Spike-based learning in VLSI networks of integrate-and-fire neurons

Giacomo Indiveri and Stefano Fusi
Institute of Neuroinformatics
University-ETH Zurich
Winterthurerstrasse 190, CH-8057
E-mail: giacomo@ini.phys.ethz.ch

***Abstract*— As the number of VLSI implementations of spike-based neural networks is steadily increasing, and the development of spike-based multi-chip systems is becoming more popular it is important to design spike-based learning algorithms and circuits, compatible with existing solutions, that endow these systems with adaptation and classification capabilities. We propose a spike-based learning algorithm that is highly effective in classifying complex patterns in semi-supervised fashion, and present neuromorphic circuits that support its VLSI implementation. We describe the architecture of a spike-based learning neural network, the analog circuits that implement the synaptic learning mechanism, and present results from a prototype VLSI chip comprising a full network of integrate-and-fire neurons and plastic synapses. We demonstrate how the VLSI circuits proposed reproduce the learning model's properties and fulfill its basic requirements for classifying complex patterns of mean firing rates.**

I. Introduction

VLSI implementations of networks of spiking neurons offer promising solutions to problems involving real–time sensory processing and on–line classification of complex patterns. A wide range of devices comprising silicon neurons and synapses has been developed. These range from sensory devices such as silicon retinas and silicon cochleas [1]–[3], to VLSI implementations of conductance-based models of neurons [4], [5], reconfigurable arrays of integrate and fire neurons [6]–[8], and learning chips implementing detailed models of spike-based synaptic plasticity [7], [9]–[12].

In parallel with the development of pulse-based VLSI devices, there have also been significant advancements in the development of asynchronous event-based communication infrastructures based on the "Address-Event Representation" (AER) [8], [13], [14]. These two factors combined led to the construction of an impressive set of pulse-based multi-chip systems. Recent examples include vision-based systems that emulate the orientation selectivity functions of the visual cortex [2], [6], [8], or large-scale systems that can perform convolution, segmentation and object tracking in natural scenes [15].

As large scale multi-chip VLSI networks of spiking neurons are becoming more and more diffuse, the development of robust spike-based learning algorithms and circuits compatible with these systems is even more important. These learning circuits should enable the multi-chip neural systems they are embedded in to adapt to the statistics of their input signals, to learn and classify complex sequences of spatio-temporal patterns (*e.g.* arising from visual or auditory signals), and eventually to interact with the user and the environment.

This latter feature implies that the learning circuits should have biologically plausible time constants (*i.e.* of the order of milliseconds), so that they are matched to the signals they process, and are inherently synchronized with the real world events. In addition if the spike-based systems need to operate in real-world scenarios, with large variation in the input signals, in the ambient temperature, and over long periods of time, it would be desirable if the learning circuits were governed by mechanisms that operate over a multitude of time-scales (ranging from fractions of milliseconds to days or even months), similar to the short/long term plasticity and homeostatic plasticity mechanisms found in biological neural systems [16].

Plasticity and learning mechanisms involving physical synapses, either biological or electronic, have to cope with two main problems: 1) how to modify the synapses in order to learn associations and generate memories, and 2) how to protect memories against the passage of time and the over-writing due to the storage of new memories. Memory protection is a serious problem because the typical memory lifetimes of realistic synapses, which have weights that vary within finite bounds, grow only logarithmically with the number of synapses [17]. Increasing the number of states each synaptic weight has, within its bounds, leads to an improvement which grows only linearly with the states, or at most quadratically if fine tuning is allowed [18]. Therefore an efficient strategy for protecting previously stored memories is to use just two stable synaptic efficacy states per synapse, rather than many, but dramatically reducing the average number of transitions made from one stable state to the other. By modifying the synaptic weight of only a small random subset of synapses, memory lifetimes increase by a factor inversely proportional to the probability of synaptic modification [17].

