LETTER

Synaptic Plasticity and Spike-based Computation in VLSI Networks of Integrate-and-Fire Neurons

Giacomo Indiveri

Institute for Neuroinformatics
UZH | ETH Zürich, Winterthurerstr. 190, Zurich, Switzerland
E-mail: giacomo@ini.phys.ethz.ch

(Submitted on December 27, 2006)

Abstract – Neuromorphic circuits are being used to develop a new generation of computing technologies based on the organizing principles of the biological nervous system. Within this context, we present neuromorphic circuits for implementing massively parallel VLSI networks of integrate-and-fire neurons with adaptation and spike-based plasticity mechanisms. We describe both analog continuous time and digital asynchronous event-based circuits for constructing spiking neural network devices, and present a VLSI implementation of a spike-based learning mechanisms for carrying out robust classification of spatio-temporal patterns, and real–time sensory signal processing. We argue that these types of devices have great potential for exploiting future scaled VLSI processes and are ideal for implementing sensory-motor processing units on autonomous and humanoid robots.

Keywords – Spike-based learning, Neuromorphic engineering, Integrate-and-fire neuron

1. Introduction

Neuromorphic circuits are a class of hybrid analog/digital circuits that implement hardware models of biological systems. Neuromorphic systems carry out sensory-signal processing tasks using computational principles analogous to the ones used by nervous systems. The styles of computation used by nervous systems are fundamentally different from those used by conventional computers: biological neural networks process information using energy-efficient asynchronous, event-driven, methods. They are adaptive, fault-tolerant, self-repairing, learn from their interactions with the environment, and can flexibly compose complex behaviors by combining multiple instances of simpler elements. These biological abilities offer an attractive alternative to conventional computing strategies. When implemented in Very Large Scale Integrated (VLSI) technology, neuromorphic systems share to a large extent the same physical constraints of their biological counterparts. Therefore they often have to use similar strategies for maximizing compactness, optimizing robustness to noise, minimizing power consumption, and increasing fault tolerance. By emulating the neural style of computation, neuromorphic VLSI architectures can exploit to the fullest potential the features of advanced scaled VLSI processes and future emerging technologies, naturally coping with the problems that characterize them, such as device inhomogeneities, and imperfections.

The greatest successes of neuromorphic systems to date have been in the emulation of peripheral sensory transduction: single chip devices, such as silicon retinas, or silicon cochleas have been successfully implemented and used in a wide variety of applications [25, 32]. In recent years a new class of neuromorphic *multi-chip* systems started to emerge [26, 10, 9, 30]. These systems typically comprise one or more neuromorphic sensors, interfaced to chips that implement general-purpose computational architectures based on networks of silicon neurons and synapses. Consistent with the neuromorphic engineering approach, the strategy used to transmit signals across chip boundaries in these types of systems is inspired from the nervous system: output signals are represented by stereotyped digital pulses (spikes), and the analog nature of the signal is typically encoded in the mean frequency of the neuron's pulse sequence (spike rates). Similarly, input signals are represented by spike trains, conveyed to the chip in the form of asynchronous digital pulses, that stimulate their target synapses on the receiving chip. The

circuits that generate and manage these streams of input/output digital pulses are fast asynchronous logic elements. Conversely, the circuits that generate the on-chip synaptic currents when stimulated by incoming spikes are slow low-power analog circuits. Typically these types of analog circuits operate in the *weak-inversion* or *subthreshold* regime [22], where current amplitudes are of the order of picoamperes and operating time-constants are of the order of milliseconds. These biologically plausible time constants are required so that the circuit response is well-matched to the sensory signals they are designed to process, and the neuromorphic system is inherently synchronized with the real world events. Between two consecutive spikes generated by a "slow" analog neuron (typically firing at rates of a few Hertz) digital pulses generated by millions of other neurons can be multiplexed on a fast digital bus, and conveyed to receiver chips. Temporal integration of input spikes inside the receiver chip, at the level of the synapses, and the spatio-temporal integration of weighted synaptic currents at the level of the neurons take place in a massively parallel fashion, and the integration time does not scale with the network size.

As VLSI networks of spiking neurons are being used more and more often in multi-chip neuromorphic systems, the development of spike-based learning circuits compatible with these systems is extremely important. Spike-based learning circuits enable these multi-chip systems 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.

In this article we present a specific instance of a generic reconfigurable neural architecture that comprises an array of spiking neurons, silicon synapses and spike-based learning circuits which implement a novel synaptic plasticity algorithm recently proposed in [8]. In the following sections we describe the asynchronous infrastructure used to construct large-scale multi-chip systems, the circuits that implement the device's neurons and synapses, the VLSI spike-based plasticity mechanism, and experimental results demonstrating how the system fulfills the learning model's requirements for classifying complex patterns of mean firing rates. In the conclusions we point out the advantages of this approach and propose possible applications for these types of devices.

2. Multi-chip Systems and the Address-Event Representation

A wide range of *spike-based* neuromorphic devices has been developed in recent years. These range from sensory devices such as spiking silicon retinas and spiking silicon cochleas [21, 10, 31], to reconfigurable arrays of integrate and fire neurons [26, 18, 9], and learning chips implementing detailed models of spike-based synaptic plasticity [18, 3, 2, 29, 7]. In parallel with the development of pulse-based VLSI devices, there have been significant advancements in the development of asynchronous event-based communication infrastructures. These led to the construction of an impressive set of pulse-based multi-chip systems, such as vision-based systems that emulate the orientation selectivity functions of the visual cortex [9, 26, 10], or large-scale systems that can perform convolution, segmentation and object tracking in natural scenes [30].

