

Comparative Evaluation of Spin-Transfer-Torque and Magnetoelectric Random Access Memory

Shaodi Wang*, Hochul Lee*, Farbod Ebrahimi*[†], P. Khalili Amiri*[†], Kang L. Wang*, *Fellow, IEEE*, and Puneet Gupta*

*Department of Electrical Engineering, University of California, Los Angeles, CA 90095 USA

[†]Inston Inc, Los Angeles, CA, 90095

Abstract—Spin-Transfer Torque random access memory (STT-RAM), as a promising non-volatile memory technology, faces challenges of high write energy and low density. The recently developed magnetoelectric random access memory (MeRAM) enables the possibility of overcoming these challenges by the use of voltage-controlled magnetic anisotropy (VCMA) effect and achieves high density, fast speed, and low energy simultaneously. As both STT-RAM and MeRAM suffer from the reliability problem of write errors, we implement a fast LandauLifshitzGilbert equation-based simulator to capture their write error rate (WER) under process and temperature variation. We utilize a multi-write peripheral circuit to minimize WER and design reliable STT-RAM and MeRAM. With the same acceptable WER, MeRAM shows advantages of 83% faster write speed, 67.4% less write energy, 138% faster read speed, and 28.2% less read energy compared with STT-RAM. Benefiting from the VCMA effect, MeRAM also achieves twice the density of STT-RAM with a 32nm technology node, and this density difference is expected to increase with technology scaling down.

Index Terms—STT-RAM, MeRAM, voltage controlled memory, MTJ, evaluation, write error rate, write speed, write energy

I. INTRODUCTION

MAGNETORESISTIVE random access memory (MRAM) [1] using magnetic tunnel junctions (MTJs) is a promising data storage technology due to its non-volatility, zero leakage power, high speed, high endurance, immunity to single-event soft error and high thermal budget [2, 3]. MTJs switched by Spin-transfer torque (STT-MTJ) [4, 5] potentially promise the speed and area of dynamic RAM (DRAM) [6]. Therefore, spin-transfer torque RAM (STT-RAM) designed with STT-MTJs is identified as a possible replacement of current memory technologies, such as static RAM (SRAM) Cache [7]–[10] and DRAM main memory [11]. In addition to the traditional uses, research also focuses on exploring new memory architectures to utilize its non-volatility (e.g., fast persistent memory system enabled by STT-RAM [12] allows to instantly recover from off state). However, STT-RAM faces the challenge of high write current (around 100 μ A at 45nm node [6]), which does not directly scale with MTJ dimension [13]. As a result, a large access transistor is always required, leading to high write energy and low density.

The recent development of voltage-controlled MTJs (VC-

MTJs) with voltage-controlled magnetic anisotropy (VCMA) provides more promising performance than STT-MTJs [14]–[20]. This technology allows for precessional switching, a process which provides flipping of the magnetization upon a voltage pulse, irrespective of the initial state. It enables the use of minimum sized access transistors, as well as precessional switching to simultaneously achieve low energy, high density and high speed magnetoelectric random access memory (MeRAM). MeRAM reduces switching energy due to reduced ohmic loss (~ 10 fJ/bit [15] for the VC-MTJs with over 100X higher resistance than STT-MTJs).

Both STT-MTJ and VC-MTJ suffer from the reliability problem of intrinsic switching failure caused by thermal fluctuation exacerbated by process variation [21, 22] and temperature variation. This problem can be quantified by write error rate (WER), which is the average number of switching failures per write. STT-RAM can simply reduce the WER by using high current and long write time. By contrast, MeRAM does not have a trivial solution, because every VC-MTJ has an optimal write pulse giving the lowest WER, and the effect of variation on the optimal pulse is less straight forward.

Considering the advantages and disadvantages, MeRAM requires a comprehensive evaluation, while STT-RAM, which is better known and has similar structure and fabrication process, is an appropriate reference. To accurately compare the reliability of the two technologies, the WER must be precisely captured. The state-of-art method is the LandauLifshitzGilbert (LLG) differential equation based Monte-Carlo simulation (e.g., [23] for STT-MTJs). However, previous implementations were too slow to be adapted for high-accuracy simulations needed for large memory array. As a result, limited samples were simulated in previous STT-RAM studies [21, 24], which could not address WER below 10^{-4} . This may lead to inappropriate designs, e.g., the WER of 10^{-8} requires 20% more write current than 10^{-4} for STT-MTJs. Moreover, the context of circuit-level optimization is also essential given that peripheral circuit can significantly affect memory performance. For instance, MRAM can leverage circuit techniques to mitigate the WER by trading off the speed and power.

In this paper, we perform the first comprehensive circuit-level comparison between MeRAM and STT-RAM with respect to reliability, power, performance, density, and scalability using a high-speed Monte-Carlo simulator. Our contributions are summarized as follows.

- The LLG equation is modified to include thermal fluctuation, temperature dependence, STT and VCMA effect.

Shaodi Wang, Hochul Lee, Farbod Ebrahimi, P. Khalili Amiri, and Puneet Gupta are with the Department of Electrical Engineering, University of California, Los Angeles, CA, 90095 USA e-mail: (shaodiwang@g.ucla.edu).

Copyright ©2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org

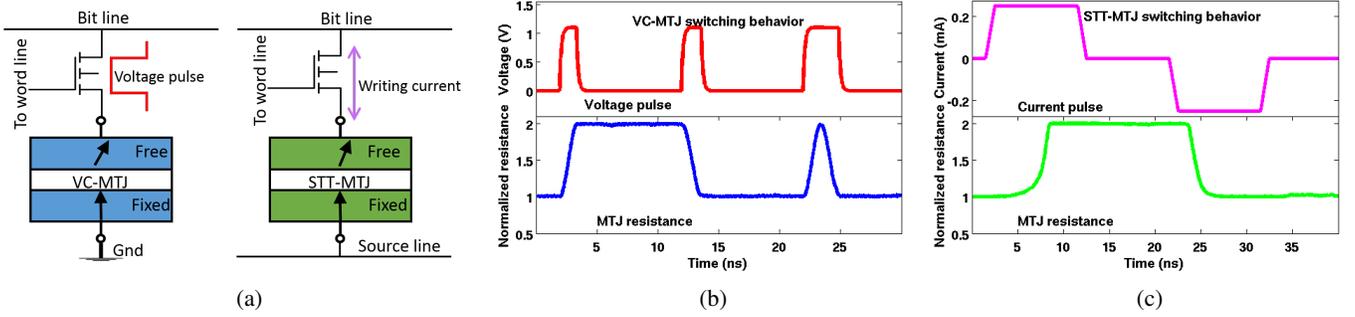


Fig. 1: (a) Writing mechanism of STT-RAM and MeRAM. (b) VC-MTJ is switched by unidirectional voltage pulses. The first two same pulses switch the resistance state of a VC-MTJ from P to AP and then back to P , the third double-width pulse make two switches continuously. (c) STT-MTJ is switched by directional current pulses, and the switching directions depends on the direction of current.

- Based on the LLG equation, a simulator is implemented in CUDA on GPU platform¹. It completes 100,000 simulations within 2s.
- The density and scalability of STT-RAM and MeRAM are analyzed based on the 32nm design rules and derived models respectively.
- The impact of process and temperature variation on STT-RAM and MeRAM is compared: for the first time the variation impact on MeRAM is analyzed; the temperature related behavior is analyzed using more accurate model than existing studies.
- The WER of both MRAMs is minimized below an acceptable rate by utilizing multiple writes enabled by the pre-read and write sense amplifier (PWSA) [25]. With this design, a fair circuit-level energy-speed comparison between STT-RAM and MeRAM is carried out.

The paper is organized as follows. Section II describes the LLG based model and simulation in detail. Section III analyzes the scalability of STT-RAM and MeRAM. Section IV designs MRAM cells under process and temperature variation and compares the cell density of two MRAMs with 32nm design rules. Section V analyzes the WER of the nominal MTJs and MRAMs with process and temperature variation separately. Section VI introduces the PWSA multi-write design and carries out a circuit-level comparison with respect to write latency, energy and MRAM failure analysis. Section VII concludes the paper.

II. MODELING AND SIMULATION

Both STT-MTJ and VC-MTJ are resistive memory devices, their resistances are determined by the magnetization directions of two ferromagnetic layers. The direction of one layer is fixed (referred to as reference layer) while the other one can be switched (referred to as free layer). A low resistance is present when magnetic directions in the two layers are parallel (referred as P state); a high resistance is present when the two directions are anti-parallel (referred as AP state). The two states are utilized to store "0" and "1". Tunnel magnetoresistance (TMR, defined as $(R_H - R_L)/R_L$) over 200% has been demonstrated, which means that the high resistance can be over 3X of the low resistance. Based on the magnetization direction of the two layers, MTJs are classified as in-plane and out-of-plane (perpendicular magnetized)

devices. Recently, STT-RAM with out-of-plane MTJs is found to have lower write current and less fabrication challenge than in-plane MTJs [26]–[29]. The magnetic anisotropy of out-of-plane MTJs is dominated by the perpendicular magnetic anisotropy (PMA). In this paper, we consider the STT-RAM and MeRAM with out-of-plane MTJs.

Although STT-MTJs and VC-MTJs share a similar device structure and data storing mechanism, their switching mechanisms differ as shown in Fig. 1, e.g., in an STT-MTJ, polarized electrons flowing from the reference layer to the free layer switch the magnetization of the free layer to P state; when electrons flow in the opposite direction, the reflected electrons from the reference layer switch the free layer to AP state.

