Spike time dependent plasticity (STDP) enabled learning in spiking neural networks using domain wall based synapses and neurons

Upasana Sahu, Aadit Pandey, Kushaagra Goyal, and Debanjan Bhowmik
Spike time dependent plasticity (STDP) enabled learning in spiking neural networks using domain wall based synapses and neurons

Upasana Sahu, Aadit Pandey, Kushagra Goyal, and Debanjan Bhowmik

AFFILIATIONS
1 Department of Electrical Engineering, Indian Institute of Technology Delhi, Delhi 110016, India
2 Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta Georgia 30332, USA
3 Department of Mechanical Engineering, Indian Institute of Technology Delhi, Delhi 110016, India

Note: This paper was presented at the 64th Annual Conference on Magnetism and Magnetic Materials.

Electronic mail: debanjan@ee.iitd.ac.in; http://web.iitd.ac.in/~debanjan/.

ABSTRACT
We have implemented a Spiking Neural Network (SNN) architecture using a combination of spin orbit torque driven domain wall devices and transistor based peripheral circuits as both synapses and neurons. Learning in the SNN hardware is achieved both under completely unsupervised mode and partially supervised mode through mechanisms, incorporated in our spintronic synapses and neurons, that have biological plausibility, e.g., Spike Time Dependent Plasticity (STDP) and homoeostasis. High classification accuracy is obtained on the popular Iris dataset for both modes of learning.

© 2019 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). https://doi.org/10.1063/1.5129729

I. INTRODUCTION
Implementing Neural Network (NN) algorithms in specialized neuromorphic hardware, including spintronic hardware, has been a heavily pursued topic of research in recent years.1–15 Spiking Neural Network (SNN) algorithms are of particular interest in this regard. Power consumption is considered to be very low in hardware implementation of SNN because computation is event based.10–15 While some SNN implementations involve training a non-spiking NN first and then converting it to SNN16,17 some other implementations attempt at training the SNN hardware itself using Spike Time Dependent Plasticity (STDP) property of synapses and homoeostasis property of neurons.18–20 Local and unsupervised nature of such STDP and homoeostasis based learning, coupled with biological plausibility of these mechanisms,18,22–24 makes such training/learning in hardware SNN very interesting.

Spintronic devices have earlier been proposed as neurons and synapses in such SNN.19,20–27 In this paper, we have designed and simulated a hardware SNN using spintronic devices (Domain Wall (DW) based devices) as synapses as well as neurons in the same network. We have showed STDP and homoeostasis enabled learning in our designed SNN. In Section II and III, we carry out micromagnetic simulation to model the DW devices. Transistor based analog circuits are designed on SPICE circuit simulator that will apply the required current pulses into DW devices to make them function as synapses and neurons.19–21 In Section IV, we incorporate the neuron and synapse characteristics in a SNN we design. We train the SNN on a popular Machine Learning (ML) dataset (Iris dataset29) both by a completely unsupervised mechanism, enabled by STDP at synapses and homoeostasis at output stage neurons,18 and a partially supervised mechanism, enabled by STDP at synapses (unsupervised local learning rule) and supervised inhibitory current at the output stage neurons.30–32 Section V concludes the paper.

II. DESIGN OF DOMAIN WALL (DW) BASED SYNAPSE
Both DW synapse and DW neuron (designed in the next section) utilize the physics of Spin Orbit Torque (SOT) driven DW motion in Heavy Metal (HM)/FerroMagnet (FM) heterostructures.
The physics has been studied extensively in the past through experiments and micromagnetic simulations. When in-plane current flows through HM layer, DW in the FM layer above it experiences SOT. If the wall has Neel chirality due to Dzyaloshinskii Moriya Interaction (DMI) at the interface, then DW motion can happen in the absence of magnetic field. We carry out micromagnetic simulations to model DW motion in two different hetero-structures- Pt (HM)/CoFe (FM)/MgO (for synapse design) and Ta (HM)/CoFe (FM)/MgO (for neuron design). The dynamics of the magnetic moments of the FM layer (in which DW moves), under the influence of vertical spin current injected into it due to in-plane charge current through the heavy metal layer below, is simulated for either system using micromagnetic package “muma×3.” Simulation parameters used are from micromagnetic models in previous works that are used to benchmark experimental data on DW motion in such systems (See Section I of supplementary material). DW velocity is plotted as function of in-plane charge current density for both the systems in Fig. 2. For both Pt/CoFe/MgO and Ta/CoFe/MgO systems, DW velocity increases linearly with current density within a certain range. However, for the same current density the velocity is much lower for Ta/CoFe/MgO compared to Pt/CoFe/MgO because the magnitude of DMI is much lower for the Ta/CoFe/MgO system (0.06 mJ/m²) than that for the Pt/CoFe/MgO system (1.2 mJ/m²).

