Design and Synthesis of Ultralow Energy Spin-Memristor Threshold Logic

Deliang Fan, Mrigank Sharad, and Kaushik Roy, *Fellow, IEEE*

***Abstract*—A threshold logic gate performs weighted sum of multiple inputs and compares the sum with a threshold. We propose spin-memeristor threshold logic (SMTL) gates, which employ a memristive cross-bar array to perform current-mode summation of binary inputs, whereas the low-voltage fast-switching spintronic threshold devices carry out the threshold operation in an energy efficient manner. Field-programmable SMTL gate arrays can operate at a small terminal voltage of ∼50 mV, resulting in ultralow power consumption in gates as well as programmable interconnect networks. We evaluate the performance of SMTL using threshold logic synthesis. Results for common benchmarks show that SMTL-based programmable logic hardware can be more than 100 × energy efficient than the state-of-the-art CMOS field-programmable gate array.**

***Index Terms*—Boolean functions, magnetic domains, memristor, nanotechnology, programmable logic arrays, threshold logic (TL).**

I. Introduction

In recent years, the "memristor" has been proposed for fabricating nanoscale programmable resistive elements [1]–[3]. The compatibility with state-of-the-art CMOS technology is of special interest to researchers, like memristor based on Ag–Si filaments [2]. Such devices can be integrated into metallic cross-bars to obtain high density memristive cross-bar arrays (MCAs). The design of multilevel, nonvolatile memory [3], [4] is enabled by continuous range of resistance values obtainable in these devices. Application of the specific device characteristics of memristors in unconventional, computing schemes like neural networks [5], [6] and threshold logic (TL) [7]–[9], has been explored in recent years.

A threshold logic gate (TLG) operation essentially constitutes of summation of weighted inputs, followed by a threshold operation [10] (1). While a memristor array can be employed to perform current mode analog summation of binary input voltage signals, the thresholding operation requires the application of a current comparator circuit. Such a comparison operation can be obtained using conventional analog circuits based on current mirrors [7] or voltage-comparators [8], [9]. However, such analog CMOS circuits often consume significant power and area, thereby eschewing the energy and density benefits of nanodevices. Rather than depending upon analog CMOS circuits for implementing current comparison, it would be desirable to explore nanodevices that can directly provide such a current mode thresholding characteristic.

Recent experiments on spin-torque devices have demonstrated high speed switching of scaled nanomagnets with small current densities [11]–[14]. Such a phenomenon can be used to design compact and low voltage current mode spintronic switches and simultaneously provide energy-efficient current-to-voltage conversion. Application of such spin-torque switches in memory [28], [29], digital [15], [30], analog [16], and neuromorphic computing applications [17] has been explored earlier. Such nanoscale, spintronic devices inherently act as compact, ultralow voltage and fast current comparators and hence, can be highly suitable for memristor-based TLG design.

In this study, we propose spin-memristor threshold logic (SMTL) design using such spin-torque switches based on magnetic domain wall (DW) motion [13]. The magneto-metallic DW switch allows ultralow voltage operation of memristive TLGs leading to low energy dissipation at the gate level. We name our proposed DW switch structure as spintronic threshold device (STD). It can facilitate ultralow voltage current mode interconnect for the design of fully programmable, large TL blocks. This helps to achieve highly reduced energy dissipation in programmable interconnects. Notably, in CMOS lookup table (LUT)-based conventional field-programmable gate arrays (FPGAs), more than 90% of energy can be ascribed to programmable switches and interconnects [19]. Further, the STD being nonvolatile magnetic switches inherently act as a latch and hence can facilitate fully pipelined connection of multiple TLG stages without the insertion of additional memory elements like flip–flops. This can provide high-performance and integration density for complex data processing blocks. The aforementioned factors combined together lead to ultralow energy consumption.

In this paper, we also present a comprehensive methodology for SMTL design, synthesis, and optimization and compare its performance with conventional CMOS FPGAs. The remainder of this paper is organized as follows. Section II discusses some of the previous work related to the design of a memristor cross-bar array for weighted sum of TL inputs. In Section III, we introduce the STD for the sign function in TLG. Circuit design and optimization for SMTL are presented in Sections IV and V, respectively. Section VI presents the SMTL synthesis methodology. The performance and prospects of SMTL are discussed in Section VII. Section VIII concludes the paper.

Manuscript received November 7, 2013; accepted February 14, 2014. Date of publication March 20, 2014; date of current version May 6, 2014. This work was supported in part by CSPIN, StarNet center, Defense Advanced Research Projects Agency, Semiconductor Research Corporation, Intel, and National Science Foundation. The review of this paper was arranged by Associate Editor A. A. Balandin.

The authors are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47906 USA (e-mail: dfan@purdue.edu; msharad@purdue.edu; kaushik@purdue.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNANO.2014.2312177

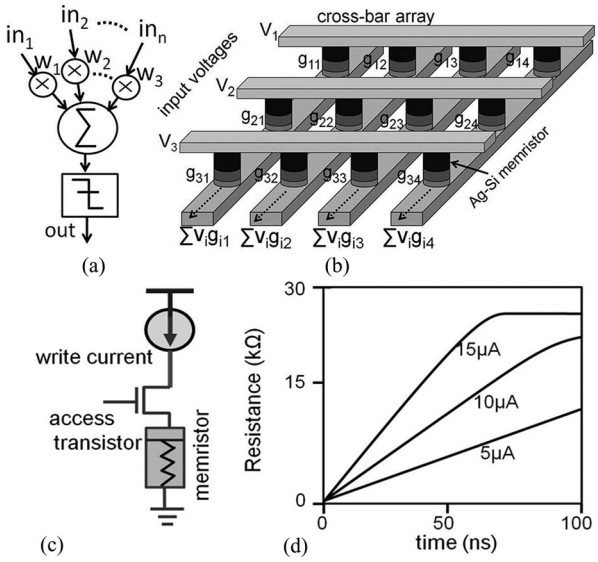


Fig. 1. (a) Schematic representation of a TLG, (b) MCA, (c) resistive memory cell with access transistors, and (d) transient change in resistance for different magnitude of programming current.

II. TLG First Stage Design Using the MCA

In this section, we review the recent progress in MCA design, programming and its application as the first stage of TL computation.

A TL operation shown in Fig. 1(a) can be expressed in the form:

$$Y = \text{sign}\left(\sum (X_i W_i + b_i)\right) \tag{1}$$

where X_is are multiple binary inputs to a threshold gate, W_is are scalar weights with which the corresponding inputs are multiplied (or scaled), and b_i is the bias for the ith gate. Note that W_i can be either positive or negative. Hence, depending upon the input combination (assuming unipolar values of inputs, i.e., 1 and 0), the summation can yield either a positive or a negative value, result of which is determined by the sign function (involving a comparison operation). The first stage of the TL computation is the scaling and summation of the inputs, which can be implemented using an MCA, as shown in Fig. 1(b). The detailed design and programming of MCA will be introduced in this section. The second stage of TL computing is a "sign" [in (1), or threshold] function, which will be implemented using the proposed STD described in Section III.

A. Multilevel MCA

Fig. 1(b) depicts an MCA that constitutes of memristors (Ag–Si) with conductivity g_{ij}, interconnecting two sets of metal bars (ith horizontal bar and jth in-plane bar). More than 8-bit write accuracy for isolated memristors have been proposed and demonstrated in literatures [3] and [4]. However, for TL design, the bit-precision requirement can be significantly less (less than 4 bit, shown in Fig. 13). The programming voltage applied across two cross-connected memristor, in a large-scale cross-bar array, results in sneak current paths through neighboring devices, which disturbs the state of unselected memristors. The application of access transistors [see Fig. 1(c)] and diodes can facilitate selective and disturb-free write operations to overcome the sneak path problem [24]. If the programming speed is not a major concern, the technique that can only program a single device at one time is also proposed in the literature [25] without access transistors or diode.

