Received 29 August 2016; revised 24 October 2016; accepted 23 November 2016. Date of publication 29 November 2016; date of current version 23 January 2017.

Digital Object Identifier 10.1109/JXCDC.2016.2633251

Non-Boolean Computing Benchmarking for Beyond-CMOS Devices Based on Cellular Neural Network

CHENYUN PAN[1] (Member, IEEE) and AZAD NAEEMI[1] (Senior Member, IEEE)

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

CORRESPONDING AUTHOR: CHENYUN PAN (e-mail: chenyun.pan@gatech.edu).

This work was supported by the Semiconductor Research Corporation NRI Theme 2624.001. This work has supplementary downloadable material available online at http://ieeexplore.ieee.org, provided by the authors. This consists of a PDF file 564 KB in size.

ABSTRACT This paper presents a uniform benchmarking methodology for non-Boolean computation based on the cellular neural network (CNN) for a variety of beyond-CMOS device technologies, including charge-based and spintronic devices. Three types of CNN implementations are investigated using analog, digital, and spintronic circuits. Monte Carlo simulations are performed to quantify the impact of the input noise, thermal noise, and the number of bits representing the weights of synapses on the overall recall probability and delay. The results demonstrate that the recall probability improves significantly as the number of synapses increase. Using a 4-b resolution for synapse weights provides the best tradeoff between the required numbers of synapses and synapse bits for a target recall rate. Finally, three types of CNN implementations with various device technologies are benchmarked for a given input noise and recall accuracy target. It is shown that spintronic devices are promising candidates to implement CNNs, where up to 3× energy-delay product improvement is predicted in domain wall devices compared to its conventional CMOS counterpart.

INDEX TERMS Beyond-CMOS technology, cellular neural network (CNN), performance benchmarking.

I. INTRODUCTION

With CMOS technology approaching its scaling limit [1], many beyond-CMOS device technologies are being considered to augment the conventional Si CMOS technology and to sustain the exponential growth of the computational power of microchips [2]. For the charge-based devices, some promising candidates include ferroelectric negative capacitance FET (NCFET) [3], [4], tunneling FET (TFET) [5]–[8], piezoelectric FET (PiezoFET) [9], graphene p-n junction switch [10], and various FETs based on 2-D materials [11], [12]. Some of these devices can potentially work with lower supply voltages as they may offer steep threshold swings. A decrease in the power supply voltage reduces the dynamic energy dissipation in devices and interconnects quadratically. Another major category of emerging devices includes spintronic devices that use magnets and spin transfer torque mechanism to store and process information [13]. All-spin logic (ASL) [14], charge-coupled spin logic [15], and domain wall logic [16] are some of the well-studied device concepts in this category. These devices can operate with supply voltages of ∼100 mV or even lower and have an additional feature of nonvolatility. Although the intrinsic energy required to switch a very stable magnet at room temperature can be as low as 40 kT, these devices are quite energy hungry because of the inefficiency of the spin transfer torque mechanism. Furthermore, the switching delay of the ferromagnet is typically in the nanosecond range compared to FETs that can switch in less than tens of picoseconds.

Recent benchmarking research based on Boolean circuits such as 32-b adders has projected a limited performance gain for only a few beyond-CMOS device candidates [2]. For spintronic devices, orders of magnitude worse performance in terms of energy-delay product (EDP) has been predicted. Research in the area of beyond-CMOS devices is progressing fast and the proposed devices are being continuously revised and reinvented. Such innovations, which are hard to predict, will with little doubt make emerging devices more competitive. However, one needs to recognize that conventional CMOS devices and their corresponding circuits and architectures have evolved together over many years. Some

of the emerging beyond-CMOS devices offer fundamentally different and in some cases unique characteristics because of which novel and nontraditional circuit concepts are needed to realize their full potential.

To better utilize emerging charge- and spin-based technologies, alternative nonBoolean platforms based on neuromorphic circuits are quite attractive [17], [18]. Biologically inspired computing platforms are highly efficient for solving many problems, particularly in the voice, image, and video processing, by taking advantage of massive parallel low-power computing blocks [19], [20]. Many proposals have studied neuromorphic systems based on spintronic devices, and they are shown to provide low energy per operation compared to the conventional CMOS technology [21], [22]. For the charge-based devices, recent studies demonstrated that using TFETs to build a cellular neural network (CNN) can potentially lower the energy per operation thanks to their low supply voltage and steep subthreshold slope [23], [24].

