# A Scaling Roadmap and Performance Evaluation of In-Plane and Perpendicular MTJ Based STT-MRAMs for High-Density Cache Memory

Ki Chul Chun, Hui Zhao, Student Member, IEEE, Jonathan D. Harms, Tae-Hyoung Kim, Member, IEEE, Jian-Ping Wang, Member, IEEE, and Chris H. Kim, Senior Member, IEEE

Abstract—This paper explores the scalability of in-plane and perpendicular MTJ based STT-MRAMs from 65 nm to 8 nm while taking into consideration realistic variability effects. We focus on the read and write performances of a STT-MRAM based cache rather than the obvious advantages such as the denser bit-cell and zero static power. An accurate MTJ macromodel capturing key MTJ properties was adopted for efficient Monte Carlo simulations. For the simulation of access devices and peripheral circuitries, ITRS projected transistor parameters were utilized and calibrated using the MASTAR tool that has been widely used in industry. 6T SRAM and STT-MRAM arrays were implemented with aggressive assist schemes to mimic industrial memory designs. A constant  $J_{C0} \cdot RA/VDD$  scaling scenario was used which to the first order gives the optimal balance between read and write margins of STT-MRAMs. The thermal stability factor ensuring a 10 year retention time was obtained by adjusting the free layer thickness as well as assuming improvement in the crystalline anisotropy. Our studies based on the proposed scaling methodology show that in-plane STT-MRAM will outperform SRAM from 15 nm node, while its perpendicular counterpart requires further innovations in MTJ material in order to overcome the poor write performance scaling from 22 nm node onwards.

*Index Terms*—Cache, macromodel, magnetic tunnel junction (MTJ), roadmap, scalability, spin torque transfer (STT), STT-MRAM, variability.

# I. INTRODUCTION

**T** O meet the ever increasing demands for higher computing throughput while curbing chip power consumption, modern processor designs have turned to micro-architectural level parallelism instead of boosting raw clock frequency. The rapid increase in the number of cores for concurrent operation necessitates a commensurate increase in cache size to fully utilize the multi-core chip architecture. 6T SRAMs have been the embedded memory of choice due to their fast memory access while 1T1C eDRAMs have also been successfully deployed

Digital Object Identifier 10.1109/JSSC.2012.2224256

in high end products due to the dense bit-cell size [1]–[6]. However, SRAMs are faced with severe scaling challenges due to the relatively large bit-cell size, large static power dissipation, and conflicting requirements for read and write [7]–[9]. 1T1C eDRAMs on the other hand, require a complex capacitor fabrication process as well as a special ultra-low leakage access transistor, and also suffers from the destructive read resulting from the charge sharing operation, making it less attractive for future cache memory.

Spin-torque-transfer magnetic RAMs (STT-MRAMs) on the other hand are gaining popularity in the research community due to their compact bit-cell structure, good scalability and non-volatility (also means that standby power is zero) [10]–[17]. A comparison between 6T SRAM, 1T1C eDRAM, and STT-MRAM is shown in Table I confirming the aforementioned features of the later memory technology for practical chip implementations. For fair comparison, the three memory circuits were evaluated in the same 65 nm process with the magnetic tunnel junction parameters extracted from recent literature [10].

We first give a brief introduction to STT-MRAMs for context. An STT-MRAM bit-cell consists of an access transistor and a magnetic tunnel junction (MTJ). The MTJ device has a free magnetic layer and a pinned magnetic layer which are separated by a thin insulator layer as shown in Fig. 1(a). To the first order, an MTJ can be thought of as a two-terminal device with a variable resistance. The relationship between resistance and write current (i.e., R-I hysteresis curve) of a typical MTJ device is shown in Fig. 1(b). Depending on the direction of the write current, magnetization of the two layers can be set to a parallel state (P: low resistance, data '0') or an anti-parallel state (AP: high resistance, data'1') using spin polarized current as illustrated in Fig. 1(c). Two kinds of STT-MTJ's exist depending on the physical origin of the free layer magnetization: in-plane and perpendicular MTJ's [18]. Magnetic anisotropy of the former is determined by the lateral shape (e.g., aspect ratio) of the MTJ device while the latter's anisotropy has no shape anisotropy, simply depending on the material choice. In-plane MTJ's are far more mature than their perpendicular counterparts, however there is growing interest in the perpendicular devices as they are believed to have a low switching current density. Read operation is accomplished by sensing the resistance difference between the two states using a small bias current in conjunction with a Sense Amplifier (S/A). Here,  $(R_{AP} - R_{P})/R_{P}$  is referred to as the tunneling magnetoresistance (TMR) which is the ratio between the resistances of the two states. A higher TMR is preferred for a reliable read operation as it will generate a larger

Manuscript received June 15, 2012; revised September 04, 2012; accepted September 19, 2012. Date of publication December 05, 2012; date of current version January 24, 2013. This paper was approved by Associate Editor Peter Gillingham.

K. C. Chun, H. Zhao, J. D. Harms, J.-P. Wang, and C. H. Kim are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: chriskim@umn.edu).

T.-H. Kim is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

|  | 6T SRAM [7]                                    | 1T1C eDRAM [5]   | <sup>(1)</sup> STT-MRAM   |  |
|--|--|--|---|--|
| Cell<br>Schematic                            | WL<br>WL<br>WL<br>WL<br>WL<br>WL<br>WL<br>WL   | <sup>₽</sup><br><sup>1</sup><br><sup>1</sup><br><sup>1</sup><br><sup>1</sup><br><sup>1</sup> | BL<br>RMTJ<br>SL  |  |
| Process                                      | Logic compatible                               | +2 (FEOL) +3 (Cap)   | +MTJ  |  |
| Bit-cell (F <sup>2</sup> )                   | 135  | 30   | 28 w/ memory process<br>48 w/ logic process                       |  |
| <sup>(2)</sup> Reported cell<br>size (ratio) | 0.46x1.24=<br>0.5704µm² (1X)                   | 0.23x0.55=<br>0.1265µm² (0.22X)  | 0.3584µm² (0.63X) [15]  |  |
| <sup>(3)</sup> Redrawn cell<br>size (ratio)  | 0.575x2.05=<br>1.179µm² (1X)                   | 0.45x0.545=<br>0.245µm² (0.21X)  | 0.52x0.455=<br>0.238µm² (0.20X)                                   |  |
| <sup>(3)</sup> Redrawn 1Mb<br>macro (ratio)  | 1.377x1.124=<br>1.548mm² (1X)                  | 0.632x0.739=<br>0.467mm² (0.30X)   | 1.394x0.379=<br>0.528mm <sup>2</sup> (0.34X)                      |  |
| Data storage                                 | Latch (Static)                                 | Capacitor (20fF)   | Magnetic anisotropy<br>(non-volatile)                             |  |
| Cell access                                  | (+) Differential read<br>(-) Ratioed operation | (-) Destructive read<br>(-) Refresh  | (+) Excellent scalability<br>of I <sub>wr</sub> , (-) Small TMR   |  |
| Random cycle                                 | <sup>(3)</sup> 1GHz                            | 500MHz   | <sup>(3)</sup> 800MHz for read<br><sup>(3)</sup> 400MHz for write |  |
| Static power<br>(ratio)                      | 1X   | 0.2X   | 0X  |  |

 TABLE I

 COMPARISON BETWEEN THREE EMBEDDED MEMORY TECHNOLOGIES: 6T SRAM, 1T1C EDRAM, AND STT-MRAM

 $^{(1)}$ STT-MRAM bit-cell: 2T1MTJ w/ W<sub>TX</sub>=12F (2x6F) and in-plane MTJ dimension: 2FxF  $^{(2)}$ All designs are in 65nm,  $^{(3)}$ Based on the same 65nm low power CMOS process



Fig. 1. (a) Magnetic tunnel junction (MTJ) stack and its equivalent circuit model, a two-terminal device with variable resistance. (b) Resistance versus write current (R-I) hysteresis curve. (c) Illustration of spin torque transfer (STT) switching principle.

signal difference between the two states. The physics and dependency of TMR on MTJ processing parameters can be found in [19]. Fig. 2 shows the STT-MRAM bit-cell schematic and the signal voltage conditions for each operating mode.

