A Neuromorphic aVLSI network chip with configurable plastic synapses

P. Camilleri[1]*, M. Giulioni[2]†, V. Dante[2], D. Badoni[3], G. Indiveri[4], B. Michaelis[1], J. Braun[1], P. del Giudice[2]

[1]: Otto-von-Guericke University, Magdeburg, Germany
[2]: Italian National Inst. of Health and INFN, Rome, Italy
[3]: INFN-RM2, Rome, Italy
[4]: Institute of Neuroinformatics, UNI | ETHZ, Zurich, Switzerland

Abstract

We describe and demonstrate the key features of a neuromorphic, analog VLSI chip (termed F-LANN) hosting 128 integrate-and-fire (IF) neurons with spike-frequency adaptation, and 16 384 plastic bistable synapses implementing a self-regulated form of Hebbian, spike-driven, stochastic plasticity. We were successfully able to test and verify the basic operation of the chip as well as its main new feature, namely the synaptic configurability. This configurability enables us to configure each individual synapse as either excitatory or inhibitory and to receive either recurrent input from an on-chip neuron or AER (Address Event Representation)-based input from an off-chip neuron. It's also possible to set the initial state of each synapse as potentiated or depressed, and the state of each synapse can be read and stored on a computer. The main aim of this chip is to be able to efficiently perform associative learning experiments on a large number of synapses. In the future we would like to connect up multiple F-LANN chips together to be able to perform associative learning of natural stimulus sets.

1. Introduction

The pioneering work of C. Mead [18] has introduced the term "neuromorphic engineering" for a growing family of analog, sub-threshold circuits, which implement the accepted equivalent circuits of biological neurons and synapses in VLSI technology. The ultimate aim of neuromorphic engineering is to mimic the capabilities of biological perception and information processing with a compact and energy-efficient platform. It is widely believed that this goal necessitates from the outset some mechanisms of "learning" that enables neuromorphic devices to adapt (or re-configure) themselves while interacting with an environment. Emulating the example of biological neurons and synapses, our neuromorphic devices attain an ability for "learning" by incorporating "Hebbian-like" mechanisms of synaptic plasticity. In the "Hebbian" scenario we adopt, the efficacy of a synapse is enhanced (i.e., its impact on the post-synaptic neuron is increased), when both the pre- and post-synaptic neurons are simultaneously highly active on a suitable time-scale, and reduced if the pre-synaptic neuron is active while the post-synaptic is not. Whether "Hebbian" learning is based on average firing rates or on individual spikes ("spike-time-dependent plasticity", or STDP) is a matter of continuing debate and a choice that strongly influences alternative designs of neuromorphic synapse circuits. The synaptic dynamics described here are spike-driven and implement a rate-based Hebbian learning, though it is compatible with some aspects of STDP.

The fact that a plausible synaptic device may assume only a limited number of alternative "states" has profound consequences for the memory capacity of a neural network. Specifically, when synaptic efficacy is bounded and changes in discrete steps, any deterministic learning rule (e.g., a "Hebbian" rule) can be shown to yield highly unfavorable scaling laws for memory capacity [1, 10]. The intuitive reason is that newly encoded memories rapidly erase earlier memories ("palimpsest property"). Perhaps surprisingly, far more favorable scaling laws may be attained with a stochastic learning rule, in which the Hebbian prescription renders synapses merely eligible for a state change but the probability of an actual change remains low. Thus, the problem of the inherently low memory capacity of biologically plausible synapses yields to a "stochastic" solution. Note, however, that this solution necessitates independent sources of "noise" at each individual synapse. Presumably not coincidentally, the necessary "noise" may be provided by the irregularity of the pre- and post-synaptic spike trains (as long as the network remains in an asynchronous activity regime).

*patrick.camilleri@ovgu.de
†giulioni@roma2.infn.it

0-7695-2946-1/07 $25.00 © 2007 IEEE
DOI 10.1109/HIS.2007.60

In short, the theoretical analysis of learning with biologically plausible synapses appears to provide a compelling computational argument for implementing neurons as spiking elements.

