A Current-Mode Implementation of A Nearest Neighbor STDP Synapse

Akwasi Akwaboah, Ralph Etienne-Cummings
Department of Electrical and Computer Engineering
Johns Hopkins University
Baltimore MD, USA
{aakwabo1, retienne}@jhu.edu

2023 21st IEEE Interregional NEWCAS Conference (NEWCAS) | 979-8-3503-0024-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/NEWCAS57931.2023.10198113

***Abstract*—The Artificial Intelligence (AI) disruption continues unabated, albeit at extreme compute requirements. Neuromorphic circuits and systems offer a panacea for this extravagance. To this effect, event-based learning such as spike-timing-dependent plasticity (STDP) in spiking neural networks (SNNs) is an active area of research. Hebbian learning in SNNs fundamentally involves synaptic weight updates based on temporal correlations between pre- and post- synaptic neural activities. While there are broadly two approaches of realizing STDP, i.e. All-to-All versus Nearest Neighbor (NN), there exist strong arguments favoring the NN approach on the biologically plausibility front. In this paper, we present a novel current-mode implementation of a postsynaptic event-based NN STDP-based synapse. We leverage transistor subthreshold dynamics to generate exponential STDP traces using repurposed log-domain low-pass filter circuits. Synaptic weight operations involving addition and multiplications are achieved by the Kirchoff current law and the translinear principle respectively. Simulation results from the NCSU TSMC $180\ \mathrm{nm}$ technology are presented. Finally, the ideas presented here hold implications for engineering efficient hardware to meet the growing AI training and inference demands.**

***Index Terms*—analog computation, STDP, learning, local synaptic plasticity, three-factor, neuromorphic**

I. Introduction

Engineering efficient hardware to address the growing artificial intelligence (AI) demands in our time is of immense interest. State-of-the-art AI models such as large language models [1] are extremely profligate in their training compute consumption [2]. While delivering phenomenal gains in recognition, their advancement using existing hardware such as graphics processing units (GPUs) or tensor processing units (TPUs), are economically and environmentally unsustainable [3]. To address this challenge, neuromorphic engineers have been emulating neural physiology in silicon with a motivation of arriving at computational primitives that possess brain-level energy efficiency (~ 20 W) [4], [5]. In this light, we sought to implement a spike-timing-dependent plasticity (STDP) circuit suitable for adoption in analog/ mixed-signal neuromorphic systems. STDP is a hebbian learning mechanism that leads to synaptic weight changes as a function of the temporal correlations between the pre- and post- synaptic neuron activities. The causal (acausal) weight updates decay exponentially with increasing pre-post (post-pre) spike intervals as shown in eq.1.

$$\Delta w(\Delta t) = \begin{cases} A_+ e^{-\frac{\Delta t}{\tau_+}} & \Delta t > 0 \\ A_- e^{\frac{\Delta t}{\tau_-}} & \Delta t < 0 \end{cases} \tag{1}$$

where, $\Delta t = t_{post} - t_{pre}$, A_+ and A_- are the intial exponential intensities for the potentiating and depressive regimes respectively. Similarly, τ_+ and τ_- are the respective time constants. t_{pre} (t_{post}) is the pre- (post-) event time. STDP can be implemented in two main ways, i.e. the classical All-to-All (A2A) approach involving the pairing of all pre- (post-) synaptic activity with all possible past and future post- (pre-) synaptic activity and the Nearest Neighbor (NN) approach where parings of pre- (post-) activity are only with respect to the nearest post- (pre-) activity. There are strong arguments [6] favoring the biological plausibility of the NN approach. These include the reset of the membrane potential in the dendritic spine by the latest postsynaptic spike thus overriding the contribution of all preceeding postsynaptic spike. Other plausible reasons include calcium saturation along with AMPA receptor desensitization arising synaptic glutamate overload from bursting presynaptic activity [7]. But also from an implementation standpoint, the A2A approach is impractical for online learning due to the dependence on future information. Secondly, naively pairing all pre-post activity presents a memory allocation challenge for time spikes in the case of an epoch-based weight update. An online STDP requires the generation of causal and acausal STDP traces, which are typically exponential [8]. Conveniently, there exists an exponential current-voltage relationship for the MOS transistor operating in subthreshold. A literature review on the exercise of designing analog synaptic circuits that realize non-integrative pulse extension and log-domain filtering can be found in [9].