Despite the recent advances in MTJ fabrication techniques, it is still unclear how the read and write performances of an STT-MRAM compares with those of conventional SRAMs or eDRAMs in future technology nodes in the presence of realistic variation effects. In this paper, we explore the scalability and variability of STT-MRAM based on a realistic MTJ and CMOS scaling scenario by comparing its read and write performances with 6T SRAM from 65 nm to 8 nm. Early works have already shown the feasibility of STT-MRAM in future technology nodes [20], [21]. The main highlights of this work compared to previous STT-MRAM research can be summarized as follows.

• A detailed scaling roadmap for in-plane and perpendicular MTJ devices is presented based on both dimensional scaling and MTJ anisotropy improvement. We show for the



|  | WL  | BL  | SL    | Comment  |  |
|--|-----|-----|-------|--|--|
| Stand-by   | 0V  | 0V  | 0V    | Zero leakage   |  |
| WR <sub>AP-P</sub>   | VPP | VDD | 0V    | VPP=2xVDD  |  |
| WR <sub>P-AP</sub>   | VPP | 0V  | VDD   | in this work   |  |
| RD<br>@T <sub>RD</sub> >>1ns   | VDD | 0V  | *∆VRD | I <sub>P-AP</sub> >I <sub>AP-P</sub> & I <sub>RD</sub> <i<sub>P-AP</i<sub> |  |
| RD<br>@T <sub>RD</sub> <1ns  | VPP | 0V  | *∆VRD | I <sub>P-AP</sub> >I <sub>AP-P</sub> & I <sub>RD</sub> >I <sub>P-AP</sub>  |  |
| *∆VRD≈0.5xI <sub>RD</sub> x(R <sub>AP</sub> -R <sub>P</sub> ) : simplified equation, |     |     |       |  |  |

(a)

does not account of MTJ or current source non-idealities (b)

Fig. 2. (a) STT-MRAM bit-cell schematic. (b) Signal voltages for each operating mode.

first time that write performance will become the scaling bottleneck for perpendicular STT-MRAMs.

- Statistical simulations are presented based on an accurate MTJ macromodel.
- For realistic transistor parameters, the ITRS MASTER<sup>TM</sup> tool was used extensively.
- Realistic STT-MRAM and 6T SRAM macros employing various assist schemes have been considered. We also use practical repair/redundant schemes in determining the cell failure rate limit.
- Specific design guidelines (e.g., assist scheme using a boosted wordline voltage) and device requirements (e.g., TMR > 200%) are provided to make STT-MRAM a viable memory technology.

The remainder of this paper is organized as follows. Section II presents the proposed STT-MTJ scaling roadmap for maintaining non-volatility and at the same time achieving optimal read and write margins. Section III describes the simulation methodology including MTJ macromodel, transistor parameters, array architectures, and variation sources. Section IV comprehensively compares the performances of 6T SRAM and STT-MRAM arrays considering variation effects with technology scaling, and conclusions are drawn in Section V.

## II. AN STT-MTJ SCALING ROADMAP

Fig. 3 shows the proposed scaling methodology for both in-plane and perpendicular STT-MTJs. The lateral dimension of the in-plane MTJ is fixed at  $2F \times F$ , where F is the half-pitch for a given process node, while the diameter of the perpendicular MTJ is fixed at F. This assumption gives the smallest bit-cell area for standalone applications as well as for embedded ones. It's worth mentioning that the bit-cell size of embedded memory cells are typically larger than that of their standalone

counterparts due to layout restrictions in a logic process; however we adhere to the aforementioned MTJ dimensions in order to evaluate the scalability for dense STT-MRAMs under the worst possible condition. The read and write performances and the non-volatile characteristics of an STT-MRAM depend on two main STT-MTJ parameters, namely the  $J_{C0} \cdot RA$  value and the thermal stability factor ( $\Delta$ ). Here,  $J_{C0}$  (unit: MA/cm<sup>2</sup>) is the critical current density at which the MTJ switching occurs, RA (unit:  $\Omega \cdot \mu m^2$ ) is the resistance area product of an MTJ, and  $\Delta$  signifies the difficulty of state reversal or the degree to which non-volatility is preserved under thermal fluctuation. Next we show a simple derivation of these two parameters which becomes the basis of our proposed scaling scenario.

# A. In-Plane STT-MTJ

Thermal stability of an in-plane STT-MTJ is primarily determined by the aspect ratio of the free layer. The governing equation for the thermal stability factor is as follows:

$$\Delta_{\rm in-plane} = \frac{E}{k_B T} = \frac{K_U V}{k_B T} = \frac{H_K M_S V}{2k_B T} \tag{1}$$

where  $K_U$  is the uniaxial anisotropy energy density, V is the volume of the free layer, T is the operating temperature,  $H_K$  is the anisotropy, and  $M_S$  is the saturation magnetization. The in-plane shape anisotropy  $H_K$  can be approximated as

$$H_K \approx \frac{8\pi M_S t (\text{AR} - 1)}{\text{wAR}} \tag{2}$$

where w, AR, and t are width, aspect ratio, and thickness of the free layer, respectively as shown in Fig. 3 [18]. Then, the thermal stability of in-plane STT-MTJs can be simplified as

$$\Delta_{\rm in-plane} \propto t^2 w ({\rm AR} - 1)$$
 (3)

As technology scales, w is shrunk to  $\alpha w$ , where  $\alpha$  is the scaling factor. Typical value for  $\alpha$  would be 0.7. This requires an increase in t by  $1/\alpha^{0.5}$  times in order to keep the thermal stability factor constant while AR remains constant. The thicker free layer increases  $J_{C0}$  which can be expressed as:

$$J_{C0} = \frac{2eaM_S t(H_K + H_{\text{ext}} + 2\pi M_S)}{\hbar\eta} \tag{4}$$

where e is the electron charge, a is the damping constant,  $\hbar$  is the reduced Planck's constant,  $H_{\text{ext}}$  is the external field, and  $\eta$ is the spin transfer efficiency [22]. If we combine (2) and (4) and assume that  $H_{\text{ext}} = 0$ , then  $J_{C0}$  scaling follows:

$$J_{C0} \propto t \left(1 + \frac{2t}{w}\right) \stackrel{\text{scaling}}{\longrightarrow} J'_{C0} \propto \frac{t}{\sqrt{\alpha}} \left(1 + \frac{2t/\sqrt{\alpha}}{\alpha w}\right)$$
(5)