Problem #1, of *how* to modify the synapses in spiking networks of neurons, has been the subject of renewed interest and has recently led to the definition of a promising class of spike-driven learning rules that are particularly well suited to VLSI implementation. A popular mechanism among these is Spike Timing Dependent Plasticity (STDP) [19]. In STDP the relative timing of pre- and post-synaptic spikes determine how

1-4244-0921-7/07 $25.00 © 2007 IEEE.

to update the efficacy of a synapse. It has been shown, both in theoretical models and VLSI implementations, that STDP can be effective in learning to classify spatio-temporal spike patterns [10], [20]. However STDP in its simplest form is not suitable for learning patterns of mean firing rates [19].

We implemented in VLSI a spike-driven learning rule, based on the timing of the pre-synaptic spike, on the membrane potential of the post-synaptic neuron and on a slow Calcium proportional to the post-synaptic neuron's mean firing rate. Such a model has been shown [21] to be able to classify patterns of mean firing rates, to capture the rich phenomenology observed in neurophysiological experiments on synaptic plasticity (see refs. therein), and to reproduce the classical STDP phenomenology. Moreover, if the patterns of firing rates are noisy, as those observed in cortical recordings *in vivo* or those obtained from AER sensors in response to real-world stimuli, this model modifies only a randomly selected subset of synapses, thus ensuring memory protection.

Here we describe the basic principles of this spike-driven learning rule (fully characterized in [21]) and present its neuromorphic VLSI implementation. The circuits developed are compatible with the ones used in the pulse-based multi-chip systems recently proposed, use the same asynchronous event-based communication infrastructure, and support long-term homeostatic plasticity mechanisms [22].

II. The spike-based learning rule

In our model every synapse has just two stable states, and each neuron is used as a classifier that learns to classify patterns of mean firing rates in a semi-supervised fashion. During training, the patterns to be classified are presented to the pre-synaptic synapses together with a teacher signal that steers the activity of the post-synaptic neuron toward the desired response. If the signal generated by the neuron in response to the input pattern, weighted by the learned synaptic weights, produces the desired response then the synapses are not modified. Otherwise the synaptic weights are updated upon the occurrence of the pre-synaptic spikes, in a stochastic manner. Specifically, if the neuron's output is low and the teacher signal is high, the synaptic efficacy is pushed toward its potentiated state, and if the neuron's output is high and the teacher signal is low, the synapse is driven toward its depressed state. In these conditions the synapse makes a transition to one of its two stable states with a probability that is proportional to the pre-synaptic firing rate. This behavior is the result of the interplay between the spike-triggered weight-update mechanism and sustained bistable synaptic dynamics, by which the synaptic efficacy is actively driven to its "low" stable state if it is below a threshold V_{wth}, and to its "high" state if it is above the same threshold. This bi-stability mechanism guarantees memory preservation in the absence of stimuli, or when the pre-synaptic activity is very low. The direction of the spike-triggered weight-update jumps (up-wards or down-wards) depend on the state of the post-synaptic membrane potential, at the time of the pre-synaptic spike arrival. Specifically, up-ward jumps are possible if the post-synaptic membrane potential is high and down-ward jumps are possible if it is low. As the state of the post-synaptic membrane potential correlates with the neuron's activity, the synapse tends to get potentiated when the neuron fires at a high rate, and depressed when it fires at a low rate. For a given pair of pre and post-synaptic firing rates, the synaptic modifications are consolidated only if the threshold V_{wth} is crossed during the presentation of the stimulus. If the pre-synaptic activity is noisy (*e.g.* a Poisson spike train), then memory consolidation occurs only with some probability, which guarantees the stochasticity of the transitions [17].

In order to determine when to stop updating the synaptic weight, we introduced a "Calcium" variable, proportional to the neuron's mean firing rate. When a neuron classifies correctly an input pattern, its Calcium variable will be either very high or very low. In these conditions we switch off the weight-update mechanism and stop learning. Indeed both conditions indicate that the input generated by the plastic synapses and the teacher signal are in agreement (*e.g.* both the weighted input and the teacher are high, and their currents sum up to drive the post-synaptic neuron to elevated mean activity). Conversely if the Calcium variable is in an intermediate range we enable learning.