In this paper, we present a post-synaptic event-driven nearest neighbor STDP synapse circuit using analog current-mode subthreshold principles. We show simulation results using the NCSU TSMC $180\ \mathrm{nm}$ technology ($V_{dd} = 1.8$ V). In our previous work [10], we implemented a digital NN STDP that kept track of the earliest and latest presynaptic events in a postsynaptic spike interval and also made accomodation for a time-constant based linear approximation of the exponential STDP trace. While tracking only two presynaptic events may seem like an oversimplification, it gets around an indefinitely

large memory allocation caveat for all possible presynaptic events within an interval. Furthermore, preliminary exploration of the algorithm showed promise on a spatiotemporal feature extraction task. Herein, with the help of analog current-mode circuits, we advance to letting capacitor saturation determine the number of presynaptic event Δw contributions to consider. We also include additional parameters to potentially allow transcending the two-factor learning (involving only pre and post STDP trace contributions) to three or even four factor and continual learning rules which may incorporate other local factors such as the stochastic synaptic neurotransmitter release probability or even meta/ global influences such as homeostatic [9] control and dopaminergic reward [11]. In what follows, we discuss each of the subcircuits, and then a top level circuit combining all these units.

II. Sub-circuits & Theory

A. Non-Integrative Postsynaptic Pulse Extender

The set-discharge synapse circuit first presented by [12] is a compact (3T) non-integrative pulse extender with transistors operating in subthreshold (shown in Fig. 1). On an input event (active low), the switching transistor M_ϕ sets the capacitive node, $V_c(t)$ to $V_{\alpha\_}$ for the duration of the pulse. Upon pulse removal, transistor M_τ linearly leaks $V_c(t)$ to V_{dd} by a rate defined by $V_{\tau\_}$. This causes the M_o drain current to decay exponentially as shown in eq.2.

$$I_{\Delta w_a}(t) = I_o \cdot exp\left(-\frac{\kappa}{V_T}(V_c(t) - V_{dd})\right) \tag{2}$$

where κ is subthreshold slope factor [13], V_T is thermal voltage (25 mV at room temperature). Transistor size of $W/L = 8/8$ and a diode-connected load are used throughout this paper unless otherwise specified. We use this circuit for generating acausal STDP trace ($I_{\Delta w_a}$), which do not require integration, i.e. every postsynaptic event initiates a new phase of Δw computation. Ultimately, the time course definition of the $I_{\Delta w_a}$ is given by,

$$I_{\Delta w_a}(t) = \begin{cases} I_{\alpha\_} & \text{(set)} \\ I_{\alpha\_} \cdot exp\left(-\frac{t}{\tau\_}\right) & \text{(discharge)} \end{cases} \tag{3}$$

where $I_{\alpha\_} = I_o \cdot exp\left(-\frac{\kappa}{V_T}(V_{\alpha\_} - V_{dd})\right)$ and $\tau\_ = \frac{V_T C}{\kappa I_{\tau\_}}$. Thus, $A\_$ and $\tau\_$ from eq. 1 are made voltage-dependent free parameters such that, $I_{\Delta w_a} \sim f\left(A\_(V_{\alpha\_}), \tau\_(V_{\tau\_})\right)$. This dependence is demonstrated in Fig 1.

B. Integrative Presynaptic Pulse Extender

For providing pulse integration, we chose the differential pair integrator (DPI) synapse circuit [9] due to its accompanying linearity improvement, as well either the ability to scale capacitor charging to allow smaller pulse widths or relative compactness compared to other linear LPF circuits [14], [15] and [16] respectively. We use the DPI for generating a causal integrative STDP trace ($I_{\Delta w_c}$). Except the switching transistors M_ϕ and M_r, all transistors have size-matching enforced as

