A Fully-Synthesized 20-Gate Digital Spike-Based Synapse with Embedded Online Learning

Charlotte Frenkel*, Giacomo Indiveri†, Jean-Didier Legat* and David Bol*
* ICTEAM Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium
† Institute of Neuroinformatics, University of Zurich and ETH Zurich, Zurich, Switzerland
Email: {charlotte.frenkel, jean-didier.legat, david.bol}@uclouvain.be, giacomo@ini.uzh.ch

***Abstract*—Neuromorphic engineering aims at building cognitive systems made of electronic neuron and synapse circuits. These emerging computing architectures have a high potential for real-world problems that are difficult to formalize and program, such as vision or sensorimotor control. In order to leverage the potential of neuromorphic engineering and study cognition principles in physical systems, the development of autonomous online learning is a key feature. However, to develop scalable systems that can be used in realistic applications, it is crucial to design compact and low-power hardware platforms. Here we analyze a spike-driven synaptic plasticity (SDSP) learning rule and show that it is particularly well suited for highly compact digital synapse implementations, especially if compared to conventional spike-timing-dependent plasticity (STDP) rules. Furthermore, we designed an asynchronous fully-synthesizable digital synapse circuit with embedded SDSP-based online learning features, and with programmability options for versatile computing. The proposed synapse implementation requires only 20 gates for a compact area of 25μm^2 in a 28nm FDSOI CMOS process.**

I. Introduction

As Moore's law is coming to an end [1], the power efficiency and scale of our modern digital computers still lags many orders of magnitude behind that of biological neural systems for tasks such as pattern recognition or real-time learning in uncontrolled environments [2]. Only a radical paradigm shift to massively parallel neuromorphic computing *in silico* offers sufficient potential to bridge this gap [3]. However, the current understanding of operating principles in the brain is not yet sufficient to determine the optimal learning methods and algorithms for general purpose applications of artificial neural network devices [4]. It is therefore important to develop efficient hardware implementations of silicon synapses with embedded online learning algorithms for exploring their properties and determining the best application areas and use cases. Several efforts have been made in the past toward this goal, driven by two main aspects. The first one is linked to the choice of the specific learning rule. While many circuits proposed to implement silicon synapses are based on the classical pair-based spike-timing-dependent plasticity (STDP) rule [5]–[11], few designs [12]–[14] explored the implementation of spike-driven synaptic plasticity (SDSP) rules based on both the spike timing and the internal variables of the synapses and neurons (e.g., see [15]), while other approaches followed biophysical descriptions based on ion channel dynamics [16], [17]. In all these approaches, the specific learning rule and resolution selected for the design determined the synapse circuit size, its task-specific learning performances and the memory lifetime of the network as a function of the number of new stimuli received (i.e. the palimpsest property) [18].

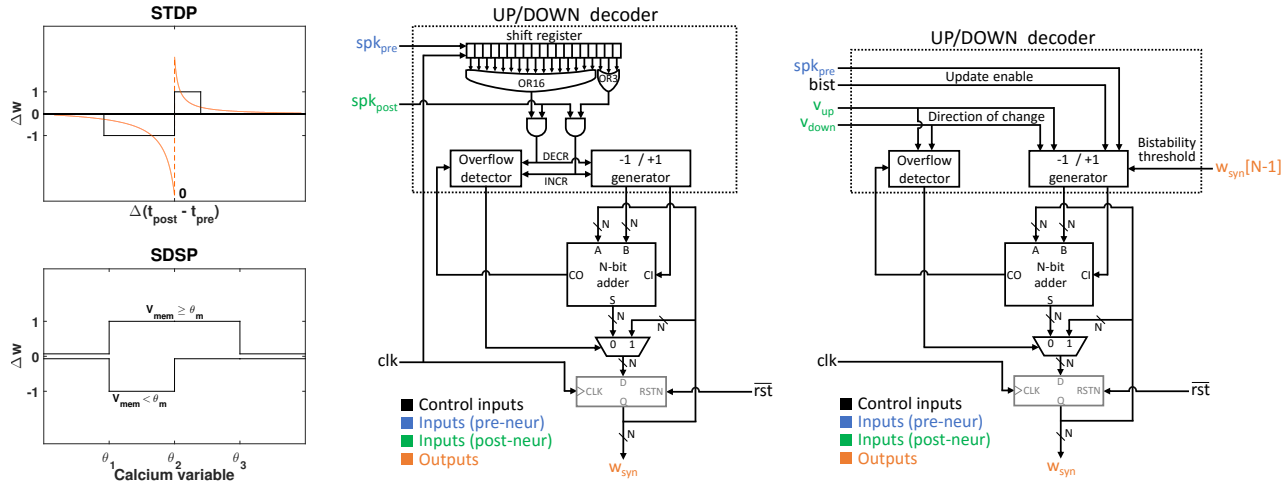
The second aspect that drives synapse design is based on the target synaptic fan-in. Biological neural networks typically have synaptic fan-ins from 100 to as high as 10000 synapses per neuron. However, in neuromorphic circuits with embedded online learning, high fan-ins often lead to designs whose synapse resources dominate neuron resources by more than one order of magnitude. For example, in the ROLLS analog neuromorphic chip [14] which implements a synaptic fan-in of 256, the area of SDSP-based synapses is 70 times higher than the area occupied by neurons. The target of high fan-ins for biophysical accuracy and high-dimensional information processing thus further stresses the importance of a compact synapse design based on an efficient learning rule implementation.