Unlike STT-MTJ, VC-MTJ utilizes an unidirectional voltage pulse to make both switches from AP to P and from P to AP . As is illustrated in Fig. 2, the energy barrier E_b separates the two stable states of the free layer magnetization (pointing up and down) when the voltage applied across the VC-MTJ is 0. The energy barrier E_b decreases with the voltage increase due to VCMA effect. When the voltage reaches V_{Write} (> 0 , see Eqn. 8), full 180° switching can be achieved by timing the precessional switching of magnetization.

In the MTJ switching simulation, the magnetization in the free layer during any short interval (e.g., 0.25ps in our setup) is described by an LLG differential equation. The entire switching is captured by iteratively solving the LLG equations in sequence. The WER is then extracted from numerous simulations in a Monte-Carlo approach. Shorter interval and more simulations can improve the accuracy at the expense of time. The LLG equation (1) describes the dynamic behavior

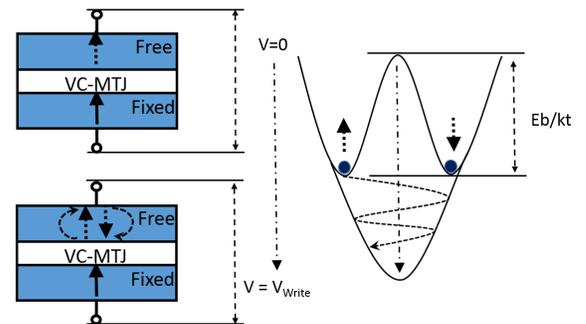


Fig. 2: VCMA-induced precessional switching. When a voltage is applied on the VC-MTJ, the energy barrier separating the two magnetization states of the free layer is reduced so that the magnetization state starts to spin.

¹Freely available at <http://nanocad.ee.ucla.edu/Main/DownloadForm>

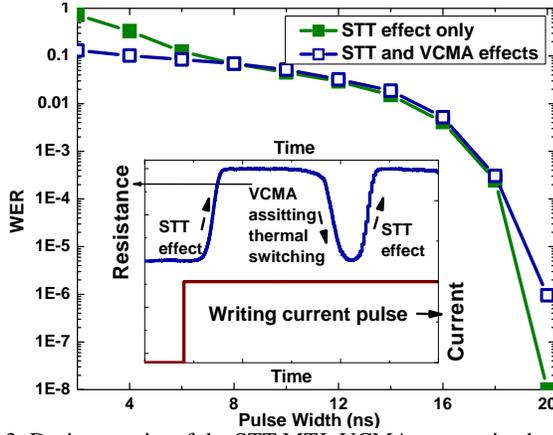


Fig. 3: During a write of the STT-MTJ, VCMA may assist the thermal activation to cause unintended switching. This effect can improve the switching probability when the write pulse width is insufficient to switch the STT-MTJ, on the other hand, may lead to switching failure when the write pulse width is sufficiently long.

of the free layer magnetization vector \mathbf{M} in the presence of an external field (\mathbf{H}_{Ext}), shape anisotropy (\mathbf{H}_{Shape}), PMA (H_{PMA}), and thermal fluctuation (\mathbf{H}_{Therm}), as follows.

$$\frac{d\mathbf{M}}{dt} = -\gamma(\mathbf{M} \times \mathbf{H}) + \frac{\alpha}{M_S} \cdot \mathbf{M} \times \frac{d\mathbf{M}}{dt} \quad (1)$$

$$+ \gamma \frac{\alpha_J(\theta)}{M_S} \mathbf{M} \times (\mathbf{M} \times \mathbf{p})$$

$$\mathbf{H} = \mathbf{H}_{Ext} + \mathbf{H}_{Shape} + H_{PMA} + \mathbf{H}_{Therm}$$

Where γ is the gyromagnetic ratio, \mathbf{H} is the effective magnetic field, α is the intrinsic damping constant, M_S is the saturation magnetization, and α_J is the amplitude of the spin-transfer torque induced by current. H_{PMA} can be reduced by voltage due to the VCMA effect, which is expressed below.

$$H_{PMA} = H_{PMA}(0) \cdot (1 - \zeta \cdot V_{MTJ}) \quad (2)$$

$$H_{PMA}(0) = 2K / (t_{FL} \cdot M_S) - M_S$$

$$\zeta = \xi / (K \cdot t_{MgO})$$

Where V_{MTJ} is the applied voltage, t_{FL} and t_{MgO} are the thickness of the free layer and MgO layer respectively, K is the anisotropy constant, ξ is the anisotropy change slope, ζ is the VCMA factor with the unit of V^{-1} . Positive V_{MTJ} causes VCMA effect to reduce H_{PMA} as well as the perpendicular magnetization to cause precessional switching [14]. An optimal applied voltage can exactly cancel out the perpendicular magnetization, and then a perfect precessional switching (controlled by the in-plane external magnetic field H_{Ext}) starts, during which the magnetization in the free layer rotates. The optimal pulse width equals to the half cycle of the precessional switching [16]. More specifically, the optimal pulse allows the magnetization to rotate exactly 180° . The VCMA effect is considered for both STT-MTJ and VC-MTJ, while previous circuit-level STT-RAM studies ignore it. As an example, the impact of VCMA effect on an STT-MTJ is shown in Fig. 3: VCMA can change the WER. When a write current is applied, the voltage drop on the STT-MTJ reduces the PMA and increases the chance of thermal activated switching. Hence when the write pulse is not long enough to guarantee a switch, the thermal activated switching assisted by VCMA increases the switching probability, but when the write pulse is long enough, it induces errors.

TABLE I: Modeling parameters at 300K.

γ [$m/(A \cdot S)$]	M_S [A/m]	ξ [$fJ/(V \cdot m)$]	α
$2.2 \cdot 10^5$	1.2×10^6	STT: 37 [30, 37], VC: 85 [38]	0.02
H_{Ext} [A/m]	K [J/m^2]	P_{SV}, P_{Tunnel}	TMR
$1.1 \cdot 10^4$	$1.068 \cdot 10^{-3}$	0.66 [35]	100%

The \mathbf{H}_{Therm} in (1) is the thermal fluctuation field and randomly determined as a variable following Normal distribution at each simulation interval (3).

$$\mathbf{H}_{Therm} = Norm3d(0, \sqrt{\frac{2k_B T}{\gamma M_S t_{FL} A}}) \quad (3)$$

Where A is the area of the MTJ, k_B is the Boltzmann constant, and T is the temperature.

Temperature significantly affects the MTJ switching behavior, e.g., the WER of an STT-MTJ can increase from 10^{-8} to 10^{-6} with temperature rising from 300K to 350K (see Fig. 9). Except \mathbf{H}_{Therm} , other terms in (1) also change with temperature as described in (4) [30], which are commonly ignored in previous large-scaled MRAM studies [7, 8]. In-situ thermal sensors [31, 32] may help to monitor MRAM temperature and modulate MTJ write schemes.

$$M_S(T) = M_S^* \left(1 - (T/T^*)^{3/2}\right) \quad (4)$$

$$K(T) = K^* \cdot (M_S(T)/M_S^*)^{2.18}$$

$$\xi(T) = \xi^* \cdot (M_S(T)/M_S^*)^{2.83}$$

Where T^* is 1120K, and M_S^* , K^* , and ξ^* are corresponding parameters at 1120K.

The spin-transfer torque effect is described in (5) [28].

$$\alpha_J(\theta) = \frac{\hbar g(\theta)}{2eM_S t_{FL}} J \quad (5)$$

Where \hbar is the reduced Plank constant, $g(\theta)$ is the spin-torque efficiency factor [33], θ is the angle between the two magnetizations of the free layer and reference layer, and J is the current density through MTJ. $g(\theta)$ can be further expanded at (6) [27, 33, 34].

$$g(\theta) = g_{Tunnel}(\theta) + g_{SV}(\theta) \quad (6)$$

$$g_{SV} = \left[-4 + (1 + P_{SV})^3 (3 + \cos \theta) / (4 \cdot P_{SV}^{3/2}) \right]^{-1}$$

$$g_{Tunnel} = 0.5 \cdot P_{Tunnel} / (1 + P_{Tunnel}^2 \cos \theta)$$

Where g_{Tunnel} and g_{SV} , as functions of θ , are polarization efficiency of tunnel current and spin valve respectively. P_{Tunnel} and P_{SV} are material-dependent polarization factors for the tunnel current and current passing through ferromagnetic layers respectively [34]. These two parameters are not necessarily equal, while we use 0.66 [35] for both of them in this paper. The required switching current (known as critical current) differs from switching directions due to the difference in polarizing efficiency [36]. The parameters used in the model and simulation are listed in Table I.

Inspired by the massive floating point calculations involved by the LLG equation and highly independent operations in Monte-Carlo simulations, we implement the switching simulator in CUDA, and it completes 100,000 simulations within 2s on NVIDIA Tesla M2070. The model has been validated. The speed improvement comes from highly parallel simulations [39]–[41].

