A Low-voltage, Low power STDP Synapse implementation using Domain-Wall Magnets for Spiking Neural Networks

Govind Narasimman*, Subhrajit Roy*, Xuanyao Fong†, Kaushik Roy†, Chip-Hong Chang * and Arindam Basu*

*School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

†School of Electrical and Computer Engineering, Purdue University, USA

Abstract— **Online, real-time learning in neuromorphic circuits have been implemented through variants of Spike Time Dependent Plasticity (STDP). Current implementations have used either floating-gate devices or memristors to implement such learning synapses together with non-volatile storage. However, these approaches require high voltages ($\approx 3-12$V) for weight update and entail high energy for learning ($\approx 4-30$pJ/write). We present a domain wall memory based low-voltage, low-energy STDP synapse that can operate with a power supply as low as 0.8V and update the weight at ≈ 40fJ/write. Device level simulations are performed to prove its feasibility. Its use in associative learning is also demonstrated by using neurons with dendritic branches to classify spike patterns from MNIST dataset.**

I. INTRODUCTION

Spiking neural network (SNN) is the 3rd generation of neural networks. It is much more bio-realistic and has higher computational power than it's predecessors. When it is realized on hardware, the circuit is activated only on sparsely distributed events, which greatly minimizes dynamic power since only spikes or events need to be communicated between neurons [1]. In addition, address event representation protocols allow for arbitrary configuration of network topology by using separate memory modules to store connection tables.

A big challenge in the field of neuromorphic systems that aim to implement low-power, brain-inspired integrated circuits to impart cognition to robots interacting with environment in real-time is the design of a self-learning synapse [2]. Since synapses outnumber neurons by a factor of 1000X, a synapse needs to be much-more compact and power-efficient than a neuron. Developing a compact learning synapse with high to moderate resolution of weights remains a bottleneck in the hardware implementation of SNN [2]. Typical CMOS solution stores only two steady states of weights [2]. For non-volatile weights that can store multiple states, the two better solutions involve usage of floating-gate (FG) devices [3] or memristors [4]. However, both FG devices and memristors require a high voltage for programming. This leads to reliability issues in the underlying CMOS, higher area required due to use of high voltage CMOS transistors as well as larger energy needed for synaptic learning.

In this paper, we present a novel spintronic domain wall magnet based compact learning synapse that requires low voltage and low power for learning. We also do a comparison with other plastic synapses, to highlight the savings in energy requirements for online modification of synaptic strengths. We

show that it can exhibit STDP or reverse STDP characteristics depending on the control signals. Furthermore, we demonstrate its use in semi-supervised learning by classifying hand written digits from the MNIST dataset [5].

II. PRELIMINARIES

A. Domain wall synapse

The output of a simple neuron (perceptron) is given by:

$$Y = G(\sum_i W_i X_i + b) \tag{1}$$

where W_i is the weight of the ith synapse, X_i is the input to the ith synapse, b is the bias of the neuron and $G(.)$ is the activation function. A domain wall nano-strip can be formed by two domains of oppositely polarized magnetic moment separated by a transition region called domain wall (DW). It has been shown in several studies [6], [7] that DW motion can be achieved with very low current densities for soft magnetic materials with high Perpendicular Magnetic Anisotropy . As suggested in [8], input charge currents injected vertically into the DW magnet enter a non-magnetic metallic channel beneath it with a thin oxide separation after being polarized. The vertically inserted charge current produces weighted spin current (WSC) in the channel, according to the position of DW measured from the center of the nanaostrip [8]. Thus, a DW nano-strip can be used as a synapse for the perceptron with the bipolar weighted term 'W' of (1) set by the displacement of DW from the center of the nano-strip.

Several studies [9], [7] have also shown that spin transfer torque can produce non-volatile memories programmable at very low power. Through spin transfer torque, the spin current injected into the metallic channel via an input magnet at one end can reverse the magnetization of a nano-magnet at the other end. Reversible magnetization switching using "non-local" spin currents, where the path for charge current and spin current are separated, has also been demonstrated in recent experimental work [9]. Such switching mechanism can be used to implement the nonlinear neuronal function G of (1). The full implementation of a neuromorphic circuit through spin mode signaling will be described in Section III-C.

