Low-Power (1T1N) Skyrmionic Synapses for Spiking Neuromorphic Systems

Tinish Bhattacharya, Sai Li, Yangqi Huang, Wang Kang, Weisheng Zhao and Manan Suri

Abstract—In this work, we propose a ‘1-transistor 1-nanotrack’ (1T1N) synapse based on movement of magnetic skyrmions using spin polarised current pulses. The proposed synaptic bit-cell has 4 terminals and fully decoupled spike transmission- and programming-paths. With careful tuning of programming parameters we ensure multi-level non-volatile conductance evolution in the proposed skyrmionic synapse. Through micromagnetic simulations, we studied in detail the impact of programming conditions (current density, pulse width) on synaptic performance parameters such as number of conductance levels and energy per transition. The programming parameters chosen used all further analysis gave rise to a synapse with 7 distinct conductance states and 1.2 fJ per conductance state transition event. Exploiting bidirectional conductance modulation the 1T1N synapse is able to undergo long-term potentiation (LTP) & depression (LTD) according to a simplified variant of biological spike timing dependent plasticity (STDP) rule. We present subthreshold CMOS spike generator circuit which when coupled with well known subthreshold integrator circuit, produces custom pre and post-neuronal spike shapes, responsible for implementing unsupervised learning with the proposed 1T1N synaptic bit-cell and consuming ~ 0.25 pJ/event. A spiking neural network (SNN) incorporating the characteristics of the 1T1N synapse was simulated for two separate applications: pattern extraction from noisy video streams and MNIST classification.

Index Terms—Skyrmion, SNN, STDP, Neuromorphic Hardware, Synapse.

I. INTRODUCTION

HARDWARE implementation of biologically inspired neuromorphic systems has gained a lot of interest in the past few years [1], [2]. Neuromorphic engineering aims to achieve intelligent computing by taking inspiration from complex neural/synaptic circuits and their biophysical mechanisms, using nano-electronic/magnetic devices [3], [4]. The nanodevices when made to follow certain conductance (weight) modulation rules within a connected network lead to equivalent trained neural networks that may be used for different training and inference tasks [5], [6]. Various emerging non-volatile nanodevices are currently being investigated for synaptic applications; Conductive-Bridge Memory (CBRAM) [5], Phase Change Memory (PCM) [7], Magnetic-Tunnel Junction (MTJ) [8], domain walls [9] and oxide based resistive switching memory (OxRAM) [6], [10]. Among emerging spintronic nanodevices, magnetic skyrmions have gained a significant amount of interest owing to their small size, topological stability and ultralow current densities [11], [12]. Skyrmions are topologically stable spin textures which have been experimentally observed in Heavy metal (HM)-Ferromagnetic Metal (FM) heterostructure and bulk ferromagnets [13], [14]. They have been proposed as possible post-Moore candidates for logic [15], [16] and storage [17] applications. They have also been proposed for emulating the properties of leaky-integrate and fire neurons in [18], [19]. However for a device to behave as analog (multilevel) synapse, it should have multiple programmable non-volatile conductance states [10]. Controlled conductance variation using skyrmions on nanotracks was first reported in [20]. However in [20] the conductance states reported were a strong function of the inter programming-pulse delay, thus making the intermediate synaptic weights transient or not truly non-volatile. Also to the best of our knowledge no work has yet shown a pathway for implementing online unsupervised learning with multilevel skyrmion on nanotrack synapses.

In this work we present a modified nanotrack structure which makes the overall synapse a 4-terminal device and completely decouples the spike transmission and programming paths. This simplifies the synaptic circuit and results in a 1T1N like configuration. We then study in detail the impact of two different programming parameters: pulse width and pulse amplitude, on the number of distinct conductance levels and energy per transition of the proposed Skyrmion on Nanotrack (Sk-N) synapse. Using this optimized nanotrack and programming parameters we propose a 1 Transistor/1 nanotrack (1T1N) Skyrmion on Nanotrack (Sk-N) based synaptic bit-cell that changes the conductance of the synapse according to a modified version of the biological unsupervised STDP learning rule [21]. We also design peripheral subthreshold CMOS neuron circuit comprising of following blocks: current scaling circuit, Differential Pair Integrator adopted from [22], [23], comparator and a custom spike generator unit. The current scaling circuit used for this work is based on the Gilbert normalizer circuit [24] and was recently shown to work with two-terminal differential memristive synapses [25]. In [25], the circuit was included inside each synapse bit-cell. In this work, we show how the same circuit can be used effectively for 4-
Terminal non-differential Sk-N based synapses by coupling it with each post-neuron rather than the synaptic bit-cells, thus saving silicon footprint. The spike generator unit is able to generate custom pre and post-neuronal spikes that lead to the modified STDP learning rule. On combining all the circuits we demonstrate spike transmission, neuronal integration and simplified STDP based conductance modulation in a single synaptic bit-cell through circuit simulations. The circuit simulations were done in CADENCE Spectre, using TSMC 65 nm technology node PDK. The characteristics of our 1T1N synaptic bit-cell and programming circuit were used in system-level simulations to demonstrate online unsupervised learning in a Spiking Neural Network (SNN) in the MATLAB based neuromorphic hardware simulator, MASTISK [26]. The task involved pattern recognition in noisy video streams and multi-class classification of handwritten digits (MNIST).

