An aVLSI Recurrent Network of Spiking Neurons with Reconfigurable and Plastic Synapses

Davide Badoni
and Massimiliano Giulioni
University Tor Vergata and INFN RM2
Roma - Italy
Email: {davide.badoni,massimiliano.giulioni}@roma2.infn.it

Vittorio Dante
and Paolo Del Giudice
Physics Lab. - ISS and Gr. Coll. Sanitá - INFN
Roma - Italy
Email: {vittorio.dante,paolo.delgiudice}@iss.infn.it

***Abstract*— We illustrate key features of an analog, VLSI (aVLSI) chip implementing a network composed of 32 integrate-and-fire (IF) neurons with firing rate adaptation (AHP current), endowed with both a recurrent synaptic connectivity and AER-based connectivity with external, AER-compliant devices. Synaptic connectivity can be reconfigured at will as for the presence/absence of each synaptic contact and the excitatory/inhibitory nature of each synapse. Excitatory synapses are plastic through a spike-driven stochastic, Hebbian mechanism, and possess a self-limiting mechanism aiming at an optimal use of synaptic resources for Hebbian learning.**

I. INTRODUCTION

Following the pioneering work of C. Mead [1] it has become customary to term 'neuromorphic devices' a growing family of analog, VLSI, sub-threshold circuits providing literal implementations of theoretical neural and synaptic models. Neuromorphic engineering aims long-term at providing autonomous, power-parsimonious, real-time, adapting devices for perception and information processing in a natural environment. Far-reaching as it is, this undertaking must encompass from the beginning a form of plasticity ("learning") dependent on the 'interaction with the environment' – external stimuli. Keeping up with the theoretical progress in theoretical neuroscience, learning is assumed to be subserved by plasticity in the synapses, and a generally adopted framework goes under the name of Hebbian learning, by which the *efficacy* of a synapse is potentiated (the post-synaptic effect of a spike is enhanced) if the pre- and post-synaptic neurons are simultaneously active on a suitable time scale. Whether or not the above condition is fully specified by the average firing rates of the two neurons, or a detailed phase relation of fired spikes is needed, has been and is still debated, and reflected in different aVLSI implementations of the synaptic elements in neuromorphic chips. The synaptic circuits described in what follows implement rate-based Hebbian learning, even though they are also compatible with most features of 'Spike-Timing-Dependent-Plasticity'.

In the last decade, it has been realized that general constraints plausibly met by any concrete implementation of a synaptic device in a neural network, bear profound consequences on the capacity of the network as a memory system. Specifically, once one accepts that a synaptic element can neither have an unlimited dynamic range (i.e. synaptic efficacy is bounded), nor can it undergo arbitrarily small changes (i.e. synaptic efficacy has a finite analog depth), it has been proven ([7], [8]) that a deterministic learning prescription implies an extremely low memory capacity, and a severe 'palimpsest' property: new memories quickly erase the trace of older ones. It turns out that a stochastic mechanism provides a general, logically appealing and very efficient solution: given the pre- and post-synaptic neural activities, the synapse is still made eligible for changing its efficacy according to a Hebbian prescription, but it actually changes its state with a given probability. The stochastic element of the learning dynamics would imply *ad hoc* new elements, were it not for the fact that for a spike-driven implementation of the synapse, the noisy activity of the neurons in the network can provide the needed 'noise generator' [6]. Therefore, for an efficient learning electronic network, the implementation of the neuron as a spiking element is not only a requirement of 'biological plausibility', but a compelling computational requirement. Learning in networks of spiking IF neurons with stochastic plastic synapses has been studied theoretically [8], [13], [9], and stochastic, bi-stable synaptic models have been implemented in silicon [6], [2]. One of the limitations so far, both at the theoretical and the implementation level, has been the artificially simple statistics of the stimuli to be learnt (e.g., no overlap between their neural representations). Very recently in [3] a modification of the above stochastic, bi-stable synaptic model has been proposed, endowed with a regulatory mechanism such that the chances the synapse has to be up- or down-regulated depend on the average activity of the post-synaptic neuron in the recent past; the basic idea is that a synapse pointing to a neuron that is found to be highly active, or poorly active, should not be further potentiated or depressed, respectively. The reason behind the prescription is essentially that for correlated patterns to be learnt by the network, a successful strategy should de-emphasize the coherent synaptic Hebbian potentiation that would result for the overlapping part of the synaptic matrix, and that would ultimately spoil the ability to distinguish the patterns. A detailed learning strategy along this line was proven in [4] to be appropriate for linearly separable patterns for a Perceptron-like network; the extension to spiking and recurrent networks is currently studied.

