Applied Physics Letters

RESEARCH ARTICLE | JUNE 27 2023

Stochastic domain wall-magnetic tu neurons for noise-resilient spikin

Thomas L. Gonon; Shaamruden Haritharan; Jean Anne C. Incorvia

Appl. Phys. Lett. 122, 262406 (2023)
https://doi.org/10.1063/5.0152211

Applied Physics Letters ARTICLE pubs.aip.org/aip/apl

Stochastic domain wall-magnetic tunnel junction artificial neurons for noise-resilient spiking neural networks

Cite as: Appl. Phys. Lett. **122**, 262406 (2023); doi: 10.1063/5.0152211
Submitted: 28 March 2023 · Accepted: 10 June 2023 ·
Published Online: 27 June 2023

Thomas Leonard, Samuel Liu, Harrison Jin, and Jean Anne C. Incorvia[a]

AFFILIATIONS

Electrical and Computer Engineering, University of Texas at Austin, Austin, Texas 78712, USA

[a]Author to whom correspondence should be addressed: incorvia@austin.utexas.edu

ABSTRACT

The spatiotemporal nature of neuronal behavior in spiking neural networks (SNNs) makes SNNs promising for edge applications that require high energy efficiency. To realize SNNs in hardware, spintronic neuron implementations can bring advantages of scalability and energy efficiency. Domain wall (DW)-based magnetic tunnel junction (MTJ) devices are well suited for probabilistic neural networks given their intrinsic integrate-and-fire behavior with tunable stochasticity. Here, we present a scaled DW-MTJ neuron with voltage-dependent firing probability. The measured behavior was used to simulate a SNN that attains accuracy during learning compared to an equivalent, but more complicated, multi-weight DW-MTJ device. The validation accuracy during training was also shown to be comparable to an ideal leaky integrate and fire device. However, during inference, the binary DW-MTJ neuron outperformed the other devices after Gaussian noise was introduced to the Fashion-MNIST classification task. This work shows that DW-MTJ devices can be used to construct noise-resilient networks suitable for neuromorphic computing on the edge.

Neuromorphic computing allows for real time analysis of unstructured information that is too complex for conventional computing. Current digital implementations of neuromorphic computing rely on large numbers of CMOS transistors to simulate artificial neural networks (ANNs), which are inefficient in terms of both area and energy.[1] For edge computing, which can benefit greatly from energy efficient implementations, analog approaches to mapping synaptic weights and neuronal activations, e.g., crossbar arrays of emerging memory devices, are promising solutions. Emerging devices can represent biological neurons and synapses on a one-to-one basis to exploit the computational dynamics found in the human brain.[2–7] Among them, spintronic devices offer a compact and energy efficient hardware-based building block for alternatives to digitally simulated neural networks on conventional CMOS devices.[8]

Several types of magnetic devices have been proposed to emulate spiking neurons. These approaches are often focused on emulating the leaky integrate and fire (LIF) model of neuron behavior via gradual switching using Hall bars fabricated on the scale of tens of micrometers.[9–12] One such approach has incorporated the use of a domain wall (DW) spin texture, however, with unfeasible area and latency.[12] Additionally, magnetic tunnel junctions (MTJs) are necessary for sufficiently large electrical readout signals. Area-scalable approaches for spintronic neurons have been proposed using alternative methods such as perpendicular magnetic tunnel junctions (MTJs)[13] and skyrmions,[14–16] where the devices operation energy reduces as size reduces and device function can be maintained down to CMOS-compatible sizes, but have not been experimentally demonstrated. A promising approach is to integrate MTJ readout with DW spin textures to represent the neuron state. These DW-MTJ devices offer a highly tunable way to emulate many neuromorphic functionalities on a monolithic platform. We have shown multi-weight (MW) artificial synapses with a single MTJ as well as directional synapses with tunable metaplasticity by varying the geometry of the DW track.[2,17] In addition, DW-MTJ devices are suited to emulate biological neurons given their intrinsic integrate-and-fire behavior, with the membrane potential represented by DW position. As current pulses are applied to the DW track, the DW can be moved using spin transfer torque (STT) or spin–orbit torque (SOT). After a certain number of pulses, the DW will pass under the MTJ, causing a resistance change, which is able to represent a neuronal spike.[18] The ability of the DW-MTJ artificial neurons to incorporate other behaviors, such as leaking,[19,20] lateral inhibition,[21,22] and edgy-relaxed neuronal behavior,[23] has also been

simulated, as well as simulations of full neural network area.[24,25] The potential of having MW synapses and LIF neurons on a monolithic platform can lead to hardware neural networks that incorporate advantages of magnetic materials for edge computing, namely, radiation hardness, speed, and energy efficiency.

