PS3-RAM: A Fast Portable and Scalable Statistical STT-RAM Reliability/Energy Analysis Method

Wujie Wen, Yaojun Zhang, Yiran Chen, Member, IEEE, Yu Wang, Senior Member, IEEE, and Yuan Xie, Senior Member, IEEE

Abstract—The development of emerging spin-transfer torque random access memory (STT-RAM) is facing two major technical challenges—poor write reliability and high write energy, both of which are severely impacted by process variations and thermal fluctuations. The evaluations on STT-RAM design metrics and robustness often require a hybrid simulation flow, i.e., modeling the CMOS and magnetic devices with SPICE and macro-magnetic models, respectively. Very often, such a hybrid simulation flow involves expensive Monte Carlo simulations when the design and behavioral variabilities of STT-RAM are taken into account. In this paper, we propose a fast and scalable semi-analytical method—PS3-RAM, enabling efficient statistical simulations in STT-RAM designs. By eliminating the costly macro-magnetic and SPICE simulations, PS3-RAM achieves more than 100,000× runtime speedup with excellent agreement with the result of conventional simulation method. PS3-RAM can also accurately estimate the STT-RAM write error rate and write energy distributions at both magnetic tunneling junction switching directions under different temperatures, demonstrating great potential in the analysis of STT-RAM reliability and write energy at the early design stage of memory or micro-architecture.

Index Terms—Process variation, reliability, spin-transfer torque random access memory (STT-RAM), statistical, thermal fluctuation, write energy.

I. INTRODUCTION

CONVENTIONAL memory technologies, i.e., SRAM, DRAM, and Flash, have achieved remarkable successes in modern computer industry. Following technology scaling, the shrunken feature size and the increased process variations impose serious power and reliability concerns on these technologies. In recent years, many emerging nonvolatile memory technologies have emerged above the horizon. As one promising candidate, spin-transfer torque random access memory (STT-RAM) has demonstrated great potentials in embedded memory and on-chip cache designs [1]–[6] through a good combination of the nonvolatility of Flash, the comparable cell density to DRAM, and the nanosecond programming time like SRAM.

In STT-RAM, the data is represented as the resistance state of a magnetic tunneling junction (MTJ) device. The MTJ resistance state can be programmed by applying a switching current, which is programmed into either of the two magnetic polarizations. Compared to the charge-based storage mechanism of conventional memories, the magnetic storage mechanism of STT-RAM shows less dependency on the device volume and hence, better scalability. Despite of these advantages, the unreliable write operation and high write energy are to be the major issues in STT-RAM designs. And these design metrics are significantly impacted by the prominent statistical factors of STT-RAM, including CMOS/MTJ device process variations under scaled technology and the probabilistic MTJ switching behaviors [7], [8]. In particular, the randomness of MTJ switching process incurred by the thermal fluctuations may generate the intermittent write failures of STT-RAM cells.

Many studies were performed to evaluate the impacts of process variations and thermal fluctuations on STT-RAM reliability [9]–[11]. The general evaluation method is as follows. First, Monte Carlo SPICE simulations are run extensively to characterize the distribution of the MTJ switching current $I$ during the STT-RAM write operations, by considering the device variations of both MTJ and MOS transistor. Then $I$ samples are sent into the macro-magnetic model to obtain the MTJ switching time ($\tau_{th}$) distributions under thermal fluctuations. Finally, the $\tau_{th}$ distributions of all $I$ samples are merged to generate the overall MTJ switching performance distribution. A write failure happens when the applied write pulse width is shorter than the needed $\tau_{th}$. Nonetheless, the costly Monte Carlo runs and the dependency on the macro-magnetic and SPICE simulations incur huge computation complexity of such a method, limiting the application of such a simulation method at the early stage STT-RAM design and optimization. Meanwhile, the modeling of write energy in STT-RAM was also studied extensively [12]. However, many such works only assume that the write energy of STT-RAM is deterministic and cannot successfully take into account its statistical characteristic induced by process variations and thermal fluctuations.

In this paper, we propose “PS3-RAM”—a fast, portable and scalable statistical STT-RAM reliability/energy analysis
method. PS3-RAM includes three integrated steps: 1) characterizing the MTJ switching current distribution under both MTJ and CMOS device variations; 2) recovering MTJ switching current samples from the characterized distributions in MTJ switching performance evaluation; and 3) performing the simulation on the thermal-induced MTJ switching variations based on the recovered MTJ switching current samples. Our major technical contributions are as follows.

1) We developed a sensitivity analysis technique to capture the statistical characteristics of the MTJ switching at scaled technology nodes. It achieves multiple orders-of-magnitude (＞10^5) run time cost reduction with marginal accuracy degradation, compared to SPICE-based Monte Carlo simulations.

2) We proposed using dual-exponential model for the fast and accurate recovery of MTJ switching current samples in statistical STT-RAM thermal analysis.

3) We released PS3-RAM from SPICE and macro-magnetic modeling and simulations, and extended its application into the array-level reliability analysis and the design space exploration of STT-RAM.

4) We introduced the concept of statistical write energy of STT-RAM and performed the statistical analysis on write energy by leveraging our PS3-RAM.

The remainder of our paper is organized as follows. Section II gives the preliminary of STT-RAM. Section III presents the details of PS3-RAM method. Section IV presents the application of our PS3-RAM on cell and array level reliability analysis and design space exploration. Section V shows the deterministic/statistical write energy analysis based on our PS3-RAM. Section VI discusses the computation complexity and Section VII concludes this paper.

II. PRELIMINARY

A. STT-RAM Basics

Fig. 1(c) shows the popular “one-transistor-one-MTJ (1T1J)” STT-RAM cell structure, which includes a MTJ and a nMOS transistor connected in series. In the MTJ, an oxide barrier layer (e.g., MgO) is sandwiched between two ferromagnetic layers. “0” and “1” are stored as the different resistances of the MTJ, respectively. When the magnetization directions of two ferromagnetic layers are parallel (anti-parallel), the MTJ is in its low (high) resistance state. Fig. 1(a) and (b) shows the low and the high MTJ resistance states, which are denoted by R_L and R_H, respectively. The MTJ switches from “0” to “1” when the switching current drives from reference layer to free layer, or from “1” to “0” when the switching current drives in the opposite direction.

B. Operation Errors of MTJ

In general, the MTJ switching time decreases when the switching current increases. A write failure happens when the MTJ switching does not complete before the switching current is removed. There are two reasons can cause this failure.

