1-1-2017

A CMOS Synapse Design Implementing Tunable Asymmetric Spike Timing-Dependent Plasticity

Robert C. Ivans
Boise State University

Kurtis D. Cantley
Boise State University

Justin L. Shumaker
US Army Research Laboratory

A CMOS Synapse Design Implementing Tunable Asymmetric Spike Timing-Dependent Plasticity

Robert C. Ivans[a)] and Kurtis D. Cantley
Department of Electrical and Computer Engineering
Boise State University
Boise, ID, USA
Email: [a)]robertivans@boisestate.edu

Justin L. Shumaker
Vehicle Technology Directorate
US Army Research Laboratory
Aberdeen Proving Ground, MD, USA

***Abstract*—A CMOS synapse design is presented which can perform tunable asymmetric spike timing-dependent learning in asynchronous spiking neural networks. The overall design consists of three primary subcircuit blocks, and the operation of each is described. Pair-based Spike Timing-Dependent Plasticity (STDP) of the entire synapse is then demonstrated through simulation using the Cadence Virtuoso platform. Tuning of the STDP curve learning window and rate of synaptic weight change is possible using various control parameters. With appropriate settings, it is shown the resulting learning rule closely matches that observed in biological systems.**

Keywords—Neuromorphic design, CMOS synapse, spike timing-dependent plasticity (STDP)

I. Introduction

The adult human neocortex is composed of trillions of synapses interconnecting billions of neurons in extremely complex structures [1]–[3]. A synapse serves to modulate the connection strength between any two neurons in the system. This is achieved by altering a pre-synaptic action potential's influence in exciting a post-synaptic neuron in proportion to a parameter called synaptic weight. Having a large weight means having a stronger connection, whereas having a small weight means that little or no propagation of a pre-synaptic signal to a post-synaptic neuron will occur. How a synaptic weight changes over time is known as the learning rule, and is some function of the activity of the associated pre- and post-synaptic neurons. In some cases, activity can refer to firing rates, but it is also known to relate to timing of individual spikes in a mechanism called Spike Timing-Dependent Plasticity (STDP) [4]–[6]. STDP can be thought of as a rule which determines synaptic weight updates as a function of timing between pre- and post-synaptic spikes. If a pre-synaptic spike is followed closely by a post-synaptic spike, the synaptic weight is increased (potentiation). In the opposite case, the weight is decreased (synaptic depression). STDP is known to be responsible for certain abilities observed across many animal species, including rapid response to threat stimuli and sound source localization [7]–[10]. It also results in the ability of networks to learn to recognize spatio- or spectro-temporal patterns [11]–[13].

For the purposes of building artificial, bio-mimetic neural networks, a simple, tunable, and repeatable synaptic implementation is needed. One such solution consists of a single device such as a memristor, the major advantage of which is an extremely high achievable synaptic density [14], [15]. However, there are many types of memristors, each requiring different fabrication methods and possessing different behaviors. There is also a lack of consensus on the ideal properties of a memristive synapse for use in a neuromorphic system. On the other hand, CMOS technologies are well-developed, ubiquitous, and continue to scale to nanometer dimensions. Extreme interconnectivity of these networks can be accomplished through careful system design. Separate cores with 2-D synaptic arrays can send and receive data through high-speed pipelines using protocols such as Address-Event Representation (AER) [16]–[18].

The idea of designing a synapse in CMOS technology is not novel [17], [19]–[21]. However, this paper presents a novel CMOS synapse design which implements tunable asymmetric STDP and is compatible with digitally spiking integrate-and-fire (I&F) neurons. This design is unique in that it achieves a more biologically realistic STDP response than [17] while using fewer components than [21]. This is accomplished by using voltage dividers, instead of amplifiers, to create the signals responsible for changing synaptic weight.

Although not yet optimized for power consumption, the design can be directly deployed into various VLSI implementations such as those based on neurosynaptic core architectures. Section II of this paper discusses general synapse operation, with detailed description of each subcircuit block. Section III demonstrates simulation of the CMOS synapse learning rules, including settings for bio-mimetic STDP. Final conclusions are presented in Section IV.