II. SKYRMION ON NANOTRACK

A. Skyrmion Physics

The creation of a skyrmion can be attributed to the competition between ferromagnetic exchange coupling and Dzyaloshinskii-Moriya interaction (DMI) in magnetic systems. These systems have breaking or lacking inversion symmetry in bulk lattices [14] or at the interface of thin films [13] respectively. The DMI interaction between two neighbouring atomic spins in a lattice having large Spin - Orbit Coupling (SOC) is given by:

$$H = -D (S_i \times S_j),$$

where D denotes the DMI vector, and $S_i$ and $S_j$ are the spins on site i and j, respectively. In this work we are mainly concerned with skyrmions created in interface of ultrathin films (also known as nanotrack) comprising of a Ferromagnetic-Metal (FM) layer deposited on a Heavy-Metal (HM) layer. In such systems spin polarised current is used to disturb the magnetization equilibrium of the metal layer and induce topological transition of the magnetic textures. This when facilitated by DMI leads to creation of skyrmions. Skyrmions on nanotacks can be moved by either Current-In-Plane (CIP) or Current-Perpendicular-to-Plane (CPP). In this work the latter case is implemented due to its energy efficiency [27]. According to this method a charge current flowing through the HM layer induces a vertical spin current due to Spin-Hall Effect (SHE) which is responsible for moving the skyrmions in the FM layer by exerting spin torque. Due to the non-linear spin texture of skyrmions, there will be definite change in the site-dependent spin mixing of magnetic states in the ferromagnetic environment. This can be detected using Tunneling Magneto Resistance (TMR) measurements [28].

B. Skyrmion on Nanotrack Synapse

The concept of skyrmion on nanotrack synapse used for this work is shown in Fig. 1. It has ultrathin metallic films comprising of a Ferromagnetic Metal (FM) layer deposited on top of a Heavy Metal (HM) layer. The FM layer has Perpendicular Magnetic Anisotropy (PMA) in the direction specified by \( \hat{y} \) (Fig. 1). A metallic capping layer with a stronger PMA than the FM layer is used as an energy barrier for the skyrmions, thus separating the nanotrack into two parts: pre-synapse and post-synapse. Please note that the terminologies pre-synapse and post-synapse have been adopted from [20]. They just refer to the two sides of the synapse separated by the PMA barrier and have no relation with pre-neuron and post-neuron. MTJ like structures are used at either ends of the nanotrack for nucleating (write-MTJ) and reading (read-MTJ) skyrmions respectively. The write-MTJ (pinned layer is the portion of FM layer below it) is used to introduce spin-polarised current into the pre-synapse region in order to create skyrmions. The read-MTJ has its free layer separated from the FM layer with an insulating magnetic material that allows magnetic coupling but prevents current flowing across it. The pinned and free layers of the Read-MTJ are separated by non-magnetic spacer. This decouples the read and write/program paths of the nanotrack completely. The benefit obtained by doing so is discussed in Section IV. The read current (\( I_{Syn} \)) flowing through the write-MTJ (shown in Fig. 1) will be proportional to the conductance that is determined by the spin configurations of the post-synaptic FM layer and the free and pinned layers of the read-MTJ. Since the spin direction of the FM layer is opposite to that of the pinned layer of read-MTJ, it is in Anti-parallel (AP) state in the absence of skyrmions. As the net spin of a skyrmion in the \( \hat{y} \) direction is zero, its presence in the post-synaptic region is equivalent to removal of a portion of the FM layer along with its spin. This leads to a less AP nature of the read-MTJ and therefore an increase in conductance. Therefore higher the number of skyrmions in the post-synapse, higher is the conductance of the post-synapse and the corresponding read current flowing through it. Programming current \( I_P \) (charge current) when passed through the HM layer, the spin current caused by it as a consequence of Spin Hall Effect (SHE) is responsible for moving the skyrmions from one end of the nanotrack to the other.

C. Micro-Magnetic Simulation Framework

The skyrmion-based artificial synaptic device was numerically simulated by 3D magnetization dynamics in the Object Oriented MicroMagnetic Framework (OOMMF) software [29] which used the extended Landau-Lifshitz-Gilbert equation including spin transfer torques as follows [30]:

$$\mathbf{M} = \mathbf{M}_0 \times \mathbf{H}_a + \mathbf{M} \times \mathbf{H}_o + \gamma \mathbf{M} \times \mathbf{B} + \alpha \mathbf{M} \times \mathbf{M} \times \mathbf{B} + \xi \mathbf{M} \times \mathbf{H}_s,$$
\[
\frac{dm}{dt} = -|\gamma_0| m \times h_{\text{eff}} + \alpha (m \times \frac{dm}{dt}) + \frac{u}{t} m \times (p \times m),
\]

where \( \gamma \) is the gyromagnetic ratio, \( m = \frac{M}{M_s} \) is the reduced magnetization, \( h_{\text{eff}} = \frac{H_{eff}}{M_s} \) is the reduced effective field, \( \alpha \) is the Gilbert Damping Factor, \( u = |\gamma_0 h| \frac{M_s}{2\mu_0} \), \( h \) is the reduced Planck constant, \( j \) is the applied current density, \( P = 0.6 \) is the spin polarization, \( \mu_0 \) is the permeability of free space, \( e \) is the electron charge, \( p \) is the unit spin polarization direction and \( p = -\hat{y} \) was set for driving the skyrmions. The parameters were extracted from experiments \([30]\) for a 0.4 nm thick Co layer on a Pt substrate.

