Design optimization for single-event upset robustness using simultaneous dual-VDD and sizing techniques

Mihir R. Choudhury Quming Zhou Kartik Mohanram
Department of Electrical and Computer Engineering
Rice University, Houston, TX 77005
{quming, mihir, kmram}@rice.edu

Abstract

An optimization algorithm for the design of combinational circuits that are robust to single-event upsets (SEUs) is described. A simple, highly accurate model for the SEU robustness of a logic gate is developed. This model – in posynomial form – is integrated with performance and power constraints into an optimization framework based on geometric programming for design space exploration. Simulation results for design optimization using simultaneous dual-$V_{\rm DD}$ and gate sizing techniques for the 70 nm process technology demonstrate the tradeoffs that can be achieved with this approach.

1. Introduction

Technology trends, including smaller feature sizes, lower voltage levels, higher operating frequencies, and reduced logic depth are projected to cause an increase in the soft error failure rate in sub-100 nm integrated circuits [1–3]. Soft errors occur as a result of single-event upsets (SEUs) caused by high-energy neutron or alpha particle strikes in integrated circuits. Although soft errors cause no permanent damage, they can severely limit the reliability of electronic systems.

Although several design as well as error detection and correction solutions for reliability to soft errors in memories, flip-flops, and latches have been proposed in literature (e.g., [4, 5]), there are relatively few techniques that are cost-effective for use in multi-level combinational logic circuits. The applicability of these techniques to combinational circuits is limited owing to (i) the irregular multi-level structure of combinational circuits that leads to very high design overhead and (ii) the high cost of error detection, correction, and recovery required to support such techniques.

Two circuit design techniques that have been used successfully to harden memories (both static and dynamic) to SEUs are the use of (i) high supply voltage $V_{\rm DD}$ and (ii) large cells with high drive strength transistors. By increasing $V_{\rm DD}$, the charge stored in the memory cell is increased. By increasing the drive strength of the transistors, the charge deposited due to a particle strike is dissipated faster. Both techniques increase SEU robustness by increasing the particle energy threshold required to cause SEUs. Such design for SEU robustness techniques are very attractive (i) since they do not incur any overhead for error detection, correction, and recovery and (ii) since they can be used to complement other techniques that enhance SEU robustness such as the use of silicon-on-insulator substrates, error detection and correction hardware, etc. Recently, transistor and gate sizing has been shown to be a promising solution for the design of logic circuits that are robust to SEUs (e.g., [6, 7]). These state-of-the-art sizing techniques for SEU robustness rely on rank-and-optimize or sensitivity-driven heuristics to reduce the soft error failure rate. Such heuristics usually optimize the cost function in a local neighborhood, and do not guarantee convergence to the optimum solution. Furthermore, they usually impose a delay penalty on the hardened design – this may be unacceptable, especially for high-performance designs.

Better design space exploration and control over design overhead can be achieved through the use of global optimization approaches that allow simultaneous tradeoffs between traditional objectives of area-delay-power and robustness to SEUs. The use of circuit optimization techniques based on geometric programming (GP) and generalized GP (GGP) can be traced back to the TILOS paper on optimum sizing for delay [8]. Such approaches have since been used with great success on problems in transistor and gate sizing, multi-$V_{\rm DD}$, and multi-$V_{\rm T}$ optimization in literature [9]. In this paper, a simple, highly accurate, and comprehensive model for the SEU robustness of a logic gate is developed. The model integrates factors such as transistor size W, supply voltage $V_{\rm DD}$, and threshold voltage $V_{\rm T}$ that are central to post-mapping transformations such as gate resizing, fanout optimization, resynthesis and remapping, etc. [10] ensuring compatibility with global optimization flows. This model is integrated with power and performance constraints into a global optimization framework based on GP for design optimization for robustness to SEUs.

To the best of our knowledge, this is the first work that integrates a SEU robustness model into GP-based global design optimization flows based on simultaneous dual-$V_{\rm DD}$ and gate sizing techniques. Such SEU robustness driven design techniques will lessen the investment in SEU analysis and hardening strategies in the latter stages of the design process. They are advantageous over rank-and-optimize and sensitivity-driven heuristics that, although effective, may not provide the best design alternatives to choose from. Simulation results for several logic circuits in the 70 nm process technology are presented to illustrate the tradeoffs that can be explored with this approach.

The paper is organized as follows. In Sec. 2, we describe the standard GP-based optimization algorithm for design using simultaneous sizing and dual-$V_{\rm DD}$ techniques. In Sec. 3, we describe the proposed model and derive closed-form size-for-robustness expressions for use in GP-based design optimization. In Sec. 4, we present and discuss simulation results. Section 5 is a conclusion.

ICCAD 2006, November 5–9, 2006, San Jose, California, USA.

2. Circuit optimization background

The use of circuit optimization techniques based on GP and GGP can be traced back to the TILOS paper on optimum sizing for delay [8]. Such approaches have since been used with success on problems in transistor and gate sizing, multi-$V_{\rm DD}$, and multi-$V_{\rm T}$ optimization in literature [9].

