SPINTRONIC DEVICES AND ITS APPLICATIONS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Mei-Chin Chen

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2020

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. Kaushik Roy, Chair

School of Electrical and Computer Engineering

Dr. Anand Raghunathan

School of Electrical and Computer Engineering

Dr. Vijay Raghunathan

School of Electrical and Computer Engineering

Dr. Zhihong Chen

School of Electrical and Computer Engineering

Approved by:

Dr. Venkataramanan Balakrishnan

Head of the School Graduate Program

This thesis is dedicated to my family and friends.

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my Ph.D. advisor, Prof. Kaushik Roy, for his patient guidance, encouragement, and support throughout my doctoral studies at Purdue University. The inspiration he gave has played significant part in shaping me over the past several years. At every stage he has motivated and encouraged me whenever I felt down and lost. I appreciate him for countless insightful discussions which help me approach complex research problems and organize my thoughts clearly and efficiently. I am truly fortunate to have him as my advisor, his enthusiasm and instruction are pivotal in making this dissertation possible.

I would also like to thank all of my doctoral dissertation committee members, Prof. Anand Raghunathan, Porf. Vijay Raghunathan, and Prof. Zhihong Chen, for providing valuable comments and suggestions to improve my research.

This dissertation would not have been possible without the help and support of the senior colleagues in Nanoelectronics Research Lab (NRL, Purdue University). I would take this opportunity to acknowledge the tremendous help and inspiration I have received from Dr. Xuanyao Fong (National University Singapore) and Dr. Yusung Kim (Intel Corp.). They have helped me built fundamental knowledge of spintronic device physics as well as circuit design. I like to thank all the former members of NRL members, especially, Dr. Karthik Yogendra (IBM), Dr. Abhronil Sengupta (Penn State University), Dr. Yong Shim (Intel Corp.), Dr. Zubair AL Azim (Intel Corp), Dr. Akhilesh Jaiswal (University of Southern California), Dr. Parami Wijesinghe (Intel Corp.), and Gopalakrishnan Srinivasan (MediaTek). I would also like to thank all the current NRL members for their support, especially, Saima Sharmin, Amogh Agrawal, Aayush Ankit, Bing Han, Minsuk Koo, Chankyu Lee, Efstathia Soufleri, and others I might have accidentally missed out.

Finally and most importantly, I would like to express my greatest gratitude to my parents for their endless love and everlasting support these years. Lastly, special thanks go to my sisters who bring joy to my life every single day.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

ABSTRACT

Chen, Mei-Chin Ph.D., Purdue University, May 2020. Spintronic Devices and Its Applications. Major Professor: Kaushik Roy.

Process variations and increasing leakage current are major challenges toward memory realization in deeply-scaled CMOS devices. Spintronic devices recently emerged as one of the leading candidates for future information storage due to its potential for non-volatility, high speed, low power and good endurance. In this thesis, we start with the basic concepts and applications of three spintronic devices, namely spin orbit torque (SOT) based spin-valves, SOT-based magnetic tunnel junctions and the magnetic skyrmion (MS) for both logic and machine learning hardware.

We propose a new Spin-Orbit Torque based Domino-style Spin Logic (SOT-DSL) that operates in a sequence of Preset and Evaluation modes of operations. During the preset mode, the output magnet is clocked to its hard-axis using spin Hall effect. In the evaluation mode, the clocked output magnet is switched by a spin current from the preceding stage. The nano-magnets in SOT-DSL are always driven by orthogonal spins rather than collinear spins, which in turn eliminates the incubation delay and allows fast magnetization switching. Based on our simulation results, SOT-DSL shows up to 50% improvement in energy consumption compared to All-Spin Logic. Moreover, SOT-DSL relaxes the requirement for buffer insertion between long spin channels, and significantly lowers the design complexity. This dissertation also covers two applications using MS as information carriers. MS has been shown to possess several advantages in terms of unprecedented stability, ultra-low depinning current density, and compact size. We propose a multi-bit MS cell with appropriate peripheral circuits. A systematic device-circuit-architecture co-design is performed to evaluate the feasibility of using MS-based memory as last-level caches for general purpose pro-

cessors. To further establish the viability of skyrmions for other applications, a deep spiking neural network (SNN) architecture where computation units are realized by MS-based devices is also proposed. We develop device architectures and models suitable for neurons and synapses, provide device-to-system level analysis for the design of an All-Spin Spiking Neural Network based on skyrmionic devices, and demonstrate its efficiency over a corresponding CMOS implementation.

Apart from the aforementioned applications such as memory storage elements or logic operation, this research also focuses on the implementation of spin-based device to solve combinatorial optimization problems. Finding an efficient computing method to solve these problems has been researched extensively. The computational cost for such optimization problems exponentially increases with the number of variables using traditional von-Neumann architecture. Ising model, on the other hand, has been proposed as a more suitable computation paradigm for its simple architecture and inherent ability to efficiently solve combinatorial optimization problems. In this work, SHE-MTJs are used as a stochastic switching bit to solve these problems based on the Ising model. We also design an unique approach to map bi-prime factorization problem to our proposed device-circuit configuration. By solving coupled Landau-Lifshitz-Gilbert equations, we demonstrate that our coupling network can factorize up to 16-bit binary numbers.

1. INTRODUCTION

For over the past five decades, the incessant down-scaling of the complementary metal-oxide-semiconductor (CMOS) has driven the evolution of the semiconductor industry. The number of on-chip transistors on a die increased significantly following Moore's law, with the scaling of CMOS technology [1].The reduction in the size of each transistor, increased integration density, lower of supply voltage and an increase in clock frequency, albeit with increase in power density, has contributed to higher processing capabilities as well as enhanced performance with more functionalities. However, further reducing the device feature size beyond 5nm may lead to significant impact on reliability, process variations, cost, and lifetime of transistors owing to potential thermal damage, increased leakage (Fig. 1.1 [2]). Moreover, as shown in Fig. 1.2, the classical technology driver, whose trend has followed Moore's Law for the past five decades, is predicted to flatten by 2025 [2]. Hence, in order to foster continued technology scaling in the absence of traditional Moore's Law, research has started in earnest to search for alternative device physics and material replacement. Moreover, for applications such as vision, optimization, pattern recognition, image/signal processing, and data classification, where the implementation in traditional von-Neumann architecture consisting of Boolean logic and memory circuits based on CMOS technology, turn out to be inefficient. This inefficiency in resources (which result in higher area requirement) and power is owing to the fundamental mismatch between the computational units and the underlying hardware.

Numerous potential solutions have been investigated to continue the performance scaling benefits. Fig. 1.3 depicts three different directions that could potentially advance the performance beyond the end of scaling limitation. In the near-term, research will be focus on developing advanced packaging and specialized architecture using existing computational units. In the mid-term, multiple beyond-CMOS devices,

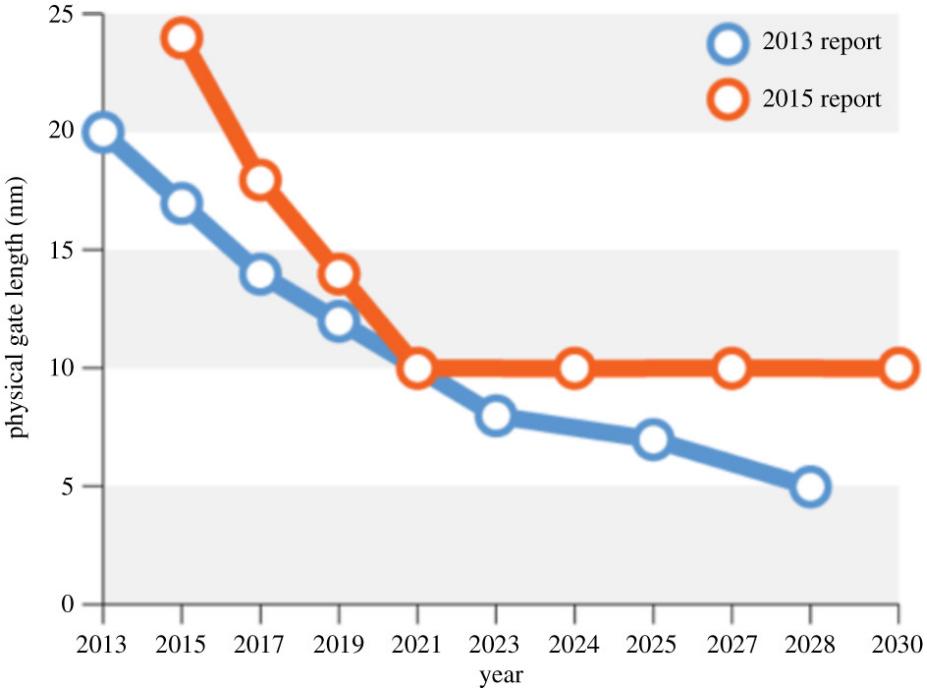

Fig. 1.1. The most recent prediction from ITRS suggests transistor scaling will end in 2021 (adapted from [2]).

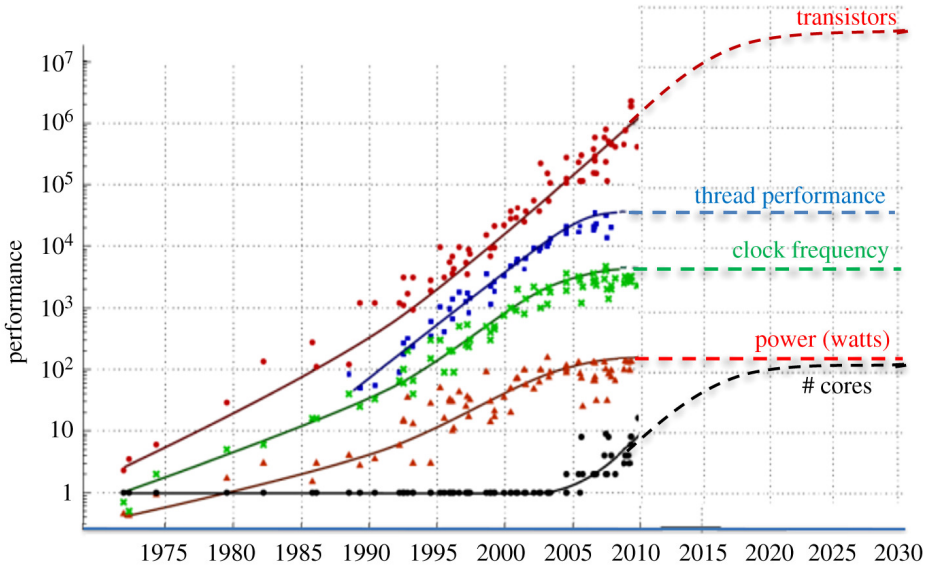

Fig. 1.2. All additional approaches to further performance will approximately end in 2025 due to the end of the road-map for improvements in semiconductor lithography. (adapted from [2])

such as graphene, carbon nanotubes, and Spintronics would be introduced to enhance the performance by creating more efficient underlying logic devices. The last direction represents using beyond CMOS technology to open new computing paradigms for non-conventional computing models such as neural network, probabilistic learning, and quantum computing, which aims to solve problems not well addressed by digital computing [2].

In order to bridge the performance gap and to keep with this scaling trend, this research exploits Spintronic devices to store data and implement computational units, exploring applications such as neural computing and combinatorial optimization. Due to the inherent non-volatility, good compatibility with the CMOS [3–6], spintronic devices are promising candidates for both storage and data processing.

The dissertation is organized as follows: The basics operation principles of various spin-based devices as storage elements are reviewed in Chapter 2. In Chapter

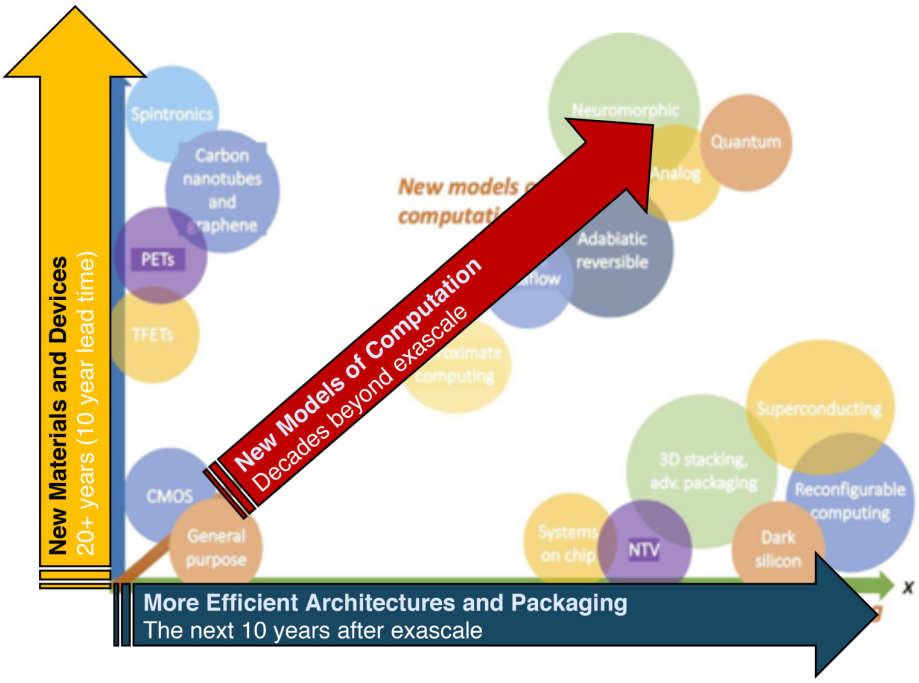

Fig. 1.3. Three potential paths forward to realize continued performance improvements for digital electronics technology (adapted from [2]).

3, we propose Spin-Orbit Torque based high-performance Domino-style Spin Logic (SOT-DSL). In Chapter 4 and 5, we move our focus to magnetic skrymion (MS), a recent exciting development in the field of spintronics for its remarkably high stability, extremely compact size, and ultra-low depinning current density. We propose the usage of MS as more robust information storage in cache memories (Chapter 4)—more specifically, a multi-bit MS cell is designed as last-level cache for general purpose processors. A device-to-system co-design is performed to evaluate the feasibility of our proposal. Moreover, we explore MS as computational units in deep spiking neural network (Chapter 5) by investigating MS-based synapses and neurons, and performing a systematic device-circuit-architecture co-design for digit recognition with the MNIST handwritten digits dataset. In Chapter 6, we introduce spintronic device as stochastic computational units to implement Ising spin model to solve bi-prime factorization. We demonstrated that our network can factorize up to 16-bit binary numbers. Finally, we summarize and conclude this dissertation in Chapter 7.

2. BASIC OPERATION OF SPIN-BASED DEVICES

This chapter provides a brief overview of several spintronic device structures, which have been widely used as storage elements or to perform specific logic functionalities. Fig. 2.1 depicts the spin-based devices whose basic functionalities will be covered in this section, including two-terminal magnetic tunnel junction (MTJ), three-terminal MTJ in which the magnetization switching is based on spin orbit torque (SOT), lateral spin valve (LSV), and magnetic skyrmion (MS).

2.1 Two-Terminal Magnetic Tunnel Junction

Magnetic tunnel junction (MTJ) is a basic storage element in almost all spin-torque based memories. As shown in Fig. 2.1(a), in the simplest description, an MTJ consists of a thin tunneling barrier (AlO or MgO) sandwiched by two ferromagnetic layers. These two layers are called the pinned layer (PL) and the free layer (FL), respectively. Normally, the PL is magnetically fixed by exchange interaction with an adjacent anti-ferromagnetic (AFM) layer so that it can be used as a reference layer. On the other hand, the magnetization of FL responds to external stimuli (such as magnetic field or spin-polarized current) and rotates freely between a direction either parallel (P) or anti-parallel (AP) to that of the PL. An MTJ presents a low resistance state (R_P) and a high resistance state (R_{AP}) for P and AP configuration, respectively [See Fig. 2.2]. These two extremely stable resistance state can be used to store binary information, and the stability of the nanomagnet in the presence of thermal noise is determined by energy barrier, E_B, and can be expressed as [7]

$$E_B = K_{u2}V \tag{2.1}$$

where K_{u2} is uni-axial anisotropy, and V is the volume of the magnet. Note that the energy barrier between two stable states are formed without application of any external voltage. Hence, a nanomagnet inherently non-volatile with no leakage power consumption. Moreover, the energy barrier between P and AP configuration is designed small enough to ensure easy writeability, but large enough to sustain thermal stability. In the absence of thermal agitation, the lifetime of information safely stored in a nanomagnet exponentially increases with the magnitude of the energy barrier, and can be expressed as

$$t_{lifetime} = \tau_0 \exp(E_B) \tag{2.2}$$

where $1/\tau_0$ is the attempt frequency of the order of 1ns. For instance, a barrier height of $40K_BT$ corresponds to a nanomagnet lifetime of $\sim$7.4 years [7].

Depending on the direction of the "easy" axis, MTJs can be simply categorized into either Perpendicular Magnetic Anisotropy (PMA) MTJ, or In-plane Magnetic Anisotropy (IMA) MTJ. Both IMA and PMA MTJs are thin-film magnets, the magnetization of which tends to point within the film owing to shape anisotropy. For IMA materials, the shape anisotropy dominates the resultant anisotropy of the nanomagnet, and hence, the "easy axis" is in-plane direction. Nanomagnets with elliptical cross-section is commonly used in this case, as the "easy axis" is in the direction of the longer axis [8–10]. On the contrary, the magnetocrystalline anisotropy is dominant in PMA nanomagnets. In order to lower the overall energy of the system, magnetization favors out-of-plane direction, and we refer the out-of-plane direction as "easy" axis direction [8, 11, 12]. Hence, PMA nanomagnets are usually of circular cross-sectional area.

The most important physical mechanism, namely, the Tunneling Magnetoresistance Ratio (TMR) and the spin-transfer torque (STT) effect will be qualitatively discussed in the following sections.

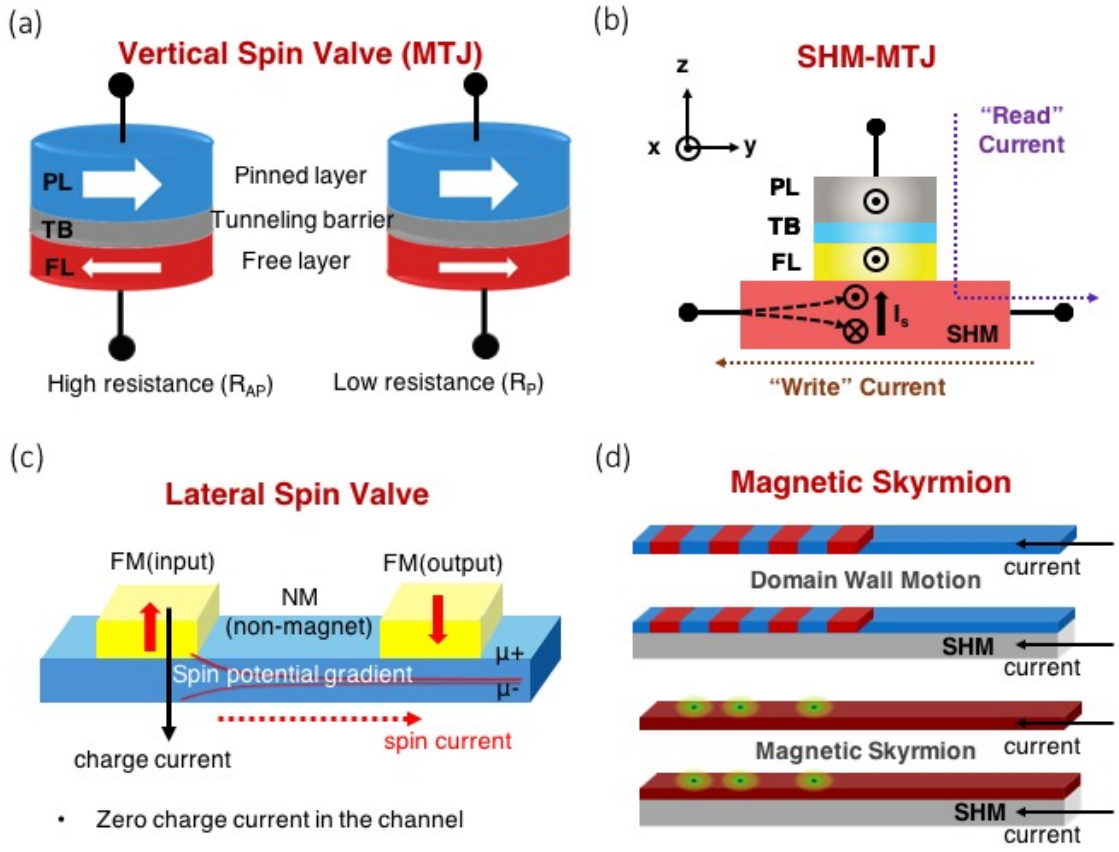

Fig. 2.1. (a) Vertical Spin Valve: A magnetic tunnel junction (MTJ) consists of a thin tunneling oxide sandwiched between two ferromagnets, namely the free layer (FL) and the pinned layer (PL). (b) 3 terminal SHM-MTJ: Device with MTJ on the top of the spin Hall metal (SHM) layer. When an injected charge current flows through the SHM, a spin-orbit torque is generated and exert a torque on the adjacent nanomagnet owing to the spin Hall effect (SHE). (c) Lateral spin valve consists of one input and one output nanomagnet connected by a non-magnetic channel. By injecting a charge current flowing through the input nanomagnet, the output nanomagnet can potentially be switched owing to non-local spin torque. The magnitude of the input spin current exponentially decreases with the connected channel length. (d) Magnetic skyrmion: Similar to domain wall motion (DWM) based device, a sequence of skyrmions can be stored in a long nanotrack and driven by either injecting a charge current to the long ferromagnetic (FM) layer or to the SHM underneath the FM layer.

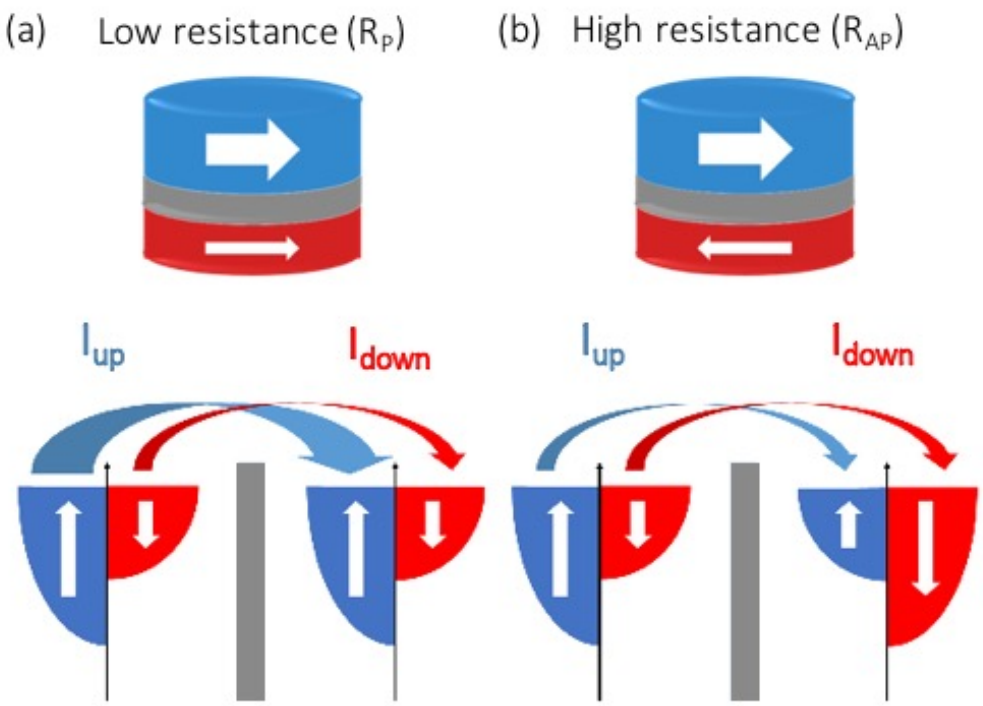

Fig. 2.2. A simplified structure of the Magnetic Tunnel Junction (MTJ) in (a) Parallel (P) and (b) Anti-parallel (AP) configuration with corresponding band diagram.

2.1.1 Tunneling Magnetoresistance Ratio (TMR)

The resistance of an MTJ is determined by the relative magnetization of two ferromagnetic layers. The Tunneling Magneto-resistance Ratio (TMR), defined by

$$TMR = \frac{R_{AP} - R_P}{R_P} \times 100\%, \tag{2.3}$$

measures the resistance difference of an MTJ between P and AP configurations. The TMR effect was first experimentally demonstrated in 1995 [1,13], and the demonstration of TMR effect up to 604 % at room temperature has been reported in 2008 [14]. TMR depends on the quality of well-crystallized MgO and interfaces [15]. High TMR ratio represents a larger gap in the resistance between the two states, and vice versa. Electrons flowing through an MTJ carry either up-spin (majority spin) or down-spin

(minority spin) [See Fig. 2.2]. In the absence of spin scattering or spin-flip processes during the tunneling process, the flow of electron carrying spin-up (I_{up}) and spin-down (I_{down}) can be treated as decoupled current paths, and the total charge current is I_{up}+I_{down}. The tunneling current of the majority and the minority current depends on the available free electrons state in the two ferromagnetic layers and the band matching of the interface. Suppose an MTJ is in the P configuration with a good band matching between the majority/minority density of states of the FL and PL, a low resistance is observed since there are sufficient states to accommodate all the available electrons for conduction. On the contrary, for an MTJ in the AP configuration, the majority (minority) spin in the FL (PL) is the minority (majority) spin in the PL (FL). This results in a large resistance as the majority/minority tunneling current is limited by either the availability of carriers in one contact or the empty states in the other contact. Note that since the mechanism of electron transport in an MTJ is the tunneling effect, the resistance of an MTJ (R_{MTJ}) depends on the thickness of the tunneling oxide as well as the applied voltage across the MTJ, which can be formulated using the Non-Equilibrium Green's Function based transport simulation framework [16].

2.1.2 Spin-Transfer Torque (STT) and magnetization dynamics

In a ferromagnetic material, the density of states for spin-up and spin-down electrons are different. Hence, a ferromagnet reorients the flow of unpolarized current to its magnetization by exerting a spin transfer torque. The opposite scenario was theoretically predicted by Slonczewski and Berger in 1998 [17, 18]— that is, when the polarized spin current is injected into a ferromagnet, the injected spin will exert a torque to the magnetization of the ferromagnet as well. To conserve the total angular momentum of the system, the change in the angular momentum of the spin current is equivalent to the change in the ferromagnet. In other words, the angular momentum lost by injected electrons (ferromagnet) must be absorbed by the ferromagnet (in-

jected electrons). This angular momentum transfer is called the spin-transfer torque (STT) effect and can be expressed as [19, 20]

$$\tau_{STT} = \gamma \frac{\hbar}{2} \frac{\eta I}{q} \frac{1}{M_s V} (m \times m_p \times m) \tag{2.4}$$

where m is the normalized magnetization vector; m_p is the fixed layer polarization; γ is the Gilbert gyromagnetic ratio; and $M_{sat}V$ is the total saturation magnetization.

To understand the magnetic dynamics of the free layer, we need to take the spin-transfer torque from the injected spin polarized current, the applied magnetic field, magnetic anisotropies, damping, and thermal fluctuations into consideration. As shown in Fig. 2.3, the precession of the magnetization is owing to an applied magnetic field, whereas the damping torque moves the magnetization toward the effective field direction. The spin-transfer torque can enlarge the damping torque, or exploit a torque in the opposite direction from the damping depending on the direction of injected current. The magnetization dynamics of the FL under an external stimuli, such as a magnetic field or spin current, at zero temperature can be obtained by solving *Landau-Lifshitz-Gilbert-Slonczewski* (LLGS) equation. [21, 22]

$$\frac{d\widehat{m}}{dt} = -\gamma(\widehat{m} \times H_{eff}) + \alpha(\widehat{m} \times \frac{d\widehat{m}}{dt}) + \frac{I_s}{qN_s}(\widehat{m} \times \sigma \times \widehat{m}), \tag{2.5}$$

where $\gamma = \frac{2\mu_B \mu_0}{\hbar}$ is the Gilbert gyromagnetic ratio for electron; $\widehat{m}$ is the unit vector of the FL; α is the Gilbert damping ratio; H_{eff} is the effective magnetic field incorporating the shape anisotropy for an elliptical disk; $N_s = \frac{M_s V}{\mu_B}$ is the number of spins in the FL occupying a definite volume V; M_s is the saturation magnetization; μ_B is the Bohr magneton; and μ_0 is the magnetic permeability.