The asynchronous communication protocol used in these types of multi-chip systems is based on the *Address-Event Representation* (AER) [20, 5]. In AER, each neuron on a sending device is assigned an address. When the neuron produces a spike its address is instantaneously put on an asynchronous digital bus (see Figure 2). In the case of single-sender/single-receiver communication, a simple handshaking mechanism ensures that all events generated at the sender side arrive at the receiver side. Event 'collisions' (cases in which sending nodes attempt to transmit their addresses at exactly the same time) are managed by on–chip arbitration schemes. Systems containing more than two AER chips can be constructed by implementing special purpose off-chip arbitration schemes.

These multiplexing strategies are very efficient because only the addresses of active elements are transmitted (as opposed to conventional scanning techniques, that allocate the same bandwidth for all the pixels, independent of their activity). The source address-events being transmitted on the digital bus can be translated, converted or remapped to multiple destinations using conventional logic and memory elements. AER infrastructures therefore allow us to construct large multi-chip networks with arbitrary connectivity, and to seamlessly reconfigure the network topology. In this way we exploit at best both the programmable and noise-insensitive properties of digital logic when transmitting signals across chips, and the size and power-efficient properties of the parallel analog neuromorphic circuits in the core of the chips.

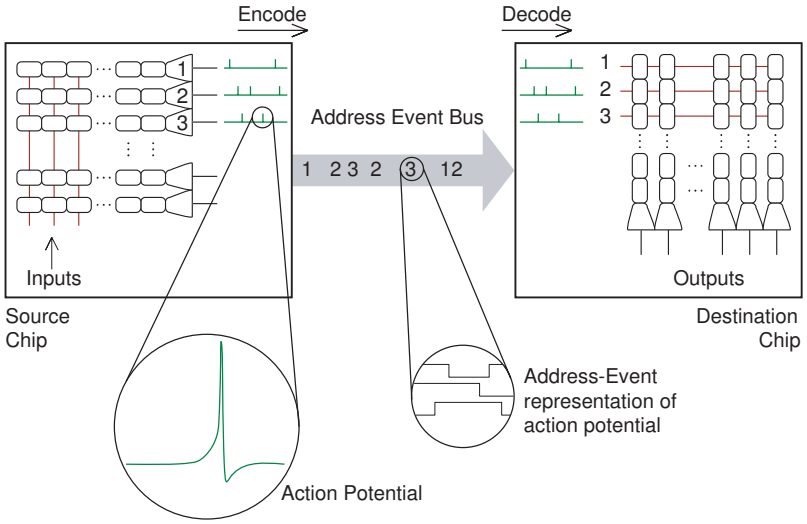


Figure 1. Asynchronous communication scheme between two chips using the Address-Event Representation (AER). When a neuron on the source chip generates an action potential, its address is placed on a common digital bus. The receiving chip decodes the address events and routes them to the appropriate synapses.

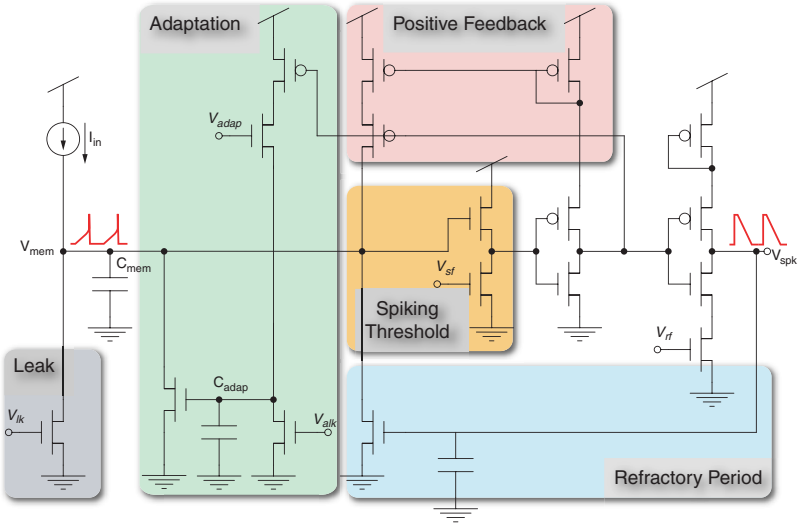


Figure 2. Schematic diagram of an integrate-and-fire neuron. The input current I_{in} is integrated onto the neuron's membrane capacitor C_{mem} until the spiking threshold is reached. At that point the output signal V_{spk} goes from zero to the power supply rail, signaling the occurrence of a spike, and the membrane capacitor is reset to zero. The "leak" module implements a current leak on the membrane. The "spiking threshold" module controls the voltage at which the neuron spikes. The "adaptation" module subtracts a firing rate dependent current from the input node. The amplitude of this current increases with each output spike and decreases exponentially with time. The "refractory period" module sets a maximum firing rate for the neuron. The "positive feedback" module is activated when the neuron begins to spike, and is used to reduce the transition period in which the inverters switch polarity, dramatically reducing power consumption. The circuit's biases (V_{lk}, V_{adap}, V_{alk}, V_{sf}, and V_{rf}) are all subthreshold voltages that determine the neuron's properties.