Here, we propose a binary DW-MTJ artificial neuron that has stochastic behavior and is fabricated from the same material as previously demonstrated metaplastic MW synapses,[2] opening the possibility for spiking neural networks (SNNs) with fully spintronic matrix operations, where both neuronal and synaptic functionalities are captured using DW-MTJ devices. The device, shown in Fig. 1, consists of a rectangular DW track with a free magnetic switching layer and a tantalum heavy metal layer beneath. The heavy metal layer allows for efficient SOT switching. At one end of the track lies an MTJ consisting of a thin MgO tunnel barrier and a synthetic antiferromagnet (SAF) pinned magnetic layer. The MTJ is positioned sufficiently far down the track to allow integration before firing. A bias field is used to assist the SOT switching to overcome the SAF pinning, but this could be removed via stack engineering. Furthermore, the device is reset using an external magnetic field after each fire event, but this could be removed by tuning the DW track geometry to facilitate intrinsic leaking. To facilitate faster resetting in practical device implementation without modifying the track dimensions, a resetting external magnetic field or a direct resetting electrical pulse can be delivered to the device after a firing event is detected, which can switch the device on the order of nanoseconds. The material stack was grown using physical vapor deposition and patterned using electron beam lithography and ion beam etching, identically to Ref. 2. We show stable resistance states and repeatable cycle-to-cycle switching. Stochastic switching is demonstrated, and switching probability is shown to be dependent on the voltage pulse amplitude. The switching probability as a function of voltage amplitude is then used to form a lookup table to simulate the dynamics of the stochastic neurons on the network level using the Norse spiking neural network framework.[26] Our online learning results show that the binary neurons perform slightly worse than the LIF baseline and the MW neuron network. However, the difference between the MW device and the proposed binary neuron becomes negligible if the hidden layer size or the number of timesteps are reduced. This suggests that the binary stochastic neuron has application in edge computing, where space and energy consumption are highly constrained. This becomes clear when the same networks are used to perform inference. Using Fashion-MNIST with Gaussian noise, we show that the binary neuron is more noise-resilient than both the MW device and the LIF baseline.

Figure 1(a) shows a schematic of the studied binary DW-MTJ artificial neuron. In comparison with the logic devices in Ref. 18, the difference is the relative location of the MTJ along the DW track. This is shown in the scanning electron microscope (SEM) image [Fig. 1(b)]. The final device is 6 μm long and 0.35 μm wide, with a 0.35 μm diameter MTJ. For a feature size F, the minimal area requirement of the device is roughly $18F^2$.[27] The material stack, detailed in supplementary material Fig. S1, and the fabrication process were the same as Refs. 2 and 18. The device operates using spin–orbit-torque (SOT) DW motion using the tantalum underlayer, and the resistance of the MTJ depends on the location of the DW within the track since the top layer of the MTJ is pinned by a synthetic antiferromagnetic layer. Two lithographically patterned notches on either side of the MTJ stabilize the DW position in one of two places along the track. To operate the device, a voltage pulse is applied from CLK to IN to nucleate, depin, and drive the DW down the track. The resistance of the MTJ is then read from IN to OUT. The Oersted (Oe) line was not used during the testing since DW nucleation occurs probabilistically during writing, and the Oe line reduces stochasticity. Sending a voltage pulse along the Oersted line generates a highly localized magnetic field due to Ampère's law, which can nucleate a DW in the notch directly below. This has been shown previously in Ref. 18 to reduce cycle-to-cycle variation by roughly an order of magnitude and achieve more deterministic behavior. Since the Oe Line is not used in this work, it is likely that the DW is nucleated in a random location either before, in, or past the notch. This greatly contributes to the stochasticity in writing.

The binary DW-MTJ stochastic neuron data are shown in Fig. 2. The high resistance antiparallel state and the low resistance parallel state are shown in the minor field loop of Fig. 2(a). The device was set to the parallel state using an external magnetic field of −200 Oe ($\sim -\hat{z}$ direction). After saturation, a bias field of 350 Oe ($\sim \hat{z}$ direction) was used. These field strengths were constant for all experiments. The bias field was not strong enough to switch the device by itself as shown in the field loop [Fig. 2(a)], but it reduced the voltage requirement to