At non-zero temperature, the magnetization switching dynamics of an MTJ is influenced by thermal noise, which is accounted as a random thermal field and can be factored into the LLGS equation by augmenting H_{eff} with a thermal field $H_{thermal}$, expressed as

$$H_{thermal} = \sqrt{\frac{\alpha}{1+\alpha^2} \frac{2K_B T}{\gamma \mu_0 M_s V \delta_t} G_{0,1}} \tag{2.6}$$

where $G_{0,1}$ is a Gaussian distribution with zero mean and unit standard deviation; K_B is the Boltzmann constant; T is the temperature; and δ_t is the simulation time

step. Hence, an MTJ exhibits stochastic switching during "write" operation in the presence of thermal noises.

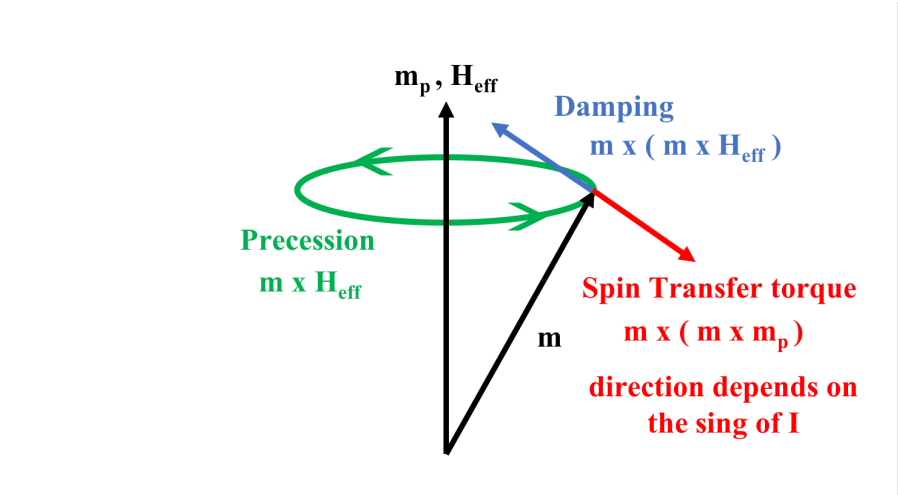

Fig. 2.3. Torques acting on a single-domain magnet.

2.1.3 Spin-Orbit Torque (SOT)

In a two terminal MTJ, FM layers serve as a spin polarizer, and the spin current is generated by passing a charge current through an FM layer. Hence, the efficiency of the generated spin current is limited by the polarization strength of the FM layer. Recently, a novel and efficient way to generate spin current has been experimentally demonstrated on I/FM/SHM (I:insulator, FM: ferromagnetc, and SHM: spin Hall metal) multilayer structures, and have opened up the possibility of much greater spin injection efficiencies owing to strong spin-orbit coupling (SOC) [23]. Spin orbit torque (SOT) can be generated at the interface of FM/SHM when a charge current passes through the underlying SHM. Specifically, efficient magnetization switching [24–30], domain wall motion [31–34], and spin-torque oscillations [35–37] were experimentally observed, which established this process as a more efficient method for magnetization switching.

Two possible mechanisms have been proposed to explain the SOT phenomenon, namely the Rashba model with broken spatial inversion symmetry [33, 38], and the spin Hall effect (SHE) [39, 40]. In this dissertation, we consider SHE to be the dominant underlying physical phenomenon. In Fig. 2.1(b), electrons with opposite spin polarization scatters on the top and bottom surfaces of SHM when passing a "write" current through the HM. The spin current, I_s, is generated by a charge current flowing through the SHM underlayer, and can be expressed as

$$I_s = \theta_{sh} \frac{A_{MTJ}}{A_{SHM}} I_e \tag{2.7}$$

where A_{MTJ} and A_{SHM} are the cross sectional areas of the MTJ and SHM, respectively, and θ_{sh} is the spin-Hall angle [39]. It is worth noting that spin current generated from SHM can potentially provide energy-efficient "write", as the spin injection efficiency can be larger than 100 % and is not limited by the polarization of the PL. Moreover, the polarization of the generated spin current is in-plane direction owing to SHE, and thereby SOT induced magnetization reversal is only possible for IMA MTJ [24, 25, 27].

2.2 Lateral Spin Valve

Figure 2.1(c) depicts a lateral spin valve (LSV) structure, consisting of an input and output nanomagnet connected by a non-magnetic channel. By injecting electrons through the input nanomagnet, electrons are spin polarized with the same magnetization of the input nanomagnet and accumulate underneath the input nanomagnet. Owing to the gradient of spin potential difference between the two sides, a spin current diffuses from the input to the output nanomagnet, and exerts a non-local spin-torque to the output nanomagnet. As a result, stored information can be transmitted from input to output nanomagnet. The non-magnetic channel here serves as a conduit for information transfer through the use of non-local spin injection. Owing to the spin-flip process in the channel, the magnitude of the injected spin current decays

exponentially with the distance between two nanomagnets. Therefore, material with larger spin-flip length would be chosen in the LSV structure [41, 42].

2.3 Magnetic Skyrmion

Figure 2.1(d) depicts multiple data bits stored in a sequence of magnetic domains, separated by domain walls, within a nanowire. DWM-based caches [43–46] have shown significant improvement in performance (with higher packing density and better energy efficiency) over other spintronic devices. However, the motion of domain walls might be pinned by the presence of defects [47], raising concerns about the feasibility of DWM-based applications. Magnetic skyrmions, on the other hand, have recently emerged as a promising alternative for future information carrier [48–51]. As shown in Fig. 2.1(d), similar to a DWM based device, a sequence of skyrmions can be stored in a long nanotrack, and driven by either injecting a current through the long nanotrack or through the SHM underneath the long nanotrack.

Magnetic skyrmions have been shown to possess several benefits over DWM based devices in terms of stability, density, and are less limited by imperfectness of the material. Specifically, its topological properties prevent the motion of skyrmions from being pinned at defect sites in a magnetic layer, and thus make them a more robust information carrier. The swirling structure of a skyrmion is characterized by the topological number defined by

$$S = \frac{1}{4\pi} \int m \cdot ((\frac{\partial m}{\partial x}) \times (\frac{\partial m}{\partial y}) dx dy) \tag{2.8}$$

as the integral of the solid angle and counts how many times $m(r) = m(x, y)$ wraps the unit sphere, m is the normalized magnetization. The topological number for a FM state ($S = 0$) is different from a skyrmion state ($S = 1$), thereby a skyrmion cannot be continuously deformed to an FM state or other magnetic state. A topological barrier is inherently built between an FM and a skyrmion state, and thus skyrmions are topologically protected and possess relatively stable structures.

As shown in Fig. 2.4, depending on the spin configuration, the whirling structure can be a chiral skyrmion, or Neel skyrmion, and its state can be explained by the presence of Dzyaloshinskii-Moriya Interaction (DMI) [52, 53] – the DMI between two atomic spins S_1 and S_2 with a neighboring atom can be expressed as $H_{DM} = -D_{1,2} \cdot (S_1 \times S_2)$ where $D_{1,2}$ is the Dzyaloshinskii-Moriya (DM) vector [48,49, 54–57]. Skyrmion lattices were first observed experimentally in B20 compounds (non-centrosymmetric bulk magnetic materials) such as MnSi, FeCoSi and FeGe, whose non-centrosymmetric crystal structure gives rise to bulk Dzyaloshinskii-Moriya Interaction (DMI) [54]. Recently, interfacial DMI have been predicted from a 3-site indirect exchange mechanism between two atomic spins S_1 and S_2 with a neighboring atom having a large spin-orbit coupling (SOC) [58]. As shown in Fig. 2.5, an interfacial DMI is generated at the interface between a thin FM layer and a SHM layer having strong SOC. This mechanism generates a DMI for the interface spins S_1 and S_2 with the DMI vector. A general advantage of skyrmions in thin films is that several parameters, such as the magnetic anisotropy and the DMI strength can be adjusted experimentally, can be modified by ion irradiation and diluted by increasing the film thickness [58], respectively.

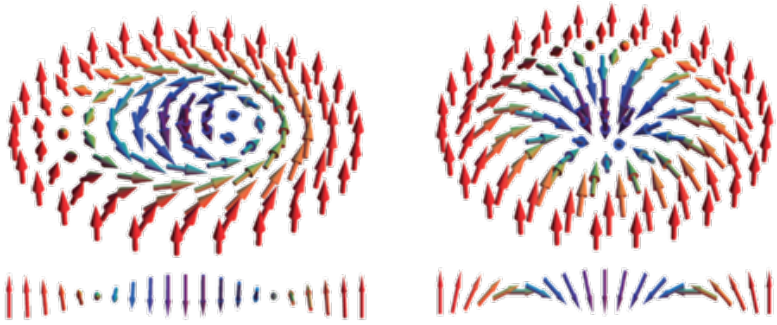

Fig. 2.4. Schematic of a chiral skyrmion (left) and a Neel skyrmion (right). The magnetization is pointing down in the center and pointing up on the edges. The magnetization rotates by 2 π along a diameter around an axis perpendicular to the diameter. [48]

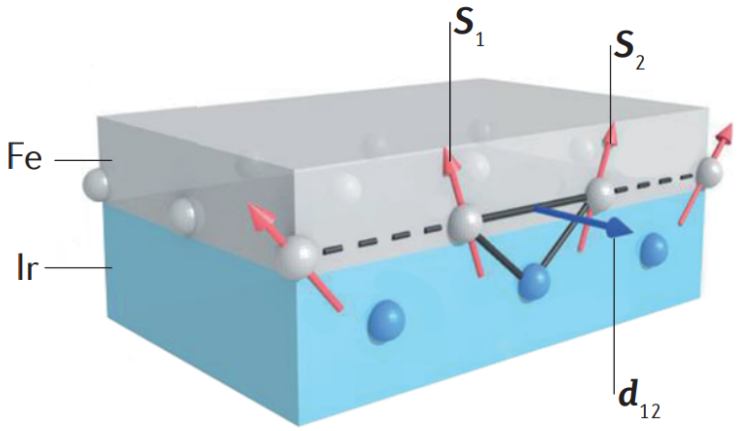

Fig. 2.5. Sketch of a DMI at the interface between a ferromagnetic metal (grey) and a metal with a strong SOC (blue). The direction of the DMI vector D_{12} is perpendicular to the plane of the triangle (black solid line), composed of two magnetic sites (grey) and an atom with strong SOC (blue). [48]

3. DOMINO-STYLE SPIN-ORBIT TORQUE-BASED SPIN LOGIC

3.1 Introduction

In quest for non-volatile and power-efficient logic styles, has started in earnest to use *spin* rather than *charge* as state variables for computation. A well-known example is the nano-magnetic logic (NML) [59] where information between clocked nano-magnets propagates through dipolar field rather than charge. Since no charge transfer occurs during the computation process, logic operations with fundamentally low energy cost can be achieved [60]. However, high energy consumption for the clocking scheme, restrictions in the number of fan-ins, and scaling challenges are some of the key obstacles to the achievement of energy-efficient digital logic circuits [60]. As a result, several approaches to reduce the clocking energy are under investigation, for example, magnets covered with enhanced permeability dielectric (EPD) films have been proposed to concentrate the external field [61].

All spin logic (ASL) [62], based on lateral spin valve and non-local spin transfer torque effect, employs nano-magnets to store information and uses pure spin current to propagate information. As a result, ASL eliminates the energy inefficiency of NML stemming from the generation of the clocking field. Moreover, ASL provides low voltage operation, high logical efficiency, and high integration density compared to its CMOS counterparts [62]. However, the operating speed of ASL is severely limited by the incubation delay of the magnetization switching. When a nano-magnet (described by the direction of its magnetization m) is interacting with the spin current (I_S), only the transverse component of the spin current ($m \times m \times I_S$) is absorbed by the nano-magnet. Therefore, when m and I_S are close to collinear, spin absorption is minimal and hence, the torque acting on the nano-magnet is close to zero. For switching

to occur, considerable amount of time is required for m to sufficiently deviate from its easy-axis and absorb enough spin current. Consequently, high speed design is significantly challenging for ASL. In order for spin-based logic to provide the same level of performance as todays charge-based digital logic, there is a strong demand for a new spin-based logic device and logic style that allows higher speed of operation.

In this chapter, we propose Spin-Orbit-Torque based Domino style Spin Logic (SOT-DSL), a device structure well-suited for higher performance logic applications. SOT-DSL operates by first presetting the output magnet to its hard-axis using energy-efficient spin orbit torque [63] prior to the evaluation mode. During the evaluation mode, information from the preceding nano-magnet is transmitted to the output magnet in the form of pure spin current. Note that the polarization of the spin current and the magnetization of the output magnet are orthogonal to each other. Therefore, when spin signal interacts with the output magnet, spin current absorption is maximal, and thus it exerts a larger torque on the magnet. Therefore, incubation delay can be eliminated and the proposed SOT-DSL is able to perform magnetization switching with much smaller input spin current while operating at higher speed. In addition, the proposed SOT-DSL can be concatenated with the addition of clock signals and can implement any set of Boolean functions. The rest of the chapter is organized as follows. Section 3.2 describes the structure and operation of our proposed SOT-DSL. Section 3.3 presents a four-phase pipelined operation required for the concatenation of SOT-DSL devices. The simulation framework used in this work is then discussed in Section 3.4. The results of our simulation and evaluation are presented in Section 3.5. Finally, Section 3.6 concludes this chapter.

3.2 SOT Based Domino Style Spin Logic

The device structure of SOT-DSL consists of a lateral spin valve with perpendicularly magnetized ferromagnets (FM) in contact with spin Hall metal (SHM), as shown in Fig. 3.1. The role of SHM is to drive the FM underlayer to its hard-axis

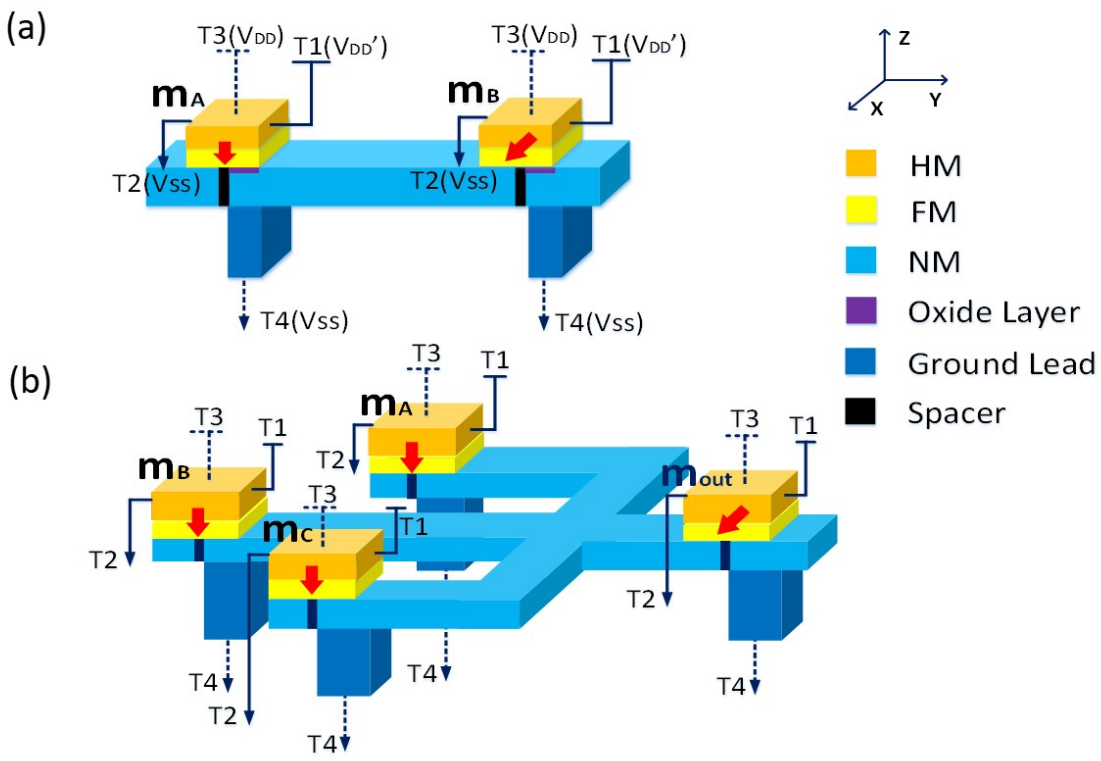

Fig. 3.1. (a) SOT buffer/inverter (b) SOT-DSL majority gate.

by utilizing spin current generated from spin Hall effect (SHE). The non-magnetic (NM) channel serves as a conduit for information transfer through the use of non-local spin injection. To ensure unidirectional information flow, the transmitting side of the FM/NM interface has high polarization (high-P) and receiving side has low polarization (low-P) with ground lead placed closer to the transmitting side [62]. The operation of SOT-DSL consists of two modes: Preset and evaluation. In order to provide these two operations, each unit of SOT-DSL is equipped with four terminals: terminal 1 and 2 (T1, T2 in Fig. 3.1) are used during preset, and terminal 3 and terminal 4 (T3, T4 in Fig. 3.1) are used during evaluation. The following subsections describe each of the two-step operation in detail.

3.2.1 Preset (Set to Hard Axis)

During the preset mode, a charge current is passed through the SHM layer to set the magnetization of output magnet into its hard axis. In the example shown in Fig. 3.1, charge current flowing through SHM in y direction (T1 to T2) generates +x directed spin current in z-direction in the bulk of SHM. The transverse component of spin current is absorbed by the FM underneath the SHM and aligns the magnetization of FM into +x (hard axis). Note that, in contrary to easy-axis switching where the spin current needs to overcome the damping torque of the nano-magnet ($\sim \alpha H_{ani}$, where α is the Gilbert damping parameter and H_{ani} is the anisotropy field), hard-axis switching requires the spin current to overcome the anisotropy field of the nano-magnet ($\sim H_{ani}$). Therefore, the required spin current for hard-axis switching is typically larger than easy-axis switching with typical α values much less than 1. However, the charge current (I_e) flowing through the SHM and the corresponding spin current(I_s) can be expressed as $I_S = \frac{A_{FM}}{A_{SHM}}\theta_{SH}[1 - sech(\frac{t_{SHM}}{\lambda_{sf}})]I_e$ [64]. Using simulation parameters listed in Table 3.1, we can obtain $I_S \cong 2.2I_e$. Therefore, spin current generated by SHE in our proposed structure has injection efficiency much larger than 100%, and hence, hard-axis switching can be performed with low energy cost.

3.2.2 Evaluation

Spin information from input magnet is transmitted to the next magnet during the evaluation step. Charge current injected from T3 of SHM to T4 develops a non-equilibrium spin potential underneath the transmitting side, which in turn generates pure spin current along the NM spin channel. In the proposed SOT-DSL device structure, the magnetization of output magnets are initialized to its hard axis direction (+x or x) prior to the evaluation step. Hence, a small spin current is sufficient to switch the output magnets into their easy-axis direction. Also note that the polarity of the voltage applied across the T3 and T4 can result in an accumulation of either the

spins in the direction of FM or the opposite spins. Therefore, the structure shown in Fig. 3.1(a) can operate as an inverter when a positive voltage is applied, and operate as a buffer when a negative voltage is applied.

Just like ASL, SOT-DSL has analog characteristic of current mode switching. Thus, multiple magnets can be connected together to implement functions like majority evaluation, as shown in Fig. 3.1(b), whose magnetization switching behavior is shown in Fig. 3.2. Moreover, such majority gate can be used to implement NOR/OR and NAND/NOR by fixing one of its inputs and applying appropriate voltages [65].

3.3 Four-Phase Pipelined SOT-DSL

Fig. 3.3 shows multiple stages of SOT-DSL concatenated using spin channels. Each black box represents a logic block, and red boxes are input or output magnets.

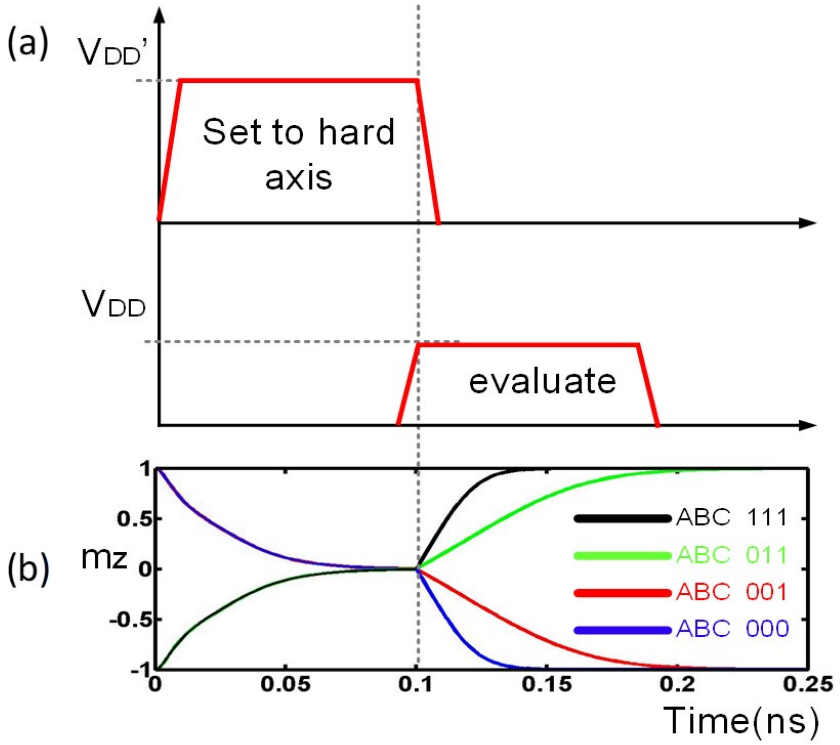

Fig. 3.2. (a) SOT buffer/inverter (b) SOT-DSL majority gate.

SHM above OUT magnets are connected to CLK1 and CLK4 whereas SHM above IN magnets are connected to CLK2 and CLK3. The operation is as follows: preset/evaluation, preset/evaluation, like a domino stack. As shown in Fig. 3.4, CLK1 and CLK3 are asserted to preset magnets to their hard axes (input current flows from T1 to T2, see Fig. 3.1) whereas CLK2 and CLK4 are asserted during logic evaluation (input current flows from T3 to T4, see Fig. 3.1). For each operation cycle, spin information stored in IN_i magnets is transmitted to the IN_{i+1} magnets. Since magnets are nonvolatile, extra latches or flip-flops are not required to implement pipeline operation.

Note that in Fig. 3.3, CLK1 presets the magnetizations of OUT_j magnets into their hard axes. Concurrently, CLK2 is asserted to pass spin information from IN_i magnets to their corresponding OUT_j magnets. This constitutes one stage in the logic pipeline. Similarly, spin information is passed to the next stage by asserting CLK3 which presets the IN_{i+1} magnet to its hard axis, and then asserting CLK4 to pass spin information from the OUT_j magnet to the IN_{i+1} magnets. The operation of each OUT_j magnet goes through a phase sequence of preset→to be evaluated→float→evaluate.

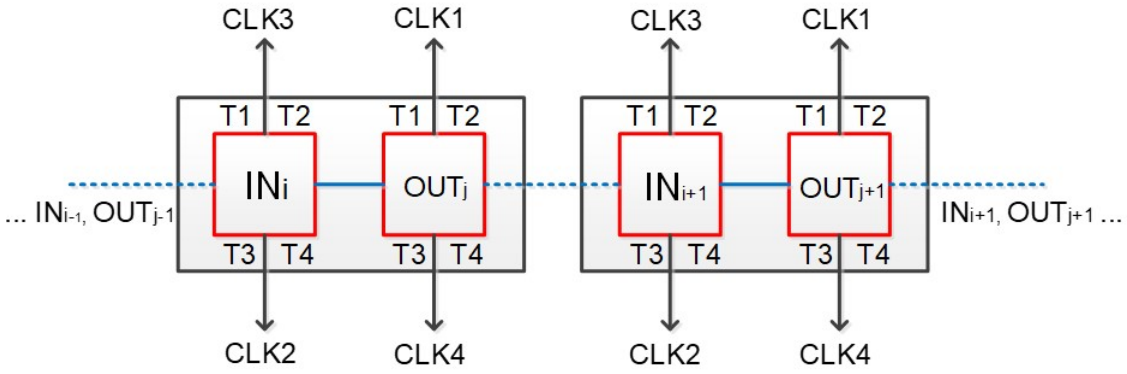

Fig. 3.3. Pipeline operation of several stages of SOT-DSL concatenated using spin channels (dash line). Black boxes represent one stage of SOT-DSL (see Fig. 3.1). Input and output magnets (red boxes) are connected by spin channel.

On the other hand, the operation of each IN magnet goes through a phase sequence of float→evaluate→preset→to be evaluated.

3.4 Modeling and Simulation Frame Work

We employ the modular approach [66] to evaluate and analyze our proposed SOT-DSL. In this approach, every part of SOT-DSL(FM, NM and interface) may be modeled as an elementary module. Each module is represented in the form of a 4-component (one charge and three spin in x, y, z directions) conductance element with series and shunt components. The branch current and node voltage are related using the relation $[I_c, I_C^z, I_C^x, I_C^y] = [G]_{4\times4}[V_c, V_C^z, V_C^x, V_C^y]$. These modules are then connected together as an inverter/buffer (as shown in Fig. 3.5). The magnetization dynamics is captured by self-consistently solving the Landau-Lifshitz-Gilbert (LLG) equation with spin diffusion and spin orbit torque terms.

Using the same approach, the spin circuit for a spin majority gate is constructed by connecting elementary modules according to its physical structure. The simula-

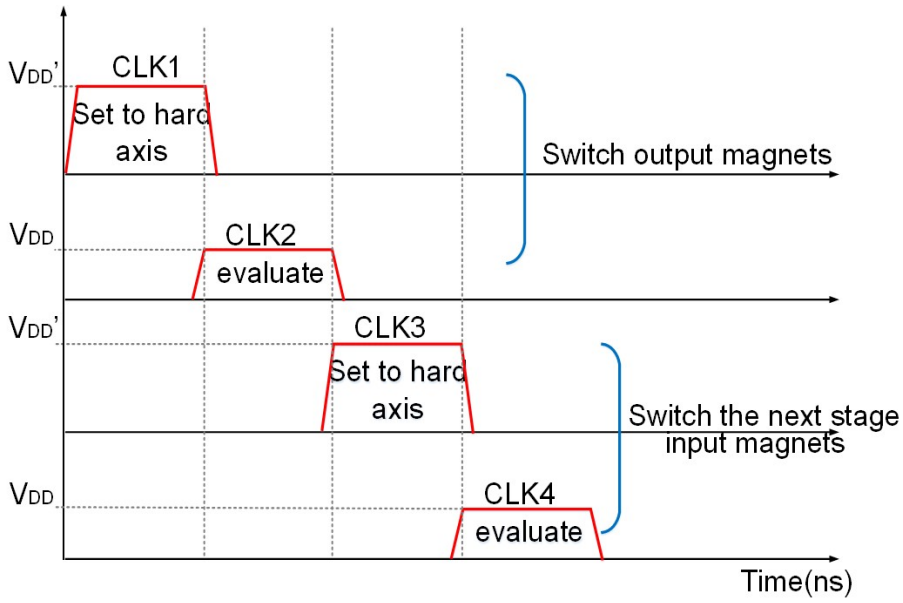

Fig. 3.4. Operation of four phase pipeline.