1) Persistent Errors: The current through the MTJ is affected by the process variations of both transistor and MTJ. For example, the driving ability of the nMOS transistor is subject to the variations of transistor channel length (L), width (W), and threshold voltage (V_th). The MTJ resistance variation also affects the nMOS transistor driving ability by changing its bias condition. The degraded MTJ switching current leads to a longer MTJ switching time and consequently, results in an incomplete MTJ switching before the write pulse ends. This kind of errors is referred to as “persistent” errors, which are mainly incurred by only device parametric variations. Persistent errors can be measured and repeated after the chip is fabricated.

2) Nonpersistent Errors: Another kind of errors is called “nonpersistent” errors, which happen intermittently and may not be repeated. The nonpersistent errors of STT-RAM are mostly caused by the intrinsic thermal fluctuations during MTJ switching [13]. Due to thermal fluctuations, the MTJ switching time will not be a constant value but rather a distribution even under a constant switching current.

III. PS3-RAM METHOD

Fig. 2 depicts the overview of our proposed PS3-RAM method, mainly including variation-aware STT-RAM cell library construction and array level analysis. The key features of our PS3-RAM method can be summarized as the sensitivity analysis for MTJ switching current (I) characterization, the switching current samples recovery, and the statistical thermal analysis of STT-RAM. In cell library construction, the first step is to configure it by inputting both the nominal design parameters and their corresponding variation statistics, like the channel length/width/threshold voltage of nMOS transistor, as well as the geometry parameters of MTJ device, such as its thickness and area. Then a multidimensional sensitivity analysis will be conducted to characterize the statistical properties of I, followed by an advanced filtering technology—smooth filter, to improve its accuracy by adaptively adjusting its resolution. After that, the write current samples can be quickly recovered based on the above characterized statistics, current distribution model and model parameter estimation equations. The write pulse distribution will be generated after mapping the switching current samples to the write pulse samples by considering the thermal fluctuations. Finally, the statistical write energy analysis and the STT-RAM cell write error rate can be performed.
TABLE I  
SIMULATION PARAMETERS AND ENVIRONMENT SETTING

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel length</td>
<td>$L = 45,\text{nm}$</td>
<td>$\sigma_L = 0.05,L$</td>
</tr>
<tr>
<td>Channel width</td>
<td>$W = 90 \sim 1800,\text{nm}$</td>
<td>$\sigma_W = 0.05,L$</td>
</tr>
<tr>
<td>Threshold voltage</td>
<td>$V_{th} = 0.462,\text{V}$</td>
<td>by eq. (2)</td>
</tr>
<tr>
<td>Mgo thickness</td>
<td>$\varphi = 2.2,\text{nm}$</td>
<td>$\sigma_\varphi = 0.02\varphi$</td>
</tr>
<tr>
<td>MTJ area</td>
<td>$A = 45 \times 90,\text{nm}^2$</td>
<td>by width/length</td>
</tr>
<tr>
<td>Resistance low</td>
<td>$R_L = 100,\Omega\text{f}$</td>
<td>by eq. (4)</td>
</tr>
<tr>
<td>Resistance high</td>
<td>$R_H = 200,\Omega\text{f}$</td>
<td>by eq. (4)</td>
</tr>
</tbody>
</table>

A. Sensitivity Analysis on MTJ Switching

In this section, we present our sensitivity model used for the characterization of the MTJ switching current distribution. We then analyze the contributions of different variation sources to the distribution of the MTJ switching current in details. The definitions of the variables used in our analysis are summarized in Table I.

1) Threshold Voltage Variation: The variations of channel length, width, and threshold voltage are three major factors causing the variations of transistor driving ability. $V_{th}$ variation mainly comes from random dopant fluctuation (RDF) and line-edge roughness (LER), the latter of which is also the source of some geometry variations (i.e., $L$ and $W$) [14], [15]. It is known that the $V_{th}$ variation is also correlated with $L$ and $W$ and its variance decreases when the transistor size increases. The deviation of the $V_{th}$ from the nominal value following the change of $L$ ($\Delta L$) can be modeled by [15]:

$$\Delta V_{th} = \Delta V_{th0} + V_{th}\exp\left(-\frac{L}{L'}\right) \cdot \frac{\Delta L}{L'}.$$  

(1)

Then the standard deviation of $V_{th}$ can be calculated as

$$\sigma_{V_{th}}^2 = \frac{C_1}{WL} + \frac{C_2}{\exp\left(L/L'\right)} \cdot \frac{W_c}{W} \cdot \sigma_L^2.$$  

(2)

Here $W_c$ is the correlation length of nonrectangular gate (NRG) effect, which is caused by the randomness in sub-wavelength lithography. $C_1$, $C_2$, and $L$ are technology dependent coefficients. The first term in (2) describes the RDFs contribution to $\sigma_{V_{th}}$. The second term in (2) represents the contribution from NRG, which is heavily dependent on $L$ and $W$. Following technology scaling, the contribution of this term becomes prominent due to the reduction of $L$ and $W$.

2) Sensitivity Analysis on Variations: Although the contributions of MTJ and MOS transistor parametric variabilities to the MTJ switching current distribution cannot be explicitly
expressed, it is still possible for us to conduct a sensitivity analysis to obtain the critical characteristics of the distribution. Without loss of generality, the MTJ switching current $I$ can be modeled by a function of $W$, $L$, $V_{th}$, $A$, and $\tau$. $A$ and $\tau$ are the MTJ surface area and MgO layer thickness, respectively. The 1st-order Taylor expansion of $I$ around the mean values of every parameter is

$$I(W, L, V_{th}, A, \tau) \approx I(\bar{W}, \bar{L}, \bar{V}_{th}, \bar{A}, \bar{\tau}) + \frac{\partial I}{\partial W}(W - \bar{W})$$

$$+ \frac{\partial I}{\partial L}(L - \bar{L}) + \frac{\partial I}{\partial V_{th}}(V_{th} - \bar{V}_{th})$$

$$+ \frac{\partial I}{\partial A}(A - \bar{A}) + \frac{\partial I}{\partial \tau}(\tau - \bar{\tau}).$$

(3)

Here $W$, $L$, and $\tau$ generally follow Gaussian distribution $[9]$. $A$ is the product of two independent Gaussian distributions, $V_{th}$ is correlated with $W$ and $L$, as shown in (1) and (2). Because the MTJ resistance $R \propto \exp (e^r)/A$ $[9]$, we have

$$\frac{\partial I}{\partial A} \Delta A + \frac{\partial I}{\partial \tau} \Delta \tau = \left(\frac{\partial I}{\partial R}\right)_{\bar{R}} \Delta A + \left(\frac{\partial I}{\partial \tau}\right)_{\bar{\tau}} \Delta \tau$$

$$= \left(\frac{\partial I}{\partial R}\right)_{\bar{R}} \Delta R.$$  

(4)