III. SCALABILITY

In this subsection, we analyze the scalability of STT-RAM and MeRAM regarding retention, write power, area, and fabrication challenges. Retention, as one of the most important metric for memory system [42], determines the available data-storing time and thus is a non-scalable parameter [6]. An MTJ with low retention is easy to flip, but high retention increases the write difficulty. Considering the trade-off, an efficient design should have its retention as low as possible but satisfy application requirement. For STT-MTJ and VC-MTJ, the retention time (mean time to false switching during idle state) τ is an exponential function of thermal stability Δ [43, 44].

$$\tau = \tau_0 \exp(\Delta) \quad (7)$$

$$\Delta = \frac{H_{K,eff} M_S A t_{FL}}{2k_B T}$$

Where $H_{K,eff}$ is the sum of perpendicular components of H_{Shape} , H_{Ext} , and H_{PMA} . Based on [14, 28], we derive the critical current of STT-MTJ and the optimal voltage of VC-MTJ as functions of Δ and MTJ area A in (8).

$$I_{STT}(A, \Delta) \approx \frac{4k_B T e}{\hbar g} \Delta \propto \frac{\Delta}{g} \quad (8)$$

$$V_{VC}(A, \Delta) \approx \frac{2k_B T \Delta}{\zeta M_S^2 t_{FL} A} \propto \frac{\Delta}{\zeta A}$$

Where e is the elementary charge, g is the spin-torque polarization efficiency, and ζ is the VCMA factor. From (8), the critical current I_{STT} of STT-MTJ does not directly depend on the MTJ dimension given that the thermal stability is constant. But as g increases with decreased A due to the sub-volume excitation for large MTJs with lateral size over 50nm [45], I_{STT} can be slightly reduced by scaling dimension. However, the reduction trend does not continue for small MTJs. The optimal voltage V_{VC} of VC-MTJ is inversely proportional to A indicating that it will increase with dimension scaling. Hence, the key for scaling both technologies is finding materials that provide more g and ζ .

With respect to the memory density, access transistors dominate the area rather than the MTJs (see Fig. 4). Because of the non-scaling critical current for small STT-MTJs, access transistors in STT-RAM have to increase the width/length ratio with dimension scaling down. MeRAM always uses minimum sized transistors and hence promises better scalability in density. Alternatively, MeRAM can be integrated in a much denser cross-bar structure, unlike STT-RAM [16].

In terms of fabrication, both STT-RAM and MeRAM face the challenge scaling MgO thickness. Scaling dimension forces STT-MTJs to reduce MgO thickness, and thin MgO may contain defects[46], such as pin-holes, which can cause MTJs to fail. Though MeRAM has thicker MgO because of the high resistance of VC-MTJs, increasing write voltage may cause MgO breakdown. Again, these challenges can be overcome by finding better materials with higher g and ζ .

IV. MRAM CELL DESIGN AND VARIATION

As discussed in Section III, both STT-MTJ and VC-MTJ face scaling problems, and enlarging MTJs exacerbates write difficulty but does not improve thermal stability due to the sub-volume excitation [48]. Considering these factors, we set the

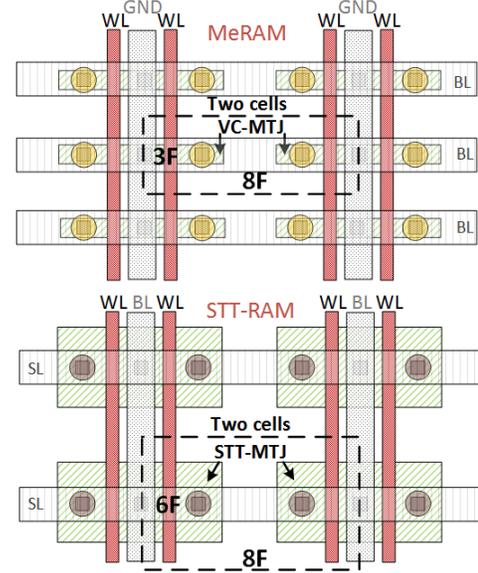


Fig. 4: Layouts of STT-RAM and MeRAM under 32nm design rules. The area of an STT-RAM cell is twice the area of a MeRAM cell, as an STT-MTJ requires a 3X wider access transistor than a VC-MTJ. Vertical transistor like nanowire may help to reduce area inefficiency [47]

diameter of STT-MTJs and VC-MTJs to 60nm (i.e., a circular MTJ structure), which has been demonstrated [49] for STT-MTJs. Access transistors and peripheral circuit are built with 32nm planar CMOS technology. The layouts of STT-RAM and MeRAM are drawn in Fig. 4 under 32nm design rules. The density of MeRAM is twice of STT-RAM for the reason that STT-MTJ needs 3X wider access transistors.

Table II lists the design parameters of MRAM cells (nominal and variation). In an MRAM cell, an MTJ is connected with an access transistor (1T1M). The MTJ resistance variation due to the MTJ shape variation was identified as a big design concern [21, 24, 54, 55]. But it is actually a secondary problem of the re-deposition of etched products on the MTJ sidewalls in current plasma-based etching system, where the re-deposition may cause an MTJ failure [56] and a shape

TABLE II: Design parameters for MTJs and access transistors. The transistors' threshold voltage variation considers the effects of line edge roughness (LER), random dopant fluctuation (RDF), and non-rectangular gate (NRG). Access transistors of MeRAM have larger threshold voltage variation because narrow transistors are affected more by NRG, RDF, and LER.

Devices	Parameters	Mean	Variation
STT-MTJ	Diameter	60nm	$\sigma=1\text{nm}$ [50]
	MgO thickness	0.7nm	$\sigma=0.001\text{nm}$ [51]
	T_{FL}	1.20nm	$\sigma=0.003\text{nm}$ [52]
	Thermal stability	71.6 (51.9@350K)	$\sigma=3.0$ (2.3@350K)
	Resistance	1K Ω / 2K Ω	dependence
	Cell area	24F ² (F: MTJ diameter)	
VC-MTJ	Diameter	60nm	$\sigma=1\text{nm}$ [50]
	MgO thickness	1.3nm	$\sigma=0.001\text{nm}$ [51]
	T_{FL}	1.19nm	$\sigma=0.003\text{nm}$ [52]
	Thermal stability	73.7 (53.6@350K)	$\sigma=3.1$ (2.3@350K)
	Resistance	100K Ω / 200K Ω	dependence
	Cell area	12F ² (F: MTJ diameter)	
Access transistor	Length	30nm	$\sigma=2.1\text{nm}$ [6]
	Width	200nm(STT) 48nm(Me)	$\sigma=2.1\text{nm}$ [6]
	Threshold voltage	493mV, LER, RDF, NRG [6, 53]	$\sigma=22.6\text{mV}$ (STT) $\sigma=42.4\text{mV}$ (Me)

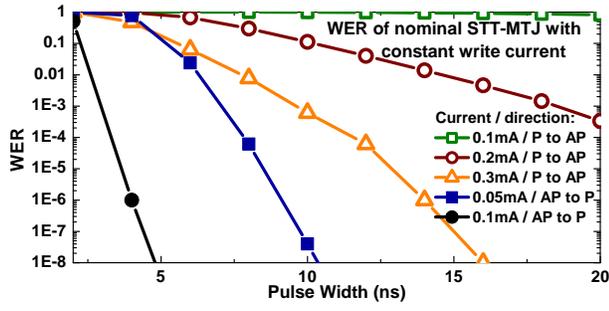


Fig. 5: WER of the nominal STT-MTJ as a function of pulse width for different perfect current pulses (constant current) and switching directions.

change. Developing selective-etching process is expected to fix both problems by forming volatile compound during a etch. Less than 4% in size variation has been shown for fabricated 50nm STT-RAM [50]. We pessimistically choose $\sigma = 1nm$ in our simulation where 6σ is 10% of the MTJ diameter.

The designed MTJs have thermal stability margins of 10 σ at 300K and 5 σ at 350K for the requirement of 40.3 [6] (i.e., 10 years retention time). External magnetic field in VC-MTJs assists the precessional switching, but reduces thermal stability, and hence the free layer thickness of VC-MTJs is set thinner to offset the thermal stability loss.

Comparing with the MTJ, CMOS variation has been well analyzed. Major variations in the 32nm planar technology are considered in Table II. FinFET and Tunneling FET technologies [57]–[59], which is possibly introduced for scaled MRAM, shows slightly smaller impact from process variation [60].

V. WRITE ERROR RATE OF MRAM

The reliability problems of MRAMs include retention error, read disturbance, read failure, and write error. We focus on the write error in this section, which is our main contribution, and the other failures are discussed in Section VI-B.

A. Write Error Rate of MTJs without Variation

Fig. 5 shows the WER of the nominal STT-MTJ. The two switching directions have different WER due to the asymmetric polarization efficiency (6). When the STT-MTJ switches from AP to P, the polarizing current changes from the majority to minority, while the polarizing current changes from the minority to majority in the opposite direction.

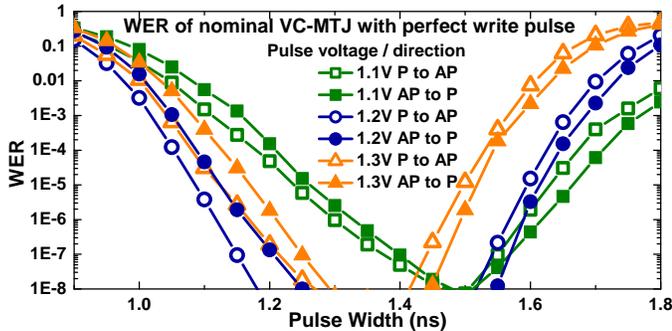


Fig. 6: The WER of the nominal VC-MTJ as a function of pulse width for different perfect voltage pulses and switching directions. A VC-MTJ has an optimal pulse, which leads to the lowest WER. The curve of 1.2V has the lower overall WER than 1.1V and 1.3V, indicating 1.2V is closer to the optimal voltage.