Associative learning in networks of spiking IF neurons with stochastic synapses has been studied both in simulation and [2, 10, 13] and in neuromorphic realizations [9, 11]. However, these first efforts were limited to artificially simplistic stimulus sets (e.g., strictly non-overlapping neural representations). To extend associative learning to more realistic stimulus sets, a further modification of the synaptic rule has been proposed, informally known as "stop-learning" [6]. In this modification, synaptic changes are additionally conditioned on average post-synaptic activity being neither too high nor too low: synapses targeting too-active neurons are not further strengthened and synapses targeting too-inactive neurons are not further weakened. This additional condition becomes crucial when partially overlapping patterns of activity are to be distinguished, as it prevents excessive potentiation of synapses in the overlapping parts, which would otherwise spoil the network's ability to distinguish these patterns. The suitability of this learning strategy was demonstrated in a Perceptron-like network for linearly separable patterns [20]. Extensions of the "stop-learning" strategy to spiking networks with recurrent connectivity are currently being pursued by several groups. We implemented a preliminary version of the "stop-learning" synapse in a previous chip [3]; the present network implementation, besides improving on several synaptic design issues, will offer a wider range of collective dynamics through a more flexibly reconfigurable architecture. A synaptic design inspired by the same "stop-learning" principles was proposed in [19].

Section 2 provides an overview of the chip architecture and Section 3 describes synapse circuits. Section 4 details the ability to configure individual synapses and shows some of our obtained results.

2. Chip architecture and main features

We describe a VLSI chip implementing a reconfigurable network of 128 integrate-and-fire neurons with spike-frequency adaptation and 16 384 (128×128) bistable, stochastic synapses implementing a Hebbian rule with "stop-learning" (see Fig. 1). The chip has a total area of 68.9 mm^2 with each synapse and neuron occupying 3,200 μm^2 and 2,400 μm^2 respectively. A standard 0.35μm CMOS technology process from *austriamicrosystems* (AMS) was used.

The synaptic matrix is configurable in such a way to support either all-to-all recurrent connectivity, or exclusively external (AER-based) connectivity, or any combination of both. In addition, the initial state of efficacy and the excitatory or inhibitory nature of synapses may be set individually for each synapse. The synaptic matrix is arranged in four identical 64×64 sub-matrices. As every signal entering a sub-matrix is properly buffered, these sub-matrices could in the future serve as building blocks for considerably larger chips. The chip is compliant with the AER asynchronous communication protocol widely used in the neuromorphic engineering community. Specifically, AER-based communication is handled through the PCI-AER board [7, 8] which allows four chips to be connected together (e.g., to implement a recurrent network of 512 neurons with a uniform 25% connectivity).

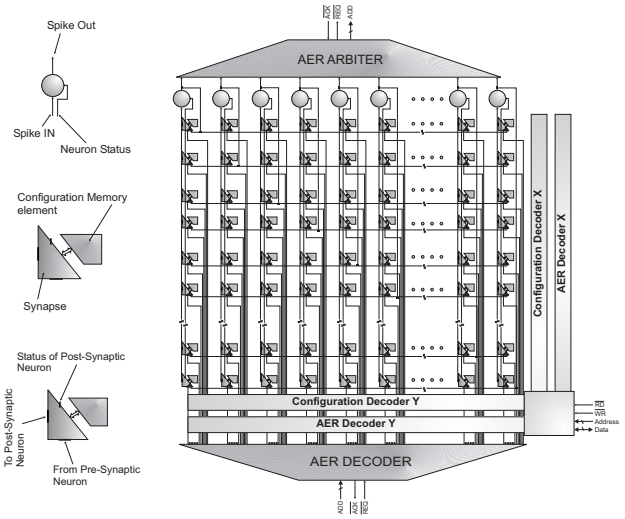

Figure 1. F-LANN chip architecture.

The neuron circuit, which implements an IF neuron with a constant leakage term and a lower bound for the membrane potential $V(t)$, was introduced in [18] and studied theoretically in [12]. An additional dynamic variable associated with the neuron reflects its recent average activity and is termed $I_{Ca}(t)$, following [6]. The variable I_{Ca}, which is incremented by each spike and decays exponentially between spikes, is implemented by a log-domain, exponential decay circuit. We use for V and I_{Ca} dynamics the low-power circuits described in [15, 16].