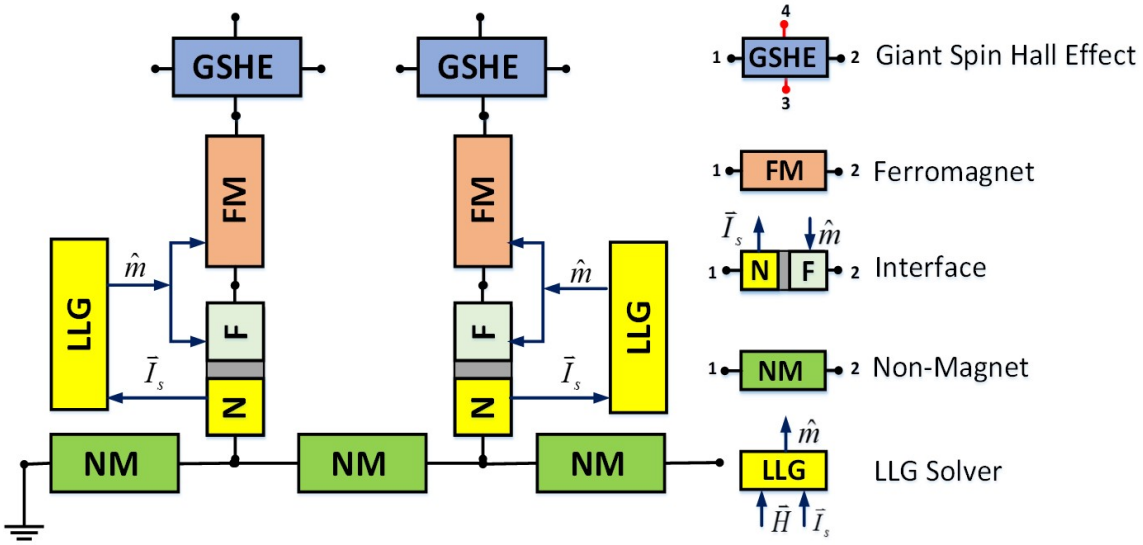

Fig. 3.5. Spin circuit model for SOT-DSL showing a lateral spin valve in contact with spin hall metal (SHM).

tion parameters are listed in Table 3.1. The results for ASL based spin majority gate are obtained by solving LLG equation and spin diffusion term. On the other hand, the results of SOT-DSL based spin majority gate are obtained by solving LLG and spin orbit torque term together with spin diffusion term. Note that the thermal fluctuation is not taken into consideration in our simulation, and as mentioned earlier in Section 3.2, high P/low P are used at the input/output interface to ensure unidirectional propagation of spin information.

3.5 Result and Discussion

In order to provide a quantitative assessment of our proposed SOT-DSL, a comparison study with ASL is carried out. The key advantage of SOT-DSL is its robustness and low energy consumption. For faster magnetization dynamics and hence higher frequency operation, the magnitude of spin current interacting with the output magnet needs to be increased by applying larger voltage at the input ports (T3 of input

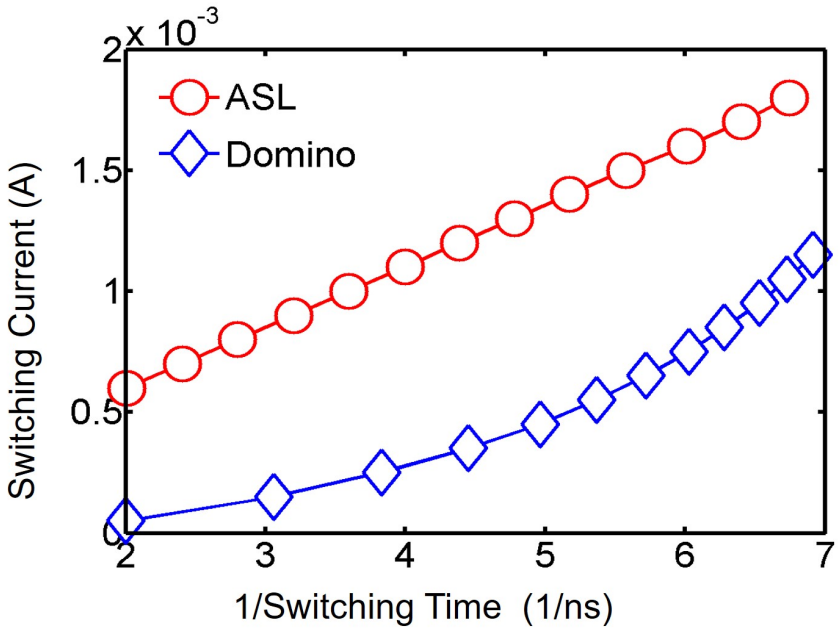

Fig. 3.6. Required input charge current for evaluation of ASL and SOT-DSL with increasing frequency of operation.

side magnets in Fig. 3.1). The corresponding input charge current that flows from the input ports into the ground lead, for both ASL and SOT-DSL majority gate during evaluation mode, are shown in Fig. 3.6. The required switching current during evaluation mode for ASL majority gate is much larger than that for SOT-DSL. This is because the spin current in ASL needs to fully switch the output magnet from one state to another, whereas in SOT-DSL, spin current only needs to bias the magnet off from its meta-stable state. Also, faster switching is possible with SOT-DSL since the magnets and spins are orthogonal to each other during its operation.

For ASL, the total energy consumption can be expressed as $E_{tot} = I_{ASL} \times V_{ASL} \times t_{SW}$, where I_{ASL} is the injected charge current flowing into the ground lead, V_{ASL} is the corresponding applied voltage, and t_{SW} represents the switching time. Since I_{ASL} is inversely proportional to t_{SW} (Fig. 3.6), the total energy consumption (E_{tot}) for an ASL majority gate increases linearly with the operating frequency (Fig. 3.7). For SOT-DSL majority gate, on the other hand, the total energy consumption comes from

preset and evaluation mode, and can be expressed as $E_{tot} = E_{PRE}(I_{SHM} \times V_{SHM} \times t_{SW1}) + E_{EVA}(I_{SOT-DSL} \times V_{SOT-DSL} \times t_{SW2})$. E_{PRE} and E_{EVA} represent the energy consumption in the preset mode and the evaluation mode, respectively. I_{SHM} is the input charge current flowing through SHM during the preset mode, and $I_{SOT-DSL}$ is the injected charge current flowing into ground lead during the evaluation mode. V_{SHM} and $V_{SOT-DSL}$ are the corresponding applied voltage. t_{SW1}, t_{SW2} represent the switching time for the preset mode and the evaluation mode, respectively. To achieve higher operating frequency, the switching time ($t_{sw1}+t_{sw2}$) needs to be reduced, which requires higher I_{SHM} and $I_{SOT-DSL}$. Note that the high resistivity of SHM requires larger applied voltage (V_{SHM}), thus higher energy would be consumed if I_{SHM} is increased to achieve higher operation frequency. As a result, a feasible way to achieve higher operation frequency would be increasing $I_{SOT-DSL}$ to reduce the switching time during the evaluation mode and keeping I_{SHM} constant. Consequently, preset mode energy consumption remains fixed while evaluation mode energy consumption increases with increasing operation frequency. At low operation frequency, the energy consumption during the preset mode dominates over the evaluation mode; however, it is still comparable to the total energy consumption of ASL. On the other hand, at high operation frequency, total energy consumption of SOT-DSL is much lower than ASL, since ASL requires much higher switching current than SOT-DSL. As shown in Fig. 3.7, SOT-DSL achieves as much as 50% improvement in energy consumption at high frequencies, as shown in Fig. 3.7.

Another benefit of SOT-DSL over ASL is the elimination of buffers in reasonably long spin channels. Note that spin signal exponentially decays along the spin channel due to spin-flip processes. Therefore, during the evaluation mode, the channel length and spin flip length of the channel material have a direct impact on the spin transfer torque acting on the target magnet. As a result, logic gates with long spin channels in ASL require insertion of buffer to satisfy target timing constraints. On the other hand, in SOT-DSL, presetting the magnetization of the output magnet to hard axis is independent of the channel length. And since initial magnetization at the onset

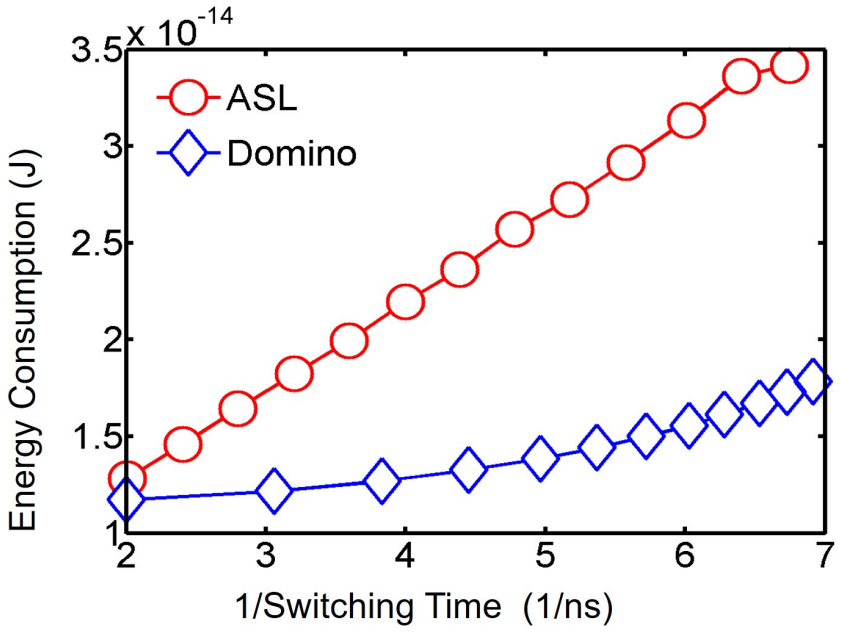

Fig. 3.7. Total energy consumption vs. frequency.

of evaluation mode is already set into hard axis, channel length has less impact on switching current. Fig. 3.8 compares the effect of channel length on the switching times of output magnet in inverters designed with ASL and SOT-DSL. Contrary to ASL, channel length in SOT-DSL has much smaller impact on the switching time of the output magnet.

3.6 Conclusion

In this chapter, we propose SOT-DSL, a pre-charge/evaluate (domino-style) spin logic using spin-orbit torque. In SOT-DSL, all nano-magnets are driven by orthogonal spins rather than collinear spins, which eliminates the incubation delay and leads to high performance. Moreover, the non-volatility of nano-magnets can facilitate pipelined operation without the need for extra latches. Multiple stages of SOT-DSL using 4-phase pipelining scheme is analyzed and compared with ASL. Our simulation results show that SOT-DSL consumes 50% less energy compared to ASL. Further-

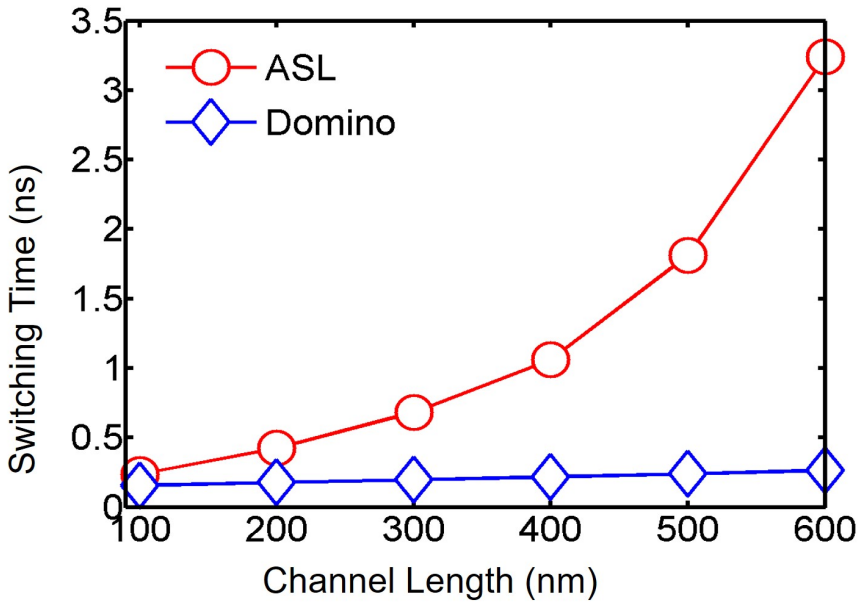

Fig. 3.8. Switching time with increasing spin channel length for ASL and SOT-DSL.

more, channel length has less impact on SOT-DSL compared to ASL, leading to robust logic.

Table 3.1.
SOT-DSL parameters

Damping Constant	0.024
Saturation Magnetization	800 KA/m
Magnetic Anisotropy	$6 \times 10^5 \mathrm{J/m^3}$
Channel Spin-Flip Length	500 nm
Heavy Metal Spin-Flip Length	1 nm
Channel Resistivity (Cu)	$6.9\ \Omega - \mathrm{nm}$
Heavy Metal Resistivity (Ta)	$200\mu\Omega - \mathrm{cm}$
Magent Resistivity (CoFeB)	$200\mu\Omega - \mathrm{cm}$
Spin Hall Angel (Ta)	0.3
SHM Dimensions	$20 \times 20 \times 2(\mathrm{nm^3})$
FM Dimensions	$20 \times 20 \times 1(\mathrm{nm^3})$
Channel Width/Thickness	20 (nm) / 50 (nm)
High P/Low P	0.49/0.1

4. CACHE MEMORY DESIGN WITH MAGNETIC SKYRMIONS IN A LONG NANOTRACK

4.1 Introduction

Increased leakage current and process variations are major challenges to memory design with deeply-scaled CMOS devices. Several emerging technologies such as phase change memory (PCM), resistive random-access memory (RRAM), spin-transfer torque Magnetic RAM (STT-MRAM), and domain wall motion (DWM) based memory have been proposed as potential substitutes for presenting the desired attributes of today's memories: the speed of Static Random Access Memory (SRAM), the density of Dynamic RAM (DRAM), and the non-volatility of Flash. One such promising high-density memory technology, namely the DWM based racetrack memory, was proposed by IBM [67]. In a racetrack memory, multiple data bits can be coded in a sequence of magnetic domains, separated by domain walls, within a nanowire. DWM-based caches [43–46] have shown significant improvement in performance (with higher packing density and better energy efficiency) over other spintronic memory devices. However, the motion of domain walls might be pinned by the presence of defects [47], raising concerns about the feasibility of DWM-based memory. Magnetic skyrmions, on the other hand, have recently emerged as a promising alternative for future information carrier [48–51]. The presence of skyrmions can be observed in non-centrosymmetric bulk magnetic materials or ultra-thin magnetic systems with breaking inversion symmetry and large spin orbital coupling. The state of a magnetic skyrmion can be explained by the presence of Dzyaloshinskii-Moriya Interaction (DMI) [52, 53] – the DMI between two atomic spins S_1 and S_2 with a neighboring atom can be expressed as $H_{DM} = -D_{1,2} \cdot (S_1 \times S_2)$, where $D_{1,2}$ is the Dzyaloshinskii-Moriya (DM) vector [48, 49, 54–57]. Magnetic skyrmions (MS) have

been shown to possess several benefits over DWM based racetrack memory in terms of stability, density, and are less limited by imperfectness of the material. Specifically, its topological properties prevent the motion of skyrmions from being pinned at defect sites in a magnetic layer, and thus skyrmions are more robust information carriers. Magnetic skyrmions can be stored as multiple bits in a long nanotrack to realize highly dense memory. Chen *et al.* [68] first demonstrated the use of magnetic skyrmions to realize on-chip caches. The work proposed the use of a shift-based write mechanism [69] for skyrmion creation. However, such approach is considered to be applicable for DWM based device, while no experimental (or simulation) results to date have demonstrated the skyrmion creation using the shift-based mechanism. In our work, a magnetic skyrmion is written (or nucleated) by injecting a local spin-polarized current in the nanotrack, whereas the read operation is performed by sensing the change in resistance arising from the presence (or absence) of skyrmion at a specific location in the nanotrack. In order to read or write a bit stored in the nanotrack, a variable number of shift operations are required depending on its location relative to the read/write port. The noticeably high density and non-volatility offered by MS-based memory are key positives for last-level on-chip cache applications.

We explore the use of magnetic skyrmions as last-level on-chip caches in general purpose processors. We propose a multi-port skyrmion-based cell and evaluate its potential in realizing an on-chip memory array. Despite possessing a number of beneficial attributes such as high stability, non-volatility, high density[1], and low leakage, magnetic skyrmions pose certain challenges: (i) the current density required for skyrmion nucleation [70] is substantially high, necessitating the need for large access transistors for writing a skyrmion and in turn limiting the density benefits. (ii) the variable access latency arising from packing multiple bits in a single nanotrack, leads to energy and performance overheads. (iii) the motion of skyrmions drifts away from the direction of the electron flow owing to the Magnus force [71]. In order to relax

[1]The sizes of skyrmions and the spacing between them can be potentially shrunk down to the nanometer scale.

the skyrmions back to the center region of the nanotrack, an idle operation time is required which leads to additional shift latency. (iv) skyrmions might suffer annihilation through the edges in the presence of large drive current density required for high-speed operation. To address these challenges, we perform a design-space exploration for the multi-bit skyrmion cell while considering the peripheral circuits required to perform these operations. We also performed layout to estimate the density benefits of the proposed multi-bit cell. To keep skyrmions enclosed in the nanotrack under high current injection, it is essential to analyze a variety of possible design choices and their impacts on system energy and performance. We develop a device-circuit-architecture framework to understand these design points for the proposed multi-bit cell.

The key contributions of this work are as follows:

- We explore the feasibility of last-level cache design for general purpose processors with magnetic skyrmion-based memory.
- We propose a magnetic skyrmion-based multi-bit cell and utilize suitable circuit and architecture optimizations that mitigate the unique challenges posed by the skyrmion structure.
- We develop a systematic device-to-architecture co-design framework and perform an in-depth analysis of the density benefits, along with the energy and performance trade-offs associated with the proposed skyrmion-based cache. Our experiments on the PARSEC benchmark suite [72] demonstrate 2.41× improvement in cache energy with 2% average degradation in cache performance over an iso-area traditional SRAM-based L2 cache.

4.2 Skyrmion-based Memory

Fig. 4.1(a) shows the proposed MS-based device structure in which skyrmions are stored in a ferromagnetic nanotrack adjacent to an SHM. To realize a bit-cell using this structure, we need to perform three different operations: (i) a write operation, (ii)

a shift and an idle operation, and (iii) a read operation. In the following paragraphs, we describe these operations in detail along with the peripheral circuits required to perform these operations

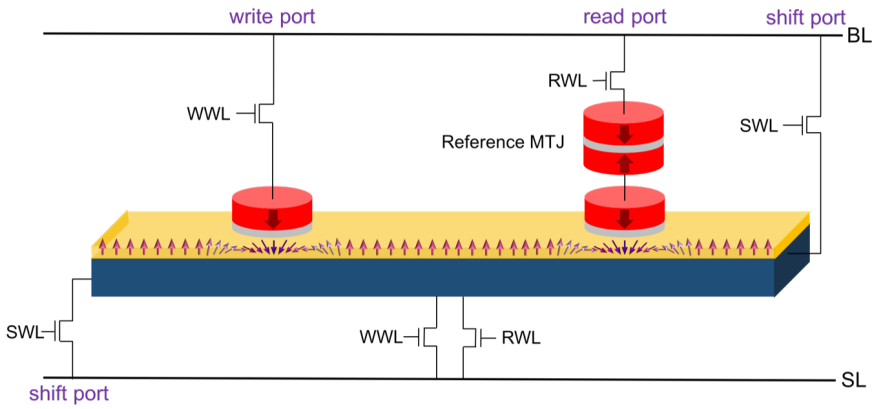

Fig. 4.1. **Schematic of MS-based device and bit-cell.** The proposed device structure can perform read/write/shift operations. A skyrmion can be nucleated in the nanotrack (yellow layer) by injecting a spin-polarized current through the left MTJ. The motion of skyrmions can be driven by utilizing vertical injection of a spin current generated from a charge current flowing through the Spin-Hall Metal (SHM) layer (blue layer). The presence of a skyrmion can be detected by sensing the voltage difference between the read MTJ and the reference MTJ

4.2.1 Nucleation of a skyrmion (Write operation)

A skyrmion is nucleated in the nanotrack by injecting a local spin-polarized current through the MTJ on the left (write MTJ). This is performed by turning ON the write access transistors, charging the bitline (BL) to V_{WRITE} and the sourceline (SL) to GND. Nucleating a skyrmion requires that the injected spin-polarized current exceeds certain threshold J_{th} [70]. We exploit spin-polarized current generated from the electrical current through a 20 nm-diameter write MTJ to create a skyrmion. The proposed device structure consists of a 0.4 nm-thick ferromagnetic nanotrack adjacent to a 3 nm-thick SHM. The material parameters used in our simulations correspond

to Co/Pt multilayers [70], and are shown in Table 6.1. In our simulation, a stable skyrmion can be nucleated in a 60 nm nanotrack by injecting a spin-polarized current through the write MTJ for 50 ps with a current density of 6.8×10^{12} A/m^2. Note that the presence of DMI necessitates the need of high current density for nucleation.

4.2.2 Motion of skyrmions (Shift operation)

Skyrmions are packed as multiple bits in a long nanotrack. Hence, in order to access a specific bit stored in a long nanotrack, the corresponding skyrmion need to be placed underneath the write (read) port via shift operations. A shift operation is accomplished by turning ON the access transistors of the shift ports and precharging BL and SL to appropriate voltage. The motion of skyrmions can be controlled by an in-plane current (CIP) flowing through the nanotrack, or by a current perpendicular to the plane (CPP). In the CPP case, since a skyrmion undergoes a larger Slonczewski in-plane torque instead of a smaller field-like out-of-plane torque, skyrmions driven by a charge current flowing through the SHM layer obtain higher velocities with lower current densities [70]. Thus, we elect to use the CPP scheme to drive the skyrmions in our proposed device structure are driven by a current perpendicular to the plane. The motion of skyrmions can be well explained by the Theile's equation [71],

$$G \times v_d - D\alpha v_d + j_{spin} = 0 \tag{4.1}$$

where j_{spin} represents the vertical spin current generated from the charge current flowing through the SHM underlayer (the blue layer in Fig. 4.1). The longitudinal and the transverse velocity can be written as

$$v_d^x = \frac{\alpha D}{G^2 + \alpha^2 D^2} j_{spin}, v_d^y = \frac{G}{G^2 + \alpha^2 D^2} j_{spin} \tag{4.2}$$

Hence, for $G \neq 0$, the motion of skyrmions deviates from the intended direction. The transverse motion of a skyrmion stops at a certain distance from the edge owing to the skyrmion-edge interaction. The final displacement with respect to the edge decreases as SHM current density increases. Skyrmions will be annihilated if the

applied charge current density is larger than a certain value, J_{ani}, which is function of the operation time. Fig. 4.2(a) shows that the critical annihilation current density can be significantly increased by reducing the shift operation time. Moreover, a high energy barrier is induced on the boundaries by adhering high-K materials at the edges, allowing skyrmions to be well confined in the nanotrack in the presence of large current injection [73]. Fig. 4.2(b) compares the critical annihilation current density (J_{ani}) for three different high-K materials (FePt, $Nd_2Fe_{14}B$, $SmCo_5$) with the edge width ranging from 1 nm to 5 nm for < 1 ns shift duration. The corresponding material parameters, adopted from [70, 74] are shown in Table 4.1. Introducing high-K materials at both edges makes switching the spin direction much harder when a skyrmion approaches the edge due to the Magnus force, thereby keeping the skyrmion in the nanotrack. The velocity of skyrmions, which increases with the current density, is therefore enhanced during shifts, leading to faster shift operations. Note that the induced energy barrier from the high-K materials is known to depend on the width of adhering high-K materials and its properties [73].

Table 4.1.
Comparison of high-K materials used in the present simulations

Material	FePt	$Nd_2Fe_{14}B$	$SmCo_5$
Anisotropy constant (MJ/m^3)	2.0	4.3	17.1
Exchange constant (pJ/m)	8	7.7	12
Saturation magnetization (MA/m)	1.1	1.28	0.84
DMI strength (mJ/m^2)	0.1	0.1	0.1

4.2.3 Detection of the presence of a skyrmion (Read operation)

Electrical detection of skyrmions at room temperature through the magnetoresistance effect have been proposed and experimentally demonstrated in recent works [75, 76]. In this work, we use this mechanism to perform the read operation. Specifically,

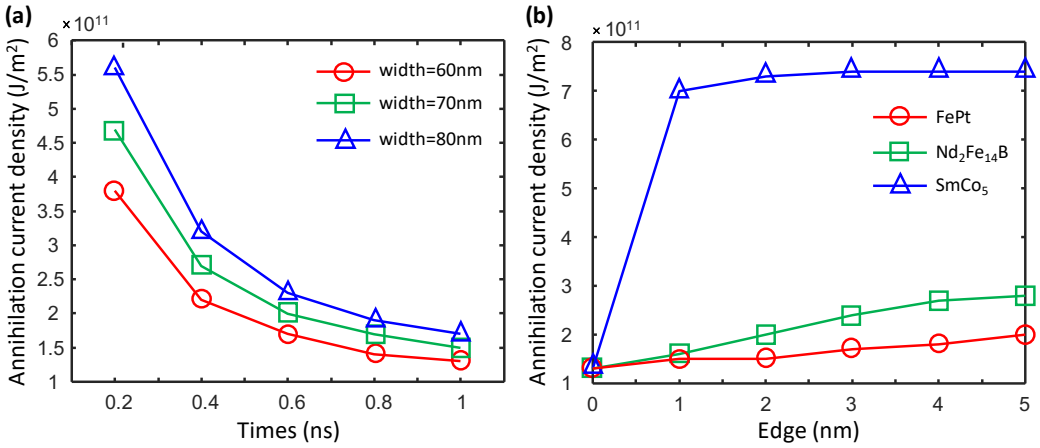

Fig. 4.2. (a) Critical annihilation current density versus various shift operation time. (b) Critical annihilation current density in a 1 ns shift operation versus various high-K materials with different width. The annihilation current density is less sensitive to the width of the high-K materials with higher anisotropy constant

we introduce a read port that includes a read MTJ and two access transistors. A read operation is performed by turning ON the two access transistors, driving BL to V_{READ} and SL to GND. The presence/absence of skyrmions alters the resistance of the read MTJ, and thus determines the magnitude of the current flowing from BL to SL. However, since the average magnetization of a skyrmion is not anti-parallel (m_z=-1) to the fixed layer (m_z=1), a smaller magnetoresistance change is obtained. To achieve sufficient resistance change for read operation, we use an MTJ of diameter 20 nm and $\sim$ 200 % magnetoresistance ratio. We also match the size of the skyrmion to the size of the read MTJ to ensure that the region captured by the read MTJ is closer to m_z=-1 (anti-parallel to the fixed layer), which in turn leads to a higher magnetoresistance change. Table 4.2 compares the voltage swing between the read and the reference MTJ by pulling up the BL voltage to 0.8 v and SL to GND. As shown in the table4.2, changing the width of the nanotrack increases the skyrmion dimension, which in turn leads to a greater voltage swing. However, this also increases

the required reliable spacing between consecutive skyrmions to free them from repulsive force between neighboring skyrmions [77]. Moreover, since the fixed layer of the read MTJ is located at the center region of the nanotrack, the deviation between the position of a skyrmion and a read port degrades the resistance change in the absence/presence of a skyrmion. Therefore, a read operation on a specific skyrmion bit requires an idle operation after each shift operation, *i.e.*, the total number of idle operations is equal to the number of shift operations required for reading a skyrmion bit. This operation relaxes the skyrmions back to the center region through edge repulsion. We achieve this by turning all access transistors OFF which stabilizes the magnetization of the nanotrack.

Table 4.2.

Read voltage difference between the read MTJ and the reference MTJ without the presence of a skyrmion for different nanotrack width and the corresponding skyrmion radius. The reliable spacing between consecutive skyrmions is also compared.

