Majority-Based Spin-CMOS Primitives for Approximate Computing

Shaahin Angizi, *Student Member, IEEE*, Honglan Jiang, *Student Member, IEEE*, Ronald F. DeMara, *Senior Member, IEEE*, Jie Han, *Senior Member, IEEE*, and Deliang Fan, *Member, IEEE*

***Abstract*—Promising for digital signal processing applications, approximate computing has been extensively considered to trade-off limited accuracy for improvements in other circuit metrics such as area, power, and performance. In this paper, approximate arithmetic circuits are proposed by using emerging nanoscale spintronic devices. Leveraging the intrinsic current-mode thresholding operation of spintronic devices, we initially present a hybrid spin-CMOS majority gate design based on a composite spintronic device structure consisting of a magnetic domain wall motion stripe and a magnetic tunnel junction. We further propose a compact and energy-efficient accuracy-configurable adder design based on the majority gate. Unlike most previous approximate circuit designs that hard-wire a constant degree of approximation, this design is adaptive to the inherent resilience in various applications to different degrees of accuracy. Subsequently, we propose two new approximate compressors for utilization in fast multiplier designs. The device-circuit SPICE simulation shows 34.58% and 66% improvement in power consumption, respectively, for the accurate and approximate modes of the accuracy-configurable adder, compared to the recently reported domain wall motion-based full adder design. In addition, the proposed accuracy-configurable adder and approximate compressors can be efficiently utilized in the discrete cosine transform (DCT) as a widely-used digital image processing algorithm. The results indicate that the DCT and inverse DCT (IDCT) using the approximate multiplier achieve ∼2x energy saving and 3x speed-up compared to an exactly-designed circuit, while achieving comparable quality in its output result.**

***Index Terms*—Approximate computing, accuracy-configurable adder, compressor, spintronic, domain wall motion device.**

I. Introduction

COMMONLY-USED multimedia applications rely on Digital Signal Processing (DSP) blocks as primary components. In such applications, low power design is an imperative requirement. Recently, approximate computing has been widely considered in algorithmic circuit design to overcome the power issue by exploiting the non-brittle perceptual abilities of human beings [1]–[3]. This means that approximate outputs can be interpreted by human senses despite being inexact. This approach may be effective in reducing circuit complexity while simultaneously addressing the problem of high energy consumption [2], [4], [5].

Various methods have been proposed for designing approximate circuits which can be categorized into two broad methodologies. The first methodology is based on voltage over scaling (VOS) such as algorithmic noise tolerance (ANT) [6] and significance driven computation (SDC) [7] for modifying or limiting the resultant errors. The second methodology approximates fundamental logic functions at the circuit-level such as a variety of approximate adder realizations [1], [8], [9].

As a basic building block in most DSP systems, the multiplier is typically located on the critical path of such systems, so it contributes significantly to the system's total power consumption and propagation delay, which greatly motivates the need for fast multiplier designs. A fast multiplication operation is usually performed in three steps, including partial product (PP) generation, PP reduction using a carry-save adder (CSA) tree and a fast carry propagation adder (CPA) for the final computation of the product [10]. Most specifically, the PP reduction circuit is crucial in determining the design complexity, latency and power consumption of a multiplier. Hence, improving the performance and energy efficiency of the PP reduction circuit using appropriate arithmetic blocks, such as compressors, can directly improve the performance and energy efficiency of a fast multiplier [5], [11]. Basically, using compressors can reduce energy dissipation by decreasing the number of PP stages in a multiplier. Optimized designs of accurate 4-2 compressors have been proposed in [10], [12]. In addition, several approximate compressors have recently been presented in the literature [13], [14].

These approximate compressors have typically been realized using Complementary Metal-Oxide-Semiconductor (CMOS) AND-OR gates that increase the design complexity and XOR gates that increase the overall switching activity. On the other hand, as we approach the physical limit of CMOS devices, an urgent need arises for a potential alternative or complementary computing technology. Among others, spintronic devices [15] have shown significant promise over the past decade because of their non-volatility, zero leakage current, high integration density, low standby power, and Back End of Line fabrication with the CMOS technology [16]. In this context, different accurate and approximate circuit designs have been presented

Manuscript received March 29, 2018; accepted May 6, 2018. Date of publication May 15, 2018; date of current version July 9, 2018. This work was partly supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada (Project Number: RES0025211) and the National Science Foundation under Grant No. 1740126 and Semiconductor Research Corporation nCORE. The review of this paper was arranged by Associate Editor W. Zhao. *(Corresponding author: Jie Han.)*

S. Angizi, R. F. DeMara, and D. Fan are with the Department of Electrical and Computer Engineering, University of Central Florida, Orlando, FL 32816 USA (e-mail: angizi@knights.ucf.edu; ronald.demara@ucf.edu; dfan@ucf.edu).

H. Jiang and J. Han are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 1H9, Canada (e-mail: honglan@ualberta.ca; jhan8@ualberta.ca).

Digital Object Identifier 10.1109/TNANO.2018.2836918

1536-125X © 2018 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.

[17]–[20]. Additionally, leveraging majority logic in nanoscale technologies can bring even higher performance and energy efficiency compared to conventional implementations of arithmetic circuits [21]–[24].

Nevertheless, a limitation of the aforementioned designs is the hardwired degree of approximation within the circuit. Therefore, the circuit can only be adjusted to meet a single quality constraint, limiting the possibility of achieving a different quality level [7], [25]. This drawback limits the circuit's practicality, since a programmable platform could facilitate execution of a range of applications with various approximations. Thus, the degree of approximation remains fluid for different applications. Jain *et al.* in [25] have proposed effective approaches to the design of quality configurable circuits through logic isolation. In another recent work, four dual-quality 4-2 compressors are presented for use in dynamic accuracy-configurable multipliers [14]. Cai *et al.* in [26] utilizes MTJ switching behavior as an innovative mechanism to switch between accurate and approximate modes.

Some preliminary results of this work have been published in [27]. In [27], a current mode spin-CMOS majority gate based on spintronic threshold device is designed. In addition, an efficient spin-CMOS accuracy-configurable adder is presented utilizing majority gates operating in two distinct modes (approximation and precision). In this paper, new designs of approximate 4-2 compressors are proposed for efficient implementations in DSP systems. As a significant extension of [27], this manuscript makes the following novel contributions:

- two distinct designs for 4-2 approximate compressors are developed based on presented scalable current mode spin-CMOS majority gate using spintronic threshold device. These designs are further leveraged for implementing fast multiplier design as a basic block in DSP hardware,
- a comprehensive evaluation framework is constructed for the proposed designs from device to application level, and
- both the accuracy-configurable adder and approximate compressors are utilized in image compression, and the resultant output quality and energy trade-offs are assessed with respect to peak signal-to-noise ratio, delay, and energy consumption.

The remainder of the paper is organized as follows. Section II introduces the spintronic threshold device structure and its modeling. Section III addresses the design and evaluation of spin-CMOS majority gate circuit. In Section IV, the majority gate-based accuracy-configurable adder is designed. Section V is dedicated to proposal of highly-efficient and low-cost approximate 4-2 compressors. Section VI discusses circuit level performance evaluation of the proposed designs. Section VII assesses the efficacy of the presented circuits in image processing applications and Section VIII concludes the paper.

II. Spintronic Threshold Device Structure

In this section, we present Spintronic Threshold Device (Spin-TD) based on a composite device structure consisting of a Domain Wall Motion (DWM) magnetic stripe and Magnetic Tunnel Junction (MTJ). The device structure for the Spin-TD is shown in Fig. 1(a) [15], [28]–[31]. It consists of a thin and short (2 nm $\times$ 20 nm $\times$ 50 nm) magnetic Domain Wall Stripe (DWS) connecting two fixed anti-parallel magnetic domains. When the electrons are injected into the lateral terminals (T1 or T2), they become spin-polarized and exert a Spin-Transfer Torque (STT) on the Domain Wall (DW) (i.e., the transition area between two domains). This spin-polarized current can move DW within DWS. A fixed small magnet and DWS beneath it form a MTJ to read the state of DWS. It is noteworthy that an MTJ [32] consists of two ferromagnetic layers (a free layer and a fixed one as shown in Fig. 1(a) with a tunneling oxide (commonly MgO) barrier sandwiched between them [15].

