This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2017.2774291, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems

Design and Evaluation of a Spintronic In-Memory Processing Platform for Non-Volatile Data Encryption

Shaahin Angizi, *Student Member, IEEE,* Zhezhi He, *Student Member, IEEE,* Nader Bagherzadeh, *Fellow, IEEE,* and Deliang Fan, *Member, IEEE*

***Abstract*—In this paper, we propose an energy-efficient reconfigurable platform for in-memory processing based on novel 4-terminal spin Hall effect-driven domain wall motion devices that could be employed as both non-volatile memory cell and in-memory logic unit. The proposed designs lead to unity of memory and logic. The device to system level simulation results show that, with 28% area increase in memory structure, the proposed in-memory processing platform achieves a write energy $\sim$ 15.6 fJ/bit with 79% reduction compared to that of SOT-MRAM counterpart while keeping the identical 1ns writing speed. In addition, the proposed in-memory logic scheme improves the operating energy by 61.3%, as compared with the recent non-volatile in-memory logic designs. An extensive reliability analysis is also performed over the proposed circuits. We employ Advanced Encryption Standard (AES) algorithm as a case study to elucidate the efficiency of the proposed platform at application level. Simulation results exhibit that the proposed platform can show up to 75.7% and 30.4% lower energy consumption compared to CMOS-ASIC and recent pipelined domain wall (DW) AES implementations, respectively. In addition, the AES Energy-Delay Product (EDP) can show 15.1% and 6.1% improvements compared to the DW-AES and CMOS-ASIC implementations, respectively.**

***Index Terms*—Domain wall motion, spin Hall effect (SHE), in-memory processing platform, AES.**

I. Introduction

NOWADAYS, while the amount of big data is dramatically rising to exascale (10^{18} bytes or flops), many challenges in hardware design remain unsolved [1]. In conventional Von-Neumann computing systems, all the digital data are maintained within the memory units separated from the processing unit. Hence, in the execution phase, either instruction or data need to be fetched from the main memories or caches, transmitted to the processor and written back afterwards [2], [3]. Keeping pace with today's big data processing, the separation of memory and computing units interconnected via buses has faced serious challenges, such as long memory access latency, considerable congestion at I/Os, limited memory bandwidth, and huge leakage power consumption in big data-driven applications [4]. To address above issues, in-memory processing architectures and devices have been presented to integrate memory and logic, leading to a much more energy efficient information processing platform [5], [6]. The basic advantageous concept behind the in-memory computing is preprocessing the data and providing intermediate result for processor rather than feeding it large volume of raw data [7]. It involves synergistic exploration spanning from device technology to architecture innovation. From device technology perspective, there are many recent and promising research works carried out by the emerging non-volatile memories (NVMs) at nano-scale for realization of such in-memory computing platforms, such as phase-change memory (PCM) [8], resistive memory (ReRAM) [4], and spintronic memory [9]–[11].

Recently, several logic-in-memory and in-memory computing architectures associated with NVMs have been presented. A new in-memory computing platform based on STT-MRAM is proposed in [12]. Different full adder designs based on new logic-in-memory architectures have been introduced in [13], [14]. The authors in [3], [13] have proposed reconfigurable in-memory logic gates based on magnetic domain-wall racetrack memory and magnetic tunnel junction devices, respectively. The authors in [4] have proposed innovative in-memory processing architecture to accelerate Neural Network applications through ReRAM-based memory banks. An alternative H-tree in-memory processing architecture has been proposed in [1] at block level that is very efficient for reducing the traffic communications. This architecture efficiently pairs each data block with in-memory logic unit. A local data processing scheme is then employed to only provide processor intermediate results, greatly reducing the communication traffic between memory and processor and improving the energy efficiency.

Spintronic devices are among the most promising alternative technologies to overcome performance and power limitations of conventional CMOS technology. Unique features of spintronic devices, such as instance wake-up, non-volatility, zero static power and high integration density, are difficult to achieve using today's CMOS technology [15], [16]. Perhaps the most intriguing feature of such new devices relies on their potential to redesign existing Boolean computing platform [17] by utilizing a completely new class of design methods such as logic-in-memory or processor-in-memory [5], which may show orders of lower power compared to CMOS counterparts.

Spin Hall effect based device is generally treated as the third generation of spintronic technology. It utilizes the large spin-orbit torque (SOT), instead of traditional spin-transfer

Partial support of this research was provided by the National Science Fundation under Grant No. 1740126

S. Angizi, Z. He and D. Fan are with the Department of Electrical and Computer Engineering, University of Central Florida, Orlando, FL, 32816. E-mail: {angizi, elliot.he}@knights.ucf.edu, dfan@ucf.edu.

N. Bagherzadeh is with the Department of Electrical Engineering and Computer Science, University of California, Irvine, CA 92697. E-mail: nader@uci.edu.

0278-0070 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2017.2774291, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems

torque (STT), to switch the adjacent free layer magnetization or induce steady domain wall motion (DWM) much more efficiently due to much larger conversion efficiency from charge current to spin current [18], [19]. Such emerging non-volatile device could be a much more energy efficient candidate for the integration of memory and logic design.

In this paper, we initially show two composite device structures employing spin Hall effect-driven domain wall motion (SHE-DWM) and Magnetic Tunnel Junction (MTJ) devices. Accordingly, we leverage them to realize an energy efficient H-tree fashion in-memory processing platform by designing both non-volatile memory cell and in-memory logic units (named TLG and XOR). The major contributions of this work are listed as follows:

- Instead of adding excessive specific-purpose logic elements to memory die, a reconfigurable H-tree fashion in-memory processing platform is designed to govern trade-offs between memory and in-memory logic efficiency. Furthermore, the reconfigurability comes from the integrated in-memory logic units performing various logics such as AND/NAND, OR/NOR, and majority function,
- Two in-memory full adder (FA) circuits are designed using hybrid spin/CMOS circuits with different methods,
- A cross-layer (device to application level) simulation framework is built for evaluation and comparison of the proposed in-memory processing platform with recently reported designs in different aspects,
- A comprehensive reliability analysis of the proposed in-memory logic is performed, considering magnetic tunnel junction conductance variation, domain wall motion strip stochastic switching effects, and CMOS peripheral circuits reliability, and
- Advanced Encryption Standard (AES) data encryption algorithm is employed as a case study to elucidate the efficiency of proposed in-memory processing platform as an in-memory data encryption engine.

For clarification, this paper is an extended version of our previously published conference papers [11], [20], in which the initial in-memory processing circuits were designed based on DWM and spin Hall effect-driven DWM devices, respectively.

The rest of this paper is organized in the following manner. Section II elucidates the device structure and modeling of SHE-DWM devices. Section III details the proposed in-memory processing platform considering non-volatile memory cell design and in-memory logic circuits. In Section IV, the comprehensive performance evaluation of the proposed platform and state-of-the-art Spin/CMOS designs are provided. Section V introduces the mapping of AES encryption algorithm to the proposed in-memory processing platform. Section VI concludes this paper.

II. Four-terminal Spin Hall Effect-Driven Domain Wall Motion Devices

Fig. 1(a) and Fig. 1(b) illustrate the presented device structures [20] referred to as 4 Terminal Spin Hall Effect-driven Domain Wall Motion (4T SHE-DWM) and 4 Terminal Spin Hall Effect-driven Differential Domain Wall Motion (4T SHE-DDWM) devices, respectively.

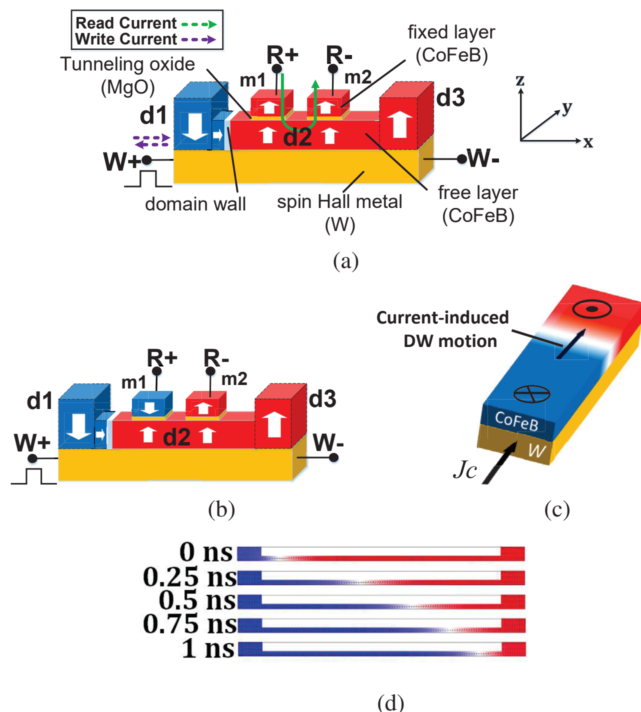


Figure 1. (a) 4T SHE-DWM device, (b) 4T SHE-DDWM device, (c) A cross-sectional schematic of the SHE-DWM device where an electrical current density J_c in the W generates a transverse spin current I_s exerting a fast speed and steady DWM, (d) Micro-magnetic simulation of spin Hall effect-driven DWM device.

Each design is a composite structure consisting of a Spin Hall Metal (SHM), a Domain Wall Motion (DWM) strip and two Magnetic Tunnel Junctions (MTJ) employing a perpendicularly magnetized W/CoFeB/MgO heterostructure [21]. DWM strip (i.e. d2) is laterally connected to two anti-parallel fixed magnetic domains (d1 and d3). MTJ is formed through a MgO layer sandwiched between fixed magnetic layer - m1 or m2 and the 'free' DWM strip - d2. The only structural difference between first and second designs is the MTJ fixed layer magnetization alignment. As is clear, the first presented design (Fig. 1(a)) consists of two MTJs with parallel fixed layers. However, the second design (Fig. 1(b)) consists of two MTJs with anti-parallel fixed layers.

The resistance of the MTJ depends on the free layer magnetization, namely the domain wall (DW, i.e. the transition area between two domains in the DWM strip) positions within the DWM strip. Electric manipulation of domain wall was typically accomplished by the current induced spin-transfer torque (STT) due to the coupling between local magnetic moments of the DW and spin-polarized currents [15]. Recently, it has been experimentally demonstrated that domain wall motion could be more energy efficiently achieved through Spin Hall Effect (SHE) [18], [21]. The authors in [21] have studied deterministic magnetic reversal of a perpendicularly magnetized CoFeB layer driven by SHE from an in-plane current flowing in an underlying W layer. In the presented device structure, when input charge current passes through the non-magnetic spin Hall metal (SHM, Tungsten in this work) in lateral paths (±x), electrons with opposite spins scatter to opposite surfaces of SHM due to strong spin-orbit coupling. Thus, a spin current (I_s) perpendicular (±z) to charge current

0278-0070 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2017.2774291, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems

Table I
DEVICE PARAMETERS USED IN SIMULATION.