While digital design is free from bias generation and benefits from technology scaling with shorter design time and low sensitivity to noise and process-voltage-temperature (PVT) variations, nearly all silicon synapses with embedded online learning proposed to date are based on analog circuits. The latter allows to emulate the different learning rule dynamics using only a few transistors around a capacitor-based or an SRAM-based weight storage. The only previously-reported fully-digital silicon synapse with online learning is proposed by Seo *et al.* in [9] and is very compact with only 8.5μm^2 and 14.5μm^2 for 1-bit and 4-bit resolutions, respectively. However, this clocked implementation of a probabilistic STDP learning rule requires heavy time window computation and random number generation inside neurons together with a custom dual-port SRAM. As the pushed rules are restricted to foudry bitcells, the DRC rules for logic must be used for custom SRAM designs, resulting in a strong area overhead. A few STDP synapse designs are also reported for FPGAs but are extremely resource-consuming, at the exception of the highly-simplified STDP learning rule proposed by Cassidy *et al.* [19].

In this paper, we propose a digital synapse with embedded SDSP online learning requiring only 20 logic gates for an area of 25μm^2 in 28nm FDSOI CMOS. A comparative analysis of the area efficiency of STDP- and SDSP-based learning rules for digital synapse design is carried out in Section II, it highlights how computational aspects of the SDSP learning rule can be off-loaded from the synapse to the neuron for compact implementations. The proposed asynchronous and fully-synthesizable 3-bit SDSP-based synapse is presented in details and validated in post-layout simulation in Section III. Finally, the obtained results are discussed in Section IV before drawing conclusions in Section V.

II. Comparison of STDP and SDSP Learning Rules

The STDP and SDSP learning rules are illustrated in Fig. 1(a). The classical STDP rule depends on the relative



(a) Digital STDP and SDSP rules. (b) Digital STDP implementation based on [19]. (c) Digital SDSP implementation.

Fig. 1: Illustration of the STDP and SDSP learning rules together with their minimal N-bit implementations using clocked digital circuits.

timing of pre- and post-synaptic spikes: if these spikes occur in causal (anti-causal) order, the synapse is potentiated (depressed) following an exponential function shown as a dotted orange line in Fig. 1(a). A strong simplification of the STDP shape is proposed by Cassidy *et al.* in [19] for digital implementations and is shown in black. On the other hand, the SDSP rule proposed by Brader *et al.* in [15] depends on the more general post-synaptic neuron state at the time of the pre-synaptic spike, not on the sole pre- and post-synaptic spike timings. If the post-synaptic membrane voltage V_{mem} is above (below) a given threshold θ_m, the synaptic weight undergoes a step increase (decrease) upon the arrival of a pre-synaptic spike. In order to improve the recognition of highly-correlated patterns, Brader *et al.* add a stop-learning rule based on the Calcium concentration Ca of the post-synaptic neuron [15]. The Calcium concentration provides an image of the post-synaptic spiking activity: if it is beyond average ranges (thresholds θ_1, θ_2 and θ_3), there is evidence that learning already occurred and that further potentiation or depression would result in overfitting. The SDSP rule with stop-learning conditions is shown in Fig. 1(a) and can be implemented following Eq. (1).