This implies that  $J_{C0}$  and write threshold current  $(I_{C0})$  will scale by  $\sim 1/\alpha^{0.5}$  and  $\sim \alpha^{1.5}$  respectively since we can assume  $t \ll w$  (e.g., t = 1 nm and w = 65 nm). In the  $J_{C0} \cdot RA$  design space, a lower  $I_{C0}$  improves the write delay of STT-MRAMs at the expense of read margin [23]. As noted earlier,  $J_{C0}$  is the critical current density in [MA/cm<sup>2</sup>] and RA is the resistance-area product in  $[\Omega \cdot \mu m^2]$ . Hence, the product of  $J_{C0}$  and RA has a unit of [V] which has the physical meaning of the MTJ's voltage headroom. Furthermore, the read margin of STT-MRAM is proportional to  $J_{C0} \cdot RA(\sim I_{RD} \cdot \Delta R)$  while the maximum write



Fig. 3. STT-MTJ scaling scenario assuming dimensional scaling and new MTJ material.

current density  $J_{WR}$  (i.e., write margin) can be expressed as  $VDD/RA = (VDD/(J_{C0} \cdot RA)) \cdot J_{C0}$ . Consequently, keeping the  $J_{C0} \cdot RA/VDD$  constant with scaling provides a reasonable balance, albeit to the first order, between read and write margins. A different scaling scenario (e.g., constant RA/VDD) could also be employed; however this will strongly favor write over read resulting in a sub-optimal design. We chose a  $J_{C0} \cdot RA/VDD$  value of 0.25 based on simulation results in 65 nm, and this value will remain constant throughout the scaling analysis. Note that in reality, the  $J_{C0} \cdot RA/VDD$  value can be further tweaked in each technology for improved read and write performances but since the goal of this work is to establish a simple and practical scaling trend, we intentionally exclude these second order effects. The STT-MTJ scaling scenario described so far is summarized in Fig. 3 (left).

# B. Perpendicular STT-MTJ

Contrary to their in-plane counterparts, perpendicular STT-MTJs possess no shape anisotropy. In other words, the thermal stability factor does not depend on the aspect ratio of the free layer material and instead, is proportional to the bulk crystalline anisotropy  $(H_K^C)$  and the volume of the free layer. The thermal stability of perpendicular STT-MTJs is given as

$$\Delta_{\text{perpendicular}} = \frac{K_U V}{k_B T} = \frac{H_K^C M_S V}{2k_B T}$$
$$= \frac{H_K^C M_S \frac{\pi}{4} w^2 t}{2k_B T} \propto H_K^C w^2 t \tag{6}$$

where the bulk crystalline anisotropy  $H_K^C$  can be expressed as  $H_K - 4\pi M_S$  for perpendicular STT-MTJs [18]. As defined earlier, w and t are the width and thickness of the free layer while  $k_B$  and T are the Boltzmann constant and absolute temperature,

respectively. With technology scaling, w decreases to  $\alpha w$  and therefore  $H_K^C \cdot t$  should be adjusted to  $H_K^C \cdot t/\alpha^2$  in order to maintain the same thermal stability factor. Consequently,  $J_{C0}$  of perpendicular STT-MTJ scales by  $1/\alpha^2$  and  $I_{C0}$  remains constant. Similar to the in-plane case, the RA value is determined such that  $J_{C0} \cdot RA/VDD$  is constant as shown in Fig. 3 (right).

# C. STT-MTJ Scaling Roadmap

Table II shows the in-plane and perpendicular STT-MTJ scaling trends using the proposed scaling scenario summarized in Fig. 3. Here, we follow Intel's server processor trends by assuming that the number of cores will double every two process generations with a commensurate increase in on-chip cache density. As for the starting point, we chose the 65 nm node with four processor cores and a 4 MB per-core L3 cache [1]–[3]. The required thermal stabilities for a 10 year retention time were calculated based on the cache densities, the access word size, and allowable chip failure rates. The chip failure rate ( $F_{chip}$ ) can be estimated by expanding the cell reversal probability as follows:

$$F_{\rm chip} = 1 - \exp\left[-m\frac{t}{\tau_0}\exp\left\{-\Delta\left(1 - \frac{I_{\rm cell}}{I_{C0}}\right)\right\}\right]$$
(7)

where m is the number of memory cells, t is the 10 years retention target,  $\Delta$  is the thermal stability factor (=  $E/(k_BT)$ ), and  $I_{cell}$  is the cell current [13]. The allowable chip failure rate is set by the capability of a typical ECC and repair/redundancy scheme. For example, in 22 nm process node where the cache density is 48 MB, the allowable  $F_{chip}$  is 7.947e-7 for a repair scheme having a redundant WL and BL per every 64 WL's and 64 BL's, respectively. The corresponding thermal stability target is 74. The J<sub>C0</sub> and RA values at 65 nm process node were calibrated and refined based on our in-house MTJ test devices as

| apping layer  | MTJ dimensio              | MTJ dimension (nm <sup>2</sup> ) |                             |
|---|---------------------------|----------------------------------|-----------------------------|
| Co <sub>60</sub> Fe <sub>20</sub> B <sub>20</sub> 1.8 nm<br>MgO 0.85 nm | Thermal sta               | Thermal stability                |                             |
| Co <sub>40</sub> Fe <sub>40</sub> B <sub>20</sub> 2.4 nm<br>Ru 0.85 nm  | 2                         | P-AP                             |                             |
| Co <sub>70</sub> Fe <sub>30</sub> 2.5 nm                                | JC0 (MA/cm <sup>2</sup> ) | AP-P                             | 2.57                        |
| PtMn 15nm   | RA (Ω•µm                  | RA (Ω•μm²)                       |                             |
| Seed layer  | TMR (%)                   | TMR (%)                          |                             |
|   | Ultra fast                | P-AP                             | 560                         |
|   | switching (ps)            | AP-P                             | 586                         |
| 20  | Optimal write             | P-AP                             | 0.706 @T <sub>WR</sub> =0.6 |
| 130X50nm-   | energy (pJ/bit)           | AP-P                             | 0.286 @T <sub>WR</sub> =1.5 |

Fig. 4. Vertical structure and SEM image of fabricated STT-MTJ (left) and summary of measured MTJ parameters (right).

| Year  |               | 2007                   | 2010  | 2012 | 2015  | 2018  | 2021  | 2024   |
|---|---------------|------------------------|-------|------|-------|-------|-------|--------|
| Technology node (nm)  |               | 65                     | 45    | 32   | 22    | 15    | 11    | 8      |
| VDD: Supply voltage (V)   |               | 1.2                    | 1.1   | 1    | 0.9   | 0.85  | 0.8   | 0.75   |
| On-chip cache memory size (MByte)                                     |               | 16                     | 24    | 32   | 48    | 64    | 96    | 128    |
| Number of cores   |               | 4                      | 6     | 8    | 12    | 16    | 24    | 32     |
| Δ: Thermal stability (for 10 yrs retention)                           |               | 72                     | 73    | 74   | 74    | 75    | 75    | 76     |
| *t: Free layer thickness  | In-plane      | 1.00                   | 1.21  | 1.44 | 1.74  | 2.11  | 2.48  | 2.91   |
| *H <sub>k</sub> <sup>c</sup> •t: Anisotropy and t                     | Perpendicular | 1.00                   | 2.11  | 4.19 | 8.93  | 19.32 | 36.24 | 68.91  |
| J <sub>C0_P-AP</sub> : Critical current density (MA/cm <sup>2</sup> ) | In-plane      | 3.00                   | 3.70  | 4.55 | 5.86  | 7.87  | 10.45 | 14.65  |
|   | Perpendicular | 1.50                   | 3.16  | 6.28 | 13.40 | 28.97 | 54.35 | 103.37 |
| ICO P.AP: Threshold write   | In-plane      | 253.5                  | 149.9 | 93.3 | 56.7  | 35.4  | 25.3  | 18.7   |
| current (µA)  | Perpendicular | 49.8                   | 50.2  | 50.5 | 50.9  | 51.2  | 51.7  | 52.0   |
| J <sub>c0</sub> •RA/VDD (MTJ voltage headroom)                        |               | 0.25 when TMR=150%     |       |      |       |       |       |        |
| R <sub>AP</sub> A: Resistance area<br>product (Ω•μm²)                 | In-plane      | 10.00                  | 7.43  | 5.49 | 3.84  | 2.70  | 1.91  | 1.28   |
|   | Perpendicular | 20.00                  | 8.71  | 3.98 | 1.68  | 0.73  | 0.37  | 0.18   |
| J <sub>RD</sub> /J <sub>C0</sub> : Read current density               |               | 1.50 (based on Fig. 8) |       |      |       |       |       |        |

