Gate-Level Mitigation Techniques for Neutron-Induced Soft Error Rate

Harmander Singh Deogun, Dennis Sylvester, David Blaauw

Department of EECS, University of Michigan, Ann Arbor, MI, US 48109
{hdeogun,dmcs,blaauw@umich.edu}

Abstract

Neutron-induced single-event upsets have become increasingly problematic in aggressively scaled process technologies due to smaller nodal capacitances and reduced operating voltages. We present a probability-based analysis of neutron strikes on combinational logic chains and investigate techniques to increase circuit robustness in terms of decreasing the probability of upsetting the capturing latch given a particle strike. We show that using a technique of inserting simple cross-coupled inverter pairs on error prone sites, as well as intelligently placing lower V_{th} devices and readjusting device width, can increase the robustness by nearly 20% thereby increasing the mean time between soft errors by almost 25%. This technique incurs substantially less overhead than traditional redundancy approaches to mitigating soft errors.

1. Introduction

Radiation-induced soft errors on large-scale integrated circuits are becoming increasingly problematic as device sizes are scaled down, operating voltages are reduced, and node capacitances shrink [1,2]. Historically, it has been known since the early 1970s that alpha particles cause single-event upsets (SEUs) in memory arrays such as DRAMs and SRAMs due to their small transistor sizes and small storage node capacitances. In general, memory arrays are protected by error correction schemes to enhance their robustness in light of alpha-particle strikes [3, 4]. More recently, random combinational logic is becoming susceptible to SEUs due to the aforementioned technology scaling trends. Cosmic rays (most prominently neutrons) can strike the silicon substrate and deposit sufficient charge at internal device nodes to cause transient signal glitches.

While both neutrons and alpha particles deliver parasitic charge, the charge deposited by neutrons (25-150 fC/μm) is much greater in magnitude than alpha-particles (4-16 fC/μm) [5]. A strike by a neutron generates enough charge to disrupt the output of random combinational logic while a strike by an alpha particle would have very little to no effect. Moreover, materials and packaging to shield against alpha particles are available whereas no practical packaging or shielding solution for neutrons exists [6]. For example, roughly one foot of concrete is required to lower the neutron flux by just 1.4X [7].

Considering that neutron flux is the primary mechanism for SEUs and that no viable means for neutron strike shielding exists, we focus our attention on circuit techniques to combat SEUs. Some general approaches [2,8-11] have been developed to address this problem from a circuit design perspective. In [2], majority voting is used with several redundant latches that sample the data input at different time points. An adaptive approach is employed in [8] to selectively provide parity, double, and triple error detection. Explicit capacitors were added on the keeper node of domino circuits in [9] to increase the critical charge (Q_{crit}) required for a single event upset. In [10], circuit duplication and time redundancy were jointly used to reduce the soft error rate. Finally, [11] examined the effect of high threshold voltages on soft error rates.

Most of these approaches rely predominantly on major tradeoffs in device area (with some approaches nearly doubling the area), with similar increases in power consumption. Delay is often substantially increased as well and there is added design complexity. In this paper our goal is to enable low overhead (in delay, power, and area) design solutions to the soft error problem using easily-adopted design techniques.

Our primary aim is to develop a systematic, probability-based approach to increasing the critical charge, Q_{crit}. We propose the selective use of cross-coupled latches to raise Q_{crit} above that of the charge deposited by a neutron strike and judicious use of low V_{th} gates to gain back performance. We also study the redistribution of circuit size, including P/N beta ratios, as another technique to improve soft error immunity. Moreover, we use a probability-based model to identify error-prone nodes that can then be addressed using the above techniques.

In Section 2, we provide background on neutron strikes and build a probability-based model. In Section 3, we apply gate-level mitigation techniques using the probability model developed in Section 2. In Section 4 we present our results and Section 5 concludes the paper.

