

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

Semi-supervised learning and inference in domain-wall magnetic tunnel junction (DW-MTJ) neural networks

Christopher H. Bennett, Naimul Hassan, Xuan Hu, Jean Anne C. Incornvia, Joseph S. Friedman, et al.

Christopher H. Bennett, Naimul Hassan, Xuan Hu, Jean Anne C. Incornvia, Joseph S. Friedman, Matthew J. Marinella, "Semi-supervised learning and inference in domain-wall magnetic tunnel junction (DW-MTJ) neural networks," Proc. SPIE 11090, Spintronics XII, 110903I (16 September 2019); doi: 10.1117/12.2530308

SPIE.

Event: SPIE Nanoscience + Engineering, 2019, San Diego, California, United States

Semi-supervised learning and inference in domain-wall magnetic tunnel junction (DW-MTJ) neural networks

Christopher H. Bennett^a, Naimul Hassan^b, Xuan Hu^b, Jean Anne C. Incorvia^c, Joseph S. Friedman^b, and Matthew J. Marinella^a

^aSandia National Laboratories, 1515 Eubank Blvd SE, Albuquerque, NM USA 87185

^bUniversity of Texas at Dallas, 800 W Campbell Rd, Richardson, TX, USA 75080

^cUniversity of Texas at Austin, 110 Inner Campus Dr., Austin, TX USA 78705

ABSTRACT

Advances in machine intelligence have sparked interest in hardware accelerators to implement these algorithms, yet embedded electronics have stringent power, area budgets, and speed requirements that may limit non-volatile memory (NVM) integration. In this context, the development of fast nanomagnetic neural networks using minimal training data is attractive. Here, we extend an inference-only proposal using the intrinsic physics of domain-wall MTJ (DW-MTJ) neurons for online learning to implement fully unsupervised pattern recognition operation, using winner-take-all networks that contain either random or plastic synapses (weights). Meanwhile, a read-out layer trains in a supervised fashion. We find our proposed design can approach state-of-the-art success on the task relative to competing memristive neural network proposals, while eliminating much of the area and energy overhead that would typically be required to build the neuronal layers with CMOS devices.

Keywords: spintronics, domain wall devices, magnetic devices, neuromorphic computing, unsupervised learning

1. INTRODUCTION

Presently, conventional/von Neumann computing devices are encountering serious difficulties in locally storing, mapping, and manipulating complex data structures, an ability which may be critical to those systems that cannot rely on external (cloud) processing for energy, distance, or trust/reliability reasons. These difficulties can broadly be divided into issues with continued processor scaling, and memory energy overhead and bandwidth issues, the latter of which is often referred to as the memory wall.¹ However, since these problems arise as a symptom of old computer architecture, the use of novel, brain-inspired (neuromorphic) architectures that exploit existing and modified logic and memory devices remains an alternate path.² In addition to the use of standard CMOS and memory devices to realize neuromorphic prototypes,³ emerging non-volatile memory (NVM) devices have recently come into consideration for in-memory computing (IMC) systems. Neuromorphic/IMC systems built with NVM devices can reach very high density in the crossbar configuration, since the cell size can scale as low as $4F^2$, where F is the base feature size, and can serve as an excellent template for performing hardware implementations of neural network learning and inference, by exploiting local physics and electrical effects.⁴ This combination of form (extreme density) and function (data transformation/analysis) opens a pathway to extreme energy efficiency in future computing systems.

Despite this promise, many types of NVM devices remain at the prototype stage, or suffer from critical issues such as low durability, slow programming/write time, or a high required programming current. Thankfully, magnetic tunnel junction (MTJ) memory devices, including typical two-terminal spin-transfer torque MTJ (STT-MTJ) and emerging three-terminal spin-orbit torque MTJ (SOT-MTJ) devices open a doorway to low current ($< 10\mu A$), fast speed ($< 10ns$), and high endurance ($> 10^{12}$ operations) dense arrays of emerging memory.^{5,6}

In order to fulfill the neuromorphic/IMC vision and perform local or in-situ learning, candidate NVM devices must be programmable according to simplified rules which are derived from either state of the art optimization techniques in machine learning, e.g. stochastic gradient descent and back-propagation, or neuroscience-derived