Equation (4) indicates that the combined contribution of $A$ and $\tau$ is the same as the impact of MTJ resistance. The difference between the actual $I$ and its mathematical expectation $\mu_I$ can be calculated by

$$I(W, L, V_{th}, R) - E(I(\bar{W}, \bar{L}, \bar{V}_{th}, \bar{R})) \approx \frac{\partial I}{\partial W} \Delta W + \frac{\partial I}{\partial L} \Delta L + \frac{\partial I}{\partial V_{th}} \Delta V_{th} + \frac{\partial I}{\partial R} \Delta R.$$  

(5)

Here we assume $\mu_I \approx E(I(\bar{W}, \bar{L}, \bar{V}_{th}, \bar{R})) = I(\bar{W}, \bar{L}, \bar{V}_{th}, \bar{R})$ and the mean of MTJ resistance $\bar{R} \approx R(\bar{A}, \bar{\tau})$. Combining (1), (2), and (5), the standard deviation of $I$ ($\sigma_I$) can be calculated as

$$\sigma^2_I = \left(\frac{\partial I}{\partial W}\right)^2 \sigma^2_W + \left(\frac{\partial I}{\partial L}\right)^2 \sigma^2_L + \left(\frac{\partial I}{\partial V_{th}}\right)^2 \sigma^2_{V_{th}}$$

$$+ \left(\frac{\partial I}{\partial R}\right)^2 \sigma^2_R$$

$$+ 2 \frac{\partial I}{\partial V_{th}} \frac{\partial I}{\partial W} \rho_1 \sqrt{\frac{C_1}{W_L}} \sigma_L + 2 \frac{\partial I}{\partial W} \frac{\partial I}{\partial V_{th}} \rho_2 \sqrt{\frac{C_1}{W_L}} \sigma_W$$

$$+ 2 \frac{\partial I}{\partial L} \frac{\partial I}{\partial V_{th}} \rho_3 \sqrt{\frac{C_1}{W_L}} \sigma_W.$$  

(6)

Here $\rho_1 = \text{cov}(V_{th0}, L)/(\sqrt{\sigma_{V_{th0}}^2 \sigma^2_L})$ and $\rho_2 = \text{cov}(V_{th0}, W)/(\sqrt{\sigma^2_{V_{th0}} \sigma^2_W})$ are the correlation coefficients between $V_{th0}$ and $L$ or $W$, respectively $[15]$, $\sigma^2_{V_{th0}} = C_1/W_L$. Our further analysis shows that the last three terms at the right side of (6) are significantly smaller than other terms and can be safely ignored in the simulations of STT-RAM normal operations.

The accuracy of the coefficient in front of the variances of every parameter at the right side of (6) can be improved by applying window based smooth filtering. Take $W$ as an example, we have

$$\left(\frac{\partial I}{\partial W}\right)_i = \frac{I(W + i\Delta W, L, V_{th}, R) - I(W - i\Delta W, L, V_{th}, R)}{2i\Delta W}.$$  

(7)

Here $i = 1, 2, \ldots, K$. Different ($\partial I/\partial W$) can be obtained at the different step $i$. $K$ samples can be filtered out by a windows based smooth filter to balance the accuracy and the computation complexity as

$$\frac{\partial I}{\partial W} = \sum_{i=1}^{K} \omega_i \left(\frac{\partial I}{\partial W}\right)_i.$$  

(8)

Here $\omega_i$ is the weight of sample $i$, which is determined by the window type, i.e., hammering window or rectangular window $[16]$.

3) Variation Contribution Analysis: The variations’ contributions to $I$ are mainly represented by the first four terms at the right side of (6) as

$$S_1 = \left(\frac{\partial I}{\partial W}\right)^2 \sigma^2_W, S_2 = \left(\frac{\partial I}{\partial L}\right)^2 \sigma^2_L, S_3 = \left(\frac{\partial I}{\partial V_{th}}\right)^2 \sigma^2_{V_{th}},$$

$$S_4 = \left(\frac{\partial I}{\partial R}\right)^2 \left(\frac{C_1}{W_L} + \frac{C_2}{\exp(L/\ell)} \cdot \frac{W c}{W} \cdot \sigma_R^2 \right).$$  

(9)

As pointed out by many prior-arts $[17]$, an asymmetry exists in STT-RAM write operations: the switching time of “0”→“1” is longer than that of “1”→“0” and suffers from a larger variance. Also, the switching time variance of “0”→“1” is more sensitive to the transistor size changes than “1”→“0.” As we shall show later, this phenomena can be well explained by using our sensitivity analysis. To the best of our knowledge, this is the first time the asymmetric variations of STT-RAM write performance and their dependencies on the transistor size are explained and quantitatively analyzed.

As shown in Fig. 1, when writing “0,” the word-line (WL) and bit-line (BL) are connected to $V_{dd}$ while the source-line (SL) is connected to ground. $V_{gs} = V_{dd}$ and $V_{ds} = V_{dd} - IR$. The nMOS transistor is mainly working in triode region. Based on short-channel BSIM model, the MTJ switching current supplied by a nMOS transistor can be calculated by

$$I = \beta \cdot \left(\frac{(V_{dd} - V_{th})}{(V_{dd} - IR) - \frac{U_0}{2}(V_{dd} - IR)^2}\right) \cdot \left(1 + \frac{1}{\text{vref}} \cdot \frac{V_{dd} - IR}{V_{dd} - IR}\right).$$  

(10)

Here $\beta = \mu_0C_{ox}/(1 + U_0 (V_{dd} - V_{th}))(W/L)$. $U_0$ is the vertical field mobility reduction coefficient, $\mu_0$ is electron mobility, $C_{ox}$ is gate oxide capacitance per unit area, $a$ is body-effect coefficient and $v_{sat}$ is carrier velocity saturation. Based on short-channel PTM model $[20]$ and BSIM model $[21]$, $[22]$, we derive ($\partial I/\partial W)^2$, ($\partial I/\partial L)^2$, ($\partial I/\partial R)^2$, and ($\partial I/\partial V_{th})^2$ as

$$(\frac{\partial I}{\partial W})^2 \approx \frac{1}{(A_1 W + B_1)^4}, \quad (\frac{\partial I}{\partial L})^2 \approx \frac{1}{(A_2 W + B_2 W + C)^2},$$

$$(\frac{\partial I}{\partial R})^2 \approx \frac{1}{(A_3 W + B_3)^4}, \quad (\frac{\partial I}{\partial V_{th}})^2 \approx \frac{1}{(A_4 W + B_4 \sqrt{W})^4}.$$
Our analytical deduction shows that the coefficients $A_{1-4}$, $B_{1-4}$ and $C$ are solely determined by $W$, $L$, $V_{th}$, and $R$. The detailed expressions of coefficients $A_{1-4}$, $B_{1-4}$ and $C$ can be found in the appendix. Here $R$ is the high resistance state of the MTJ, or $R_L$. For a nMOS transistor at “0”→“1” switching, the MTJ switching current is

$$I = \frac{\beta}{2a} \left[ (V_{dd} - IR - V_{th}) - \frac{I}{WC_{ox}^{2}\sigma_a^2} \right]^2.$$  