B. Spiking Neural Network (SNN) and Spike Timing Dependent Plasticity (STDP)

SNN consists of spiking neurons that do not generate outputs at each time step like other artificial neutral networks. Instead, a neuron produces a spike asynchronously when its membrane potential (V_m) reaches a specific value. The information transfer in SNN hence takes place through precise spiking time or rates of spikes. STDP is one of the biologically
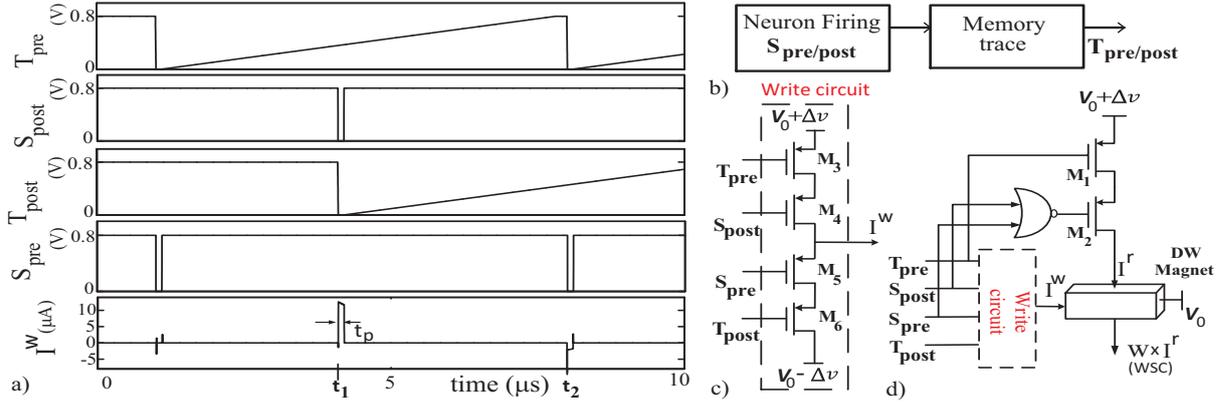
Fig. 1: a) Write current generation from sampling of $T_{pre(post)}$ by $S_{post(pre)}$. At instance t_1, a post synaptic spike S_{post} happens which samples T_{pre} waveform. The resulting write current creates a positive ΔW according to Δt. On the other hand, a negative current is generated by a pre-synaptic spike S_{pre}, at time instance t_2, to decrease the weight 'W' of the corresponding synapse. b) Generation of $T_{pre/post}$ according to the memory trace. c) PMOS stack (M_3-M_6) used to generate the write current to obtain the desired ΔW. V_0 and Δv are 750mV and 50mV respectively. d) Compact representation of the whole DW magnet based synapse completed with the write circuit.

plausible learning rules commonly used for modifying the synaptic strengths in SNN [10]. In the STDP learning rule, the arrival of pre-synaptic spike before postsynaptic action potentials leads to long-term potentiation of the synapse. Conversely, the arrival of pre-synaptic spike after postsynaptic spikes leads to long-term depression of the same synapse. This implies that the change in weight ΔW, should be a function of the difference in arrival time between the pre- and post-synaptic spikes, Δt [10].

Various non-volatile elements like floating gate transistor, memristor [3], [4] etc. have been used to replicate the STDP properties of a plastic synapse in the recent past. However, the high voltage requirement for programming the weights reduce the reliability of the circuit and increase the energy needed for on-chip learning. In the next section, we show how such function can be implemented with a DW synapse with a few transistors to overcome these problems.

III. DW MAGNET BASED SYNAPSE

A. Characterization of DW motion

A thin permalloy strip of cross section (60nm × 4nm) is used as the DW synapse. The DW motion can be controlled by varying the amplitude and duration of the current pulses. The parameters of DW strip are listed in Table I, which are in close agreement with [7]. Periodic triangular notches are added to the DW strip at an interval of 80nm. These constrictions act as local pinning sites to prevent the motion of DW due to thermal fluctuation [7].