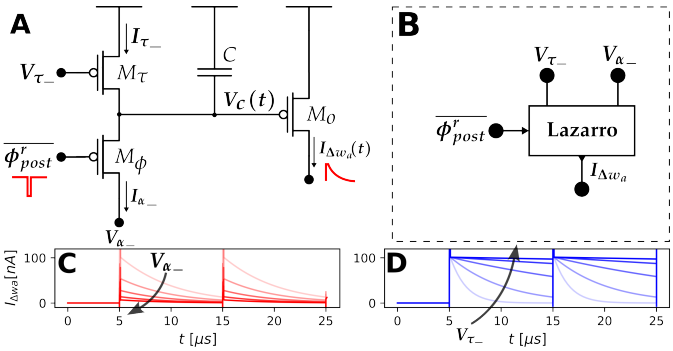

Fig. 1. **A**. The non-integrative set-discharge circuit schematic by [12] used for generating acausal Δw contribution, $I_{\Delta w_a}$.**B**: Schematic symbol. Simulation results for intensity, $A\_(V_{\alpha\_})$ variation in **C**. $V_{\alpha\_}$ was swept from 1.33 V to 1.45 V in steps of 30 mV for $V_{\tau\_} = 1.55$ V. **D** shows the time-constant, $\tau\_(V_{\tau\_})$ variation by sweeping $V_{\tau\_}$ from 1.5 to 1.7 V in steps of 50 mV while $V_{\alpha\_} = 1.33$ V. Common parameters values used in **C** and **D**: $C =$ 100 fF, pulse width of 100 ns and 10 μs period.

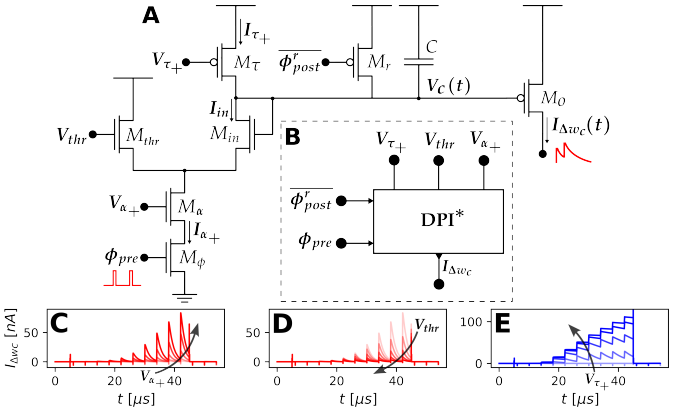

Fig. 2. **A**. Circuit schematic of the DPI circuit by [9] modified to include a postsynaptic event-controlled reset transistor M_r necessary for realizing the nearest neighbor feature. The corresponding schematic symbol is shown in **B**. The influence of V_{α_+} and V_{thr} are shown in **C** and **D** respectively for $V_{\tau_+} = 1.5$ V. In **C**, the more sensitive V_{α_+} is swept from 410 mV to 420 mV in steps of 2.5 mV, while $V_{thr} = 1.34$ V. In **D**, V_{thr} is swept from 1.34 V to 1.5 V in steps of 40 mV. **E** depicts current integration V_{τ_+} is varied from 1.5 V to 1.7 V in steps of 50 mV. Common parameter values for **C**, **D** and **E** are $C = 100$ fF, pulse width of 100 ns, ϕ^r_{post} period of 40 μs and ϕ_{pre} period of 4 μs

well as operate in subthreshold. The DPI operates as follows: an input event completes the path from V_{dd} to ground through transistors M_{in}, M_α and M_ϕ and ultimately sets the initial amplitude of the exponential trace. This amplitude is a function of V_{thr} and V_{α_+}, which intuitively controls the strength of the current I_{in}. An elevated V_{thr} chokes the I_{in} differential pair branch of the tail current set by V_{α_+}. Exploiting the exponential drain current and gate voltage relationship in transistor M_o along with capacitor current relation, the first order differential LPF equation shown in eq. 4 can be obtained.

$$\tau_+ \frac{dI_{\Delta w_c}}{dt} + I_{\Delta w_c} = \frac{I_{gain} I_{\alpha_+}}{I_{\tau_+}} \tag{4}$$

where $I_{gain} = I_o \cdot exp\left(-\frac{\kappa}{V_T}(V_{thr} - V_{dd})\right)$. The reader

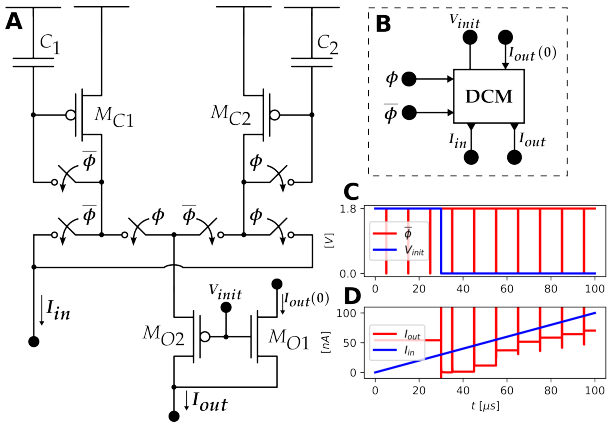

Fig. 3. **A**. Circuit Schematic for the dual-phase Dynamic Current Mirror used in memorizing input current. **B**. Schematic Symbol. **C**. shows the events $\overline{\phi}$ as the switching signal. The initialization control signal V_{init} is also shown. **D**. shows the initialization in the first 30 μs ($I_{out}(0) = 50$ nA) and subsequent input (ramp) current memorization. Parameter values are follows: $C = 100$ fF, pulse width of 100 ns and period of 10 μs