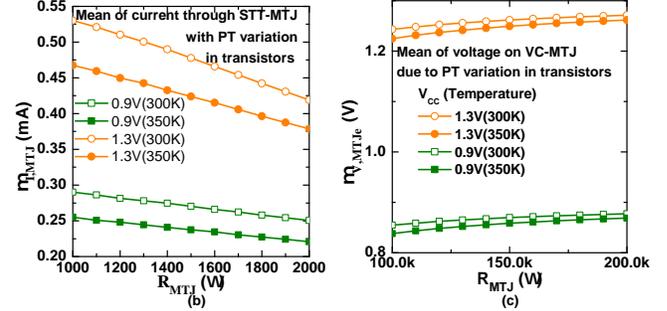
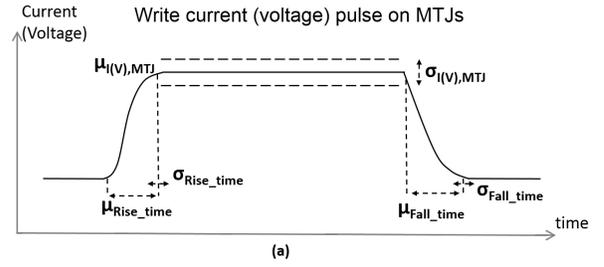


Fig. 7: (a) Write current (voltage) pulse on STT-MTJs (VC-MTJs). The rise and fall time are measured by the time while voltage is rising and falling between 10% and 90% of the peak voltage respectively. (b) Mean of write current on STT-MTJs as a function of MTJ resistance. (c) mean of write voltage on VC-MTJs as a function of MTJ resistance.

The WER of the nominal VC-MTJ is shown in Fig. 6. The curve of 1.2V is observed to have lower overall WER (for different pulse widths) than 1.1V and 1.3V indicating that it is closer to the optimal voltage. A non-optimal voltage either under-compensate or over-compensate the PMA, resulting in an imperfect precessional switching and thus a higher WER. As can be seen in Fig. 6, the low WER region of 1.3V averagely locates left (shorter pulse width) to 1.2V and 1.1V, as 1.3V over-compensates the PMA more to result in a faster precessional switching. Small WER asymmetry is observed for the two switching directions, because the write voltage induces leakage current and corresponding STT effect, which assists the switching from P to AP but resists the switching from AP to P.

B. Write Error Rate of MRAM Array

To estimate the WER of an entire array with temperature and process variations, WER must be simulated for different cells that have varying design parameters.

The variations of access transistors result in variation of write pulse voltage, rise and fall time as is shown in Fig.7 a. We obtain the distribution of the write pulse using Monte-Carlo SPICE simulations. In simulations, an access transistor is connected with a resistor and a capacitor (a lumped model TABLE III: Summary of write pulse variation due to transistor process variation at temperature of 300°C and 350°C. Mean shift is the percentage change of parameters' mean between high and low MTJ resistance states

MTJ	Parameters	mean shift	σ/μ
STT-MTJ	I_{MTJ}	< 26.5%	< 16%
	Rise time	< 7.0%	< 10.6%
	Fall time	< 11.5%	< 14.1%
VC-MTJ	V_{MTJ}	< 3.5%	< 1.0%
	Rise time	< 3.6%	< 11.1%
	Fall time	< 88.1%	< 7.3%

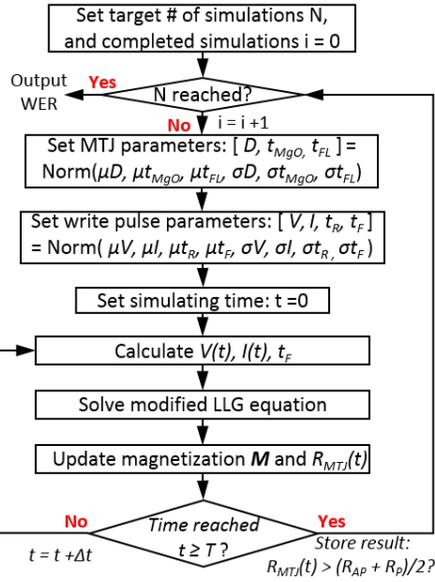


Fig. 8: A Monte-Carlo simulation flow to obtain the WER of 1T1M MRAM array. N is the sample size, and T is the simulation time including a writing time and a waiting time (for the MTJ to settle down, e.g., waiting time is 20ns in the simulations).

for the MTJ [61]). The parameters of access transistors are randomly determined based on Table II. Then the distribution of pulse current (current through STT-MTJs), pulse voltage (voltage on VC-MTJs), pulse rise time, and pulse fall time are statistically extracted from 100,000 simulations for each MTJ resistance state (i.e., resistance changes during switching) and V_{CC} (the supply voltage between 0.9V to 1.3V, which drops over an access transistor and an MTJ in series). The standard deviation (σ), mean (μ) of pulse current and voltage vary with MTJ resistance. As is shown in Fig. 7 b and c, the μ of pulse current changes up to 26.5% with STT-MTJ resistance, whereas the μ of pulse voltage only changes below 3.5% with VC-MTJ resistance because that the high resistance of VC-MTJs drops more than 95% of the V_{CC} . The σ/μ of pulse current in STT-RAM is up to 16%, whereas the σ/μ of pulse voltage in MeRAM is below 1% for the reason that in STT-RAM the pulse current is mainly controlled by access transistors and thus suffers more impact from transistor variation. The σ , μ of pulse rise time are mainly determined by access transistors, which barely do not depend on MTJ resistance. For a given V_{CC} , the μ varies within 7% with resistance, and the σ/μ is around 10%. By contrast, the fall time is mostly determined by the leaking current through MTJs. In STT-RAM the μ of pulse fall time varies between 61.6ps and 70.6ps with MTJ resistance, while the μ for MeRAM is much longer and varies between 128ps and 248ps because of the high resistance of VC-MTJs. The σ/μ of pulse fall time due to transistor variation is around 9% (maximum 14.1%) for STT-RAM and 5% (maximum 7.3%) for MeRAM for different V_{CC} and temperatures. We summarize the write pulse variation in Table III.

The σ , μ of pulse current/voltage, rise/fall time are fitted to polynomial models of MTJ resistance, which are accurate enough as their dependence on resistance is nearly linear. The models are inputs to the CUDA simulator. The Monte-Carlo simulation flow is shown in Fig. 8. At the beginning of a

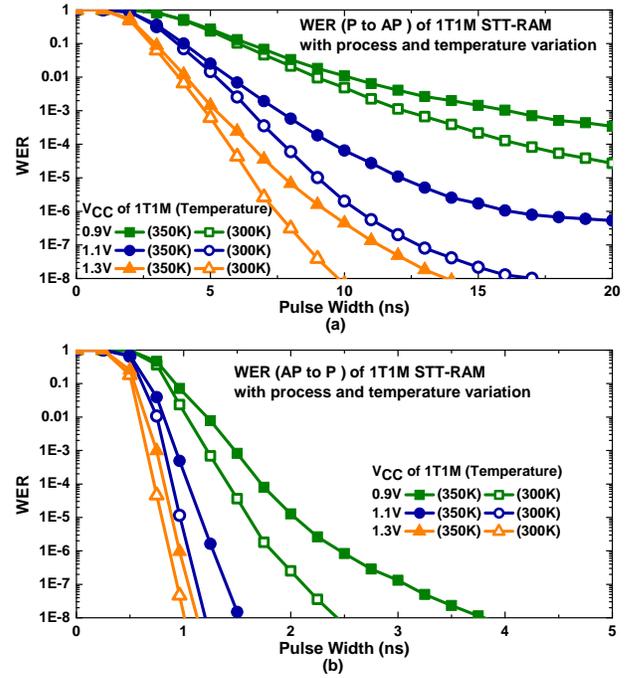


Fig. 9: The WER of an STT-RAM under process and temperature variation for different write pulses and switching directions (a: P to AP , b: AP to P).

simulation, pulse voltage/current, rise/fall time, and MTJ parameters are generated following Normal distribution. During the simulation, pulse voltage, current, and fall time are updated according to the MTJ resistance state.

The WER of a 1T1M STT-RAM is shown in Fig. 9. As expected, the WER of STT-RAM decreases monotonically with increasing V_{CC} and pulse width, and the WER increases with temperature. The switching from P to AP shows higher WER due to the asymmetry in spin-torque polarization efficiency.

The WER of a 1T1M MeRAM is shown in Fig. 10. Unlike the results of the nominal VC-MTJ, the MeRAM with process variation cannot achieve WER below 10^{-8} because there is no common optimal voltage for all VC-MTJs in the MeRAM. Temperature also has significant impact on the MeRAM: the WER of the pulse (1.26V/1.42ns), which gives the lowest WER at 300K, increases 1000X at 350K; the voltage that gives the lowest WER changes from 1.26V at 300K to 1.20V at 350K because high temperature leads to low thermal stability

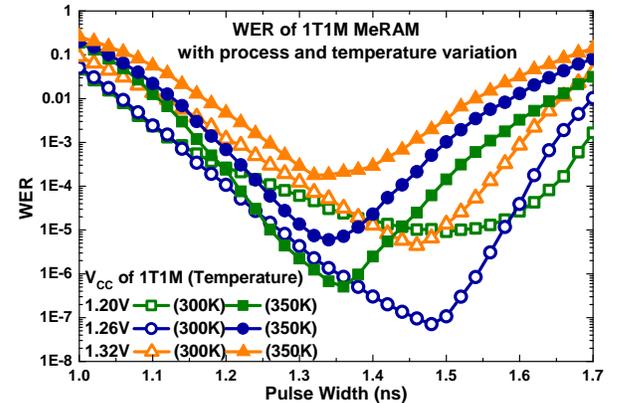


Fig. 10: WER of an MeRAM under process and temperature variation for different write pulses. The WER is averaged over two switching directions.

