Temperature Impact Analysis and Access Reliability Enhancement for 1T1MTJ STT-RAM

Bi Wu, *Student Member, IEEE,* Yuanqing Cheng, *Member, IEEE,* Jianlei Yang, *Member, IEEE,* Aida Todri-Sanial, *Member, IEEE,* and Weisheng Zhao, *Senior Member, IEEE*

***Abstract*—Spin-Transfer Torque Magnetic Random Access Memory (STT-RAM) is a promising and emerging technology due to its many advantageous features such as scalability, non-volatility, density, endurance and fast access speed. However, the operation of STT-RAM is severely affected by environmental factors such as process variations and temperature. As the temperature rockets up in modern computing systems, it is highly desirable to understand thermal impact on STT-RAM operations and reliability. In this paper, a thermal-aware MTJ model, calibrated and validated by experimental measurements, is proposed as the basis for thoroughly thermal aware analysis of 1T1MTJ STT-RAM cell structure. Using this model, we investigate temperature effect on memory cell access behavior in terms of access latency, energy and reliability on 45nm technology node. Thermal impact on more advanced 11nm technology node is also evaluated in the paper. Additionally, we propose a thermal-aware design for STT-RAM sensing circuit using body-biasing technique, which can enlarge read margin dramatically to enhance read reliability under temperature variations. Moreover, our proposed technique can suppress read disturbance effectively as well. Experimental results show that our proposed sensing circuit can enlarge read margin by 2.47X when reading '0' and 3.15X when reading '1', and reduce read disturbance error rate by 55.6% on average.**

***Index Terms*—Reliability, STT-RAM, Body-biasing, Temperature, Thermal modeling**

$STT-RAM$	Spin-Transfer Torque Random Access Memory
MTJ	Magnetic Tunnel Junction
TMR	Tunnel Magneto-resistive Ratio
J_c	MTJ switching current density
J_{c0}	MTJ switching current density at 0K
k_B	Boltzmann constant
T	Absolute temperature
E_b	Energy barrier height
Δ	Thermal stability of MTJ
t_{PW}	Switching current pulse width
τ_0	Inverse of attempt frequency
V	Free layer volume
H_K	Effective anisotropy field
ψ	Magnetization angle difference between free layer and pinned layer
θ	Magnetization angle between free (pined) layer with Z axis
ψ	Magnetization angle between free (pinned) layer with X axis
W	MTJ width
L	MTJ length
K_u	Magnetic anisotropy
α	Magnetic damping constant
M_s	Saturation magnetization
t_{ox}	Oxide barrier thickness
t_F	Free layer thickness
γ	Gyromagnetic ratio
V_{sb}	Transistor source-body voltage
V_{th0}	Transistor threshold voltage when $V_{sb}=0$
γ	Body effect coefficient
Φ_F	Surface potential
T_r	Room temperature

Bi Wu and Yuanqing Cheng have equal contributions in this work. Weisheng Zhao is the corresponding author.

Bi Wu, Yuanqing Cheng and Weisheng Zhao are with the School of Electrical and Information Engineering, Beihang University, Beijing, 100191 China (e-mail: bi.wu@buaa.edu.cn; yuanqing@ieee.org; weisheng.zhao@buaa.edu.cn).

Jianlei Yang is with the School of Computer Science and Engineering, Beihang University, 100191 China (e-mail: jerryyangs@gmail.com)

Aida Todri-Sanial is with LIRMM Laboratory, CNRS/University of Montpellier, Montpellier, 34095 France (e-mail: aida.todri@lirmm.fr).

This work is supported in part by China NSFC No. 61401008, Beijing NSF No. 4154076, and EU H2020 CONNECT grant agreement No. 688612

I. Introduction

TECHNOLOGY node aggressive shrinking has made static power consumption a severe challenge to CMOS circuit due to sub-threshold leakage [1]. Memory systems including on-chip cache and off-chip main memory suffer from leakage power as their capacities get larger to meet ever-increasing memory bandwidth requirements. Even worse, leakage power consumption has gradually dominated total power budget [2]. To solve the elevated "power wall" problem [3], emerging non-volatile memory (NVM) technologies are proposed for ultra low power memory design. Among them, STT-RAM is one of the most promising candidates to replace conventional SRAM and DRAM due to its fast access speed comparable with SRAM, easy integration with CMOS process, and non-volatility [4]–[10].

Although STT-RAM has many advantages over traditional memory technologies such as SRAM and DRAM, it also faces many challenges hindering the widespread use in commercial market. Among them, read/write reliability is a big concern [11], [12]. As TMR usually lies between 100% and 200% [13], the resistance difference between R_{ap} and R_p is small and it is difficult to sense the resistance difference reliably. The transistor in series with MTJ in 1T1MTJ STT-RAM cell makes this issue even worse due to its non-negligible resistance. The main sources affecting STT-RAM reliability

come from both process variations and environmental factors, such as temperature, magnetic field, etc.

Compared to extensive literatures investigating process variation effects on STT-RAM read/write reliability [14]–[16], temperature impact on STT-RAM read/write operations, however, lack of in-depth study. Although some papers investigated thermal impact on MTJ theoretically or experimentally [17]–[19], the lack of thermal model for electrical simulation prevents the evaluation of thermal impact on circuit and architecture level design.

In the paper, we build a thermal-aware MTJ model for electrical simulation, which can capture TMR and switching current variations of MTJ with temperature fluctuations. After that, we use the MTJ model to investigate thermal impact on STT-RAM cell access behavior and its thermal sensitivity. The main contributions of this work are listed as follows:

- We develop a thermal-aware MTJ model for electrical simulation and the model is calibrated and validated against measurements from real STT-RAM prototypes.
- We thoroughly investigate temperature impact on access behavior of the typical 1T1MTJ cell structure from latency, energy and reliability perspectives. The temperature effect with technology scaling is also extensively investigated and some important observations and design inspirations are presented to assist future STT-RAM chip design.
- In order to improve read reliability of STT-RAM cell, we propose a body-biased sensing scheme that can enlarge read margin and suppress read disturbance effectively with only negligible area overhead.

The rest of paper is organized as follows. Section II presents preliminaries of STT-RAM and MTJ temperature dependency, which motivates our work. Section III illustrates our thermal model development for hybrid CMOS/MTJ electrical simulation. Section IV analyzes temperature effect on 1T1MTJ cell access behavior in 45nm technology node. We also evaluate thermal impact on the more advanced 11nm technology node to predict thermal scalability of STT-RAM in this section. Based on these analyses, we propose a sensing circuit adopting body-biasing technique to improve read margin and suppress read disturbance at different temperatures in Section V. Experimental results are shown in this section as well. Section VI presents the related work, and Section VII concludes the paper.

II. Preliminaries of STT-RAM and Thermal Dependency of MTJ Device

A. Introduction to STT-RAM

STT-RAM uses MTJ as data storage device depending on MTJ resistance. It is mainly composed of three layers, i.e., pinned layer, insulating barrier, and free layer. The nanopillar resistance(R_p, R_{ap}) depends on the corresponding state of the magnetization of the two ferromagnetic layers Parallel(P, denoted as data '0') or Anti-Parallel(AP, denoted as data '1'). The resistance difference is characterized by TMR=$(R_{ap} - R_p)/R_p$.

According to the magnetizations of free layer and pinned layer, MTJ can be classified into in-plane and perpendicular

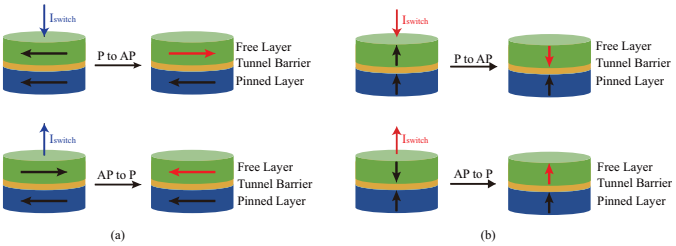


Fig. 1: Illustration of MTJ structure. Data can be stored in the cell by switching magnetization of free layer to be parallel or anti-parallel to pinned layer. (a) In-plane MTJ: magnetizations of free and pinned layers are parallel to MTJ pillar surface. (b) Perpendicular MTJ: magnetizations of free and pinned layers are perpendicular to MTJ pillar surface.

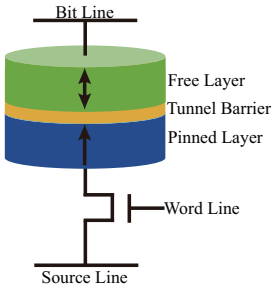


Fig. 2: 1T1MTJ cell structure consisting of one transistor and one MTJ.

ones. The former MTJ, which is shown in Fig. 1(a), retains data by shape anisotropy and requires large current density to switch MTJ state. Compared to its in-plane counterpart, free layer magnetization of the latter one is perpendicular to MTJ pillar surface and requires much lower switching current density thus having better scalability than its in-plane counterpart [20]. Fig. 1(b) shows perpendicular MTJ and its two different states. In this article, we use in-plane MTJ technology in 45nm technology node [21] and perpendicular MTJ technology in 11nm technology node [22] for investigation.

The typical STT-RAM cell consists of 1 transistor and 1 MTJ called 1T1MTJ structure as shown in Fig. 2. It is widely used to construct STT-RAM memory array due to its high integration density and simplicity for fabrication [23], [24]. To read the cell, word line is enabled, and a read voltage can be applied between bit line and source line. The resulting sensing current will be compared to that of reference cell to distinguish '0' from '1' being read out. The reference cell consists of two MTJs, one in parallel state and the other in anti-parallel state. To write data into memory cell, word line is enabled and a voltage is applied between bit line and source line to generate sufficient switching current. Depending on polarity of writing current, a '0' or '1' can be written. Next, we will analyze temperature impact on STT-RAM cell, which motivates our work.