Symbol	Quantity	Values
α	Damping coefficient	0.3
K_u	Uniaxial anisotropy constant	$3.5 \times 10^5 J/m^3$
M_s	Saturation magnetization	$6.8 \times 10^5 A/m$
A_{ex}	Exchange stiffness	$1.1 \times 10^{-11} J/m$
t_{MgO}	MgO thickness	1 nm
RA	MTJ Resistance area product	$2.38\Omega\mu m^2$
TMR_{AP}	Tunnel Magneto resistance	168%
ρ	Resistivity of magnet	$170\Omega nm$
ρ_{SHM}	Resistivity of SHM (W)	$200\mu\Omega cm$ [9]
θ_{SHM}	Spin Hall angle	0.3 [22]
$(L.W)_{MTJ}$	MTJ dimention	$20 \times 20nm^2$
$(L.W.t)_{DWM}$	DWM strip dimension	$100 \times 20 \times 1nm^3$
$(L.W.t)_{SHM}$	SHM dimension	$120 \times 20 \times 2.8nm^3$

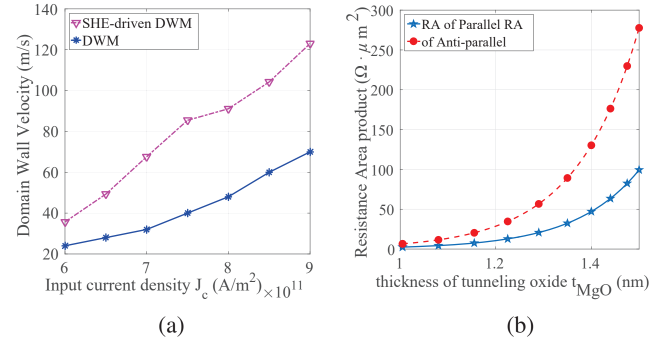


Figure 2. (a) DW velocity vs. lateral current density of SHE-driven DW and conventional domain wall motion, (b) Resistance-area product vs. the thickness of tunneling oxide in AP and P states.

(I_c) is generated owing to the SHE [15], which will be leveraged to induce steady and fast domain wall motion along the input current direction [21] as shown in Fig. 1(c).

The magnetization dynamics, m, of a nano-magnet with $N_s x$ as the number of spins per domain in x-direction in the presence of an effective magnetic field, H_{eff}, and a spin current, I_s, is modeled using the Landau–Lifshitz–Gilbert (LLG) equation as follow [15]:

$$\frac{\partial m}{\partial t} = -|\gamma|(m \times H_{eff}) + \alpha(m \times \frac{\partial m}{\partial t}) - \frac{I_s}{qN_{sx}}\frac{\partial m}{\partial x} + \xi\frac{I_s}{qN_{sx}} m \times \frac{\partial m}{\partial x} \quad (1)$$

In this equation, the first two terms denote the usual precession and damping terms, respectively. The third term is the local tracking of conduction electrons to local magnetization and the fourth term describes a phenomenological non-adiabatic spin-transfer term whose strength is described by ξ. According to SHE, the generated spin current can be written as

$$I_s = \theta_{SHM}\frac{A_s}{A_c}I_c\sigma \quad (2)$$

where θ_{SHM} is the spin Hall angle characterizing the strength of SHE in SHM, σ is the polarization of the electron spin and A_s and A_c are the cross-sectional areas through which spin and charge current flow, respectively.

In the proposed device structures, the free layer dimension is ($100nm \times 20nm \times 1nm$), so a Néel type DW is formed due to the small strip width (20nm) [15]. The larger thickness at the edges of the DWM strip is used to stabilize the DW at an intermediate position [15], [23]. In addition, in order to have a better controllability and thermal stability over DW movement within domain wall strip, three artificial trapping sites can be considered in the left, middle and right end of DWM strip [6]. MTJ is employed to read the state of DWM strip. The transient micro-magnetic simulation of DW position (achieved from OOMMF [24]) is illustrated in Fig. 1(d), using device dimension listed in Table I, from 0 to 1ns. Since the magnetization of DWM strip beneath the two MTJs is fully switched at 1ns, the intrinsic threshold current (I_{th}) of this device can be considered $39.2\mu A$ within 1ns corresponding to DW velocity of $\sim 75m/s$. Fig. 2(a) depicts and compares the DW velocity vs. lateral current density of SHE-driven DW presented herein and conventional domain wall motion device (with same dimension without SHM layer) based on micro-magnetic simulations. It can be seen that the application of SHE induces a much higher DW velocity at the same input current density.

The presented heterostructures can function as 4-terminal devices with completely decoupled write and read terminals. For the lateral write path ($\pm x$), the magnetization of DWM strip can be identified anti-parallel (AP) or parallel (P) to the fixed layer of MTJs (m1 and m2) by injecting a small current (larger than critical current) along SHM terminals (W+ to W- or vice-versa). Hence, the resistance states are different values based on either high (corresponding to AP) or low (corresponding to P) configuration of MTJs and can be read by injecting a small sensing current through R+ to R- terminals employing a sense circuit. It is worth pointing out that the sense current ($\sim 1\mu A$) is significantly less than the DW critical depinning current so that the state of the read MTJs is not disturbed during read operation. MTJ resistance can be expressed in terms of voltage, tunneling oxide thickness (t_{MgO}), and the angle between free layer and fixed layer magnetizations. The atomistic level experimental benchmarked simulation framework based on Non-Equilibrium Green's Function (NEGF) formalism [23] is used to evaluate the MTJ resistance. The MTJ resistance-area product vs. MgO thickness in AP and P states is plotted in Fig. 2(b) with a constant voltage of 50mV. It shows that resistance-area product exponentially increases with the increase of tunnel oxide thickness.

The presented SHE-DWM devices supports following pivotal operations: (1) DW motion can be achieved through SHE, (2) DW motion can be precisely controlled by the magnitude and direction of the laterally applied current, with the assistance of notches on the domain wall nano-strip and (3) the MTJs mounted on top of the domain wall strip can have configurable resistances within the sensing path. The aforementioned design concepts have been experimentally demonstrated by the following works [18], [21], [25], [26]. In order to simulate the presented devices with CMOS interface circuits in SPICE, 4T SHE-DWM device is modeled as two MTJs with variable resistance depending on DW positions and equivalent resistive network can be written as Eq. (3).

$$R_{MTJ} = R_{m1} + R_{FL} + R_{m2} = \begin{cases} 2RA_P/A_{MTJ} + R_{FL} & \text{DW at left,} \\ 2RA_{AP}/A_{MTJ} + R_{FL} & \text{DW at right,} \\ (RA_{AP} + RA_P)/A_{MTJ} + R_{FL} & \text{DW at middle.} \end{cases} \quad (3)$$

0278-0070 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2017.2774291, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems

where R_{m1} and R_{m2} represent resistance of two read MTJs which are shown by m1 and m2, respectively, R_{FL} is the lateral free layer resistance between the two read MTJs, RA_{AP} and RA_P denote MTJ Resistance-Area (RA) product for anti-parallel and parallel configurations, respectively, obtained in the NEGF based MTJ model [23] and A_{MTJ} is the read MTJ area. As is clear in Eq. (3), the output resistance can have three different values based on the DW positions. However, the output resistance (from R+ to R-) of 4T SHE-DDWM device can also be expressed as follows:

$$R_{MTJ} = R_{m1} + R_{FL} + R_{m2} = \begin{cases} (RA_P + RA_{AP})/A_{MTJ} + R_{FL} & \text{DW at left (right),} \\ 2RA_P/A_{MTJ} + R_{FL} & \text{DW at middle.} \end{cases} \quad (4)$$

It can be seen that second device only has two distinct resistance levels based on its DW positions and the output resistances are identical when the DW is positioned in the left and right end. Compared to existing 3-terminal SHE-DWM device structure [19], the proposed 4T SHE-DWM devices with 2 mounted MTJs can provide three and two levels of resistance shown in Eq. (3) and (4) (compared to only two-level resistance of 3-terminal SHE-DWM). Therefore, they can be potentially leveraged to implement hybrid spin-CMOS logic circuits that could not be implemented by previous 3-terminal device as detailed in following sections.

III. Proposed In-memory Processing Platform

In this section, we propose a new reconfigurable architecture for realizing a streamlined and efficient in-memory processing platform for non-volatile data-encryption. This new platform is a distinct solution from either early or recent in-memory processing works [1], [4]. Instead of adding excessive specific-purpose logic elements to memory die, in our design, a part of in-memory logic units can also be employed as memory units to increase the memory capacity, leading to ultra-small area overhead. Generally speaking, a higher in-memory logic to memory units ratio offers higher execution throughput by sacrificing the storage efficiency and vice versa. This can be accomplished by simple modification of the improvised in-memory logic peripheral circuits. As shown in Fig. 3, each H-tree fashion in-memory processing subarray is divided into two data/logic blocks with eight embedded units. Each block mainly consists of four memory units as well as four in-memory logic units (i.e. TLG and XOR). The presented in-memory Threshold Logic Gate (TLG) units can work in two distinct operation modes i.e. Computing Mode and Memory Mode. In the computing mode, these units can efficiently serve as functional cells to perform basic logic operations within memory along with in-memory XOR unit without integrating complex logic circuits into memory. We have developed specialized in-memory XOR units to handle dominant XOR/XNOR operations in encryption and decryption algorithms. In the memory mode, either the memory units (M) or in-memory TLG units (TLG) have the storage capability. acting as typical non-volatile memory array. The reconfigurability can be used to govern the ratio of the logic and data

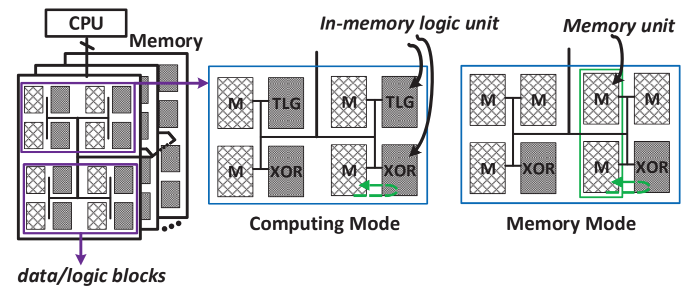


Figure 3. Architecture of the proposed in-memory processing platform in two different operation modes (i.e. computing and memory).

storage resources according to the application requirements. The following subsections elaborates the micro-architecture and circuit design of the integrated units in the proposed in-memory processing platform.

A. Memory Unit Design Based on 4T SHE-DWM Device

In this subsection, an efficient one-bit memory cell based on 4T SHE-DWM device is designed. Note that only two resistance levels (corresponding to DW located in the left end and right end) are used in memory cell design as shown in Fig. 4(a). The read MTJ resistivity status is utilized for representation of stored data. In this way, when DW is located in the leftmost end, it stores "1" (low resistance). When DW is positioned in the rightmost end of DWM strip, it stores "0" (high resistance). We have considered a decoupled write and read terminals for our presented memory cell to overcome the reliability issues associated with traditional STT-MRAM with shared read and write path design [10]. Besides, two mounted MTJs can provide $2R_{AP(/P)}$+R_{FL} resistances in sensing path compared to $R_{AP(/P)}$ +R_{FL} in conventional 3-terminal design [19] that increases sensing margin to reliably distinguish the resistance level.