Send correspondence to C.H.B: cbennet@sandia.gov

rules, such as spike-timing dependent plasticity (STDP). While both approaches have so far been experimentally demonstrated using alternative emerging NVM devices,^{7–9} demonstrations of online learning using spintronic NVM candidates such as SOT-MTJ and STT-MTJ are sparse, or rely on significant CMOS-based circuits to operate.^{10,11} An additional challenge is that, while emerging NVM devices are typically highly analog, MTJ variant devices are typically binary between their Antiparallel (AP) and parallel (P) states. This requires the use of stochastic behaviors as a stand-in for physically analog behavior, and can result in promising results.¹² In addition, spintronic memories are an exciting alternative for future neuromorphic systems since they may avoid some of the signature issues that plague neuromorphic implementations with existing filamentary or phase-change devices, e.g. characteristic non-linearity of write-mode asymmetry relating to physical challenges in moving ions. The combination of non-linearity and asymmetry can be devastating to efficient NVM online learning systems, so finding building blocks where this issue can be obviated is a promising path forward.¹³

We focus on an especially promising spintronic device: a three-terminal domain-wall MTJ (DW-MTJ) devices with intrinsic analog behavior and the potential for attoJoule switching energies.¹⁴ While experimentally demonstrated only in the logic context so far,¹⁵ these devices constitute a possible implementation for not only synapse but neuron in emerging nanomagnetic hardware neural network proposals. Specifically, an array of DW-MTJ devices can be used to implement the essential features of a population or layer of spiking neurons- leak, integrate and firing- along with the inhibition behavior which is critical to unsupervised learning.¹⁶ In this work we consider how effectively a nanomagnetic neural network learning with DW-MTJ output and hidden layers can perform on a state of the art machine learning task by deriving these learned weights *in-situ*. Specifically, we consider a dual-phase , or semi-supervised, approach to online learning. In this approach, an unsupervised system/layer first adapts and is later co-integrated with a read-out system. Both systems can be physically realized using an adaptive crossbar system with emerging memory devices and realized at terabit or higher density scales for efficient edge computing applications. In the following sections, we detail this design and preliminary results when using it on a standard task.

2. DW-MTJ DEVICE FULLY ONLINE LEARNING SYSTEM

2.1 Realistic neuronal emulation by DW-MTJ nanotracks

As described in,¹⁴ the motion of a domain-wall logic or memory devices are physically modeled and explained by the spin-polarization of injected current in a soft ferromagnetic wire which sits between an input and output MTJ device; antiferromagnetic contacts placed on the ends of the wire create an exchange bias which pins the magnetic field at either end and drives domain wall motion back towards the Input MTJ, or forward towards the Output MTJ. This complex physical system can be notably simplified by noting that the current must be above a certain critical threshold I_{th} in a certain direction to drive the DW 'forward', or 'backward'.

Recently, the DW-MTJ device was further engineered with the addition of a heavy ferromagnetic underlayer such that, when no current is applied, the position (x) of the DW will tend to naturally relax towards the Input MTJ.¹⁶ In this same work the ability of transverse magnetic fields from a given DW-MTJs track to inhibit its neighbors was explored and confirmed via micromagnetic simulations. Taken together, the operations of leak, integrate and fire along with inhibition create a physical basis for realistic emulation of winner-take-all circuits. Besides an underlayer, an alternative physical method to implement intrinsic leaking of DW-MTJs is the use of a gradient anisotropy.¹⁷

2.2 Realizing fully online learning DW-MTJ neural networks

2.2.1 First (un-supervised) layer

Winner-take-all circuits are a staple of neural computing, and it has been shown that soft-winner-take all gates, *i.e.* more than one 'winning' or spiking neuron per layer/population, can serve as a foundation for universal function approximation.¹⁸ However, when WTA networks use plastic weight updates governed by the biological learning rule STDP, winner take all circuits have been demonstrated to possess powerful emergent computational properties such as approximating hidden Markov models.¹⁹ In addition, the use of plastic updates and lateral inhibition in the context of vector-matrix multiplies and subsequent updates in a dense memory array has been suggested to implement the powerful algorithm known as Expectation Maximization (EM).²⁰ Due to these

