Spin-Hall Effect MRAM Based Cache Memory: A Feasibility Study

Jongyeon Kim, Bill Tuohy, Cong Ma, Won Ho Choi, Ibrahim Ahmed, David Lilja, and Chris H. Kim
Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455
Phone: 612-280-6796, Email: kimx2889@umn.edu

One of the key objectives of STT-MRAM research has been on minimizing switching current while maintaining the required nonvolatility. To address this challenge, non-traditional MRAMs based on novel switching mechanisms have been proposed. In particular, spin-Hall effect (SHE) which utilizes large spin currents generated in the direction transverse to the charge current have been recently drawing attention [1]. Despite early promises such as lower switching current by means of efficient spin generation (i.e. I_{spin}/I_{charge}>100%) and longer device lifetime owing to the decoupled read and write paths, there is still a lack of a comprehensive study for benchmarking SHE-MRAM against other memory technologies. In this work, we explore the trade-off points across different levels of design abstraction (i.e. device, circuit, and architecture) to evaluate the feasibility of SHE-MRAM for large on-die cache memory.

Fig. 1 illustrates the generation of pure spin current by SHE, along with the cell structure of a SHE-MRAM which requires two transistors for separate read and bidirectional write operation. In Fig. 2, the target thermal stability (Δ) criterion is set by considering a fixed bit-cell failure rate under a 10 year data retention time [2]. The maximum read current is also determined under the same degree of read disturbance failure rate. The material parameters and device dimensions used in this study are listed in Table 1, which are determined based on the thermal stability (Δ) requirement at the 22nm node. Note that an in-plane MTJ is used as a storage element of SHE-MRAM since the direction of spin from SHM is aligned with the lateral dimension. We consider a Tungsten (W) based spin-Hall metal (SHM) with a 2.2nm thickness where the maximum spin generation takes place as shown in Fig. 3. In order to conduct a circuit-level evaluation, a SPICE-compatible SHE-MTJ model was implemented by incorporating the spin current from SHM into LLG equation as shown in Fig. 4 [3]. Then, we use the model to simulate the read and write circuitry in Fig. 6. Here, 22nm high performance (HP) CMOS transistors from a publically available predictive technology model (PTM) were used for the circuit simulation [4]. We constructed a realistic memory macro including bi-directional write current drivers and dual-voltage WL drivers to ensure a robust write operation. To maximize the read sensing margin, we adopted dummy MTJ cells for generating a reference current corresponding to the average value of the parallel and anti-parallel stage currents (i.e. $I_{Ref} = (I_{AP}+I_P)/2$) [5][6]. The read current is applied in the antiparallel to parallel direction to minimize read disturbance issues [7]. For estimating the bit-cell area, a FinFET based layout is considered as shown in Fig. 6 [8]. Here, we use 2 fins for read and write transistors which makes the cell area of SHE-MRAM comparable to that of a standard STT-MRAM cell. Compared to an SRAM cell, both SHE-MRAM and STT-MRAM are roughly 3x denser.

Using the proposed simulation setup, we first compare the performance of a single 256Kbit subarray in Table 2 which shows that SHE-MRAM has a 4.7x shorter write time and 1.3x shorter read delay as compared to a standard STT-MRAM with the same cell size. These results indicate that SHE-MRAM will always outperform STT-MRAM regardless of the cache size. Next we evaluate SHE-MRAM for replacing SRAM in large caches. For L3 or L4 cache, we can expect that SHE-MRAM, which has the same cell size as STT-MRAM, to have a shorter access latency than SRAM since the access time of these larges caches is dominated by the global interconnect delay rather than the single subarray delay [9]. So the more interesting question was whether SHE-MRAM can outperform SRAM for smaller L2 caches with densities in the order of 1Mbit. Standard STT-MRAM could not outperform SRAM for smaller caches due to the long write delay, but the faster write coupled with the shorter global interconnect delay could potentially make SHE-MRAM a viable option for L2. One unique advantage of SHE-MRAM is that the read delay can be reduced without a commensurate increase in write delay by simply increasing the thermal stability Δ as shown in Figs. 7 and 8. This is contrary to standard STT-MRAM which has an inherent conflict between read and write delays. Therefore, we assumed an SHE-MRAM with a higher thermal stability compared to STT-MRAM in our analysis. We used CACTI, a widely accepted architecture simulator, for extracting the power and performance numbers [10]. As shown in Table 3, SHE-MRAM based L2 cache with a higher Δ has a read latency comparable to that of SRAM but with a lower leakage power and denser area. It should be noted however that a higher TMR and efficient sensing circuits are necessary to reduce the high read energy incurred by the current-forcing read of SHE-MRAM.