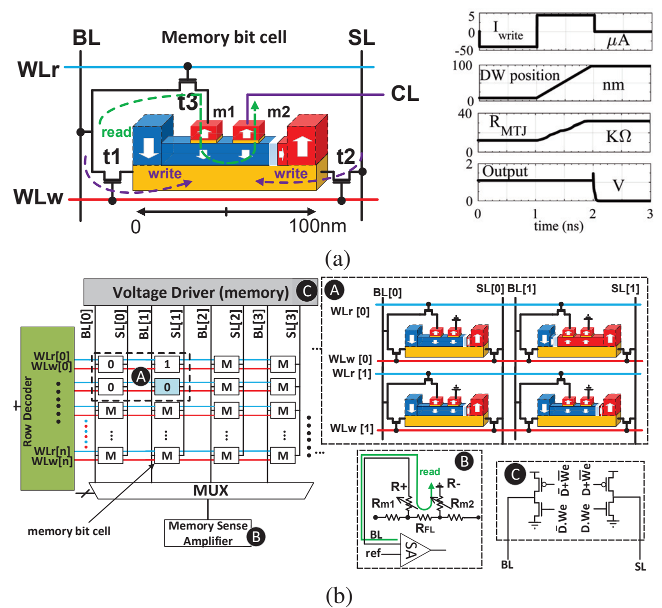


Figure 4. (a) Proposed memory cell design based on 4T SHE-DWM device and transient plot, (b) Illustration of memory sub-array architecture used in memory unit (M unit) with peripheral circuitry.

0278-0070 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2017.2774291, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems

As depicted in Fig. 4(a), the write and read operations can be accomplished by applying appropriate voltages to source-line (SL), the bit-line (BL), word-line write (WLw), and word-line read (WLr) where Common-line (CL) is grounded. Fig. 4(b) shows the memory sub-array architecture used in memory units along with a 2×2 array structure (A). The detailed peripheral circuitry used in read and write paths are also shown in Fig. 4(b) (B) and Fig. 4(b) (C), respectively. In order to accomplish *read operation*, the state of m1 and m2 MTJs should be read by setting WLr to a high voltage through Row Decoder and the read voltage (V_{read}) is applied via BL. Since m2 is grounded and WLw is low, two write access transistors (t1 and t2) are turned off and a sense current (I_{read}) will flow from m1 to m2. Eventually, the memory cell content is determined after comparing the sensing current to a reference current by a memory sense amplifier (SA) unit (Fig. 4(b) (B)). The sense current in read operation is unidirectional, while a bidirectional current is required to write data. In order to perform *write operation*, WLw is applied with VDD to turn on two write access transistors (t1 and t2) and a write voltage is applied across the BL or SL using write circuitry (Fig. 4(b) (C)). To write “0” (i.e. $V_{BL} - V_{SL} = V_{wr} - 0$), the write current is injected from W+ to W- and the DW will be pushed to the right end. To write “1” (i.e. m1 and m2 are in parallel states, DW located in left end), the voltage polarity across the bit-cell is reversed (i.e. $V_{BL} - V_{SL} = 0 - V_{wr}$). Based on our micromagnetic simulation, a $\pm 39.2\mu A - 1ns$ current pulse can move the DW from one end to the other. As shown in transient plot of Fig. 4(a), for writing “0”, a $+42.1\mu A (> 39.2\mu A)$ current is applied to the device, accordingly DW position is changed from left to right end (from 0 to 100nm) leading to a change in amount of R_{MTJ} (in Eq. (3)). It can be seen that the memory content is read using memory sense amplifier after 2ns. The complete operation of the presented memory cell is tabulated in Table II.

Table II
COMPLETE OPERATION OF MEMORY-BIT CELL.

Operation	WLw	WLr	BL	SL	CL	t1	t2	t3
Read	0	VDD	I_{sen}	0	0	off	off	on
Write 0	VDD	0	V_{wr}	0	0	on	on	off
Write 1	VDD	0	0	V_{wr}	0	on	on	off

B. In-Memory Logic Units Based on 4T SHE-DWM and 4T SHE-DDWM Devices

1) *Hybrid Spin-CMOS Threshold Logic Gate:* In this part, we present a hybrid spin-CMOS threshold logic gate (TLG) design employing 4T SHE-DWM device to efficiently implement Boolean functions such as 3-input majority gate (MG), 2-input AND/NAND, and 2-input OR/NOR gates. A TLG essentially constitutes of summation of weighted inputs, followed by a thresholding operation as expressed in Eq. (5). [6].

$$Y = \begin{cases} 1, & if \ \sum_i^n IN_i \cdot W_i - \theta \geq 0 \\ 0, & if \ \sum_i^n IN_i \cdot W_i - \theta < 0 \end{cases} \quad (5)$$

In Eq. (5), IN_i's are binary inputs, W_i's are scalar weights and θ is the threshold. The TLG output is “1” only if the weighted summation of binary inputs is greater or equal than the threshold. The same TLG circuit can implement different Boolean functions by reconfiguring the weights, threshold, or both. Table III shows the truth table of a 3-input TLG. Considering $W_i = 1$, $\theta = 2$ and input A as a “Bias” pin, different logic gates such as 3-input majority gate (MG), 2-input OR/AND gates can be implemented using the same 3-input TLG circuit. For instance, a 2-input AND gate can be readily implemented by setting Bias=“0” and a 2-input OR gate can be achieved by setting Bias=“1”.

Table III
TRUTH TABLE OF 3-INPUT TLG.

Inputs			3-input TLG		
A/Bias	B	C	$\sum$	3-input MG	
0	0	0	0<2	0	**2-input AND**
0	0	1	1<2	0	
0	1	0	1<2	0	
0	1	1	$2 \geqslant 2$	1	
1	0	0	1<2	0	**2-input OR**
1	0	1	$2 \geqslant 2$	1	
1	1	0	$2 \geqslant 2$	1	
1	1	1	$3 \geqslant 2$	1	

Fig. 5 shows memory sub-array architecture used in TLG units. In order to provide TLG unit with proper functioning and also enable morphing between computing and memory modes, peripheral circuitry is modified and enhanced compared to M unit. Sensing component consists of a Differential Latch (used in computing mode) and a Sense Amplifier (used in memory mode) that could be selected according to decoded command coming from control unit (Fig. 5 (A)). Voltage driver integrated into each memory sub-array is adjusted by control unit to assign different voltages to BL and SL by voltage multiplexing. Besides, control unit can configure memory columns so that more than one BL/SL can be simultaneously selected. We used the approach proposed in [27] for multiple line selecting shown in Fig. 5 (B). The mechanism to write (/load) data into memory cells is similar to M units. It can be accomplished by activating WLw (/Load in Fig. 5) and accordingly applying different voltage polarities to SL and BL. After loading data, 3 consecutive memory cells located in a memory row (Fig. 5 (C)) can be selected by row decoder for computation. As shown in Fig. 5(C) and (D), our presented 3-input spin-CMOS TLG circuit mainly consists of two components, corresponding to two steps of a TLG, namely *Weighted Summation* and *Thresholding*. The TLG can be efficiently executed in the following steps:

(1) Weighted Summation: When Load is high (Load=VDD), the operand A (or Bias), B and C can be loaded into the corresponding 4T SHE-DWM devices (indicated by Device #1 to #3) using the method and circuit described earlier similar to that of writing into memory cells. Hence, the conductance of the 4T SHE-DWM devices can be either high or low based on the operand values. After the operands are loaded, a voltage pulse (Clk1, 0.4V-1ns) is applied on the R+ terminal of each device through the BL, simultaneously. Therefore, the current coming out of R- is the weighted current, either low or high depending on operand values. For example, as shown for Device #1, “A=0”, so the corresponding conductance from R+ to R- is low, leading to “$I_A = low$”. For Device #2,

0278-0070 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2017.2774291, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems

Table IV
TRUTH TABLE OF THE TWO-INPUT AND/OR GATES EMPLOYING TLG UNIT AND THE CORRESPONDING RESISTANCE VALUES FOR EACH LEVEL.

	Inputs			Weighted Summation Component				Thresholding	
	A/Bias	B	C	Device #1	Device #2	Device #3	I_{sum}	Device #4	Latch Output (Out)
2-input AND gate	0	0	0	$2R_{AP}+R_{FL}$	$2R_{AP}+R_{FL}$	$2R_{AP}+R_{FL}$	Low	$R_{MTJ}<R_{ref}$	0
	0	0	1	$2R_{AP}+R_{FL}$	$2R_{AP}+R_{FL}$	$2R_P+R_{FL}$	Low	$R_{MTJ}<R_{ref}$	0
	0	1	0	$2R_{AP}+R_{FL}$	$2R_P+R_{FL}$	$2R_{AP}+R_{FL}$	Low	$R_{MTJ}<R_{ref}$	0
	0	1	1	$2R_{AP}+R_{FL}$	$2R_P+R_{FL}$	$2R_P+R_{FL}$	High	$R_{MTJ}>R_{ref}$	1
2-input OR gate	1	0	0	$2R_P+R_{FL}$	$2R_{AP}+R_{FL}$	$2R_{AP}+R_{FL}$	Low	$R_{MTJ}<R_{ref}$	0
	1	0	1	$2R_P+R_{FL}$	$2R_{AP}+R_{FL}$	$2R_P+R_{FL}$	High	$R_{MTJ}>R_{ref}$	1
	1	1	0	$2R_P+R_{FL}$	$2R_P+R_{FL}$	$2R_{AP}+R_{FL}$	High	$R_{MTJ}>R_{ref}$	1
	1	1	1	$2R_P+R_{FL}$	$2R_P+R_{FL}$	$2R_P+R_{FL}$	High	$R_{MTJ}>R_{ref}$	1

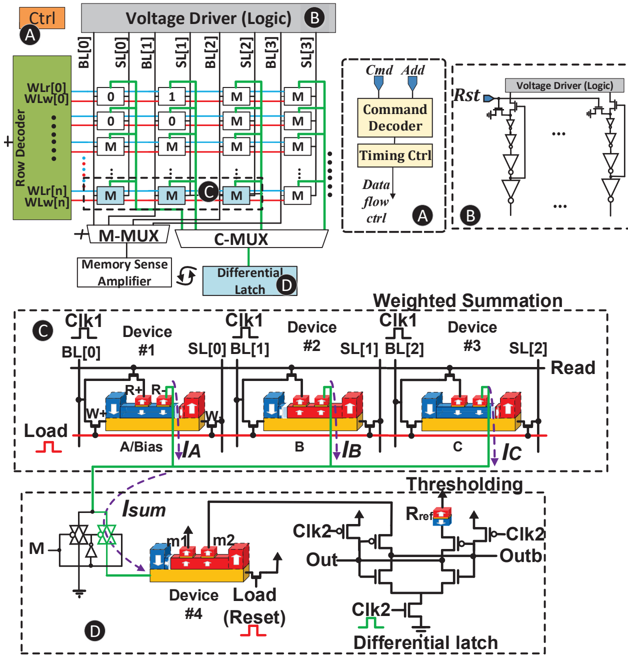


Figure 5. Illustration of memory sub-array architecture used in TLG unit with modified peripheral circuitry for realization of Hybrid spin-CMOS 3-input TLG design.

