Two-Dimensional Markov Chain Analysis of Radiation-Induced Soft Errors in Subthreshold Nanoscale CMOS Devices

Pooya Jannaty, Florian C. Sabou, Matthew Gadlage, R. Iris Bahar, Joseph Mundy, William Patterson, Robert A. Reed, Senior Member, IEEE, Robert A. Weller, Ronald D. Schrimpf, and Alexander Zaslavsky

Abstract—Radiation-induced soft errors have been a reliability concern for logic integrated circuits since their emergence. Feature-size and supply-voltage reduction require the analysis of soft-error sensitivity as a function of technology scaling. In this paper, an analytical framework based on Markov chains and queue theory is presented for computation of alpha-particle-induced soft-error rates of a flip-flop operated in the subthreshold regime. The proposed framework is capable of reflecting the technology parameters such as supply voltage \( V_{dd} \), channel length, process-induced threshold variation, and operating temperature. As an example, the framework is used to investigate the mean time to error of flip-flops built in a 32 nm fully-depleted silicon-on-insulator technology operating in the subthreshold regime subject to two limiting fluxes of alpha particle radiation: high at 100 \(( \alpha / \hbar c {m}^2 \)) and ultra-low alpha (ULA) emission 0.002 \(( \alpha / \hbar c {m}^2 \)).

Index Terms—Alpha-particle radiation effects, CMOS devices, logic devices, radiation effects in devices, reliability, single event upset, soft errors, SRAMs.

I. INTRODUCTION

INTEREST in ultra-low-voltage digital circuits has been growing in the past years, prompted by the need to reduce the overall energy consumption [1], [2]. If the power reduction requirement is particularly stringent in some specific application, such as biomedical implants, environmental-monitoring devices or space systems, reducing \( V_{dd} \) into the subthreshold regime could prove useful [3], [4]. This reduction in the operating voltage raises the need for probabilistic frameworks capable of analyzing the effect of noise sources on low-power devices. The error rate estimates arising from such models can serve as a guideline for designing logic circuits operated at ultra-low or subthreshold \( V_{dd} \).

In our previous work [5]–[7], a probabilistic framework for the analysis of thermal-noise-induced variations in the logic stability of memory circuits such as flip-flops has been developed. As shown in Fig. 1, noise in a flip-flop operated in the subthreshold regime can be represented as a Poisson process with independent distributions for the arrival and departure of carriers at a drain or source node [8]. The node-capacitor charge is modeled as a queue where the transistor current fluctuations are considered to be random carrier arrival and departure events. The result is a two-dimensional (2D) queue that contains all the available states for the system considered in terms of the number of electrons \( k_1, k_2 \) on the two inverters.

The two valid logic states, ‘0’ and ‘1’, correspond to states in opposite corners of the 2D queue. Considering only the effects of thermal noise, previous work [5], [6] has shown that thermally-induced transitions between the stable states are exponentially rare and occur predominantly on a diagonal path between the stable corner states. Recently we have generalized our approach to a numerical solution of the full 2D queue [7]. As described in the next section, radiation events create a wide range of charge amounts at either electronic node. This sudden change in the electron population on the two inverters also changes the state of the system as represented in the 2D queue, often moving the system far from the stable corner states. The cumulative effects of the thermal broadening of the logic states and radiation render all states in the 2D queue to have a non-negligible transient occupation probability. Therefore, a complete solution of the full 2D queue is required to investigate the soft-error rates induced by radiation.

Manuscript received March 27, 2010; revised June 09, 2010; accepted August 04, 2010. Date of publication October 04, 2010; date of current version December 15, 2010. The work at Brown was supported in part by NSF NIRT award CCF-0506732, NSF award ECCS-0701635, and by the Defense Threat Reduction Agency, Basic Research Award HDTRA1-10-1-0013. The work at Vanderbilt was supported by AFOSR through the MURI program and by the Defense Threat Reduction Agency, Basic Research Award HDTRA1-10-1-0013. P. Jannaty, F. C. Sabou, R. I. Bahar, J. Mundy, W. Patterson, and A. Zaslavsky are with the Department of Physics and the School of Engineering, Brown University, Providence, RI 02912 USA (e-mail: pooya@brown.edu).

M. Gadlage, R. A. Reed, R. A. Weller, and R. D. Schrimpf are with the Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37235 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNS.2010.2068561

Fig. 1. The modeled flip-flop circuit. Capacitors \( C_1 \) and \( C_2 \) represent the node capacitances associated with each inverter. For each transistor, the charging and discharging rates \( \lambda \) and \( \mu \) determine the electron populations on the node capacitances.