Both the stochasticity supported by the synaptic dynamics, and this stop learning mechanism allow our model neurons to linearly classify separable patterns of mean firing rates [23]. Non-linearly separable patterns can be classified by simply using more than one output unit per class [21].

III. The neuromorphic VLSI implementation

To test the learning capabilities of a real physical implementation of the model described in Section II we fabricated an AER trans-ceiver chip comprising an array of 16 integrate-and-fire (I&F) neurons and 2048 synaptic circuits: 128 per neuron, of which 120 are plastic, 4 excitatory non-plastic and 4 inhibitory, non-plastic. The chip, produced using a standard 0.35μm CMOS technology, occupies an area of $6.1mm^2$.

Input spike patterns are provided to the synapses via the asynchronous AER interfacing circuits [14]. Each synapse circuit uses a diff-pair integrator (a recent current-mode low-pass filter described in [24]) to generate an output current proportional to the mean rate of its input spike train, modulated by its synaptic weight. The I&F neurons (described in [7]) integrate the input currents produced by the synapses and generate output spike trains with mean firing rates proportional to their input currents. The teacher signal is provided to the neurons by using one of the 4 non-plastic excitatory synapses, with a fixed synaptic weight.

The spike-based learning circuits that implement the model of Section II can be subdivided into two main blocks: a post-synaptic stop-learning control module, present in the neuron's soma, and a spike-triggered weight-update module with bistable synaptic dynamics, present in each plastic synapse.

The post-synaptic stop-learning control module, shown in Fig. 1(a), is responsible for generating the two global signals V_{UP} and V_{DN}, shared among all synapses belonging to the same

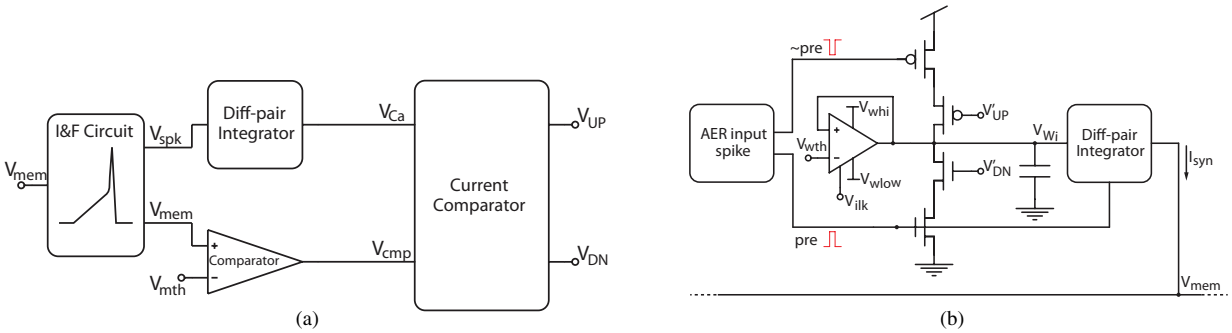

Fig. 1. (a) Post-synaptic weight control module. This module comprises a low-power I&F neuron circuit, a diff-pair integrator low-pass filter, a voltage comparator and a current comparator. (b) Pre-synaptic weight update module. An AER input block comprises the interfacing circuits that generate the *pre* and $\sim$ *pre* pulses. An amplifier in positive-feedback configuration drives the weight voltage V_{Wi} toward one of the two stable states V_{wlow} or V_{whi}. The transistors driven by the signals *pre* and ~*pre*, and the ones controlled by V'_{UP} and V'_{DN} form the weight-update circuit. The diff-pair integrator block generates the final synaptic current I_{syn}, summed to the currents generated by all other synapses into the V_{mem} node of the I&F circuit.