GP for minimum power: It is possible to formulate the problem of design optimization for minimum power using gate sizing and dual-$V_{\rm DD}$ techniques – subject to performance constraints on delay $T_{\rm spec}$ at the primary outputs – as follows. We term this algorithm $\mathcal{PD}$ for power-delay optimization.

$$\mathcal{PD}: \text{Minimize design power } \sum_{i=1}^{n} P_{{\rm dyn},i} + P_{{\rm stat},i} \qquad (1)$$

$$\text{where } \left\{ \begin{array}{c} P_{{\rm dyn},i} = \alpha_{{\rm sw},i}\, C_{{\rm out},i} V_{{\rm DD},i}^2 \\ P_{{\rm stat},i} = W_i\, I_{{\rm leak},i} V_{{\rm DD},i} \end{array} \right.$$

Subject to the following constraints

$$\left. \begin{array}{l} W_{{\rm min},i} \le W_i \le W_{\rm max} \\ V_{{\rm DD,min}} \le V_{{\rm DD},i} \le V_{\rm DD,max} \end{array} \right\} \quad i = 1, 2, ..., n$$

$$\begin{array}{ll} V_{{\rm DD},i} \ge V_{{\rm DD},j} & j \in \text{fanout}(i) \text{ and } i = 1, 2, ..., n \\ \delta_i + T_j \le T_i & j \in \text{fanin}(i) \text{ and } i = 1, 2, ..., n \\ T_i \le T_{\rm spec} & i \in \text{primary outputs} \end{array}$$

where

1. $P_{{\rm dyn},i}$ and $P_{{\rm stat},i}$ are the dynamic and static power components of the $i^{\rm th}$ gate,
2. $\alpha_{{\rm sw},i}$ is the switching activity of the $i^{\rm th}$ gate,
3. $C_{{\rm out},i}$ is the total capacitance (load and parasitic) at the $i^{\rm th}$ gate given by
$$C_{{\rm out},i} = \left(W_i C_{{\rm int},i} + \sum\nolimits_{j \in \text{fanout}(i)} C_{{\rm in},j} W_j + C_{\rm L} \right),$$
4. W_i is the size of the $i^{\rm th}$ gate,
5. $C_{{\rm int},i}$ is the output capacitance of the unit-scaled gate,
6. $C_{{\rm in},j}$ is the input capacitance of the unit-scaled gate,
7. $C_{\rm L}$ is the load capacitance if the $i^{\rm th}$ gate is a primary output and 0 otherwise,
8. $V_{{\rm DD},i}$ is the supply voltage of the $i^{\rm th}$ gate,
9. $I_{{\rm leak},i}$ is the leakage current of the unit-scaled gate,
10. δ_i is the delay of the $i^{\rm th}$ gate given by
$$\delta_i = R_i C_{{\rm int},i} + (R_i / W_i) \left(\sum\nolimits_{j \in \text{fanout}(i)} C_{{\rm in},j} W_j + C_{\rm L} \right),$$
11. T_i is the arrival time at the output of the $i^{\rm th}$ gate,
12. $T_{\rm spec}$ is a specified circuit delay, and
13. R_i is the resistance of the unit-scaled delay-calibrated gate.

Here, W_i and $V_{{\rm DD},i}$ are the variables of algorithm $\mathcal{PD}$. The arrival times T_i are intermediate variables used to express delay constraints. The GP formulation requires that dynamic power, static power, and delay be expressible as posynomial functions in the variables of the GP [9]. Although we refer to transistor sizes and use W_i in the formulation, we limit ourselves to symmetric gate sizing in this paper. Thus, scaling a single transistor through W_i is equivalent to scaling all transistors (nMOS and pMOS) in the gate by the same ratio. Also, the solution of the above GP formulation results in the supply $V_{{\rm DD},i}$ assuming continuous values over the range. Our implementation of $\mathcal{PD}$ uses standard branch-and-bound techniques from literature to solve this GP problem to obtain discrete values for $V_{{\rm DD},i}$ [11, 12].

3. Closed-form circuit-level SEU model

In this section, we use linear gate models to derive closed-form expressions for the waveform of the SEU-induced transient. We extend this model to derive closed-form expressions for the gate size and supply required to limit the magnitude of the SEU-induced transient to less than a pre-specified value at the site of the strike, say, $\eta V_{\rm DD}$. These closed-form expressions yield posynomial SEU robustness constraints as functions of W_i and $V_{{\rm DD},i}$ that are the variables of the GP algorithm $\mathcal{PD}$. The constraints are integrated into algorithm $\mathcal{PD}$ to optimize the design globally for SEU robustness (algorithm $\mathcal{PDS}$, Sec. 3.3).

Consider a gate driving one or more identical gates in its transitive fanout to two (or more) levels of logic that approximate loading conditions. A SEU-induced transient to logic 1 (logic 0) refers to the case when the steady-state logic value at the output of the gate is logic 0 (logic 1) in the fault-free case and a SEU generates a positive (negative) transition to logic 1 (logic 0). The worst-case transient occurs when the site for the particle strike is the gate output, since transients at internal nodes are reduced in severity before they propagate to the output of the gate. Without loss of generality, the analysis in this section discusses only $0 \rightarrow 1$ SEU-induced transients; $1 \rightarrow 0$ SEU-induced transients can be analyzed in a similar manner by symmetry.

The charge deposition due to a particle strike at the output of the gate is modeled by a parameterized, double-exponential current pulse $I_{\rm in}(t)$ at the output [13]:

$$I_{\rm in}(t) = \frac{Q}{(\tau_\alpha - \tau_\beta)} \left({\rm e}^{-t/\tau_\alpha} - {\rm e}^{-t/\tau_\beta} \right) \qquad (2)$$

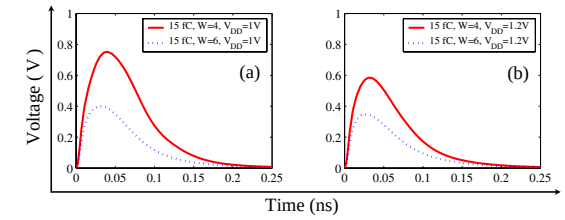
where Q is the charge (positive or negative) deposited as a result of the particle strike, τ_α is the collection time-constant of the junction, and τ_β is the ion-track establishment time-constant. τ_α and τ_β are constants that depend on process-related factors. Note that $\lim_{t\rightarrow\infty} \int_0^t I_{\rm in}(x){\rm d}x$ equals Q for conservation of charge. Note also that this model can be replaced by the weighted, single-pole current pulse model for SEUs [14] without loss of generality.

