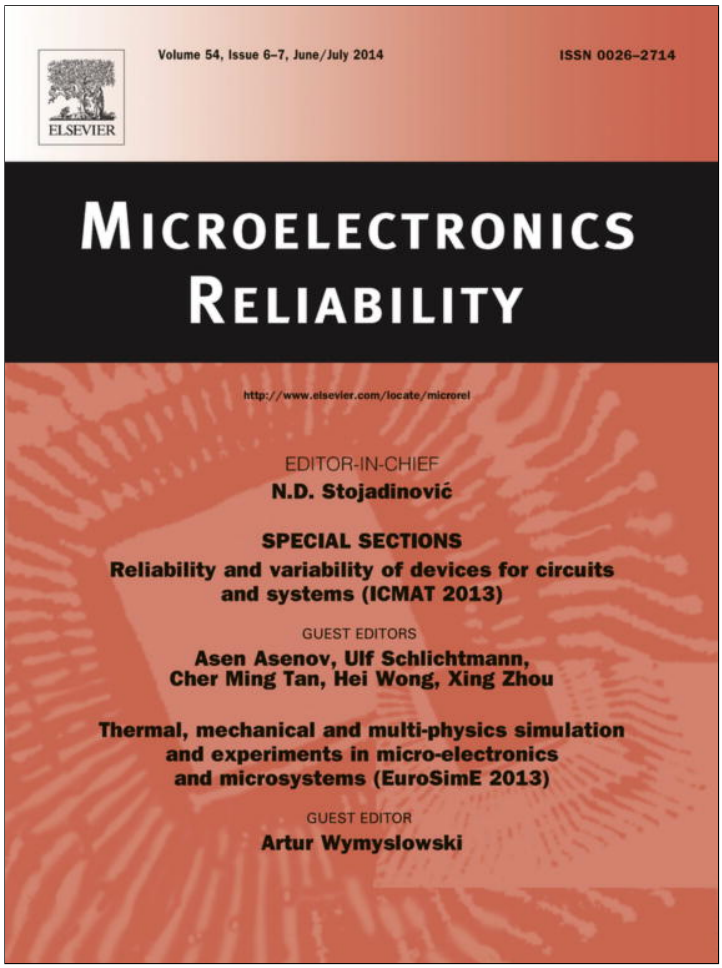

Microelectronics Reliability 54 (2014) 1412–1420

Contents lists available at ScienceDirect

Microelectronics Reliability

journal homepage: www.elsevier.com/locate/microrel

Soft error estimation and mitigation of digital circuits by characterizing input patterns of logic gates

Siavash Rezaei [a,*], Seyed Ghassem Miremadi [a], Hossein Asadi [a], Mahdi Fazeli [b]

[a] Department of Computer Engineering, Sharif University of Technology, Tehran, Iran
[b] Department of Computer Engineering, Iran University of Science and Technology, Tehran, Iran

ARTICLE INFO

Article history:
Received 29 May 2013
Received in revised form 4 February 2014
Accepted 3 March 2014
Available online 3 April 2014

Keywords:
Soft errors
Single event transient
Multiple event transient
Radiation hardening
Gate sizing

ABSTRACT

Soft errors caused by particles strike in combinational parts of digital circuits are a major concern in the design of reliable circuits. Several techniques have been presented to protect combinational logic and reduce the overall circuit *Soft Error Rate* (SER). Such techniques, however, typically come at the cost of significant area and performance overheads. This paper presents a low area and zero-delay overhead method to protect digital circuits' combinational parts against particles strike. This method is made up of a combination of two sub-methods: (1) a SER estimation method based on signal probability, called *Estimation by Characterizing Input Patterns* (ECIP) and (2) a protection method based on gate sizing, called *Weighted and Timing Aware Gate Sizing* (WTAGS). Unlike the previous techniques that either overlook internal nodes signal probability or exploit fault injection, ECIP computes the sensitivity of each gate by analytical calculations of both the probability of transient pulse generation and the probability of transient pulse propagation; these calculations are based on signal probability of the whole circuit nodes which make ECIP much more accurate as well as practical for large circuits. Using the results of ECIP, WTAGS characterizes the most sensitive gates to efficiently allocate the redundancy budget. The simulation results show the SER reduction of about 40% by applying the proposed method to ISCAS'89 benchmark circuits while imposing no delay overhead and 5% area overhead.

1. Introduction

The collision of alpha and neutron particles to a semiconductor device may generate a track of electron–hole pairs. The possible existence of electric field in off-state transistors causes electrons and holes move in the opposite directions [1–3]. This generates an unwelcome transient current pulse which may charge or discharge the load capacitance of a single or multiple gates causing either *Single Event Transient* (SET) or *Multiple Event Transients* (METs) [4]. If a particle directly strikes the sensitive nodes of memory elements, it may change the stored value depending on the amount of deposited charge, resulting in either *Single Event Upset* (SEU) or *Single Event Multiple Upset* (SEMU) [5]. Any unwelcome bit-flip in memory elements due to these phenomena is referred to as soft error.

Although SEUs were previously the major concern in digital circuits, with emerging nanoscale dimensions, SETs and METs have become the dominant threat to the reliability of digital circuits due to two main reasons. First, the occurrence rate of SEUs in sequential parts (memories and latches) has been approximately constant over several technology generations; while technology scaling extremely increases the occurrence probability of SETs/METs in combinational parts [6]. Second, there are two main approaches that can effectively protect sequential parts against SEUs: (a) *Error Detection And Correction* (EDAC) codes which are a viable solution to protect sequential parts such as cache or register file against SEUs [4,7], and (b) hardening approach such as hardened latch [8], hardened flip-flop [9], and hardened SRAM [10]; the former approach, i.e., the EDAC codes, cannot be applied to detect SETs/METs and the latter approach, i.e., the use of radiation hardened elements, would result in a significant area, power, and delay overheads since the number of gates in a digital circuit is much greater than sequential cells.

Hardening of combinational parts can be done by two main approaches: (1) enhancement of the inherent masking capabilities of circuits, i.e., logical masking, electrical masking, and latching-window masking. In this approach, the masking factors inherently prevent SETs/METs either propagating in combinational logic or being latched in sequential elements [11]; the main shortcoming of this approach is imposing significant performance degradation.

* Corresponding author. Tel.: +98 9183668185.
E-mail addresses: srezaei@ce.sharif.edu (S. Rezaei), miremadi@sharif.edu (S.G. Miremadi), asadi@sharif.edu (H. Asadi), m_fazeli@iust.ac.ir (M. Fazeli).

http://dx.doi.org/10.1016/j.microrel.2014.03.003
0026-2714/

(2) selectively hardening of a subset of elements to prevent transient pulses to being generated [12,13,6]. In this approach, to achieve the best hardened circuit for a certain amount of overheads, an accurate comparative estimation of gates sensitivity as well as an appropriate allocation of redundancy budget to the selected elements are required. To the best of our knowledge, none of the previous selective hardening methods has proposed a practical approach to compute the probability of transient pulse generation in large circuits. The only well-known approach to compute this probability is based on fault injection which is not tractable for large circuits.

This paper presents an analytical *Soft Error Rate* (SER) estimation method followed by a low area and zero-delay overhead method to protect combinational logic against particles strike. The main contribution of this paper can be discussed in two major parts as in the following.