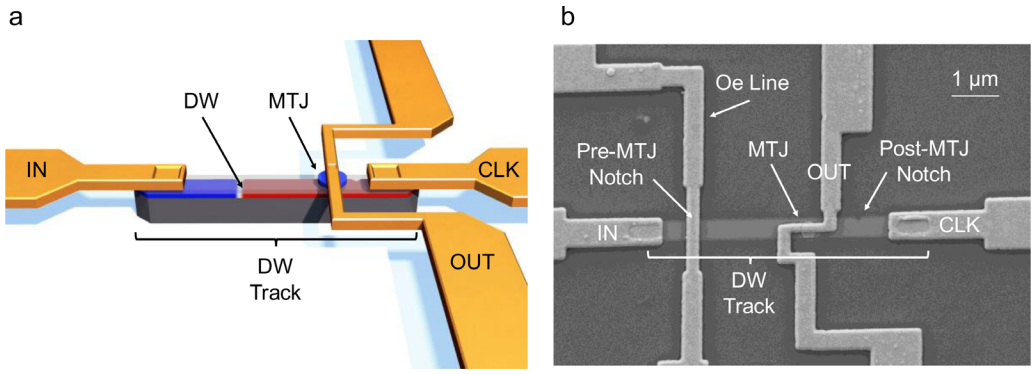

FIG. 1. Device schematic. (a) 3D rendering of the binary DW-MTJ stochastic neuron. Blue/red represent up/down magnetic domains with white DW. Gray represents Tantalum underlayer. The three critical terminals are IN, CLK, and OUT. (b) SEM image of the binary DW-MTJ stochastic neuron. The post-MTJ notch is visible; the pre-MTJ notch is under the Oe line.

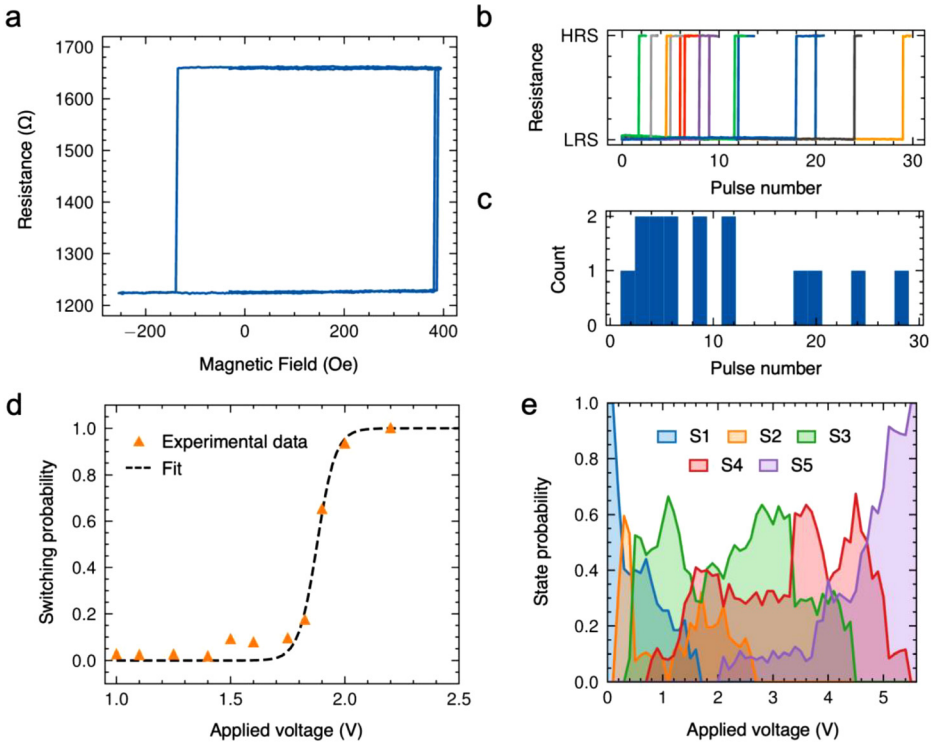

FIG. 2. Experimental results. (a) Binary neuron MTJ resistance (two-point resistance from IN to OUT) as function of out-of-plane magnetic field. (b) MTJ resistance vs pulse number for constant pulse amplitude of 1.75 V repeatably applied from CLK to IN for 15 cycles. (c) Count of number of times each pulse number switches the binary DW-MTJ neuron, for 1.75 V constant amplitude. (d) Switching probability vs applied voltage for respective (b) and (c) data obtained for 11 voltage amplitudes. (e) Analogous measured data for a MW DW-MTJ neuron from Ref. 2 which has five lithographic notches and hence a probability in being in one of five states, S1–S5, for a given applied volage pulse amplitude.

move the DW using SOT. This bias field could be removed through film stack optimization.[28] After the bias field was set, constant voltage pulses were applied from CLK to IN to nucleate, depin, and move the DW down the track. The pulse lengths were fixed at 30 ns in length with 10 ns rise/fall time. Figure 2(b) shows switching data for 15 cycles at a single voltage amplitude of 1.75 V. The switching behavior is highly stochastic since the Oersted line is not used to nucleate the DW within the pre-MTJ notch each cycle, allowing the DW to be nucleated in a random position using current pulses along the DW track. This cycling experiment was repeated for 11 different voltage amplitudes between 1.0 and 2.2 V (supplementary material Fig. S2). After cycling the device, the data were tabulated as the number of times each pulse number switched the device for that voltage amplitude, shown in Fig. 2(c) for 1.75 V. The switching probability per pulse for each voltage amplitude is plotted in Fig. 2(d), showing a sigmoid fit.