 TABLE II

 IN-PLANE AND PERPENDICULAR STT-MTJ PARAMETERS BASED ON SCALING TRENDS IN FIG. 3

\*t and Hk<sup>c</sup>•t are normalized to 65nm technology node.

well as the previously reported MTJ data in [10], [11]. The fabricated MTJ structure, SEM image, and summary of measured data are shown in Fig. 4 [24].

# III. MTJ MACROMODEL AND SIMULATION STRATEGY

Comparing the performances between STT-MRAM with 6T SRAM considering variation effects across different technology generations is an overwhelming task. Care must be taken to make sure that proper assumptions are made for a fair comparison without letting the design space explode due to the numerous device and circuit parameters. This section presents the simulation methodology used in this work including the MTJ macromodel, ITRS predicted transistor parameters, sub-array architecture, assist and repair techniques, and variation sources.

# A. Review of MTJ Macromodel Used in This Work

For efficient Monte Carlo simulations, an accurate MTJ macromodel capable of capturing important MTJ properties such as hysteresis, TMR dependency on bias voltage, and the relationship between  $I_{WR}$  and  $T_{WR}$  was developed in [25]. We

opt to use this semi-empirical model over a full LLG solver based physical model [26] due to its simplicity without compromising accuracy. Next, we briefly introduce the macromodel for completeness and for further insight. Fig. 5(a) shows the simplified block diagram of the proposed MTJ macromodel comprising HSPICE subcircuits based on behavioral models and internal current and voltage sources. The model consists of an MTJ electrode, a decision circuit, a bistable circuit, and a curve-fitting circuit. The MTJ electrode is a two terminal Voltage Controlled Resistor (VCR) device where the resistance is determined by the bias voltage applied to the MTJ (i.e.,  $V_{\rm TMR}$ ). The decision circuit determines the MTJ switching time ( $t_p$ ) which can be described using the thermal stability factor and the write threshold current:

$$t_p = t_0 \exp\left[\frac{E}{k_B T} \left(1 - \frac{I_{\rm MTJ}}{I_{C0}}\right)\right]$$
(8)

where  $I_{MTJ}$  is the current flowing through the MTJ and  $t_0$  is the nominal switching time. To model this transient behavior in HSPICE with physical relevance, we used the charging of a



Fig. 5. (a) Proposed MTJ macromodel for HSPICE simulations based on measured MTJ parameters. For simplicity, we list equations for the thermal activation condition only. (b) Fitting examples with the previous work of [10] and our measured TMJ data.

capacitor to determine when the MTJ should switch states. The capacitor charging process can be expressed as

$$V = \frac{1}{C}\frac{dQ}{dt} = \frac{I \times t}{C}$$
(9)

If we solve for I by combining (8) and (9) while assuming that V = 1 V, C = 1 nF, and  $t = t_p$ , then the capacitor charging current through the MTJ ( $I_1$  or  $I_2$  in the decision circuit) can be expressed as

$$I = \exp\left[-\frac{E}{k_B T} \left(1 - \frac{I_{\rm MTJ}}{I_{C0}}\right)\right]$$
(10)

Now, the MTJ switching time can be modeled in HSPICE as a capacitor charging time, namely the required time for a 1 nF capacitor to be charged such that  $\Delta V = 1$  V. Note that the capacitor charging current in (10) includes physical MTJ parameters such as thermal stability, MTJ current, and write threshold current that emulates an actual MTJ switching event. The buffered output signal  $V_{\text{decision}}$  is set to be +1 V when  $V_1 \ge 1$  V (P-to-AP switching) and -1 V when  $V_2 \le -1$  V (AP-to-P switching), respectively, and otherwise remains at 0 V. For precessional switching, extra fitting parameters were added to (8):

$$t_p = t_0 \exp\left[\frac{E}{k_B T} \left(1 - \frac{I_{\rm MTJ}}{I_{C0}}\right)\right] + t_{t \to p} (I_{t \to p}/I_{\rm MTJ})^2$$
(11)

where  $t_{t \to p}$  is the user defined pulse width where switching transitions from being dominated by thermal activation to precessional switching and  $I_{t\to p}$  is the macromodel calculated thermal to precession transition current using (8). The bistable circuit consists of a bistable multivibrator with amplitude control that can be initialized to a particular state. The amplitude control is implemented using behavioral description of an ideal amplifier. There are two capacitors in the feedback path that provide the initial condition of the state of the MTJ. The bistable circuit accepts  $V_{\text{decision}}$  as input and incorporates the bistable multivibrator circuits to replicate the hysteresis behavior of the MTJ. The output of the multivibrator ( $V_{\text{state}}$ ) is set to be +10 V when the last  $V_{\text{decision}}$  is -1 V (P state) and -10 V when the last  $V_{\text{decision}}$  is +1 V (AP state), respectively. The output of the bistable circuit is shifted and scaled, so that the signal ( $V_{\text{ctrl}}$ ) is bound between 1 (P state) and 1 + TMR<sub>ratio</sub> (AP state). To model the TMR dependency on bias voltage, we used the following Gaussian function:

$$V_{\rm TMR} = V_{\rm ctrl} \times \exp(-V_{\rm MTJ}^2/2c^2)$$
(12)

The curve-fitting function is a Gaussian approximation and is a simplified version of the approach taken in [27]. The value of the fitting parameter c is the bias voltage that causes the TMR to drop by 30%.  $V_{\rm TMR}$  is the final output of the control circuit. Fig. 5(b) shows fitting results of the MTJ macromodel showing good agreement with experimental data.