1. SER estimation: this paper presents a detailed study supplemented with extensive simulation results to demonstrate the inaccuracy due to ignoring the effect of input patterns in transient pulses generation. Based on this fact, we propose an analytical method to determine circuits' SER by characterizing the circuits' nodes signal probability. We call this method *Estimation by Characterizing Input Patterns* (ECIP).
2. SER mitigation: we have demonstrated that the signal probability of internal nodes has a significant effect on selective hardening methods. Hence, we leverage this fact to determine the most sensitive circuit's gates and protect them by using our proposed gate sizing algorithm (called *Weighted and Timing Aware Gate Sizing* (WTAGS) which is a combination of the best characterizations of previous gate sizing methods). Finally, we investigate the impact of the input reordering technique beside WTAGS with different orders to achieve higher level of protection.

The main aim of the proposed method is to enhance the reliability of target circuits considering limited area and/or delay overheads. To this end, we try to efficiently assign the available overhead to achieve the highest possible reliability. The proposed method is applied to ISCAS'89 benchmark circuits using the Nangate 45 nm technology library [14]. The efficiency of the proposed method is evaluated by using a combination of HSPICE simulations and statistical analysis. The results demonstrate on average 40% SER reduction with 5% area overhead and no performance penalty.

The rest of this paper is organized as follows. In Section 2, we present an overview of previous research on SET modeling and the previous work on SER reduction. Section 3 presents the proposed method, i.e., ECIP and WTAGS. The simulation setup and the simulation results are given in Sections 4 and 5, respectively. Section 6 discusses the limitations of the proposed method and possible extension of this work. Finally, Section 7 concludes the paper.

2. Related work

Before discussing the previous works on SER mitigation techniques, it is necessary to explain how the effect of a particle strike on a sensitive node of a circuit can be modeled. This effect can be modeled as a single or a double exponential time-dependent current pulse (injecting into the victim transistors drain) [15]. The double exponential model has been widely used to model an alpha particle strike [16] while the single exponential model is more accurate to model a neutron particle strike [17]. We have used the single exponential model shown in (1) in our experiments, however, this does not affect the effectiveness of the proposed methods.

$$I(t) = \frac{2 \times Q}{\tau_\alpha \times \sqrt{\pi}} \times \sqrt{\frac{t}{\tau_\alpha}} \times \left(e^{\frac{-t}{\tau_\alpha}}\right) \quad (1)$$

In this equation, Q is the amount of charge deposited by the strike of a particle and τ_α is the charge collection time constant of the p–n junction (a CMOS technology process-related factor).

Generally, SET mitigation can be done using the following approaches:

1. Reducing the probability that transient pulses result in soft errors, i.e., trying to prevent transient pulses to be latched by circuit bistables. This can be achieved mainly by increasing the capability of a circuit to mask transient pulses by logical, electrical, or latching-window masking factors.
2. Reducing the probability of transient pulse generation. This approach includes gate or transistor resizing and/or reordering techniques.

In the next subsections, we review these two categories in detail.

2.1. SER Mitigation by Enhancing Masking Factors

Inserting a filtering circuit into some paths of a digital circuit is the main idea of the circuit level methods presented in [18–23]. This increases the effect of electrical masking in the modified paths. A major disadvantage of such methods is introducing new susceptible regions to the combinational parts. Furthermore, inserting the filtering circuits in an internal node of a circuit imposes significant performance degradation if the target nodes rely on the circuit critical path.

The method presented in [9] have focused on increasing the probability of latching-window masking by proposing a circuit inserted in the clock input of memory elements. The proposed circuit prevents a SET with a pulse width less than a certain threshold value to be latched by regulating the clock edge timing. However, regulating the clock edge timing at sub-threshold voltages may make the design unreliable. In [24], a method has been presented which increases the effect of latching-window masking by using data multiple clocking. This method is based on a *Triple Modular Redundancy* (TMR) technique that votes between three different memory elements taking three different samples of data in different time slices. However, producing shifted clock pulses for redundant memory elements needs a relatively complex circuit. In addition, the voter circuitry in this method is a single point of failure. Lastly, this method introduces a significant amount of performance degradation and area overhead to the circuit.

2.2. SER Mitigation by Reducing Pulse Generation Probability

The most common point in fault avoidance based methods is gate sizing. When the dimension of a transistor is increased, because of enlargement of parasitic capacitances and augmentation of transistor current drive, the critical charge (Q_{Crit}) of transistor increases; consequently, the device would become more robust against particles strike. The critical charge of a transistor in a logic gate is the minimum amount of charge that if injected into the drain of that transistor, a transient voltage pulse is generated at the output of the gate. However, applying the gate sizing method to all logic gates imposes a significant amount of area and performance overhead. On the other hand, it has been shown that the origin of more than 80% of soft errors is only 50% of the gates [25]. Therefore, the gate sizing method can be used selectively. Thus, one of the most problematic challenges in utilizing the gate sizing method is determining the critical gates to achieve the maximum reduction of SER for limited amount of area and/or performance overhead budget.

There are several works that have tried to reduce the SER of combinational circuits using the gate sizing method

[25,12,6,13,26]. In all proposed techniques, a subset of the most sensitive gates is selected for gate sizing process. However, the main difference between these techniques is the way the sensitivity of logic gates is extracted. In some of these techniques, the sensitivity of gates is extracted by considering parameters such as logical masking effect, the number of primary outputs in fan-out cone, observability, and *Error Propagation Probability* (EPP) [12,6,13]. These parameters only focus on propagation probability of transient pulse from the fault site, i.e., the output of the gate that a particle strike has occurred. Such parameters, however, do not take into account the probability of a pulse generation at the fault site. To accurately identify the most vulnerable gates in a circuit, both probability of pulse generation and the probability of error propagation are essential.

In the method presented in [25], the sensitivity of each gate is computed by considering the probability of both SET generation and SET propagation. In this method, however, the probability of SET generation is pre-computed considering the type and the size of each gate. The main drawback of this technique is that it has ignored the effect of input patterns on the probability of pulse generation and pulse propagation. The method presented in [26] has considered input patterns; however, it uses a fault injection method which makes it impractical for large circuits.

A zero overhead technique at the circuit-level is presented in [27]. This technique reorders the gates' inputs to achieve the minimum probability of SET generation that can be obtained by each gate in an unprotected circuit. This is motivated by the fact that the critical charge of a transistor depends on the state of that transistor as well as the transistor location in the gate structure.

3. The proposed method

The proposed soft error estimation and mitigation method consists of three main parts: (1) vulnerability identification, (2) weighted and timing aware gate sizing, and (3) input reordering.