0018-9499/$26.00 © 2010 IEEE
The paper is organized as follows: Section II details the simulation process for the alpha particle radiation. Section III briefly reviews the formulation for the 2D Markov chain (fully described in [7]). Section IV presents the soft error rates for the modeled flip-flop. Finally, Section V contains the conclusions.

II. ALPHA PARTICLE RADIATION

The interaction of alpha particles with semiconductor devices can have a significant impact on electronic system reliability. Alpha particles can be emitted by trace impurities (such as uranium) in packaging materials [9], and can be byproducts of neutrons produced by cosmic radiation [10]. An ionizing particle, such as an alpha particle, can induce what is known as a single-event upset (SEU)—a change in the state of a node due to a single ionizing event. An alpha particle depositing charge in an SRAM cell and causing the bit to flip is a classic example of an SEU. As integrated circuits (ICs) shrink in feature size, the charge needed to cause an SEU also decreases, and as a result, single events are projected to become one of the biggest reliability issues for advanced electronics.

Charge generated by an alpha particle has to be collected at a circuit node to cause an SEU. For the radiation simulations in this paper, the volume in which charge can be collected is approximated as the body of the fully-depleted silicon-on-insulator (FD-SOI) device. This technology was chosen for analysis since the charge-collection volumes are relatively simple and it also holds the promise to become a mainstream technology for general applications. For the simulations in this paper, we used a 32 nm FD-SOI technology described in [11] with an effective gate length $L_G \sim 25 \text{ nm}$ and Si channel thickness of 10 nm.

To gain insight into the amount of charge an alpha particle can generate in a highly-scaled FD-SOI SRAM cell, simulations were performed using the Monte Carlo radiative energy deposition (MRED) software [12]. MRED is a simulation tool, based on the Geant4 libraries [13], for quantifying the energy deposited by radiation in microelectronic devices. A $10^4 \text{ nm} \times 10^4 \text{ nm} \times 7 \times 10^2 \text{ nm}$ silicon block was simulated for a range of sensitive volumes, as seen in Fig. 2. The sensitive volume (representing the volume for which charge is collected, as described above) is the region in which the deposited energy is recorded by the MRED simulation and corresponds to the channel region of one transistor in the FD-SOI SRAM cell.

The average charge generated in the sensitive volume is plotted versus incident particle energy in Fig. 3 for $1.3 \times 10^6$ alpha particles normally incident on the top of the silicon block. The charge generated in the sensitive volume is obtained from the MRED simulations based on the average energy needed to create an electron hole pair in silicon ($\sim 3.6 \text{ eV}$ at room temperature). If the width of a transistor in an FD-SOI SRAM doubles, the area under the gate (i.e., the body region) also doubles. Therefore, for a given particle fluence, the number of alpha particles that deposit energy in the body increases proportionally. To confirm this, the sensitive volume was increased from $25 \text{ nm} \times 25 \text{ nm} \times 10 \text{ nm}$ to $200 \text{ nm} \times 25 \text{ nm} \times 10 \text{ nm}$. The simulations for 5 MeV alpha particles in Fig. 2 show that the number of events scales linearly with the size of the sensitive volume. This property will be used later when transistors with larger sensitive volumes are analyzed, both for normal and isotropic alpha particle incidence.

III. FORMULATION OF TWO-DIMENSIONAL MARKOV CHAIN

In this section, an outline of the probabilistic formulation of a 2D Markov chain is presented. Both the steady-state and transient behavior of the system are formulated in terms of probability density functions (PDFs) and the results are used in Section IV.

For generality, a 2D chain of $N+1$-by-$N+1$ states, extending from the stable state $(N,0)$ to the other stable state $(0,N)$ is assumed, see Fig. 4, where state $(m,n)$ corresponds to the flip-flop having $k_1 = m$ electrons on one inverter and $k_2 = n$ electrons on the other. The rates of transmission from any state to its adjacent states are given in 4 matrices of the same size named $u$, $d$, $r$ and $l$, for up, down, right and left, respectively.
and are four independent, generating events according and since the system has no fluctuations.

Fluctuations

A. Calculating the Mean Time to Failure Due to Thermal Fluctuations

As mentioned earlier, in the subthreshold regime, the random sources governing the transitions have Poisson distributions, having their mean equal to the corresponding transition rates [8]. Associated to each state \((m, n)\) are four independent Poisson sources for each main direction with the corresponding rates of \(u_{m,n}, d_{m,n}, r_{m,n}\) and \(l_{m,n}\), generating events according to their respective means. They can be written as

\[
P_{u0}(t) = \text{Prob}('\text{up' event happens indepentently)}/dt = u e^{-ut} \quad (1)
\]

and similar equations for moving in the other three directions.