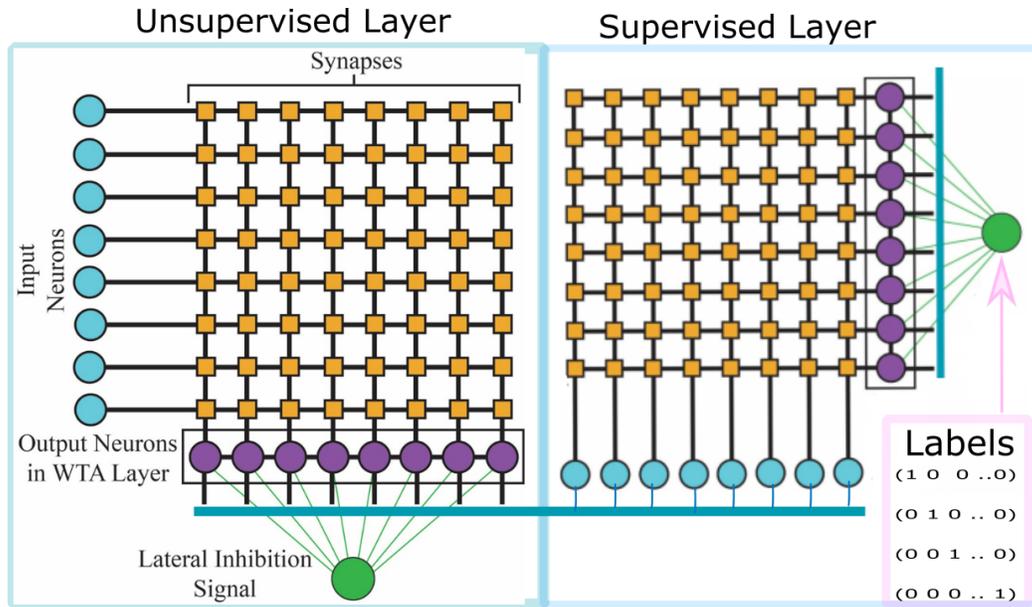


Figure 1. The two systems, which can both be conveniently realized in a shared nanofabric, both use DW-MTJ devices at the hidden and output layer, notably using the lateral inhibition function to implement the k-WTA property and the max-out property, respectively. The yellow crosspoints, here representing adaptive synapses, can in principle be any two-terminal NVM device. For the first layer, we have considered them as both analog filamentary and binary magnetic devices; for the second layer, we have considered them as only analog devices.

mutually powerful effects, we use a first layer of DW-MTJs learning in two cases: with un-adaptive synapses but active neuron effects still in place (random weights), and with plastically adaptive synapses along with the active neuron effects (STDP). In addition, we have considered two variants of STDP to determine the variety of plastic updates needed to implement this additional computational power. On one hand, we considered a binary approximation of STDP (B-STDP), which following¹² relies upon the naturally stochastic behavior of two-terminal STT-MTJ junctions, and moves NVM devices in the first layer between only extrema states (P and AP, in the case of STT-MTJ devices). Second, analog STDP (A-STDP), in which multi-bit analog NVM devices switch between multiple states, as proposed in²¹ and realized experimentally in.²² As in these references, we have assumed the naturally exponential (non-linear) nature of the device's state changes.

2.2.2 Second (supervised) layer

As noted in,²³ such an unsupervised layer can be difficult to properly read-out in the context of standard machine learning tasks, creating the need for a companion supervised learning system. Although the companion supervised learning system does not increase the computational capacity of the network, it serves as a bridge to implement an easily interpretable threshold-based learning system. Following hardware realizations that have realized efficient sign-based implementation of the classic delta or Widrow-Hoff rule,²⁴ we realize weights in pairs of NVM devices, and adjust weights at every weight pair according to:

$$\Delta W_{i,k} = \Delta G \text{sign}(X_i(T_k - O_k)), \quad (1)$$

where $\Delta W_{i,k}$ is the weight change at each programming step for the synapse connecting input i to output k , T_k is the expected output, O_k the actual output, X_i the input value, and ΔG the conductance change. The teacher signal, in the case of our network are the labels for the actual image class (*e.g.* digit '1') during the teaching/training moment, but may in principle be any real-time signal directing the system's adaptation.

