpha flux in both space and time, the SER contribution is linearly dependent on the diffusion area and the window of vulnerability associated with the node. The third and dominant factor is the critical charge, because the probability of the collected charge depends exponentially on the amount of the collected charge in the case of alpha particle hits. To first order, the critical charge is linearly proportional to the node capacitance. The node capacitance is the sum of gate capacitance (of other inputs connected to this node), diffusion capacitance (of the drain regions connected to this node) and routing capacitance (of connections between this node and other inputs) [32]. Not only is the fan-out capacitance equal to one gate capacitance of the minimum inverter, but the routing capacitance is also reduced to the minimum.

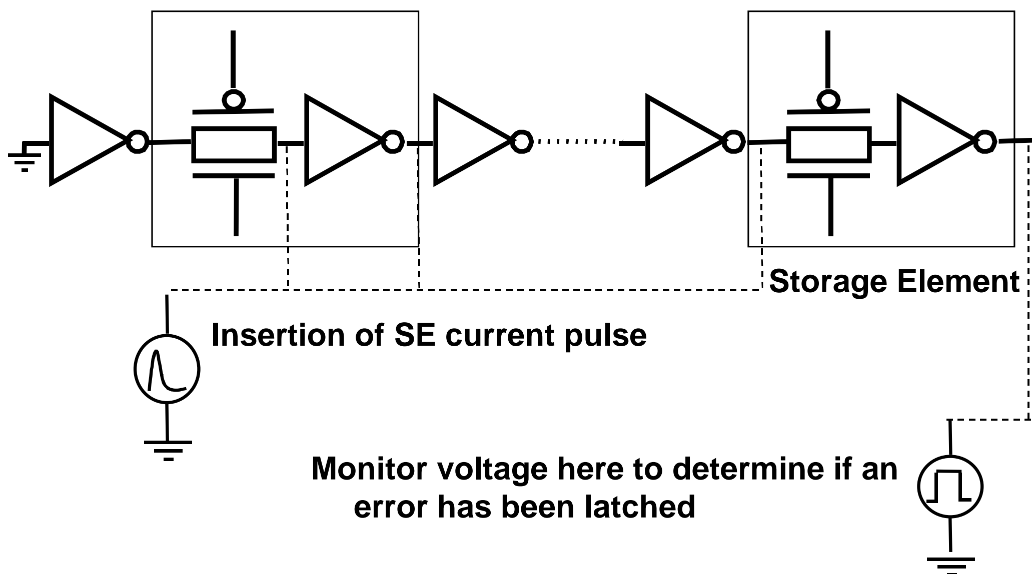


Figure 19: Simulation setup of evaluating the SE vulnerability of the circuit in SPICE

Following the computational method described above, the simulations are set up to calculate the soft error rates according to the equation (4-13). The simulation of a single event hit at the circuit level using SPICE is illustrated in Figure 19. For each data-path, the SE susceptibility of each circuit node is evaluated. As a benchmark, the clock signal controlling the data-path is divided into 1000 intervals. For simulations at low frequencies, a finer temporal mesh is used. The transient SE current pulse is inserted at each interval, and the error is detected at the storage element at the end of the data-path. The critical charge of the node at each time interval is found

iteratively by changing the charge Q_i contained in the transient SE current pulse. The flow chart of the simulation method to find the critical charge is illustrated in Figure 20.

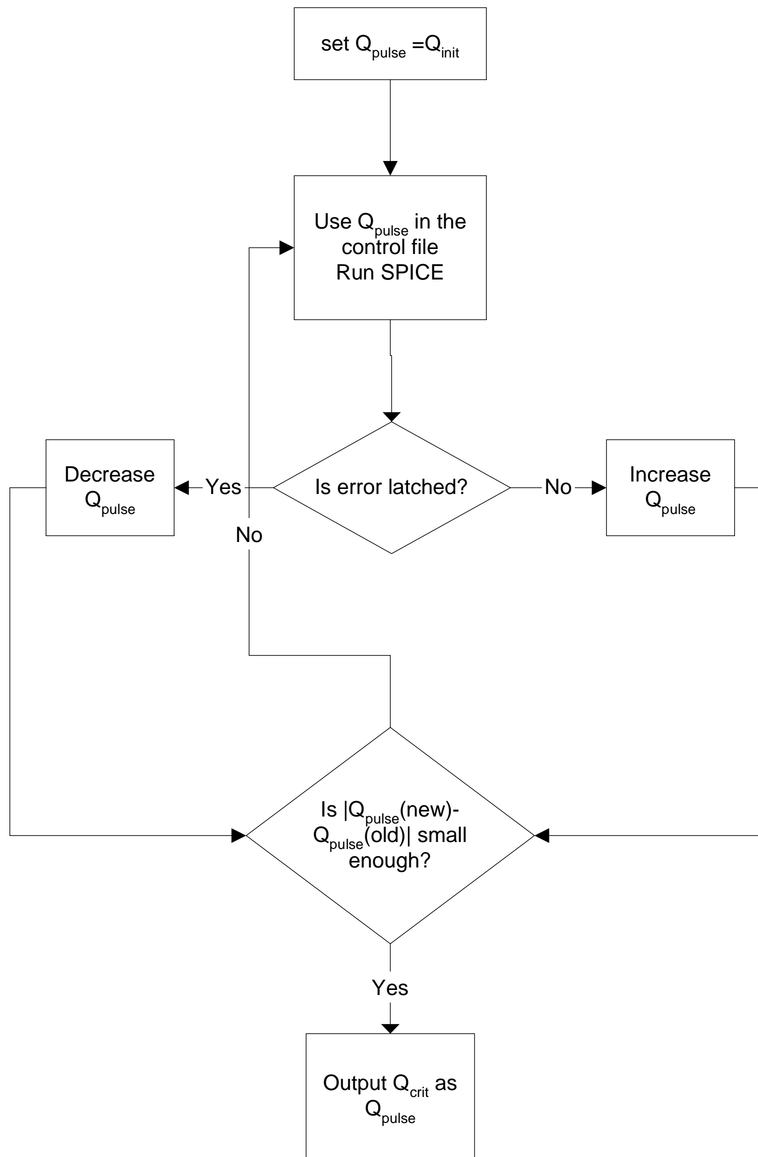


Figure 20: Flow chart of the simulation method to find the critical charge

Experimental Details

Alpha particles emitted from the trace radioactive material used for device fabrication has been a primary terrestrial radiation source and received a lot of attention over the years. Americium 241 is a preferred alpha particle source to simulate alpha emission from lead-based

solder compounds, because the alpha particles emission from Americium has an energy range similar to those alphas resulting from the decay of Uranium and Thorium, the two major radioactive contamination elements in the device material. Another reason to use accelerate alpha particle testing to assess the SER susceptibility of the microprocessor is availability of the alpha particle source at the lab [33]. Therefore, Americium 241 foils were used for the experimental study of this work. Please note that all the results presented in this work are under the alpha particle environment.

Americium 241 decays to Np 237 via a 5 MeV alpha particle with a half-life of 433 years [34]. The range of the 5 MeV alpha particle is about 5 cm in air [35]. The material holding the Americium absorbs some of the energy from decays occurring deeper within the material so there is a large spread of energies and ranges of the alpha particles that reached the chip. The dimension of the foil is 1.25 inch by 1.25 inch, the same size as the lid of the chip. Therefore, during the test the lid of a wire-bonded Alpha microprocessor is replaced with a radioactive Am 241 foil.

The alpha-particle induced failure rates were measured using a Schlumberger ITS9000 tester, which is a system for testing high performance microprocessors and high-end logic devices. The expected outputs (vectors) are stored in the memory of the computer that controls the testing. Therefore, a mismatch of the logic value on the output pins of the chip can be readily detected while the program is run on the chip. The background rate was determined for all products when the α-particle source was removed and was in all cases zero. In order to address experimentally the relative susceptibilities of the cache and core logic circuits to SEU, the die were covered partially with either epoxy or paper to block alpha particles from reaching the covered part (see Figure 21). The difference in SER with and without epoxy is a measure of the susceptibility of those areas to α-particle induced failures.