"B=1", so the corresponding conductance from R+ to R- is high, leading to "$I_B = high$". Then the weighted summation current ($I_{sum} = I_A + I_B + I_C$) is achieved at the common node.

(2) Thresholding: I_{sum} flows into the W+ terminal of the 4T SHE-DWM thresholding device (Device #4), programming the DW position (i.e. read MTJs resistance). In order to provide proper reconfigurability for TLG unit, a 2:1 MUX is embedded in Thresholding component (Fig. 5 (D)) to connect CL of selected memory cells either to GND (for memory mode) or thresholding device (for computing mode) based on decoded command coming from control unit. Considering M selector is set to VDD, Thresholding component is activated. Assuming the DW is initially located in the left end, thus the two read MTJs are in parallel states ($2R_P$). If I_{sum} is greater than the critical current (i.e. minimum current required to move the domain wall from left end to right end within 1ns), both two read MTJs switch to anti-parallel states ($2R_{AP}$). If not, two read MTJs are either both in parallel states ($2R_P$) or only m1 switches to anti-parallel state ($R_{AP} + R_P$) based on the current magnitude.

A differential latch is then used to read the states of read MTJs (R_{MTJ}) with one current branch connecting from m1 to m2 and the other current branch through a reference MTJ (with resistance value of $1.5R_{AP}+0.5R_P$). It is worth pointing out that proper clocking of circuit prevents backward injection of current from differential latch to weighted summation devices by removing the required potential to move the DW. In summary, the latch output (out) is high when I_{sum} is greater than the critical current ($\sim 39.2\mu A$) and vice versa. In this manner, 3-input MG and 2-input AND/OR gates can be implemented. Truth table of the two-input AND/OR gates employing presented TLG structure and the corresponding resistance values of each level is tabulated in Table IV. Obviously the complementary output of differential latch (Outb as shown in Fig. 5) can be readily used for implementing NAND/NOR gates. It is worth mentioning that the discussed approach for realizing the 3-input TLG can be generalized for higher input TLG circuits without any circuit's parameter modifications. In this way, the number of device used in Weighted Summation level needs to be increased according to the desired input number and the Thresholding component remains unchanged.

2) *Hybrid Spin-CMOS XOR/XNOR Design:*

In this subsection, we show a compact in-memory XOR logic gate using a single 4T SHE-DDWM device as shown in Fig. 6, greatly reducing the latency, power consumption and area. In order to have a precise control of DW pinning/depinning operation and good thermal stability, three notches are manufactured within the free layer magnetic nanostrip, located in left, middle and right ends [26].

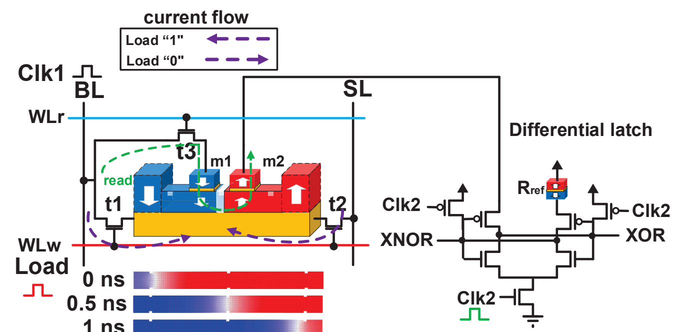


Figure 6. The presented 2-input XOR/XNOR gate based on 4T SHE-DDWM device and micro-magnetic simulations.

Based on our micro-magnetic simulation, a $\pm 39.2\mu A - 0.5ns$ current pulse can move DW from one pinning site to the neighboring pinning site. We define the current flow from W+ to W- to write "0" and current flow from W- to W+ to write

0278-0070 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2017.2774291, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems

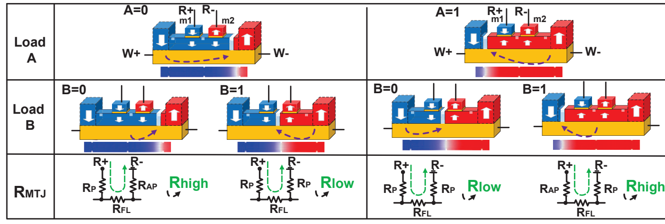


Figure 7. Intuitive illustration of micromagnetic simulations of presented 2-input XOR/XNOR gate.

"1" as shown in Fig. 6. DW is initially located in the middle notch (corresponding to parallel configurations for both m1 and m2). Input operands are sequentially loaded to 4T SHE-DDWM device, by applying $\pm 39.2\mu A - 0.5ns$ current pulse along the lateral writing path. For instance, if "(A,B)=(0,1)", the DW firstly moves to the right pinning site due to loading "A=0". Then, the spin current due to loading "B=1" will push DW to its left neighboring pinning site. As shown in Fig. 7, the final DW positions for 4 possible input combinations AB=(00, 01, 10, 11) are (right, middle, middle, left) corresponding to R= $(R_{high}, R_{low}, R_{low}, R_{high})$. R_{high} and R_{low} are achieved based on m1 and m2 MTJs configurations as mentioned in (4). A differential latch is then used to read the two read MTJ resistance states with one current branch from m1 to m2 and the other current branch through a reference MTJ. Therefore, the latch output is (0, 1, 1, 0), successfully realizing a 2-input XOR gate. Table V shows the truth table of the two-input XOR gate and the corresponding resistance values for each input combination.

Table V
TRUTH TABLE OF THE TWO-INPUT XOR GATE AND THE CORRESPONDING RESISTANCE VALUES FOR EACH INPUT COMBINATION.

Inputs		Corresponding device's resistance			
A	B	Load A	Load B	R_{MTJ}	XOR Output
0	0	$R_P + R_{FL} + R_{AP}$	$R_P + R_{FL} + R_{AP}$	$R_{MTJ} > R_{ref}$	0
0	1	$R_P + R_{FL} + R_{AP}$	$2R_P + R_{FL}$	$R_{MTJ} < R_{ref}$	1
1	0	$R_{AP} + R_{FL} + R_P$	$2R_P + R_{FL}$	$R_{MTJ} < R_{ref}$	1
1	1	$R_{AP} + R_{FL} + R_P$	$R_{AP} + R_{FL} + R_P$	$R_{MTJ} > R_{ref}$	0

3) *Hybrid Spin-CMOS Non-Volatile Adder:* In this part, we propose two different design approaches for realization of in-memory addition using the proposed platform:

(1) In the first design approach shown in Fig. 8(a), a hybrid spin-CMOS Full Adder (FA) is designed employing the presented in-memory XOR structure in conjunction with 3-input TLG circuit. As expressed in Eq. (6) and Eq. (7), the Carry output (C_{out}) can be directly obtained using 3-input TLG circuit representing a 3-input MG and Sum output can be achieved using two cascaded 2-input XOR gates, respectively. Due to the non-volatility of the proposed 1-bit full adder, an N-bit serial adder connecting the carry-out to carry-in can be readily designed as shown in Fig. 8(b) [14]. Such design does not sacrifice the operation latency due to the fact that the higher bit should wait the carry-out signal from low bit. Thus, an N-bit adder can be implemented by employing only one single 1-bit non-volatile full adder without extra overheads, leading to greatly reduced area and power consumption, while maintaining almost same throughput [14].

$$C_{out} = AB + AC + BC \tag{6}$$

$$Sum = A \oplus B \oplus C \tag{7}$$

(2) Fig. 8(c) depicts a simplified illustration of the second approach for in-memory addition only employing presented in-memory TLG unit. Letting $M1, M2$, and $M3$ as the inputs, the principle Boolean expressions of FA Sum can be derived using only the 3- and 5-input MGs as follows:

$$C_{out} = MG(M1, M2, M3) \tag{8}$$

$$Sum = MG(M1, M2, M3, \overline{C_{out}}, \overline{C_{out}}) \tag{9}$$

In-memory addition can be performed through a four-step process based on aforementioned equation using two rows of in-memory TLG unit. (1) The majority function of data in the three cell (M1,M2 and M3) in the first row is computed and stored in M/F4 cell which represents C_{out} of FA. (2) Read operation is performed to readout the M/F4 cell content using differential latch. (3) The complementary output of latch (Outb) is simultaneously used to inverse the C_{out} and the result is consequently written to Ma and Mb cells in the second row. (4) The majority function of five cells (M1, M2, M3, Ma and Mb) is computed and stored in the second row's M/F4 cell representing Sum of FA.

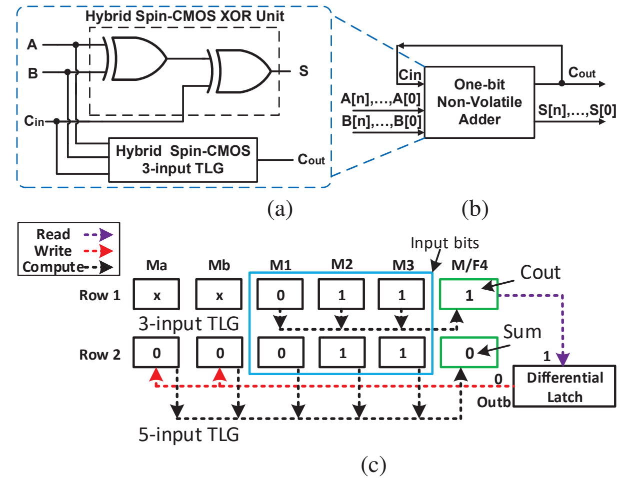


Figure 8. (a) Schematic representation of a non-volatile full adder employing both 4T SHE-DWM and 4T SHE-DDWM devices (first design approach), (b) N-bit serial adder structure based on the proposed one-bit adder, (c) Four-step addition employing only 4T SHE-DWM device (second design approach).

IV. PERFORMANCE EVALUATION

In order to evaluate the performance of the proposed in-memory processing platform, we come up with a comprehensive simulation framework as shown in Fig. 9. For device level simulation, we benchmarked the spin Hall effect-based domain wall motion dynamics with experimental data [28] utilizing Object Oriented MicroMagnetic Framework (OOMMF) [24]. The MTJ (constituted of DWM strip, tunneling oxide layer and fixed ferromagnetic layer) is modeled in Verilog-A, using

0278-0070 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2017.2774291, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems

NEGF-LLG (non-equilibrium Green's function and Landau-Lifshitz-Gilbert equations) solution for spin to charge interface and calibrated with data in [23]. For the circuit level simulation, a Verilog-A model of 4T SHE-DWM and 4T SHE-DDWM devices is developed to co-simulate with the interface CMOS circuits in Cadence Spectre and SPICE. 45nm North Carolina State University (NCSU) Product Development Kit (PDK) library [29] is used in SPICE to verify the proposed design and acquire the performance (Energy dissipation and reliability analysis) of designs.