3. SIMULATION METHODS AND RESULTS

3.1 Description of dual learning operation simulations

In our simulations, we have co-integrated a model of the interacting DW-MTJs during the un-supervised learning as a time-step simulator using forward Euler integration, along with a simplified model of lateral inhibition, with $t = 200$ time steps during which the first set of parallel DW-MTJs, DW_M operate based on the input currents provided at that moment. At the end of the last time step, the spike(s) are collected by the simulator and passed forward to implement the plasticity operation (if it is being implemented in that first network W_{us}). Following this operation, another example is presented and fed to the DW-MTJs, for a total number of L_{us} unsupervised learning samples. After this, the weights of the first memory matrix/crossbar are frozen and remain so during the second phase. During the second phase of the dual learning approach, training inputs are again fed to the input network (W_{us}), spikes are collected from the DW-MTJ interaction sub-simulator, and these spikes are propagated to the second network (W_s) in a logical sign-symmetric manner as suggested in.²⁵ As pictured in Fig. 2(a), the output from this second crossbar is used to produce a set of spike outputs from the second set of DW-MTJs DW_N , and collected in a logic cell or local memory. Subsequently, these spike outputs in comparison to a teaching signal is then used to implement the on-chip learning rule. Meanwhile, as in 2(b), the programming occurs via simultaneous voltage pulses which must bypass the DW-MTJ devices, in the case of the post-synaptic pulse. This operation occurs a total of L_s times using the frozen weights already obtained in the first phase.

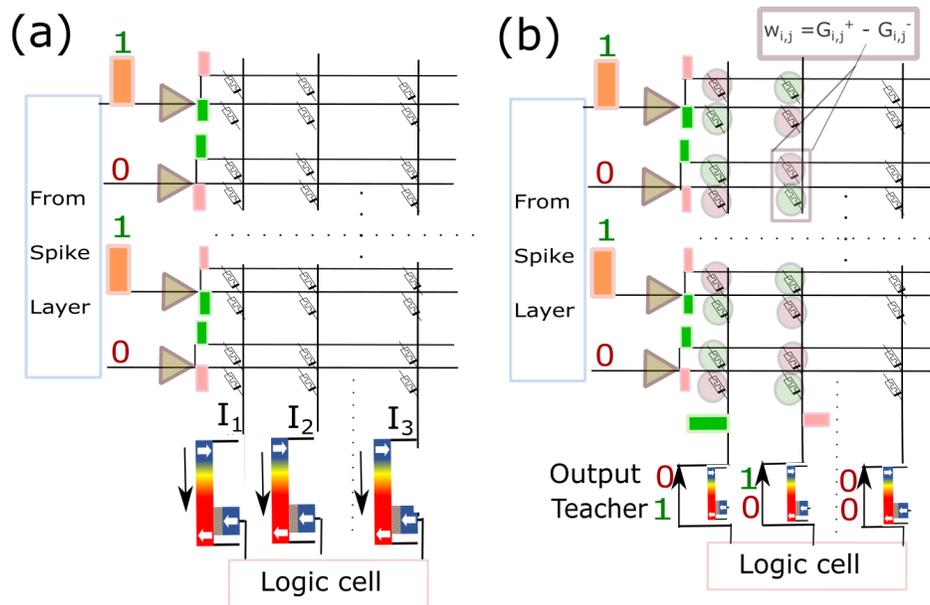


Figure 2. (a) depicts the vector-matrix multiply (VMM) and max-out operation of W_s , and (b) depicts the physical mapping of a learning rule (binary Widrow-Hoff algorithm). Note that, the actual comparison of the output and teacher signal must be stored and compared in a logic cell made of standard CMOS, or logic domain-wall devices, and can't be made all at once due to the staged operation.

3.2 Preliminary simulation results

We have tested this procedure and our simulator on the classic MNIST task of handwritten digits. Our preliminary simulations suggest that with $DW_M = 200$, that is with 200 competing hidden units, and with $DW_M = 10$ as in¹⁶ at the final output layer, we achieve 88% accuracy on the test-set with plastic binary synapses and 92% with plastic analog synapses, following $L_{us} = 500$ and then $L_s = 50000$. In order to verify that the weight adaptations in the unsupervised layer are of value, we also set the first layer weights as constant (that is, skipped the first of the dual phases mentioned earlier). Considering random binary and analog devices, we achieve 78% and 79% respectively using this approach; this result unambiguously confirms the value of lateral inhibition combined with plastic updates.

