Reconfigurable Codesign of STT-MRAM Under Process Variations in Deeply Scaled Technology

Wang Kang, Student Member, IEEE, Liuyang Zhang, Jacques-Olivier Klein, Member, IEEE, Youguang Zhang, Member, IEEE, Dafine Ravelosona, and Weisheng Zhao*, Senior Member, IEEE

Abstract—Recently, spin-transfer torque magnetic random access memory (STT-MRAM) has been considered as a promising universal memory candidate for future memory and computing systems, thanks to its nonvolatility, high speed, low power, good endurance, and scalability. However, as technology scales down, STT-MRAM suffers from serious process variations and thermal fluctuations, which greatly degrade the performance and stability of STT-MRAM. In general, the optimization and robustness of STT-MRAM under process variations often require a hybrid design flow and multilevel codesign strategies. In this paper, we quantitatively analyze the impacts of process variations and thermal fluctuations on the STT-MRAM performances from physics, technology, and circuit design point of views. Based on the analyses, we found that readability is becoming the newest challenge for deeply scaled STT-MRAM due to the conflict between sensing margin and read disturbance. To deal with this problem, a novel reconfigurable design strategy from device, circuit, and architecture codesign perspective is then proposed. Finally, a conceptual hybrid magnetic/CMOS design flow is also proposed for STT-MRAM in deeply scaled technology nodes.

Index Terms—Magnetic tunnel junction (MTJ), nonvolatile memory, process variations, reconfigurable design, spin-transfer torque magnetic random access memory (STT-MRAM).

I. INTRODUCTION

THE increasing power consumption because of the leakage currents has become one of the major obstacles that prevent the continuous miniaturization of conventional CMOS-based memories (e.g., SRAM and DRAM) [1], [2]. Many novel devices have, therefore, been presented recently as alternatives to help the CMOS technology continue to downscale [3]–[5]. Among them, spin-transfer torque-based magnetic tunnel junction (STT-MTJ) is considered, at least in the short term, as one of the most powerful candidates to proceed the Moore’ law beyond the CMOS scaling limits, thanks to its nonvolatility, high access speed, low power, good endurance, and scalability [2], [6]. Based on the STT-MTJ device, the emerging spin-transfer torque magnetic random access memory (STT-MRAM) has become a promising universal memory candidate for future memory and computing systems [7]–[10].

A typical STT-MRAM bit cell consists of an MTJ connected in series with an nMOS transistor between a bit line (BL) and a source line, named 1T1MTJ cell structure, as shown in Fig. 1. The MTJ is used as a storage element, while the transistor acts as an access device controlled by a word line (WL). An MTJ can present two stable resistance states (i.e., low-resistance $R_P$ and high-resistance $R_{AP}$) depending on the relative magnetization orientation [parallel (P) or antiparallel (AP)] of the two FM layers [5]. The resistance difference between the two stable resistance states of the MTJ can be characterized by the tunnel magnetoresistance (TMR) ratio $\text{TMR} = (R_{AP} - R_P)/R_P$. To write data bit into the MTJ, only a bidirectional current (through STT mechanism [11]) is required. Due to the TMR effect, the digital information stored in the STT-MRAM bit cell can be readout by distinguishing the two different resistances of the MTJ with voltage or current sensing techniques.

In spite of the various promising merits and simple bit-cell structure of STT-MRAM, designing high-density STT-MRAM chips with high performance and stability is still challenging, especially as technology scales down to the nanometer regime. The increasing process variations and environment fluctuations may cause cell-to-cell...
and cycle-to-cycle variations, posing bit threats to STT-MRAM designs [12]–[15]. Process variations are introduced by the uncertainties and imperfections during the manufacturing process, thereby the device parameters are fixed and their impacts (cell-to-cell) on circuits are deterministic after chip fabrications, whereas the environment (e.g., thermal) fluctuations vary (cycle-to-cycle) during the chip lifetime [16], [17]. For example, the random dopant fluctuations (RDFs) and line-edge roughness (LER) may lead to CMOS transistor parameter variations, such as the threshold voltage, channel width, and length [18], [19]. Similarly, the thin-film deposition flatness and lithography roughness may result in the MTJ parameter variations, e.g., the oxide barrier thickness and cross-sectional area [13], [16]. In addition, random thermal fluctuations may cause STT-MRAM cycle-to-cycle variations, because the current drivability of the CMOS transistor, the resistance, and the thermal stability of the MTJ are dependent on temperature [20], [21].

