

An Expanded Benchmarking of Beyond-CMOS Devices Based on Boolean and Neuromorphic Representative Circuits

CHENYUN PAN (Member, IEEE) , and AZAD NAEEMI (Senior Member, IEEE)

School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA.

CORRESPONDING AUTHOR: C. Pan (chenyun.pan@gatech.edu).

This work was supported by the Semiconductor Research Corporation under Grant NRI Theme 2624.001. This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the author.

ABSTRACT The latest results of benchmarking research are presented for a variety of beyond-CMOS charge- and spin-based devices. In addition to improving the device-level models, several new device proposals and a few majorly modified devices are investigated. Deep pipelining circuits are employed to boost the throughput of low-power devices. Furthermore, the benchmarking methodology is extended to interconnect-centric analyses and non-Boolean logic applications. In contrast to Boolean circuits, non-Boolean circuits based on the cellular neural network demonstrate that spintronic devices can potentially outperform conventional CMOS devices.

INDEX TERMS Beyond-CMOS technology, cellular neural network (CNN), domain wall motion, ferroelectric FET, interconnect, magnetoelectric, non-Boolean computing, spin diffusion, spin hall effect (SHE), spintronics, throughput, tunneling FET (TFET).

I. INTRODUCTION

Faced with the challenges and limitations of CMOS scaling, there is a global search for beyond-CMOS device technologies that are capable of augmenting or even replacing conventional Si CMOS technology and sustaining Moore's law [1]–[4]. There is an increasing need for a uniform benchmarking methodology to capture and evaluate the latest research and development for various beyond-CMOS proposals. Such research is critical in identifying the key limiting factors for promising devices and in guiding future research directions through modification or even reinvention of the proposed devices.

Beyond-CMOS benchmarking (BCB) efforts have continued for several years with three major releases. The first one originating in 2010, BCB 1.0, was led by Bernstein *et al.* [2] and used unmodified device inputs from various NRI groups. It was followed by two sequential uniform benchmarking works led by Nikonov and Young [3], [4], BCB 2.0 and 3.0. They treated a broader range of devices and circuits using a consistent, transparent, and physics-based methodology [4].

In this paper, we have added two recently proposed voltage-controlled spintronic devices, magnetoelectric magnetic tunneling junction (MEMTJ), and composite-input

magnetoelectric-based logic technology (CoMET). More elaborate device-level modeling approaches are applied to many spintronic devices. In addition, major modifications for several spintronic devices have been proposed and evaluated. Major updates have also been applied for charge-based FETs to reflect the latest research developments in the past two years, i.e., tunneling FET (TFET), ferroelectric-based FET, graphene p-n-junction (GPNJ) device, 2-D material-based devices, and negative differential resistance (NDR) devices.

At the circuit level, the arithmetic logic unit (ALU) circuit is adopted from the previous Boolean logic benchmarking [4]. We include the deep pipelining analysis to take advantage of the inherent memory feature of some of the low-power devices and the supply clocking that has to be used to eliminate standby power dissipation in current-driven devices. Since this approach is somewhat similar to dynamic logic, CMOS implementation of dynamic logic has been added to the reference benchmarking data points. To account for the fact that interconnects pose major limitations on the state-of-the-art VLSI systems [5], [6], interconnect-centric performance benchmarking is also included in this paper. It covers multiple key interconnect metrics, such as the optimal delay and energy of a long interconnect with repeaters

and the span of control. We then explore the benefits and limitations of emerging charge- and spin-based technologies from the perspective of interconnect design. Following BCB 3.0, we choose a simple analytical approach to capture the key advantages, challenges, and limitations of various emerging technologies.

Previous benchmarking results for a 32-bit ALU have shown that only a few devices can potentially outperform CMOS in terms of energy-delay product (EDP) [4]. Most devices are worse in terms of delay and energy per ALU operation, especially for spintronic devices, where orders of magnitude larger EDPs are projected. Therefore, it is crucial to search for nontraditional circuits where beyond-CMOS devices can realize their full potential. During the past few years, research has shown that alternative computing paradigms, such as non-Boolean circuits and systems, are potentially capable of taking advantage of the unique physical properties of novel devices [7], [8]. In this paper, we choose the cellular neural network (CNN) as the benchmarking circuit because of the following:

- 1) It performs many tasks in the areas of sound, image, and video processing quite efficiently [9], [10].
- 2) It has a well-established theory [11].
- 3) It can be implemented by a wide range of emerging technologies for both charge- and spin-based devices [9], [10], [12].

Furthermore, a recent work has shown that CNN can be efficiently used to create a convolutional neural network that is widely used in deep-learning applications [13]. In this paper, we investigate three types of CNN implementations, i.e., analog, digital, and spintronic circuits, covering a broad range of charge- and spin-based beyond-CMOS devices. All results shown in this paper are reproducible, and the MATLAB code is publicly available [14]. The user has the flexibility of adding new devices and performing benchmarking against other beyond-CMOS technologies for a variety of applications.

The rest of this paper is organized as follows. Section II introduces several new devices and latest device-level models to be covered in this new release of benchmarking. Section III shows the circuit-level benchmarking methodology for both Boolean and non-Boolean computing. Benchmarking results and discussions are presented in Section IV. Finally, conclusions are made in Section V.

II. DEVICE-LEVEL MODELS

A. CHARGE-BASED DEVICES

1) TUNNELING FETs

The intrinsic capacitance and on current of several TFETs have been modified significantly for 2-D heterojunction interlayer TFET (ThinTFET), gallium nitride TFET (GaNTFET), and transition metal dichalcogenide TFET (TMDTFET) [15]–[17] (Fig. 1 in supplementary material). These modifications are based on atomistic simulations performed at the Low Energy Systems Technology (LEAST) center, a research center sponsored by SRC and DARPA through

STARnet [18]. Substantial performance improvements are observed for the following reasons. The wave function coupling between the two 2-D materials in ThinTFET has been adjusted to the best-known value. An updated simulation of the charge in the device has changed the gate capacitance. The 2015 NEMO simulation [19] of the inline GaNTFET at 0.4 V has resulted in a much larger saturation current than the comparable TCAD simulation. In addition to improvements in the atomistic simulations using NEMO, there have been changes in materials, structures, stress, and doping of TMDTFET. The simulations have been performed for 15 device options, and the best ones are selected in this paper [18]. The baseline TFET devices, i.e., the homogeneous and heterogeneous TFETs, are taken from BCB 3.0 [4].

2) FERROELECTRIC-BASED FETs

Three ferroelectric-based FETs from BCB 3.0 are included in this paper: negative-capacitance FET (NCFET), metal-insulator transition FET, and ferroelectric FET [4], [20], [21]. The updated IV characteristics for the NCFET are obtained from the LEAST center [22], [23]. Following [4], a partial polarization intrinsic switching time of 10 ps is added on top of the intrinsic switching delay of NCFET. One change made in this benchmarking release is that only one ferroelectric switching delay is added on a different logic function, such as an inverter and nand gate, assuming the ferroelectrics on the gates of nMOS and pMOS transistors are switched in parallel.