$$\begin{cases} w \to w+1 & \text{if } V_{\text{mem}}(t_{\text{pre}}) \geq \theta_m,\ \theta_1 \leq \text{Ca}(t_{\text{pre}}) < \theta_3 \\ w \to w-1 & \text{if } V_{\text{mem}}(t_{\text{pre}}) < \theta_m,\ \theta_1 \leq \text{Ca}(t_{\text{pre}}) < \theta_2 \end{cases} \quad (1)$$

Finally, a bistability mechanism [15] is added to the SDSP description of Eq. (1): if the synaptic weight is above (below) a given threshold θ_w, it is attracted toward a high (low) state in the long term. This mechanism is important for two reasons [18]: first, it prevents undesired fluctuations resulting from spontaneous activity in the network. Second, it results in a natural stochastic selection mechanism for long-term potentiation (LTP) and long-term depression (LTD) as the spike activity in neural networks follows a Poisson distribution. The long-term learning performances of SDSP are similar to those of STDP but present better bio-physical accuracy and generalization properties [15].

The optimal synchronous digital implementation for STDP proposed by Cassidy *et al.* is shown in Fig. 1(b), we refer the reader to [19] for operation details. As the STDP learning rule cannot be implemented efficiently with asynchronous logic and would require long delay chains for time window generation around pre-synaptic ($\texttt{spk}_{\texttt{pre}}$) and post-synaptic ($\texttt{spk}_{\texttt{post}}$) spikes in the UP/DOWN decoder, we present a synchronous digital SDSP synapse circuit in Fig. 1(c) in order to carry a fair comparison. It reveals the key advantage of the SDSP learning rule: the main synaptic computations are off-loaded to the post-synaptic neuron, which computes the SDSP learning conditions $\texttt{v}_{\texttt{up}}$ and $\texttt{v}_{\texttt{down}}$ based on Eq. (1) for all synapses. The post-synaptic neuron must thus compute the Calcium concentration, which can be achieved using a simple counter. The associated resources are amortized for an entire dendritic tree (i.e. synaptic fan-in), avoiding replication in all synapses. A trigger signal `bist` is also provided as a time reference for the bistability mechanism, the associated bistability threshold is given by the most significant bit of the synaptic weight $\texttt{w}_{\texttt{syn}}$.

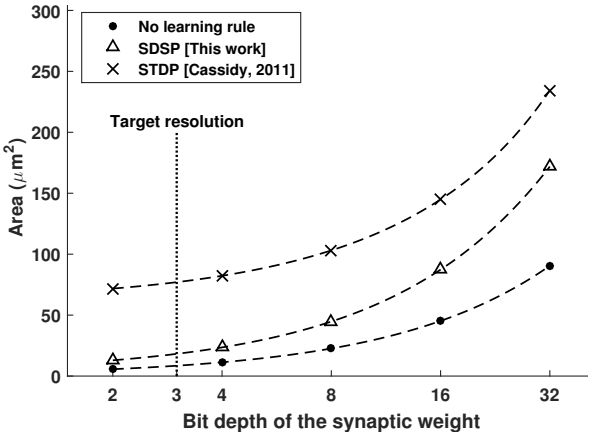


Fig. 2: Area comparison of synapses with STDP, SDSP and without learning.

The STDP and SDSP implementations are compared in Fig. 2 for different numbers of bits N in the synaptic weight, a synapse without learning rule made only of flip-flops for weight storage is provided as a reference. As expected, STDP suffers from a high fixed cost, which is especially critical for