To address the process variations and thermal fluctuations, the worst case corner is generally considered for the STT-MRAM standard design strategy, which is quite robust; however, it results in serious performance overhead. To solve this problem, various attempts have been made at, respectively, device, circuit, and architecture levels beyond optimizing the manufacturing process [22]–[24]. They can indeed achieve some improvements, but inevitably lead to some other overheads.

In this paper, we propose a new reconfigurable STT-MRAM design strategy, from the perspective of device, circuit, and architecture codesign to address the process variations. We quantitatively analyze the impacts of process variations (of both MTJ and CMOS transistor) and thermal fluctuations on the STT-MRAM design. Based on the analysis results, we found that readability, rather than writability, has become a newest barrier for the deeply scaled STT-MRAM. Differential sensing is a viable solution to solve this problem, but it results in serious hardware overhead because two MTJs are required to store only one single bit of data. To take advantage of differential sensing and to deal with the die-to-die (D2D) process variations, we propose a reconfigurable architecture by exploring the benefits of different memory bit-cell structures. Finally, for evaluation and optimization of STT-MRAM under process variations, we propose a conceptual hybrid magnetic/CMOS statistical design flow from multilevel codesign perspective.

The remainder of this paper is organized as follows. Section II analyzes the impacts of process variations and thermal fluctuations on the STT-MRAM performance. In Section III, we demonstrate the proposed reconfigurable codesign strategy to deal with the D2D process variations. Section IV illustrates the hybrid magnetic/CMOS statistical design flow. Finally, the conclusion is drawn in Section V.

II. IMPACTS OF PROCESS VARIATION AND THERMAL FLUCTUATION ON STT-MRAM PERFORMANCE

The first step to deal with process variations and thermal fluctuations is to analyze their impacts on the STT-MRAM performance. As we all know, an STT-MRAM chip is composed of MTJs and CMOS transistors (in memory bit cells and peripheral circuits). Both the MTJ and the transistor parameter variations caused by manufacturing processes, e.g., sputtering, subwavelength lithography, and etching [16]–[19], may affect the STT-MRAM products. In this section, we will present a quantitative analysis of the impacts of process variations and thermal fluctuations on the STT-MRAM performance.

A. Process Variation and Thermal Fluctuation on MTJ

An MTJ is mainly composed of several ultrathin (~1 nm) films. The aggressive technology scaling has inevitably led to drastic increase in process variations on MTJ devices, such as oxide barrier thickness, shape, and cross-sectional area. These device variations affect the magnetic and electrical parameters of STT-MRAM. As we all know, the MTJ acts as a variable resistor in STT-MRAM design. The critical parameters of MTJ include: 1) critical switching current \( I_{C0} \) in write operations; 2) resistances \( R_{MTJ} = R_P \) or \( R_{AP} \) in read operations; and 3) thermal stability factor \( \Delta \) in retention.

The STT switching statics of the perpendicular magnetic anisotropy (PMA) MTJ is mainly based on the critical switching current \( I_{C0} \), which can be expressed by [25], [26]

\[
I_{C0} = a \frac{e}{2 \mu_B} (\mu_0 M_S) H_K V = 2a \frac{g e}{2 \mu_B} E 
\]

\[
E = (\mu_0 M_S) H_K V / 2 
\]

where \( a \) is Gilbert damping coefficient, \( g \) is the gyromagnetic constant, \( e \) is the magnitude of the electron charge, \( \mu_B \) is the Bohr magneton constant, \( g \) is the spin polarization efficiency factor, \( \mu_0 \) is the permeability in free space, \( M_s \) is the saturation magnetization, \( H_K \) is the anisotropy field, \( V \) is the volume of the free layer, and \( E \) is the oxide barrier energy.

The MTJ resistance, \( R_P \) and \( R_{AP} \), can be expressed as [25]

\[
R_P = \frac{t_{ox}}{F \times \phi^{1/2} \times Area} \exp(1.025 \times \phi_{ox} \times \phi^{1/2}) 
\]

\[
R_{AP} = R_P \times (TMR + 1) 
\]

where \( t_{ox} \) is the oxide barrier thickness, \( \phi = 0.4 \) is the potential barrier height of MgO, \( F \) is a fitting constant, and area is the MTJ cross-sectional area.