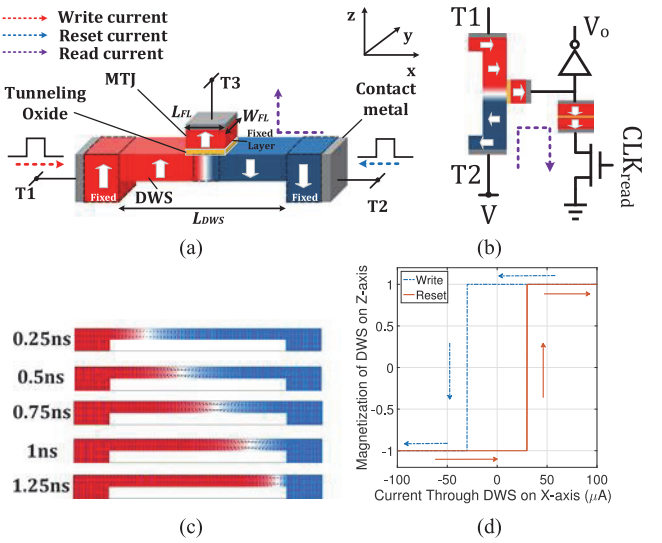


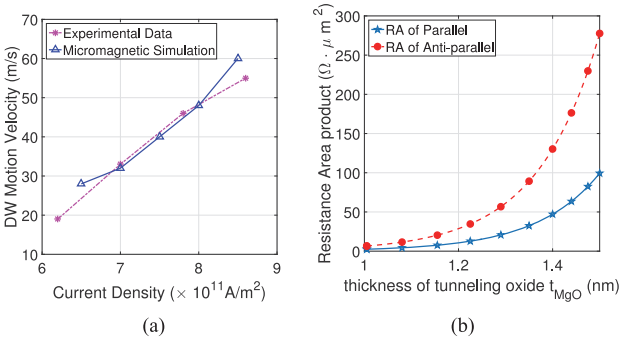
Fig. 1. (a) Spintronic threshold device (Spin-TD) structure. (b) Spin-TD sense circuit. (c) Micro-magnetic simulation for the DW position, (d) Spin-TD transfer function and reset.

The fixed layer of sense MTJ in Spin-TD is very small (20 nm $\times$ 20 nm). The magnetization of DWS can be identified anti-parallel (AP) or parallel (P) to the fixed layer by injecting a current (larger than critical current) along it from its terminals (T1 to T2) or vice-versa [33]. Hence, the Spin-TD can detect the polarity of current flow at its input node, acting as an ultra-low voltage and compact current comparator. The resistance states are binary, i.e. either high (corresponding to AP configuration) or low (corresponding to P configuration) and can be read employing the Spin-TD sense circuit as shown in Fig. 1(b). The threshold of Spin-TD, i.e. the minimum current magnitude required to switch the DWS magnetization (move DW from one end to the other end), is determined by the critical current density and DW velocity.

The transient micro-magnetic simulation of DW position (achieved from OOMMF [34]) is illustrated in Fig. 1(c), using device dimension shown in Table I, from 0.25 ns to 1.25 ns. Since the magnetization of DWS beneath the MTJ is fully switched at 1ns, the Spin-TD intrinsic threshold (I_{th}) of this device can be considered 30 μA within 1 ns corresponding to DW velocity of $\sim$50 m/s. Fig. 1(d) describes DWS magnetization switch corresponding to the applied current pulse

TABLE I
DEVICE PARAMETERS USED IN SIMULATION

Symbol	Quantity	Values
α	Damping coefficient	0.02
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	$1.1 \times 10^{-11} J/m$
P	Polarization	0.6
t_{MgO}	MgO thickness of MTJ	1.5 nm
$(L.W.t)_{DWS}$	DWS dimension	$50 \times 20 \times 2nm^3$



(a) (b)

Fig. 2. (a) Simulated DW motion velocity vs. lateral current density, showing a good match with experimental data reported in [35]. (b) Resistance-area product vs. the thickness of tunneling oxide in AP and P state (with 50 mv constant voltage).

(1 ns). A hysteresis effect can be observed due to DWM critical current density. The device parameters used in the simulation are listed in Table I. We benchmarked the micro-magnetic simulation with the experimental data in [35] (the same nano-stripe width of 20 nm is fabricated) and it shows a good match as shown in Fig. 2(a). The MTJ is modeled using NEGF-LLG solution (non-equilibrium Green's function and Landau-Lifshitz-Gilbert equations) for spin to charge interface and calibrated with experimental data in [35], [36]. Resistance-area (RA) product vs. the thickness of tunneling oxide in AP and P states in this work considering a constant voltage of 50 mv is plotted in Fig. 2(b). Basically, the resistance-area (RA) product of the MTJ, which corresponds to the thickness of the MTJ tunneling oxide and the reliability of the MTJ, needs to meet the design specifications. Otherwise an accident write of MTJ may occur when the current flowing through the MTJ, is more than threshold current, I_{th}, during read operation. It may occur when a thinner t_{MgO} is used, which further leads to logic failure. Our simulations showed that 1.5 nm thickness provides the circuit with a favorable reliability during sensing.

The effective resistance of the MTJ formed between DWS and fixed layer (T3 side) is smaller when they have the identical magnetization and vice versa. The ratio of two resistances is defined in terms of Tunneling Magneto Resistance ratio (TMR). As shown in Fig. 1(b), Spin-TD forms a voltage divider with a fixed reference MTJ to sense the resistance state. Static current in the voltage divider can be minimized by increasing the MTJ oxide thickness. For a 1 ns clock cycle, the oxide thickness in this work is chosen to be 1.5 nm that results in a total power dissipation of ~ 1 μW for the sensing circuit (including the clocking power). It is worth noting that in the sense circuit,

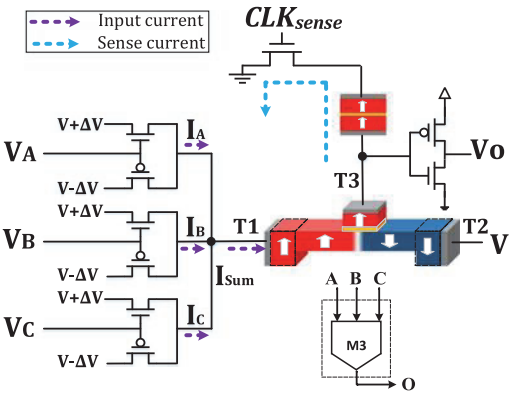


Fig. 3. Spin-CMOS implementation of three-input majority gate.

the transient current with short duration (1 ns) and low magnitude (~ 2 μA) flows from T2 to T3, which will not disturb the state of DWS (domain wall position). The sense current can be further reduced to less than 1 μA by increasing the oxide thickness [33].

III. SPIN-CMOS MAJORITY GATE CIRCUIT DESIGN

In this section, we present a highly-scalable spin-CMOS majority gate circuit design based on Spin-TD. The output of an *n*-input Majority Gate (MG) (*n is odd*) is determined by the majority of its inputs. For instance, the output is asserted to be logic value "1" only when more than $(\frac{n-1}{2})$ of the inputs are "1".

The proposed three-input MG circuit employing Spin-TD is shown in Fig. 3. As shown, the input terminal (T1) is connected to a network consisting of 3 pairs of NMOS-PMOS input transistors, in which all of the input transistors work as Deep Triode region Current Sources (DTCS) by applying $V + \Delta V =$ 550 mv and $V - \Delta V = 450$ mv to the source and drain, respectively. The proposed circuit is controlled by two clock signals (CLK_{compute} and CLK_{sense}) and each clock period is set to be 1 ns to synchronize with next stage circuits (discussed thoroughly in Section VI). Note that, T2 of Spin-TD is connected to a constant voltage of $V = 500$ mv and the voltage difference is $\Delta V = 50$ mv, leading to an ultra-small voltage drop and correspondingly-low power consumption.

During the computation clock interval, the binary input voltages (VDD, GND) are applied at the gate of the input transistors, leading to input current flowing into (positive) or out of (negative) the connected Spin-TD. According to the principle of conservation of electric charge, the direction and magnitude of total current at intersection node depend on the algebraic sum of the input currents (I_A, I_B and I_C herein). This summation current (I_{Sum}) determines the position of DW in the DWS as described in Section II. By properly sizing the input transistors, the current flowing to T1 from each input branch is either $+30$ μA or -30 μA corresponding to input gate voltages as high ("1") or low ("0"), respectively. For instance, the input combination of (A, B, C) = (0, 1, 1) leads to (I_A, I_B, I_C) = (-30 μA, $+30$ μA, $+30$ μA) and the total current flowing into T1 is $+30$ μA. Such current is equal to the threshold current of the Spin-TD and

TABLE II
THREE-INPUT SPIN-CMOS MG CURRENT SUMMATION AT T1 AND CORRESPONDING DOMAIN WALL POSITION

Input Currents (μA)			Summation Current (μA)	Initial DW position		Final DW position
I_A	I_B	I_C	I_{Sum}	@Right	@Left	
-30	-30	-30	-90	Right	Right	
-30	-30	$+30$	-30	Right	Right	
-30	$+30$	-30	-30	Right	Right	
-30	$+30$	$+30$	+30	Left	Left	
$+30$	-30	-30	-30	Right	Right	
$+30$	-30	$+30$	+30	Left	Left	
$+30$	$+30$	-30	+30	Left	Left	
$+30$	$+30$	+30	+90	Left	Left	

TABLE III
TRUTH TABLE FOR ACCURATE AND APPROXIMATE FAS

Inputs			Acc. Outputs		App. Outputs			
A	B	C_{in}	C_{out}	Sum	C_{out}		Sum	
0	0	0	0	0	0	✓	1	✘
0	0	1	0	1	0	✓	1	✓
0	1	0	0	1	0	✓	1	✓
0	1	1	1	0	1	✓	0	✓
1	0	0	0	1	0	✓	1	✓
1	0	1	1	0	1	✓	0	✓
1	1	0	1	0	1	✓	0	✓
1	1	1	1	1	1	✓	0	✘

relocates the domain wall towards the T1 side, further resulting in the sense MTJ in an anti-parallel high resistance state. During the sense phase, when the CLK_{sense} is high, a voltage divider between Spin-TD's MTJ and a fixed reference MTJ is formed to sense the resistance state of spin-CMOS 3-input MG to produce reliable output voltage right after the inverter. In this case, the sensing circuit will generate a high output representing logic "1".