In this paper, for the first time, uniform nonBoolean benchmarking is performed for a variety of beyond-CMOS devices based on the CNN architecture. The CNN is a suitable platform for the purpose of benchmarking, because a variety of charge- and spin-based devices can be used to implement CNN circuits efficiently [23]–[25]. Moreover, the mathematical framework for CNN circuits is well defined and understood, which facilitates benchmarking various implementations for a given task and desired accuracy. Furthermore, there has been a great deal of research on both digital and analog implementation of CNN circuits with CMOS devices [26], [27]. The benchmarking in this paper covers three CNN implementations based on analog and digital charge-based switches and spintronic devices. The performance is compared in terms of the energy and delay for a given associative memory application with a certain accuracy target and input noise level. It is crucial to understand and identify the advantage and drawback of each device technology by means of a rigorous and fair benchmarking to guide device researchers to develop a device for optimal circuit performance.

The rest of this paper is organized as follows. Section II describes the design methodology and the modeling approach for three types of CNNs, including the analog, digital, and spintronic implementations. Section III shows the functional demonstration and the benchmarking results, comparing the recall accuracy, energy, and delay for different CNN implementations using various emerging beyond-CMOS devices. The conclusions are drawn in Section IV.

II. CNN CELL IMPLEMENTATIONS AND MODELING APPROACHES

A. OVERVIEW

The CNN is a non-Boolean computing architecture that contains an array of computing cells that are connected to nearby cells. Since interconnects are major limitations in modern VLSI systems, CNN systems can take advantage of the local communication and encounter fewer constraints imposed by interconnects. The CNN can be considered as a brain-inspired computing architecture that relies on neurons to integrate the incoming currents. The accumulated and activated output signal drives nearby neurons through weighted synapses. The underlying mathematics of a CNN was proposed by Chua and Yang [28] and the dynamic state equation of each CNN cell circuit is written as

$$C_f \frac{dx_{ij}}{dt} = -\frac{1}{R_f} x_{ij} + \sum_{kl \in S_{ij}} A_{ij,kl} y_{kl} + \sum_{kl \in S_{ij}} B_{ij,kl} u_{kl} + I_{ij}$$
$$y_{ij} = f(x_{ij}) \quad (1)$$

where x_{ij} is the state voltage of the cell, R_f and C_f are the linear resistance and capacitance of each cell, y_{kl} and u_{kl} are the outputs and inputs of neighboring cells, respectively, $f(x)$ is the sigmoid function that describes the characteristic between the output voltage and cell state voltage, A_{kl} and B_{kl} are templates of each cell, whose values represent the weights of synapses connecting two nearby cells, and I_{ij} is the input bias current of each cell.

B. ANALOG IMPLEMENTATION

Many prior publications have focused on the CNN implemented by analog circuits using CMOS transistors. A widely used implementation is based on operational amplifiers and operational transconductance amplifiers (OTAs) as neurons and synapses, respectively [26], [27]. Some recent work has also investigated CNN using beyond-CMOS charge-based devices, such as TFETs, to potentially improve the energy efficiency [23], [24] thanks to their steep subthreshold slope and low operating voltage.

In this paper, the delay per operation is numerically solved based on CNN dynamics shown in (1), where the linear feedback capacitance, C_f, is the summation of the input and output capacitances of nearby OTAs to reliably sink current, the feedback resistance, R_f, is set as $2/G_m$ to achieve a stable output, where G_m is the summation of the conductance of input synapses. Since the synapses are realized with OTAs [23], the conductance of the synapse is equal to the transconductance of transistor connecting to the input, which is written as

$$g_m = \frac{i_b \ln 10}{\mathrm{SS}} \quad (2)$$

where i_b is the bias current of the device and SS is the subthreshold slope in mV/Dec at the bias current point. In this paper, the bias current of the transistor is set as the geometric mean current of ON and OFF currents, which are adopted from the previous benchmarking work [2], shown in Fig. 1. The transistor width is set as 10 F, where F is the minimum feature size of 15 nm.

Fig. 2 shows the comparison of subthreshold slope and transit frequency among various charge-based device technologies. The subthreshold slope is estimated as the average slope based on the ON and OFF currents, assuming the maximum threshold voltage is 0.3 V. The transit frequency is

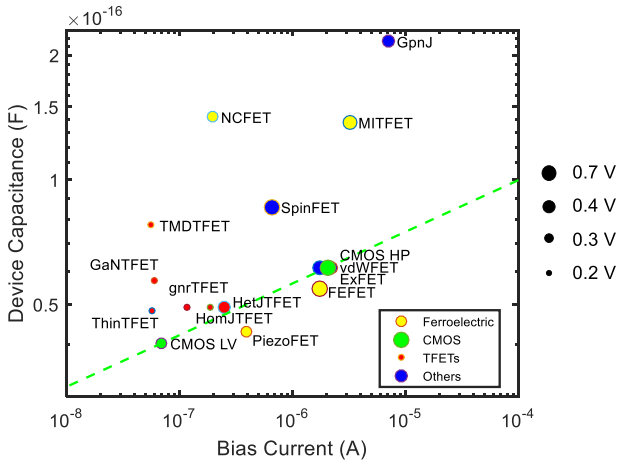

FIGURE 1. Device input capacitance versus the bias current for a variety of charge-based device technologies. Right: size of the circle represents the supply voltage. Results are adopted from [2]. The bottom-right is the preferred corner.

