Exploring potentials of perpendicular magnetic anisotropy STT-MRAM for cache design

Xiaolong Zhang, Yuanqing Cheng, Weisheng Zhao, Youguang Zhang, Aida Todri-Sanial

EXPLORING POTENTIALS OF PERPENDICULAR MAGNETIC ANISOTROPY STT-MRAM FOR CACHE DESIGN

Xiaolong Zhang*, Yuanqing Cheng*, Weisheng Zhao*†, Youguang Zhang* and Aida Todri-Sanial‡
*School of Electronic and Information Engineering, Beihang University, Beijing, China 100191
†IEF, Université Paris-Sud / CNRS, Orsay, 91405, France
‡LIRMM – Université Montpellier 2 / CNRS, Montpellier, 34095, France
Corresponding Author's Email: yuanqing@ieee.org

***Abstract*—Traditional CMOS integrated circuits suffer from elevated power consumption as technology node advances. A few emerging technologies are proposed to deal with this issue. Among them, STT-MRAM is one of the most important candidates for future on-chip cache design. However, most STT-MRAM based architecture level evaluations focus on in-plane magnetic anisotropy effect. In the paper, we evaluate the most advanced perpendicular magnetic anisotropy (PMA) STT-MRAM for on-chip cache design in terms of performance, area and power consumption perspectively. The experimental results show that PMA STT-MRAM has higher power efficiency compared to SRAM as well as desirable scalability with technology node shrinking.**

I. Introduction

As the technology node continuously shrinks, CMOS based VLSI circuits suffer from severe static power consumption challenge due to the aggravating sub-threshold leakage. To tackle this problem, some non-volatile technologies, such as STT-MRAM [1], PCRAM [2] and ReRAM [3], have emerged to reduce leakage power consumption. Among them, STT-MRAM (Spin Transfer Torque MRAM) is one of the most promising candidates for on-chip cache design because of its fast access speed, near zero leakage and unlimited read/write endurance [4].

STT-MRAM stores data by magnetic tunneling junction (MTJ). When spinpolarized current passes through MTJ, the magnetic direction of the free layer can be switched to parallel or antiparallel to that of the fixed layer resulting different MTJ resistances. Depending on the low/high resistance status of MTJ, a '0' or '1' can be recorded. Above 90nm technology node, in-plane magnetic anisotropy (i.e., magnetic directions of both free layer and fixed layer are within MTJ surface) dominates MTJ switching. Faber et al. proposed a compact model of in-plane STT-MRAM, which can capture its electrical behavior accurately [5]. Based on such compact models, many STT-MRAM based on-chip cache organizations are proposed to exploit its low power, non-volatility and fast read speed from the architecture perspective [6].

However, when the MTJ feature size scales down to 40nm, perpendicular magnetic anisotropy (PMA) dominates MTJ switching and manifests distinct electrical characteristics. PMA STT-MRAM has smaller switching current, regular shape (circle instead of ellipse) and consumes lower power. Therefore, it has desirable scalability with technology node. Unfortunately, there are few investigations focusing on perpendicular magnetic anisotropy STT-MRAM, especially at the architecture level due to the lack of low level device model capable of describing PMA MTJ switching mechanism accurately until recently. In the paper, we evaluate the effectiveness of PMA STT-MRAM as the on-chip cache based on the PMA STT-MRAM compact model proposed by Zhang et al. [7]. Our main contributions are as follows,

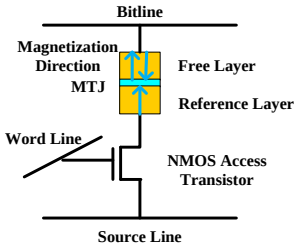

Fig. 1. PMA STT-MRAM Cell Structure

- To the best of our knowledge, this is the first work to evaluate PMA STT-MRAM effectiveness for on-chip cache design;
- Extensive simulations are performed to compare PMA STT-MRAM with SRAM technologies in terms of area, power and performance;
- Design space suitable for PMA STT-MRAM cache design are explored to instruct designer to take its full advantages.

The rest of the paper is organized as follows. Section II introduces PMA STT-MRAM briefly as well as the 40nm compact PMA STT-MRAM model used in this paper. Section III presents the architecture level evaluation method and several important considerations for parameter settings. Experimental results and analysis are shown in Section IV. Section V concludes the paper.

