CMOS-Free Multilayer Perceptron Enabled by Four-Terminal MTJ Device

W. H. Brigner[1], N. Hassan[1], X. Hu[1], C. H. Bennett[2], F. Garcia-Sanchez[3,4], M. Marinella[2], J. A. C. Incorvia[5], J. S. Friedman[1]

[1]Electrical and Computer Engineering, University of Texas at Dallas, Richardson, TX, United States

[2]Sandia National Laboratories, Albuquerque, NM, United States

[3]Department of Applied Physics, University of Salamanca, Salamanca, Salamanca, Spain

[4]Istituto Nazionale di Ricerca Metrologica, Turin, Piedmont, Italy

[5]Electrical and Computer Engineering, University of Texas at Austin, Austin, TX, United States

Abstract—**Neuromorphic computing promises revolutionary improvements over conventional systems for applications that process unstructured information. To fully realize this potential, neuromorphic systems should exploit the biomimetic behavior of emerging nanodevices. In particular, exceptional opportunities are provided by the non-volatility and analog capabilities of spintronic devices. While spintronic devices have previously been proposed that emulate neurons and synapses, complementary metal-oxide-semiconductor (CMOS) devices are required to implement multilayer spintronic perceptron crossbars. This work therefore proposes a new spintronic neuron that enables purely spintronic multilayer perceptrons, eliminating the need for CMOS circuitry and simplifying fabrication.**

Keywords—artificial neural network, neuromorphic computing, leaky integrate-and-fire neuron, multilayer perceptron

I. INTRODUCTION

According to neuroscientists, a brain consists of neurons and synapses. A neuron integrates a series of electrical spikes through its axons and, when enough of these current pulses have been integrated, it releases a spike of its own from the soma (cell body) and through its dendrites to the axons of other neurons. Synapses, on the other hand, provide electrical connectivity between the axons of one neuron and the dendrites of other neurons.

To mimic the brain, neuromorphic systems typically contain artificial neuron and synapse analogs. Such systems can be emulated using software on von Neumann computers [1],[2]; however, due to the fact that complementary metal oxide semiconductor (CMOS) technology, which is the primary technology used in standard von Neumann computers, was not specifically designed to implement the required behavior, these systems consume considerably more energy than an actual human brain [3]. Novel CMOS-based systems can also be used to implement neuromorphic systems [4],[5], but again, CMOS was not specifically designed to match the needs for neuromorphic computing and such chips are highly inefficient.

To resolve the issue of inefficiency and high power consumption, neuromorphic-inspired beyond-CMOS devices have garnered a considerable amount of attention in the scientific community. Since synapses only require non-volatility and a variable resistance, and several beyond-CMOS synapses have already been proposed [6]-[9]. On the other hand, fewer spintronic neurons have been proposed due to the unique challenges in emulating the leaky integrate-and-fire (LIF) neuron model [10]-[12]. While these recently proposed spintronic neurons efficiently provide leaking, integrating, firing, and lateral inhibition, electrical connectivity between the input and output ports results in a need for CMOS circuitry between perceptron layers.

This paper therefore proposes a novel spintronic LIF neuron that enables a monolithic, purely spintronic neuromorphic architecture utilizing previously proposed spintronic synapses. A brief background into the field of neuromorphic computing is provided in section II, and background on spintronic synapses and our previously proposed neurons is presented in section III. Section IV discusses the new spintronic neuron, section V proposes the CMOS-free spintronic multilayer perceptron, and conclusions are provided in section VI.

II. BACKGROUND

Due to the fact that novel technologies and architectures are still required to be compatible with currently existing fabrication techniques, many beyond-CMOS technologies, including the neuromorphic structure proposed in this work, are designed using a crossbar array structure.

A. Crossbar Array

An NxM crossbar array consists of N horizontal wires (word lines) and M vertical wires (bit lines). For a layer in a neuromorphic system, a single input neuron will be attached to each word line and a single output neuron will be attached to each bit line, resulting in N+M total neurons. Likewise, a single synapse exists at each word and bit line intersection to provide connectivity between the input and output neurons for a total of N*M synapses. The resistance states of these synapses, which are capable of being finely tuned during training, determine the weighting between the input and output neurons and, ultimately, the overall functionality of the neuromorphic layer [13]-[16].