3. A Low-power Silicon Neuron Circuit

A spiking neuron model that allows us to implement large, massively parallel networks of neurons is the Integrate-and-Fire (I&F) model. I&F neurons integrate pre-synaptic input currents and generate a voltage pulse analogous to an action potential when the integrated voltage reaches a spiking threshold. Networks of I&F neurons

have been shown to exhibit a wide range of useful computational properties, including feature binding, segmentation, pattern recognition, onset detection, input prediction, *etc.* [23]. Many variants of these circuits had been built during the 50's and 60's using discrete electronic components. The first simple VLSI version was probably the *Axon-Hillock* circuit, built by Mead in the late eighties [28]. In this circuit, a capacitor that represents the neuron's membrane capacitance integrates current input to the neuron. When the capacitor potential crosses the spiking threshold a pulse is generated and the membrane potential is reset. This circuit captures the basic principle of operation of biological neurons, but cannot faithfully reproduce all of the dynamic behaviors observed in real neurons.

More elaborate models of neurons, that take into account the biophysical properties of the voltage dependent conductances and currents present in real neurons have also been proposed. The first and most influential conductance-based silicon neuron is perhaps that of Douglas and Mahowald [24]. This silicon neuron is composed of connected compartments, each of which is populated by modular sub-circuits that emulate particular ionic conductances. This circuit can reproduce in great detail many of the behaviors observed in real neurons, but their overall size and circuit complexity is significantly larger than the one of the simpler Axon-Hillock circuit.

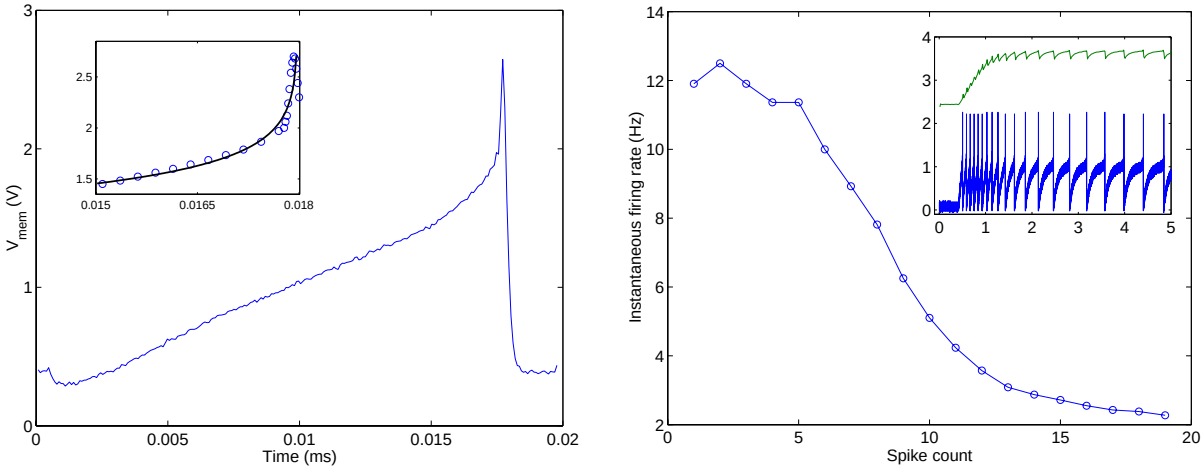
A compromise between the above two approaches is provided by a more elaborate I&F neuron circuit that implements additional neural characteristics, such as spike-frequency adaptation properties and refractory period mechanisms. An example of such a circuit is shown in Figure 3. In addition to implementing the basic behavior of integrating input currents and producing output pulses at a rate that is proportional to the amplitude of its input, this low-power I&F neuron [18] implements a *leak* mechanism (as in leaky I&F neuron models); an adjustable *spiking threshold* mechanism for adapting or modulating the neuron's spiking threshold; a *refractory period* mechanism for limiting the maximum possible firing rate of the neuron; and a spike-frequency *adaptation* mechanism, for modeling some of the adaptation mechanisms observed in real neurons. The circuit's low-power dissipation properties derive from the positive feedback block, that drastically reduces the switching time of the neuron's first inverter.

Figure 3 shows experimental measurements from one of these silicon neurons, in response to a constant current input. Figure 3(a) shows a single spike, where the effect of positive feedback is evidenced in the figure's inset. The data in the figure inset is fitted with an equation analytically derived from the circuit schematics, and fully characterized in [17]. Figure 3(b) shows the effect of the spike-frequency adaptation block, when the neuron is stimulated with a constant current step. After an initial transient, the mean firing rate of the neuron decreases to a low steady-state value. This mechanism ensures that the neuron responds mainly to changes in the input, and minimizes the bandwidth requirements for data transmission.