3.1 Motivation

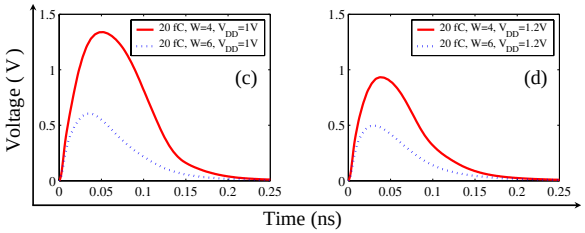
We begin by presenting the results of SPICE simulations to motivate the advantages of using gates with larger size and supply voltage to achieve SEU robustness. Consider a fanout-of-2 chain of 2-input nand gates. The output response of the nand gate to SEUs that deposit 15 fC and 20 fC of charge to produce $0 \rightarrow 1$ transients – for combinations of gate size and supply voltage – is presented in Fig. 1. τ_α and τ_β were 50 ps and 1 ps for the simulations.

Sub-figures 1(a) and 1(b) correspond to 15 fC charge deposition. Sub-figures 1(c) and 1(d) correspond to 20 fC charge deposition. In each sub-figure, it is clear that the magnitude and duration of the SEU-induced transient diminishes as the gate size increases from 4 to 6. Further, for a particular gate size, use of a high supply voltage of 1.2 V reduces the magnitude and duration of the SEU-induced transient at the output of the gate. This is observed by comparing the corresponding waveforms between sub-figures 1(a) and 1(b), and sub-figures 1(c) and 1(d) respectively. Both sizing and high-$V_{\rm DD}$ increase the drain current $I_{\rm D}$ through the nMOS transistors and thus increase the robustness of the gate to SEUs.

Although it is possible to use SPICE simulations to determine optimum size and $V_{\rm DD}$ assignments to a gate in isolation, such an approach is not only computationally expensive but also sub-optimal. This is because sizing as well as $V_{\rm DD}$ assignments affect both the fanin and the fanout from delay, power, and SEU robustness stand-points. Whereas the standard posynomial models for delay and power account for these effects and have been widely



1(a) and 1(b): 15 fC particle strike



1(c) and 1(d): 20 fC particle strike

Figure 1: SPICE simulations of particle strikes of 15 fC and 20 fC charge at a nand gate, whose size and $V_{\rm DD}$ are varied.

studied in literature, analogous models for SEU robustness have never been explored. In the next sub-section, a linear modeling approach to evaluate the SEU robustness as a function of charge, gate size, supply voltage, and load capacitance is developed. Such a model finds direct application in traditional GP-based design optimization flows described in Sec. 2.

3.2 Closed-form circuit-level SEU model

The differential equation of the SEU-induced transient at the output of the gate is given by

$$C_{\rm out}\frac{{\rm d}V_{\rm out}}{{\rm d}t} = I_{\rm in}(t) - I_{\rm D}(t) \tag{3}$$

where $C_{\rm out}$ is the total capacitance (load and parasitic) at the output node, $I_{\rm in}(t)$ is the model for the SEU, and $I_{\rm D}(t)$ is the current in the transistor network restoring the output to its fault-free value.

Linear gate modeling techniques are widely used in delay and power modeling, and can be used to derive closed-form expressions for the waveform and peak of the SEU-induced voltage transients. This approach is effective and appropriate when the primary goal is to determine the minimum gate size that reduces the magnitude of the SEU-induced transient at the output of the gate to less than a pre-specified value $\eta V_{\rm DD}$. Note that the nMOS transistor network of the gate is in the linear region of operation as long as $\eta V_{\rm DD}$ is less than $(V_{\rm DD} - V_{\rm T})$, where $V_{\rm GS}$ is assumed to be $V_{\rm DD}$. Hence the current $I_{\rm D}(t)$ can be approximated by $V_{\rm out}/R_{\rm SEU}$, where $R_{\rm SEU}$ is the equivalent resistance of the nMOS transistor network in the linear region of operation. Based upon this, the solution to Eqn. (3) is given by

$$V_{\rm out}(t) = \frac{QR_{\rm SEU}}{\tau_\alpha - R_{\rm SEU}C_{\rm out}}\left({\rm e}^{-t/\tau_\alpha} - {\rm e}^{-t/R_{\rm SEU}C_{\rm out}}\right). \tag{4}$$

Note that τ_β adds an exponential term to Eqn. (4) that is negligible, and is ignored from here on for simplicity. The peak value of the SEU-induced voltage transient $V_{\rm max}$ is obtained by differentiating the above solution, solving for $t_{\rm max}$, and by resubstitution in Eqn. (4). Let $\mu \equiv \tau_\alpha/R_{\rm SEU}C_{\rm out}$ for ease of notation. Then,

$$V_{\rm max} = \frac{QR_{\rm SEU}}{\tau_\alpha}\mu^{\frac{1}{1-\mu}}. \tag{5}$$

Parameterized model for $V_{\rm max}$: Parameterization of this model is achieved based upon simplification of the $\mu^{\frac{1}{1-\mu}}$ term and an equivalent linear model for the resistance $R_{\rm SEU}$ of the nMOS network. The function $\mu^{\frac{1}{1-\mu}}$ is well approximated by $k_\mu\mu^\gamma$ with a goodness of fit value of 0.98 when $\gamma = 0.4$ and $0.25 \le \mu \le 5$. This range of values for μ corresponds to a gate delay of 10–200 ps, which is valid for current process technologies. In the linear regime, the resistance $R_{\rm SEU}$ of the nMOS network is given by