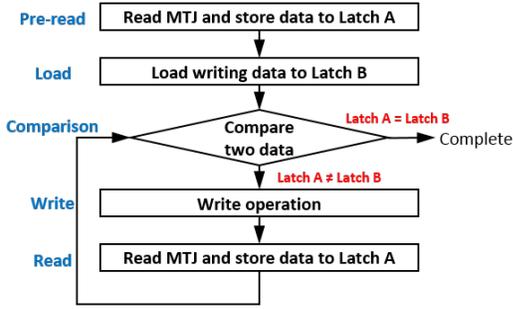


Fig. 11: Data program flow for the PWSA multi-write design.

and thus low required voltage (8); the pulse width giving the lowest WER for a given voltage decreases with temperature for the reason that higher temperature reduces the horizontal demagnetization field, then the external field is less canceled and drives the precessional switching faster. Despite of the high WER, MeRAM shows clear speed advantage.

VI. CIRCUIT-LEVEL EVALUATION

A. MRAM Write/Read with PWSA Multi-write Design

As well desired, peripheral circuit can improve the reliability of MRAMs at the expense of speed, power, and area [62]. Memory with multi-write schemes can significantly reduce write errors, e.g., incremental step pulse programming for Flash technology [63]. We utilize PWSA [25] to reduce the WER of MRAMs, where PWSA is a peripheral circuit designed for STT-RAM's and MeRAM's write and read operations. As MeRAM uses an unidirectional pulse to write both "1" and "0", PWSA uses an operation called pre-read to check the stored data prior to a write, and no write is performed if the stored data matches the writing one. In addition to pre-read, PWSA also enables multi-write policy which can perform additional write after a write error. The data flow in Fig. 11 illustrates the PWSA multi-write design. It is noticed that read failure has been well analyzed in [21, 55, 64], which is mainly caused by process variation and is a permanent failure (not like write error). Read failure can be eliminated by chip test at the expense of yield loss. All pre-read, comparison, and read share one sense amplifier and are assumed to be error-free operations in the PWSA multi-write design.

We divide the multi-write flow into four steps: read (also pre-read, including pre-charge and sensing), load, comparison, and write (including control of logic circuit and MTJ switching). To obtain reasonable delay and energy consumption of the peripheral circuit, each sense amplifier is connected to a bit-line of 256 1T1M cells and a ref bit-line. The delay and energy for these steps are extracted from Spectre simulation of PWSA circuit [25] using 32nm PTM HP model [65] and are listed in Table IV. To guarantee a good pulse shape for the write in MeRAM, a pre-charge operation is performed to raise the bit-line voltage to V_{CC} prior to turning on access transistors, where the pre-charge takes around 0.15ns. Though STT-MTJs do not have strict requirement on the pulse shape, STT-RAM needs to raise bit-line voltage to bias access transistors to offer required write current, which takes similar delay and consumes more power due to its larger access transistors. Conversely, the read energy of STT-RAM is lower, because the low resistance of STT-MTJ allows lower read

TABLE IV: Energy and delay for operations in the PWSA multi-write circuit at 300K temperature.

Operations	Energy (fJ)		Delay (ns)	
	STT-RAM	MeRAM	STT-RAM	MeRAM
Read (Pre-read)	54.7	91.0	1.8	2.0
Load	122.4	122.4	1.3	1.3
Comparison	15.2	15.2	0.4	0.4
Write (logic)	691	318	0.5	0.5
Write (MTJ)	680/ns	10.1	≥ 3	1.42

voltage than VC-MTJ. The difference of read delay is within 0.2ns. MeRAM shows great advantage in the MTJ switching energy, whereas this energy is a bottleneck for STT-RAM due to the high leakage current caused by the low resistance of STT-MTJs. The switching energy can be reduced by pre-read, though pre-read is not mandatory for STT-RAM (i.e., the STT-MTJ can be directly written with directional current).

We utilize the PWSA multi-write design to achieve reliable STT-RAM and MeRAM with the same acceptable WER and then compare the expected latency and energy of writing a word. The acceptable WER is $< 10^{-23}$ for a cell in one multi-write operation, which is slightly smaller than the soft error rate in DRAM technology [66] and can be handled by error-correction code (ECC) designs. A word is a set of bits to be written in parallel, and the size of word varies from storage devices, e.g., a main memory usually writes 64 bits in parallel. The expected write latency/energy is the sum of products of the delay/energy and probability for all possible scenarios (i.e., different numbers of writes for different number of bits). In our calculations, we assume the memory system to store "1"s and "0"s in balance, which means there is 50% probability that the bit to be written equals to the bit stored in the target MRAM cell. As the multi-write design writes only to the cells that fail in the pre-read check or previous writes, so the expected energy does not count the write energy of the cells that do not need writes (i.e., energy of the pre-read and comparison is always counted). There is a maximum allowed write times per multi-write operation to avoid infinite writes

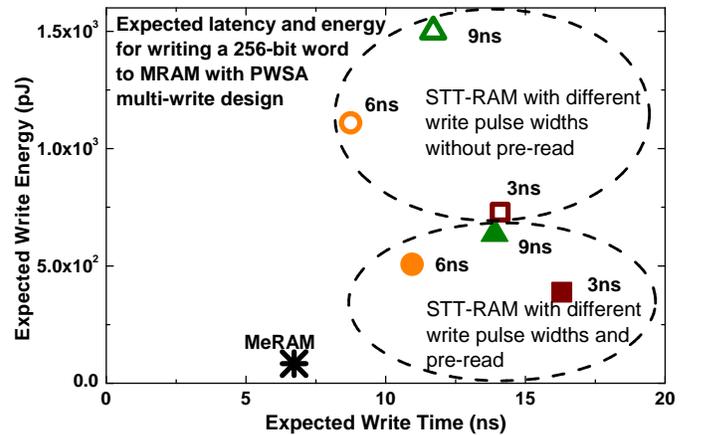


Fig. 12: Expected word-write energy and latency for MeRAM and STT-RAM with PWSA multi-write circuit. The word size is 256bits, which are the number of bits being simultaneously written. The WER/bit after multiple writes is minimize below 10^{-23} . The labels 3ns, 6ns, and 9ns on STT-RAM are single write pulse widths. The pre-read is not mandatory for STT-RAM. The top circled designs are STT-RAMs without pre-read operation. The bottom circled ones are STT-RAMs with pre-read operation, which count the overhead of pre-read but save unnecessary write.

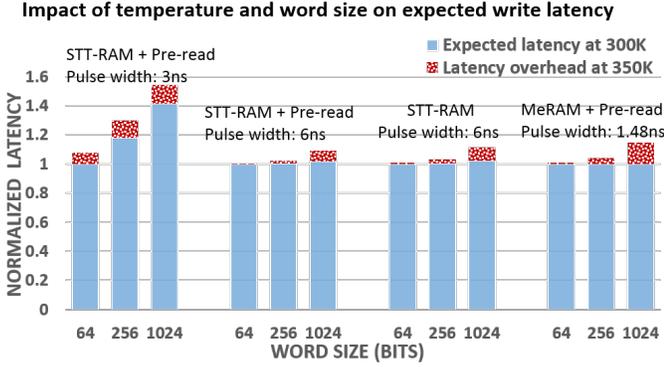


Fig. 13: Impact of word size and temperature on the expected write latency and energy of MRAMs. All bars are normalized within each group to the expected write latency of MRAM at 300K with 64-bit word size.

in case that permanently failed cells exist, which is calculated by the number of writes to achieve the acceptable WER of 10^{-23} , e.g., three for MeRAM at 300K.

We tried multiple write configurations for STT-RAM to explore tradeoffs between write latency and energy: with pre-read, without pre-read, and different write pulse widths for switching from P to AP including 3ns, 6ns, and 9ns. For switching from AP to P , the pulse width is set to 3ns, which exactly achieves the acceptable WER. The V_{CC} for STT-RAM is chosen to 1.3V, which is the most efficient one in Fig. 10a. The write pulse used for MeRAM is the optimized pulse at 300K (1.26V/1.42ns, see Fig. 10b). The pre-read is mandatory for MeRAM.

Fig. 12 shows the expected write energy and latency for STT-RAM and MeRAM with PWSA multi-write design at 300K. Benefiting from the fast and energy-efficient switching of VC-MTJs, MeRAM shows substantial advantages of both speed and energy against STT-RAM. Among configurations for STT-RAM, the one with the shortest pulse width of 3ns and pre-read gives the lowest expected energy, as most cells can pass the comparison check in the pre-read and the first write. Nevertheless, it has the longest expected latency for that the pre-read adds latency and its high WER leads to the most write errors and write iterations. The STT-RAM with 6ns pulse without pre-read shows the fastest speed, as a comparison, the 9ns pulse has lower WER but higher latency and energy, because the benefit of the lower WER does not compensate the overhead brought by its longer pulse. This also indicates that directly using a long pulse in STT-RAM to guarantee zero WER is not an energy-latency efficient design. The energy-latency Pareto fronts of STT-RAMs are 3ns with pre-read, 6ns without pre-read, and 6ns with pre-read.