Table II lists eight possible input current combinations and the corresponding summation current. The last two columns of Table II list the DW position before and arrival of the computation clock. It is clear that the proposed 3-input spin-CMOS MG does not require an additional reset clock, since the final DW position is solely determined by the summation current direction and the initial DW position does not have an effect on the final DW position. As an instance, when the I_{Sum} is equal to or greater than $+30$ μA, either the DW's initial position is at the right or left side, it will either be pushed towards or remain on the left side. It is worth pointing out that 2-input AND or OR gates can be efficiently designed just by setting one of the three MG inputs to GND or VDD, respectively. In addition, the proposed MG circuit readily allows for the scaling of input fan-in. It means that the 3-input MG circuit design can be effectively extended for implementing *n*-bit MGs. To do so, the connected input branches are increased. For instance, a 5-input MG will be obtained by employing five pairs of NMOS-PMOS input transistors without changes in circuit parameters. Note that, in order to produce a highly reliable complementary output voltage, we can also add an additional cascaded inverter to the sensing circuit right after Vo in Fig. 3. In the following two sections, the proposed spin-CMOS MG is used to implement an accuracy-configurable adder and two approximate compressors.

IV. SPIN-CMOS ACCURACY-CONFIGURABLE ADDER

A. Functionality Analysis

A full adder (FA) is one of the most frequently-used components in arithmetic circuitry. In addition to its regular use for addition, it is employed in other arithmetic operations such as subtraction, multiplication, and division [37]. For instance, multiplication has been implemented using successive additions. Moreover, FA is the key component and optimization target of many DSP algorithms. Hence, in order to obtain a high performance DSP system, we need to design energy efficient and low complexity adders [5]. While extensive work has been done in designing approximate adders [38], [39], the research efforts on accuracy-configurable approximate adders are limited. Let A, B and C_{in} be inputs of an accurate full adder, the principle Boolean expression of Carry out (C_{out}) and accurate Sum (Sum_{acc}) of FA cell are as follows:

$$C_{\text{out}} = AB + AC_{\text{in}} + BC_{\text{in}} = M3(A, B, C_{\text{in}}) \quad (1)$$

$$Sum_{\text{acc}} = ABC_{\text{in}} + \bar{A}\bar{B}C_{\text{in}} + \bar{A}B\bar{C_{\text{in}}} + A\bar{B}\bar{C_{\text{in}}} \quad (2)$$

Some Boolean expressions for Sum_{acc} and C_{out} of FA based on inverters and MGs have been reported in [27], [40], [41]. As can be seen in (1), C_{out} can be readily derived with a 3-input MG. Alternatively, Sum_{acc} can be obtained by using 3- and 5-input MG functions as (3).

$$\begin{aligned} Sum_{\text{acc}} &= ABC_{\text{in}} + (\overline{AB}.\overline{AC_{\text{in}}}.\overline{BC_{\text{in}}})(A + B + C_{\text{in}}) \\ &= ABC_{\text{in}} + \overline{M3}.(A + B + C_{\text{in}}) \\ &= ABC_{\text{in}} + \overline{M3}.(A + B + C_{\text{in}}) + \overline{M3}M3 \\ &= M5(A, B, C_{\text{in}}, \overline{M3}, \overline{M3}) \\ &= M5(A, B, C_{\text{in}}, \overline{C_{\text{out}}}, \overline{C_{\text{out}}}) \end{aligned} \quad (3)$$

Table III shows the truth table of an FA. A close observation clarifies that six of eight outputs are correct if we make $Sum = \overline{C_{\text{out}}}$. Based on this observation, we propose a streamlined and cost-effective approximate FA circuit comprising one 3-input MG and one cascaded inverter. The approximate Sum output (Sum_{app}) of this adder is given by:

$$Sum_{\text{app}} = \overline{C_{\text{out}}} = \overline{M3(A, B, C_{\text{in}})} \quad (4)$$

B. Spin-CMOS Implementation

The proposed spin-CMOS implementation of the accuracy-configurable FA cell is shown in Fig. 4 consisting of two stages: Stage 1 to generate C_{out} and Sum_{app} and Stage 2 to generate Sum_{acc}. The first stage consists of a spin-CMOS MG realizing an approximate FA (App. FA) according to (1) and (4). As shown in Fig. 4, this circuit is designed with an appropriate fan-out for producing Sum_{app} output after one add-on inverter, while C_{out} is already achieved according to the Boolean expression in (1).

Meanwhile, the $\overline{C_{\text{out}}}$ (or Sum_{app}) produced in Stage 1 is then connected to a similarly scaled input transistor network

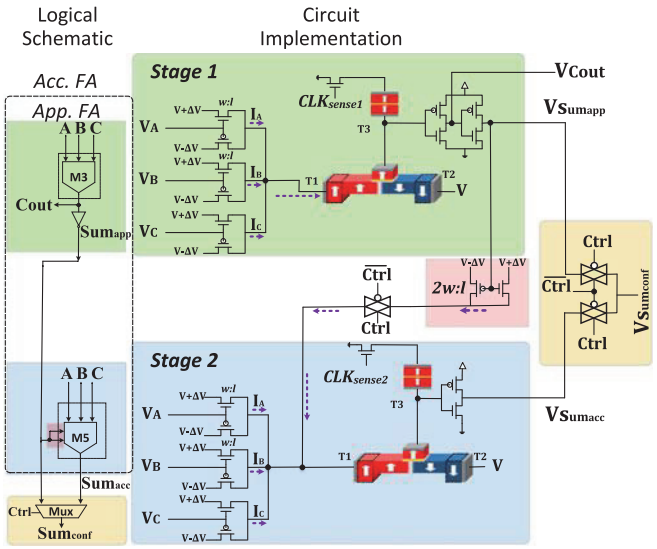


Fig. 4. Logical schematic and circuit implementation of Spin-CMOS accuracy-configurable FA. When Ctrl knob is high, the circuit functions as an accurate FA and when Ctrl knob is low, the circuit functions as an approximate adder.

but with a $\frac{2w}{l}$ ratio to provide a double weighted current as expressed in (3). The double weighted current in conjunction with the sum of three primary inputs flow towards the T1 of the Stage 2's MG (realizing a 5-input MG as depicted in the logical schematic in Fig. 4). Consequently, the output voltage of this stage is Sum_{acc} realizing an accurate FA (Acc. FA). To provide the circuit with a proper and streamlined configurability, the wire connection between these two stages is regulated using a CMOS transmission gate (TG). Furthermore, the sum outputs of both stages are laterally connected to a 2:1 CMOS multiplexer implemented utilizing two TGs to produce configurable sum (Sum_{conf}). Accordingly, the proposed spin-CMOS accuracy-configurable circuit operates in two different modes i.e. precision and approximation. In the precision mode, the control knob (*Ctrl*) is high, so the intermediate TG is ON and the double weighted current is routed to the second stage MG. Consequently, the circuit functions as an accurate adder since the second input of the multiplexer will be transmitted to the output ($Sum_{\text{conf}} = Sum_{\text{acc}}$). In the approximation mode, the *Ctrl* is low and the double weighted branch is disconnected avoiding any switching activity in second stage. Therefore, the Stage 1's circuit works as a low power approximate adder when $Sum_{\text{conf}} = Sum_{\text{app}}$. Timing diagram and analysis are shown later in Fig. 9.

V. Spin-CMOS Approximate Compressors

A fast multiplier typically consists of three primary modules: (1) a Partial product generator, (2) a Carry save adder (CSA) tree for reducing the partial products, and (3) a Carry propagation adder (CPA) for final computation. The second module dominates the circuit complexity, delay, and power consumption of a multiplier. The main idea behind utilizing multi-operand CSA is to reduce n numbers to two numbers; that is why $n-2$ compressor blocks have been widely explored in computer arithmetic

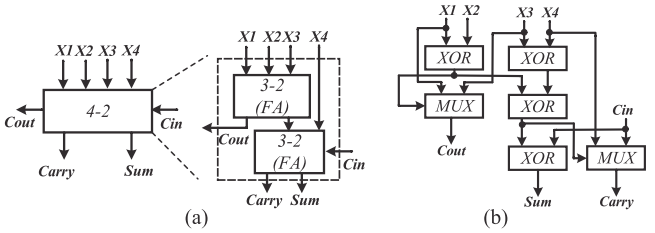


Fig. 5. (a) 4-2 compressor using two FAs. (b) Optimized 4-2 compressor [10].