3) OTHER CHARGE-BASED DEVICES

There is a major update on the modeling approach of the GPNJ devices. Instead of the analytical angular dependent transmission probability model used in previous benchmarking, a more realistic model based on ray-tracing approach is employed [24]. The results are more accurate and consistent with rigorous NEGF simulations. The on-off ratio of the device predicted by the new model degrades by 10 \times even at a large device width because of multiple reflections of electron beams at graphene edges and junctions. Two configurations of GPNJ devices are investigated with different input back-gates.

Two NDR devices, bilayer pseudospin FET (BisFET) and interlayer TFET (ITFET), from BCB 3.0 are included in this paper [4], [25]. After voltage signals reach the input of logic gate, complementary supply voltages are applied to perform the logic computation and lock the output according to the inputs. The supply voltage needs to be held in order to lock the output voltage. Meanwhile, the logic device consumes static power until the supply voltage returns to zero. More detailed device- and circuit-level operations are illustrated in the supplementary material. Note that the exciton condensation in bilayer graphene in BisFET has not been observed in experiments yet. The benchmarking results are based on the theoretical modeling. The 2-D material-based van der Waals FET device is updated based on a new channel material, black phosphorus, which has a large field-effect mobility and highly anisotropic band structure [26].

B. SPINTRONIC DEVICES

Another category of devices is the spintronic devices. These devices are promising candidates to complement conventional CMOS devices as they provide new features, such as nonvolatility and low-voltage operation [27]. One set of the spintronic devices are current driven, and some of the well-studied device concepts in this category include all-spin logic (ASL) [28], charge-coupled spin logic (CSL) [29], and domain wall logic (mLogic) [30]. Another set of spintronic devices is based on voltage-controlled switching of magnets. These devices can potentially improve energy efficiency because they do not need a large current and avoid the energy associated with the Joule heating and the leakage. The representative devices included in this paper are the MEMTJ device, the spin wave device (SWD), and CoMET.

MEMTJ and CoMET are two new devices that are added to the benchmarking. Furthermore, several modified technology options for MEMTJ and CSL devices with more advanced device materials and structures have been evaluated. More accurate modeling approaches have also been used for the existing devices, such as ASL, CSL, mLogic, and SWD. The updated modeling approach for each device is described as follows.

1) ALL-SPIN LOGIC (ASL)

The original ASL device was proposed in [28]. Compared to the model used in BCB 3.0, the new model takes into account spin relaxation during the spin diffusion along the metallic channel. The spin polarized current density received at the output magnet is [31]

$$J_s = \frac{\beta J_c}{\frac{\sinh(l_c/l_{sf}) \cosh(l_g/l_{sf})}{\sinh(l_g/l_{sf})} + \cosh(l_c/l_{sf})} \quad (1)$$

where J_c is the charge current density, l_c is the channel length, l_g is the length of the ground path, β is the spin injection coefficient, and l_{sf} is the spin diffusion length. The spin diffusion length in copper has been obtained using compact physical models that account for surface and ground boundary scatterings in nanoscale wires [32]. The spin injection coefficient, critical switching current, and magnet switching delay are adopted from previous work [4]. In addition to the in-plane magnet, perpendicular magnetic anisotropy (PMA) magnets are also investigated in ASL devices. PMA magnets require much lower critical switching currents, allowing for more energy-efficient computing [33]. One novel PMA magnetic material, Heusler alloy, is also added into the benchmarking plot. It has the advantage of a large anisotropy value of 2.6×10^6 J/m³, allowing a small magnet size without sacrificing thermal stability and enabling a fast switching time [34].

2) CHARGE-SPIN LOGIC (CSL)

CSL was originally proposed in [29], as shown in Fig. 1(a). The magnetic orientation of the bottom free magnet is controlled by the spin orbital torque generated by passing a charge current through a heavy metal, namely, spin Hall

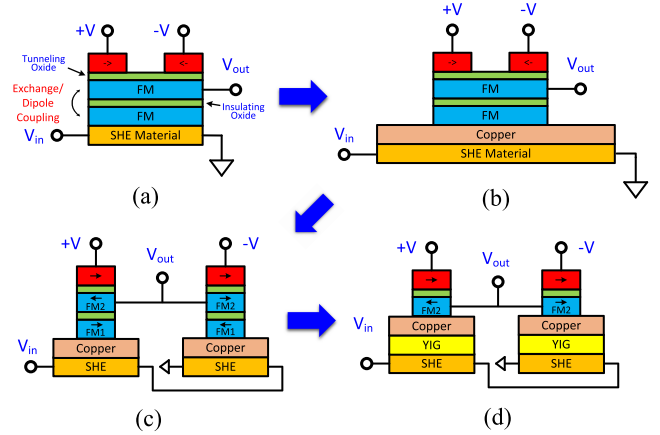


FIGURE 1. Schematics of four CSL devices (a) with originally proposed device, (b) that use copper collector to boost spin current, (c) that separate the pull-up and pull-down networks, and (d) that add a YIG layer between the copper collector and SHE material.

effect (SHE). The magnetic orientation of the top magnet is controlled by the bottom magnet via dipole coupling. The two magnets are electrically isolated which ensures input–output isolation. The magnetization of the top magnet determines the polarity of the charge current transmitted to the next stage via the tunneling magnetoresistance effect. Several updates have been made for the CSL device: the magnets were made $3 \times$ bigger to fit two MTJs; the magnet parameters were updated to ensure perfect dipole coupling [35]; and the output resistance network is included to estimate the driving current to the next stage, as shown in the supplementary material.

To further improve the spin current injected into the bottom magnet, a copper layer has been proposed to collect spins from a large surface area and funnel them to the magnet via spin diffusion [36], as shown in Fig. 1(b). For the benchmarking results shown in Section IV, an enhancement factor of 2 is considered to show the potential improvement by adding the copper collector (supplementary material). This factor has been calculated by accounting for the fact that the electrical current in the heavy metal gets partly shunted by the copper layer. Using a thinner copper layer decreases the shunted electrical current; however, the spin diffusion length in the copper layer decreases due to size effects [32]. Hence, there is an optimal Cu thickness that maximizes the spin transfer torque for a given electrical current.

For the device structure shown in Fig. 1(a) and (b), one major fabrication challenge is to create two fixed magnets side by side whose magnetizations point in opposite directions. To address this challenge, we propose and investigate a new CSL structure by breaking the device into two complementary pull-up and pull-down networks, as shown in Fig. 1(c). In this complementary device, all fixed magnets point in the same direction. The last advancement we investigated is adding an Yttrium iron garnet (YIG) layer between the SHE material and the copper collector, as shown in Fig. 1(d). This creates an insulating layer that electrically

isolates the input and the output, eliminating the need for dipole coupling between two free magnets [37]. In addition, the YIG layer prevents the parasitic current path through the copper collector and further improves the spin injection by an extra 50% [36]. The comparison among four different CSL devices with advanced materials and structures will be illustrated in Section IV.

3) MAGNETIC DOMAIN WALL MAGNETIC LOGIC (mLogic)

The magnetic domain wall-based logic device, mLogic, is included in the updated benchmarking work. It is considered as the same device as the STT-DW device in BCB 3.0 but with an updated complementary logic implementation and numerical simulation for the domain wall velocity. Unlike other spintronic devices relying on majority-gate logic, mLogic devices perform computation with complementary logic circuits that are similar to CMOS circuits. The output voltage depends on the pull-up and pull-down resistance networks that are set according to the input current generated by the previous stage. The modeling approach, such as the domain wall speed versus the input current density, follows the previous work [30].