In order to calculate the conductance of the post-synaptic region Tunneling Magneto Resistance (TMR) has been employed. This is done by dividing the read - MTJ into multiple 2 nm \( \times \) 2 nm cells in the x-y plane and calculating the conductance of each cell by Julliere’s model \([31]\):

\[
G = G_0 \left( \frac{1 + p^2 \cos \theta}{1 + p^2} \right),
\]

putting \( \theta \Rightarrow \pi - \theta \)

\[
G = G_0 \left( 1 - \frac{p^2 \cos \theta}{1 + p^2} \right),
\]

where \( G_0 \) is the conductance when the magnetization is perfectly parallel to the reference layer, \( p \) is the spin polarization and \( \theta \) is the magnetization of each cell with respect to the reference layer. However since in this work the spins of the reference layer of MTJ and FM layer are oppositely oriented, \( \theta \) is replaced by \( \pi - \theta \) leading to the form shown in Eq. \([3]\).

As a result the introduction of skyrmions in the post-synapse leads to \( \theta \) being less than \( \pi \) which gives a higher value of \( G \) according to Eq. \([4]\).

| TABLE I: Device parameters for the HM-FM heterostructure |
| --- | --- |
| Parameter | Description | Value |
| \( M_s \) | Saturation Magnetization | 580 kA/m |
| A | Exchange Constant | 15 pJ/m |
| D | DMI Factor | 3 mJ/m² |
| \( \alpha \) | Gilbert Damping Factor | 0.3 |
| K | Magnetic anisotropy | 0.8 MJ/m⁴ |
| P | Spin Polarization | 0.6 |
| t | Thin Film Thickness | 1 nm |
| 1 \( \times \) w | Nanotrack Length and Width | 820 nm \( \times \) 280 nm |
| \( \rho \) | Resistivity of HM layer | 100 \( \mu \)Ω·cm |

III. NANOTRACK PROGRAMMING PARAMETERS

In the nanotrack simulations it was found that when skyrmions try to flow across the PMA barrier, there exists a competition between: driving force due to current, repulsive force due to nanotrack edge, PMA barrier and inter-skyrmion (Sk-Sk) interactions. If repulsive and driving forces are not optimized it leads to leaky movement of skyrmions across the barrier after the programming pulses are removed. This makes the conductance levels dependent on the inter-pulse delay as was the case in \([20]\) and thus cannot be used for stable temporal long-term (LTP/LTD) non-volatile synaptic emulation. Each skyrmion crossing over to the post-synapse region increases the conductance of the read path. Thus, maximum number of conductance levels that can be achieved is the total number of skyrmions nucleated in pre-synapse. If multiple skyrmions enter the post-synapse under the influence of a single pulse then the effective number of distinct conductance states is reduced. The maximum number of stable skyrmions that are created during nucleation depends on the width of nanotrack \([20]\). Therefore the number of conductance levels is proportional to the width of the nanotrack. Increasing the length, allows more room for sequential movement of skyrmions in response to driving currents. However a large width and length would also incur synaptic area overheads. Therefore dimension of the nanotrack was fixed at: 820 nm \( \times \) 280 nm \( \times \) 1 nm. It allowed nucleation of 17 skyrmions in the pre-synapse during initiation. The spin polarisation value was set to 0.6 so as to improve the control of driving current on the movement of skyrmions across the PMA barrier. The micromagnetic simulation parameters are listed in Table \([1]\).

The variation of conductance levels and programming energy per pulse with current density is shown in Fig. \([2]\) (a). As the current density increases, the number of conductance states decrease. This is because with higher values of current, the events when multiple skyrmions cross the barrier increases.
However if the current density is lower than a certain threshold ($I_{th}$) then it fails to move the skyrmions across the barrier. The current pulse width was kept at a constant value of 2 ns. The energy values are calculated using Equation 5:

$$E_{\text{Spike}} = \rho \times t \times l \times w \times J^2 \times T_W,$$

where $\rho$ is the resistivity of Pt thin film [32], $t$ and $w$ are the thickness and width of HM layer and $J$ and $T_W$ are the current density amplitude and width of the pulses used for programming. The variation of conductance levels and energy for different programming pulse widths is shown in Fig. 2(b). Current density of the pulses used was 5 MA/cm$^2$. As the pulse width was decreased, the time window allowed for multiple skyrmions to cross the barrier decreased and therefore the number of distinct conductance levels increased. It was also found that just after a pulse ended, there was some transient component in the skyrmions’ velocity. The maximum inter pulse delay required for the skyrmion movement across the barrier to stabilise was found to be 5 ns. This puts a maximum limit on the operating frequency of neurromorphic systems built with this nanotrack. In all the simulations only those conductance states are reported which were non-volatile and did not involve any skyrmion crossing barrier after removal of programming pulse. The programming parameters used in this work are: current density = 5 MA/cm$^2$ and pulse width = 2 ns and has been circled in red in Fig. 2. Using these parameters, consecutive potentiating (depressing) pulses which moved skyrmions from pre to post-synaptic (post to pre-synaptic) region were applied. The position of skyrmions in the nanotrack after different pulses are shown in Fig. 3(a) whereas the conductance and post-synaptic skyrmion population variation with pulses is shown in Fig. 3(b). The conductance modulation curve is nearly linear and symmetrical in both LTP and LTD phases, which is a desirable property in electronic synapses [33].