TABLE IV
Truth Table for Accurate and Approximate Compressors

Inputs					Acc Output			Design I			Design II		
Cin	X_4	X_3	X_2	X_1	Cout	Carry	Sum	Cout	Carry'	Sum'	Cout	Carry	Sum'
0	0	0	0	0	0	0	0	0	0	1✘	0	0	1✘
0	0	0	0	1	0	0	1	0	0	1	0	0	1
0	0	0	1	0	0	0	1	0	0	1	0	0	1
0	0	0	1	1	1	0	0	1	0	1✘	1	0	1✘
0	0	1	0	0	0	0	1	0	0	1	0	0	1
0	0	1	0	1	1	0	0	1	0	1✘	1	0	1✘
0	0	1	1	0	1	0	0	1	0	1✘	1	0	1✘
0	0	1	1	1	1	0	1	1	0	1	1	0	1
0	1	0	0	0	0	0	1	0	1✘	0✘	0	0	1
0	1	0	0	1	0	1	0	0	1	0	0	1	0
0	1	0	1	0	0	1	0	0	1	0	0	1	0
0	1	0	1	1	1	0	1	1	0	1	1	0	1
0	1	1	0	0	0	1	0	0	1	0	0	1	0
0	1	1	0	1	1	0	1	1	0	1	1	0	1
0	1	1	1	0	1	0	1	1	0	1	1	0	1
0	1	1	1	1	1	1	0	1	0✘	1✘	1	1	0
1	0	0	0	0	0	0	1	0	1✘	0✘	0	0	1
1	0	0	0	1	0	1	0	0	1	0	0	1	0
1	0	0	1	0	0	1	0	0	1	0	0	1	0
1	0	0	1	1	1	0	1	1	0	1	1	0	1
1	0	1	0	0	0	1	0	0	1	0	0	1	0
1	0	1	0	1	1	0	1	1	0	1	1	0	1
1	0	1	1	0	1	0	1	1	0	1	1	0	1
1	0	1	1	1	1	1	0	1	0✘	1✘	1	1	0
1	1	0	0	0	0	1	0	0	1	0	0	1	0
1	1	0	0	1	0	1	1	0	1	0✘	0	1	0✘
1	1	0	1	0	0	1	1	0	1	0✘	0	1	0✘
1	1	0	1	1	1	1	0	1	1	0	1	1	0
1	1	1	0	0	0	1	1	0	1	0✘	0	1	0✘
1	1	1	0	1	1	1	0	1	1	0	1	1	0
1	1	1	1	0	1	1	0	1	1	0	1	1	0
1	1	1	1	1	1	1	1	1	1	0✘	1	1	0✘

[13], [37]. As shown in Fig. 5(a), a widely-used 4-2 compressor receives 4 primary inputs ($X1-X4$) and one carry bit (C_{in}) from the lower position block, then it produces 2 primary outputs ($Carry$ and Sum)) and sends one carry bit (C_{out}) to the higher position block. Fig. 5(b) depicts the design of an accurate 4-2 compressor based on the so-called CMOS XOR-XNOR gates [10].

In this section, we propose two designs for approximate 4-2 compressors based on accurate and approximate FAs proposed in Section IV-A. Intuitively, in order to design an approximate 4-2 compressor (with the truth table shown in Table IV), it is possible to replace the accurate full-adder cells by approximate cells. In other words, two cascaded approximate 3-2 compressors can be readily employed to realize an approximate 4-2 compressor (such as the first design presented in [38]). However, this solution has not been very popular so far due to the high error rate of basic modules such that it shows 53% error rate (with at least 17 incorrect results out of 32 possible outputs). Note that herein the error rate is defined as the ratio of number of erroneous outputs to the total number of outputs.

A. Design I

The gate level structure of the first proposed approximate 4-2 compressor is depicted in Fig. 6(a). As can be seen, only two approximate FAs (App. FA) are cascaded to realize such

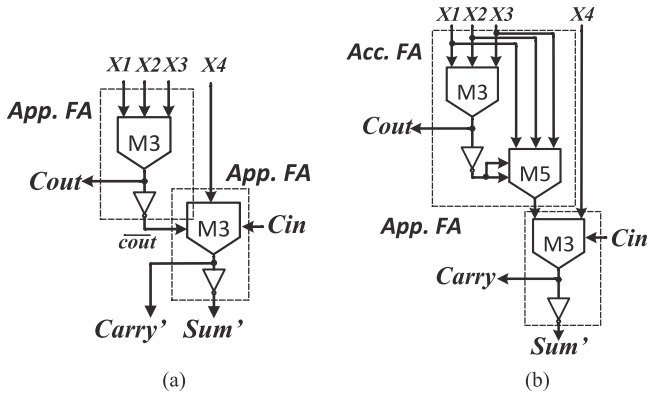


Fig. 6. The proposed approximate 4-2 compressors. (a) Design I employs two approximate FAs. (b) Design II employs one accurate and one approximate FA.

a low-complexity design. $X1$-$X3$ inputs are assigned to the first App. FA and $X4$, C_{in} along with $\overline{C_{\text{out}}}$ are connected to the second App. FA. In this way, C_{out} can be obtained accurately for all input combinations using (5). $Carry'$ is given in (6) with only 4 incorrect outputs as tabulated in Table IV. Sum' is accordingly derived in (7) by inverting the result of $Carry'$ with 12 incorrect output out of 32 possible outputs. Overall, Design I yields an error rate of 37.5% that is smaller than the error rate of employing the best approximate FA [38] and the same as that of the first design presented in [13]. Furthermore, Design I shows significant improvement for the critical delay (2Δ[1]) compared to the first approximate design in [13] (3Δ) and optimized design in [10] (3Δ).

$$C_{\text{out}} = M3(X1, X2, X3) \tag{5}$$

$$Carry' = M3(\overline{C_{\text{out}}}, X4, C_{\text{in}}) \tag{6}$$

$$Sum' = \overline{Carry'} \tag{7}$$

B. Design II

Fig. 6(b) depicts the second proposed design employing one approximate FA (App. FA) and one accurate FA (Acc. FA) cell. Applying an accurate FA cell in the first level ensures that, in addition to C_{out} (5), $Carry$ output can be achieved correctly for all input combinations as tabulated in the last few columns in Table IV. This design generates 8 erroneous outputs for Sum', therefore the error rate is now reduced to 25%. As a trade-off between accuracy and circuit delay/complexity, design II incurs (3Δ) as the critical path delay with an additional 5-input MG compared to design I.

The proposed compressors are readily implemented in hybrid spin-CMOS circuits as shown in the logical diagrams in Fig. 6 based on spin-CMOS MG shown in Fig. 3. Fig. 7 shows Design I implementation by using 2 DWSs and 4 MTJs. Design II is similarly implemented using 3 DWSs and 6 MTJs.

[1] Δ is defined as gate delay [13]

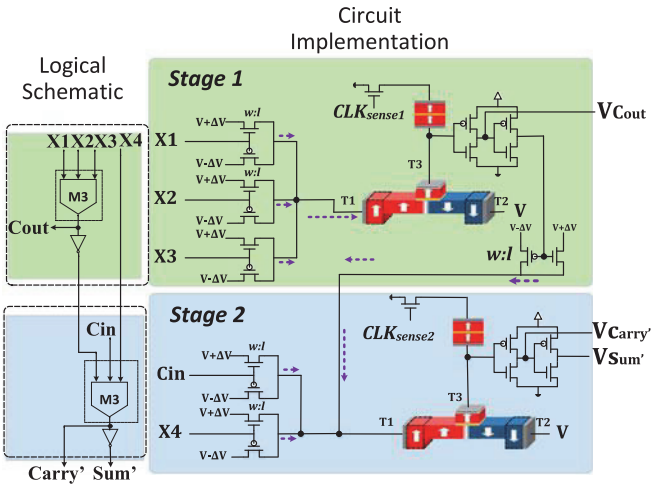


Fig. 7. Logical schematic and circuit implementation of Spin-CMOS compressor based on Design I.

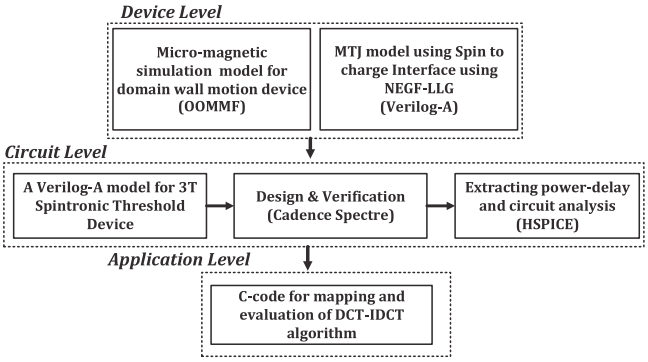


Fig. 8. Device to application level co-simulation framework.