For the system to move to, say, the upper state at time \(t\), firstly, an ‘up‘ event should happen, secondly, no other events, namely ‘down‘, ‘right‘ or ‘left‘ should have happened until time \(t\). This can be written mathematically as

\[
P_{u\alpha}(t) = \text{Prob('up' event no other events until time t)}/dt
\]

\[
= P_{u0}(t) \left( 1 - \int_0^t P_{\alpha0}(t')dt' \right)

\times \left[ 1 - \int_0^t P_{\alpha0}(t')dt' \right]^{-1}

\times \left( 1 - \int_0^t P_{\alpha0}(t')dt' \right)

= u e^{-(u+d+r+l)t} \quad (2)
\]

For an arbitrary direction \(x\), one has

\[
P_{xx}(t) = xe^{-(u+d+r+l)t}, \quad x = u,d,r,l. \quad (3)
\]

The transit time from the state \((m, n)\) to the final state can be represented as the sum of transit time to an adjacent state and transit time from there to the final state. The sum of two transit times multiplied by their joint probability, integrated over all possible values for them, applied for all directions, gives the mean transit time:

\[
T_{m,n} = \sum_{x=i,d,r,l} \int_0^\infty \int_0^\infty (t_{ix} + t_{xf}) P_{ixf}(t_{ix}, t_{xf}) dt_{ix} dt_{xf} \quad (4)
\]

where \(t_{ix}\) is the transit time from the initial state to the neighbor state \(x\) and \(t_{xf}\) is the transit time from the neighbor state \(x\) to the final state or sink. The index \((m, n)\) is implicit in all variables although dropped for better readability.

By the definition of a Markov chain, the probability distribution for \(t_{xf}\) is independent of \(t_{ix}\); since the system has no memory of which path it has taken so far. As a result:

\[
T_{m,n} = \sum_{x=i,d,r,l} \int_0^\infty \int_0^\infty (t_{ix} + t_{xf}) P_{ix}(t_{ix})P_{xf}(t_{xf}) dt_{ix} dt_{xf} \quad (5)
\]

As described in [7], (5) can be solved using a recursive approach, which reduces the memory requirements and the computation time significantly. Starting from the bottom-right corner in the 2D queue, \(T_{0, N}\) is expressed in terms of the transit times of the preceding diagonal, \(T_{1,N}\) and \(T_{0,N-1}\). The transit times of this diagonal are in turn expressed in terms of those of the next diagonal, and so on, until all the diagonals are recursively scanned and the other corner of queue is reached. Once the value of \(T_{N,0}\) is determined, it can be back-substituted into the recursive equations derived, until the values for all the transit times are calculated.

B. Steady-State Thermal Distribution

The same numerical approach introduced in [7] can be used to derive the steady-state thermal distribution of a 2D system. This distribution will be used later in estimating the radiation-induced soft-error rates in a flip-flop.

Given an initial probability distribution for the states of the system, the continuity equation governs the dynamics of probability flow between the states:

\[
-\alpha \nabla \cdot \mathbf{P}^\text{ss}(t) = 0 \quad (6)
\]

in which \(\alpha\) is a constant of proportionality and \(\nabla \cdot \mathbf{P}^\text{ss}(t)\) indicates the net probability flow out of the state \((m, n)\). As \(t \to \infty\) the system acquires a global steady state and the probabilities lose their time dependence, making the left hand side of (6) zero. In this condition, expanding the discrete divergence in the right hand side of the same equation reads

\[
(u_{m,n} + d_{m,n} + r_{m,n} + l_{m,n}) P^\text{ss}_{m,n} = d_{m+1,n} P^\text{ss}_{m+1,n}

+ u_{m-1,n} P^\text{ss}_{m-1,n} + l_{m,n+1} P^\text{ss}_{m,n+1} + r_{m,n-1} P^\text{ss}_{m,n-1}. \quad (7)
\]
C. System Transient Behavior