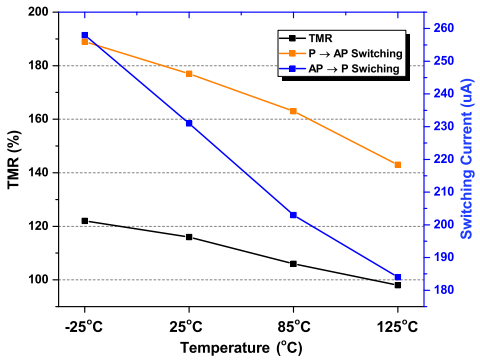


Fig. 3: TMR and switching current variations at different temperatures [21].

B. Thermal dependency of 1T1MTJ STT-RAM cell

As computing enters into multi-core and many core era, power density on chip increases sharply and makes thermal issue a big problem of modern processors [25]. STT-RAM integrated as on chip cache, which is one of its most possible applications, suffers from high temperature as well. In fact, as the core component of STT-RAM, MTJ itself is highly temperature dependent. Fig. 3 shows TMR dependence on temperature measured from a STT-RAM prototype [21]. The research from Qualcomm also got similar results [17].

As shown in the figure, TMR decreases with increased temperature. Since TMR has dominant impact on STT-RAM cell sensing reliability, it is imperative to consider temperature when investigating STT-RAM cell access behavior. On the other hand, Fig. 3 also indicates that MTJ switching threshold current decreases with temperature. It is because E_b decreases with temperature as observed in [17], and MTJ switching current density depends on E_b following the formula shown below,

$$J_c = J_{c0}[1 - \frac{k_b T}{E_b} ln(\frac{t_{PW}}{\tau_0})] \quad (1)$$

It indicates that switching current decreases with E_b. From Fig. 3, we can observe the switching current variation is coincide with the formula prediction. The reduced MTJ switching current is due to degraded MTJ's thermal stability in high temperature. It can be calculated by the following formula [26],

$$\Delta = \frac{E}{k_B T} = \frac{H_K M_S}{k_B T} V \quad (2)$$

From (2) we can observe that thermal stability decreases with temperature increases. In other words, it requires less current to switch MTJ state as temperature increases. On the balance, both TMR and switching current, which determine access behavior of MTJ, are affected by temperature variations. The thermal impact on transistor has been well studied. When temperature increases, carrier mobility degrades accordingly as shown in the following expression, μ is the carrier mobility of transistor, and depends on the temperature indicated by the following relation:

$$\mu(T) = \mu(T_r)(\frac{T}{T_r})^{-k_u} \quad (3)$$

Therefore, the driving ability of transistor is affected by temperature as well. In summary, it is necessary to consider both MTJ and transistor thermal dependency when analyzing the access behavior of STT-RAM cell at different temperatures.

III. MTJ Modeling Technique Considering Temperature Effect for Hybrid CMOS/MTJ Electrical Simulation

Although there are several MTJ models available for circuit level simulation, such as [27]–[29], they can not capture temperature effects on several MTJ properties simultaneously or support direct hybrid CMOS/MTJ electrical simulation. To facilitate thermal analysis of STT-RAM cell at circuit level and above, we develop a thermal model of MTJ based on MTJ model proposed by [30] and fit our model through measured data from a STT-RAM prototype [21].

Firstly, in order to capture temperature effect on TMR, we use the following approximation formula to calculate TMR at zero biasing voltage, i.e., $TMR(0)$,

$$TMR(0) = a_1 + a_2 cos(T \times b_2) + b_1 sin(T \times b_2) \quad (4)$$

where a_1, a_2, b_1, b_2 are fitting parameters to match measurements from [17].

Note that TMR depends not only on temperature but also on bias voltage as in [31]. We use the following formula to calculate TMR under different biasing voltages:

$$TMR(V_{bias}) = \frac{TMR(0)}{1 + V_{bias}^2 / V_h^2} \quad (5)$$

Therefore, TMR can be modeled as a function of both temperature and biasing voltage. MTJ thermal stability Δ depends on M_s as shown below,

$$\Delta = \frac{M_s H_K V cos^2(\theta)}{k_B T} \quad (6)$$

Note that M_s is temperature dependent as well according to [32]. We use the following formula to define M_s such that thermal stability values at different temperatures fit with measurement values from [17]:

$$M_s = p_1 T^3 + p_2 T^2 + p_3 T + p_4 \quad (7)$$

where p_1, p_2, p_3 and p_4 are fitting parameters. Based on (6) and (7), we can obtain thermal stability at different temperatures.

After that, we use the work flow similar to that proposed in [30] to integrate temperature effect in MTJ model as shown in Fig. 4. The inputs of our model include MTJ size, barrier thickness, temperature, initial magnetization angles of free layer (θ and ϕ). As shown in the work flow, the temperature and bias voltage dependent TMR can be derived by (4) and (5) while temperature dependent thermal stability can be calculated from (6) and (7). Then, we can derive the MTJ switching current and spin torque induced by the injected spin polarized current. The spin torque together with uniaxial anisotropy torque, and shape anisotropy are fed into RHS of LLG equation [33] to update magnetization angle of free layer, i.e., θ and ϕ, until MTJ approaches a steady state (anti-parallel state or parallel state). Since both in-plane and perpendicular

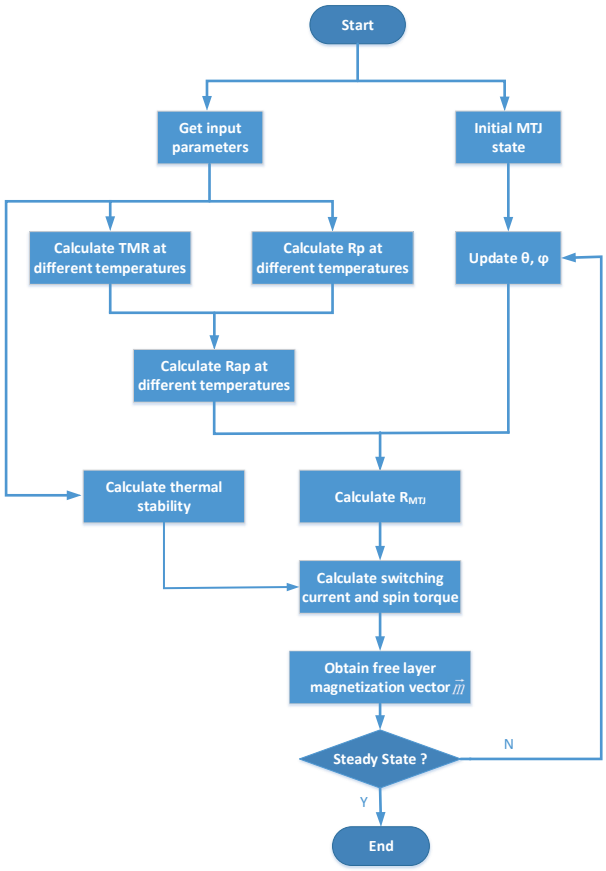


Fig. 4: The work flow of our thermal model

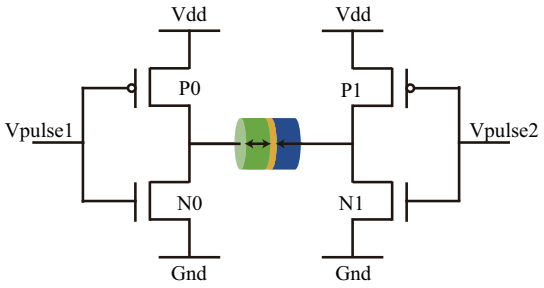


Fig. 6: Write circuit schematic of 1T1MTJ cell.

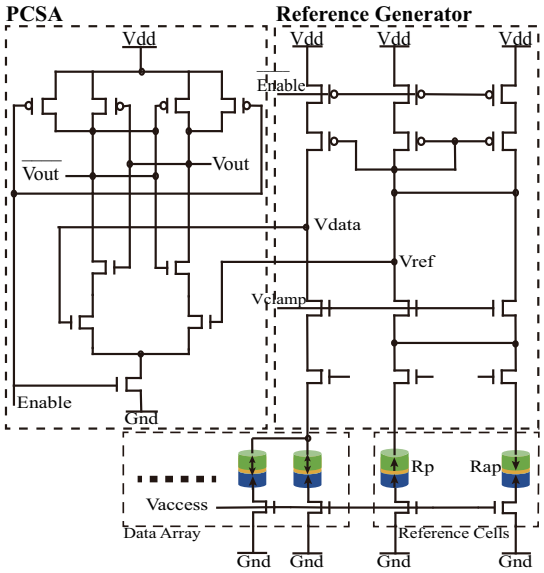


Fig. 7: Read circuit schematic of 1T1MTJ cell.

MTJ switching procedures are governed by LLG equation, our thermal model can be applied to both cases. The thermal model is implemented in Verilog-AMS and can be used for electrical simulation. Fig. 5 compares R_p, R_{ap} and switching current derived by our model to measurements from a prototype chip [21] in different biasing voltages and temperatures. It shows that our thermal model can capture TMR and switching current variations accurately at different temperatures. The modeling errors of R_p, R_{ap} and switching currents under different temperatures are within 10% of experimental meansurements. In next section, we will use this model to analyze temperature impact on read/write operations of a typical 1T1MTJ STT-RAM cell.

IV. Temperature Impact on 1T1MTJ STT-RAM Cell Operations

A. Introduction to Read/Write Circuit of 1T1MTJ STT-RAM Cell

We adopt the writing circuit proposed by [31] to perform write operation of 1T1MTJ STT-RAM cell. The circuit schematic is drawn in Fig. 6. By enabling transistor pair in diagonal, we can inject current in different directions, and write data '0' or '1' into the cell accordingly. For example, transistors P0 and N1 are enabled when writing '0' since the injected current flowing from upper left V_{dd} to lower right ground can switch MTJ from anti-parallel state to parallel state. To write '1', we need to enable P1 and N0 to reverse switching current direction.