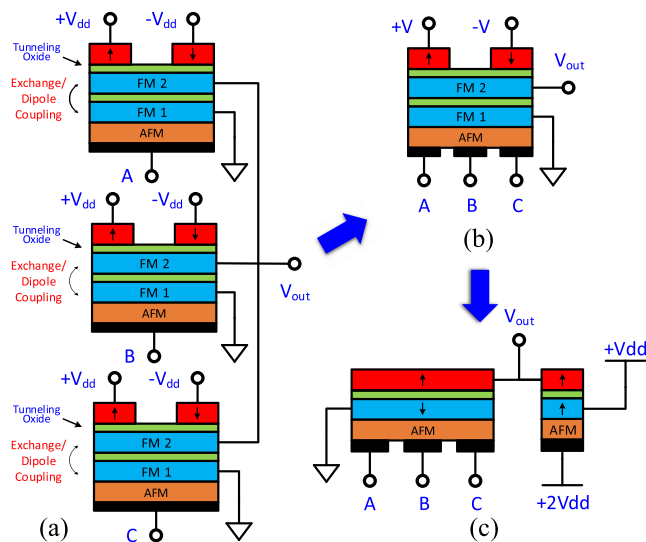


FIGURE 2. Schematics of three MEMTJ devices. (a) Standard MEMTJ. (b) Compact MEMTJ device with the assumption of a single-domain magnet. (c) Preset-based MEMTJ device without the dipolar coupling.

4) MAGNETOELECTRIC MAGNETIC TUNNELING JUNCTION (MEMTJ) DEVICE

The proposed stand-alone voltage-controlled MEMTJ device is shown in Fig. 2(a). The original MEMTJ logic concept was originated in [38] and [39]. The basic building block consists of a magnetoelectric antiferromagnetic (AFM) layer stacked with an MTJ. Chromia (Cr_2O_3) provides an exciting opportunity in this regard. The boundary magnetization of Cr_2O_3 can be isothermally controlled via an applied electric field and the generated voltage-controlled perpendicular exchange bias can be used to switch an adjacent

ferromagnetic layer [40]–[43]. The magnetization of the free magnet determines the output MTJ resistance. Using a MOSFET at the output, this MTJ resistance is converted back to the voltage and drives the next stage. Building upon this design, we propose a stand-alone voltage-controlled magnetoelectric device to address the following challenges and limitations. First, each device needs multiple dedicated MOSFETs to drive the next stage. Second, a preset and clocking scheme is required to perform logic functions since the output voltage is only positive. Third, devices are very sensitive to the insulator thickness variability because the voltage division between the FET and the MTJ determines the output voltage. Any variation in the insulator thickness changes the MTJ resistance exponentially, and consequently shifts the output voltage significantly.

The proposed MEMTJ is similar to the CSL device in which the current controlled write element (SHE) has been replaced with a voltage-controlled magnetoelectric element. Like CSL, it satisfies all five essential requirements of general logic applications: nonlinearity, gain, concatenability, feedback prevention, and a complete set of Boolean operations based on the majority gate and inverter. However, the magnetoelectric effect is far more energy efficient compared to spin transfer torque. The proposed device can directly drive the next stage and eliminate the need for any auxiliary FETs. Two matched MTJs are built at the output stage so that the output voltage is not sensitive to the MTJ insulator thickness variability. This is because the pull-up and pull-down resistances change proportionally as the MTJ insulator thickness changes. A major challenge for this device is to ensure perfect coupling between the write and read magnets via dipolar coupling. MEMTJ uses PMA magnets and the design space for perfect coupling of PMA magnets via dipole coupling is quite narrow [44].

One can connect the outputs of three MEMET to create a majority gate. However, a more compact MEMTJ device can be created assuming a single-domain magnet similar to the previous proposal [41], [42] [Fig. 2(b)]. In this case, the AFM is gated with three inputs and the magnetization of the free bottom magnet depends on the majority of the inputs. This compact device option also improves the voltage swing of the output because the pull-up and pull-down units only have one MTJ. In addition, the input capacitance is smaller compared to the case where the outputs of three inverters [Fig. 2(a)] are connected in parallel.

To address the challenge with the dipole coupling and also to eliminate the need for fixed magnets pointing in opposite directions, a third MEMTJ device is proposed in Fig. 2(c). The detailed operation and configuration during the preset and computation period are described in the supplementary material.

5) SPIN WAVE DEVICE (SWD)

In SWD, a voltage is applied across a piezoelectric material to create strain and change the magnetization of the magnet through a magnetostrictive effect. This creates a spin wave

that propagates through the magnetic channel toward the output magnet. The output magnet is preset at the metastable condition until the spin wave arrives. After that, a phase-dependent deterministic switching is realized by modifying the energy landscape and shifting the location of the saddle point [45]. A clocking scheme enables the correct detection and transmission of spin wave signals and guarantees nonreciprocity [46]. The majority of the delay is associated with the magnet switching from the metastable state to the steady state. The intrinsic switching energy of the ME cell dominates the overall energy dissipation. The energy associated with the clocking is small because one clocking transistor can drive hundreds of ME cells in the same stage. The detailed modeling approaches to calculate the energy and delay per operation is described in [47].

6) COMPOSITE-INPUT MAGNETOELECTRIC-BASED LOGIC TECHNOLOGY (CoMET)

CoMET is another new voltage-controlled device that has been added into the benchmarking. It enables low-voltage-induced domain wall nucleation based on the magneto-electric effect. The fast domain propagation is realized by passing a charge current through a heavy metal laid underneath the PMA magnetic channel (SHE). The delay and energy dissipation are dominated by the domain wall nucleation/propagation and the switching energy of CMOS transistors, respectively. The modeling approaches to calculate the delay per operation is adopted from the numerical simulation in [48] with updated energy dissipation calculation for the dynamic switching energy of the transistors, leakage energy of CMOS inverters, and Joule heating energy.

C. ARCHIVED DEVICES

Some of the devices from BCB 3.0 are archived due to their uncompetitive performance as well as the lack of activity in various research centers. These devices include graphene nanoribbon FET (GNR TFET), SpinFET, piezoelectric FET, excitonic FET, spin majority gate, nanomagnetic logic, and spin torque oscillator (STOlogic).

III. BENCHMARKING METHODOLOGY

A. BOOLEAN LOGIC BENCHMARKING

The 32-bit ALU from BCB 3.0 [4] is adopted for the circuit-level benchmarking of the Boolean circuits. One major update has been made for the clocking scheme of NDR devices (i.e., BisFET and ITFET). Complementary logic gates based on NDR devices take their input at the rising edge of the supply voltage and their output will not change until the supply voltage falls to zero again. A multiphase clocking scheme has been proposed to perform logic computation and propagation in a multistage Boolean circuit, such as the ripple carry adder [49]. To ensure all SUM bits are available at the end of the 32-bit addition, BCB 3.0 assumes that all NDR devices are constantly clocked until all 32 bits are computed. This constant clocking means that each bit of sum

is calculated 32 times, and it is used only once at the end. Alternatively, in this paper, we disable the clocking after each logic gate finishes the computation, and only hold the clock on for the last logic gates (xor) of each full adder to lock the SUM bit. This reduces the dynamic power dissipation by $32\times$ but increases the leakage power of the last logic gates. The implications of this tradeoff will be discussed in Section IV.