Simulation framework employed here is similar to that of [8]. The 4-component spin transport model of nano-magnets is self-consistently solved with Landau-Lifshitz-Gilbert (LLG) equation as explained in [8]. The one-dimensional model from [6] was adopted, for DW motion along a nano-strip under the influence of current pulse (Equations (2) and (3)). These 3 blocks were simulated with circuit elements in a SPICE platform. A system level model for the neuromorphic circuit was obtained from this simulation. The DW motion in a nanostrip can be characterized by incorporating good estimates of anisotropy field H_k and DW width Δ into the 1-D model

TABLE I: Magnetic parameters of DW strip

Parameter	value
Exchange Energy Constant (A)	$1 * 10^{-11} J/m$
Damping parameter (α)	0.04
Saturation Magnetization.-(M_s)	860kA/m
Nano strip dimensions (Length /Breadth/thickness)	1.3μm x 60nm x 4nm
Non-adiabatic Parameter (β)	0.01
Uniaxial Anisotropy-(H_k)	2200 A/m

as follows.

$$(1+\alpha^2)\frac{\dot{Z}}{\Delta} = \alpha\gamma_0(H_e + H_p(Z) + H_{th}(t))$$
$$+ \frac{1}{2}\gamma_0 H_k sin(2\Phi) - (1+\alpha\beta)\frac{b_J}{\Delta} \quad (2)$$

$$(1+\alpha^2)\dot{\Phi} = \alpha(H_e + H_p(Z) + H_{th}(t))$$
$$- \frac{1}{2}\alpha\gamma_0 H_k sin(2\Phi) - (\alpha - \beta)\frac{b_J}{\Delta} \quad (3)$$

where $Z = Z(t)$ is the position of the DW's center, $\Phi = \Phi(t)$ is the tilt of DW magnetization, β is the 'non-adiabatic constant, α is the 'damping parameter, $H_p(Z) = -\frac{1}{2\mu_0 M_s L_y L_z}\frac{\partial V_{pin}}{\partial Z}$ in the vicinity of notch, Δ is the DW width, V_{pin} is a measure of notch strength, and $b_J = \frac{J_a \mu_B P}{e M_s}$ is the term used to account for the current density J_a and polarization P of the ferromagnet. The model is benchmarked against the results of [7]. The peak pinning potential, V_{pin}, has a value of 10^{-20} J which is equal to the effective energy barrier micromagnetically estimated in [7]. The results of [7] reveal that an initially pinned DW can be expelled away from the constriction with a current pulse of sufficient amplitude. By increasing the amplitude of the current pulses further, more constrictions can be surpassed. The width of the current pulse (t_p) used here is 100ns. On release of the current pulse, the DW experiences an attractive force $H_p(Z)$, which pulls the DW towards the nearest pinning site. The final displacements of DW (ΔW) will be quantized in terms of the interval of notches, and as a function of the current pulse amplitude (J_a). For learning the synaptic weights W, these DW displacements have to be controlled by the write currents.

B. Enabling STDP with DW synapse

For generating these write currents as a function of Δt, a circuit akin to [11] was used. The spike waveforms of the pre- and post-synaptic neurons are denoted by S_{pre} and S_{post}, respectively. The time since last spike of the pre- and post-synaptic neurons can be encoded in two waveforms, T_{pre} and T_{post}, respectively (see Fig. 1(b)). These exponentially or linearly decaying waveform can be generated by a discharging capacitor through a resistor or a current source, respectively. The magnitude of the waveform encodes the time since the last event or spike, hence the name memory trace.

Fig. 1(a) shows the generation of write current by sampling $T_{pre}(T_{post})$ by $S_{post}(S_{pre})$. When a post-synaptic spike S_{post} happens at t_1 after the pre-synaptic spike, S_{post} samples the T_{pre} waveform at instant t_1 to generate a write current. Similarly, when a pre-synaptic spike happens at t_2 after the post-synaptic spike, S_{pre} samples the T_{post} waveform at time instant t_2 to generate a negative write current. So the STDP characteristics can be changed according to the memory trace. The STDP characteristics will be expressed in terms of $\epsilon = T_{pre}/T_{post}$ henceforth.

