Impact of Dynamic Voltage Scaling and Thermal Factors on SRAM Reliability

F.R. Rosa^{a,*}, R. M. Brum^b, G. Wirth^{a,b}, F. Kastensmidt^a, L.Ost^c, R. Reis^a

^a*Instituto de Informtica PGMICRO/PPGC, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.*

^b*Escola de Engenharia PPGE, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil*

^c*Department of Engineering, University of Leicester, Leicester, United Kingdom*

Abstract

This work investigates the effects of temperature and voltage scaling in neutron-induced bit-flip in SRAM memory cells. Proposed approach allows determining the critical charge according to the dynamic behaviour of the temperature as a function of the voltage scaling. Experimental results show that both temperature and voltage scaling can increase in at least two times the susceptibility of SRAM cells to Soft Error Rate (SER). In addition, a model for electrical simulation for soft error and different voltages was described to investigate the effects observed in the practical neutron irradiation experiments. Results can guide designers to predict soft error effects during the lifetime of SRAM-based devices considering different power supply modes.

Keywords: SRAM, critical charge, supply voltage, critical charge profile, dynamic voltage scaling, semi dynamic method, soft error, dependent critical charge curve.

1. Introduction

Static Random Access Memory (SRAM) is widely deployed in almost all microelectronics devices occupying up to 90% of System on Chip (SoC) area [1]. The increasing chip power densities of SoCs allied to the continuous technology shrink are making SRAM cells more susceptible to soft errors, such as the ones caused by radiation effects [2], particularly when it comes to neutron or alpha particles strikes [3, 4]. Soft errors or Single Event Upset (SEU) induced by neutron may cause critical failures on system behaviour, which can lead to financial or human life losses [5]. With the advent of multiprocessor SoCs, researchers are investigating different alternatives to achieve both system performance and energy consumption. Among well-known techniques, dynamic voltage scaling (DVS) has been pointed as an efficient technique for balancing energy and performance gains. DVS reduces energy consumption by lowering the supply voltage and system operating frequency. However, while reducing the supply voltage decreases energy consumption, it also reduces

the critical charge, which increases the sensitivity of SRAM cells to SEUs.

The impact of using DVS on SRAM reliability has been addressed over the last years, considering different aspects. In [6] Baumann analyzes the effect of technology scaling on soft-error rates in SRAMs. SRAM soft error rate under PVT variations is investigated in [7, 8]. Underlying approaches consider only static variations of supply voltage and temperatures. Therefore, when analyzing the memory cells susceptibility to SEUs both supply voltage adjustment and resulting temperature must be taken into account. To overcome the aforementioned constraint, this paper contributes by proposing a methodology that allows determining critical charge according to the dynamic behaviour of the temperature as a function of the voltage scaling.

To the best of our knowledge, promoted methodology, described in Section 2, is the first to inspect the SRAM vulnerability considering critical charge under dynamic temperature and DVS variations. In order to demonstrate the effectiveness of our approach, Section 3 presents a trustworthy number of SPICE simulations (up to 400), which were performed considering four advanced SRAM technologies: 16nm, 22nm, 32nm, and 45nm. Afterwards, conclusions and perspectives are discussed in Section 4.

*Corresponding author at: Instituto de Informtica - Universidade Federal do Rio Grande do Sul - Caixa Postal 15064 - 91501-970 Porto Alegre - RS - Brasil

Email address: frdarosa@inf.ufrgs.br (F.R. Rosa)

2. Proposed Methodology

Proposed methodology consists of pre-characterizing the SRAM bit-cells sensibility to SEUs under a broad range of supply voltage and temperature operation conditions. This sensitivity is characterized by the critical charge (Q_{crit}), defined as the minimum amount of collected charge in the transistor junction needed to cause a bit-flip.

Figure 1. Proposed methodology for profiling the critical charge: (a) Bit-cell simulation with SEU emulated by current injection; (b) Relationship between critical charge and the injected current; (c) Binary search for the critical charge; (d) Resulting critical charge profile.