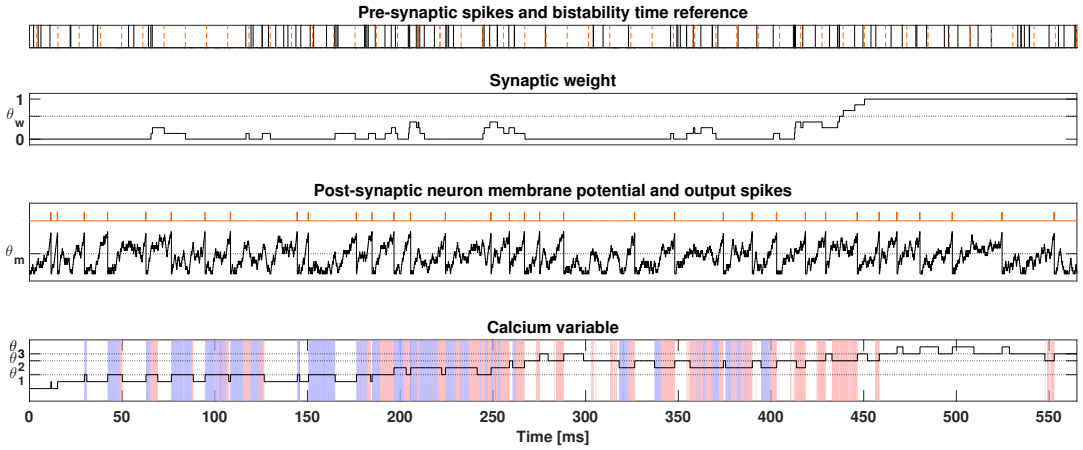


Fig. 4: Post-layout simulation of the proposed asynchronous SDSP-based synapse. The bistability `bist` reference trigger is shown as dotted orange lines next to the pre-synaptic spikes. The SDSP conditions for potentiation (v_{up}) and depression (v_{down}) from the post-synaptic neuron are shown next to the Calcium variable in red and blue, respectively. Non-colored parts of the Calcium concentration are associated with stop-learning conditions.

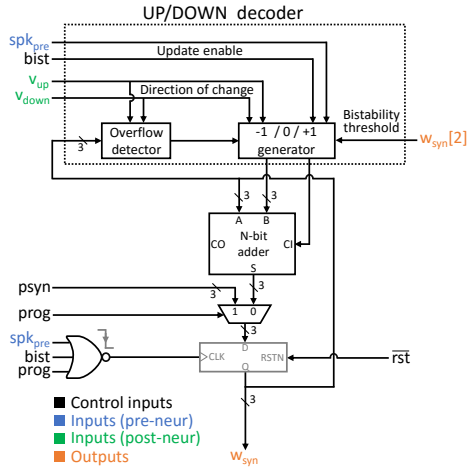


Fig. 3: Proposed SDSP-based 3-bit asynchronous programmable synapse.

low resolutions. As resolutions beyond 4 bits are of marginal interest for STDP [20] and as 3 bits are sufficient for SDSP to obtain the same dynamics as those presented in [15], the SDSP synapse offers strong potential for low-area implementations compared to STDP synapses.

III. Proposed Asynchronous SDSP Synapse

The synchronous SDSP-based synapse design of Fig. 1(c) was useful only for comparison purposes. As this model lends itself naturally to clockless event-driven operation, we present here an asynchronous implementation (see Fig. 3) using a 3-bit resolution to reproduce weight update dynamics similar to those of the original SDSP description [15]. The asynchronous mode of operation is more energy-efficient as clock distribution is eliminated. It also allows to spare synchronization resources around the asynchronous address-event representation (AER) bus [21], which is a popular communication infrastructure in spiking neural networks.

In the proposed design, we added programmability features: if the `prog` signal is asserted, the SDSP online learning is stopped through gating of pre-synaptic and bistability events. The `prog` signal also selects the externally-programmed 3-bit synaptic weight `psyn` at the input of the state element, storing it in the synaptic weight only after a negative edge. This programmability makes the proposed silicon synapse an ideal investigation platform as it allows to carry arbitrary experiments using an external workstation. Post-layout simulations of SDSP-based learning are shown in Fig. 4, where an LTP transition occurs after $t = 400$ms. The red and blue shades superimposed over the Calcium variable illustrate the SDSP signals that enable synaptic potentiation (v_{up}) and depression (v_{down}), as prescribed by Eq. (1), while bistability events induce updates independently. An 8-bit leaky integrate-and-fire (I&F) neuron extended with Calcium concentration is used as the post-synaptic neuron. This neuron extension required by SDSP occupies only $65\mu m^2$ in 28nm FDSOI CMOS, which is negligible as it is amortized over the number of synapses in the neuron dendritic tree. The proposed clockless implementation of a 3-bit SDSP-based synapse with programmability requires only 20 gates for an area of $25\mu m^2$ in 28nm FDSOI CMOS.