The dendritic tree of each neuron is composed of 128 synapses. Each synapse accepts as input, spikes from either internal or external neurons. In the latter case the spikes come in the form of AER events which are addressed to the correct synapses by the X-Y decoder. Excitatory synapses are plastic, inhibitory are fixed.

Even if in principle, recurrent connectivity can also be achieved by looping through the AER, the ability to reconfigure synapses as either recurrent or AER-based allows adequate flexibility to optimally balance AER bandwidth requirements and complexity of design.

Another XY-decoder allows synapses to be independently addressed and configured. In addition, dedicated hardware circuits have been added to directly set and read the internal state of selected synapses.

The AER input block, responsible for the communication handshaking, was designed for a multi-chip system. To avoid a single incorrect AER transaction blocking the AER bus, the latter is released without waiting for an acknowledgment from the target synapse. To this end, a transparent latch array stores the AER address as soon as it enters the bus. Similarly, an internal neuron contributing a spike to the AER bus does not wait for an external acknowledgment but resets immediately. Although this approach introduces a small possibility that some AER events are lost, it ensures that AER delays do not disrupt internal network activity.

All spikes generated within the chip are arbitered for access to the AER bus.

3. Synapse and Calcium Circuit

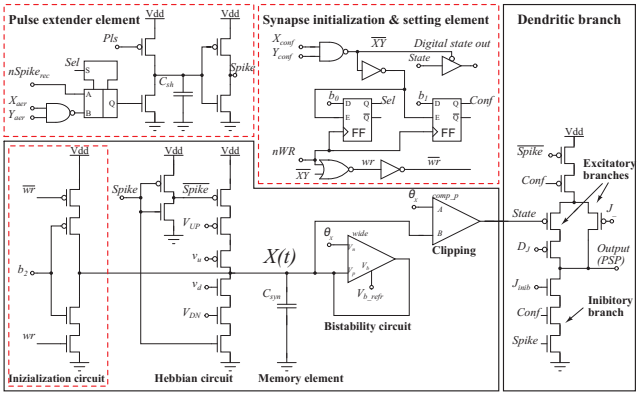

Figure 2. Synapse circuit main blocks.

Fig. 2 and Fig. 3 illustrate the synaptic circuit and the comparator system needed to implement the model described in [6] and briefly motivated in the introduction. Following the arrival of a pre-synaptic spike, X jumps upward or downward, depending on the following conditions on the post-synaptic state: $X(t) \rightarrow X(t) + a$ if $V(t) > \theta_p$ and $I_{TH1} < I_{Ca} < I_{TH3}$; $X(t) \rightarrow X(t) - b$ if $V(t) \leq \theta_p$ and $I_{TH1} < I_{Ca} < I_{TH2}$ where a and b are the tunable amplitudes of the jumps. In the absence of pre-synaptic spikes, if $X(t) > \theta_X$ ($X(t) < \theta_X$) X relaxes towards the upper (lower) barrier and the efficacy of the bistable synapse is set as 'potentiated' ('depressed'). The synaptic efficacy changes only when $X(t)$ crosses θ_X.

The Bistability sub-circuit (see Fig. 2) is a wide output-range transconductance amplifier with positive feedback: it attracts $X(t)$ towards the upper or lower stable value depending on the comparison with the threshold θ_X, which also determines, through the Clipping block (a two-stage open-loop comparator), the efficacy value ($J\_$ – 'depressed' or $J\_ + DJ$ – 'potentiated'). The UP and $DOWN$ signals on the left, coming from the Calcium block, exclusively enable the branches of the Hebbian circuit and inject or subtract a current regulated by v_u and v_d. The Dendritic branch is triggered by the pre-synaptic spike and generates the up/down jump in the post-synaptic $V(t)$ according to the configuration bit $Conf$ which sets the synapse as excitatory or inhibitory.