The SEU fault injection was modeled by injecting a transient current pulse in one of the SRAMs sensitive nodes, as seen in Fig. 1 (a). We employ a double exponential model [9] for the current source, as defined by:

$$I_p(t) = I_0(e^{\frac{t}{\tau_{fall}}} - e^{\frac{t}{\tau_{rise}}}) \quad (1)$$

The amount of charge provided by this current source can be obtained by directly integrating $I_p(t)$ over time. By fixing I_0 and making τ_{RISE} a function of τ_{FALL} , we can further simplify this equation to depend on one parameter only. As a τ_{FALL} dependent function, we test several parameters of τ_{FALL} . When τ_{FALL} increases the $I_p(t)$ shape changes, consequently its integral value. We then set out to find the smallest value of τ_{FALL} and, thus, the smallest amount of charge that triggers a bit-flip. This search is done by employing a binary search method, as shown in Fig. 1 (c). At each step the method halves

Figure 2. Deriving time-dependent critical-charge curves using static and semi-dynamic methods.

the interval computing the midpoint, leaving two possibilities: the subinterval between lower bound and midpoint has opposite signs or subinterval between midpoint and upper bound has opposite signs. The method selects the subinterval that has opposite signs as the new interval, which reduces the interval width by half at each iteration. The process continues until the interval is sufficiently small, we adopt a 5% margin. Underlying process is then repeated for each supply voltage and temperature condition, allowing us to construct a profile similar to the one represented by Fig. 1 (d), where a darker shade of red represents the smaller critical charge and, consequently, a more sensible cell. While the example shown in this picture is symmetrical with respect to the main diagonal, temperature and voltage may affect the critical charge differently, as shown in the next section.

Either way, the worst case will be located in the

upper-left corner, as higher temperatures and lower voltages are expected to reduce the critical charge. However, seldom chips will work in such critical conditions, at least not for a long time. To estimate the timeframe, hereby appointed vulnerability window, during which the circuit is exposed to this situation, we may combine the critical charge profile with time-dependent supply voltage and temperature curves.

The simplest way to do this is to plot the critical charge parameterized by the voltage and the temperature, which are then a function of the time. This is illustrated by Fig. 2. We refer to this approach as static method, as the voltage and the temperature are independent of each other and supplied as input. The static method helps us to determine the vulnerability window introduced by the supply voltage and temperature transitions, providing that the voltage and temperature traces supplied as input are accurate.

As an improvement to this method, we introduce a simple self-heating model for the transistors, similar to the one proposed by Bielefeld et al. [10]. Its schematic is shown in Fig. 3.

Figure 3. Self-heating transistor model. A fifth terminal T is added to the standard four-terminal model. Its voltage represents the device instantaneous temperature.

Passive elements in this model are related to the transistor's physical layers. Resistive elements relate the thermal conductivity, whereas capacitors are related to the specific heat and mass of these materials. Active elements will be related to the power-temperature relationship of the transistor. To obtain accurate numerical results, a proper calibration of these elements would be required. For the purpose of proving that DVS impacts on the circuit's reliability, we're only interested in the ratio between the time constants associated with the thermal circuit ($RC_{temperature}$) and the supply voltage circuit. By using this model, the associated effect on the bit

cell temperature can be obtained and crossed with the critical-charge profile. To account for the power supply capacitance, we introduce a lumped capacitor in the power line. We refer to this method as the semi-dynamic method, as we now consider the effect of the supply voltage over the transistor temperature. Still, we rely on the pre-characterized critical charge profile to obtain the time-dependent critical charge curve. This method is an alternative to a complete dynamic approach, in which the temperature would be recalculated at each timestep as a function of the transistor power consumption. While such an implementation could provide more accurate results, it is not compatible with commercial, well-established tools such as Cadence Spectre and Synopsys Hspice.

The promoted semi-dynamic method works as follows. First, a source current is embedded in the SRAM description to model the SEU. Thus, our methodology searches the Q_{crit} for each temperature and voltage condition according to the desired granularity, generating a Q_{crit} profile matrix. Afterward, we attach the power model to the original SRAM description, which is simulated in two distinct phases. In the first phase, a high throughput memory access in overdrive Vdd is triggered; driving more current, and therefore increasing the SRAM cell temperature. Our methodology uses a 0.5 activity factor, a good average approximation (i.e. neither conservative nor pessimistic) for the least significant bits in a cache memory [11]. In this context, the cell temperature reaches a plateau after a hundred of nanoseconds.

In the second phase, the memory voltage supply is reduced to a low power level, which is defined according to the adopted technology (e.g. 0.5V for 32nm). The initial temperature acquired during the high memory operation dissipates through the energy transported to the package. At this point, the SRAM Cell has a low voltage supply and a high temperature, therefore, rising by several times its vulnerability to SEU. The period where the Q_{crit} is lower than expected bounds the vulnerability window, which is defined according to the confluence of temperature and low Vdd. The energy dissipation occurs according to the $RC_{temperature}$ and its ratio with the $RC_{electric}$, which is determined by the technology construction.

