

A VLSI spike-driven dynamic synapse which learns only when necessary

Srinjoy Mitra, Stefano Fusi, and Giacomo Indiveri

Institute of Neuroinformatics

University of Zurich - ETH Zurich

Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

Email: [srinjoy|giacomo|fusi]@ini.phys.ethz.ch

Abstract— We describe an analog VLSI circuit implementing spike-driven synaptic plasticity, embedded in a network of integrate-and-fire neurons. This biologically inspired synapse is highly effective in learning to classify complex stimuli in semi-supervised fashion. The circuits presented are designed in sub-threshold CMOS consuming extremely low power. The pulse-based neural network communicates with the outside world using the Address Event Representation in an asynchronous fashion. We present measurements from a test chip, characterizing all the modules of the circuit and show how they match well with theoretical expectations. We finally demonstrate that the learning mechanism of the synapse is fully functional by stimulating it with Poisson distributed spike trains.

I. INTRODUCTION

Recent developments in asynchronous communication infrastructures based on the *Address-Event Representation* (AER) [1], [2] enabled researchers to build large scale pulse-based neural networks implemented as multi-chip VLSI systems. Within this context, several multi-neuron *transceiver* chips have been proposed [3]–[5]. These types of neuromorphic chips comprise arrays of neurons and synapses implemented with analog circuits and interfaced to asynchronous digital circuits used to deliver input pulses to the network’s synapses, and to transmit output spikes to the AER bus. One of the most interesting features of these multi-chip neural systems is their ability to implement spike-driven learning algorithms. In this paper we present a VLSI synapse that implements a recently proposed model of spike-driven synaptic learning which has been successfully used to classify complex patterns in a semi-supervised fashion [6]. The chip that hosts the synaptic circuits proposed here, has been designed in parallel with a device that has both similar and complementary features [4], in an effort to build components for constructing large-scale multi-chip systems capable of learning and classifying complex spatio-temporal sequences in real-time.

II. LEARNING RULE AND SYNAPTIC MODEL

In these types of spike-based learning systems, during training, the activity pattern to be classified is presented to the neuron’s plastic synapses while a teaching signal steers the post-synaptic activity in the direction of the desired output pattern (i.e. the pattern representing the class to which the input belongs). The activity of each post-synaptic neuron is composed of an input generated component and of the

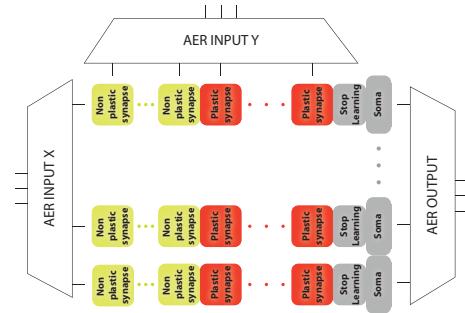


Fig. 1. Floor-plan of the multi-neuron transceiver test chip. The test chip contains 32 neurons and 8192 synapses. Each neuron receives input currents from 240 plastic synapses and 16 non-plastic synapses.

teacher’s signal component. When both components agree, the output neuron is driven either to its maximum or to its minimum activity. These two firing regimes are interpreted as an indication that the neuron would classify correctly the pattern even in the absence of the teacher signal and that the synapses should not be modified any further. The importance of this stop-learning criterion is at least double: 1) it allows to classify highly correlated patterns [7], and 2) it can greatly reduce the average number of synaptic modifications during the learning phase. The second point is particularly important for bistable synapses (such as the one we propose), as they can preserve memory of only the last synaptic modifications [8].

In order to guarantee convergence of this learning process, the stop-learning condition must be complemented by a mechanism which consolidates the synaptic modification on long time scales in a randomly chosen small fraction of synapses [7]. We implement this mechanism by using Poisson distributed spike-trains as input signals. Upon the arrival of a pre-synaptic spike, the synapse reads the post-synaptic membrane potential and the synaptic weight is updated (the weight goes up/down when the post-synaptic membrane potential is high/low). During the intervals between two pre-synaptic spikes the synaptic weight is slowly driven toward the closest of two stable states corresponding to the minimal and the maximal values of the synaptic efficacy. Long term potentiation occurs when both the pre- and the post-synaptic mean firing rates are high enough to drive the synaptic weight in the band where it is attracted toward the high stable state. This