Width (nm)	Radius (nm)	ΔV (V)	Reliable spacing (nm)
60	13.5	0.125	$\sim$ 74
70	15.5	0.133	$\sim$ 77
80	18.5	0.137	$\sim$ 81

4.3 Multi-bit Skyrmion Cell Design

Fig. 4.3 shows the logical representation of data stored along the nanotrack. Depending on the existence of a skyrmion, different logic values can be stored along the nanotrack as multiple bits. The presence of a skyrmion denotes the logic "1", while its absence denotes the logic "0". A current injected into the SHM (the blue layer in Fig. 4.3) from the right can shift skyrmions to the right-hand side of the nanotrack, and vice versa. The logical views of a multi-bit cell with a single write/read port

and a cell with single write and multiple read ports are shown in Fig. 4.3(a) and (b), respectively. Note that the read ports can be placed at any location along the nanotrack, however, the write port is placed at the end of the long nanotrack to ensure simplicity for write operation. Consider Fig. 4.3(a) as an example. A write port at address "0x0" and a read MTJ at address "0x7" with a sequence of 0's and 1's stored in the cell is presented. In the first write cycle, "0" is written into the address "0x0", and subsequently shifted right to the next address "0x1". "1" is written into the address "0x0" during the next write cycle, and then the data in the nanotrack is again shifted to the right. By repeatedly writing data into the address "0x0" and subsequently right shifting all stored data to the next address, a sequence of bits can be written to the nanotrack. To read the stored data at address "0x5", for example, the bit is shifted right by two positions to reach the location under the read MTJ. Similarly, to write data at a specific address, we first shift the bit to the position where the write port is located. Before writing a new data into the address, the previously stored data is cleared by injecting a current with spin polarization in the opposite direction to the magnetization of the skyrmion center. To prevent stored data in the nanotrack overflowing during shift operations, we extend the nanotrack by having extra data bits (the light yellow part in Fig. 4.3). In the worst-case scenario for this example, to access the stored data at address "0x0", the bit is required to be shifted right by seven positions. Thus, seven extra bits are required to avoid the loss of stored data from address "0x1" to "0x7". The write/read latency is dependent on the location where a bit is stored. However, the average read latency can be alleviated by introducing multiple read ports, as shown in Fig. 4.3(b). The current location of the read port is referred to as the current port status. In order to access a bit from a multi-bit cell, a shift controller determines the appropriate read port and calculates the number of shift operations required by comparing the input address bits with the current port status. This results in a reduction of the number of extra bits required to avoid data loss. Table 4.3 lists the bias voltage conditions for write/shift/read/clear/idle operations.

Table 4.3.
Bias voltage conditions for various operations

	RWL	WWL	SWL	BL	SL
Read	V_{DD}	0	0	V_{READ}	0
Shift Left	0	0	V_{DD}	0	V_{SHIFT}
Shift Right	0	0	V_{DD}	V_{SHIFT}	0
Write	0	V_{DD}	0	V_{WRITE}	0
Clear	0	V_{DD}	0	0	V_{WRITE}
Idle	0	0	0	0	0

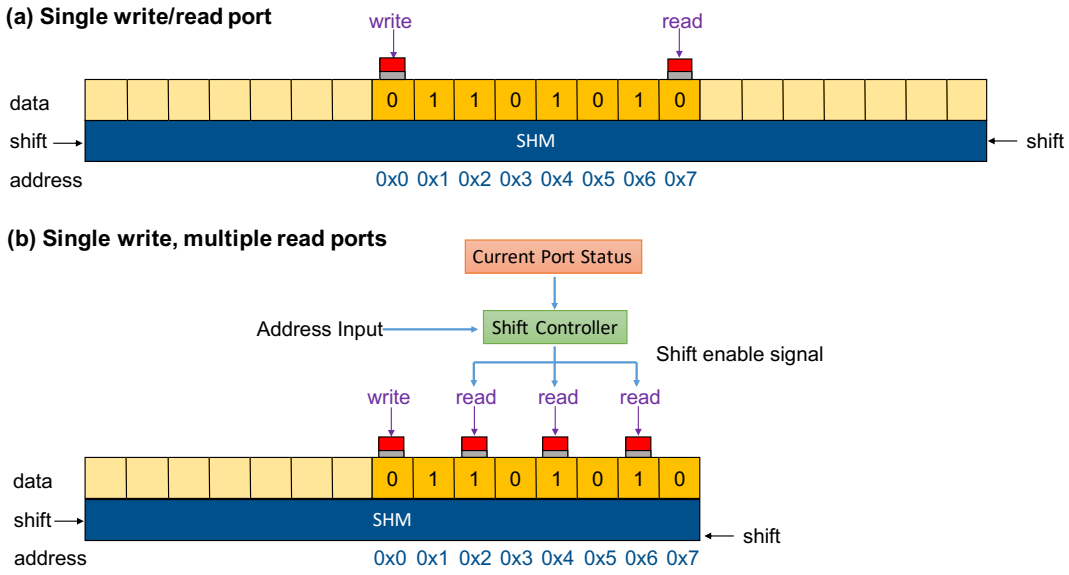

Fig. 4.3. Logical view of a multi-bit MS-based cell with (a) single write/read port or (b) single write and multiple read ports. A sequence of bits are stored in the nanotrack.

Density of the skyrmion based multi-bit MS cell

Fig. 4.4 shows the layout of an 8/16/32-bit MS cell with a single write/read port. As discussed in Section 4.2, the current required for the write operation is considerably

higher than that for the read and shift operations. Hence, as shown in Fig. 4.4(a), for an 8-bit MS cell with single write/read port, the cell area is dominated by the peripheral write transistors since the dimension of the write transistors are much larger than the nanotrack. Note that the length of the nanotrack is determined by the number of stored bits and the read ports. The total length of the nanotrack can be reduced by having multiple read/write ports as fewer extra bits are required to prevent the stored data from being destroyed during shift operations (light yellow part in the Fig. 4.3). For the 8-bit MS cell case with a single read port, the write transistors dominate the total cell area. Thus, the density (i.e., cell area per bit) of the 8-bit MS cell cannot be further improved by introducing more read ports (as presented in Fig. 4.5). On the other hand, for the 16/32-bit MS cell, since the nanotrack dominates the cell area with one read port, the density can be improved by packing more bits within a smaller area. Hence, as shown in Fig. 4.5, at the 45nm technology node (F), for an 16-bit cell with one/two read ports, the CMOS transistors require 135.59 F^2/bit and 112.67 F^2/bit, respectively. Fig. 4.5 compares the cell size of the proposed MS cells, i.e., 8/16-bit MS cells with one write port and 32-bit MS cell with both a single write port and two write ports, while varying the number of read ports. We also show the cell size of SRAM (triangle) and 1T-1R STT-MRAM (star) in Fig. 4.5 for reference. The total area of a multi-bit MS cell is determined by the number of read and write transistors, as well as the length of the nanotrack. For an 8-bit MS cell, the cell size is dominated by the write transistors when the total number of read ports is less than 3. Although having more read ports beyond 3 shortens the nanotrack with fewer extra bits required, the area of read peripheral transistors inevitably increases too. Similarly, for the 16-bit MS cell case with less than 5 read ports, the bit-cell area is mainly determined by the nanotrack itself. In the case of 32-bit MS cell, the multi-bit cell area is further dominated by the nanotrack; therefore, having an extra write port helps reduce the 15 extra bits required, thereby improving the density (i.e., cell area per bit).

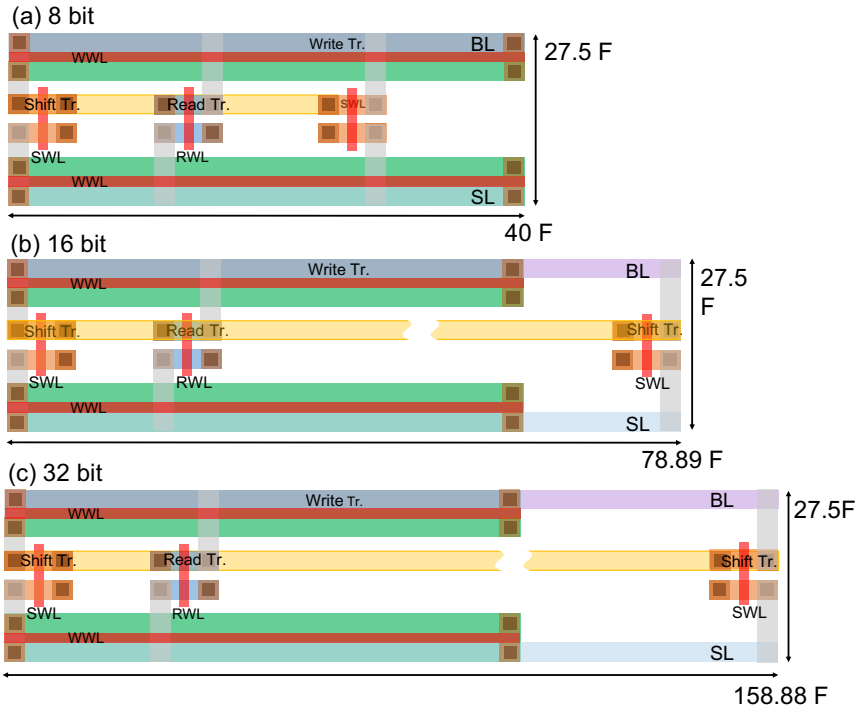

Fig. 4.4. Layout of a 8/16/32-bit MS cell with single write/read port at the 45nm technology node (F)

4.4 Array Organization

Fig. 4.6 shows the memory array organization with the proposed multi-bit cell. The wordlines for performing read, write and shift operations (i.e., RWLs, WWLs, SWLs) are shared among all the multi-bit cells placed in a row. The BL and the SL are shared among all the multi-bit cells placed in a column. In this architecture, multiple words can be placed on the same row and accessed independently. The address decoder is used to select a multi-bit skyrmion cell in the array, with the shift control logic selecting the appropriate word. Note that the sense amplifier shared across the entire column detects the output signal as logic '0' or '1'.

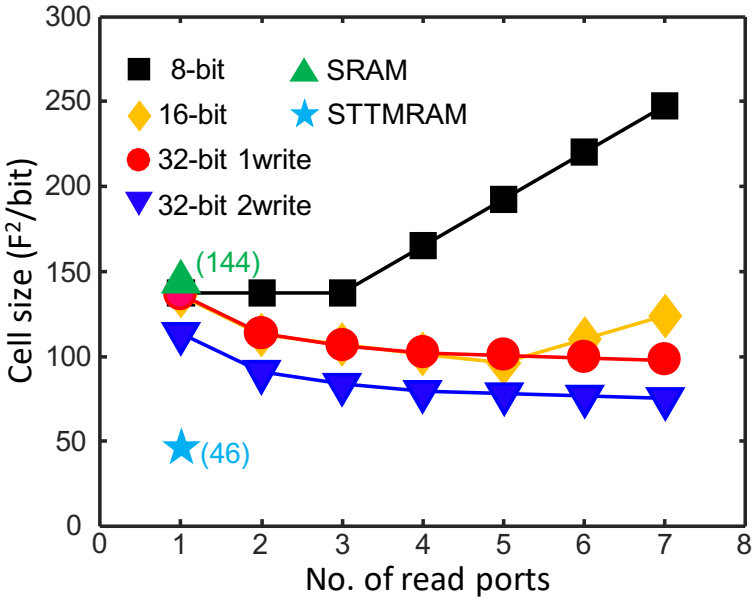

Fig. 4.5. Bit-cell area comparison for different multi-bit designs

Skyrmion-based cache design

To evaluate the benefits of the proposed memory array at the application level, we integrate it as a last-level cache in the memory hierarchy of a general purpose processor. Towards this end, we follow the DWM-based hybrid cache organization presented in TapeCache [43], i.e., the tag array is designed with SRAM to avoid variable access latency during performance-critical tag lookup operations, and the data array is realized using the proposed multi-bit skyrmion array. The data array is further composed of randomly addressable clusters, each of which stores multiple cache blocks. We assume a bit-interleaved mapping of the cache blocks in each cluster, such that a given cache block can be accessed in parallel after performing an appropriate number of shift operations to all the nanotracks within a cluster. The addressing policy and the cache management policies are also assumed to be similar to that of TapeCache.

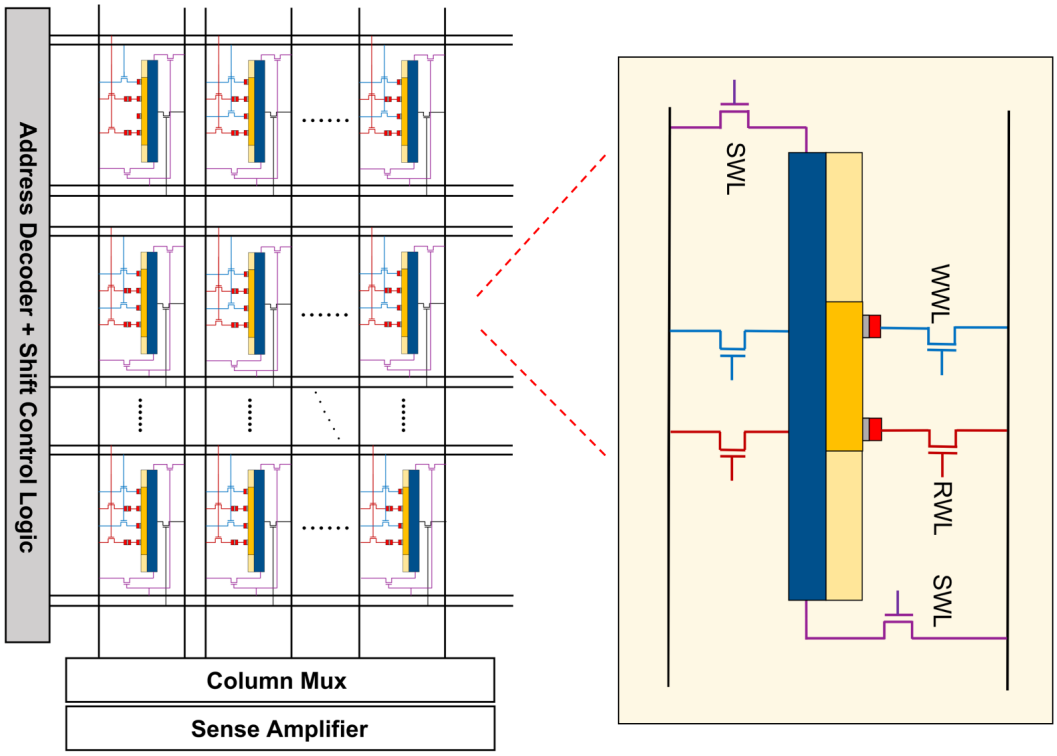

Fig. 4.6. The memory array organization of skyrmion based multi-bit cells

4.5 Experimental Methodology

In this section, we present a brief description of the simulation framework and present the experimental setup used to evaluate our proposal.

4.5.1 Simulation framework

Micromagnetic simulations of the skyrmion device are performed using the tool Mumax3 [78, 79]. The magnetization dynamics of magnetic skyrmions driven by vertical current can be expressed by

$$\begin{aligned} \tau &= \frac{\gamma}{1+\alpha^2}(m \times H_{eff} + \alpha(m \times (m \times H_{eff}))) + \tau_{SL} \\ \tau_{SL} &= \beta\frac{\epsilon - \alpha\epsilon\prime}{1+\alpha^2}(m \times (m_p \times m)) - \beta\frac{\epsilon\prime - \alpha\epsilon}{1+\alpha^2} m \times m_p \\ \beta &= \frac{j_z \hbar}{M_{sat} e d} \\ \epsilon &= \frac{P\Lambda^2}{(\Lambda^2+1) + (\Lambda^2-1)(m \cdot m_p)} \end{aligned} \tag{4.3}$$

where m is the normalized magnetization vector; m_p is the fixed layer polarization; γ is the Gilbert gyromagnetic ratio; α is the Gilbert damping parameter; H_{eff} is the effective field; j_z is the current density along the z axis; M_{sat} is the saturation magnetization; e is the elementary charge; d is the skyrmion layer thickness; P is the polarization of conduction electron; the Slonczewski Λ parameter characterizes the spacer layer; and $\epsilon\prime$ is the secondary spin transfer term. The material parameters used in our simulations correspond to Co/Pt multilayers [80], and are shown in Table 6.1. We consider a 0.4 nm thick Co nanotrack with perpendicular magnetic anisotropy on a 3 nm Pt substrate inducing DMI. The sample is discretized into an element size of $1 \times 1 \times 0.4$ nm^3. The Non-equilibrium Green's Function (NEGF) based spin transport simulation has been used in order to obtain the resistance of the MTJ [16]. The charge current (I_e) flowing through the SHM and the corresponding spin current (I_s) are calculated using [81]

$$I_s = \theta_{sh}\frac{A_{MTJ}}{A_{SHM}} I_e \tag{4.4}$$

where A_{MTJ} and A_{SHM} are the cross sectional areas of the MTJ and SHM, respectively, and θ_{sh} is the spin-Hall angle. The spin current from Eq.(4) is used to analyze the magnetization dynamics with the generalized LLGS equation. Magnetization dynamics simulations are performed using the Mumax3 platform [78, 79].

Table 4.4.
Material parameters used for simulation

Parameter	**Value**
Saturation magnetization (M_{sat})	580 kA/m
PMA anisotropy constant (K_u)	0.8 MJ/m^3
Exchange constant (A)	15 pJ/m
Dzyaloshinskii-Moriya interaction (DMI) strength (D)	3 mJ/m^2
Gilbert damping constant (α)	0.1
Spin polarization (P)	0.4
Spin Hall Angle (θ_{sh})	0.07
Nanotrack width and thickness	60 nm × 0.4 nm
SHM width and thickness	60 nm × 3 nm
MTJ diameter	20 nm

4.5.2 Experimental setup

The device parameters obtained with the proposed simulation framework are used as technology parameters in a modified version of CACTI [82] to evaluate the read/write characteristics of the skyrmion-based cache. The array-level characteristics are then reflected in the GEM5 architectural simulator that models the skyrmion cache architecture to evaluate the proposed design as an L2 cache [83]. In our experiments, we perform an iso-area replacement of L2 cache and compare the energy and performance of the proposed design with that of SRAM-based and STT-MRAM-based caches. All the memory technologies considered in the evaluation are based on a 45 nm technology node. The CMOS baseline system configuration used in our analysis is shown in Table 4.5. We perform a full-system simulation for one billion instructions in the regions of interest for caches across a suite of multi-threaded benchmarks from PARSEC [72].

Table 4.5.
System configuration

Processor Core	Alpha, out-of-order processor, 4 cores at 2 GHz
L1 I/D-cache	16KB per core, 2 way-set associative, 64B line size
L2 unified cache	2MB shared, 16 way-set associative, 64B line size
Cache latency	L1 cache: 2-cycle, L2 cache: 9-cycle

4.6 Results and Discussions

4.6.1 Device and circuit level results

As we discussed in Section 4.3, shift operations are involved in both write and read operations. However, during shift operations, the trajectory of skyrmions in the nanotrack bends away from the center as a result of Magnus force. Thus, an idle operation is required to relax skyrmions back to the center region through edge repulsion after every shift operation. Fig. 4.7 compares the required relaxation time and the longitudinal shift distance within various operation times for a current of 1.44×10^{-5} A and 5.76×10^{-5} A, respectively, with and without adhering high-K materials at both edges. High-K materials are adhered to prevent skyrmions from annihilation under a current of 5.76×10^{-5} A. The longitudinal velocity is proportional to the injection current density, and the required relaxation time is related to the transverse shift distance, which increases with both the drive current and the operation time. Since skyrmions in the nanotrack stop at a certain distance to the edge owing to the skyrmion-edge interaction, the required relaxation time is the same after 1.2 ns and 0.8 ns operation time under a current of 1.44×10^{-5} A and 5.76×10^{-5} A, respectively. With the aid of adhering high-K materials, skyrmions can be operated under a higher current density, and thus higher transverse velocity can be reached. However, a higher relaxation time is also required which increases the shift latency. Since the reliable spacing in our case is ~ 74 nm, a current of 1.44×10^{-5} A

for 1 ns and 5.76×10^{-5} A for 0.2 ns is required during the shift operation, leading to a 0.9 ns and 1.3 ns relaxation time, respectively. We compare the performance and energy consumption for 8/16/32 bit-MS with either high-K materials at both edges (for a current of 5.76×10^{-5} A) or the absence of high-K materials (for a current of 1.44×10^{-5} A).

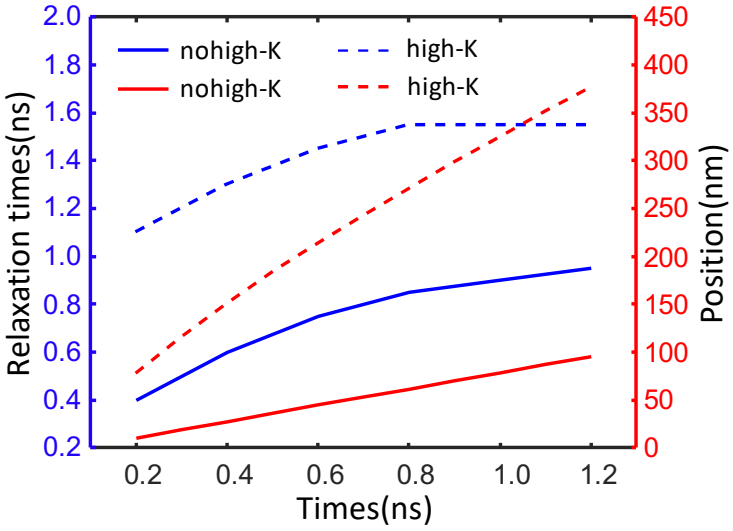

Fig. 4.7. Relaxation times and final position comparison under a current of 1.44×10^{-5} A and 5.76×10^{-5} A. High-K materials are adhered under the current of 5.76×10^{-5} A to avoid skyrmions from annihilation from edges.

4.6.2 System level results

In this section we present the array-level analysis of the proposed MS-based cache design and then evaluate the impact on system performance and cache energy.

Array level energy

Fig. 4.8 compares different energy components of the proposed MS-based array with an STT-MRAM array. The MS-based array is realized using an 8-bit multi-bit

cell. As shown in the figure, the write energy for the MS-based array is 1.78× higher than STT-MRAM. This is because of the high current requirements for skyrmion nucleation, resulting in wider access transistor for the bit-cell which in turn leads to higher capacitance. On the other hand, the read energy is identical for both STT-MRAM and the proposed MS-based array due to similar read mechanisms. Apart from the read and write energies, the MS array also consumes shift energy during read/write operations. The shift energy is a function of the length of the nanotrack, the number of read/write ports in the bit-cell, and the high-K material on the edges of the nanotrack. In our evaluations, the shift energy per operation was found to be 9.51×10^{-4} and 8.68×10^{-5} pJ for the 8-bit based memory array with and without high-k material on the nanotrack edges, respectively.

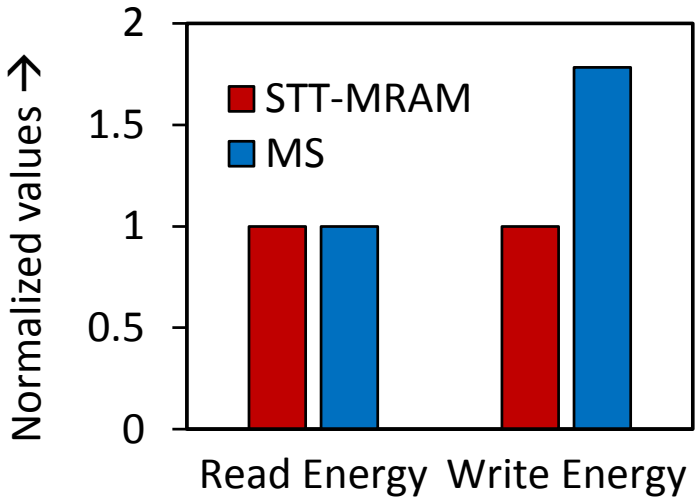

Fig. 4.8. Array-level comparison of read and write energies with iso-area STT-MRAM

Performance evaluation

Fig. 4.9 compares the instructions per cycle (IPC) for six different skyrmion-based cache configurations with SRAM and STT-MRAM caches. We consider eight different L2 cache designs under iso-area conditions: (1) a 2MB SRAM cache with 1 read/write port, (2) an 8MB STT-MRAM cache with 1 read/write port, (3) two

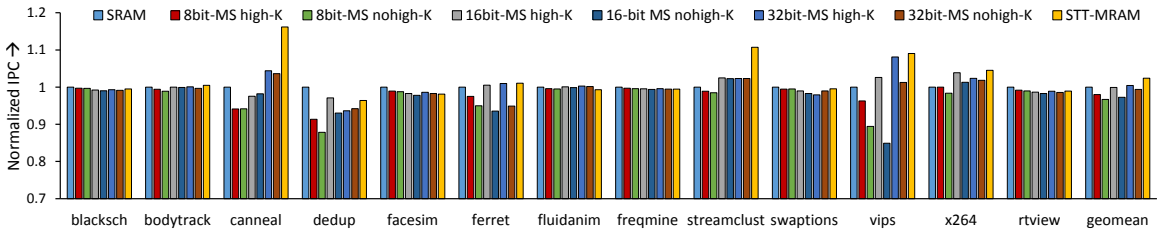

Fig. 4.9. L2 cache performance comparison across different memory technologies

2MB MS-based cache designs with 3 read ports and 1 write port, storing 8 bits in the nanotrack, with and (4) without high-K material at both edges (8bit-MS high-K and 8bit-MS nohigh-K), (5) two 4MB MS-based cache designs with 3 read ports and 1 write port, storing 16 bits in the nanotrack, with and (6) without high-k material at both edges (16bit-MS high-K and 16bit-MS nohigh-K), and (7) two 4MB MS-based cache designs with 8 read ports and 2 write ports, storing 32 bits in the nanotrack with and (8) without high-K material at both edges (32bit-MS high-K and 32bit-MS nohigh-K). The IPC is normalized to the 2MB SRAM-based cache design. Across all benchmarks, the 8bit-MS high-K design leads to an average degradation of 2.0% and 4.3% in performance compared to the SRAM and STT-MRAM designs. This degradation is primarily due to two factors: (a) reduced cache capacity (iso-capacity with respect to SRAM and 0.25× capacity with respect to STT-MRAM), and (b) shift overhead arising from the memory structure. In contrast, for the 8bit-MS nohigh-K design, the system performance further reduces by 3.3% and 5.6% compared to the SRAM and the STT-MRAM cache as a result of additional shift latency incurred for each cache access in the absence of high-K material at the edges.

On the other hand, the two 16-bit MS configurations (16bit-MS high-K and 16bit-MS nohigh-K) degrade the performance by 0.1% and 2.7% compared to the SRAM cache on average, respectively. Furthermore, we observe a 2.4% and 5.0% reduction in performance for the two designs when compared with the STT-MRAM cache design. The smaller degradation in the performance over the SRAM and STT-MRAM cache

is attributed to the 2× higher cache capacity offered by the 16-bit configuration. For the 32bit-MS designs, the performance improves by 0.4% with the high-K cache design, and degrades by 0.6% for the design with no high-K material, over the SRAM cache. This improvement is mainly because of a reduced number of shift operations performed on average with the aid of higher number of read and write ports in the 32-bit design. Note that the performance reduces by 0.6% and 3.0%, respectively, over the STT-MRAM cache design, since the overall cache capacity does not increase with the 32bit-MS design as discussed earlier.