TABLE I: MTJ parameters used in our simulations

Symbol	Value
L	40nm
W	100nm
K_u	$5 \times 10^5 A/m$
α	0.03
M_s at 25°C	$3.68 \times 10^3 A/m$
t_{ox}	0.85nm
t_F	33.55nm
γ	$1.76 \times 10^7 rad/(s \cdot T)$

Considering read operations, we adopt the widely used two-stage sensing scheme proposed by [8] and [34] as shown in Fig. 7. The first stage shown on the right part of the figure compares sensing current of data cell with that of reference cell. A clamping voltage is applied on both data cell and reference cell to make sure both sensing paths have the same voltage. Then, the sensing current difference between data cell path and reference path is converted to voltage difference via uploading transistors. Since the sensing voltage difference is usually only tens of millivolts being too small to be recognized by the following logic stage, it needs to be amplified by the second stage to generate full-swing signals, which is shown

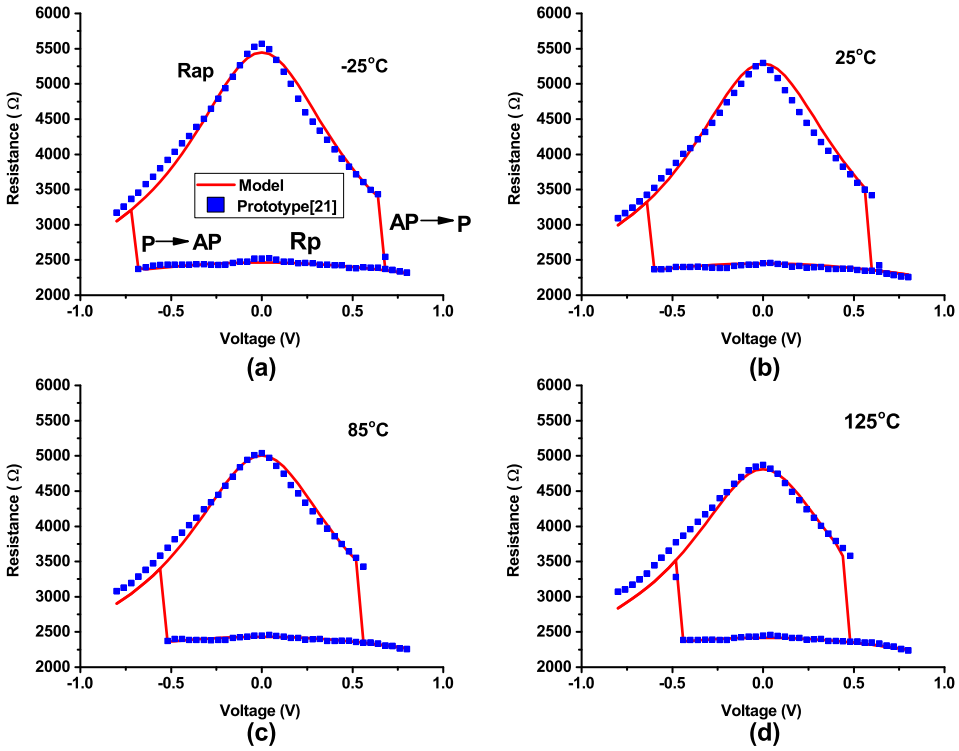


Fig. 5: Validation of our thermal-aware MTJ model against measurements from [21] (a) comparison of -25°C case. (b) comparison of 25°C case. (c) comparison of 85°C case. (d) comparison of 125°C case.

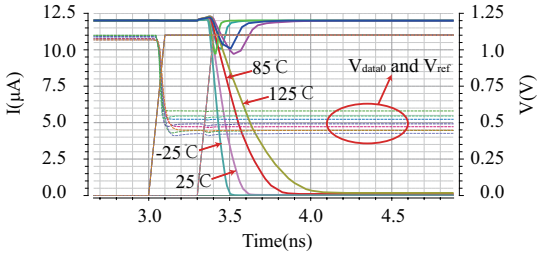
on the left part of Fig. 7. In the following subsections, we use 45nm PTM CMOS model [35] and MTJ thermal model developed above to analyze temperature impact on reading and writing behaviors of STT-RAM cell. The parameters of MTJ model used for evaluation are listed in Table I.

B. Temperature Effect on Reading Operations

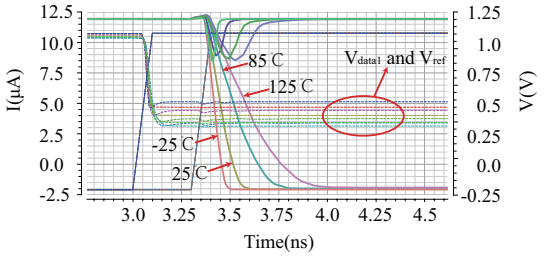
In the following, we evaluate temperature effect on STT-RAM cell's reading behavior in terms of performance, energy and reliability using Cadence Spectre [36].

Fig. 8 plots sensing amplifier output voltage waveforms at four different temperatures (i.e., -25°C, 25°C, 85°C and 125°C). The read pulse applied between bit line and source line has 2ns width to make sure data can be sensed reliably. Fig. 8a shows output waveforms of the second stage of sensing amplifier when '0' is read. Fig. 8b shows output waveforms when read '1' at different temperatures. Considering reading '0', read latency increases with temperature. For instance, it is 510ps at -25°C while it becomes 1.06ns at 125°C. Reading '1' operation indicates the similar trend. By examining sensing signals, we observe that read latency increase is mainly due to two factors. First of all, read margin shrinks with temperature (showed in Table II). Read margin is defined as the difference between reading signal and reference signal. Refer to Fig. 5, R_p of MTJ remains almost the same at different temperatures while R_{ap} becomes smaller at higher temperature. The data sensing voltage output from the first stage of sensing circuit can be expressed as:

$$V_{data} = V_{dd} - \frac{V_{clamp}}{R_{mtj}} \times R_{load}$$



(a) Read '0' latencies at four temperatures.



(b) Read '1' latencies at four different temperatures.

Fig. 8: Read latencies of 1T1MTJ cell at different temperatures

Reference sensing voltage output is,

$$V_{ref} = V_{dd} - (\frac{V_{clamp}}{R_{ap}} + \frac{V_{clamp}}{R_p})/2 \times R_{load}$$

where V_{clamp} is clamping voltage applied over data cell and reference cell. R_{load} is resistance of uploading transistor. When sensing '0', $R_{mtj} = R_p$, and read margin $|V_{data} - V_{ref}| = V_{clamp}R_{load}(\frac{1}{2R_p} - \frac{1}{2R_{ap}})$. As shown in Fig.

TABLE II: Read margins and energy consumptions at different temperatures at 45nm technology node.

	-25°C	25°C	85°C	125°C
Read margins at different temperatures. (unit: mV)				
Read '0'	35.0	32.3	26.3	22.1
Read '1'	40.3	32.0	20.2	14
Read energy at different temperatures. (unit: fJ)				
Read '0'	18.4	18.9	19.7	21.1
Read '1'	17.7	18.5	19.5	19.9

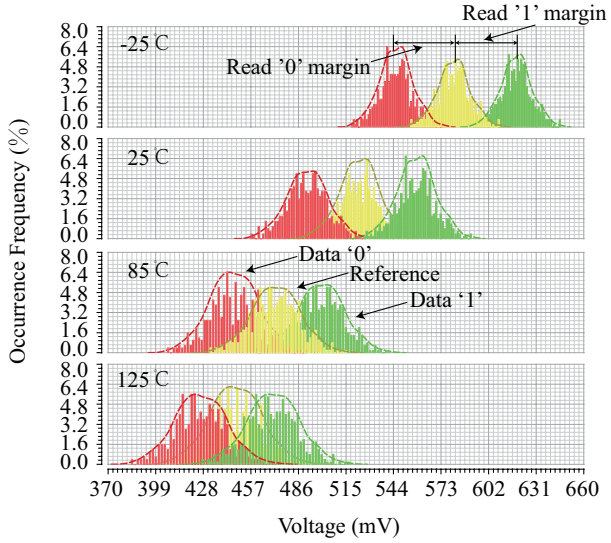


Fig. 9: Sensing and reference voltage distributions at different temperatures through 1000 Monte Carlo simulations.

5, R_{ap} decreases with temperature while R_p remains almost unchanged. Therefore, read margin is smaller in higher temperature making sensing time much longer. Likewise, when reading '1', sensing margin also becomes smaller in higher temperature. Hence, sensing latency increases with temperature. Another reason of read latency increasing with temperature is that transistor driving ability also degrades with temperature. Since the second stage relies on discharging speed difference between two discharging paths, degrading current driving ability makes data sensing more difficult in higher temperature. Table II lists read margins at different temperatures when read '1' and '0' respectively, which validates our theoretical analysis. As shown in the table, when temperature increases from -25°C to 125°C, read margin of reading '1' decreases by 65.3%, and that of reading '0' decreases by 36.9%. Therefore, temperature has significant effect on read performance, and should be thoroughly considered during STT-RAM cell design such that it can achieve given performance specification across the whole working temperature range.

Read energy at different temperatures is also tabulated in Table. II, which is consistent with data in [37]. Read energy increases by 14.7% from -25°C to 125°C when reading '0'. Reading '1' shows similar energy increasing with temperature. Based on these results, we can observe that although driving current reduces as temperature, increased read latency has a larger impact on read energy consumption.

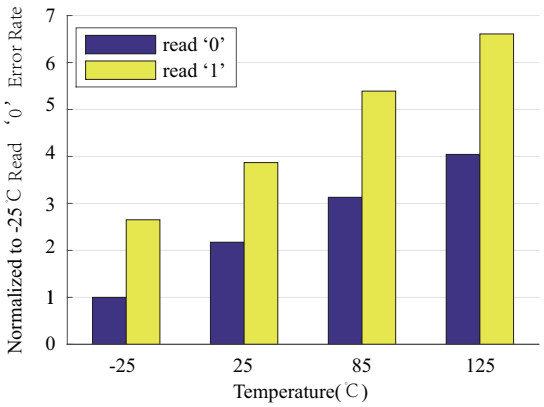


Fig. 10: Read error rate comparisons at different temperatures in 45nm technology node. Both reading '0' and '1' error rates increase with temperature, and reading '1' error rate is higher due to R_{ap} variation with temperature.