$$R_{\rm SEU} = \frac{k_{\rm nMOS}}{(V_{\rm DD} - V_{\rm T})W} \tag{6}$$

where W is the gate size, $V_{\rm T}$ is the threshold voltage, and $k_{\rm nMOS}$ is a constant. Resubstitution of these expressions for $\mu^{\frac{1}{1-\mu}}$ and $R_{\rm SEU}$ in Eqn. (5) for the peak value of the SEU-induced voltage transient at the output gives

$$V_{\rm max} = k_\mu Q\tau_\alpha^{\gamma-1}\left(\frac{k_{\rm nMOS}}{(V_{\rm DD} - V_{\rm T})W}\right)^{1-\gamma} C_{\rm out}^{-\gamma}. \tag{7}$$

Conditions for SEU robustness: Without loss of generality, let $V_{\rm max} \le \eta V_{\rm DD}$ for the SEU-induced transient to be dissipated locally without sufficient magnitude and duration to propagate through the fanout gates. Upon simplification of the expression (7) using this inequality, the condition for SEU robustness is given by

$$kW^{-1}(V_{\rm DD} - V_{\rm T})^{-1}\left(\frac{Q}{V_{\rm DD}}\right)^{1+\beta} C_{\rm out}^{-\beta} \le 1 \tag{8}$$

where $\beta = \frac{\gamma}{1-\gamma}$ and $k = k_{\rm nMOS}\tau_\alpha^{-1}(k_\mu/\eta)^{1+\beta}$. Based upon this, the constraints for SEU robustness that can be integrated into algorithm $\mathcal{PD}$ described in Sec. 2 will be given by Eqn. (8) for all (or a subset) of the gates in the design. In order to increase the accuracy of the expressions in Eqn. (8) in comparison to SPICE simulations, a three-parameter model for Eqn. (8) given by

$$kW^{-1}V_{\rm DD}^{-1}\left(\frac{Q}{V_{\rm DD}}\right)^{1+\beta_0} C_{\rm out}^{-\beta_1} \le 1 \tag{9}$$

is derived using SPICE-based calibration runs. Note that $(V_{\rm DD} - V_{\rm T})^{-1}$ is approximated by $V_{\rm DD}^{-1}$ to the first order in Eqn. (9). The parameters k, β_0 and β_1 are obtained by data fitting the simulation results obtained from SPICE. The parameters β_0 and β_1 were found to lie in the interval $[0.5\,, 0.8]$ in our simulations. The main difference between the expressions in Eqn. (8) and Eqn. (9) is that the parameters in Eqn. (9) are further tuned by SPICE simulations to increase the accuracy, with a maximum error of 0.5 times the size of a unit-scaled gate (Sec. 4.1).

Compact robustness model: In order to determine the constraints for SEU robustness, it is necessary to use a robustness charge $Q_{\rm rob}$ as a figure-of-merit during design analysis and optimization. $Q_{\rm rob}$ could be the nominal or the maximum charge deposited by particle strikes for a process technology, and can be determined using actual measurements with test structures or by 3-dimensional device simulations. For a given $Q_{\rm rob}$, the optimum size W and supply voltage $V_{\rm DD}$ can be determined during design-space exploration using dual-$V_{\rm DD}$ and gate sizing. Thus, for a given $Q_{\rm rob}$, the compact model for SEU robustness is given by the following expression:

$$k'W^{-1}(V_{\rm DD})^{-2-\beta_0}(C_{\rm out})^{-\beta_1} \le 1 \tag{10}$$

where k' equals $k(Q_{\rm rob})^{1+\beta_0}$. This expression is used to derive robustness constraints for all (or a subset of) the gates in the design for incorporation into the $\mathcal{PD}$ algorithm, as described in Sec. 3.3.

3.3 Optimization for SEU robustness

The SEU robustness constraints for minimum gate size derived in the previous section can be incorporated into the $\mathcal{PD}$ optimization framework to obtain the power-delay-SEU ($\mathcal{PDS}$) optimization algorithm as follows. For every gate in the design, consider an additional constraint of the form

$$k_i' W_i^{-1} (V_{\mathrm{DD},i})^{-2-\beta_0} (C_{\mathrm{out},i})^{-\beta_1} \leq 1 \qquad i = 1, 2, ..., n \quad (11)$$

where k_i' is a constant for each type of gate (inverter, 2-input nand, etc.), $V_{\mathrm{DD},i}$ is the supply voltage of the i^{th} gate, and $C_{\mathrm{out},i}$ is the total capacitance at the output of the i^{th} gate. $C_{\mathrm{out},i}$ is given by

$$C_{\mathrm{out},i} = \left(W_i C_{\mathrm{int},i} + \sum\nolimits_{j \in \mathrm{fanout}(i)} C_{\mathrm{in},j} W_j + C_{\mathrm{L}} \right)$$

that is an affine function. Note that C_{L} is non-zero only for primary outputs and is included here without loss of generality. Since $\beta_1 > 0$, Eqn. (11) for SEU robustness *does not* have a posynomial form for direct integration into the GP-based algorithm $\mathcal{PD}$.

