Spintronic foundation cells for large-scale integration

Zhihua Xiao[1,2*], Yaoru Hou[1,2], and Qiming Shao[1,2#], *Senior Member, IEEE*

[1]Department of Electronic and Computer Engineering, HKUST, Hong Kong, China, *zxiaoam@connect.ust.hk
[2]ACCESS – AI Chip Center for Emerging Smart Systems, InnoHK Centers, Hong Kong, China, #eeqshao@ust.hk

As Moore's law reaches its practical limits and computing paradigms become increasingly complex, there is a growing demand for innovative device technologies. However, completely replacing complementary metal-oxide-semiconductor (CMOS) technology is not considered feasible in the foreseeable future. In this paper, we present how a promising emerging technology, spintronic devices, can be integrated with CMOS to function as foundational building blocks for scalable memory and computing systems.

***Index Terms*—Spintronics, CMOS, Neuromorphic Computing**

I. Introduction

Recent advancements in intelligent systems have transformed the computing paradigm by reshaping both computational architectures and resource demands. The traditional Von Neumann architecture and CMOS technology, which underpinned information processing during the 20th century, now face significant challenges such as the memory wall and intrinsic physical limitations. These issues have spurred interest in novel memory technologies and alternative computing paradigms that address modern system demands.

Emerging memory technologies promise to overcome the limitations of conventional architectures by addressing performance bottlenecks and reducing latency. New memory modalities—including phase-change memory, resistive random-access memory (RAM), and magnetic RAM—offer substantially lower latency and nonvolatility compared to traditional DRAM, thereby reducing CPU idle times and enhancing throughput. Advances in non-volatile memory improve energy efficiency by lowering power consumption for data retention and enable integration of persistent storage within computing systems. Moreover, these technologies support high-bandwidth memory architectures that boost data transfer rates and parallelism, facilitating rapid bulk data movement and improved processing efficiency. Collectively, such innovations provide scalable, high-performance solutions tailored to the needs of modern applications and big data analytics.

Alternative computing paradigms, including neuromorphic, probabilistic, and quantum computing, also rely on emerging device technologies for enhanced efficiency. Notably, spintronic devices—exploiting electron spin in addition to charge—offer novel functionalities and dynamic behaviors. In the past two decades, spintronic mechanisms have evolved from spin-transfer torque (STT) to spin–orbit torque (SOT) and voltage-controlled magnetic anisotropy (VCMA). The fabrication of these devices now incorporates advanced materials such as antiferromagnetic, topological, and two-dimensional (2D) substances. Research into these new physical phenomena and device architectures is continuously driving improvements in overall computing performance [1].

Integrating spintronic devices into large-scale architectures necessitates compatibility with CMOS technology[2].

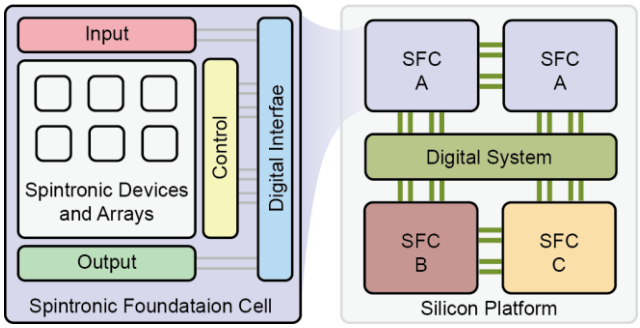

Fig. 1. Schematic diagram of a spintronic foundational cell. The scalable system consists of multiple spintronic foundational cells of various types, which are interconnected via digital interfaces.

Conventional CMOS logic relies on integrating fundamental cells such as logic gates, and these principles remain vital as emerging technologies develop. Novel devices and mechanisms must form foundational cells that interface seamlessly with CMOS or similar scalable architectures. To this end, spintronic foundational cells (SFCs) are designed to integrate key components—such as memory or crossbar arrays and probabilistic bits—with CMOS peripheral circuits for device control. Digital interfaces in these cells facilitate interaction among spintronic modules and existing digital circuits (Fig. 1).