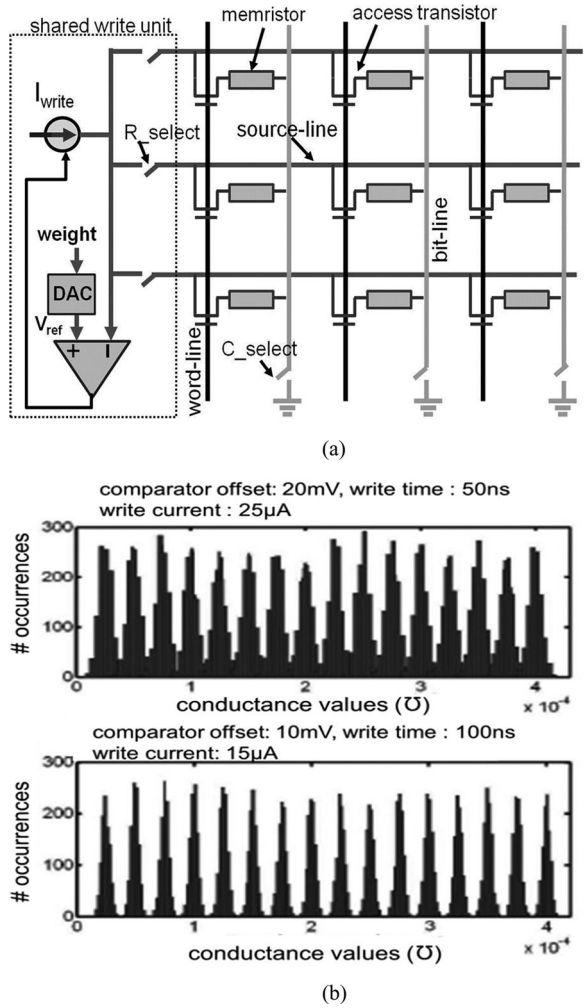


Fig. 2. (a) Resistive memory array with multilevel programming periphery. (b) Simulation results for feedback-based write show that higher write precision can be obtained by employing higher resolution comparator and longer write-time. These trends have been obtained using analytical model for memristors [4].

A multi-bit memristor programming array-level scheme employing adjustable pulse width is shown in Fig. 2(a) [4], [33]. In this scheme, when programming one specific memristor cell in the array, the corresponding set of the word line, the source-line, and the bit line will be selected. In Fig. 2(a), only a single write unit is shared among all of the rows for infrequent write operations, while a dedicated programming cell can be assigned to each row for maximum write speed. This would allow writing of one column at a time, by selecting a particular world line. During the writing operation, a constant current will be injected into the selected cell and the voltage developed on the source line is compared with a comparator threshold. A digital-to-analog converter (DAC) is used to set the threshold proportional to the target resistance. As soon as the accessed memristor is programmed as

the target value, the current source is disconnected. More precise tuning of memristor value can be achieved by applying a lower value of write current resulting in slower ramp in the resistance value, as shown in Fig. 1(d). The write precision in the method described earlier is mainly limited by the random offset of the comparator, inaccuracy in the current source and DAC. Larger accuracy would entail higher design complexity for these blocks and lower write speed [see Fig. 2(b)] [33]. The analytical model for memristor has been used for simulation in this study [4]. The memristive devices (including Ag–Si) do exhibit a finite-write threshold for an applied current/voltage, below which there is negligible change resistance [26]. Since the application of spin-based current comparator facilitates ultralow voltage (and hence low current) operation of the memristors for computing as will be described in the following sections, the state of memristor in the MCA will not be disturbed for reading.

B. TL Computation Using the MCA

For a TLG, the scaling and summation operations can be implemented using an MCA, as shown in Fig. 1(b). If we assume that the outward terminals of the in-plane bars are connected to ground potential, for a given set of binary voltage inputs V_i, the resulting current flowing out of the in-plane bars can be visualized as the dot product of the inputs voltages and the conductance values [3], [10].

The previous principle can be exploited in realizing current mode analog scaling (multiplication) and summation that corresponds to the first stage operation of a TLG. Several authors have proposed the design of hybrid TLG hardware based on MCAs and analog CMOS circuits, where analog circuits are employed to perform the second stage operation of the TLG, namely, thresholding [7]–[9]. For instance, application of analog current mirrors has been proposed for implementing memristor-based hybrid TLGs in [7]. However, such a design requires additional interconnect networks to realize fully programmable logic modules. Notably, energy consumption of interconnects dominates the total power budget of an FPGA [19]. Authors in [8] and [9] applied CMOS voltage comparators for realizing the thresholding operation for memristor-based TLGs. Application of analog amplifiers and comparators may lead to significant energy consumption. Authors in [10] recently demonstrated the use of a simple CMOS latch for thresholding operation. Such a scheme would need large voltage inputs (resulting in large current) to the memristors, so that a digital latch can directly sense the voltage mode output of a TLG. This would result in power hungry TLG blocks that may not be suitable for large-scale integration.

Thus, although memristors can provide an efficient mapping of the first stage operation of a TLG (namely current-mode scaling/multiplication and summation), the second operation, namely, the current mode thresholding, does not have a likewise "matching" device. The aforementioned inefficiencies could be eliminated if an alternate device structure could be found that could perform the current mode thresholding operation in an energy efficient way. In the next section, we present a spin-torque-based device that can be ideally suitable for this purpose.

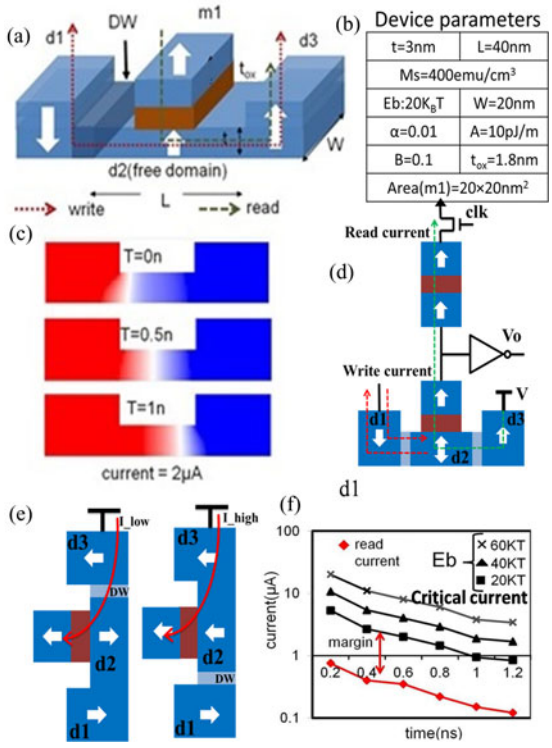


Fig. 3. (a) Device structure for STD. (b) Device parameters table. (c) Transient micromagnetic simulation plots. Red color represents the "down spin" corresponding to d1. Blue color represents the "up spin" in d3. White color shows the magnetic DW. (d) STD state sense circuit, (e) read current for different d2 states, and (f) read current margin to critical current.

III. TLG Second Stage Design Using STD

In this section, we present the STDs, based on magnetic DW, suitable for the design of energy efficient SMTL. This STD design will serve as the second stage of TL computing, which is a thresholding ("sign") function.

The device structure for the STD is shown in Fig. 3(a). It constitutes of a thin and short ($20 \times 40 \times 3$ nm^3) nanomagnet domain, d2 connecting two antiparallel nanomagnet domains of fixed polarity, d1 and d3. Domain-1 forms the input port, whereas domain-3 is grounded. Spin polarity of the free layer (d2) can be written parallel to d1 or d3 by injecting a small current along it from d1 to d3 and vice versa [15]–[17], [32]. A fixed polarity magnet m1 and d2 form a magnetic tunnel junction (MTJ) to read the state of d2. Thus, the STD can detect the polarity of the current flow at its input node. Note that STD acts as an ultralow voltage and compact current comparator that can be employed in the design of current mode TL.