(11)

Here $R$ is the low resistance state of the MTJ, or $R_L$. We have

$$\left( \frac{\partial I}{\partial W} \right)_{1}^{2} \approx \frac{1}{(A_{5}W + B_{5})^4}, \quad \left( \frac{\partial I}{\partial L} \right)_{1}^{2} \approx \frac{1}{(A_{6}W + B_{6})^2} \quad (12)$$

$$\left( \frac{\partial I}{\partial R} \right)_{1}^{2} \approx \frac{1}{(A_{7}W + B_{7})^4}, \quad \left( \frac{\partial I}{\partial V_{th}} \right)_{1}^{2} \approx \frac{1}{(A_{8}W + B_{8})^2}.$$  

Again, $A_{5-8}$ and $B_{5-8}$ can be expressed as the function of $W$, $L$, $V_{th}$, and $R$ and the detailed expressions of those parameters can be found in the appendix.

In general, a large $S_1$ corresponds to a large contribution to $I$ variation, where $W$ is approaching infinity, only $S_3$ is nonzero at “1”→“0” switching while both $S_2$ and $S_3$ are nonzero at “0”→“1” switching. It indicates that the residual values of $S_1 - S_4$ at “0”→“1” switching is larger than that at “1”→“0” switching when $W \rightarrow \infty$. In other words, “0”→“1” switching suffers from a larger MTJ switching current variation than “1”→“0” switching when nMOS transistor size is large.

4) Simulation Results of Sensitivity Analysis: Sensitivity analysis [23] can be used to obtain the statistical parameters of MTJ switching current, i.e., the mean and the standard deviation, without running the costly SPICE and Monte Carlo simulations. It can be also used to analyze the contributions of different variation sources to $I$ variation in details. The normalized contributions ($P_i$) of variation resources, i.e., $W$, $L$, $V_{th}$, and $R$, are defined as

$$P_i = \frac{S_i}{\sum_{i=1}^{4} S_i}, i = 1, 2, 3, 4.$$  

(12)

Figs. 3 and 4 show the normalized contributions of every variation source at “0”→“1” and “1”→“0” switching’s, respectively, at different transistor sizes. We can see that $L$ and

B. Write Current Distribution Recovery

After the $I$ distribution is characterized by the sensitivity analysis, the next question becomes how to recover the distribution of $I$ from the characterized information in the statistical analysis of STT-RAM reliability. We investigated the typical distributions of $I$ in various STT-RAM cell designs and found that dual-exponential function can provide the excellent accuracy in modeling and recovering these distributions. The dual-exponential function we used to recover the $I$ distributions can be illustrated as

$$f(I) = \begin{cases} a_1 e^{b_1(l-u)} & I \leq u \\ a_2 e^{b_2(u-I)} & I > u. \end{cases}$$  

(13)

Here $a_1$, $b_1$, $a_2$, $b_2$, and $u$ are the fitting parameters, which can be calculated by matching the first and the second order moments of the actual $I$ distribution and the dual-exponential function as

$$\begin{align*}
    \int f(I)dl &= 1 \\
    \int I f(I)dl &= E(I) \\
    \int I^2 f(I)dl &= E(I^2) = \sigma_I^2.
\end{align*}$$  

(14)

Here $E(I)$ and $\sigma_I^2$ are obtained from the sensitivity analysis. The recovered $f$ distribution can be used to generate the MTJ switching current samples, as shown in Fig. 5. At the beginning of the sample generation flow, the confidence interval for STT-RAM design is determined, e.g., $[\mu_1 - 6\sigma_1, \mu_1 + 6\sigma_1]$ for a six-sigma confidence interval. Assuming we need to generate $N$ samples within the confidence interval, say, at the point of $I = I_i$, a switching current sequence of $[NP_{I_i}]$ samples must be generated. Here $P_{I_i} \approx f(I_i) \Delta$, $\Delta$ equals $(12\sigma_I/N)$, or the step of sampling generation. $f(I_i)$ is the dual-exponential function.
C. Statistical Thermal Analysis

The variation of the MTJ switching time (τth) incurred by the thermal fluctuations follows Gaussian distribution when τth is below 10 ∼ 20 ns [17]. In this range, the distribution of τth can be easily constructed after the I is determined. The distribution of MTJ switching performance can be obtained by combining the τth distributions of all I samples.

IV. APPLICATION 1: WRITE RELIABILITY ANALYSIS

In this section, we conduct the statistical analysis on the write reliability of STT-RAM cells by leveraging our PS3-RAM method. Both device variations and thermal fluctuations are considered in the analysis. We also extend our method into array-level evaluation and demonstrate its effectiveness in STT-RAM design optimizations.

A. Reliability Analysis of STT-RAM Cells

The write failure rate PWF of a STT-RAM cell can be defined as the probability that the actual MTJ switching time τth is longer than the write pulse width Tw, or PWF = P(τth > Tw). τth is affected by the MTJ switching current magnitude, the MTJ and MOS device variations, the MTJ switching direction, and the thermal fluctuations. The conventional simulation of PWF requires costly Monte Carlo runs with hybrid SPICE and macro-magnetic modeling steps. Instead, we can use PS3-RAM to analyze the statistical STT-RAM write performance. The corresponding simulation environment is also summarized in Table I.

Figs. 9 and 10 depict the PWF’s simulated by PS3-RAM for both switching directions at 300 K. For comparison purpose, the Monte Carlo simulation results are also presented. Different Tw’s are selected at either switching directions due to the asymmetric MTJ switching performances [17].

Note that N determines both the analysis granularity and the level of the estimated error rate.

Fig. 6 shows the relative errors of the mean and the standard deviation of the recovered I distribution with respect to the results directly from the sensitivity analysis [see (5) and (6)]. The maximum relative error < 10^{-2}, which proves the accuracy of our dual-exponential model.

Figs. 7 and 8 compare the probability distribution functions (PDFs) of I from the SPICE Monte Carlo simulations and from the recovery process based on our sensitivity analysis at two switching directions. Our method achieves good accuracy at both representative transistor channel widths (W = 90 nm or W = 720 nm).

C. Statistical Thermal Analysis

The variation of the MTJ switching time (τth) incurred by the thermal fluctuations follows Gaussian distribution when


i.e., $T_w = 10, 15, 20$ ns at “0”→“1” and $T_w = 6, 8, 10, 12$ ns at “1”→“0.” Our PS3-RAM results are in excellent agreement with the ones from Monte Carlo simulations.