3.1. Vulnerability identification

As mentioned, the most important and challenging part of a soft error mitigation technique is to find the most vulnerable gates/paths in circuits. It should be noted that in a soft error mitigation technique based on selective protection of gates, the exact SER of gates is not desired rather we need to accurately rank the gates based on their contribution in the overall circuit SER. Therefore, the factors that have very slight or similar effect on all logic gates can be neglected. As described in [16], the SER of a circuit due to SETs can be computed according to (2).

$$\mathrm{SER} = \sum \mathrm{SER}(G_i) \tag{2}$$

The SER of gate G_i, i.e., $\mathrm{SER}(G_i)$, can be computed according to (3) [25].

$$\mathrm{SER}(G_i) = \mathrm{PGP}(G_i) \times \mathrm{EPP}(G_i) \times LP \tag{3}$$

In this equation, $\mathrm{PGP}(G_i)$ is the *Pulse Generation Probability* (PGP) at the output of gate G_i due to a particle strike to the transistors of the gate G_i. $\mathrm{EPP}(G_i)$ is the propagation probability of transient pulses from the output of gate G_i to at least one of the circuit sequential elements, and LP is the latching probability of a transient pulse in the sequential elements. As described in [15], due to low logic depth and hence high operational frequency in today's digital circuits, the effect of latching-window masking on the overall SER of the circuit has significantly decreased. On the other hand, the latching-window masking effect highly depends on the pulse width and the clock period. This means that the location of a particle strike does not considerably affect the circuit LP [12]. Thus, the latching probability of a transient pulse with a specific width is almost the same for all gates. However, if we intend to accurately measure the SER, this assumption would be a source of a negligible inaccuracy, but this inaccuracy would not affect the outcome of our ranking.

The other factors, i.e., the pulse generation probability and error propagation probability play an important role in the SER of logic gates and their corresponding SER rankings. Following we will explain how we estimate these two factors.

3.1.1. Probability of transient pulse generation

The most important contribution of this paper is the way we measure the probability of SER generation at the output of a gate. In addition, here, we present a detailed study supplemented with extensive simulation results to demonstrate the inaccuracy due to ignoring the effect of input patterns in transient pulse generation.

The pulse generation probability of gate G_i is computed according to (4), where n is the number of gate's inputs.

$$\mathrm{PGP}(G_i) = \sum_{v=0}^{2^n-1} \mathrm{PGP}_v(G_i) \times P_v \tag{4}$$

In (4), $\mathrm{PGP}_v(G_i)$ is the pulse generation probability at the output of gate G_i when its input value is equal to v ($0 \leqslant v < 2^n$). In this equation, P_v is the probability that the input value of gate G_i is equal to v. Using this equation, the effect of input values on the probability of pulse generation at the output of the gate is accurately considered. The probability of pulse generation at the output of gate G_i can be calculated by (5) [28], where m is the number of transistors in gate G_i.

$$\mathrm{PGP}_v(G_i) = \sum_{f=0}^{m-1} A_d(T_f) \times F \times K \times e^{-\frac{Q_{Crit(v)}(T_f)}{Q_s}} \tag{5}$$

In (5), $A_d(T_f)$ is the drain area of transistor T_f, F is the neutron flux, K is a constant, independent to the supply voltage and doping profiles, $Q_{Crit(v)}(T_f)$ is the critical charge of transistor T_f when the input value of gate G_i is v, and Q_s is the charge collection slope which strongly depends on the supply voltage and doping. The parameters F and K are common for all transistors of a circuit. The critical charge of a transistor in a logic gate depends on the state of the transistor and its position in the gate. Thus, to extract the critical charge of a transistor in a gate, it is necessary to take into account the effect of all possible input values of the gate. The critical charge of a gate's transistor highly depends on the value of the gate's inputs. This will result in different vulnerability for the gate. Fig. 1 shows the vulnerabilities of some basic gates of a standard library when different input values are applied. As shown in this figure, input values have a significant effect on the vulnerability of a gate. For example, a 2-input XOR gate is not vulnerable to particles strike when its input value is equal to either 1 (in binary "01") or 2 (in binary "10") in the decimal coding.

Based on (4) and considering different values of PGP_v for different values of v, it is obvious that the probability of transient pulse generation (PGP) strongly depends on the probability of input values (P_v). A straightforward way to consider the effect of input patterns when estimating the SER of a circuit is to assume that the probability that a line holds logical value of "1" or "0" is 0.5, i.e., the signal probability of a line in the circuit is assumed to be 0.5. To investigate the validity of this assumption, we have carried out a set of simulations for all ISCAS'89 benchmark circuits. For each circuit, we have performed two different experiments. In the first experiment, we have assumed that the signal probability of all circuit primary inputs are 0.5, i.e., it is assumed that input patterns have been distributed uniformly. In the second experiment, we have assumed that the signal probability of all circuit

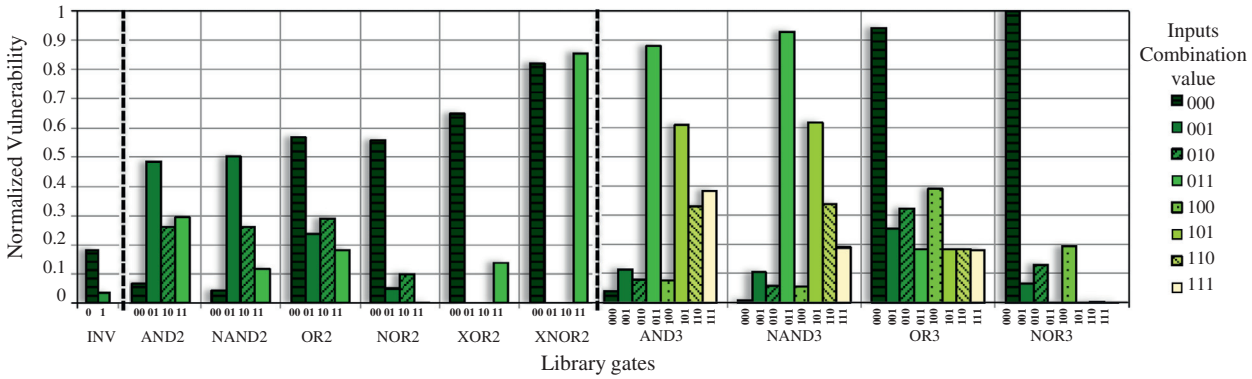

Fig. 1. Normalized vulnerability of standard library gates for different input values.

primary inputs are equal to either 0.1 or 0.9 with the same probability. For the sake of paper brevity, we only report the simulation results for six sample circuits in Fig. 2. This figure consists of sub-figures representing the distribution of the signal probability of internal nodes in two cases. Fig. 2a demonstrates this distribution when the signal probability of primary inputs is equal to 0.5 while in Fig. 2b, the signal probability of primary inputs is randomly set to either 0.1 or 0.9. The vertical axis in all sub-figures shows the signal probability and the horizontal axis represents the entire nodes of the circuit.

Two key points can be observed from the results reported in Fig. 2. The first observation is that even if one assumes a uniform input pattern for the primary inputs, the input pattern of the internal gates would not be uniform at all. Therefore, assuming that the probability of all possible input patterns in all gates of a circuit are equal will result in a significant inaccuracy for estimating the probability of SET generation at the gate outputs. The second observation point is that the signal probability of internal nodes are close to either 1 or 0 in both experiments. This case is more pronounced in the second experiment where we have assumed that the primary inputs may have a signal probability of 0.1 or 0.9. As an example considering *s*1494, when the signal probability of primary inputs is equal to 0.5 (as shown in Fig. 2a), only about 10% of internal nodes would have a signal probability greater than 0.2 and smaller than 0.8 while in the second experiment for this circuit (shown in Fig. 2b), more than 99% of the internal nodes have signal probability either greater than 0.8 or smaller than 0.2. This means that the probability that an internal gate has an unbalanced signal probability is significantly high. This means that almost always one specific input pattern is much more frequent at the input of internal gates. As we will see later in this section, we can reorder the inputs of a gate to have a highest possible critical charge with respect to the most probable input pattern of a logic gate. For the sake of clarity, here we present an example in which the probability of pulse generation at the output of a 2-input NAND gate is calculated considering two different values for inputs signal probabilities. Suppose that the signal probability of the first input of a 2-input NAND gate is 0.05 and the signal probability of the second input is 0.1. The normalized vulnerability of this gate can be computed according to (4) and Fig. 1 as follows:

$$P_0 = 0.95 \times 0.9 = 0.855 \quad P_1 = 0.05 \times 0.9 = 0.045$$

$$P_2 = 0.95 \times 0.1 = 0.095 \quad P_3 = 0.05 \times 0.1 = 0.005$$

$$\mathrm{PGP(NAND2)} = \sum_{a=0}^{3} \mathrm{PGP}_a(G_i) \times P_a = 0.040 \times 0.855 + 0.500 \times 0.045 + 0.261 \times 0.095 + 0.116 \times 0.005 = 0.082$$