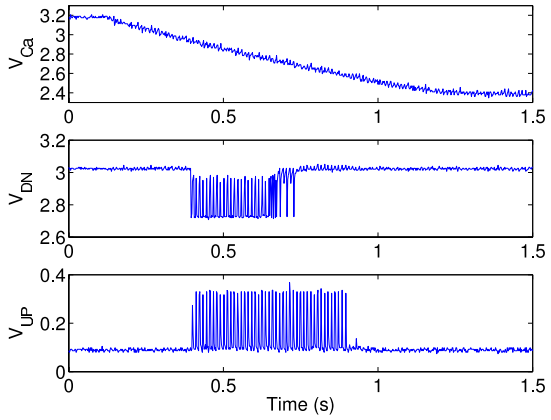

Fig. 2. Post-synaptic circuit data. (a) State of the V_{UP} and V_{DN} voltages are active only if the V_{Ca} signal is in an intermediate range.

dendritic tree, that enable positive and negative weight updates respectively. Post-synaptic spikes V_{spk}, generated by the I&F neuron are integrated by a diff-pair integrator. The integrator produces a V_{Ca} signal, related to the Calcium concentration in real neurons, that is compared to three different thresholds (V_{th1}, V_{th2}, and V_{th3}) by a comparator block made using three winner-take-all circuits. In parallel, the neuron's membrane potential V_{mem} is compared to a fixed threshold V_{mth}. The values of V_{UP} and V_{DN} depend on the state of the neuron's membrane potential and its Calcium concentration (see Fig. 2). Specifically if $V_{th1} < V_{Ca} < V_{th3}$ and $V_{mem} > V_{mth}$ then increases in synaptic weights ($V_{UP} < V_{dd}$) are enabled. And if $V_{th1} < V_{Ca} < V_{th2}$ and $V_{mem} < V_{mth}$, then decreases in synaptic weights ($V_{DN} > 0$) are enabled. Otherwise no changes in the synaptic weights are allowed ($V_{UP} = V_{dd}$, and $V_{DN} = 0$).

The pre-synaptic weight-update module is shown in Fig. 1(b). It comprises four main blocks: an input AER interfacing circuit [14], a bi-stability weight refresh circuit, a spike-triggered weight update circuit and a current-mode diff-pair integrator circuit [24]. Upon the arrival of an input address-event, the AER circuits produce an active-high pulse *pre*, and a complementary active-low pulse ~*pre*. These pulses trigger the weight update block. The *pre* pulse is also used to drive the plastic synapse's diff-pair integrator. This circuit generates an output current I_{syn} with exponential temporal dynamics, proportional to the input firing rate, modulated by its weight bias V_{Wi} [24]. The bi-stability weight refresh circuit is a positive-feedback amplifier with very small "slew-rate" (set by the V_{ilk} bias) that compares the weight voltage V_{Wi} to a set threshold V_{wth}, and slowly drives it toward one of the two rails V_{whi} or V_{wlow}, depending whether $V_{Wi} > V_{wth}$ or $V_{Wi} < V_{wth}$ respectively. This bistable drive is continuous and its effect is superimposed to the one from the spike-triggered weight update circuit.

If during a pre-synaptic spike the V_{UP} signal from the post-synaptic stop-learning control module is enabled ($V_{UP} < V_{dd}$), the synapse's weight V_{Wi} undergoes an instantaneous increase. Similarly, if during a pre-synaptic spike the V_{DN} signal from the post-synaptic weight control module is high, V_{Wi} undergoes an instantaneous decrease. If the weight increases bring V_{Wi} above the V_{wth} threshold, the bi-stability block will slowly drive V_{Wi} toward V_{whi} (thus consolidating the potentiated state of the synapse). Conversely, if the weight decreases bring V_{Wi} below the V_{wth} threshold, the bi-stability block will consolidate the synapse's depressed state, driving V_{Wi} to the V_{wlow} stable state.