4. A Compact Silicon Synapse Circuit

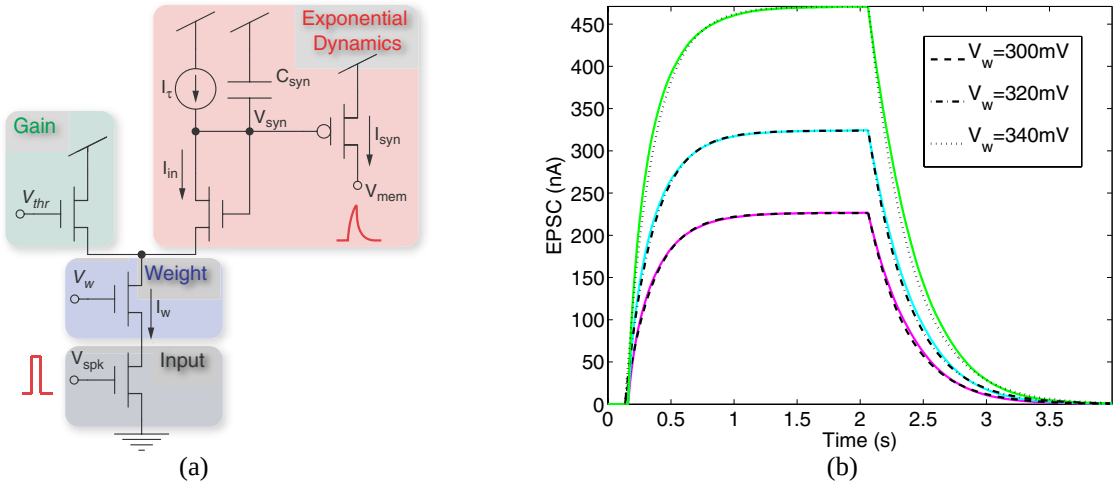
Synapses are highly specialized structures that represent the input terminals of a neuron, and are responsible for converting sequences of digital spikes into analog currents. When a spike generated by a source neuron reaches a pre-synaptic terminal, a cascade of events leads to the production of ionic currents that flow into or out of the post-synaptic neuron's membrane. These excitatory or inhibitory post-synaptic currents (EPSC or IPSC respectively) have temporal dynamics with a characteristic time course that can last up to several hundreds of milliseconds [12]. In neuromorphic chips, the detailed dynamics of post-synaptic currents can be readily emulated using dedicated subthreshold analog circuits. An example of an excitatory synapse circuit is shown in Figure 4. This circuit, dubbed the "Diff-Pair Integrator" (DPI), implements a log-domain temporal filter that reproduces the temporal properties of real synapses and accounts for the linear summation property of post-synaptic currents [4]. The DPI has independent control of time constant, synaptic weight, and synaptic scaling parameters; moreover it supports a wide range of synaptic properties, ranging from short-term depression to conductance-based EPSC generation, to synaptic plasticity. The time course and duration of the synaptic currents is set by the circuit's time constant parameter I_τ, while their amplitude can be modulated locally by changing the synaptic weight bias V_w, or globally by changing the common synaptic scaling bias V_{thr}.

Typically, in a neural-network chip, many synaptic circuits are connected to an I&F neuron (*e.g.* see Figure 5(a)). The output currents of all the afferent synapses are summed in parallel onto the neuron's membrane capacitor node, following Kirchhoff's current law. If multiple neurons are present on the chip, as is usually the case (*e.g.* see Figure 2), they also integrate in parallel the net input currents, and produce spike trains proportional to the weighted sum of their synaptic currents.



(a) (b)

Figure 3. (a) Membrane potential V_{mem} of the silicon neuron circuit in response to a constant input current. The inset shows the effect of the positive feedback block, when V_{mem} reaches the spiking threshold. The solid line represents a fit of the data derived analytically from the circuit equations. (b) Instantaneous firing rate as a function of spike count. This data shows the effect of the spike-frequency adaptation block, when the neuron is stimulated with an input current step. The inset shows how the individual spikes increase their inter-spike interval, with time.



(a) (b)

Figure 4. (a) The Diff-Pair Integrator circuit. The circuit's analog output current I_{syn} is an EPSC that encodes the frequency of the input spikes, represented as digital voltage pulses arriving at the V_{spk} node. The circuit's time constant is set by adjusting the value of the I_τ current. The V_{thr} bias in the "gain" module can be used to set the circuit's gain. Similarly the synaptic efficacy can be set with the V_w bias voltage in the "weight" module. Additional circuits can be connected to the V_w node to implement short and/or long term plasticity mechanisms, and to locally store or refresh the value of the weight. (b) Measured EPSC in response to an input step current, for different values of the synaptic weight. The shaded curves show the DPI response, and the superimposed dashed lines represent the fits of the data with the circuit's impulse response derived from the circuit equations [4].

5. Plasticity, Learning, and Classification

One of the key properties of biological synapses is their ability to exhibit different forms of *plasticity*. The type of plasticity that we consider here is "long-term" plasticity [1]. This mechanism produces long term changes in the synaptic strength of individual synapses in order to form memories or learn about the statistics of the input stimuli.

In neuromorphic VLSI chips, implementations of *long–term* plasticity mechanisms allow us to implement learning algorithms and set synaptic weights automatically, without requiring dedicated pins or wires for each individual synapse. Moreover, plasticity mechanisms based on the timing of the spikes (or Address-Events) map very effectively onto silicon neuromorphic devices [18, 6, 14, 16]. A popular class of spike-driven learning rules that has recently been the subject of renewed interest is the one based on the Spike Timing Dependent Plasticity (STDP) [27, 1]. In STDP the relative timing of pre- and post-synaptic spikes determine how to update the efficacy of a synapse. Several examples of STDP leaning chips have been fabricated [18, 2, 6], and a wide range of theoretical models proposed. It has been shown, both in theoretical models and VLSI implementations, that STDP can be effective in learning to classify spatio-temporal spike patterns [15, 2]. However, the STDP algorithm in its simplest form has the problem of not being suitable for learning different patterns of mean firing rates [1].

An additional problem that arises when considering physical implementations of synapses, either biological or electronic, has to do with the fact that the synaptic weights are bounded (they cannot grow indefinitely or assume negative values). This imposes strong constraints on the network's capacity to preserve memories, stored in the synaptic weights: in our VLSI implementation we have to make sure we don't over-write the stored synaptic values due to the storage of new memories. Furthermore we have to implement long-term storage of the synaptic weight voltage biases, making sure they don't decay with the passage of time. It has been shown that an efficient strategy for protecting previously stored memories is to use two stable synaptic efficacy states per synapse and very low average number of transitions from one stable state to the other [13]. By modifying only a random subset of the synapses with a small probability, memory lifetimes increase by a factor inversely proportional to the probability of synaptic modification [13]. As for spike sequences observed in cortical recordings *in vivo*, patterns obtained from AER sensors in response to real-world stimuli are noisy. We decided to use a bistable synapse circuit and exploit the noise in the input spike patterns to implement the stochastic update mechanism required to increase memory lifetimes. Using just two stable synaptic states solves efficiently also the problem of long-term storage: it is sufficient to use a bi-stable circuit that restores the synaptic state to either its high rail or its low one, depending if the weight is above or below a set threshold. In this way memory preservation is guaranteed also in the absence of stimuli, or when the pre-synaptic activity is very low.