Afterwards, the simulation generates a temporal voltage and temperature trace. This trace is analyzed at each time interval, consulting the profile matrix to acquire the Q_{crit} at this specific temperature

Figure 4. Critical charge profile obtained for a 6T-SRAM cell built for the PTM 16nm (a), 22nm (b), 32nm (c), and 45nm (d) technologies. The sensibility to SEU's (i.e. Q_{crit}) worsens from blue to red.

and voltage. In the next section, we present some of the obtained results for the **Predictive Technology Model (PTM) from Arizona State University [12]** 16nm, 22nm, 32nm and 45nm technologies, employing both the static and the semi-dynamic methods.

3. Experimental Setup and Results

Cadence SPECTRE simulator is employed to perform all experiments in an Intel Xeon L5520 2.27 GHz with 32 GB RAM. 6T-SRAM cells using two inverters and two passing gates are used as the case study. Fig. 4 shows the critical charge profile obtained for the four PTM technologies. This profile has a resolution of 0.05 V and 0.1 C, covering temperatures ranging from -40C to 125C and supply voltages ranging from 0.35 V to 1.2 V depending on the technology.

To better understand this profile, we can plot horizontal or vertical slices of one matrix, by setting one of the variables as a parameter and plotting the critical charge as a function of the remaining variable. Fig. 5 shows the critical charge as a function of the temperature for various supply voltages for the

32nm technology, whereas Fig. 6 shows the critical charge as a function of the supply voltage for selected temperatures. From these profiles, it becomes clear that the supply voltage has a stronger influence than the temperature over the considered operating ranges. Notwithstanding, while operating at higher voltages, the circuit sensibility is strongly influenced by the temperature: moving from room temperature to 125C can almost halve the critical charge of the bit-cell.

The results for other technologies present a similar trend, thus the same conclusions from the remaining experiments. Indeed, higher supply voltage improves radiation resilience, reaching five times between extreme points. High Vdd voltage improves the noise margin, and consequently the amount of charge necessary to provoke a memory state flip.

Nevertheless, dynamic and static power dissipation effects are directly related to the supply voltage, therefore reducing the battery life in mobile applications for exemple. When applying DVS, both effects of temperature and supply voltage are seen in a dynamic way where there are periods when the voltage

Figure 5. Critical charge over temperature variation with fixed supply voltage.

Figure 6. Critical charge over supply voltage variation with fixed temperature.

is constant, and the temperature behave according to a model and periods that the supply voltage and temperature are changing to adjust to the DVS.

Fig. 7(a) shows the Q_{crit} behaviour in dynamic Vdd voltage scenarios using our static method applied to the 32nm technology. The black line represents the voltage supplied, at beginning the voltage supply is the nominal value, 0.9V, the next 30 points represent a linear voltage increase up to 1.1V and maintaining it until the point 50. Thus, to reproduce a dynamic voltage scaling, dropping the voltage aggressively, however reaming operational. The dotted four curves represent the Q_{crit} behaviour for four at fixed temperatures.

Fig. 7(b) shows the dynamic behavior of the temperature as a function of the voltage scaling in the same scenario using the proposed semi-dynamic method. The Q_{crit} worst-case happens right after completing the voltage downscaling, as the circuit is still in the process of dissipating the heat induced by the higher supply voltage previous period, increas-

ing the circuit sensitivity to radiations effects. Additionally, after the voltage downscaling, the aforementioned vulnerability window emerges augmenting more the sensitivity.

Figure 7. Critical-charge as a function of time, obtained for a 32nm bit-cell using (a) the static method and (b) the semi-dynamic method.

In order to investigate different $\frac{RC_{temperature}}{RC_{electric}}$ ratios, different heat dissipation conditions and several ratios were simulated. Fig. 8 displays (from top to down) the voltage wave emulating a voltage transition, the resulting temperature from this context according our model, Q_{crit} duration, and vulnerability window zoom for ratios from 1 to 100K considering the four tech technologies. As expected, the Q_{crit} presents lower value than expect bounds the vulnerability window. As this RC ratio increases the heat dissipates faster, consequently, attenuating the vulnerability window with low voltage and high temperature. Fig. 8 shows the Q_{crit} for 16nm (a), 22nm

Figure 8. shows (from top to down) voltage waveform, temperature, critical charge, and vulnerability window for the 16nm (a), 22nm (b), 32nm (c), and 45nm (d) PTM technology employing our proposed semi-dynamic method.