In what follows we report the design, the implementation

0-7803-9390-2/06/$20.00 ©2006 IEEE

and first experiments for a chip composed of IF neurons and self-regulating plastic synapses; furthermore, for this chip it is possible at any time to set individually each synapse, i.e. to decide if two neurons are synaptically connected or not and, if yes, the excitatory or inhibitory nature of the synapse.

Another contribution to this same conference ([15]) reports an alternative implementation of a similar model for the single synapse; the two parallel, and coordinated, developments are carried out as part of a joint EU project.

II. CHIP ARCHITECTURE AND MAIN FEATURES

We describe a chip implementing a recurrent network of 32 integrate-and-fire neurons with spike-frequency adaptation and bi-stable, stochastic, Hebbian synapses(see Fig.1). A completely reconfigurable synaptic matrix supports up to all-to-all recurrent connectivity, and AER-based external connectivity. Besides establishing an arbitrary synaptic connectivity, the excitatory/inhibitory nature of each synapse can also be set.

The implemented neuron is the IF neuron with constant leakage term and a lower bound for the 'membrane potential' $V(t)$ introduced in [1] and studied theoretically in [5]. The circuit is borrowed from the low-power design described in [14], [11], to which we refer the reader for details. Only 2 neurons can be directly probed (i.e., their 'membrane potential' sampled), while for all of them the emitted spikes can be monitored via AER.

The dendritic tree of each neuron is composed of up to 31 activated recurrent synapses and up to 32 activated external, AER ones. For the recurrent synapses, each impinging spike triggers short-time (and possibly long-term) changes in the state of the synapse, as detailed in Section IV. Spikes from neurons outside the chip come in the form of AER events, and are targeted to the correct AER synapse by the X-Y Decoder. AER synapses which are set to be excitatory are plastic as the recurrent ones (inhibitory synapses are fixed). Spikes generated by the neurons in the chip are arbitered for access to the AER bus for monitoring and/or mapping to external targets.

III. CHIP MISMATCHES *vs* MONTECARLO SIMULATION

It is well known that mismatches due to the fabrication process introduce variability in the actual sub-threshold working of circuits with identical schematics. When it comes to chips hosting big networks, it is important to have predictive tools allowing to assess the expected level and type of variability in the fabricated chip. Furthermore, one needs to identify, out of the plethora of conceivable tests of this kind, those which are informative of critical aspects of the network's dynamics. For instance, neurons receiving nominally identical inputs will fire at different rates; such firing rates distribution would affect both the match between the chip's dynamics and theoretical predictions, for fixed synaptic efficacies, and the synaptic dynamics.

We performed a Montecarlo analysis of the effect of mismatches in the neurons implementation, and check against our measurements. Neurons on chip were driven by the same DC current, all the synapses were configured to have zero efficacy, and the distribution of inter-spike intervals (ISI) across neurons was sampled (Fig.2, left panel). The Montecarlo prediction is shown in Fig.2, right panel.

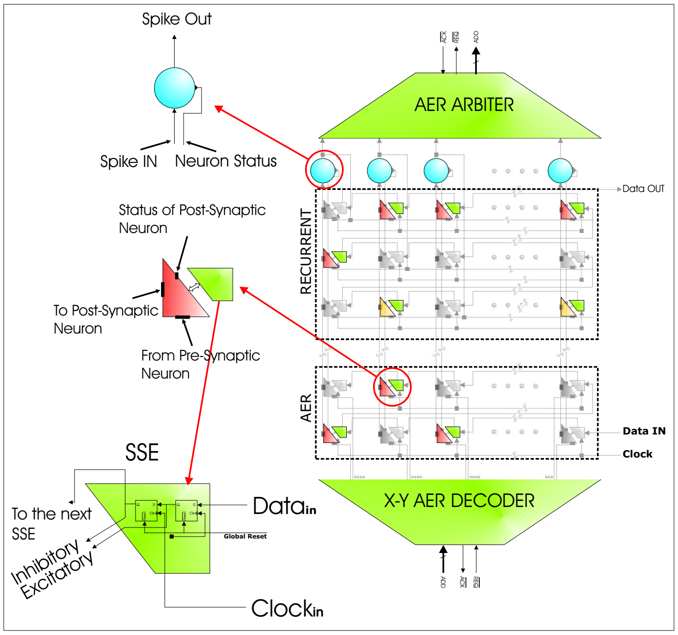

Fig. 1. Scheme of the configurable synaptic connectivity.