VI. Performance Evaluation

In order to evaluate the performance of the proposed circuits, we designed a comprehensive simulation framework as shown in Fig. 8. This bottom-up simulation framework can be divided into three main levels:

1) Device level: For device level simulation, we benchmarked the domain wall motion dynamics with experimental data [35] utilizing Object Oriented MicroMagnetic Framework (OOMMF) [34]. The MTJ (composed of a DWS, a tunneling oxide layer and a fixed ferromagnetic layer) is modeled in Verilog-A, using NEGF-LLG (non-equilibrium Green's function and Landau-Lifshitz-Gilbert equations) solution for spin to charge inter-face and calibrated with the experimental data in [36].
2) Circuit level: For the circuit level simulation, a Verilog-A model of 3T-Spin-TD is developed to co-simulate with the interface CMOS circuits in Cadence Spectre and SPICE. 45 nm North Carolina State University (NCSU) Product Development Kit (PDK) library [42] is used in SPICE to verify the proposed design and acquire the performance (power, delay, etc.) of designs.
3) Application level: We consider a widely-used image compression algorithm, the Discrete Cosine Transform (DCT),

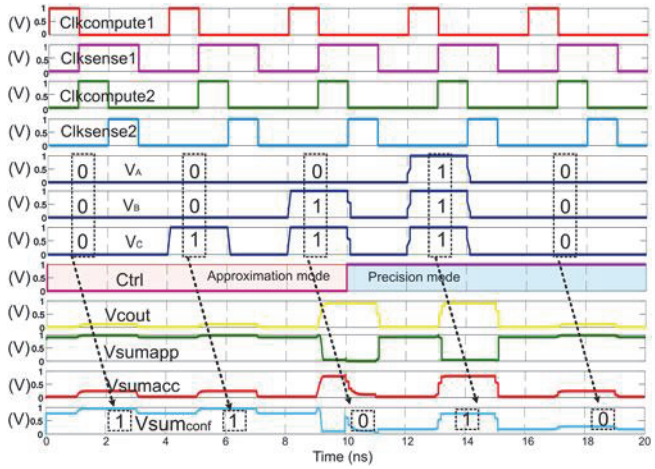


Fig. 9. Transient voltage analysis of the proposed accuracy-configurable FA cell.

to show the results of using the proposed accuracy-configurable adder and approximate compressor-based multipliers at the application level.

This section deals with device and circuit-level evaluations; however, Section VII is fully dedicated to application level evaluations.

A. Accuracy-Configurable Adder

Fig. 9 depicts waveforms of transient voltage analysis of the proposed accuracy-configurable FA cell. A 3 ns period is considered as a full computation cycle for the circuit. Both stages use identical pulse widths of 1 ns for CLK_{compute}. Stage 1 uses a 2 ns CLK_{sense1} signal for proper implementation of sensing and Stage 2 uses 1 ns CLK_{sense2}. Since C_{out} in Stage 1 is used in the next stage MG, it should last 2 ns to be synchronized with the sum generated in Stage 2. Four input combinations regardless of the sequence (000, 001, 011, and 111) are considered as input vectors (where V_A, V_B and V_C are A, B, and C voltages, respectively). Moreover, V_{Cout}, $V_{Sum_{\text{app}}}$, and $V_{Sum_{\text{acc}}}$ stand for C_{out}, approximate sum, and accurate sum voltages, respectively.

In the approximation mode (*Ctrl* $= 0$), when $Clk_{compute1}$ is high, the input voltages are applied to Stage 1 circuit for 1 ns. Clk_{sense1} is then activated leading to generate the first stage output voltages (V_{Cout} and $V_{Sum_{\text{app}}}$). As is clear in Fig. 9, for three input combinations of (000, 001, and 011), the final Sum signal $V_{Sum_{\text{conf}}}$ is (1, 1 and 0) corresponding to $V_{Sum_{\text{app}}}$. It is noteworthy that in the approximation mode, besides switching off the intermediate TGs connecting Stage 1 to Stage 2, power gating is also employed to reduce the power consumption of Stage 2. In the precision mode (*Ctrl* $= 1$), the input voltages are applied to Stage 1 and Stage 2 in two consecutive nanoseconds when $Clk_{compute1}$ and $Clk_{compute2}$ are respectively high. After the computation clock of Stage 1, Clk_{sense1} should be activated for 2 ns in a manner such that the required inputs are fed to the second stage and synchronized outputs are provided for the FA. As is clear in Fig. 9, the valid results can be obtained after applying Clk_{sense2} so that for two input combinations of (000 and 111), the final Sum signal $V_{Sum_{\text{conf}}}$ is 0 and 1 corresponding to $V_{Sum_{\text{acc}}}$.

Comparison results between the proposed adder and previously published CMOS- [1], [43], MTJ- [26], [43], Spin Hall Effect (SHE)- [20] and Domain Wall Motion (DWM)- [19] based FAs are summarized in Table V. Various metrics including the device count, total power consumption, and delay are considered for the comparison. In addition, the important approximate computing metric, Error Distance (ED) [44] is used for approximate adders' evaluation. Basically, in any approximate circuit, the inexact output a and accurate output b is compared arithmetically for all possible combination inputs bit by bit: $ED(a,b) = |a-b| = \left|\sum_i a[i]\cdot 2^i - \sum_j b[j]\cdot 2^j\right|$, where i and j are the indices for the bits in a and b [5], [45]. Here, we report Error Rate (ER), Mean Error Distance (MED), as the average of the error distances across all possible input vectors, and Mean Relative Error Distance (MRED) for different designs. The MRED is computed by averaging all possible absolute relative error distances (RED) (i.e., $RED = \left|\frac{ED}{b}\right|$), where the RED is not considered when the accurate output b is 0.

As shown in Table V, the proposed design in approximation mode shows smaller ER, MED and MRED compared to the approximate designs in [26]. However, it shows identical values to the proposed designs in [1], [23]. Since the design proposed in [23] was implemented in NML technology and there was no performance metrics reported in this reference, so the power/delay analysis of the design is inevitably left for future investigations.

Based on Table V, the accuracy-configurable circuit in this work along with the presented designs in [26] are the only adders with the approximation configurability. For a fair comparison, since most of the counterpart designs were designed and evaluated in 180 nm, we scaled ours and others to this process node. We have done fixed-voltage scaling by using the appropriate scaling factor, which is $(1/S^2)$ for area and $(1/S)$ for energy [46]. In addition, CMOS FAs contain one output register along with FA cell since non-volatile designs also have memory functions.

The results clearly show that the proposed accuracy-configurable adder consumes smaller power than the other designs in [19], [20], [26], [43]. For instance, 34.58% and 66% improvement in power consumption can be reported for the precision and approximation modes, respectively, over the best DWM-based FA design in [19]. In addition, compared to the recently-published work by Roohi *et al.* in [20], the proposed FA in precision mode can show $\sim$12.7$\times$ and 2.3$\times$ smaller power and delay, respectively.

The area-efficient accuracy-configurable adder also exhibits $\sim 18\%$ reduction in circuit complexity over the accurate CMOS-based FA design in [43]. However, the proposed design utilizes 28 MOS transistors, which is more than the designs in [1], [19], [26]. It is worth pointing out that the device count can offer a representative estimation of the area overhead since the proposed full adder is more compactly implemented than a CMOS implementation [19], [43].

TABLE V
COMPARISON OF FA DESIGNS

Designs	Type	ER[1](%)	MED[2]	MRED[3] (%)	Device count[4]	Power[5]	Delay[6]	Conf.[8]
CMOS [43]	Accurate	0	0	0	42T	71.1μW+ 0.9nW	2200ps	No
CMOS [1]	Approximate	25	0.25	4.17	14T	32.5μW+ 2.1nW	645ps	No
MTJ-based [43]	Accurate	0	0	0	34T+4M	2100 μW + 0nW	10200ps	No
MTJ-based [26]	Approximate	50	0.5	29.17	21T+4M	1702.6μW + 329.5pW	3016.22ps	Yes
MTJ-based [26]	Accurate	0	0	0	25T+4M	1895.1μW + 401.6pW	3019.3ps	Yes
MTJ-based [26]	Approximate	50	0.5	31.25	25T+4M	784.5μW +77.91pW	3152.7ps	Yes
SHE-based [20]	Accurate	0	0	0	23T+3SM	710μW + 0nW	7000ps	No
HPM DWM [19]	Accurate	0	0	0	20T+4M+2D	1364μW + 0nW	269ps	No
LPM DWM [19]	Accurate	0	0	0	20T+4M+2D	85μW+ 0nW	877ps	No
Prop. FA	Accurate	0	0	0	28T+4M+2D	55.6μW+ 0nW	3000ps[7]	Yes
Prop. FA	Approximate	25	0.25	4.17	28T+4M+2D	28.9μW+ 0nW	2000ps*	Yes

Note: To attain a fair comparison, technology scaling is applied. (1) Error Rate. (2) Mean Error Distance. (3) Mean Relative Error Distance. (4) T: MOS Transistor, M: MTJ, SM: SHE-MTJ, D: DW. (5) Total power including write and read operations: dynamic power + static power. Power must be supplied to keep data in CMOS-based storage circuit at any time. However, it can be cut-off in the non-volatile designs. (6) Total delay including write and read operations. (7) 1000ps considering the pipeline technique. (8) Provision of approximation configurability.

TABLE VI
COMPARISON OF ACCURATE AND APPROXIMATE COMPRESSOR DESIGNS

Designs[1]	Device count	Power (μW)	Delay[2] (ns)
MTJ [43]	68T+8M	4200	20.4
HPM DWM [19]	46T+8M+4D	2728	2.54
LPM DWM [19]	46T+8M+4D	170.2	3.7
Design I	22T+4M+2D	57.8	3
Design II	33T+6M+3D	84.5	4

(1) Accurate compressors are designed based on the FAs in the references.
(2) Total delay including write and read operations.