However, regarding the value of analog v. binary plastic updates, we remain ambivalent about whether our early results suggest that one is definitely better than the other. Since the binary updates rely upon stochastic effects to perform properly, we hope to better benchmark whether different probability switching distributions on each device may help to close this gap. Nevertheless, at the moment we can observe notably richer analog filters in the first layer as compared to binary ones, as in Fig. 3. Lastly, we have eliminated the first layer entirely and found that a perceptron with max-out DW-MTJs and following the weight updates as in Fig. 2 achieve 84% on the test set, an intermediate result between the random weight input layer and the plastic input layer.

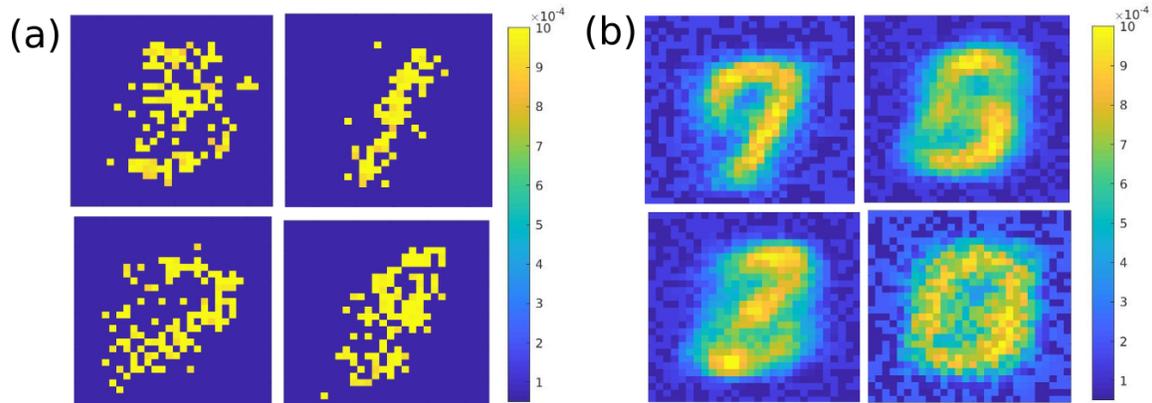


Figure 3. (a) shows several hidden neuron’s weight representations, reconstructed as a 28x28 grid, for the binary STDP updates case, while (b) shows the same but for the analog update case. The colorbars and color coding depicts physical conductance values between the minima and maximum extrema, not the logical weight values.

4. DISCUSSION

4.1 Comparison to state-of-the-art approaches

While^{10,11} previously proposed nanomagnetic neural networks incorporating realistic leaky-integrate-fire and STDP learning approaches, respectively, neither incorporated the inhibition effect which is critical to k-WTA learning. By combining STDP and k-WTA and demonstrating that these two effects strictly outperform either alone, our work inspires a new avenue of research to explore the maximum computational power possible with locally learning arrays. Meanwhile, in comparison to other proposals combining WTA and stochasticity in the context of non-magnetic memristive/NVM devices,²⁶ we have achieved unambiguously better numerical results on the task while also massively reducing CMOS reliance at the neuron device level.²⁶ While not a state-of-the-art result on the ML task, our approach yields a strong performance when compared to other competing neural network schemes using spintronics devices even using fully feature backprop. For instance, this performance is superior to various in-situ learning styles with two-terminal magnetic devices alone.²⁷ Beyond this, we hope to achieve state-of-the-art by increasing the small number of DW-MTJ tracks operating at the hidden layer, or by demonstrating that hidden layers of DW-MTJs may be stacked in order to approach performances derived using standard convolutional neural networks, e.g. > 99% on the test-set.²⁸

4.2 Lateral inhibition and physical constraints

Consistent with,¹⁶ our simulations calculate the lateral inhibition assuming a small inter-DW-MTJ distance ΔD . A parameter analysis suggests that when this distance is expanded and the magnetic lateral force B_F exerted upon neighbor wires is too low, too many DW-MTJs devices tend to spike. This degrades the system’s performance during both inference and k-WTA operation; at the worst case almost every other neuron will spike, and performance is very poor. While we calculated a favorable value for ΔD and hence B_F which resulted in on average on 30% or less of the neurons in hidden layer DW_M firing each example, a priority is to integrate more physically realistic modeling into the neural solver to see if this is realistic. Optimizing our system to allow for a large enough B_F value may be the most challenging task in later physically realizing a version of this system.