Energy comparison

Fig. 4.10 compares the L2 cache energy consumption between our proposed cache designs and the iso-area SRAM and STT-MRAM caches. The cache energy is normalized to the energy consumed by the STT-MRAM design. On average, we observe a 2.41× and 2.45× reduction in cache energy for the 8bit-MS high-K and 8bit-MS nohigh-K designs over the SRAM cache, attributed to the reduced leakage energy consumption with non-volatile magnetic skyrmions. The energy benefits are slightly higher for the 8bit-MS nohigh-K design because of lower shift energy consumed with no high-K material on the nanotrack edges. For the 16-bit designs, the energy benefits were found to be 2.37× and 2.41× for the 16bit-MS high-K and 16bit-MS nohigh-K designs, respectively. The energy benefits are moderately lower for the 16-bit configurations over the 8-bit designs due to a higher energy consumed by the shift operations. Note that for a subset of benchmarks (canneal, ferret, streamclust and vips), the 16-bit configurations have a lower energy compared to the 8-bit designs. This is because of the lower capacity misses observed in the 16-bit designs that eventually leads to lower write energy. In these benchmarks, the benefits in write energy outweigh the increase in shift energy, thereby leading to improved cache energy. The energy benefits over SRAM reduce to 2.27× and 2.31× in the 32bit-MS high-K and 32bit-MS nohigh-K designs, respectively. The benefits in energy are lower than the other two

designs (8-bit and 16-bit configurations) since the resistance offered by the nanotrack increases, which in turn increases the energy consumed for each shift operation.

In contrast, the energy consumed by the MS-based cache designs is higher than the baseline iso-area STT-MRAM cache in all cases. Specifically, the 8-bit and 16-bit designs consume 1.29× and 1.27×, 1.30× and 1.28× higher energy than the STT-MRAM cache. Similarly, the 32-bit designs consume 1.37× and 1.34× energy over the STT-MRAM cache. This increase in energy is because of the additional shift energy overheads and the reduced cache capacity arising from the larger write transistor requirements for the multi-bit MS cell.

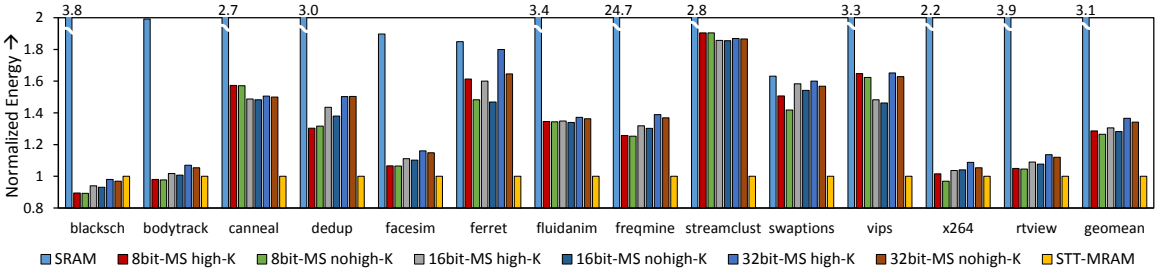

Fig. 4.10. Energy trends across different memory technologies

In summary, our results show that skyrmion-based caches offer small improvements in performance with substantial energy reduction over an iso-area SRAM-based cache. They also point to key avenues for improvement in skyrmion-based memory—specifically, the high nucleation energy for skyrmions leads to large write transistors and curtails density benefits, while the latency due to shift operations limits performance.

4.7 Conclusion

In this chapter, we explored magnetic skyrmions to design last-level caches. We propose a multi-bit skyrmion-based cell design that packs multiple bits in a nanotrack. Since the size and spacing of skyrmions can be down to nanometer scale, the

skyrmion-based nanotrack has the potential to provide significant density benefits compared to other memory technologies. However, the high current requirements for skyrmion nucleation is a bottleneck for achieving significant density benefits. We analyzed different device parameters and design tradeoffs associated with the proposed bit-cell and evaluated the area, performance, and energy benefits while accounting for the peripheral circuit requirements. We designed a device-circuit-architecture framework to evaluate the system-level benefits of the proposed design. Our experiments reveal considerable benefits over an iso-area SRAM cache. However, the energy and performance is lower than an iso-area STT-MRAM cache, suggesting the need for mechanisms to lower the current density requirements for skyrmion nucleation.

5. MAGNETIC SKYRMION FOR SPINTRONIC DEEP LEARNING SPIKING NEURON PROCESSOR

5.1 Introduction

Deep neural networks, inspired from neuronal arrangements of cortical visual information processing networks, are currently one of the most popular architectures for natural image classification, video and speech analysis tasks [84]. To overcome the high energy consumption of deep networks, low-power event-driven sparse communication and processing enabled Spiking Neural Networks (SNNs) have gained great interests as an alternative computational paradigm [85]. However, since a direct mapping to operations involved in neural and synaptic computations is missing, CMOS implementations of such SNNs entail significantly high power consumption and area overhead. Besides, energy consumption caused by synaptic memory access for each processed spike further exacerbates the situation. As a result, emerging post-CMOS technologies such as multilevel Ag-Si memristors [86], Phase Change Memories (PCM) [87, 88] and multi-layer spintronic devices [89, 90] that can mimic the functionalities of such "non-von Neumann" computing platforms are currently under exploration. Researchers have recently proposed "All-Spin" neuromorphic systems based on magnetic structures with domain wall as a neuronal and synaptic unit that can be manipulated by ultra-low voltage and can be switched by ultra-low currents [91, 92]. However, the motion of domain walls might be pinned by the presence of defects, raising concerns about the feasibility of DWM-based neurons or synapses [47]. Magnetic skyrmion, a particle-like chiral spin texture with reversed magnetization in nanomagnets, possesses several benefits over DWM-based devices in terms of high stability, ultra-small size, and is less limited by imperfectness of the material [93, 94]. Specifically, topological properties prevent the motion of skyrmions

from being pinned at defect sites in a magnetic layer, and thus skyrmions are considered as more robust information carriers [48, 95]. In this chapter, we explore the possibility of a Spiking Neural Network design where the core functionalities can be emulated by skyrmion-based neurons/synapses, a ferromagnet (FM) - spin Hall metal (SHM) multilayer structure. We also present design schemes to alleviate one of the potential limitations of skyrmionic synaptic devices—low resistance ratio between the two extreme stable resistive states of the device. In order to increase the range of the synaptic plasticity, a skyrmion-based synapse is designed to have multiple branches with various conductance ranges. The synaptic weight of our proposed device can be modulated by the presence/absence of a skyrmion under the sensing magnetic tunneling junction (MTJ) of each branch. The behavior of an integrate-and-fire (IF) neuron is described by the skyrmion motion dynamics under a vertical injection of a spin current.

There have been prior proposals of neuromorphic devices based on skyrmion motion. For instance, Huang *et al.* [96] considered a skyrmion-based synaptic device structure where multiple skyrmions are used to modulate the device conductance. Nevertheless, unlike our proposal, they do not address the issue of limited conductance range available in such devices. He *et al.* [97] presented a tunable-threshold artificial neuron based on voltage controlled pinning phenomenon, and introduced a Skyrmion Neuron Cluster (SNC) design by connecting a total number of N skyrmion neurons in parallel to implement various neuron activation functions. However, achieving better recognition accuracy requires larger SNC size, thereby increasing the total energy consumption. Moreover, the work is based on non-spiking neural network designs, and therefore do not pertain to our focus on the implementation of deep spiking neural systems. Li *et al.* [98] proposed a neuronal device structure with varying anisotropy along the nanotrack to mimic a Leaky-Integrate-Fire Spiking Neuron. In this proposal, possessing such a variable anisotropy along the length of the device could be potentially difficult from the fabrication perspective. In contrast to these aforementioned findings, we propose simple device structures based on experiments to

implement Integrate-Fire Spiking Neurons and demonstrate high accuracy digit recognition systems based on such skyrmionic neurons. Furthermore, prior works either focus solely on the neuron or the synapse, while a holistic design of a skyrmion-based neuromorphic system is still missing. In this chapter, we provide a device, circuit, and system level analysis for the design of an All-Spin Spiking Neural Network based on skyrmionic devices and demonstrate its efficiency over a corresponding CMOS implementation. The key contributions of this work are as follows:

- We propose a Deep Spiking Neural Network architecture where the underlying computation units can be realized by spintronic devices. An ultra-low power "in-memory" neural computing platform can be realized by using spintronic synaptic crossbar arrays interfaced with spintronic neurons.
- We develop a "top-down" (algorithm) and "bottom-up" (device modeling) simulation framework to evaluate the performance of such a spintronic neural processor.

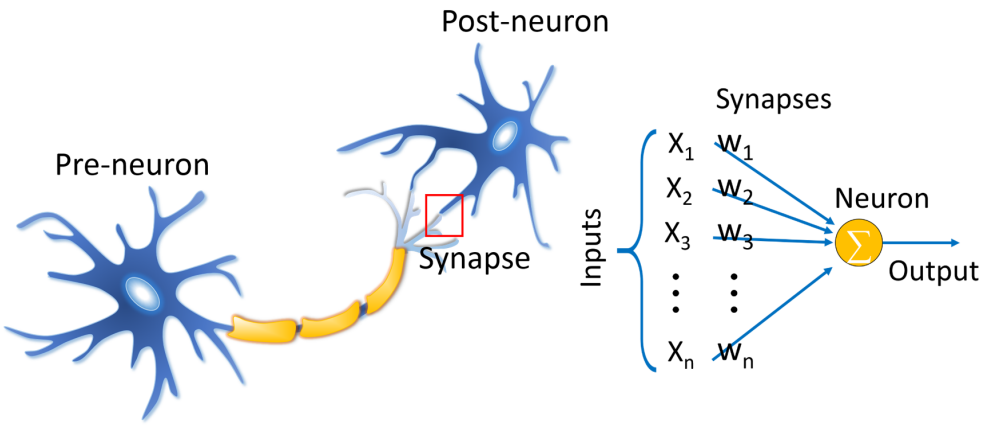

Fig. 5.1. Illustration of a biological neuron network. A biological pre-neuron receives, processes and transmits information to a post-neuron via synapse.

5.2 Deep Spiking Neural Networks

Neurons and synapses are two essential and fundamental computational units of Deep Spiking Neural Networks. As shown in Fig. 5.1, the neuron transmitting signals is known as the pre-neuron and the one receiving signals is the post-neuron. Spikes (information) are transferred from one neuron to another via junctions termed as synapse. Weighted synaptic summation of inputs from pre-neurons is received by a post-neuron. In this work, the integrate-and-fire (IF) spiking neuron computing model is used owing to its popularity in training and implementing Deep Spiking Neural Nets [99]. The skyrmion motion dynamics in a nanotrack, driven by a charge current, can well perform the functionalities of an IF neuron. The evolution of the membrane potential v_{mem} depending on the timing of its spiking inputs can be expressed as

$$\frac{dv_{mem}(t)}{dt} = \sum_{i} \sum_{t_f,i} w_i \cdot \delta(t - t_{f,i}) \tag{5.1}$$

where $t_{f,i}$ is the spike timing of the i_{th} input pre-neuron; w_i is the corresponding weight of the incoming synapse; and $\delta(\cdot)$ is the delta function. Post-neurons generate an output spike once the membrane potential crosses the spiking threshold (v_{th}) and get reset to a reset potential (v_{res}).

Currently, Deep Learning architectures are utilized to implement complex pattern recognition problems. Such networks, inspired from neural networks observed in the biological nervous system, is composed of alternating layers of convolution (C) and spatial subsampling(S) operations. It is worth noting here that both these operations can be mapped to a dot product operation between the inputs and the synaptic weights for a particular layer. The output dot product of both operations are then processed by the spiking neurons. Readers are referred to Ref. [99] for details on the training mechanism of Deep Spiking Neural Networks being considered in this work. The hardware mapping of such networks will be performed by spintronic devices which will be described in the following section.

5.3 Underlying Physics of the Proposed Spintronic Device

Magnetic skyrmions can be observed in ultra-thin magnetic systems with breaking inversion symmetry and large spin orbital coupling, and the state of a magnetic skyrmion can be explained by the presence of Dzyaloshinskii-Moriya Interaction (DMI) [52, 53] – the DMI between two atomic spins S_1 and S_2 with a neighboring atom [48, 49, 54–57] can be expressed as $H_{DM} = -D_{1,2} \cdot (S_1 \times S_2)$ where $D_{1,2}$ is the Dzyaloshinskii-Moriya (DM) vector. Here, we propose a three-terminal skyrmion-based synapse/neuron with decoupled programming-current (write) and spike-transmission (read) paths as a basic building block for the Deep Spiking Neural Network. The operation principles and underlying physics are explained as follows.

Due to the topologically protected property, a skyrmion cannot be created by continuous deformation from FM state or other magnetic state, and thus a localized spin-polarized current is required to create the singularity [70]. Note that only one-time nucleation operation is required owing to its nonvolatile properties. Detecting operation is performed by sensing the change in resistance arising from the presence/absence of skyrmion at a specific location in the nanotrack [100]. However, since the average magnetization of a skyrmion is not anti-parallel (m_z=-1) to the fixed layer (m_z=1) of the MTJ, a smaller magnetoresistance change is obtained. The change is directly proportional to the diameter of the skyrmion, and is inversely proportional to the cross-sectional area of the MTJ. Therefore, if the average magnetization of the region captured by the detecting MTJ is closer to m_z=-1 (anti-parallel to the fixed layer), a higher magnetoresistance change could be achieved. The skyrmions in our proposed device structure can be displaced by a charge current flowing through the SHM. Note that the motion of skyrmions can be controlled by a current in-plane (CIP) flowing through the nanotrack, or by a current perpendicular to the plane (CPP) [24, 26]. In the CPP case, since a skyrmion undergoes a larger Slonczewski in-plane torque instead of a smaller field-like out-of-plane torque, skyrmions driven by a charge current flowing through the SHM layer obtain higher velocities with lower cur-

rent densities. Moreover, in the CPP case, we assume that most of the input charge current flow through the SHM underlayer, since the resistance of the FM is around an order higher than that of the SHM. The motion of skyrmions in the nanotrack can be well explained by the Theile's equation

$$G \times v_d - D\alpha v_d + j_{spin} = 0 \tag{5.2}$$

where j_{spin} represents the vertical spin current generated from the charge current passing through the SHM underlayer. The motion of skyrmion drifts away from the center region owing to the Magnus force [71] and moves toward either the upper or the lower edge depending on the direction of a charge current passing through the SHM. Owning to the increasing skyrmion-edge interaction while approaching the boundary, a skyrmion stops at a certain distance from edge. The final displacement with respect to the edge decreases as the SHM current density increases, and thus a skyrmion is annihilated from boundaries if the drive current exceeds a certain value. Here, a higher energy barrier is induced by attaching high-K materials at boundaries to well-confine skyrmions in the nanotrack under a high level current injection, leading to a higher speed. The magnetization dynamics of magnetic skyrmions driven by vertical current can be expressed by

$$\begin{aligned}
\tau &= \frac{\gamma}{1+\alpha^2}(m \times H_{eff} + \alpha(m \times (m \times H_{eff}))) + \tau_{SL} \\
\tau_{SL} &= \beta\frac{\epsilon - \alpha\epsilon\prime}{1+\alpha^2}(m \times (m_p \times m)) - \beta\frac{\epsilon\prime - \alpha\epsilon}{1+\alpha^2} m \times m_p \\
\beta &= \frac{j_z \hbar}{M_{sat} e d} \\
\epsilon &= \frac{P\Lambda^2}{(\Lambda^2+1) + (\Lambda^2-1)(m \cdot m_p)}
\end{aligned} \tag{5.3}$$

5.4 Proposed Device as Synapse and Neuron

5.4.1 Skyrmion-based synapse

Synapses are junctions to transmit weighted signals from a pre-neuron to a post-neuron. The functionality of synapses can be performed by programmable resistors.

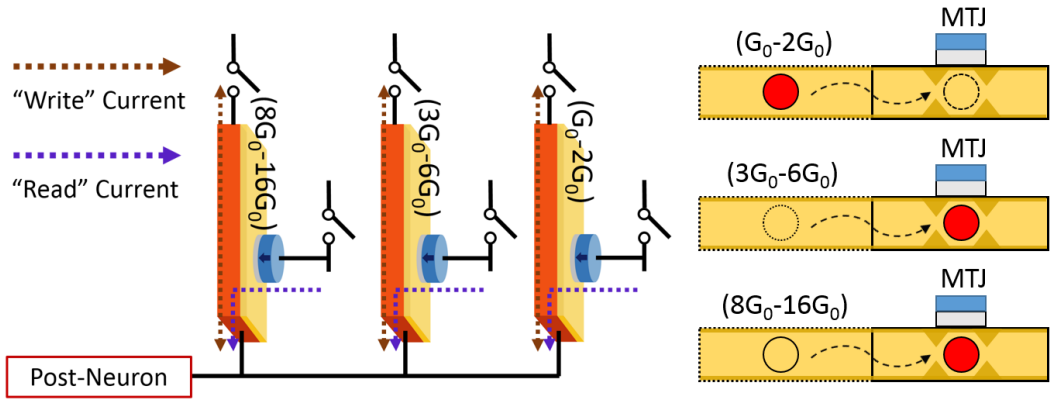

Fig. 5.2. Skyrmion-based synaptic device structure and the corresponding schematic. A vertical injection of a spin-polarized current is utilized to drive a skyrmion in the nanotrack. Notches with high-K material are attached at both sides of the read position to prevent skyrmion displacement. Besides, high-K materials are attached at both upper and lower boundaries to avoid skyrmion annihilation from edges.

Given a particular operating voltage, the amount of current flowing through the device is weighted by the device conductance, thereby emulating the synaptic scaling operation. Fig 5.2 shows the device structure of our proposed synapse, a skyrmion is initially nucleated at the left side and a read MTJ is located at the right side. The magnitude of the "write" current passing through the SHM (between terminals T2 and T3) determines the final position of the skyrmion, while the magnitude of the "read" current (between terminals T1 and T3) is modulated by the conductance of the read MTJ. The synaptic weights can be appropriately set by shifting or not shifting the skyrmion to the position underneath the sensing MTJ. Here, the FM layer at the read region serves as the "free layer" of the read MTJ, and the conductance of the read MTJ is denoted as G_{sk} (G_p) with the presence (absence) of a skyrmion under the read position. Considering just a single device, the range of the two conductance states $|G_{sk} - G_p|$ is determined by the oxide thickness and the Tunneling Magnetoresistance

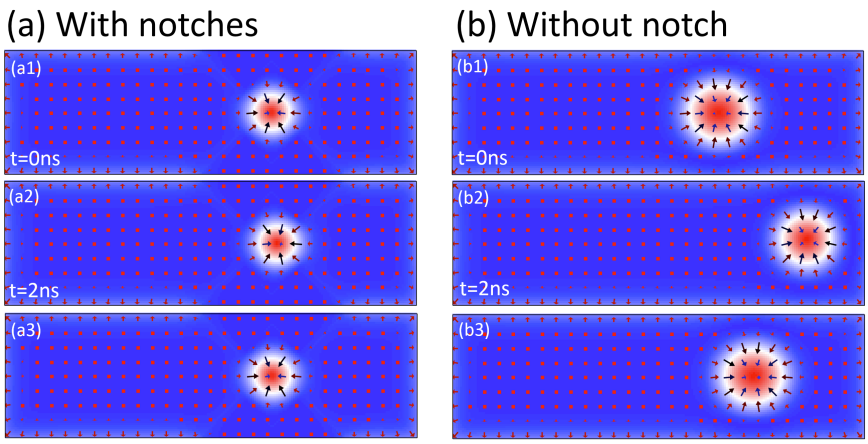

Fig. 5.3. Motion of a skyrmion under a read current of 56 μA in a nanotrack w/(with) and w/o(without) notches. (a1) and (b1) show the initial position of a skyrmion under the read MTJ. (a2) and (b2) represent the final position after 2 ns under the read current. (a3) and (b3) are the results of 1 ns relaxation time after turning the read current OFF.

Ratio (TMR). As discussed earlier, the synaptic conductance range is even lower than the TMR since the proportion of "free layer" domain anti-parallel to the "fixed layer" domain is limited by the cross-sectional area of the skyrmion. In order to improve the accuracy in complex pattern recognition tasks, a larger ratio of maximum to minimum synaptic conductance is required and can be achieved by having multiple branches with various conductance ranges, as shown in Fig. 5.2. The figure demonstrates the implementation of a 16-level synapse where the ratio of the maximum to minimum conductance of the equivalent synapse is 16. This is realized using a set of 3 skyrmionic devices where the oxide thickness of each branch is varied to ensure a different range of synaptic conductance. Each branch is also associated with a switch to ensure that all discrete conductances in the range G_0-$16G_0$ (G_0 being a constant) can be realized. For a fixed voltage applied between the terminals T1 and T3 of each branch, the equivalent synaptic weight is the combination of the conductance of read MTJs from active branches. For instance, a synaptic weight of $5G_0$ is obtained by turning ON the first two branches and turning OFF the third branch and the weight is the

combination of the read MTJ conductances from the first (without skyrmion) and the second (with skyrmion) branch.

As shown in Fig. 5.2, notches with high-K material are implemented at both sides of the read position for the following two reasons. First, since only an ultra-low current density is required for skyrmion displacement, notches at the right side of the read position are used to prevent the skyrmion from shifting during the read current injection. Second, due to a repulsive force from the right notches, the stable position of a skyrmion under a current injection is different from the position after the current is turned OFF. Fig. 5.3 shows the skyrmion motion under a read current density of 56 μA in a nanotrack w/(with) and w/o(without) notches. Device dimensions and material parameters are listed in Table. 6.1. With the aid of high-K notches, the skyrmion under the read MTJ can be well confined under the injection of a read current, as shown in Fig. 5.3a. Nevertheless, the skyrmion moves rightward without the help of high-K notches and stops at a certain position instead of pushing out due to the repulsive force from the edge, as shown in Fig. 5.3(b). Fig. 5.3(a3), (b3) shows the stable skyrmion position after turning the read current OFF for 1 ns. As shown in Fig. 5.3(b3), the skyrmion in a nanotrack shifts leftward owing to the skyrmion-edge interaction. Hence, a notch depth of 20 nm is needed in our simulation to avoid the skyrmion right shift under a required maximum read current. Also, a depth of 20 nm notch at the other side is implemented to free a skyrmion from back shift after turning the read current OFF. Moreover, as we mentioned previously, high-K materials ($SmCo_5$) with an anisotropy constant of 17.1 MJ/m^3 are also attached at both boundaries to prevent skyrmion from annihilation under a high-level programming current injection.

5.4.2 Skyrmion-based neuron

subsec:neuron The functionality of a spiking IF neuron can be exhibited by the proposed device structure, as shown in Fig. 5.4. Similar to our proposed synapse

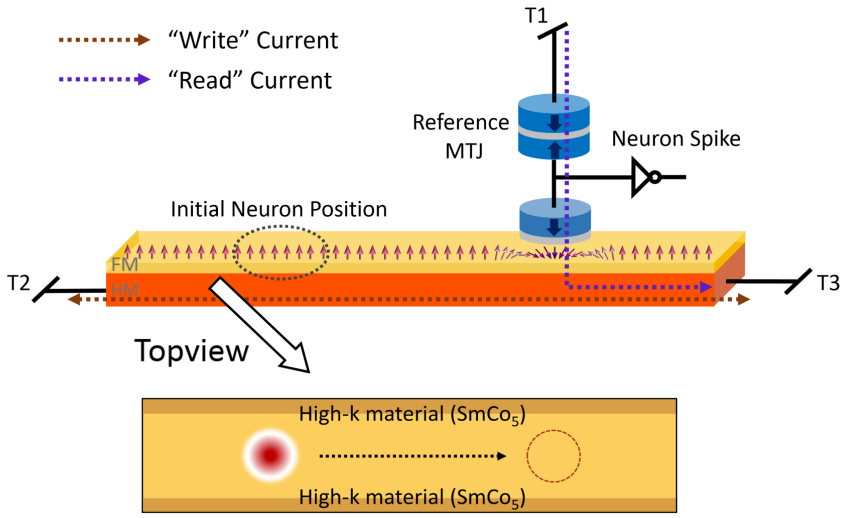

Fig. 5.4. Device structure of the proposed spintronic spiking neuron and the top view of the nanotrack. High-K materials with a width of 1 nm are attached to both edges for better skyrmion confinement. The width of the nanotrack is 60 nm, and the length is 200 nm.

device structure, only onetime nucleation process is needed to generate a skyrmion at the left end of the FM. High-K materials are also attached at both edges to ensure that the skyrmion is well-confined in the nanotrack. Membrane potential of a neuron increases as the skyrmion moves forward by the synaptic current passing through the SHM (T2 to T3). The amount of skyrmion displacement is related to the magnitude of an incoming charge current, which is modulated by the synaptic weight. The "integrate" property of the neuron can be realized by mapping the motion of a skyrmion in such magnetic multilayer structures. An MTJ in series with another reference MTJ at the end of the FM is used to detect whether a skyrmion reaches the region underneath the sensing MTJ. The FM layer at the read region serves as the "free layer" of the sensing MTJ. The resistance change arising from the presence of a skyrmion, in turn drives an inverter and "fires" an output spike in case the skrmion reaches the opposite edge of the "free layer". Finally, the neuron is reset by

driving the skyrmion backward to the origin position using a charge current flow in the opposite direction (T3 to T2).

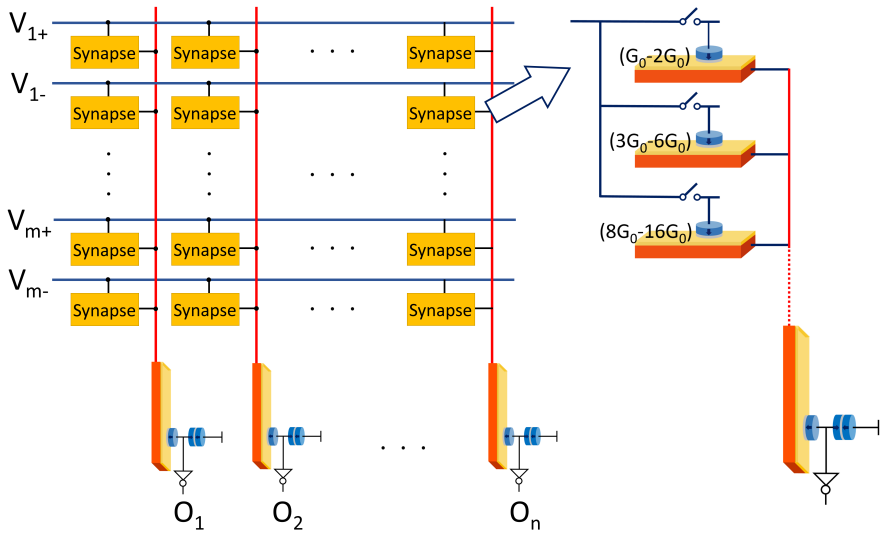

Fig. 5.5. Hardware mapping of an All-Spin Deep Spiking neural architecture. Skyrmion-based synapses in the crossbar array encode the synaptic weight, and provide a corresponding synaptic current to the interfaced skyrmion-based neurons.

5.4.3 All spin neuromorphic architecture

The core computational unit in any Deep Convolutional architecture is a massively parallel dot-product implementation between the inputs of a particular layer and the corresponding synaptic weights of different neurons in that corresponding layer, followed by neural processing. This operation can be mapped to a synaptic crossbar array architecture, as shown in Fig. 5.5, where the synapses drive magneto-metallic skyrmionic neurons. Each time-step of operation of the SNN can be resolved in two successive "write"-"read" cycles for the skyrmionic neurons. In the "write" phase, input voltages drive different rows of the synaptic crossbar array corresponding to those rows which receive an input spike. Since the input "write" resistance

of the skyrmionic neuron (the SHM underlayer) is relatively low, the voltage drop across the neuron is much lower in comparison to the voltage drop across the synaptic crossbar array. Consequently, the neurons receive a resultant synaptic current that is the weighted summation of the spike voltage inputs for each neuron. In the successive "read" cycle, the neurons generate an output spike in that time-step only if the skyrmion has reached the opposite edge of the device and triggered the output inverter. The energy efficiency of such All-Spin crossbar based neuromorphic architectures stems from two perspectives: (i) Compared to other emerging post-CMOS devices with higher switching voltage, low current requirement of magneto-metallic skyrmion-based neurons enable the low voltage operation (can be down to a few tens to hundreds of mV) of the large skyrmion device based crossbar arrays, and (ii) the non-volatile skyrmion device based "in-memory" computing architectures alleviates the memory-fetch bottleneck prevalent in CMOS based neuromorphic architectures.