TABLE III: Monte Carlo simulation parameters

Parameters	Mean	Std.Dev
Channel Length	45nm	4%
Channel Width	135nm	4%
Threshold Voltage	0.4V	16mV
MTJ Width	40nm	4%
MTJ Length	100nm	4%

Next, we evaluate read reliability of 1T1MTJ cell at different temperatures. To obtain sensing and reference voltage distributions, we perform 1000 Monte Carlo simulations for reading '0' and reading '1' respectively. The simulation settings are shown in Table.III. The process variations of transistor gate length, threshold voltage and MTJ size are assumed to obey Gaussian distributions, and σ/μ of each parameter distribution is assumed as 4%. Sensing voltage distributions at different temperatures are plotted in Fig. 9. From the figure, we have two important observations. Firstly, as the temperature increases, overlapping of read voltage and reference voltage distributions becomes more significant, which is attributed to shrinking read margin. Overlapping area denotes where read error may occur. Secondly, the reference voltage drifts with temperature instead of a fixed value. Since resistances of uploading PMOS transistors increase with temperature while MTJ resistance remains almost the same (for R_p) or becomes smaller (R_{ap}), voltage drop on uploading PMOS transistors becomes larger in higher temperature. Therefore, both reference and data sensing voltage reduce with temperature. Fig. 10 dictates read error rates at different temperatures normalized to that of -25°C. We can observe that read error rate increases with temperature, and error rate at 125°C increases by 4X compared to that of -25°C when reading '0' and increases by 3.3X when reading '1'. Note that R_{ap} decreasing with temperature on both data sensing path and reference path makes sensing '1' more error-prone. In summary, shrinking read margin degrades sensing reliability of STT-RAM, i.e., given the same sensing period, it is expected to incur more read errors at higher temperature. Therefore, the sensing circuit design for STT-RAM should take temperature into account

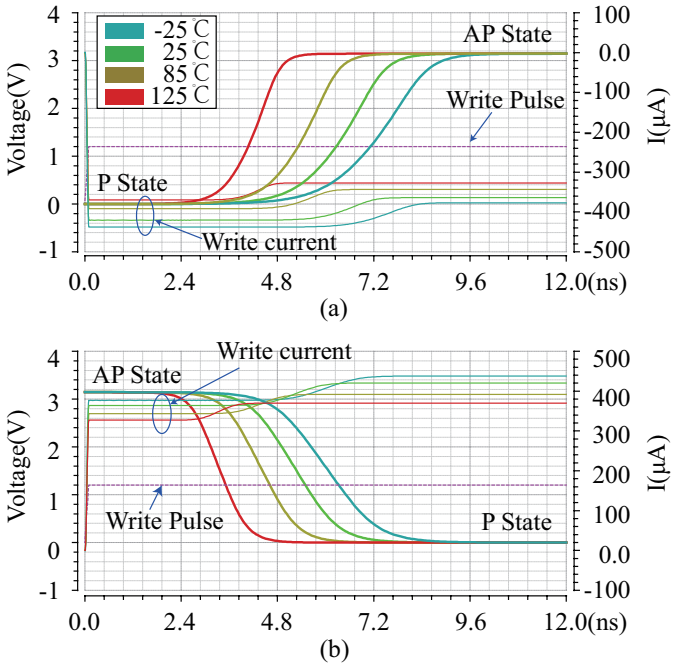


Fig. 11: Write currents at different temperatures. (a) write current decreases with temperature when MTJ is switched from P state to AP state due to reduced thermal stability. (b) write '0' current decreases with temperature as well due to the same reason.

TABLE IV: Write latencies and energy at different temperatures

Write latency (ns)	-25°C	25°C	85°C	125°C
write '0'	8.81	7.60	6.41	5.16
write '1'	10.02	8.75	7.37	5.64
Write energy (pJ)	-25°C	25°C	85°C	125°C
write '0'	3.70	3.27	2.83	2.28
write '1'	4.10	3.88	3.61	2.91

to ensure it can work reliably across the whole working temperature range without violating reliability specification.

C. Temperature Effect on Write Operations

We use the write circuit shown in Fig. 6 in our simulation. Fig. 11 plots write current of writing '0' and '1' waveforms respectively at different temperatures. As shown in the figure, write current decreases with temperature for both writing '0' and '1' cases. For example, writing '1' current drops from $380\mu A$ in -25°C to $330\mu A$ in 125°C. As mentioned in Section II, MTJ's switching current depends on thermal stability which reduces with temperature. As a result, switching current decreases with temperature as well. Table IV lists write latencies and energy at different temperatures. We can observe that write latency decreases with temperature implying that MTJ can be switched more easily due to reduced thermal stability. As both write latency and current become smaller in higher temperature, MTJ switching energy also decreases with temperature as shown in Table IV. Note that writing '1' energy

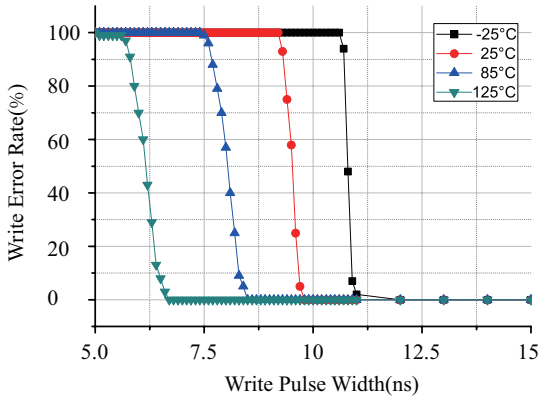


Fig. 12: Write error rate at different temperatures. Write latency decreases with temperature and write error rate curve shifts to the left with temperature increasing.

is higher than writing '0' energy since P $\rightarrow$ AP switching is more difficult than AP $\rightarrow$ P [21].

Next, we consider write error rate at different temperatures. The write pulse width is swept from 6ns to 11ns which covers typical write latencies of STT-RAM. The simulation results are plotted in Fig. 12. We can observe that write error rate curve shifts to right remarkably as the temperature decreases. For instance, at 125°C, 6ns is enough to ensure write error free while the write pulse width should be larger than 11ns when temperature decreases to -25°C. Additionally, since write current lies in hundreds of μA while read current is only tens of μA in 45nm technology node, read disturbance can be negligible[1]. Write errors mostly come from insufficient write pulse width.

In summary, high temperature incurs high read latency and energy. Read margin also shrinks with temperature making read reliability a challenge in high temperature. Whereas, write pulse can be shortened in high temperature and less error-prone. Additionally, due to large difference between write current and read current, read disturbance is not significant at 45nm technology node. The write error is mainly caused by insufficient write pulse width or magnitude.

D. Temperature impact on STT-RAM technology scalability

As technology node shrinks, difference between write current and read current becomes smaller and read reliability will be a severe challenge for reliable operation of STT-RAM [11]. In this subsection, we will evaluate temperature impact on STT-RAM access operations at 11nm technology node. We calibrate our MTJ model with measured data from [22] to match switching current, energy barrier, TMR and MTJ resistance of 11nm MTJ at room temperature. Due to the lack of published measurements of MTJ parameters at different temperatures, we assume TMR and resistance variation trend with temperature at 11nm technology node are similar to those of 45nm technology node due to their similar material and stacking structure. Then, we can scale our thermal model to

[1]Read disturbance means erronously switching of MTJ during read procedure due to large read current.

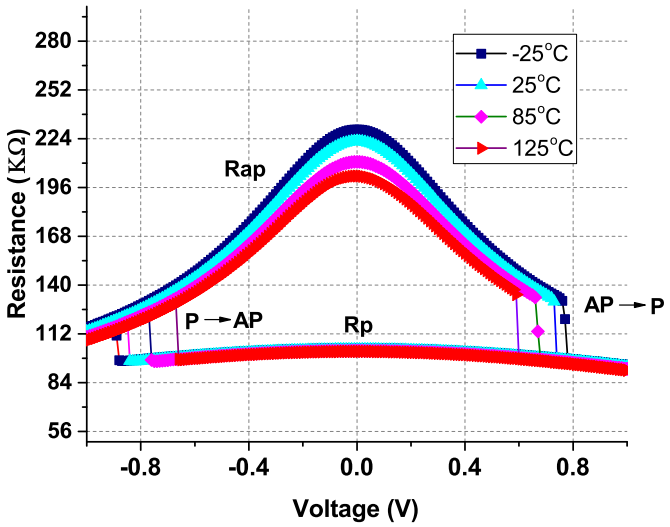


Fig. 13: Temperature effect on switching current and MTJ resistance at 11nm technology node. Switching current, TMR and MTJ resistance at room temperature are calibrated with measurements presented in [22].

11nm accordingly. The relationship of MTJ resistance and biasing voltage at different temperatures derived from HSPICE simuations are plotted in Fig. 13. The X axis represents MTJ biasing voltage in Volt, and the Y axis is resistance of MTJ in $K\Omega$. As shown in Fig. 13, Rp and Rap derived by our thermal model are $224K\Omega$ and $103K\Omega$ respectively. The maximum error compared to data from [22] is 3%. The switching current from [22] at 11nm technology node is below $10\mu A$ while that derived from our thermal model is about $7\mu A$ which matches with experimental result as well. Therefore, our scaled thermal model can be used to predict 1X nm STT-RAM thermal behaviors accurately.

From the figure, we can observe several interesting points. First, both R_p and R_{ap} are much larger than those of 45nm MTJ. The reason is that it is preferable to maintain a roughly constant RA (the product of MTJ resistance and area) across different generations of MTJ fabrication technology. The MTJ area shrinks from $40nm \times 100nm$ to $\pi \times 5.5nm^2$, which makes MTJ resistance larger to maintain RA value unchanged. Secondly, although 11nm MTJ switching current is much smaller than that of 40nm MTJ, considering sharply increase of MTJ resistance, the supply voltage of 1X nm CMOS technology may be not enough to provide enough switching current for MTJ switching. Therefore, we have to apply a higher voltage in 11nm technology node to write MTJ successfully. New material or novel MTJ structure should be explored to reduce MTJ switching current further for compatible hybrid CMOS/Magnetic 1T1MTJ cell design at more advanced technology node as mentioned in [22]. However, this topic is out of the scope of this article and we will take this issue as our future work.