2. Neutron strike induced failures

When a neutron strikes the silicon substrate of a device, it fissures the silicon nucleus, resulting in the generation of electron-hole pairs [12]. These mobile carriers in the substrate can lead to current pulses with sufficient amplitude and duration to cause the output to unexpectedly switch. This transient glitch on the combinational circuit can be propagated through the logic path and erroneously latched into a sequential element. However, there are three types of natural deterrents that help to attenuate the propagation of such transient glitches [13]:

1) **Temporal Masking –** when a transient glitch reaches the latch but is not captured due to the latch being in its opaque state.

2) **Logical Masking –** when a transient glitch reaches a logic gate where the output is determined only by its other inputs.

3) **Electrical Masking –** when a transient glitch is electrically attenuated by subsequent logic gates.

Temporal masking is becoming a less effective method in screening out transient glitches due to heavy pipelining and the resulting very short cycle times for which sequential elements are

Proceedings of the Sixth International Symposium on Quality Electronic Design (ISQED'05)
0-7695-2301-3/05 $ 20.00 IEEE

transparent over a larger fraction [14]. Logical masking is inherent in most circuits and to increase it, the logic in the circuit would need to be changed. Since temporal masking is becoming less effective and since we do not want to change the logic in a circuit, we focus on improving the effectiveness of electrical masking by increasing node Q_{crit}, and thereby increasing the attenuation of marginal transient glitches.

A substantial portion of our work relies on the flux of sea-level neutrons and the extracted probabilities from such data. This data is available from the Joint Electron Device Engineering Council (JEDEC) Solid State Technology Association standard JESD89 [15]. Using the data points from JESD89, an energy vs. neutron flux plot can be generated as shown in Figure 1. We have taken data points for neutron energies of 1-6 MeV and fit the flux to a decaying exponential as given by,

$$f(e) = A\exp\left[\frac{-e}{t}\right] + fo \qquad (1)$$

where A, t and f_o are constants. We obtain an R^2 value of 0.998 using this model, implying a good fit to the data.

With knowledge of the flux of neutrons between 1 and 6 MeV, we convert the data in Figure 1 into a distribution of neutron charge vs. probability. We approximate that, for every 1 MeV of neutron energy, a strike will incur approximately 20 fC of charge [2] and that the charge increases linearly with neutron energy. We assume that the probability of charge below 20 fC is zero since a neutron strike depositing less than 20 fC of charge does not disrupt circuits in any appreciable way. The probability of charge above 120 fC (6 MeV) is also assumed to be zero since the flux of neutrons at this energy level is at least 10X less than that at the 1 MeV level and thus more than 10X less probable.

We note that although our formulation is somewhat approximate, more accurate relationships between energy, charge, and probability models are not readily available or easily demonstrable. Our estimates are based on data taken from the relevant literature [2,5,7,12,16].

Since the neutron flux is a known quantity, we convert the relative fluxes into relative probabilities using the curve fit to the data in Figure 1. Next, by bounding the charges at 1 MeV (20 fC) on the lower end and 6 MeV (120 fC) at the upper end, we can integrate this curve to find the total area, or probability, which we then normalize to a value of 1. Thus, our total charge probability lies in the range of 20-120 fC, as shown in Figure 2.

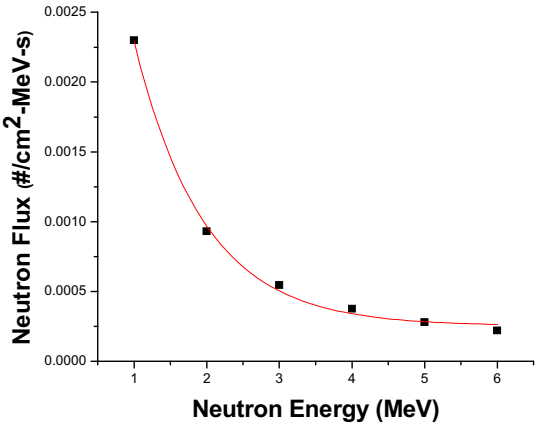

Figure 1: Flux as a function of neutron energy.