II. Circuit and Subcircuits Operation

The results in this work were generated using the Cadence Virtuoso (6.1.7-64b) design suite and the NCSU Cadence Design Kit (CDK 1.6.0.beta). This design kit included the MOSIS models for CMOS devices which are extremely accurate over a wide range of operating conditions. The overall synapse design currently utilizes a total of 41 transistors and three capacitors. Associated layouts have been created and

This work was supported in part by the Air Force Office of Scientific Research under award number FA9550-14-1-0188, and also by an appointment to the U.S. Army Research Lab at the Aberdeen Proving Ground administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Army Research Lab.

submitted for fabrication and future testing. In the ON Semiconductor C5 process, the circuit occupies an area of approximately 200 × 300 μm^2, which is comparable to other approaches [17], [22]. Future work includes fully investigating scalability of the design and its power consumption. Currently, energy consumption per spike ranges from approximately 23 pJ to 1.5 μJ for spike pairs with pulse widths of 1 ms. Pulses generated by all neurons are presumed asynchronous and digital, meaning they may occur at any time and alternate between values of 0 V (inactive) and 5 V (during an action potential). All pulses in the system are of a set duration.

There are three total connections between the synapse and the two neurons it connects: two inputs are for spikes received from the output of both the pre- and post-synaptic neurons, and the synapse output is connected to the input of the post-synaptic neuron. A diagram containing the three different subcircuit blocks of the synapse is shown in Fig. 1. The synapse requires four control voltages to set the STDP characteristics: V_{pre_leak}, V_{post_leak}, V_{inc_th}, and V_{red_th}. Although not demonstrated in this paper, a biasing circuit can be used to create them from V_{dd}.

A. Race Condition Discriminator Circuit

Within the synapse, the race condition discriminator circuit (RCD) handles the situation in which pre- and post-synaptic spikes overlap. The RCD output (V_{rcd} in Fig. 1) and its inverse control a PMOS device in each of the two Gauntlet circuits (M4 in Fig. 3). Providing these two particular PMOS devices with opposing signals prevents overlapping spikes from influencing the synaptic core at the same time.

In order to produce V_{rcd}, the RCD uses cross-connected outputs to suppress propagation of competing input signals, as shown in Fig. 2a. Initially, nodes V_{rcd} and V_{rcd}' are both at 0 V, placing M1 and M3 in saturation and M2 and M4 in cutoff. If a pre-synaptic pulse arrives at the V_{pre} input before a post-synaptic pulse arrives at the V_{post} input, then the voltages at A and B lower, causing node V_{rcd} to rise to 5 V, which in turn causes M3 to cutoff and M4 to saturate, forcing V_{rcd}' to 0 V and preventing secondary signal propagation from C to D. A similar series of events occurs if a post-synaptic pulse arrives at the V_{post} input before a pre-synaptic pulse arrives at the V_{pre} input which forces V_{rcd} to 0 V, preventing signal propagation from Node A to Node B. Effectively, the RCD serves to pass signals from V_{pre} to V_{rcd} unless a signal from V_{post} precedes and overlaps it (Fig 2b).

B. Gauntlet Circuit

Fig. 3a shows the schematic of the Gauntlet Circuit. The Gauntlet Circuit's purpose is to facilitate STDP in the synapse by providing a tunable window within which pre- and post-synaptic spikes can influence synaptic weight. The diode-connected PMOS, M1, allows 5 V digital pulses, applied to V_2, to quickly charge capacitor C_1 without also quickly discharging via the input after the pulse ends. A tunable discharge path for C_1 is provided by M2, with the discharge rate controlled by V_{leak}. The resulting exponentially decaying analog signal V_{delay}, whose time constant is determined by the value of V_{leak}, is applied to the gate of M3 (see top trace of Fig. 3b). M3 uses V_{delay} to alter the magnitude of digital pulses applied to V_1 before they reach the Synaptic Core. M4 uses the V_{not_pass} signal from the RCD to ensure that only one V_{change} signal reaches the Synaptic core at a time. M5, M6, M7, and M8 provide a low resistance path to ground, in the absence of a pulse at V_1, to discharge trapped charge on either side of M8.

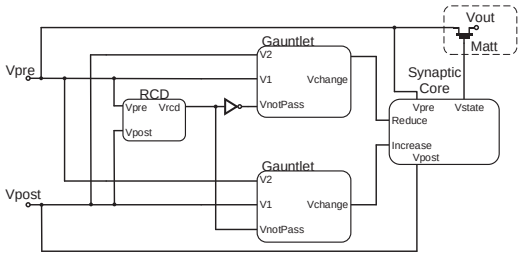

Fig. 1. A block diagram showing the connections between the subcircuits within the CMOS synapse. V_{post} is feedback from the output of the post-synaptic neuron, whereas V_{pre} is connected to the output of the pre-synaptic neuron. V_{out} is the modulated version of V_{pre} which is fed to the input node of the post-synaptic neuron circuit.