The energy dissipation required for a write pulse can be estimated from this distribution. For a switching probability of around 50%, 1.8 V is applied across the heavy metal wire, which has a two-point resistance of 1029 Ω. The four-point resistance of the device is 365 Ω, indicating that 0.64 V is dissipated across the DW track. As a result, performing the energy calculation $E = \int_0^t (V^2/R)\,dt$ with a 30 ns pulse and a 10 ns rise/fall time, the estimated energy per sample is calculated to be around 44.9 pJ. This energy is competitive with digital static random-access memory (SRAM) implementations of spiking neurons at greatly reduced area and without standby energy dissipation from leakage current.[29,30] To further decrease the energy dissipation, shorter pulse length voltage pulses can be used. Additionally, the DW-MTJ neurons benefit from scaling, where the threshold voltage would decrease proportionally to a decrease in width of the DW track.[31,32]

Figure 2(e) is data from our previous measurements of a multi-weight (MW) DW-MTJ device shown in Ref. 2 that is used as a comparison to the binary DW-MTJ measured in Figs. 2(a)–2(d). Because the binary device presented here only has two states, integration is not emulated in device operation by the artificial neuron. The MW device was operated as a synapse in Ref. 2 [schematic and SEM image in supplementary material Figs. S3(b) and S3(c)], but the device can also be used as a neuron by reducing the size of the MTJ to cover only a small region just before the last notch, shown in supplementary material Fig. S3(d). The measured synapse device data are used to characterize the behavior of the hidden state of the simulated neuron, thereby emulating integration behavior. The MW DW-MTJ was tested using ramping voltages from 0 to 5.5 V for ten cycles, and the voltage-dependent probability that the device is in any of the five states is tabulated. The results show that the MW DW-MTJ device combines quantized integration behavior with stochastic switching, showing the flexibility of the device platform.

In this section, we show how the binary stochastic DW-MTJ neuron can be used to build noise-resilient SNNs for training and inference on the edge. The architecture of the network is a multilayer perceptron (MLP) with two hidden layers with the stochastic neuron activations, shown in Fig. 3(a). The output layer is a leaky integration layer to convert spike trains to scalar values, allowing for supervised learning through backpropagation to occur. The role of the two hidden layers is to increase the learning capacity and complexity of the neural network by adding neurons and synapses in the form of the two layers labeled "Dense" in Fig. 3(a). To regulate the spiking frequency through each layer, a batch normalization layer is added after each hidden layer. The SNN is implemented in the Norse framework[26] with custom

