# Computing-in-Memory Architecture based on Field-Free SOT-MRAM with Self-Reference Method

Chao Wang<sup>1,2</sup>, Zhaohao Wang<sup>1,3,4</sup>, Yansong Xu<sup>1,2</sup>, Jianlei Yang<sup>4,5</sup>, Youguang Zhang<sup>1,2</sup>, and Weisheng Zhao<sup>1,3,4</sup>

<sup>1</sup>Fert Beijing Research Institute, Beihang University

<sup>2</sup>School of Electronics and Information Engineering, Beihang University

<sup>3</sup>School of Microelectronics, Beihang University

<sup>4</sup>Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University

<sup>5</sup>School of Computer Science and Engineering, Beihang University

Beijing 100191, China. Email: zhaohao.wang@buaa.edu.cn

Abstract—On the current computing platforms, the memory wall between processor and memory has become the toughest challenge for the traditional Von-Neumann computer architecture. Computing-in-Memory (CIM) is taken as a promising approach to solving the above bottleneck in computing systems. In this paper, we propose a CIM platform with field-free spinorbit torque magnetic random access memory (SOT-MRAM). The self-reference (SelfRef) method is designed to enhance the read reliability and directly obtain logic results through memory-like read operations without adding logic cells. Memory read/write and logic operations, including NOT, AND/NAND and OR/NOR, can be implemented in the same SOT-MRAM chip. The speed and power penalties caused by SelfRef scheme are acceptable thanks to the ultrafast switching of the SOT. The read reliability and logic correctness of the proposed CIM are demonstrated by hybrid simulation on a 40 nm technology node.

# I. INTRODUCTION

As the CMOS technology node is shrinking continuously, increasing leakage current is becoming a major miniaturization bottleneck in CMOS memories [1]. Moreover, the conventional Von-Neumann architecture, with memory and computing units separated and interconnected via buses, is faced with huge challenges, including the limitation of data transfer bandwidth and the ever-increasing speed gap between the processor and memory [2].

Recently, emerging non-volatile (NV) memories have been carried out to design Computing-in-Memory (CIM) platforms which provide the possibility of reducing the data movement overhead and static power consumption [3]. Among them, spin-transfer torque magnetic random access memory (STT-MRAM) has been considered one of the most promising candidates due to its long retention, nearly infinite endurance and great CMOS compatibility [4]. Therefore, STT-MRAM has received extensive attention in both academia and industry [5]. However, STT-MRAM has serious disadvantages in write speed, power consumption and reliability due to physical shortcomings. To overcome the STT write bottlenecks, the spin-orbit torque (SOT) has been investigated to offer a superior write approach with short write latency and low write power consumption [6]. Nonetheless, traditional SOT is unable to induce deterministic data writing without an assisted STT current or an external field. Recently, a new field-free SOT

mechanism was demonstrated to implement ultrafast write operations independently [7], [8]. At the architecture level, a number of CIM platforms based on STT-MRAM and SOT-MRAM have been proposed [9]–[13]. By implementing computation and storage operations simultaneously or providing the processor with pre-computed results, the CIM platforms can reduce the frequency of memory access and data transfer. Most of the current MRAM based CIM designs implement Boolean logic by modifying the peripheral circuits of the memory to combine cells and adjust the reference circuit.

In this paper, we propose a CIM implementation based on field-free SOT-MRAM. The characteristics of field-free SOT allow us to achieve high-speed write operations and highreliability read operations through a self-reference (SelfRef) scheme. By connecting three cells in parallel and then utilizing the SelfRef read method, the logical operation can be directly obtained. Besides, on the proposed CIM platform, memory read/write and logic NOT, AND/NAND and OR/NOR can be implemented by using basically identical operations. Such logic implementation with operator stored in the cell is difficult to be achieved for conventional STT-MRAM and SOT-MRAM without complementary structures [10], [11].

The rest of this paper is organized as follows. Section II briefly introduces the preliminaries of field-free SOT-MRAM. In Section III, we present the SelfRef scheme and the CIM paradigm. The simulation results are given in Section IV. Finally, Section V concludes this paper.

# II. FIELD-FREE SOT-MRAM

As shown in Fig. 1(a), the basic storage element of mainstream SOT-MRAM is a perpendicular magnetic anisotropy magnetic tunnel junction (p-MTJ) composed of two ferromagnetic layers, i.e., free layer (FL) and pinned layer (PL), sandwiching an oxide tunnel barrier [4]. Each MTJ can exhibit two stable resistivity states which represent data '0' and '1', respectively. The resistance difference between P and AP states affects the read performance, as measured by the tunnel magnetoresistance (*TMR*) ratio, which can be calculated by  $TMR = (R_{AP} - R_P)/R_P$  [14].