For the system level simulations, we employ a modified self-consistent NVSim [30] along with an in-house developed C++ code to verify the performance of memory mode of proposed in-memory processing platform and to report the accurate power and area. For the application level, we take the Advanced Encryption Standard (AES) algorithm as an example and elucidate the operations of proposed in-memory processing platform in comparison to previously reported designs employing Synopsys Design Compiler, system level processor power evaluating tool McPAT [31] and cycle-accurate architecture simulator gem5 [32].

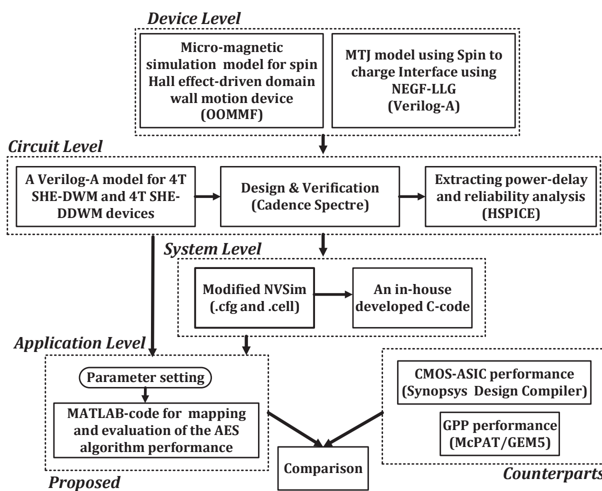


Figure 9. Device to application level co-simulation framework.

A. Device and Circuit Level Evaluation

We evaluate and compare the device and circuit level performance of presented in-memory processing model with the recently reported designs with respect to the logic and memory.

1) *Memory Unit:* Table VI tabulates the quantitative comparison of presented memory-bit cell and recently published works in different point of views. The presented memory cell exhibits three distinct advantages over the previously reported MTJ or DWM-based designs. First and foremost, the write operation with SHE is much more energy-efficient [9], since the spin current generation efficiency is much larger. The device-circuit SPICE simulation results listed in Table VI indicate that the write energy of the presented memory cell is $\sim$ 15.6 fJ/bit which is one order less than standard STT-MRAM in [10]. It also shows a 79% reduction in write energy compared to the recent SOT-MRAM design [9] while keeping the identical 1ns writing speed. We have also compared the write energy of presented memory cell with the DWM-based RAM design in [11] with similar dimensions. As expected, the application of SHE could reduce the write energy by 27.4%. Note that, the read energy and latency of these four different non-volatile 1-bit memory cell designs are almost similar due to the same sensing scheme, namely reading the MTJ resistance. Second, the write current directly flows through SHM rather than the tunneling oxide of MTJ. As a result, high write current can be injected to obtain fast switching, but avoiding reliability concerns associated with the tunnel barrier [9]. Third, in comparison to STT-MRAM design, a smaller access transistor width is required for providing the write current since the SHM has a lower resistance than MTJ. It is noteworthy that in STT-MRAM cell [10], high write speed (1ns) design requires much larger access transistor width (1.035μm) and boosted voltage (V_{wr}=1.2V) to provide sufficient write current. However, in our presented SHE-DWM based memory cell, a much smaller transistor width (115 nm) and low write voltage (V_{wr}=0.4V) is enough to provide the required write current (39.2μA) for 1ns write time.

Table VI
PERFORMANCE COMPARISON OF MEMORY CELLS.

Memory Attributes	Proposed	SOT-MRAM [9]	STT-MRAM [10]
Bit cell TMR	166% (current sensing)	172% (voltage sensing)	205% (current sensing)
Read/Write Voltage	1.0 V/0.4 V	1.0 V/0.4 V	0.25V/1.2 V
Switching Current Density	7.5 MA/cm^2 @ 1ns	7 MA/cm^2 @ 1ns	7.4 MA/cm^2 @ 5ns
Read Frequency	1 GHZ	1 GHZ	1 GHZ
Access Transistor width	115 nm	120 nm	1035 nm
Read Energy/Bit @ Tread= 1ns	1.1 fJ	1.1 fJ	0.9 fJ
Write Energy/Bit @ Twrite= 1ns	15.6 fJ	77 fJ	744 fJ

2) *TLG and XOR Units:* Fig. 10 shows the SPICE transient simulation of the presented hybrid spin-CMOS 3-input TLG. Three complementary clocks with 1ns pulse width are used in the circuit level simulation. For each input combination, we evaluate four distinct parameters: (1) summation of current or I_{sum} at intersection point, (2) domain wall position, (3) R_{MTJ} (i.e. the series resistance of m1 and m2 in 4T SHE-DWM thresholding device) and (4) output indicating the differential latch output (out). The first clock is "load" clock, which is used to load the operand "A", "B" and "C" by programming the corresponding 4T SHE-DWM device conductance. Meanwhile, a reset current ($-39.2\mu A$) is also generated to initialize the DW into the left end of the 4T SHE-DWM thresholding device, ready for next cycle operation. When "Clk1" is on, a small voltage ($\sim$400 mV) is applied at the R+ terminals of 4T SHE-DWM weighing devices and I_{sum} is determined by the operand (A, B, C) values, leading to the change of R_{MTJ}. As shown in the transient simulation, for operand "A", "B" and "C" with values of (000, 001, 110, 111), the corresponding I_{sum} are (24.4, 32.1, 42.2, 51.8)μA. Since the threshold current is 39.2μA, the corresponding R_{MTJ} are (12.22, 19.36, 27.86, 31.92)$k\Omega$, leading to the differential latch output as (0, 0, 1, 1) in the next "Clk2" cycle.

As mentioned earlier, the presented 3-input TLG design can be easily reconfigured just by considering "A" as a "Bias"

0278-0070 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2017.2774291, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems

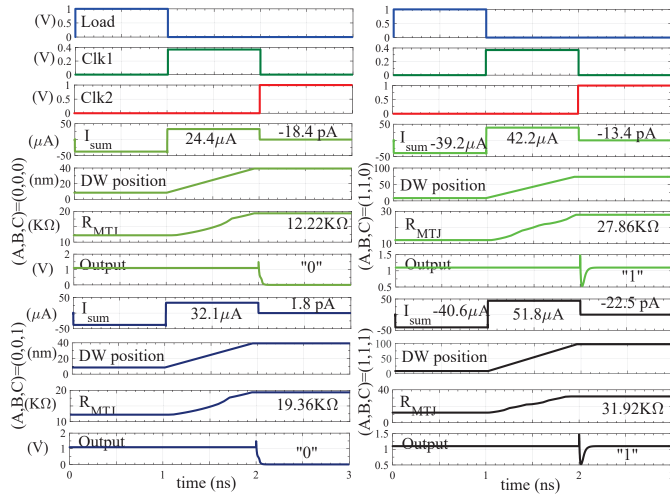


Figure 10. SPICE transient simulation of 3-input TLG.

shown in Table IV. Comparison result between the recent in-memory non-volatile 2-input AND gates referred to as NVLs (non-volatile logic) and the CMOS counterpart is listed in Table VII. Note that, NVL1 [3] and NVL3 [13] are domain wall racetrack-based logics and NVL2 [33] is an MTJ-based logic.

Table VII
PERFORMANCE OF 2-INPUT AND GATES.

	Operation Energy		Operation Speed	
Circuits	Computation	Computation & Read	Computation	Computation & Read
Proposed	25.1 fJ	26.2 fJ	2 ns	3 ns
NVL1	66.41 fJ	67.72 fJ	∼1 ns	1.12 ns
NVL2	121.5 fJ	125.85 fJ	∼1.02 ns	∼1.18 ns
NVL3	∼501 fJ	504.36 fJ	∼2.02 ns	∼2.14 ns
CMOS	-	6.69 fJ	-	62 ps

The operation energy (speed) of the non-volatile logic can be referred to the energy (time) to write the data to the NVMs and to perform the computation. However, a more accurate evaluation should consider the time required to read the data from non-volatile device. Therefore, we have compared the performance of different circuits in two distinct cases (i.e. computation/computation and read). As it can be seen, the proposed in-memory TLG unit improves the operating energy by 61.3% as compared with the best reported memory-based NVL in [3]. However, the operation speed is obviously less than domain wall racetrack-based logics due to their intrinsically faster computation (only two write operations) and shift.

A comparison between CMOS and Diode-GSHE (Giant Spin Hall Effect) full adders presented in [14] at 22nm technology and our SHE-DWM based full adders is presented in Table VIII. The performance comparison shows that the power consumption of the proposed FAs based on 1st and 2nd design approach is less than the other circuits. Such considerable power efficiency improvement mainly comes from the low voltage used in computation and small critical current of 4T SHE-DWM and 4T SHE-DDWM devices. Besides, our simple XOR gate design only requires one single 4T SHE-DDWM device. Note that, owing to the back end of line fabrication process of hybrid spin-CMOS circuits, we only consider the transistor count to approximate the area of each design shown in Table VIII.

Table VIII
PERFORMANCE EVALUATION OF FA CELLS.

Parameters	1st design	2nd design	CMOS [14]	Diode-GSHE [14]
Power Consumption (μW)	14.65	13.1	49.4	15.6
Complexity (transistor count)	37	33	42	20

3) *Reliability Analysis:* In this subsection, we will analyze the reliability of the presented in-memory TLG circuit, in which process variation or thermal noise can be contributed from both device (MTJ's conductance/ DWM strip) and peripheral CMOS circuits (transistors). As discussed in the mathematical expression of TLG in Eq. (5), the output of hybrid spin-CMOS 3-input TLG is "1" when the summation of current (I_{sum}) is larger than or equivalent to the intrinsic threshold of SHE-DWM device. Therefore, the probability of an erroneous output is much higher when (I_{sum}) is equal or close to (I_{th}) owning to device variation or thermal noise. A countermeasure to variation tolerance close to the threshold is mentioned in [34] by considering defect tolerance factors as expressed below:

$$Y = \begin{cases} 1, & if \ \sum_i^n IN_i \cdot W_i - \theta \geq \delta_{on} \\ 0, & if \ \sum_i^n IN_i \cdot W_i - \theta \leq -\delta_{off} \end{cases} \quad (10)$$

where δ_{on} and δ_{off} denote defect tolerances that should be considered due to temperature variation, manufacturing defects, and and etc. Correspondingly, in our work, we define the current margins between I_{th} and I_{sum} as $I_{delta_{on}}$ and $I_{delta_{off}}$, indicated in Fig. 11. It shows a possible I_{sum} and I_{th} distribution due to the process variation and thermal noise. The overlap of such distribution causes erroneous outputs. Ideally, if $I_{delta_{on}}$=$I_{delta_{off}}$, the variation tolerance could be maximized.

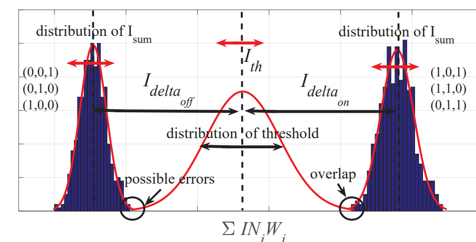


Figure 11. Representation of normal distribution of summation current (I_{sum}) for more sensitive input combinations considering threshold current (I_{th}), $I_{delta_{on}}$ and $I_{delta_{off}}$.