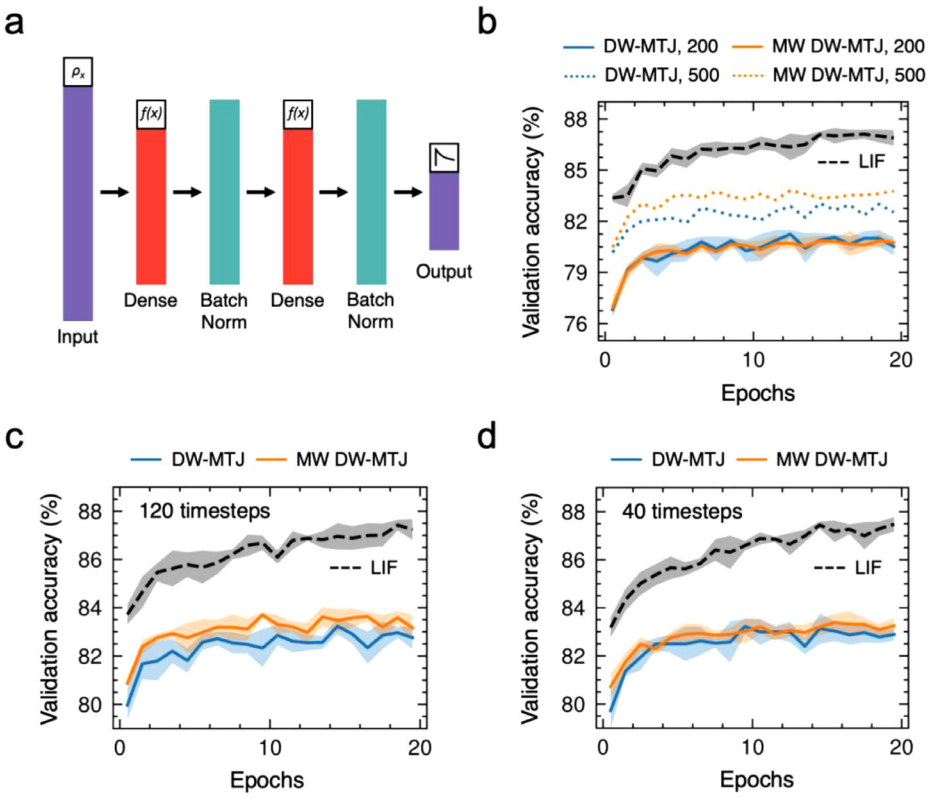

FIG. 3. Spiking neural network architecture and training performance on Fashion-MNIST. (a) Schematic of multilayer perceptron with two hidden layers. (b) Validation accuracy of networks consisting of binary DW-MTJ (blue) and MW DW-MTJ (orange) stochastic neurons for hidden layer sizes of 200 units (solid line) and 500 units (dotted line). LIF neuron SNN validation accuracy is also shown (dashed black line). Clouds depict the standard deviation of accuracy across five seeds. Validation accuracy of three network types with 500 hidden units trained for (c) 120 timesteps and (d) 40 timesteps per image shown.

neuron modules using lookup tables consisting of the experimental data collected from the binary DW-MTJ [Fig. 2(d)] and MW DW-MTJ [Fig. 2(e)] stochastic neurons. To isolate the performance impact of the neuron types, synaptic connections were modeled as ideal double-precision weights. To demonstrate supervised training performance, the spiking MLP was applied on the Fashion-MNIST clothing article classification task.[33] The Fashion-MNIST dataset was encoded in the frequency domain using a Poisson spiking encoding scheme, limiting the maximum spiking frequency of the brightest pixels to the sampling speed of the network, chosen to be 10 MHz. The error function of the network was calculated by taking the maximum value across the timesteps of each output channel; then, a softmax was applied to normalize. The Adam optimizer[34] was used to calculate the weight updates, using a learning rate of 0.001. The training was done with a batch size of 100 images.

Several networks with different activations were evaluated to compare the training performance of the binary DW-MTJ and MW DW-MTJ stochastic neurons, compared to a baseline of an ideal LIF neuron. The validation accuracy figure of merit is the classification accuracy of the 10 000 images in the validation set, distinct from the 60 000 images in the training set. Figure 3(b) shows SNNs consisting of the three types of neurons with 200 (solid lines) and 500 units (dotted lines) per hidden layer. The network was sampled for 80 timesteps per image. The training performance results for the baseline LIF network are of a network with 500 units per hidden layer, though reducing the number of units per hidden layer to 200 did not significantly affect the validation accuracy. Because the MW DW-MTJ had multiple levels, a form of quantized integrate-and-fire behavior was emulated. The extra levels added complexity to the activation function, allowing the MW DW-MTJ neuron to slightly outperform the simpler binary DW-MTJ neuron when the network had 500 units per hidden layer. However, in the network with 200 units per hidden layer, there was no significant difference in training performance between the two networks, suggesting that the expressivity of the extra levels in the MW DW-MTJ device does not significantly improve training performance for smaller networks. This indicates that the simpler fabrication of the binary DW-MTJ neuron can still be desirable for resource-limited edge computing.

This is further corroborated by the validation accuracy results where networks with 500 hidden units per layer were sampled for 120 [Fig. 3(c)] and 40 [Fig. 3(d)] timesteps. In the network sampled for 120 timesteps, the MW DW-MTJ device network slightly outperformed the binary DW-MTJ in terms of validation accuracy. However, when the network was sampled for 40 timesteps, both types of DW-MTJ devices had similar validation accuracy. As a result, for smaller networks sampled with less timesteps, it can be beneficial to reduce fabrication complexity by choosing the binary DW-MTJ instead of the MW device. Due to the small feature size of the notches, choosing the binary DW-MTJ neurons can also result in favorable scalability.

It is well known in the literature that when noise is introduced during the training phase, the resulting network is less affected by noisy input data.[35] This is because the optimization algorithm will find a network state that will minimize the error while random noise is