C. Synaptic Core

Fig. 4 depicts a schematic of the Synaptic Core circuit. The Synaptic Core produces V_{state}, which is roughly analogous to the synaptic weight. V_{state} is produced by the movement of charge on to, or off of, the state storage capacitor C_{state}. This is accomplished via M5 and M8, respectively. When one of these devices is turned on, charge must also flow through the two optional MOSFETs M6 and M7, whose sole purpose is to help to reduce leakage current from C_{state} through M5 and M8. The amount of directed charge is controlled by two active element voltage dividers that enable fine tuning of the STDP characteristics of the CMOS synapse. One voltage divider, formed by M1 and M2 in Fig. 4, allows for control over the amount of charge directed into C_{state} for a given signal applied to $V_{increase}$. This is done by limiting the drain current via V_{inc_th}, so that increasing V_{inc_th} reduces the amount of directed charge for a given signal applied to $V_{increase}$. The other voltage divider (M9 and M10 in Fig. 4) allows for control over the amount of charge directed out of C_{state} for a given signal applied to V_{reduce}. This is accomplished by limiting the drain current via V_{red_th}. The result is that decreasing V_{red_th} reduces the amount of directed charge for a given signal applied to V_{reduce}.

For initial testing, the synapse was designed such that its conductance was controlled by applying V_{state} to the gate of a MOSFET (M_{att} in Fig. 1). The issue with this is that values of C_{state} above M_{att}'s threshold voltage do not cause a proportional

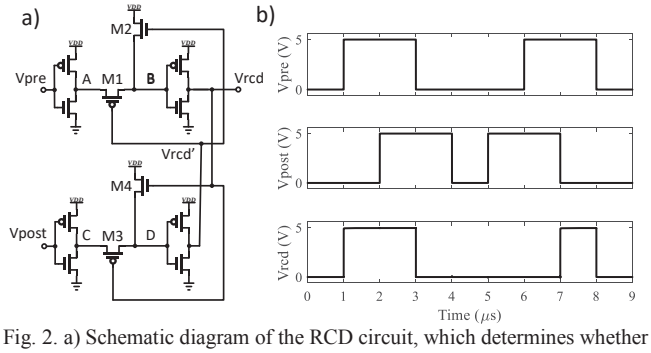

Fig. 2. a) Schematic diagram of the RCD circuit, which determines whether increase or decrease signals should be admitted to the Synaptic Core. All PMOS and NMOS are sized W/L = 30/4 and 10/4 respectively. b) The simulated response of the RCD circuit. When V_{pre} and V_{post} overlap, it is observed that V_{rcd} is V_{pre} unless V_{post} arrives first and blocks V_{pre}.

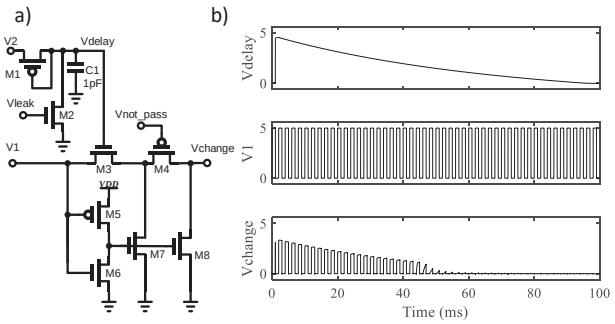

Fig. 3. a) The Gauntlet circuit schematic. The Gauntlet circuit helps to facilitate STDP by shaping V_{pulse} into V_{change} through V_{delay}. M5, M6, M7, and M8 help to drain charge trapped on either side of M2. All PMOS and NMOS are sized W/L = 30/4 and 10/4, respectively. b) Gauntlet circuit response to stimulus. A single 5 V pulse 1 ms wide digital pulse is applied to V_2 at 1 ms. V_{leak} is set to 433 mV. V_1 is supplied by a 5 V square wave with a period of 2 ms; this is atypical and solely for illustrative purposes. V_{not_pass} has been tied to ground to ensure that the difference between V_{change} and V_1 is due exclusively to V_{delay}. Notice that the magnitude of V_{change} decreases as V_{delay} decays. This decrease in magnitude helps to create STDP in the synapse.

change in signal attenuation because the MOSFET will operate in saturation. In future work, M_{att} will be replaced with a voltage controlled current source with a gain controlled by V_{state}.