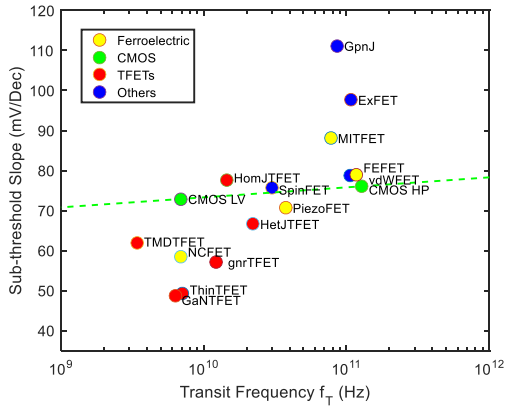

FIGURE 2. Subthreshold slope versus the transit frequency for a variety of charge-based device technologies. The bottom-right is the preferred corner.

defined as $f_T = g_m/2\pi C_o \propto 1/RC$, where C_0 is the input capacitance of the transistor. Since the settling time of the system to reach the equilibrium state is proportional to the RC time constant according to the CNN dynamic equation (1), the inverse of the transit frequency represents the speed of the CNN operation. In general, the TFET has a steeper subthreshold slope compared to the conventional CMOS HP and LV devices. In addition, the TFET operates at a low supply voltage. However, the low ON current limits the bias current, leading to a slow operation speed. The energy dissipation is written as $E = V_{\rm dd} I_b \cdot t_d \propto V_{\rm dd} \cdot {\rm SS} \cdot C_0$, where I_b is the total biasing current of synapses and neurons. The neuron is built with a two-stage differential-input single-ended output 7-transistor operational amplifier [29]. The bias current of the neuron is set as the maximum current from incoming synapses to provide a large driving capability. Based on the analysis above, a device with a large bias current and a small subthreshold slope and capacitance can potentially be faster; a device with a small subthreshold slope, supply voltage, and capacitance may potentially consume less energy. These trends can be observed based on more rigorous numerical simulation results of CNN dynamics shown in Section III.

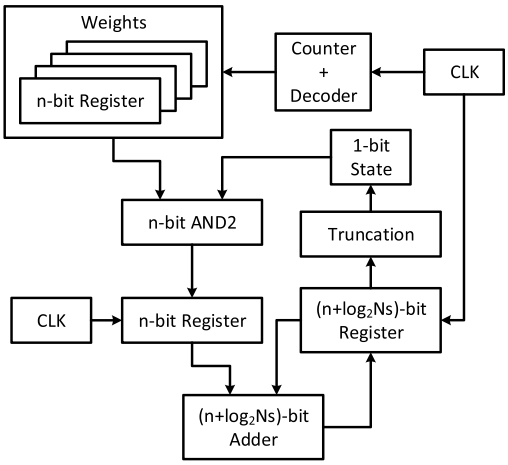

FIGURE 3. Diagram of the CNN cell implementation based on the digital circuit.

C. DIGITAL IMPLEMENTATION

The diagram of the digital CNN cell implementation is illustrated in Fig. 3. The n-b weights of synapses connecting nearby cells are stored in n-b registers. The following operations are performed in each cell in parallel.

1) The counter is activated by the clock and the output of the decoder will select one weight.
2) The weight is multiplied by the corresponding one-bit state from either the current cell or one of the nearby cells. Multiplication is performed by the n-b two-input AND gates, whose outputs are stored in the n-bit register.
3) At the next clock cycle, the weighted state gets added by the $(n + \log_2 {\rm Ns})$-bit adder, and the output is stored in an $(n + \log_2 {\rm Ns})$-bit register, where Ns is the number of synapses connecting to the cell, namely, the number of weights.
4) After all the weighted states are summed, the final weighted state is truncated and updates the one-bit state in the current cell. After all cell states in the CNN system are updated, the system goes to the next time step. This iteration continues until the CNN system reaches the steady state.

For the performance modeling, the delay and energy dissipation of the register, counter, decoder, two-input AND gate, and adder follow the previous benchmarking work for the Boolean logic circuits [2].