### IV. 1T1N Synapse and Simplified STDP

The proposed 1T1N synaptic circuit along with the pre and post-neurons are shown in Fig. 4(a). The proposed pre/post-neuronal spikes (programming methodology) are shown in Fig. 5 In our proposed synaptic design there are four possible modes of operation; (1) idle, (2) spike transmission, (3) potentiation and (4) depression. The different components of a neuron used in this work are: Current scaling circuit, integrator, comparator and a spike generator. The circuit and functioning of each of these components are discussed in detail in Section V. Spike transmission occurs when there is a pre-spike i.e. $V_{S1}^{post}$ is HIGH. The pre-spike voltage applied across the terminals B and C result in a current proportional to the conductance of the synapse ($I_{syn}$), flowing out of C and entering the post neuron through the current scaling circuit. Consequently a current $I_{Spike}$ enters the integrator and keeps getting integrated in the form of voltage ($V_{mem}$) across a capacitor. When $V_{mem}$ crosses a threshold, two different spikes are generated: $V_P - V_N$ which is the post-neuronal spike for the synapse under consideration and $V_{S1}^{post}$ which is the pre-neuronal spike for the next layer of synapses. Although during spike transmission the gate of T1 receives a HIGH voltage, if there is no post-neuronal spike it will be in cutoff and no significant current will flow between A and D. In the case when only post spike occurs, the synaptic circuit remains in idle mode. This is because there is no current flow through the synapse as the transistor T1 is off. Plasticity (LTP or LTD) occurs when both the pre and post-spikes have temporal coincidence. When the positive part of post-spike coincides with the pre-spike as shown in Fig. 5(a), LTP occurs (conductance of read-MTJ increases). At the onset of the pre-spike, T1 is in cutoff and spike transmission takes place from B to C. As soon as the post-spike arrives, T1 starts conducting and a current (whose magnitude depends on HM layer resistance and $V_P - V_N$) flows from A to D, moving the skyrmions to the right. Its worth noting that spike transmission takes place between B and C even when the post-neuronal spike is driving skyrmions across the barrier in the HM layer. This is the unique advantage provided by our 4 Terminal decoupled read-program nanotrack. In case of a 3-terminal nanotrack [9], [20] separate transistors would be required for switching off spike transmission during programming mode, since both operations take place through a common node. LTD occurs when the negative part of the post-spike coincides with the pre-spike (Fig. 5(b)). This results in current flowing from D to A, thus moving the skyrmions from post to pre-synaptic region and thus reducing the conductance of read-MTJ. For certain range of temporal spacing between pre and post-neuronal spikes, portions of both the positive and negative parts of the post-neuronal spike coincides with the pre-neuronal spike (shown in Fig. 5(c)). In such a case both LTP and LTD would take place in degrees proportional to their extent of their overlap. Based on the micromagnetic simulations and nanotrack parameters described in the previous section, the characteristics of pre/post neuron spikes used are: $t_{S2} = 2$ ns, $t_{S1} = 22$ ns and $t_a = 17$ ns. Therefore the temporal conditions for which LTP, LTD and both occur are: $3 \leq \Delta t < 22$, $-21 < \Delta t \leq -2$ and $-2 < \Delta t < 3$ respectively, where $\Delta t = t_{post} - t_{pre}$. The resultant STDP characteristic curve showing the percentage change in conductance for different values of $\Delta t$ is shown in Fig. 6. The chosen spike scheme leads to synaptic programming energy consumption of $\sim 1.2 \ \mu J$ per spike, (Equation 5). The arrangement of our proposed 1T1N synaptic bit-cell in

---

Fig. 3: (a) Micromagnetic simulations showing non-volatile conductance modulation of our skyrmionic synapse. First 7 pulses (5 MA/cm$^2$, 2 ns width) move the skyrmions from pre to post-synapse region (LTP) and the next 7 vice-versa (LTD). (b) Evolution of post-synapse conductance.
crossbar arrangement is shown in Fig. 4 (b). All bit-cells in a column share the same post-neuron whereas all those in a row share the same pre-neuron. Therefore currents from all the synaptic bit-cells in a single column that receive pre-neuronal spike contribute to $I_{Syn}$ that enters the post-neuron below. All the bit-cells in a column receive post-neuronal spike across $V_P$ and $V_N$, but only those are programmed whose transistor is ON due to simultaneous presence of pre-spike. Also note that due to the decoupled write and read paths in our proposed synaptic structure, conductance modulation and spike transmission can both occur simultaneously, thus leading to faster unsupervised learning. Whereas in the structure proposed in [20] due to a shared node between the two paths only one function (spike transmission or conductance modulation) could occur at a time. Considering that the same learning rule is being implemented, an extra transistor would be required to switch off the spike transmission path when the programming pulses are being applied in that case.

V. SUBTHRESHOLD CMOS NEURON CIRCUIT

In this section we present the schematics, functioning and outputs of different building blocks of the CMOS based neuron circuit. All simulations were done in CADENCE Spectre, using TSMC 65 nm technology node PDK. The various circuit parameters used for simulation has been detailed in Table II.