In the following, we adopt 16nm CMOS PTM model [35] and 11nm MTJ for circuit simulations. Using the similar sensing circuit as that for 45nm case, we evaluate temperature effect on read behavior of 1Xnm 1T1MTJ STT-RAM. The

TABLE V: Read performance and energy at different temperatures in 11nm MTJ technology node

Read latency (ps)	-25°C	25°C	85°C	125°C
read '0'	410	420	440	450
read '1'	413	421	430	439
Read current (μA)	-25°C	25°C	85°C	125°C
read '0'	2.76	2.46	2.13	1.93
read '1'	1.88	1.73	1.59	1.49
Read energy (fJ)	-25°C	25°C	85°C	125°C
read '0'	3.92	3.47	3.08	2.87
read '1'	3.61	3.18	2.76	2.56

TABLE VI: Write latency and energy at different temperatures of 11nm 1T1MTJ memory cell

Write latency (ns)	-25°C	25°C	85°C	125°C
write '0'	7.86	6.69	5.38	3.98
write '1'	9.09	7.68	6.3	4.65
Write energy (fJ)	-25°C	25°C	85°C	125°C
write '0'	81.7	68.3	53.0	37.9
write '1'	102	84.6	66.6	47.3

read latencies, currents and energy consumptions at different temperatures are listed in Table. V. From the table, we can observe that MTJ read latency becomes much smaller in 11nm technology node compared to that of 45nm case, which shows a good scalability of MTJ technology. The read latency of 45nm MTJ is 1ns at 125°C while it is about 0.4ns for 11nm MTJ. On the other hand, as temperature increases, read latency of 11nm memory cell also increases accordingly because of shrinking read margin, which has the similar trend as that of 45nm case. Considering read current, it is much smaller in 11nm case, which only requires several μA for reading. The read current decreases with temperature, which is similar as 45nm case. It is caused by degraded driving ability of CMOS transistor at high temperature. However, reading energy decreases with temperature as shown in Table. V. It is because decreasing read current dominates the read energy in 11nm technology node, while increasing read latency dominates the read energy in 45nm technology node. In summary, compared to Table II, both read latency and energy reduce significantly at 11nm node which indicates good scalability of MTJ with technology node shrinking.

Table VI tabulates write energy and latencies of 11nm STT-RAM cell working at different temperatures. From the table, we observe that write latencies become much smaller than those of 45nm STT-RAM cell. Write energy also reduces compared to 45nm case. The reduction of write energy and latency is attributed to reducing thermal stability of MTJ at 11nm technology [22]. As shown in the table, write energy is reduced by two orders of magnitude compared to 45nm case which indicates the good scalability of STT-RAM technology. As the temperature increases, both write latency and write energy decrease which are similar to 45nm case because of reduced thermal stability in high temperature. Moreover, for 11nm technology node, latencies of writing '1' are significantly larger than those when writing '0'. It is because the switching asymmetry becomes more severe due to large difference between R_p and R_{ap} at 11nm technology node.

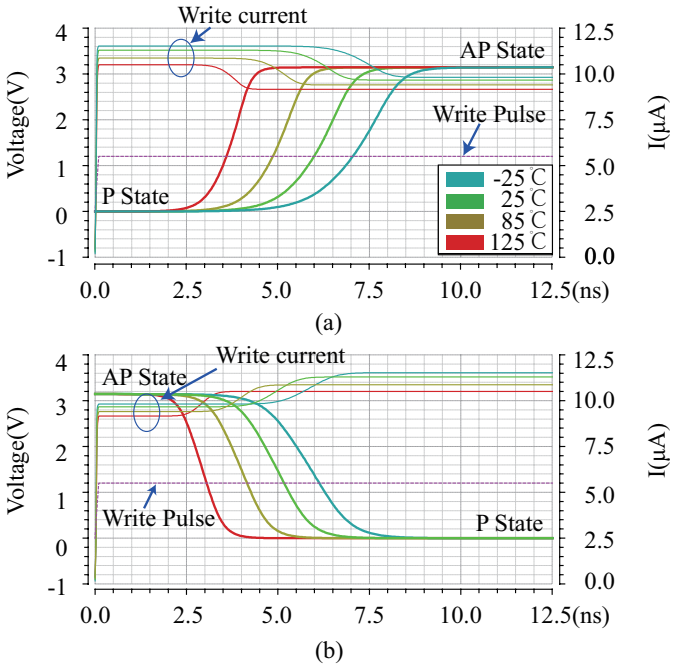


Fig. 14: Write current variations with temperature of 11nm MTJ. (a) write '1' current decreases from $11.5\mu A$ to $10\mu A$ when temperature elevates from -25°C to 125°C. (b) write '0' current decreases from $10\mu A$ to $9\mu A$ when temperature elevates from -25°C to 125°C.

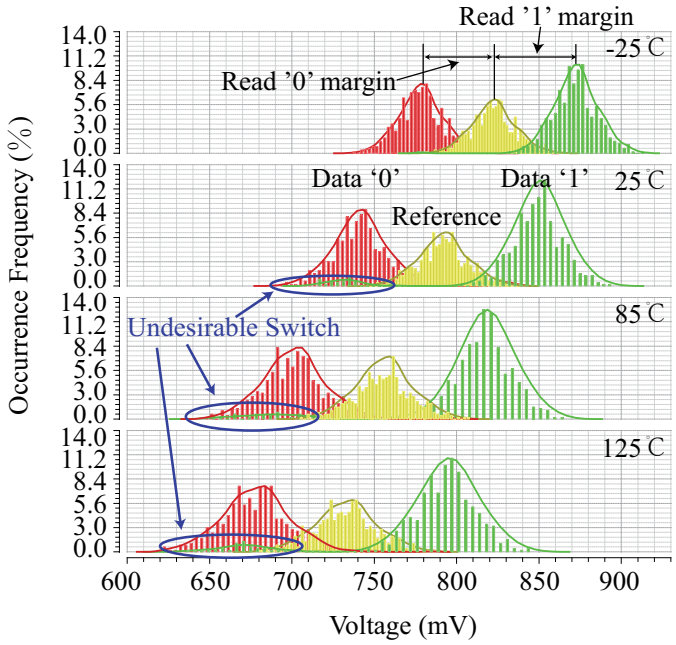


Fig. 15: Read voltage distributions of 11nm STT-RAM cell at different temperatures. In addition to read margin shrinking with temperature similar to 45nm case. When reading '0' from 11nm MTJ, it indicates considerable read disturbance occurrences.

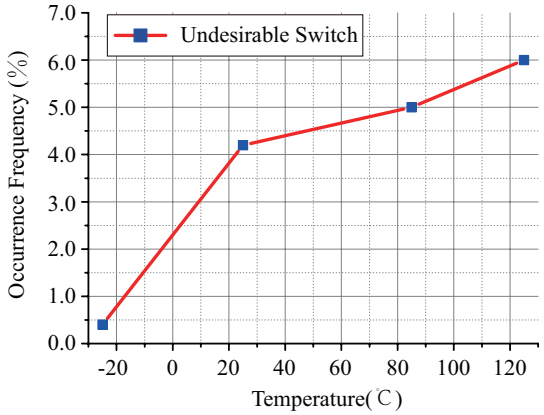


Fig. 16: Read error rate due to read disturbance at different temperatures.

The write currents at different temperatures are plotted in Fig. 14. As shown in the figure, write '1' current decreases from $11.5\mu A$ to $10\mu A$ when temperature elevates from -25°C to 125°C while write '0' current decreases from $10\mu A$ to $9\mu A$ when temperature elevates from -25°C to 125°C. Additionally, at the same temperature, it requires higher switching current for P $\rightarrow$ AP switching. For instance, at 25°C, the write current is about $9\mu A$ to switch MTJ from AP state to P state while switching from P to AP needs $11\mu A$ current.

Although the reducing switching current is helpful to reduce writing energy, it also introduces read disturbance in advanced technology node as read current approaches write current more closely. To evaluate read reliability, we perform 1000 Monte Carlo simulations at -25°C, 25°C, 85°C and 125°C respectively. The σ/μ of CMOS transistor threshold voltage, gate length and MTJ radius distributions are assumed to be 4% in the simulation. Refer to read circuit used in the paper shown in Fig. 7, only '1' $\rightarrow$ '0' switching can occur during reading procedure. The sensing and reference voltage distributions are illustrated in Fig. 15. Except for read margin shrinking with temperature, we can observe read disturbance occurs during reading '1', which is circled in the figure. Compared to 45nm case, the read disturbance is much more severe. To show it more clearly, we plot read error caused by read disturbance at different temperatures in Fig. 16. As shown in the figure, the read error rate caused by read disturbance increases from 0.5% to 6% when temperature increases from -25°C to 125°C due to reduced thermal stability of MTJ at high temperature. Therefore, read disturbance will be a big concern when technology node shrinks below 1X nm regime, and has strong dependence on temperature.

V. Improving Reading Reliability of STT-RAM with Temperature Variations By Body-Biasing Technique

From the above analyses, we find that read margin shrinks with temperature making sensing reliability a big challenge for STT-RAM especially at high temperature. Additionally, read disturbance aggravates as temperature increases as well. To

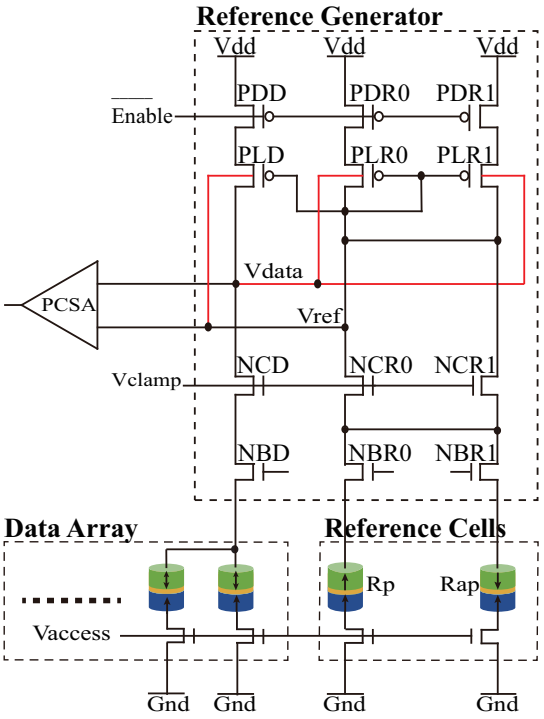


Fig. 17: Body biased sensing circuit schematic.

overcome these problems, we propose a body-biasing technique to adjust reference voltage dynamically such that read margin can be enlarged and read reliability can be improved. Moreover, if read margin remains unchanged, we can reduce the sensing current accordingly, and limit read disturbance at very deep sub-micron regime.