Since “0”→“1” is the limiting switching direction for STT-RAM reliability, we also compare the $P_{WF}$’s of different STT-RAM cell designs under different temperatures at this switching direction in Fig. 11. The results show that PS3-RAM can provide very close but pessimistic results compared to those of the conventional simulations. PS3-RAM is also capable to precisely capture the small error rate change incurred by a moderate temperature shift (from $T = 300$ K to $T = 325$ K).

It is known that prolonging the write pulse width and increasing the MTJ switching current (by sizing up the nMOS transistor) can reduce the $P_{WF}$. In Fig. 12, we demonstrate an example of using PS3-RAM to explore the STT-RAM design space: the tradeoff curves between $P_{WF}$ and $T_w$ are simulated at different $W$’s. For a given $P_{WF}$, for example, the corresponding tradeoff between $W$ and $T_w$ can be easily identified on Fig. 12.

**B. Array Level Analysis and Design Optimization**

We use a 45 nm 256 Mb STT-RAM design [24] as the example to demonstrate how to extend our PS3-RAM into array-level analysis and design optimizations. The number of bits per memory block $N_{bit} = 256$ and the number of memory blocks $N_{word} = 1 M$. ECC is applied to correct the random write failures of memory cells. Two types of ECCs with different implementation costs are being considered, i.e., single-bit-correcting hamming code and a set of multibits-correcting Bose-Chaudhuri-Hocquenghem (BCH) codes. We use $(n, k, t)$ to denote an ECC with $n$ codeword length, $k$ bit user bits being protected ($256$ bit here) and $t$ bits being corrected. The ECCs corresponding to the error correction capability $t$ from 1 to 5 are hamming code ($265$, $256$, 1) and four BCH codes—BCH1 ($274$, $256$, 2), BCH2 ($283$, $256$, 3), BCH3 ($292$, $256$, 4) and BCH4 ($301$, $256$, 5), respectively.

The write yield of the memory array $Y_{wr}$ can be defined as

$$Y_{wr} = P(n_e \leq t) = \sum_{i=0}^{t} C_n^i P_{WF}^i (1 - P_{WF})^{n-i}. \quad (15)$$

Here, $n_e$ denotes the total number of error bits in a write access. $Y_{wr}$ indeed denotes the probability that the number of error bits in a write access is smaller than the error correction capability.

Fig. 13 depicts the $Y_{wr}$’s under different combinations of ECC scheme and $W$ when $T_w = 15$ ns at “0”→“1” switching. The ECC schemes required to satisfy $\sim 100\%$ $Y_{wr}$ for different $W$ are: 1) hamming code for $W = 630$ nm; 2) BCH2 for $W = 540$ nm; and 3) BCH4 for $W = 480$ nm. The total memory array area can be estimated by using the STT-RAM cell size equation $A_{cell} = 3 (W/L + 1) (F^2)$ [25]. Calculation shows that combination 3) offers us the smallest STT-RAM array area, which is only 88% and 95% of the ones of 1) and 2), respectively. We note that PS3-RAM can be seamlessly embedded into the existing deterministic memory macro
models [25] for the extended capability on the statistical reliability analysis and the multidimensional design optimizations on area, yield, performance and energy. Fig. 14 illustrates the STT-RAM design space in terms of the combinations of $Y_{wr}$, $W$, $T_{sw}$, and ECC scheme. After the pair of $(Y_{wr}, T_{w})$ is determined, the tradeoff between $W$ and ECC can be found in the corresponding region on the figure. The result shows that PS3-RAM provides a fast and efficient method to perform the device/circuit/architecture co-optimization for STT-RAM designs.

V. APPLICATION 2: WRITE ENERGY ANALYSIS

In addition to write reliability analysis, our PS3-RAM method can also precisely capture the write energy distributions influenced by the variations of device and working environment. In this section, we first prove that there is a sweet point of write pulse width for the minimum write energy without considering any variations. Then we introduce the concept of statistical write energy of STT-RAM cells considering both process variations and thermal fluctuations, and perform the statistical analysis on write energy using our PS3-RAM method.

A. Write Energy Without Variations

The write energy of a STT-RAM cell during each programming cycle without considering process and thermal variations is deterministic and can be modeled by (16) as

$$E_{av} = I^2 R \tau_{th}. \quad (16)$$

Here $I$ denotes the switching current at either "0"→"1" or "1"→"0" switching, $\tau_{th}$ is the corresponding MTJ switching time and $R$ is the MTJ resistance value, i.e., $R_{th}$ ($R_H$) for "0"→"1" ("1"→"0") switching. As discussed in prior art [17], the switching process of an STT-RAM cell can be divided into three working regions

$$I = \begin{cases} 
I_{th} \left(1 - \frac{\ln(\tau_{th}/\tau_0)}{\Delta}\right), & \tau_{th} > 10 \text{ ns} \\
I_{th} + C \ln \left(\pi \frac{\tau_{th}}{2\theta}\right)/\tau_{th}, & \tau_{th} < 3 \text{ ns} \\
P_{th} + Q, & 3 \leq \tau_{th} \leq 10 \text{ ns}.
\end{cases} \quad (17)$$

Here $I_{th}$ is the critical switching current, $\Delta$ is thermal stability, $\tau_0 = 1$ ns is the relax time, $\theta$ is the initial angle between the magnetization vector and the easy axis, $C, P, Q$ are fitting parameters.

For a relatively long switching time range ($\tau_{th} \approx 10 \sim 300$ ns), the undistorted write energy $E_{av}$ can be calculated as

$$E_{av} = I_{th}^2 \left(1 - \frac{\ln(\tau_{th})}{\Delta}\right)^2 R \tau_{th}$$

$$= \frac{I_{th}^2 R}{\Delta^2} (\Delta - \ln \tau_{th})^2 \tau_{th}. \quad (18)$$

In the long switching time range, we have $\ln \tau_{th} < 0$. Thus, $(\Delta - \ln \tau_{th})^2 \tau_{th}$ or $E_{av}$ monotonically raises as the write pulse $\tau_{th}$ increases and the minimized write energy $E_{av}$ occurs at $\tau_{th} = 10$ ns.