The resolution of the device, i.e., the minimum current magnitude required to switch the free layer, is determined by the critical current density for DW motion. Several recent experiments have achieved subnanosecond DW motion, with current density of $\sim 10^6$ A/cm^2 [11]. Magnetic domain with perpendicular magnetic anisotropy can provide scaled device dimensions (thickness $\sim$3 nm and width $<$50 nm) as well as relatively lower critical current density [12]–[14]. More recently, application of spin orbital coupling has been explored for reducing the required current for a given speed of DW motion by an order of

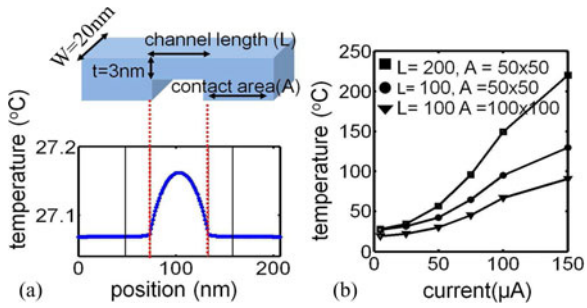


Fig. 4. (a) COMSOL simulation for temperature rise in the STD device for different device dimensions. (b) Plot showing temperature profile along the device for a small input current.

magnitude [14]. These device optimizations can be used to engineer current thresholds of the order of $\sim$2 μA for 1 ns switching. Fig. 3(c) shows the transient micromagnetic simulation plots for the proposed STD design using object-oriented micromagnetic framework [31] when supplied with a 2 μA current. It can be seen that the magnetic DW moves from the left free domain boundary to the right boundary within 1 ns. We will analyze the effect of STD resolution on the energy efficiency of SMTL in Section VII.

The effective resistance of the MTJ formed between m1 and d2 is smaller when they have the same spin polarity and vice versa [design parameter in Fig. 3(b)]. The ratio of the two resistances is defined in terms of tunnel magneto resistance ratio (TMR). STD forms a voltage divider with a fixed reference MTJ, as shown in Fig. 3(d). A TMR of $\sim$400% can provide a voltage swing close to $V_{\rm dd}$/2 that can be detected using a simple CMOS inverter. Static current in the voltage divider can be minimized for a given operation speed by increasing the MTJ oxide thickness. For 500-MHz clock frequency, the oxide thickness was determined to be $\sim$1.8 nm that resulted in a total power dissipation of $\sim$0.15 μW for the sensing unit (including the clocking power), for a supply voltage of 0.6 V.

Note that in the detection circuit, the terminal d3 of the STD is connected to $V_{\rm dd}$. Hence, the transient evaluation current flows from d3 to d2 as shown in Fig. 3(e). The current required for the DW motion increases proportional to the switching speed. Since the transient read current flows only for a short duration and the magnitude is lower than the critical current to move the DW, it does not disturb the state of d2. The red margin can be seen in Fig. 3(f). Apart from device scaling, the STD critical current can also be lowered by manipulating other device parameters, like the anisotropy energy (E_b) of the magnet [see Fig. 3(f)].

The reliability of a magnetic DW motion device is excellent. The velocity of DW and critical current are not sensitive to the external magnetic field or temperature [27]. 10-year retention time at 150 °C and 1×10^{14}-times write endurance for the Co/Ni wire are also reported in [27]. Another critical reliability assessment of the STD is the heating effect on the magneto-metallic STD. The effect of Joule heating in the STD was simulated using finite-element simulation through COMSOL [28]. The thin and short central free domain is the most critical portion with respect to current driven heating [see Fig. 4(a)]. Fig. 4(b) shows that larger contact area of the two fixed domains and shorter free domain can reduce the heating in the device. Thus, the current handling capacity of the device can be increased by appropriate structural optimization [33].

In general, Fig. 3(d) circuit forms the "sign" function in (1). The STD works as a current comparator and its input is the output current of the first stage MCA. If the input current to STD is larger than the critical current, the output of the inverter in Fig. 3(d) is high (vice versa). Next, we describe circuit design for combining the MCA and STD to implement TL array design.

IV. Design of SMTL Array

Fig. 5(a) and (b) shows two threshold logic networks (TLNs) for ISCAS85 benchmarks, C-17 and C-432 [7], obtained using the threshold logic synthesis (TELS) technique presented in [18]. The meaning of the symbols is explained in the caption. C-17 is a simple TLN, while C-432 is a much larger scale TLN. In order to show our design is compatible to large-scale TLN mapping, we will use C432 as a design example in the following paper.

TLN constitutes of a network of TLGs which can be divided into multiple stages. Each circle in the plot represents one TLG and the TLG in the same columns will be mapped to the same MCA stage. The connections between the TLGs are implemented by the MCA described in Section II, whereas the conductance of memristor corresponds to the synthesized weights. In such a multistage logic scheme, each MCA stage would comprise a number of TLGs receiving inputs from its previous stage and communicating their outputs to the next stage. Let us consider the design of such a stage using the MCA and the STD device.

Fig. 5(c) shows the circuit realization of a single MCA stage that contains N number of TLGs based on STD. Each stage has a maximum of M inputs (which can be set as a parameter during the implemented MCA mapping tool), and N STDs, forming the N TLGs. The ith input to the MCA may connect to the jth STD (i.e., jth TLG) with either a positive, negative or zero weight. This is achieved by programming either of $G_{ij}+$ or $G_{ij}-$ to the corresponding weight value. (The bias of each TLG can be viewed as the weight of an extra input whose value is always high.) For zero weight (i.e., no connectivity), both $G_{ij}+$ and $G_{ij}-$ are driven to high resistance off state. The input signal to the MCA is received through PMOS transistors with source terminals connected to a potential $V + \Delta V$ (for positive weights) and $V - \Delta V$ (for negative weights) where ΔV can be less than $\sim$50 mV. These input transistors act as deep triode region current sources (DTCS) [16]. The STD is connected to a dc supply V. This effectively clamps the potential of all the vertical metal bars in Fig. 5(c) to the same potential (due to small resistance of the magneto-metallic STD). The static current employed in computing therefore flows across a small terminal voltage of ΔV, resulting in small static power consumption. Moreover, the dynamic power dissipation on the metallic interconnects forming the programmable cross-bar is also largely reduced due to ultra-small voltage swing. The direction of current flow at the input of a STD, and hence the output of a TLG, would depend upon