The proposed adder does not improve delay compared to the previous designs in [1], [19], nonetheless it can achieve higher speed and throughput using pipeline techniques without any additional clock control circuit. A fully pipelined design can be realized by alternately applying two clock signals on neighboring stages, for instance, in an *n*-bit adder structure. Hence, the proposed adder's throughput can be considerably increased to one output set per 1 ns, which leads to an equivalent 1 ns delay. A larger current injection to the MG could lead to a higher computation speed, but it also leads to a higher power consumption. Furthermore, an embedded buffer can be presumed for spintronic devices due to their non-volatility characteristic; however, such a buffer should be inserted between every other logic gates working at different operational phases in a CMOS design. The designs in [19] also lack the appropriate input circuit such that driving transistors are needed for cascading to other cells. This point is also taken into account in the design of compressors using cascaded FAs, as evaluated next.

B. Approximate Compressors

We have evaluated the performance of proposed approximate 4-2 compressors in terms of device count, total power and delay. Three different accurate spintronic FAs (i.e. MTJ-based [43], LPM-DWM [19] and HPM-DWM [19]) listed in Table V are used for constructing accurate 4-2 compressors as Fig. 5(a). To make the counterpart designs cascadable, appropriate input transistors are added. Table VI compares their simulation results with the proposed hybrid spin-CMOS approximate compressors (as delineated via Designs I and II). It can be seen that Design I shows significant reduction in power consumption compared to other designs, with ∼66%, 97.8% and 98.6% less power than LPM-DWM [19], HPM-DWM [19] and MTJ-based compressors [43], respectively. In addition, ∼19% speed-up is achieved compared to LPM-DWM based compressor.

VII. APPLICATIONS

In this section, we focus on image compression algorithms and show the results of using accuracy-configurable adder and approximate compressor-based multipliers in such applications. Most of DSP algorithms use two basic operations: additions and multiplications. Thus, we expect that leveraging the proposed majority-based primitives could provide limited accuracy loss for improvements in other circuit metrics such as power and speed. The Discrete Cosine Transform (DCT) and Inverse DCT (IDCT) are the kernel of the international standard lossy image compression algorithm referred to as JPEG [47]. The interesting feature of DCT is that, for a typical image, most of the visually important information is concentrated in a few coefficients of DCT. One-dimensional integer DCT for an 8-point sequence $x(i)$ is given by

$$y(k) = \sum_{i=0}^{7} f(k,i)x(i), k = 0,1,2,...,7 \tag{8}$$

We assess the output quality of the decoded image after IDCT employing the well-known metric of peak signal-to-noise ratio (PSNR) which is based on the mean square error (MSE):

$$\text{MSE} = \frac{1}{mp}\sum_{i=0}^{m-1}\sum_{j=0}^{p-1}[I(i,j) - F(i,j)]^2 \tag{9}$$

$$\text{PSNR} = 10\log_{10}\left(\frac{{MAX_I}^2}{MSE}\right) \tag{10}$$

In (9), m and p denote terms for the image dimensions; $I(i,j)$ and $F(i,j)$ are the exact and computed values of each pixel, respectively. In (10), MAX_I represents the maximum value of each pixel.

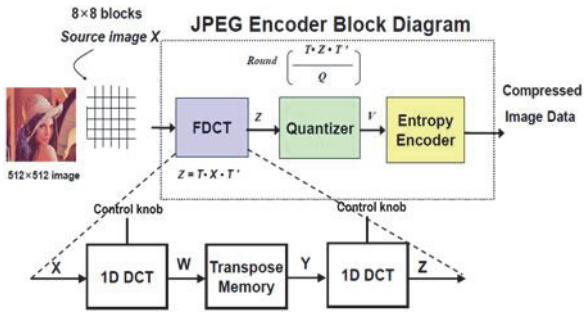


Fig. 10. System block diagram of DCT/IDCT architecture.

A. Accuracy-Configurable Adder

To efficiently implement DCT-IDCT employing the proposed accuracy-configurable adder, each $f(k,i)$ (i.e. cosine functions) in (8), is converted into an integer [38]. As thoroughly discussed in [1], [48], the integer output $y(k)$ is accordingly right shifted to produce the actual DCT output. An identical expression is also presented in [48] for 1-D integer IDCT. We change the integer coefficient $f(k,i)$ for $k = 1, 2, \ldots, 7$ in order that the multiplication between $f(k,i)$ and $x(i)$ can be equivalently implemented by two left-shifts and an addition. The most significant coefficient $f(0,i)$ is left unchanged. In this way, $f(0,i)x(i)$ is basically the sum of 4 terms, so it can be implemented with a CSA tree by a 4-2 compressor followed by a Ripple Carry Adder (RCA). In addition, every DCT/IDCT output is the addition of eight terms that can be computed employing a CSA tree (implemented by an 8-2 compressor) followed by an RCA. Therefore, the entire DCT–IDCT system can be implemented employing RCAs and CSAs and can be approximated using the proposed adder.

We use the approximation mode of the proposed accuracy-configurable FA only in the LSBs of adders in a 20-bit DCT-IDCT architecture while exploiting the precision mode in MSBs. Accordingly, as depicted in Fig. 10, the output quality can be controlled in DCT blocks using the control knob regulating the operation mode of the proposed adders. The simulation results are obtained by using Matlab with an Intel Core i7 processor and 4GB RAM. Fig. 11 shows the processing quality of the examined image in the base case (i.e., 20-bit in precision mode), 8-, 10-, and 12-LSB cases. As shown, there is some loss of quality in the reconstructed image in Fig. 11(c) using approximate adders at 10 LSBs with the PSNR (26.93 dB), however the image is still well recognizable. Fig. 12(a) shows the output quality for the base case and five different degrees of approximations in PSNR. It can be seen that by increasing the approximation degree from the base case to 8 LSBs, the PSNR only drops by 2.93 dB.

The power consumption of the DCT-IDCT circuit is evaluated using Synopsys Design Compiler for both pure-CMOS and spin-CMOS circuits as depicted in Fig. 12(b). For pure-CMOS and spin-CMOS circuits, a Verilog code describing the truth table in Table III is considered for implementing the approximate adder based on existing and developed cell libraries, receptively, which is then used in 8-12 LSBs of a 20-bit DCT-IDCT architecture. Simulation results show that for all cases the power

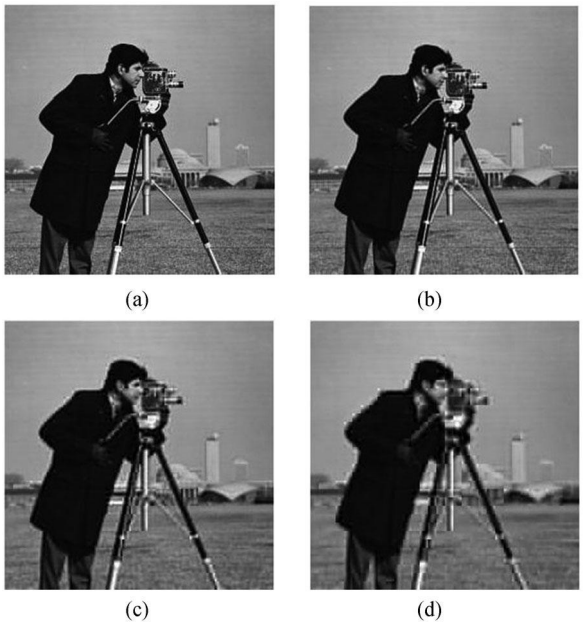


Fig. 11. Compressed images and corresponding PSNR. (a) Base case (33.73 dB). (b) 8 LSBs (30.82 dB). (c) 10 LSBs (26.93 dB). (d) 12 LSBs (23.75 dB).

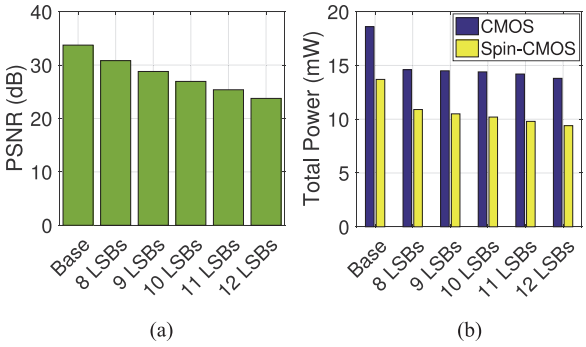


Fig. 12. (a) Output quality comparison of different approximations. (b) Power consumption comparison of CMOS and spin-CMOS DCT-IDCT.

dissipation of the proposed spin-CMOS architecture is smaller than the CMOS counterpart. Evidently, by changing the degree of approximation, the power consumption of the entire system is changed. For instance, 31.33% power saving is obtained for the spin-CMOS architecture with 12 approximate LSBs in comparison with the base case, although the output quality is degraded to a PSNR of 23.75 dB. In a similar scenario, 8 approximate LSBs provide power saving of 20.4%, although the output quality is slightly degraded to 30.82 dB.