References: [1] L. Q. Liu, *Science*, 2012. [2] R. Takemura, *JSSC*, 2010. [3] Y. Kim, *EDL*, 2013. [4] W. Zhao, *TED*, 2006. [5] G. D. Sandre, *ISSCC*, 2010. [6] D. Gogl, *JSSC*, 2005. [7] T. Kawahara, *JSSC*, 2008. [8] A. Shafaei, *ICCD*, 2014. [9] K. C. Chun, *JSSC*, 2013. [10] N. Muralimanohar, *HP Lap. Tech. Rep.* HPL-2009-85, 2009.

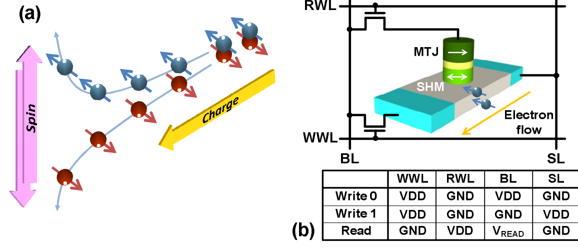


Fig. 1. (a) Transverse spin current generation by SHE. (b) SHE-MRAM cell configuration.

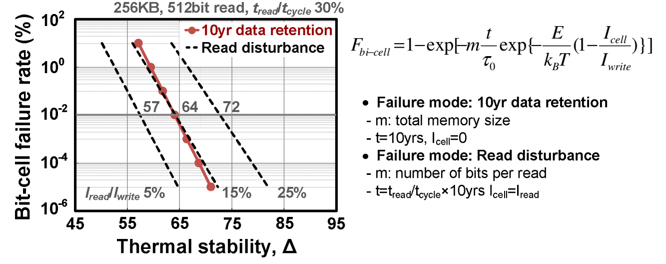


Fig. 2. Thermal stability requirement estimation considering 10 year data retention and read disturbance.

Parameters	STT-MRAM	SHE-MRAM
MTJ type	Interface perpendicular	In-plane
Thermal stability factor	65	65
Free layer material	CoFeB	CoFeB
Free layer dimensions, $W_F \times L_F \times t_F$	40nm×40nm×1.34nm	22nm×77nm×2.7nm
Saturation magnetization, M_s	1.077×10^3A/m	1.077×10^3A/m
Damping factor, α	0.018	0.006
Polarization factor, P	0.63	0.63
Critical thickness, t_c	1.5nm	-
TMR	130%	130%
RA ($\Omega\cdot\mu m^2$)	5	5.5
SHM dimensions, $W_{SHM} \times L_{SHM} \times t_{SHM}$	-	77nm×44nm×2.2nm
SHM spin diffusion length, λ_{ch}	-	1.5nm
SHM resistivity, ρ_{ch}	-	200μΩ·cm² (for W)
Spin Hall angle, θ_{SH}	-	0.3 (for W)

* Δ and material parameters are extracted based on 85°C.
* RA of SHE-MRAM includes R_{SHM}. Thickness dependency of α is also considered.

Table 1. Material parameters and device dimensions used for STT- and SHE-MRAM.

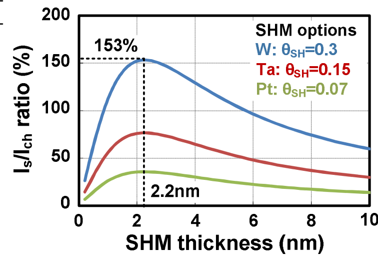


Fig. 3. Spin hall metal thickness versus spin current generation.

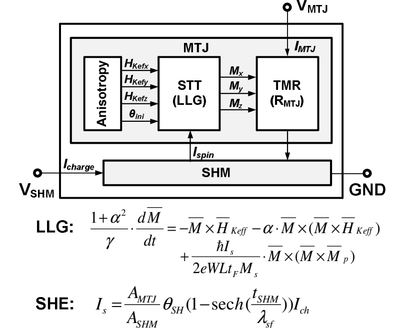


Fig. 4. SPICE simulation framework.