The SER measurement acceleration can be obtained using an Alpha particle source of high flux since SER is linearly proportional to the particle flux. However, in this work the acceleration factor provided by the Am 241 foil was not sufficient to study the core logic SER with a cost-efficient use of the tester. Further SER test acceleration was achieved by reducing the supply voltage during the SER measurement. However, reducing the supply voltage is equivalent to reducing the critical charge of the chip; therefore, the acceleration factor is exponentially dependent on the V_{DD} reduction since in the V_{DD} range in the work, the SER increases

exponentially with decreasing V_{DD}. For both generations, a low-frequency (20MHz) V_{DD} scan was first performed. However, reducing V_{DD} also limits the maximum operational frequency of the microprocessor. Therefore, a tradeoff between the V_{DD} and operating frequency range has to be made. The supply voltage during the frequency scan measurement is 1.5 V for the 21164 Alpha microprocessor, and 1.1 V for the 21264 Alpha microprocessor. The SER at the nominal operating supply voltage can be back calculated, as will be shown in the result section.

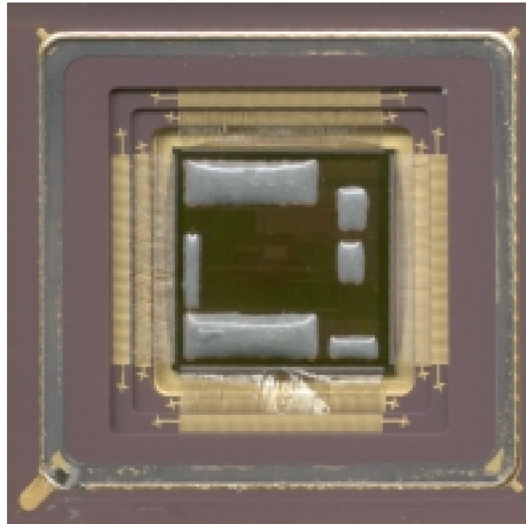


Figure 21: 21164 Alpha microprocessor with several memory arrays covered with epoxy (six gray rectangles in the middle of the chip) to mask the alpha particle

The cache and core logic failure rates were measured using the inherent testing capabilities of the Alpha microprocessor. During test two different patterns were run on the microprocessors: a) the RETENTION pattern and b) the TOAST pattern. The RETENTION pattern executes rather simple tasks and mainly involves filling the data cache with '1's and then reading the cache content one by one. Therefore, not much of the core logic circuitry is utilized by RETENTION pattern. On the other hand, the TOAST pattern is designed to maximize the node toggle activity of the core logic in an Alpha microprocessor. The name TOAST pattern was inspired from the maximum power consumed by the chip while executing the vector sets contained in the TOAST pattern. At the same time, it is important to note that the TOAST pattern also utilizes the caches since the program itself is stored in the instruction cache and data cache is also in use. Therefore, the overall chip-level SER measured when the TOAST pattern is run defines an upper bound rather than a typical number for the chip-level SER, since the SER measured under this condition includes contributions from both the cache and the core logic.

When studying the core logic SER, all caches and TAG arrays had to be covered with epoxy to prevent alpha particles from inducing failures in the memory cells in order to ensure that the SER reflects the failure trend of the core logic [18]. In the case of the 21264 microprocessor, the soft-error protection circuitry (ECC) in the caches was intentionally disabled when the cache SER was investigated.

In most cases, the experimental measurements are repeated three times. The standard deviation is within 10% of the mean value for the RETENTION pattern testing. In the case of TOAST pattern testing, the worst standard deviation is 35% (the uncertainty of the mean is 17%).

The initialization process such as ramping V_{DD} up and starting the clock is done before a pattern is run. Any failure during the initialization is ignored, because the chip must execute several instructions to configure itself and start properly. The origin of an initialization failure is nearly impossible to identify. Our observed data show an initialization fail is rare and occurs 0.05% of the time. However, any pattern failure occurring after the initialization process is recorded as single failure when the tester notices a pattern mismatch.

The concern of having multiple SER failures during a pattern run was not observed in our data. Extremely long patterns can be prone to this. By design, the test patterns are very short and reduce the possibility of multiple SER failures during a single pattern run. In the simple RETENTION pattern, verification of the pin failures indicates single cache failure and extremely rare two cache failures. In the more complex TOAST pattern, the same is true because once the chip deviates from the expected data, the input data will eventually cause a "Machine Check" and cause a halt in the processor. Machine Check is terminology for the Alpha chip. A certain type of output can be recognized as "Machine Check" from the Alpha chip. The Alpha chip received an invalid instruction, which then outputs a machine check. Usually, this means the chip got bad data, via input or internal sources. Thus, one SER failure per pattern is a conservative estimate.

CHAPTER V

RESULTS AND DISCUSSION

Device Level – Q_{crit} and Charge Collection Efficiency

Device scaling leads to a reduced critical charge; on the other hand, technical advances in the fabrication process results in reduced charge collection efficiency. The question to be answered at the device level is which effect, the reduced Q_{crit} or reduced Q_{coll} dominates.

Figure 22 shows measured accelerated SRAM failure rates per bit as a function of power supply voltage (V_{DD}) in arbitrary unit for the two Alpha technologies under study. The SER of the SRAM cell is approximately an order of magnitude lower for the 21264 than it is for the 21164 over the entire V_{DD} range of measurements. Moreover, the extrapolated SER value at their respective nominal supply voltage follow this trend. Not unexpectedly, both voltage scaling and process technology have a pronounced effect on the SER.

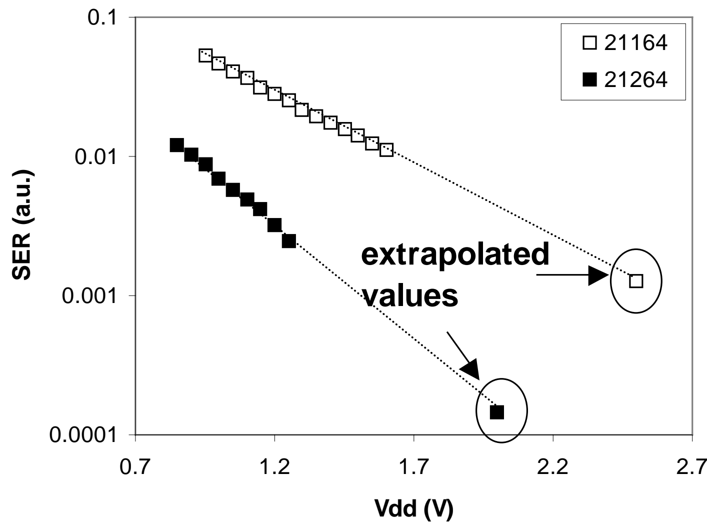


Figure 22: Measured SER of on-chip cache as a function of operating voltage for two different technologies. Circled values indicate the extrapolated values to the respective nominal supply voltages

The simulated critical charges for 6T SRAM upset for both generations of the Alpha microprocessors are shown in Figure 23. Compared to the Alpha 21164 (0.35 μm CMOS), the critical charge for 6T SRAM upset in the Alpha 21264 (0.25 μm CMOS) is 10%~20% lower at the same supply voltage due to process and geometry scaling. The exponential dependence of the

SER on the power supply voltage, also observed in experimental data [13][18][27], suggests a sharp increase of the SER for an SRAM cell in the Alpha 21264 process technology (lower operating voltage). However, the much lower SER of the 21264 SRAM cell in Figure 22 clearly indicates a much reduced charge collection efficiency in the 0.25 μm CMOS bulk technology than in the previous generation. Therefore, process improvements outpace the reduction of supply voltage and capacitance, and the error rate per information bit decreases instead of increasing with technology scaling. This has been observed and commented on in DRAM technologies [36].

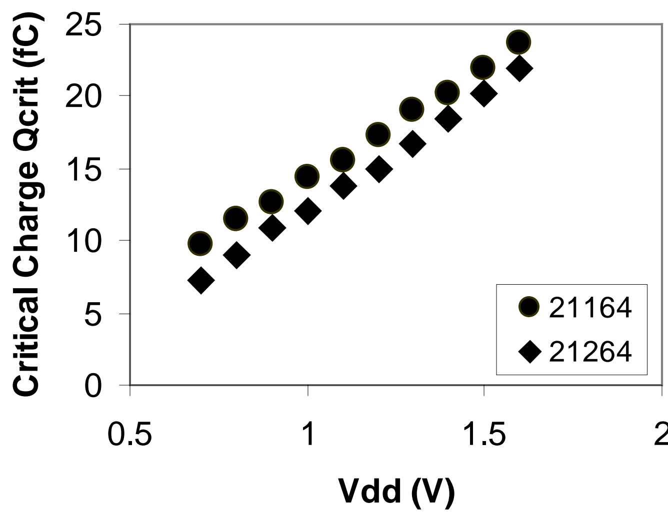


Figure 23: Simulated critical charge of the SRAM cell for two different technologies

<u>Circuit Level – Circuit Design Style</u>

The analysis above indicates the process scaling can improve SEU immunity. Contrary to what people have feared, the error rate per bit decreases, rather than increases. However, when a system migrate from one generation to the next, the size of the on-chip cache also tends to increase. In addition, in order to accommodate the new features in microprocessor architecture such as out-of-order execution or branch prediction, the complexity of the core logic circuitry also increases. The overall SER of a chip can not be extrapolated from the SER of the regular SRAM arrays. In each generation of the Alpha microprocessor, the sophisticated circuit design techniques have been used to increase the chip performance further than that brought by the

process scaling. For example, the latch library of the 21164 and 21264 Alpha microprocessors are quite different. It is advantageous to investigate the impact of different circuit design techniques on the SER.