The thermal stability factor \( \Delta \) can be expressed as

\[
\Delta = (\mu_0 M_S) H_K V / 2k_B T = E / k_B T
\]

where \( T \) is temperature. As can be seen, these parameters are affected by process variations and thermal fluctuations as

\[
I_{C0} \propto V 
\]

\[
R_{MTJ} \propto t_{ox} \exp(t_{ox})/Area 
\]

\[
\Delta \propto V / T.
\]

The impact of thermal fluctuations on the MTJ switching can be modeled by adding a thermal-induced random field \( (H_{fluc}) \) in the Landau–Lifshitz–Gilbert (LLG) equation [27], [28]

\[
dm/dt = \gamma m_f \times (H_{eff} + H_{fluc}) - \alpha m_f \times (m_f \times (H_{eff} + H_{fluc})) + J(\theta)(m_f \times m_f \times m_p) 
\]
where $m_f$ and $m_p$ are the unit magnetization vectors of the free layer and pinned layer, respectively, $H_{\text{eff}}$ is the effective magnetic field, and $J(\theta)$ is the coefficient of the STT term, which is dependent on the initial magnetization angle. Based on the above LLG equation, the current-driven magnetization switching of the MTJ can be divided into three: 1) thermal activation; 2) dynamic reversal; and 3) precessional switching, as shown in Fig. 2 [8].

The impact of thermal fluctuations on the MTJ resistance can be evaluated with the compact model provided in [29]. The conductance of the MTJ can be calculated by

$$G(\theta) = G_T(1 + \cos(\theta) P_1 P_2) + G_{\text{SI}}$$

where $\theta = 0^\circ$ or $180^\circ$ denotes the MTJ magnetization states, P or AP, respectively. $G_T = G_0 CT / \sin(CT)$ is the prefactor for direct elastic tunneling, here $G_0$ is a constant and $C = 1.387 \times 10^{-3} / \sqrt{\varphi}$, with the oxide barrier width ($d$) in Å and barrier height ($\varphi$) in electronvolt, $P_1$ and $P_2$ are the effective tunneling electron spin polarizations of the FM's, which are temperature-dependent parameters, as $P(T) = P_0 (1 - \beta T^{1.5})$, here $\beta$ is a material-dependent constant. $G_{\text{SI}}$ is a smaller second conductance, $G_{\text{SI}} = 0.5(G(0^\circ) + G(180^\circ)) - G_T$. In addition, the MTJ resistance in different states has different temperature dependences ($R_P$ is weaker than $R_A$); therefore, the TMR ratio is also temperature-dependent.

Fig. 3 shows the MTJ parameter dependences on process variations (the MTJ size and thin-film height) and temperature [30]–[32]. For example, as shown in Fig. 3(c), the MTJ resistance increases (or decreases) as the oxide barrier thickness grows (or reduces), represented by the blue solid line (or green dotted line). Thereby, as the oxide barrier thickness variation increases, the resistance ($R_P$) variation may be even larger than the TMR ratio under bias condition (inset).

### B. Process Variation and Thermal Fluctuation on Transistor

The scaling of technology results in also CMOS transistor parameter variations. Among all the variations, RDFs and LER represent the major variation sources, which lead to the variations of CMOS transistor parameters, such as threshold voltage ($V_{\text{th}}$), channel width ($W$), and length ($L$), finally affecting the output current of the transistor [16], [18], [33].

The threshold voltage ($V_{\text{th}}$) under RDF can be expressed as

$$V_{\text{th0}} = \varphi_{\text{MS}} + 2Q_{\text{dep}} / C_{\text{ox}}$$

$$\varphi_F = (k_B T / q) \ln(N_{\text{sub}} / n_i)$$

$$Q_{\text{dep}} = \sqrt{2q\varepsilon_{\text{si}} |\varphi_F| N_{\text{sub}}}$$

where $\varphi_{\text{MS}}$ is the difference between the work functions of the polysilicon gate and the silicon substrate, $Q_{\text{dep}}$ is the charge in depletion region, $C_{\text{ox}}$ is the gate oxide capacitance per unit area, $N_{\text{sub}}$ is the doping concentration, $n_i$ is the carrier concentration, and $\varepsilon_{\text{si}}$ is the dielectric constant of silicon. According to the theoretical models [18], [33], the deviations of $V_{\text{th}}$ due to RDF is inversely proportional to the square root of the transistor size