5.5 Simulation Framework and Results

In order to simulate the implementation of All-Spin Deep SNN based on the proposed magnetic skyrmionic spin device, a hybrid device-circuit-algorithm co-simulation framework is utilized. A "top-down" (algorithm) and "bottom-up" (device modeling) simulation framework is developed to evaluate the performance of such All-Spin SNNs based on skyrmionic devices over a standard digit recognition problem on the MNIST dataset [101]. Micromagnetic simulations of the skyrmion device are performed using the tool Mumax3 [78, 79], a GPU accelerated micromagneitc simulation framework. Subsequently, a SPICE model is developed for the neurocomputing fabric. The circuit-level simulation is based on a 45 nm technology node. The Deep SNN used for this work, consists of two convolution layers and two subsampling layers arranged alternatively ($28 \times 28 - 6c5 - 2s - 12c5 - 2s - 10o$). Readers are directed to Ref. [99] for details on the supervised training process of such Deep SNNs. Our design falls into the category of offline supervised learning where the synaptic

weights are learnt off-chip by backpropagation and are subsequently programmed to corresponding resistive states of the spintronic crossbar array.

Behavioral device models are developed for system level simulations of the Deep Spiking Neural Network. The material parameters used in our simulations correspond to Co/Pt multilayers adopted from Ref. [70], and are shown in Table 6.1. We consider a 0.4 nm thick Co nanotrack with perpendicular magnetic anisotropy on a 3 nm Pt substrate inducing DMI. The sample is discretized into an element size of 1 nm $\times$ 1 nm $\times$ 0.4 nm. For a device with dimensions of 200 nm $\times$ 60 nm, a current of 17.5 μA is required to displace a skyrmion from initial position to the position underneath the read MTJ in a duration of 2 ns and 0.5 ns relaxation time. However, as shown in Fig. 5.6(b), the annihilation of a skyrmion takes place under a current of 50 μA in 2 ns. High energy barriers can be induced at both edges by attaching high-K materials, and thus, the critical annihilation current can be increased to 72 μA. In order to obtain the resistance of read MTJ, Non-equilibrium Green's Function (NEGF) based spin transport simulation has been used in our simulations. The charge current (I_e) flowing through the SHM and the corresponding spin current (I_s) are calculated using $I_s = \theta_{sh}\frac{A_{MTJ}}{A_{SHM}}I_e$ [81], where A_{MTJ} and A_{SHM} are the cross sectional areas of the MTJ and SHM, respectively, and θ_{sh} is the spin-Hall angle. The spin current is used to analyze the magnetization dynamics with the generalized Landau-Lifshitz-Gilbert-Slonczewski equation. Note that skyrmions might be pinned while shifting in a thin nanotrack [102, 103]. However, this effect is not taken into consideration in our work, as we believe its impact on the performance (classification accuracy) of the neural network would not be significant due to the inherent error-resiliency of such brain-inspired computing platforms.

The bit discretization level in the neuron and synapse has great impact on the accuracy of the neural network. It was observed that 16 (4 bit) intermediate levels in the synapse and 4 (2 bit) intermediate levels in the neuron results in insignificant degradation in classification accuracy for the particular digit recognition framework under consideration. Assuming that a skyrmion over a distance of 20 nm can be

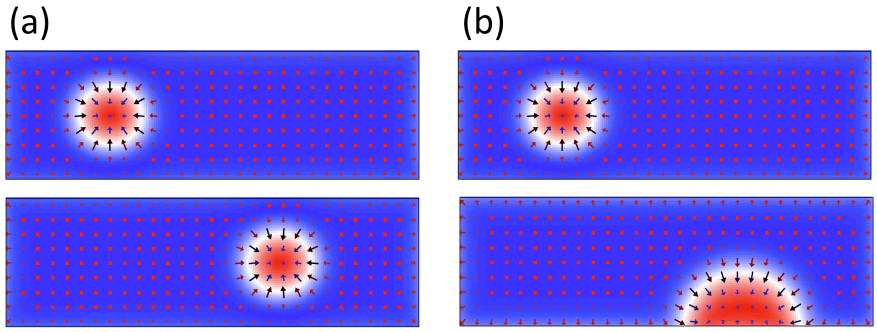

Fig. 5.6. (a) shows the motion of a skyrmion under a current magnitude of 70 μA for 2 ns and 0.5 ns relaxation time with high-K materials adhered on both edges. (b) shows a skyrmion is annihilated in 2 ns under a current of 50 μA for 2 ns without the aid of high-K materials

Table 5.1.
Material parameters used in simulation

Parameters	**Value**
Saturation magnetization (M_{sat})	580 kA/m
PMA anisotropy constant (K_u)	0.8 MJ/m^3
Exchange constant (A)	15 pJ/m
Dzyaloshinskii-Moriya interaction (DMI) strength (D)	3 mJ/m^2
Gilbert Damping Factor (α)	0.3
Spin Hall Angle (θ_{sh})	0.07
Neuron length and width	200 nm × 60 nm
Synapse length and width	120 nm × 60 nm
Nanotrack thickness	0.4 nm
SHM thickness	3 nm
MTJ diameter	20 nm

sensed, and the minimum distance from the center of a stable skyrmion to the edge is $\sim$ 50 nm owing to the skyrmion-edge interaction, the length of neurons were chosen

to be 200 nm. Considering the minimum distance from a skyrmion to the edge, and assuming a 20 nm separation from initial position to the final position, the length of synapses were chosen to be 120 nm. Fig. 5.7 depicts the variation of the classification accuracy as a function of the number of time-steps of operation of the SNN. A competitive accuracy of 96.02 % (measured over the entire testing set) was achieved at the end of 50 time-steps by simulating the entire All-Spin skyrmionic SNN architecture.

Circuit level simulations indicate that the proposed skyrmionic SNN is able to achieve an energy improvement by $\sim 117\times$ in comparison to a baseline iso-network CMOS implementation in 45 nm technology. The CMOS baseline is equipped with power gating functionalities to exploit energy reductions that can be obtained due to sparse firing of neural events. Note that this energy calculation includes only the logic power consumption of the CMOS neural network. Energy consumption required to fetch the synaptic weights from SRAM would involve an additional energy overhead and further improve the energy efficiency of our proposed skyrmion "All-Spin" SNN. An intuitive understanding of the energy efficiency of such All-Spin skyrmionic networks can be obtained from device-level simulations. Micromagnetic simulations reveal that $\sim 17.5\ \mu A$ current is required to displace the skyrmion from one edge of the neuron to another (dimension 200 $nm \times 60\ nm$) in a duration of 2 ns, resulting in an energy consumption of 0.14 fJ (I^2Rt energy consumption). In addition, the spintronic synapses providing input currents to each neuron are operated at ultra-low terminal voltages of 100 mV due to the low switching current requirement for these magneto-metallic skyrmionic neurons.

5.6 Conclusion

In this work, we propose a magnetic skyrmion based device to emulate the core functionality of neurons and synapse for an All-Spin Spiking Deep Neural Network. The synaptic weight can be adjusted by the number of skyrmions under the read

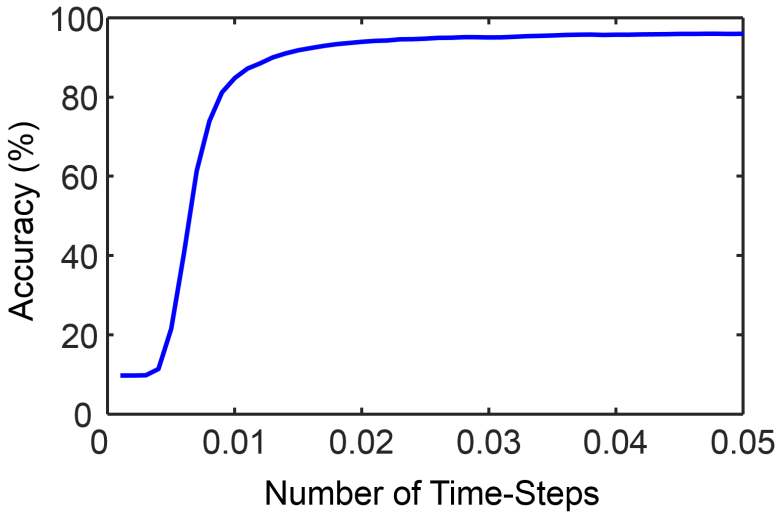

Fig. 5.7. Temporal variation of the classification accuracy of the "All-Spin" skyrmionic SNN as a function of the number of time-steps.

MTJs, and the resolution can be improved by having multiple branches with various conductance range. Moreover, ultra-low current switching of skyrmionic neurons have the potential to provide ultra-low voltage operation for skyrmionic synapses, suggesting new possibilities for exploiting skyrmion-based devices in low-power Deep Spiking neural architectures.

6. ISING COMPUTATION BASED BI-PRIME FACTORIZATION

6.1 Introduction

The quest for efficient computing algorithm and hardware for solving combinatorial optimization problem has long been considered a popular but challenging topic. Formulated as the search of a set of discrete variables among all the possible combinations to minimize or maximize specific functions, these problems are the foundation of a variety of large-scale, complex real-world systems. Biprime factorization, for example, is at the heart of secure data transmission (such as RSA encryption [104] [1]), in which the security of the data is guaranteed since no classical algorithm can factorize a large number in polynomial time. Using conventional computing method to solve these optimization problems, it demands a sequence of data fetching, decoding, and executing in each operation cycle. Hence, considering the energy and power requirements, finding an optimum solution seems more infeasible when the required amount of resources for solving these problems get larger. On the other hand, quantum computers are theoretically proven capable of solving such problems efficiently [105], but the largest number demonstrated yet is still relatively limited, as even the most powerful quantum device has only tens of physical qubits at present.

Ising model emerges as a potential candidate and a special-purpose device to efficiently solve several kinds of combinatorial problems [106]. Through mapping the problem into the form of an Ising Hamiltonian (i.e., pairwise spin interactions in the presence of external driving fields), the optimization procedures now amount to finding the minimum energy state for the associated Hamiltonian. This can be realized

[1] RivestShamirAdleman (RSA) algorithm is one of the most widely used encryption techniques. In RSA algorithm, Alice randomly picks two large, prime numbers p and q to generate a public key pq for other users to encrypt message.

by physically setting up the coupling coefficients between Ising cells and the local fields, and simply observing an initial state gradually decreasing to the lower-energy state. The introduction of the stochastic behaviour to the Ising cell can avoid the local minima trap and guarantee the ground state can be reached. Owing to its simple architecture and inherent ability to solve optimization problems, Ising model has been researched extensively [107, 108]. An Ising model based on CMOS implementation has also been proposed by Yamaoka *et al.* [109]. However, to realize some primitive functionalities of the Ising model, such as stochastic behavior and the updating process, complex CMOS circuits designs such as random number generator and majority gate are introduced to their design.On the contrary, using beyond CMOS devices whose fundamental properties can mimic the functionalities required in the Ising model opens up new possibilities for hardware implementation. In this work, we propose a spin Hall Metal (SHM) underlying a magnetic tunnel junction (MTJ) as a non-volatile Ising cell (SHM-MTJ). The magnetization direction represents the state of each Ising cell, and the stochastic switching dynamics can be controlled by the magnitude of charge current flowing through the SHM. The device used in this work has been experimentally demonstrated in Ref. [24], and proposed in Ref. [110] to solve Maximum-cut and Graph coloring problem. Recently, a similar spin-based Ising cell with a low energy barrier ($< 5KT$) between two possible state has been proposed in which the magnetization switching is also induced by spin Hall effect (SHE) [111]. However, the devices are exponentially vulnerable to process and temperature variations and the stored data could be destructed during the read operation owing to the low energy barrier between the two states. Another work suggesting that Ising cell can be realized by high energy barrier nanomagnet has been proposed in Ref. [112], where the magnetization of the nanomagnet is initially set on its hard axis direction. The requirement for the dipole coupling of nanomaget increases the device complexity, and details regarding peripheral circuits to realize given functions are missing. In this work, SHE-MTJs are used as stochastic switching bits, and the simplicity of our proposed device, along with its peripheral CMOS circuits, is well-suited

for Ising cell construction. The stochastic magnetization dynamics can be realized by *Landau-Lifshitz-Gilbert-Slonczewski* (LLGS) equations. We design a unique approach to map bi-prime factorization problem to our proposed device-circuit configuration. By solving coupled LLGS equations, we demonstrate that our coupling network can factorize up to 16-bit binary numbers. Beyond 16-bits, the run-time to simulate the operations increases a lot. Note, however, when the proposed model is implemented in hardware, larger numbers can be factorized.

6.2 Ising model

In this section, we start with a general introduction of the Ising model. Fig. 6.1(a) represents a schematic of an Ising model, which consists of individual spins (x_i), the coupling coefficients between neighboring spins (J_{ij}), and local magnetic fields (h_i). The system Hamiltonian (H), total energy, is defined as below,

$$H = -\sum_{i<j} J_{ij} s_i s_j - \sum_{i=1}^{N} h_i s_i \tag{6.1}$$

Each spin can be in one of the two states, where "+1" represents spin pointing up and "-1" represents spin pointing down. The next state of each spin is determined by the effective force from adjacent spins and its local field. Take Fig. 6.1(a) as a example. Each spin interacts with four adjacent neighbors and its local magnetic field. The center spin, denoted as s_0, interacts with a local magnetic field (h_0) and connects with spin s_1, s_2, s_3, s_4, with coupling coefficient J_{01}, J_{02}, J_{03}, J_{04}, respectively. A positive (negative) J_{0j} implies the neighboring spin s_j tend to align the center spin parallel (anti-parallel) to its direction, while the tendency is determined by the magnitude of J_{0j}. Similarly, a positive (negative) external magnetic field h_i attempts to align a given spin up (down), with the tendency determined by the magnitude of h_i. For the center spin s_0, one has to consider the contributions from all neighboring spins and the local magnetic field. Assume we would like to know how a particular spin s_i will evolve in the process. Given its connection with a neighboring spins s_j with a

coupling coefficients of J_{ij}, we could calculate the effective force for a specific combination of (s_j, J_{ij}) as the product of the neighboring spin state and the associated coupling coefficient, i.e., $s_j \times J_{ij}$. This number is unitless, with its magnitude only showing the relative strength with respect to the other interactions. For example, for the combinations of (+1,+3) and (-1,-4), we could calculate an effective force of "+3" and "+4" on spin s_i, respectively. For both cases, the spin s_i will tend to point up, but shows a stronger tendency to be +1 in the latter combination. Including the contributions from other neighboring spins and the local magnetic field h_i, we could calculate the final effective force for the spin s_i by summing all combinations of (s_j, J_{ij}) and h_i, and map this number to an injected current, which corresponds to a switching probability for this spin state. Detailed discussions covering this mapping process will be covered in the latter section. As a result, numerous combinatorial optimization problems can be mapped to the Ising model with appropriate interconnection weights. The spin configuration evolves through interaction with adjacent spins and automatically converges toward the lowest energy state. The solution (i.e., global minima) is the spin status when the system reaches the steady state.

Fig. 6.1(b) illustrates an example of the energy profile for different combinations of spin states. This particular profile indicates that the system has several local minimas and one global minimum. As the spin states evolve to minimize the associated Hamiltonian (i.e. converge toward the lowest energy state), the system could be stuck at some local minimum states. Several approaches have been proposed to prevent the system from being trapped at some local minima, such as "simulated annealing" and "quantum annealing". Simulated annealing is one of the oldest algorithms to solve Ising model, in which the system moves toward the global minimum by gradually decreasing the temperature. On the other hand, quantum annealing relies on tunneling effects to escape from local minima. CMOS architecture, as discussed earlier, can perform the required updating and annealing process with additional circuit design, but at the expense of higher energy consumption and extra area. Our proposed spintronic device, on the other hand, is inherently suitable for implementation of Ising

model. Its stochastic switching behavior in the presence of thermal noises can readily realize the annealing process, while the local fields and the interactions between spin states can be set up to mimic the minimization process. In the next section, we will describe the device structure and the operation of our proposed Ising cell.

6.3 Underlying physics of the proposed Ising cell

Fig. 6.2 shows our proposed three-terminal Ising cell, composed of a conventional MTJ in contact with a SHM layer. The structure of an MTJ consists of a tunneling barrier (TB) sandwiched by two ferromagnetic (FM) layers. One FM layer is engineered to have its magnetization pinned (pinned layer, or PL) in one direction so as to be used as a reference layer, whereas the magnetization of the other FM, free layer (FL), is reversible under an incoming spin current from the SHM. Depending on the relative magnetization of the FL layer to the magnetization of the PL (either parallel

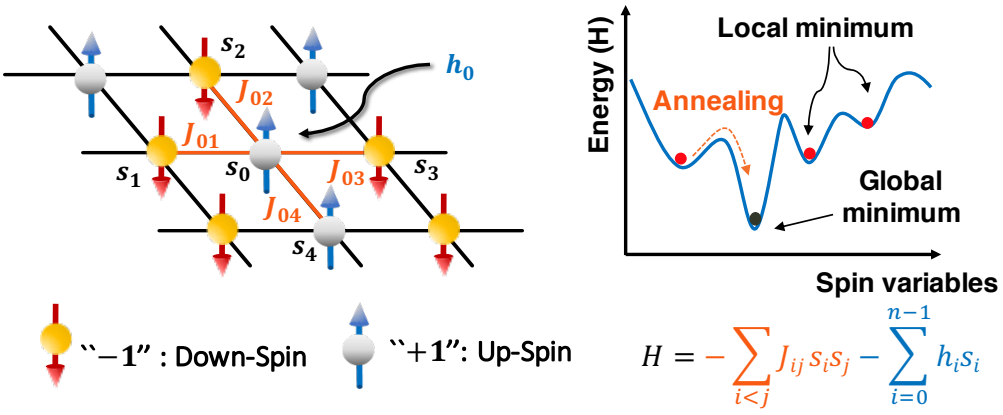

Fig. 6.1. (a) Ising model (b) Energy profile and annealing process

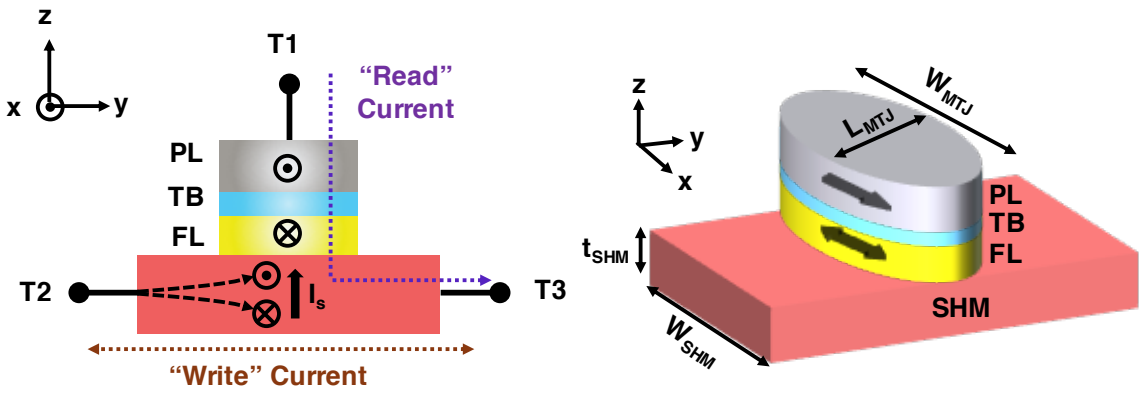

Fig. 6.2. Three-terminal SHM-MTJ device, with a SHM underlying an in-plan magnetic anisotropy (IMA) MTJ. The magnetization of the FL can be written by injecting a current through the SHM, and read by sensing the resistance of the MTJ. Write/read operation can be optimized separately as the write/read path is decoupled, and hence, the design is more flexible in such three-terminal device.

or anti-parallel), an MTJ can exhibit a low resistance state (R_P) or a high resistance state (R_{AP}), respectively.

It has been experimentally demonstrated that spin current can be generated more efficiently by injecting current through an SHM with strong spin-orbit coupling (SOC) [23]. One of the proposed mechanism to explain such phenomenon is the spin Hall effect (SHE) [39,40]. As described in Fig. 6.2, when a charge current flowing through the SHM, electrons with opposite spins are deflected to opposite surfaces of the SHM. A spin current is generated and flowing to the FL. The polarization (σ) of the spin current can be described by the relation between the spin (J_s) and charge current density (J_c), $J_S = \theta_{sh}(\sigma \times J_q)$, where θ_{sh} is the spin Hall angle. The spin torque exerting on the FL depends on the direction of charge current flowing through the SHM. When passing a charge current from terminal T2 to T3, this spin torque switches its magnetization to become parallel to that of the PL. Whereas, if the charge current injecting from terminal T3 to T2, the spin torque will be anti-parallel to the

magnetization of the PL. Hence, the "write" current flowing from terminal T2 to T3 switches the resistance of MTJ from R_{AP} to R_p, and vice versa. The resistance of the MTJ (either R_{AP} or R_p) can be sensed by applying a fixed voltage between the terminal T1 and T2/T3, and checking the "read" current flowing through the MTJ. The proposed MTJ-SHM multilayer structure provides some desirable characteristics as an Ising cell. First, the decoupled path of "write" and "read" operation provides design flexibility. Second, energy-efficient spin current can be generated by injecting current through the SHM underlying an MTJ, resulting in lower write energy. Third, switching probability can be controlled by altering the input current. The details regarding how our proposed device can efficiently perform the underlying functionalities of an Ising model will be described in the next section.

Fig. 6.3(a) shows the switching probability (P_{sw}) of an MTJ in response to a current flowing through the underlying SHM. A "write" current with a pulse width 1 ns is injected to the SHM underlayer, and an additional 1 ns of relaxation time is required to stabilize the magnetization. The amount of input current varies from 40 μA to 160 μA with a 1 μA step. A switching probability is obtained by executing 10^4 simulations for each current step. Fig. 6.3(b) illustrates the magnetization dynamics when the input charge current is 55 μA, 85 μA and 115 μA for 1 ns, and followed by 1 ns relaxation time. For better visualization, only 50 out of 10^4 simulations are plotted here, and the results can be predicted from Fig. 6.3(a).

Generally, the annealing process in an Ising model involves the random perturbation of spin states to prevent the system from trapping at some local minimum energy states. In this work, a nanomagnet can be used as a randomizer to realize the "annealing process" in the Ising model, owing to its stochastic flipping nature in the presence of thermal noise. Besides, the desired switching probability can be controlled by the magnitude of the "write" current.

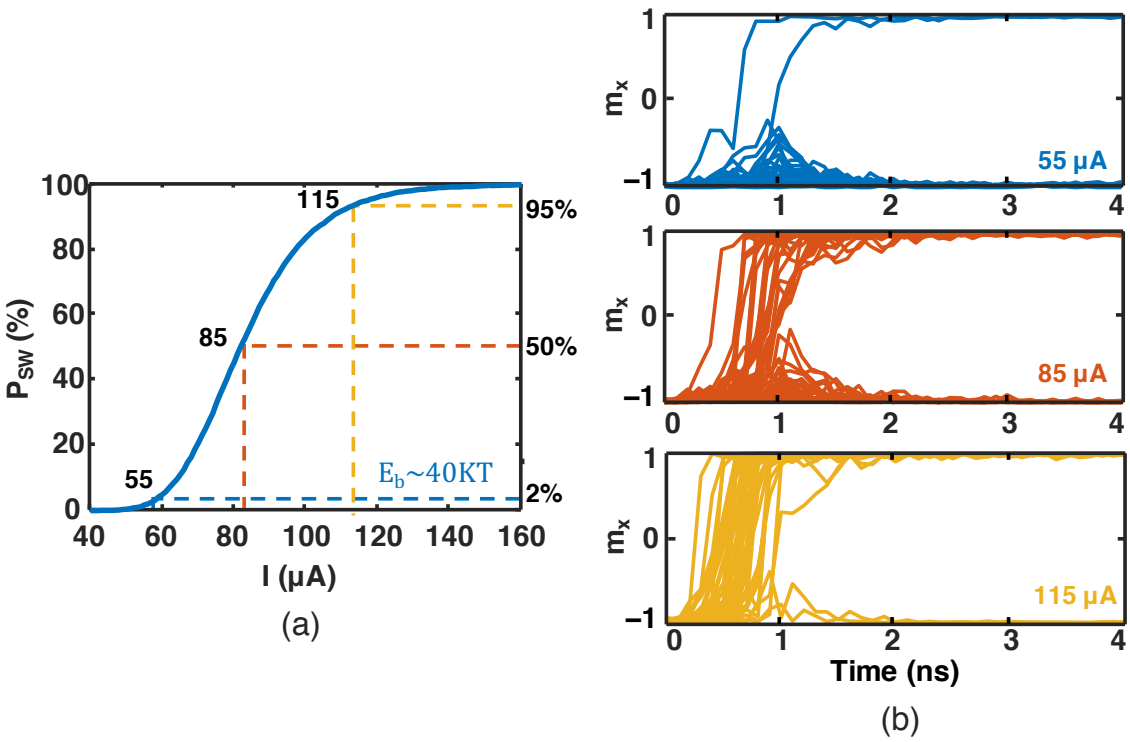

Fig. 6.3. (a) Switching probability of an MTJ in response to a current pulse flowing through the SHM. 10^4 simulations were performed for each current step. The current with different magnitude (varies from 40 μA to 160 μA with 1 μA per step) is applied for 1 ns, and followed by a 1 ns relaxation time. (b) Magnetization switching from -1 to +1 under an input current of 55 μA, 85 μA and 115 μA for 1 ns. For better visualization, only 50 out of 10^4 simulations are plotted here, and the results are obtained from Fig.6.3.(a)

6.3.1 Device operation

The underlying functionalities of the Ising model can be categorized into two parts, "annealing process" and "updating process". In this section, we will describe how our proposed device can be used to mimic these two operations. Fig. 6.4(a) shows our proposed SHM based Ising cell circuit that can efficiently implement the updating process. Multiple current sources and switches here are used to control the total current received by SHM. The magnitude of each current source from a neighboring spin (s_j) and the local field (h_i), depends on the value of $|J|$,$|h|$, respectively. The status of switches (ON/OFF) is determined by the state of neighboring spins and the coupling coefficient. Fig. 6.4(b) depicts the corresponding P_{sw} under a specific total current injection. Here, we assume operation current is limited within 55 μA to 115 μA, and this range (60 μA) is divided among the current sources from neighbors and local field. The minimal operation current, 55 μA, is provided by the I_{BIAS} current source, and corresponds to a switching probability of 2%. Assume the spin state to be updated is currently pointing up (down), a negative (positive) effective force from neighbors/local fields will tend to flip this spin, and thus we turn ON the corresponding switch for current injection. While for a positive (negative) force, we turn OFF the switch and the system will only have a limited switching probability due to the bias current (I_{BIAS}). For example, if a particular spin (s_i) currently points up and all the effective forces are positive, the total operation current becomes 55 μA, suggesting a switching probability of 2 % only. Similarly, if all the effective forces are negative, then the total current now becomes 115 μA, which leads to 95 % switching probability. This behavior can be directly used for performing the update process in an Ising model. Also, as we discussed in the previous section, inherent stochastic behavior mimics the natural annealing process.