As mentioned previously, the 21164 and 21264 Alpha cache sense and read circuitry share a lot of similarities, while the latch design, the key component of the core logic, is totally different. The study of the impact of the circuit design techniques on the SER is done through studying the core logic SER of the two generations of the Alpha microprocessor. The data are taken with all the memory arrays on chip shielded from the radioactive source and with the TOAST pattern run. Since it is impossible to compare the SE vulnerability of different circuit design styles without specifying their operating frequency, the frequency dependence of SER is an issue on both the circuit level and system level. In this section, we will include the SER frequency dependence of the core logic with the discussion of the circuit design style impact on the core logic SER. In the next section, we will concentrate on the analysis on the frequency dependence of the cache and overall chip-level SER.

The sensitivity of the core logic to SEU is higher in the Alpha 21164 than in the 21264. In Figure 24, the experimental SER of the core logic in these two generations is shown. As will be shown later, the SER of the core logic is much lower than the contribution from the cache. As a result, in order to collect statistically meaningful data while maintaining the economic use of the tester time, the supply voltage has to be reduced. However, the chips could not operate up to the target frequency due to the reduced supply voltage. The 21164 data was taken with supply voltage of 1.5 V, and the maximum frequency under this supply voltage is 400 MHz. The supply voltage is reduced to 1.1 V for the 21264 data, and the maximum frequency under this supply voltage is only 167 MHz. To compare the SE sensitivity of the core logic at the nominal operating voltages, the data shown in Figure 24 include the acceleration factors for both generations. Although the experimental data show different frequency trends, it is clear that the core logic of the Alpha 21264 is much less sensitive than that of the previous generation. The main reason for this improvement lies in the latch design techniques employed in the later generation. Comparable improvements in SER were observed for the new Alpha design fabricated in the older generation process [20], so improvements in the design methodologies primarily impact the improved SER.

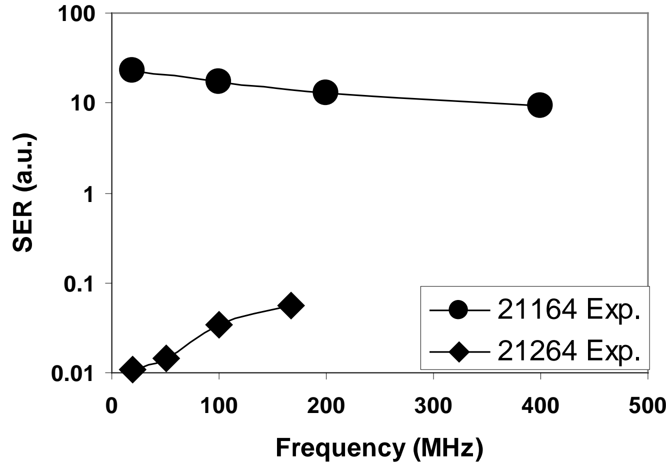


Figure 24: Experimental accelerated SER of core logic in two generations of Alpha microprocessors

We first identify the circuit types that show the highest degree of sensitivity to SE, and then analyze on these circuits. In the data path depicted in Figure 7, both the size of the inverters and latches are set to be the minimum size found in the real implementation of the Alpha microprocessor design. Therefore, the inverters and the latch in Figure 7 represent the most sensitive part of combinational logic circuits and sequential circuits respectively. By comparing the critical charge of the latch nodes and the inverter nodes, one should be able to know whether the SE originated from the combinational logic or the direct hit on the latch dominate the core logic SER.

Table 5 lists the critical charge of the inverters and the latches normalized to the SRAM cell in the same technology. The critical charge of the static inverter is five times higher than that of the latch in the 21164 Alpha design built in 0.35 μm technology, and is three times as high as the Q_{crit} of the latch in the 21264 Alpha design built in 0.25 μm technology. As discussed in the methodology section, the SER of a circuit is the sum of the contributions from each node. Since the number of combinational circuit nodes is on the same order as the latch nodes, and since, in the frequency range considered, the window of vulnerability of the latch nodes is higher than that of combinational circuit nodes, the exponential dependence of the SER on the power supply voltage observed in experiments indicates that the contribution from those nodes with the lowest Q_{crit} will dominate the SER of the whole circuits. Therefore, the SE due to the direct hit on the latch nodes dominate the SER of core logic, while the SE originated from combinational circuit is negligible.

Table 5: Critical charge of comparison (normalized unit)

	21164	21264
Inverter node	3.29	3
Latch node	0.55	1
SRAM	1	1

The fact that the latch nodes dominate the SER of core logic holds true for both generations of the Alpha microprocessors. The transmission gate latches in the Alpha 21164 microprocessor store information on floating nodes. In the case of the Alpha 21264 microprocessor, the sense amplified based latches are static for pre-charged inputs or can be made quasi-static with the addition of keeper (weak pull-up or pull-down) devices. However, the latch nodes still pose higher SE vulnerability than the combinational logic nodes, and dominate the core logic SER.

Core Logic SER: Latch

Experimental Results:

In Figure 24 the frequency dependence is shown for the alpha-particle induced accelerated SER in arbitrary units for both generations of the Alpha microprocessor when the TOAST pattern that maximizes the core logic activity is run. The data reflect the SER trend of the core logic, since all major memory arrays have been protected. From the experimental results, one can clearly see that the measured SER of the 21164 Alpha microprocessor core logic is much higher than that of the 21264 generation. This is due to the large number of floating latch nodes present in the 21164. Furthermore, the measured core logic SER of the 21164 Alpha microprocessor decreases with increasing frequency, while an increase in the SER with increasing frequency is observed for 21264 Alpha microprocessor.

It should be noted that the 21164 core logic SER data seems to be contradictory to Buchner's results [6]. However, Buchner et al. did not investigate data-paths that involve floating nodes, where "floating" denotes dynamic nodes with no feedback devices (keepers) attached. In contrast, the utilization of level sensitive transmission gate type latches in the core logic of 21164 Alpha microprocessors results in a large number of floating nodes when the TOAST pattern was

running. The truly floating internal node of the transmission gate latch shown in Figure 12 is responsible for the relatively high soft error (SE) susceptibility of the core logic of the 21164 Alpha microprocessor and for the observed clock frequency trend of the SER. In the next generation of the Alpha microprocessors this design style is not used anymore, and is replaced by a sense-amplifier based latch (SABL) design, mainly driven by latency requirements. The reduced core logic SER of 21264 Alpha microprocessor and its opposite frequency trend stem from the more static latch design adopted in its core logic design. We will see further below that this is the key to understanding the observed experimental results.

Simulation Results:

Figure 25: depicts the critical charge of the floating latch node and static node between inverters in 21164 Alpha microprocessor core logic data path as a function of time. One can clearly see that the critical charge of the static combinational node dips around the clock edge minus the time it takes the fault to propagate to the second latch, whereas Q_{crit} of the floating node is fairly constant during the clock phase when the node is floating (i.e. transmission gate is off). The WOV is therefore much larger in the case of the floating node.