Utilizing the physics of DW motion, a Magnetic Tunnel Junction (MTJ) device has been proposed as a synapse in analog hardware NN. The FM layer in which the DW moves is the free layer of the MTJ device. As the in-plane current flowing through the HM layer below the free layer moves the DW, average magnetization inside the free layer changes and hence conductance of the MTJ changes. Change in conductance (ΔG) of the synaptic device is related to the change in weight the synapse stores (Δw) as follows:

$$\Delta G = \frac{G_{max} - G_{min}}{w_{max} - w_{min}} \Delta w$$  \hspace{1cm} (1) \hspace{1cm}

where $G_{max}$ and $G_{min}$ are the maximum and minimum conductances of the MTJ. Pt/CoFe/MgO system is chosen for DW synapse. Lat-  

eral dimensions of the synaptic DW device simulated here are 500 by 50 nm. Based on the values of R-A product and Tunneling Magneto-Resistance (TMR) of the MTJ, $G_{max} = 2.9 \times 10^{-3} \Omega^{-1}$, $G_{min} = 6.1 \times 10^{-3} \Omega^{-1}$ (Fig. 3). $w_{max}$ and $w_{min}$ are the maximum and minimum values that the weight of any synapse in the SNN can take. Antiferromagnetic regions at the edges prevent the DW from getting destroyed.

Since velocity of the DW is proportional to the current density, when in-plane current pulse of a fixed duration (3 ns), also known as “write” current pulse, is applied on the device, conductance of the MTJ changes by a magnitude proportional to that of the current pulse (Fig. 3). Hence magnitude of the “write” current pulse ($I_{write}$) needed to bring about a certain change in conductance ($\Delta G$) is given as follows:

$$I_{write} = \frac{\partial I_{write}}{\partial G} \Delta G$$  \hspace{1cm} (2) \hspace{1cm}

where $\frac{\partial I_{write}}{\partial G}$ is the slope of straight line that fits the conductance vs write current characteristic of the DW device (Fig. 3) = 1.63 × 10⁵ μA −Ω. In the SNN we design in Section IV, a neuron of the input layer (pre-neuron) is connected to a neuron of the output layer (post-neuron) through a synapse and the weight of the synapse is updated by the STDP rule:

$$\Delta w = \Gamma e^{-\frac{t_{pre}}{\tau_1}} - \Gamma e^{-\frac{t_{post}}{\tau_2}} \text{ if } t_{pre} > t_{post}; \text{ if } t_{pre} < t_{post}$$  \hspace{1cm} (3) \hspace{1cm}

where $\Delta w$ is the change in weight of the synapse, $t_{pre}$ is the time when the pre-neuron spikes, $t_{post}$ is the time when the post-neuron spikes, $\Gamma$ is a dimensionless constant of proportionality equal to 7, and $\tau_1$ and $\tau_2$ are the two time constants for the synapse.

From equation (1), (2) and (3), write current $I_{write}$ needs to be applied on the DW synapse as follows in order to update the weight

---

**FIG. 1.** General schematic of a SOT driven DW device. The MTJ needed for synapse or neuron functionality is not shown here.

**FIG. 2.** DW velocity vs. charge current density for Pt/CoFe/MgO and Ta/CoFe/MgO.

**FIG. 3.** Conductance of MTJ in DW synapse vs magnitude of “write” current pulse.
and train the SNN:

\[ I_{\text{write}} = I_0 e^{-\left(\frac{t_{\text{post}} - t_{\text{pre}}}{\tau_1}\right)} \quad \text{if } t_{\text{post}} > t_{\text{pre}} \]

\[ I_{\text{write}} = -I_0 e^{-\left(\frac{t_{\text{pre}} - t_{\text{post}}}{\tau_2}\right)} \quad \text{if } t_{\text{post}} < t_{\text{pre}} \]  

(4)

where \( I_0 \approx 5 \mu A \) (from equation (1) and (2)). For that purpose, the DW synapse is integrated with a transistor based circuit operating in sub-threshold regime, which injects appropriate “write” current pulse into DW synapse, as shown in Fig. 4, Fig. 5 and Fig. 6.