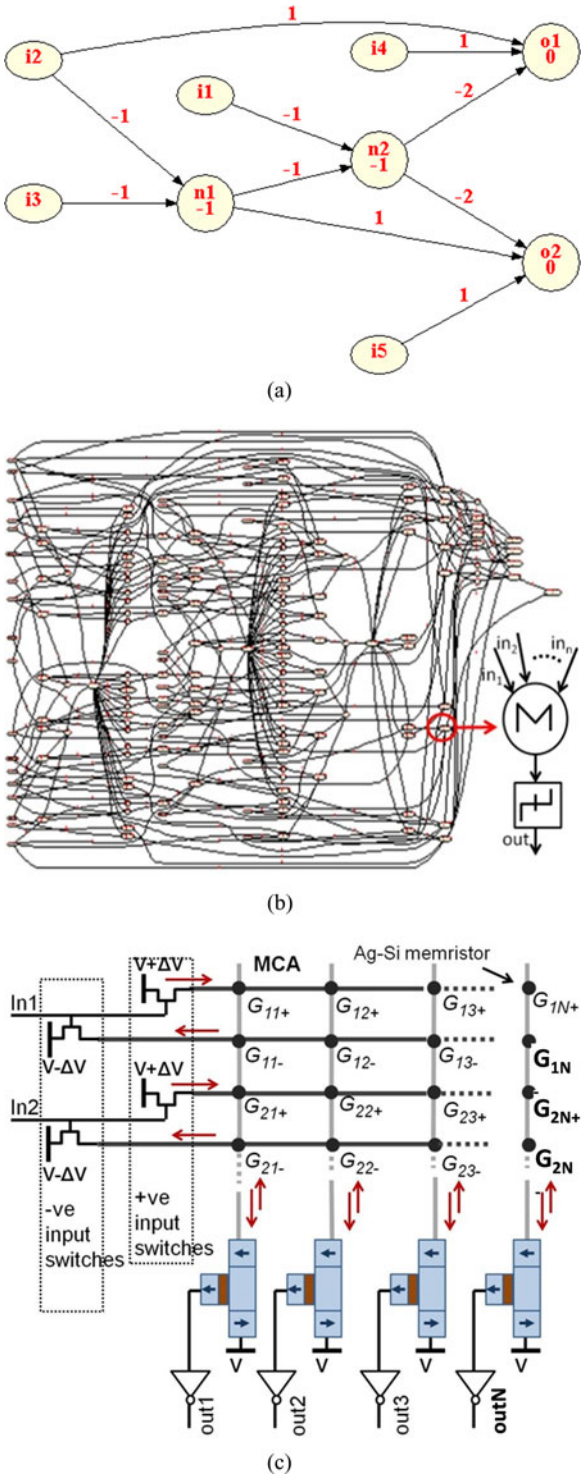


Fig. 5. (a) Synthesized ISCAS85 benchmark C17 TLN. Each circle represents one TLG. The connections between each TLG are the fan-ins and fan-outs. The node without fan-ins is the input node. The node without fan-outs is the output node. The weights are labeled along the connections. i_1–i_5 are the input nodes, n_1 and n_2 are TLGs, and o_1 and o_2 are the output nodes. The bias values are labeled inside the TLGs. The nodes in the same column are in the same stage. (b) Synthesized ISCAS85 benchmark-C432 (27-channel interrupt controller) TLN. The weight range is shown in Fig. 13. This synthesized TLN consists of 15 stages and 122 TLGs, while each stage is comprised of Ni TLGs. The maximum fun-in for each TLG is 4. (c) Circuit of one single TL stage using MCA and STD.

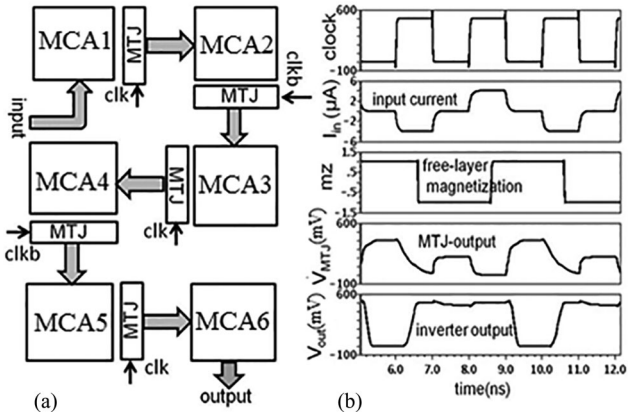


Fig. 6. (a) Two-phase pipelined MCA blocks for large-scale logic design, and (b) transient simulation plots for a single TLG.

the input data and the corresponding weights (determined by the programmed memristor conductance). Note that the resistance values for the memristors can be chosen large enough to avoid inaccuracy due to resistive voltage division between the DTCS transistors and the memristors in a given row. The output of the MTJ-based detection circuit associated with each TLG, in turn, drives a corresponding DTCS transistor that communicates the outputs of the TLGs to the next stage.

Due to the nonvolatility of the STD, the MCA design described earlier can be extended to realize a two-phase pipelined architecture composed of large number of such hybrid arrays without inserting the CMOS latches, as shown in Fig. 6(a). In such a design, consecutive MCAs operate with complementary clock phases. For instance, in Fig. 6(a), when the clock is high, MCA1 is driving MCA2, and MCA3 is driving MCA4. When the clock goes low, the driver and driven MCAs exchange roles. The exemplary simulation plots for a single TLG are shown in Fig. 6(b).

Next, we discuss an optimal pipelining and partitioning scheme for the mapping of large logic blocks on to the SMTL array.

V. Optimal Pipelining and Partitioning of SMTL Arrays for Logic Mapping

A. Pipeline Optimization

As mentioned earlier, each STD acts as a nonvolatile latch, and hence, a multistage MCA can be pipelined without insertion of additional CMOS latches. However, logic paths in the TLN of a generic logic block [like for C432 shown in Fig. 5(a)] may be unequal. Hence, "buffer-nodes" need to be inserted to make them equal and to facilitate fine-grained pipelining. The number of buffers needed depends upon the granularity of pipelining. In case, each MCA stage is pipelined, the number of buffers is the maximum. Fully pipelined TLN for C432 is shown in Fig. 7(a). In such a TLN, each stage is mapped into a separate MCA stage. For a given switching speed of the STD, this configuration yields maximum throughput. However, the total energy consumption also depends upon the total number of TLG nodes.

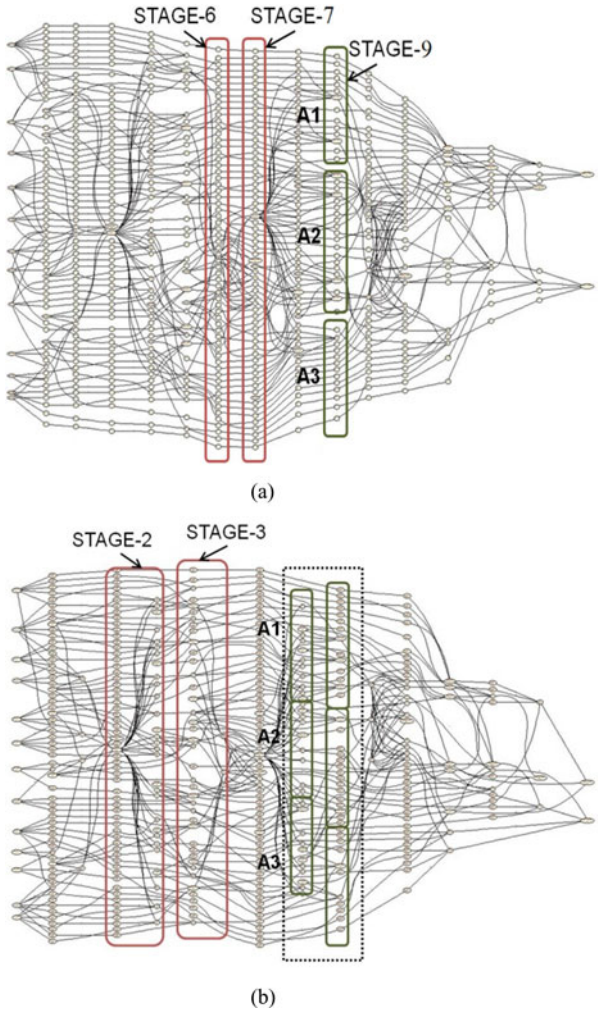


Fig. 7. Synthesized C432 pipelined TLN. (a) Fully pipelined architecture. (b) Two TLG stages combined with one pipeline stage. Each circle represents one TLG and the TLGs in the same column are in the same stage.

Combining two MCA stages to form a single pipelined stage [see Fig. 7(b)] reduces throughput by half; however, the total number of nodes for most benchmarks was found to reduce by a larger factor, which leads to reduced energy consumption. Note that despite using multiple MCA layers per pipeline stage, the same throughput can be maintained by increasing the current injection, i.e., the switching speed of the STD.