Now suppose that the signal probability of the first input of 2-input NAND gate is 0.95 and the signal probability of the second input is 0.1. Using the same calculations, the normalized vulnerability is computed as 0.442. As can be seen in this example, the inputs signal probability would significantly affect the probability of transient pulse generation.

3.1.2. Error propagation probability (EPP)

Electrical and logical masking are two factors affecting the propagation of a transient pulse. However, in order to rank the gates based on their sensitivity, computing only logical masking is sufficient. This is because in nanometer technology, the effect of electrical masking has been significantly decreased due to reduced nodal capacitances and circuits supply voltages [29]. In addition, both logical and electrical masking factors of a gate depend on the distance of the gate to the primary outputs. This means that the closer a gate to the primary outputs, the more its logical and electrical masking factors. In fact, it is uncommon that a gate has a high probability of logical masking while having low probability of electrical masking and vice versa [12]. To compute the probability of logical masking effect, we use a statistical analysis method presented in [13,30]. In this method, a set of probabilities is propagated from each gate towards primary outputs and memory elements. These probabilities are:

- P_0: the probability that the node has the correct logic value of 0.
- P_1: the probability that the node has the correct logic value of 1.
- P_a: the probability that the node has an erroneous value that is propagated from the error site within an even number of inversions.
- $P_{\bar{a}}$: the probability that the node has an erroneous value that is propagated from the error site within an odd number of inversions.

For the output of a gate, these probabilities are computed according to the gate type and the set of probabilities related to the gate's inputs [13].

In the following subsections, our proposed protection technique is presented. Briefly, the proposed protection technique consists of two parts. The first part is based on a gate sizing approach and the second part is based on an input reordering approach.

3.2. Weighted and timing aware gate sizing process

In our work, we have assumed that before protecting a circuit, the designer identifies the maximum allowable area and performance overhead. The problem statement here is how the allowable

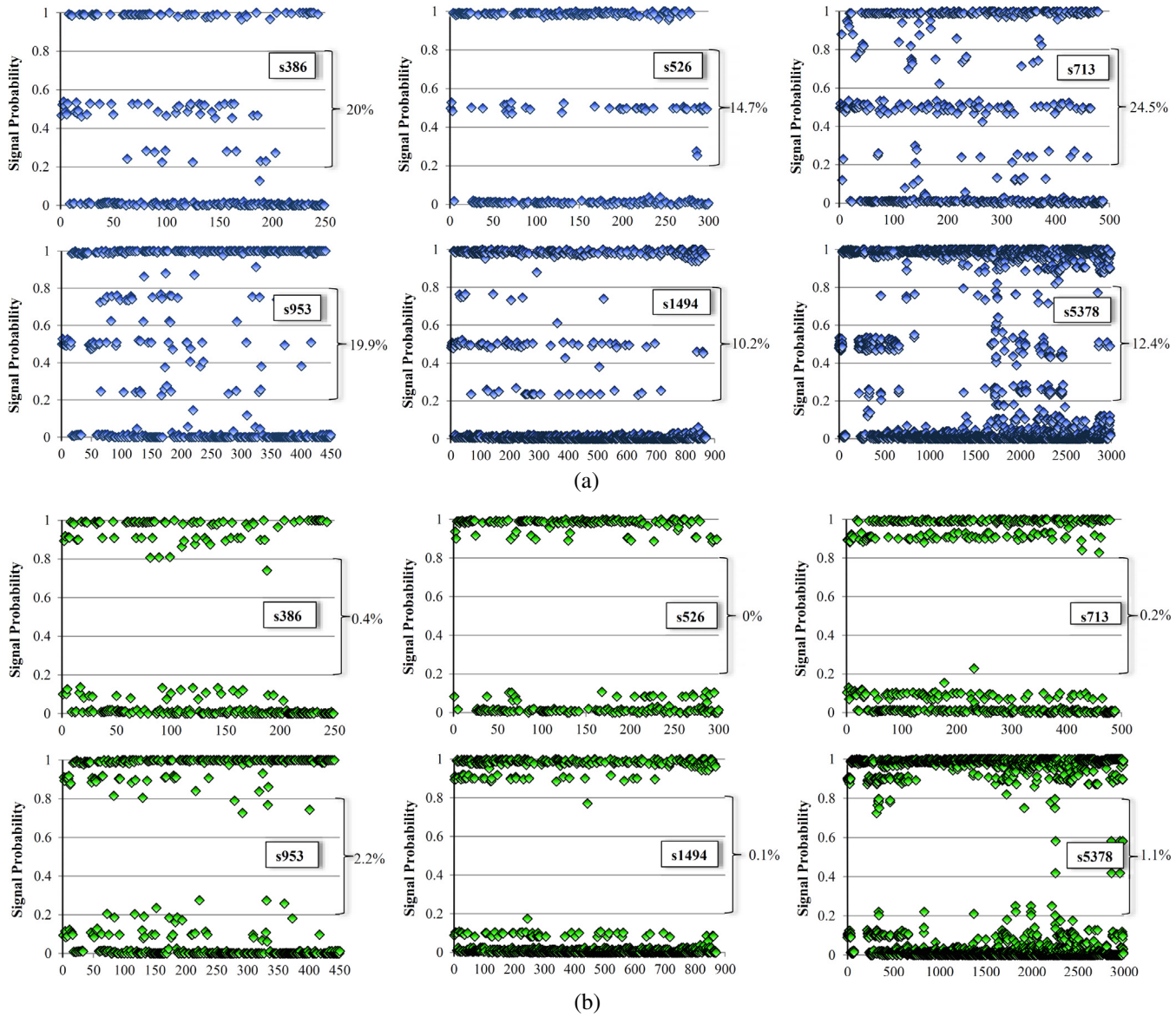

Fig. 2. Signal probabilities of internal nodes of six ISCAS'89 benchmark circuits for two situations: (a) the signal probability of primary inputs is set to 0.5, and (b) the signal probability of primary inputs is set to 0.1 or 0.9.

area overhead can be shared among the vulnerable gates in order to gain maximum robustness against particles strike using gate sizing approach.