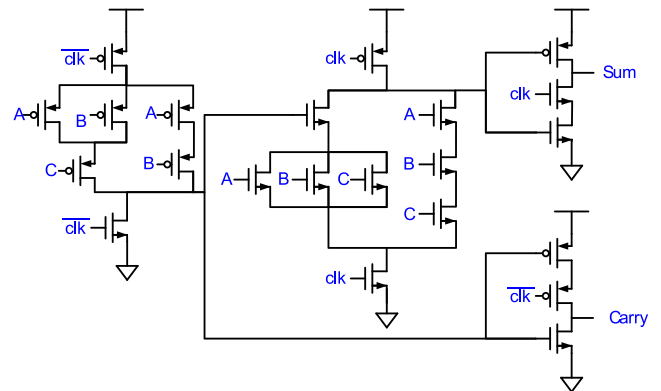


FIGURE 3. Circuit diagram of a full-bit adder using the standard N-P domino logic [50].

To better utilize the power density budget and quantify the benefit of the deep pipelining, the standard N-P domino logic is implemented to boost the throughput of low-power FETs, as shown in Fig. 3 [50]. The delay is estimated based on the worst-case input combinations, and the energy is estimated based on the switching probability of inputs as well as internal nodes, such as the \overline{Carry} and \overline{Sum} .

For the NDR devices, BisFET and ITFET, as well as the spintronic devices, the supply clocking that needs to be used to ensure functionality or to eliminate standby power enables a similar pipelining where the logic depth becomes equal to one. All these devices are intrinsic memory elements. Magnetic devices are nonvolatile, and NDR devices latch the input signal only at the rising edge of the supply voltage (clock). The value of the input has no effect on the output so long as the supply voltage remains high.

B. INTERCONNECT CENTRIC ANALYSIS

Interconnects impose a major limitation on the state-of-the-art integrated circuits. Previous studies have shown that interconnects account for more than half of dynamic power dissipation and critical path delay [5]. More than 50% of the logic cells on a chip may be used as repeaters for long interconnects [6]. As wire dimensions scale, the size effect significantly increases wire resistivity below sub-20-nm nodes [51]. Therefore, it is crucial to evaluate the implications of various novel device proposals. The energy and delay of an interconnect with the optimal repeater insertion are

derived as [52]

$$t_{\text{wire}} = 1.4\sqrt{R_0 C_0 r_w c_w} \cdot l + 2\sqrt{(0.7R_0 C_0 + t_p)0.4r_w c_w} \cdot l \quad (2)$$

$$E_{\text{wire}} = \frac{1}{2}c_w l \left(1 + \sqrt{\frac{0.4R_0 C_0}{0.7R_0 C_0 + t_p}} \right) V_{dd}^2 \quad (3)$$

respectively, where r_w and c_w are the resistance and capacitance per unit length of the interconnect, l is the length of the interconnect, t_p is the extra polarization switching time of ferroelectric devices, R_0 and C_0 are the output resistance and input capacitance of a minimum-sized repeater, respectively, and V_{dd} is the supply voltage.

The span of control, originally proposed in [53], addresses the communication among logic switches by measuring the number of accessible devices within one clock cycle. This metric is a rough indicator of the circuit block size for each device technology beyond which interconnects become a major limitation. In this work, the clock period is assumed to be 30 times the intrinsic device delay. Here, only static logic circuits are considered. Results based on the latest device-level models in Section II will be discussed in Section IV.

C. NEUROMORPHIC BENCHMARKING CIRCUITS

Despite the large research efforts in the Boolean logic domain, few device concepts have better or comparable performance compared to the conventional CMOS technology [4]. Recent studies have shown that non-Boolean logic can better utilize emerging technologies, such as spintronics, and achieve a better computing energy efficiency [7], [8].

Some non-Boolean computing architectures, such as the CNN [11], are promising candidates to provide higher energy efficiencies due to their massively parallel processing capability. In this paper, we follow the same methodology presented in our previous work to benchmark a variety of charge- and spin-based devices [54]. For charge-based devices, both analog and digital implementations are investigated. Results are simulated based on the updated device-level characteristics, i.e., the bias current, subthreshold slope, and supply voltage. For the spintronic CNN, devices with new materials and structures are included, including ASL devices with Heusler alloy magnets and CSL devices with copper collector and YIG. The associative memory application is investigated for three types of CNN implementations using 4-bit weight synapses to achieve 90% recall accuracy for a given input noise of 10%.

IV. RESULTS AND DISCUSSION

In this section, the benchmarking results are demonstrated based on the device-, interconnect-, and circuit-level models described in Sections II and III.

A. 32-BIT ADDER

1) ENERGY VERSUS DELAY

Results of a 32-bit ALU are shown in Fig. 4 for a variety of charge- and spin-based devices. In general, spintronic devices

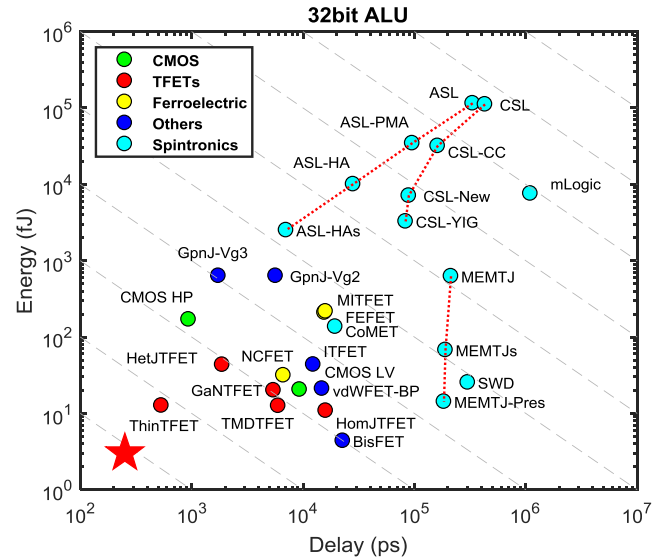


FIGURE 4. Energy versus delay of a 32-bit ALU for a variety of charge- and spin-based devices. Here, ASL-HA and ASL-HAS stand for ASL devices using the Heusler alloy with nominal and improved saturation magnetization values of 4×10^5 and 10^5 A/m, respectively; CSL, CSL-CC, CSL-New, and CSL-YIG correspond to device structures shown in Fig. 1(a)–(d), respectively; MEMTJ, MEMTJs, and MEMTJ-preset correspond to device structures shown in Fig. 2(a)–(c), respectively. The red star indicates the preferred corner.

consume more energy and delay per operation due to the large switching delay of nanomagnets as well as the large Joule heating of the current-driven devices.