Concerning the weight-update rule, rather than using a simple form of STDP we implemented the spike-triggered plasticity mechanisms described in [8], which depends on the post-synaptic membrane potential and on a slow "calcium" variable that represents the neuron's recent spiking activity. This model [8] has been shown to be able to classify patterns of mean firing rates, to capture the rich phenomenology observed in neurophysiological experiments on synaptic plasticity, and to reproduce the classical STDP phenomenology.

The circuits that implement this plasticity mechanism are shown in Figure 5. A spike-triggered weight-update module is implemented in every synapse (see Figure 5(b)), while a post-synaptic stop-learning control module (shown in Figure 5(c)) is implemented only at the neuron level.

5.1 Spike-triggered weight-update module

This module comprises four main blocks: an input AER interfacing circuit [5], a bi-stability weight refresh circuit, a spike-triggered weight update circuit and a DPI circuit, of the type described in Section 4. Upon the arrival of an input address-event, the AER circuits produce an active-high pulse pre, and a complementary active-low pulse $\sim pre$. These pulses trigger the weight update block. The pre pulse is also used to drive the plastic synapse's diff-pair integrator. 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 active, 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.

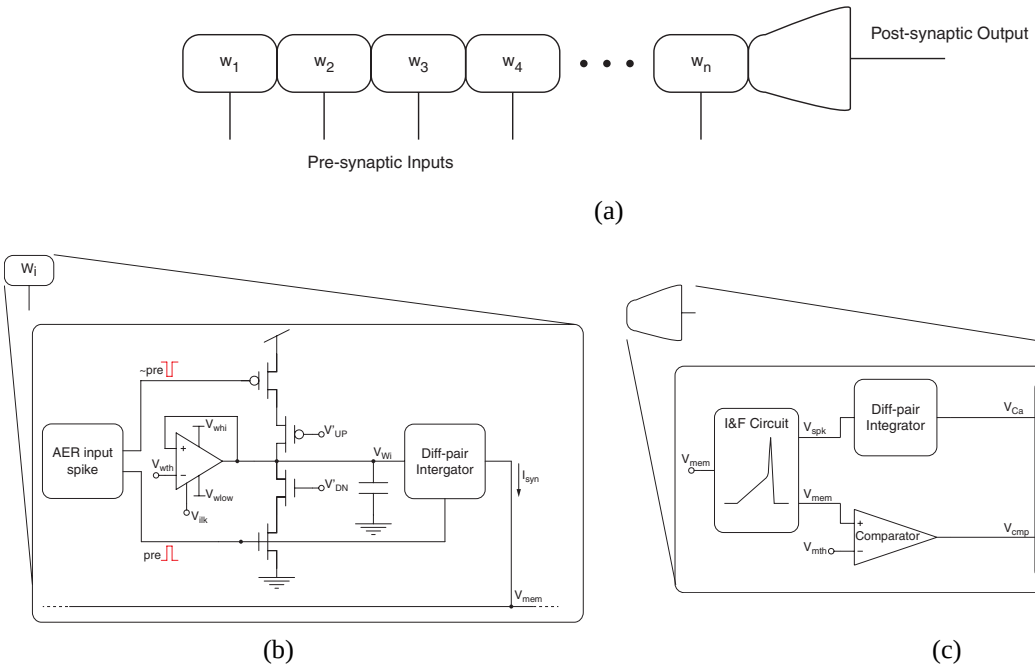


Figure 5. (a) Schematic diagram of a typical VLSI learning neuron: multiple instances of synaptic circuits source in parallel their output currents into the I&F neuron's membrane capacitance. The I&F neuron integrates the weighted sum of the currents and produces sequences of spikes in output. (b) Synapse with pre-synaptic weight update module. An AER asynchronous logic block receives input spikes and generates the pre and $\sim pre$ pulses. An amplifier in positive-feedback configuration slowly drives the weight voltage V_Wi toward one of the two stable states V_wlow or V_whi. The transistors driven by the pre and $\sim pre$ pulses, together with the ones controlled by the V'_UP and V'_DN signals implement the weight update. The diff-pair integrator block generates the output synaptic current I_syn that is sourced into the V_mem node of the I&F circuit. (c) Neuron with post-synaptic weight control module. An I&F neuron circuit, integrates the input synaptic currents and produces a spike train in output. A diff-pair integrator filter generates the V_Ca signal, encoding the neuron's mean firing rate. Voltage comparator and a current comparator circuits determine whether to update the synaptic weights of the afferent synapses, and whether to increase or decrease their value (see Section 5.2 for details).

