Design of adaptive nano/CMOS neural architectures

T. Serrano-Gotarredona and B. Linares-Barranco
Instituto de Microelectrónica de Sevilla
Spanish Research Council IMS-CNM-CSIC-US
Seville, Spain
terese, bernabe @ imse-cnm.csic.es

***Abstract*— Memristive devices are a promising technology to implement dense learning synapse arrays emulating the high memory capacity and connectivity of biological brains. Recently, the implementation of STDP learning in memristive devices connected to spiking neurons have been demonstrated as well as the dependency of the form of the learning rule on the shape of the applied spike. In this paper, we propose a fully CMOS integrate-and-fire neuron generating a precisely shaped spike that can be tuned through programmable biases. The implementation of STDP learning is demonstrated through electrical simulations of a 4x4 array of memristors connected to 4 spiking neurons.**

I. INTRODUCTION

Human brains are highly-parallel systems composed of about 10^{10} neurons, where each neuron is connected to 10^3-10^4 other neurons through learning connections or synapses. This huge parallel capacity and learning capability helps in making biological brains outperform the most modern supercomputing systems when cognitively processing and learning sensory data coming from the real world. Emulating the biological brains unique capability to extract and integrate vast amounts of sensory stimuli into meaningful categorizations has been a subject of intense research.

However, there are several major unsolved challenges when trying to build brain-inspired artificial computing systems as their massive parallelism, massive interconnectivity as well as the plasticity of the interconnections.

Recently, scalable systems based in the modular interconnection of neuron populations through address-event-representation links have been reported [1]-[4] as well as other approaches based on building large neuron populations in a wafer scale integration [5]. However, the limited connectivity imposed by the 2-dimensional nature of CMOS devices as well as the difficulty in implementing CMOS dense analog plastic devices limit the performance of these systems.

In this context, two terminal memristive devices exhibiting non-volatile continuous memory have appeared as a promising technology to implement dense arrays of plastic synaptic devices. It has been demonstrated that these devices when connected with spiking neurons exhibit biological spike-timing-dependent-plasticity (STDP) type of learning [14]-[16]. The use of memristors as dense synaptic plastic devices would allow the implementation of large scale brain inspired computing systems.

In this paper, we propose hybrid architectures composed of CMOS spiking neurons combined with memristive devices arrays. The shape of the pulses generated by the CMOS neurons can be programmed allowing tuning and manipulating the STDP learning rules [14]-[16]. The functionality of the proposed circuits is demonstrated through circuit level Spectre simulations.

II. STDP/ANTI-STDP

Spike-time-dependent-plasticity is a neural learning mechanism originally postulated [6] in the context of artificial machine learning algorithms (or computational neuroscience) exploiting spike-based computations (as in brains). It has been proven successful to learn hidden spiking patterns [7] or to perform competitive spike pattern learning [8]. Astonishingly, experimental evidences of biological STDP have later been reported by several neuroscience groups worldwide [9]. Let us call w the synaptic weight connection between two spiking neurons (the pre-synaptic neuron and the post-synaptic neuron). In STDP the change in synaptic weight Δw is expressed as a function ξ of the time difference ΔT between the post-synaptic spike at t_{pos} and the pre-synaptic spike at t_{pre}. Fig. 1(a) illustrates the situation where the pre-synaptic spike occurs before the post-synaptic spike, thus ΔT is positive, whereas Fig. 1(b) illustrates the situation where the post-synaptic spike occurs before the pre-synaptic spike, thus yielding a negative ΔT. Fig. 2 plots experimental data from Bi and Poo [9] and the shape of the interpolated STDP function. Specifically, $\Delta w = \xi(\Delta T)$, with $\Delta T = t_{pos} - t_{pre}$. For positive ΔT there will be a potentiation of synaptic weight $\Delta w > 0$, which will be stronger as $|\Delta T|$ reduces. For negative ΔT there will be a depression of synaptic weight $\Delta w < 0$, which will be stronger