Compared to the previous benchmarking work, the data points for several TFETs have moved considerably toward the preferred corner because of more accurate modeling approaches and improved material and structures. Compared to other TFET devices, ThinTFET provides the best performance thanks to its steep subthreshold slope and large on current at a small supply voltage. The ultrathin channel structure leads to a strong gate control over the channel. Layered 2-D crystals provide a sharp turn on of density of states at the band edges and have no surface dangling bonds. This potentially enables a low interfacial density of state, which are highly desired for achieving a steep subthreshold slope [55]. GPNJ devices consume a much larger energy per operation because the new modeling approach based on ray tracing and NEGF has demonstrated that a larger device size is required to achieve a reasonable on–off ratio, leading to large dynamic switching energy. For BisFET and ITFET, a new clocking scheme is applied, as described in Section III-A. Since the supply clocking of the logic gate is disabled once the computation completes, the dynamic energy is reduced significantly. However, with a low on–off ratio of 10, BisFET and ITFET suffer from large leakage energy, which contributes to the majority of the energy dissipation.

For spintronic devices, the data points for the original ASL and CSL device proposals have moved away from the

preferred corner because of the more realistic modeling approach described in Section II. However, many recent advancements in both device structures and materials improve the switching delay and reduce the critical switching current requirements, leading to a continuous performance improvement toward the preferred corner. One can observe that voltage-controlled spintronic devices, such as MEMENTJ, SWD, and CoMET, have a great advantage in terms of energy dissipation compared to their current-driven counterparts. In the supplementary material, we have shown a detailed comparison between BCB 3.0 and the latest results for both charge- and spin-based devices on two separate plots.

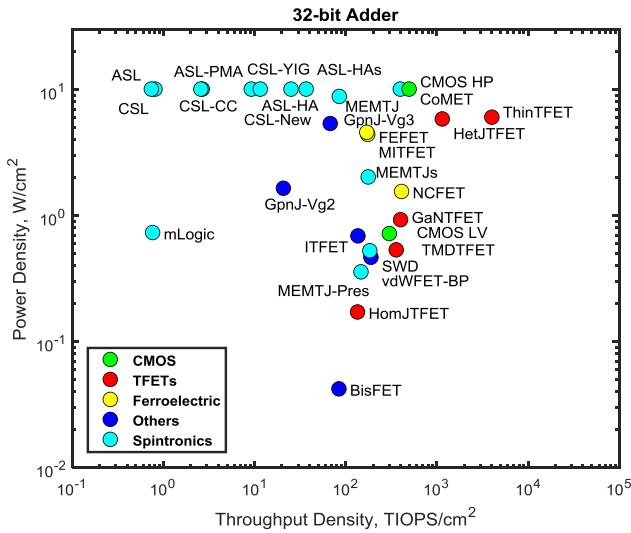


FIGURE 5. Power density versus throughput density for a variety of charge- and spin-based devices.

2) PERFORMANCE UNDER THE POWER DENSITY CONSTRAINT

For many computing applications, power density is a critical constraint that limits the maximum operation speed of a chip. Therefore, it is desirable to investigate the throughput density under a fixed power density. Fig. 5 shows the constrained throughput of a 32-bit adder under the power density limit of 10 W/cm² for a variety of devices. For low-power devices, the throughput is limited by the delay. The CMOS HP device cannot fully utilize its speed advantage due to the power density cap, and the slower but more energy-efficient HetJTFET has a better throughput.

To further improve the throughput of low-power devices, such as TFETs and voltage-controlled spintronic devices, ultradeep pipelining circuit techniques are employed. For charge-based devices, standard N-P domino logic is implemented to enable the pipeline circuit as described in Section III-A. For NDR devices and spintronic devices, supply clocking is used to achieve ultradeep pipelining and to boost the throughput. The comparison of various technologies using deep-pipelined circuits is shown in Fig. 6. One clear trend is that low-power devices shift significantly

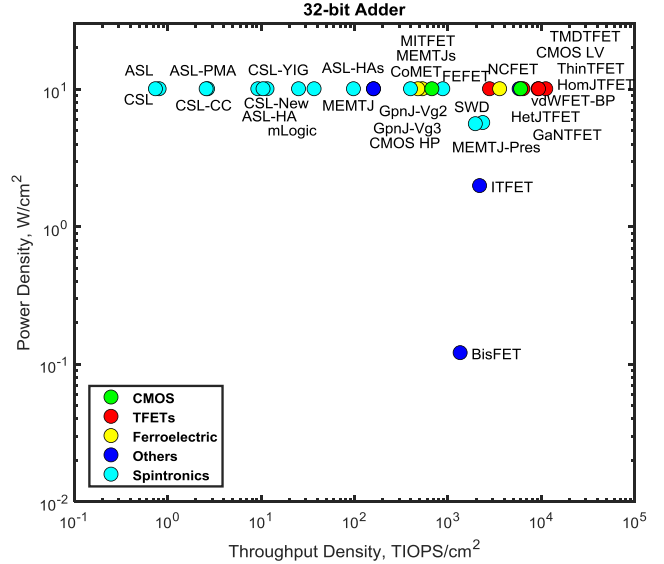


FIGURE 6. Power density versus throughput density for a variety of charge- and spin-based devices with ultradeep pipelining. The charge-based FETs are implemented with the standard N-P domino logic except for NDR devices, which inherently have a memory feature.

to the top right corner. Devices closer to the power density cap benefit less from the pipelined circuit. Most charge-based devices and three voltage-controlled spintronic devices provide better throughput than the CMOS HP.

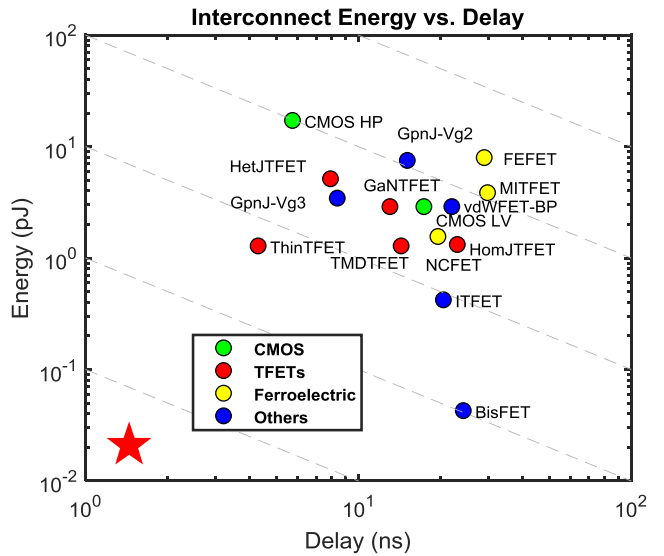


FIGURE 7. Energy versus delay of a 10- μ m interconnect with repeater insertion using charge-based devices, where the red star indicates the preferred corner.