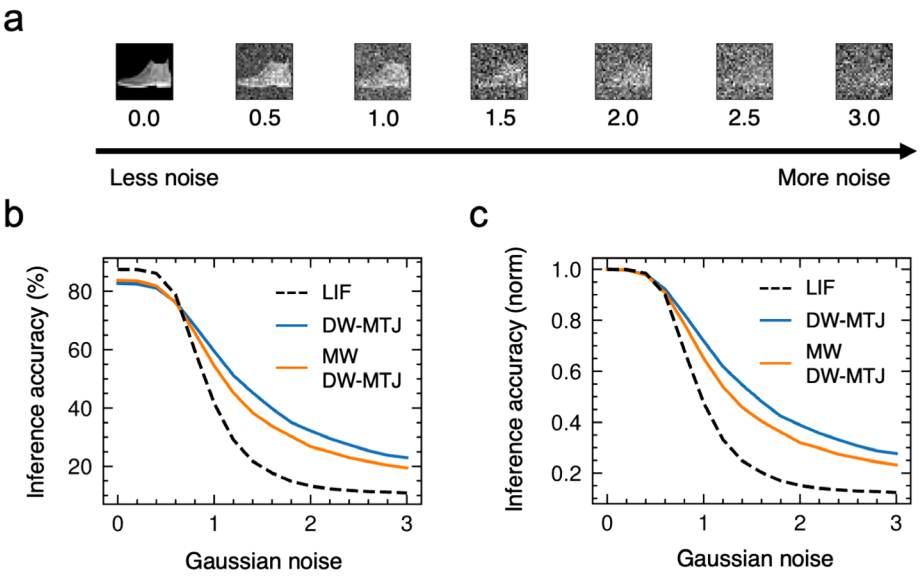

FIG. 4. Inference accuracy on Fashion-MNIST data with (a) added Gaussian noise. (b) Raw and (c) normalized test accuracy for added Gaussian noise with standard deviation from 0 to 3.

included during training. Due to the probabilistic nature of switching, both types of DW-MTJ neurons are hypothesized to be able to fulfill the same functionality of increasing the noise-resilience of a network. To characterize this, varying degrees of Gaussian noise were added to the 10 000 images in the Fashion-MNIST test set, shown in Fig. 4(a). Inference is then performed on the noisy dataset using the highest accuracy models of each type trained on the non-noisy dataset in the previous paragraph. The results in Fig. 4(b) show that while the LIF-based network maintained a higher starting accuracy than the other two networks, there was a sharp drop off at a Gaussian noise magnitude of 0.7 to below that of the stochastic DW-MTJ neurons. This point is also where the inference accuracy of the network composed of MW DW-MTJ devices dropped below that of the simpler binary DW-MTJ devices. At a Gaussian noise magnitude of 3.0, the LIF network accuracy became insignificant compared to a random guess (∼10%) while both DW-MTJ device-based networks are still able to accomplish inference at a better accuracy than a random guess. The advantage of the DW-MTJ stochastic neurons becomes clearer when viewing the data normalized to the accuracy of the networks on pristine data [Fig. 4(c)]. In practical implementation, this noise resilience may not be an acceptable trade-off for accuracy. However, in deep neural networks for edge applications, a promising implementation is to integrate stochastic neurons in one layer to introduce noise-resilience while maintaining high classification accuracy through LIF neurons in other layers.

In summary, scaled DW-MTJ artificial neurons were fabricated, and stochastic switching behavior was characterized. The resulting distribution of spiking probability as a function of input voltage for the binary DW-MTJ device was approximately sigmoidal, with a low energy dissipation of 44.9 pJ per sample for an individual device. An alternative binarized device architecture was also proposed, with multiple notches to implement quantized integration states. The experimental data for both devices was then mapped into lookup tables to simulate supervised learning of a MLP for SNNs. Online learning results show that while there was a slight (∼1%) advantage in validation accuracy to training with the MW DW-MTJ due to the additional expressivity enabled by the extra notches. However, at smaller network sizes and fewer time-steps, the binary DW-MTJ device can match the performance of the MW DW-MTJ device. When performing inference on noisy datasets, the DW-MTJ devices were able to outperform the ideal LIF neurons due to the introduced noise resilience in networks that were trained using the stochastic devices. Where the LIF neuron network inference accuracy rapidly decayed with increasing noise in the dataset, the networks composed of the DW-MTJ neurons of both types were able to maintain reasonable performance above that of the LIF network. These results indicate that the proposed devices are well suited for energy efficient, resource-limited hardware for neuromorphic computing at the edge, making a further case for analog deep neural network (DNN) accelerators with spintronic weights and activations.

See the supplementary material for thin film stack information, additional experimental cycling data, and a detailed description of the MW-DW-MTJ device used in the neural network simulation.

The authors acknowledge research support from the National Science Foundation under CCF Award No. 1910997 as well as support from the National Science Foundation graduate research fellowship program under Grant Nos. 2020307514 and 2021311125. The work was done at the Texas Nanofabrication Facility supported by NSF Grant No. NNCI-1542159 and at the Texas Materials Institute (TMI). The authors acknowledge computing resources from the Texas Advanced Computing Center (TACC) at the University of Texas at Austin (http://www.tacc.utexas.edu).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Thomas Leonard and Samuel Liu contributed equally to this work.