In the ultrashort switching time range ($\tau_{th} < 3$ ns), $E_{av}$ can be obtained as

$$E_{av} = \left[I_{th} + C \ln \left(\frac{\pi}{2\theta}\right)/\tau_{th}\right]^2 R \tau_{th}$$

$$= 2I_{th}R C \ln \left(\pi \frac{\tau_{th}}{2\theta}\right) + \frac{C^2 \ln^2 \left(\pi / 2\theta\right) R}{\tau_{th}}$$

$$\geq 2I_{th}R C \ln \left(\frac{\pi}{2\theta}\right) + 2 \sqrt{\frac{I_{th}^2 R^2 C^2 \ln^2 \left(\pi / 2\theta\right)}{\tau_{th}}}$$

$$\geq 4I_{th}R C \ln \left(\frac{\pi}{2\theta}\right). \quad (19)$$

As (19) shows, the minimum of $E_{av}$ can be achieved when $\tau_{th} = (C \ln (\pi/2\theta))/I_{th}$. However, for the ultrashort switching time range (usually $(C \ln (\pi/2\theta))/I_{th} > 3$ ns), $E_{av}$ monotonically decreases as $\tau_{th}$ increases.

Similarly, in the middle switching time range ($3 \leq \tau_{th} \leq 10$ ns), $E_{av}$ can be expressed as

$$E_{av} = \left(\frac{P}{\tau_{th}} + Q\right)^2 R \tau_{th}$$

$$= \left(\frac{P}{\sqrt{\tau_{th}}} + Q\sqrt{\tau_{th}}\right)^2 R$$

$$\geq 4PQR. \quad (20)$$

Again, the minimized $E_{av}$ occurs at $\tau_{th} = (P/Q)$. Here $(P/Q) \geq 10$ ns based on our device parameters characterization [17]. Thus, the write energy $E_{av}$ in this range monotonically decreases as $\tau_{th}$ grows.

According to the monotonicity of $E_{av}$ in the three regions, the most energy-efficient switching point of $E_{av}$ should be at
Fig. 16. Average write energy versus write pulse width under different temperature.

\[ \tau_{th} = 10 \text{ ns}. \] To validate above theoretical deduction for the sweet point of \( E_{av} \), we also conduct the SPICE simulations. Here the STT-RAM device model without considering process and thermal variations is also adopted from [17].

Fig. 15 shows the simulated write energy \( E_{av} \) over different write pulse at \( "0" \rightarrow "1" \) switching. As Fig. 15 shows, \( E_{av} \) monotonically decreases in the ultrashort switching range and continues decreasing in the middle range, but becomes monotonically increasing after entering the long switching time range. The sweet point of \( E_{av} \) occurs around \( \tau_{th} = 10 \text{ ns} \), which validates our theoretical analysis for the write energy without considering any variations.

We also present the simulated \( E_{av} - \tau_{th} \) curve under different temperatures in Fig. 16. The trend and sweet point of \( E_{av} - \tau_{th} \) curves remain almost the same when the temperature increases from \( T = 300 \text{ K} \) to \( T = 400 \text{ K} \). In fact, the write energy \( E_{av} \) decreases a little bit as the temperature increases. The reason is that the driving ability loss of the nMOS transistor (\( I \)) dominates \( E_{av} \) though the MTJ switching time (\( \tau_{th} \)) increases when the working temperature raises.

B. PS3-RAM for Statistical Write Energy

As discussed in Section V-A, the write energy of a STT-RAM cell can be deterministically optimized when all the variations are ignored. However, since the switching current \( I \), the resistance \( R \), and the switching time \( \tau_{th} \) in (16) may be distorted by CMOS/MTJ process variations and thermal fluctuations, the deterministic value will no longer be able to represent the statistic nature of the write energy of a STT-RAM cell. Accordingly, the optimized write energy at sweet point (\( \tau_{th} = 10 \text{ ns} \)) shown in Fig. 15 should be expanded as a distribution.

Similar to the write failure analysis in Section IV, we conduct the statistical write energy analysis using our PS3-RAM method. We choose the mean of nMOS transistor width \( W = 540 \text{ nm} \). The remained device parameters and variation configurations keep the same as Table I.

Figs. 17 and 18 show the simulated statistical write energy by PS3-RAM for both switching directions at 300 K. For comparison, the SPICE simulation results are also presented. As shown in the figures, the distribution of write energy captured by our PS3-RAM method are in excellent agreement with the results from SPICE simulations at both \( "1" \rightarrow "0" \) and \( "0" \rightarrow "1" \) switching’s.

VI. COMPUTATION COMPLEXITY EVALUATION

We compared the computation complexity of our proposed PS3-RAM method with the conventional simulation method. Suppose the number of variation sources is \( M \), for a statistical analysis of a STT-RAM cell design, the numbers of SPICE simulations required by conventional flow and PS3-RAM are \( N_{std} = N_{s}^M \) and \( N_{PS3-RAM} = 2K + 1 \), respectively. Here \( K \) denotes the sample numbers for window based smooth filter in sensitivity analysis, \( N_{s} \) is average sample number of every variation in the Monte Carlo simulations in conventional method, \( K \ll N_{s} \). Note that our switching current sample recovery flow does not require any extra Monte Carlo simulations. The speedup \( X_{\text{speedup}} \approx (N_{s}^M / 2KM) \) can be up to multiple orders of magnitude: for example, if we set \( N_{s} = 100 \), \( M = 4 \), (note: \( V_{th} \) is not an independent variable) and \( K = 50 \), the speed up is around \( 2.5 \times 10^5 \).

VII. CONCLUSION

We developed a fast and scalable statistical STT-RAM reliability/energy analysis method called PS3-RAM. PS3-RAM can simulate the impact of process variations and thermal fluctuations on the statistical STT-RAM write performance or write energy distributions, without running costly Monte Carlo simulations on SPICE and macro-magnetic models. Simulation results show that PS3-RAM can achieve very high

\[ \text{Fig. 17. Statistical write energy versus write pulse width at "1"→"0."} \]

\[ \text{Fig. 18. Statistical write energy versus write pulse width at "0"→"1."} \]
TABLE II
PARAMETER DEFINITION

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( U_0 )</td>
<td>Vertical field mobility reduction coefficient</td>
</tr>
<tr>
<td>( \mu_0 )</td>
<td>Electron mobility</td>
</tr>
<tr>
<td>( C_{ox} )</td>
<td>Gate oxide capacitance per unit area</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>Body-effect coefficient</td>
</tr>
<tr>
<td>( \nu_{sat} )</td>
<td>Carrier velocity saturation</td>
</tr>
</tbody>
</table>

...accuracy compared to the conventional simulation method, while achieving a speedup of multiple orders of magnitude. The great potentials of PS3-RAM in the application of the device/circuit/architecture co-optimization of STT-RAM designs are also demonstrated.

APPENDIX

In this appendix, we give the details on the model deduction in sensitivity analysis and the summary of the analytic results involved in the PS3-RAM development. We also present the validation of our analytic results based on Monte Carlo simulations. Table II [21] summarizes some additional parameters used in this section.