may refer to [9] for derivation and operating conditions. Ultimately, the time course definition of $I_{\Delta w_c}$ is given by:

$$I_{\Delta w_c}(t) = \begin{cases} \frac{I_{gain} I_{\alpha_+}}{I_{\tau_+}} + I^{\delta}_{\Delta w_c}(t) & \text{(charge)} \\ I^{+}_{\Delta w_c} \cdot exp\left(-\frac{t-t_i^+}{\tau_+}\right) & \text{(discharge)} \end{cases} \quad (5)$$

where $I^{\delta}_{\Delta w_c}(t) = \left(I^{-}_{\Delta w_c} - \frac{I_{gain} I_{\alpha_+}}{I_{\tau_+}}\right) exp(-\frac{t-t_i^-}{\tau_+})$ is the transient response and $\tau_+ = \frac{C V_T}{\kappa I_{\tau_+}}$. t_i^- and t_i^+ are the pulse start and end times respectively and $I^{-}_{\Delta w_c}$ and $I^{+}_{\Delta w_c}$ are the corresponding output current values. Similarly, A_+ and τ_+ from eq. 1 are voltage dependent free parameters such that, $I_{\Delta w_c} \sim f\left(A_+(V_{\alpha_+}, V_{thr}), \tau_+(V_{\tau_+})\right)$ This dependence is demonstrated in Fig.2.

C. Supporting Circuits

1) Dynamic Current Mirror: The dynamic current mirror (DCM) acts as a current-mode memory element deployed in latching synaptic weight and weight update current values in this work. The dual-phase DCM is suitable for simultaneous input memorizing (reading) and output sourcing (writing). Transfer of learned input current to the output is facilitated via switching ϕ_1 and ϕ_2 in a complementary non-overlapping manner. For simplicity, we set $\phi_1 = \phi$ and $\phi_2 = \overline{\phi}$, where ϕ is a postsynaptic-event-derived pulse. PMOS transistors are used here, hence switching is active low. The reader may refer to [17] for a detailed description of the DCM. We include a branch on the output to switch between sourcing a user-defined initial current, $I_{out}(0)$ and the learned current from M_{C1} post-initialization. Initialization is set by V_{init} high. In all (see Fig 6), two DCMs are used in the STDP cell – one for achieving $I_{\Delta w}(i+1)$ from $I_{\Delta w}(i)$ and the other for achieving $I_w(i+1)$ from $I_w(i)$. i is the postsynaptic interval index. See the current initialization and memorization in Fig.3.

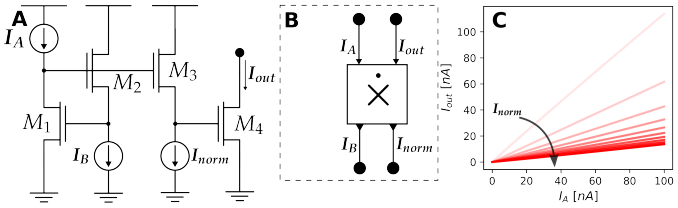

Fig. 4. **A**. Circuit Schematic for Translinear Multiplier. **B**. Schematic Symbol. **C**. shows simulation results of a dc sweep for I_A over the range of 0−100 nA with $I_B = 1$ nA. The effect of I_{norm} on I_{out} is demonstrated by stepping 1 nA to 10 nA in 1 nA increment