D. SPINTRONIC IMPLEMENTATION

In this section, CNN implementations based on three major types of current-driven spintronic devices are investigated, including the spin diffusion, spin Hall effect (SHE), and domain wall devices. Magnet switching dynamics follow the Landau–Lifshitz–Gilbert equation with a spin-transfer-torque

term [30], [31]. The magnetization of a magnet, $\vec{m}$, under a perpendicular spin-polarized current, $\vec{I}_{S,\perp}$, is written as

$$\frac{d\vec{m}}{dt} = -\gamma\mu_0[\vec{m}\times\vec{H}_{\text{eff}}] + \alpha\left[\vec{m}\times\frac{d\vec{m}}{dt}\right] + \frac{\vec{I}_{S,\perp}}{qN_s} \quad (3)$$

where $\vec{H}_{\text{eff}}$ is the effective field, γ is the gyro ratio, μ_0 is the permittivity, α is the damping factor, q is the elementary charge, N_s is the number of magnetons, and $\vec{I}_{S,\perp}$ can be expressed as $i_0 \cdot (\sum_{kl\in S_{ij}} A_{ij,kl}y_{kl} + \sum_{kl\in S_{ij}} B_{ij,kl}u_{kl} + I_{ij})$, where i_0 is the unit spin-polarized current when the template value is unity. The amplitude and the direction of the spin-polarized current depend on the output and input voltage polarities of nearby cells and the weights of synapses connecting those cells.

1) SPIN DIFFUSION-BASED DEVICE

The CNN using spin diffusion-based devices has been investigated thoroughly in a previous study [25]. It relies on the ASL as the basic building block, where PMA magnets are assumed in the simulation. For the CNN benchmarking in this paper, IMA magnets are also included for the spin diffusion-based devices. The CNN design parameters, such as magnet dimensions and material properties, are listed in Table 1.

TABLE 1. Spintronic CNN Design Parameters

Parameters	Values
Minimum Feature Size F (nm)	15
CMOS Driving Voltage V_{read} (V)	0.5
Transistor Resistance @ W = 4F (KΩ)	5
Transistor Capacitance @ W = 4F (aF)	60
Damping Coefficient α	0.01
IMA Saturation Magnetization Ms (A/m)	10^6
PMA Saturation Magnetization Ms (A/m)	3×10^5
Perpendicular Anisotropy Ku (J/m^3)	6×10^4
IMA Magnet Dimension (nm^3)	15×60×2
PMA Magnet Dimension (nm^3)	30×30×2
Ground Resistance (Ω)	50
Spin Hall Angle	0.3
Spin Injection Coefficient β	0.5

2) SPIN HALL EFFECT-BASED DEVICE

Fig. 4(a) shows the schematic of a CNN cell implemented by magnetic tunnel junction (MTJs) as synapses and SHE-based device as the neuron. In this example, each 3-b synapse has two fixed resistances and three MTJs that are digitally programmable to achieve eight different combinations of parallel and antiparallel states. Each combination represents one weight, and overall eight quantized output currents can be generated accordingly, shown in the bar chart in Fig. 4(b). The resistances are adjusted such that the current linearly increases with the weight. Depending on the input voltage polarity, both positive and negative weights can be realized.

During the CNN operation, multiple synapses connecting nearby cells are connected to the neuron, and the net charge

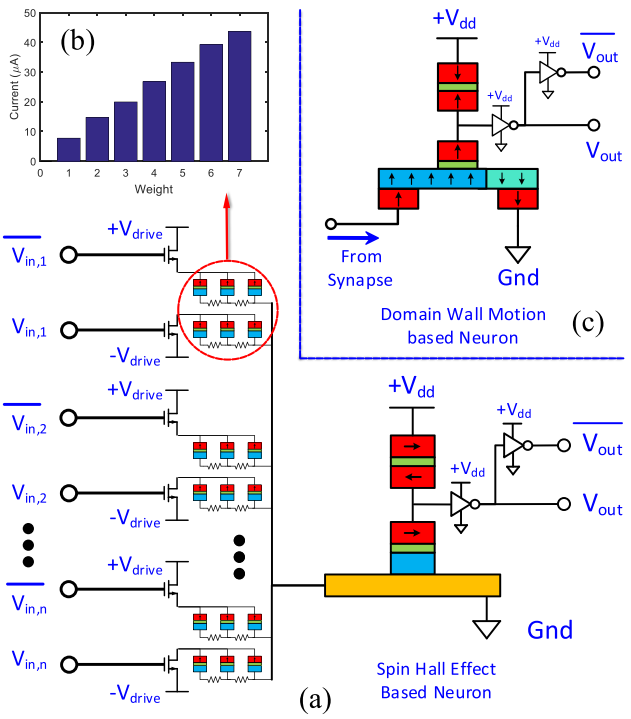

FIGURE 4. Schematics of spintronic CNN implementations. (a) SHE-based neuron with MTJ-based synapses, (b) quantized current versus the weights of the digitally programmable synapses, and (c) domain wall motion based neuron.

current flowing through the SHE material in the neuron is converted into spin currents due to the spin orbit coupling [32]. The spin-polarized current density is $J_s = J_c\theta$, where θ is the spin Hall angle at a value of 0.3 [33], J_c is the charge current density. The dimension of the SHE material is $150 \times 60 \times 2$ nm^3 with a resistance of 60 Ω. To further improve the spin current received in the magnet, recent work has shown that adding an extra layer of the copper plate between the SHE material and the free magnet can enhance the spin injection through the lateral diffusion [34]. For the benchmarking results shown in Section III, an enhancement factor of 2 is considered to show the potential improvement by using the copper collector.