$$\sigma V_{\text{th}} \propto 1/\sqrt{WL}.$$  

Furthermore, due to the nonrectangular gate (NRG) and drain-induced barrier lowering effects [34], [35], the standard deviation of $V_{\text{th}}$ is also dependent on the channel size. Therefore, the total variations of $V_{\text{th}}$ under RDF and LER can be expressed as

$$\sigma V_{\text{th}}^2 = \frac{C_1}{W} + \frac{C_2}{\exp(L'/l')} + \frac{R_C}{W} \cdot \sigma_L^2$$

where $R_c$ is the correlation length of NRG; $C_1$, $C_2$, and $l'$ are the technology-dependent coefficients; and $\sigma_L^2$ is the variation of $L$. The first and second terms describe, respectively, the contributions of RDF and LER to the $V_{\text{th}}$ variations. With technology scaling, the contribution of the second term becomes prominent due to the reduction of CMOS transistor size.

In STT-MRAM, CMOS transistor exists in both the memory bit cells (access device) and the peripheral circuits. The process variations on the access CMOS transistors in bit cells mainly affect their current driving capability. The drain current ($I_D$) of the CMOS transistor can be expressed as

$$I_D = \mu C_{\text{ox}} W L_{\text{eff}} [(V_{GS} - V_{\text{th}}) V_{DS} - 0.5 V_{DS}^2]$$

Fig. 2. Spin current-driven MTJ magnetization switching phase diagram [8].

Fig. 3. MTJ parameter dependence on the size and temperature. (a) Critical switching current with respect to the MTJ size. (b) Thermal stability factor with respect to the MTJ size. (c) Resistance with respect to the oxide barrier thickness. (d) Resistance and TMR ratio with respect to temperature [30]–[32].
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

IEEE TRANSACTIONS ON ELECTRON DEVICES

4

Fig. 4. CMOS transistor parameter dependence on the process variations. (a) Threshold voltage variation due to RDF. (b) Threshold voltage variation due to LER. (c) Conduction and leakage currents with respect to the technology node. (d) Current variation with respect to the technology node [18], [19], [33].

where $\mu$ is the carrier motility, $L_{\text{eff}}$ is the effective channel length, $V_{\text{GS}}$ and $V_{\text{DS}}$ are, respectively, the gate–source voltage and drain–source voltage of the transistor. As can be seen, $I_D$ is affected by both RDF and LER through $V_{\text{th}}$ and transistor size. Fig. 4 shows the CMOS transistor parameter dependences on the process variations [18], [19], [33]. In addition, since the carrier mobility decreases as temperature increases, the drivability of the CMOS transistor degrades at higher temperature with the same biasing condition. On the other hand, the MTJ resistance decreases as temperature increases, therefore, the driving current in the 1T1MTJ bit cell is generally independent on temperature. Process variations on the CMOS transistors of the peripheral circuits (e.g., write and read circuits) may also affect the STT-MRAM performances. Taking the read circuit, for example, if there exits device mismatch between the load transistors of the data and the reference cells, the sensing margin (SM) will be reduced, degrading the sensing reliability.

C. Process Variation and Thermal Fluctuation on Bit Cell

By integrating both the process variations (including the oxide barrier thickness 10%, FM layer thickness 10%, and resistance 5%) and the thermal fluctuations (10%) into a compact PMA STT-MTJ SPICE model [25], meanwhile employing a commercial CMOS 40-nm design kit (with process variations $3\sigma$, for the channel length, width, and threshold voltage) [36], Monte Carlo statistical simulations (1000 runs) are performed to evaluate the impacts of process variations and thermal fluctuations on the STT-MRAM 1T1MTJ bit cell. Some critical parameters are listed as follows: 1) supply voltage $V_{\text{dd}} = 1$ V; 2) critical switching current (standard) of the MTJ is $I_{\text{C0}} = 55$ $\mu$A; 3) MTJ size $= 40$ nm; and 4) the resistance. Area product $R.A = 5 \Omega \cdot \mu m^2$ and TMR $= 100\%$. Fig. 5 shows the dc characteristic of the MTJ (four runs). As can be seen, the practical critical switching current varies under process variations and thermal fluctuations. Fig. 6(a) and (b) shows the resistance and switching time statics of the 1T1MTJ bit cell.