The expression for $C_{\mathrm{out},i}$ can be bounded using the well-known inequality that the arithmetic mean (AM) $\geq$ geometric mean (GM). Let m be the number of terms in $C_{\mathrm{out},i}$. For an internal gate, m is $|\mathrm{fanout}|+1$ and for a primary output, m is $|\mathrm{fanout}|+2$ since there is an additional load term C_{L}. Since $C_{\mathrm{out},i}$ has strictly positive terms, the AM $\geq$ GM inequality applied to $C_{\mathrm{out},i}$ gives

$$\frac{W_i C_{\mathrm{int},i} + \sum_{j \in \mathrm{fanout}(i)} C_{\mathrm{in},j} W_j + C_{\mathrm{L}}}{m} \geq \left(W_i C_{\mathrm{int},i} \cdot \prod\nolimits_{j \in \mathrm{fanout}(i)} C_{\mathrm{in},j} W_j \cdot C_{\mathrm{L}} \right)^{1/m} \quad (12)$$

Rearranging terms to recover $C_{\mathrm{out},i}$ from Eqn. (12),

$$C_{\mathrm{out},i} \geq \left(m \left(W_i C_{\mathrm{int},i} \cdot \prod\nolimits_{j \in \mathrm{fanout}(i)} C_{\mathrm{in},j} W_j \cdot C_{\mathrm{L}} \right)^{1/m} \equiv M_i \right)$$

Since $C_{\mathrm{out},i} \geq M_i$ and since k_i' and β_1 are positive,

$$k_i' W_i^{-1} (V_{\mathrm{DD},i})^{-2-\beta_0} C_{\mathrm{out},i}^{-\beta_1} \leq k_i' W_i^{-1} (V_{\mathrm{DD},i})^{-2-\beta_0} M_i^{-\beta_1} \quad (13)$$

and the SEU robustness conditions in Eqn. (11) can be rewritten in the form

$$k_i' W_i^{-1} (V_{\mathrm{DD},i})^{-2-\beta_0} C_{\mathrm{out},i}^{-\beta_1} \leq k_i' W_i^{-1} (V_{\mathrm{DD},i})^{-2-\beta_0} M_i^{-\beta_1} \leq 1 \quad (14)$$

Since M_i is a monomial, the AM $\geq$ GM transformation allows the original SEU robustness constraints in terms of $C_{\mathrm{out},i}$ to be re-expressed using the monomial M_i. The SEU robustness constraints derived from Eqn. (14) in the form $k_i' W_i^{-1} (V_{\mathrm{DD},i})^{-2-\beta_0} M_i^{-\beta_1} \leq 1$ are posynomial. If these constraints are satisfied, it follows from Eqn. (14) that the original SEU robustness constraints in Eqn. (11) are also satisfied. The posynomial constraints can be integrated into the basic $\mathcal{PD}$ algorithm, and we term this algorithm $\mathcal{PDS}$ for power-delay-SEU optimization.

An error is introduced by the AM $\geq$ GM inequality. The gap between the AM and the GM depends on the distribution of the data points in the inequality. In the SEU constraints, the data points are the W_i and W_j terms in $C_{\mathrm{out},i}$. Our simulations indicate that this approximation is highly accurate in practice since design optimization tends to size gates evenly across the design. As a result, the sizes W_i and W_j are clustered close enough to the average that over-optimization is minimal. The average over-approximation across all the benchmark circuits in our simulations was 2.5%.

4. Results

The geometric programming framework for circuit optimization was implemented using the optimization tool MOSEK [15]. The SPICE library for the 70 nm technology node was obtained from the Berkeley predictive technology model [16]. Twelve combinational benchmark circuits were chosen from the ISCAS85 and LGSynth91 suite [17]. We used $\tau_\alpha = 50$ ps and SEU robustness charges of 15 fC and 20 fC in all our simulations. We built a technology library that comprised inverters, and 2-input and 3-input nand and nor gates of different drive strengths for initial synthesis of the benchmarks. The optimization for SEU robustness was performed on these synthesized netlists. η was set to 0.5, so $V_{\max}$ ($V_{\min}$) was $0.5V_{\mathrm{DD}}$ for $0 \rightarrow 1$ ($1 \rightarrow 0$) SEU-induced transients.

4.1 Model validation

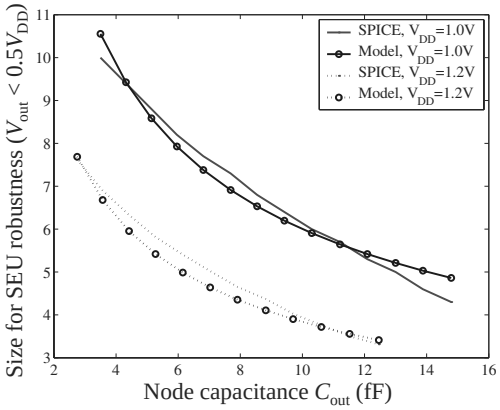


Figure 2: This figure illustrates the accuracy of the compact robustness model given by Eqn. (10) for a 2-input nand gate for a dual-V_{DD} technology. It is clear that the optimum size for SEU robustness is estimated to within 0.5 times the size of the unit-scaled gate over the range of circuit configurations.

Fig. 2 presents the results of simulations that were performed to validate the compact robustness model given by Eqn. (10). Simulations were performed on 2-input nand gates over a range of load capacitance and two supply voltages (1.0 V and 1.2 V). In all cases, the gate size required to limit the peak of the SEU-induced voltage transient to $0.5V_{\mathrm{DD}}$ was determined using the compact models as well as using SPICE simulations. The minimum and maximum load capacitances chosen for model validation include fanout-of-1 to fanout-of-4 circuits with gate sizes ranging from 2 to 10 units. The maximum error in size for SEU robustness determined using the model was 0.5 times the size of the unit-scaled 2-input nand gate. Similar results were observed for the other logic gates that were used for synthesis of the benchmarks.