B. Leaky Integrate-and-Fire Neuron

A fairly biologically-accurate neuron model is the leaky integrate-and-fire (LIF) neuron model, which is a considerable improvement over the previous integrate-and-fire (IF) neuron [1]. LIF neurons should implement three primary functionalities. As implied by the name, these
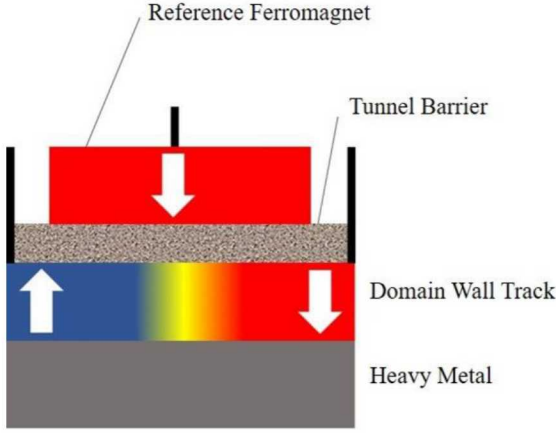
Fig. 1. DW-MTJ synapse with wide tunnel barrier for analog resistance states.

functionalities are integrating, leaking, and firing. When an energy spike is fed into the input of an LIF neuron, the neuron should store the energy from the spike. However, when no energy is applied to the input of the neuron, it should gradually dissipate the stored energy. Finally, once the stored energy has reached a certain threshold, the neuron should suddenly release an output energy spike and reset its state to the initial state.

However, integrating, leaking, and firing are not the only functions an LIF neuron should be able to implement. One such secondary functionality is lateral inhibition, where one neuron that has integrated more than nearby neurons should inhibit integration by its neighbors. This functionality can be extended to implement a winner take all (WTA) system, where a single neuron firing causes a reset of all other neurons, allowing only one neuron to fire for a given data set [10].

III. DOMAIN WALL MAGNETIC TUNNEL JUNCTION

A domain wall magnetic tunnel junction (DW-MTJ) consists of a "free" ferromagnetic layer containing two magnetic domains separated by a domain wall (DW) such that the position of the DW wall determines the MTJ resistance. When the DW, which can be moved using a current, passes underneath, the MTJ switches between the parallel (conductive) and anti-parallel (resistive) states.

A. DW-MTJ Synapse & Neuron Integration

This synapse consists of a long tunnel barrier that covers a large portion of the DW track [8], [9], allowing the device to exhibit analog resistance states (Fig. 1). When a current flows through the DW track, the DW shifts from one end of the track to the other. When used in a DW-MTJ neuron, this DW motion represents integration through an increase in stored energy.

B. Leaking with a DW-MTJ Neuron

Leaking can be achieved using one of three methods. With the first method, a ferromagnet is placed underneath the DW track, as illustrated in Figure 2a. When an external magnetic field is applied parallel to a magnetic domain's magnetization

state, the field causes the domain to expand. Conversely, if the same magnetic field is applied antiparallel to a magnetic domain's magnetization state, the field will cause the domain to shrink. Therefore, a magnetic field applied to a DW track along the same axis as the two antiparallel magnetic domains will cause the DW to shift from one side of the track to the other. The dipolar coupling field provided by the aforementioned ferromagnet allows the DW to shift from one end of the track to the other in the absence of any current. As illustrated in the micromagnetic simulation results of Fig. 3a, when a current is applied from right to left through the DW track, the DW shifts from left to right. However, when no current is applied, the DW gradually shifts from right to left [10].