A. Current Scaling Circuit

The CMOS circuit involved in scaling and normalizing the current entering the post-neuron during spike transmission is shown in Fig. 7. The total current from synapses $I_{Syn}$ in the previous layer enter the post-neuron via this current scaling circuit block as $I_{Spike}$. The transistors N2-N6 operate in their subthreshold regime. This is ensured by making the bias current ($I_B$) controlled by $V_B$ and is very small, of the order of a few hundreds of nano-amperes. In that case the currents $I_{Spike}$, $I_{ref}$ and $I_B$ are given by equation 6. The currents $I_{Syn}$ and $I_{Spike}$ are determined by equation 6:

$$\begin{align*}
I_{Spike} &= I_0 e^{\frac{k(V_p - V_c)}{VT}} \\
I_{ref} &= I_0 e^{\frac{k(V_{ref} - V_c)}{VT}} \\
I_B &= I_0 e^{\frac{kV_B}{VT}}
\end{align*}$$

(6)
On applying Kirchoff’s law at node C, one can eliminate \( V_C \) to get the relation between \( I_{\text{Spike}1} \) and node voltage \( V_x \), as shown in equation [7]

\[
I_{\text{Spike}1} = I_b \left( \frac{e^{\frac{kV_x}{T}}}{e^{\frac{kV_{VT}}{T}} + e^{\frac{kV_{m}}{T}}} \right)
\]  

Voltage \( V_x \) is a function of the current \( I_{\text{syn}} \) and the exact relation between them depends on whether the transistor N1 operates in subthreshold or superthreshold regime. The PMOS transistors N5 and N6 are used as to make a current mirror that converts the sinking current flowing through N2 into a sourcing current flowing through N6 and therefore \( I_{\text{Spike}1} \approx I_{\text{Spike}} \).

As \( I_{\text{Spike}} \) gets integrated in the neuron’s integrator block, its input impedance keeps increasing. Therefore in order to prevent the transistor N6 from going into cutoff region while trying to source the same amount of current into a larger impedance load, its supply voltage (Vdd2) was kept at a slightly higher level (650 mV) as compared to Vdd1 (600 mV).

![Fig. 8: Current response of the scaling and normalizing block for different values of cumulative resistance of the synapses connecting it to neurons in previous layer.](image)

Voltage \( V_x \) is a function of the current \( I_{\text{syn}} \) and the exact relation between them depends on whether the transistor N1 operates in subthreshold or superthreshold regime. The PMOS transistors N5 and N6 are used as to make a current mirror that converts the sinking current flowing through N2 into a sourcing current flowing through N6 and therefore \( I_{\text{Spike}1} \approx I_{\text{Spike}} \).

As \( I_{\text{Spike}} \) gets integrated in the neuron’s integrator block, its input impedance keeps increasing. Therefore in order to prevent the transistor N6 from going into cutoff region while trying to source the same amount of current into a larger impedance load, its supply voltage (Vdd2) was kept at a slightly higher level (650 mV) as compared to Vdd1 (600 mV).

B. Integrator, Comparator and Spike Generator

The CMOS circuit that constitutes the remaining of our post-neuron is shown in Fig. [9](a). It starts with a Differential Pair Integrator filter for integrating incoming current signal (M1-M3). It forms the integrator block of our post-neuron. The voltage accumulated across the capacitance \( C_{\text{mem}} \) acts as the neuron’s membrane potential. The inverters constituted by M5-M6 and M7-M8 play the role of comparator and spike event generating block. The circuit beyond it comprising of the Flip-Flop, AND gates, M9-M10 and R1-R2 is responsible for generating the spikes as specified in Fig. [5] whenever a spike event is encountered.

The expression governing the dynamics of a generic non-linear integrate and fire system is given by equation [8][34],

\[
\frac{dv}{dt} = F(u) + G(u)I
\]

Assuming all transistors to be in subthreshold saturation and applying translinear principle, one gets the expression for \( I_{\text{out}} \) as shown in equation [10] where the currents \( I_{\text{out}} \), \( I_t \), \( I_e \) are given by equation [9] and \( \tau = C_{\text{mem}}V_T/kI_r \).

\[
I_{\text{out}} = I_0 e^{\frac{kV_{\text{mem}}}{V_T}} \quad I_t = I_0 e^{\frac{kV_{\text{thr}}}{V_T}} \quad I_e = I_0 e^{\frac{kV_{\text{min}}}{V_T}}
\]

On replacing \( I_{\text{out}} \) with \( I_0 e^{\frac{kV_{\text{mem}}}{V_T}} \) (see equation [9]), one obtains the temporal dynamics of \( V_{\text{mem}} \) shown in equation [11].

\[
\frac{dv_{\text{mem}}}{dt} = \frac{I_{\text{I}}I_{\text{Spike}}}{I_r(I_{\text{in}} + I_t)} - \frac{V_T}{k}
\]

It is worth noting here that \( I_{\text{Spike}} \) is the input current to the DPI integrator block and \( I_{\text{out}} \) is a function of \( V_{\text{mem}} \). Therefore both equation [10] and [11] represent non-linear integrate and fire dynamics (see equation [8]) in variables \( I_{\text{out}} \) and \( V_{\text{mem}} \) respectively.

The threshold voltage of the neuron is determined by the switching voltage of the inverter (M5-M6). This can be controlled by changing the dimensions (W/L ratio) of M5 and M6. Therefore as \( V_{\text{mem}} \) approaches the threshold, the output of the first inverter changes drastically to 0. In order to invert this output and provide spike events such that a high voltage is given by equation [9] and \( \tau = C_{\text{mem}}V_T/kI_r \).
and post-synaptic regions of our nanotrack synapse. Therefore the differential voltage \( V_P - V_N \) represent the post-neuronal spike, used to drive skyrmions in the synapses in the previous layer. As we saw in Section III, different current amplitudes and pulse widths may lead to different conductance levels and energy dissipation in the synapse, one can control the amplitude of \( V_P - V_N \) by varying the values of R1, R2 and W/L ratios of M9 and M10, whereas the temporal parameters of the spikes \( t_{S1}, t_{w}, t_{S2} \) discussed in Section IV can be controlled by the pulse widths and duty ratios of CLK, POS, NEG. The time evolution of global clock signals, spike events and output spikes \( (V_{S1}, V_{P} - V_{N}) \) are shown in Fig. 9 (b). The temporal parameters of CLK, POS and NEG (shown in Fig. 9 (b)) were chosen so as to generate spikes with parameters as discussed in Section IV.