The PMOS transistor stack (Fig. 1(c)) was employed to generate the write currents I^w for obtaining the STDP characteristics shown in Fig. 3(a). The transistors, $M_3 - M_6$, are sized such that a $\Delta V = 50mV$ with $V_0 = 0.75V$ is sufficient to produce the maximum write current I_{max} for ΔW_{max} in Fig. 3(a). The same T_{pre} signal is converted into an input read current I^r to the synapse through a transistor M_1 operating in the deep triode region. The read and write operations are coupled by an active low signal generated from the NOR-operation of S_{post} and S_{pre}. As a result, transistor M_2 is turned off when the DW is moved by the write current. This decouples the learning operation, which is localized at a $S_{pre/post}$ pulse, from the regular operation (Section III-C), which is guided by the read currents that span a longer time duration governed by $T_{pre/post}$.

C. System integration

As shown in Fig. 2(a), input charge currents are injected vertically into the DW magnets to produce the corresponding WSC in the channel. These WSC form a net spin current (I_{net}^{spin}) in the channel which is non-locally absorbed by the output magnet. This net spin current exerts a spin torque on the output nano-magnet while the charge current in the channel flows into the ground lead, which is biased at V_0 volts [8]. The magnetization of the output magnet can be sensed by forming a magnetic tunnel Junction (MTJ) stack [8] with the output magnet as its free-layer. By employing bennett clocking scheme, the current I_{net}^{spin} required to reverse the output magnet's magnetization can be reduced. This can be achieved by adding a preset magnet, which has its magnetization directed orthogonal to the output magnet, to the structure (see Fig. 2(b)). Upon injection of a current pulse through the preset magnet, the output magnet aligns itself to the hard axis. The final state of the output magnet is decided by the net spin current present in the channel on the release of preset pulse. The change in resistance of the MTJ stack can be sensed by comparing it with the resistance of a reference MTJ in
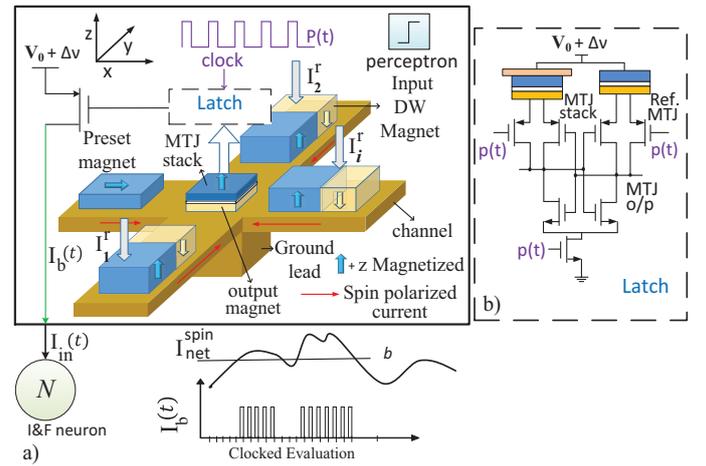

Fig. 2: a) Signal flow in neuromorphic circuit: Inputs $I_i^r(t)$ are propagated through the neuromorphic circuit to produce synchronous current pulses $I_b(t)$. The symbol for one such block of perceptron is shown on the top right corner. $I_{in}(t)$ is the total current to the I&F neuron by combining p dendritic branches. For this figure, $p = 1$. b) Clocked differential latch employed to detect change in MTJ resistance[8].

a differential latch (see Fig 2(b)) to produce a binary output [8]. Thus the MTJ output Y connected to k DW synapses can be expressed as:

$$Y(t) = u(I_{net}^{spin}) = u(\sum_{i}^{k} W_i I_i^r(t) + b) \qquad (4)$$

where $u(.)$ denotes a heaviside function, I_i^r(t) is the input current to the i-th DW magnet and W_i is its weight. Here +z magnetized output magnet produces $Y = 1$. This equation is similar to (1) where the heaviside function represents the neuronal nonlinearity.