5.2 Post-synaptic stop-learning control module

This module, shown in Figure 5(c), is responsible for generating the two global signals V_{UP} and V_{DN} that enable positive and negative weight updates respectively. The signals V'_{UP} and V'_{DN} are buffered copies of V_{UP} and V_{DN}, that are shared among all synapses belonging to the same dendritic tree. 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 represents the neuron's mean firing rate. Learning is enabled only if the neuron's mean firing rate is in an intermediate range. Indeed, if the rate is either very high or very low, this indicates that the current synaptic weights already allow the neuron to classify correctly the input pattern [8, 19], so the weight update should be switched off. This stop-learning condition is evaluated in our circuits by comparing V_{Ca} to three different thresholds (V_{th1}, V_{th2}, and V_{th3}) using winner-take-all circuits as current comparators [19]. 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 both on the state of the neuron's membrane potential V_{mem} and its Calcium concentration V_{Ca}. Specifically if $V_{th1} < V_{Ca} < V_{th3}$ and $V_{mem} > V_{mth}$, then the arrival of a pre-synaptic spike at a synapse produces an increases in its synaptic weight. And if $V_{th1} < V_{Ca} < V_{th2}$ and $V_{mem} < V_{mth}$, then the pre-synaptic spike produces a decreases in the synapse's weight. Otherwise no changes in the synaptic weights are allowed.

This stop-learning mechanism, together with the bi-stability circuit, the spike-driven weight update circuits, and the stochasticity in the AER spike trains allow us to faithfully implement the learning model proposed in [8].

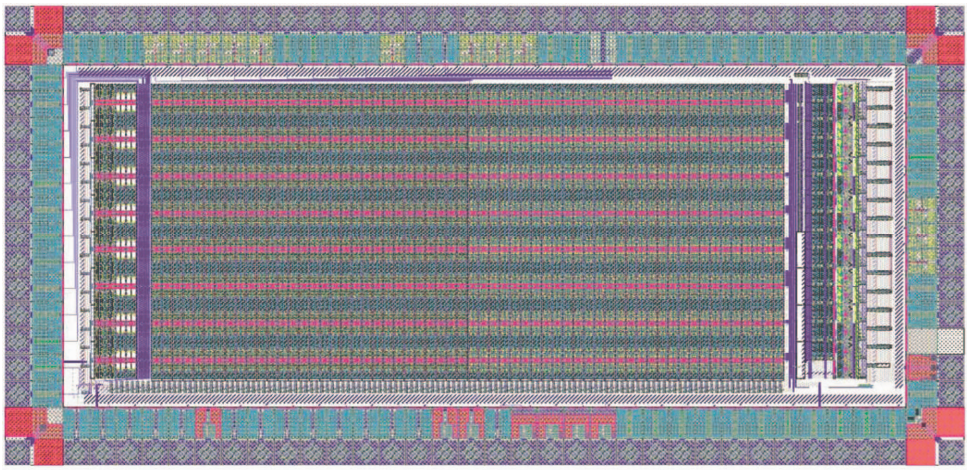


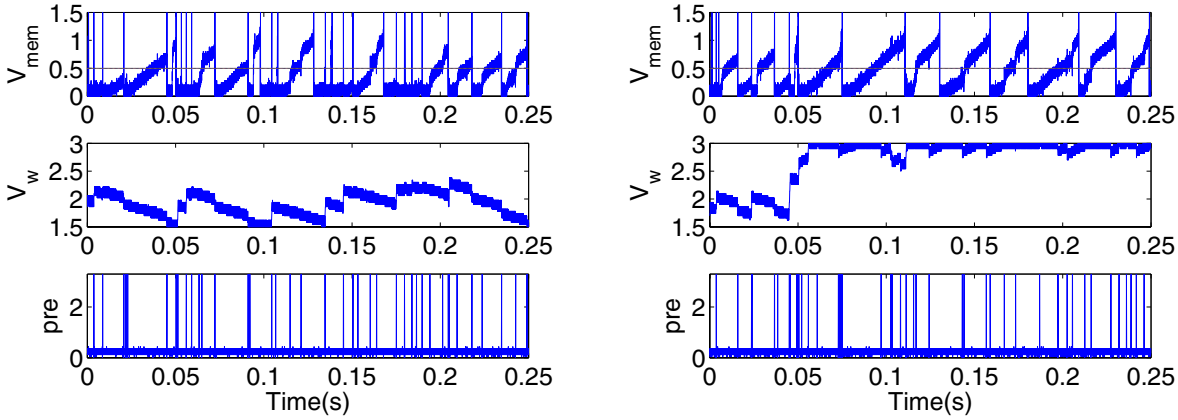
Figure 6. Layout of a $6.1mm^2$ neuromorphic chip comprising an array of 16 I&F neurons and 2048 synaptic circuits. The vast majority of the silicon area is occupied by plastic synapses. The row of 16 neurons is layed out on the far right of the chip. The AER digital interfacing circuits are layed out on the periphery of the chip.

6. Experimental Results

We designed an AER trans-ceiver chip comprising an array of 16 I&F neurons and 2048 synaptic circuits: 128 per neuron, of which 120 plastic (as described in Figure 5(b)), 4 excitatory non-plastic (as described in Figure 4(a)) and 4 inhibitory, non-plastic. The chip, fabricated using a standard 0.35μm CMOS technology, occupies an area of $6.1mm^2$ (see Figure 5.2).