4.2 Benchmark cu

Before presenting the results of optimization on all benchmarks, we present a case study of the power versus delay curves for the benchmark cu in 70 nm technology in Fig. 3. The results of optimization in both $\mathcal{PD}$ and $\mathcal{PDS}$ using continuous values for V_{DD} are in solid lines. The results of optimization for discrete V_{DD} is presented in dashed lines. The conventional power-delay curve is labeled the $\mathcal{PD}$ curve. The two $\mathcal{PDS}$ curves are obtained when cu is optimized for SEU robustness with charges Q_{rob} of 15 fC and 20 fC respectively.

At large values of delay Δ, the gates in $\mathcal{PD}$ are near the minimum size and there is a power overhead to achieving SEU robust-

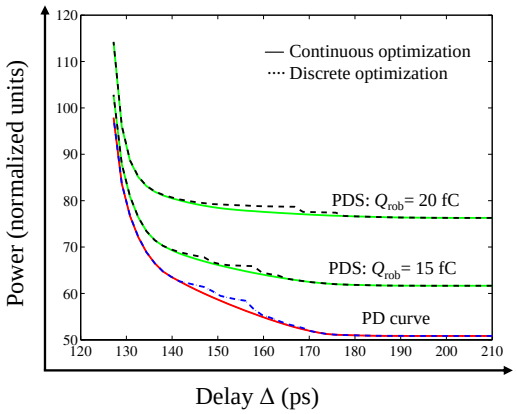


Figure 3: This figure illustrates the power versus delay curves for the benchmark circuit cu in 70 nm technology. The solid (dashed) curve corresponds to the continuous (discrete) version of $\mathcal{PD}$ and $\mathcal{PDS}$ in each case. For large delay Δ, the power of the design optimized for SEU robustness by $\mathcal{PDS}$ is greater than that of the design optimized for power-delay by $\mathcal{PD}$. For the same delay Δ, the overhead for a SEU robustness charge $Q_{\rm rob}$ of 15 fC is less than that for a $Q_{\rm rob}$ of 20 fC. As Δ decreases, the delay constraints dominate the SEU robustness constraints and all the curves converge.

ness with $\mathcal{PDS}$. Both $\mathcal{PD}$ and $\mathcal{PDS}$ use sizing exclusively, and only low-$V_{\rm DD}$ assignments are made in this region. Note also that for the same delay Δ, the power overhead for a $Q_{\rm rob}$ of 15 fC is less than the power overhead for a $Q_{\rm rob}$ of 20 fC. As delay Δ is continuously reduced, the gates start increasing in size and both $\mathcal{PD}$ and $\mathcal{PDS}$ begin high-$V_{\rm DD}$ assignments. There are jumps between the discrete and continuous versions of both $\mathcal{PD}$ and $\mathcal{PDS}$ in this range of delay Δ in all three cases when mixed $V_{\rm DD}$ assignments are made. This is because the continuous algorithm always returns a lower bound on the power by making $V_{\rm DD}$ assignments over the available range. As delay Δ is further reduced, one or more arrival time constraints in the optimization formulation begin to dominate the SEU robustness constraints at one or more gates. Over this range of delay Δ, both $\mathcal{PD}$ and $\mathcal{PDS}$ prefer high-$V_{\rm DD}$ assignments. Thereafter, after the flexibility afforded by high-$V_{\rm DD}$ is exhausted, both use sizing exclusively to meet delay constraints. As a result, the continuous and discrete versions of both $\mathcal{PD}$ and $\mathcal{PDS}$ converge since all supply voltages are set to high-$V_{\rm DD}$ values. On the far left, i.e., at very low delay Δ, the delay constraints dominate the SEU robustness constraints at all the gates and the SEU robustness constraints are trivially satisfied. At this point, the three curves converge as seen in the figure. Since the overhead for 20 fC $Q_{\rm rob}$ is higher than the overhead for 15 fC $Q_{\rm rob}$, the curve for 15 fC $Q_{\rm rob}$ converges with the base $\mathcal{PD}$ curve earlier.

4.3 Optimization results

The results for design optimization of twelve logic circuits from the ISCAS85 and LGSynth91 suite [17] is presented in Fig. 4. Note that we chose circuits that were purely logic or a mixture of logic and control for the experiments.

Fig. 4 presents power and area overhead (in %) when the benchmarks are optimized using $\mathcal{PD}$ and $\mathcal{PDS}$ to meet a delay constraint of $T_{\rm spec}$, $1.15\,T_{\rm spec}$, and $1.3\,T_{\rm spec}$ on all outputs using simultaneous dual-$V_{\rm DD}$ and gate sizing. The value of $T_{\rm spec}$ for each benchmark was set to $\Delta_{\rm min} + 0.1(\Delta_{\rm max} - \Delta_{\rm min})$, where $\Delta_{\rm min}$ and $\Delta_{\rm max}$ are the minimum and maximum delays for the design [18]. The power and area overhead for SEU robustness for the high performance design that meets $T_{\rm spec}$ is reported in the first row of the figure. The $T_{\rm spec}$ values are then relaxed by 15% and 30% from this optimum and the power and area overhead for SEU robustness are reported in the second and third rows of the figure respectively. In all three cases, the overhead for SEU robustness is reported with respect to the total power and area of the design optimized using the $\mathcal{PD}$ algorithm for that value of $T_{\rm spec}$. At $T_{\rm spec}$, $1.15\,T_{\rm spec}$, and $1.3\,T_{\rm spec}$, the average power overhead for 15 fC $Q_{\rm rob}$ was 5.7 %, 12 %, and 20.6 %; the average area overhead was 7.0 %, 10.6 %, and 28.2 %. Similarly, the average power overhead for 20 fC $Q_{\rm rob}$ was 23.8 %, 33.9 %, and 55.0 %; the average area overhead was 26.4 %, 32.5 %, and 53.2 % respectively.