In the second method, the DW track contains a magnetocrystalline anisotropy gradient, which can be implemented using a thickness and/or composition gradient [17]-[19]. When the DW is in a region of the track with a higher anisotropy value, it exists in a higher energy state, and when the DW is in a region of the track with a lower anisotropy value, it exists in a lower energy state. Therefore, the anisotropy gradient generates an energy landscape that is more favorable to the DW existing on one side of the track than the other. Fig. 3b demonstrates micromagnetic simulation results of the combined leaking and integrating
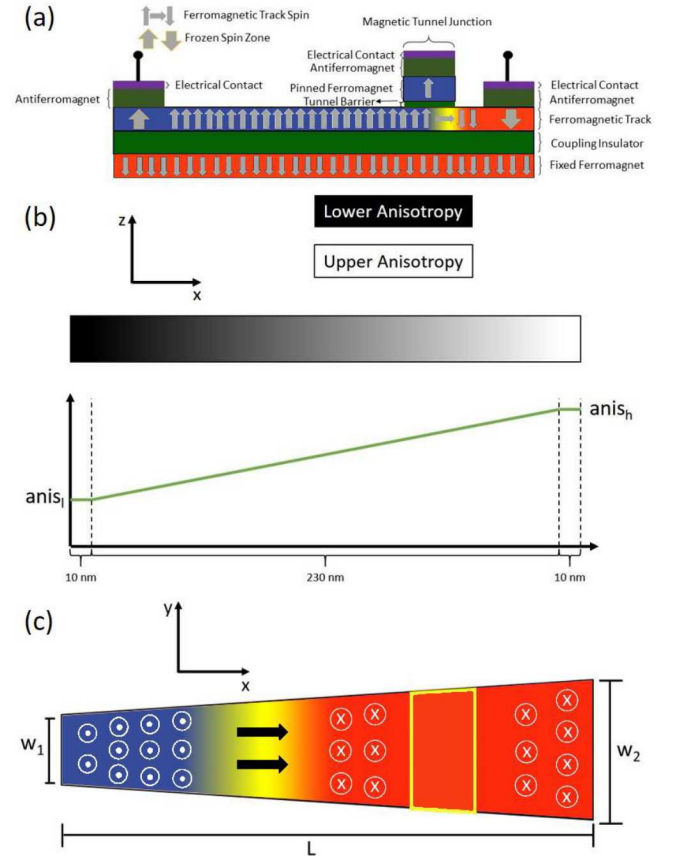

Fig. 2. (a) Side view of the neuron with dipolar coupling field. (b) Side view of the neuron with graded anisotropy illustrating the anisotropy gradient instead of magnetization. (c) Structure of the neuron with shape-based DW drift.
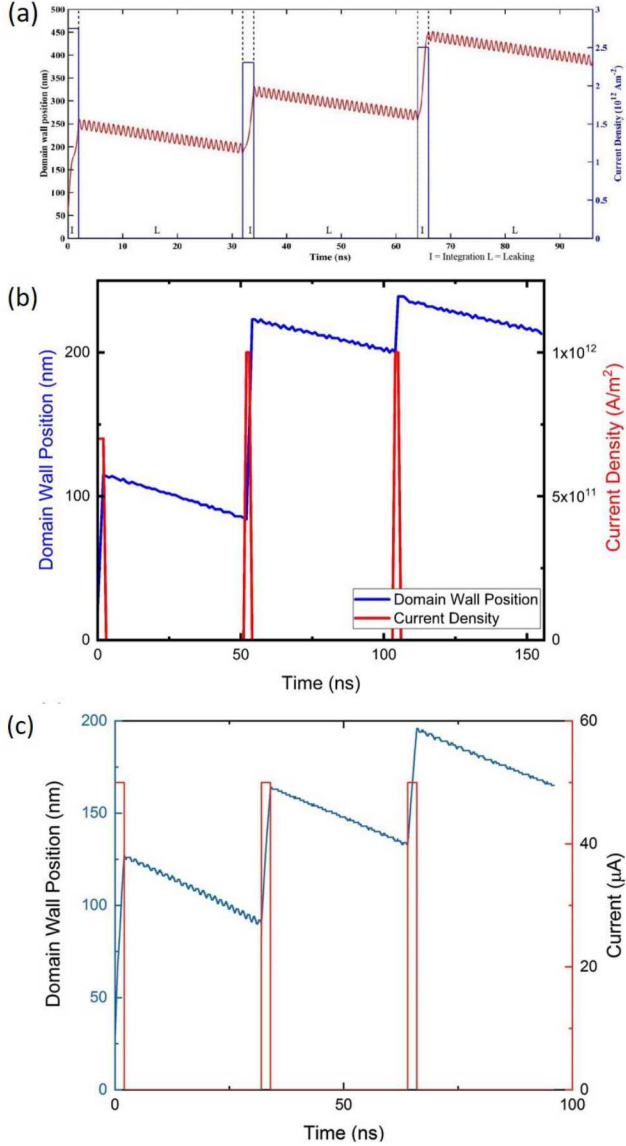
Fig. 3. (a) Combined integrating and leaking characteristics of a neuron implemented using a dipolar coupling field. (b) Combined integrating and leaking characteristics of a neuron with anisotropy gradient. (c) Combined integrating and leaking characteristics of a neuron with shape-based DW drift.