(b), 32nm (c), and 45nm (d) for the 1 to 100K RC ratio range. A fast heat dissipation reduces the coincidence between low voltage supply and high temperature. As consequence, the vulnerability window decreases.

4. Conclusion and Future Directions

In this paper, we devised an approach to determine the critical charge of SRAM bit-cells under voltage scaling, considering transient temperature effects. By introducing a simple self-heating model, we demonstrated that memory circuits are more sensitive to radiation when the supply voltage is transitioning from a higher to a lower voltage, reducing the Q_{crit} up to 5 times. The heat dissipation in these devices typically happens at a much slower rate than their capacity to transition from one supply voltage level to another. DVS technique improves SoC performance and power consumption trade-off. Additionally, this technique and similar ones are deployed automatically by the processor management unit, relying on external factors as temperature and workload. This way is impossible correct predicts it employment and duration, exposing more certain code

fragments than others. Requiring by the system engineer a greater understanding and study of performance enhancement techniques to analyse the soft-error rate for a given system. Solutions for this issue may include delaying supply voltage transition times for radiation-hardened circuits or a gradual, step-based supply voltage transition. In the future, we intend to study new approaches to radiation-aware voltage scaling.

5. References

- [1] ITRS, International Technology Roadmap for Semiconductors (2013).
- [2] R. Baumann, Radiation-induced soft errors in advanced semiconductor technologies, IEEE Transactions on Device and Materials Reliability 5 (3) (2005) 305–316. doi:10.1109/TDMR.2005.853449.
- [3] K. Rodbell, D. Heidel, H. Tang, M. Gordon, P. Oldiges, C. E. Murray, Low-Energy Proton-Induced Single-Event-Upsets in 65 nm Node, Silicon-on-Insulator, Latches and Memory Cells, IEEE Transactions on Nuclear Science 54 (6) (2007) 2474–2479. doi:10.1109/TNS.2007.909845.
- [4] P. Hazucha, C. Svensson, Impact of CMOS technology scaling on the atmospheric neutron soft error rate, IEEE Transactions on Nuclear Science 47 (6) (2000) 2586–2594. doi:10.1109/23.903813.
- [5] Toyota Case: Single Bit Flip That Killed | EE Times.

URL http://www.eetimes.com/document.asp?doc_id=1319903

- [6] R. Baumann, Soft errors in advanced computer systems, *IEEE Design Test of Computers* 22 (3) (2005) 258–266. doi:10.1109/MDT.2005.69.
- [7] G. Tsiligiannis, E. Vatajelu, L. Dilillo, A. Bosio, P. Girard, S. Pravossoudovitch, A. Todri, A. Virazel, F. Wrobel, F. Saigne, SRAM soft error rate evaluation under atmospheric neutron radiation and PVT variations, in: *On-Line Testing Symposium (IOLTS)*, 2013 IEEE 19th International, 2013, pp. 145–150. doi:10.1109/IOLTS.2013.6604066.
- [8] E. Vatajelu, J. Figueras, Robustness analysis of 6t SRAMs in memory retention mode under PVT variations, in: *Design, Automation Test in Europe Conference Exhibition (DATE)*, 2011, 2011, pp. 1–6. doi:10.1109/DATE.2011.5763159.
- [9] G. Hubert, L. Artola, Single-Event Transient Modeling in a 65-nm Bulk CMOS Technology Based on Multi-Physical Approach and Electrical Simulations, *IEEE Transactions on Nuclear Science* 60 (6) (2013) 4421–4429. doi:10.1109/TNS.2013.2287299.
- [10] J. Bielefeld, G. Pelz, H. Abel, G. Zimmer, Dynamic SPICE-simulation of the electrothermal behavior of SOI MOSFET's, *IEEE Transactions on Electron Devices* 42 (11) (1995) 1968–1974. doi:10.1109/16.469405.
- [11] H. Homayoun, M. Rahmatian, V. Kontorinis, S. Golshan, D. Tullsen, Hot peripheral thermal management to mitigate cache temperature variation, in: *2012 13th International Symposium on Quality Electronic Design (ISQED)*, 2012, pp. 755–763. doi:10.1109/ISQED.2012.6187576.
- [12] Y. K. Cao, What is Predictive Technology Model (PTM)?, *SIGDA Newsl.* 39 (3) (2009) 1–1. doi:10.1145/1862891.1862892.