The maximum runtime for i10, the largest circuit with 2950 gates, was 118 minutes (50 seconds) for discrete (continuous) optimization on a 2.4 GHz Opteron processor with 4 GB of memory. The maximum runtime usually occurs when the design is optimized for $1.15\,T_{\rm spec}$, since there is extensive use of dual-$V_{\rm DD}$ optimization without saturation to exclusive use of low-$V_{\rm DD}$ or high-$V_{\rm DD}$ across the design. The discrete optimization algorithm has a longer runtime since the branch-and-bound technique makes several iterations of the continuous optimization algorithm to discretize all $V_{\rm DD}$ assignments.

The power and area overhead required for SEU robustness increases from high performance (delay = $T_{\rm spec}$) to low performance (delay = $1.3\,T_{\rm spec}$) designs. This is because when the design is optimized for $T_{\rm spec}$, a significant number of gates in the design have larger sizes and high $V_{\rm DD}$ in the base case. Hence, the overhead required to satisfy SEU robustness constraints is a smaller fraction of the power and area of the baseline design. As we relax $T_{\rm spec}$, there is a decrease in the average size of the gates and fewer gates use high $V_{\rm DD}$ to meet delay constraints when $\mathcal{PD}$ is run. In the term $k_i' W_i^{-1} (V_{\rm DD})^{-2-\beta_0} (C_{{\rm out},i})^{-\beta_1}$ at a gate in the design, smaller gates and low-$V_{\rm DD}$ on average implies that $W_i^{-1} (V_{\rm DD})^{-2-\beta_0} (C_{{\rm out},i})^{-\beta_1}$ is larger on average across the design after $\mathcal{PD}$ optimization. When $\mathcal{PDS}$ is run, it has to increase the W_i and make more assignments to high-$V_{\rm DD}$ to meet SEU robustness requirements. This is observed in the larger power and area overhead with respect to the baseline case for slow designs (large $T_{\rm spec}$).

Note that one constraint on $V_{\rm DD}$ assignment at all gates i in both $\mathcal{PD}$ and $\mathcal{PDS}$ is of the form $V_{{\rm DD},i} \geq V_{{\rm DD},j}$ where $j \in {\rm fanout}(i)$. This ensures a clean partitioning of high- and low-$V_{\rm DD}$ gates for compatibility with standard flows for dual-$V_{\rm DD}$-based design.

Finally, it is clear that the proposed algorithm provides the designer with several alternatives to choose from. A significant advantage is that $\mathcal{PDS}$ optimization for SEU robustness combines optimization for delay and power globally across the design, leading to optimal or near-optimal designs.

Selective optimization: The optimization technique described in this paper targets all the gates in the design, regardless of their contribution to the overall soft error failure rate of the logic circuit. It is possible to leverage the asymmetry in the soft error failure rates of gates – due to masking factors as well as local input vector biasing – to reduce the impact of the proposed technique. The asymmetry can be exposed by running complete soft error failure rate estimation, or by using cheaper fault simulation and analysis techniques. Such approaches have been explored in literature (e.g., [6,7,19,20]), and the proposed technique can be guided using these metrics to selectively optimize a subset of the gates for robustness to SEUs.