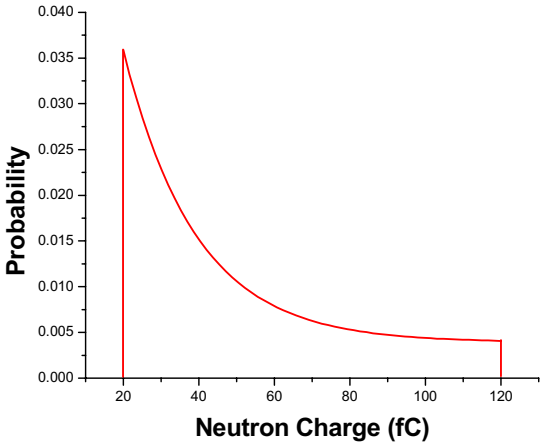

Figure 2: Probability distribution of charge deposition.

We simulate a neutron strike by an exponentially pulsed current source with a fast rise time and slower decay time [17]. Varying the amount of pulsed current, we can dictate the exact amount of charge injected. In general, our sweep of injected charge is in steps of 1.17 fC.

3. Circuit techniques

3.1 Preliminary analysis

To increase the nodal Q_{crit} required for a soft error to occur, we analyze several different protection schemes: 1) inserting a cross-coupled pair (CCP) on the node, 2) increasing the width of the gate driving the victim node, 3) use of low-V_{th} (LVT) devices, and 4) a combination of these methods. As a precursor to our primary analysis, we used the above techniques on a two-stage 2-input NAND circuit to determine their efficacy in an isolated topology.

This preliminary test was set up as shown in Figure 3. A 2-input NAND gate has one of its inputs tied to V_{DD} and the other to a DC input. The output of this NAND gate serves as the controlling input to the second, equally sized, NAND gate. Charge is injected on the node connecting the two NAND gates and the output voltage is observed. We assume a soft error occurs when the output height reaches a value of $V_{DD}/2$, or 0.6 V in our case.

We found that, in general, with increasing charge on the input the output voltage height rises slowly below 0.2 V and above 1.0 V, but has a steep slope in the 0.2-1.0V range. It was seen that once the critical charge threshold is reached, failure happens rapidly and with just a very small increment in injected charge (in the 5 fC range). Adding a protection scheme reduces the steepness of this slope, which is where potential gains in Q_{crit} are made.

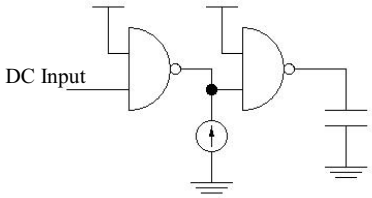

Figure 3: 2-input NAND structure used in the preliminary analysis.

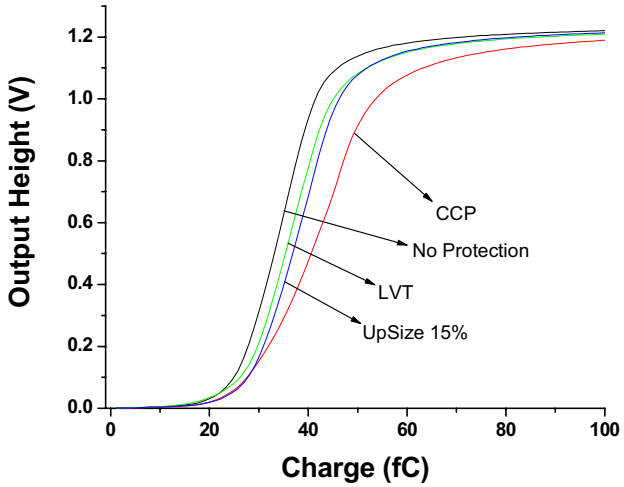

Figure 4: Output height vs. charge for various protection methods.

Figure 4 shows this phenomenon for the above mentioned cases of 1) no protection, 2) CCP protection, 3) up-sizing (increasing width by 15%), and 4) using LVT devices for both NANDs.