In order to demonstrate the working of the proposed 1T-1N STDP powered skyrmionic synapse with neuron circuit, we considered a single synaptic bit-cell between a pre and post-neuron. The pre-neuron was modeled as a pulse source, generating \( V_{PRE}^{S1} \) with pulse width 23 ns and a frequency \( \sim 27 \) MHz. The post-neuron was simulated with the circuit discussed in this section. The HM layer was modeled by a resistance of 3 k\( \Omega \). The time evolution of \( V_{POST}^{S1}, V_{mem} \) (membrane potential), programming current flowing through HM layer and resistance of read MTJ is shown in Fig. 10. It is worth noting here that the sign of the programming current shown in Fig. 10 depends on the time \( (\Delta t) \) with which the pre and post-neuronal spikes are separated (exact dependence has been discussed in Section IV). Accordingly current which is positive, negative or both might flow through the synapse, resulting in LTP, LTD or a combination of LTP and LTD respectively. The energy consumed in the neuron circuit (including the current scaling block, integrator, comparator and spike generation blocks) was found to be 0.25 pJ/spike for a supply voltage \( (V_{dd}) \) of 600 mV and neuron firing rate of \( \sim 2.3 \) MHz.

**VI. UNSUPERVISED LEARNING WITH SKYRMION SYNAPSE**

In order to validate the working of our proposed skyrmion on nanotrack synaptic bit-cell for unsupervised learning we
simulated two applications. In the first, we simulated a fully connected feed forward two-layer Spiking Neural Network (SNN) (see Fig. 11) powered by Spike-Timing Dependent Plasticity (STDP). All simulations were done on the MATLAB based neuromorphic simulator named MASTISK [26]. The network has 90,000 spiking pixels in the first layer and 2 LIF neurons in the output layer. The two layers are connected by excitatory synapses in an all-to-all fashion. The output neurons are connected to each other through inhibitory synapses. This is done to implement Winner-Take-All mechanism [35]. Video stream comprising of binary 300 × 300 frames were used for training the SNN as shown in Fig. 11 (b)-(d). The video stream comprises of mostly gaussian noise frames (see Fig. 11 (b)) with two complex patterns (IIT-D logo and BUAA logo) embedded at different timestamps (Fig. 11 (c)-(d)). The input layer encodes the frames into spikes using poisson spike encoding [39] with the mean firing rate proportional to the intensities of the pixels. Fig. 12 (a) shows the raster plot of input neurons during the entire training period. Spiking activity of the neurons are denoted by dots in Fig. 12 (a) and the color denotes the kind of frame that was presented. Black dots represent noisy frame whereas the red and blue dots represent the frames containing the IIT-D logo and BUAA logo respectively. The input video frames at specific timestamps are illustrated in Fig. 12 (b). Fig. 12 (c) and (d) show the evolution of conductance of the synapses connecting the output neurons 1 and 2 to the input layer respectively, thus depicting the representations learned by each of the output neurons. Neuron 1 became selective to BUAA logo, while neuron 2 got selective to IIT-D logo. Initially the output neurons fire randomly, however from 1000 ns onwards as the occurrence of patterns increases the spiking becomes more specific to occurrence of a particular input pattern (Fig. 12 (c)). In order to study the degree of selectiveness of the neurons to different patterns, we separately noted the conductance of the synapses connecting the output neurons to different types of input pixels. The input pixels could either be pattern pixels and carry information for the two different patterns or be background pixels. The input pixels could either be pattern pixels and carry information for the two different patterns or be background pixels. The input pixels could either be pattern pixels and carry information for the two different patterns or be background pixels. The input pixels could either be pattern pixels and carry information for the two different patterns or be background pixels.

The averaged conductance of the synapses between the input pixels of the three types and two output neurons are shown in Fig. 13. The logo with maximum averaged conductance is the one for which a particular neuron gets selective. The difference between the conductances of the two patterns for a particular neuron depicts how well it can differentiate between the two patterns. The low conductance of the background synapses (noise) shows that the network is able to get selective to patterns and not background noise. The system achieves a low false positive spike rate of 6.5 × 10⁻⁴ (inset of Fig. 13). Since output neuron firing activity starts around 350 ns there is a transient in the false positive spike rate of either neuron. Ultra-low on-line unsupervised learning synaptic programming power consumption of ~ 1 nW / synapse and neuron firing power consumption of ~ 1.64 μW / neuron was achieved (Total programming events: ~ 4.6 × 10⁵; Total post-neuron firing events: 40; synaptic programming energy: 1.2 fJ; CMOS neuron energy per firing event: 0.25 pJ; duration: 3050 ns; total synapses: 1.8 × 10⁵). The power takes in account both: energy spent in moving skyrmions in the nanotrack, and integration of I_{Spike} in DPI integrator, comparator and the spike generation circuit over the entire duration of learning.