A detailed reliability analysis over the performance of hybrid spin-CMOS 3-input TLG is performed in 3 steps.

(1) The main goal of first step is to observe the functional Error Rate (ER) coming from the variation of I_{sum} due to process variation of MTJs and then find the optimized defect tolerance factor. In order to evaluate the ER, three primary Boolean functions listed in Table III are taken into account.

Monte Carlo statistical analysis is performed using HSPICE tool with a Gaussian distributed variation (3σ = 0% to 70%) added to weighted summation device's conductance (i.e. MTJ). In this way, 36 equally-distributed samples of conductance variation between 0% and 70% are selected. Accordingly, we run the Monte Carlo simulation for 1000 times for each

0278-0070 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2017.2774291, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems

variation sample considering all possible input combinations. The variation simulation result is depicted in Fig. 12(a). As shown, there is no erroneous output when the conductance variation is less than 5% where the average ER of different functions increases with rising of conductance variation. We should make sure that considering 10% variation (typical MTJ conductance variation [35]), the error rate for different Boolean functions is still zero. As the inset histogram plot shows in Fig. 12(a), the original I_{th} is 39.2μA. It leads to unequal values of $I_{delta_{on}}$ and $I_{delta_{off}}$, which is the main reason a relative larger error rate is observed. There are two solutions to address these issues. The first one is to shift I_{th} to achieve a balanced $I_{delta_{on}}$ and $I_{delta_{off}}$. It could be seen in Fig. 12(a) that the error rate reduces greatly after shifting I_{th}. A zero error rate is achieved at typical 10% MTJ conductance variation. Since I_{th} is mainly determined by the SHE-DWM device dimension and parameters, it is difficult to adjust after fabrication. Another feasible solution which is applied in this work is to tune the computation voltage (i.e. V_{Clk1}) to shift the distribution of I_{sum} as shown in Fig. 12(b). Specifically, increasing the computation voltage from 400mV to 412mV could achieve balanced $I_{delta_{on}}$ and $I_{delta_{off}}$.

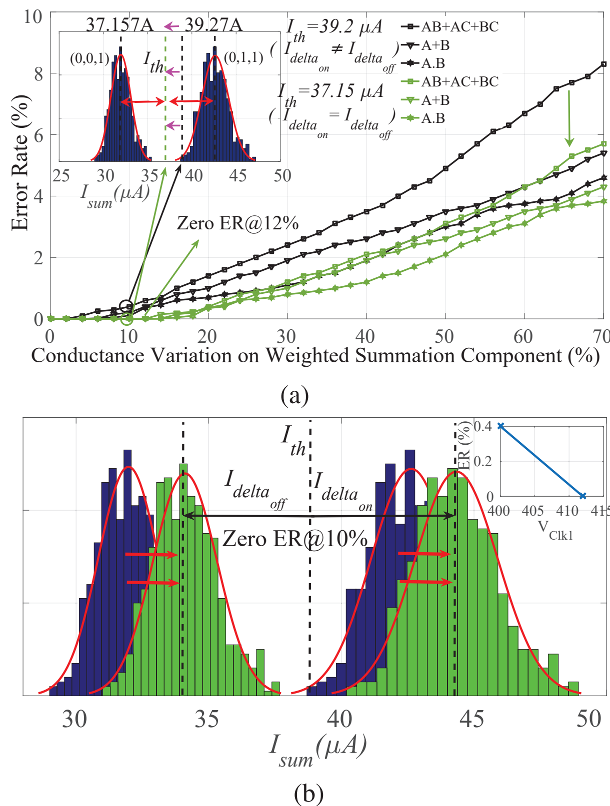


Figure 12. The average Error Rate (ER) of TLG-based logic functions vs. conductance variation of weighted summation component. The inset histogram plots show the distribution of I_{sum} for two input combinations. By shifting I_{th}, the corresponding ER is greatly reduced, (b) A balanced $I_{delta_{on}}$ and $I_{delta_{off}}$ is achieved by increasing the computation voltage.

(2) The second step considers both MTJ conductance variation as well as DWM strip stochastic switching effects. As thoroughly explained in [36], when the driving current of DWM strip is well above the deterministic intrinsic threshold, thermal perturbations has a negligible effect on the DW velocity that is found to depend approximately linearly on driving current similar to zero temperature case. In our presented logic circuit, the threshold current is the minimum current required to move domain wall from one end to the other end, leading to a domain wall velocity of ∼75m/s. Such driving current is much larger than the DWM critical current. The thermal noise effect on our presented circuit reliability is significantly smaller than the process variation of MTJs. Still, we consider 5% variation on threshold current of DWM to incorporate the effect of thermal noise. A similar 1000 times Monte Carlo simulation was conducted and the simulation results shows almost **zero** error rate at a typical 10% MTJ variation.

(3) The circuit level thermal noise can be defined as $I_n^2 = 4KTf/R$ [37], where K is the Boltzmann constant ($1.38 \times 10^{-23} J/K$), T is the temperature ($= 300K$), R is the resistance of the component and f is the frequency of spintronic thresholding component, which is $1/3$ GHz in our design. Therefore, for hybrid spin-CMOS 3-input TLG, the root mean square of noise current $I_{rms}(\sigma)$, at spintronic thresholding component is $\sim 0.95\mu A$. The current margin achieved in this work is about $5\mu A$ which is $\sim 5\times I_{rms}$.

The reliability analysis confirms that considering MTJ conductance variation, DWM strip stochastic switching effects and circuit level thermal noise, the presented in-memory TLG unit shows almost **zero** error rate at a typical 10% MTJ variation. Obviously, this robustness is even more crucial when it comes to large scale in-memory computations such as in-memory data encryption presented in the next section.

B. System Level Evaluation

Fig. 13(a) depicts the memory array organization modeled in this work. In order to evaluate the memory performance of proposed in-memory processing platform, we configure the memory chip by dividing it into multiple Banks consisting of multiple Mats. Each Mat includes multiple sub-arrays organized in an H-tree routing manner discussed in Section III. The external sensing scheme using shared SAs is then employed to improve the area efficiency of our design.

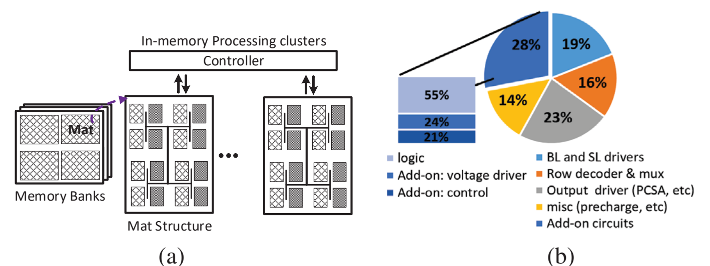


Figure 13. (a) Memory organization of the proposed platform, (b) Area overhead in MAT structure.

For the simulation, we employ modified self-consistent NVSim [30] along with an in-house developed C++ code to verify the architecture level performance and to report the accurate power and area. Table IX lists our proposed in-memory processing platform working in memory mode validation for a sample 128 KB memory with 512 word-width in 45nm process node.

0278-0070 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2017.2774291, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems

Table IX
PROPOSED MEMORY MODEL EVALUATION.

Metrics	Write	Read
Dynamic Energy	880.395pJ	890.174pJ
Mat Dynamic Energy	19.997pJ per mat	22.442pJ per mat
Subarray Dynamic Energy	4.893pJ per active subarray	5.504pJ per active subarray
Leakage Power	45.510mW per mat	
Area Overhead	18.450mm^2	

Fig. 13(b) shows the breakdown of area overhead for the proposed in-memory processing platform in a Mat structure. There is 28% area increase in memory die for supporting in-memory logic. The proposed reconfigurability offers ∼30 % area saving as compared to H-tree in-memory processing model in [1].

V. IN-MEMORY DATA-ENCRYPTION AS A CASE STUDY

In this section, we take the Advanced Encryption Standard (AES) algorithm as an example to elucidate the mapping of transformations in the proposed in-memory platform, which reveals its benefits of energy-efficiency and high throughput for in-memory data encryption applications. AES is an iterative symmetric-key cipher where both sender and receiver units use a single key for encryption and decryption. AES basically works on the standard input length of 16 bytes (128 bits) data organized in a 4×4 matrix (called the state matrix (S_M)) while using three different key lengths (128, 192, and 256 bits) [38]. For 128-bit key length, AES encrypts the input data after 10 rounds of consecutive transformations. These transformations as depicted in the flowchart in Fig. 14 are enumerated as SubBytes, ShiftRows, MixColumns, and AddRoundKey.

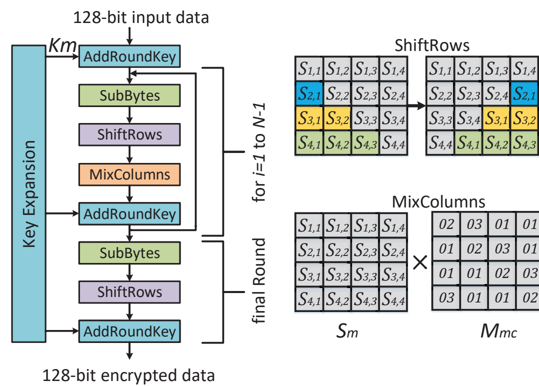


Figure 14. AES block diagram along with schematic representation of ShiftRows and MixColumns transformations.

A. Mapping of AES Transformations

To facilitate working with input data as depicted in Fig. 15(a), each byte in the input data is distributed into 8-bit. In order to increase the parallelism for the AES algorithm, we found that the number of employed memory units as well as in-memory XOR units need to be increased. Therefore, the memory mode of in-memory processing platform is activated. Three levels of parallelism, referred to as $P1, P2$, and $P4$ henceforth, can be considered in the proposed platform according to the state matrix's data organization. In the least possible parallelism ($P1$) shown in Fig. 15(a), 16 rows of one memory unit are filled with the state matrix's data where each memory row is filled with one byte. For $P2$, the data organization is different where two memory units are simultaneously occupied by 2 rows of S_M.

1) *SubBytes:* In SubBytes stage, each byte of S_M will undergo a Look-up table (LUT) based transformation using S-box and will be independently updated by a nonlinear transformation f ($S_{i,j} \leftarrow f(S_{i,j})$). As depicted in Fig. 15(b), input of S-Box LUT (16×16 memory array) is essentially a Byte which is divided into two 4-bit data patterns. Each pattern yields a row or a column index for the decoders reaching target cell in S-box. Then, the addressed data byte in S-Box is written back to the memory unit and substitutes the original data. As shown in $P2$ of Fig. 15(b), by employing two LUTs at the same time both read and both write back operations can be accomplished, simultaneously. The number of cycles (i.e. t_{SB}) used in the SubBytes stage based on our proposed in-memory processing platform can be derived as follows:

$$t_{SB} = (t_{read} + t_{LUT} + t_{write}) \times \frac{16}{N_M} \tag{11}$$

where the t_{read}, t_{write}, and t_{LUT} are the read, write, and LUT access latency, respectively (in terms of clock cycles). The $N_M \in \{1, 2, 4\}$ represents the number of specific-purpose memory units (shown by M in Fig. 15(b)) that can be accessed simultaneously.