Figure 4 shows that for a single 2-input NAND gate with no protection, LVT and up-sizing, the critical charge lies in the 35 fC range. The two schemes of using LVT and upsizing the device width by 15% have modest effects in raising Q_{crit} of the gate by about 5-10%. Using the CCP protection method, we are able to raise the Q_{crit} by approximately 25%.

Figure 5 shows the power and delay tradeoffs for the various protections schemes. From this initial analysis, we draw some important conclusions regarding the various proposed protection schemes. First, we note that all three of the protection schemes trade off power for improved robustness, as expected. Next, increasing device width or using LVT gates is better in terms of the delay penalties. The CCP scheme has a substantial increase in delay but is the only scheme that demonstrates substantial Q_{crit}

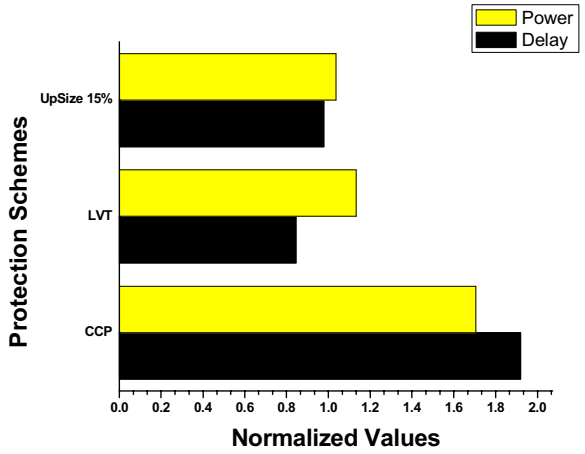

Figure 5: Power and delay tradeoffs, normalized to the unprotected case, for different protection schemes using two 2-input NAND gates.

gains; the other two provide relatively minor Q_{crit} gains. We can decrease the size of the CCP to decrease the delay (since there will be less contention at the node) but this also decreases the achieved gains in Q_{crit}.

From this preliminary analysis, we make a supposition that a better protection scheme would rely on a combination of these three schemes. In such a hybrid scheme, an optimally-sized CCP would contribute to raising the Q_{crit} substantially. The use of LVT gates and device width upsizing could additionally raise Q_{crit}, but would primarily help by ameliorating the delay penalty of the CCP. Note that device width redistribution and V_{th} assignment are known power-delay tradeoff optimizations implying that design automation solutions exist to leverage these techniques for improving circuit robustness. Similarly, the insertion of CCPs is not disruptive to current design practices and would not greatly increase design complexity.

Based on these initial protection schemes, we now describe our examinations of tapered and non-tapered NAND chains.

3.2 Analysis of 10-stage 2-input NAND chain

The primary analysis is performed on two separate 10-stage 2-input NAND chains designed in industrial 0.13µm technology as a design vehicle as shown in Figure 6. The first chain is tapered-up in size to drive a specific output load while the second chain has fixed size NAND gates throughout. The current injection node is swept across all 10 internal node locations since a neutron strike can occur with equal probability on any location of the NAND chain.

Additionally, we simulated both a positive and negative injected current pulse at both DC high and low inputs to take into account all neutron strike and circuit state possibilities. We now go over the two NAND chain approaches, starting with the tapered-up NAND chain.

The tapered-up NAND chain was designed such that the device widths of the gates would increase gradually to drive a 20 fF load capacitance at the output. We use a taper factor of ~1.2 to limit the transistor widths of the final gate to 4 µm.

Before investigating the combined protection schemes as mentioned near the end of the previous section, we simulated the tapered-up NAND chain using only one protection scheme at a time to inspect the behavior of a logic chain, rather than just two NAND gates.

At this point we incorporate the probability model developed in Section 2. At each current injection point, we sweep the injected charge from 20 fC to 120 fC and measure the output height at the output node (OUT). Once we determine whether or not a specific charge at a given internal node location causes a failure, we determine the probability of a strike having that amount of charge

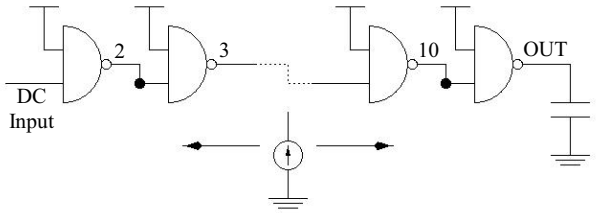

Figure 6: 10-stage 2-input NAND chain used in primary analysis.