Fig. 1(b) shows the device of SOT-MRAM with the heavy metal (HM) strip under the FL. The current flowing through



Fig. 1. (a) Sandwich structure and resistance states of p-MTJ. (b) Structure of three-terminal SOT-MRAM. (c) Simulation results of time-dependent z-component magnetization  $(m_z)$  for the field-free SOT mechanism.

the HM can generate an SOT for magnetization switching. SOT-induced magnetization dynamics can be described by an extended Landau-Lifshitz-Gilbert (LLG) equation:

$$\frac{\partial \vec{m}}{\partial t} = -\gamma \mu_0 \vec{m} \times \vec{H}_{\text{eff}} + \alpha \vec{m} \times \frac{\partial \vec{m}}{\partial t} + \vec{\tau}_{\text{DL}} + \vec{\tau}_{\text{FL}} 
\vec{\tau}_{\text{DL}} = -\lambda_{\text{DL}} J_{\text{SOT}} \xi \vec{m} \times (\vec{m} \times \vec{\sigma}) 
\vec{\tau}_{\text{FL}} = -\lambda_{\text{FL}} J_{\text{SOT}} \xi \vec{m} \times \vec{\sigma}$$
(1)

where  $\vec{m}$  and  $\vec{\sigma}$  are the unit vectors of FL magnetization and SOT-induced spin polarization, respectively.  $\vec{H}_{\rm eff}$  is the sum of effective magnetic fields,  $\gamma$  is the gyromagnetic ratio,  $\mu_0$  is the vacuum permeability, and  $\alpha$  is the Gilbert damping constant.  $\lambda_{\rm DL}$  and  $\lambda_{\rm FL}$  represent the strengths of the damping-like and field-like torques, respectively.  $J_{\rm SOT}$  is the SOT current density, and  $\xi$  is a device-dependent parameter.

Since the effective torques of SOT are orthogonal to the anisotropy axis ( $\vec{m} \times \vec{\sigma} \gg 0$ ) for p-MTJ, the SOT mechanism can eliminate the incubation delay of the STT to achieve ultrafast switching, but cannot implement deterministic switching independently. A new SOT mechanism dominated by stronger field-like torque was recently demonstrated to implement the ultrafast field-free switching of MTJ [15]. Take  $\lambda_{\rm FL}/\lambda_{\rm DL} = 3$  for example, the magnetization of FL is always switched to the opposite state at sub-nanosecond speed once an SOT current with appropriate amplitude is applied as shown in Fig. 1(c). It should be noted that the SOT write operation is independent of the current polarity and the write operation inevitably changes the data. Therefore, the actual implementation of the associated SOT-MRAM requires a read-before-write operation [7].

## III. FIELD-FREE SOT-MRAM BASED CIM

#### A. Read and write

In this paper, we propose a CIM platform based on the above-mentioned field-free SOT-MRAM and a novel read scheme. The detailed read circuits are illustrated in Fig. 2, two sampling capacitors  $C_0$  and  $C_1$  are used to store the sensing voltage signals under the control of switches  $S_0$  and  $S_1$ . The read operations are described as follows:



Fig. 2. (a) Read/write schemes for the proposed CIM platform. (b) Schematic of the used sense amplifier.

**Step-1:** The RWL and WWL are activated to select a bit cell and then the reading current  $I_{\text{read}}$  generated from RBL flows through the selected cell, producing a corresponding voltage signal. At the same time,  $S_0$  is turned on so that  $C_0$  can sample and store the voltage signal, which is treated as  $V_{\text{data}}$ , then  $S_0$  and RWL are turned off in sequence.

**Step-2:** The WWL remains active and then a unidirectional write current pulse  $I_{\text{write}}$  generated from WBL is applied to the selected cell. Under the effect of field-free SOT, the resistance of MTJ transfers from  $R_{\text{P}}$  to  $R_{\text{AP}}$  or from  $R_{\text{AP}}$  to  $R_{\text{P}}$ .

**Step-3:** A similar approach as step-1 is performed, and the difference is that  $S_1$  is turned on so that the voltage signal can be sampled and stored by  $C_1$ , which is recorded as  $V_{\text{data}}$ #. QE is set high to release the pull-up voltage which will be used in the next step.