After computing the sensitivity of logic gates, we determine a threshold, called *sensitivity threshold*, such that only the gates having higher sensitivity than the sensitivity threshold are considered for protection. The sensitivity threshold is defined as a fraction of the highest measured gate sensitivity. Selecting an effective sensitivity threshold will be discussed in Section 5. The allowable area overhead should be shared among gates having a sensitivity greater than the sensitivity threshold. However, it should be noted that there is always a limitation on the amount of gate upsizing. This is because of three main reasons that each one determines a restricting factor for upsizing. These three factors define a maximum level of upsizing for a gate, called *Maximum allowable Upsizing Factor* (MUF) as detailed in the following:

1. Upsizing a gate would increase the amount of its inputs capacitance. This would violate the allowable output capacitance of the gates in fan-in defined in the technology library. To calculate the MUF of a gate for acceptable fan-out capacitance of its fan-in gates (MUF1 in Fig. 3), for each fan-in gate, we determine the maximum capacitance which can be added. Then, MUF1 is determined by tracing different upsizing factors starting from the area budget of the target gate using the binary search algorithm to find the proper value which does not violate the maximum determined capacitances.
2. Upsizing a gate would have a negative impact on the circuit paths delay. To determine the MUF of a gate such that the maximum acceptable delay (MUF2 in Fig. 3) is not violated, we need to upsize the gate for different factors and then check whether the delay restrictions have been violated or not. We also use the binary search algorithm started from the given area budget of the target gate to find the most proper value of MUF2.
3. The SER of a gate becomes saturated as the gate size increases, i.e., after a specific threshold, upsizing would not decrease the gate SER tangibly. The saturation threshold for each library gate is pre-computed by applying different upsizing factors to the library gates.

Therefore, it is probable that the available area budget that can be assigned to a gate for upsizing is greater than its MUF. In such

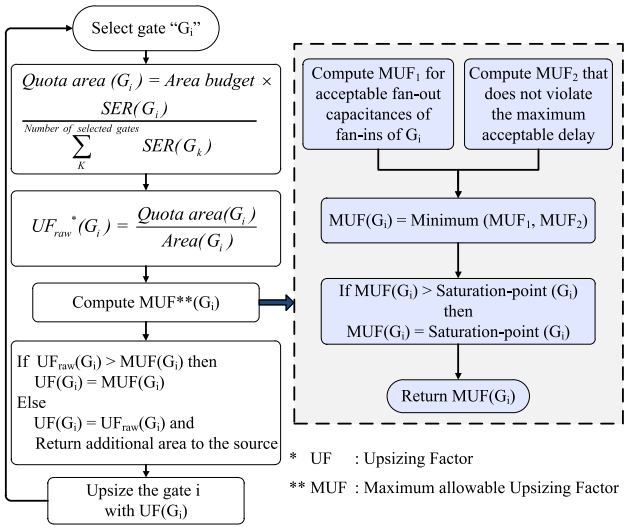

Fig. 3. Flowchart of the proposed gate sizing algorithm.

cases, the additional area is returned to the source to be assigned to other gates. Fig. 3 shows the upsizing process of an arbitrary gate.

3.3. Input reordering

It has been shown in [27] that the critical charge of a gate depends on the spatial order of its inputs, i.e., input reordering in a gate would have impact on the gate SER. In other words, an effective inputs reordering of a gate can reduce its PGP.

To achieve more reduction in SER without additional overhead, we combine the proposed method with the input reordering technique proposed in [27]. Reordering gates' inputs has, however, two main challenges:

1. Input reordering can only be applied to symmetric gates, since reordering the inputs of a non-symmetric gate may change the gate functionality.
2. The effect of latching-window masking may be altered since the propagation delay of paths may be affected. This challenge appears when two different transient pulses converge together.

In the proposed technique, we have applied the input reordering technique to only symmetric gates. In addition, we have ignored the effect of input reordering on paths propagation delay, as this effect is not significant. This technique can be employed either before or after applying the proposed method. The results of these two cases are presented in Section 5.

4. Simulation setup

To implement the proposed method, an automated tool has been developed in C programming language. This tool gets the following files and information as inputs: (1) Verilog description of the target circuit, (2) pre-characterization results of standard cells of technology library extracted by HSPICE simulations, (3) the standard cells information extracted from technology library, and (4) the desired area and timing overheads. The pre-characterization results include the critical charge of standard cells' transistors for different values of upsizing factor, output capacitance, and input patterns. These values are used to compute the SER of each gate before and after the hardening process. Since the values of upsizing factor and the output capacitance in a circuit are correlated, the sensitivity of gates is computed using the linear interpolation method.

Nangate 45 nm technology library [14] is used as our target library. Because of nanometer dimensions of the used technology library, the *Non-Linear Delay Model* (NLDM) has been employed for calculating the delay of circuits. Moreover, we have used the charge collection slope (Q_S) of 10.48fC reported in [31] for 45 nm technology size. In addition, ISCAS'89 benchmark circuits are used as testbench in the experiments.

To validate the proposed method, we have also developed a reference model employing the Monte-Carlo simulation based on fault injection experiments. In the reference model, we have used a *Statistical Fault Injection* (SFI) engine based on the Monte-Carlo simulation which has been developed in [30]. In this engine, for each simulation iteration, a random pulse width is injected at the output of a random gate with an arbitrary value of circuit's primary inputs (at a random time during the clock period). This is done for numerous number of glitches. Then, the timing simulation determines if the injected pulse is propagated and captured in any flip-flop. To determine circuit's SER, for each fault injection, the probability of transient pulse generation for a corresponding random primary input vector is also determined. Then, the summation of transient pulse generation probability for each pulse being propagated and captured in a flip-flop is computed. Note that three masking factors, logical, electrical, and latching-window masking have been incorporated in the SFI engine (for more information please see [30]). Thus, the results obtained by SFI are considered as reference model. Finally, it is notable that for all results presented in Section 5, a uniform signal probability is supposed for the primary inputs of the circuits.

5. Simulation results

Fig. 4 shows the normalized SER of ISCAS'89 benchmark circuits computed by our proposed method (ECIP) and those extracted by the SFI engine. The comparison of ECIP and the reference model reveals that ECIP has an inaccuracy up to 12% of the results provided by the reference model. Based on the following analytical study, the slight SER difference between ECIP and SFI can be related to electrical and latching-window masking.

It is notable that for each gate, the probability of transient pulse generation for ECIP and SFI for a large number of iterations (for computing signal probabilities in ECIP and fault injection in SFI) is the same based on (6).

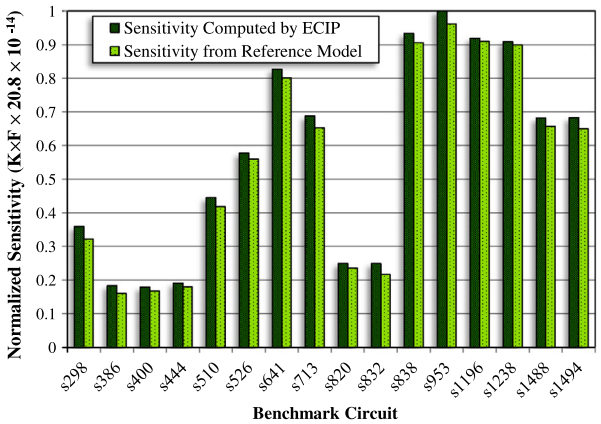

Fig. 4. Comparing ECIP with the reference model (based on the Monte-Carlo simulation and fault injection).