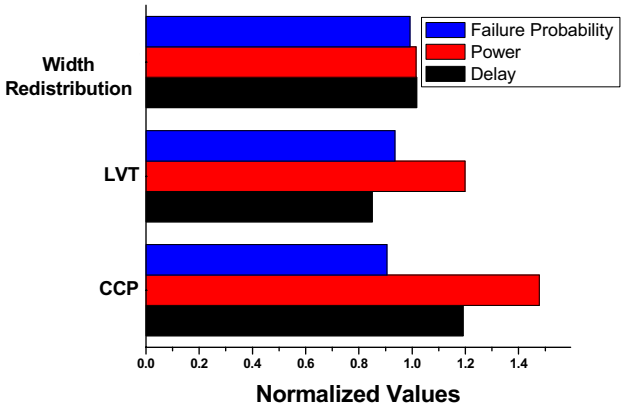

Figure 7: Normalized power, delay and probability for different protection schemes, for a tapered-up NAND chain.

and being in that node location (with all locations having equal likelihoods of being struck). Figure 7 shows the delay, power, and *total failure probability* for the protection schemes of CCP, LVT gates, and device width reallocation (the victim node gate width is upsized by 20%), normalized to the nominal (no-protection) case. For the CCP case, we insert three CCPs on the later stages of the NAND chain in an alternating fashion such that locations 5, 7, 9 and OUT have CCPs. The reason for choosing these locations is explained below. Note that the total failure probability signifies the probability the circuit will fail *given* the occurrence of a neutron strike. The inverse of total failure probability can be used to reflect mean time to failure; for example a 33% reduction in total failure probability implies a 50% longer time between soft errors.

We see from Figure 7 that the total failure probability does not decrease significantly, even in the CCP case. This can be explained by the nature of the 10-stage NAND chain. Figure 8 plots the probability of failure for each of the ten node positions.

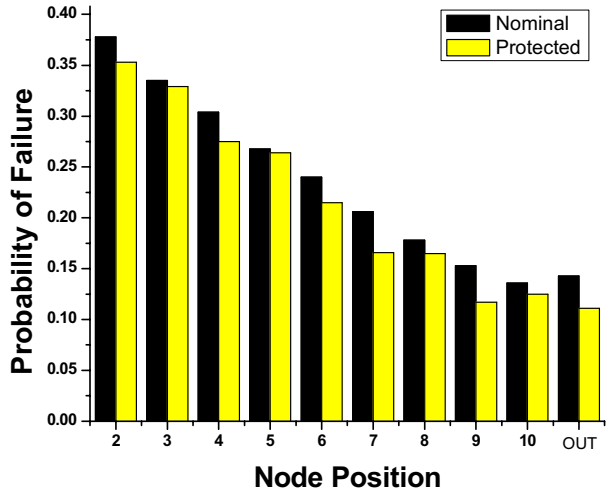

Figure 8: Probability of failure for each node position for a tapered-up NAND chain.

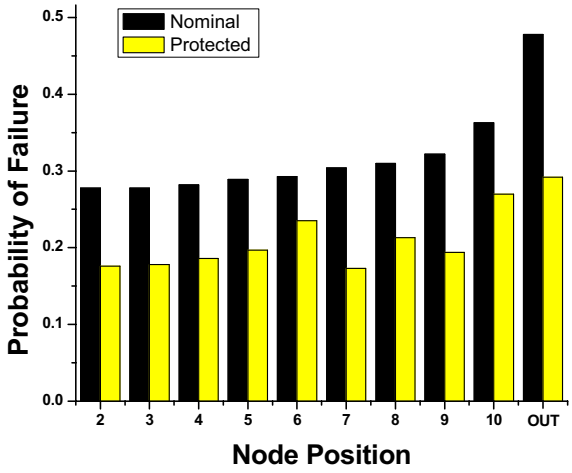

Figure 9: Probability of failure for each node position for a fixed size NAND chain.