The read-out circuitry is identical to the spin diffusion based CNN by using two MTJs [25]. The voltage at the input of the inverter becomes low and high when the bottom MTJ is at parallel and antiparallel configurations, respectively. The complementary voltages are generated to drive the synapses in nearby cells. The dynamic of the magnet orientation of each cell is numerically solved based on (3).

3) DOMAIN WALL MOTION-BASED DEVICE

Another CNN implementation is based on using a domain wall device as the neuron, shown in Fig. 4(c). By replacing the SHE material with the domain wall magnet, the resistance of the bottom MTJ depends on the position of the domain wall underneath the fixed magnets. For instance, if the input electrons flow to the right direction, the domain wall moves to

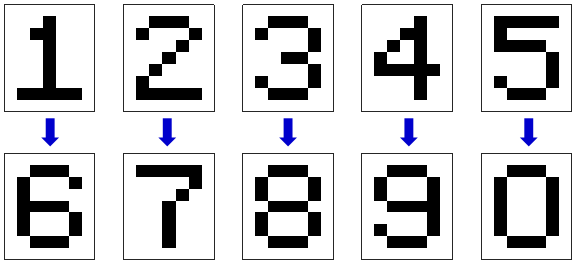

FIGURE 5. **Training patterns of the associative memory application, where digital numbers '1'–'5' are associated with '6'–'0'.**

the right and the bottom MTJ is at the parallel configuration, lowering the voltage at the input of the inverter, and the V_{out} rises to V_{dd}. The domain wall magnet size is $150 \times 30 \times 2$ nm^3 with a resistance of 150 Ω, and the relation between the domain wall speed c_{dw} and the input current density are adopted from [16].

III. BENCHMARKING METHODOLOGY AND RESULTS

A. FUNCTIONAL DEMONSTRATION

The associative memory application is widely used in tasks of voice and image recognition, which can be efficiently performed in the CNN architecture [35], [36]. In this section, three types of CNN implementations are investigated to perform the pattern recall task, shown in Fig. 5. The top five digital numbers, '1' – '5', are associated with the bottom five numbers, '6' – '0'. The training method used for storing patterns is adopted from the Hebbian learning algorithm [37]. It can be applied for a large number of free parameters in a space-varying template used in the associative memory with a fair computational cost and good convergence speed. Once the training is finished, a digital pattern with certain random noisy pixels is set as the input. Here, a spintronic CNN based on spin diffusion as the writing mechanism using PMA magnets is shown in Fig. 6 as an example to demonstrate the functionality of the application. A noisy pattern '1' is used as the input and the output is expected to be associated with the pattern '6'. The simulations are performed at room temperature, and the thermal noise is taken into account. Depending on the random input noise and the thermal noise, the output has a probability of successful recall. In addition, the delay per CNN operation is dependent on the input pattern, input noise, and the thermal noise.

B. ISO-ACCURACY ANALYSIS

Since different CNN implementations may have different accuracies, iso-accuracy analyses are important to achieve a fair benchmark among various types of CNNs. For each input pattern shown in Fig. 5, 100 Monte Carlo simulations are performed for the associate memory application with a given number of random noisy pixels. The recall accuracy is defined as the number of output patterns that completely match with the associated patterns during the training.

Fig. 7 shows the comparison of the recall accuracy versus the number of synapses for four CNN implementations with

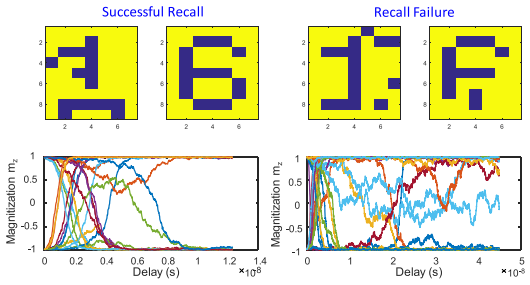

FIGURE 6. **Functional demonstration of a spintronic CNN using ASL devices with PMA magnets as the basic building blocks. (a) Successful recall and (b) failure recall using pattern '1' as the input with 10% noisy pixels.**