# B. Transistor Scaling Trends

As for the access devices and peripheral circuitry, transistor parameters from ITRS [28] were adopted from 65 nm down to 8 nm. Based on the high performance logic transistor roadmap from ITRS, we reproduced core NMOS parameters using the MASTAR tool that has been extensively used by ITRS to predict electrical characteristics of future CMOS devices. The resulting  $I_{dsat}$ 's of the core NMOS transistors were linearly



Fig. 6. High performance (HP) transistor scaling trend based on ITRS.

extrapolated for a smooth scaling trend. This prevents the general performance trends of SRAMs and STT-MRAMs from being distorted by any abrupt change in transistor parameters (e.g., high-k, FinFETs). The  $V_{thsat}$ 's of the core PMOS transistors are identical to the NMOS counterparts' while the  $I_{dsat}$ 's of PMOS were determined based on the  $I_{on,n}/I_{on,p}$  ratio projected by ITRS. An aggressive assist scheme that boosts the wordline voltage to  $2 \times VDD$  during write was adopted for robust switching with a commensurate increase in  $T_{OX}$  to ensure oxide reliability and a 1.2X longer gate length  $(L_{gate})$ to account for short channel effect concerns. During read, a lower bitline voltage that satisfies  $J_{RD}/J_{C0} = 1.5$  is used to prevent read disturbance (further details are given in the next section). The transistor parameters of the special thick oxide device were also extracted using the MASTAR tool. Further increase in the boosted voltage will result in oxide reliability issues and difficulty in generating a boosted level stably as experienced in standalone DRAM designs. The  $I_{dsat}$  and  $V_{thsat}$ trends of the core thin and the special thick  $T_{OX}$  devices for the STT-MRAM implementation are shown in Fig. 6.

#### C. Sub-Array Architecture and Variation Sources

The 6T SRAM used for comparison has the following transistor dimensions:  $W_{PU} = W_{min}, W_{PD} = 2 \times W_{min}$ and  $W_{ACCESS} = W_{min}$ , with all devices having a minimum channel length. This is the standard sizing method and extensive Monte Carlo simulations were performed to verify good read and write margins. The width of the STT-MRAM access device  $(W_{TX})$  is chosen as 12F based on the cell layout style in [13]. This makes the STT-MRAM cell size comparable to that of an eDRAM (in a memory specific process) or 3X denser when compared to an SRAM cell in a generic logic process. Fig. 7 shows the 128 kb sub-array architectures of 6T SRAM and STT-MRAM that has been extensively used in this work for performance evaluations. The unit 128 kb sub-array can be tiled to construct larger caches with mega byte densities. The layout dimension denoted in the figure shows that STT-MRAM is roughly 3X denser that 6T SRAM including all control circuitries in a 65 nm low power generic logic process. The



Fig. 7. 128 kb sub-array architectures of (a) SRAM and (b) STT-MRAM.

SRAM array employs assist schemes to achieve good read and write margins where the power supply level ( $V_{\rm SRAM}$ ) is dynamically controlled on a column-by-column basis [29]. More specifically,  $V_{\rm SRAM}$  is controlled to be a 0.1 V higher than the wordline voltage level ( $V_{\rm WL}$ ) during read to minimize read disturbance while during write,  $V_{\rm WL}$  is 0.1 V higher than  $V_{\rm SRAM}$  for robust operation. Similarly, advanced circuit techniques such as dummy cell averaging with disturb-free [16] and localized write drivers [13] were implemented for optimal read and write performances for STT-MRAM.

Variation sources present in practical industry designs have been included in our analysis as summarized in Table III. This includes process variation in the memory cells and the S/A as well as realistic variation for the resistances, capacitances, reference biases and supply levels. Here, a gradual scaling of  $\sigma_{Vt}$ and  $C_{BL}$  is again assumed to prevent the performance scaling trends from being distorted with any abrupt change in the transistor parameters and the parasitic capacitances although this can happen in real situations. Fig. 8 shows simulation results of  $J_{RD}/J_{C0}$  versus read disturb rate at  $T_{RD} = 2$  ns with the proposed scaling scenario and simulation methodology. This indicates that read disturb worsens for  $J_{RD}/J_{C0} > 2$ . The simulation result agrees with the previous conclusions in [20] showing that  $J_{RD}/J_{C0} < 2$  is required for a disturb-free read operation

TABLE III Simulation Set-Up for Evaluating SRAM and STT-MRAM Variability

|  | 6T SRAM   | STT-MRAM   |  |  |
|--|---|--|--|--|
| Power supply noise                       | -10% to account for supply noise                                    |  |  |  |
| Bit-cell                                 | Device mismatches   |  |  |  |
| Parasitic capacitance (C <sub>BL</sub> ) | $\sigma/\mu$ =5%: each $\mu$ are calculated based on sub-array size |  |  |  |
| Resistance area product                  | -   | σ/μ=5%   |  |  |
| Sense Amplifier (S/A)                    | Voltage S/A pair<br>mismatches                                      | I-applying and V-sensing method<br>(AP direction read) + Voltage S/A<br>: I <sub>REF</sub> σ/μ=2.5%, S/A pair mismatches |  |  |
| Reference cell -                         |   | Reference cell averaging<br>scheme with MTJ replica cells  |  |  |
| Write threshold current                  | -   | σ/u=5%   |  |  |

\* Mismatches are based on inverse square root relationship of devices' areas.

\* µ(C<sub>BL</sub>) is assumed to be scaled proportional to scaling factor.



Fig. 8. Simulated read disturb rate for different  $J_{RD}/J_{C0}$  ratios.

under a 50% read duty cycle (i.e.,  $T_{RD}/T_{CYCLE}$ ). So we chose a  $J_{RD}/J_{C0}$  value of 1.5 in this work.

## IV. COMPARISON WITH SRAM

In order to demonstrate the potential of STT-MRAM as an alternative for high density on-chip cache memory, this section presents macro level performance comparisons with 6T SRAM down to the 8 nm technology node. Extensive Monte-Carlo simulations were performed on megabit density SRAM and STT-MRAM arrays to estimate their performance under a practical scaling scenario [30], [31].

# A. Macro Level Performance

Despite having a longer access time, dense memories such as eDRAMs or STT-MRAMs are preferred for low level caches due to their smaller memory footprint which results in a shorter global interconnect delay. Since STT-MRAMs have a 3–5X bit-cell density advantage over SRAMs, their system level performance is expected to be higher even with a longer sensing delay. We first verify this conventional wisdom by comparing the latencies between several embedded memory options (Fig. 9). Critical path delay simulation results show that the normalized WL-to-BL read sensing delays of 6T SRAM, 1T1C eDRAM [6], 2T eDRAM [32], STT-MRAM are approximately 1X, 5X, 2X, and 3X, respectively. For small



Fig. 9. Latency comparison between several embedded memory options with (a) 1 Mb and (b) 64 Mb densities. The longer interconnect delay (even with repeaters) make dense memories such as eDRAM and STT-MRAM more attractive for larger caches.

caches (e.g., L1), the bitline sensing delay becomes a greater portion of the total cache latency as shown in Fig. 9(a). As a result, SRAM achieves the shortest cache latency. For larger caches (e.g., L2 or L3) however, the global interconnect delay, even with repeaters, dominates the cache latency making dense memories such as eDRAM or STT-MRAM attractive over SRAM as shown in Fig. 9(b). This is the reason why 1T1C eDRAM replaced SRAM in several processor designs [4]–[6] and the same principle can be applied for STT-MRAMs. We estimate that as long as STT-MRAMs have a bitline delay that is 3–5X longer than that of SRAMs, they can outperform SRAMs for cache densities greater than 64 Mb (= 8 MByte).

<sup>\*</sup> Based on historic data, we assume  $\sigma_{v_i}/F$  is constant with technology scaling



Fig. 10. In-plane STT-MRAM write delay scaling trends: (a) Actual delay values. (b) Delay normalized to that of SRAM.

In the rest of this paper, we will assume that an STT-MRAM with a 3X longer bitline access time compared to an SRAM will have an iso-latency, which is a conservative assumption since cache densities projected in Table II are much greater than 8 MByte.