We see that the earlier nodes are more likely to fail given a strike, primarily due to the earlier stages having smaller widths. The first few stages are 3-4X smaller in width than the last few stages which make them more susceptible to failure due to having a smaller Q_{crit}.

Additionally, placing CCPs at the beginning of the NAND chain will help in lowering the probability of failure at two or three crucial nodes in the beginning, but they will do nothing to help attenuate glitches at later stages. For example, placing a CCP at Nodes 2, 3 and 4 acts to increase the Q_{crit} at those nodes. However, since a neutron has equal probability of landing on any of the nodes, the remaining 7 nodes, or 70% of the circuit, are still unprotected. Although the absolute failure probabilities at each individual node of the remaining seven are smaller, the cumulative probability of failure is greater than the savings achieved from using the CCPs at the three nodes in the beginning. Therefore, the placement of CCPs at the beginning of the chain is undesirable. For these reasons, a single protection scheme is not sufficient to provide a significant improvement in Q_{crit}.

Using the same approach developed for the tapered-up NAND chain, we run simulations on a non-tapered NAND chain. The output capacitance was reduced and all gates were equally sized. Figure 9 shows the probability of failure for each node location, for the fixed size chain. Unlike for the tapered-up chain, the probability of each node failing increases as the node is traversed. Since each NAND gate is fixed in size, the gates in the front will not intrinsically be more error-prone since these gates will still have same sized capacitances as the rest of the gates. However, later nodes have greater probability of failure since they capture the cumulative probability of the previous nodes failing. In this case, using CCPs at the end of the chain is beneficial.

3.3 Combined protection scheme

We now focus on using a combination of the proposed protection schemes. We use four CCPs at node locations 5, 7, 9, and OUT for the tapered-up chain and three CCPs at node locations 7, 9, and OUT for the fixed size chain. To recover the

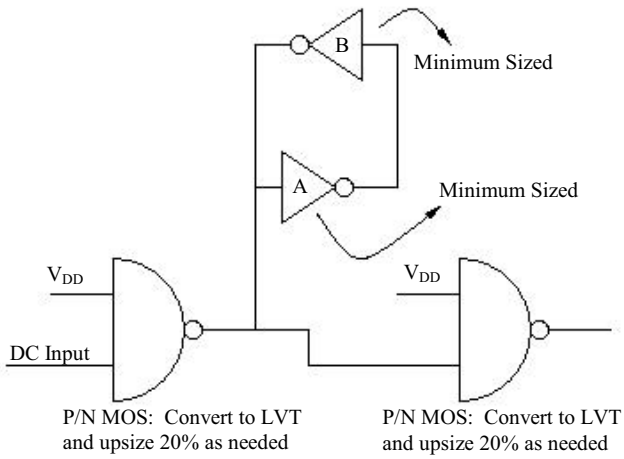

Figure 10: Cross-coupled pair, LVT gate and sizing scheme for a generic logic chain.

delay lost in adding CCPs, we convert the gate following a CCP to an all LVT gate as needed. This reduces the delay penalty as well as adds some measure of robustness as was shown in Figure 4. Next, we readjust the device widths in each gate shown in Figure 10, if necessary. In a CCP, inverter A contributes to the improved robustness only in a second-order fashion by increasing capacitive load on the victim node. Therefore, we make this inverter minimum sized as it does not need appreciable drive strength. For inverter B, we also make this gate minimum sized to reduce contention between it and the node it is protecting. Additionally, LVT gates are used to reduce the delay penalty as much as possible.

3.4 Skewed protection scheme

We simulated both positive and negative injected current pulses at both DC high and low inputs to consider all neutron strike and circuit state possibilities. However, it can be shown that internal nodes have a tendency to stay in one state or another with high probabilities, even when high or low inputs are equally likely.