II. PMA STT-MRAM Compact Model

Instead of storing data by charge, STT-MRAM store information by electron's spin property. The core component

TABLE I. PARAMETERS INVOLVED IN THE PAPER

Parameters already defined in PMA MTJ compact model		
Symbol	Definition	Value
ϕ	Potential barrier height of MgO	0.4
H_k	Anisotropy field	$113.0 \times 10^3 A/m$
M_s	Saturation magnetization	$456.0 \times 10^3 A/m$
t_{ox}	Oxide barrier height	$0.85nm$
μ_0	Permeability in free space	$1.2566 \times 10^{-6} H/m$
μ_B	Bohr magneton	$9.274 \times 10^{-24} J/T$
γ	Gyromagnetic ratio	$1.76 \times 10^7 rad/(s \cdot T)$
g	Spin polarization efficiency factor	Depending on technology node
e	Electron charge	$1.6 \times 10^{-19} C$
α	Magnetic damping constant	0.027
F	Material dependent constant used for R_P calculation	664
$R \cdot A$	MTJ resistance area product	$5\Omega \cdot \mu m^2$

of STT-MRAM is It consists of two ferromagnetic layers (usually using CoFeB), and one insulating layer (usually made by MgO), which is sandwiched between ferromagnetic layers as shown in Fig.1. Bottom ferromagnetic layer pinned to a fixed magnetization direction is called reference layer while the magnetization direction of top layer, which is called free layer, can be switched by passing through spin-polarized current. When magnetization directions of reference layer and free layer are the same, MTJ manifests low resistance. Otherwise, it has high resistance. The information stored in MTJ can be sensed or changed through a CMOS access transistor, which forms the commonly referred 1T-1MTJ structure. PMA STT-MRAM has the magnetization direction perpendicular to MTJ surface instead of within the surface plane. The shape of MTJ for PMA STT-MRAM is circular whereas that of in-plane STT-MRAM is elliptical.

Our work is based on a compact model characterizing sub-50nm MTJ electromagnetic behaviors [7]. In this region, enhanced thermal stability can be calculated by the following formula,

$$E = \frac{\mu_0 M_s \times V \times H_k}{2} \tag{1}$$

The intrinsic critical current can be derived by,

$$I_{c0} = \alpha \frac{\gamma e}{\mu_B g} (\mu_0 M_s) H_K V \tag{2}$$

See Table I for the symbol meanings of above two formulas. Comparing with experimental results, the compact model can capture PMA MTJ electrical behavior accurately according to [7]. Our following exploration of PMA STT-MRAM cache design will base on this compact model.

III. EVALUATION METHOD & CRITICAL PARAMETER SETTINGS

In the paper, we use NVSim to explore the PMA STT-MRAM based cache design space [8]. NVSim is an architecture level simulator for emerging non-volatile memory design optimizations. It requires some technology node and memory cell structure parameters as inputs, and outputs optimal cache/memory organization for a specified optimization goal, such as performance, area and power consumption. Important parameters used for simulating PMA STT-MRAM device based on-chip cache design are derived as follows.

MTJ resistance If two ferromeganetic layers have the same magnetization direction, MTJ is in low resistance state.

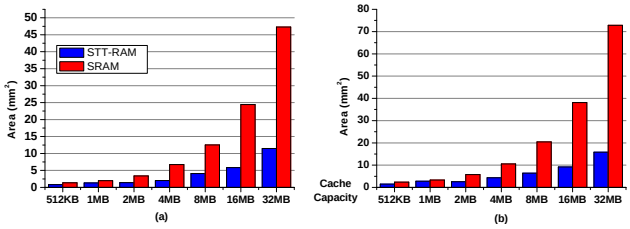

Fig. 2. Area comparisons of PMA STT-MRAM and SRAM with different technology nodes: (a) 32nm (b) 40nm

The resistance can be calculated by [7],

$$R_P = \frac{t_{ox}}{F \times \phi^{-\frac{1}{2}} \times A} \times e^{1.025 \times t_{ox} \times \phi^{-\frac{1}{2}}} \tag{3}$$

Please refer Table I for symbol meanings and values in the formula.