$$
\begin{aligned}
SER_{SFI}(G_i) &= \frac{A_0 \times PGP_0(G_i) + A_1 \times PGP_1(G_i) + \ldots + A_{2^k-1} \times PGP_{2^k-1}(G_i)}{N} \\
&= \frac{A_0 \times PGP_0(G_i) + A_1 \times PGP_1(G_i) + \ldots + A_{2^k-1} \times PGP_{2^k-1}(G_i)}{A_0 + A_1 + \ldots + A_{2^k-1}} \\
&\times \frac{A_0 + A_1 + \ldots + A_{2^k-1}}{N} = \left(\frac{A_0}{A_0 + A_1 + \ldots + A_{2^k-1}} \times PGP_0(G_i) \right. \\
&+ \frac{A_1}{A_0 + A_1 + \ldots + A_{2^k-1}} \times PGP_1(G_i) + \ldots + \frac{A_{2^k-1}}{A_0 + A_1 + \ldots + A_{2^k-1}} \\
&\left. \times PGP_{2^k-1}(G_i) \right) \times \frac{A_0 + A_1 + \ldots + A_{2^k-1}}{N} = (P_0 \times PGP_0(G_i) + P_1 \\
&\times PGP_1(G_i) + \ldots + P_{2^k-1} \times PGP_{2^k-1}(G_i)) \times \frac{A_0 + A_1 + \ldots + A_{2^k-1}}{N} \\
&= \left(\sum_{v=0}^{2^k-1} PGP_v(G_i) \times P_v \right) \times \frac{A_0 + A_1 + \ldots + A_{2^k-1}}{N} \qquad (6)
\end{aligned}
$$

where each $A_0, A_1, \ldots, A_{2^k-1}$ is the number of times which the pulse width is propagated and captured in at least one flip-flop as the input vector of gate G_i is 0, 1, …, $2^k - 1$, respectively. In this equation, k is the number of the gate's inputs, and N is the number of iterations. This equation shows that the amount of transient pulse generation probability computed by SFI is equal to that computed by the use of signal probabilities. Thus, for a large number of iterations, transient pulse generation probabilities computed by signal probabilities can also be used as a reference model to evaluate the transient pulse generation probability in SER estimation methods. In order to accurately extract the inaccuracy of previous SER estimation techniques which employ the uniform pulse generation probability, one should develop the reference SER estimation method such that it employs exactly the same error propagation approach. In this way, the discrepancy between the reference model and the previous SER estimation technique can be directly associated to the uniform pulse generation approach employed in the SER estimation technique. Likewise, it can be shown that for large number of iterations, the amount of logical masking for ECIP and SFI is also the same.

Fig. 5 presents the inaccuracy of SER estimation due to ignoring the input patterns probability of logic gates. Thus, to have a fair comparison, we have to use a reference method with two main features:

1. The same computation algorithm for transient pulse propagation probability with ECIP.
2. Pre-computing the probability of transient pulse generation for each gate instead of considering the probability of input patterns which is different across circuits.

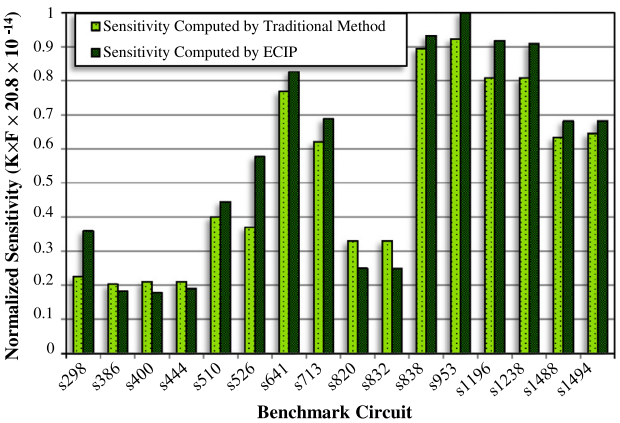

Fig. 5. Normalized sensitivities of the traditional [13,25] and the proposed method.

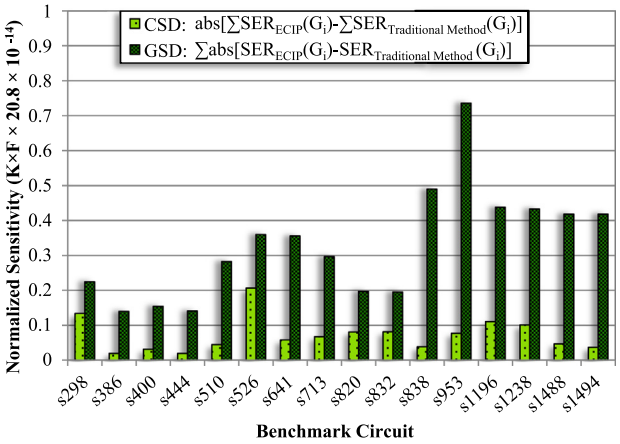

Fig. 6. Difference of normalized sensitivities computed by the traditional method [13,25] and the proposed method.

Unfortunately, we could not find any method in previous works with the above features. As such, to provide a fair comparison, we have combined two previous methods to measure the impact of considering input patterns on circuits SER. The transient pulse propagation and the transient pulse generation probability computation methods have been proposed in [13,25], respectively.

It can be concluded from Fig. 5 that there is a noticeable difference up to 38% between the SERs computed by considering gates' input patterns (ECIP) and ignoring those (we refer this method as the traditional method in this figure). It is notable that here all parts of the two methods are the same except ignoring and considering the gates' input patterns in computing PGP.

In each circuit, the SERs estimated by the traditional method for individual gates are either lower or higher than those estimated by ECIP. When the main objective is computing the total SER of a circuit, these higher and lower differences partially cancel the effect of each other. However, for the purpose of SER mitigation, the SER value of each individual gate is important. This is more vital for the selective hardening methods. Fig. 6 presents the amount of the cancellation effect for some ISCAS'89 benchmark circuits. This figure consists of two bars for each circuit. The first bar presents the difference between circuit SERs computed by the traditional method and ECIP, i.e., $abs[\sum SER_{ECIP}(G_i) - \sum SER_{Traditional\ method}(G_i)]$; we refer it as *Circuit SERs Difference* (CSD). The second bar is $\sum d(G_i)$, where $d(G_i)$ is equal to $abs[SER_{ECIP}(G_i) - SER_{Traditional\ method}(G_i)]$; we refer $\sum d(G_i)$ as *Gate SER Difference* (GSD). As it can be seen, for majority of the circuits the value of GSD is much greater than the value of CSD. This means that it is probable that the total SER of a circuit by two methods be approximately equal due to the cancellation effect of the SER of individual gates, however, as the bars related to GSD in Fig. 6 show, the SER of individual gates computed by the traditional method and ECIP are quite different. In other words, the vulnerable gate ranking performed by the traditional method is considerably different with that performed by ECIP. In fact, for the circuits in which the ratio of GSD/CSD is greater, the vulnerable gate ranking performed by the traditional method can be more inaccurate.