III. Learning Rule Demonstration

A. Varying Circuit Parameters

Pair-based STDP curves were created to demonstrate the effects of varying circuit parameters on the synapse. Each STDP data point was collected from a 110 ms transient simulation which contained only one pre- and one post-synaptic spike. For each simulation the synaptic weight was initially set to one half of V_{dd} (V_{state}=2.5 V). The timing difference between the rising edges of pre- and post-synaptic spikes ($\Delta t=t_{post}-t_{pre}$) was recorded as the x-coordinate. Then, since V_{state} only changes due to pairs of spikes, and only changes on the second spike in the pair, the change in V_{state}, between just before and just after the second spike, was recorded as the y-coordinate. Finally, the resulting x- and y-coordinate pair was plotted.

Fig. 5a depicts the effects of varying V_{pre_leak} and V_{post_leak}, which control the decay times of the two gauntlet circuits (see Fig. 3). The left and right sides of the figure (for negative and positive Δt, respectively) can be independently controlled by the

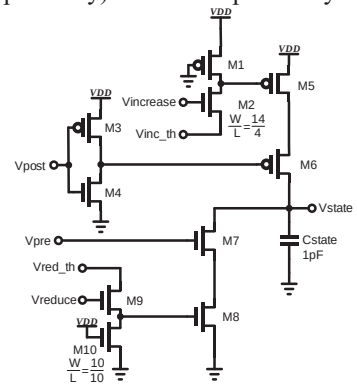

Fig. 4. The Synaptic Core schematic. V_{red_th} and V_{inc_th} control the magnitude by which the charge in the capacitor can change to allow fine control of the STDP curve. All PMOS and NMOS are sized W/L = 30/4 and 10/4 respectively, except where otherwise indicated.

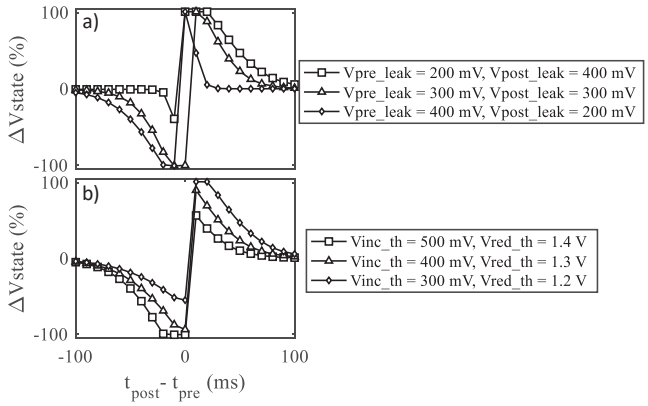

Fig. 5. a) The effects of varying V_{pre_leak} and V_{post_leak} on the STDP behavior of the synapse. When pre- and post-synaptic pulses are applied to the CMOS synapse, it is observed that the amount of change that occurs in V_{state} (ΔV_{state}) is related to the difference in time between the spikes ($\Delta t=t_{post}-t_{pre}$), and the settings of V_{pre_leak} and V_{post_leak}. Notice that as V_{pre_leak} and V_{post_leak} are increased, the STDP curve narrows. This plot was made using V_{inc_th} = 300 mV and V_{red_th} = 1.4 V. Input pulse widths were 1 ms. b) The effects of varying V_{inc_th} and V_{red_th} on the STDP behavior of the synapse. When pre- and post-synaptic pulses are applied to the CMOS synapse, it is observed that the amount of change that occurs in V_{state} (ΔV_{state}) is related to the difference in time between the spikes ($t_{post}-t_{pre}$), and the settings of V_{inc_th} and V_{red_th}. Notice that, as V_{inc_th} is increased and V_{red_th} is reduced, the magnitude of change is reduced. This plot was made using V_{pre_leak} = 200 mV and V_{post_leak} = 200 mV. Input pulse widths were 1 ms.

two voltages. Increasing V_{pre_leak} or V_{post_leak} will shorten the corresponding learning window for positive and negative Δt. When the two values are equal, the STDP curve will essentially be symmetrical for both positive and negative Δt, exemplified by the curve marked by triangle symbols in Fig. 5.