5. CONCLUSION

Domain-wall magnetic tunnel junction (DW-MTJ) are an extremely promising platform for neuromorphic computing as they can be used at the synapse and neuron level, have rich analog behavior, and a very low switching energy requirement. When co-integrated in an ultra-dense array, they can stand in for biorealistic populations of neurons by implementing the key functions of leak, integrate and fire (LIF). We have demonstrated that beyond inference or prediction, this property can be used in the online learning context to implement a powerful semi-supervised learning strategy. This strategy, which uses both general plastic updates in the first system/layer as well as specifically directed updates in the second system/layer, has already achieved promising results compared to other semi-supervised and/or spintronic neuromorphic proposals.

ACKNOWLEDGMENTS

C. H. Bennett and M. Marinella acknowledge support from Sandias Laboratory-Directed Research and Development program. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energys National Nuclear Security Administration under contract DE-NA0003525. This research is also sponsored in part by the National Science Foundation under CCF award 1910800.

REFERENCES

- [1] Horowitz, M., “1.1 computing’s energy problem (and what we can do about it),” in [2014 *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*], 10–14, IEEE (2014).
- [2] Furber, S., “Large-scale neuromorphic computing systems,” *Journal of neural engineering* **13**(5), 051001 (2016).
- [3] Akopyan, F., Sawada, J., Cassidy, A., Alvarez-Icaza, R., Arthur, J., Merolla, P., Imam, N., Nakamura, Y., Datta, P., Nam, G.-J., 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).
- [4] Burr, G. W., Shelby, R. M., Sebastian, A., Kim, S., Kim, S., Sidler, S., Virwani, K., Ishii, M., Narayanan, P., Fumarola, A., et al., “Neuromorphic computing using non-volatile memory,” *Advances in Physics: X* **2**(1), 89–124 (2017).
- [5] Khalili Amiri, P., Zeng, Z., Langer, J., Zhao, H., Rowlands, G., Chen, Y.-J., Krivorotov, I., Wang, J.-P., Jiang, H., Katine, J., et al., “Switching current reduction using perpendicular anisotropy in cofeb–mgo magnetic tunnel junctions,” *Applied Physics Letters* **98**(11), 112507 (2011).
- [6] Cubukcu, M., Boule, O., Drouard, M., Garello, K., Onur Avci, C., Mihai Miron, I., Langer, J., Ocker, B., Gambardella, P., and Gaudin, G., “Spin-orbit torque magnetization switching of a three-terminal perpendicular magnetic tunnel junction,” *Applied Physics Letters* **104**(4), 042406 (2014).
- [7] Covi, E., Brivio, S., Serb, A., Prodromakis, T., Fanciulli, M., and Spiga, S., “Analog memristive synapse in spiking networks implementing unsupervised learning,” *Frontiers in neuroscience* **10**, 482 (2016).
- [8] Prezioso, M., Kataeva, I., Merrih-Bayat, F., Hoskins, B., Adam, G., Sota, T., Likharev, K., and Strukov, D., “Modeling and implementation of firing-rate neuromorphic-network classifiers with bilayer pt/al2o3/tio2-x/pt memristors,” in [2015 *IEEE International Electron Devices Meeting (IEDM)*], 17–4, IEEE (2015).
- [9] Burr, G. W., Shelby, R. M., Sidler, S., Di Nolfo, C., Jang, J., Boybat, I., Shenoy, R. S., Narayanan, P., Virwani, K., Giacometti, E. U., et al., “Experimental demonstration and tolerancing of a large-scale neural network (165 000 synapses) using phase-change memory as the synaptic weight element,” *IEEE Transactions on Electron Devices* **62**(11), 3498–3507 (2015).
- [10] Srinivasan, G., Sengupta, A., and Roy, K., “Magnetic tunnel junction based long-term short-term stochastic synapse for a spiking neural network with on-chip stdp learning,” *Scientific reports* **6**, 29545 (2016).