B. Approximate Compressor-Based Multipliers

As mentioned earlier, in the DCT-IDCT computation, the multiplication operations can be implemented by the approximate compressor-based multipliers, while the additions remain accurate. As the DCT coefficients are in the range of $(-1, 1)$, they are multiplied by 2^{15} to be converted into 16-bit signed binary numbers in 2's complement representation. Hence, the matrix multiplication in DCT and IDCT are implemented by

TABLE VII
COMPARISON OF THE ACCURATE AND APPROXIMATE COMPRESSOR-BASED MULTIPLIERS FOR DCT-IDCT

Design	Accurate	Approximate (32 bits)		Approximate (16 bits)		Approximate (13 bits)		Approximate (12 bits)	
		Design I	Design II	Design I	Design II	Design I	Design II	Design I	Design II
PSNR (dB)	Inf	4.0948	4.0948	13.0542	14.1232	37.0205	37.8094	50.2156	50.9583
Delay reduction (ns)	-	108.89	102.19	85.64	80.44	75.08	73.25	69.56	66.71
Energy reduction (mJ)	-	140.24	118.05	91.12	89.99	80.16	77.29	74.44	71.73

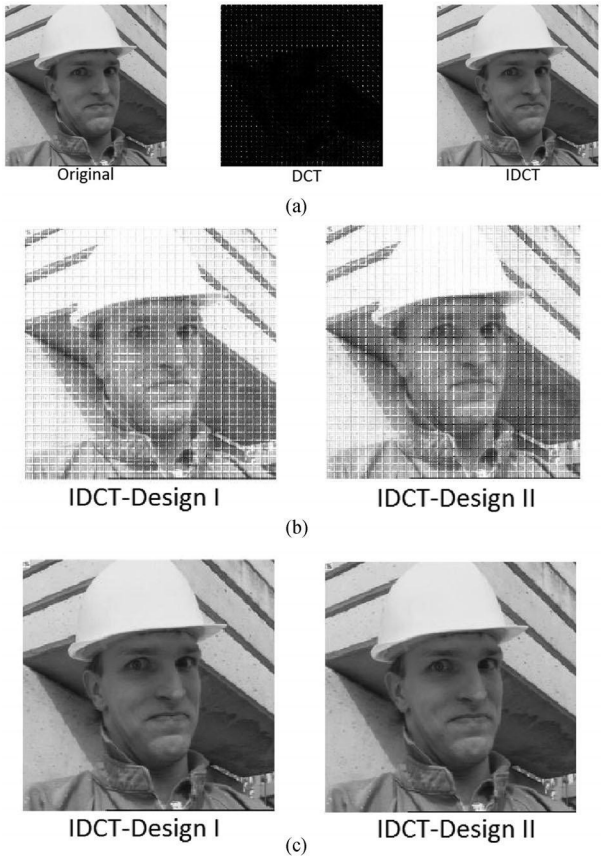


Fig. 13. DCT-IDCT results of using (a) accurate compressor, (b) approximate compressors in half LSBs (16 bits), and (c) approximate compressors for 12 LSBs.

16 × 16 approximate signed multipliers. To obtain the best trade-off, different configurations of 16 × 16 approximate signed multipliers are employed for the matrix multiplication in the DCT and IDCT algorithms. A configuration means using the proposed approximate 4-2 compressors for the accumulation of a different number of columns of least significant partial product bits. The signed multiplier is implemented by using the Baugh-Wooley algorithm, thus, a similar partial product array is obtained as the unsigned multiplier. As in [13], the partial products of the signed multiplier are accumulated by a Dadda tree. The accurate addition is implemented using the proposed accuracy-configurable adder in precision mode. We run the experiment using the approximate compressors at all, half (16), 13 and 12 LSBs of the multipliers. Fig. 13(a) shows the accurate results for the DCT-IDCT implementation. The results of using approximate compressors on half and 12 LSBs are shown in Fig. 13(b) and (c), respectively.

The reconstructed images reveal that using the approximate compressors for all partial product bits or half LSBs cause image distortion, while the reconstructed images using approximate compressors on 12 LSBs show a similar quality with the accurate result. The defects in the image generated by the multiplier using approximation on the half [Fig. 13(b)] and 13 LSBs are visible after zooming in. The PSNR values provided in Table VII indicate the same conclusion. The delay and energy reduction of using approximate compressor-based multipliers compared to accurate MTJ-based multiplier [43] are also listed in Table VII. The total number of approximate compressors used in different configurations is obtained to evaluate the respective energy reduction. As for delay reduction, the total number of approximate compressors in the critical path is obtained. The results indicate that the DCT/IDCT systems using the approximate compressor-based multipliers achieve ∼50% reduction in energy consumption and 3x speed-up compared to the exact circuit with a comparable output quality. Obviously, by sacrificing the quality, system attains even higher energy-efficiency and speed-up. It is noteworthy that in all cases, the multiplier which is based on Design I has provided better result in terms of energy and delay with lower PSNR as compared to that of Design II.

VIII. CONCLUSION

In this paper, a compact and energy-efficient accuracy-configurable adder design and two approximate compressors based on a composite spintronic device structure have been developed and assessed. Based on the majority logic, the proposed designs can be effectively utilized to trade off computation energy and solution quality in DSP systems. A device-to-application simulation framework has been constructed and shown to be effective to evaluate the proposed hybrid spin-CMOS circuits. Furthermore, the proposed accuracy-configurable adder and approximate compressors are efficiently-utilized in a DCT block to fully-realize a widely-used digital image processing algorithm. The results indicate that the DCT/IDCT using an approximate multiplier achieves ∼50% energy consumption while attaining roughly 3x speed-up compared to the exact MTJ-based design with a comparable accuracy.

REFERENCES

[1] V. Gupta, D. Mohapatra, A. Raghunathan, and K. Roy, "Low-power digital signal processing using approximate adders," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 32, no. 1, pp. 124–137, Jan. 2013.
[2] H. Jiang, J. Han, F. Qiao, and F. Lombardi, "Approximate radix-8 booth multipliers for low-power and high-performance operation," *IEEE Trans. Comput.*, vol. 65, no. 8, pp. 2638–2644, Aug. 2016.
[3] B. Li, P. Gu, Y. Shan, Y. Wang, Y. Chen, and H. Yang, "RRAM-based analog approximate computing," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 34, no. 12, pp. 1905–1917, Dec. 2015.

[4] Y. Kim, S. Venkataramani, K. Roy, and A. Raghunathan, "Designing approximate circuits using clock overgating," in *Design Automation Conference (DAC), 2016 53nd ACM/EDAC/IEEE*, 2016, pp. 1–6: IEEE.

[5] J. Han and M. Orshansky, "Approximate computing: An emerging paradigm for energy-efficient design," in *Proc. 18th IEEE Eur. Test Symp.*, 2013, pp. 1–6.

[6] B. Shim, S. R. Sridhara, and N. R. Shanbhag, "Reliable low-power digital signal processing via reduced precision redundancy," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 12, no. 5, pp. 497–510, May 2004.

[7] D. Mohapatra, G. Karakonstantis, and K. Roy, "Significance driven computation: A voltage-scalable, variation-aware, quality-tuning motion estimator," in *Proc. 2009 ACM/IEEE Int. Symp. Low Power Electron. Des.*, 2009, pp. 195–200.

[8] H. Jiang, C. Liu, L. Liu, F. Lombardi, and J. Han, "A review, classification, and comparative evaluation of approximate arithmetic circuits," *ACM J. Emerg. Technol. Comput. Syst. (JETC)*, vol. 13, no. 4, p. 60, 2017.

[9] A. K. Verma, P. Brisk, and P. Ienne, "Variable latency speculative addition: A new paradigm for arithmetic circuit design," in *Proc. Conf. Des., Autom. Test Eur.*, 2008, pp. 1250–1255.

[10] C.-H. Chang, J. Gu, and M. Zhang, "Ultra low-voltage low-power CMOS 4-2 and 5-2 compressors for fast arithmetic circuits," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 10, pp. 1985–1997, Oct. 2004.

[11] M. Moaiyeri, F. Sabetzadeh, and S. Angizi, "An efficient majority-based compressor for approximate computing in the nano era," *Microsyst. Technol.*, vol. 24, no. 3, pp. 1589–1601, 2017.

[12] D. Baran, M. Aktan, and V. G. Oklobdzija, "Energy efficient implementation of parallel CMOS multipliers with improved compressors," in *Proc. ACM/IEEE Int. Symp., Low-Power Electron. Des.*, 2010, pp. 147–152.

[13] A. Momeni, J. Han, P. Montuschi, and F. Lombardi, "Design and analysis of approximate compressors for multiplication," *IEEE Trans. Comput.*, vol. 64, no. 4, pp. 984–994, Apr. 2015.

[14] O. Akbari, M. Kamal, A. Afzali-Kusha, and M. Pedram, "Dual-quality 4: 2 compressors for utilizing in dynamic accuracy configurable multipliers," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 25, no. 4, pp. 1352–1361, Apr. 2017.

[15] X. Fong *et al.*, "Spin-transfer torque devices for logic and memory: Prospects and perspectives," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 35, no. 1, pp. 1–22, Jan. 2016.

[16] Y. Gang, W. Zhao, J.-O. Klein, C. Chappert, and P. Mazoyer, "A high-reliability, low-power magnetic full adder," *IEEE Trans. Magn.*, vol. 47, no. 11, pp. 4611–4616, Nov. 2011.