B. INTERCONNECT

1) INTERCONNECTS WITH REPEATER INSERTIONS

Figs. 7 and 8 show the delay and energy of passing data through a 10- μ m-length interconnect using charge- and

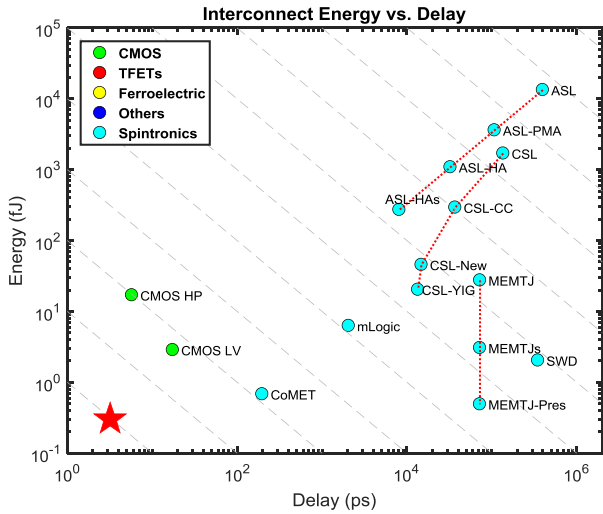


FIGURE 8. Energy versus delay of a 10- μm interconnect with repeater insertion using spintronic devices, where the red star indicates the preferred corner.

spin-based technologies, respectively. BisFET and ITFET are much closer to the preferred corner than other charge-based devices. This is because in the 32-bit ALU benchmarking, the leakage energy contributes to the majority of energy dissipation, as shown in Fig. 4. For the interconnect application, the dynamic energy becomes more dominant; BisFET and ITFET have the advantage of ultralow supply voltage, leading to a much lower energy.

Due to the limited magnet switching speed, spintronic interconnects are much slower compared to charge-based interconnects, as shown in Fig. 8. Compared to results for the 32-bit ALU, the gap between the charge- and spin-based interconnects is even larger because the majority-gate-based spintronic logic is very efficient to perform a full adder. Only one majority gate is required in the critical path to generate the carry bit.

2) SPAN OF CONTROL

Fig. 9 shows the maximum numbers of reachable NAND2 gates per clock cycle for various emerging technologies. For instance, the GPNJ device has a relatively fast intrinsic speed and a short clock cycle; however, it has a large span and reaches more gates than most of the other devices despite a large footprint area. This is mainly because of a low resistance enabled by the high mobility of graphene. Ferroelectric- and piezoelectric-based devices, marked by yellow circles, offer the largest span of control because of their long clock cycles due to the extra polarization switching time. It should be noted that the span of control is not a performance metric; rather, it is an indicator of the circuit size beyond which interconnects impose severe limits. Obviously, an intrinsically fast device is more susceptible to performance degradation due to interconnects.

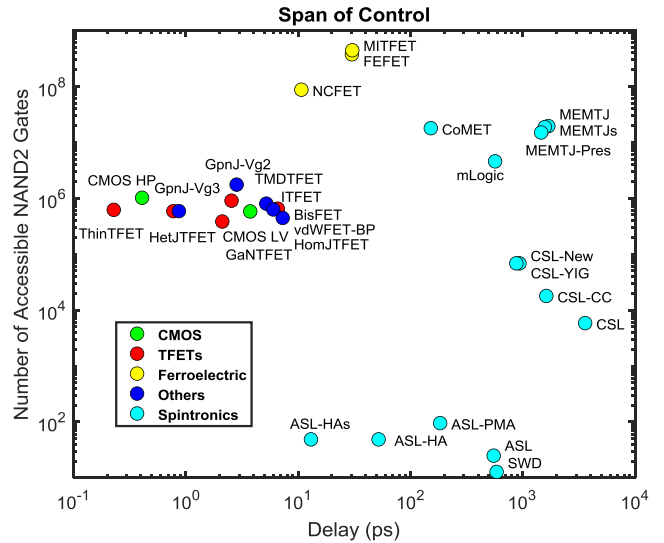


FIGURE 9. Span of control versus intrinsic delay for a variety of charge- and spin-based devices.

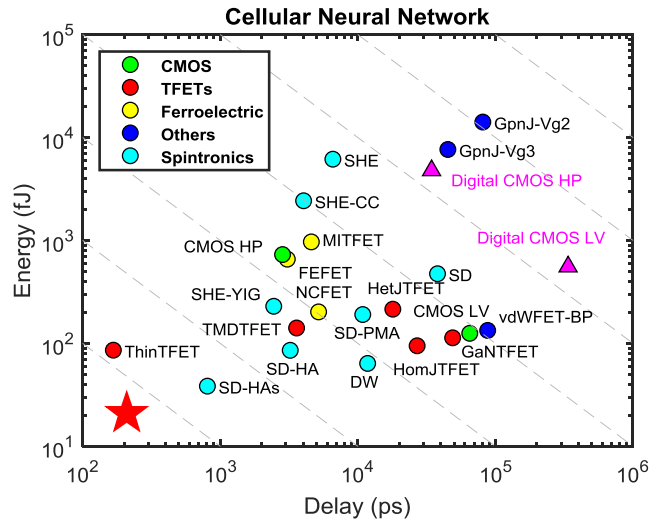


FIGURE 10. Energy versus delay per memory association operation using CNN for a variety of charge- and spin-based devices, where the red star indicates the preferred corner.

C. NON-BOOLEAN COMPUTING BENCHMARKING BASED ON CELLULAR NEURAL NETWORK (CNN)

Following our previous non-Boolean computing benchmarking for a CNN [54], the updated results are shown in Fig. 10 with the latest device-level models described in Section II. Note that to achieve 90% recall accuracy for a given input noise of 10%, the number of connectivity required is about 30, which may impose constraints on the routing. However, the benchmarking methodology in this paper can be further extended to a convolutional neural network based on CNN, where the connectivity requirement is 9 [13].

In Fig. 10, triangle markers show the digital CNN implementation based on CMOS HP and LV devices. Compared to analog implementation (green markers), digital CNNs require

multiple addition and multiplication operations for each time step, which is time and energy consuming. Therefore, all the other emerging charge-based devices are implemented based on the analog circuits. TFETs have significant performance improvement because of their steep subthreshold slopes and large driving currents at ultralow supply voltages. Compared to the previous benchmarking work, spintronic devices have shifted much closer to the preferred corner. This is because for spin diffusion and SHE-based CNNs, a single magnet can mimic the functionality of a neuron to perform the integration; for the domain-wall-based CNNs, the integration is performed by moving the domain wall inside the free magnet, which is very energy efficient due to a small critical switching current. For charge-based CNNs, an operational amplifier is required for each neuron and synapse, which consumes more power and requires a large footprint area.

V. CONCLUSION

In this paper, a new release of the uniform benchmarking methodology for beyond-CMOS device is presented for both Boolean and non-Boolean logic applications. More realistic modeling approaches are included, and more advanced device material and structures are investigated. In general, spintronic devices are slower than charge-based devices because of the limited ferromagnet switching speed and domain wall propagation speed. Voltage-controlled spintronics devices are more energy efficient than current-driven ones. Three types of CNN implementations have been investigated based on a uniform benchmarking methodology. Spintronic devices show great performance in neuromorphic computing circuits, which differ significantly from their results in Boolean circuits, such as a 32-bit ALU. This indicates that new devices need to be complemented with novel circuits to achieve their full potential.