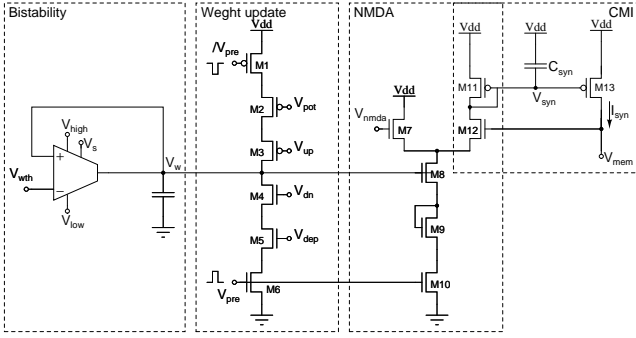


Fig. 2. Schematic diagram of the plastic synapse circuit.

occurs with a certain probability because the intervals between successive pre-synaptic spikes are random [8]. Analogously, long term depression occurs probabilistically when the pre-synaptic neuron has elevated firing rate and the post-synaptic neuron has low activity. Reading the post-synaptic membrane potential allows to encode the mean firing rates and at the same time, preserves the fluctuations due to the stochasticity of the intervals between pre-synaptic spikes. Moreover it allows to make the synapse sensitive to higher order statistics of the pre and post-synaptic activity (*e.g.* variability of the inter-spike interval, synchronization, and exact temporal relation between pre and post-synaptic spikes). However, as this mechanism alone cannot encode mean frequencies on a wide range, to stop learning when the post-synaptic frequency is too high or too low, we introduced an extra element that integrates spikes on a longer time scales.

III. CIRCUIT DESCRIPTION

The test chip comprising the circuits described in this paper was fabricated using a standard $0.35\mu\text{m}$ CMOS technology, and occupies an area of 12^2mm . It implements an array of 32 neurons and 8192 synapses (see Fig.1). Thanks to the AER communication protocol, neural networks of arbitrary topology can be implemented by simply re-routing output spikes to input synapses of the same chip, or of other AER multi-neuron chips.

Each neuron in the array is connected to 256 synaptic circuits. There are 240 excitatory plastic synapses and 16 non-plastic ones, 8 of which are excitatory and 8 inhibitory. The weights of the non-plastic synapses are constant and can be set via external voltage references. All synapses can be addressed individually through the AER communication protocol. This transceiver chip can receive AER spikes via the X and Y decoders (see Fig.1) and it can transmit the activity of the neurons to the outside world asynchronously using an arbitrated communication scheme.

The circuits implementing the integrate-and-fire neuron (labeled “soma” in Fig.1) have been described in [5]. Here we focus on the plastic synapses and stop-learning module.

A. The Plastic Synapse

The schematic diagram of this circuit is shown in Fig.2. The synapse is composed of four main blocks: a current-

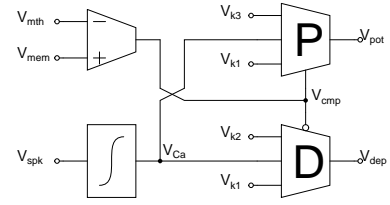


Fig. 3. Block diagram of the stop learning module. Two dual threshold comparator circuits P and D enable weight updates based on the integrator (bottom left) output V_{Ca} , and on the result of the comparison between V_{mem} and V_{mth} .

mirror integrator (CMI), the NMDA comparator, the weight update part, and the bi-stability circuit. The input signal is represented by V_{pre} , which produces an active high pulse when a pre-synaptic spike arrives. The width of the V_{pre} pulse can be controlled by a pulse extender circuit (not shown) that decouples the fast AER spikes from the analog synaptic circuit.

Each time a V_{pre} pulse is generated, the CMI subtracts a set amount of charge, set by V_w , onto C_{syn} to produce an excitatory post-synaptic current I_{syn} with biologically plausible dynamics. A detailed analysis of this circuit has been recently presented in [9]. The ‘NMDA’ part of the synapse gates the charge that is subtracted from C_{syn} by comparing the post-synaptic neuron’s membrane potential V_{mem} to a fixed threshold V_{mnda} . This circuit models the functionality of voltage-dependent NMDA receptors in real synapses [10].

The ‘weight update’ part receives signals from both pre and post-synaptic neurons and updates the weight on each pre-synaptic spike. If the V_{pot} signal is low during a pre-synaptic spike (*i.e.* when V_{pre} is low), the node V_w receives a charge packet whose magnitude depends on the bias V_{up} and its potential increases accordingly. Similarly, if V_{dep} is high during a pre-synaptic spike, the weight V_w decreases. In the condition in which V_{pot} is high and V_{dep} is low, the weight is not modified, irrespective of the state of the pre-synaptic pulse.