In this work, we present design examples of spintronic foundational cells specifically tailored for scalable memory and computing systems, highlighting their potential to meet the challenges posed by next-generation intelligent systems.

II. Memory Foundation Cell with Quantum Materials

The schematic of foundation cell of SOT-MRAM is shown in Fig. 2a, in which the memory unit is the SOT-magnetic tunnel junction (MTJ) (Fig. 2b). To improve energy efficiency of SOT-MRAM, promising routes include increasing SOT efficiency and removing external field [3]. We report an all–van der Waals heterostructure integrating a type-II Weyl semimetal (Fig. 2c), $TaIrTe_4$, with an above-room-temperature ferromagnet, Fe_3GaTe_2, that enables robust field–free magnetization switching via an unconventional SOT [4]. The reduced crystal symmetry at the $TaIrTe_4$ surface produces a significant out–of–plane spin polarization and thus induces energy-efficient field-free switching when current is applied along its a–axis (Fig. 2d). The field-free switching polarity maintains until an in–plane

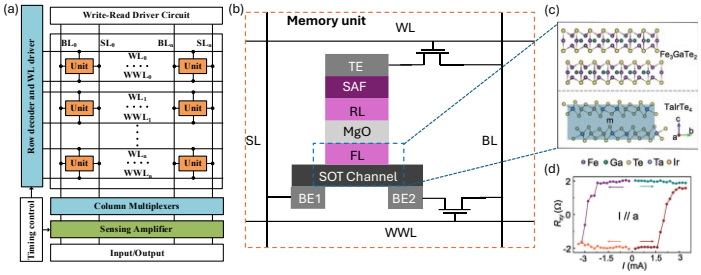

Fig. 2. (a) Schematic of the foundation cell of SOT-MRAM. (b) Schematic of the memory unit of SOT-MTJ. (c) Schematic of $TaIrTe_4$ and Fe_3GaTe_2. (d) Field-free SOT switching when the current is applied along the a-axis of $TaIrTe_4$.

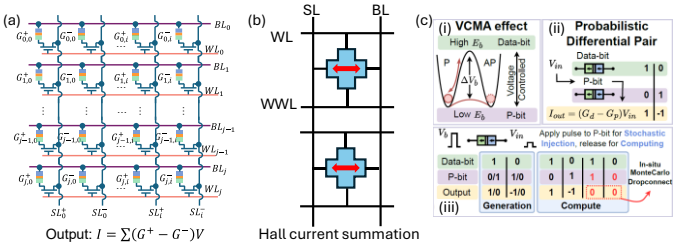

Fig. 3. (a) Schematic of an AiMC crossbar array. (b) Schematic of two Hall bar devices for realizing Hall current summation. (c) Schematics of (i) data bit, p-bit and (ii) their differential pair based on VCMA-MTJs. (iii) Possible outputs of the probabilistic differential pair for weight and dropconnect. (a)

magnetic field of 252 mT is applied. Macrospin simulations incorporating thermal fluctuations corroborate that larger spin canting angles reduce the required switching currents and improve stability. This robust, energy–efficient switching and high magnetic field resilience highlight the $TaIrTe_4/Fe_3GaTe_2$ heterostructure's potential for non–volatile memory and scalable spintronic applications.

III. Analog in Memory Computing Foundation Cells

Memristors have emerged as promising devices for efficient multiply-accumulate (MAC) operations in crossbar array-based foundation cells (Fig. 3a)—a capability that is critical for advancing analog in-memory computing (AiMC). However, device and circuit variations inherent in memristors can compromise the accuracy of analog computations. To construct a scalable AiMC system, it is essential to address and leverage these variations. Traditionally, on-chip training has been used to mitigate such issues; however, this approach is challenging for memristors due to their limited endurance.

In our recent study[5], [6], we introduce a hardware–software co-design framework that employs MTJ-based off-chip calibration for AiMC, achieving software-level accuracy without the overhead associated with expensive on-chip training. Experimentally, our results demonstrate that MTJ devices exhibit ultralow cycle-to-cycle variations, a finding validated through tests on over one million mass-produced devices. On the software side, the high degree of hardware uniformity enables us to develop an off-chip training strategy that fine-tunes deep neural network parameters, resulting in highly precise AiMC inference. This approach allows the system to be scaled by replicating multiple identical AiMC spintronic foundation cells, without being limited by device variations.