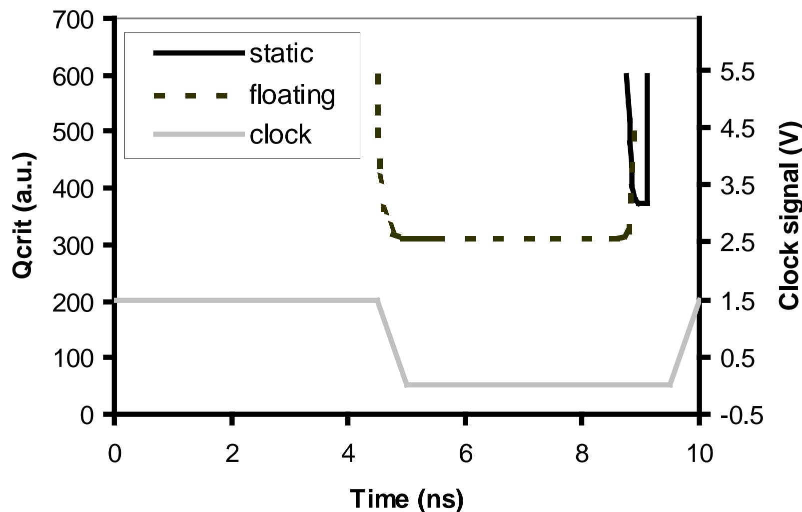


Figure 25: Simulated critical charges are plotted for two different nodes of the 21164 core logic data-path in one clock cycle

The results of our simulations suggest that the SER increases with clock frequency when combinational static nodes (the input/output nodes between inverters in the inverter chain in Figure

7) are upset, in agreement with Buchner's work. According to Figure 25: , static nodes contribute to the SER only at the clock edges. Therefore, the smaller the cycle time, the more frequently the clock fires and the higher the probability that a fault is latched. However, since the critical charge of the combinational logic is considerably larger than that of the dynamic node in a 21164 type latch and because the WOV is smaller for hits in the combinational logic (Figure 25:), the SER contribution of the combinational logic turns out to be negligible.

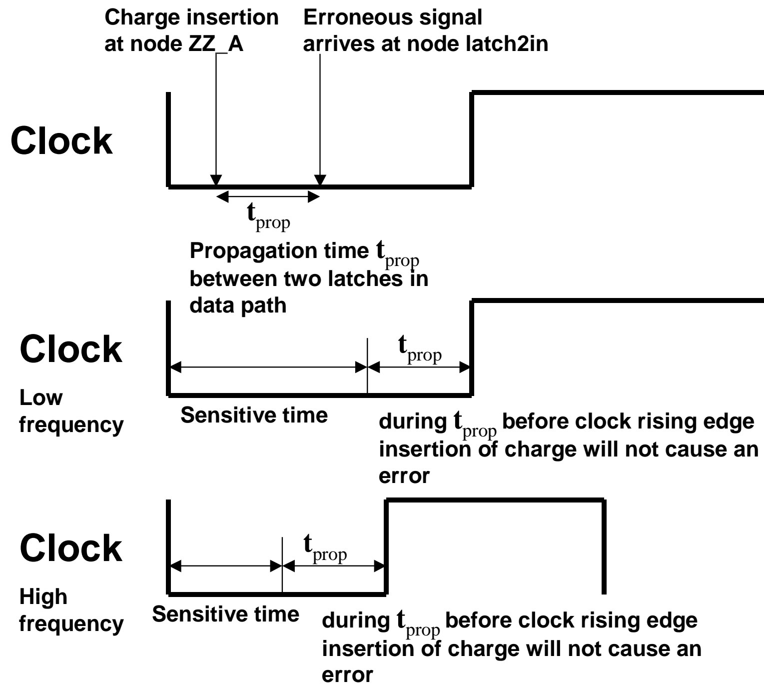


Figure 26: The WOV decreases with increasing clock frequency, which results in a corresponding decrease in SER

The main reason for the SER of the 21164 core logic to decrease with clock frequency lies in the fact that upsets of dynamic nodes with no feedback devices are persistent and, in contrast to static nodes, automatically satisfy the setup- and hold time criteria of latches. There exists a fixed time window (t_{prop}) during which the injection of charge at the dynamic node out (Figure 12) would not be able to arrive at the second latch before the clock fires (see Figure 26). t_{prop} equals the time it takes for the glitch to propagate to the second latch. The relation between the WOV and the SER can be expressed as

$$SER \propto \frac{WOV}{T_{cycle}} = \frac{\frac{1}{2} T_{cycle} - t_{prop}}{T_{cycle}}. \quad (5\text{-}1)$$

The SER therefore decreases as the clock frequency increases. This is true for all floating nodes located in either the latches or in the combinational logic. The percentage of floating nodes in the combinational logic is too small, however, to account for the observed SER trend of the core logic. Therefore, the observed frequency trend of the core logic SER stems from latch nodes as opposed to floating nodes in the combinational logic.

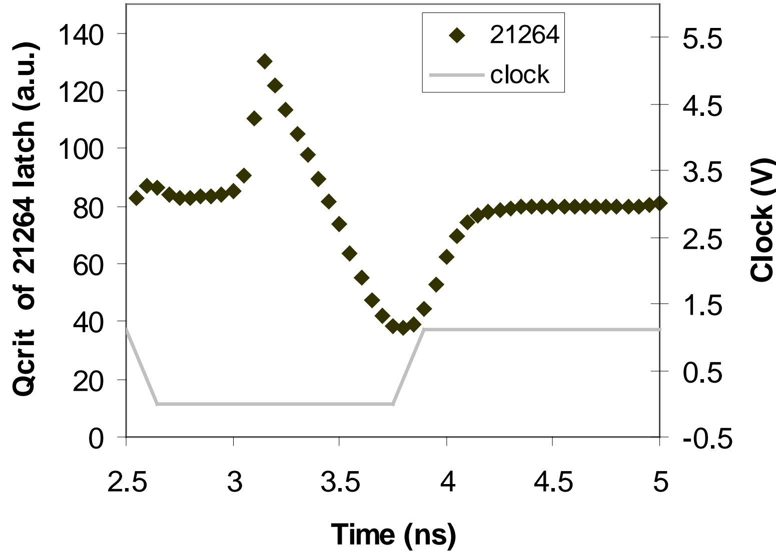


Figure 27: Simulated critical charge of latch designs in 21264 Alpha microprocessor core logic

In the case of the 21264 Alpha microprocessor, a more robust latch design was adopted. First, no truly floating nodes are present in the 21264 die, which results in higher critical charges. Second, the sense amp based design results in a lower WOV. These combined effects lead to a much smaller SER compared to that of the 21164 type latch (see Figure 24). The D_H node (see Figure 13) was identified to be the most sensitive 21264 latch node and dominates the SER frequency trend of the core logic of the 21264 Alpha microprocessor. Figure 27 plots the critical charge of the latch node D_H in the 21264 Alpha microprocessor core logic data path as a function of time. In contrast to the transmission gate latch design, D_H is particularly vulnerable to SE at the clock rising edge, when its critical charge reaches the lowest value. The relation between the WOV and the SER can be expressed as

$$SER \propto \frac{WOV}{T_{cycle}} = \frac{const}{T_{cycle}}. \qquad (5\text{-}2)$$

Therefore, the SER increases linearly with clock frequency.

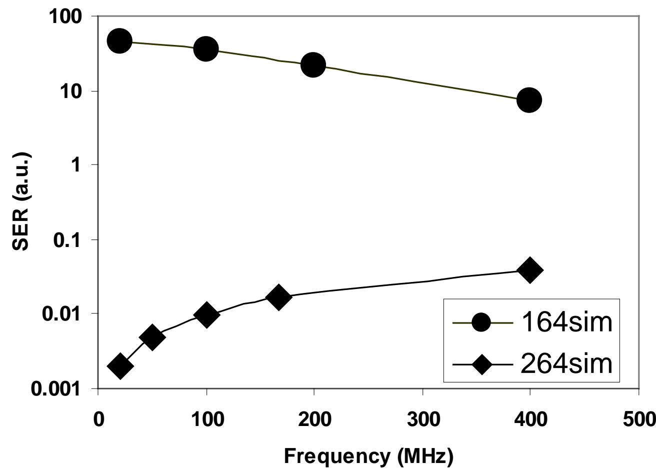


Figure 28: Computed failure rates are shown as a function of clock frequency for two types of latch designs

The computed core logic SER of the two generations of the Alpha core logic at their respective operating voltages as a function of frequency is shown in Figure 28. In the case of the dynamic transmission gate type design, the SER *decreases* with clock speed. On the contrary, the sense-amplifier based latch design exhibited an opposite trend; the SER *increases* with increasing frequency. Since the core logic SER is dominated by the contribution of the latches in both cases (the 21164 and the 21264 Alpha microprocessors), the computed core logic SER frequency trend is in agreement with our experimental observations in Figure 24.

Core Logic SER: Combinational circuit

The previous analysis on 21164 Alpha microprocessor transmission gate latch indicates that the most sensitive circuit nodes are the floating ones, not only because they have low critical charge, but also they have a long window of vulnerability (WOV). The high degree of vulnerability of floating nodes comes from the fact that they have no active path to either V_{DD} or GND. In contrast, static combinational circuit nodes are less vulnerable due to their inherent regenerative circuit path. With the presence of the keeper device, the dynamic logic circuit nodes are no longer floating. Therefore, the charge loss on the dynamic node is restored through the keeper device, and the node voltage recovers eventually to its original value. However, keeper devices are usually sized much smaller than the normal device. As a result, the process of charge restoration is much slower than in the case of static nodes.