Fig. 7 shows the percentage of SER mitigation of 17 ISCAS'89 benchmark circuits for different values of sensitivity threshold. In this experiment, for each value of sensitivity threshold, the percentage of SER mitigation has been measured for 1%, 3%, and 5% area overheads and no delay overhead. The curves show that in sensitivity threshold of 10%, the maximum SER mitigation is achieved. Consequently, in the subsequent experiments, the sensitivity threshold is fixed to 10%. As it can be inferred from Fig. 7,

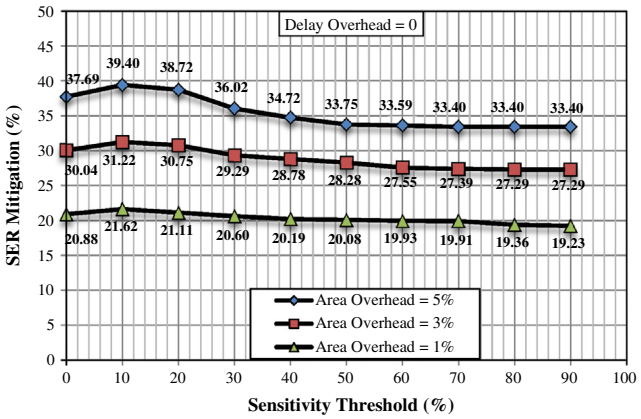

Fig. 7. SER mitigation for different values of sensitivity threshold for 1%, 3%, and 5% area overheads and no delay overhead.

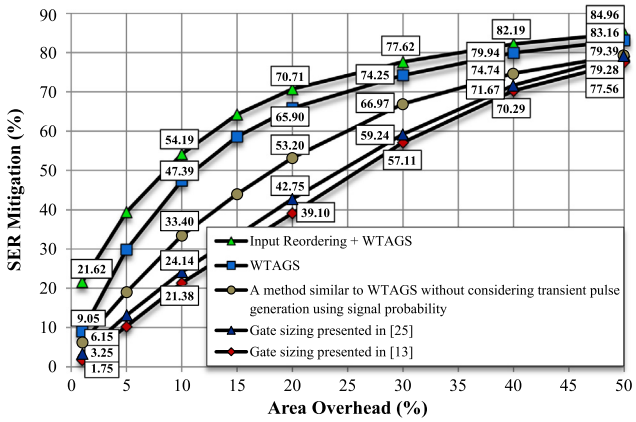

Fig. 9. The amount of SER mitigation for different values of area overhead and no delay overhead for the following methods: (1) the gate sizing method presented in [13], (2) the gate sizing method presented in [25], (3) a method similar to WTAGS without considering signal probability of circuit nodes in the computation of transient pulse probability, (4) WTAGS, (5) the combination of input reordering technique and WTAGS.

using a sensitivity threshold greater than 70% and 60%, the percentage of SER mitigation is saturated for the curves related to 3% and 5% area overheads, respectively. For the sake of clarity, we will explain this behavior in an example. Suppose that the area overhead is fixed to 5% and the sensitivity threshold is fixed to 90%. In this condition, even complete hardening of the selected subset of gates (with considering the maximum allowable upsizing factor) cannot consume the total area budget, because the number of selected gates is not enough. To utilize the unused area, our tool hardens the most sensitive gates that are not hardened yet. This continues until the whole redundant area is consumed. This process leads to the same results of SER mitigation for high values of sensitivity threshold.

Fig. 8 presents the amount of SER mitigation for three methods: (1) input reordering, (2) WTAGS, and (3) a combination of input reordering and WTAGS with two different possible orders. Here, the area and delay overheads are fixed to 5% and zero, respectively. On average, the input reordering method and WTAGS each solely mitigates the SER by 13.59% and 29.86%, respectively. Employing the input reordering method after WTAGS reduces SER by 38.5%. This reduction further improves to 39.4% by changing the order of these two methods, i.e., applying input reordering and then WTAGS. In fact, applying input reordering at the first helps WTAGS to select the most sensitive gates among the gates which have their minimum possible sensitivity leading to a more appropriate subset of gates.

Fig. 9 shows the percentage of SER mitigation for different values of area overhead and no delay overhead for 17 circuits of ISCAS'89 benchmark suite. The results in this figure have been reported for the following methods:

1. The gate sizing method presented in [13]. This method uses EPP to select more sensitive gates and allocate the area overhead budget as un-weighted.
2. The gate sizing method presented in [25]. This method uses logical masking probability and a uniform signal probability for all internal nodes to determine the sensitivity of logic gates and allocate the area overhead budget as un-weighted.
3. A gate sizing method exactly similar to WTAGS with the only difference that it uses a uniform signal probability for all internal nodes to determine the sensitivity of logic gates.
4. WTAGS.
5. The combination of input reordering technique and WTAGS.

It is notable that for all methods presented in this figure, the allocation of area overhead is accomplished with considering MUF which helps to have a fair comparison. The following items are the key observations that can be inferred from Fig. 9.

1. The SER reduction of WTAGS (and the combinational method) is greater than that of the other methods. This is because the proposed method offers a more effective selection of the suitable gates for protection and more effective allocation of the area budget. Thus, for the same overheads, WTAGS achieves more reduction in SER as compared to the previous methods. This reduction is more considerable for small amounts of overheads.
2. The percentage of SER mitigation of WTAGS and the other methods approximately reaches to the same point as the area overhead increases. This is because the number of intersection

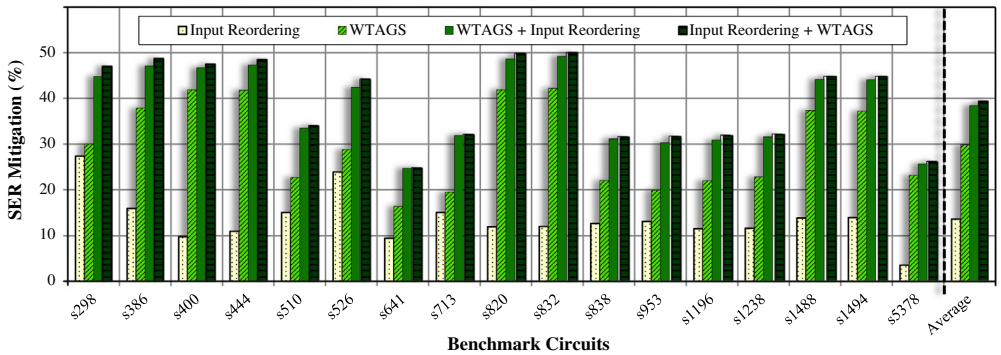

Fig. 8. SER mitigation for input reordering, WTAGS, and two possible combinations of input reordering and WTAGS for 5% area overhead and no delay overhead.

gates between the selected subset by all methods becomes greater when the amount of area overhead increases. This is mainly due to the ability of our tool to add more gates to the selected subset while the selected gates do not completely consume the given area overhead budget. Also, for high amounts of area overhead, WTAGS and the combinational method converge together since with increasing the amount of area overhead, the allocated area to logic gates increases and their sensitivity decreases exponentially. Thus, the difference between the sensitivities of a gate for different input values before upsizing is much greater than the difference between those of that gate after upsizing.

6. Discussion

Emerging deep sub-micron technologies and the integration of more cells in today's chips have caused higher occurrence probability of METs. It is notable that our proposed method is not limited to SETs and can be generalized to METs too. To protect a circuit against METs, it is more efficient to consider the circuit as neighboring regions after the placement in the chip, instead of considering the gates independently. This requires extracting neighboring gates of circuits using the layout information. Characterizing input patterns also plays an inevitable role in the computation of regional sensitivities of the chip; this can be done by considering the signal probability of the whole circuit nodes. Extracting the sensitivity of neighboring regions needs two steps: (1) extracting the occurrence probability of transient pulse generation at the output of gates in each region at the layout level according to the input patterns probability of logic gates, and (2) estimating the propagation probability of multiple pulses from neighboring gates at the gate level. METs modeling to propose an efficient method to protect circuits against this phenomenon is a part of our ongoing research and will be addressed in our future work.