**Step-4:** DE is activated and then the  $V_{data}$  stored in  $C_0$  and  $V_{data}$ # in  $C_1$  are synchronized to the sensing amplifier (SA). Then DE is turned off while SAE is activated so that data in MTJ can be read out in OUT by comparing  $V_{data}$  with  $V_{data}$ # while OUT# can be considered as NOT logic.

**Step-5:** The data stored in MTJ has been changed in step-2, thus we have to repeat step-2 in order to rewrite the MTJ back to its original state.

Due to the destructive write mechanism, the field-free SOT-MRAM requires a read-before-write operation to determine whether a subsequent write operation is necessary. In the proposed structure, the write operation does not need to be performed after completing a full five-steps read operation. If the read data in step-4 is the opposite of write data, the writing process will end directly, otherwise step-5 will continue. Therefore, the write operation can be completed in four or five steps.

#### B. AND and OR

The core method of performing bitwise logic operations in SOT-MRAM is to generate and differentiate multiple resistors combinations. As shown in Fig. 3, the proposed CIM platform extends the states of the resistors through connecting three resistors in parallel similar to [10], [11], one of which is used to determine the type of operation. When performing bitwise logic operations between the second and third rows, the corresponding  $R(W)WL_{[0]}$ ,  $R(W)WL_{[1]}$  and  $R(W)WL_{[2]}$ 



Fig. 3. Schematic of the proposed CIM paradigm for bitwise logic operations.

 TABLE I

 TRUTH TABLE FOR THE PROPOSED CIM IMPLEMENTATION

Logic	Cell <sub>0</sub>	Cell <sub>1</sub>	Cell <sub>2</sub>	OUT
AND	$0 (P \leftrightarrow AP)$	$0 (P \leftrightarrow AP)$	$0 (P \leftrightarrow AP)$	0
	$0 (P \leftrightarrow AP)$	$0 (P \leftrightarrow AP)$	1 (AP $\leftrightarrow$ P)	0
	$0 (P \leftrightarrow AP)$	1 (AP $\leftrightarrow$ P)	$0 (P \leftrightarrow AP)$	0
	$0 (P \leftrightarrow AP)$	$1 (AP \leftrightarrow P)$	$1 (AP \leftrightarrow P)$	1
OR	1 (AP $\leftrightarrow$ P)	$0 (P \leftrightarrow AP)$	$0 (P \leftrightarrow AP)$	0
	1 (AP $\leftrightarrow$ P)	$0 (P \leftrightarrow AP)$	1 (AP $\leftrightarrow$ P)	1
	1 (AP $\leftrightarrow$ P)	1 (AP $\leftrightarrow$ P)	$0 (P \leftrightarrow AP)$	1
	$1 (AP \leftrightarrow P)$	$1 (AP \leftrightarrow P)$	$1 (AP \leftrightarrow P)$	1

should be activated for parallelizing Cell<sub>0</sub>, Cell<sub>1</sub> and Cell<sub>2</sub>, then the logic operations can be implemented by comparing  $R_{\text{Cell0}} || R_{\text{Cell1}} || R_{\text{Cell2}}$  and  $R_{\text{Cell0}} # || R_{\text{Cell1}} # || R_{\text{Cell2}} #$  using the above read scheme. Specifically, as shown in Table I, if the data in Cell<sub>0</sub> is '1', OUT is the bitwise OR of the data in Cell<sub>1</sub> and Cell<sub>2</sub>. Conversely, if the data in Cell<sub>0</sub> is '0', OUT is the bitwise AND of the data in Cell<sub>1</sub> and Cell<sub>2</sub>. OUT# can be considered as NOR/NAND logic as above.

In this paper, decoders for activating one R(W)WL in access operations are enhanced to activate three R(W)WLs simultaneously during logic operations. In addition, the read and write drivers have also been improved to provide a large enough current to switch three MTJs in step-2 and 5, as well as to obtain an effective read voltage in parallel during step-1 and 3. It can be seen that the proposed CIM can realize switchable memory and logic modes by only modifying the peripheral circuits. Read/write and logic operations are performed using identical operation steps and latency, which facilitates the top-level timing optimization.