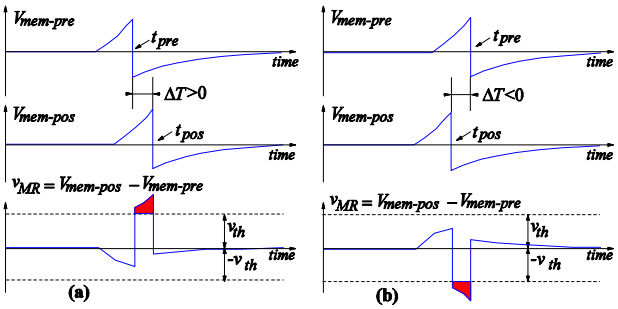

Fig. 1. Pre a Post Synaptic Membrane Voltages for the situations of (a) Positive ΔT and (b) Negative ΔT. Voltage V_{MR} is the difference between post-synaptic membrane voltage and pre-synaptic membrane voltage

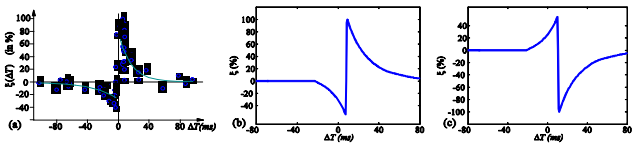

Fig. 2. STDP characterization in biological synapses. Vertical axis is synaptic strength change and horizontal axis is time delay between pre- and post-synaptic spikes

This work has been supported by Andalusian research grant NANONEURO (TIC-2010-6091), Spanish grants VULCANO TEC2009-10639-C04-01 and BIOSENSE TEC2012-37868-C04-01 (with support of the European Regional Development Fund), and European ERANET grant PNEUMA PRI-PIMCHI-2011-0768

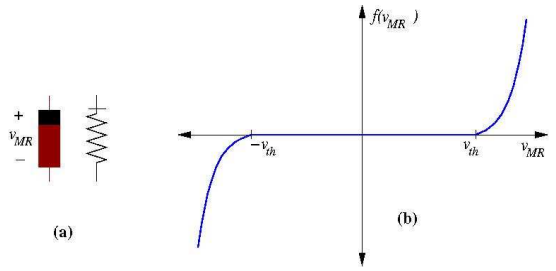

Fig. 3. (a) Memristor symbols and (b) typical characteristic function f(v_{MR}) with exponential growth beyond a threshold

as $|\Delta T|$ reduces. Bi and Poo concluded that they had observed an asymmetric critical window for ΔT of about 40-80ms for synaptic modification to take place. Mathematically, this $\xi(\Delta T)$ STDP learning function is described by computational neuroscientists as

$$\zeta(\Delta T) = \begin{cases} a^{+} e^{-\Delta T/\tau^{+}} if \quad \Delta T < 0 \\ -a^{-} e^{-\Delta T/\tau^{-}} if \quad \Delta T < 0 \end{cases} \tag{1}$$

where a^{+} and a^{-} are parameters controlling the maximum learning rate, and τ^{+} and τ^{-} define the time extension of the learning window.

III. MEMRISTORS AND STDP

Memristance has been recently demonstrated in nanoscale two-terminal devices, such as certain titanium-dioxide [10]-[11] and amorphous Silicon [12] cross-point switches and BFO materials [17]. Memristance arises naturally in nanoscale devices because small voltages can yield enormous electric fields that produce the motion of charged atomic or molecular species, changing structural properties of a device (such as its conductance) while it operates. In this paper we will restrict our discussion to voltage-controlled two-terminal passive memristors of the form [13]

$$\begin{aligned} i_{MR} &= G(w, v_{MR}) v_{MR} \\ \dot{w} &= f(v_{MR}) \end{aligned} \tag{2}$$

where G is its (nonlinear) conductance and w is some structural parameter of the device that controls directly its conductance. In general, we may assume that G is monotonically increasing with w. In memristive nanoscale devices, function f may describe ionic drift under electric fields. Although a linear dependence of f with voltage v_{MR} yields memristive behavior [10], it is more realistic for f to grow exponentially and/or include a threshold barrier v_{th}, as is shown in Fig. 3(b). Fig. 3(a) shows two possible symbols for the memristor two-terminal passive device. Note that it is an asymmetric device, and hence its polarity needs to be indicated explicitely in the symbol: an increase of v_{MR} beyond threshold v_{th} will produce an increase of w (and G). By turning the memristor upside-down, the same increase of v_{MR} would produce a decrease of w (and G).