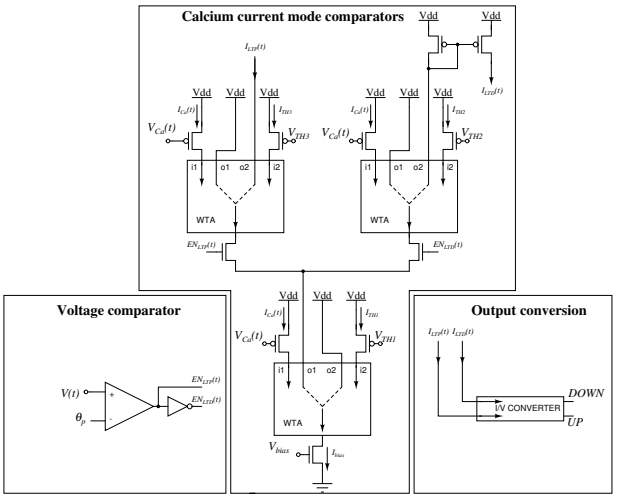

Figure 3. Comparators system.

The "stop-learning" mechanism relies on the "calcium" variable of the post-synaptic neuron. This variable, represented by the current $I_{Ca}(t)$, is incremented by each post-synaptic spike and decays exponentially between spikes. Accordingly, its value integrates the post-synaptic spiking activity in the recent past. Together with suitable thresholds, it determines which synaptic changes will be allowed to occur. For example, it can prevent an upward jump of $X(t)$ when the post-synaptic neuron is already very active, thus lowering the probability of synaptic potentiation.

The synapse accepts AER events (the 'AND' of X_{AER} and Y_{AER} signals in Fig. 2) or recurrent spikes $nSpike_{rec}$, depending on the configuration bit Sel. The event triggers the pulse extender circuit which generates a pulse $spike$ with a duration determined by an external bias voltage Pls. In typical conditions an AER event lasts around 200ns while the recurrent spike only 10-20ns. This circuit equalizes the recursive and AER pulse durations extending them to a few microseconds. This makes sure that the Hebbian circuit (see Fig. 2) is enabled for the same amount of time irrespective of whether the impinging spike was generated recursively or through the AER bus. This "long" interval of time al-

lows, together with parameters v_u and v_d a fine tuning of the amount of charge injected or subtracted from the synaptic capacitor C_{syn}, giving rise to the jumps in X. The same interval of time determines the duration of the induced synaptic current on the post-synaptic neuron.

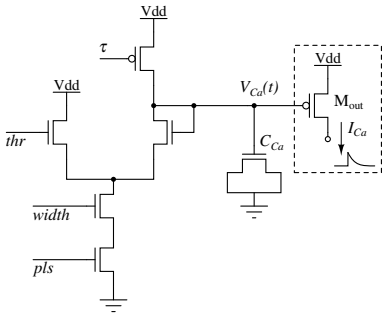

Figure 4. Diff-pair log-domain calcium circuit. The M_{out} MOSFET is part of the circuit shown in figure 3.

The circuit which generates the calcium variable I_{Ca}, (see Fig. 4) is a diff-pair integrator (DPI) implementing a log-domain filter. The output of this module, described in [4, 5], is a current which increases suddenly upon the arrival of impinging spikes and exponentially decays between two spikes. For constant average firing of the neuron, the average I_{Ca} current level is proportional to the firing rate. The M_{out} MOSFET is part of the WTA comparators system reported in Fig. 3.

I_{Ca} is compared to three thresholds I_{TH1}, I_{TH2}, and I_{TH3} in the module in Fig. 3 to generate the two signals UP and $DOWN$ shared among all synapses belonging to the same dendritic tree. The comparison is performed by three current-mode winner-take-all circuits (WTA) [14, 17]. In parallel, the instantaneous voltage value of the post-synaptic neuron potential $V(t)$ is compared to a threshold θ_P (see Fig. 3). Depending on the outcome of these comparisons, the current-comparator produces either an output current $I_{LTP} = I_{bias}$ enabling an upward jump for $X(t)$, a current $I_{LTD} = I_{bias}$ enabling a downward jump, or no output current at all ($I_{LTP} = I_{LTD} = 0$). Two corresponding voltages UP and $DOWN$ are produced by current-conveyors and broadcasted along the neuron's dendritic tree. This system of comparators implements the inequalities above for the dynamics of $X(t)$.