Fig. 8 shows the power consumption for C432 for different number of MCA levels (note, single MCA level for a pipelined stage implies maximum pipeline granularity) in a single pipelined stage. The power component due to the detection unit ("Power_det" due to MTJ voltage divider, clock, and inverter) reduces with reducing pipeline granularity, because of reduction in total number of TLG nodes in the resulting TLN [see Fig. 8(b)]. However, to maintain the same throughput, larger currents need to be supplied by the DTCS transistors, which lead to increase in static power consumption in the MCA ["Power_MCA" in Fig. 8(a)]. For most ISCAS85 benchmarks, a pipelined stage with two-MCA levels yielded optimal results [see Fig. 8(b)].

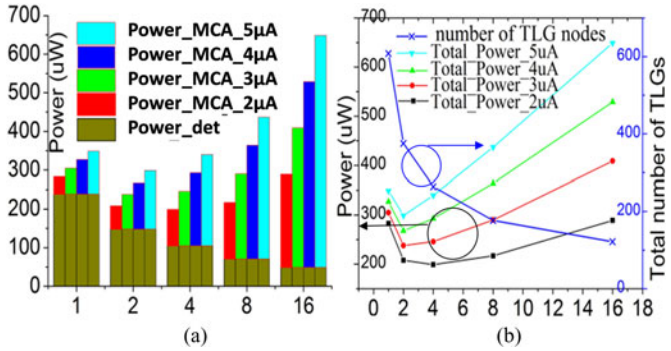


Fig. 8. (a) Power consumption of different pipeline configurations and (b) tradeoff between power and area. "Power_MCA_5 uA" represents the power of memristor cross-bar array when the DTCS current is 5 μA. "Power_det" is the power of detection module including MTJ-voltage divider, clock and inverter.

B. Partition and Interconnects

So far, we assumed that each stage of the pipelined TLN is assigned to a single large dimension MCA. In such a design, no additional interconnect network is required, as the outputs of the Nth MCA stage can directly connect to the inputs of the $(N+1)$th MCA stage using the scheme shown in Fig. 7. Due to the absence of additional interconnect power dissipation, this leads to the minimum energy solution [see Fig. 10(a)]. However, in this case, the MCAs have sparse connectivity (due to having large number of inputs but each input connecting to only few outputs, determined by the fan-in limitation) due to which the overall area efficiency is significantly sacrificed, as shown in Fig. 10(b). To reduce the overall area, each pipeline stage can be divided into multiple smaller dimension subarrays [A_is shown in Fig. 7(b) and an enlarged version in Fig. 9(a)]. In this case, some of the interlayer connections can still be directly routed to the next stage [see Fig. 9(a)]. However, some others (between nodes that are not located on directly opposite MCAs) need to be routed through an additional routing network. Such a design scheme is shown in Fig. 9(b). For reducing MCA dimensions (implying the use of large number of smaller MCA modules in a single stage), the usage of the interconnect network increases. This also necessitates larger and longer interconnect array, leading to larger parasitic resistance drops along the current signal paths, mandating the use of larger voltage. As a result, energy component due to interconnect increases. Fig. 10 shows the tradeoff between area and power of SMTL with respect to the size of the sub-MCA array size. A design choice can be made based on priority.

VI. Simulation and Synthesis Algorithm

In this section, we discuss the synthesis scheme used in this study to assess the performance of SMTL.

Fig. 11 shows the high-level overview of the SMTL synthesis and hardware mapping methodology employed in this study. We employed the TELS algorithm proposed in [18] to do the initial synthesis, which reads a logic description and generates the functionally equivalent threshold network. Some important

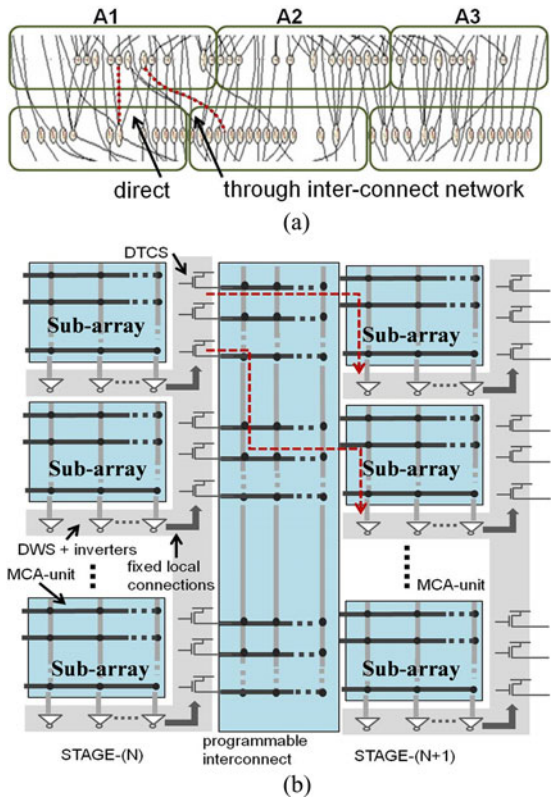


Fig. 9. (a) Enlarged green square part of Fig 7(b). (b) SMTL network partition architecture.

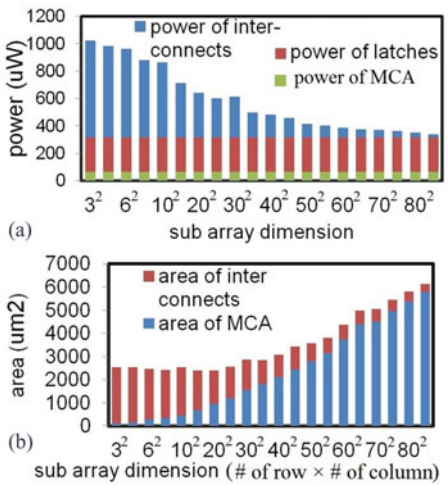


Fig. 10. Relationship between (a) power, (b) area and subarray dimension (larger dimension implies lower number of subarrays needed).

parameters like the fan-in restriction of TLGs and defect tolerance in the weights can be preset as parameters [18].

The SMTL mapping algorithm proposed and implemented in this paper, shown in Fig. 12, reads the synthesized TLG network and maps it to SMTL hardware. The tool first reorders the positions of TLGs in each stage so as to minimize the use of the interconnect network. This is achieved by placing the TLGs in the subarrays such that the use of direct links between face-to-face MCAs [as depicted in Fig. 9(a)] is maximized.

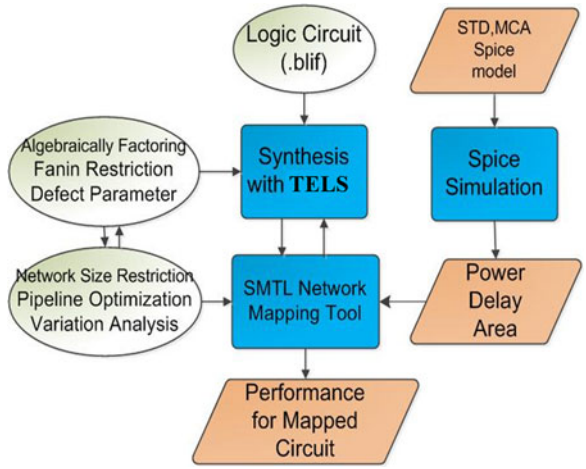


Fig. 11. Proposed design methodology.