For 50% input state probabilities, we analyzed ten ISCAS85 benchmark circuits [18] to determine the fraction of internal nodes having state probabilities of less than 20% or greater than 80%. Table shows these results. Summing these probabilities,

Table 1: Fraction of highly skewed internal nodes in ISCAS85 benchmark circuits.

Circuit	Fraction in Skewed State
c432	0.4068
c499	0.2601
c880	0.283
c1355	0.2311
c1908	0.2986
c2670	0.1675
c3540	0.4468
c5315	0.2385
c6288	0.2312
c7552	0.1204

we find that on average 27% of all internal nodes are in one state ≥80% of the time. We apply this result to our circuit technique by predicating that one polarity of the neutron strike will cause less damage than another. This is because if, for example, one node is in the logic high state for 80% of the time, then a neutron strike generating a positive voltage on that node will typically not cause a logical upset. We incorporate this by assuming that one state is vulnerable 80% of the time and the other only 20% of the time. This is a realistic and reasonable assumption since most logic gates (such as NANDs and NORs) prefer one output state over the other.

4. Results

Using the circuit technique developed in Section 3, we simulated a neutron strike on the ISCAS85 c17 [18] circuit to determine the efficacy of our method. C17 is used despite its small size since very computationally expensive SPICE simulations must be used to generate the results. In particular, Monte Carlo application of current pulses (with varying magnitudes) to all circuit nodes must be performed. We achieve an 18% increase in total circuit robustness with a worst-case delay penalty of just over 20%. Due to the insertion of cross-coupled pairs, as well as the use of LVT gates and device width upsizing for performance and protection as needed, there is a 45% total power penalty (62% penalty in dynamic power, less than 10% in static power) relative to the nominal unprotected case. Incorporating skew of internal nodes, we find an increase in robustness for the ISCAS'85 c17 circuit [18] of almost 20%.

We note that power penalties in the 45% range may seem considerably high, however, redundant schemes [10] can double or triple the area and add extra control logic thereby increasing the power by over 200%. Finally, we remark that the inverse of the total failure probability can be used to reflect mean time to failure, where with a 20% increase in robustness (or, conversely, a 20% decrease in errors), the time between soft errors is increased by 25%. Figure 11 summarizes these results, normalized to the nominal, unprotected case.

5. Conclusions

A charge-based probability model was developed and applied to neutron strikes on a circuit. From this model, we were able to find the failure probabilities of a circuit given a neutron strike. We then introduced a circuit technique that places a cross-coupled pair on specific nodes in a path to increase the nodal Q_{crit} and thereby increasing circuit robustness. This was then combined with device width readjustment and selective use of low V_{th} gates. Additionally, we found that often internal nodes are greatly skewed towards one state or the other. We combined this with our techniques to further enhance the protection scheme.

Applying these techniques to a small circuit, we found that robustness increased by about 20%, considering skewed internal nodes. This is equivalent to an increase of ~25% in mean time between soft errors. The newly proposed techniques have much lower power penalties compared to traditional redundancy-based approaches and thus provide a more fine-grained SEU reduction technique. In particular, we have substantially reduced the area overhead compared to some of the mentioned techniques as well as greatly decreasing design complexity.

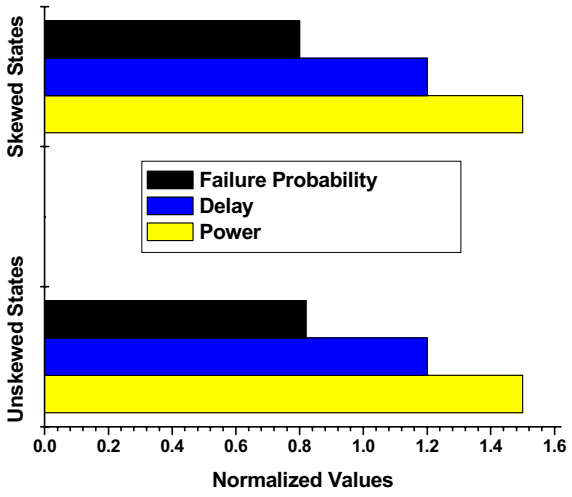

Figure 11: ISCAS85 c17 circuit for both skewed and unskewed internal node states.