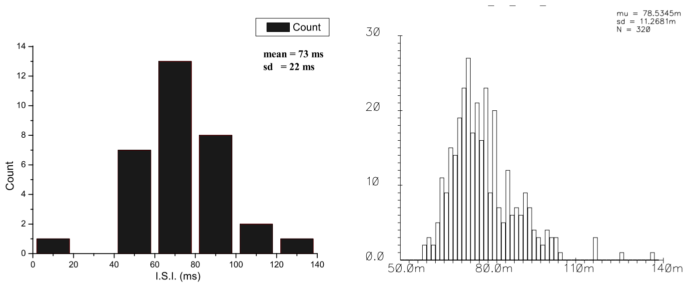

Fig. 2. Measurements *vs* Montecarlo. Left panel: Measured ISI distribution. Right panel: Simulated ISI distribution from Montecarlo.

The reported, preliminary, test supports the predictive value of Montecarlo simulations; the information derived on the firing rates distribution will be used to *a priori* tune the parameters of the synaptic dynamics in order to achieve the desired speed of Hebbian, rate-driven learning, and assess the feasibility of different learning scenarios in bigger chips.

IV. A SYNAPSE THAT KNOWS WHEN TO STOP LEARNING.

Fig.3 illustrates the synaptic circuit that implements the model proposed in [3] and briefly motivated in the Introduction. The instantaneous state of the synapse is given by the analog variable $X(t)$, subjected to short-term, spike-driven dynamics; a threshold mechanism acting on $X(t)$ determines the synaptic efficacy, which is preserved on long time scales in the absence of intervening spikes; the synapse possesses only two states of efficacy (a bi-stable device).

Upon the arrival of a pre-synaptic spike, which acts as a trigger for synaptic changes, X is candidate for an upward or downward jump, depending on the comparison between the instantaneous value of the post-synaptic potential and a threshold. As an additional regulatory element, a further dynamic variable is associated with the post-synaptic neuron, which essentially measures the average level of firing activity in the recent past. Following [3], because of an analogy with the role played by the intracellular concentration of calcium ions near the emission of a spike, we will call it a 'calcium variable' $C(t)$. $C(t)$ undergoes an upward jump when the post-synaptic neuron emits a spike, and linearly decays between two spikes. It therefore integrates the spikes sequence and, when compared to suitable thresholds as detailed below, it determines which candidate synaptic changes will be allowed to occur; for example, it can instruct the synapse to stop up-regulating because the post-synaptic neuron is already very active. In the absence of spikes X is forced to drift towards a 'high' or 'low' value depending on whether the last jump left it above or below another threshold, respectively, thereby determining the synaptic efficacy (details are found in [6]).

The Bistability sub-circuit (see Fig.3) 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 $V(t)$ determined by b_0/b_1. Fig.4, left, describes the circuit for the calcium variable. The topmost mosfet is gated by the post-synaptic spike, letting a current I_{ca} regulated by the mid mosfet, charge the capacitance C_{ca}; in the absence of post-synaptic spikes, C_{ca} discharges with a rate determined by the bottom mosfet. Fig.5 illustrates the design of one of the gated comparators in Fig.4: a two-stage open-loop comparator with two additional mosfets which, for digital En_1 and En_2 signals, act as switches. The comparator is enabled by (En_1 `NOR` En_2). The system of comparators for the calcium variable implements the following conditions: $X(t) \rightarrow X(t) + a$ if $V_p(t) > \theta_p$ and $K_1^{up} < C(t) < K_2^{up}$ $X(t) \rightarrow X(t) - b$ if $V_p(t) \leq \theta_p$ and $K_1^{down} < C(t) < K_2^{down}$. Otherwise $X(t)$ is linearly driven towards its upper (lower) bound.

Fig.6 illustrates the working of the calcium circuit: the bottom and top traces are the potential $V(t)$ of two neurons connected by the synapse whose $X(t)$ variable is shown in the second trace from below. The second trace from top is the calcium $C(t)$ variable, and for illustrative purposes only the K_2^{up} threshold is set. $X(t)$ undergoes up and down jumps according to the instantaneous value of the post-synaptic $V(t)$, as long as $C(t) < K_2^{up}$; when $C(t) > K_2^{up}$ synaptic jumps are prevented, and $X(t)$ drifts towards the appropriate stable value.