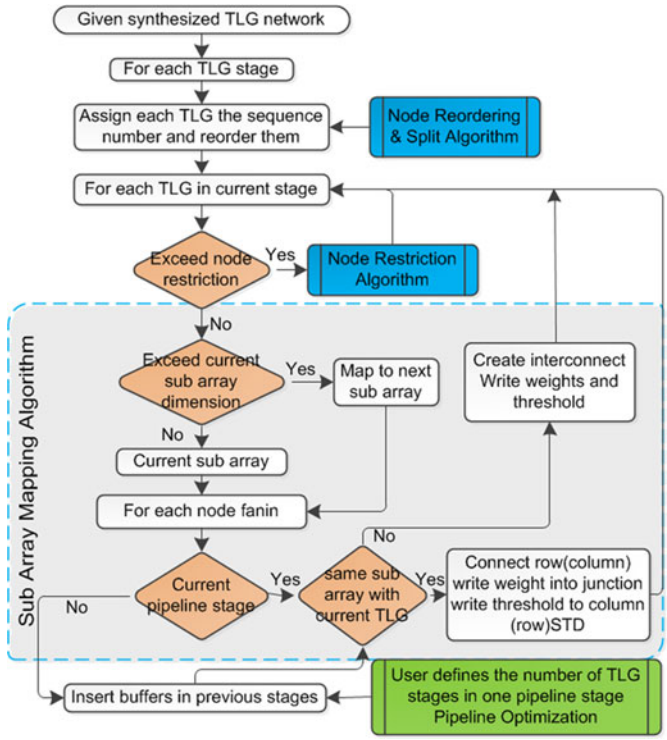


Fig. 12. SMTL network mapping algorithm.

Next, if the number of nodes in the current stage exceeds the restriction (number of MCA in a given stage times MCA size), one or more nodes are moved to next stage. This is done in a way that minimizes the number of intermediate buffers. The nodes without fan-out to next one stage are selected with highest priority, following which, the nodes with minimum fan-ins are shifted.

Some of the layers in the SMTL netlist may have very small number of nodes, for which, the use of a separate MCA unit may be wasteful. In TELS, such nodes are incorporated in the MCA units corresponding to the previous stage, through the provision of a small numbers of programmable backward connections (from output of an MCA back to its input).

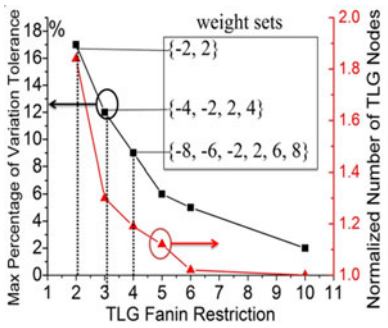


Fig. 13. Relationship between variation tolerance, TLG fan-in restriction and number of TLGs.

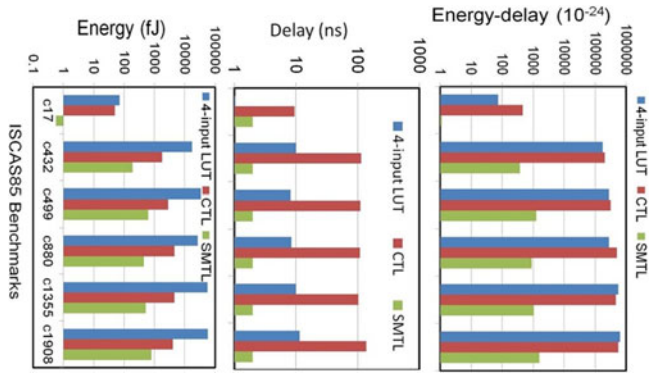


Fig. 14. (a) Computation energy, (b) delay and (c) energy-delay product of SMTL compared with four-input LUT-based FPGA [7] and CTL [7] for ISCAS85 benchmarks. (CTL: a CMOS-based implementation style for TLG [7].)

The fan-out number of some nodes can be very large. Such TLGs communicate evenly to all the MCAs in the next level, making heavy use of the interconnect network. Such high loading can lead to significant voltage division between the DTCS source and the receiving memristors, leading to significant lowering of the input voltage and the current for the loads. A simple way to address this issue is to split the large fan-out nodes into multiple smaller nodes.

Larger TLG fan-in generates denser SMTL network with smaller number of TLG nodes. This can provide larger area and energy efficiency. However, simulations show that larger fan-in restriction leads to reduced variation tolerance for memristor values, as seen in Fig. 13(b). In this plot, variation tolerance is defined as the standard-deviation (σ) value for which total 10^5 test vector simulation gave zero errors. The variation tolerance increases for lower fan-in restriction, but the use of lower fan-in TLGs results in larger number of nodes, leading to increase in overall area [see Fig. 13]. In this study, we choose the fan-in restriction to be 4 [leading to a variation tolerance of $\sim$9%, as shown in Fig. 13]. There are only six different levels of memristor conductance needed for mapping the TLG weighs; therefore, the programming bit resolution for memristor is 3 bit. Note that in this paper, we have assumed that the memristor programming thresholds are large enough, such that passing small computing currents (few μA) does not significantly disturb their state [2].

Next, we discuss the performance of SMTL and compare it with conventional CMOS programmable logic based on CMOS LUTs.

VII. Performance and Prospects

In the conventional FPGA-based TLG design, the interconnect power is the bottleneck of the total power consumption. Note that more than 90% of energy can be ascribed to programmable switches and interconnects [19]. The reason is the fact that the FPGA interconnect circuit has an extremely low utilization rate ($\sim$12%) for purpose of programmability. The energy and delay of four-input LUT-based FPGA for ISCAS85 benchmark using 45 nm technology is shown in Fig. 14. While in our proposed SMTL design, the energy efficiency mainly comes from four aspects: 1) The interconnect energy dissipation in the metallic cross-bars as well as the interconnect network is drastically lowered due to ultralow voltage ($\sim$50 mV), current mode signaling between the MCA layers, which comes from low voltage, low current operation of spin-torque-based TLGs. The STD device can sense and compare the ultralow current (few μA) enabling ultralow voltage biasing of the MCA and hence, low voltage operation of the threshold gates. As a result, the static power consumption, due to direct current paths, is largely reduced. Note that in the SMTL design, memristors play the dual role of computing elements as well as programmable interconnects. This can be contrasted with earlier approaches where memristors were employed only as programmable interconnects [21] or only as computing elements [7]. 2) In our proposed TLN design, the output inverters of a particular MCA layer drives only the DTCS transistors that in turn supply current to the next MCA stage. The MCA itself is operated across a small terminal voltage ΔV, thereby reducing the CV^2f dynamic power consumption in large number of programmable interconnects. Such low voltage operation of the MCA can also significantly reduce the disturb rate of the programmed memristors and can enhance the retention time of the hardware. 3) The STD achieves energy efficient current to voltage conversion with the help of MTJ-based voltage divider. This eliminates the need of analog transimpedance circuits based on current mirrors and amplifier, leading to high energy and area efficiency. 4) Due to the nonvolatility of STD, the proposed SMTL design can be extended to realize a pipelined architecture without inserting the CMOS latches. The throughput of the design is determined by a single stage delay. This delay in turn is limited by the switching speed of the STD device. As discussed earlier, larger current per input can be used to increase the STD switching speed. DW velocities of more than 400 m/s have been demonstrated in the literature [22]; hence, for a 40 nm long free domain, more than 1 GHz processing speed may be achievable. In this study, a clock frequency of 500 MHz has been used, corresponding to STD switching time of 1 ns. Recently application of spin-hall effect has been explored for bringing large reduction in DW current thresholds [13]. Such phenomena can be exploited in improving the resolution of scaled STD devices.