# IV. SIMULATION

To validate the effectiveness of the proposed CIM platform, we performed extensive simulations using the Cadence Specter with a 40 nm CMOS design kit and a PMA SOT-MTJ compact

TABLE II DEVICE PARAMETERS USED IN OUR SIMULATIONS

Parameter	Value 40 nm×40 nm×π/4 40 nm×60 nm×2 nm			
MTJ area				
HM dimension				
Saturation magnetization	$1 \times 10^{6}$ A/m			
Effective anisotropy	$1.33 \times 10^5$ A/m			
Spin Hall angle	0.3			
Gilbert damping constant	0.1			
Thickness of MgO Barrier $(t_{OX})$	0.85 nm			
TMR	120%			
Thickness of FL $(t_{\rm FL})$	1 nm			
Resistance-area product	$10 \ \Omega \cdot \mu m^2$			
Resistivity of HM	200 $\mu\Omega$ ·cm			
0 Iread Iwrite 0 Steady S	State Transition state			
	QE DE SAE			
Vdata	ata#			
1 0 1 0				
0 05 1 15	2 25 3			

Fig. 4. Transient simulation waveforms of the SelfRef read operation.

model [16]. Table II lists the key simulation parameters of the MTJ and we consider the *TMR*,  $t_{\rm FL}$  and  $t_{\rm OX}$  with  $3\sigma$  and 1% variations for verifying the reliability of the read scheme.

Fig. 4 shows the transient simulation waveforms of the Self-Ref read operation. Thanks to the field-free-SOT writing, each step is achieved within sub-nanoseconds and write current is also greatly reduced. Fig. 5 shows the resistance margins (RMs) of the proposed scheme compared to the SOT-MRAM with  $R_{\text{REF}} = (R_{\text{AP}} + R_{\text{P}})/2$  as the reference (HalfRef) and complementary reference (ComRef) [6]. The ideal RM of our design is almost twice as large as that of HalfRef and 1000 Monte Carlo simulations performed on the above three read schemes illustrate that ComRef and SelfRef have no error while the error rate of HalfRef is 5.1%. Since the deviation direction for  $R_{\rm P}$  and  $R_{\rm AP}$  in SelfRef scheme is consistent, even the RM in extreme conditions is close to the ideal. Therefore, the proposed SelfRef can reveal higher reliability than ComRef under more serious process variation. Table III demonstrates more access performance results compared to mainstream cell structures. The proposed design achieves high read reliability at the cost of acceptable latency and power consumption without increasing the area.

Fig. 6 describes the transient simulation waveforms of the proposed CIM implementation and initial data in  $\text{Cell}_{0,1,2}$  are all set to '0'. Fig. 6(a) and (b) show that  $\text{Cell}_{1,2}$  are constantly written different data to verify all logic operation functions. As shown in Fig. 6(c) and (d), the data in  $\text{Cell}_0$  is written as '0', and all bitwise AND/NAND operations are performed.



Fig. 5. The resistance margins (RMs) of  $R_{\rm P}$ ,  $R_{\rm REF}$  and  $R_{\rm AP}$ .

TABLE III Access performance at different structures.

MRAM tpye Read scheme <sup>a</sup>	STT HalfRef	STT ComRef	SOT HalfRef	SOT ComRef	Proposed SelfRef
Unit area <sup>b</sup>	1T	2T	2T	4T	2T
Write latency (ns)	4.94	4.94	0.55	0.55	2.5
Write energy (fJ)	503.1	1006	34.18	67.68	48.97
Read latency (ns)	1	1	1	1	2.5
Read energy (fJ)	20.72	19.13	20.69	19.92	52.73
Error rate	4.7%	0%	5.1%	0%	0%

<sup>a</sup>Read performance simulations are performed by the same read circuit. <sup>b</sup>T represents transistor, and *n*T means *n* transistor(s) per bit.



Fig. 6. Transient waveforms of the logic operations performed on the proposed CIM platform.

Fig. 6(e) and (f) depicts all bitwise OR/NOR operations when the data in Cell<sub>0</sub> is written as '1'. The output results are consistent with Table I, which verifies the correctness of the proposed CIM. Compared with the ComRef STT-MRAM based CIM [10], [11], our design has an advantage on write performance without using more transistors.

# V. CONCLUSION

In this paper, we present a CIM platform using field-free SOT mechanism and SelfRef read scheme. Compared to the HalfRef and ComRef solutions, our novel read circuit is less affected by process variations. The proposed CIM platform performs bitwise logic operations by utilizing peripheral circuitry embedded in memory, without adding any processing units. Circuit-level simulation results show that the proposed read scheme improves the read reliability at an acceptable cost, and the proposed CIM platform implements fast logic operations, including NOT, AND/NAND, and OR/NOR. By addressing memory wall issues, our proposed design can provide powerful help for big data applications.

#### ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grant 61704005.

# REFERENCES

- N. S. Kim, T. Austin, D. Blaauw, T. Mudge, J. S. Hu, M. J. Irwin, M. Kandemir, V. Narayanan *et al.*, "Leakage current: Moore," *computer*, no. 12, pp. 68–75, 2003.
- [2] S. W. Keckler, W. J. Dally, B. Khailany, M. Garland, and D. Glasco, "GPUs and the future of parallel computing," *IEEE Micro*, vol. 31, no. 5, pp. 7–17, 2011.
- [3] S. Li, C. Xu, Q. Zou, J. Zhao, Y. Lu, and Y. Xie, "Pinatubo: A processing-in-memory architecture for bulk bitwise operations in emerging non-volatile memories," in *Proceedings of the 53rd Annual Design Automation Conference*, 2016, p. 173.
- [4] M. Wang, W. Cai, K. Cao, J. Zhou, J. Wrona, S. Peng, H. Yang, J. Wei, W. Kang, Y. Zhang *et al.*, "Current-induced magnetization switching in atom-thick tungsten engineered perpendicular magnetic tunnel junctions with large tunnel magnetoresistance," *Nature communications*, vol. 9, no. 1, p. 671, 2018.
- [5] L. Wei et al., "13.3 A 7Mb STT-MRAM in 22FFL FinFET technology with 4ns read sensing time at 0.9V using write-verify-write scheme and offset-cancellation sensing technique," in 2019 IEEE International Solid-State Circuits Conference - (ISSCC), Feb 2019, pp. 214–216.
- [6] Z. Wang, Z. Li, Y. Liu, S. Li, L. Chang, W. Kang, Y. Zhang, and W. Zhao, "Progresses and challenges of spin orbit torque driven magnetization switching and application (invited)," in 2018 IEEE International Symposium on Circuits and Systems (ISCAS), May 2018, pp. 1–5.
- [7] Z. Wang, B. Wu, Z. Li, X. Lin, J. Yang, Y. Zhang, and W. Zhao, "Evaluation of ultrahigh-speed magnetic memories using field-free spinorbit torque," *IEEE Transactions on Magnetics*, vol. 54, no. 11, pp. 1–5, 2018.
- [8] N. Hassan, S. P. Lainez-Garcia, F. Garcia-Sanchez, and J. S. Friedman, "Toggle spin-orbit torque MRAM with perpendicular magnetic anisotropy," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits.*
- [9] S. Jain, A. Ranjan, K. Roy, and A. Raghunathan, "Computing in memory with spin-transfer torque magnetic RAM," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 26, no. 3, pp. 470–483, 2017.
- [10] W. Kang, H. Wang, Z. Wang, Y. Zhang, and W. Zhao, "In-memory processing paradigm for bitwise logic operations in STT-MRAM," *IEEE Transactions on Magnetics*, vol. 53, no. 11, pp. 1–4, 2017.
- [11] L. Zhang, E. Deng, H. Cai, Y. Wang, L. Torres, A. Todri-Sanial, and Y. Zhang, "A high-reliability and low-power computing-in-memory implementation within STT-MRAM," *Microelectronics Journal*, vol. 81, pp. 69–75, 2018.
- [12] M. Zabihi, Z. Zhao, D. Mahendra, Z. I. Chowdhury, S. Resch, T. Peterson, U. R. Karpuzcu, J. Wang, and S. S. Sapatnekar, "Using spin-hall MTJs to build an energy-efficient in-memory computation platform," in 20th International Symposium on Quality Electronic Design (ISQED), March 2019, pp. 52–57.
- [13] F. Parveen, S. Angizi, Z. He, and D. Fan, "Low power in-memory computing based on dual-mode SOT-MRAM," in 2017 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED), July 2017, pp. 1–6.
- [14] G. Wang, Y. Zhang, J. Wang, Z. Zhang, K. Zhang, Z. Zheng, J. Klein, D. Ravelosona, Y. Zhang, and W. Zhao, "Compact modeling of perpendicular-magnetic-anisotropy double-barrier magnetic tunnel junction with enhanced thermal stability recording structure," *IEEE Transactions on Electron Devices*, vol. 66, no. 5, pp. 2431–2436, 2019.
- [15] W. Legrand, R. Ramaswamy, R. Mishra, and H. Yang, "Coherent subnanosecond switching of perpendicular magnetization by the field-like spin-orbit torque without an external magnetic field," *Physical Review Applied*, vol. 3, no. 6, p. 064012, 2015.
- [16] Z. Wang, W. Zhao, E. Deng, J.-O. Klein, and C. Chappert, "Perpendicular-anisotropy magnetic tunnel junction switched by spinhall-assisted spin-transfer torque," *Journal of Physics D: Applied Physic*s, vol. 48, no. 6, p. 065001, 2015.