In the circuit of Fig. 4, drain current through transistor T4 corresponds to \( I_{\text{write},1} \), and drain current through T8 corresponds to \( I_{\text{write},2} \). Based on SPICE simulations of the circuit in Fig. 4, we carry out on Cadence Virtuoso simulator, we plot \( I_{\text{write},1} \) and \( I_{\text{write},2} \) as a function of time for spiking pattern 1 (Fig. 5) and spiking pattern 2 (Fig. 6). Circuit parameters are chosen such that \( \tau_1 = 1.5 \mu s \) for all synapses in our designed SNN. For spiking pattern 1, pre-neuron spikes once \( (t_{\text{pre}}) \) followed by several post-neuron spikes \( (t_{\text{post}}) \). Hence \( t_{\text{post}} > t_{\text{pre}} \) here. Since pre-neuron spike corresponds to a sharp drop in gate voltage of T1 followed by its rise to original value \( (\text{Fig. 5(a)}) \), T2 being pMOS first turns on resetting voltage across capacitor C1 to 0.8 V (lower electrode of C1 in Fig. 4) considered positive terminal) and then turns off. For all time \( t \) after \( t_{\text{pre}} \) C1 is charged through T1. So voltage across C1, and hence gate voltage of T3, increases linearly with \( t - t_{\text{pre}} \) with a slope inversely proportional to value of capacitance C1 \( (\text{Fig. 5(c)}) \). When post-neuron spikes at \( t_{\text{pre}} \) (sharp rise in gate voltage of transistor T4 turning on T4 - Fig. 5(b)), since T3 is designed to operate in sub-threshold region, drain current through T3 and hence T4 \( (I_{\text{write},2}) \) is proportional to \( e^{-\left(\frac{t_{\text{pre}} - t_{\text{post}}}{\tau_2}\right)} \) for spiking pattern 1 since \( t_{\text{post}} > t_{\text{pre}} \) \( (\text{Fig. 5(d)}) \). The STDP time constant \( \tau_1 \) is directly proportional to sub-threshold swing of T3 and capacitance C1. Unlike C1, charging/discharging does not happen for C2 for spiking pattern 1 \( (\text{Fig. 5(c)}) \). This is because the equivalent of T2 for C2 (T6) is connected to post-neuron spike instead of pre-neuron spike and the equivalent of T4 for C2 (T8) is connected to pre-neuron spike instead of post-neuron spike. So \( I_{\text{write},2} = 0 \) for spiking pattern 1 since \( t_{\text{post}} > t_{\text{pre}} \) as desired. For spiking pattern 2 \( (\text{Fig. 6}) \), post-neuron spikes once \( (\text{Fig. 6(b)}) \) and pre-neuron spikes several times after that \( (\text{Fig. 6(a)}) \). Hence \( t_{\text{post}} < t_{\text{pre}} \) here. In this case, voltage across C2 is reset after the post-neuron spike and then it drops linearly with \( t - t_{\text{post}} \) due to discharge through T5, with a slope inversely proportional to value of capacitance C2 \( (\text{Fig. 6(c)}) \). T7 operates in sub-threshold region here and T8 is pMOS transistor here as opposed to T4 (nMOS). Hence when pre-neurons spikes at \( t_{\text{post}} \) T8 turns on and current through T8 \( (I_{\text{write},2}) \) is proportional to \( e^{-\left(\frac{t_{\text{pre}} - t_{\text{post}}}{\tau_2}\right)} \) and has a negative sign \( (\text{Fig. 6(d)}) \). Thus, \( \tau_2 \) is now linearly proportional to sub-threshold swing of T7 and
However dependence of $\tau$ the STDP rule (equation (3)) is implemented for our DW synapse.

The waveforms in Fig. 5 and Fig. 6 are for specific values of C1 and C2 in the circuit of Fig. 4 which lead to $\tau_1 = \tau_2 = 1.5 \mu$s. However dependence of $\tau_1$ on C1 and $\tau_2$ on C2, as obtained from multiple SPICE simulations of the STDP circuit in Fig. 4 for different values of C1 and C2, are plotted in Section 4 of supplementary material. Variation in values of C1 and C2 due to circuit imperfections leads to variation in values of $\tau_1$ and $\tau_2$. However if $\tau_1$ and $\tau_2$ lie within a certain range the designed SNN can be accurately trained on the given datasets. We discuss this in Section IV where we present the performance metrics for learning/training using our designed SNN.