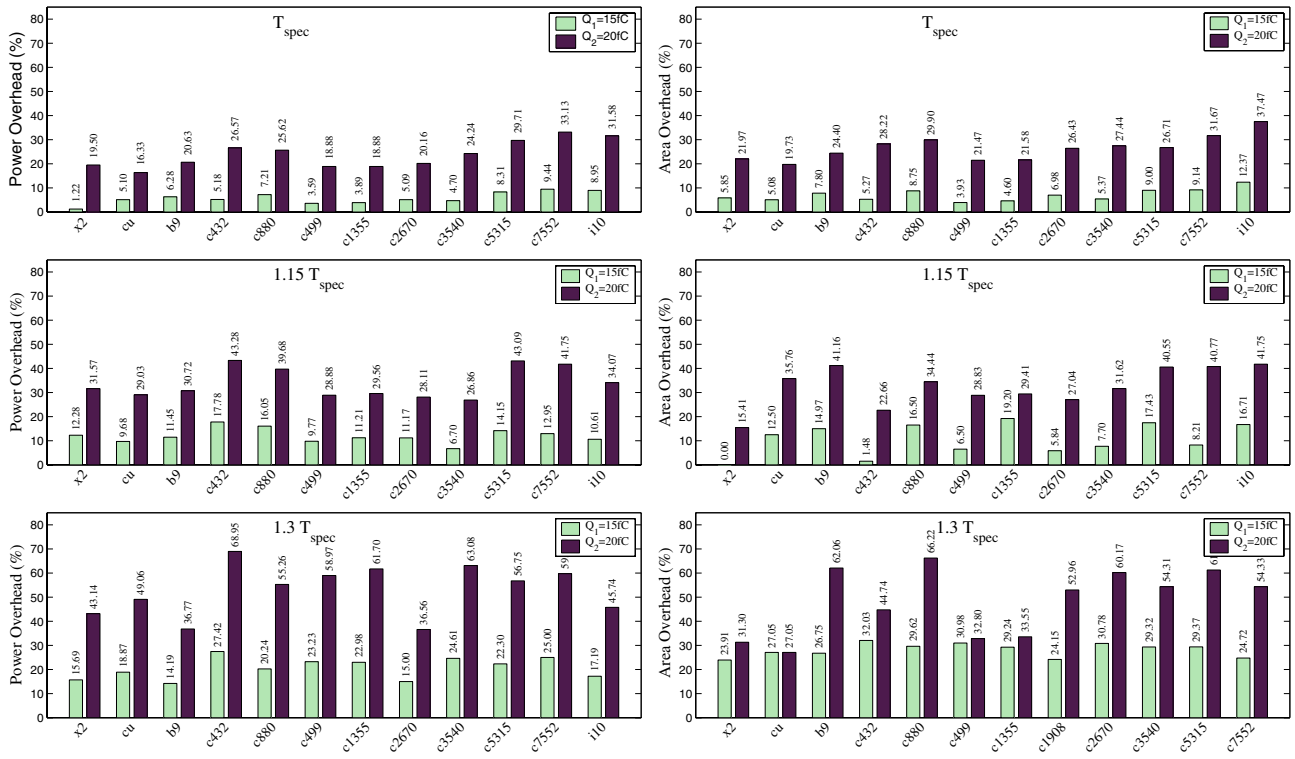


Figure 4: This figure presents power and area overhead (in %) when the benchmarks are optimized to meet a delay constraint of $T_{\rm spec}$, $1.15\,T_{\rm spec}$, and $1.3\,T_{\rm spec}$ on all outputs using simultaneous dual-$V_{\rm DD}$ and gate sizing. The power and area overhead required for SEU robustness increases from high performance ($\Delta = T_{\rm spec}$) to low performance ($\Delta = 1.3\,T_{\rm spec}$) designs (explained in Sec. 4.3).

5. Conclusion

An efficient global optimization technique that uses gate size and $V_{\rm DD}$ as design parameters to realize SEU-robust designs was described in this paper. The proposed approach incorporates SEU robustness constraints into a traditional area-delay-power optimization framework based on geometric programming. The results motivate further research into approaches based on (i) selective optimization for SEU robustness and (ii) exploration of the synergy and tradeoffs between the proposed technique and techniques that explicitly target SEU robustness in latches and flip-flops.

6. References

[1] J. F. Ziegler *et al.*, "IBM experiments in soft fails in computer electronics (1978–1994)," *IBM Journal of Research and Development*, vol. 40, pp. 3–18, Jan. 1996.

[2] H. H. K. Tang and K. P. Rodbell, "Single-event upsets in microelectronics: fundamental physics and issues," *Materials Research Society Bulletin*, vol. 28, pp. 111–116, Feb. 2003.

[3] R. Baumann, "Soft errors in advanced computer systems," *IEEE Design and Test of Computers*, vol. 22, pp. 258–266, May 2004.

[4] D. G. Mavis and P. H. Eaton, "Soft error rate mitigation techniques for modern microcircuits," in *Proc. Intl. Reliability Physics Symposium*, pp. 216–225, 2002.

[5] S. Mitra *et al.*, "Robust system design with built-in soft error resilience," *IEEE Computer*, vol. 38, pp. 43–52, Feb. 2005.

[6] J. M. Cazeaux *et al.*, "On transistor level gate sizing for increased robustness to transient faults," in *Proc. Intl. On-line Testing Symposium*, pp. 23–28, 2005.

[7] Q. Zhou and K. Mohanram, "Gate sizing to radiation harden combinational logic," *IEEE Trans. Computer-aided Design*, vol. 25, pp. 155–166, Jan. 2006.

[8] J. P. Fishburn and A. E. Dunlop, "TILOS: A posynomial programming approach to transistor sizing," in *Proc. Intl. Conference Computer-aided Design*, pp. 326–328, 1985.

[9] S. Boyd *et al.*, "Digital circuit optimization via geometric programming," *Operations Research*, vol. 53, pp. 899–932, Nov.–Dec. 2005.

[10] R. Murgai, "Technology-based transformations," in *Logic synthesis and verification* (S. Hassoun, T. Sasao, and R. K. Brayton, eds.), ch. 6, Kluwer Academic Publishers, Boston, MA, 2002.

[11] E. L. Lawler and D. E. Wood, "Branch-and-bound methods: A survey," *Operations Research*, vol. 14, pp. 699–719, Jul.–Aug. 1966.

[12] S. Boyd *et al.*, "A tutorial on geometric programming." To appear *Optimization and Engineering*. Please visit the URL http://www.stanford.edu/ boyd/gp_tutorial.html for further details.

[13] G. C. Messenger, "Collection of charge on junction nodes from ion tracks," *IEEE Trans. Nuclear Science*, vol. 29, pp. 2024–2031, Dec. 1982.

[14] L. B. Freeman, "Critical charge calculations for a bipolar SRAM array," *IBM Journal of Research and Development*, vol. 40, pp. 119–129, Jan. 1996.

[15] "MOSEK ApS." Details are available at http://www.mosek.com/.

[16] Y. Cao *et al.*, "New paradigm of predictive MOSFET and interconnect modeling for early circuit design," in *Proc. Custom Integrated Circuits Conference*, pp. 201–204, 2000.

[17] S. Yang, "Logic synthesis and optimization benchmarks user guide," Tech. Rep. 1991-IWLS-UG-Saeyang, MCNC, Research Triangle Park, 1991.

[18] J. Singh *et al.*, "Robust gate sizing by geometric programming," in *Proc. Design Automation Conference*, pp. 315–320, 2005.

[19] C. Zhao, Y. Zhao, and S. Dey, "Constraint-aware robustness insertion for optimal noise-tolerance enhancement in VLSI circuits," in *Proc. Design Automation Conference*, pp. 190–195, 2005.

[20] B. Zhang, W.-S. Wang, and M. Orshansky, "FASER: Fast analysis of soft error susceptibility for cell-based designs," in *Proc. Intl. Symposium on Quality Electronic Design*, pp. 755–760, 2006.