Figure. 6.5 shows the overall device-circuit configuration of our proposed Ising cell. Read operation is performed by a resistance in series with the MTJ. Depending on whether the MTJ is in high resistance state or low resistance state, the inverter

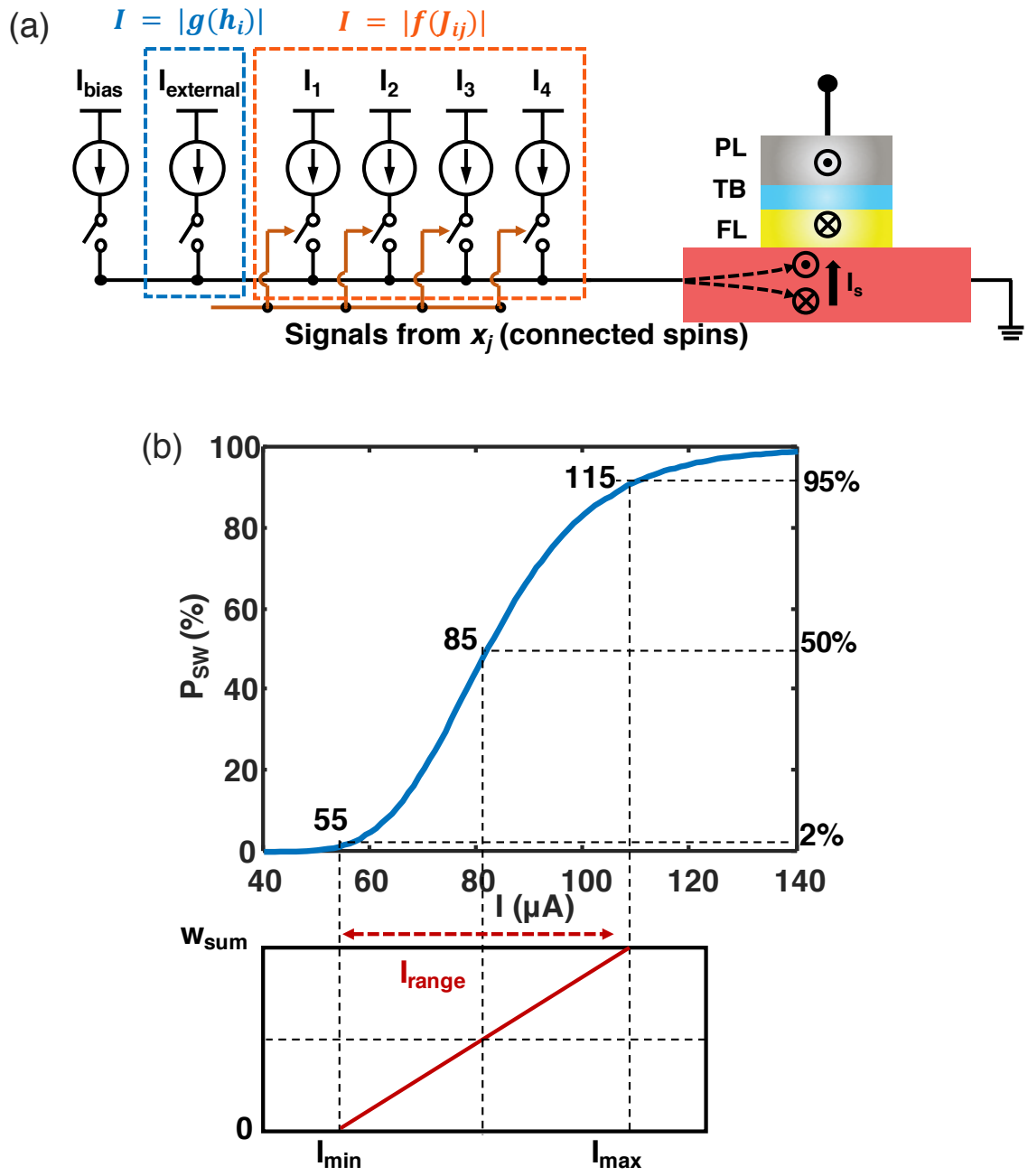

Fig. 6.4. (a) The evolution of a particular spin (s_i) is determined by the final effective force from all the adjacent neighbors and its local field. This updating process is mapped to current-dependent switching probability of the magnetization of the FL. To control the amount of total current flowing through the SHM, multiple current sources and switches are introduced. The magnitude of each current source from a neighboring spin (s_j) and the local field (h_i), depends on the value of $|J|$,$|h|$, respectively. (b) Magnetization switching probability can be controlled by the amount of input current.

outputs a logic "0" or "1", respectively. The R_{ref} is set between two stable states of the MTJ. In order to decouple the interaction with neighboring spin and evaluate the energy of current spin configuration, the spin state is stored in the following D-latch. For the write operations, the direction of the current injection depends on the spin state at present, which can be obtained from the "next state" signal. There are multiple branches of stacked transistors to perform the updating process. One transistor is used for biasing and the other served as switches. XNOR/XOR logic gates are used to combine information from neighbors (the sign of the neighboring spin state and the associated coupling coefficient), and NOR/OR gates are used to control the signal from local field. Therefore, the probability of switching to the opposite direction or remaining in the present state depends on the overall injecting current.

6.3.2 Modeling and Simulation

The magnetization dynamics of the FL under an external stimuli, such as a magnetic field or spin current, at zero temperature can be obtained by solving *Landau-Lifshitz-Gilbert-Slonczewski* (LLGS) equation.

$$\frac{d\widehat{m}}{dt} = -\gamma(\widehat{m} \times H_{eff}) + \alpha(\widehat{m} \times \frac{d\widehat{m}}{dt}) + \frac{I_s}{qN_s}(\widehat{m} \times \sigma \times \widehat{m}) \tag{6.2}$$

where $\widehat{m}$ is the unit vector of the FL; $\gamma = \frac{2\mu_B\mu_0}{\hbar}$ is the Gilbert gyromagnetic ratio for electron; α is the Gilbert damping ratio; H_{eff} is the effective magnetic field incorporating the shape anisotropy for an elliptical disk; $N_s = \frac{M_s V}{\mu_B}$ is the number of spins in the FL occupying a definte volume V; M_s is the saturation magnetization; μ_B is Bohr magneton; and μ_0 is the magnetic permeability. The spin current, I_s, is generated by a charge current flowing through the SHM underlayer, and can be expressed as

$$I_s = \theta_{sh}\frac{A_{MTJ}}{A_{SHM}}I_e \tag{6.3}$$

where A_{MTJ} and A_{SHM} are the cross sectional areas of the MTJ and SHM, respectively, and θ_{sh} is the spin-Hall angle. The device parameters we used for benchmark

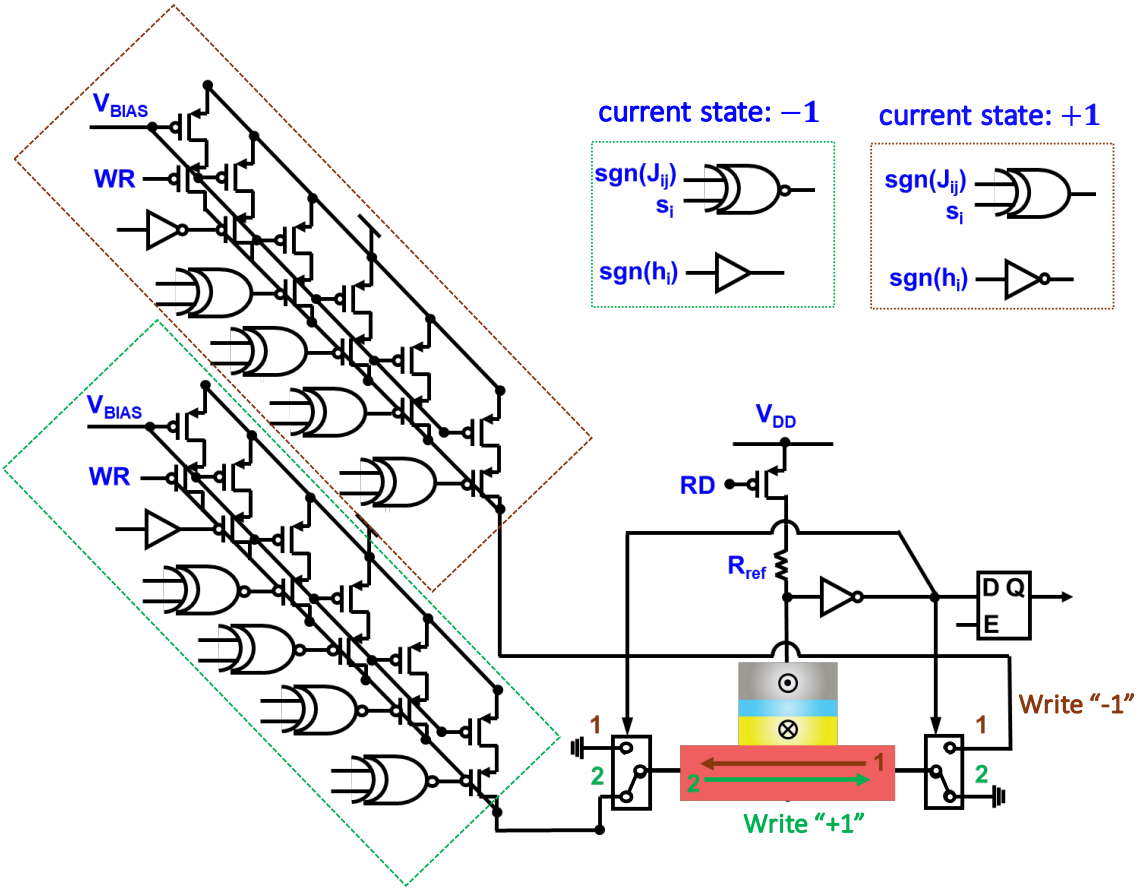

Fig. 6.5. The proposed device-circuit configuration for single Ising model. Peripheral CMOS logic gates are used to perform the updating process. An inverter in series with other transistors and reference resistors is used to convert the spin state to binary voltage value.

have been shown in Table 1. It is worth noting that spin current generated from SHM can potentially provide energy-efficient "write", as the spin injection efficiency can be larger than 100 % and is not limited by the polarization of the PL.

Finally, at non-zero temperature, the magnetization switching dynamics of an MTJ is influenced by thermal noise, which is accounted as a random thermal field and can be factored into the LLGS equation by augmenting H_{eff} with a thermal field, $H_{thermal}$

$$H_{thermal} = \sqrt{\frac{\alpha}{1+\alpha^2}\frac{2K_BT}{\gamma\mu_0 M_s V \delta_t}G_{0,1}} \tag{6.4}$$

where $G_{0,1}$ is a Gaussian distribution with zero mean and unit standard deviation; K_B is Boltzmann constant; T is the temperature; and δ_t is the simulation time step. Hence, an MTJ exhibits stochastic switching during "write" current operation in the presence of thermal noise.

Table 6.1.
Material parameters used in simulation

Parameters	**Value**
Free layer area	1000 KA/m
Free layer thickness	1000 KA/m
Spin Hall metal thickness	2.3 KA/m
Saturation magnetization (M_{sat})	1000 KA/m
Spin Hall Angle (θ_{sh})	0.3
Gilbert Damping Factor (α)	0.3
Energy barrier (E_B)	40 KT
MgO thickness (t_{MgO})	1.4 nm
Resistivity of SHM (ρ_{SHM})	20
Write pulse width (t_{pw})	3 ns
Temperature (α)	300 K
Supply voltage (V_{DD})	1 V

6.4 Problem mapping and results

In this section, we first elaborate how to map a bi-prime factorization problem to our proposed Ising cell. Fig. 6.6 shows the hardware implementation of our proposed SHM-MTJ devices. A SHM layer underneath an MTJ with peripheral CMOS circuits can represent an Ising cell. Given a specific problem to solve, we program the coupling coefficients of the adjacent spins and local fields. Starting from an initial spin configuration and its corresponding connections, the system starts evolving toward the minimum energy state as we update the current on each spin sequentially, by checking the local field as well as the neighboring spin states and their coupling coefficients. The overall effective current for write operation is obtained through SPICE simulations. The resultant current is then fed into the LLGS solver to analyze the magnetization dynamics. We apply a 1 ns write current pulse to the SHM, and examine the spin state after a 1 ns relaxation time. This process is repeated until all the spins in the system are updated.

The aforementioned methodology was then applied to solve the biprime factorization problem. Finding the prime factors of an integer can be mapped to the final Ising Hamiltonian of the general form $H = -\sum_{i<j} J_{ij} s_i s_j - \sum_{i=1}^{N} h_i s_i$. In order to factor $N = pq$, where both p and q are prime numbers, we assume the bits (length) of p and q are m and n in binary representation, respectively. We take $p = (x_{m-1}x_{m-2}...x_1 1)$, $q = (x_{n-1}x_{n-2}...x_1 1)$, where x_i is binary digit, and thus we have $p = \sum_{i=1}^{m-1} 2^i x_i + 1$ and $q = \sum_{i=1}^{n-1} 2^i x_i + 1$. We define the cost function $f(x_1, x_2, ..., x_{m+n-2}) = (N - pq)^2$. Note that, in this work, we only focus on the case where N can be factorized into p and q of similar bit-length, since it is harder for classical computer to solve. In this case, p and q are roughly the square root of N. Thus, if we want to represent them in the binary form, their bit-lengths are half of the binary digit of N. Consider $N = 15$ as an example. We could express the two prime factors as: $p = (x_1 1)_2 = 2 \times x_1 + 1$, $q = (x_3 x_2 1)_2 = 2^2 \times x_3 + 2 \times x_2 + 1$, $x_i \in \{0, 1\}$. The corresponding cost function is

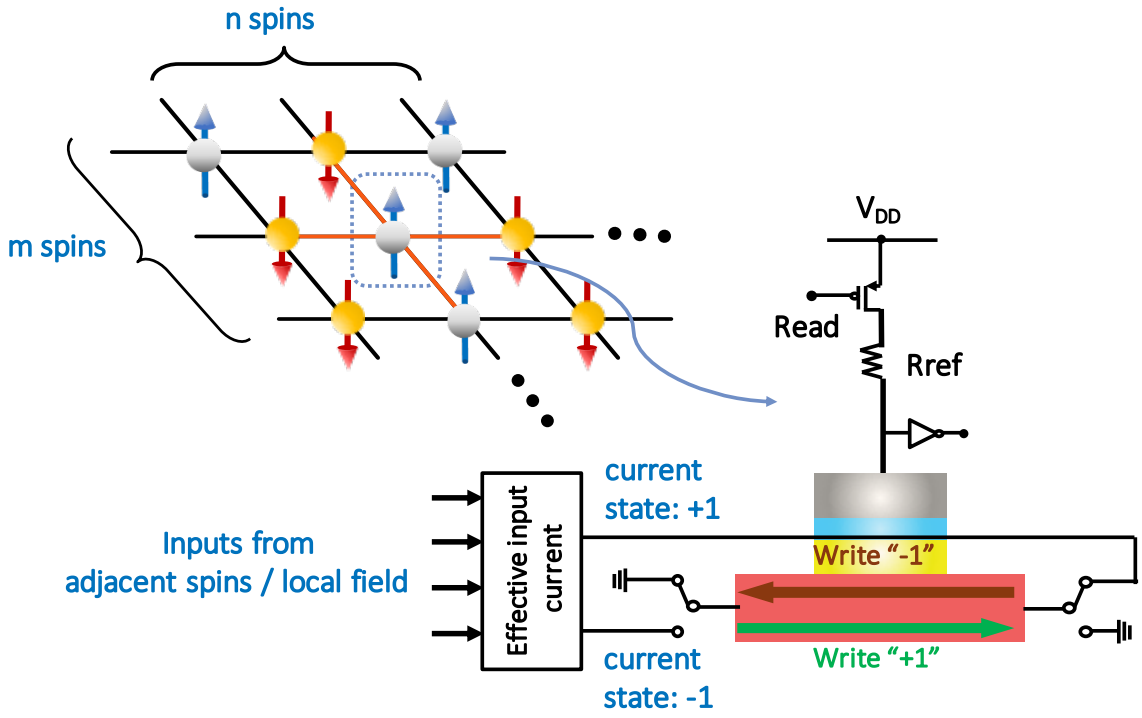

Fig. 6.6. Hardware implementation of an Ising model based on the proposed SHM-MTJ devices as a computational unit.

$$\begin{aligned} f(x) &= 128x_1x_2x_3 - 56x_1x_2 - 48x_1x_3 \\ &+16x_2x_3 - 52x_1 - 52x_2 - 96x_3 + 196 \end{aligned} \tag{6.5}$$

Since our Ising model relies on interactions between two spins at a time, we have to reduce the 3-local terms in Eq. 6.5 to 2-local terms. Utilizing the following equivalent relations: for $x, y, z \in \{0, 1\}$, (i) $xy = z$ iff $xy - 2xz - 2yz + 3z = 0$, and (ii) $xy \neq z$ iff $xy - 2xz - 2yz + 3z > 0$. Therefore, we can have

$$\begin{aligned} &x_1x_2x_3 = x_4x_3 + 2(x_1x_2 - 2x_1x_4 - 2x_2x_4 + 3x_4) \\ &\text{if } \ x_4 = x_1x_2 \\ &x_1x_2x_3 < x_4x_3 + 2(x_1x_2 - 2x_1x_4 - 2x_2x_4 + 3x_4) \\ &\text{if } \ x_4 \neq x_1x_2 \end{aligned}$$

$$\min(x_1x_2x_3) = \min(x_4x_3 + 2(x_1x_2 - 2x_1x_4 - 2x_2x_4 + 3x_4))$$

Replacing x_1x_2 with x_4, we can transform the undesired term $x_1x_2x_3$ into a quadratic form. Thus, we can rewrite the Eq. 6.5 as

$$\begin{aligned} &f'(x) = \\ &200x_1x_2 - 48x_1x_3 - 512x_1x_4 + 16x_2x_3 - 512x_2x_4 \\ &+128x_3x_4 - 52x_1 - 52x_2 - 96x_3 + 768x_4 + 196 \end{aligned} \tag{6.6}$$

, where

$$\min_{x_1x_2=x_4} f(x_1, x_2, x_3) = \min f'(x_1, x_2, x_3, x_4)$$

Since our variables x_i here are binary digits while each spin s_i is either +1 or -1, we perform transformation of variables using $x_i = \frac{1-s_i}{2}$, $i = 1, 2, 3, 4$

$$\begin{aligned} &f'(x_1, x_2, x_3, x_4) = \\ &200\tfrac{1-s_1}{2}\tfrac{1-s_2}{2} - 48\tfrac{1-s_1}{2}\tfrac{1-s_3}{2} - 512\tfrac{1-s_1}{2}\tfrac{1-s_4}{2} + 16\tfrac{1-s_2}{2}\tfrac{1-s_3}{2} \\ &-512\tfrac{1-s_2}{2}\tfrac{1-s_4}{2} + 128\tfrac{1-s_3}{2}\tfrac{1-s_4}{2} - 52\tfrac{1-s_1}{2} - 52\tfrac{1-s_2}{2} \\ &-96\tfrac{1-s_3}{2} + 768\tfrac{1-s_4}{2} + 196 \\ &= 116s_1 + 100s_2 + 24s_3 - 160_4 + 50s_1s_2 - 12s_1s_3 - 128s_1s_4 \\ &+4s_2s_3 - 128s_2s_4 + 32s_3s_4 + 298 \\ &= 2g(s_1, s_2, s_3, s_4) \end{aligned} \tag{6.7}$$

Hence, the function g is in the form of Ising Hamiltonian, and can be readily mapped into our system, where the local fields h_i and coupling coefficients J_{ij} are

$$[h_i]^T = \begin{pmatrix} 58 & 50 & 12 & -80 \end{pmatrix} \tag{6.8}$$

$$[J_{ij}] = \begin{pmatrix} & 25 & -6 & -64 \\ & & 2 & -64 \\ & & & 16 \\ & & & \end{pmatrix} \tag{6.9}$$

6.4.1 Neighboring and external interactions of biprime factorization in Ising formulation

In this section, we will discuss how to map the obtained Ising function with local external field h and coupling J onto our proposed Ising model. Fig. 6.4(b) illustrates the mapping of overall weight from the neighbors and the local field to a corresponding input current. Consider the spin state at s_i, we here define a term *total connection* W_{sum} from all the neighbors and its local field, by summing all the entries in row i of matrix $|h|$ and $|J|$ (h and J defined in Eq. 6.8 and 6.9), namely *weights*. For example, $W_{sum}(s_1)$ (i.e., total connection for spin s_1) is $58+(25+6+64)$. As mentioned in the previous section, the write operation current is limited to a range of 60μA (I_{range}). The magnitude of each current source from the local field and neighbors depends on the corresponding connection coefficient, and can be calculated as $I_{range} * |h_i|/W_{sum}$ and $I_{range} * |J_{ij}|/W_{sum}$, respectively. The current flow direction to write a spin to point up (+x direction, or '+1' here) or down (-x direction, or '-1' here) is opposite. As shown in Fig. 6.5, a current will flow from path 2 to switch a spin from '+1' to '-1', while follows path 1 to write a spin from '-1' to '1'. XNOR gates are used to control the switch while updating a spin at '+1' state; on the other hand, XOR gates are used for updating a spin at '-1' state. Take spin state (s_1) at '+1' and '-1' state again for example. The exerting force from local field and neighbors are +58,-25,+6,-64,

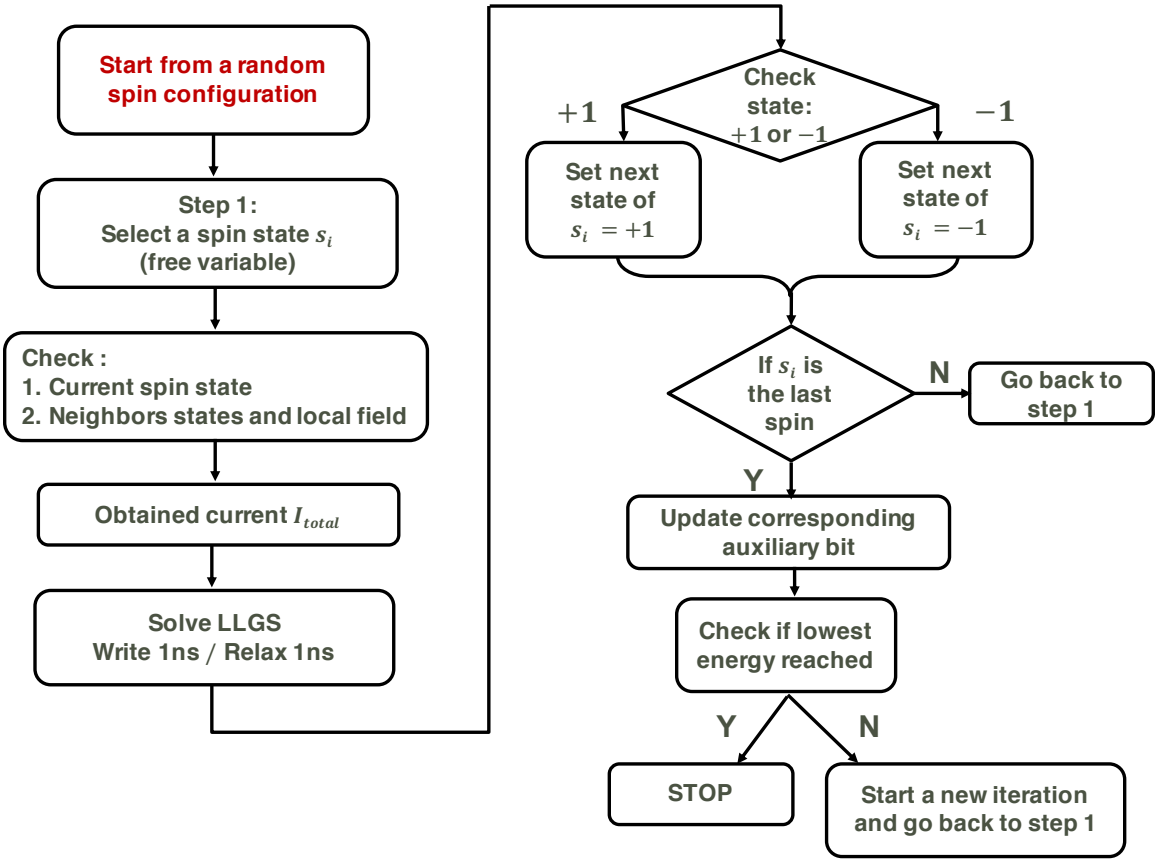

Fig. 6.7. The process of spin updates.

respectively. If spin s_1 is at '+1' state, current will flow through path 2 direction in Fig. 6.5, XOR gates are turned on only for current source associated with negative force, and current from only those branches will flow through the SHM. Likewise, if the spin state s_1 is at '-1', injection current will follow the path 1 direction, and XNOR gates are only turned on for the current sources corresponding to positive exerting force. Finally, the total injected current will have a one-to-one mapping onto the switching probability curve (P_{sw}), shown in Fig. 6.4(b). Noted that as mentioned in previous section, we bias a current of I_{min} corresponding the a switching probability 2 % to avoid system from trapping at local minimum state. The time evolution of the update process is shown in Fig.6.7.

N = P * Q	143 (8 bit)	10403 (14 bit)	47053 (16 bit)	154433 (18 bit)
# of spins	15	55	71	109
spin (variable)	6	13	15	17
spin (auxiliary)	9	42	56	72
Avg. iterations	6.4	81	752	10961
Pure stochastic	~10	~150	~2450	~7500

Fig. 6.8. Simulation result of our proposed Ising cell to solve biprime factorization.

Fig. 6.8 shows our simulation results for biprime factorization, and the number of Ising cells required for different N in our system. The simulations were conducted 1000 times for each problem to obtain an average number of iterations to reach the minimum energy state. The transition of the system energy is monitored by checking the states of all the spins after each epoch to verify if the system found a solution. By mapping the bi-prime factorization to the Ising hardware, our simulation results highlight significant improvement compared to pure stochastic method [2] (i.e. random search) for the N binary bits (length) up to 16.

Finally, let us briefly discuss the energy consumption of each Ising cell used our proposed SHM-MTJ device. The operation of each iteration can be divided into three parts: write operation, relaxation time, and read operation. We assign a time period of 1 ns to each operation. By applying a voltage (V_{DD}) of 1 V for write operation, the energy consumption is evaluated to be $\sim$ 0.08 pJ. When considering the read current to be $\sim$ 38 μA and the applied sensing voltage (V_{DD}) to be 1V, the average energy consumption is $\sim$ 0.01 pJ. Also, $\sim$ 0.01 pJ energy consumption is contributed from the relaxation mode and CMOS switches.

[2]Pure stochastic means, we simply assume each spin state has 50% pointing up and 50% pointing down for each iteration.

6.5 Conclusion

In this chapter, we present factorization of a bi-prime number on an Ising hardware, in which the computational unit is realized using our proposed SHM-MTJ devices. By leveraging the stochastic switching behavior SOT-MTJ, the proposed device inherently mimics the required operation of an Ising model. Using coupled magnetization dynamics, our coupling network surpasses pure stochastic method in the number of iterations to factorize numbers up to 16 binary digits. For larger numbers ($\geq$18 bits), our scheme can still outpace other general computing methods from the perspective of hardware implementation (such as energy consumption and total runtime) despite requiring more iterations. These findings, along with other inherent properties such as simple circuit structure, scalability, and low-power dissipation, can potentially pave a way for future hardware realization of complex optimization problems.

7. SUMMARY

With an increasing demand for fast, energy-efficient, and low-cost memory in scaled technology, a variety of beyond-CMOS technologies have been widely researched. In this dissertation, we have proposed multiple spin-based devices for logic applications, memory design, and stochastic computing.

We first proposed lateral spin valve as a spin logic. SOT-DSL was investigated and we found higher performance can be achieved, leading to robust logic device. Subsequently, we shifted our research focus to the field of magnetic skyrmions. We explored magnetic skyrmions to design last-level caches and to emulate the core functionality of neurons and synapse for an All-Spin Spiking Deep Neural Network. In a multi-bit skyrmion-based cell design, the size and spacing of skyrmions can be down to nanometer scale, however, the high current density requirements for skyrmion nucleation is a bottleneck to achieve significant density benefits. We developed a device-to-system framework to evaluate the system-level benefits of the proposed design. Our experiments reveal considerable benefits over an iso-area SRAM cache. However, the performance is lower than an iso-area STT-MRAM cache, suggesting the need for mechanisms to lower the current density requirements for skyrmion nucleation. On the other hand, using magnetic skyrmion as a spiking neuron processor, we only need onetime nucleation in the proposed skyrmion-based neuron/synapse structure. To "write" a skyrmionic neuron/synapse is solely based on shifting a skyrmion in the long nanotrack, which only consumes a small amount of energy. As a result, we believe the ultra-low current switching of skyrmionic neuron/synapses, representing new possibilities for exploiting skyrmion-based devices in low-power Deep Spiking neural architectures.