A. Sensitivity Analysis Model Deduction

The sensitivity analysis model is developed based on the electrical MTJ model and the simplified BSIM model [21], [22]. At “1”→“0” switching, the MTJ switching current supplied by an nMOS transistor working in the triode region is

\[
I = \frac{\beta}{2a} \left[ (V_{dd} - V_{th}) (V_{dd} - IR) - \frac{1}{2} (V_{dd} - IR)^2 \right], \tag{21}
\]

Here \( \beta = (\mu_0 C_{ox})/(1 + U_0 (V_{dd} - V_{th}))(W/L) \). As summarized in Table II, \( U_0 \) is the vertical field mobility reduction coefficient, \( \mu_0 \) is electron mobility, \( C_{ox} \) is gate oxide capacitance per unit area, \( \alpha \) is body-effect coefficient and \( \nu_{sat} \) is carrier velocity saturation. The MTJ is in its high resistance state, or \( R = R_H \). Based on PTM [20] and BSIM [21], the partial derivatives in (5) can be calculated by ignoring the minor terms in the expansion of (21) as

\[
\left( \frac{\partial I}{\partial W} \right)_0 \approx \frac{1}{(A_1 W + B_3)^4}, \quad \left( \frac{\partial I}{\partial L} \right)_0 \approx \frac{1}{(A_3 W + B_3 W + C)^2} \tag{22}
\]

\[
\left( \frac{\partial I}{\partial R} \right)_0 \approx \frac{1}{(A_4 W + B_3)^4}, \quad \left( \frac{\partial I}{\partial V_{th}} \right)_0 \approx \frac{1}{(A_5 \sqrt{W} + B_3 \sqrt{W})^4}.
\]

Here

\[
A_1 = \sqrt{\frac{\mu_0 C_{ox} V_{dd} (V_{dd} - V_{th})}{L}} \cdot \frac{1}{R} \quad B_1 = \sqrt{\frac{\mu_0 C_{ox} V_{dd} (V_{dd} - V_{th})}{L}} \quad A_2 = \frac{\mu_0 C_{ox} V_{dd} (V_{dd} - V_{th})}{L^2}
\]

At “0”→“1” switching, the nMOS transistor is working in the saturation region. The current through the MTJ is

\[
I = \frac{\beta}{2a} \left[ (V_{dd} - IR - V_{th}) - \frac{I}{WC_{ox}\nu_{sat}^2} \right]^2. \tag{22}
\]

The MTJ is in its low resistance state, or \( R = R_L \). The derivatives can be also calculated as

\[
\left( \frac{\partial I}{\partial W} \right)_1 \approx \frac{1}{(A_3 W + B_3)^4}, \quad \left( \frac{\partial I}{\partial L} \right)_1 \approx \frac{1}{(A_3 W + B_3 W + C)^2} \tag{22}
\]

\[
\left( \frac{\partial I}{\partial R} \right)_1 \approx \frac{1}{(A_4 W + B_3)^4}, \quad \left( \frac{\partial I}{\partial V_{th}} \right)_1 \approx \frac{1}{(A_5 \sqrt{W} + B_3 \sqrt{W})^4}.
\]

by ignoring the minor terms in the expansion of (22). Here

\[
A_5 = \sqrt{\frac{2C_{ox} \nu_{sat} \mu_0}{La + \mu_0 (V_{dd} - V_{th})} \cdot \frac{1}{R}} \quad B_5 = \frac{2C_{ox} \nu_{sat} |La + \mu_0 (V_{dd} - V_{th})|}{\mu_0} \quad A_6 = \frac{2a C_{ox} \nu_{sat}^2}{R \mu_0} \quad B_6 = \frac{R \mu_0}{\nu_{sat}} \quad A_7 = \sqrt{\frac{2C_{ox} \nu_{sat}}{La v_{sat} + \mu_0 (V_{dd} - V_{th})} \cdot \frac{1}{R}} \quad B_7 = \sqrt{\frac{2C_{ox} \nu_{sat}}{La v_{sat} + \mu_0 (V_{dd} - V_{th})} \cdot \frac{1}{R}}
\]

The contributions of different variation sources to \( I \) are represented by

\[
S_1 = \left( \frac{\partial I}{\partial W} \right)^2 \sigma_W^2, \quad S_2 = \left( \frac{\partial I}{\partial L} \right)^2 \sigma_L^2, \quad S_3 = \left( \frac{\partial I}{\partial R} \right)^2 \sigma_R^2 \quad S_4 = \left( \frac{\partial I}{\partial V_{th}} \right)^2 \left( C_1 \frac{W}{L} + C_2 \frac{W}{L} \exp \left( \frac{L}{I} \right) \cdot \frac{W}{L} \right) \sigma_L^2.
\]

Here \( S_1, S_2, S_3, \) and \( S_4 \) denote the variations induced by \( W, \) \( L, R \) (\( R_H \) or \( R_L \)), and \( V_{th} \), respectively.
TABLE III
SUMMARY OF VARIATION CONTRIBUTION

<table>
<thead>
<tr>
<th>Variation</th>
<th>Monotony</th>
<th>( \min S_1 = 0 )</th>
<th>( \max S_2 = \left( \frac{V_{dd}}{R_H} \sigma L \right)^2 )</th>
<th>( \max S_3 )</th>
<th>( \max S_4 = \left( \frac{V_{dd} - V_{th}}{V_{dd} - V_{th}} \sigma R_H \right)^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \downarrow )</td>
<td>( \frac{W}{ \infty} \rightarrow 0 )</td>
<td>( \frac{W}{ \infty} \rightarrow 0 )</td>
<td>( W \rightarrow \infty )</td>
<td>( W \rightarrow \infty )</td>
</tr>
<tr>
<td>( S_1 )</td>
<td>( \downarrow )</td>
<td>( S_2 )</td>
<td>( S_3 )</td>
<td>( S_4 )</td>
<td></td>
</tr>
</tbody>
</table>

\( S_1 \) \( \downarrow \) \( \max S_2 = \left( \frac{V_{dd}}{R_H} \sigma L \right)^2 \) \( W \rightarrow \infty \) \( S_1 \rightarrow 0 \)

\( S_2 \) \( \uparrow \downarrow \) \( \frac{W}{ \infty} \rightarrow 0 \)

\( S_3 \) \( \uparrow \) \( \max S_3 \) \( W \rightarrow \infty \) \( \max S_3 \)

\( S_4 \) \( \uparrow \downarrow \) \( \frac{W}{ \infty} \rightarrow 0 \)

\( S_1 \) \( \downarrow \) \( \min S_1 = 0 \) \( \frac{W}{ \infty} \rightarrow 0 \)

\( S_2 \) \( \uparrow \) \( \max S_2 \) \( \frac{W}{ \infty} \rightarrow 0 \)

\( S_3 \) \( \uparrow \) \( \max S_3 \) \( \frac{W}{ \infty} \rightarrow 0 \)

\( S_4 \) \( \uparrow \downarrow \) \( \frac{W}{ \infty} \rightarrow 0 \)

**B. Analytic Results Summary**

Table III shows the monotonicity and the upper or lower bounds of the variation contributions \( S_1 \) \( \cdots \) \( S_4 \) as the transistor channel width \( W \) increases. Here, \( \uparrow \), \( \downarrow \), and \( \uparrow \downarrow \) denotes monotonic increasing, monotonic decreasing and changing as a convex function. \( K_1 = \left( \frac{C_1}{L} + \frac{C_2 W \sigma^2}{\exp(L/\ell)} \right) \).