Table III compares our proposed skyrmion synapse with other nanodevice based synapses. It clearly highlights the ultra-low energy (1.2 fJ/event) and sub-nanosecond time (2 ns) consumed in changing conductance states in our synapse. Since synapses based on nanotrack type structure (domain wall or skyrmion based) usually have 3 or more terminals, it becomes difficult to use nanotrack based synapses in 1R configuration (synaptic bit-cell containing only nanotrack and no transistor). For instance domain wall on nanotrack based synapse proposed in [9] had 3 transistors apart from the nanotrack in their synaptic bit-cell. However we show that using our proposed skyrmion on nanotrack configuration and neuron circuit one can implement a modified version of STDP learning rule with only 1 transistor along with the nanotrack in a bit-cell. A limitation of our synapse is the large dimensions of the nanotrack (1.5 % larger than [9]). The length of the nanotacks determines the degree with which the skyrmions can freely move in the nanotrack without interacting with each other. It also has an impact on number of distinct conductance levels.

**TABLE III: MNIST Classification Accuracy for different synapses**

| Synapse Type | Accuracy  |
|--------------|-----------|
| Redundancy = 2 | 73.8 %  |
| Redundancy = 4 | 82.17 % |
| Redundancy = 6 | 84.97 % |
| Ideal Synapse | 85.5 %   |

For the second application, we simulated a 3-layer SNN comprising of input, output (excitatory) and inhibitory neurons respectively inspired from [37]. The network was trained on 60,000 training images of the MNIST database. In the interference mode, the skyrmion on nanotrack characterized by its properties depicted in Fig. 3 was used as synapse. We varied the number of nanotacks connected in parallel to constitute a single synapse in order to see how performance varied with the redundancy. The different levels of redundancy and their corresponding classification accuracies have been shown in Table III. The synapse with redundancy = 6 almost reaches the performance level of ideal synapse (infinite dynamic range and conductance levels).

**VII. Discussion**

Table IV compares our proposed skyrmion synapse with other nanodevice based synapses. It clearly highlights the ultra-low energy (1.2 fJ/event) and sub-nanosecond time (2 ns) consumed in changing conductance states in our synapse. Since synapses based on nanotrack type structure (domain wall or skyrmion based) usually have 3 or more terminals, it becomes difficult to use nanotrack based synapses in 1R configuration (synaptic bit-cell containing only nanotrack and no transistor). For instance domain wall on nanotrack based synapse proposed in [9] had 3 transistors apart from the nanotrack in their synaptic bit-cell. However we show that using our proposed skyrmion on nanotrack configuration and neuron circuit one can implement a modified version of STDP learning rule with only 1 transistor along with the nanotrack in a bit-cell. A limitation of our synapse is the large dimensions of the nanotrack (1.5 % larger than [9]). The length of the nanotacks determines the degree with which the skyrmions can freely move in the nanotrack without interacting with each other. It also has an impact on number of distinct conductance levels.
TABLE IV: Comparison with other synapses

| Device                        | Dimension     | Programming Energy | Programming Time | Synapse Configuration | Terminals | Type         | Programming Neuron Spikes |
|-------------------------------|---------------|--------------------|------------------|-----------------------|-----------|--------------|--------------------------|
| Ag-Si memristor [38]          | 100 nm x 100 nm | 25.2 pJ - 403.2 pJ | 300 us           | 1R                    | 2         | Experimental | Identical                |
| PCM [39]                      | length: 500 nm, BE diameter: 75 nm | LTD (avg) = 13.5 pJ LTP (avg) = 50 pJ | LTD transition: 75 ns LTP transition: 5 us | 1R        | 2         | Experimental | Non-Identical            |
| AlOx/HfO2 Bilayer RRAM [40]   | 21 nm thick   | LTP (avg) = 3.24 nJ LTD (avg) = 4 nJ | 100 us           | 1T1R                  | 3         | Experimental | Identical                |
| Domain Wall synapse [9]       | 320 nm x 20 nm | 48 J/event         | 1 ns             | 3T1R                  | 5         | Simulation   | Pres-Spikes = Complex exponential, Post-Spikes = identical pulses |
| 1T1N Skyrion Synapse [This Work] | 820 nm x 280 nm | 1.2 J/event       | 2 ns             | 1T1R                  | 4         | Simulation   | Identical                |

levels that can be achieved. In order to achieve the same number of conductance states with shorter length it would require scaling down skyrmion sizes. In our simulations we used skyrmions with diameter ~ 15 nm. Recently studies have shown that size of skyrmions can even be reduced to 1-3 nm [30], [41]. This opens possibilities of reducing the nanotrack length further without losing number of conductance levels.

This work is based on simulations done holistically at the device, circuit and architectural level with parameters calibrated to experimental results [30]. It clearly highlights the benefits of neuromorphic systems built with hybrid CMOS-skyrmion circuits.

VIII. CONCLUSION

We illustrate an approach for practical realization of multi-level synapse using hybrid skyrmion-CMOS based spiking neuromorphic systems based on simulations done at device, circuit and system level. Firstly through micromagnetic simulations we study in detail the impact of different programming parameters of a HM/FM nanotrack on various synaptic performance parameters and demonstrate true non-volatile multi-level conductance states, independent of inter-spike delay or frequency. We use a read-MTJ on top of the post-synaptic region of the nanotrack separated by insulating magnetic material from the FM layer beneath it. This makes our synapse 4-terminal with completely decoupled read and program paths. Leveraging the conductance modulation, we propose a 1T1N synaptic architecture and programming methodology to implement a modified version of the biological STDP rule, while consuming ~ 1.2 fJ energy per programming event. We design a custom subthreshold synchronous spike generator circuit which when coupled with a current scaling circuit, Differential Pair Integrator and inverter based thresholding circuit, can perform the desired functionalities of a Leaky-Integrate and Fire neuron at extremely low energy (0.25 pJ/output spike). Unsupervised learning is demonstrated by simulating feed-forward SNN for pattern extraction and multi-class classification application.