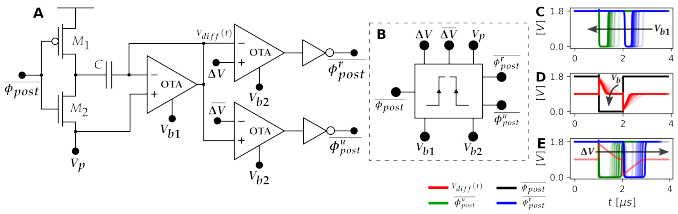

Fig. 5. The Pulse edge detection circuit shown in **A** (symbol in **B**) used for generating two non-overlapping edge-coupled daughter pulses for separating the DCM output current update and net STDP trace reset. **C**. shows how the two pulselets can be narrowed by increasing the bias voltage V_{b_1} set by current mirror input current I_{b_1}. I_{b1} is swept from 100 nA to 300 nA in 40 nA steps while $\Delta V = 0.7$ V. The effect of this on the decay time constant of follower-integrator output, V_{diff} is shown in **D**. **E**: A widening effect on the pulselets can also be achieved by increasing ΔV. I_{b_1} is set to 100 nA while ΔV is swept from 0.4 V to 0.8 V in steps of 80 mV steps. Common paramters are as follows: $C = 100$ fF, $V_p = 0.9$ V, pulse width of 1 μs and period 4 μs.

2) Translinear Multiplier: This 4T circuit translates the log-domain voltage addition (from a translinear loop) into a normalized current multiplication of input currents I_A and I_B. i.e, $I_{out} \approx \frac{I_A \cdot I_B}{I_{norm}}$. The reader may refer to [18] for details on the circuit operation. Fig.4 shows the circuit schematic and DC response. The required normalization current, I_{norm}, offers an extra free parameter for realizing beyond the classical two-factor learning.

3) Pulse Edge Detector: We adopted a follower-differentiator (diff1) circuit [4], [13] for detecting rising and falling edge of the postsynaptic events. This was necessary for performing I_w updates and resets on separate phases – update on first and reset on second phase respectively. The idea here is to leverage the bipolar exponential potential decays arising from switching. By thresholding at ΔV and $\overline{\Delta V} = V_{dd} - \Delta V$ with the help of two comparators, the rising and falling edge-induced decays produce two edge-coupled daughter pulses $\overline{\phi^u_{post}}$ and $\overline{\phi^r_{post}}$ from the original postsynaptic pulse $\overline{\phi_{post}}$ (See Fig.5).

III. The NN STDP Synapse Cell (Top Level)

Bringing it all together, the to-be-latched weight update current, $I_{\Delta w}(i+1)$ is computed with the inclusion of a reference current I_{ref} to circumvent the unidirectional current constraint. Thus yielding;

$$I_{\Delta w}(i+1) = I_{\Delta w_c}(t) + I_{ref} - I_{\Delta w_a}(t) \quad (6)$$