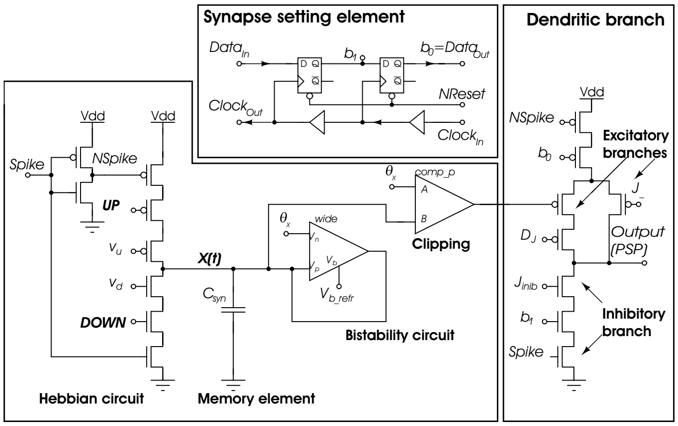

Fig. 3. Synaptic circuit.

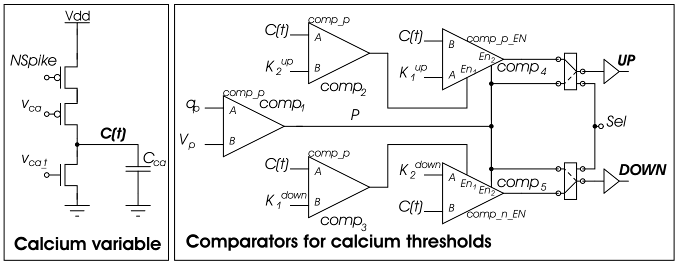

Fig. 4. Circuits for the calcium variable.

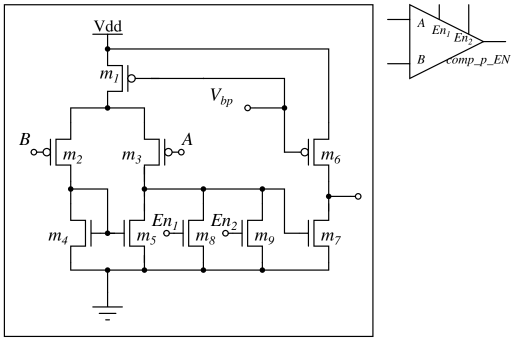

Fig. 5. p Comparator circuit

V. ON-CHIP SYNAPTIC MATRIX RECONFIGURATION.

The Synapse Setting Element (SSE) in Fig.3 allows to 1) set each possible synaptic contact as active or non-active, 2) for an activated synapse, selectively activate one of the dendritic branches, making the synapse excitatory or inhibitory. The configuration of the 2048 synapses (1024 recurrent, 1024 AER) is encoded in a 4096 bit word which is serially fed into the 2048 SSE, each hosting two flip-flop, which globally

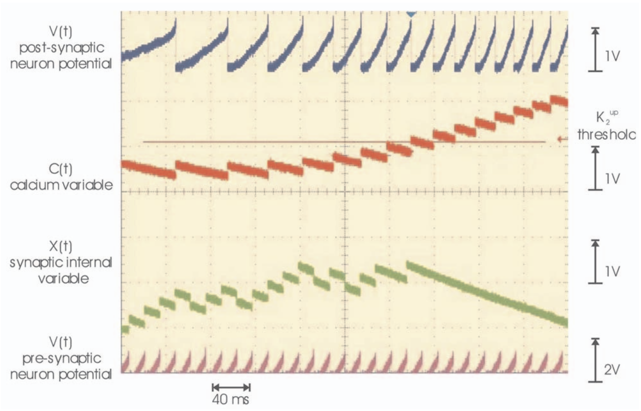

Fig. 6. Traces illustrating the working of the calcium variable circuit. $C(t) > K_2^{up}$ prevents synaptic jumps.

constitute a 4096-bit shift register. A complete configuration of the synaptic connectivity requires 4096 clock cycles (with an approximate upper bound of 10 MHz for the clock frequency). Two digital signals are input into each SSE (see Fig.3).

The CLOCK$_{in}$ global signal synchronizes the DATA$_{in}$ signal which is propagated in pipeline through the sequence of SSE: at each rising edge of CLOCK$_{in}$, the content b_1 is transferred into b_0 and b_0 is transferred into the next SSE. For each SSE b_0 and b_1 activate the excitatory and inhibitory branch, respectively. In Fig.1, highlighted synapses are set as active. The main blocks active in the configuration cycle are detailed.