Fig. 5 illustrates the effect of the Calcium circuit on $X(t)$. Thresholds were set to have $I_{TH3} > I_{TH1} = I_{TH2}$ (which for the corresponding voltages applied to the gates of the p-MOSFETs implies $V_{TH3} < V_{TH1} = V_{TH2}$ – see dashed lines in the figure). The synapse is initially set depressed. The post-synaptic neuron is excited by a train of AER spikes (via a different synapse) with increasing frequency (corresponding upward jumps are visible in the $V_{\rm post}$. As long as $I_{TH1} = I_{TH2} > I_{Ca}$ ($V_{TH1} = V_{TH2} < V_{Ca}$), i.e. the post-synaptic neuron activity is low, neither up nor down transitions of X are allowed, and $X(t)$ stays fixed at its lower value, until I_{Ca} crosses $I_{TH1} = I_{TH2}$, when upward jumps of X become allowed. Upon crossing θ_X, the slope of the current attracting X towards the upper value is activated (this is when the synaptic efficacy gets potentiated – not shown). X undergoes upward jumps until $I_{TH3} < I_{Ca}$ ($V_{TH3} > V_{Ca}$), and upward jumps are forbidden. At this point X is driven towards its upper value and stays there.

When $I_{TH3} < I_{Ca} < I_{TH2}$ only downward jumps are allowed and $X(t)$ is driven towards its lower bound. When $I_{Ca} > I_{TH2}$ $X(t)$ jumps are forbidden. In Fig. 5 we report the voltages V_{TH2} and V_{TH3} applied to the gates of the p-MOSFETs which control I_{TH2} and I_{TH3}.

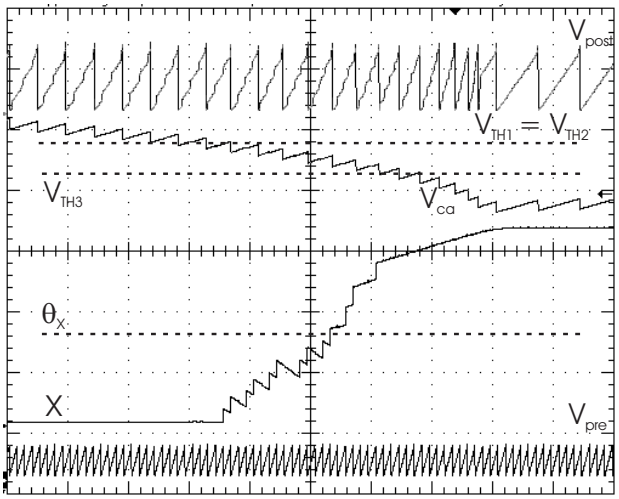

Figure 5. Illustrative example of the 'stop-learning' mechanism (see text). Top to bottom: the post-synaptic neuron potential, the voltage V_{Ca} controlling the calcium variable, the internal synaptic variable X, and the pre-synaptic neuron potential. The voltage values corresponding to the thresholds I_{TH1}, I_{TH2} and I_{TH3} are indicated by the horizontal lines, together with the threshold θ_X.

4. Synapse Configuration

The dashed regions in Fig. 2 highlight the main new features introduced with respect to the previous C-LANN chip [3]. A 4-bit bus ($b2$, $b1$, $b0$, nWR) is used to control the configuration and initialization of all the individual

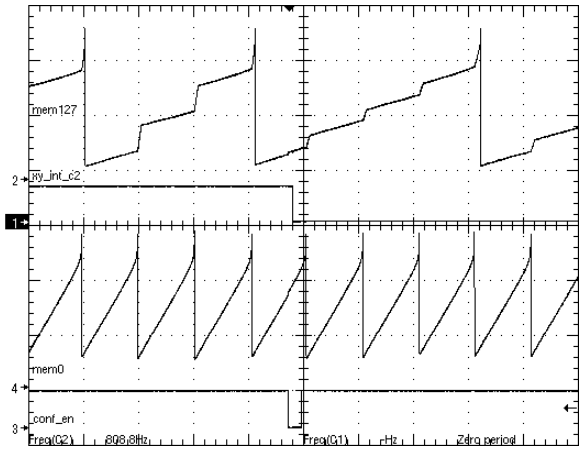

Figure 6. A potentiated synapse is set to the depressed state. Top to bottom: $V_{\rm post}$, X, $V_{\rm pre}$, configuration digital signal.