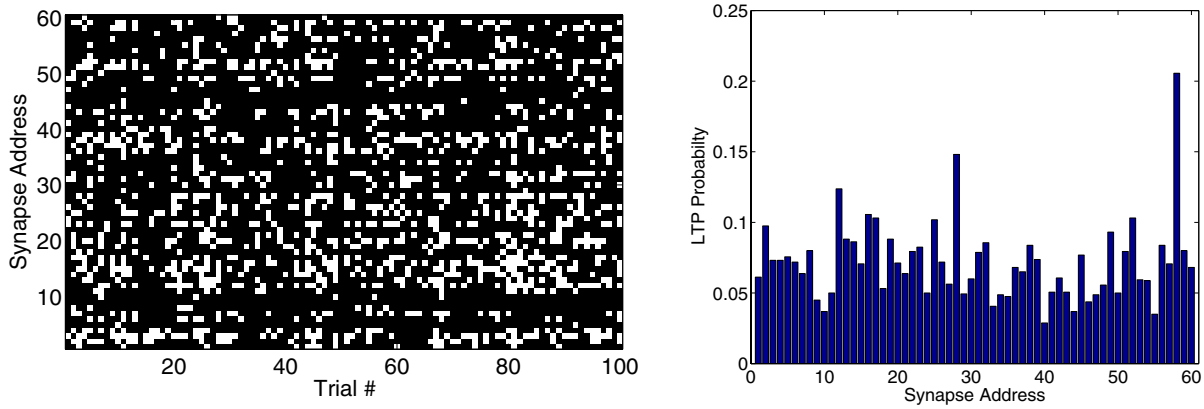
To characterize the plasticity circuits and test the the stochastic nature of the weight update mechanism, we stimulated a plastic synapses and observed its weight voltage as it was being updated. Specifically, we first made a post-synaptic neuron fire at an average frequency of 80Hz, by stimulating one of its excitatory non-plastic synapses. Once the the V_{Ca} voltage was in the relevant range, as specified in Section 5.2, we stimulated also one of the plastic synapses, with Poisson distributed spike trains with a mean firing rate of 100Hz. Figure 7(a) shows an example of a stimulation session where the weight was increased several times during the trial, but never consolidated a Long Term Potentiation (LTP) transition. In Figure 7(b) we show another instance of a stimulation session, in which we used Poisson spike trains with analogous pre- and post-synaptic firing rates. In this case an LTP transition was made and consolidated (at $t = 0.05s$). 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 induce an LTP transition, while others don't. The probability of inducing long-term potentiation, or long-term depression can be easily controlled by changing the bias parameters of the learning circuits (such as V_{wth}, V_{mth}, *etc.*), as well as the mean frequencies of the pre-synaptic and post-synaptic firing rates.

To further evaluate the stochastic nature of LTP transitions at the network level we stimulated 60 plastic synapses of all 16 neurons present in the chip over 100 trials. In Figure 8(a) we show the response of 60 synapses of one of the network's neurons over all trials. In the experiment, at each trial, all synapses were reset to an initial low state (black pixel). After the stimulation we tested the synapse to check the state of its weight, and assigned it a 1 (white pixel) if it had undergone an LTP transition. The figure shows that all the 60 synapses exhibit the stochastic transition nature, and that all share similar statistics. In Figure 8(b) we show the LTP probability averaged over all 16 neurons, for the same bias settings used in Figure 8(a). We tuned the chip to produce low LTP probabilities, by manipulating the height of the weight "up" and "down" jumps. The value of the average LTP probability can be well controlled also by changing the mean frequencies of the input spike trains, the current comparator thresholds and/or the voltage comparator threshold.



(a) (b)

Figure 7. Synaptic weight updates. The non-plastic synapse is stimulated with Poisson distributed spikes, making the post-synaptic neuron fire at an average rate of 80Hz (top row). The pre-synaptic synapse is stimulated with Poisson distributed spike trains with a mean firing rate of 100Hz (bottom row). Depending on the state of V_{mem}, the weight V_w is either increased or decreased with every pre-synaptic spike (middle row). The bi-stability circuit continuously drives V_w to the synapse's low or high stable states. (a) The updates in the synaptic weight did not produce an LTP transition during the 250ms stimulus presentation. (b) The updates in the synaptic weight produced an LTP transition that remains consolidated. The transition threshold V_{wth} is set to $V_{wth} = 2.5V$ in this example.



(a) (b)

Figure 8. (a) LTP transitions of 60 synapses over 100 trials; (b) LTP transition probability of all synapses in the array (16x60)

7. Discussion and Conclusions

We presented circuits for implementing neural architectures distributed across multiple chips, that use computational principles analogous to the ones used in the nervous system. We described the circuit implementation of a recently proposed spike-driven plasticity mechanism [8], to endow these architectures with learning capabilities, and showed that the circuits fulfill all of the model's requirements for learning to classify complex patterns of mean firing rates. We showed data that characterizes the basic neuron, synapse and plasticity mechanisms. Real world applications are currently being implemented. Examples include real–time learning of auditory data extracted from a silicon cochlea, or on-line classification of spike data obtained in neuro-physiological experiments (*e.g.* to drive actuators or build neural prostheses devices) [11].

Given their adaptive properties, the low-power and size characteristics, and the flexibility offered by the AER infrastructure, these types of systems are particularly suitable for implementing processing units on autonomous and humanoid robots. For example, we are at a point now where building a system comprising two silicon cochleas,

two silicon retinas, and a number of similar AER neural network devices responsible for processing auditory signals, computing motion, implementing selective attention, and so forth is viable.

The types of devices described here are also in a position to best exploit the features of advanced scaled CMOS VLSI processes, without being dramatically affected by the device mismatch, leakage current, and unreliability problems that characterize them: the theoretical model that the plasticity circuits implement explicitly prescribes the use of redundancy and stochastic selection mechanisms. The neural architectures we propose are therefore fault-tolerant *by construction*. The conditions required by the learning theory on the VLSI design are in perfect agreement with the ones required for achieving fault-tolerance, minimizing the effect of device mismatch, avoiding the problems of fast digital clock logic design, *etc.* in advanced VLSI processes.