Fig. 13 shows the impact of word size and temperature on the write latency. A longer word usually leads to more write iterations as more cells are written giving rise to more write errors. The STT-RAM with the 3ns pulse is affected most by the word size due to its high WER. The STT-RAM with 3ns pulse and MeRAM are affected most by temperature, as their WERs increase the most from 300K to 350K (i.e., the WER of a MeRAM cell increases from 9×10^{-8} to 2.2×10^{-4} , and the WER of an STT-RAM cell with 3ns pulse increases from 0.06 to 0.09). Moreover, the temperature induced overhead

TABLE V: Failure types and FIT for a 16MB memory bank. 10^9 reads and 10^9 writes in a bank-hour are assumed. The read disturbance rate is extrapolated from simulations. As a comparison, the FIT of single-bit fault in a 16MB bank is about $2 \cdot 10^{-4}$ [66], and the FIT of DDR bus errors is about 100 [68].

Failures	write errors	retention error	read failure	*read disturbance
Types	non-persistent	non-persistent	persistent	non-persistent
MeRAM	$< 10^{-5}$	< 0.58	< 0.0029	$4 \cdot 10^{-6}$
STT-RAM	$< 10^{-5}$	< 3.4	$< 10^{-15}$	$3 \cdot 10^{-43}$

increases with word size. Illustrated from the comparison of two STT-RAMs with 6ns pulse, pre-read can mitigate the impact of temperature variation for the reason that about 50% writes are saved if pre-read is enabled. For all MRAM designs, maximum 15% latency increase is shown due to temperature variation, whereas energy only shows maximum 2.3% increase at 350K against 300K. Among the energy overhead, a big portion comes from the energy of bit-line charging (i.e., 4% increase in this energy from 300K to 350K). Again, the small energy overhead is because that most cells only need one write.

The delay and power overhead of pre-read and comparison are simulated and listed in Table IV. In this section, we analyze the area overhead of PWSA. One PWSA contains 37 transistors and one regular STT-RAM sense amplifier contains 8 transistors [67]. Considering the bit-line size of 256 cells and four bit-lines sharing one sense amplifier, the area overhead is 2.7%. However, the area of design also depends on size of transistors. The transistors of sense amplifier for STT-RAM are much larger than those for MeRAM given that STT-RAM requires larger write current. Indeed the PWSA (37 transistors) for MeRAM occupies 20% less area than the regular sense amplifier (8 transistors) for STT-RAM.

B. Failure Analysis and Error Correction

As is listed in Table V, memory failures in STT-RAM and MeRAM are classified into four types: write errors, retention errors, read failure, and read disturbance. We use failure-in-time (FIT, average number of failures in a billion-device-hours) to represent the error rate for these failure types.

Write errors have been analyzed in the Sections V and VI-A. The multi-write scheme significantly reduces write error rate. Its FIT for an 16-MB MeRAM bank decreases from over 10^{10} without multi-write scheme to below 10^{-5} with multi-write scheme.

False switching of MTJs during idle state is called retention error. As is mentioned in Section III, the VC-MTJs and STT-MTJs have been designed with enough margin in thermal stability to minimize the retention error. The FIT of retention error in a 16-MB MRAM is calculated according to Table II and is listed in Table V.

Read failure due to the MTJ resistance variation [21, 55, 64] is a persistent memory failure which stay in memory and frequently produces errors. More specifically, large shape variation of an MTJ can lead to significant resistance change which decreases sensing margin and results in read errors. Read errors are produced in all reads to a failed MTJ (with large resistance change) indicating that the multi-write design also creates write errors due to the involved read step. However, the multi-write design does not increase read error rate for the fact that such write errors only occur on failed MTJs, and the

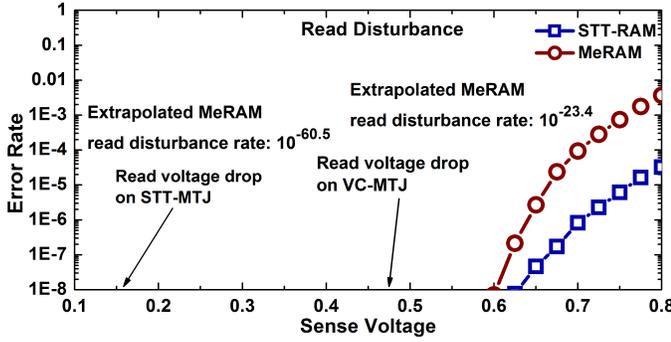


Fig. 14: Read disturbance rate as a function of read voltage. The read disturbance rate for MeRAM and STT-RAM are extrapolated to the read voltage drop on MTJs (0.48V and 0.15V are respectively for VC-MTJs and STT-MTJs).

failed MTJs are always read out incorrectly. The resistance change due to shape variation is mainly caused by wafer-level process variation [1], which can be minimized by increased TMR or recently developed peripheral circuit designs, e.g., the local-reference reading scheme [69] and the self-reference scheme [70]. In our experimental setup, the AP and P resistance for STT-RAM (MeRAM) are $2,000\Omega$ ($200,000\Omega$) and $1,000\Omega$ ($100,000\Omega$) respectively, and reference resistors are $1,500\Omega$ ($150,000\Omega$). Reference resistors are fabricated with traditional CMOS process, and their variation are negligible compared to MTJs. The MTJ resistance variation is assumed to follow Gaussian distribution [21, 55]. In [51], standard deviation of MTJ resistance is measured as 1.5% of mean resistance from a 4-Mb MRAM array. Accordingly, we calculate its MgO thickness variation in Table II and estimate the resistance standard deviation of 2.6% for STT-MTJs and VC-MTJs with the diameter of 60nm. By setting 0.05V sensing margin for sense amplifiers to operate functionally (i.e., 0.05V is enough for the limited variation of large sized sense amplifiers), our sensing scheme (using PWSA and 2ns sensing time) can tolerate 17.5% resistance variation (i.e., STT-RAM: 20% in AP and 45% in P, MeRAM: 17.5% in AP and 40% in P). It is noticed that we consider both access transistor variation and MTJ shape variation, but the access transistor variation has negligible impact due to the low sensing voltage (0.2V for STT-RAM and 0.48V for MeRAM), where sensing current is dominated by MTJs. The read failure rate of an MTJ due to resistance variation is $1.75 \cdot 10^{-10}$, which gives rise to 99.22% yield for a 16-MB bank array. Redundancy technique of sparing columns is a common technology for yield improvement. By adding one sparing column (every column has 256 cells) to every mat (contains multiple rows and columns, e.g., a 16-MB bank has 64 mats, and every mat has 256 rows and 8192 columns), the yield of a 16-MB memory bank is improved to 99.994% with 0.01% area overhead.

The failure rate of read disturbance is close to zero when short read pulse ($< 2ns$) and low read voltage (0.48V on VC-MTJs, and 0.15V on STT-MTJs) are used [24]. We have simulated the read disturbance of MeRAM and STT-RAM as functions of read voltage with the precision of 10^{-9} using the CUDA LLG-based model and extrapolated the read disturbance to our read voltage using polynomial models as shown in Fig. 14.

Error-correction-code (ECC) is a common technique to

TABLE VI: Write/read latency/energy for one write/read in a x8 16-MB STT-RAM and MeRAM banks. One write/read operates on 64 bits (72bits in memory banks for in-memory ECC detection and correction) in a row in burst mode.

Memory	Write		Read		Area
	latency	energy	latency	energy	
MeRAM	9.4 ns	271.0 pJ	5.0 ns	210.3 pJ	9.5 mm^2
STT-RAM	17.2 ns	831.7 pJ	11.9 ns	293.2 pJ	17.0 mm^2

protect memory from memory errors. We use MEMRES (a fast system-level memory reliability simulator) [71] to simulate and analyze the impact of MRAM introduced failures on an 8-GB memory. The memory is comprised of 512 16-MB banks and is protected by in-memory ECC (SECDED) (i.e., locates in memory banks) and in-controller SECDED/Chipkill (i.e., locates in memory controller) [72, 73]. MRAM introduced failures (listed in Table V) are included in MEMRES simulations in addition to typical memory logic-circuit induced failures (e.g., bank failure, row failure, column failure, and etc.[71]). Based on simulated results, the probability that MRAM introduced failures cause an ECC uncorrectable error in an 8-GB memory is $< 0.0001\%$ for 5-year operating time (i.e., no such error is found in 10,000,000 5-year memory reliability simulations), indicating that traditional ECC designs are strong enough to handle failures in MeRAM and STT-RAM with PWSA multi-write design.

C. Latency, Energy, and Area of a 16-MB MRAM Bank

In order to include the energy and latency of ECC designs, we compare STT-RAM and MeRAM in memory-bank level. With inputs of MTJ cell area (see Table II) and bit-line write/read latency/energy (see Table IV), we use NVSIM [74] to obtain the area, energy, and latency of a 16-MB STT-RAM bank and a 16-MB MeRAM bank. In-controller ECC commonly exists in current server-class processors for DRAM error detection and correction. In-memory ECC is a new technology, which correct errors individually in memory banks. We only count the power and latency of in-memory ECC in our STT-RAM and MeRAM comparison, because in-memory ECC can correct all MRAM introduced failures, and in-controller ECC is already used for current memory technologies. The in-memory ECC detection and correction latency are about 0.34ns and 4.4ns respectively [75], and encoding latency is assumed to be 0.3ns (i.e., should be little shorter than detection). The energy of encoding, detection, and correction is below 1 pJ per access, which are ignored compared to other memory components. The area overhead of in-memory ECC is about 12.5%.