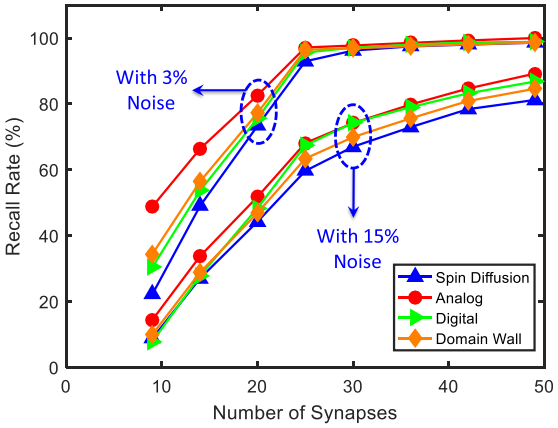

FIGURE 7. **Recall accuracy versus the number of synapses using input patterns with and without noise for four CNN implementations based on analog, digital, PMA, and domain wall devices.**

3% and 15% noisy pixels at the input patterns. By increasing the number of synapses for each neuron, the recall accuracy can be increased significantly. This is because each cell can reach and communicate with more nearby cells and improve the probability of the successful recall. This improvement comes at the cost of a larger footprint area, energy dissipation, and training cost. For a given number of synapses, the recall accuracy differs among various CNN implementations because of the differences in the CNN characteristics, such as the sigmoid function, the dynamic behavior of the magnets, and the feedback of the analog integrators. The analog CNN provides a better recall accuracy for processing input patterns with few noisy pixels. This advantage may come from the unique feedback mechanism of the analog operational amplifier. For the CNN used in this paper, the dynamic of each cell state depends not only on the inputs of the system but also on its own cell and other cell states. This differs from other types of feedforward neural networks without the feedback. For a given recall accuracy of 80%, Fig. 8 shows the required number of synapses for four CNN implementations, and the analog CNN requires fewer synapses compared to other CNNs.

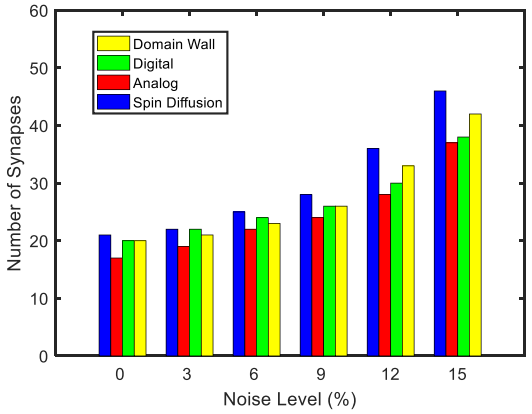

FIGURE 8. **Required number of synapses versus noise levels of the input patterns for four CNN implementations based on analog, digital, PMA, IMA, and domain wall devices at 80% recall accuracy.**

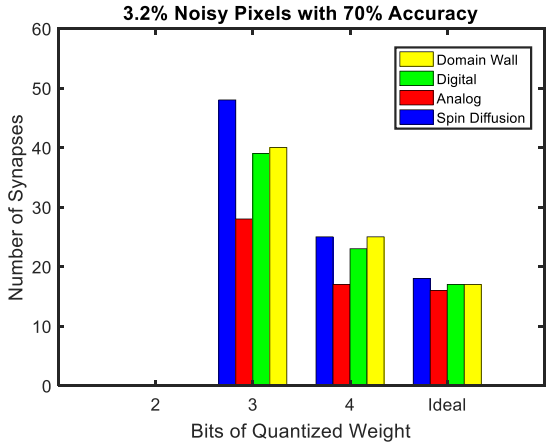

FIGURE 9. **Required number of synapses versus the number of bits used in quantized weights with 3% noisy pixels at the input patterns for four CNN implementations based on analog, digital, spin diffusion, and domain wall devices at 70% recall accuracy.**

The results shown above are based on synapses with ideal weights. To quantify the impact of the finite resolution of synapse weights on the recall accuracy, the numbers of required synapses for different CNN implementations to achieve 70% recall accuracy are shown in Fig. 9. Here, the input noise level is set as 3%. One can observe that there is a tradeoff between the number of synapses and the number of bits representing the weights of synapses. With a 2-b weight, no CNN implementation can reach a 70% accuracy even if all the cells are connected with each other. As the number of bits representing the synapse weight increases, the required number of synapses keeps decreasing. Compared to CNNs using 4-bsynapses, ∼50% more synapses are required for those using 3-b synapses. Therefore, using 4-b weights provides a good tradeoff and also imposes small overheads compared to the ideal weights.