# B. In-Plane STT-MRAM versus 6T SRAM

Fig. 10(a) and (b) show the actual and normalized values of  $6\sigma$  write delays for SRAM and in-plane STT-MRAM, respectively. Here, the write time  $T_{WR}$  is defined as the delay from WL activation to the point when the cell node flips for SRAMs, and to the point when the MTJ switches for STT-MRAMs. With technology scaling,  $T_{WR}$  of 6T SRAM degrades due to the reduced supply voltage level and the ratioed operation even after applying a write assist scheme [29]. In contrast,  $I_{C0}$  of in-plane STT-MRAMs scales by roughly  $\alpha^{1.5}$  improving the MTJ switching time with scaling. Unfortunately, this is not enough for STT-MRAMs to outperform SRAMs in terms of write latency even for an 8 nm process node. If the supply voltage of STT-MRAMs can be increased by 0.3 V, STT-



Fig. 11. Write delay distributions of SRAM and in-plane STT-MRAM (P-to-AP) for a 1 Mb macro in 15 nm.

MRAMs can outperform SRAMs from 15 nm on wards when following the constant  $J_{C0} \cdot RA/VDD$  scaling scenario. Unlike in STT-MRAMs where the standby power is zero, increasing the supply voltage is difficult in SRAMs as it will directly impact the leakage power consumption. Fig. 11 shows the detailed  $T_{WR}$  distributions of SRAM and in-plane STT-MRAM in 15 nm obtained by running 2<sup>20</sup> Monte-Carlo runs in HSPICE, which represents the cell-to-cell variation of a 1 Mb memory macro.

Fig. 12 shows the  $6\sigma$  read sensing delay comparison between SRAM and in-plane STT-MRAM. Again, absolute values are shown in Fig. 12(a) whereas normalized values are shown in Fig. 12(b). Both SRAM and STT-MRAM arrays have a 1 Mb macro density and a 256 cells-per-BL architecture. Here, the read sensing delay  $T_{RD}$  is defined as the delay from WL activation to the time when  $\Delta BL = VBL(D1) - VBL(D0)$ reaches 50 mV for SRAMs, and to the time when VBL-VREF (or VREF-VBL) reaches 25 mV for STT-MRAMs, respectively. Due to the single-ended sensing nature and the limited TMR, it is not practical to enforce the same BL voltage difference requirement for the two memory types. Instead, we assume a more robust S/A design for STT-MRAM such as the Negative Resistance Read Scheme (NRRS) [14] or pre-amplifier technique which of course will result in a larger S/A area. Alternatively, a S/A with a 4X larger input pair or an offset cancellation scheme may be considered to cope with the reduced sensing voltage difference (e.g., 25 mV for STT-MRAMs compared to 50 mV for SRAMs). The layout dimensions of the 128 kb sub-arrays in Fig. 7 already take this area overhead into consideration. Note that STT-MRAM is still 3X denser than SRAM after accounting for all control circuits and special S/A's owing to the  $\sim 5X$  smaller bit-cell." The  $6\sigma$  $T_{RD}$  comparison in Fig. 12 based on the above assumptions indicates that a TMR greater than 200% is required in order for STT-MRAMs to be advantageous over SRAMs. Considering that today's state-of-the-art MTJ devices have a TMR in the range of 100-150%, our results show that further improvement in TMR is needed for STT-MRAMs to become practical. Fig. 13 shows the T<sub>RD</sub> distributions of SRAM and in-plane STT-MRAM for a 1 Mb array in 15 nm. Even with the reduced



Fig. 12. In-plane STT-MRAM read sensing delay scaling trends: (a) Actual delay values. (b) Delay normalized to that of SRAM.



Fig. 13. Read sensing delay distributions of SRAM and in-plane STT-MRAM for a 1 Mb macro in 15 nm.

BL voltage difference of 25 mV, STT-MRAM suffers from read failure due to the small TMR ratio. This requirement can be relaxed by increasing  $J_{C0} \cdot RA$  since the fast write performance in Fig. 10 can be traded off for better read margin.



Fig. 14. J<sub>C0</sub> and RA scaling trends of in-plane and perpendicular STT-MTJs.

#### C. In-Plane STT-MRAM versus Perpendicular STT-MRAM

Fig. 14 shows the scaling trends of  $J_{C0}$  and RA for both in-plane and perpendicular STT-MTJs. Due to the different physical origin of magnetic anisotropy, the scaling trend of perpendicular MTJs is drastically different from that of their in-plane counterparts. Namely,  $J_{C0}$  scales by  $1/\alpha^2$  ( $\alpha$  is the scaling factor) which results in a constant  $I_{C0}$  value irrespective of the technology node [18]. The popular belief is that  $I_{C0}$ of perpendicular STT-MTJs is smaller than that of in-plane devices for the same thermal stability factor. Our projection based on the scaling methodology in Table II shows that this hypothesis is true only down to the 22 nm node. Since  $I_{C0}$ of in-plane STT-MRAMs scales by roughly  $\alpha^{1.5}$  while I<sub>C0</sub> remains constant for perpendicular STT-MRAMs, the  $I_{C0}$ values will cross over at around 15 nm. Under a constant  $J_{C0} \cdot RA/VDD$  scaling scenario, the read margin of perpendicular STT-MRAMs would be similar to that of in-planes as shown in Fig. 15(a) while write performance will become worse as shown in the  $T_{WR}$  scaling trend in Fig. 15(b). This is primarily due to the lower drive current of the access transistor devices (even with a wordline voltage boosted up to  $2 \times VDD$ ) resulting from the smaller width and lower VDD. Note that the RA of a perpendicular MTJ needs to be scaled exponentially in order to maintain a constant  $J_{C0} \cdot RA/VDD$  ratio as shown in Fig. 14. Our projections show that the required RA at the 8 nm node should be less than  $0.2\Omega \cdot \mu \text{ m}^2$  which could lead to severe reliability issues in the thin MTJ tunneling barrier as well as imposing limits on the achievable TMR value. This suggests that significant reduction in  $J_{C0}$  must be achieved through MTJ device engineering for perpendicular STT-MRAMs to become a viable option.

## V. CONCLUSION

Zero standby power dissipation and high bit cell density are the two obvious advantages of STT-MRAM. However, performance metrics (e.g., cache latency and access time) or scaling trend of this emerging memory technology have not been studied thoroughly in the past. This work explores the scalability and variability of in-plane and perpendicular MTJ based STT-MRAMs by comparing their read and write performances with those of SRAM. We consider realistic MTJ properties as

TABLE IV QUALITATIVE SUMMARY OF THIS WORK

|              | In-plane STT-MRAM   | Perpendicular STT-MRAM   |  |  |
|--------------|---|--|--|--|
| Obvious      | Up to 5X higher bit cell density, zero static power dissipation, nominally-off instant-on operation   |  |  |  |
| Good<br>news | Cache latency shorter than that of<br>6T SRAM for 15 nm and beyond  | One time improvement in<br>write performance over in-plane<br>MTJ's  |  |  |
| Bad<br>news  | 1. TMR>200% required<br>2. Aggressive assist scheme needed<br>(e.g. WL boosted to 2xVDD with<br>thick T <sub>ox</sub> access device)<br>3. Robust sense amps required | <ol> <li>Write performance degrades<br/>beyond 15nm due to constant I<sub>C0</sub></li> <li>Extremely low RA required →</li> <li>likely to cause MTJ reliability issues</li> <li>Robust sense amps required</li> </ol> |  |  |