IV. Weight-Update Measurements

To test the weight update mechanism, we made the post-synaptic neuron fire at an average frequency of 80Hz, using one of the non-plastic synapses, and stimulated the pre-synaptic synapse with Poisson distributed spike trains with a mean firing rate of 100Hz for a period of 250ms. Figure 3 shows an example of a stimulation session where the weight was increased several times during the trial, but never crossed the V_{wth} threshold, and therefore never made an Long Term

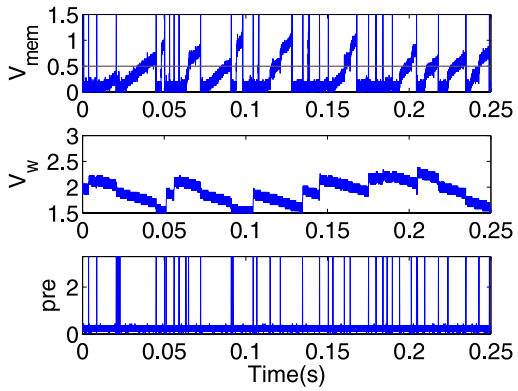

Fig. 3. Synaptic weight updates. Depending on the state of V_{mem}, the weight V_w is either increased or decreased with every pre-synaptic spike. Simultaneously, the bi-stability circuit drives V_w to the synapse's low stable state as long as $V_w < V_{wth}$ ($V_{wth} = 2.5V$ in this example).

Potentiation (LTP) transition. Due to stochastic nature of the pre- and post-synaptic spiking activity, some instances of pre- and post-synaptic spiking patterns with the same mean firing rates can induce an LTP transition, while others don't. The stochastic nature of LTP is an essential feature of the learning model described in Section II that our circuits fulfill. The probability of inducing long-term potentiation, or long-term depression can be easily controlled by acting on the bias parameters of the learning circuits (such as V_{wth}, V_{mth}, *etc.*), as well as the mean frequencies of the input and teacher signals.

V. Conclusions

We proposed a set of circuits for implementing an elaborate spike-driven synaptic learning rule and presented its neuromorphic VLSI implementation. We verified experimentally the functional behavior of the model proposed in [21] and showed how the plasticity circuits proposed, meet the specifications of the model. As all requirements necessary for learning complex patterns are met by the circuits proposed, the chip described is well suited to classification tasks on real-world problems using real-time spike data, *e.g.* obtained from AER sensory devices such as silicon retinas or silicon cochleas [1], [3].

Acknowledgment

The design of the VLSI chip and the experimental measurements were done together with Srinjoy Mitra. This work was supported by the EU grants ALAVLSI (IST-2001-38099), the Swiss National Science Foundation grant no. PP00A–106556, and by the ETH TH0–20174–04 grant.

References

[1] P. Lichtsteiner, C. Posch, and T. Delbrück, "A 128×128 120dB 30mW asynchronous vision sensor that responds to relative intensity change," in *2006 IEEE ISSCC Digest of Technical Papers*. IEEE, 2006, pp. 508–509.

[2] T. Y. W. Choi, P. A. Merolla, J. V. Arthur, K. A. Boahen, and B. E. Shi, "Neuromorphic implementation of orientation hypercolumns," *IEEE Transactions on Circuits and Systems I*, vol. 52, no. 6, pp. 1049–1060, 2005.

[3] A. van Schaik and S.-C. Liu, "AER EAR: A matched silicon cochlea pair with address event representation interface," in *IEEE International Symposium on Circuits and Systems*, vol. V, May 2005, pp. 4213–4216.

[4] L. Alvado, J. Tomas, S. Saighi, S. Renaud-Le Masson, T. Bal, A. Destexhe, and G. Le Masson, "Hardware computation of conductance-based neuron models," *Neurocomputing*, vol. 58–60, pp. 109–115, 2004.

[5] C. Rasche and R. Douglas, "An improved silicon neuron," *Analog Integrated Circuits and Signal Processing*, vol. 23, no. 3, pp. 227–36, 2000.

[6] U. Mallik, R. J. Vogelstein, E. Culurciello, R. Etienne-Cummings, and G. Cauwenberghs, "A real-time spike-domain sensory information processing system," in *Proceedings of IEEE International Symposium on Circuits and Systems*, vol. 3, 2005, pp. 1919–1922.