IV. Discussion

To carry out pattern recognition on images and real-world data with neuromorphic processors, silicon neurons are required to have a very large number of synapses. It is therefore crucial that the synapse circuits occupy the smallest possible area. To fulfill this requirement, it is important to choose the proper design style (analog/digital, synchronous/asynchronous, etc.), given a chosen technology node. While in conservative technologies such as 0.35μm or 0.18μm CMOS processes analog circuits would clearly outperform digital ones in terms of area requirements for implementing synaptic dynamics and learning features, in more advanced processes the distinction between analog and digital is not as clear cut. This is one of the first studies that demonstrates how it is possible to obtain synaptic dynamics and learning features analogous to those obtained with compact analog neuromophic circuits (e.g., as those presented in [14]) in a more advanced process, using significantly less area. As shown in Table I, other attempts have been made in the past (e.g., in [9]). However, as opposed to [9], which requires a custom dual-port SRAM design with both row and column accesses in order to carry STDP updates for pre- and post-synaptic events respectively, the state flip-flops

TABLE I: State of the art of silicon synapses with embedded online long-term learning rules, adapted and completed from [17].

	Technology	Plasticity	Asynchronous	Fully-synthesizable	Weight storage (precision)	Estimated area
Schemmel, 2006 [5]	0.18μm	STDP	✓	✗	SRAM (4 bits)	100μm^2
Arthur, 2006 [6]	0.25μm	STDP	✓	✗	SRAM (1 bit)	440μm^2
Tanaka, 2009 [7]	0.25μm	STDP	✓	✗	Capacitor (Analog)	5000μm^2
Cruz-Albrecht, 2012 [8]	90nm	STDP	✓	✗	Capacitor (Analog)	4823μm^2
Seo, 2011 [9]	45nm SOI	Probabilistic STDP	✗	✗†	SRAM (1 bit)	8.5μm^2 *
Seo, 2011 [9]	45nm SOI	Probabilistic STDP	✗	✗†	SRAM (4 bits)	14.5μm^2 *
Ramakrishnan, 2011 [10]	0.35μm	Weight-dependent STDP	✓	✗	Floating gate (Analog)	100μm^2
Bamford, 2012 [11]	0.35μm	Weight-dependent STDP	✗	✗	Capacitor (Analog)	400μm^2
Maldonado Huayaney, 2016 [16]	0.18μm	Calcium-based STDP	✓	✗	Capacitor (Bistable)	1638μm^2
Giulioni, 2008 [12]	0.35μm	SDSP	✓	✗	Capacitor (Bistable)	3140μm^2
Mitra, 2009 [13]	0.35μm	SDSP	✓	✗	Capacitor (Bistable)	3000μm^2
Qiao, 2015 [14]	0.18μm	SDSP	✓	✗	Capacitor (Bistable)	252μm^2
Proposed synapse	28nm FDSOI	SDSP	✓	✓	Flip-flops (3 bits)	25μm^2

* Estimated with the assumption that two thirds of the neuron area are occupied by time window generation and random number generation for the probabilistic STDP rule.
† Requires a custom SRAM design.

of the proposed SDSP-based synapse are updated only after pre-synaptic events. Therefore, the weights of the proposed synapse could be stored in a standard SRAM based on a 6-T foundry bitcell with densities as low as 0.120μm^2/bit in 28nm FDSOI CMOS [22]. For example, considering a fan-in of 256, the density of the proposed 3-bit 20-gate SDSP synapse in standard SRAM is expected to be reduced to only 0.67μm^2 as the remaining 17 gates (15μm^2) and the neuron Calcium extension (65μm^2) will be amortized over the synaptic fan-in.

V. Conclusion

We proposed a novel digital synapse circuit which exhibits three key characteristics that are desirable in neuromorphic systems: (i) asynchronous communication for low-power operation and efficient AER interfacing, (ii) SDSP online learning and programmability features, and (iii) compact design for low resolutions, requiring only 20 logic gates for an area of 25μm^2 in 28nm FDSOI CMOS. Furthermore, in the proposed synapse, the synaptic weights can be off-loaded to a standard SRAM, which opens the doors for highly compact digital neuromorphic circuits with embedded online learning.

Acknowledgment

The authors would like to thank Ning Qiao and Hesham Mostafa from the Institute of Neuroinformatics (INI) for fruitful discussions.

C. Frenkel is with Université catholique de Louvain as a Research Fellow from the National Foundation for Scientific Research (FNRS) of Belgium.