A. The working principle of our proposed body biasing sensing circuit

Based on the scheme we used for read operation evaluations above, which is proposed by [8], we propose a body-biasing technique shown in Fig. 17 to enlarge read margin of 1T1MTJ STT-RAM cell. The main idea is that through adjusting body voltage of uploading PMOS transistors, their threshold voltages and driving abilities can be enhanced or suppressed dynamically to enlarge voltage difference between data sensing path and reference path. In the proposed read circuit, data output in the first sensing stage is connected to body terminals of two uploading PMOS transistors (PLR0 and PLR1) on reference branches while reference output is connected to the body terminal of uploading PMOS transistor on data sensing branch PLD) as shown in the figure. The working detail of body-biasing sensing circuit is analyzed as follows.

Considering body effect, the threshold voltage of PMOS transistor is given by the following equation,

$$|V_{th}| = |V_{th0}| + |\gamma|(\sqrt{|2\Phi_F + V_{SB}|} - \sqrt{|2\Phi_{F|}}) \quad (8)$$

As the body voltage changes, the V_{SB} and V_{th} vary accordingly indicated by (8). Therefore, driving current of the uploading transistor can be tuned through body voltage

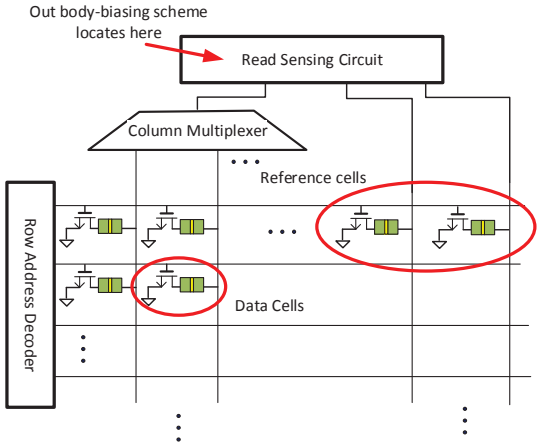


Fig. 18: The location of body-biasing sensing circuit in STT-RAM sub-array.

variations. For example, if the data cell stores '0', i.e., the MTJ has a low resistance, I_{data0} will be larger than I_{ref}. Consequently, the voltage drop over uploading transistors in data branch (i.e., PDD and PLD) is larger than those in reference branch (i.e., PDR0 and PLR0 or PDR1 and PLR1). Then, V_{data0} will be lower than V_{ref}. As a result, source-body voltage of PLD would be smaller than that of PLR0/1. So driving current of PLR increases while those of PDR0 and PDR1 decreases. Consequently, the difference between data sensing voltage and reference voltage will be enlarged further, and read margin can be improved. The case when sensing data '1' can be analyzed similarly. The detailed experimental results are presented in the next subsection.

Fig.18 shows how our proposed sensing circuit can be integrated in the STT-RAM sub-array. As shown in the figure, due to sensing amplifiers are shared among multiple column cells and our proposed technique only changes the body-biasing connections to uploading transistors in sensing circuit, the incurred area overhead is negligible.

B. Read reliability improvement by our proposed body-biasing based sensing circuit design

To evaluate read margin and reliability enhancement, we use 1T-1MTJ memory cell structure mentioned above in the following electrical simulations. Fig. 19 plots waveforms of V_{data} and V_{ref} during reading '0' and '1' when using our body-biasing technique compared to the case without using it at 25°C in 45nm technology node. It shows that voltage difference is magnified dramatically by body-biasing. The read margin increases by nearly 3X, for both reading '0' and reading '1', which is helpful to reduce read error rate for the second sensing stage. The experimental results on 11nm 1T1MTJ STT-RAM also show the similar trend. Then, we plot the read margin comparisons at different temperatures in Fig. 20. As shown in the figure, sensing margins of our method are significantly higher than those of [8] over the temperature range from -20°C to 100°C. Note that read margin of reading '1' is much larger than that of reading '0' by body-biasing technique. The reason is that when reading '0', the reference

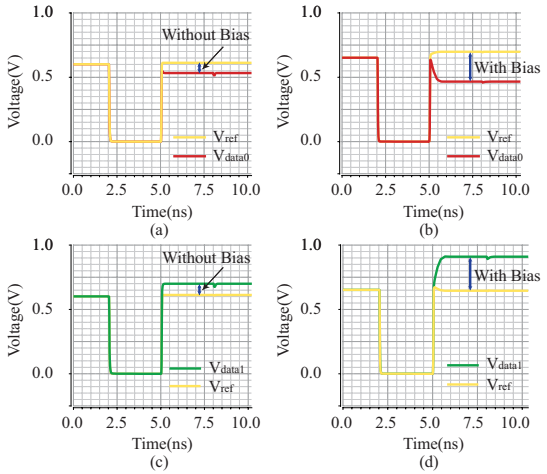


Fig. 19: Sensing signal waveforms when reading '0' and reading '1'. (a) The sensing '0' voltage waveforms when using the circuit proposed by [8]. (b) The sensing '0' voltage waveforms when using our proposed technique. (c) The sensing '1' voltage waveforms when using the circuit proposed by [8]. (d) The sensing '1' voltage waveforms when using our proposed technique.

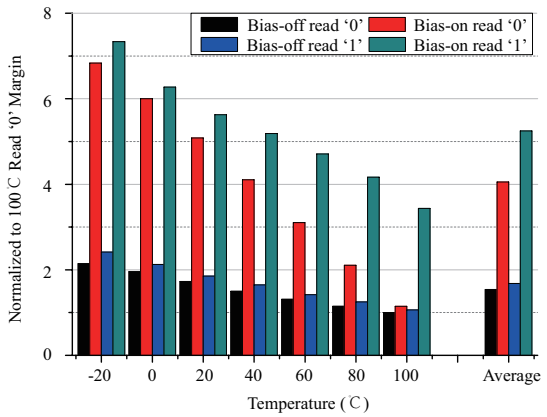


Fig. 20: Read margin improvements by body-biasing technique at different temperatures when read '0' and read '1' respectively.

voltage is higher than data voltage, and approaches nominal body voltage of PMOS transistor more closely. Therefore the tuning head room is reduced for reading '1' case. Another interesting point is that read margin improvement shrinks with temperature increasing. This is attributed to the fact that MOS transistor carrier mobility deteriorates significantly at high temperature making body-biasing less effective on tuning transistor driving currents. However, on average, our proposed sensing scheme can enlarge read margin by 2.47X when reading '0' and 3.15X when reading '1'. The enhanced read margin guarantees that the data can be sensed more reliably and read error rate can be reduced accordingly. Then, we evaluate the read error rate when using our proposed sensing circuit at different temperatures in 45nm technology node. We obtain the simulation result through 1000 Monte Carlo simulations with 4% MOS transistor and MTJ size variations,

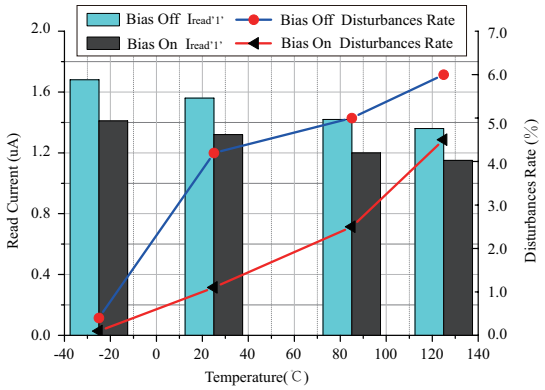


Fig. 21: Read current reduction due to body-biasing technique and read disturbance improvements at different temperatures at 11nm technology node.

and there is no read error rate any more due to enlarged read margin until 100°C where body-biasing technique can not tune the driving ability of transistor effectively.

As mentioned in Section IV.D, read disturbance becomes significant when MTJ size shrinks from 40nm to 1Xnm. Since body-biasing technique can be helpful on enlarging read margin, we can reduce the read current with the same read margin constraint. Fig. 21 depicts the read current reduction due to body-biasing technique and read disturbance improvement at different temperatures on 11nm MTJ technology node. We can observe that read current of body-biasing sensing circuit can reduce read current by 20% on average with the same read margin of original sensing circuit design[2]. Therefore, read disturbance can be alleviated due to reduced read current. Fig. 21 also shows read disturbance reductions at different temperatures when compared to those of the original sensing circuit. It clearly indicates that read error rate due to read disturbance can be reduced effectively. On average, the read disturbance error rate can be reduced by 55.6% when using our proposed body-biasing technique.