7. Conclusion

Technology scaling increases soft errors caused by particles strike in today's digital circuits. One of the most effective approaches to mitigate these types of errors is selective hardening, i.e., hardening a subset of gates to achieve the best protected circuit for a certain amount of overheads. In this paper, we proposed an analytical method to estimate the sensitivity of logic gates by characterizing the corresponding inputs signal probabilities. We showed that considering input patterns is extremely important for the selective hardening approach. We also proposed a weighted and timing aware gate sizing algorithm to harden an appropriate subset of gates. Finally, we combined the proposed method with a zero overhead technique, input reordering, to achieve further reduction of circuit SERs. The simulation results show that the SER is reduced, on average, by 40% while the area overhead is about 5% and the delay overhead is zero.

References

[1] Hsieh CM, Murley PC, O'Brien R. A field-funneling effect on the collection of alpha-particle-generated carriers in silicon devices. IEEE Trans Electron Dev Lett (TEDL) 1981;2(4):686–93.

[2] Weaver C, Emer J, Mukherjee SS, Reinhardt SK. Techniques to reduce the soft error rate of a high-performance microprocessor. In: Proceedings of IEEE international symposium on computer architecture (ISCA'04); 2004. p. 264–75.

[3] Ramanarayanan R, Degalahal VS, Krishnan R, Kim J, Narayanan V, Xie Y, et al. Modeling soft errors at the device and logic levels for combinational circuits. IEEE Trans Depend Secure Comput (TDSC) 2009;6(3):202–16.

[4] Ziegler JF et al. Ibm experiments in soft fails in computer electronics (1978–1994). IBM J Res Develop 1996;40(1):3–18.

[5] Nicolaidis M. Design for soft error mitigation. IEEE Trans Dev Mater Reliab (TDMR) 2005;5(3):405–18.

[6] Rao RR, Blaauw D, Sylvester D. Soft error reduction in combinational logic using gate resizing and flipflop selection. In: Proceedings of IEEE/ACM international conference on computer-aided design (ICCAD'06); 2006. p. 502–9.

[7] Reed IS, Solomon G. Polynomial codes over certain finite fields. SIAM J Appl Math 1960;8(2):300–4.

[8] Fazeli M, Patooghy A, Miremadi SG, Ejlali A. Feedback redundancy: a power efficient seu-tolerant latch design for deep sub-micron technologies. In: Proceedings of IEEE international conference on dependable systems and networks (DSN'07); 2007. p. 276–85.

[9] She X, Li N, Carlson RM, Erstad DO. Single event transient suppressor for flip-flops. IEEE Trans Nucl Sci (TNS) 2010;57(4):2344–8.

[10] Calin T, Vargas F, Nicolaidis M, Velazco R. A low-cost, highly reliable seu-tolerant sram: prototype and test results. IEEE Trans Nucl Sci (TNS) 1995;42(6):1592–8.

[11] Sheng W, Xiao L, Mao Z. Soft error optimization of standard cell circuits based on gate sizing and multi-objective genetic algorithm. In: Proceedings of IEEE design automation conference (DAC'09); 2009. p. 502–7.

[12] Zhou Q, Mohanram K. Gate sizing to radiation harden combinational logic. IEEE Trans Comput – Aided Des Integr Circ Syst (TCAD) 2006;25(1):155–66.

[13] Asadi H, Tahoori MB. Soft error modeling and remediation techniques in asic designs. Elsevier Microelectron J (MEJ) 2010;41(8):506–22.

[14] Nangate. nangate open cell library; 2008, [http://openeda.si2.org/projects/nangatelib/].

[15] Mavis DG, Eaton PH. Seu and set modeling and mitigation in deep submicron technologies. In: Proceedings of IEEE international reliability physics symposium (IRPS'07); 2007. p. 293–305.

[16] Dabiri F, Nahapetian A, Massey T, Potkonjak M, Sarrafzadeh M. General methodology for soft-error-aware power optimization using gate sizing. IEEE Trans Comput – Aided Des Integr Circ Syst (TCAD) 2008;27(10):1788–97.

[17] Freeman L. Critical charge calculations for a bipolar sram array. IBM J Res Develop 1996;40(1):119–29.

[18] Almukhaizim S, Makris Y. Soft error mitigation through selective addition of functionally redundant wires. IEEE Trans Reliab (TR) 2008;57(1):23–31.

[19] Sasaki Y, Namba K, Ito H. Soft error masking circuit and latch using schmitt trigger circuit. In: Proceedings of international symposium on defect and fault tolerance in VLSI systems (DFT'06); 2006. p. 327–35.

[20] Deogun HS, Sylvester D, Blaauw D. Gate-level mitigation techniques for neutron-induced soft error rate. In: Proceedings of international symposium on quality electronic design (ISQED'05); 2005. p. 175–80.

[21] Krishnamohan S, Mahapatra N. Slack redistribution in pipelined circuits for enhanced soft-error rate reduction. In: Proceedings IEEE international system-on-chip (SoC) conference (SoCC'08); 2008. p. 159–62.

[22] Hill EL, Lipasti MH, Saluja KK. An accurate flip-flop selection technique for reducing logic ser. In: Proceedings of IEEE international conference on dependable systems and networks (DSN'08); 2008. p. 128–36.

[23] Haghi M, Draper J. Single-event transient mitigation in sub-micron combinational circuits. In: Proceedings of international midwest symposium on circuits and systems (MWSCAS'11); 2011. p. 1–4.

[24] Avirneni NDP, Somani AK. Low overhead soft error mitigation techniques for high-performance and aggressive designs. IEEE Trans Comput (TC) 2012;61(4):488–501.

[25] Srinivasan V, Sternberg AL, Duncan A, Robinson WH, Bhuva BL, Massengill LW. Single-event mitigation in combinational logic using targeted data path hardening. IEEE Trans Nucl Sci (TNS) 2005;52(6):2516–23.

[26] Sootkaneung W, Saluja KK. On techniques for handling soft errors in digital circuits. In: Proceedings of IEEE international test conference (ITC'10); 2010. p. 1–9.

[27] Sootkaneung W, Saluja KK. Gate input reconfiguration for combating soft errors in combinational circuits. In: Proceedings of international conference on dependable systems and networks workshops (DSN-W'10); 2010. p. 107–12.

[28] Hazucha P, Svensson C. Impact of cmos technology scaling on the atmospheric neutron soft error rate. IEEE Trans Nucl Sci (TNS) 2000;47(6):2586–94.

[29] Shivakumar P, Kistler M, Keckler WW, Burger D, Alvisi L. Modeling the effect of technology trends on the soft error rate of combinational logic. In: Proceedings of IEEE international conference on dependable systems and networks (DSN'02); 2002. p. 389–98.

[30] Asadi H, Tahoori MB, Fazeli M, Miremadi S. Efficient algorithms to accurately compute derating factors of digital circuits. Elsevier Microelectron Reliab (MR) 2012;52(6):1215–26.

[31] Peng HK, Wen CHP, Bhadra J. On soft error rate analysis of scaled cmos designs-a statistical perspective. In: Proceedings of IEEE/ACM international conference on computer-aided design (ICCAD'09); 2009. p. 157–63.