Fig. 10 shows the performance comparison among three types of CNN implementations using 4-b weight synapses to

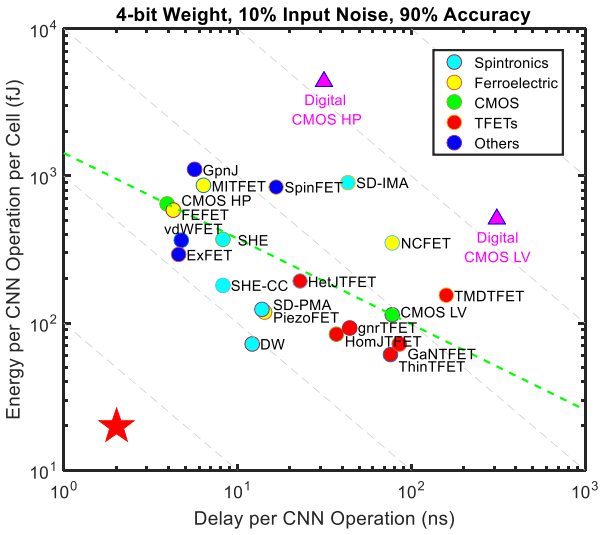

FIGURE 10. **Comparison of energy and delay per operation among various beyond-CMOS technologies based on analog, digital, and spintronic implementations. Triangle and circular points of charge-based devices represent the digital and analog CNN implementation, respectively. For the text labels of spintronic CNN implementation, SD, SHE, and DW stand for spin diffusion, SHE, and domain wall motion, respectively, and CC represents the copper collector.**

achieve 90% recall accuracy for a given input noise of 10%. For the charge-based CNN implementation, CMOS HP and LV devices are employed to quantify the performance of the digital CNN and to compare against their analog counterparts. It is shown that the digital CNNs are quite power hungry and slow. This is because multiple cycles are required to read out the weights from the register and perform the summation in the adder, which is energy- and time- consuming. In general, the analog CNNs implemented by TFETs dissipate less energy thanks to their steep subthreshold slope and lower supply voltage, shown in Figs. 1 and 2.

For devices with an extra ferroelectric switching time, such as FEFET, MITFET, PiezoFET, and NCFET, their relative positions in the energy-delay plot shift largely toward the preferred corner compared to the results shown in the previous Boolean logic benchmarking [2]. The reason is that the extra polarization switching time of the ferroelectric material can be comparable or even larger than the intrinsic switching delays of FETs in a Boolean circuit; however, for the CNN application, the settling time is dominated by the product of the feedback resistance and capacitance.

In contrast to Boolean circuits, spintronic devices are more competitive compared to charge-based devices. This is because a single magnet can mimic the functionality of a neuron, and these spintronic devices operate at a low supply voltage. For the domain wall device, it provides the best performance in terms of the EDP thanks to its low critical current requirement. The spin diffusion-based CNN with IMA magnets consumes more energy due to the large critical current required to switch the magnet.

IV. CONCLUSION

In this paper, a uniform benchmarking methodology is presented for the non-Boolean computation based on the CNN architecture. A variety of beyond-CMOS device technologies, including charge-based and spintronic devices, are compared based on three types of CNN implementations, using analog, digital, and spintronic circuits. The impact of the input noise, thermal noise, and the number of bits of synapses on the system-level recall probability and delay are quantified based on Monte Carlo simulations. As the number of synapses increases, the recall probability improves significantly. The tradeoff between the required number of synapses and the resolution of the synapse weight is quantified, showing the benefit of using 4-b weights. At the end, three types of CNN implementations with various charge- and spin-based devices are benchmarked in a single plot for a given input noise and recall accuracy target. The results demonstrate that TFET-based CNNs, in general, consume less energy thanks to their steep threshold slope and low supply voltage. CNNs implemented by digital CMOS perform worse compared to their analog counterpart due to the large energy and delay from multiplying and adding synapse weights. Spintronic devices are promising candidates for implementing CNN. Up to 3× EDP improvement is observed for the CNN implemented by domain wall devices compared to its conventional CMOS counterpart. The results of this benchmarking are in sharp contrast to those for Boolean functions, such as a 32-b adder or an ALU, where spintronic devices performed significantly worse compared to CMOS devices.

REFERENCES

[1] K. J. Kuhn, "CMOS scaling for the 22nm node and beyond: Device physics and technology," in *Proc. Int. Symp. VLSI Technol. Syst. Appl. (VLSI-TSA)*, 2011, pp. 1–2.

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

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

[4] S. Salahuddin and S. Datta, "Use of negative capacitance to provide voltage amplification for low power nanoscale devices," *Nano Lett.*, vol. 8, no. 2, pp. 405–410, 2007.

[5] A. C. Seabaugh and Q. Zhang, "Low-voltage tunnel transistors for beyond CMOS logic," *Proc. IEEE*, vol. 98, no. 12, pp. 2095–2110, Dec. 2010.

[6] S. Das, A. Prakash, R. Salazar, and J. Appenzeller, "Toward low-power electronics: Tunneling phenomena in transition metal dichalcogenides," *ACS Nano*, vol. 8, no. 2, pp. 1681–1689, Jan. 2014.