III. DESIGN OF DOMAIN WALL (DW) BASED NEURON

In the Leakly Integrate Fire (LIF) model of neuron,\textsuperscript{18,23} its membrane potential $v(t)$ is governed by the following equation:

$$C \frac{dv(t)}{dt} = -G_L(v(t) - E_L) + I(t)$$

where $I(t)$ is input current to the neuron, $G_L$ is membrane conductance, $E_L$ is resting potential of neuron. Once $v(t)$ reaches the threshold potential ($V_{th}$), the neuron generates a spike and $v(t)$ drops to $E_L$. For our designed SNN in next section, we take $G_L = 30$ pS, $E_L = -70$ mV, $C = 1200$ pF and $V_{th} = 20$ mV for the LIF model of our neurons. Time between two consecutive spikes generated by the neuron is plotted as a function of input dc current to the neuron after solving equation (6) for the given parameters (Fig. 7(a)). Our choice of LIF parameters is such that the time between two consecutive spikes of the neuron is in the order of the time constant of the STDP synapse (1.5 $\mu$s) and hence several orders higher than duration of the “write” current pulse into the DW synapse, which is same as the duration of each spike in our designed SNN (3 ns). This is a requirement for STDP based learning to work (Fig. 5, Fig. 6).

The DW neuron, integrated with a transistor based circuit, we design in Fig. 7(b) satisfies the desired LIF characteristic described above (Fig. 7(a)). Working principle of DW neuron is as follows: input current flowing through HM layer moves the DW in the FM layer from one end of the device to another, much like the DW synapse. However unlike the DW synapse, the MTJ is located only at the other end. When the DW reaches the other end, TMR of the MTJ changes abruptly and spike is generated.\textsuperscript{24} Thus time period between two consecutive spikes is equal to the time taken by the DW to traverse the length of the device (Fig. 7(a)). Since this time period must be several orders higher than the duration of “write” current pulse for DW synapse (3 ns), which is equal to the time taken by the DW to traverse the length of the synapse, the length of the DW neuron (6 $\mu$m) is chosen to be much higher than that for the DW synapse (500 nm). Width is also chosen to be higher for the DW neuron (600 nm) compared to the DW synapse (500 nm) because for the same magnitude of input current, higher width corresponds to smaller current density and hence lower velocity of DW. Ta/CoFe/MgO system is chosen for DW neuron instead of Pt/CoFe/MgO since DW velocity is found to be lower in Ta/CoFe/MgO than Pt/CoFe/MgO for the same current density (Fig. 2). The MTJ in the DW neuron has dimensions of 600 nm by 100 nm.

Once the DW reaches the other end, TMR of the MTJ increases due to switching of moment in the free layer. As a result, gate voltage of the transistor T1 in Fig. 7(b) increases, turning it on. This gate voltage is the output of the neuron ($V_{out}$) if it is a post-neuron since the spike required for post-neuron is positive (Fig. 5(b), Fig. 6(b)). For pre-neuron another two transistors based standard inverter circuit is connected to the gate voltage of this transistor T1 (Section 3 of supplementary material). The output of the inverter circuit shows a negative spike when domain wall in the neuron device reaches the MTJ, as required for a pre-neuron (Fig. 5(a), Fig. 6(a)). The ON current of transistor T1 flows in the reverse direction of input current in the DW neuron and moves DW to its initial position. This is equivalent to $v(t)$ in LIF model of the neuron dropping to $E_L$ after a spike. Additional components are added to the transistor based circuit of Fig. 7(b), as discussed in Section 2 of supplementary material, to incorporate homoeostasis property in the DW neuron if required.

IV. DESIGN OF SNN AND TRAINING ON IRIS DATASET

We next design a SNN with a layer of pre-neurons (input stage) connected to a layer of post-neurons (output stage) through synapses (Fig. 8). Based on the spike times of pre-neurons and post-neurons, weights of the synapses are updated by STDP rule (equation (4)) for every input/sample in each epoch. Then
the process is repeated for several epochs to achieve learning/training.\textsuperscript{18,30,31} Currents proportional to the 16 different features of each input/flower in the Iris training and test sets (after basic feature engineering on the 4 features in the original dataset\textsuperscript{25–31}) are applied on the pre-neurons. The label/type the input/flower belongs to is determined by the post-neuron that fires most for that input. Since there are 3 types of flowers in the dataset, we use 3 post-neurons. The post-neurons are connected inhibitorily with each other to implement the “Winner Take All” (WTA) mechanism.\textsuperscript{18,30,31} We implement the WTA mechanism among the post-neuron circuits of Fig. 7(b) through an additional transistor based circuit.\textsuperscript{18} The circuit schematic of the implementation and corresponding SPICE simulations are shown in Section 5 of supplementary material. Alternatively, WTA can be implemented through dipole coupling among the ferromagnetic layers of the domain wall neuron devices themselves, through which the domain walls move.\textsuperscript{18} Then the additional transistors we have used to implement WTA mechanism will not be required.