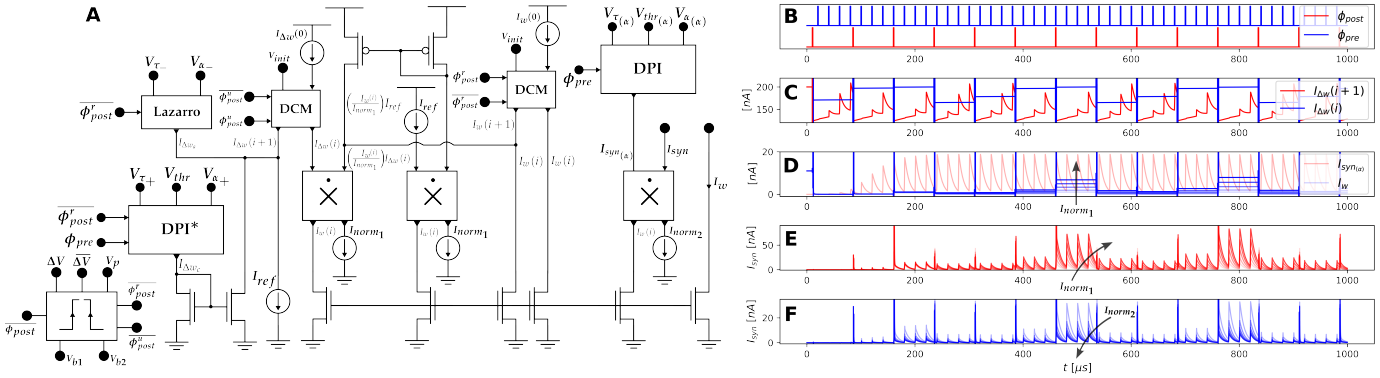

Fig. 6. **A**. Top-level NN STDP cell circuit. Simulation results: **B** – pre- and post- synaptic activity of periods of 20 μs and 75 μs respectively, pulse width of 1 μs, $V_{\alpha\_} = 1.332$ V, $V_{\tau\_} = 1.7$ V, $V_{\alpha_+} = 350$ mV, $V_{thr} = 1.415$ V, $V_{\tau_+} = 1.57$ V, $V_{\alpha_{(\alpha)}} = 320$ mV, $V_{\tau_{(\alpha)}} = 1.57$ V, $V_{thr_{(\alpha)}} = 1.42$ V, $V_p = 900$ mV, $\Delta V = 700$ mV, $I_{b_1} = 100$ nA, $I_{b_2} = 100$ nA, $I_{\Delta w}(0) = 1$ nA, $I_w(0) = 10$ nA. All but the DCM capacitor are 100 fF, DCM capacitor is 400 fF. The effect of I_{norm_1} is shown in **E** and **F**. The product of I_w and $I_{syn_{(\alpha)}}$ is shown in **E**. Similarly, I_{norm_2} effect is shown in **F**. In both, 2 nA to 10 nA sweep at 2 nA step.

Addition is conveniently achieved by the Kirchoff Current Law (KCL). The influence of the I_{ref} offset on $I_{\Delta w}$ (in eq. 6) is subtracted from the synaptic weight current, I_w as shown in eq. 7. I_w is non-negative during operation. This is neurophysiologically consistent – Chemical synapses are often designated either excitatory (AMPAergic and NMDAergic) or inhibitory (GABAergic) [11] – e.g., pyramidal cortical neurons and chandelier cells respectively.

$$I_w(i+1) = I_w(i) + \left(\frac{I_{\Delta w}(i)}{I_{norm_1}}\right) I_w(i) - \left(\frac{I_w(i)}{I_{norm_1}}\right) I_{ref} \quad (7)$$

Finally, another translinear multiplication can be performed between $I_w(i)$ and a preferred synaptic circuit (DPI used here) current, $I_{syn(\alpha)}$ (eq. 8). This offers another normalization current, I_{norm_2} that can be adopted in a meta-level optimization alongside the three-factor learning.

$$I_{syn}(t) = \frac{I_w(i) \cdot I_{syn(\alpha)}(t)}{I_{norm_2}} \quad (8)$$

A top-level schematic integrating all above-mentioned subcircuits is presented in Fig.6A. Simulation results showing potentiation and depression arising from pre- and post-synaptic activity are also shown in Fig.6B-F. Post-synaptic-event-driven sourcing of the memorized current $I_{\Delta w}(i)$ from the net STDP trace $I_{\Delta w}(i+1)$ can be observed in Fig.6C. The product of the $I_w(i)$ and a template synaptic current, $I_{syn(\alpha)}$ (in Fig.6 D) is shown in Fig.6 E. The effects of I_{norm_1} and I_{norm_2} ,potentially useful for three-factor learning, are shown in Fig.6E and F respectively.

Estimates for average power and energy per spike[1] for the parameter case shown in Fig.6 are 1.77 μW and 27.9 pJ/spike. For context, similar work by [19] (in TSMC 250 nm, $V_{dd} = 3.3$ V) achieved an estimated performance in the range of $42 - 83$ pJ/spike. A much efficient implementation [20] (in 90 nm, $V_{dd} = 0.6$ V) reported ~ 0.4 pJ/spike. It is worth pointing out that energy dissipation in this study, is a function of event rate and the STDP parameters. For instance, keeping all other parameter values used in Fig.6 and setting pre period to 3 μs, $V_{\tau_+} = 1.5$ V, and $V_{\alpha_+} = 320$ mV yields 1.72 μW and 5 pJ/spike. Thus, finding the optimal parameters for energy-efficient operation is subject to future work.

[1]In this work, energy/spike = average power/(#pre-spikes/s + #post-spikes/s)

IV. Discussion

SNNs are rivialed by conventional recurrent neural networks and the more recent transformer networks that power the much talked-about GPT-3 [1]. However, SNNs are well suited for computational efficiency that arises from their biologically plausible sparsity. This study presented a current-mode implementation of a postsynaptic-event-driven NN STDP. Simulation results based on the NCSU TSMC 180 nm technology in Cadence were presented. We repurposed analog synaptic-current-producing circuits for generating exponential STDP traces as well as translinear multiplication which would have otherwise been extravagant to achieve in digital logic. More so, additional parameters arising from the required normalization currents in the translinear multipliers offers the opportunity to explore three-factor and meta learning rules, which are of immense interest in recent times for continual learning [21], [22]. It is worth mentioning that the simulation results may suffer non-idealities such as transistor mismatch and thermal-induced excursions post-fabrication. Such perturbations can, however, be leveraged to arrive at a robust learning mechanism. Future work include circuit optimization to reduce transistor count as well as a subsequent chip-tapeout of an array of such synaptic units for the purpose of integration into a silicon neural array. This will ultimately be deployed in robust event-based recognition tasks.