characteristics of a graded anisotropy neuron. As with the neuron with dipolar coupling neuron, a current passed from right to left in the device causes a rapid shift in the DW position from left to right. When no current is applied, the DW slowly leaks from right to left [11].

Finally, with the third method, the sides of the DW track are angled to form a trapezoidal track instead of a rectangular track. Similar to the anisotropy gradient neuron, the DW exists in a higher energy state when it is in the wider region than when it is in the narrower region. Therefore, the DW will shift from the wide portion of the DW track to the narrow portion without any externally-applied stimuli. When applying a current from right to left through the device, as in Fig. 3c, the DW shifts from left to right. When no current is passed

through the device, the energy landscape present in the DW track forces the DW to shift from right to left [12].

IV. FIRING WITH FOUR-TERMINAL DW-MTJ NEURON

In order to avoid electrical connectivity between the input and output ports of the DW-MTJ neuron, an addition free ferromagnetic layer is added in the novel four-terminal DW-MTJ LIF neuron proposed in Fig. 3. When the DW passes underneath the electrically-isolated MTJ, the free layer of the MTJ reorients itself to match the magnetization of the DW track underneath it due to dipole coupling, thereby switching the state of the MTJ. This effect, in combination with a voltage applied at T3, provides the firing functionality without any electrical connectivity between the MTJ and DW track.

V. CMOS-FREE MULTILAYER SPINTRONIC PERCEPTRON

This electrical isolation within the four-terminal DW-MTJ neuron enables the first CMOS-free spintronic multilayer perceptrons, as shown in Fig. 4. While spintronic neural networks have previously been considered [9], significant CMOS circuitry was required in addition to the spintronic synapses and neurons in order to interconnect the layers, implement the leaking, and provide lateral inhibition.

As shown in Fig. 5, the electrical isolation provided by the four-terminal DW-MTJ allows an M×N crossbar layer to be connected to an N×O crossbar layer while maintaining unidirectional signal flow. This CMOS-free spintronic structure can be extended to deep neural networks with numerous layers.

VI. CONCLUSIONS

This work proposes three novel DW-MTJ neurons that, in conjunction with DW-MTJ-based synapses, enable a purely spintronic neural network. This LIF neuron intrinsically provides the leaking, integrating, and firing capabilities, thereby eschewing the need for additional CMOS circuitry. Therefore, we have the capability of implementing monolithic CMOS-free multilayer spintronic perceptrons. By exploiting biomimetic synapses and neurons in a simplified fabrication, this multilayer perceptron concept promises significant advances for efficient machine learning and artificial intelligence.
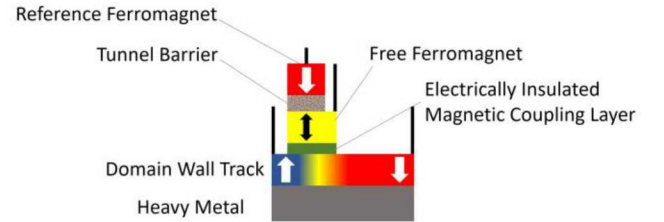