The bi-stability circuit continuously and actively drives the weight V_w toward the stable state V_{high} or V_{low} , depending whether $V_w \geq V_{wth}$, at a rate set by V_s . The constant bias V_{wth} represents the synaptic state transition threshold. The slew rate of the bi-stability amplifier is typically set to be very low, such that the drive toward one of the two stable states is slow. The capacitance in the testability circuit has to be approximately 100fF. As the voltage V_w is always close to ground, we used a p-type MOSCAP to implement the capacitor. The total area occupied by the synapse (including the capacitor) is $630^2\mu\text{m}$.

B. Stop Learning Module

The stop learning module is composed of a standard comparator, a linear integrator, and two low-power dual threshold comparators [11] (see Fig.3).

The amplifier in the left top corner of Fig.3 compares the neuron’s membrane potential V_{mem} to the threshold V_{mth} and produces a digital signal V_{cmp} used to enable or disable the comparators P and D in the right part of the figure. The

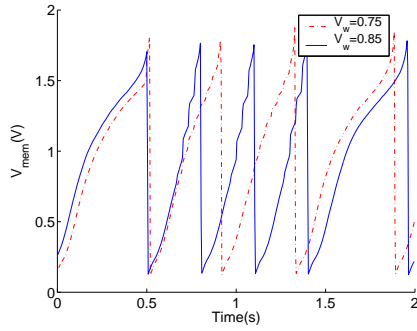


Fig. 4. Response of the neuron to both constant current injection and synaptic stimulation for different values of V_w . Synaptic stimulation occurs between 0.5s and 1.5s

linear integrator, in the bottom left corner of Fig.3, integrates the spikes from the neuron to generate the voltage V_{Ca} , functionally equivalent to the calcium concentration in real neurons. The integrated signal V_{Ca} is harnessed to stop learning when the neuron responds correctly (*i.e.* when it fires at either very high, or very low rates). When the teacher signal and the current generated by the synaptic input disagree, the post-synaptic neuron fire at intermediate frequencies, far from the minimum and the maximum. In this condition V_{Ca} fluctuates in an intermediate range, indicating that learning should occur. Specifically, the dual threshold comparators P and D compare the signal V_{Ca} to two threshold signals for defining when the synapse should be allowed to increase its weight (comparator P), or when it should be allowed to decrease it (comparator D). If $V_{k1} < V_{Ca} < V_{k3}$ and P is on, the signal V_{pot} switches to its active low state and enables a positive weight update of amplitude set by V_{up} (see Fig.2). If $V_{k1} < V_{Ca} < V_{k2}$ and D is on, V_{dep} switches to its active high state and enables a negative weight update of amplitude set by V_{dn} .

Functionally, enabling positive or negative updates of synaptic weight depending on the state of the post-synaptic neuron's membrane potential is compatible with spike time dependent plasticity (STDP) [8], [12].

IV. RESULTS

We characterized the different components of the plastic synapse and stop-learning modules both as isolated circuits and as full functional blocks. The data of Fig.4 shows the membrane potential of the post-synaptic neuron in response to a constant current superimposed to synaptic stimulation. When the neuron receives only constant current input (from 0s to 0.5s, and from 1.5s on), it integrates, reaches the spiking threshold, generates a spike (and an AER event) and is reset. The integration is not perfectly linear due to the fact that the circuit implements a leaky integrate-and-fire neuron with spike-frequency adaptation properties [5]. During synaptic stimulation (from 0.5s to 1.5s in Fig.4) the soma receives an additional excitatory post-synaptic current (EPSC), as long as V_{mem} is above the NMDA threshold, and the neuron's firing rate increases with increasing synaptic weight values V_w .

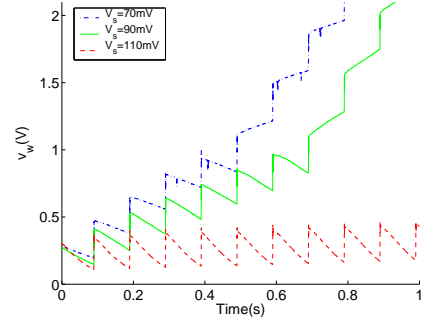


Fig. 5. Synaptic weight dynamics, below and above the transition threshold for different bi-stability slope settings. As long as the weight voltage lies below the transition threshold (900mV in this experiment), the bi-stability circuit drives V_w toward its low asymptote. As soon as the weight crosses the transition threshold, the bistable circuit drives V_w to the high stable state.