References

[1] M. Horowitz, "1.1 Computing's energy problem (and what we can do about it)," *Proc. of IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 10-14, 2014.

[2] F. Sandin et al., "Concept learning in neuromorphic vision systems: What can we learn from insects?," *Journal of Software Engineering and Applications*, vol. 7, no. 5, pp. 387-395, 2014.

[3] C. S. Poon and K. Zhou, "Neuromorphic silicon neurons and large-scale neural networks: challenges and opportunities," *Frontiers in neuroscience*, vol. 5, p. 108, 2011.

[4] G. Indiveri, E Chicca and R. J. Douglas, "Artificial cognitive systems: from VLSI networks of spiking neurons to neuromorphic cognition," *Cognitive Computation*, vol. 1, no. 2, pp. 119-127, 2009.

[5] J. Schemmel et al., "Implementing synaptic plasticity in a VLSI spiking neural network model," *Proc. of IEEE International Joint Conference on Neural Network (IJCNN)*, 2006.

[6] J. Arthur and K. Boahen, "Learning in silicon: Timing is everything," in *Advances in Neural Information Processing Systems (NIPS)*, Eds. Cambridge, MA: MIT Press, vol. 18, pp. 75-82, 2006.

[7] 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.

[8] 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.

[9] J.-S. Seo et al., "A 45nm CMOS neuromorphic chip with a scalable architecture for learning in networks of spiking neurons," *IEEE Custom Integrated Circuits Conference (CICC)*, 2011.

[10] S. Ramakrishnan, P. E. Hasler and C. Gordon, "Floating gate synapses with spike-time-dependent plasticity," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 5, no. 3, pp. 244-252, 2011.

[11] S. A. Bamford, A. F. Murray and D. J. Willshaw, "Spike-timing-dependent plasticity with weight dependence evoked from physical constraints," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 6, no. 4, pp. 385-398, 2012.

[12] M. Giulioni et al., "A VLSI network of spiking neurons with plastic fully configurable "stop-learning" synapses," *Proc. of IEEE Int. Conference on Electronics, Circuits and Systems (ICECS)*, pp. 678-681, 2008.

[13] S. Mitra, S. Fusi and G. Indiveri, "Real-time classification of complex patterns using spike-based learning in neuromorphic VLSI," *IEEE Trans. on Biomedical Circuits and Systems,* vol. 3, no. 1, pp. 32-42, 2009.

[14] N. Qiao et al., "A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128K synapses," *Frontiers in neuroscience*, vol. 9, no. 141, 2015.

[15] J. M. Brader, W. Senn and S. Fusi, "Learning real-world stimuli in a neural network with spike-driven synaptic dynamics," *Neural computation*, vol. 19, no. 11, pp. 2881-2912, 2007.

[16] F. L. Maldonado Huayaney and E. Chicca, "A VLSI Implementation of a calcium-based plasticity learning model," *Proc. of IEEE International Symposium on Circuits and Systems (ISCAS)*, 2016.

[17] M. R. Azghadi et al., "Spike-based synaptic plasticity in silicon: design, implementation, application, and challenges," *Proceedings of the IEEE*, vol. 102, no. 5, pp. 717-737, 2014.

[18] S. Fusi, "Hebbian spike-driven synaptic plasticity for learning patterns of mean firing rates," *Biological Cybernetics*, vol. 87, pp. 459-470, 2002.

[19] A. Cassidy, A. G. Andreou and J. Georgiou, "A combinational digital logic approach to STDP," *Proc. of IEEE International Symposium of Circuits and Systems (ISCAS)*, pp. 673-676, 2011.

[20] T. Pfeil et al., "Is a 4-bit synaptic weight resolution enough? – Constraints on enabling spike-timing dependent plasticity in neuromorphic hardware," *Frontiers in Neuroscience*, vol. 6, no. 90, 2012.

[21] K. A. Boahen, "Point-to-point connectivity between neuromorphic chips using address events," *IEEE Transactions on Circuits and Systems II*, vol. 47, no. 5, pp. 416-434, 2000.

[22] O. Thomas et al., "Dynamic single-p-well SRAM bitcell characterization with back-bias adjustment for optimized wide-voltage-range SRAM operation in 28nm UTBB FD-SOI," *Proc. of IEEE International Electron Devices Meeting (IEDM)*, 2014.