Fig. 4. Schematic of the four-terminal DW-MTJ LIF neuron. When a current flowing from T1 to T2 shifts the DW underneath the MTJ, the magnetization of the free ferromagnet switches, thereby switching the MTJ resistance state. With a constant supply voltage applied at T3, this state change results in increased current output through T4.
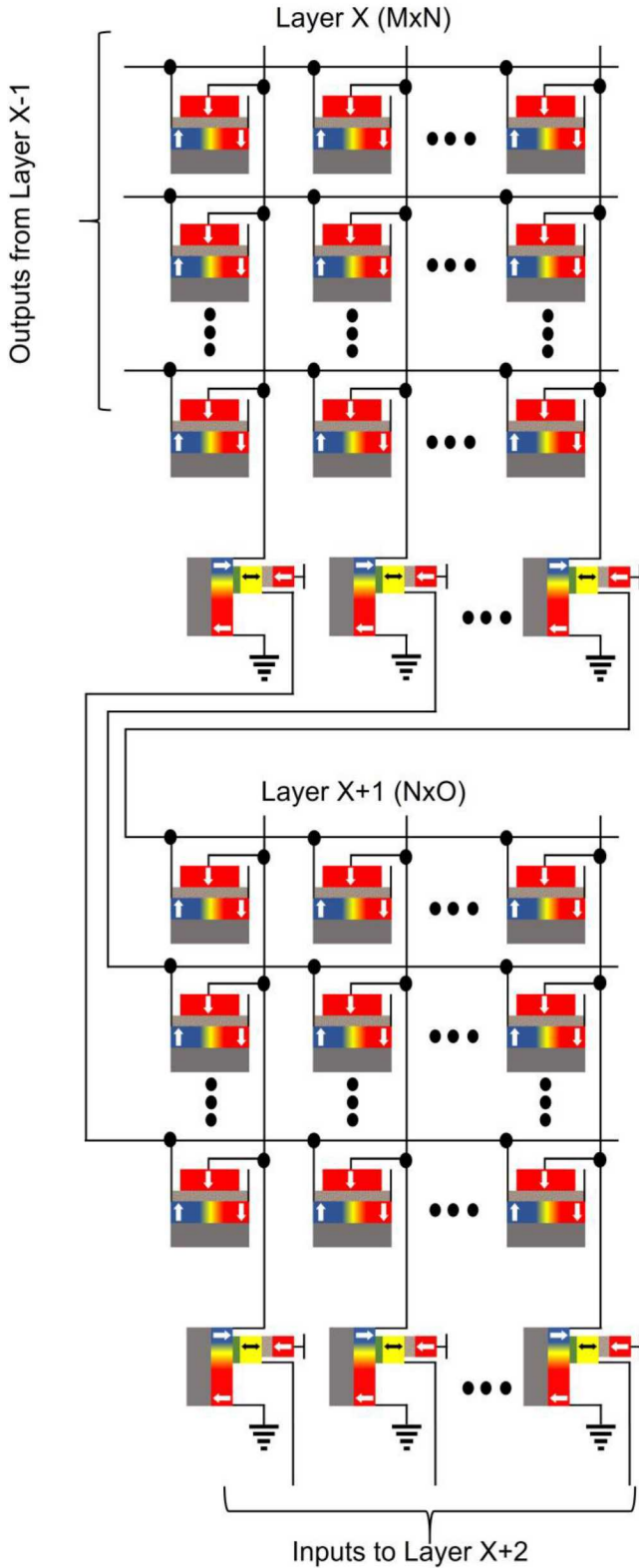
Fig. 5. Multilayer perceptron using the four-terminal DW-MTJ neuron of Fig. 3 and the three-terminal DW-MTJ synapse of Fig. 1. The synapses at the intersections of word and bit lines provide weights between two individual neurons, and the neurons of one layer provide the inputs to another layer.

REFERENCES

[1] Delorme, A., Gautrais, J., Rullen, R. V., Thorpe, S., "SpikeNET: A simulator for modeling large networks of integrate and fire neurons," *Neurocomputing* **26-27**, 989–996 (1999).

[2] Han, S., Mao, H., Dally, W. J., "Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization, and Huffman Coding," *ICLR* 2016.

[3] Sengupta, B., Stemmler, M. B., "Power Consumption During Neuronal Computation," *Proceedings of the IEEE* **102**(5), 738–750 (2014).

[4] Akopyan, F., Sawada, J., Cassidy, A., Alvarez-Icaza, R., Arthur, J., Merolla, P., Imam, N., Nakamura, Y., Datta, P., et al., "TrueNorth: Design and Tool Flow of a 65 mW 1 Million Neuron Programmable Neurosynaptic Chip," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **34**(10), 1537–1557 (2015).