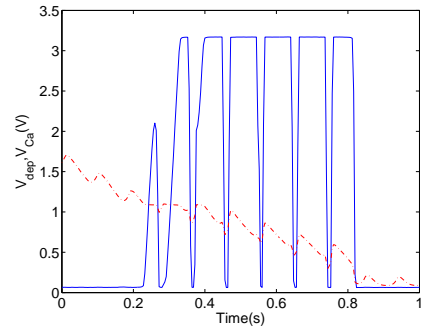


Fig. 6. State of the V_{dep} signal (solid line) as a function of the integrated post-synaptic spiking activity V_{Ca} (dashed line). When V_{Ca} is within the right range the V_{dep} signal switches between low and high values, depending on the sign of $(V_{mem} - V_{mth})$.

To characterize the bi-stability and weight update parts of the plastic synapse circuit, we set the signals V_{pot} and V_{dep} to constant values. When both V_{pot} and V_{dep} are set low, the weight-update part of the synapse increases the signal V_w with every pre-synaptic spike. In Fig.5 we show the evolution of V_w in response to a series of pre-synaptic spikes, and as a function of different slew rates of the bi-stability circuit: as long as V_w is below the synaptic state transition threshold V_{wth} , the bi-stability circuit drives V_w back toward V_{low} , but as soon as V_w crosses V_{wth} the synaptic weight V_w is quickly brought to the V_{high} stable state, given that both bi-stability and weight-update circuits act in the same direction. In the bottom trace of Fig.5 the slew rate is high enough to keep V_w below the threshold.

In Fig.6 we plot data showing the time evolution of V_{Ca} and of V_{dep} , as the mean firing rate of the post-synaptic neuron decreases. When V_{Ca} falls within the right region (here approximately between 0.2V and 1.1V), V_{dep} starts switching between its low and high states, depending on the outcome of the comparison between V_{mem} and V_{mth} . The behavior of the V_{pot} signal is functionally equivalent, and hence not shown.

The up and down jump probabilities of V_w as well as its transition probability depend on higher order statistics of the pre and post-synaptic neuron frequency. In Fig.7, we show the

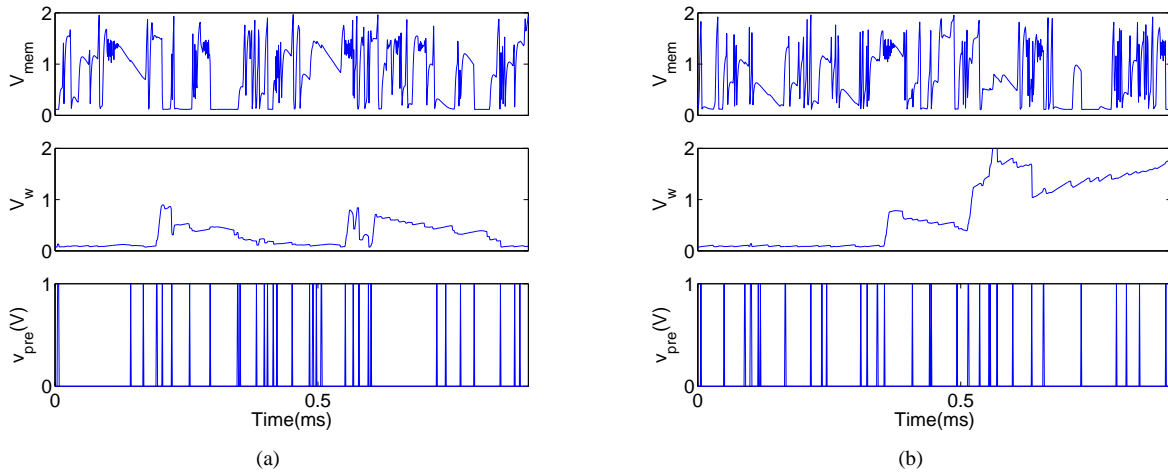


Fig. 7. Post-synaptic membrane potential (top trace), synaptic weight (middle trace), and pre-synaptic spikes (bottom trace). Despite the fact that both pre- and post-synaptic neurons have same mean firing rates (in the left and right panels), the weight undergoes no transitions (left case) or one transition (right case), depending on the precise timing of the pre- and post-synaptic spikes.