V. Acknowledgements

We would like to acknowledge discussions at the Telluride Neuromorphic Cognition Workshop 2022. We would also like to acknowledge Dr. Philippe Pouliquen for the insights from his class. This work was supported by NSF through Grant #2223725.

REFERENCES

[1] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[2] D. Amodei and D. Hernandez, "Ai and compute," Nov 2018. [Online]. Available: https://openai.com/blog/ai-and-compute/

[3] N. C. Thompson, K. Greenewald, K. Lee, and G. F. Manso, "The computational limits of deep learning," *arXiv preprint arXiv:2007.05558*, 2020.

[4] C. Mead and M. Ismail, *Analog VLSI implementation of neural systems*. Springer Science & Business Media, 1989, vol. 80.

[5] C. Mead, "How we created neuromorphic engineering," *Nature Electronics*, vol. 3, no. 7, pp. 434–435, 2020.

[6] E. M. Izhikevich and N. S. Desai, "Relating stdp to bcm," *Neural computation*, vol. 15, no. 7, pp. 1511–1523, 2003.

[7] R. Dingledine, K. Borges, D. Bowie, and S. F. Traynelis, "The glutamate receptor ion channels," *Pharmacological reviews*, vol. 51, no. 1, pp. 7–62, 1999.

[8] G.-q. Bi and M.-m. Poo, "Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type," *Journal of neuroscience*, vol. 18, no. 24, pp. 10 464–10 472, 1998.

[9] C. Bartolozzi and G. Indiveri, "Synaptic dynamics in analog vlsi," *Neural computation*, vol. 19, no. 10, pp. 2581–2603, 2007.

[10] A. Akwaboah and R. Etienne-Cummings, "Lodenns: A linearly-approximated and optimized dendrocentric nearest neighbor stdp," in *Proceedings of the International Conference on Neuromorphic Systems 2022*, 2022, pp. 1–8.

[11] P. Sterling and S. Laughlin, *Principles of neural design*. MIT press, 2015.

[12] J. Lazzaro and J. Wawrzynek, *Low-power silicon neurons, axons and synapses*. Springer, 1994.

[13] S.-C. Liu, J. Kramer, G. Indiveri, T. Delbrück, and R. Douglas, *Analog VLSI: circuits and principles*. MIT press, 2002.

[14] J. V. Arthur and K. Boahen, "Recurrently connected silicon neurons with active dendrites for one-shot learning," in *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*, vol. 3. IEEE, 2004, pp. 1699–1704.

[15] J. V. Arthur and K. A. Boahen, "Synchrony in silicon: The gamma rhythm," *IEEE Transactions on Neural Networks*, vol. 18, no. 6, pp. 1815–1825, 2007.

[16] Z. Shi and T. Horiuchi, "A summating, exponentially-decaying cmos synapse for spiking neural systems," *Advances in neural information processing systems*, vol. 16, 2003.

[17] R. J. Baker, *CMOS: circuit design, layout, and simulation*. John Wiley & Sons, 2019.

[18] A. G. Andreou and K. A. Boahen, "Translinear circuits in subthreshold mos," *Analog Integrated Circuits and Signal Processing*, vol. 9, pp. 141–166, 1996.

[19] H. Tanaka, T. Morie, and K. Aihara, "A cmos spiking neural network circuit with symmetric/asymmetric stdp function," *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 92, no. 7, pp. 1690–1698, 2009.

[20] J. M. Cruz-Albrecht, M. W. Yung, and N. Srinivasa, "Energy-efficient neuron, synapse and stdp integrated circuits," *IEEE transactions on biomedical circuits and systems*, vol. 6, no. 3, pp. 246–256, 2012.

[21] F. Zenke, B. Poole, and S. Ganguli, "Continual learning through synaptic intelligence," in *International conference on machine learning*. PMLR, 2017, pp. 3987–3995.

[22] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, "Continual lifelong learning with neural networks: A review," *Neural networks*, vol. 113, pp. 54–71, 2019.