[5] Merolla, P. A., Arthur, J. V., Alvarez-Icaza, R., Cassidy, A. S., Sawada, J., Akopyan, F., Jackson, B. L., Imam, N., Guo, C., et al., "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science* **345**(6197), 668–673 (2014).

[6] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," *Nature*, vol. 453, pp. 80–83, May 2008.

[7] X. Chen, W. Kang, D. Zhu, X. Zhang, N. Lei, Y. Zhang, Y. Zhou, and W. Zhao, "A compact skyrmionic leaky–integrate–fire spiking neuron device," *Nanoscale*, vol. 10, no. 13, pp. 6139–6146, 2018.

[8] S. Dutta, S. A. Siddiqui, F. Buttner, L. Liu, C. A. Ross, and M. A. Baldo, "A logic-in-memory design with 3-terminal magnetic tunnel junction function evaluators for convolutional neural networks," *NANOARCH*, 2017.

[9] A. Sengupta, Y. Shim, and K. Roy, "Proposal for an all-spin artificial neural network: Emulating neural and synaptic functionalities through domain wall motion in ferromagnets," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 6, pp. 1152–1160, 2016.

[10] N. Hassan *et al.*, "Magnetic domain wall neuron with lateral inhibition," *J. Appl. Phys.*, vol. 124, no. 15, p. 152127, Oct. 2018.

[11] W. H. Brigner, X. Hu, N. Hassan, C. H. Bennett, J. A. C. Incorvia, F. Garcia-Sanchez, and J. S. Friedman, "Graded-Anisotropy-Induced Magnetic Domain Wall Drift for an Artificial Spintronic Leaky Integrate-and-Fire Neuron," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, vol. 5, no. 1, pp. 19–24, 2019.

[12] W. H. Brigner, N. Hassan, L. Jiang-Wei, X. Hu, D. Saha, C. H. Bennett, M. J. Marinella, J. A. C. Incorvia, F. Garcia-Sanchez, and J. S. Friedman, "Shape-Based Magnetic Domain Wall Drift for an Artificial Spintronic Leaky Integrate-and-Fire Neuron," *IEEE Transactions on Electron Devices* (accepted).

[13] Lin, Y.-P., Bennett, C. H., Cabaret, T., Vodenicarevic, D., Chabi, D., Querlioz, D., Jousselme, B., Derycke, V., Klein, J.-O., "Physical Realization of a Supervised Learning System Built with Organic Memristive Synapses," *Scientific Reports* **6**(1) (2016).

[14] C. H. Bennett, J. A. C. Incorvia, X. Hu, N. Hassan, J. S. Friedman, M. M. Marinella, "Semi-Supervised Learning and Inference in Domain-Wall Magnetic Tunnel Junction (DW-MTJ) neural Networks," *Proc. SPIE Spintronics XII*, Aug. 2019 (invited).

[15] Sharad, M., Fan, D., Aitken, K., Roy, K., "Energy-Efficient Non-Boolean Computing With Spin Neurons and Resistive Memory," *IEEE Transactions on Nanotechnology* **13**(1), 23–34 (2014).

[16] Sharad, M., Fan, D., Aitken, K., Roy, K., "Energy-Efficient Non-Boolean Computing With Spin Neurons and Resistive Memory," *IEEE Transactions on Nanotechnology* **13**(1), 23–34 (2014).

[17] Y. Sun, C. R. Sullivan, W. Li, D. Kopp, F. Johnson, S. T. Taylor, "Soft magnetic properties of obliquely deposited Co–Zr–O films", *IEEE Trans. Magn.*, vol. 43, no. 12, pp. 4060-4063, Dec. 2007.

[18] N. N. Phuoc, L. T. Hung, C. K. Ong, "FeCoHfN thin films fabricated by co-sputtering with high resonance frequency", *J. Alloys Compounds*, vol. 509, no. 9, pp. 4010-4013, 2011.

[19] S. Li, Z. Huang, J.-G. Duh, M. Yamaguchi, "Ultrahigh-frequency ferromagnetic properties of FeCoHf films deposited by gradient sputtering", *Appl. Phys. Lett.*, vol. 92, no. 9, 2008.