As the circuits proposed are extremely low power and compact, they can be used to implement very large, massively parallel, and redundant arrays of neurons and synapses. In principle these types of neural networks can scale up to any arbitrary size. In practice the network size is limited by the maximum silicon area and AER bandwidth available. Given the current speed and specifications of the AER interfacing circuits, and the availability of silicon VLSI technology, there is room for developing neuromorphic networks with millions of neurons, and billions of synapses.

Acknowledgment

A large part of the work presented here was done in collaboration with C. Bartolozzi, S. Mitra, S. Fusi, and R. Douglas. This work was supported by the EU grants ALAVLSI (IST-2001-38099), DAISY (FP6-2005-015803), the Swiss National Science Foundation grant no. PP00A–106556, and by the ETH TH0–20174–04 grant.

References

[1] L.F. Abbott and S.B. Nelson. Synaptic plasticity: taming the beast. *Nature Neuroscience*, 3:1178–1183, November 2000.

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

[3] 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*, pp. 1227–1230. IEEE, IEEE, May 2006.

[4] C. Bartolozzi and G. Indiveri. Synaptic dynamics in analog VLSI. *Neural Computation*, 19(10):2581-2603, 2007.

[5] K. A. Boahen. Point-to-point connectivity between neuromorphic chips using address-events. *IEEE Transactions on Circuits and Systems II*, 47(5):416–34, 2000.

[6] A. Bofill-i Petit and A. F. Murray. Synchrony detection and amplification by silicon neurons with STDP synapses. *IEEE Transactions on Neural Networks*, 15(5):1296–1304, September 2004.

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

[8] J. Brader, W. Senn, and S. Fusi. Learning real world stimuli in a neural network with spike-driven synaptic dynamics. *Neural Computation*, 2007. (In press).

[9] 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*, 54(5):981-993, May 2006.

[10] 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*, 52(6):1049–1060, 2005.

[11] A. Cohen, R. Douglas, R. Etienne-Cummings, P. Hasler, T. Horiuchi, G. Indiveri, T. Sejnowski, S. Shamma, and A. van Schaik, editors. *Telluride Workshop on Neuromorphic Engineering Report*, 2006. http://ine-web.org/workshops/past-workshops/index.html.

[12] A. Destexhe, Z.F. Mainen, and T.J. Sejnowski. *Methods in Neuronal Modelling, from ions to networks*, chapter Kinetic Models of Synaptic Transmission, pp. 1–25. The MIT Press, Cambridge, Massachussets, 1998.

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

[14] S. Fusi, M. Annunziato, D. Badoni, A. Salamon, and D. J. Amit. Spike–driven synaptic plasticity: theory, simulation, VLSI implementation. *Neural Computation*, 12:2227–58, 2000.

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

[16] P. Häfliger, M. Mahowald, and L. Watts. A spike based learning neuron in analog VLSI. In M. C. Mozer, M. I. Jordan, and T. Petsche, editors, *Advances in neural information processing systems*, Vol.9, pp. 692–698. MIT Press, 1997.

[17] G. Indiveri. A low-power adaptive integrate-and-fire neuron circuit. In *Proc. IEEE International Symposium on Circuits and Systems*, pp. IV–820–IV–823. IEEE, May 2003.

[18] 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*, 17(1):211–221, Jan 2006.

[19] G. Indiveri and S. Fusi. Spike-based learning in VLSI networks of integrate-and-fire neurons. In *IEEE International Symposium on Circuits and Systems, ISCAS 2007*, pp. 3371-3374, 2007.

[20] J. Lazzaro, J. Wawrzynek, M. Mahowald, M. Sivilotti, and D. Gillespie. Silicon auditory processors as computer peripherals. *IEEE Transactions on Neural Networks*, 4:523–528, 1993.

[21] 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*, pages 508–509. IEEE, 2006.

[22] S.-C. Liu, J. Kramer, G. Indiveri, T. Delbrück, and R. Douglas. *Analog VLSI:Circuits and Principles*. MIT Press, 2002.

[23] W. Maass and C. M. Bishop. *Pulsed Neural Networks*. MIT Press, 1998.

[24] M. Mahowald and R. Douglas. A silicon neuron. *Nature*, 354:515–518, 1991.

[25] M. Mahowald and C. Mead. *Analog VLSI and Neural Systems*, chapter Silicon Retina, pages 257–278. Addison-Wesley, Reading, MA, 1989.

[26] 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, pp. 1919–1922, 2005.

[27] H. Markram, J. Lübke, M. Frotscher, and B. Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, 275:213–215, 1997.

[28] C.A. Mead. *Analog VLSI and Neural Systems*. Addison-Wesley, Reading, MA, 1989.

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

[30] 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 S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems*, volume 15. MIT Press, Dec 2005.

[31] 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*, volume V, pp. 4213–4216, May 2005.

[32] L. Watts, D.A. Kerns, R.F. Lyon, and C.A. Mead. Improved implementation of the silicon cochlea. *IEEE J. Solid State Circuits*, 27:692–700, 1992.

Giacomo Indiveri is an Assistant Professor at the Institute of Neuroinformatics of the Swiss Federal Institute and the University of Zurich. He is an Electrical Engineer by training, and obtained his Master and PhD degrees from the University of Genoa. In 2006 he received the "Habilitation" in Neuromorphic Engineering from the Swiss Federal Institute and the University of Zurich. His current research interests include the design and implementation of neuromorphic systems for exploring the computational properties of networks of silicon integrate and fire neurons, for emulating spike-based plasticity mechanisms, and implementing real-time autonomous systems potentially suitable for engineering applications.