Fig. 14(a) compares the computation energy of the proposed SMTL design with that of four-input LUT-based FPGA and with capacitive threshold logic (CTL, a CMOS-based

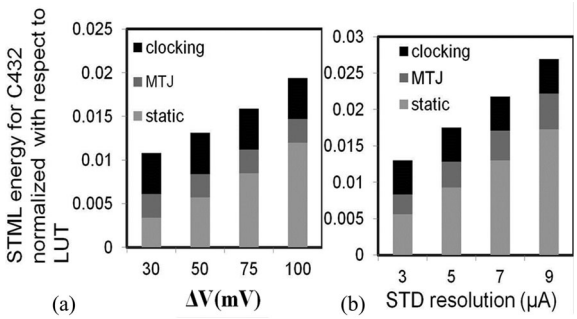


Fig. 15. SMTL energy for C432 normalized with respect to four-input LUT for the case of (a) increasing ΔV, (b) increasing STD threshold for a fixed ΔV of 50 mV; LUT delay is $\sim$10 ns.

TABLE I
DESIGN PARAMETERS

Free domain size	3×20×40nm^3	MTJ t_{ox}	1.8nm
Ms	400emu/cm^3	R_{MTJ} (parallel)	300K Ω
Ku_2V	$20K_BT$	MTJ TMR	400%
β (nonadiabatic const.)	0.1	MTJ area (nm^2)	20×20
α (damping coeff.)	0.01	Memristor values (Ω)	50K to 1M
$I_{threshold}$ for STD	2μA	ΔV	50mV
V	0.6V	CMOS tech.	45nm

implementation style for TLG [7]). It shows about two orders of magnitude lower computing energy for the proposed design as compared to the LUT-based FPGA TLG. SMTL also shows much smaller delay compared with LUT and CTL, as shown in Fig. 14(b). Results in Fig. 14(c) show around three orders of magnitude lower energy delay product as compared to both the CMOS-based schemes.

The energy efficiency of the proposed design is dependent on two critical design parameters. First is the minimum achievable ΔV (voltage swing across MCA) in such a hybrid circuit. Fig. 15(a) shows that increasing ΔV increases the static power consumption due to current mode computing in MCAs (strength of DTCS transistors is reduced to keep the current drive constant). The second important parameter is the resolution of the STD device. As mentioned earlier, a poor resolution would require larger current per-input for a TLG. Corresponding results are shown in Fig. 15(b), showing almost linear increase in computation energy with reducing resolution.

Integration of Ag–Si memristors with CMOS has been demonstrated in recent years [2], [3]. The same is true with magnetic DW-based memory cells [13], [23], [27]. However, integrating two novel technologies with CMOS to realize the proposed SMTL scheme can be significantly more challenging, especially when scaled dimensions of STD devices, such as used in this study, is targeted. However, the possibility of large energy benefits of the proposed design can be a motivating factor.

Some critical design parameters used in this study are given in Table I. The device characteristics for STD were obtained using the micromagnetic simulation framework for DW magnet presented in [23]. Behavioral model based on statistical characteristics of the device was used in SPICE simulation to assess the system-level functionality.

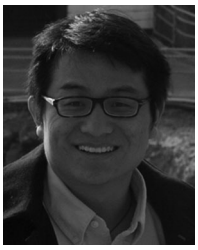
VIII. CONCLUSION

STD can be combined with CMOS compatible Ag–Si memristors for designing ultralow energy SMTL. Such hardware can achieve more than 100 × improvement in energy and 1000 × improvement in energy-delay product, as compared to state-of-the-art CMOS FPGA-based TLG, due to low voltage, low current computing facilitated by a spin-torque device.

REFERENCES

[1] S. H. Jo, K. H. Kim, and W. Lu, "High-density cross-bar arrays based on a SI memristive system," *Nano Lett.*, vol. 9, no. 2, pp. 870–874, 2009.
[2] S. H. Jo and W. Lu, "CMOS compatible nanoscale nonvolatile resistance switching memory," *Nano Lett.*, vol. 8, no. 2, pp. 392–397, 2008.
[3] L. Gao, F. Alibart, and D. B. Strukov, "Analog-input analog-weight dot-product operation with Ag/a-Si/Pt memristive devices," in *Proc. VLSISOC*, 2012, pp. 88–93.
[4] S. Shin, K. Kim, and S. M. Kang, "Memristor applications for programmable analog ICs," *IEEE Trans. Nanotechnol.*, vol. 10, no. 2, pp. 266–274, Mar. 2011.
[5] K. Likharev, "Biologically inspired computing in CMOL crossnets," in *Proc. AAAI Fall Symp.: Biol. Inspired Cogn. Architectures*, 2009, pp. 90.
[6] J. Turel, J. H. Lee, X. Ma, and K. K. Likharev, "Neuromorphic architectures for nanoelectronic circuits," *Int. J. Circ. Theor. Appl.*, vol. 32, no. 5, pp. 277–302, 2004.
[7] J. Rajendra, H. Manem, R. Karri, and G. S. Rose, "An energy-efficient memristive threshold logic circuit," *IEEE Trans. Comput.*, vol. 61, no. 4, pp. 474–487, Apr. 2012.
[8] T. Tran, A. Rothenbuhler, E. Smith, V. Saxena, and K. A. Campbell, "Reconfigurable threshold logic gates using memristive devices," in *Proc. IEEE Subthreshold Microelectron. Conf.*, 2012, pp. 1–3.
[9] D. Chabi, W. Zhao, D. Querlioz, and J. O. Klein, "Robust neural logic block (NLB) based on memristor cross-bar array," in *Proc. IEEE/ACM Int. Symp. Nanoscale Architectures*, 2011, pp. 137–143.
[10] L. Gao, F. Alibart, and D. B. Strukov, "Programmable CMOS/memristor threshold logic," *IEEE Trans. Nanotechnol.*, vol. 12, no. 2, pp. 115–119, Mar. 2012.
[11] C. K. Lim, T. Devolder, C. Chappert, J. Grollier, V. Cros, A. Vaures, A Fert, and G. Faini, "Domain wall displacement induced by subnanosecond pulsed current," *Appl. Phys. Lett.*, vol. 84, no. 15, pp. 2820–2822, 2004.
[12] D-T. Ngo, N. Watanabe, and H. Awano, "CoB/Ni-based multilayer nanowire with high-speed domain wall motion under low current control," *Jpn. J. Appl. Phys.*, vol. 51, 09302 (4 pp.), 2012.
[13] S. Fukami, T. Suzuki, K. Nagahara, N. Ohshima, Y. Ozaki, S. Saito, R. Nebashi *et al.*, "Low-current perpendicular domain wall motion cell for scalable high-speed MRAM," in *Proc. VLSI Technol. Symp.*, 2009, pp. 230–231.
[14] A. V. Khvalkovskiy, V. Cros, D. Apalkov, V. Nikitin, M. Krounbi, K. A. Zvezdin, A. Anane, J. Grollier, and A. Fert, "Matching domain-wall configuration and spin-orbit torques for efficient domain-wall motion," *Phys. Rev. B*, vol. 87, 020402 (5 pp.), 2013.
[15] D. Morris, D. Bromberg, J. G. Zhu, and L. Pileggi, "mLogic: Ultra-low voltage non-volatile logic circuits using STT-MTJ devices," in *Proc. Design Autom. Conf.*, 2012, pp. 486–491.
[16] M. Sharad, C. Augustine, and K. Roy, "Boolean and non-Boolean computation with spin devices," in *Proc. IEEE Int. Electron Devices Meeting*, 2012, pp. 11–16.
[17] M. Sharad, D. Fan, and K. Roy, "Ultra low power associative computing with spin neurons and resistive crossbar memory," in *Proc. ACM Design Autom. Conf.*, 2013, vol. 107, pp. 107-1–107-6.
[18] R. Zhang, P. Gupta, L. Zhong, and N. K. Jha, "Synthesis and optimization of threshold logic networks with application to nanotechnologies," in *Proc. Design Autom. Test Eur. Conf. Exhib.*, 2004, vol. 2, pp. 904–909.
[19] Y. Lin, F. Li, and L. He, "Routing track duplication with fine-grained power-gating for FPGA interconnect power reduction," in *Proc. Asia South Pacific Design Autom. Conf.*, 2005, vol. 1, pp. 645–650.
[20] J. A. Currivan, Y. Jang, M. D. Mascaro, M. A. Baldo, and C. A. Ross, "Low energy magnetic domain wall logic in short, narrow, ferromagnetic wires," *IEEE Magn. Lett.*, vol. 3, 3000104 (4 pp.), 2012.
[21] Q. Xia, W. Robinett, M. W. Cumbie, N. Banerjee, T. J. Cardinali, J. J. Yang, W. Wu, X. Li, W. M. Tong, D. B. Strukov, G. S. Snider, G. M.