We summarize the area, latency, and energy of one access to 16-MB STT-RAM and MeRAM banks in Table VI (every access is 64 bits in burst mode). Benefited by the smaller size of VC-MTJs, MeRAM has smaller bank area, shorter interconnect, and smaller sized peripheral circuits, which turn into less energy and shorter latency in the logic operations like row decoding and MUX selection. The bank read latency and energy are dominated by these logic operations, thus MeRAM shows faster read speed and less read energy. For write operation, VC-MTJs' small cell size, shorter write pulse, and less write energy jointly build the advantages of both write energy and latency.

VII. CONCLUSION

We comprehensively compare the two promising non-volatile magnetic memory technologies, STT-RAM and MeRAM, in the circuit context with respect to reliability, energy, speed, area, and scalability. MeRAM has higher WER than STT-RAM under process and temperature variation, but by utilizing a multi-write design, both MRAMs are able to achieve an acceptably low WER. With clear advantages of MTJ switching delay and energy, MeRAM outperforms STT-RAM by 83% in write speed, 67.4% in write energy, 138% in read speed, and 28.2% in read energy. In terms of density, VCMA allows to use minimum sized access transistors, which helps MeRAM to achieve twice the density of STT-RAM at 32nm node, and the density advantage is expected to increase at smaller nodes indicating that MeRAM has better density scalability. With respect to challenge of technology scaling down, simply shrinking dimension does not save energy and introduces fabrication defects for both technologies; more effort should be spent on discovering materials with higher polarization efficiency.

ACKNOWLEDGMENT

The authors would like to thank Richard Dorrance, Dejan Markovic, Juan Alzate, Jayshankar Nath, Dheeraj Srinivasan, and Daniel Matic for their help and contribution in developing the LLG equation based model and simulator. This work was supported in part by the NSF Engineering Research Center on Translational Applications of Nanoscale Multiferroic Systems (TANMS), and in part by a Phase II SBIR award from the National Science Foundation.

REFERENCES

- [1] Said Tehrani et al. "Progress and outlook for MRAM technology". *Magnetics, IEEE Transactions on* 35.5 (1999), pp. 2814–2819.
- [2] B Fang et al. "Tunnel magnetoresistance in thermally robust Mo/CoFeB/MgO tunnel junction with perpendicular magnetic anisotropy". *AIP Advances* 5.6 (2015), p. 067116.
- [3] Witold Skowroński et al. "Underlayer material influence on electric-field controlled perpendicular magnetic anisotropy in CoFeB/MgO magnetic tunnel junctions". *Physical Review B* 91.18 (2015), p. 184410.
- [4] C Heide. "Spin currents in magnetic films". *Physical review letters* 87.19 (2001), p. 197201.
- [5] DC Worledge et al. "Spin torque switching of perpendicular Ta form". *Applied Physics Letters* 98.2 (2011), p. 2501.
- [6] ITRS. <http://www.itrs.net/about.html>. 2008,2011.
- [7] Zhenyu Sun et al. "Multi retention level STT-RAM cache designs with a dynamic refresh scheme". *proceedings of the 44th annual IEEE/ACM international symposium on microarchitecture*. ACM. 2011, pp. 329–338.
- [8] Clinton W Smullen et al. "Relaxing non-volatility for fast and energy-efficient STT-RAM caches". *High Performance Computer Architecture (HPCA), IEEE 17th International Symposium on*. IEEE. 2011, pp. 50–61.
- [9] Wei Xu et al. "Design of last-level on-chip cache using spin-torque transfer RAM (STT RAM)". *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on* 19.3 (2011), pp. 483–493.
- [10] Adwait Jog et al. "Cache revive: architecting volatile STT-RAM caches for enhanced performance in CMPs". *Proceedings of the 49th Annual Design Automation Conference*. ACM. 2012, pp. 243–252.
- [11] Emre Kultursay, Mahmut Kandemir, Anand Sivasubramaniam, and Onur Mutlu. "Evaluating STT-RAM as an energy-efficient main memory alternative". *ISPASS*. IEEE. 2013, pp. 256–267.
- [12] Steven Pelley, Peter M Chen, and Thomas F Wenisch. "Memory persistency". *Proceeding of the 41st annual international symposium on Computer architecture*. IEEE Press. 2014, pp. 265–276.
- [13] Ki Chul Chun et al. "A scaling roadmap and performance evaluation of in-plane and perpendicular MTJ based STT-MRAMs for high-density cache memory". *Solid-State Circuits, IEEE Journal of* 48.2 (2013), pp. 598–610.
- [14] P Khalili Amiri, P Upadhyaya, JG Alzate, and KL Wang. "Electric-field-induced thermally assisted switching of monodomain magnetic bits". *Journal of Applied Physics* 113.1 (2013), p. 013912.
- [15] Juan G Alzate et al. "Voltage-induced switching of nanoscale magnetic tunnel junctions". *Electron Devices Meeting (IEDM), IEEE International*. IEEE. 2012, pp. 29–5.
- [16] Richard Dorrance et al. "Diode-MTJ crossbar memory cell using voltage-induced unipolar switching for high-density MRAM". *Electron Device Letters, IEEE* 34.6 (2013), pp. 753–755.
- [17] S Kanai et al. "Electric field-induced magnetization reversal in a perpendicular-anisotropy CoFeB-MgO magnetic tunnel junction". *Applied Physics Letters* 101.12 (2012), p. 122403.
- [18] Yoichi Shiota et al. "Pulse voltage-induced dynamic magnetization switching in magnetic tunneling junctions with high resistance-area product". *Applied Physics Letters* 101.10 (2012), p. 102406.
- [19] Yoichi Shiota et al. "Induction of coherent magnetization switching in a few atomic layers of FeCo using voltage pulses". *Nature materials* 11.1 (2012), pp. 39–43.
- [20] Wei-Gang Wang, Mingen Li, Stephen Hageman, and CL Chien. "Electric-field-assisted switching in magnetic tunnel junctions". *Nature materials* 11.1 (2012), pp. 64–68.
- [21] Jing Li, Charles Augustine, Sayeef Salahuddin, and Kaushik Roy. "Modeling of failure probability and statistical design of spin-torque transfer magnetic random access memory (STT MRAM) array for yield enhancement". *Design Automation Conference, 2008. DAC 2008. 45th ACM/IEEE*. IEEE. 2008, pp. 278–283.
- [22] Zihan Xu et al. "Compact modeling of STT-MTJ devices". *Solid-State Electronics* 102 (2014), pp. 76–81.