Given a perturbation to the system, for example due to an incident alpha particle, the system changes its state abruptly, most likely to a state away from the two stable states. Such a volatile system will eventually relax to either of its stable states via a transient route on the 2D queue. Depending on the strength of the initial perturbation and the probabilistic nature of the system in responding to it, the final state might not be identical to the initial one, leading to a soft error. It is desired to know the probability of collapse to either of the stable states, given an arbitrary initial state anywhere on the 2D queue. The equations to be solved are

\[ P_{m,n}^{tr} = z_{m,n} (u_{m,n} P_{m+1,n}^{tr} + d_{m,n} P_{m-1,n}^{tr} + l_{m,n} P_{m,n-1}^{tr} + r_{m,n} P_{m,n+1}^{tr}) \]  

(8)

where \( P_{m,n}^{tr} \) denotes the probability of collapse to the state \((0,N)\), let the system initiate from the state \((m,n)\). The probability of collapse to the state \((N,0)\) is \(1 - P_{m,n}^{tr}\) for obvious reasons. Eq. (8) can be solved using the same technique described in [7].

The algebraic nature of the computational framework makes it possible to quantify the effects of important varying parameters such as the manufacturing-process-induced threshold variations of devices in integrated circuits. The variations in the threshold voltage for cutting edge technologies are approximately 10–15% of the nominal \(V_T\) [14]. The method described here accounts for such threshold variations by adjusting the transition rates between neighboring states to reflect the mismatch between the transistors in the flip-flop. To mirror the known spread in threshold voltage variations, FD-SOI CMOS flip-flops with perfectly matched and 15% mismatched \(V_T\) have been investigated and the results are shown in Fig. 6. It can be observed that threshold variations have an important effect in determining the stability domains for given devices with respect to transient perturbations, such as the ones created by alpha particle strikes.

The sharp switching staircase in Fig. 6 reflects the probability domains for single flip-flops with a mismatch of either zero or 15% in \(V_T\) between the inverters. For a large ensemble of devices, the distribution of threshold voltage variations would lead to a smoother average switching probability curve falling between the extreme cases illustrated in Fig. 6.

IV. Estimating Radiation-Induced Soft Error Rates

Equipped with the solutions for probability moment distributions, as well as the steady-state and transient behavior of the 2D system for each of its individual states, a systematic approach for a first-order estimation of the PDF of failure for a flip-flop under radiation is established in this section.

Assume that the function \( event(q) \) contains the number of single particle events out of \(N_{inc} \) incident particles, as a function of deposited charge \(q\) on a sensitive volume as discussed in
previous sections. If the flip-flop system is in state \((m, n)\) not far from \((N, 0)\) and some charge \(q\) is deposited on its first capacitive inverter, then the system moves to the state \((m + q, n)\). Therefore the probability for the system to relax to the state \((0, N)\), leading to a soft error, is equal to \(P_{\text{tr}}^{m+q,n}\) which is previously obtained as the solution of (8). This probability, averaged on the two inverters, times \(\text{event}(q)\), integrated over charge \(q\), times the steady-state probability \(P_{\text{ss}}^{m,n}\), summed over all the states \((m, n)\) with condition \(m + n < N\), gives the number of soft-error events due to \(N_{\text{inc}}\) incident particles. This number divided by \(N_{\text{inc}}\) gives the probability for a single particle to cause an error:

\[
P_{\text{single}} = \frac{1}{N_{\text{inc}}} \sum_{m+n<N} P_{\text{ss}}^{m,n} \int_{0}^{q} \frac{P_{\text{tr}}^{m+q,n} + P_{\text{tr}}^{m,n+q} \text{event}(q) dq}{2}
\]

(9)

Implicit in this equation is that the distribution \(P_{\text{ss}}^{m,n}\) is renormalized over the states \(m + n < N\).

Assuming that the radiation flux on a single device is known, the average time interval between two incident particles is \(T = 1/flux\). For a soft-error event to happen between the times \(t\) and \(t + dt\), none of the \(t/T\) incident particles arriving before time \(t\) should cause any events, while the next particle arriving between the aforementioned interval, with probability \(dt/T\), should cause an event. This, by definition, gives the time-dependent PDF for soft-error events in a power-law form:

\[
P_{\text{error}}(t) = \frac{1}{T} P_{\text{single}}(1 - P_{\text{single}})^{t/T}.
\]

(10)

In obtaining (9) and (10), the following assumptions were made:

- Incident particles were assumed to arrive one at a time on a single device within a device relaxation time frame. This is a legitimate assumption for almost all practical purposes: even with a flux of 100 (alpha/hour·cm²), the average time between two arrivals on an area of 1 cm² is 36 seconds which is orders of magnitude larger than device relaxation times (\(\sim 1\) ns, [7]).
- In (9) it was assumed that the deposited charge will not cause the system to move beyond its available states. Mathematically, this means that \(q\) is always such that \(0 \leq m + q \leq N\) and \(0 \leq n + q \leq N\). For instances that this condition does not hold, the corresponding \(P_{\text{tr}}\) in (9) can be replaced by unity as a worst case scenario, to enforce a soft error. However, the violation of this condition is not a concern for the states \((m, n)\) far from the stable state \((N, 0)\) as the overall effect will be exponentially attenuated by the term \(P_{\text{ss}}^{m,n}\) for such states.
- The detailed physics of charge deposition was ignored in the derivations above and it was assumed that all the charge deposited in the sensitive volume is collected by the device. This assumption is based on the fact that the sensitive volume coincides with the transistor channel.

To illustrate this method, consider the event-charge curves for alpha particles of different energies shown in Fig. 7. To obtain each of these curves, \(10^8\) alpha particles were incident on an area of 100 \(\mu m^2\) and only the fraction of charge deposited in a sensitive volume representing the transistor channel was recorded. All the curves in Fig. 7(a) correspond to normally-incident radiation whereas Fig. 7(b) compares the normally-incident and isotropically-incident 5 MeV alpha particles over \(2\pi\) steradians.

The soft error rates for alpha particles with various incident energies are calculated and shown in Fig. 8 in units of failures in time (FITs)\(^3\) for an FD-SOI CMOS flip-flop of gate length \(L_G = 25 \mu m\) and \(W/L\) ratio of 15 for its NMOS transistors and 60 for its PMOS transistors (for matched current drive). At \(V_{dd} = 0.3\) V, such a flip-flop contains \(\sim 1000\) electrons. The error rates were calculated for two alpha particle fluxes, 100 (alpha/hour·cm²) and 0.002 (alpha/hour·cm²) [15], the latter being the onset of ultra low alpha (ULA) emission.

As Fig. 8 suggests, the error rates for incident alpha-particle energies less than 3 MeV are roughly constant and independent

\(^3\)FIT is the number of failures in 1 billion device-operation hours.
of radiation energy. This is because the deposited charge at these energies is usually more than half the charge capacity of the flip-flop, see Fig. 3. Therefore, as illustrated in Fig. 6, almost all the incident particles will translate the system to the domain of the alternative stable state, causing a soft error. The mean time to failure becomes of the order of the average time between arrival of particles. For normally-incident alpha particles with energies greater than 3 MeV however, the number of particles that produce upsets is so small that their contribution to the FIT rate is negligible. However, in the case of isotropic radiation, the error rates are much higher, as shown in Fig. 8, due to the broader distribution of deposited charge.

The error rates presented in Fig. 8 agree with conventional analyses, which show that the radiation-induced soft error mechanism is strongly dominated by a critical charge collection value, at least for a simple system, like the flip-flop example used in this paper. The precise value of the critical collection threshold is influenced by several factors, such as device technology node and geometry, threshold variation, and operating voltage and temperature. Equally critical for the accurate prediction of radiation-induced soft errors is the correct evaluation of the interplay between the steady state, small thermal fluctuations that result in the broadening of the logic states, as illustrated in Fig. 5, and the large, transient fluctuations driven by the collected charge resulted after a radiation event. The computational framework presented in this paper, stemming from our previous work on thermal noise [5], [7], treats both fluctuation sources in a unified way.

V. CONCLUSION

In this paper, a probabilistic framework that has been applied in previous work to investigate the thermal-noise-induced soft errors has been extended to analyze the effects of radiation events on the logic stability of CMOS devices. While the thermal-induced variations have been shown in [5], [6] to be mostly confined to the diagonal region of the 2D queue, the electron population perturbations created by a particle strike span the entire 2D queue, which can be solved numerically [7].

As a predictive tool, the framework has been used to analyze the effects of alpha particle radiation on the logic stability of flip-flops built in a 32 nm silicon-on-insulator technology operating in the subthreshold regime. Two limiting values for the flux of alpha particle radiation have been considered: 100 $(\alpha / \text{cm}^2 \cdot \text{s})$ and ultra low alpha emission $0.002 (\alpha / \text{cm}^2 \cdot \text{s})$. For high levels of normally-incident alpha particle radiation with incident energies below 4 MeV, the mean time to logic error is on the order of $10^{10}$ s. Considering the large number of memory devices on integrated circuits, such error rates, though small on an individual device basis, represent a significant reliability concern. For isotropically-incident alpha particles, higher energies are also problematic, due to the broader distribution of deposited charge.

REFERENCES