The effects on the STDP curve of varying V_{inc_th} and V_{red_th} are depicted in Fig. 5b. These two values control the maximum change in the weight for a pre-post or a post-pre pair (the ΔV_{state} values nearest to Δt=0). For increased values of V_{inc_th} (and decreased values of V_{red_th}), the weight will change more drastically for presentation of a single pair, but only to a maximum of ±100%, at which point the weight saturates. When saturation occurs, it does not change the difficulty for the next (oppositely alternating) pair to change the state back to some intermediate value. In other words, there is no "memory" or other driving force pushing the state toward one extreme or the other. However, in the absence of spiking, subthreshold conduction through M5, M6, M7, and M8 in Fig. 4, will cause V_{state} to trend toward some value near $V_{dd}/2$ over a period of approximately ten seconds. Some form of long-term motion of V_{state} is common with all synaptic circuits that use MOSFETs to control the charge on a capacitor. In this case, if spike pairs are presented with regularity (at least a few times per second), the STDP learning will overcome the very slow state change.

B. Fitting Biological Data

By choosing appropriate V_{pre_leak}, V_{post_leak}, V_{inc_th}, and V_{red_th} values, the STDP curve of the synapse can be tuned to fit a wide range of models with biphasic decaying exponential form. Fig. 6 demonstrates the CMOS synapse tuned to approximate STDP

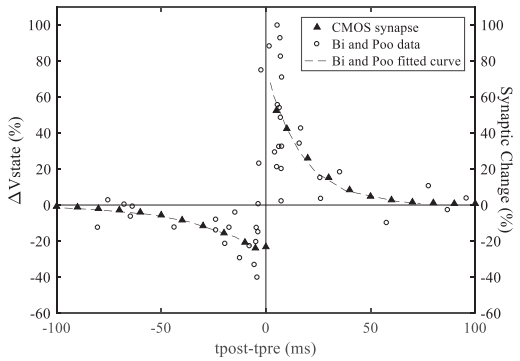

Fig. 6. Adjusting V_{pre_leak}, V_{post_leak}, V_{inc_th}, and V_{red_th} allows the STDP curve of the CMOS Synapse to be adjusted such that it can be fitted to biological data. In this figure, the CMOS synapse has been adjusted such that its STDP curve aligns with biological synapse data collected by Bi and Poo [5]. The settings used to create this plot are: V_{pre_leak} = 270 mV, V_{post_leak} = 300 mV, V_{inc_th} = 540 mV, and V_{red_th} = 1.08 V. Input pulse widths were 1 ms.

data measured from a biological synapse [5].

C. Power Consumption

The power consumed by the synapse is dependent upon the initial state of the synapse, the magnitude of the weight change, and whether the weight is increasing or decreasing. Fig. 7 depicts the energy consumed by the synapse as a function of the temporal difference between pre- and post-synaptic spikes. Each point represents the result of a simulation of a single pair of pre- and post-synaptic spikes with V_{state} initialized to 2.5 V, V_{pre_leak} = 270 mV, V_{post_leak} = 300 mV, V_{inc_th} = 540 mV, and V_{red_th} = 1.08 V. Input pulse widths were 1 ms. With these settings and an initial V_{state} of 2.5 V the energies used to decrease and increase synaptic weight are about 23 nJ and 1 µJ respectively.

IV. Conclusions

A novel CMOS synapse implementation with a tunable pair-based STDP learning rule has been demonstrated through simulation. The synapse circuit is compatible with many VLSI neuromorphic designs that incorporate spiking neurons. Three primary subcircuit blocks (Gauntlet, RCD, and Synaptic Core) with very specific functionality are used to realize STDP. In addition, four control voltages provide the ability to tune the STDP curve learning window and learning rate over a large range of operation. One specific example is fitting to biologically measured STDP data. Future directions for this research will be to optimize the circuit layout, to explore circuit simplification, and to investigate pattern recognition and classification tasks with large networks of spiking neurons.

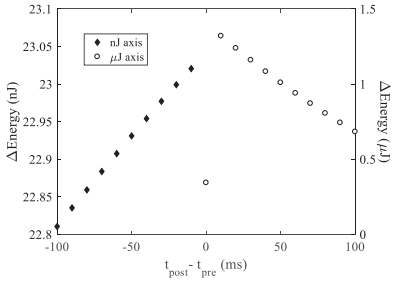

Fig. 7. Energy consumption by the synapse as a function of the temporal difference between pre- and post-synaptic spikes. The settings used to create this plot are: V_{pre_leak} = 270 mV, V_{post_leak} = 300 mV, V_{inc_th} = 540 mV, and V_{red_th} = 1.08 V. Pulse widths were 1 ms. V_{state} = 2.5 V.

References

[1] J. Hawkins and S. Blakeslee, *On Intelligence*. New York: Henry Holt and Company, 2004.