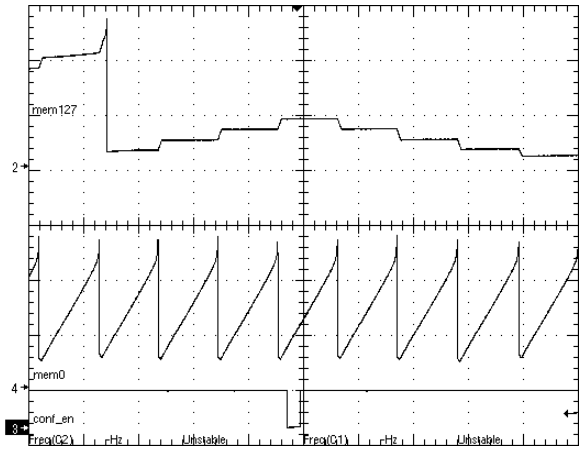

Figure 7. An excitatory synapse is set to be inhibitory. Top to bottom: $V_{\rm post}$, $V_{\rm pre}$, configuration digital signal.

synapses. The selection of the synapses is done with the help of the row-column selection lines X_{conf}, Y_{conf}.

The state of the selected synapse is available at the output of a tri-state buffer. The SISE (Synapse Initialization and Setting Element) is the digital control element which contains the memory (2 bits) for the configuration and reads the state of the synapse. The control signal nWR performs both the loading of $b0$ and $b1$ in the respective FF (Flip-Flop) and the initialization phase of the synapse. Bit $b0$ loaded in the first FF produces the Sel signal which configures the synapse as either recurrent or AER, through the

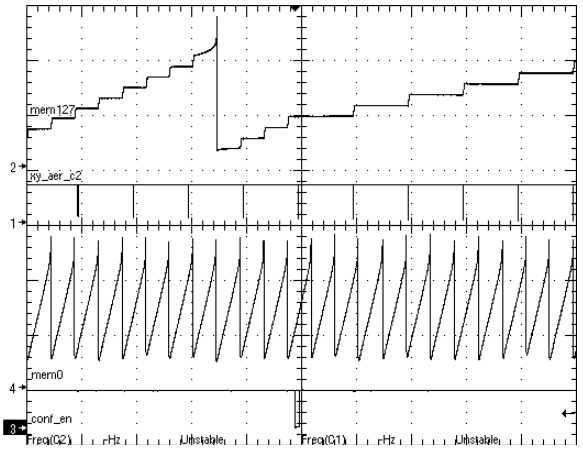

Figure 8. A recurrent synapse is set to be AER. Top to bottom: $V_{\rm post}$, $AER_{\rm pre}$, $V_{\rm pre}$, configuration digital signal.

MUX visible in the Pulse extender block in Fig. 2. Bit $b1$ loaded in the second FF produces the $Conf$ signal which sets the synapse as excitatory or inhibitory through the dendritic branch. Bit $b2$ is a global signal over the synaptic array, and it decides whether the selected synapse is forced toward a potentiated or depressed state when nWR is enabled through the initialization circuit. The pulse extender element regulates the duration of the spike, controlled by the voltage Pls and triggered by the incoming spike, either AER or recurrent.

It is possible to continuously monitor the synaptic state without affecting the chip dynamics.

Decoders are used to access the synapses to configure them as excitatory or inhibitory, and recursive or AER. Other decoders are also used when addressing the synapses in case of AER spiking activity. 7-to-128 bit decoders were implemented to address the 128×128 synaptic matrix, using standard cells from *austriamicrosystems* (AMS) and automatic place-and-route tools supplied by CADENCE. These cells should lower noise and reduce ground bounce and voltage drops.

In Fig. 6, Fig. 7, and Fig. 8 we illustrate the relevant aspects of synaptic configuration. Fig. 6 shows the post-synaptic manifestation of a potentiated synapse being set as depressed (larger to smaller jumps induced in the post-synaptic potential).