The MTJ output is evaluated by a clocked latch and hence becomes a synchronous signal. Since each perceptron only outputs a 1-bit signal, a set of p parallel perceptrons is used to boost its classification power. This structure is also analogous to the biological dendritic branches [12]. Denoting the clock pulses by $p(t)$, the synchronous branch current I_b(t) (see Fig. 2(a)) for any dendritic branch can be written as $I_b(t) = Y \times p(t)$. Many such currents (I_b(t)) are combined to give the total input current $I_{in}(t)$ to the Integrate and Fire (I&F) neuron. The evolution of membrane potential V_m of the I&F neuron can be expressed as:

$$\tau_m \frac{d(v_m)}{dt} + v_m(t) = RI_{in}(t) = R\sum_{j}^{p} I_b(t) \qquad (5)$$

where $\tau_m = RC$ of leaky integrator [10] and p denotes the number of branches connected to the neuron. Whenever $V_m > V_{th}$, the neuron asynchronously fires a post-synaptic spike S_{post} of duration t_p and then resets to zero. The I&F neuron circuit from [13] is used for SPICE simulation, with the memory trace $T_{pre/post}$ generated by $T_{pre/post} = K (t) * S_{pre/post}$ where $K(t)$ is the spike convolution kernel.

The circuit is thus Globally Asynchronous Locally Synchronous since the dendritic branch outputs are synchronous but the neuronal firings are asynchronous.

IV. Results

A. STDP curves from DW synapse

The STDP curves obtained from the DW synapse are shown in Fig 3. These STDP curves can be tuned by adjusting T_{pre} and T_{post} or the maximum write current I_{max}. Increasing I_{max} increases ΔW_{max}. By interchanging T_{pre} with T_{post} and S_{pre} with S_{post} in Fig. 1(c), the Reverse STDP (R-STDP) curve is obtained in Fig. 3(b). Anti-hebbian rules like R-STDP are also observed in biology [12].

The average write energy/spike required for the DW synapse is ≈ 40fJ from the SPICE simulation based on 65nm CMOS process. This is nearly three orders of magnitude lower than the 30pJ of write energy consumed by a memristive synapse (with $\Delta W_{max}/W = 0.25$) [4] and two orders of magnitude lower than 4.5pJ consumed by a floating gate (FG) based synapse [3] for the same STDP. Moreover, the proposed neuromorphic circuit can work with power supply as low as 0.8V compared to 12/4.5V required for tunneling/injection in FG based synapse.

B. Network simulation: handwritten digit classification

The online learning capability of our system is demonstrated by binary classification of different handwritten digits from the MNIST database [5]. As a proof of concept, only binary classification is shown. A network setup depicted in Fig. 4(a) is implemented in software where each component follows the restrictions of our neuromorphic circuit. $d = 784$ input dimensions of a pattern are fed to each excitatory neuron through p perceptrons that mimic the dendritic branches. Each perceptron or dendritic branch has k random synaptic contacts, where $k << d$. Having few synapses on a branch is useful since the spin diffusion length is quite small and cannot support full connectivity per branch. T is the time duration of one pattern and $k = p = 28$ in this network. We employ a R-STDP based learning rule similar to [12], to train this network. After training the network for 3 epochs, the resultant receptive field formed is shown in Fig. 4(b). The testing accuracy averaged over 10 trials for classifying digit pairs into two classes is 95.4667% (SD=2.948%), which proves the viability of this method.

V. Conclusion and Discussion

We presented a DW magnet based synapse that can exhibit STDP behavior while operating at a low voltage of 0.8V
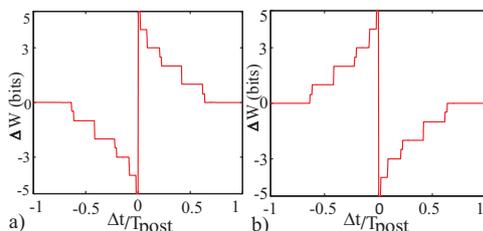
Fig. 3: STDP curves: a) STDP curve obtained from SPICE simulation of the circuit shown in Fig. 1(d) based on a 65nm CMOS process. b) R-STDP curve obtained by interchanging T_{pre} and S_{post} with T_{post} and S_{pre}, respectively in Fig. 1(c). Both curves are obtained from an exponentially decaying memory trace.
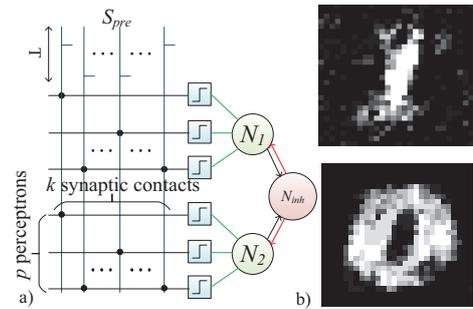
Fig. 4: a) Neural network architecture used for the simulation with each perceptron mimicking a dendritic branch. b) Receptive fields obtained for hand written digits *0* and *1* after training with the network.