To identify the impact of keeper design on the circuit SER vulnerability, we have evaluated the SE vulnerability of the dynamic inverter output node D_{out} without a keeper, with a small keeper and with a large keeper (see Figure 11). The critical charges of dynamic nodes under these three conditions are plotted in Figure 29. The dynamic node with no keeper exhibits the same Q_{crit} characteristics as the floating node in transmission gate latch design; after all, the dynamic node with no keeper is left floating during the evaluation phase when the NMOS device N1 is off (see Figure 11). The SE vulnerability of the dynamic node is drastically reduced when a big keeper is added, and is very similar to that of the static node, a much higher critical charge and a much shorter WOV. The keeper device has two main functions. First, the presence of the keeper restores the charge on the dynamic node shortly after a single event hit, thus reducing the WOV. Second, because of the regenerative circuit path introduced by the keeper, a single event current pulse with higher magnitude is needed to discharge the dynamic node, thus increasing Q_{crit}.

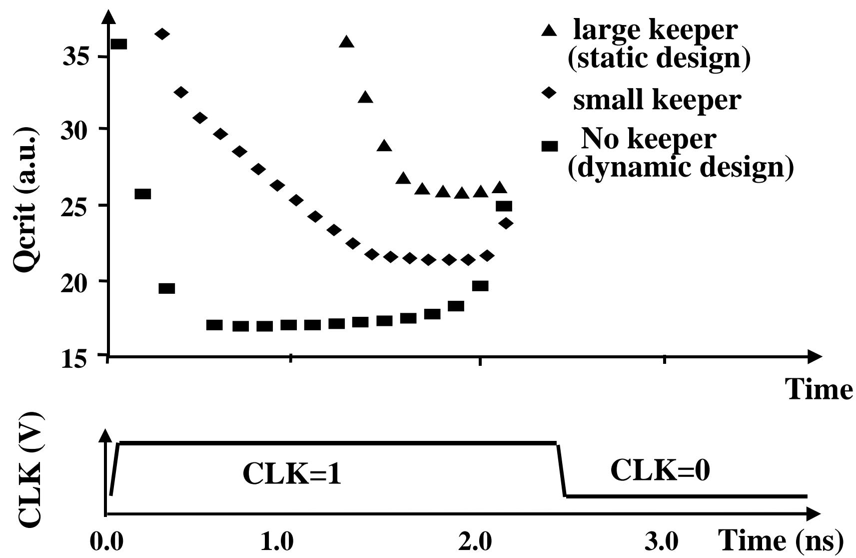


Figure 29: Simulated critical charge of various keeper designs

The efficiency of the regenerative path is increased by increasing the size of the keeper. The plot of the critical charge of the dynamic node with a small keeper clearly demonstrates the impact of the keeper device size on SE vulnerability. The larger keeper not only makes the circuit harder to upset, but the increased efficiency leads to a faster recovery and a narrower WOV.

In Figure 30 the computed SER contributed from the dynamic nodes under the three conditions is shown as a function of clock frequency. SER contributed by the floating nodes is the highest and decreases with clock frequency. The main reason for this trend lies in the fact that upsets of dynamic nodes with no feedback devices are persistent and, in contrast to static nodes, automatically satisfy the setup- and hold time criteria of latches. When the regenerative circuit path is established by the keeper, not only is the SER sensitivity of the dynamic node reduced, but the SER frequency trend on this node also changes. The introduction of the strong regenerative circuit path, either in the case of static nodes, or in the case of the dynamic node with a large keeper, enables the information restoration, and SER contributed by this type of node increases with frequency. The slope of the SER frequency trend is determined by the size of the keeper device.

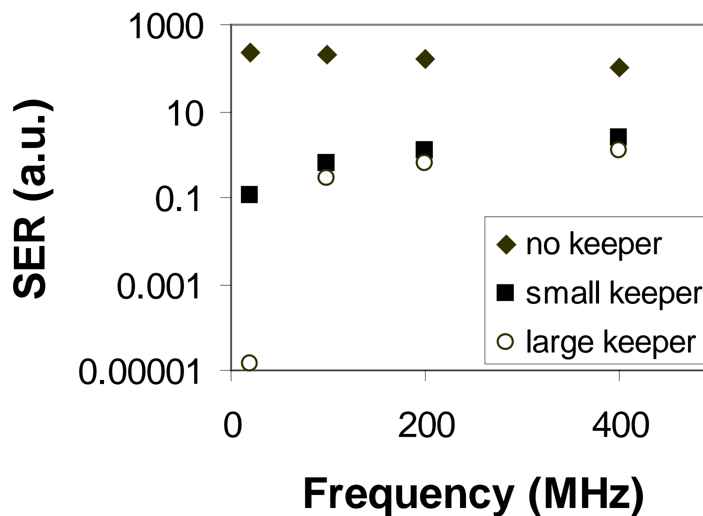


Figure 30: Computed failure rates are shown as a function of clock frequency for various keeper designs

The frequency trends shown in Figure 30 seem to suggest that at high frequencies, the dynamic design might end up having a lower SER value than the static design (when these lines cross). This will not occur. Assuming the transistors of similar sizes are used, the SER from the dynamic design is much higher than the static one at low frequencies. Because the dynamic design not only has a lower Q_{crit}, at low frequencies it also has a larger WOV. The reason for the SER of dynamic design to decrease with increasing frequency lies in the fact that the propagation delay of the data path is independent of frequency and causes a reduced WOV, as described by

equation (5-1). However, the WOV also correlates the time period during which the valid data could successfully transfer from the first latch to the second one on this data path. At very high frequencies when the propagation delay equals half the clock cycle, although WOV is reduced to zero, the data path itself also ceases to function. Therefore, at high frequencies in the future technologies, the WOV of the dynamic circuits might reduce to a value comparable to the static circuits, but the SER of the dynamic design will still be higher than its equivalent static design due to its lower Q_{crit}.

Frequency Dependence

The demand for higher performance has resulted in clock speeds of high performance microprocessors of the order of several GHz. The frequency dependence of SER, therefore, has important reliability implications for future technologies. SER frequency dependence due to different circuit styles in the core logic was discussed in detail at the circuit level previously. In this section we will focus on the frequency dependence of the on-die cache.

Experimental Results:

Figure 31: depicts the frequency dependence of the SER when the cache intensive RETENTION pattern is run on the 21164 Alpha microprocessor. One can clearly see that the SER does not approach zero for low clock frequencies. This background level is due to the SER contributions from the SRAM cells of the caches. The linear increase in frequency stems from upsets generated during the reading of the content of the cache cells, as we will discuss in the next section.

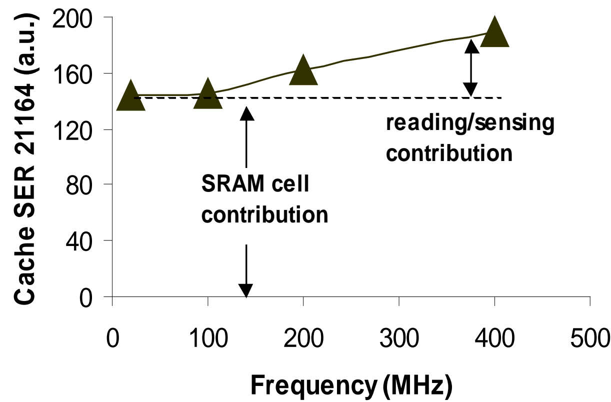


Figure 31: Experimental accelerated SER as a function of clock frequency for the RETENTION pattern for 21164 Alpha microprocessor

Plotted in Figure 32 is the cache SER as a function of frequency when the cache intensive RETENTION pattern is run on the 21264 Alpha microprocessor with everything covered except the data cache. Similarly to the 21164 Alpha cache SER, a background SER when frequency approaches zero is observed from the 21264 data and 21264 cache SER increases with clock speed. However, the cache SER starts saturating at 200 MHz. This phenomenon can be explained by the simulation results.