Fig. 7 shows the effect of changing the synapse from being excitatory to inhibitory (upward to downward jumps in the post-synaptic potential).

Fig. 8 shows a recurrent synapse being set as AER (post-synaptic jumps are first locked to the recurrently transmitted spikes, then become locked to the AER spikes).

5. Conclusions

We report here an analog VLSI chip (termed F-LANN) for a neuromorphic network for associative learning. Neuromorphic neurons and synapses feature adaptive and self-regulating properties designed for the associative learning of complex and partly correlated patterns. Although the F-LANN incorporates 128 neurons and 16 384 synapses, significantly greater numbers of neurons and synapses will be needed for associative learning with natural stimulus sets. An attractive route to larger networks is to link multiple VLSI chips via an AER-based communication infrastructure. For this reason, the F-LANN implements an AER-compliant chip design in which each neuron features an AER segment on its dendritic tree, which stands ready to accept spikes from external sources. The external source may either be another VLSI chip or a software simulation. To achieve maximal flexibility in setting a connection architecture, each synapse can be individually configured to be either recurrent or AER-based, either excitatory or inhibitory, and of either high or low initial efficacy. In addition, selected synapses may be read and set without impeding spike traffic on the AER bus. In summary, the F-LANN represents a critical step toward flexible multi-chip systems that perform associative learning of natural stimulus sets with biologically plausible components.

6. Acknowledgment

This work was supported by the ALAVLSI (IST2001-38099) EU grant.

References

[1] D.J. Amit and S. Fusi. *Neural Computation*, 6:957, 1994.

[2] D.J. Amit and G. Mongillo. *Neural Computation*, 15:565, 2003.

[3] D. Badoni, M. Giulioni, V. Dante, and P. Del Giudice. In *Proc. IEEE International Symposium on Circuits and Systems ISCAS06*, 2006.

[4] C. Bartolozzi and G. Indiveri. In *Proc. Brain Inspired Cognitive Systems*, 2006.

[5] C. Bartolozzi and G. Indiveri. Synaptic dynamics in analog vlsi. *Neural Computation (in press)*, 2006.

[6] J.M. Brader, W. Senn, and S. Fusi. *Neural Computation (in press)*, 2007.

[7] E. Chicca, V. Dante, A.M. Whatley, P. Lichtsteiner, T. Delbruck, G. Indiveri, P. Del Giudice, and R. J. Douglas. *IEEE Transactions on Circuits and Systems I*, 54(5):981–993, 2007.

[8] V. Dante, P. Del Giudice, and A. M. Whatley. The neuromorphic engineer newsletter. 2005.

[9] E. Chicca et al. *IEEE Transactions on Neural Networks*, 14(5):1297, 2003.

[10] S. Fusi. *Biological Cybernetics*, 87:459, 2002.

[11] S. Fusi, M. Annunziato, D. Badoni, A. Salamon, and D.J. Amit. *Neural Computation*, 12:2227, 2000.

[12] S. Fusi and M. Mattia. *Neural Computation*, 11:633, 1999.

[13] P. Del Giudice, S. Fusi, and M. Mattia. *Journal of Physiology Paris*, 97:659, 2003.

[14] G. Indiveri. *Analog Integrated Circuits and Signal Processing*, 28(3):279, 2001.

[15] G. Indiveri. In *Proc. IEEE International Symposium on Circuits and Systems*, 2003.

[16] G. Indiveri, E. Chicca, and R. Douglas. *IEEE Transactions on Neural Networks*, 2005.

[17] J. Lazzaro, S. Ryckebusch, M.A. Mahowald, and C.A. Mead. *Advances in neural information processing systems*, 2:703, 1989.

[18] C. Mead. *Analog VLSI and neural systems*. Addison-Wesley, 1989.

[19] S. Mitra, S. Fusi, and G. Indiveri. In *Proc. IEEE International Symposium on Circuits and Systems ISCAS06*, pages 2777–2780, 2006.

[20] W. Senn and S. Fusi. *Neural Computation*, 17:2106, 2005.