Thomas Leonard: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal). **Samuel Liu:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Software (lead); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Harrison Jin:** Investigation (supporting); Methodology (supporting); Software (supporting). **Jean Anne C. Incorvia:** Investigation (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]V. Milo, G. Malavena, C. M. Compagnoni, and D. Ielmini, "Memristive and CMOS devices for neuromorphic computing," Materials **13**, 166 (2020).
[2]T. Leonard, S. Liu, M. Alamdar, H. Jin, C. Cui, O. G. Akinola, L. Xue, T. P. Xiao, J. S. Friedman, M. J. Marinella, C. H. Bennett, and J. A. C. Incorvia, "Shape-dependent multi-weight magnetic artificial synapses for neuromorphic computing," Adv. Electron. Mater. **8**, 2200563 (2022).
[3]D. Kireev, S. Liu, H. Jin, T. Patrick Xiao, C. H. Bennett, D. Akinwande, and J. A. C. Incorvia, "Metaplastic and energy-efficient biocompatible graphene artificial synaptic transistors for enhanced accuracy neuromorphic computing," Nat. Commun. **13**, 4386 (2022).
[4]G. Zhou, Z. Wang, B. Sun, F. Zhou, L. Sun, H. Zhao, X. Hu, X. Peng, J. Yan, H. Wang, W. Wang, J. Li, B. Yan, D. Kuang, Y. Wang, L. Wang, and S. Duan, "Volatile and nonvolatile memristive devices for neuromorphic computing," Adv. Electron. Mater. **8**, 2101127 (2022).
[5]S. Kumar, X. Wang, J. P. Strachan, Y. Yang, and W. D. Lu, "Dynamical memristors for higher-complexity neuromorphic computing," Nat. Rev. Mater. **7**, 575–591 (2022).
[6]D. Ielmini and H.-S. P. Wong, "In-memory computing with resistive switching devices," Nat. Electron. **1**, 333–343 (2018).
[7]S. Liu, T. P. Xiao, J. Kwon, B. J. Debusschere, S. Agarwal, J. A. C. Incorvia, and C. H. Bennett, "Bayesian neural networks using magnetic tunnel junction-based probabilistic in-memory computing," Front. Nanotechnol. **4**, 1021943 (2022).
[8]J. Grollier, D. Querlioz, K. Y. Camsari, K. Everschor-Sitte, S. Fukami, and M. D. Stiles, "Neuromorphic spintronics," Nat. Electron. **3**, 360–370 (2020).
[9]Q. Yang, R. Mishra, Y. Cen, G. Shi, R. Sharma, X. Fong, and H. Yang, "Spintronic integrate-fire-reset neuron with stochasticity for neuromorphic computing," Nano Lett. **22**, 8437–8444 (2022).
[10]S. Yang, J. Shin, T. Kim, K. W. Moon, J. Kim, G. Jang, D. S. Hyeon, J. Yang, C. Hwang, Y. J. Jeong, and J. P. Hong, "Integrated neuromorphic computing networks by artificial spin synapses and spin neurons," NPG Asia Mater. **13**, 11 (2021).
[11]C. Wang, C. Lee, and K. Roy, "Noise resilient leaky integrate-and-fire neurons based on multi-domain spintronic devices," Sci. Rep. **12**, 8361 (2022).
[12]D. Wang, R. Tang, H. Lin, L. Liu, N. Xu, Y. Sun, X. Zhao, Z. Wang, D. Wang, Z. Mai, Y. Zhou, N. Gao, C. Song, L. Zhu, T. Wu, M. Liu, and G. Xing, "Spintronic leaky-integrate-fire spiking neurons with self-reset and winner-takes-all for neuromorphic computing," Nat. Commun. **14**, 1068 (2023).
[13]A. Sengupta, P. Panda, P. Wijesinghe, Y. Kim, and K. Roy, "Magnetic tunnel junction mimics stochastic cortical spiking neurons," Sci. Rep. **6**, 30039 (2016).
[14]P. Jadaun, C. Cui, S. Liu, and J. A. C. Incorvia, "Adaptive cognition implemented with a context-aware and flexible neuron for next-generation artificial intelligence," PNAS Nexus **1**, pgac206 (2022).
[15]M. A. Azam, D. Bhattacharya, D. Querlioz, and J. Atulasimha, "Resonate and fire neuron with fixed magnetic skyrmions," J. Appl. Phys. **124**, 152122 (2018).
[16]H. Vakili, J.-W. Xu, W. Zhou, M. N. Sakib, M. G. Morshed, T. Hartnett, Y. Quessab, K. Litzius, C. T. Ma, S. Ganguly, M. R. Stan, P. V. Balachandran, G. S. D. Beach, S. J. Poon, A. D. Kent, and A. W. Ghosh, "Skyrmionics—Computing and memory technologies based on topological excitations in magnets," J. Appl. Phys. **130**, 070908 (2021).