VI. Related Work

Since STT-RAM has many advantages to make it a promising candidate for future memory technologies, it has attracted much attentions from both academia and industry. The STT-RAM research can be divided into different levels. At device level, a compact model of MTJ is essential for upper level simulation and analyses. Sun investigated spin current induced MTJ switching and described switching behavior by Landau-Lifshitz-Gilbert equation accurately making it a base for following compact model development [38]. Faber et al. proposed a compact model of MTJ for in-plane STT-RAM, which can capture both static and dynamic characteristics [6]. Guo et al. investigated MTJ switching induced by Spin Transfer torque effect and built a SPICE compact model for STT-RAM simulation [27]. Shang et al. developed an SPICE-like simulation tool integrating MTJ switching model. This

[2]The relative small current reduction of read current compared to read margin amplification is that the resistance of 11nm MTJ is much larger than 45nm MTJ case. The large voltage difference only converted to much smaller read current variation.

tool models MTJ switching behavior using electrical devices such as current source, resistance, capacitance, etc, and it supports hybrid CMOS/MTJ circuit simulation. However, due to the complex equivalent circuit, simulation overhead is non-negligible [39]. Zhao et al. developed a model for thermally assisted MTJ model [28]. Zhang et al. developed a compact model for perpendicular anisotropy MTJ using Verilog-A at 40nm technology node [28]. With this model, authors explored tradeoffs involved in designing non-volatile flip-flops. Based on this work, Wang et al. enhanced the compact model to capture MTJ stochastic switching property and evaluate MgO barrier breakdown effect on MTJ reliability [29]. Fong et al. developed a verilog-A MTJ model based on LLG equation and Green function method which can model both in-plane and perpendicular MTJ switching procedure accurately [30]. Unfortunately, none of above models can capture temperature effect on MTJ and STT-RAM access behavior effectively. However, as the on-chip power consumption and temperature increases sharply, it is imperative to investigate temperature effect on STT-RAM access performance, power consumption and reliability to derive an optimal design across the whole operation temperature range of the specified STT-RAM product.

Shang et al. investigated temperature effect on spin polarized tunneling of MTJ built with different ferromagnetic materials [18]. The authors observed that both junction resistance and magnetoresistance decreases with temperature increases. Hagler et al. explored temperature dependence of MTJ with Al_2O_3 barrier experimentally [19], which also revealed that MTJ magnetoreistance reduction as temperature increased while TMR value showed a decreasing trend. Lee et al. emphasized importance of thermal aware design of STT-RAM and illustrated temperature effect on MTJ resistance and switching current [17]. A 45nm prototype STT-RAM from Lin et al. also demonstrated that temperature has significant impact on MTJ switching current and anti-parallel resistance [21]. However, these research efforts only focus on the device level and lack of temperature effect evaluation on circuit and architecture level. Considering the sensing circuit design, J. Kim proposed a degeneration loading scheme to mitigate process variation impact on STT-RAM sensing circuit [40]. Based on this sensing structure, they improved its sensing reliability further by self body-biasing technique in [40]. However, uploading transistors on both data and reference paths have the same body potential and can not benefit from dynamically changed reference voltage level to achieve larger read margin. Additionally, they did not analyze temperature variations on MTJ access behavior in detail. Bi et. al. investigated the temperature dependence of in-plane STT-RAM cell, which is the most relevant literature to our work. However, they only considered in-plane STT-RAM [41]. To fill the gap between device and upper level study concerning thermal effect, we build a thermal-aware MTJ model for hybrid MTJ/CMOS electrical simulation to capture temperature factors during MTJ switching and use it for circuit and architecture level analyses of STT-RAM.

VII. Conclusion & Future Work

As the integration density of conventional CMOS transistor sharply increases, power consumption of memory system makes it a road block for high performance exascale computing. Several emerging non-volatile memory technologies are proposed to replace traditional SRAM or DRAM to achieve low power and memory consistency. Among them, STT-RAM is a promising candidate as on-chip cache or main memory due to its high integration density, high access speed, and compatibility with CMOS fabrication process. However, MTJ as the core component of STT-RAM is sensitive to temperature fluctuations. Therefore, it is imperative to investigate performance and reliability of STT-RAM taking thermal factor into account. In the article, we build a thermal model which can capture thermal behavior of MTJ and support hybrid MTJ/CMOS electrical simulations. Based on the model, we evaluate thermal effect on 1T1MTJ memory cell from performance, latency and reliability perspectives. We also predict the temperature effect at more advanced 1Xnm technology node to analyze scalability of STT-RAM technology. After that, we propose a body-biasing sensing scheme to improve read margin and read reliability over the whole working temperature range. Experimental results show that our proposed technique can improve read reliability effectively. Additionally, with the same read margin constraint, it can reduce read disturbance significantly due to smaller sensing current at more advanced technology node. Our future work will be investigation of design techniques that ensure sensing reliability at very high temperature ($> 100°C$) when it is difficult to adjust driving abilities of transistors by body-biasing.

References

[1] N. S. Kim, T. Austin, D. Baauw, T. Mudge, K. Flautner, J. S. Hu, M. J. Irwin, M. Kandemir, and V. Narayanan, "Leakage current: Moore's law meets static power," *IEEE Computer*, vol. 36, no. 12, pp. 68–75, 2003.

[2] M. B. Taylor, "A landscape of the new dark silicon design regime," *IEEE Micro*, vol. 33, no. 5, pp. 8–19, 2013.

[3] H. Esmaeilzadeh, E. Blem, R. S. Amant, K. Sankaralingam, and D. Burger, "Dark silicon and the end of multicore scaling," in *Proc. IEEE/ACM International Symposium on Computer Architecture (ISCA'11)*, San Jose, USA, Jun.4–8 2011, pp. 365–376.

[4] J. Li, P. Ndai, A. Goel, S. Salahuddin, and K. Roy, "Design paradigm for robust spin-torque transfer magnetic ram (STT MRAM) from circuit/architecture perspective," *IEEE Trans. VLSI Syst.*, vol. 18, no. 12, pp. 1710–1723, 2010.

[5] W. Zhao and G. Prenat, *Spintronics-based Computing*. Springer: Berlin, Germany, 2015.

[6] L.-B. Faber, W. Zhao, J.-O. Klein, T. Devolder, and C. Chappert, "Dynamic compact model of spin-transfer torque based magnetic tunnel junction (MTJ)," in *Proc. IEEE Design & Technology of Integrated Systems in Nanoscale Era (DTIS'09)*, Cairo, Egypt, Apr.6–9 2009, pp. 130–135.

[7] L. Su, Y. Zhang, J.-O. Klein, Y. Zhang, A. Bournel, A. Fert, and W. Zhao, "Current-limiting challenges for all-spin logic devices," *Scientific Reports*, vol. 5, p. 14905, 2015.

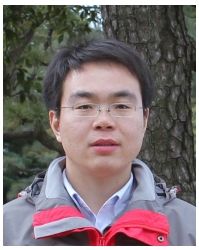
[8] J. Kim, K. Ryu, S. H. Kang, and S.-O. Jung, "A novel sensing circuit for deep submicron spin transfer torque MRAM (STT-MRAM)," *IEEE Trans. VLSI Syst.*, vol. 20, no. 1, pp. 181–186, 2012.

[9] H. Noguchi, K. Ikegami, K. Kushida, K. Abe, S. Itai, S. Takaya, N. Shimomura, J. Ito, A. Kawasumi, H. Hara, and S. Fujita, "A 3.3ns-access-time 71.2μw/mhz 1mb embedded STT-MRAM using physically eliminated read-disturb scheme and normally-off memory architecture," in *Proc. IEEE International Solid- State Circuits Conference (ISSCC'15)*, San Francisco, USA, Feb.22–26 2015, pp. 136–138.