In the circuit-level design perspective, the impacts of process variations and thermal fluctuations on STT-MRAM can be classified into two parts: 1) write operation and 2) read operations. Previous efforts are concentrated mainly on improving the writability of STT-MRAM. Fortunately, thanks to the rapid progress and the improvements of physics and electronics, the various methods and techniques from, device level, circuit level, and system level have been proposed to improve the STT-MRAM write performance [37]–[40]. For example, PMA MTJ has been widely investigated to lower the switching current and latency. Meanwhile, many adaptive write algorithms and circuits have been proposed to deal with the cell-to-cell and cycle-to-cycle variations. In addition, the continuous technology scaling results in dramatically the reduction of MTJ critical switching current [Fig. 3(a)], which alleviates greatly the drivability of the CMOS transistor. However, with technology shrinking, the readability, rather than writability, may become an ultimate bottleneck of
STT-MRAM because of the increased process variations in scaled technology nodes.

D. Technology Scaling on Readability

As technology scales, on the one hand, the reduction of the MTJ critical switching current will increase the probability of read disturbance (RD) during the read operations, since the read current ($I_{\text{read}}$) may approach the MTJ critical switching current, as shown in Fig. 7(a) [41]. On the other hand, the process variations of both MTJ and CMOS transistor increase rapidly, significantly degrading the SM, as shown in Fig. 7(b). This is because process variations may induce: 1) device mismatch between the data and the reference cells and 2) input offset of the read circuit. In general, a high read current is required to improve the SM; however, it inevitably leads to high probability of RD as

$$P_{\text{dis}} = 1 - \exp \left( -\frac{t_{\text{read}}}{\tau_0} \exp \left[ -\frac{\Delta}{1 - \frac{I_{\text{read}}}{I_{C0}}} \right] \right)$$  \hspace{1cm} (17)$$

where $P_{\text{dis}}$ denotes the RD probability, $\tau_0$ is the attempt period ($\sim$ 1 ns), $I_{\text{read}}$ and $t_{\text{read}}$ are the read current pulse amplitude and width. Therefore, there exists a conflict between SM and RD. Fig. 8 shows the read error rate (due to the limited SM) and RD probability with respect to the read current amplitude (read current pulse width is set as 0.5 ns). This conflict has become a critical challenge for deeply scaled STT-MRAM design. To solve this conflict, eliminating process variations by optimizing the manufacturing process turns to be impractical, because it is relatively expensive and ineffective in nanoscale nodes. Alternative solutions by design from device level, circuit level, and system level are currently preferred, such as the stacked bit cell, triple-stage offset cancellation read circuit, and error correction codes [14], [24], [42]. These solutions can indeed reduce the RD and/or improve the SM to some extent; however, they inevitably lead to some other overheads. The conflict cannot be well solved with only one single-level design techniques, as the critical switching current of the MTJ reduces, whereas the process variations grow, following the continuous scaling of the technology. In this case, multilevel codesign solutions are strongly required.

III. RECONFIGURABLE CODESIGN STRATEGY

In general, process variations consist of D2D and within-die (WID) variations. D2D variation denotes the parameter fluctuations across the dies and wafers, whereas WID variation refers to parameter variations within a single die. In general, the D2D process variation exhibits as a random distribution and is relatively large among different dies, while the WID process variation shares similar design parameter variations within a die and is relatively small due to spatial correlations [43]. Previous efforts focus mainly on dealing with the WID process variation with device-level and circuit-level design techniques; however, they are inefficient for the D2D process variation. As technology scales down, both D2D and WID process variations increase, degrading significantly the STT-MRAM chip yield and robustness. This section presents a novel reconfigurable codesign strategy to deal with both the D2D and the WID process variations. Considering that the device-level and circuit-level variation-aware design techniques are rather popular, this section concentrates mainly on the reconfigurable architecture.

A. Motivation

As discussed above, the 1T1MTJ bit-cell faces readability challenges in deeply scaled technology nodes (e.g., <30 nm) due to the conflict between the SM and the RD. Differential sensing is a good choice to improve SM without increasing RD. In this scenario, two MTJs are used to store only one data bit (one MTJ for the data bit and the other one for its component, as shown in Fig. 9, named 1T2MTJ and 2T2MTJ bit cells) and the read operation is completed by sensing the resistance difference between the two MTJs rather than comparing the data cell with a reference cell [44]–[46]. Differential sensing is able to double the SM compared with standard sensing scheme with the same read current configurations, but it has two main drawbacks, i.e., high hardware and power consumption overheads.