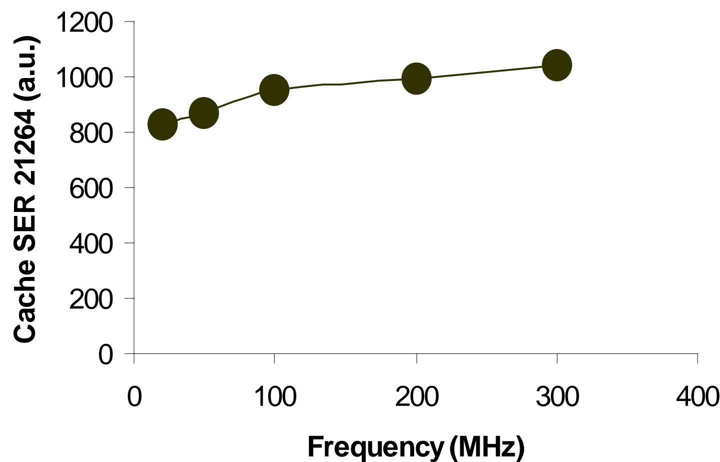


Figure 32: Experimental accelerated SER as a function of clock frequency for the RETENTION pattern for 21264 Alpha microprocessor

Simulation Results:

Table 6 summarizes the simulation results in terms of minimum critical charges (Q_{crit}) and corresponding WOV $\Delta\tau$ for all the circuit nodes of the cache read path of the 21164 Alpha microprocessor shown in Figure 15.

Table 6: Summary of simulation results at 20 MHz for nodes shown in Figure 15

Node	Min Qcrit (a.u.)	$\Delta\tau$ (a.u.)	SER (a.u.)	Node type	f-trend for increasing f
sram_h (wl=0)	1.0	1.0	1.0	Static	Independent
sram_h	0.79	0.524	0.46	Static	Increasing
Bitline	>10	0.036	~ 0	static	Increasing
sa_h	2.44	0.002	~ 0	Dynamic	Increasing
Storage_h	>10.0	N/A	~ 0	static	Independent

Among all the nodes in the SRAM reading/sending circuitry, the internal SRAM cell node sram_h appears to be the most vulnerable one. Therefore, the frequency dependence of sram_h dominates the frequency trend of the cache. Our simulation results shown in Figure 33 explain why the SRAM/Cache type circuit shows an increased SER trend with increasing clock rate and why the cache SER saturates at high frequencies. The Q_{crit} of sram_h is constant when the word line is not asserted, thus the errors occurring in those SRAM cells not being read are independent of clock frequency, marked as the SRAM cell contribution in Figure 31: .

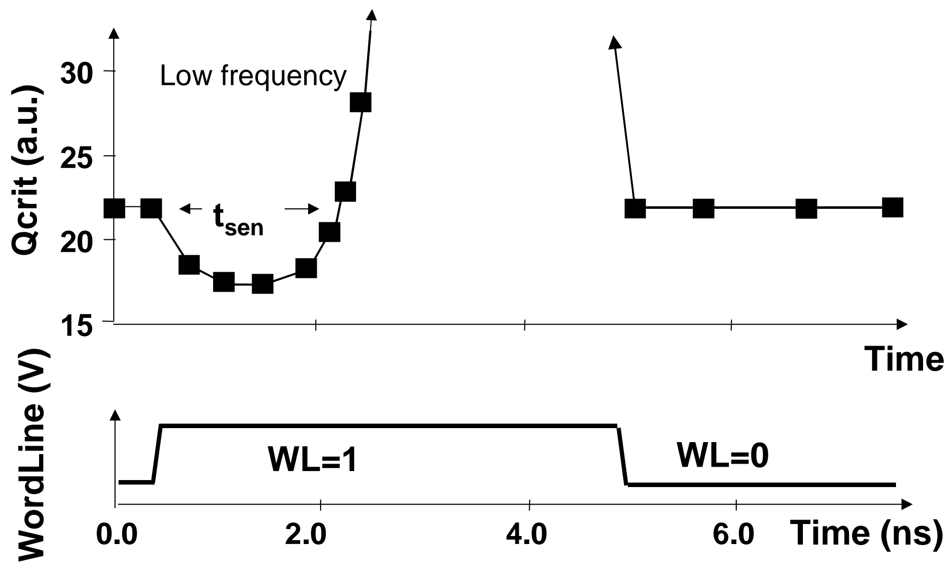


Figure 33: Simulated critical charge of internal node of SRAM cell under reading/sensing condition at low frequency

When the SRAM cell is being read, both bitlines are pre-charged to V_{DD} just before the word line is enabled. Charge sharing results in a lowered Q_{crit} of the SRAM cell during the reading/sensing process. However, once one of the bitlines is discharged low by the SRAM cell, the vulnerability of the SRAM cell decreases due to the large capacitance on the bitlines. Therefore, the WOV is roughly constant during reading at low frequencies, and the SER of the SRAM cell increases linearly with clock frequency.

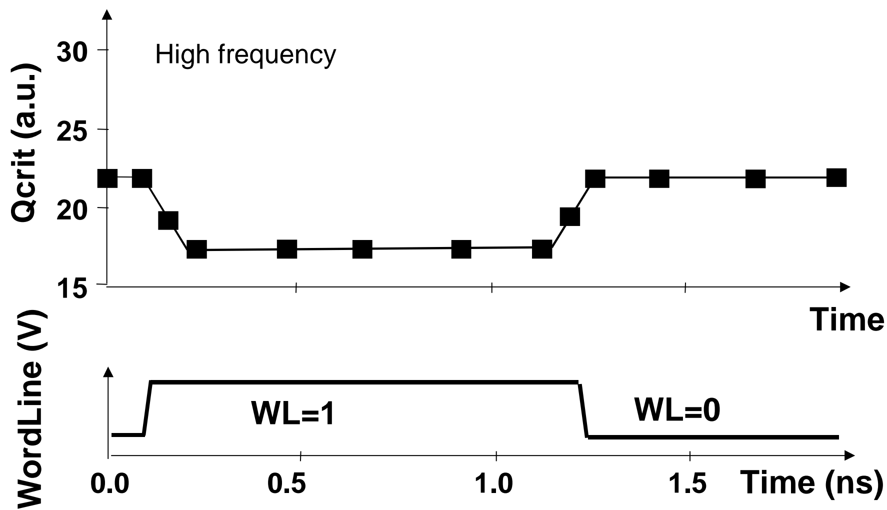


Figure 34: Simulated critical charge of internal node of SRAM cell under reading/sensing condition at high frequency

At high frequencies, the cycle time is reduced. The period during which the SRAM cell is less susceptible to SE due to the large capacitance of the bitlines disappears. Figure 34 plots the

critical charge of internal node of SRAM cell under reading condition at high frequency. Q_{crit} is lower during the whole phase when the word line is high. Since the word line is just a delayed version of the clock signal, the duty cycle (t_{sen} vs. cycle time) remains constant at higher frequencies.

Overall Chip-level SER

The chip-level SER is a sum of the contribution from the cache and core logic of the microprocessor. The experimental SER of the TOAST pattern in two generations with and without the memory array covered is shown in Figure 35 and Figure 36 respectively. In order to obtain a relative comparison of the overall chip-level SER between two generations, the error rates in both graphs are normalized to the same scale. The migration of technology from the 0.35 µm node to the 0.25 µm node, combined with the improvement from circuit design, reduces the overall chip level SER by an order of magnitude.

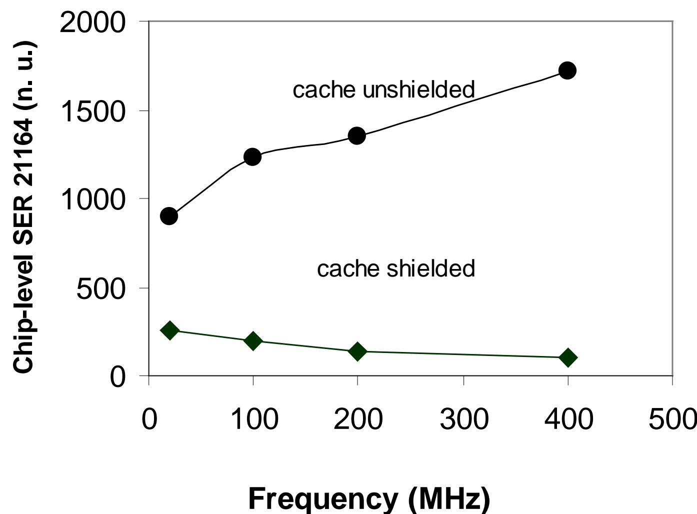


Figure 35: Experimental accelerated SER of the TOAST pattern run on 21164 Alpha microprocessor is shown as a function of clock frequency with and without memory arrays covered

In addition, the contribution from the cache dominates the overall chip level SER in both generations of the Alpha microprocessor. This result demonstrates how ECC can significantly reduce the overall chip-level SER by eliminating the dominating contribution from the cache.