- [11] Sengupta, A., Shim, Y., and Roy, K., "Proposal for an all-spin artificial neural network: Emulating neural and synaptic functionalities through domain wall motion in ferromagnets," *IEEE transactions on biomedical circuits and systems* **10**(6), 1152–1160 (2016).
- [12] Vincent, A. F., Larroque, J., Locatelli, N., Romdhane, N. B., Bichler, O., Gamrat, C., Zhao, W. S., Klein, J.-O., Galdin-Retailleau, S., and Querlioz, D., "Spin-transfer torque magnetic memory as a stochastic memristive synapse for neuromorphic systems," *IEEE transactions on biomedical circuits and systems* **9**(2), 166–174 (2015).
- [13] Agarwal, S., Plimpton, S. J., Hughart, D. R., Hsia, A. H., Richter, I., Cox, J. A., James, C. D., and Marinella, M. J., "Resistive memory device requirements for a neural algorithm accelerator," in [2016 International Joint Conference on Neural Networks (IJCNN)], 929–938, IEEE (2016).
- [14] Currivan, J. A., Jang, Y., Mascaro, M. D., Baldo, M. A., and Ross, C. A., "Low energy magnetic domain wall logic in short, narrow, ferromagnetic wires," *IEEE Magnetics Letters* **3**, 3000104–3000104 (2012).
- [15] Currivan-Incorvia, J., Siddiqui, S., Dutta, S., Evarts, E., Zhang, J., Bono, D., Ross, C., and Baldo, M., "Logic circuit prototypes for three-terminal magnetic tunnel junctions with mobile domain walls," *Nature communications* **7**, 10275 (2016).
- [16] Hassan, N., Hu, X., Jiang-Wei, L., Brigner, W. H., Akinola, O. G., Garcia-Sanchez, F., Pasquale, M., Bennett, C. H., Incorvia, J. A. C., and Friedman, J. S., "Magnetic domain wall neuron with lateral inhibition," *Journal of Applied Physics* **124**(15), 152127 (2018).
- [17] Brigner, W. H., Hu, X., Hassan, N., Bennett, C. H., Incorvia, J. A. C., Garcia-Sanchez, F., and Friedman, J. S., "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* **5**(1), 19–24 (2019).
- [18] Maass, W., "On the computational power of winner-take-all," *Neural computation* **12**(11), 2519–2535 (2000).
- [19] Kappel, D., Nessler, B., and Maass, W., "Stdp installs in winner-take-all circuits an online approximation to hidden markov model learning," *PLoS computational biology* **10**(3), e1003511 (2014).
- [20] Querlioz, D., Bichler, O., Vincent, A. F., and Gamrat, C., "Bioinspired programming of memory devices for implementing an inference engine," *Proceedings of the IEEE* **103**(8), 1398–1416 (2015).
- [21] Querlioz, D., Bichler, O., Dollfus, P., and Gamrat, C., "Immunity to device variations in a spiking neural network with memristive nanodevices," *IEEE Transactions on Nanotechnology* **12**(3), 288–295 (2013).
- [22] Serb, A., Bill, J., Khiat, A., Berdan, R., Legenstein, R., and Prodromakis, T., "Unsupervised learning in probabilistic neural networks with multi-state metal-oxide memristive synapses," *Nature communications* **7**, 12611 (2016).
- [23] Querlioz, D., Zhao, W., Dollfus, P., Klein, J.-O., Bichler, O., and Gamrat, C., "Bioinspired networks with nanoscale memristive devices that combine the unsupervised and supervised learning approaches," in [2012 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)], 203–210, IEEE (2012).
- [24] Lin, Y.-P., Bennett, C. H., Cabaret, T., Vodenicarevic, D., Chabi, D., Querlioz, D., Jousset, B., Derycke, V., and Klein, J.-O., "Physical realization of a supervised learning system built with organic memristive synapses," *Scientific reports* **6**, 31932 (2016).
- [25] Bennett, C. H., Parmar, V., Calvet, L. E., Klein, J.-O., Suri, M., Marinella, M. J., and Querlioz, D., "Contrasting advantages of learning with random weights and backpropagation in non-volatile memory neural networks," *IEEE Access* (2019).
- [26] Al-Shedivat, M., Naous, R., Cauwenberghs, G., and Salama, K. N., "Memristors empower spiking neurons with stochasticity," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* **5**(2), 242–253 (2015).
- [27] Mondal, A. and Srivastava, A., "In situ stochastic training of mtj crossbars with machine learning algorithms," *ACM Journal on Emerging Technologies in Computing Systems (JETC)* **15**(2), 16 (2019).
- [28] Ciresan, D. C., Meier, U., Gambardella, L. M., and Schmidhuber, J., "Convolutional neural network committees for handwritten character classification," in [2011 International Conference on Document Analysis and Recognition], 1135–1139, IEEE (2011).