REFERENCES

[1] G. W. Burr, R. M. Shelby, S. Sidler, C. di Nolfo, J. Jang, I. Boybat, R. S. Shenoy, P. Narayanan, K. Virwani, E. U. Giacometti, B. N. Kurdi, and H. Hwang, “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, vol. 62, no. 11, pp. 3498–3507, Nov 2015, doi: 10.1109/TED.2015.2439635.

[2] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. L. S. K. Esser, A. Appuswamy, Y. Tabba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, and D. S. Modha, “A million spiking-neuron integrated circuit with a scalable communication network and interface," Science, vol. 345, no. 6197, pp. 668–673, Aug 2014, doi: 10.1126/science.1254642.

[3] G. Indiveri, E. Chicca, and R. Douglas, “A visi array of low-power spiking neurons and bistable synapses with spike-time dependent plasticity," IEEE Transactions on Neural Networks, vol. 17, no. 1, pp. 211–221, Jan 2006, doi: 10.1109/TNN.2005.868050.

[4] C. Mead, “Neuromorphic electronic systems," Proceedings of the IEEE, vol. 78, no. 10, pp. 1629–1636, Oct 1990, doi: 10.1109/5.583855.

[5] M. Suri, O. Bichler, D. Querlioz, G. Palma, E. Vianello, D. Vuillaume, C. Gamrat, and B. DeSalvo, "Ocm devices as binary synapses for low-power stochastic neuromorphic systems: Auditory (cochlea) and visual (retina) cognitive processing applications," in 2012 Internationa Electron Devices Meeting, Dec 2012, pp. 10.3.1–10.3.4, doi: 10.1109/IEDM.2012.6479017.

[6] S. Yu, Y. Wu, R. Jeyasingh, H. S. P. Wong, "An electronic synapse device based on metal oxide resistance switching memory for neuromorphic computation," IEEE Transactions on Electron Devices, vol. 58, no. 8, pp. 2729–2737, Aug 2011, doi: 10.1109/TED.2011.2147791.

[7] M. Suri, O. Bichler, D. Querlioz, O. Cueto, L. Perniola, V. Sousa, D. Querlioz, C. Gamrat, and B. DeSalvo, "Phase change memory as synaptic element in stochastic spiking neural networks with on-chip stdp learning," Scientific reports, vol. 6, p. 29545, July 2016, doi: 10.1038/srep29545.

[8] S. Xinivasan, A. Sengupta, and K. Roy, “Magnetic tunnel junction based long-term short-term stochastic synapse for a spiking neural network," Scientific reports, vol. 6, p. 29545, July 2016, doi: 10.1038/srep29545.

[9] A. Fert, V. Cros, and J. Sampaio, "Skyrmions on the track," Nature nanotechnology, vol. 8, no. 3, pp. 152–156, Mar 2013, doi: 10.1038/nnano.2013.29.

[10] S. Heinze, K. Von Bergmann, M. Menzel, J. Brede, A. Kubetzka, W. K. H. Muller, and G. Indiveri, “Spin transfer torque devices for spiking neural networks," in Proceedings of the IEEE, vol. 104, no. 10, pp. 2040–2061, Oct 2016, doi: 10.1109/JPROC.2016.2591578.

[11] M. Sondhi, V. Parmar, and M. Suri, "Mastisk: Simulation framework for design exploration of neuromorphic hardware," 2018 International Joint Conference on Neuronal Networks (IJCNN), in press.

[12] J. Sampaio, V. Cros, S. Rohart, A. Fert, "Nucleation, stability and current-induced motion of isolated magnetic skyrmions in nanostructures," Nature nanotechnology, vol. 8, no. 11, pp. 839–844, 2013.

[13] C. Griffiths, D. A. C. Hall, R. A. Buhrman, "Spin torque study of the spin hall conductivity and spin diffusion length in platinum thin films with varying resistivity," Phys. Rev. Lett., vol. 116, p. 126601, Mar 2016, doi: 10.1103/PhysRevLett.116.126601.

[14] S. Ito, K. Gomi, T. Kawai, K. Kondo, and H. Higuchi, "Improvement of on-chip performance of large-scale neural network (165 000 synapses) using phase-change memory as the synaptic weight element," Nano letters, vol. 10, no. 4, pp. 1297–1301, 2010.

[15] P. U. Diehl and M. Cook, "Unsupervised learning of digit recognition using pulse-coupled oscillators," Physical Review E, vol. 48, no. 2, p. 1483, 1993.

[16] D. Heeger, "Poisson model of spin generation," Handout, University of Standford. IEEE, 2009, pp. 1054–1060.

[17] D. Heeger, "Poisson model of spin generation," Handout, University of Standford. IEEE, 2009, pp. 1054–1060.

[18] D. Heeger, "Poisson model of spin generation," Handout, University of Standford. IEEE, 2009, pp. 1054–1060.

[19] D. Heeger, "Poisson model of spin generation," Handout, University of Standford. IEEE, 2009, pp. 1054–1060.

[20] D. Heeger, "Poisson model of spin generation," Handout, University of Standford. IEEE, 2009, pp. 1054–1060.