[17]S. Liu, T. P. Xiao, C. Cui, J. A. C. Incorvia, C. H. Bennett, and M. J. Marinella, "A domain wall-magnetic tunnel junction artificial synapse with notched geometry for accurate and efficient training of deep neural networks," Appl. Phys. Lett. **118**, 202405 (2021).
[18]M. Alamdar, T. Leonard, C. Cui, B. P. Rimal, L. Xue, O. G. Akinola, T. Patrick Xiao, J. S. Friedman, C. H. Bennett, M. J. Marinella, and J. A. C. Incorvia, "Domain wall-magnetic tunnel junction spin–orbit torque devices and circuits for in-memory computing," Appl. Phys. Lett. **118**, 112401 (2021).
[19]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 J. Explor. Solid-State Comput. Devices Circuits **5**, 19–24 (2019).
[20]W. H. Brigner, N. Hassan, X. Hu, C. H. Bennett, F. Garcia-Sanchez, C. Cui, A. Velasquez, M. J. Marinella, J. A. C. Incorvia, and J. S. Friedman, "Domain wall leaky integrate-and-fire neurons with shape-based configurable activation functions," IEEE Trans. Electron Devices **69**, 2353–2359 (2022).
[21]N. Hassan, X. Hu, L. Jiang-Wei, W. H. Brigner, O. G. Akinola, F. Garcia-Sanchez, M. Pasquale, C. H. Bennett, J. A. C. Incorvia, and J. S. Friedman, "Magnetic domain wall neuron with lateral inhibition," J. Appl. Phys. **124**, 152127 (2018).
[22]C. Cui, O. G. Akinola, N. Hassan, C. H. Bennett, M. J. Marinella, J. S. Friedman, and J. A. C. Incorvia, "Maximized lateral inhibition in paired magnetic domain wall racetracks for neuromorphic computing," Nanotechnology **31**, 294001 (2020).
[23]S. Liu, C. Bennett, J. Friedman, M. Marinella, D. Paydarfar, and J. A. Incorvia, "Controllable reset behavior in domain wall–magnetic tunnel junction artificial neurons for task-adaptable computation," IEEE Magn. Lett. **12**, 20943805 (2021).
[24]P. O. Mathews, C. B. Duffee, A. Thayil, T. E. Stovall, C. H. Bennett, F. Garcia-Sanchez, M. J. Marinella, J. A. C. Incorvia, N. Hassan, X. Hu, and J. S. Friedman, "High-speed CMOS-free purely spintronic asynchronous recurrent neural network," APL Mach. Learn. **1**, 016107 (2023).
[25]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. **10**, 1152–1160 (2016).
[26]C. Pehle and J. E. Pedersen (2021). "Norse—A deep learning library for spiking neural networks" Zenodo. https://doi.org/10.5281/zenodo.4422025
[27]D. E. Nikonov and I. A. Young, "Uniform methodology for benchmarking beyond-CMOS logic devices," in *International Electron Devices Meeting* (IEEE, 2012).
[28]K. Garello, F. Yasin, H. Hody, S. Couet, L. Souriau, S. H. Sharifi, J. Swerts, R. Carpenter, S. Rao, W. Kim, J. Wu, K. K. V. Sethu, M. Pak, N. Jossart, D. Crotti, A. Furnemont, and G. S. Kar, "Manufacturable 300 mm platform solution for field-free switching SOT-MRAM," in *Symposium on VLSI Circuits* (IEEE, 2019).
[29]C. Frenkel, M. Lefebvre, J.-D. Legat, and D. Bol, "A 0.086-mm^2 12.7-pJ/SOP 64k-synapse 256-neuron online-learning digital spiking neuromorphic processor in 28 nm CMOS," IEEE Trans. Biomed. Circuits Syst. **13**, 18413831 (2018).
[30]M. Davies, N. Srinivasa, T.-H. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain, Y. Liao, C.-K. Lin, A. Lines, R. Liu, D. Mathaikutty, S. McCoy, A. Paul, J. Tse, G. Venkataramanan *et al.*, "Loihi: A neuromorphic manycore processor with on-chip learning," IEEE Micro **38**, 82–99 (2018).
[31]G. S. D. Beach, M. Tsoi, and J. L. Erskine, "Current-induced domain wall motion," J. Magn. Magn. Mater. **320**, 1272–1281 (2008).
[32]K.-S. Ryu, L. Thomas, S.-H. Yang, and S. Parkin, "Chiral spin torque at magnetic domain walls," Nat. Nanotechnol. **8**, 527–533 (2013).
[33]H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms," arXiv:1708.07747 (2017).
[34]D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv:1412.6980 (2014).
[35]H. Song, M. Kim, D. Park, Y. Shin, and J.-G. Lee, "Learning from noisy labels with deep neural networks: A survey," IEEE Trans. Neural Networks Learn. Syst. **2022**, 1–19.