We assume Tunnel MagnetoResistance (TMR) value to be 120% according to [7]. The antiparallel resistance can be derived from TMR and R_P. The insulator layer thickness is assumed to be 0.85nm as shown in Table I.

NMOS transistor aspect ratio To determine NMOS access transistor's aspect ratio, we construct a PMA STT-MRAM cell using Cadence Virtuoso and our compact device model. Then, we use Cadence Spectre for circuit simulation to derive aspect ratio of NMOS transistor with enough driving ability for MTJ switching. The simulation results show that W/L should be larger than 2 for fast and reliable switching. Therefore, we choose $W/L = 3$ in the following evaluation.

Write energy of a single PMA STT-MRAM cell The supply voltage of PMA STT-MRAM cell is set to be 1.2V to ensure MTJ switching. Then, the MTJ write power and NMOS access power can be calculated by NMOS V_{ds} and MTJ I_{c0}. As a result, the write energy of a single cell is 0.51pJ at 40 nm technology node and 0.15pJ at 32nm technology node respectively.

IV. EXPERIMENTAL RESULTS

A. Area

We compare cache area of PMA STT-MRAM and SRAM with different capacities for 32nm and 40nm technology nodes. The results are plotted in Fig.2. As shown in the figure, PMA STT-MRAM occupies much smaller area compared to SRAM with the same capacity. The reason is that a STT-MRAM cell size is $12F^2$ while a SRAM cell size is $146F^2$(F is the feature size of fabrication process). However, since peripheral circuitry also occupies some area, the area ratio between STT-MRAM and SRAM is smaller than $146/12$. As the cache capacity increases, PMA STT-MRAM based cache can save more area compared to SRAM. For example, at 40nm technology node, STT-MRAM can save nearly 80% area compared with SRAM cache indicating high integration density advantage brought by PMA STT-MRAM.

B. Power Consumption

As for in-plane STT-MRAM, write energy dominates the dynamic power consumption and is usually higher than that

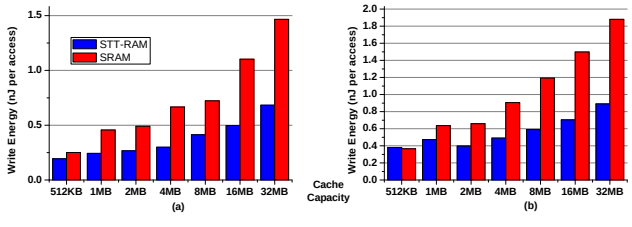

Fig. 3. Write dynamic power comparisons between PMA STT-MRAM and SRAM under (a) 32nm (b) 40nm technology nodes respectively

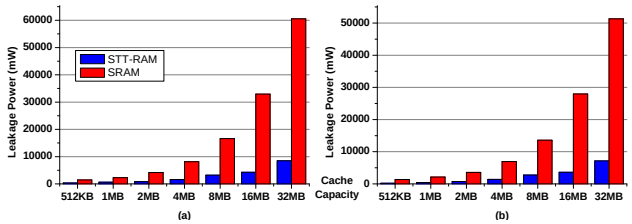

Fig. 4. Leakage power comparisons between PMA STT-MRAM and SRAM under (a) 32nm (b) 40nm technology nodes respectively

of SRAM. In contrast, PMA STT-MRAM shows a significant advantage in term of write power as shown in Fig.3. Although the power for writing a single STT-MRAM cell is higher than that for a single SRAM cell, PMA STT-MRAM has much smaller area thus much lower H-Tree interconnect power. Moveover, PMA STT-MRAM write latency for a single cell is comparable to SRAM. Therefore, PMA STT-SRAM shows overall higher power efficiency than SRAM, and a desirable scalability with technology node shrinking. The static power comparison is shown in Fig. 4. Thanks to non-volatile property, PMA STT-MRAM has negligible leakage power compared to SRAM similar to its in-plane counterpart. As the technology node shrinks and capacity increases, the SRAM leakage power rises dramatically compared to PMA STT-MRAM, which indicates huge power benefits of PMA STT-MRAM as on-chip cache.