In Fig.7 we show sample traces illustrating the synaptic configuration in action. Two synaptically connected neurons are driven by DC current: when the pre-synaptic neuron (second trace from top, high rate) fires, the post-synaptic neuron (topmost trace, low rate) undergoes an upward or downward jump in the potential $V(t)$, depending on the excitatory or inhibitory character of the synapse. The CLOCK signal (bottom trace), starts the synaptic configuration cycle. Pairs of successive bits along the DATA stream (second trace from below) define the state of a SSE. In the case shown, the synapse is excitatory before starting the reconfiguration cycle which turns it into an inhibitory one.

VI. CONCLUSIONS

We have shown the working of the building blocks of a recurrent network hosting a self-regulating synapse designed to allow the learning of correlated patterns. One of the innovative features of the chip is that it offers the possibility of reconfiguring the synaptic connectivity and the excitatory/inhibitory components of the synaptic matrix (even ‘on the fly’ with In-Site Programming, if one can afford the transient perturbation of the network’s dynamics during the configuration cycle).

The number of neurons is small, and the on-chip network is unlikely to support interesting dynamics. However, the chip is AER compliant, and each neuron has an AER segment of its dendritic tree, ready to transmit spikes coming from outside the chip (either from other neuromorphic chips, or from a simulated source); excitatory AER synapses are plastic as the recurrent ones.

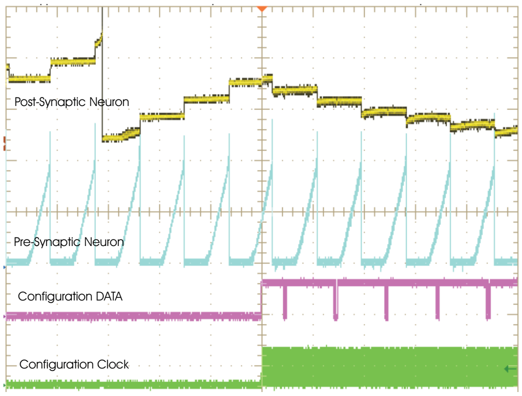

Fig. 7. Traces illustrating the working of the synapse configuration circuit. The configuration cycle turns the excitatory synapse into an inhibitory one.

To increase the number of neurons and reach interesting computational capabilities, one has either to design and fabricate much bigger chips, or put several chips together. The latter alternative, besides lowering the needed investment, is more appealing because of modularity and scalability. We recently made critical steps towards a flexible AER-based infrastructure supporting multiple chips, in the form of a ‘PCI-AER’ board [10]; we are presently testing an extension of that approach, implementing an autonomous, AER-based multi-chip systems.

ACKNOWLEDGMENT

We acknowledge the EU support through the IST2001-38099 grant.

REFERENCES

[1] C. Mead, Analog VLSI and neural systems, Addison-Wesley 1989
[2] E. Chicca et al., IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 14, NO. 5, pag. 1297, 2003
[3] J. Brader, W. Senn and S. Fusi, submitted to Neural Computation (2005)
[4] W. Senn and S. Fusi, Neural Computation 17, 2106, 2005
[5] S. Fusi and M. Mattia Neural Computation 11, 633, 1999
[6] S. Fusi, M. Annunziato, D. Badoni A. Salamon and D.J. Amit, Neural Computation 12, 2227, 2000
[7] D.J. Amit and S. Fusi, Neural Computation 6, 957, 1994
[8] S. Fusi, Biol. Cybern. 87, 459, 2002.
[9] D.J. Amit and G. Mongillo, Neural Computation 15, 565, 2003
[10] V. Dante, P. Del Giudice and A. M. Whatley, The Neuromorphic Engineer Newsletter, 2005. Available from `http://ine-web.org/fileadmin/templates/_docs/NME3.pdf`
[11] G. Indiveri, E. Chicca, and R. Douglas, IEEE Trans. on Neural Networks, 2005, (in press).
[12] S. R. Deiss, R. J. Douglas, and A. M. Whatley, in Pulsed Neural Networks, W. Maass and C. M. Bishop, Eds. MIT Press, 1998, ch. 6, pp. 157178.
[13] P. Del Giudice, S. Fusi and M. Mattia, Journal of Physiology Paris, 97, 659, 2003
[14] G. Indiveri, In Proc. IEEE International Symposium on Circuits and Systems, 2003.
[15] S Mitra, S. Fusi and G. Indiveri, submitted to IEEE International Symposium on Circuits and Systems, 2006.