Fig. 15. Perpendicular STT-MRAM scaling trends. (a) Sensing delay comparison with SRAM. (b) Write time comparison with SRAM.

well as ITRS projected CMOS device scaling trends for demonstrating the potential of STT-MRAM. We propose an STT-MTJ scaling scenario based on dimensional scaling in conjunction with MTJ material innovations (e.g., RA,  $J_{C0}$ ). A constant  $J_{C0} \cdot RA/VDD$  ratio was assumed for optimal read and write performances while the thermal stability factor was determined for a 10 year data retention. The simulation methodology utilizes an efficient MTJ macromodel based on experimental data, ITRS projected transistor parameters for access devices and peripheral circuitries, state-of-the-art sub-array architectures and assist schemes, and variation sources present in practical industry designs. Our studies based on extensive Monte Carlos simulations show that the in-plane STT-MRAM is a promising alternative for future high density cache memories, outperforming SRAMs from the 15 nm process node. We show that a TMR ratio greater than 200% is needed in conjunction with an aggressive assist scheme employing special thick oxide access transistors and a boosted WL voltage of 2 × VDD. Perpendicular STT-MRAMs on the other hand suffer from poor write performance scaling trends due to the difficulty in scaling  $I_{C0}$ . Without a significant reduction in  $J_{C0}$  through MTJ device innovations, write performance is expected to become the key bottleneck for perpendicular MTJ's in future technology nodes. Table IV provides a qualitative summary of this work.

### ACKNOWLEDGMENT

The author would like to thank the Intel post-CMOS circuits and architecture program for financial support and technical feedback, and a scholarship from Samsung Electronics.

#### REFERENCES

- S. Rusu et al., "A 45 nm 8-core enterprise Xeon® processor," IEEE J. Solid-State Circuits, vol. 45, no. 1, pp. 7–14, Jan. 2010.
- [2] S. Rusu et al., "A 65-nm dual-core multithreaded Xeon® processor with 16-MB L3 cache," *IEEE J. Solid-State Circuits*, vol. 42, no. 1, pp. 17–25, Jan. 2007.
- [3] R. J. Riedlinger et al., "A 32 nm 3.1 billion transistor 12-wide-issue Itanium® processor for mission-critical servers," in *IEEE ISSCC Dig. Tech. Papers*, 2011, pp. 84–85.
- [4] R. Kalla, B. Sinharoy, W. J. Starke, and M. Floyd, "POWER7: IBM's next-generation server processor," *IEEE Micro*, vol. 30, no. 2, pp. 7–15, Mar.–Apr. 2010.
- [5] J. Barth et al., "A 500 MHz random cycle, 1.5 ns latency, SOI embedded DRAM macro featuring a three-transistor micro sense amplifier," *IEEE J. Solid-State Circuits*, vol. 43, no. 1, pp. 86–95, Jan. 2008.
- [6] J. Barth et al., "A 45 nm SOI embedded DRAM macro for the power™ processor 32 MByte on-chip L3 cache," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 64–75, Jan. 2011.
- [7] K. Zhang et al., "SRAM design on 65-nm CMOS technology with dynamic sleep transistor for leakage reduction," *IEEE J. Solid-State Cir*cuits, vol. 40, no. 4, pp. 895–901, Apr. 2005.
- [8] F. Hamzaoglu et al., "A 153 Mb-SRAM design with dynamic stability enhancement and leakage reduction in 45 nm high-k metal-gate CMOS technology," in *IEEE ISSCC Dig. Tech. Papers*, 2008, pp. 376–377.
- [9] P. Packan *et al.*, "High performance 32 nm logic technology featuring 2nd generation high-k + metal gate transistors," in *IEEE IEDM Dig. Tech. Papers*, 2009, pp. 659–662.

- [10] M. Hosomi *et al.*, "A novel nonvolatile memory with spin torque transfer magnetization switching: Spin-RAM," in *IEEE IEDM Dig. Tech. Papers*, 2005, pp. 459–462.
- [11] T. Kishi et al., "Lower-current and fast switching of a perpendicular TMR for high speed and high density spin-transfer-torque MRAM," in *IEEE IEDM Dig. Tech. Papers*, 2008, pp. 309–312.
- [12] T. Kawahara et al., "2 Mb SPRAM (SPin-transfer torque RAM) with bit-by-bit bi-directional current write and parallelizing-direction current read," *IEEE J. Solid-State Circuits*, vol. 43, no. 1, pp. 109–120, Jan. 2008.
- [13] R. Takemura et al., "A 32-Mb SPRAM with 2T1R memory cell, localized bi-directional write driver and '1'/'0' dual-array equalized reference scheme," *IEEE J. Solid-State Circuits*, vol. 45, no. 4, pp. 869–879, Apr. 2010.
- [14] D. Halupka *et al.*, "Negative-resistance read and write schemes for STT-MRAM in 0.13 μm CMOS," in *IEEE ISSCC Dig. Tech. Papers*, 2010, pp. 256–257.
- [15] K. Tsuchida et al., "A 64 Mb MRAM with clamped-reference and adequate-reference schemes," in *IEEE ISSCC Dig. Tech. Papers*, 2010, pp. 258–259.
- [16] J. P. Kim et al., "A 45 nm 1 Mb embedded STT-MRAM with design techniques to minimize read-disturbance," in Proc. VLSI Circuits Symp., 2011, pp. 296–297.
- [17] K. Lee and S. H. Kang, "Development of embedded STT-MRAM for mobile system-on-chips," *IEEE Trans. Magnetics*, vol. 47, no. 1, pp. 131–136, Jan. 2011.
- [18] D. Apalkov et al., "Comparison of scaling of in-plane and perpendicular spin transfer switching technologies by micromagnetic simulation," *IEEE Trans. Magnetics*, vol. 46, no. 6, pp. 2240–2243, June 2010.
- [19] J. Hayakawa, S. Ikeda, F. Matsukura, H. Takahashi, and H. Ohno, "Dependence of giant tunnel magnetoresistance of sputtered CoFeB/ MgO/CoFeBMagnetic tunnel junctions on MgO barrier thickness and annealing temperature," *Jpn. J. Appl. Phys.*, vol. 44, pp. L587–L589, 2005.
- [20] A. Raychowdhury, D. Somasekhar, T. Karnik, and V. K. De, "Modeling and analysis of read (RD) disturb in 1T-1STT MTJ memory bits," in *IEEE DRC Dig. Tech. Papers*, 2010, pp. 43–44.
- [21] R. Dorrance, R. Fenbo, Y. Toriyama, A. A. Hafez, C.-K. K. Yang, and D. Markovic, "Scalability and design-space analysis of a 1T-1MTJ memory cell for STT-RAMs," *IEEE Trans. Electron Devices*, vol. 59, no. 4, pp. 878–887, 2012.
- [22] Z. Diao et al., "Spin-transfer torque switching in magnetic tunnel junctions and spin-transfer torque random access memory," J. Phys.: Condensed Matter, vol. 19, no. 16, pp. 165209-1–165209-13, Apr. 2007.
- [23] 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 *IEDM Dig. Tech. Papers*, 2009, pp. 707–710.
- [24] H. Zhao et al., "Low writing energy and sub nanosecond spin torque transfer switching of in-plane magnetic tunnel junction for spin torque transfer random access memory," J. Appl. Phys., vol. 109, no. 7, pp. 07C720-1–07C720-3, Mar. 2011.
- [25] J. D. Harms, F. Ebrahimi, X. Yao, and J.-P. Wang, "SPICE macromodel of spin-torque-transfer-operated magnetic tunnel junctions," *IEEE Trans. Electron Devices*, vol. 57, no. 6, pp. 1425–1430, June 2010.
- [26] G. Panagopoulos, C. Augustine, and K. Roy, "A framework for simulating hybrid MTJ/CMOS circuits: Atoms to system approach," *Design Automation and Test Europe*, pp. 1443–1446, 2012.
- [27] S. Lee, S. Lee, H. Shin, and D. Kim, "Advanced HSPICE macromodel for magnetic tunnel junctions," *Jpn. J. Appl. Phys.*, vol. 44, no. 4B, pp. 2696–2700, Apr. 2005.
- [28] International Technology Roadmap for Semiconductors 2009 [Online]. Available: http://www.itrs.net, [Online]. Available:
- [29] K. Zhang et al., "A 3-GHz 70-Mb SRAM in 65-nm CMOS technology with integrated column-based dynamic power supply," *IEEE J. Solid-State Circuits*, vol. 41, no. 1, pp. 146–151, Jan. 2006.
- [30] K. Agarwal and S. Nassif, "The impact of random device variation on SRAM cell stability in sub-90-nm CMOS technologies," *IEEE Trans. Very Large Scale Integration (VLSI) Syst.*, vol. 16, no. 1, pp. 86–97, Jan. 2008.
- [31] R. Beach et al., "A statistical study of magnetic tunnel junctions for high-density spin torque transfer-MRAM (STT-MRAM)," in IEEE IEDM Dig. Tech. Papers, 2008, pp. 306–308.
- [32] K. Chun, P. Jain, T. Kim, and C. H. Kim, "A 667 MHz logic-compatible embedded DRAM featuring an asymmetric 2T gain cell for high speed on-die caches," *IEEE J. Solid-State Circuits*, vol. 47, no. 2, Feb. 2012.