2) *ShiftRows:* In ShiftRows stage, S_M will undergo a cyclically shift operation by a certain offset. Algorithmically, the i-th row of S_M will be cyclically left shifted by i-1 bytes. Accordingly, the first row of state matrix is left unchanged. For the second to fourth rows, each byte is shifted by offsets of one to three, respectively. To perform the shift operation, one of the memory units is considered as a buffer to temporary save the readout data. In this way, after reading the data from second to fourth row (3 rows), they can be easily rewritten to the memory with desired order. Considering the write and read operations are not performed concurrently, the number of cycles used in the ShiftRows stage, (i.e. t_{SR}), can be achieved as follow:

$$t_{SR} = (2 \times t_{read} + 2 \times t_{write}) \times X \tag{12}$$

where $X \in \{4, 8, 12\}$ is the required coefficient corresponding to $P4, P2$ and $P1$ levels of parallelism.

3) *MixColumns:* In MixColumns stage, the state matrix will be multiplied by a preset matrix depicted in Fig. 14. The four bytes of each column of S_M are combined using an invertible linear transformation ($S_{i,j} \leftarrow M_{mc} \times S_{i,j}$). The prerequisite operations for this stage are addition, multiplication by two (times2) and multiplication by three (times3). The addition could be efficiently executed using our proposed in-memory XOR unit where 3 cycles are required for each computation. To maximize the efficiency of AES performance, we use a LUT-based transformation followed by XOR operations to implement times2, similar to the design in [7]. The times3 operation is defined as time2 result XOR with the original

0278-0070 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2017.2774291, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems

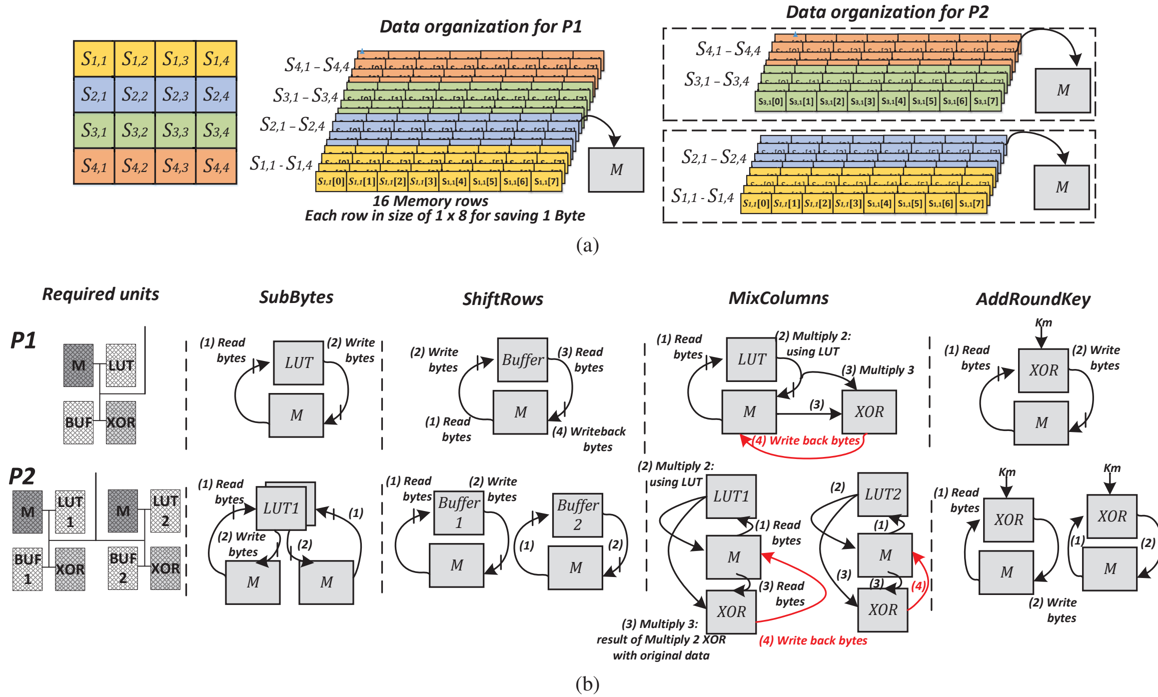


Figure 15. (a) Data organization for 2 levels of parallelism (i.e. $P1$ and $P2$) (b) Mapping of four AES transformations to the proposed in-memory processing platform corresponding to $P1$ and $P2$. As shown, the required units (including memory and XOR) for each level of parallelism are also indicated.

value as shown in Fig. 15(b). The number of clock cycle used in computation can be written as follow:

$$t_{MC} = (t_{read} + t_{LUT} + 3 \times t_{XOR} + t_{write}) \times \frac{16}{N_M} \tag{13}$$

4) *AddRoundKey:* In AddRoundKey stage, the subkey is combined with the state matrix. For each round, key expansion unit produces a subkey derived from the main key using Rijndael's key schedule [38]. The 16-byte round keys are organized in a similar 4×4 array (K_M) as the state matrix with $K_{i,j}$ as matrix entry. In this process, each byte of state matrix will be replaced by bit-wise XOR result of $S_{i,j}$ and $K_{i,j}$ (subkey's corresponding bit). This stage can be easily performed using the in-memory XOR unit. The number of cycles consumed in this stage can be calculated as follow:

$$t_{AK} = (t_{read} + t_{XOR} + t_{write}) \times \frac{16}{N_M} \tag{14}$$

As depicted in Fig.15(b), $P2$ takes the advantage of using two parallel XOR units, so each two Byte data (16-bit) of S_M can be processed concurrently.

B. AES Performance Evaluation

The performance comparison of different AES implementations with the proposed in-memory AES is tabulated in Table X. For evaluation of AES performance in general purpose processor (GPP), we use the similar method in [7]. AES C code is firstly extracted from [39] and compiled, then cycle-accurate architecture simulator gem5 [32] is employed to take AES binary and accordingly system level processor power evaluating tool McPAT [31] is used to estimate the power dissipation. For evaluation of AES in CMOS ASIC, Synopsys Design Compiler tool is used to run an in-house developed AES Verilog code. Note that, the same 30MHz clock frequency is used in all the hardware implementations listed in Table X, while the AES C code on GPP is at 2GHz. For fair comparison, we have done fixed-voltage scaling of the results obtained from our work to 32nm by using the appropriate scaling factor- which is ($1/S^2$) for area and ($1/S$) for energy [40], here S= L/32nm, where L=45nm. As is clear, the different levels of parallelism significantly improves the data encryption performance by having less energy consumption in comparison to GPP-, ASIC-, CMOL- and DW-based AES implementations. This significant improvement mainly comes from our proposed in-memory processing energy-efficient operations. It is noteworthy that the optimal trade-off among energy, area and speed can be achieved according to the optimization target. For instance, as shown in Table X, $P1$ exhibits 75.7%, 33.3%, 30.4%, and 40.7% improvements over energy metric in comparison to CMOS-ASIC, Baseline DW, Pipelined DW, and Multi-issue DW implementations, respectively, however it consumes more clock cycles.

Table X
THE PERFORMANCE COMPARISON OF 128-BIT AES IMPLEMENTATIONS

Platforms	Energy (nJ)	Cycles	Area (μm^2)
GPP [39]	460	2309	2.5e+6
ASIC [41]	6.7	336	4400
CMOL [42]	10.3	470	320
Baseline DW [7]	2.4	1022	78
Pipelined DW [7]	2.3	2652	83
Multi-issue DW [7]	2.7	1320	155
Proposed (P1)	1.6	4176	127
Proposed (P2)	1.74	2168	272
Proposed (P4)	1.92	1084	508

Fig. 16 depicts the energy-delay product (EDP) of different AES implementations. Based on this plot, $P4$ implementation

0278-0070 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2017.2774291, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems

shows 15.1% and 6.1% improvements over Baseline DW [7] (shown by B-DW) and ASIC implementations, respectively. It should be noted that the area overhead of $P4$ is still more than Baseline DW implementation [7] and less than ASIC one.

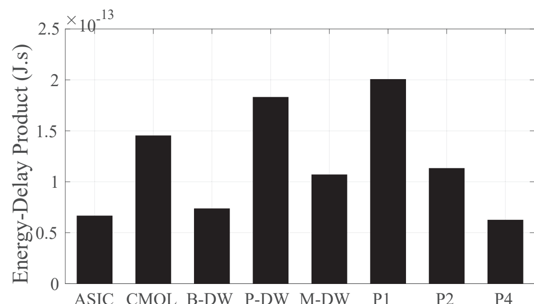


Figure 16. Energy-Delay Product (EDP) comparison of different AES implementations.

VI. Conclusion

In this paper, we proposed an ultra-energy efficient in-memory processing platform utilizing two new spin Hall effect-driven domain wall motion devices to realize both non-volatile memory cell and in-memory logic designs. The device to application level simulation results showed that, with 28% area increase, the proposed platform achieves the write energy $\sim$ 15.6 fJ/bit which is more than one order lower than that of standard STT-MRAM counterpart while keeping the identical 1ns writing speed. In addition, its logic scheme improves the operating energy by 61.3% as compared with the conventional nonvolatile in-memory logic designs. In order to show the efficiency of the proposed platform at application level AES cryptography algorithm was taken into consideration where simulation results exhibited that at a certain degree of parallelism, it can show up to 75.7% and 30.4% lower energy consumption compared to CMOS-ASIC and recent pipelined domain wall (DW) AES implementations, respectively.

References

[1] Y. Wang, H. Yu, L. Ni, G.-B. Huang, M. Yan, C. Weng, W. Yang, and J. Zhao, "An energy-efficient nonvolatile in-memory computing architecture for extreme learning machine by domain-wall nanowire devices," *IEEE Transactions on Nanotechnology*, vol. 14, no. 6, pp. 998–1012, 2015.

[2] Y. Zhang, C. Zhang, J. Nan, Z. Zhang, X. Zhang, J.-O. Klein, D. Ravelosona, G. Sun, and W. Zhao, "Perspectives of racetrack memory for large-capacity on-chip memory: From device to system," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 63, no. 5, pp. 629–638, 2016.

[3] K. Huang and R. Zhao, "Magnetic domain-wall racetrack memory-based nonvolatile logic for low-power computing and fast run-time-reconfiguration," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, no. 9, pp. 2861–2872, 2016.

[4] P. Chi, S. Li, Z. Qi, P. Gu, C. Xu, T. Zhang, J. Zhao, Y. Liu, Y. Wang, and Y. Xie, "Prime: A novel processing-in-memory architecture for neural network computation in reram-based main memory," in *Proceedings of ISCA*, vol. 43, 2016.