Table III also gives the maximum and minimum values of \( S_i \) \( (i = 1 \cdots 4) \) and their corresponding \( W \)'s.

**C. Validation of Analytic Results**

As (23) shows, \( \left( \partial I / \partial W \right)^2 \), \( \left( \partial I / \partial L \right)^2 \), and \( \left( \partial I / \partial R \right)^2 \) solely determine the trends of \( S_1 \), \( S_2 \), \( S_3 \), respectively, when \( W \) increases at both switching directions. The corresponding Monte Carlo simulation results of \( S_1 \), \( S_2 \), \( S_3 \) are shown in Figs. 19–21, respectively.

Fig. 19 shows \( S_1 \) monotonically decreases to zero as \( W \) increases to infinity at both switching directions. Its value at \( "1" \rightarrow "0" \) switching is always greater than that at \( "0" \rightarrow "1" \) switching because \( A_1 < A_3 \).

Fig. 20 shows that the variation contribution of \( L \) at \( "0" \rightarrow "1" \) switching is always larger than that at \( "1" \rightarrow "0" \) switching. The gap between them reaches the maximum when \( W \rightarrow \infty \).

Fig. 21 shows that the contribution from MTJ resistance \( R \) becomes dominant in the MTJ switching current distribution when \( W \) is approaching infinity. Because \( \left( \frac{V_{dd} - V_{th}}{R_H} \sigma R_H \right)^2 \), the normalized contribution of \( R \) is always larger at \( "1" \rightarrow "0" \) switching than that at \( "0" \rightarrow "1" \) switching.

We note that the additional coefficient \( (C_1/WL + (C_2/\exp(L/\ell))(Wc/W)) \) at the right side of (23) after the \( \left( \partial I / \partial V_{th} \right)^2 \) results in the different features of \( \left( \partial I / \partial V_{th} \right)^2 \) from \( S_4 \) in our simulations.
Fig. 23. Contributions from $V_{th}$.

Fig. 22 shows the values of $(\partial I/\partial V_{th})^2$ at both switching directions. At “0”→“1” switching, $(\partial I/\partial V_{th})^2$ increases monotonically when $W$ grows. At “1”→“0” switching, $(\partial I/\partial V_{th})^2$ increases first, then quickly decays to zero after reaching its maximum. These trends follow the expressions of $(\partial I/\partial V_{th})^2$ at either switching directions very well.

However, because of the additional coefficient on the top of $(\partial I/\partial V_{th})^2$, $S_4$ does not follow the same trend of $(\partial I/\partial V_{th})^2$ at either switching directions. Fig. 23 shows that at “0”→“1” switching, $S_4$ increases first and then slowly decreases when $W$ rises. At this switching direction, $S_4$ will become zero when $W \to \infty$ due to the existence of the additional coefficient $(C_1/WL) + (C_2/exp(L/1))(W/\sqrt{W_0})^2$. All these above results are well consistent with our analytic analysis in Table III.

REFERENCES


Wujie Wen received the B.S. and M.S. degrees in electrical engineering from the Beijing Jiaotong University, Beijing, China, and Tsinghua University, Beijing, in 2006 and 2010, respectively. He is currently pursuing the Ph.D. degree from the ECE Department, University of Pittsburgh, Pittsburgh, PA, USA.

His current research interests include design automation of statistical reliability analysis for emerging nonvolatile memories (ENVM), error-correction codes for ENVMs, and low-power design and computer architecture.

Dr. Wen was the recipient of the 49th DAC A. Richard Newton Graduate Scholarship, the 49th Young Student Support Program Award, and Graduate Travel Award of 2012 Nonvolatile Memories Workshop in San Diego, CA, USA.
Yaojun Zhang received the B.S. and M.S. degrees in electrical engineering from Shanghai Jiaotong University, Shanghai, China, and the University of Pittsburgh, Pittsburgh, PA, USA, in 2008 and 2010, respectively, where he is currently pursuing the Ph.D. degree from the ECE Department. His current research interests include emerging nonvolatile memory design for high performance, low power and scalable computer architectures, statistical reliability analysis of memory technology, and customized array and cell level circuit design for STT-RAM technology.

Yiran Chen (M’05) received the B.S. (Hons.) and M.S. (Hons.) degrees from the Tsinghua University, Beijing, China, and the Ph.D. degree from the Purdue University, West Lafayette, IN, USA, in 2005. After five years in industry, he joined the University of Pittsburgh, Pittsburgh, PA, USA, in 2010, as an Assistant Professor and was promoted to Associate Professor with the ECE Department, in 2014. He has published one book, several book chapters, and over 180 technical publications. He has been granted 83 U.S. and international patents with other 17 pending applications.

Yu Wang (SM’14) received the B.S. and Ph.D. (Hons.) degree in electronics engineering from Tsinghua University, Beijing, China, in 2002 and 2007, respectively. He is an Associate Professor with the EE Department, Tsinghua University, Beijing, China. His current research interests include parallel circuit analysis, application specific hardware computing (especially on the brain-related problems), and power and reliability aware system design methodology. He has authored and co-authored over 100 papers in refereed journals and conferences.

Dr. Wang was the recipient of IBM X10 Faculty Award in 2010, the Best Paper Award in ISVLSI 2012, the Best Poster Award in HEART 2012, and six best paper nominations in ASPDAC/CODES/ISLPED. He was the TPC Co-Chair of ICFPT 2011, a Finance Chair of ISLPED 2013/2014, and serves as a TPC member in several EDA/FPGA conferences.

Yuan Xie (SM’07) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, the M.S. and the Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, USA, in 1997, 1999, and 2002, respectively. He is currently a Professor with the ECE Department, University of California at Santa Barbara (UCSB), Santa Barbara, CA, USA. Before joining UCSB in 2014, he was a Professor at the Pennsylvania State University, State College, PA, USA, since 2003. He was also with IBM and AMD during 2002–2003 and 2012–2013, respectively. His current research interests include computer architecture, EDA, and VLSI design.