To leverage the benefits and drawbacks of the differential sensing, we propose a reconfigurable architecture based on the
following two facts. First, a general computing system needs to deal with various applications with different performance requirements. For example, banking services require 100% data accuracy and are relatively insensitive to storage density. In contrast, video services usually require high data throughput, but occasional erroneous frames or pixels are generally not noticeable. Second, due to the D2D variation, the dies (chips) on a wafer exhibit different process variations. Based on these, the dies with relatively small process variations (or used in high-throughput applications) can be configured into standard sensing, while the dies with relatively higher process variations (or used in high-performance applications) can be configured into differential sensing. Simultaneously, both device-level and circuit-level variation-tolerant techniques, e.g., reference signal calibration and offset compensation, are then employed for dealing with the WID process variations. In S-mode, the target memory bit cell is read by comparing its sensed voltage with a reference signal ($V_{\text{ref}}$); while in D-mode, two adjacent MTJs are used to store only one single data bit and differential sensing is used to read the data stored in the two complementary MTJs. As shown in Fig. 10, compared with the standard architecture, the main hardware overhead of the proposed reconfigurable one includes the MUX and TG. Note that the MUXs and TGs are shared by the whole BL or WL, therefore, their direct overheads can be negligible for the entire STT-MRAM chip.

For comparison among the three structures, R-2 has the highest hardware efficiency, since two MTJs share one nMOS transistor. However, it leads to some write and read latencies when it operates in the S-mode, as the two MTJs cannot be accessed simultaneously with one access transistor similar to the multilevel-cell concept. R-3 is the same as R-1 in S-mode, but it has better hardware efficiency in D-mode, since it can store two data bits with three MTJs by differential sensing with each other. Unfortunately, R-3 needs system-level data coding and new sensitive read circuits in D-mode, since the data are represented by the combinations of the MTJs (the details are not the scope of this paper and will be included in another publication). Therefore, R-1 is suggested as the most potential candidate for the reconfigurable design because of its simple structure and easy implementation.

C. Evaluation

Using the SPICE STT-MTJ compact model presented in Section II and a CMOS design kit, we took R-1, for example, to evaluate the proposed reconfigurable design with Monte Carlo simulations (2000 runs) in 40-nm technology node. In the simulations, we consider both the process variations of the MTJs and the CMOS transistors. The parameters are the same, as listed in Section II-C. Fig. 11 shows the SM and read error rate of the proposed design in S-mode and D-mode with respect to the technology nodes. Because the read current amplitude in D-mode can be half of that in S-mode to achieve the same SM, the RD probability can be significantly reduced, as shown in Fig. 12. In our simulations, the read current is carefully selected to tradeoff between SM and RD.
Actually, no RD was detected in our simulations. Fig. 13 shows the average read power and the latency of the proposed design between S-mode and D-mode. Both read power and speed in D-mode are higher than those in S-mode due to the differential sensing. The average write power in D-mode (∼0.76 pJ at 40 nm) is nearly double of that in S-mode (∼0.40 pJ at 40 nm), since two MTJs are required to be programmed in D-mode. The write latency is not so obvious between the two operating modes thanks to the P write scheme. In summary, STT-MRAM operating in D-mode has better reliability and latency than in S-mode by sacrificing hardware and power consumption. These results are consistent with the above analyses. Please note that only circuit-level simulations are performed to evaluate the bit-cell performance here. These results can reveal the basic concepts and merits of our proposed reconfigurable design. Nevertheless, further detailed system-level simulations or case studies are expected (which are undergoing in our lab) in the future works to assess the overall system performance.