ACKNOWLEDGMENT

The authors would like to thank colleagues in SRC STARnet and NRI: J. Nahas, J. Appenzeller, S. Datta, C. Kim, S. Sapatnekar, M. Mankalale, Z. Liang, J. Wang, P. Dowben, S. Hu, M. Niemier, V. Narayanan, S. Salahuddin, A. Seabaugh, F. Register, A. Marshall, R. Lake, S. Sylvia, A. Ghosh, and M. Elahi. They would like to thank the NRI/STARnet Benchmarking Steering Committee members, I. Young and D. Nikonov from Intel Co., Santa Clara, CA, USA, S. Kramer from Micron, Boise, ID, USA, W. Haensch from IBM, Armonk, NY, USA, and J. Herbsommer from Texas Instruments, Dallas, TX, USA, for many useful discussions during the quarterly meetings. They would also like to thank current and former colleagues at Georgia Tech, Atlanta, GA, USA, S. Chang, S. Dutta, N. Kani, R. Irai, V. Huang, and C. Hsu, for their contributions.

REFERENCES

[1] T. N. Theis and H.-S. P. Wong, "The end of Moore's Law: A new beginning for information technology," *Comput. Sci. Eng.*, vol. 19, no. 2, pp. 41–50, Apr. 2017.

[2] K. Bernstein, R. K. Cavin, III, W. Porod, A. Seabaugh, and J. Welser, "Device and architecture outlook for beyond CMOS switches," *Proc. IEEE*, vol. 98, no. 12, pp. 2169–2184, Dec. 2010.

[3] D. E. Nikonov and I. A. Young, "Uniform methodology for benchmarking beyond-CMOS logic devices," in *IEDM Tech. Dig.*, Dec. 2012, pp. 25.4.1–25.4.4.

[4] D. E. Nikonov and I. A. Young, "Benchmarking of beyond-CMOS exploratory devices for logic integrated circuits," *IEEE J. Exploratory Solid-State Comput. Devices Circuits*, vol. 1, no. 1, pp. 3–11, Dec. 2015.

[5] N. Magen, A. Kolodny, U. Weiser, and N. Shamir, "Interconnect-power dissipation in a microprocessor," in *Proc. Int. Workshop Syst. Level Interconnect Prediction*, Feb. 2004, pp. 7–13.

[6] P. Saxena, N. Menezes, P. Cocchini, and D. A. Kirkpatrick, "Repeater scaling and its impact on CAD," *IEEE Trans. Comput.-Aided Design Integr.*, vol. 23, no. 4, pp. 451–463, Apr. 2004.

[7] S. G. Ramasubramanian, R. Venkatesan, M. Sharad, K. Roy, and A. Raghunathan, "SPINDLE: SPINtronic deep learning engine for large-scale neuromorphic computing," in *Proc. Int. Symp. Low Power Electron. Design*, 2014, pp. 15–20.

[8] A. Sengupta, S. H. Choday, Y. Kim, and K. Roy, "Spin orbit torque based electronic neuron," *Appl. Phys. Lett.*, vol. 106, no. 14, p. 143701, 2015.

[9] A. R. Trivedi and S. Mukhopadhyay, "Potential of ultralow-power cellular neural image processing with Si/Ge tunnel FET," *IEEE Trans. Nanotechnol.*, vol. 13, no. 4, pp. 627–629, Jul. 2014.

[10] I. Palit, X. S. Hu, J. Nahas, and M. Niemier, "TFET-based cellular neural network architectures," in *Proc. Int. Symp. Low Power Electron. Design*, Sep. 2013, pp. 236–241.

[11] L. O. Chua and L. Yang, "Cellular neural networks: Theory," *IEEE Trans. Circuits Syst.*, vol. CS-35, no. 10, pp. 1257–1272, Oct. 1988.

[12] C. Pan and A. Naeemi, "A proposal for energy-efficient cellular neural network based on spintronic devices," *IEEE Trans. Nanotechnol.*, vol. 15, no. 5, pp. 820–827, Sep. 2016.

[13] A. Horváth, M. Hillmer, Q. Lou, X. S. Hu, and M. Niemier, "Cellular neural network friendly convolutional neural networks—CNNs with CNNs," in *Proc. Design, Autom. Test Europe Conf. Exhib. (DATE)*, Mar. 2017, pp. 145–150.

[14] (2018). *Benchmarking of Devices in the Nanoelectronics Research Initiative*. Available: <https://nanohub.org/tools/nribench/browser/trunk/src>

[15] H. Ilatikhameneh, Y. Tan, B. Novakovic, G. Klimeck, R. Rahman, and J. Appenzeller, "Tunnel field-effect transistors in 2-D transition metal dichalcogenide materials," *IEEE J. Exploratory Solid-State Comput. Devices Circuits*, vol. 1, no. 1, pp. 12–18, Dec. 2015.

[16] M. O. Li, D. Esseni, J. J. Nahas, D. Jena, and H. G. Xing, "Two-dimensional heterojunction interlayer tunneling field effect transistors (thin-TFETs)," *IEEE J. Electron Devices Soc.*, vol. 3, no. 3, pp. 200–207, May 2015.

[17] A. Seabaugh *et al.*, "Steep subthreshold swing tunnel FETs: GaN/InN/GaN and transition metal dichalcogenide channels," in *IEDM Tech. Dig.*, Dec. 2015, pp. 35.6.1–35.6.4.

[18] J. Nahas, private communication, May 2017.

[19] S. Steiger, M. Povolotskiy, H.-H. Park, T. Kubis, and G. Klimeck, "NEMOS: A parallel multiscale nanoelectronics modeling tool," *IEEE Trans. Nanotechnol.*, vol. 10, no. 6, pp. 1464–1474, Nov. 2011.

[20] S. L. Miller and P. J. McWhorter, "Physics of the ferroelectric non-volatile memory field effect transistor," *J. Appl. Phys.*, vol. 72, no. 12, pp. 5999–6010, 1992.

[21] K.-S. Li *et al.*, "Sub-60mV-swing negative-capacitance FinFET without hysteresis," in *IEDM Tech. Dig.*, Dec. 2015, pp. 22.6.1–22.6.4.

[22] J. P. Duarte *et al.*, "Compact models of negative-capacitance FinFETs: Lumped and distributed charge models," in *IEDM Tech. Dig.*, Dec. 2016, pp. 30.5.1–30.5.4.

[23] A. I. Khan *et al.*, "Negative capacitance in short-channel FinFETs externally connected to an epitaxial ferroelectric capacitor," *IEEE Electron Device Lett.*, vol. 37, no. 1, pp. 111–114, Jan. 2016.

[24] S. Chen *et al.*, "Electron optics with p-n junctions in ballistic graphene," *Science*, vol. 353, no. 6307, pp. 1522–1525, 2016.

[25] X. Mou, L. F. Register, A. H. MacDonald, and S. K. Banerjee, "Quantum transport simulation of exciton condensate transport physics in a double-layer graphene system," *Phys. Rev. B*, vol. 92, p. 235413, Dec. 2015.

[26] S. S. Sylvia, K. Alam, and R. K. Lake, "Uniform benchmarking of low-voltage van der Waals FETs," *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 2, pp. 28–35, 2016.