Activating ECC in the Alpha 21264 further reduced the chip-level experimental SER by two orders of magnitude.

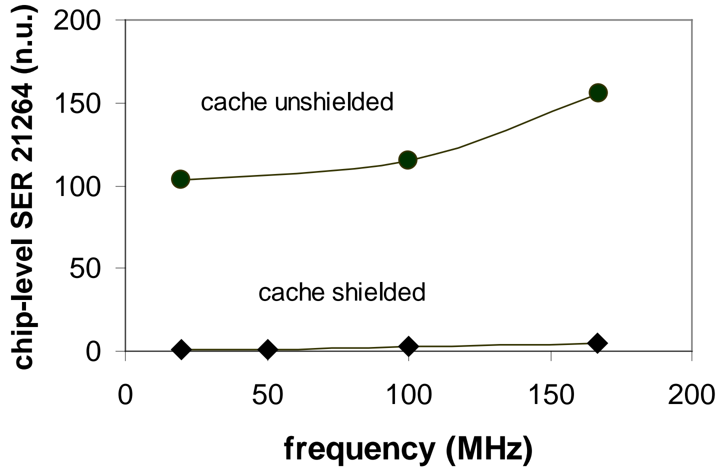


Figure 36: Experimental accelerated SER of the TOAST pattern run on 21264 Alpha microprocessor is shown as a function of clock frequency with and without memory arrays covered

The final FIT rates of the product are influenced by many factors. Environment plays an important role since the FIT rate is affected by both the flux and the spectrum of the particles. Besides the improvements from the process scaling such as the reduced charge collection efficiency, the architectural decision (ECC), the circuit design (latch design) and packaging (lid coat, flip chip) also play an important role. Figure 37 demonstrates the importance of these factors. The FIT rates of three generations of Alpha microprocessors manufactured in various processes due to alpha particle hits are plotted. The nomenclature used in this graph is summarized in Table 7, where the process and the supply voltage corresponding to each chip are also listed [20].

Table 7: Nomenclature used in Figure 37 [20]

Chip-Version	Name used	Ldrawn (μm)	Cache (KBytes)	V_{DD} (V)
21064A	EV45	0.5	32	3.3
21164	EV5	0.5	112	3.3
21164A	EV56	0.35	120	2.5
21264	EV6	0.35	128	2.2
21264A	EV67	0.25	128	2.0
21264B	EV68	0.18	128	1.6

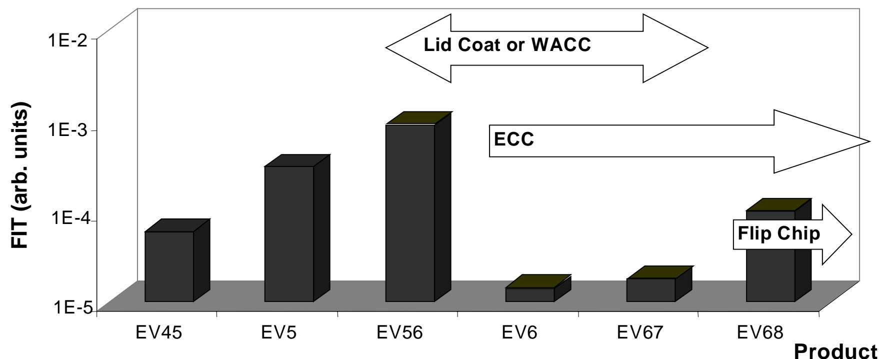


Figure 37: The FIT rates due to α-particle strikes are plotted for several different Alpha microprocessor generations [20]

Historical trend in alpha-particle induced soft error rates of the Alpha microprocessor has not been a monotonic one, because the final FIT rates displayed in Figure 37 exhibited a combined effect of the process influence, system level and design decision, and environment change. Before the appearance of the ECC protection, the chip-level SER had steadily increased. This result is related to the fact that on-chip cache increases from EV45 to EV56, and further confirms the dominating contribution of SER from the cache. For EV56, EV6 and EV67 Alpha microprocessors, a lid coat or a wirebond attached chip capacitor (WACC) was placed in between the die and the package. In those cases the main source of α-particles was from the interconnect stack. The α-particle flux was reduced by an order of magnitude with respect to EV45 and EV5, where the package provided the major source of α-particles. The two generations of the Alpha microprocessors studied in details in this work correspond to EV56 and EV67, respectively. The ECC protection of the data cache and reduction of the α-particle flux due to the use of a lid coat has dramatically decreased the SER for EV6 and EV67. Since then the chip-level SER has steadily increased again and is dominated by core logic SER rather than cache SER as in the case of the EV4 and EV5 generations. The introduction of flip chip technology significantly increased the α-particle flux due to the lead present in the bumps. However, even for flip chip technology, the chip-level SER is still more than an order of magnitude smaller than the α-particle induced SER for EV56 [20].

As the technology and process scaling lead to higher soft error susceptibility due to the reduced noise margin, single event effects are taking a prominent position in the mainstream

integrated circuit industry. However, design and process measures have been taken in the past to keep the chip-level SER of the Alpha microprocessors very small. The results obtained from this study have demonstrated that the SER *decreases* across the generations on the device, circuit, and system levels. Although this conclusion based on the study of Alpha microprocessors, it is universally applicable for those microprocessors using the similar process and design measures. Therefore, the SE issue is well contained with the technology and process scaling.

CHAPTER VII

CONCLUSION

In this work we have investigated the impact of technology trends (both technology and circuit design) on alpha particle induced soft error rates in state-of-the-art commercial microprocessors.

At the device level, both critical charge and charge collection efficiency decrease as technologies move to the next generation. For the two technology nodes studied in this work, process improvements outpace the reduction of supply voltage and capacitance. As a result, the error rate per information bit decreases with technology scaling. This result is consistent with the previous observations by other groups.

At the circuit level, latch design has a profound impact on the SER contribution from the core logic part of the microprocessor. The transmission gate latches in the Alpha 21164 store information on floating nodes. Floating internal latch nodes are particularly sensitive to alpha particle hits due to the low critical charge and long window of vulnerability. Therefore, elimination of floating nodes is the key to improving the SER susceptibility of the chip. By comparison, the sense-amplifier based latch design used in the Alpha 21264 is more robust. The complementary signals are stored in the latch by the cross-coupled configuration. Therefore, all the latch nodes are actively driven in all circuit operating modes. The high degree of SE sensitivity of dynamic circuits has made them unattractive in most previous space applications. However, as commercial chips find their way into space-borne systems due to the performance requirement and cost considerations, it is advantageous to quantify the SE vulnerability of this type of circuit.

The circuit type not only has direct impact on the SER vulnerability of the chip but also has great influence on the frequency behavior of that vulnerability. The opposite frequency dependence trends observed in the core logic circuitry in the two generations of Alpha microprocessors have been explained using critical charge circuit simulations of the two latch circuit types. A *decrease* in the SER with increasing frequency is observed in the 21164 core logic; the dynamic nodes with no feedback devices attached (floating nodes) pose the highest SE sensitivity. The upsets on the floating nodes are persistent and can be captured by the next latch

easily. Because the error propagation delay from the floating node to the next latch is independent of the clock frequency, the SER due to the floating nodes decreases with the clock speed. This surprising discovery seems to be contradictory to the common belief that SER either increases with or is independent of frequency, and is well explained by the simulations done in this work. In the case of the 21264 Alpha core logic, the sense amplifier based latch design is particularly vulnerable to SE at latch evaluation edge, which is proportional to the clock frequency. Therefore, the SER *increases* with clock frequency. The keeper device introduces a regenerative circuit path in the dynamic circuits and enables the information restoration. The slope of the SER frequency trend is determined by the size of the keeper device. The analysis in this work reveals the root cause of the opposite frequency trends of the SER between the dynamic and static circuits. Although the function of the keeper devices is to prevent the charge leakage problem in the dynamic circuit, it could also be applied as a SEU hardening technique.

Similar read/sensing circuits are used in both the 21164 and 21264 Alpha microprocessor on-chip cache. When the SRAM cell is being read, both bitlines are pre-charged to V_{DD} just before the word line is enabled. Charge sharing results in a lowered Q_{crit} of the SRAM cell during the reading/sensing process. However, once one of the bitlines is discharged low by the SRAM cell, the vulnerability of the SRAM cell decreases due to the large capacitance on the bitlines. Therefore, the window of vulnerability is roughly constant during reading at low frequencies, and the cache SER increases linearly with clock frequency.