IV. HYBRID MAGNETIC/CMOS DESIGN FLOW

Unlike the conventional memory technologies, STT-MRAM involves magnetic and CMOS parts; therefore, the optimization and robustness of STT-MRAM design under process variations and environment fluctuations require a hybrid magnetic/CMOS design flow. Fig. 14 shows the overview of our proposed design flow, including device-level, circuit-level, and architecture-level variation-aware design considerations.
The key feature of our proposed design flow is the reconﬁgurable codesign strategy. In the device-level design, based on the design speciﬁcations, the STT-MTJ SPICE model is constructed and the critical device parameters, such as the critical switching current, TMR ratio, and resistances, are obtained. According to the current drivability requirement of the STT-MTJ, the nMOS access transistor size of the memory bit cell can then be determined and the bit-cell library can be constructed with the nominal process parameters and the corresponding variations. Finally, sensitivity and robustness analyses under different corners can be performed to characterize the statistical features of the memory bit cell. In circuit-level design, memory array conﬁguration and peripheral circuits, e.g., write/read circuits are determined, considering the process variations and environment ﬂuctuations. At this level, by estimating the parasitic parameters (resistance and capacitance) of the STT-MRAM array, the block and bank conﬁgurations can also be determined. Meanwhile, variations-aware design techniques, such as transistor sizing, body biasing, reference calibration, input offset compensation, and WL/BL voltage boosting, are employed. Monte Carlo statistical simulations are carried out to evaluate the overall circuit performances, such as hardware area, latency, reliability, and power. In architecture-level design, the above reconﬁgurable design strategy is considered. After chip fabrication, different architectures can be conﬁgured depending on the chip testing results or the application-oriented performance requirements.

V. CONCLUSION

In this paper, we aim to deal with the process variations and thermal ﬂuctuations of STT-MRAM with multilevel codesign strategy. By analyzing the impacts of process variations of both MTJ and CMOS transistor and thermal ﬂuctuations on the STT-MRAM designs, we found that readability becomes a newest challenge and barrier for STT-MRAM in deeply scaled technology nodes. Differential sensing is a viable solution to alleviate this issue, but it may induce serious hardware and power overheads. By exploring the beneﬁts of differential sensing and taking into consideration the application-oriented performance requirements, a reconﬁgurable codesign strategy, combining device-level, circuit-level, and architecture-level techniques, was proposed to deal with the D2D and WID process variations and thermal ﬂuctuations. Finally, a robust hybrid magnetic/CMOS design ﬂow was proposed to optimize the overall reliability and yield of STT-MRAM under process variations and thermal ﬂuctuations.

REFERENCES


Wang Kang (S’12) was born in China in 1987. He received the B.S. degree from the School of Electronic and Information Engineering, Beihang University, Beijing, China, in 2009. He is currently pursuing the Ph.D. degree in microelectronics with Beihang University, and the University of Paris-Sud, Orsay, France.

Liuyang Zhang was born in China in 1988. He received the B.S. degree from the School of Information Engineering, The PLA Information Engineering University, Zhengzhou, China, in 2011, and the M.S. degree from the School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou, China, in 2014. He is currently pursuing the Ph.D. degree in microelectronics with Beihang University, Beijing, China.

Jacques-Olivier Klein (M’90) was born in France in 1967. He received the Ph.D. degree and the Habilitation degree in electronic engineering from the University of Paris-Sud, Orsay, France, in 1995 and 2009, respectively.

He is currently a Professor with the Institute of Fundamental Electronics, University of Paris-Sud, where he leads the Nanocomputing Research Group.

Youguang Zhang (M’13) was born in China in 1963. He received the M.S. degree in mathematics from Peking University, Beijing, China, in 1987, and the Ph.D. degree in communication and electronic systems from Beihang University, Beijing, in 1990. He is currently a Professor with the School of Electronic and Information Engineering, Beihang University. His current research interests include microelectronics and wireless communication.

Dafiné Ravelosona received the Ph.D. degree in solid states physics with the Centro Nacional de Microelectrónica, Madrid, Spain, in 1995. He was an Invited Scientist with the Research Center, Hitachi Global Storage Technology, San Jose, CA, USA, from 2004 to 2005. He is currently an Experimentalist Physicist and the Head of the Nanospintronics Group with the Institute of Fundamental Electronics, University of Paris-Sud, Orsay, France.

Weisheng Zhao (M’06–SM’14) was born in China in 1980. He received the M.S. degree in electrical engineering from Ecole Nationale Supérieure d’Electrotechnique, d’Électronique, d’Informatique, d’Hydraulique et des Télécommunications Engineering School, Toulouse, France, in 2004, and the Ph.D. degree in physics from the University of Paris-Sud, Orsay, France, in 2007. He is currently a Professor with the Spintronics Interdisciplinary Center, Beihang University, Beijing, China.