The DW devices designed in previous sections, with the same parameters stated there, are used to model the neurons and synapses in our designed SNN. Learning is achieved through weight update of synapses due to “write” current pulses applied on the DW synapses, following equations (1)–(4). STDP always leads to some degree of unsupervised nature to the learning. However based on how we control the spiking of post-neurons, we have two different learning schemes here–completely unsupervised and partially supervised. Under the completely unsupervised scheme, the post-neurons have an additional homoeostasis property (Section 2 of supplementary material). During training, when input of one type makes a post-neuron spike, its spiking threshold $V_{th}$ in the LIF model goes up by 7 mV, followed by a decay with time constant $\tau_{\text{homeo}} = 15 \mu s$ (Section 2 of supplementary material). Because of the increased threshold, only when another input is of the same type, the incoming current to that post-neuron is large enough for the neuron to spike. For input of other types it does not spike. Thus classification is achieved without a learner.\textsuperscript{33} Under the partially supervised scheme, the post-neurons do not have homoeostasis property. Instead, during the training process, for an input of a particular type, inhibitory currents are applied on all post-neurons except the post neuron, which we want to spike most for any input of that type.\textsuperscript{25,31} For either scheme, high training accuracy (45 train samples) and test accuracy (105 test samples) are obtained after $\approx 15$ epochs (Fig. 9). The net energy dissipated in all the synapses because of Joule heating for “write current” pulses during the learning process is calculated to be in the range of 50–200 $fJ$.

The classification accuracy for the SNN depends on values of $\tau_1$ and $\tau_2$ since they control the STDP based synaptic weight update rule (equation 3), upon which training/learning of the SNN is based. We observe from our simulations that both train and test accuracy continue to be around 90 percent as desired as long as $\tau_1$ is in between 1.3 $\mu$s and 1.9 $\mu$s and $\tau_2$ is in between 1 $\mu$s and 1.7 $\mu$s, given other parameters in SNN are not changed. From Section 4 of supplementary material, this corresponds to variation of capacitance $C_1$ between 0.7 $pF$ and 1.05 $pF$ and $C_2$ between 2.5 $pF$ and 4.8 $pF$ in the STDP exhibiting synapse circuit of Fig. 4. Variation of capacitances within this range due to circuit imperfections will hence not affect the performance of the designed SNN.

V. CONCLUSION

We have simulated STDP enabled learning under two different schemes in SNN hardware using DW devices both as synapses and neurons. We have obtained high classification accuracy on a popular ML dataset–the Iris dataset. Training it on larger and more complicated datasets however involves many more STDP synapse circuits since the number of pre-neurons and post-neurons goes up. If the STDP rule in equation (3) is simplified by eliminating the exponential characteristic,\textsuperscript{18} number of transistors in the STDP circuit will go down making the overall circuit simpler and more scalable. Training our designed SNN on larger and more complicated datasets will be the subject of our future study.

SUPPLEMENTARY MATERIAL

See supplementary material for circuit implementations of homoeostasis and Winner Take All (WTA) mechanism.

ACKNOWLEDGMENTS

Debanjan Bhowmik thanks Department of Science and Technology (DST), India for INSPIRE Faculty Award and Science and Engineering Research Board (SERB), India for Early Career Research (ECR) Award, which helped fund this research.

REFERENCES

1. H. Tsai, S. Ambrogio, P. Narayanan, R. M. Shelby, and G. W. Burr, J. Phys. D Appl. Phys. 51, 283001 (2018).
2. A. Sebastian, M. L. Gallo, G. W. Burr, S. Kim, M. BrightSky et al., J. Appl. Phys. 124, 111101 (2018).
3. C. Li, D. Belkin, Y. Li, P. Yan, M. Hu et al., Nat. Commun. 9, 2385 (2018).
4. E. Cai, J. M. Correll, S. H. Lee, Y. Lim, V. Bothra et al., Nat. Electronics 2, 290–299 (2019).
5. J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa et al., Nature 547, 428–432 (2017).
6. M. Azam, D. Bhattacharya, D. Quehlouz, and J. Atulasimha, J. Appl. Phys. 124, 152122 (2018).
7. A. Sengupta, Y. Shim, and K. Roy, IEEE Trans. Biomed. Circuits Syst. 10, 1152 (2016).
8. U. Saxena, D. Kaushik, M. Bansal, U. Sahu, and D. Bhowmik, IEEE Trans. Magn. 54, 1 (2018).
9. D. Bhowmik, U. Saxena, A. Dankar, A. Verma, D. Kaushik et al., J. Magn. Magn. Mater. 489, 165434 (2019).
10. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada et al., Science 345(6197), 668 (2014).