- [27] S. A. Wolf, J. Lu, M. R. Stan, E. Chen, and D. M. Treger, "The promise of nanomagnetism and spintronics for future logic and universal memory," *Proc. IEEE*, vol. 98, no. 12, pp. 2155–2168, Dec. 2010.
- [28] B. Behin-Aein, D. Datta, S. Salahuddin, and S. Datta, "Proposal for an all-spin logic device with built-in memory," *Nature Nanotechnol.*, vol. 5, no. 4, pp. 266–270, Apr. 2010.
- [29] S. Datta, S. Salahuddin, and B. Behin-Aein, "Non-volatile spin switch for Boolean and non-Boolean logic," *Appl. Phys. Lett.*, vol. 101, no. 25, p. 252411, 2012.
- [30] D. Morris, D. Bromberg, J.-G. Zhu, and L. Pileggi, "mLogic: Ultra-low voltage non-volatile logic circuits using STT-MTJ devices," in *Proc. 49th Annu. Design Autom. Conf.*, 2012, pp. 486–491.
- [31] C. Pan, S.-C. Chang, and A. Naeemi, "Performance analyses and benchmarking for spintronic devices and interconnects," in *Proc. IEEE Int. Interconnect Technol. Conf. (IITC)*, May 2016, pp. 56–58.
- [32] S. Rakheja, S.-C. Chang, and A. Naeemi, "Impact of dimensional scaling and size effects on spin transport in copper and aluminum interconnects," *IEEE Trans. Electron Devices*, vol. 60, no. 11, pp. 3913–3919, Nov. 2013.
- [33] D. C. Worledge *et al.*, "sullivan, and R. Robertazzi, "Spin torque switching of perpendicular TaCoFeB/MgO-based magnetic tunnel junctions," *Appl. Phys. Lett.*, vol. 98, no. 2, p. 022501, 2011.
- [34] C. H. Kim and S. S. Sapatnekar, private communication, May 2017.
- [35] N. Kani, S.-C. Chang, S. Dutta, and A. Naeemi, "A model study of an error-free magnetization reversal through dipolar coupling in a two-magnet system," *IEEE Trans. Magn.*, vol. 52, no. 2, Feb. 2016, Art. no. 1300112.
- [36] S. Sayed, V. Q. Diep, K. Y. Camsari, and S. Datta, "Spin funneling for enhanced spin injection into ferromagnets," *Sci. Rep.*, vol. 6, p. 28868, 2016.
- [37] J. Appenzeller, private communication, Aug. 2017.
- [38] C. Binek and B. Doudin, "Magnetoelectronics with magnetoelectrics," *J. Phys., Condens. Matter*, vol. 17, no. 2, p. L39, 2004.
- [39] M. Bibes and A. Barthélemy, "Towards a magnetoelectric memory," *Nature Mater.*, vol. 7, no. 6, pp. 425–426, 2008.
- [40] X. He *et al.*, "Robust isothermal electric control of exchange bias at room temperature," *Nature Mater.*, vol. 9, pp. 579–585, Jun. 2010.
- [41] N. Sharma, A. Marshall, J. Bird, and P. Dowben, "Magneto-electric magnetic tunnel junction logic devices," in *Proc. 4th Berkeley Symp. Energy Efficient Electron. Syst. (ES)*, 2015, pp. 1–3.
- [42] N. Sharma, A. Marshall, J. Bird, and P. Dowben, "Magneto-electric magnetic tunnel junction as process adder for non-volatile memory applications," in *Proc. IEEE Dallas Circuits Syst. Conf. (DCAS)*, Oct. 2015, pp. 1–4.
- [43] N. Sharma, J. Bird, P. Dowben, and A. Marshall, "Compact-device model development for the energy-delay analysis of magneto-electric magnetic tunnel junction structures," *Semicond. Sci. Technol.*, vol. 31, p. 065022, 2016.
- [44] N. Kani, private communication, May 2017.
- [45] S. Dutta, D. E. Nikonov, S. Maniaturuni, I. A. Young, and A. Naeemi, "Phase-dependent deterministic switching of magnetoelectric spin wave detector in the presence of thermal noise via compensation of demagnetization," *Appl. Phys. Lett.*, vol. 107, no. 19, p. 192404, 2015.
- [46] S. Dutta *et al.*, "Non-volatile clocked spin wave interconnect for beyond-CMOS nanomagnet pipelines," *Sci. Rep.*, vol. 5, May 2015, Art. no. 9861.
- [47] S. Dutta *et al.*, "Impact of spintronics transducers on the performance of spin wave logic circuit," in *Proc. IEEE 16th Int. Conf. Nanotechnol. (IEEE-NANO)*, Aug. 2016, pp. 990–993.
- [48] M. G. Mankalale, Z. Liang, Z. Zhao, C. H. Kim, J.-P. Wang, and S. S. Sapatnekar, "CoMET: Composite-input magnetoelectric-based logic technology," *IEEE J. Explor. Solid-State Computat. Devices Circuits*, vol. 3, pp. 27–36, 2017.
- [49] L. F. Register, D. Basu, and D. Reddy, "From coherent states in adjacent graphene layers toward low-power logic circuits," *Adv. Condensed Matter Phys.*, vol. 2011, Aug. 2010, Art. no. 258731.
- [50] F. Lu and H. Samueli, "A 200 MHz CMOS pipelined multiplier-accumulator using a quasi-domino dynamic full-adder cell design," *IEEE J. Solid-State Circuits*, vol. 28, no. 2, pp. 123–132, Feb. 1993.
- [51] C. Pan and A. Naeemi, "A paradigm shift in local interconnect technology design in the era of nanoscale multigate and gate-all-around devices," *IEEE Electron Device Lett.*, vol. 36, no. 3, pp. 274–276, Mar. 2015.
- [52] C. Pan and A. Naeemi, "Interconnect design and benchmarking for charge-based beyond-CMOS device proposals," *IEEE Electron Device Lett.*, vol. 37, no. 4, pp. 508–511, Apr. 2016.
- [53] D. Matzke, "Will physical scalability sabotage performance gains?" *Computer*, vol. 30, no. 9, pp. 37–39, Sep. 1997.
- [54] C. Pan and A. Naeemi, "Non-Boolean computing benchmarking for beyond-CMOS devices based on cellular neural network," *IEEE J. Explor. Solid-State Computat. Devices Circuits*, vol. 2, pp. 36–43, 2016.
- [55] M. O. Li, R. Yan, D. Jena, and H. G. Xing, "Two-dimensional heterojunction interlayer tunnel FET (thin-TFET): From theory to applications," in *IEDM Tech. Dig.*, Dec. 2016, pp. 19.2.1–19.2.4.



CHENYUN PAN (S'11–M'15) received the B.S. degree in microelectronics from Shanghai Jiao Tong University, Shanghai, China, in 2010, and the M.S. and Ph.D. degrees from the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA, in 2013 and 2015, respectively.

He is currently a Research Engineer with the School of Electrical and Computer Engineering, Georgia Institute of Technology. His current research interests include the device-, circuit-, and system-level modeling and optimization for Boolean and non-Boolean computing platforms based on various emerging device and interconnect technologies.



AZAD NAEEMI (S'99–M'04–SM'04) received the B.S. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 1994, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2001 and 2003, respectively.

He was a Research Engineer with the Micro-Electronics Research Center, Georgia Institute of Technology, from 2003 to 2008, where he joined the Electrical and Computer Engineering Faculty in 2008. He is currently an Associate Professor involving in nanotechnology solutions to the challenges facing giga- and tera-scale systems.