[5] S. Angizi, Z. He, F. Parveen, and D. Fan, "Rimpa: A new reconfigurable dual-mode in-memory processing architecture with spin hall effect-driven domain wall motion device," in *VLSI (ISVLSI), 2017 IEEE Computer Society Annual Symposium on*. IEEE, 2017, pp. 45–50.

[6] Z. He and D. Fan, "Energy efficient reconfigurable threshold logic circuit with spintronic devices," *IEEE Transactions on Emerging Topics in Computing*, 2016.

[7] Y. Wang, L. Ni, C.-H. Chang, and H. Yu, "Dw-aes: A domain-wall nanowire-based aes for high throughput and energy-efficient data encryption in non-volatile memory," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 11, pp. 2426–2440, 2016.

[8] H.-S. P. Wong *et al.*, "Phase change memory," *Proceedings of the IEEE*, vol. 98, no. 12, pp. 2201–2227, 2010.

[9] Y. Kim, S. H. Choday, and K. Roy, "Dsh-mram: differential spin hall mram for on-chip memories," *IEEE Electron Device Letters*, vol. 34, no. 10, pp. 1259–1261, 2013.

[10] Y. Kim, S. K. Gupta, S. P. Park, G. Panagopoulos, and K. Roy, "Write-optimized reliable design of stt mram," in *Proceedings of the 2012 ACM/IEEE international symposium on Low power electronics and design*. ACM, 2012, pp. 3–8.

[11] D. Fan, "Low power in-memory computing platform with four terminal magnetic domain wall motion devices," in *Nanoscale Architectures (NANOARCH), 2016 IEEE/ACM International Symposium on*. IEEE, 2016, pp. 153–158.

[12] S. Jain, A. Ranjan, K. Roy, and A. Raghunathan, "Computing in memory with spin-transfer torque magnetic ram," *arXiv preprint arXiv:1703.02118*, 2017.

[13] H.-P. Trinh, W. Zhao, J.-O. Klein, Y. Zhang, D. Ravelsona, and C. Chappert, "Magnetic adder based on racetrack memory," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 6, pp. 1469–1477, 2013.

[14] Y. Zhang, B. Yan, W. Wu, H. Li, and Y. Chen, "Giant spin hall effect (gshe) logic design for low power application," in *Proceedings of the 2015 Design, Automation & Test in Europe Conference & Exhibition*. EDA Consortium, 2015, pp. 1000–1005.

[15] X. Fong, Y. Kim, K. Yogendra, D. Fan, A. Sengupta, A. Raghunathan, and K. Roy, "Spin-transfer torque devices for logic and memory: Prospects and perspectives," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 35, no. 1, pp. 1–22, 2016.

[16] S. Angizi, Z. He, R. F. DeMara, and D. Fan, "Composite spintronic accuracy-configurable adder for low power digital signal processing," in *Quality Electronic Design (ISQED), 2017 18th International Symposium on*. IEEE, 2017, pp. 391–396.

[17] J. Kim, A. Paul, P. A. Crowell, S. J. Koester, S. S. Sapatnekar, J.-P. Wang, and C. H. Kim, "Spin-based computing: device concepts, current status, and a case study on a high-performance microprocessor," *Proceedings of the IEEE*, vol. 103, no. 1, pp. 106–130, 2015.

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

[19] 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 transactions on biomedical circuits and systems*, vol. 10, no. 6, pp. 1152–1160, 2016.

[20] S. Angizi, Z. He, and D. Fan, "Energy efficient in-memory computing platform based on 4-terminal spin hall effect-driven domain wall motion devices," in *Proceedings of the on Great Lakes Symposium on VLSI 2017*. ACM, 2017, pp. 77–82.

[21] J. Torrejon, J. Kim, J. Sinha, S. Mitani, M. Hayashi, M. Yamanouchi, and H. Ohno, "Interface control of the magnetic chirality in cofeb/mgo heterostructures with heavy-metal underlayers," *Nature communications*, vol. 5, 2014.

[22] C.-F. Pai, L. Liu, Y. Li, H. Tseng, D. Ralph, and R. Buhrman, "Spin transfer torque devices utilizing the giant spin hall effect of tungsten," *Applied Physics Letters*, vol. 101, no. 12, p. 122404, 2012.

[23] X. Fong, S. K. Gupta, N. N. Mojumder, S. H. 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 *2011 International Conference on Simulation of Semiconductor Processes and Devices*. IEEE, 2011, pp. 51–54.

[24] http://math.nist.gov/oommf/.

[25] D. Bromberg, M. Moneck, V. Sokalski, J. Zhu, L. Pileggi, and J.-G. Zhu, "Experimental demonstration of four-terminal magnetic logic device with separate read-and write-paths," in *Electron Devices Meeting (IEDM), 2014 IEEE International*. IEEE, 2014, pp. 33–1.

[26] S. J. Noh, Y. Miyamoto, M. Okuda, N. Hayashi, and Y. K. Kim, "Effects of notch shape on the magnetic domain wall motion in nanowires with in-plane or perpendicular magnetic anisotropy," *Journal of Applied Physics*, vol. 111, no. 7, p. 07D123, 2012.

[27] S. Li, C. Xu, Q. Zou, J. Zhao, Y. Lu, and Y. Xie, "Pinatubo: A processing-in-memory architecture for bulk bitwise operations in emerging non-volatile memories," in *Design Automation Conference (DAC), 2016 53nd ACM/EDAC/IEEE*. IEEE, 2016, pp. 1–6.

0278-0070 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

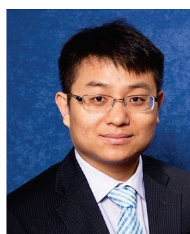
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2017.2774291, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems

[28] S. Fukami *et al.*, "20-nm magnetic domain wall motion memory with ultralow-power operation," in *2013 IEEE International Electron Devices Meeting*, 2013.
[29] (2011) Ncsu eda freepdk45. [Online]. Available: http://www.eda.ncsu.edu/wiki/FreePDK45:Contents
[30] X. Dong, C. Xu, N. Jouppi, and Y. Xie, "Nvsim: A circuit-level performance, energy, and area model for emerging non-volatile memory," in *Emerging Memory Technologies*. Springer, 2014, pp. 15–50.
[31] S. Li *et al.*, "Mcpat: an integrated power, area, and timing modeling framework for multicore and manycore architectures," in *MICRO*. ACM, 2009, pp. 469–480.
[32] N. Binkert *et al.*, "The gem5 simulator," *SIGARCH*, vol. 39, pp. 1–7, 2011.
[33] K. Huang, R. Zhao, and Y. Lian, "Stt-mram based low power synchronous non-volatile logic with timing demultiplexing," in *Proceedings of the 2014 IEEE/ACM International Symposium on Nanoscale Architectures*. ACM, 2014, pp. 31–36.
[34] R. Zhang, P. Gupta, L. Zhong, and N. Jha, "Synthesis and optimization of threshold logic networks with application to nanotechnologies," in *Design, Automation and Test in Europe Conference and Exhibition, 2004. Proceedings*, vol. 2. IEEE, 2004, pp. 904–909.
[35] H. Noguchi, K. Ikegami, K. Abe, S. Fujita, Y. Shiota, T. Nozaki, S. Yuasa, and Y. Suzuki, "Novel voltage controlled mram (vcm) with fast read/write circuits for ultra large last level cache," in *Electron Devices Meeting (IEDM), 2016 IEEE International*. IEEE, 2016, pp. 27–5.
[36] E. Martinez, "The stochastic nature of the domain wall motion along high perpendicular anisotropy strips with surface roughness," *Journal of Physics: Condensed Matter*, vol. 24, no. 2, p. 024206, 2011.
[37] "Op amp noise relationships: 1/f noise, rms noise, and equivalent noise bandwidth, analog devices. [online]."
[38] N.-F. Standard, "Announcing the advanced encryption standard (aes)," *FIPSP*, vol. 197, 2001.
[39] K. Malbrain, "Byte-oriented-aes: a public domain byte-oriented implementation of aes in c," 2009.
[40] Z. Abbas and M. Olivieri, "Impact of technology scaling on leakage power in nano-scale bulk cmos digital standard cells," *Microelectronics Journal*, vol. 45, no. 2, pp. 179–195, 2014.
[41] S. Mathew *et al.*, "340 mv–1.1 v, 289 gbps/w, 2090-gate nanoaes hardware accelerator with area-optimized encrypt/decrypt gf (2 4) 2 polynomials in 22 nm tri-gate cmos," *IEEE J. Solid-State Circuits*, vol. 50, no. 4, pp. 1048–1058, 2015.
[42] Z. Abid, A. Alma'Aitah, M. Barua, and W. Wang, "Efficient cmol gate designs for cryptography applications," *IEEE transactions on nanotechnology*, vol. 8, no. 3, pp. 315–321, 2009.

Shaahin Angizi (S'15) received his B.Sc. in Computer Engineering, Hardware from South Tehran Branch of IAU, Tehran, Iran in 2012 and his M.Sc. in Computer Engineering, Computer Systems Architecture from Science and Research Branch of IAU, Tabriz, Iran in 2014. He is currently working toward the Ph.D. degree in Computer Engineering at University of Central Florida, Orlando, USA. His research interests include in-memory computing, deep learning, low power VLSI designs, Spin-based computing and QCA.

Zhezhi He (S'16) received his B.S. degree in Information Science and Engineering from Southeast University, Nanjing, China, in 2012, and M.E. degree in Electrical and Computer Engineering from Oregon State University, Corvallis, OR, USA, in 2015. Currently he is pursuing Ph.D degree in Electrical Engineering at University of Central Florida. His research interests include neuromorphic computing, analog and mixed signal circuit design, and emerging technology.

Nader Bagherzadeh (F'14) received the PhD degree from the University of Texas at Austin in 1987. He is a professor of computer engineering at the Department of Electrical Engineering and Computer Science, University of California, Irvine, where he served as a chair from 1998 to 2003. He has been involved in research and development in the areas of computer architecture, reconfigurable computing, VLSI chip design, network-on-chip, 3D chips, sensor networks, computer graphics, memory, and embedded systems. He is a fellow member of the IEEE. He has published more than 250 articles in peer-reviewed journals and conferences. His former students have assumed key positions in software and computer systems design companies in the past 25 years. He has been a PI or Co-PI on more than $10 million worth of research grants for developing next generation computer systems for applications in general purpose computing and digital signal processing as well as other related areas.



Deliang Fan (M'15) received his B.S. degree in Electronic Information Engineering from Zhejiang University, China, in 2010. He received M.S. and Ph.D. degree in Electrical and Computer Engineering from Purdue University, West Lafayette, IN, USA, in 2012 and 2015, respectively. He joined the Department of Electrical and Computer Engineering at University of Central Florida, Orlando, FL, as an Assistant Professor in 2015. His primary research interest lies in Ultra-low Power Brain-inspired (Neuromorphic), Non-Boolean and Boolean Computing Using Emerging Nanoscale Devices like Spin-Transfer Torque Devices and Memristors. His other research interests include nanoscale physics based spintronic device modeling and simulation, low power digital and mixed-signal CMOS circuit design.

0278-0070 (c) 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.