Finally, we proposed a SHM-MTJ based Ising cell to solve bi-prime factorization problem. We demonstrated a novel approach to map this problem into an Ising

model. By leveraging the stochastic switching behavior of a nanomagnet driven by energy-efficient spin-orbit torque, our proposed device inherently mimics the underlying behaviors of an Ising computing primitive. The simple architecture and intrinsic connection to the Ising model empowers this computing model to be an attractive and implementable system for solving combinatorial optimization problems. The feasibility of the system has been demonstrated by factorizing up to 16-bit binary number in a short amount of time compared to random search method.

REFERENCES

REFERENCES

[1] G. E. Moore *et al.*, "Progress in digital integrated electronics," in *Electron devices meeting*, vol. 21, 1975, pp. 11–13.

[2] J. Shalf, "The future of computing beyond moores law," *Philos. Trans. R. Soc. A*, vol. 378, no. 2166, p. 20190061, 2020.

[3] K. Wang, J. Alzate, and P. K. Amiri, "Low-power non-volatile spintronic memory: Stt-ram and beyond," *J. Phys. D*, vol. 46, no. 7, p. 074003, 2013.

[4] D. Apalkov, A. Khvalkovskiy, S. Watts, V. Nikitin, X. Tang, D. Lottis, K. Moon, X. Luo, E. Chen, A. Ong *et al.*, "Spin-transfer torque magnetic random access memory (stt-mram)," *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, vol. 9, no. 2, pp. 1–35, 2013.

[5] S. Kang and K. Lee, "Emerging materials and devices in spintronic integrated circuits for energy-smart mobile computing and connectivity," *Acta Mater.*, vol. 61, no. 3, pp. 952–973, 2013.

[6] 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 *2009 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 2009, pp. 1–4.

[7] L. Sun, Y. Hao, C.-L. Chien, and P. C. Searson, "Tuning the properties of magnetic nanowires," *IBM Journal of Research and Development*, vol. 49, no. 1, pp. 79–102, 2005.

[8] X. Fong, Y. Kim, R. Venkatesan, S. H. Choday, A. Raghunathan, and K. Roy, "Spin-transfer torque memories: Devices, circuits, and systems," *Proc. IEEE*, vol. 104, no. 7, pp. 1449–1488, 2016.

[9] A. Driskill-Smith, D. Apalkov, V. Nikitin, X. Tang, S. Watts, D. Lottis, K. Moon, A. Khvalkovskiy, R. Kawakami, X. Luo *et al.*, "Latest advances and roadmap for in-plane and perpendicular stt-ram," in *2011 3rd IEEE International Memory Workshop (IMW)*. IEEE, 2011, pp. 1–3.

[10] G. Jeong, W. Cho, S. Ahn, H. Jeong, G. Koh, Y. Hwang, and K. Kim, "A 0.24-/spl mu/m 2.0-v 1t1mtj 16-kb nonvolatile magnetoresistance ram with self-reference sensing scheme," *IEEE J. Solid-State Circuits*, vol. 38, no. 11, pp. 1906–1910, 2003.

[11] S. Ikeda, K. Miura, H. Yamamoto, K. Mizunuma, H. Gan, M. Endo, S. Kanai, J. Hayakawa, F. Matsukura, and H. Ohno, "A perpendicular-anisotropy cofeb–mgo magnetic tunnel junction," *Nat. Mater.*, vol. 9, no. 9, pp. 721–724, 2010.

[12] M. Gajek, J. Nowak, J. Sun, P. Trouilloud, E. Osullivan, D. Abraham, M. Gaidis, G. Hu, S. Brown, Y. Zhu *et al.*, "Spin torque switching of 20 nm magnetic tunnel junctions with perpendicular anisotropy," *Appl. Phys. Lett.*, vol. 100, no. 13, p. 132408, 2012.

[13] T. Miyazaki and N. Tezuka, "Giant magnetic tunneling effect in fe/al2o3/fe junction," *J. Magn. Magn. Mater.*, vol. 139, no. 3, pp. L231–L234, 1995.

[14] S. Ikeda, J. Hayakawa, Y. Ashizawa, Y. Lee, K. Miura, H. Hasegawa, M. Tsunoda, F. Matsukura, and H. Ohno, "Tunnel magnetoresistance of 604% at 300 k by suppression of ta diffusion in co fe b/ mg o/ co fe b pseudo-spin-valves annealed at high temperature," *Appl. Phys. Lett.*, vol. 93, no. 8, p. 082508, 2008.

[15] W. Butler, X.-G. Zhang, T. Schulthess, and J. MacLaren, "Spin-dependent tunneling conductance of fe— mgo— fe sandwiches," *Phys. Rev. B*, vol. 63, no. 5, p. 054416, 2001.

[16] X. Fong, S. Gupta, N. Mojumder, S. Choday, C. Augustine, and K. Roy, "Knack: A hybrid spin-charge mixed-mode simulator for evaluating different genres of spin-transfer torque mram bit-cells," in *Simulation of Semiconductor Processes and Devices (SISPAD), 2011 International Conference on.* IEEE, 2011, pp. 51–54.

[17] J. C. Slonczewski *et al.*, "Current-driven excitation of magnetic multilayers," *J. Magn. Magn. Mater.*, vol. 159, no. 1, p. L1, 1996.

[18] L. Berger, "Emission of spin waves by a magnetic multilayer traversed by a current," *Phys. Rev. B*, vol. 54, no. 13, p. 9353, 1996.

[19] D. C. Ralph and M. D. Stiles, "Spin transfer torques," *J. Magn. Magn. Mater.*, vol. 320, no. 7, pp. 1190–1216, 2008.

[20] J. Slonczewski, "Currents, torques, and polarization factors in magnetic tunnel junctions," *Phys. Rev. B*, vol. 71, no. 2, p. 024411, 2005.

[21] J. Z. Sun, "Spin-current interaction with a monodomain magnetic body: A model study," *Phys. Rev. B*, vol. 62, no. 1, p. 570, 2000.

[22] J. C. Slonczewski, "Conductance and exchange coupling of two ferromagnets separated by a tunneling barrier," *Phys. Rev. B*, vol. 39, no. 10, p. 6995, 1989.

[23] A. Brataas and K. M. Hals, "Spin–orbit torques in action," *Nat. Nanotechnol.*, vol. 9, no. 2, p. 86, 2014.

[24] L. Liu, C.-F. Pai, Y. Li, H. Tseng, D. Ralph, and R. Buhrman, "Spin-torque switching with the giant spin hall effect of tantalum," *Science*, vol. 336, no. 6081, pp. 555–558, 2012.

[25] L. Liu, O. Lee, T. Gudmundsen, D. Ralph, and R. Buhrman, "Current-induced switching of perpendicularly magnetized magnetic layers using spin torque from the spin hall effect," *Phys. Rev. Lett.*, vol. 109, no. 9, p. 096602, 2012.

[26] G. Finocchio, M. Carpentieri, E. Martinez, and B. Azzerboni, "Switching of a single ferromagnetic layer driven by spin hall effect," *Appl. Phys. Lett.*, vol. 102, no. 21, p. 212410, 2013.

[27] G. Yu, P. Upadhyaya, Y. Fan, J. G. Alzate, W. Jiang, K. L. Wong, S. Takei, S. A. Bender, L.-T. Chang, Y. Jiang *et al.*, "Switching of perpendicular magnetization by spin–orbit torques in the absence of external magnetic fields," *Nat. Nanotechnol.*, vol. 9, no. 7, p. 548, 2014.

[28] I. M. Miron, G. Gaudin, S. Auffret, B. Rodmacq, A. Schuhl, S. Pizzini, J. Vogel, and P. Gambardella, "Current-driven spin torque induced by the rashba effect in a ferromagnetic metal layer," *Nat. Mater.*, vol. 9, no. 3, pp. 230–234, 2010.

[29] T. Suzuki, S. Fukami, N. Ishiwata, M. Yamanouchi, S. Ikeda, N. Kasai, and H. Ohno, "Current-induced effective field in perpendicularly magnetized ta/cofeb/mgo wire," *Appl. Phys. Lett.*, vol. 98, no. 14, p. 142505, 2011.

[30] X. Fan, J. Wu, Y. Chen, M. J. Jerry, H. Zhang, and J. Q. Xiao, "Observation of the nonlocal spin-orbital effective field," *Nat. Commun.*, vol. 4, no. 1, pp. 1–7, 2013.

[31] K.-S. Ryu, L. Thomas, S.-H. Yang, and S. Parkin, "Chiral spin torque at magnetic domain walls," *Nat. Nanotechnol.*, vol. 8, no. 7, p. 527, 2013.

[32] S. Emori, U. Bauer, S.-M. Ahn, E. Martinez, and G. S. Beach, "Current-driven dynamics of chiral ferromagnetic domain walls," *Nat. Mater.*, vol. 12, no. 7, pp. 611–616, 2013.

[33] I. M. Miron, T. Moore, H. Szambolics, L. D. Buda-Prejbeanu, S. Auffret, B. Rodmacq, S. Pizzini, J. Vogel, M. Bonfim, A. Schuhl *et al.*, "Fast current-induced domain-wall motion controlled by the rashba effect," *Nat. Mater.*, vol. 10, no. 6, pp. 419–423, 2011.

[34] P. Haazen, E. Murè, J. Franken, R. Lavrijsen, H. Swagten, and B. Koopmans, "Domain wall depinning governed by the spin hall effect," *Nat. Mater.*, vol. 12, no. 4, pp. 299–303, 2013.

[35] R. Liu, W. Lim, and S. Urazhdin, "Spectral characteristics of the microwave emission by the spin hall nano-oscillator," *Phys. Rev. Lett.*, vol. 110, no. 14, p. 147601, 2013.

[36] V. E. Demidov, S. Urazhdin, H. Ulrichs, V. Tiberkevich, A. Slavin, D. Baither, G. Schmitz, and S. O. Demokritov, "Magnetic nano-oscillator driven by pure spin current," *Nat. Mater.*, vol. 11, no. 12, pp. 1028–1031, 2012.

[37] L. Liu, C.-F. Pai, D. Ralph, and R. Buhrman, "Magnetic oscillations driven by the spin hall effect in 3-terminal magnetic tunnel junction devices," *Phys. Rev. Lett.*, vol. 109, no. 18, p. 186602, 2012.

[38] V. M. Edelstein, "Spin polarization of conduction electrons induced by electric current in two-dimensional asymmetric electron systems," *Solid State Commun.*, vol. 73, no. 3, pp. 233–235, 1990.

[39] J. Hirsch, "Spin hall effect," *Phys. Rev. Lett.*, vol. 83, no. 9, p. 1834, 1999.

[40] A. Hoffmann, "Spin hall effects in metals," *IEEE Trans. Magn.*, vol. 49, no. 10, pp. 5172–5193, 2013.

[41] Y. Ji, A. Hoffmann, J. Jiang, and S. Bader, "Spin injection, diffusion, and detection in lateral spin-valves," *Appl. Phys. Lett.*, vol. 85, no. 25, pp. 6218–6220, 2004.

[42] Y. Fukuma, L. Wang, H. Idzuchi, S. Takahashi, S. Maekawa, and Y. Otani, "Giant enhancement of spin accumulation and long-distance spin precession in metallic lateral spin valves," *Nat. Mater.*, vol. 10, no. 7, pp. 527–531, 2011.

[43] R. Venkatesan, V. Kozhikkottu, C. Augustine, A. Raychowdhury, K. Roy, and A. Raghunathan, "Tapecache: a high density, energy efficient cache based on domain wall memory," in *Proceedings of the 2012 ACM/IEEE international symposium on Low power electronics and design*. ACM, 2012, pp. 185–190.

[44] Z. Sun, W. Wu, and H. Li, "Cross-layer racetrack memory design for ultra high density and low power consumption," in *Proceedings of the 50th Annual Design Automation Conference*. ACM, 2013, p. 53.

[45] A. Ranjan, S. Ramasubramanian, R. Venkatesan, V. Pai, K. Roy, and A. Raghunathan, "Dyrectape: A dynamically reconfigurable cache using domain wall memory tapes," in *Proceedings of the 2015 Design, Automation & Test in Europe Conference & Exhibition*. EDA Consortium, 2015, pp. 181–186.

[46] R. Venkatesan, S. Ramasubramanian, S. Venkataramani, K. Roy, and A. Raghunathan, "Stag: Spintronic-tape architecture for gpgpu cache hierarchies," in *Computer Architecture (ISCA), 2014 ACM/IEEE 41st International Symposium on*. IEEE, 2014, pp. 253–264.

[47] A. Thiaville, Y. Nakatani, J. Miltat, and Y. Suzuki, "Micromagnetic understanding of current-driven domain wall motion in patterned nanowires," *EPL (Europhysics Letters)*, vol. 69, no. 6, p. 990, 2005.

[48] A. Fert, V. Cros, and J. Sampaio, "Skyrmions on the track," *Nat. Nanotechnol.*, vol. 8, no. 3, p. 152, 2013.

[49] N. Nagaosa and Y. Tokura, "Topological properties and dynamics of magnetic skyrmions," *Nat. Nanotechnol.*, vol. 8, no. 12, p. 899, 2013.

[50] R. Wiesendanger, "Nanoscale magnetic skyrmions in metallic films and multilayers: a new twist for spintronics," *Nat. Rev. Mater.*, vol. 1, no. 7, p. 16044, 2016.

[51] R. Tomasello, E. Martinez, R. Zivieri, L. Torres, M. Carpentieri, and G. Finocchio, "A strategy for the design of skyrmion racetrack memories," *Sci. Rep.*, vol. 4, p. 6784, 2014.

[52] I. Dzyaloshinsky, "A thermodynamic theory of weak ferromagnetism of antiferromagnetics," *Journal of Physics and Chemistry of Solids*, vol. 4, no. 4, pp. 241–255, 1958.

[53] T. Moriya, "New mechanism of anisotropic superexchange interaction," *Phys. Rev. Lett.*, vol. 4, no. 5, p. 228, 1960.

[54] S. Mühlbauer, B. Binz, F. Jonietz, C. Pfleiderer, A. Rosch, A. Neubauer, R. Georgii, and P. Böni, "Skyrmion lattice in a chiral magnet," *Science*, vol. 323, no. 5916, pp. 915–919, 2009.

[55] X. Yu, Y. Onose, N. Kanazawa, J. Park, J. Han, Y. Matsui, N. Nagaosa, and Y. Tokura, "Real-space observation of a two-dimensional skyrmion crystal," *Nature*, vol. 465, no. 7300, p. 901, 2010.

[56] X. Yu, N. Kanazawa, Y. Onose, K. Kimoto, W. Zhang, S. Ishiwata, Y. Matsui, and Y. Tokura, "Near room-temperature formation of a skyrmion crystal in thin-films of the helimagnet fege," *Nat. Mater.*, vol. 10, no. 2, p. 106, 2011.

[57] S. Heinze, K. Von Bergmann, M. Menzel, J. Brede, A. Kubetzka, R. Wiesendanger, G. Bihlmayer, and S. Blügel, "Spontaneous atomic-scale magnetic skyrmion lattice in two dimensions," *Nature Physics*, vol. 7, no. 9, p. 713, 2011.

[58] A. Fert, "Magnetic and transport properties of metallic multilayers," in *Materials Science Forum*, vol. 59. Trans Tech Publ, 1990, pp. 439–480.

[59] A. Imre, G. Csaba, L. Ji, A. Orlov, G. Bernstein, and W. Porod, "Majority logic gate for magnetic quantum-dot cellular automata," *Science*, vol. 311, no. 5758, pp. 205–208, 2006.

[60] M. T. Alam, M. J. Siddiq, G. H. Bernstein, M. Niemier, W. Porod, and X. S. Hu, "On-chip clocking for nanomagnet logic devices," *IEEE Trans. Nanotechnol.*, vol. 9, no. 3, pp. 348–351, 2010.

[61] P. Li, G. Csaba, M. Niemier, X. Sharon Hu, J. Nahas, W. Porod, and G. H. Bernstein, "Power reduction in nanomagnet logic using high-permeability dielectrics," *J. Appl. Phys*, vol. 113, no. 17, p. 17B906, 2013.

[62] B. Behin-Aein, D. Datta, S. Salahuddin, and S. Datta, "Proposal for an all-spin logic device with built-in memory," *Nat. Nanotechnol.*, vol. 5, no. 4, pp. 266–270, 2010.

[63] C. Augustine, G. Panagopoulos, B. Behin-Aein, S. Srinivasan, A. Sarkar, and K. Roy, "Low-power functionality enhanced computation architecture using spin-based devices," in *2011 IEEE/ACM International Symposium on Nanoscale Architectures*. IEEE, 2011, pp. 129–136.

[64] S. Manipatruni, D. E. Nikonov, and I. A. Young, "Energy-delay performance of giant spin hall effect switching for dense magnetic memory," *Appl. Phys. Express*, vol. 7, no. 10, p. 103001, 2014.

[65] V. Calayir, D. E. Nikonov, S. Manipatruni, and I. A. Young, "Static and clocked spintronic circuit design and simulation with performance analysis relative to cmos," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 2, pp. 393–406, 2013.

[66] K. Y. Camsari, S. Ganguly, and S. Datta, "Modular approach to spintronics," *Sci. Rep.*, vol. 5, p. 10571, 2015.

[67] S. Parkin, M. Hayashi, and L. Thomas, "Magnetic domain-wall racetrack memory," *Science*, vol. 320, no. 5873, pp. 190–194, 2008.

[68] F. Chen, Z. Li, W. Kang, W. Zhao, H. Li, and Y. Chen, "Process variation aware data management for magnetic skyrmions racetrack memory," in *Design Automation Conference (ASP-DAC), 2018 23rd Asia and South Pacific*. IEEE, 2018, pp. 221–226.

[69] R. Venkatesan, M. Sharad, K. Roy, and A. Raghunathan, "Dwm-tapestri-an energy efficient all-spin cache using domain wall shift based writes," in *Proceedings of the Conference on Design, Automation and Test in Europe*. EDA Consortium, 2013, pp. 1825–1830.

[70] J. Sampaio, V. Cros, S. Rohart, A. Thiaville, and A. Fert, "Nucleation, stability and current-induced motion of isolated magnetic skyrmions in nanostructures," *Nat. Nanotechnol.*, vol. 8, no. 11, p. 839, 2013.

[71] A. Thiele, "Steady-state motion of magnetic domains," *Phys. Rev. Lett.*, vol. 30, no. 6, p. 230, 1973.

[72] C. Bienia, S. Kumar, J. Singh, and K. Li, "The PARSEC Benchmark Suite: Characterization and Architectural Implications," in *Proc. PACT*, 2008, pp. 72–81.

[73] P. Lai, G. Zhao, H. Tang, N. Ran, S. Wu, J. Xia, X. Zhang, and Y. Zhou, "An improved racetrack structure for transporting a skyrmion," *Sci. Rep.*, vol. 7, p. 45330, 2017.

[74] G. Zhao, X. Zhang, and F. Morvan, "Theory for the coercivity and its mechanisms in nanostructured permanent magnetic materials," *Rev. Nanosci. Nanotechnol.*, vol. 4, no. 1, pp. 1–25, 2015.

[75] C. Hanneken, F. Otte, A. Kubetzka, B. Dupé, N. Romming, K. Von Bergmann, R. Wiesendanger, and S. Heinze, "Electrical detection of magnetic skyrmions by tunnelling non-collinear magnetoresistance," *Nat. Nanotechnol.*, vol. 10, no. 12, p. 1039, 2015.

[76] D. Maccariello, W. Legrand, N. Reyren, K. Garcia, K. Bouzehouane, S. Collin, V. Cros, and A. Fert, "Electrical detection of single magnetic skyrmions in metallic multilayers at room temperature," *Nat. Nanotechnol.*, vol. 13, no. 3, p. 233, 2018.

[77] X. Zhang, G. Zhao, H. Fangohr, J. P. Liu, W. Xia, J. Xia, and F. Morvan, "Skyrmion-skyrmion and skyrmion-edge repulsions in skyrmion-based racetrack memory," *Sci. Rep.*, vol. 5, p. 7643, 2015.

[78] M. Najafi, B. Krüger, S. Bohlens, M. Franchin, H. Fangohr, A. Vanhaverbeke, R. Allenspach, M. Bolte, U. Merkt, D. Pfannkuche *et al.*, "Proposal for a standard problem for micromagnetic simulations including spin-transfer torque," *J. Appl. Phys*, vol. 105, no. 11, p. 113914, 2009.

[79] A. Vansteenkiste, J. Leliaert, M. Dvornik, M. Helsen, F. Garcia-Sanchez, and B. Van Waeyenberge, "The design and verification of mumax3," *AIP Adv.*, vol. 4, no. 10, p. 107133, 2014.

[80] P. Metaxas, J. Jamet, A. Mougin, M. Cormier, J. Ferré, V. Baltz, B. Rodmacq, B. Dieny, and R. Stamps, "Creep and flow regimes of magnetic domain-wall motion in ultrathin pt/co/pt films with perpendicular anisotropy," *Phys. Rev. Lett.*, vol. 99, no. 21, p. 217208, 2007.

[81] L. Liu, T. Moriyama, D. Ralph, and R. Buhrman, "Spin-torque ferromagnetic resonance induced by the spin hall effect," *Phys. Rev. Lett.*, vol. 106, no. 3, p. 036601, 2011.

[82] "CACTI, www.hpl.hp.com/research/cacti."

[83] N. Binkert, B. Beckmann, G. Black, S. Reinhardt, A. Saidi, A. Basu, J. Hestness, D. Hower, T. Krishna, S. Sardashti, R. Sen, K. Sewell, M. Shoaib, N. Vaish, M. Hill, and D. Wood, "The gem5 simulator," *SIGARCH Computer Arch. News*, vol. 39, no. 2, pp. 1–7, Aug. 2011.

[84] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[85] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura *et al.*, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, vol. 345, no. 6197, pp. 668–673, 2014.

[86] S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, "Nanoscale memristor device as synapse in neuromorphic systems," *Nano Lett.*, vol. 10, no. 4, pp. 1297–1301, 2010.

[87] D. Kuzum, R. G. Jeyasingh, B. Lee, and H.-S. P. Wong, "Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing," *Nano Lett.*, vol. 12, no. 5, pp. 2179–2186, 2011.

[88] T. Tuma, A. Pantazi, M. Le Gallo, A. Sebastian, and E. Eleftheriou, "Stochastic phase-change neurons," *Nat. Nanotechnol.*, vol. 11, no. 8, pp. 693–699, 2016.

[89] A. Sengupta, Z. Al Azim, X. Fong, and K. Roy, "Spin-orbit torque induced spike-timing dependent plasticity," *Appl. Phys. Lett.*, vol. 106, no. 9, p. 093704, 2015.

[90] A. Sengupta, A. Banerjee, and K. Roy, "Hybrid spintronic-cmos spiking neural network with on-chip learning: Devices, circuits, and systems," *Phys. Rev. Appl.*, vol. 6, no. 6, p. 064003, 2016.

[91] A. Sengupta and K. Roy, "A vision for all-spin neural networks: A device to system perspective," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 63, no. 12, pp. 2267–2277, 2016.

[92] A. Sengupta, Y. Shim, and K. Roy, "Proposal for an all-spin artificial neural network: Emulating neural and synaptic functionalities through domain wall motion in ferromagnets," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 6, pp. 1152–1160, 2016.

[93] J. Iwasaki, M. Mochizuki, and N. Nagaosa, "Current-induced skyrmion dynamics in constricted geometries," *Nat. Nanotechnol.*, vol. 8, no. 10, pp. 742–747, 2013.

[94] J. Iwasaki, M. Mochizuki, and N. Nagaosa, "Universal current-velocity relation of skyrmion motion in chiral magnets," *Nat. Commun.*, vol. 4, no. 1, pp. 1–8, 2013.

[95] O. Boulle, J. Vogel, H. Yang, S. Pizzini, D. de Souza Chaves, A. Locatelli, T. O. Menteş, A. Sala, L. D. Buda-Prejbeanu, O. Klein *et al.*, "Room-temperature chiral magnetic skyrmions in ultrathin magnetic nanostructures," *Nat. Nanotechnol.*, vol. 11, no. 5, pp. 449–454, 2016.

[96] Y. Huang, W. Kang, X. Zhang, Y. Zhou, and W. Zhao, "Magnetic skyrmion-based synaptic devices," *Nanotechnology*, vol. 28, no. 8, p. 08LT02, 2017.

[97] Z. He and D. Fan, "A tunable magnetic skyrmion neuron cluster for energy efficient artificial neural network," in *2017 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2017, pp. 350–355.

[98] S. Li, W. Kang, Y. Huang, X. Zhang, Y. Zhou, and W. Zhao, "Magnetic skyrmion-based artificial neuron device," *Nanotechnology*, vol. 28, no. 31, p. 31LT01, 2017.

[99] P. U. Diehl, D. Neil, J. Binas, M. Cook, S.-C. Liu, and M. Pfeiffer, "Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing," in *Neural Networks (IJCNN), 2015 International Joint Conference on*. IEEE, 2015, pp. 1–8.

[100] K. Sawada, T. Uemura, M. Masuda, K.-i. Matsuda, and M. Yamamoto, "Tunneling magnetoresistance simulation used to detect domain-wall structures and their motion in a ferromagnetic wire," *IEEE Trans. Magn.*, vol. 45, no. 10, pp. 3780–3783, 2009.

[101] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[102] S.-Z. Lin, C. Reichhardt, C. D. Batista, and A. Saxena, "Driven skyrmions and dynamical transitions in chiral magnets," *Phys. Rev. Lett.*, vol. 110, no. 20, p. 207202, 2013.

[103] S.-Z. Lin, C. Reichhardt, C. Batista, and A. Saxena, "Particle model for skyrmions in metallic chiral magnets: Dynamics, pinning, and creep," *Phys. Rev. B*, vol. 87, no. 21, p. 214419, 2013.

[104] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Commun. ACM*, vol. 21, no. 2, pp. 120–126, 1978.

[105] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM Review*, vol. 41, no. 2, pp. 303–332, 1999.

[106] H. Goto, K. Tatsumura, and A. R. Dixon, "Combinatorial optimization by simulating adiabatic bifurcations in nonlinear hamiltonian systems," *Sci. Adv.*, vol. 5, no. 4, p. eaav2372, 2019.

[107] F. Barahona, "On the computational complexity of ising spin glass models," *J. Phys. A*, vol. 15, no. 10, p. 3241, 1982.

[108] B. A. Cipra, "An introduction to the ising model," *The American Mathematical Monthly*, vol. 94, no. 10, pp. 937–959, 1987.

[109] M. Yamaoka, C. Yoshimura, M. Hayashi, T. Okuyama, H. Aoki, and H. Mizuno, "A 20k-spin ising chip to solve combinatorial optimization problems with cmos annealing," *IEEE J. Solid-State Circuits*, vol. 51, no. 1, pp. 303–309, 2015.

[110] Y. Shim, A. Jaiswal, and K. Roy, "Ising computation based combinatorial optimization using spin-hall effect (she) induced stochastic magnetization reversal," *J. Appl. Phys*, vol. 121, no. 19, p. 193902, 2017.

[111] B. Sutton, K. Y. Camsari, B. Behin-Aein, and S. Datta, "Intrinsic optimization using stochastic nanomagnets," *Sci. Rep.*, vol. 7, no. 1, pp. 1–9, 2017.

[112] B. Behin-Aein, V. Diep, and S. Datta, "A building block for hardware belief networks," *Sci. Rep.*, vol. 6, no. 1, pp. 1–10, 2016.

VITA

VITA

Mei-Chin Chen received the B.S. degree in Electrophysics from National Chiao Tung University, Taiwan, in 2012. Currently, she is pursuing Ph.D. degree in electrical engineering at Purdue University, under the guidance of Prof. Kaushik Roy since Fall 2013. Her primary research focus is in the area of device/circuit/architecture co-design of embedded systems using emerging technologies. She is also exploring the possible applications of magnetic skyrmion-based devices in memory and neuromorphic computing. Her work has resulted in five journal publications and some conference presentations. She was selected as a recipient of the Studying Abroad Scholarship from the Education Ministry of Taiwan for her academic performance.