[17] H. Cai, Y. Wang, L. A. Naviner, Z. Wang, and W. Zhao, "Approximate computing in MOS/spintronic non-volatile full-adder," in *Proc. IEEE/ACM Int. Symp., Nanoscale Architectures.*, 2016, pp. 203–208.

[18] E. Deng *et al.*, "Robust magnetic full-adder with voltage sensing 2t/2MTJ cell," in *Proc. 2015 IEEE/ACM Int. Symp., Nanoscale Architectures*, 2015, pp. 27–32.

[19] A. Roohi, R. Zand, and R. F. DeMara, "A tunable majority gate-based full adder using current-induced domain wall nanomagnets," *IEEE Trans. Magn.*, vol. 52, no. 8, pp. 1–7, Aug. 2016.

[20] A. Roohi, R. Zand, D. Fan, and R. F. DeMara, "Voltage-based concatenatable full adder using spin hall effect switching," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 36, no. 12, pp. 2134–2138, Dec. 2017.

[21] V. Pudi, K. Sridharan, and F. Lombardi, "Majority logic formulations for parallel adder designs at reduced delay and circuit complexity," *IEEE Trans. Comput.*, vol. 66, no. 10, pp. 1824–1830, Oct. 2017.

[22] Z. Rouhani, S. Angizi, M. Taheri, K. Navi, and N. Bagherzadeh, "Towards approximate computing with quantum-dot cellular automata," *J. Low Power Electron.*, vol. 13, no. 1, pp. 29–35, 2017.

[23] C. Labrado, H. Thapliyal, and F. Lombardi, "Design of majority logic based approximate arithmetic circuits," in *Proc. IEEE Int. Symp., Circuits Syst.*, 2017, pp. 1–4.

[24] S. Angizi, Z. He, N. Bagherzadeh, and D. Fan, "Design and evaluation of a spintronic in-memory processing platform for non-volatile data encryption," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, 2017. doi: 10.1109/TCAD.2017.2774291.

[25] S. Jain, S. Venkataramani, and A. Raghunathan, "Approximation through logic isolation for the design of quality configurable circuits," in *Proc. Des., Autom. Test Eur*. Conf. Exhib., 2016, pp. 612–617.

[26] H. Cai, Y. Wang, L. A. D. B. Naviner, and W. Zhao, "Robust ultra-low power non-volatile logic-in-memory circuits in fd-soi technology," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 64, no. 4, pp. 847–857, Apr. 2017.

[27] S. Angizi, Z. He, R. F. DeMara, and D. Fan, "Composite spintronic accuracy-configurable adder for low power digital signal processing," in *Proc. 18th Int. Symp. Qual. Electron. Des.,* 2017, pp. 391–396.

[28] D. Fan, "Ultra-low energy reconfigurable spintronic threshold logic gate," in *Proc. 26th Ed. Great Lakes Symp. VLSI*, 2016, pp. 385–388.

[29] S. Gu, E. H.-M. Sha, Q. Zhuge, Y. Chen, and J. Hu, "Area and performance co-optimization for domain wall memory in application-specific embedded systems," in *Proc. 52nd Annu. Des. Autom. Conf*, 2015, p. 20: ACM.

[30] W. Zhao, D. Ravelosona, J. Klein, and C. Chappert, "Domain wall shift register-based reconfigurable logic," *IEEE Trans. Magn.*, vol. 47, no. 10, pp. 2966–2969, Oct. 2011.

[31] J. Kim *et al.*, "Spin-based computing: Device concepts, current status, and a case study on a high-performance microprocessor," *Proc. IEEE*, vol. 103, no. 1, pp. 106–130, Jan. 2015.

[32] Y. Wang *et al.*, "Compact model of dielectric breakdown in spin-transfer torque magnetic tunnel junction," *IEEE Trans. Electron Devices*, vol. 63, no. 4, pp. 1762–1767, Apr. 2016.

[33] D. Fan, M. Sharad, and K. Roy, "Design and synthesis of ultralow energy spin-memristor threshold logic," *IEEE Trans. Nanotechnology*, vol. 13, no. 3, pp. 574–583, May 2014.

[34] 1999. [Online]. Available: http://math.nist.gov/oommf/

[35] S. Fukami *et al.*, "20-nm magnetic domain wall motion memory with ultralow-power operation," in *Proc. IEEE Int., Electron Devices Meeting*, 2013, pp. 3–5.

[36] 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 *Proc. Int. Conf. Simul. Semicond. Processes Devices*, 2011, pp. 51–54.

[37] B. Parhami, *Computer Arithmetic*. vol. 20, Oxford, U.K.: Oxford Univ. Press, 1999.

[38] V. Gupta, D. Mohapatra, S. P. Park, A. Raghunathan, and K. Roy, "Impact: Imprecise adders for low-power approximate computing," in *Proc. 17th IEEE/ACM Int. Symp. Low-power Electron. Des.*, 2011, pp. 409–414.

[39] H. Jiang, J. Han, and F. Lombardi, "A comparative review and evaluation of approximate adders," in *Proc. 25th Ed. Great Lakes Symp. VLSI*, 2015, pp. 343–348.

[40] R. Zhang, K. Walus, W. Wang, and G. A. Jullien, "Performance comparison of quantum-dot cellular automata adders," in *Proc. IEEE Int. Symp., Circuits Syst*, 2005, pp. 2522–2526.

[41] M. R. Azghadi, O. Kavehie, and K. Navi, "A novel design for quantum-dot cellular automata cells and full adders," *Journal of Applied Sciences*, vol. 7, pp. 3460–3468, 2007.

[42] Ncsu eda freepdk45. 2011. [Online]. Available: http://www.eda.ncsu.edu/wiki/FreePDK45:Contents

[43] S. Matsunaga *et al.*, "Fabrication of a nonvolatile full adder based on logic-in-memory architecture using magnetic tunnel junctions," *Appl. Phys. Express*, vol. 1, no. 9, 2008, Art. no. 091301.

[44] S. Dutt, S. Nandi, and G. Trivedi, "Analysis and design of adders for approximate computing," *ACM Trans. Embedd. Comput. Syst. (TECS)*, vol. 17, no. 2, p. 40, 2018.

[45] J. Liang, J. Han, and F. Lombardi, "New metrics for the reliability of approximate and probabilistic adders," *IEEE Trans. Comput.*, vol. 62, no. 9, pp. 1760–1771, Sep. 2013.

[46] Z. Abbas and M. Olivieri, "Impact of technology scaling on leakage power in nano-scale bulk CMOS digital standard cells," *Microelectronics J.*, vol. 45, no. 2, pp. 179–195, 2014.

[47] G. K. Wallace, "The jpeg still picture compression standard," *IEEE Trans. Consumer Electron.*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.

[48] G. Karakonstantis, D. Mohapatra, and K. Roy, "System level DSP synthesis using voltage overscaling, unequal error protection & adaptive quality tuning," in *Proc. IEEE Workshop, Signal Process. Syst.*, 2009, pp. 133–138.

Shaahin Angizi (S'15) received the B.Sc. degree in computer engineering, hardware from South Tehran Branch Islamic Azad University (IAU), Tehran, Iran in 2012 and the M.Sc. degree in computer engineering, computer systems architecture from Science and Research Branch, IAU, Tabriz, Iran in 2014. He is currently working toward the Ph.D. degree in computer engineering at University of Central Florida, Orlando, FL, USA. His research interests include in-memory computing, deep learning, low power VLSI designs, spin-based computing, and quantum-dot cellular automata.

Honglan Jiang (S'14) received the B.S. and Master's degrees in instrument science and technology from Harbin Institute of Technology, Harbin, China, in 2011 and 2013, respectively. Since September 2013, she has been working toward the Ph.D. degree at the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. Her current research interests are approximate computing and stochastic computing.

Ronald F. DeMara (SM'05) received the Ph.D. degree in computer engineering from the University of Southern California, LA, CA, USA, in 1992. Since 1993, he has been a Full-Time Faculty Member with the University of Central Florida Orlando, FL, USA, where he is a Professor and a Computer Engineering Program Coordinator. His research interests include computer architecture with emphasis on evolvable hardware and emerging devices, on which he has authored and coauthored approximately 225 articles. He was an Editorial Boards of the IEEE TRANSACTIONS ON VLSI SYSTEMS, *ACM Transactions on Embedded Systems, Journal of Circuits, Systems, and Computers*, the *Journal Microprocessors and Microsystems*, various conference program committees, and is currently a Topical Editor of the IEEE TRANSACTIONS ON COMPUTERS. He received the Joseph M. Biedenbach Outstanding Engineering Educator Award in 2008, the highest educational honor from IEEE in the Southeast United States.

Jie Han (SM'16) received the B.Sc. degree in electronic engineering from Tsinghua University, Beijing, China, in 1999 and the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 2004. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. His research interests include approximate computing, stochastic computation, reliability and fault tolerance, nanoelectronic circuits and systems, and novel computational models for nanoscale and biological applications.

Deliang Fan (M'15) received the B.S. degree in electronic information engineering from Zhejiang University, Hangzhou China, in 2010. He received the M.S. and Ph.D. degrees in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2012 and 2015, respectively.

In 2015, he was an Assistant Professor with the Department of Electrical and Computer Engineering, University of Central Florida, Orlando, FL,USA. His primary research interest include ultralow 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.