- [23] Peiyuan Wang et al. “A thermal and process variation aware MTJ switching model and its applications in soft error analysis”. *Proceedings of the International Conference on Computer-Aided Design*. IEEE. 2012, pp. 720–727.
- [24] Yaojun Zhang, Xiaobin Wang, and Yiran Chen. “STT-RAM cell design optimization for persistent and non-persistent error rate reduction: a statistical design view”. *Proceedings of the International Conference on Computer-Aided Design*. IEEE Press. 2011, pp. 471–477.
- [25] H. Lee et al. “Design of a Fast and Low-Power Sense Amplifier and Writing Circuit for High-Speed MRAM”. *Magnetics, IEEE Transactions on* 51.5 (2015), pp. 1–7.
- [26] R Sbiaa et al. “Reduction of switching current by spin transfer torque effect in perpendicular anisotropy magnetoresistive devices”. *Journal of Applied Physics* 109.7 (2011), p. 07C707.
- [27] Yue Zhang et al. “Compact modeling of perpendicular-anisotropy CoFeB/MgO magnetic tunnel junctions”. *Electron Devices, IEEE Transactions on* 59.3 (2012), pp. 819–826.
- [28] KJ Lee, Olivier Redon, and Bernard Dieny. “Analytical investigation of spin-transfer dynamics using a perpendicular-to-plane polarizer”. *Applied Physics Letters* 86.2 (2005), p. 022505.
- [29] Yiming Huai. “Spin-transfer torque MRAM (STT-MRAM): Challenges and prospects”. *AAPPS Bulletin* 18.6 (2008), pp. 33–40.
- [30] Juan G Alzate et al. “Temperature dependence of the voltage-controlled perpendicular anisotropy in nanoscale MgO—CoFeB—Ta magnetic tunnel junctions”. *Applied Physics Letters* 104.11 (2014), p. 112410.
- [31] Xiaoyang Zhang et al. “An electrothermal/electrostatic dual driven MEMS scanner with large in-plane and out-of-plane displacement”. *Optical MEMS and Nanophotonics (OMN), 2013 International Conference on*. IEEE. 2013, pp. 13–14.
- [32] Xiaoyang Zhang, Boxiao Li, Xingde Li, and Huikai Xie. “A robust, fast electrothermal micromirror with symmetric bimorph actuators made of copper/tungsten”. *Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS), 2015 Transducers-2015 18th International Conference on*. IEEE. 2015, pp. 912–915.
- [33] John C Slonczewski. “Current-driven excitation of magnetic multilayers”. *Journal of Magnetism and Magnetic Materials* 159.1 (1996), pp. L1–L7.
- [34] GD Fuchs et al. “Adjustable spin torque in magnetic tunnel junctions with two fixed layers”. *Applied Physics Letters* 86.15 (2005), p. 152509.
- [35] Jack C Sankey et al. “Measurement of the spin-transfer-torque vector in magnetic tunnel junctions”. *Nature Physics* 4.1 (2008), pp. 67–71.
- [36] M Hosomi et al. “A novel nonvolatile memory with spin torque transfer magnetization switching: Spin-RAM”. *Electron Devices Meeting (IEDM), IEEE International*. IEEE. 2005, pp. 459–462.
- [37] F Bonell et al. “Large change in perpendicular magnetic anisotropy induced by an electric field in FePd ultrathin films”. *Applied Physics Letters* 98.23 (2011), p. 232510.
- [38] Takayuki Nozaki et al. “Voltage-Induced Magnetic Anisotropy Changes in an Ultrathin FeB Layer Sandwiched between Two MgO Layers”. *Applied Physics Express* 6.7 (2013), p. 073005.
- [39] Wei Wu et al. “Exploiting parallelism by data dependency elimination: A case study of circuit simulation algorithms” (2012).
- [40] Xiaoming Chen et al. “An escheduler-based data dependence analysis and task scheduling for parallel circuit simulation”. *Circuits and Systems II: Express Briefs, IEEE Transactions on* 58.10 (2011), pp. 702–706.
- [41] Wei Wu et al. “FPGA accelerated parallel sparse matrix factorization for circuit simulations”. *Reconfigurable Computing: Architectures, Tools and Applications*. Springer, 2011, pp. 302–315.
- [42] Richard F Freitas and Winfried W Wilcke. “Storage-class memory: The next storage system technology”. *IBM Journal of Research and Development* 52.4.5 (2008), pp. 439–447.
- [43] Anurag Nigam et al. “Delivering on the promise of universal memory for spin-transfer torque RAM (STT-RAM)”. *Proceedings of the 17th IEEE/ACM international symposium on Low-power electronics and design*. IEEE Press. 2011, pp. 121–126.
- [44] ND Rizzo et al. “Thermally activated magnetization reversal in submicron magnetic tunnel junctions for magnetoresistive random access memory”. *Applied Physics Letters* 80.13 (2002), pp. 2335–2337.
- [45] Satoshi Ohuchida, Kenchi Ito, and Tetsuo Endoh. “Impact of sub-volume excitation on improving overdrive delay product of sub-40 nm perpendicular magnetic tunnel junctions in adiabatic regime and its beyond”. *Japanese Journal of Applied Physics* 54.4S (2015), p. 04DD05.
- [46] Sidi Fu, V Lomakin, A Torabi, and B Lengsfeld. “Modeling perpendicular magnetic multilayered oxide media with discretized magnetic layers”. *Magnetics Conference (INTERMAG), 2015 IEEE*. IEEE. 2015, pp. 1–1.
- [47] Wei-Che Wang and Puneet Gupta. “Efficient layout generation and evaluation of vertical channel devices”. *Proceedings of the 2014 IEEE/ACM International Conference on Computer-Aided Design*. IEEE Press. 2014, pp. 550–556.
- [48] JZ Sun et al. “Effect of subvolume excitation and spin-torque efficiency on magnetic switching”. *Physical Review B* 84.6 (2011), p. 064413.
- [49] T Ohsawa et al. “1Mb 4T-2MTJ nonvolatile STT-RAM for embedded memories using 32b fine-grained power gating technique with 1.0 ns/200ps wake-up/power-off times”. *VLSI Circuits (VLSIC), Symposium on*. IEEE. 2012, pp. 46–47.

- [50] KT Nam et al. "Switching properties in spin transfer torque MRAM with sub-50nm MTJ size". *Non-Volatile Memory Technology Symposium*. IEEE. 2006, pp. 49–51.
- [51] Renu W Dave et al. "MgO-based tunnel junction material for high-speed toggle magnetic random access memory". *Magnetics, IEEE Transactions on* 42.8 (2006), pp. 1935–1939.
- [52] JM Slaughter et al. "Magnetic tunnel junction materials for electronic applications". *JOM(USA)* 52.6 (2000), p. 11.
- [53] Yun Ye, Frank Liu, Sani Nassif, and Yu Cao. "Statistical modeling and simulation of threshold variation under dopant fluctuations and line-edge roughness". *Design Automation Conference, 2008. DAC 2008. 45th ACM/IEEE*. IEEE. 2008, pp. 900–905.
- [54] EY Chen et al. "Comparison of oxidation methods for magnetic tunnel junction material". *Journal of Applied Physics* 87.9 (2000), pp. 6061–6063.
- [55] Jing Li, Haixin Liu, Sayeef Salahuddin, and Kaushik Roy. "Variation-tolerant Spin-Torque Transfer (STT) MRAM array for yield enhancement". *Custom Integrated Circuits Conference, 2008. CICC 2008. IEEE*. IEEE. 2008, pp. 193–196.
- [56] Jong-Yoon Park et al. "Etching of CoFeB Using CO/NH₃ in an Inductively Coupled Plasma Etching System". *Journal of The Electrochemical Society* 158.1 (2011), H1–H4.
- [57] G. Leung et al. "An Evaluation Framework for Nanotransfer Printing-Based Feature-Level Heterogeneous Integration in VLSI Circuits". *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on* PP.99 (2015), pp. 1–13. ISSN: 1063-8210. DOI: 10.1109/TVLSI.2015.2477282.
- [58] Shaodi Wang, Andrew Pan, Chi On Chui, and Puneet Gupta. "PROCEED: A pareto optimization-based circuit-level evaluator for emerging devices". *Design Automation Conference (ASP-DAC), 2014 19th Asia and South Pacific*. IEEE. 2014, pp. 818–824.
- [59] S. Wang, A. Pan, C.O. Chui, and P. Gupta. "PROCEED: A Pareto Optimization-Based Circuit-Level Evaluator for Emerging Devices". *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on* (2015). ISSN: 1063-8210. DOI: 10.1109/VeryLargeScaleIntegration(VLSI) Systems, IEEETransactionson.2015.2393852.
- [60] Shaodi Wang et al. "Evaluation of digital circuit-level variability in inversion-mode and junctionless FinFET technologies". *Electron Devices, IEEE Transactions on* 60.7 (2013), pp. 2186–2193.
- [61] Roy Scheuerlein et al. "A 10 ns read and write non-volatile memory array using a magnetic tunnel junction and FET switch in each cell". *IEEE International Solid-State Circuits Conference*. IEEE. 2000, pp. 128–129.
- [62] Wei Wu, Fang Gong, Gengsheng Chen, and Lei He. "A fast and provably bounded failure analysis of memory circuits in high dimensions". *Design Automation Conference (ASP-DAC), 2014 19th Asia and South Pacific*. IEEE. 2014, pp. 424–429.
- [63] Kang-Deog Suh et al. "A 3.3 V 32 Mb NAND flash memory with incremental step pulse programming scheme". *Solid-State Circuits, IEEE Journal of* 30.11 (1995), pp. 1149–1156.
- [64] Richard Dorrance et al. "Scalability and design-space analysis of a 1T-1MTJ memory cell for STT-RAMs". *Electron Devices, IEEE Transactions on* 59.4 (2012), pp. 878–887.
- [65] *PTM model*. <http://ptm.asu.edu/>.
- [66] Vilas Sridharan and Dean Liberty. "A study of DRAM failures in the field". *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*. IEEE Computer Society Press. 2012, p. 76.
- [67] Chia-Tsung Cheng, Yu-Chang Tsai, and Kuo-Hsing Cheng. "A high-speed current mode sense amplifier for Spin-Torque Transfer Magnetic Random Access Memory". *Circuits and Systems (MWSCAS), 53rd IEEE International Midwest Symposium on*. IEEE. 2010, pp. 181–184.
- [68] Noriyuki Miura, Kazutaka Kasuga, Mitsuko Saito, and Tadahiro Kuroda. "An 8Tb/s 1pJ/b 0.8 mm²/Tb/s QDR Inductive-Coupling Interface Between 65nm CMOS GPU and 0.1 μm DRAM". *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International*. IEEE. 2010, pp. 436–437.
- [69] UK Klostermann et al. "A perpendicular spin torque switching based MRAM for the 28 nm technology node". *Electron Devices Meeting (IEDM), IEEE International*. IEEE. 2007, pp. 187–190.
- [70] Yiran Chen et al. "A nondestructive self-reference scheme for spin-transfer torque random access memory (STT-RAM)". *Proc. DATE*. IEEE. 2010, pp. 148–153.
- [71] Shaodi Wang, Henry Hu, Hongzhong Zheng, and Puneet Gupta. "MEMRES: A Fast Memory System Reliability Simulator". *SELSE: the 12th Workshop on Silicon Errors in Logic - System Effects*. IEEE. 2015.
- [72] Mario Blaum, Rodney Goodman, and Robert McEliece. "The reliability of single-error protected computer memories". *Computers, IEEE Transactions on* 37.1 (1988), pp. 114–119.
- [73] Timothy J Dell. "A white paper on the benefits of chipkill-correct ECC for PC server main memory". *IBM Microelectronics Division* (1997), pp. 1–23.
- [74] Xiangyu Dong, Cong Xu, Yuan Xie, and Norman P Jouppi. "Nvsim: A circuit-level performance, energy, and area model for emerging nonvolatile memory". *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on* 31.7 (2012), pp. 994–1007.
- [75] Henry Duwe, Xun Jian, and Rakesh Kumar. "Correction prediction: Reducing error correction latency for on-chip memories". *High Performance Computer Architecture (HPCA), 2015 IEEE 21st International Symposium on*. IEEE. 2015, pp. 463–475.