C. Access Latency

The read latency comparisons at 32nm and 40nm technology nodes are shown in Fig.5. It shows that SRAM read access latency is smaller than PMA STT-MRAM when the cache capacity is small (e.g., smaller than 4MB). As the capacity increases, PMA STT-MRAM shows lower latency due to the reducing H-Tree interconnect delay. The write latency of PMA STT-MRAM is larger than SRAM as shown in Fig. 6. However, write path usually doesn't lie in the critical path

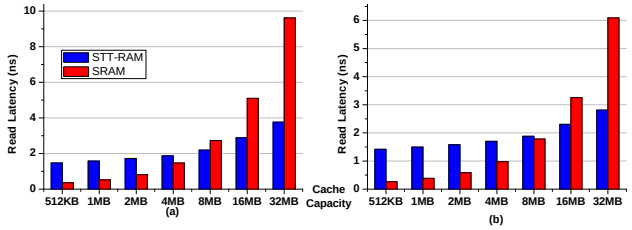

Fig. 5. Read latency comparisons between PMA STT-MRAM and SRAM under (a) 32nm (b) 40nm technology nodes respectively

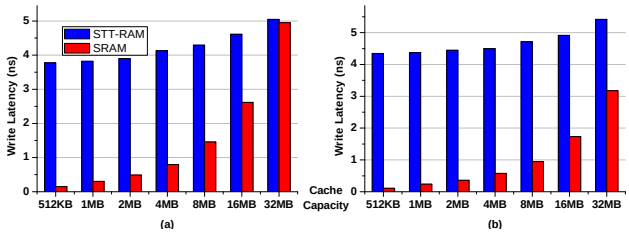

Fig. 6. Write latency comparisons between PMA STT-MRAM and SRAM under (a) 32nm (b) 40nm technology nodes respectively

thereby will not impact cache performance significantly.

In summary, although PMA STT-MRAM has relative larger write latency, it outperforms SRAM in other metrics such as area overhead, power consumption and read latency. Therefore, it is advantageous to use PMA STT-MRAM for future on-chip cache design.

V. Conclusions

As integration density sharply increases with technology node shrinking, CMOS technology suffers from severe power consumption. To overcome this problem, some emerging technologies are proposed including PCRAM, ReRAM and STT-MRAM. STT-MRAM is a competitive candidate to be used for cache design. However, existing research focuses on in-plane STT-MRAM. When MTJ feature size scales down to 40nm and below, perpendicular magnetic anisotropy effect becomes dominant and requires new compact model as well as architecture level evaluations based on it. In this paper, we evaluate benefits brought by PMA STT-MRAM compared to SRAM from power, performance and area perspectives with the aid of a PMA STT-MRAM compact model. The simulation results show that PMA STT-MRAM can achieve better performance, lower power consumption and smaller area overhead and pave the way for next generation cache design.

References

[1] Claude Chappert et al., "The emergence of spin electronics in data storage," *Nature Materials*, vol. 6, no. 11, pp. 813–823, 2007.

[2] S. Raoux et al., "Phase-change random access memory: A scalable technology," *IBM Journal of Research and Development*, vol. 52, no. 4.5, pp. 465–479, 2008.

[3] J. Joshua Yang et al., "Memristive devices for computing," *Nature Nanotechnology*, vol. 8, no. 1, pp. 13–24, 2012.

[4] W. Kang et al., "An overview of spin-based integrated circuits," in *Proceedings of 19th Asia and South Pacific Design Automation Conference (ASP-DAC)*, Singapore, Jan. 20-24 2014, pp. 676–683.

[5] L. B. Faber et al., "Dynamic compact model of spin-transfer torque based magnetic tunnel junction (MTJ)," in *Proceedings of IEEE 4th International Conference on Design & Technology of Integrated Systems in Nanoscal Era*, Cairo, 6-9 April 2009, pp. 130–135.

[6] G. Sun et al., "A novel architecture of the 3d stacked mram l2 cache for CMPs," in *Proceedings of IEEE International Symposium on High Performance Computer Architecture*, Raleigh, USA, 2009, pp. 239–249.

[7] Y. Zhang et al., "Compact modeling of perpendicular-anisotropy CoFeB/MgO magnetic tunnel junctions," *IEEE Transactions on Electron Devices*, vol. 59, no. 3, pp. 819–826, 2012.

[8] X. Dong et al., "Nvsim: A circuit-level performance, energy, and area model for emerging nonvolatile memory," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 31, no. 7, pp. 994–1007, 2012.