with a write energy that is at least $\approx 100X$ lower than other methods based on floating-gates or memristors. The synapses are embedded in a neural network where neurons with dendritic branches are used. As each dendritic branch is connected to few inputs only, the limited spin diffusion length is not a problem for this architecture. As a proof of concept, binary classification of handwritten digits is shown. In future, we will extend this work to multi-class classification and perform a more extensive design space exploration.

References

[1] S. C. Liu, T. Delbruck, G. Indiveri, A. Whatley, and R. Douglas, *Event-Based Neuromorphic Systems*. U.K: John Wiley & Sons, 2014.

[2] M. R. Azghadi, N. Iannella, S. F. Al-Sarawi, G. Indiveri, and D. Abbott, "Spike-based synaptic plasticity in silicon: Design, implementation, application, and challenges," *Proc. of IEEE*, vol. 102, no. 5, pp. 717–737, May 2014.

[3] S. Brink, S. Nease, P. Hasler, S. Ramakrishnan, R. Wunderlich, A. Basu, and B. Degnan, "A learning-enabled neuron array ic based upon transistor channel models of biological phenomenon," *IEEE Trans. on Biomedical Circuits and Syst.*, vol. 7, no. 1, pp. 71–81, Feb. 2013.

[4] S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, "Nanoscale memristor device as synapse in neuromorphic systems," *Nano Lett.*, vol. 10, no. 4, pp. 1297–1301, 2010.

[5] Y. LeCun, C. Cortes, and C. Burges, "The mnist database of handwritten digits, 1998," *Available electronically at http://yann. lecun. com/exdb/mnist*, 2010.

[6] E. Martinez, "The stochastic nature of the domain wall motion along high perpendicular anisotropy strips with surface roughness," *Journal of Physics: Condensed Matter*, vol. 24, no. 2, p. 024206, 2012.

[7] E. Martinez, L. Lopez-Diaz, O. Alejos, L. Torres, and M. Carpentieri, "Domain-wall dynamics driven by short pulses along thin ferromagnetic strips: Micromagnetic simulations and analytical description," *Physical Review B*, vol. 79, no. 9, p. 094430, 2009.

[8] M. Sharad, C. Augustine, G. Panagopoulos, and K. Roy, "Spin-based neuron model with domain-wall magnets as synapse," *IEEE Trans. on Nanotechnology*, vol. 11, no. 4, pp. 843–853, 2012.

[9] T. Yang, T. Kimura, and Y. Otani, "Giant spin-accumulation signal and pure spin-current-induced reversible magnetization switching," *Nature Physics*, vol. 4, no. 11, pp. 851–854, 2008.

[10] J. P. Pfister, "Theory of non-linear spike-time dependent plasticity," Ph.D. dissertation, Swiss Federal Institute of Technology, Lausanne, Switzerland, 2006.

[11] G. Indiveri, E. Chicca, and R. Douglas, "A vlsi array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity," *IEEE Trans. on Neural Net.*, vol. 17, no. 1, pp. 211–221, 2006.

[12] S. Hussain, S. Liu, and A. Basu, "Hardware-amenable structural learning for spike-based pattern classification using a simple model of active dendrites," *Neural Computation*, vol. 27, no. 4, pp. 845–897, 2015.

[13] A. Banerjee, S. Kar, S. Roy, A. Bhaduri, and A. Basu, "A current-mode spiking neural classifier with lumped dendritic nonlinearity," in *Proc. of IEEE, Int. Symp. on Circuits and Syst.* Lisbon, Portugal: IEEE, May 2015, pp. 714–717.