Furthermore, implementing AiMC under cryogenic temperatures can enhance performance and energy efficiency in computation-intensive environments, particularly in applications such as quantum control[7]. Magnetic topological insulators (MTIs) offer a promising path toward reliable AiMC by facilitating the summation of the anomalous Hall current[8], enabled by their large anomalous Hall resistance range and high noise-to-signal ratio (Fig. 3b).

In addition to stable memory states, probabilistic bits (p-bits) can be constructed using MTJs [9]. Conventionally, the generation and computation of random bits are separated by sample and transfer operations between analog and digital systems. We propose to use the dual feature – stable data bit and p-bit – of MTJs to realize an in-memory probabilistic computing scheme (Fig. 3c) [10]. We use probabilistic differential pairs to construct a Monte Carlo dropconnect Bayesian neural network to directly calculate the stochastic differential equations for image generation.

IV. Acknowledgements

These works are partially supported by ACCESS – AI Chip Center for Emerging Smart Systems, the InnoHK initiative of the Innovation and Technology Commission of the Hong Kong Special Administrative Region Government, ITC ITSP seed program (ITS/153/22), and RGC Theme-based Research Scheme (T46-705/23-R).

References

[1] Q. Shao, Z. Wang, Y. Zhou, S. Fukami, D. Querlioz, and L. O. Chua, "Spintronic memristors for computing," *npj Spintron.*, vol. 3, no. 1, p. 16, May 2025, doi: 10.1038/s44306-025-00078-z.

[2] Q. Shao, K. Garello, and J. Tang, "Spintronic foundation cells for large-scale integration," *Nat. Rev. Electr. Eng.*, Oct. 2024, doi: 10.1038/s44287-024-00106-w.

[3] Q. Shao *et al.*, "Roadmap of Spin-Orbit Torques," *IEEE Trans. Magn.*, vol. 57, no. 7, pp. 1–1, 2021, doi: 10.1109/TMAG.2021.3078583.

[4] Y. Zhang *et al.*, "Robust Field-Free Switching Using Large Unconventional Spin-Orbit Torque in an All-Van der Waals Heterostructure," *Adv. Mater.*, 2024, doi: 10.1002/adma.202406464.

[5] Z. Xiao, V. B. Naik, J. H. Lim, Y. Hou, Z. Wang, and Q. Shao, "Adapting magnetoresistive memory devices for accurate and on-chip-training-free in-memory computing," *Sci. Adv.*, vol. 10, no. 38, Sep. 2024, doi: 10.1126/sciadv.adp3710.

[6] Z. Xiao *et al.*, "Device Variation-Aware Adaptive Quantization for MRAM-based Accurate In-Memory Computing Without On-chip Training," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, vol. 2022-Decem, pp. 1051–1054, 2022, doi: 10.1109/IEDM45625.2022.10019482.

[7] Y. Liu *et al.*, "Cryogenic in-memory computing using magnetic topological insulators," *Nat. Mater.*, Jan. 2025, doi: 10.1038/s41563-024-02088-4.

[8] K. Qian *et al.*, "Cryogenic In-Memory Computing Circuits with Giant Anomalous Hall Current in Magnetic Topological Insulators for Quantum Control," *Int. Electron Devices Meet.*, 2024.

[9] S. Chowdhury *et al.*, "A Full-Stack View of Probabilistic Computing With p-Bits: Devices, Architectures, and Algorithms," *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 9, no. 1, pp. 1–11, 2023, doi: 10.1109/JXCDC.2023.3256981.

[10] Z. Xiao *et al.*, "In-Memory Neural Stochastic Differential Equations with Probabilistic Differential Pair Achieved by In-situ P-bit using CMOS Integrated Voltage-Controlled Magnetic Tunnel Junctions," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, 2024, doi: 10.1109/IEDM50854.2024.10873318.