At the system level, the contribution from the cache dominates the overall chip level SER. However, with ECC protection the SER contribution from the core logic is becoming more important to chip level SER. Activating ECC in the Alpha 21264 further reduced the chip-level experimental SER by two orders of magnitude. The migration of technology from the 0.35 μm node to the 0.25 μm node, combined with the improvement from circuit design, reduces the overall chip level SER by an order of magnitude. The SER increases exponentially with critical charge ($\approx C_{node} \times V_{DD}$) reduction if everything else is held constant, so the projected SER based only on the lower supply voltage and node capacitance depicts a very scary picture for future technologies. However, in this work, we have shown the improvement of the SER from process scaling and circuit techniques in these two generations of the Alpha microprocessors, and large mitigation of the chip-level SER due to ECC. Although the process and design measures taken have successfully contained the SE issue, the continuous scaling of the node capacitance and

supply voltage continue to make the SE issue an important reliability concern. It is not guaranteed that process improvements will continue to compensate for power supply voltage reduction and device scaling in the future as it did in the past. Should it not be the case, chip-level or system-level redundancy might be inevitable.

REFERENCES

[1] "The national technology roadmap for semiconductors," Tech. Rep., Semiconductor Industry Association, 2000

[2] E. Normand, "Single event upset at ground level," IEEE Trans. Nucl. Sci., vol. NS-43, No. 6, pp. 2742 -2750, 1996.

[3] J.W. Howard Jr., K.A. LaBel, M.A. Carts, R. Stattel, C.E. Rogers, and T. Irwin, "Update on total dose and single event effects testing of the Intel Pentium III (P3) and AMD K7 microprocessors," Military and Aerospace Applications of Programmable Logic Devices Conference, Laurel, MD, 2001.

[4] D. Binder, E. C. Smith, and A. B. Holman, "Satellite anomalies from galactic cosmic rays," IEEE Trans. Nucl. Sci., vol. NS-22, No. 6, pp. 2675-2680, 1979.

[5] T.C. May and M.H. Woods, "Alpha-particle-induced soft errors in dynamic memories," IEEE Trans. Electron Devices, vol. ED-26, No. 1, pp. 2-9, 1979.

[6] J. F. Ziegler, "Terrestrial cosmic rays," IBM Journal of Research and Development, Vol. 40, No. 1, pp. 19-39, 1996.

[7] J. F. Ziegler, "IBM experiments in soft fails in computer electronics (1978-1994)," IBM Journal of Research and Development, Vol. 40, No. 1, pp. 19-39, 1996.

[8] M. Baze, S. Buchner, "Attenuation of single event induced pulses in combinational logic," IEEE Trans. Nucl. Science, vol. NS-44, No. 6, pp. 2217, 1997.

[9] S. Diehl and J. E. Vinson, "Considerations for single event immune VLSI logic," IEEE Trans. Nucl. Sci., vol. 30, no. 6, pp. 4501-4507, 1984.

[10] L. Massengill, "SEU modeling and prediction techniques," 1993 IEEE NSREC Short Course, Snowbird, UT.

[11] P. E. Dodd, F. W. Sexton, "Critical charge concepts for CMOS SRAMs," IEEE Trans. Nucl. Sci., vol. 42, pp. 1764, 1995.

[12] S. Buchner, M. Baze, D. Brown, D.McMorrow, and J. Melinger, "Comparison of error rates in combinational and sequential logic," IEEE Trans. Nucl. Science, vol. NS-44, No. 6, pp. 2209, 1997.

[13] S. M. Kang, Y. Leblebici, "CMOS digital integrated circuits," McGraw-Hill, 1998.

[14] P. E. Gronowski, W. J. Bowhill, R. P. Preston, M. K. Gowan, and R. L. Allmon, "High performance microprocessor designing," IEEE J. Solid-State Circuits, vol. 33, No. 5, pp. 676, 1998.

[15] L.W. Massengill, "Single event modeling in emerging commercial technologies," Military and Aerospace Applications of Programmable Logic Devices Conference, Greenbelt, MD, 2000.

[16] R. Ronen, A. Mendelson, K. Lai, Lu Shih-Lien, F. Pollack, J. P. Shen, "Coming challenges in microarchitecture and architecture," Proceedings of the IEEE, vol. 89, No. 3, pp. 325, 2001.

[17] Tanay Karnik, Bradley Bloechel, K. Soumyanath, Vivek De, Shekhar Borkar, "Scaling trends of cosmic rays induced soft errors in static latches beyond 0.18 μ," Proceedings of the 2001 Symposium on VLSI Technology Digest of Technical Papers, pp. 61-62, 2001.

[18] N. Kaul, B.L. Bhuva, S.E. Kerns, "Simulation of SEU transients in CMOS ICs," IEEE Trans. on Nucl. Sci., NS-38, pp. 1514 -1520, 1990.

[19] K. C. Holland, J. G. Tront, "Probability of latching single event upset errors in VLSI circuits," IEEE Proceeding of Southeast conference, vol. 1, pp.109-113, 1991.

[20] N. Seifert, D. Moyer, N. Leland, R. Hokinson, "Historical trend in alpha-particle induced soft error rates of the Alpha microprocessor," Proceedings of the International Physics Reliability Symposium (IRPS), pp. 259 –265, 2000.

[21] B. Benschneider et al., "A 300-MHz 64-b quad-issue CMOS RISC microprocessor," IEEE J. Solid-State Circuits, vol. 30, pp. 1203–1214, 1995.

[22] A. Chandrakasan, W. J. Bowhill, F. Fox, "Design of high-performance microprocessor circuits," IEEE Press, pp. 223, 2001

[23] Jan M. Rabaey, "Digital integrated circuits: a design perspective," Prentice Hall, 1996

[24] E. L. Petersen, J. C. Pickel, J. H. Adams, "Rate prediction for single event effects," IEEE Trans. on Nucl. Sci., NS-39, pp. 1577 -1599, 1992.

[25] T. V. Rajeevakumar, N. C. Lu, W. H. Henkels, W. Hwang, and R. Franch, "A new failure mode of radiation-induced soft error in dynamic memories," IEEE Elec. Dev. Lett., EDL-9, pp. 644 -646, 1988.

[26] J.M. Palau, G. Hubert, K. Coulie, B. Sagnes, M-C, Calvet, S. Fourtine, "Device simulation study of the SEU sensitivity of SRAMs to internal ion tracks generated by nuclear reactions," IEEE Trans. on Nucl. Sci., NS-48, no.2, pp. 225 -231, 2001.

[27] C. Dai, N. Hakim, S. Hareland, J. Maiz, and SW Lee, "Alpha-SER modeling & simulation for sub-0.25 um CMOS technology," Proceedings of the 1999 Symposium on VLSI Technology Digest of Technical Papers, pp. 81-82, 1999.

[28] S. Hareland, J. Maiz, M. Alavi, K. Mistry, S. Walsta, and C. Dai, "Impact of CMOS process scaling and SOI on the soft error rates of logic processes," Proceedings of the 2001 Symposium on VLSI Technology Digest of Technical Papers, pp. 73-74, 2001.

[29] N. Seifert, X. Zhu, D. Moyer, R. Mueller, N. Leland, R. Hokinson, M. Shade, L. Massengill, "Frequency dependence of soft error rates for sub-micron CMOS technologies," Proceedings of the International Electron Devices Meeting, Vol. 9, No. 12, pp. 644-646, 2001.

[30] L.W. Massengill, A.E. Baranski, D.O. van Nort, J. Meng, and B.L. Bhuva, "Analysis of single event effects in combinational logic -- simulation of the AM2901 bitslice processor," IEEE Trans. on Nucl. Sci., vol. 47, no. 6, pp. 2609-2615, 2000.

[31] Norbert Seifert, private communication.

[32] N. H. E. Weste, K. Eshraghian, "Principles of CMOS VLSI design," Addison-Wesley, 1993.

[33] R. Baumann, T. Hossain, S. Murata, H. Kitagawa, "Boron compounds as a dominant source of alpha particles in semiconductor devices," Proceedings of the International Physics Reliability Symposium (IRPS), pp. 297 –302, 1995.

[34] F.F. Chen, "Introduction to Plasma Physics," Plenum Press, chap. 4, 1976.

[35] J. F. Ziegler, J. P. Biersack, U. Littmark, in "The stopping and range of ions in solids, Volume 1 of the stopping and ranges of ions in matter," Pergamon Press, New York, 2nd Edition, 1996.

[36] L. Massengill, "Cosmic and terrestrial single-event radiation effects in dynamic random access memories," IEEE Trans. Nucl. Sci., vol. 43, pp. 576-593, 1996.