Ki Chul Chun received the B.S. degree in electronics engineering from Yonsei University, Seoul, Korea, in 1998, the M.S. degree in electrical engineering from KAIST, Daejeon, Korea, in 2000, and the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis, in 2012. In 2000, he joined the Memory Division, Samsung Electronics, Gyeonggi-Do, Korea, where he has been involved in DRAM circuit design. After his Ph.D. study at the U of MN, he rejoined Samsung Electronics in 2012, where he has worked for

Low-Power DRAM development.

Dr. Chun is the recipient of a Samsung Ph.D. Scholarship for outstanding employees and an ISLPED 2009 Low Power Design Contest Award. His research interests include digital, mixed-signal and memory circuit designs with special focus on DRAM, PRAM, and STT-MRAM in scaled technologies.



**Hui Zhao** (S'10) is currently pursuing the Ph.D. degree in electrical engineering at the University of Minnesota, Minneapolis, MN. She received the B.S. degree in optical information science and technology and M.S. degree in optics from Fudan University, Shanghai, China, in 2005 and 2008 respectively.

Her research focuses on developing MgO magnetic tunnel junctions (MTJ) with low critical current and fast switching speed for STT-RAM application. More specific work includes the MTJ material stack design and cell nanofabrication, MTJ cell

and CMOS circuit integration, advanced device characterization and physics study. In 2012, she was a summer intern in Seagate, working on the read head ferromagnetic resonance characterization.



**Jonathan D. Harms** is a graduate student at University of Minnesota—Twin Cities, where he is expecting to finish his M.S. in electrical and computer engineering. (F'12) He worked on magnetic tunnel junction (MTJ) based logic device, magnetic quantum cellular automata (MQCA), and other non-volatile logic and storage devices. His research focused on architectural designs and simulations of logic-in-memory devices.

He worked as an intern at Seagate in the Head-

Disk Integration Group on characterization of head fly height. In the fall of 2012 he took a job at Micron Technology working in the Emerging Memory Group on STT-RAM.



**Tae-Hyoung Kim** (M'06) received the B.S. and M.S. degrees in electrical engineering from Korea University, Seoul, Korea, in 1999 and 2001, respectively. He received the Ph.D. degree in electrical and computer engineering from the University of Minnesota, Minneapolis, Minnesota, USA, in 2009.

He joined the Device Solution Network Division, Samsung Electronics, Yong-in, Korea, in 2001. From 2001 to 2005, he performed research on the design of high-speed SRAM memories. In summer 2007 and 2008, he was with IBM T. J. Watson Research Center,

Yorktown Heights, NY, where he worked on NBTI/PBTI-induced circuit reliability measurement circuits. In summer 2009, he was an intern at Broadcom where he performed research on ultra-low power SRAM design. He joined the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore in 2009 where he is currently an assistant professor.

Prof. Kim is the recipient of a best paper award at ISOCC 2011, the 2008 AMD/CICC Student Scholarship Award, 2008 DAC/ISSCC Student Design Contest Award, 2008 Departmental Research Fellowship from University of Minnesota, 2008 Samsung Humantec Thesis Award (2008, 2001, and 1999), and 2005 ETRI Journal Paper of the Year Award. His research interests include ultra-low power and high performance integrated circuits including low voltage circuits, silicon and non-silicon memories, and energy efficient circuits and systems.



Jian-Ping Wang (M'97) is the Distinguished McKnight University professor at the University of Minnesota. He is the faculty member at the Electrical and Computer Engineering Department also a graduate faculty member of Departments of Physics, Chemical Engineering and Materials Science and Biomedical Engineering at the University of Minnesota. He is the associate director of the Center for Micromagnetics and Information Technologies (MINT). His current research programs focus on searching, fabricating and fundamentally

understanding new nanomagnetic and spintronic materials and devices. He received his Ph.D. degree from Institute of Physics, Chinese Academy of Sciences, in 1995. He was the founding program manager for the Magnetic Media and Materials program in Data Storage Institute, Singapore, from 1998 to 2002. He received the INSIC technical award in 2006 for his pioneering experimental work on exchange coupled composite magnetic media and the 2011 College of Science and Engineering Outstanding Professor Award for his dedication to teaching undergraduates. He has authored and co-authored more than 200 publications in peer-reviewed top journals and conferences and holds 17 patents.



**Chris H. Kim** (M'04–SM'10) received his B.S. and M.S. degrees from Seoul National University and a Ph.D. degree from Purdue University. He spent a year at Intel Corporation where he performed research on variation-tolerant circuits, on-die leakage sensor design and crosstalk noise analysis. He joined the electrical and computer engineering faculty at the University of Minnesota, Minneapolis, MN, in 2004 where he is currently an associate professor.

Prof. Kim is the recipient of an NSF CAREER Award, a Mcknight Foundation Land-Grant Pro-

fessorship, a 3M Non-Tenured Faculty Award, DAC/ISSCC Student Design Contest Awards, IBM Faculty Partnership Awards, an IEEE Circuits and Systems Society Outstanding Young Author Award, ISLPED Low Power Design Contest Awards, and an Intel Ph.D. Fellowship. He is an author/coauthor of 100+ journal and conference papers and has served as a technical program committee chair for the 2010 International Symposium on Low Power Electronics and Design (ISLPED). His research interests include digital, mixed-signal, and memory circuit design in silicon and non-silicon (such as organic TFT and spin) technologies.