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

[8] E. Chicca, A. M. Whatley, V. Dante, P. Lichtsteiner, T. Delbrück, P. Del Giudice, R. J. Douglas, and G. Indiveri, "A multi-chip pulse-based neuromorphic infrastructure and its application to a model of orientation selectivity," *IEEE Transactions on Circuits and Systems I, Regular Papers*, 2006, (In press).

[9] D. Badoni, M. Giulioni, V. Dante, and P. Del Giudice, "An aVLSI recurrent network of spiking neurons with reconfigurable and plastic synapses," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, IEEE. IEEE, May 2006, pp. 1227–1230.

[10] J. Arthur and K. Boahen, "Learning in silicon: Timing is everything," in *Advances in Neural Information Processing Systems 18*, Y. Weiss, B. Schölkopf, and J. Platt, Eds. Cambridge, MA: MIT Press, 2006.

[11] H. Riis and P. Hafliger, "Spike based learning with weak multi-level static memory," in *Proceedings of IEEE International Symposium on Circuits and Systems*. IEEE, 2004, pp. 393–396.

[12] A. Bofill-i Petit and A. Murray, "Learning temporal correlations in biologically-inspired aVLSI," in *Proceedings of IEEE International Symposium on Circuits and Systems*, vol. V. IEEE, 2003, pp. 817–820.

[13] F. Gomez-Rodriguez, R. Paz, A. Linares-Barranco, M. Rivas, L. Miro, S. Vicente, G. Jimenez, and A. Civit, "AER tools for communications and debugging," in *Proceedings of IEEE International Symposium on Circuits and Systems*. IEEE, 2006, pp. 3253–3256.

[14] 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. 416–34, 2000.

[15] R. Serrano-Gotarredona, M. Oster, P. Lichtsteiner, A. Linares-Barranco, R. Paz-Vicente, F. Gómez-Rodríguez, H. Kolle Riis, T. Delbrück, S. C. Liu, S. Zahnd, A. M. Whatley, R. J. Douglas, P. Häfliger, G. Jimenez-Moreno, A. Civit, T. Serrano-Gotarredona, A. Acosta-Jiménez, and B. Linares-Barranco, "AER building blocks for multi-layer multi-chip neuromorphic vision systems," in *Advances in Neural Information Processing Systems*, S. Becker, S. Thrun, and K. Obermayer, Eds., vol. 15. MIT Press, Dec 2005.

[16] G. Turrigiano and S. Nelson, "Homeostatic plasticity in the developing nervous system," *Nature Reviews Neuroscience*, vol. 5, pp. 97–107, February 2004.

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

[18] S. Fusi and L. Abbott, "Limits on the memory storage capacity of bounded synapses," *Nature Neuroscience*, 2006, (In press).

[19] L. Abbott and S. Nelson, "Synaptic plasticity: taming the beast," *Nature Neuroscience*, vol. 3, pp. 1178–1183, November 2000.

[20] R. Gütig and H. Sompolinsky, "The tempotron: a neuron that learns spike timing–based decisions," *Nature Neuroscience*, vol. 9, pp. 420–428, 2006.

[21] J. Brader, W. Senn, and S. Fusi, "Learning real world stimuli in a neural network with spike-driven synaptic dynamics," *Neural Computation*, 2006, (Submitted).

[22] C. Bartolozzi and G. Indiveri, "Silicon synaptic homeostasis," in *Brain Inspired Cognitive Systems 2006*, October 2006, pp. 1–4.

[23] W. Senn and S. Fusi, "Learning Only When Necessary: Better Memories of Correlated Patterns in Networks with Bounded Synapses," *Neural Computation*, vol. 17, no. 10, pp. 2106–2138, 2005.

[24] C. Bartolozzi and G. Indiveri, "Synaptic dynamics in analog VLSI," *Neural Computation*, 2007, (In press).