Ribeiro, and R. S. Williams, "Memristor-CMOS hybrid integrated circuits for reconfigurable logic," *Nano Lett.*, vol. 9, no. 10, pp. 3640–3645, 2009.
[22] P. J. Metaxas *et al.*, "High domain wall velocities via spin transfer torque using vertical current injection," *Sci. Rep.*, vol. 3, 1829 (6 pp.), 2013.
[23] C. Augustine, A. Raychowdhury, B. Behin-Aein, S. Srinivasan, J. Tschanz, V. De, and K. Roy, "Numerical analysis of domain wall propagation for dense memory arrays," in *Proc. IEEE Int. Electron Devices Meeting*, 2011, pp. 17.6.1–17.6.4.
[24] H. Manem and G. S. Rose, "A read-monitored write circuit for 1T1M multi-level memristor memories," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2012, pp. 2938–2941.
[25] C. Jung, J. Choi, and K. Min, "Two-step write scheme for reducing sneak-path leakage in complementary memristor array," *IEEE Trans. Nanotechnol.*, vol. 1, no. 3, pp. 611–618, May 2012.
[26] K. Kim, S. Gaba, D. Wheeler, J. M. Cruz-Albrecht, T. Hussain, N. Srinivasa, and W. Lu, "A functional hybrid memristor crossbar-array/CMOS system for data storage and neuromorphic applications," *Nano Lett.*, vol. 12, no. 1, pp. 389–395, 2011.
[27] S. Fukami, M. Yamanouchi, H. Ohno *et al.*, "High-speed and reliable domain wall motion devices: Material design for embedded memory and logic application," in *Proc. Symp. VLSI Technol.*, 2012, pp. 61–62.
[28] G. Zhang, Q. Zhang, C. T. Bui, G. Q. Lo, and B. Li, "Thermoelectric performance of silicon nanowires," *Appl. Phys. Lett.*, vol. 94, 213108 (3 pp.), 2009.
[29] W. S. Zhao, S. Chaudhuri, C. Accoto, J. Klein, C. Chappert, and P. Mazoyer, "Cross-point architecture for spin transfer torque magnetic random access memory," *IEEE Trans. Nanotechnol.*, vol. 11, no. 5, pp. 907–917, Sep. 2012.
[30] X. F. Yao, J. Harms, A. Lyle, Y. Zhang, and J. P. Wang, "Magnetic tunnel junction-based spintronic logic units operated by spin transfer torque," *IEEE Trans. Nanotechnol.*, vol. 11, no. 1, pp. 120–126, Jan. 2012.
[31] OOMMF [Online]. Available: http://math.nist.gov/oommf/
[32] M. Sharad, D. Fan, and K. Roy, "Ultra-low energy, high performance and programmable magnetic threshold logic," arXiv:1308.4169 [cs.ET] 2013.
[33] M. Sharad, D. Fan, K. Aitken, and K. Roy, "Energy-efficient non-Boolean computing with spin neurons and resistive memory," *IEEE Trans. Nanotechnol.*, vol. 13, no. 1, pp. 23–34, Jan. 2014.



Deliang Fan received the B.S. degree in electronic information engineering from Zhejiang University, Hangzhou, China, in 2010 and the M.S. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2012. He is currently a Graduate Research Assistant of Professor Kaushik Roy and is working toward the Ph.D. degree in electrical and computer engineering at Purdue University.

His primary research interest lies in cross-layer (algorithm/architecture/circuit) codesign for low-power Boolean, non-Boolean, and neuromorphic computation using emerging technologies like spin transfer torque devices. His past research interests include cross-layer digital system optimization and imperfection-resilient scalable digital signal processing algorithms and architectures using significance driven computation.

Mrigank Sharad received the B.Tech. and M.Tech. degrees in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, India, in 2010, where he specialized in microelectronics and very large scale integration design and is currently working toward the Ph.D. degree in electrical and computer engineering at Purdue University, West Lafayette, IN, USA.

He was a Research Intern with IBM T. J. Watson Research Centre in the summer of 2012. He has authored more than 30 papers in refereed journals and conferences. His primary research interests include low-power digital/mixed-signal circuit design. His current research is focused on device-circuit codesign for low-power logic and memory, with emphasis on exploration of post-CMOS technologies like spin-devices. He has also worked on application of spin-torque devices in approximate computing hardware, interconnect and memory design.

Mr. Sharad received the Prime Minister of India Gold Medal for his academic performance at IIT Kharagpur. He received Andrews Fellowship from Purdue University in 2010.

Kaushik Roy (F'02) received the B.Tech. degree in electronics and electrical communications engineering from the Indian Institute of Technology (IIT), Kharagpur, India, and the Ph.D. degree from the Electrical and Computer Engineering Department, University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 1990.

He was with the Semiconductor Process and Design Center of Texas Instruments, Dallas, where he worked on field-programmable gate array architecture development and low-power circuit design. He joined the electrical and computer engineering faculty at Purdue University, West Lafayette, IN, USA, in 1993, where he is currently Edward G. Tiedemann Jr. Distinguished Professor. His research interests include spintronics, device-circuit codesign for nanoscale silicon and non-silicon technologies, low-power electronics for portable computing and wireless communications, and new computing models enabled by emerging technologies. He has published more than 600 papers in refereed journals and conferences, holds 15 patents, graduated 60 Ph.D. students, and is coauthor of two books on Low Power CMOS VLSI Design (John Wiley and McGraw Hill).

Dr. Roy received the National Science Foundation Career Development Award in 1995, IBM Faculty Partnership Award, ATT/Lucent Foundation Award, 2005 SRC Technical Excellence Award, SRC Inventors Award, Purdue College of Engineering Research Excellence Award, Humboldt Research Award in 2010, 2010 IEEE Circuits and Systems Society Technical Achievement Award, Distinguished Alumnus Award from the IIT, Kharagpur, Fulbright-Nehru Distinguished Chair, and best paper awards at 1997 International Test Conference, IEEE 2000 International Symposium on Quality of IC Design, 2003 IEEE Latin American Test Workshop, 2003 IEEE Nano, 2004 IEEE International Conference on Computer Design, 2006 IEEE/ACM International Symposium on Low Power Electronics & Design, and 2005 IEEE Circuits and system society Outstanding Young Author Award (Chris Kim), 2006 IEEE Transactions on VLSI Systems best paper award, 2012 ACM/IEEE International Symposium on Low Power Electronics and Design best paper award, 2013 IEEE Transactions on VLSI Best paper award. He was a Purdue University Faculty Scholar (1998–2003). He was a Research Visionary Board Member of Motorola Labs (2002) and held the M.K. Gandhi Distinguished Visiting faculty at IIT, Bombay. He has been in the editorial board of IEEE Design and Test, IEEE Transactions on Circuits and Systems, IEEE Transactions on VLSI Systems, and IEEE Transactions on Electron Devices. He was Guest Editor for Special Issue on Low-Power VLSI in the IEEE Design and Test (1994) and IEEE Transactions on VLSI Systems (June 2000), *IEE Proceedings–Computers and Digital Techniques* (July 2002), and IEEE Journal on Emerging and Selected Topics in Circuits and Systems (2011).