6. Acknowledgement

The authors would like to thank Dongwoo Lee and Robert Baumann for their insightful comments and assistance in this work. Funding for this work was provided in part by the NSF.

7. References

[1] P. Hazucha, et al. "Neutron Soft Error Rate Measurements in a 90-nm CMOS Process and Scaling Trends in SRAM from 0.25-µm to 90-nm Generation". *International Electron Devices Meeting,* 2003.

[2] D. Mavis and P. Eaton. "Soft Error Rate Mitigation Techniques for Modern Microcircuits". *International Reliability Physics Symposium,* 2002.

[3] N. Seifert, D. Moyer, N. Leland and R. Hokinson. "Historical Trend in Alpha-Particle induced Soft Error Rates of the Alpha Microprocessor". *International Reliability Physics Symposium,* 2001.

[4] T. Karnik, et al. "Scaling Trends of Cosmic Rays Induced Soft Errors in static latches beyond 0.18µ". *Symposium on VLSI Circuits Digest of Technical Papers,* 2001.

[5] R. Baumann. "Soft Errors in Advanced Semiconductor Devices—Part I: The Three Radiation Sources". *IEEE Transactions on Device and Materials Reliability*, March 2001.

[6] L. Wissel, et al. "Managing Soft Errors in ASICs". *Custom Integrated Circuits Conference*, 2002.

[7] J. F. Ziegler. "Terrestrial Cosmic Rays". *IBM Journal of Research and Development*, Jan 1996.

[8] L. Li, N. Vijaykrishnan, M. Kandemir and M.J. Irwin. "Adaptive Error Protection for Energy Efficiency". *International Conference on Computer Aided Design*, 2003.

[9] T. Karnik, et al. "Selective Node Engineering for Chip-Level Soft Error Rate Improvement". *Symposium on VLSI Circuits Digest of Technical Papers*, 2002.

[10] L. Anghel., D. Alexandrescu and M. Nicolaidis. "Evaluation of a Soft Error Tolerance Technique Based on Time and/or Space Redundancy". *13th Symposium on Integrated Circuits and Systems Design,* 2000.

[11] V. Degalahal, et al. "The Effect of Threshold Voltages on the Soft Error Rate". *International Symposium on Quality Electron Design*, 2004.

[12] J. F. Ziegler, et al. "IBM Experiments in Soft Fails in Computer Electronics (1978-1994)". *IBM Journal of Research and Development*, Jan 1996.

[13] P. Lidén, P. Dahlgren, R. Johansson and J. Karlsson. "On Latching Probability of Particle Induced Transients in Combinational Networks". *24th Symposium on Fault-Tolerant Computing*, 1994.

[14] P. Shivakumar, et al. "Modeling the Effect of Technology Trends on the Soft Error Rate of Combinational Logic". *International Conference on Dependable Systems and Networks*, 2002.

[15] JEDEC JESD89. "Measurement and Reporting of Alpha Particles and Terrestrial Cosmic Ray-Induced Soft Errors in Semiconductor Devices". Joint Electron Device Engineering Council, Solid State Technology Association, Aug 2001.

[16] G. R. Srinivasan, H. K. Tang and P. C. Murley. "Parameter-Free, Predictive Modeling of Single Even Upsets Due to Protons, Neutrons, and Pions in Terrestrial Cosmic Rays". *IEEE Transactions on Nuclear Science,* Dec 1994.

[17] L. B. Freeman. "Critical Charge Calculations for a Bipolar SRAM Array". *IBM Journal of Research and Development,* Jan 1996.

[18] F. Brglez and H. Fujiwara. "A Neutral Netlist of 10 Combinational Benchmark Circuits". *International Symposium on Circuits and Systems,* 1985.