[10] Z. Sun, X. Bi, H. Li, W.-F. Wong, and X. Zhu, "STT-RAM cache hierarchy with multiretention MTJ designs," *IEEE Trans. VLSI Syst.*, vol. 22, no. 6, pp. 1281–1293, 2014.
[11] R. Wang, L. Jiang, Y. Zhang, L. Wang, and J. Yang, "Selective restore: an energy efficient read disturbance mitigation scheme for future STT-MRAM," in *Proc. IEEE/ACM Design Automation Conference (DAC'15)*, San Francisco, USA, Jun.7–11 2015, pp. 1–6.
[12] Z. Pajouhi, X. Fong, and K. Roy, "Device/circuit/architecture co-design of reliable STT-MRAM," in *Proc. IEEE/ACM Conference on Design, Automation and Test in Europe (DATE'15)*, Grenoble, France, Mar.9–13 2015, pp. 1437–1442.
[13] Y. Lu, T. Zhong, W. Hsu, S. Kim, X. Lu, J. J. Kan, C. Park, W. C. Chen, X. Li, X. Zhu, P. Wang, M. Gottwald, J. Fatehi, L. Seward, J. P. Kim, N. Yu, G. Jan, J. Haq, S. Le, Y. J. Wang, L. Thomas, J. Zhu, H. Liu, Y. J. Lee, R. Y. Tong, K. Pi, D. Shen, R. He, Z. Teng, V. Lam, R. Annapragada, T. Torng, P.-K. Wang, and S. H. Kang, "Fully functional perpendicular STT-MRAM macro embedded in 40 nm logic for energy-efficient IOT applications," in *Proc. IEEE International Electron Devices Meeting (IEDM'15)*, Washington, USA, Dec. 7–9, 2015, pp. 26.1.1–26.1.4.
[14] W. Kang, L. Zhang, J.-O. Klein, Y. Zhang, D. Ravelosona, and W. Zhao, "Reconfigurable codesign of STT-MRAM under process variations in deeply scaled technology," *IEEE Trans. Electron Devices*, vol. 62, no. 6, pp. 1769–1777, 2015.
[15] A. Vatankhahghadim, W. Song, and A. Sheikholeslami, "A variation-tolerant MRAM-backed-SRAM cell for a nonvolatile dynamically reconfigurable FPGA," *IEEE Trans. Circuits Syst. II*, vol. 62, no. 6, pp. 573–577, 2015.
[16] J. Li, H. Liu, S. Salahuddin, and K. Roy, "Variation-tolerant spin-torque transfer (STT) MRAM array for yield enhancement," in *Proc. IEEE Custom Integrated Circuits Conference (CICC'08)*, San Jose, CA, Sep.21–24 2008, pp. 193–196.
[17] K. Lee and S. H. Kang, "Design consideration of magnetic tunnel junctions for reliable high-temperature operation of STT-MRAM," *IEEE Trans. Magn.*, vol. 46, no. 6, pp. 1537–1540, 2010.
[18] C. H. Shang, J. Nowak, R. Jansen, and J. S. Moodera, "Temperature dependence of magnetoresistance and surface magnetization in ferromagnetic tunnel junctions," *Physical Review B*, vol. 58, pp. R2917–R2920, 1998.
[19] T. Hagler, R. Kinder, and G. Bayreuther, "Temperature dependence of tunnel magnetoresistance," *Journal of Applied Physics*, vol. 89, no. 11, p. 7570, 2001.
[20] T. Kishi, H. Yoda, T. Kai, T. Nagase, E. Kitagawa, M. Yoshikawa, K. Nishiyama, T. Daibou, M. Nagamine, M. Amano, S. Takahashi, M. Nakayama, N. Shimomura, H. Aikawa, S. Ikegawa, S. Yuasa, K. Yakushiji, H. Kubota, A. Fukushima, M. Oogane, T. Miyazaki, and K. Ando, "Lower-current and fast switching of a perpendicular TMR for high speed and high density spin-transfer-torque MRAM," in *Proc. IEEE International Electron Devices Meeting (IEDM'08)*, San Francisco, USA, Dec.15–17 2008, pp. 1–4.
[21] C. J. Lin, S. H. Kang, Y. J. Wang, K. Lee, X. Zhu, W. C. Chen, X. Li, W. N. Hsu, Y. C. Kao, M. T. Liu, W. C. Chen, Y. Lin, M. Nowak, N. Yu, and L. Tran, "45nm low power CMOS logic compatible embedded STT MRAM utilizing a reverse-connection 1T/1MTJ cell," in *Proc. IEEE International Electron Devices Meeting (IEDM'09)*, Baltimore, USA, Dec.7–9 2009, pp. 256–259.
[22] S. Ikeda, H. Sato, H. Honjo, E. C. I. Enobio, S. Ishikawa, M. Yamanouchi, S. Fukami, S. Kanai, F. Matsukura, T. Endoh, and H. Ohno, "Perpendicular-anisotropy CoFeB-MgO based magnetic tunnel junctions scaling down to 1x nm," in *Proc. IEEE Electron Devices Meeting (IEDM'14)*, San Francisco, USA, Dec.15-17 2014, pp. 33.2.1–33.2.4.
[23] G. Jeong, C. Wooyoung, S. Ahn, J. Hongsik, G. Koh, H. Youngnam, and K. Kim, "A 0.24-μm 2.0-v 1T1MTJ 16-kb nonvolatile magnetoresistance RAM with self-reference sensing scheme," *IEEE J. Solid-State Circuits*, vol. 38, no. 11, pp. 1906–1910, 2003.
[24] T. W. Andre, J. J. Nahas, C. K. Subramanian, B. J. Garni, H. S. Lin, A. Omair, and W. L. Martino, "A 4-Mb 0.18-μm 1T1MTJ toggle MRAM with balanced three input sensing scheme and locally mirrored unidirectional write drivers," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 301–309, 2005.
[25] K. Skadron, M. R. Stan, W. Huang, V. Sivakumar, S. Karthik, and D. Tarjan, "Temperature-aware microarchitecture," in *Proc. IEEE/ACM International Symposium on Computer Architecture (ISCA'03)*, San Diego, USA, Jun.9–11 2003, pp. 2– 13.
[26] A. Raychowdhury, D. Somasekhar, T. Karnik, and V. De, "Design space and scalability exploration of 1T-1STT MTJ memory arrays in the presence of variability and disturbances," in *Proc. IEEE International Electron Devices Meeting (IEDM'09)*, Baltimore, USA, Dec. 7–9 2009, pp. 1–4.
[27] W. Guo, G. Prenat, V. Javerliac, M. E. Baraji, N. de Mestier, and C. . D. B. Baraduc, "Spice modelling of magnetic tunnel junctions written by spin-transfer torque," *Journal of Physics D: Applied Physics*, vol. 43, p. 215001, 2010.
[28] W. Zhao, J. Duval, J. O. Klein, and C. Chappert, "A compact model for magnetic tunnel junction (MTJ) switched by thermally assisted spin transfer torque (TAS + STT)," *Nanoscale Research Letters*, vol. 6, no. 1, p. 368, 2011.
[29] Y. Wang, Y. Zhang, E. Y. Deng, J. O. Klein, L. A. B. Naviner, and W. S. Zhao, "Compact model of magnetic tunnel junction with stochastic spin transfer torque switching for reliability analyses," *Microelectronics Reliability*, vol. 54, no. 9-10, pp. 1774–1778, 2014.
[30] X. Fong, S. H. Choday, P. Georgios, C. Augustine, and K. Roy, "Purdue nanoelectronics research laboratory magnetic tunnel junction model," Oct 2014. [Online]. Available: https://nanohub.org
[31] Y. Zhang, W. Zhao, Y. Lakys, J. O. Klein, J.-V. Kim, D. Ravelosona, and C. Chappert, "Compact modeling of perpendicular-anisotropy CoFeB/MgO magnetic tunnel junctions," *IEEE Trans. Electron Devices*, vol. 59, no. 3, pp. 819–826, 2012.
[32] S. Ikeda, K. Miura, H. Yamamoto, K. Mizunuma, H. D. Gan, M. Endo, S. Kanai, J. Hayakawa, F. Matsukura, and H. Ohno, "A perpendicular-anisotropy CoFeBMgO magnetic tunnel junction," *Nature Materials*, vol. 9, no. 9, pp. 721–724, 2010.
[33] T. L. Gilbert, "Classics in magnetics a phenomenological theory of damping in ferromagnetic materials," *IEEE Trans. Magn.*, vol. 40, no. 6, pp. 3443–3449, 2004.
[34] W. Zhao, C. Chappert, V. Javerliac, and J. P. Noziere, "High speed, high stability and low power sensing amplifier for MTJ/CMOS hybrid logic circuits," *IEEE Trans. Magn.*, vol. 45, no. 10, pp. 3784–3787, 2009.
[35] Predictive technology model. [Online]. Available: http://ptm.asu.edu
[36] Cadence. Spectre circuit simulator. [Online]. Available: http://www.cadence.com
[37] X. Dong, X. Wu, G. Sun, Y. Xie, H. Li, and Y. Chen, "Circuit and microarchitecture evaluation of 3D stacking magnetic ram MRAM as a universal memory replacement," in *Proc. IEEE/ACM Design Automation Conference (DAC'08)*, Austin, USA, Jun.5–9 2008, pp. 554–559.
[38] J. Z. Sun, "Spin-current interaction with a monodomain magnetic body: A model study," *Physical Review B*, vol. 62, no. 1, p. 570, 2000.
[39] Y. Shang, W. Fei, and H. Yu, "Fast simulation of hybrid CMOS and STT-MTJ circuits with identified internal state variables," in *Proc. IEEE/ACM Asia and South Pacific Design Automation Conference (ASP-DAC'12)*, Sydney, Australia, Jan.30–Feb. 2 2012, pp. 529 – 534.
[40] J. Kim, K. Ryu, J. P. Kim, S. H. Kang, and S.-O. Jung, "STT-MRAM sensing circuit with self-body biasing in deep submicron technologies," *IEEE Trans. VLSI Syst.*, vol. 22, no. 7, pp. 1630–1634, 2014.
[41] X. Bi, H. Li, and X. Wang, "STT-RAM cell design considering CMOS and MTJ temperature dependence," *IEEE Trans. Magn.*, vol. 48, no. 11, pp. 3821–3824, 2012.

Bi Wu received his B.S. and M.S. degree from China University of Mining and Technology, Xuzhou, China, and Beihang University, Beijing, China respectively. He is persuing Ph.D. degree of Electrical Engineering at Beihang University. His research interests include circuit level design and optimization of STT-RAM, STT-RAM reliability analysis and improvement, etc.

Yuanqing Cheng (S'11, M'13) received his Ph.D. degree from the Key Laboratory of Computer System and Architecture, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. After spending one year post-doc study at LIRMM, CNRS, France, he joined Beihang University, China as an assistant professor. His research interests include VLSI design for 3D integrated circuits considering thermal and defect issues, as well as spintronics computing system architecture design. He is currently an IEEE member and ACM member.



Jianlei Yang (S'12) received the B.S. degree in microelectronics from Xidian University, Xi'an, China, in 2009, and the Ph.D. degree in computer science and technology with Tsinghua University, Beijing, China, in 2014. Dr. Yang is currently a postdoctoral researcher with the Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania, United States. From 2013 to 2014, he was a research intern at Intel Labs China, Intel Corporation. His current research interests include numerical algorithms for VLSI power grid analysis and verification, spintronics and neuromorphic computing. Dr. Yang was the recipient of the first place on TAU Power Grid Simulation Contest in 2011, and the second place on TAU Power Grid Transient Simulation Contest in 2012. He was a recipient of IEEE ICCD Best Paper Award in 2013, and ACM GLSVLSI Best Paper Nomination in 2015.

Aida Todri-Sanial (M'03) received the B.S. degree in electrical engineering from Bradley University, IL in 2001, the M.S. degree in electrical engineering from Long Beach State University, CA, in 2003 and Ph.D. degree in electrical and computer engineering from the University of California, Santa Barbara, in 2009. She is currently a research scientist for French National Center of Scientific Research (CNRS) attached to Laboratoire dInformatique de Robotique et de Microlectronique de Montpellier (LIRMM). Previously she was an RD Engineer for Fermi National Accelerator Laboratory, IL where she was the recipient of John Bardeen Fellow in Engineering in 2009. She has also held visiting positions at Mentor Graphics, Cadence Design Systems, STMicroelectronics and IBM TJ Watson Research Center.

Weisheng Zhao (M'06, SM'14) received the Ph.D. degree in physics from the University of Paris-Sud, France, in 2007. From 2004 to 2008, he investigated Spintronic devices based logic circuits and designed a prototype for hybrid Spintronic/CMOS (90 nm) chip in cooperation with STMicroelectronics. Since 2009, he joined the CNRS as a tenured research scientist and his interest includes the hybrid integration of nano-devices with CMOS circuit and new non-volatile memory (40 nm technology node and below) like MRAM circuit and architecture design. He has authored or co-authored more than 150 scientific papers (e.g Advanced Material, Nature Communications, IEEE Transactions etc.); he is also the principal inventor of 4 international patents. From 2014, he becomes a youth 1000 plan distinguished professor in Beihang University, Beijing, China, and the associated editor for IEEE Transactions on Nanotechnology.