[7] M. O. Li, D. Esseni, G. Snider, D. Jena, and H. G. Xing, "Single particle transport in two-dimensional heterojunction interlayer tunneling field effect transistor," *J. Appl. Phys.*, vol. 115, no. 7, p. 074508, 2014.

[8] W. Li *et al.*, "Polarization-engineered III-nitride heterojunction tunnel field-effect transistors," *IEEE J. Explor. Solid-State Computat. Devices Circuits*, vol. 1, no. 1, pp. 28–34, Dec. 2015.

[9] D. Newns, B. Elmegreen, X. H. Liu, and G. Martyna, "A low-voltage high-speed electronic switch based on piezoelectric transduction," *J. Appl. Phys.*, vol. 111, no. 8, p. 084509, 2012.

[10] C. Pan and A. Naeemi, "Device- and system-level performance modeling for graphene P-N junction logic," in *Proc. 13th Int. Symp. Quality Electron. Design (ISQED)*, Mar. 2012, pp. 262–269.

[11] J. Son, S. Rajan, S. Stemmer, and S. J. Allen, "A heterojunction modulation-doped Mott transistor," *J. Appl. Phys.*, vol. 110, no. 8, p. 084503, 2011.

[12] L. Liu, Y. Lu, and J. Guo, "On monolayer field-effect transistors at the scaling limit," *IEEE Trans. Electron Devices*, vol. 60, no. 12, pp. 4133–4139, Oct. 2013.

[13] S. A. Wolf, J. Lu, M. R. Stan, E. Chen, and D. M. Treger, "The promise of nanomagnetics and spintronics for future logic and universal memory," *Proc. IEEE*, vol. 98, no. 12, pp. 2155–2168, Dec. 2010.

[14] 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.

[15] 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.

[16] D. Morris, D. Bromberg, J.-G. J. 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.

[17] H. Markram, "The blue brain project," *Nature Rev. Neurosci.*, vol. 7, no. 2, pp. 153–160, 2006.

[18] P. A. Merolla *et al.*, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, vol. 345, no. 6197, pp. 668–673, Aug. 2014.

[19] D. Monroe, "Neuromorphic computing gets ready for the (really) big time," *Commun. ACM*, vol. 57, no. 6, pp. 13–15, 2014.

[20] S. Venkataramani, A. Ranjan, K. Roy, and A. Raghunathan, "AxNN: Energy-efficient neuromorphic systems using approximate computing," in *Proc. Int. Symp. Low Power Electron. Design*, 2014, pp. 27–32.

[21] 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.

[22] 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.

[23] 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.

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

[25] 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. 1–8, Aug. 2016.

[26] L. Yang, L. O. Chua, and K. R. Krieg, "VLSI implementation of cellular neural networks," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 1990, pp. 2425–2427.

[27] E. Y. Chou, B. J. Sheu, and R. C. Chang, "VLSI design of optimization and image processing cellular neural networks," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 44, no. 1, pp. 12–20, Jan. 1997.

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

[29] P. R. Gray, P. Hurst, R. G. Meyer, and S. Lewis, *Analysis and Design of Analog Integrated Circuits*. Hoboken, NJ, USA: Wiley, 2001.

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

[31] D. V. Berkov and J. Miltat, "Spin-torque driven magnetization dynamics: Micromagnetic modeling," *J. Magn. Magn. Mater.*, vol. 320, no. 7, pp. 1238–1259, 2008.

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

[33] C.-F. Pai, L. Liu, Y. Li, H. W. Tseng, D. C. Ralph, and R. A. Buhrman, "Spin transfer torque devices utilizing the giant spin Hall effect of tungsten," *Appl. Phys. Lett.*, vol. 101, no. 12, p. 122404, 2012.

[34] 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, Jul. 2016.

[35] M. Namba and Z. Zhang, "Cellular neural network for associative memory and its application to Braille image recognition," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2006, pp. 2409–2414.

[36] D. Liu and A. N. Michel, "Cellular neural networks for associative memories," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 40, no. 2, pp. 119–121, Feb. 1993.

[37] P. Szolgay, I. Szatmari, and K. Laszlo, "A fast fixed point learning method to implement associative memory on CNNs," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 44, no. 4, pp. 362–366, Apr. 1997.

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

He is currently a Research Engineer with the Institute for Electronics and Nanotechnology, Georgia Institute of Technology. His 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 Sharif University, Tehran, Iran, in 1994, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Georgia Institute of Technology (Georgia Tech), Atlanta, GA, USA, in 2001 and 2003, respectively.

He was a Research Engineer with the Microelectronics Research Center, Georgia Tech, from 2003 to 2008 and joined the Electrical and Computer Engineering faculty in 2008. He is currently an Associate Professor, exploring nanotechnology solutions to the challenges facing giga- and tera-scale systems.