results of a learning experiment where we tried to induce long-term potentiation on the plastic synapse. The top trace of the figure shows the post-synaptic neuron's membrane potential and the bottom trace shows the input pre-synaptic spikes. To set the mean firing rate of the post-synaptic neuron to a specific value (in order to bring V_{Ca} in the right regime) we stimulated its non-plastic AER synapses with an artificial Poisson spike train input. Even though both the experiments (a) and (b) have equal mean pre-synaptic frequency and nearly equal post-synaptic frequency, two completely different temporal dynamics of V_w (middle trace) can be observed. The jumps may or may not consolidate into a transition (middle trace of Fig.7(b) and (a) respectively) depending on the specific realization of the stochastic pre- and post-synaptic activity. The transition probability increases with the degree of synchrony of the pre and post-synaptic activity. At parity of mean firing rates, the synaptic modification is consolidated only in a small fraction of cases, which depends on the second and higher order statistics of the pattern of activities [8].

V. CONCLUSION

We proposed a VLSI implementation of a spike based learning algorithm, mapped onto an AER network of integrate-and-fire neurons. We presented experimental results characterizing both the individual components and the full functional aspect of the synaptic learning mechanism. The algorithm that our implementation is based upon has successfully been used to classify complex patterns [6]. Our results confirm that the silicon model of synaptic plasticity is functionally equivalent to the original algorithmic one. Therefore we plan to test spike-driven learning at the network level and apply the chip to real-world classification problems.

ACKNOWLEDGMENT

This work was funded by the EU grants ALAVLSI (IST-2001-38099) and DAISY (FP6-2005-015803), and by the ETH

TH0-20174-04 grant. We thank P. Del Giudice and V. Dante, of the Italian ISS, for the original design of the PCI-AER board used to test the chip.

REFERENCES

- [1] S. R. Deiss, R. J. Douglas, and A. M. Whatley, "A pulse-coded communications infrastructure for neuromorphic systems," in *Pulsed Neural Networks*, W. Maass and C. M. Bishop, Eds. MIT Press, 1998, ch. 6, pp. 157–178.
- [2] K. Boahen, "A burst-mode word-serial address-event link III: Analysis and test results," *IEEE Trans. Circuits Syst. I*, vol. 51, pp. 1292–1300, July 2004.
- [3] R. J. Vogelstein, U. Mallik, and G. Cauwenberghs, "Silicon spike-based synaptic array and address-event transceiver," in *Proceedings of IEEE International Symposium on Circuits and Systems*. IEEE, May 2004, pp. V385–V388 Vol. 5.
- [4] 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, 2006, submitted.
- [5] 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. on Neural Networks*, Vol. 17, Num. 1, pp. 1–12, January 2006.
- [6] J. Brader, W. Senn, and S. Fusi, "Learning real world stimuli in a neural network with spike-driven synaptic dynamics," *Neural Computation*, 2005, submitted.
- [7] W. Senn and S. Fusi, "Learning only when necessary: better memories of correlated patterns in networks with bounded synapses," *Neural Computation*, vol. 17, pp. 2106–2138, 2005.
- [8] S. Fusi, "Hebbian spike-driven synaptic plasticity for learning patterns of mean firing rates," *Biological Cybernetics*, vol. 87, pp. 459–470, 2002.
- [9] E. Chicca, G. Indiveri, and R. Douglas, "An adaptive silicon synapse," in *Proc. IEEE International Symposium on Circuits and Systems*. IEEE, May 2003, pp. I-81–I-84.
- [10] E. R. Kandel, J. Schwartz, and T. M. Jessell, *Principles of Neural Science*. Mc Graw Hill, 2000.
- [11] S. Mitra and G. Indiveri, "A low-power dual-threshold comparator for neuromorphic systems," in *2005 PhD Research in Microelectronics and Electronics*, vol. 2. Lausanne: IEEE, Jul 2005, pp. 402–405.
- [12] M. Coath, J. M. Brader, S. Fusi, S. L. Denham, "Multiple views of the response of an ensemble of spectro-temporal features support concurrent classification of utterance, prosody, sex and speaker identity," *Network: Computation in Neural Systems*, vol. 16, Num. 2-3, pp. 285-300.