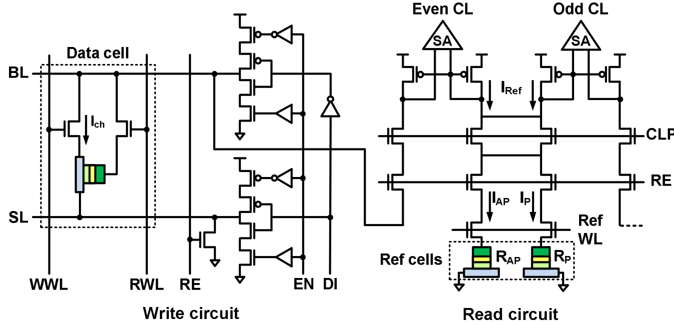


Fig. 5. Read and write circuitry for SHE-MRAM.

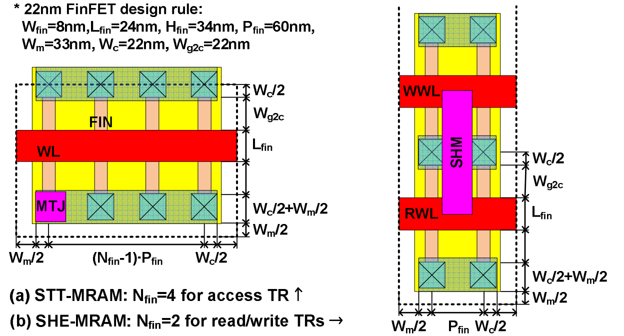
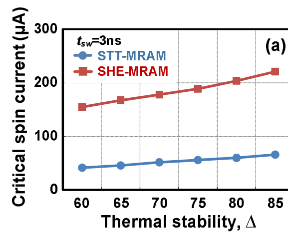


Fig 6. 22nm FinFET-based layout (3x denser than SRAM).

Metrics	STT-MRAM	SHE-MRAM
I_{spin} at t_{sw}=3ns (μA)	73	134
Read delay (ns)	0.42	0.33
Read energy (fJ)	4.5	7.1
Write delay (ns)	6.6	1.4
Write energy (fJ)	720	208
Area (μm²)	0.029	0.029

Table 2. Single 256Kbit sub-array performance (Δ=65 @ 85°C, 512 cells/BL, 512 cells/WL).



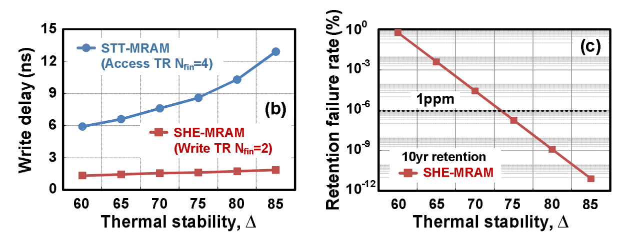


Fig. 7. Impact of thermal stability factor Δ on (a) critical spin current, (b) write delay, and (c) retention failure rate.

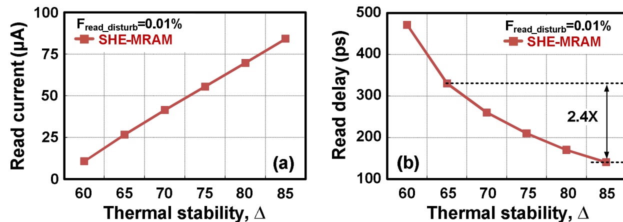


Fig. 8. Impact of thermal stability factor Δ on (a) read current and (b) read delay of SHE-MRAM for a 0.01% read disturbance failure rate.

Metrics	SRAM	STT-MRAM	SHE-MRAM
Thermal stability (@85°C)	-	65	85
Bit-cell failure rate (%)	-	10^{-2}	10^{-11}
Read latency (ns)	0.42	0.71	0.43
Read energy (nJ)	0.07	0.22	0.44
Write latency (ns)	0.42	6.77	1.95
Write energy (nJ)	0.10	0.41	0.21
Leakage power (mW)	39.5	4.96	4.96
Area (mm²)	0.55	0.16	0.16

Table 3. L2 cache performance summary (1Mbit, 8-way associativity, private bank, CACTI simulator [10]).