[2] B. Pakkenberg and H. J. G. Gundersen, "Neocortical neuron number in humans: Effect of sex and age," *J. Comp. Neurol.*, vol. 384, no. 2, pp. 312–320, 1997.

[3] Y. Tang, J. R. Nyengaard, D. M. G. DeGroot, and H. J. G. Gundersen, "Total regional and global number of synapses in the human brain neocortex," *Synapse*, vol. 41, no. 3, pp. 258–273, 2001.

[4] H. Markram, J. Lubke, M. Frotscher, and B. Sakmann, "Regulation of Synaptic Efficacy by Coincidence of Postsynaptic APs and EPSPs," *Science (80-.).*, vol. 275, no. 5297, pp. 213–215, 1997.

[5] G.-Q. Bi and M.-M. Poo, "Synaptic modification by correlated activity: Hebb's postulate revisited," *Annu. Rev. Neurosci.*, vol. 24, pp. 139–166, 2001.

[6] S. Song, K. D. Miller, and L. F. Abbott, "Competitive Hebbian learning through spike-timing-dependent synaptic plasticity," *Nat. Neurosci.*, vol. 3, no. 9, pp. 919–926, 2000.

[7] W. Bialek, F. Rieke, R. de Ruyter van Steveninck, and D. Warland, "Reading a neural code," *Science (80-.).*, vol. 252, no. 5014, pp. 1854–1857, Jun. 1991.

[8] C. E. Carr and M. Konishi, "A circuit for detection of interaural time differences in the brain stem of the barn owl," *J. Neurosci.*, vol. 10, no. 10, pp. 3227–3246, 1990.

[9] B. Glackin, J. A. Wall, T. M. McGinnity, L. P. Maguire, and L. J. McDaid, "A spiking neural network model of the medial superior olive using spike timing dependent plasticity for sound localization," *Front. Comput. Neurosci.*, vol. 4, no. 18, 2010.

[10] J. A. Wall, L. J. McDaid, L. P. Maguire, and T. M. McGinnity, "Spiking Neural Network Model of Sound Localization Using the Interaural Intensity Difference," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 23, no. 4, pp. 574–586, 2012.

[11] R. Guyonneau, R. VanRullen, and S. J. Thorpe, "Neurons Tune to the Earliest Spikes Through STDP," *Neural Comput.*, vol. 17, no. 4, pp. 859–879, 2005.

[12] T. Masquelier, R. Guyonneau, and S. J. Thorpe, "Spike Timing Dependent Plasticity Finds the Start of Repeating Patterns in Continuous Spike Trains," *PLoS One*, vol. 3, no. 1, 2008.

[13] R. C. Ivans, K. D. Cantley, and E. M. Vogel, "Spatio-Temporal Pattern Recognition in Circuits with Memory-Transistor-Driven Memristive Synapses," in *International Joint Conference on Nerual Networks*, 2017.

[14] L. Chua, "Memristor-The missing circuit element," *IEEE Trans. Circuits Theory*, vol. 18, no. 5, pp. 507–519, 1971.

[15] L. Chua and S. M. Kang, "Memristive devices and systems," *Proc. IEEE*, vol. 64, no. 2, pp. 209–223, 1976.

[16] K. A. Boahen, "Point-to-point connectivity between neuromorphic chips using address events," *IEEE Trans. Circuits Syst. II*, vol. 47, no. 5, pp. 416–434, 2000.

[17] 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. Neural Networks*, vol. 17, no. 1, pp. 211–221, 2006.

[18] P. A. Merolla *et al.*, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science (80-.).*, vol. 345, no. 6197, pp. 668–673, Aug. 2014.

[19] S. G. Pradyumna and S. S. Rathod, "Analysis of CMOS Synapse generating Excitatory Postsynaptic Potential using DC Control Voltages," 2015, no. Proceedings of 2015 Global Conference on Communication Technologies (GCCT 2015), pp. 433–436.

[20] V. Saxena, "US 2015/0278681 A1," 2015.

[21] J. M. Cruz-Albrecht, M. W. Yung, and N. Srinivasa, "Energy-efficient neuron, synapse and STDP integrated circuits," *IEEE Trans. Biomed. Circuits Syst.*, vol. 6, no. 3, pp. 246–256, 2012.

[22] S. Fusi, M. Annunziato, D. Badoni, A. Salamon, and D. J. Amit, "Spike-Driven Synaptic Plasticity: Theory, Simulation, VLSI Implementation," *Neural Comput.*, vol. 12, no. 10, pp. 2227–2258, 2000.