As demonstrated in [14]-[15], when a memristor is stimulated with pre-synaptic and post-synaptic spikes as shown in Fig. 1, according to function f in Fig. 3(b), weight update will take place only if v_{MR} exceeds threshold v_{th}, as indicated by the red shaded areas in the bottom plot of Fig. 1. By integrating the bottom equation in eq. (2) we can compute the weight update

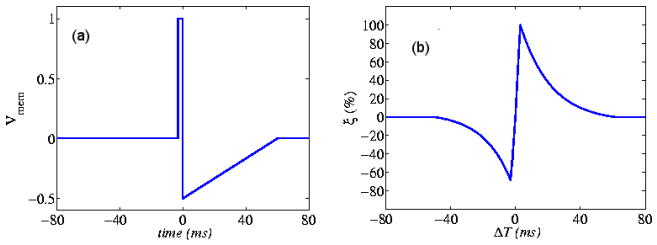

Fig. 4. (b) Resulting STDP function $\xi(\Delta T)$ for the action potential shape shown in (a).

$$\Delta w(\Delta T) = \int f(v_{MR}(t, \Delta T))dt \tag{3}$$

which is the area of the red shaded regions in Fig. 1 previously amplified exponentially through function $f(\ )$. Positive areas (above v_{th} , when $\Delta T > 0$) yield increments for w ($\Delta w > 0$), while negative areas (below $-v_{th}$, when $\Delta T < 0$) result in decrements for w ($\Delta w < 0$). As $|\Delta T|$ approaches zero, the peak of the red area in v_{MR} is higher. Since this peak is amplified exponentially, the contribution for incrementing/decrementing w will be more pronounced as $|\Delta T|$ is reduced. It follows indeed the same behavior of the interpolated STDP function ξ obtained by Bi and Poo from physiological experiments, shown in Fig. 1(a). Consequently, this parameter w can be identified with the synaptic weight in the neural connections with a biological STDP-type learning rule defined in the previous section.

As can be deduced form equation (3), the shape of the spike strongly influences the shape of the resulting STDP function [15]-[16]. Fig. 4(b) illustrates the STDP learning function computed for the spike shape shown in Fig. 4(a) using equation (3). As can be observed, for this spike shape, the STDP learning rule closely resembles the biological STDP function shown in Fig. 2(b).

IV. CMOS IMPLEMENTATION OF THE SPIKING NEURON

Synchronous multi-phase memristive STDP learning architectures have been proposed in the literature [18]. In this paper, we use an alternative fully asynchronous structure composed of memristors and CMOS spiking neurons [14]-[16].

We first need a neural circuit that integrates spikes until a threshold is reached. At that moment, it should provide a spike of the desired shape. A possible schematic diagram for a leaky integrate-and-fire (I&F) neuron block is shown in Fig. 5. The neurons need to include a current summing and sinking input terminal so that in the absence of spike output the integral of input current spike signals can be computed, while maintaining the input node tied to a fixed voltage. This can be done by using an integrator with a clamped voltage input. The

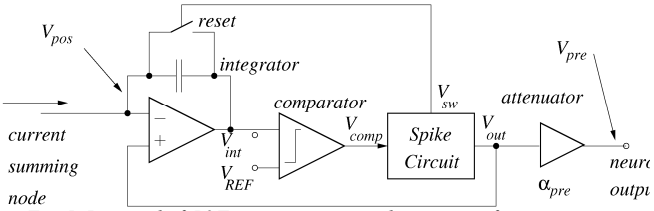

Fig. 5. Proposal of I&F neuron circuit implementation for memristance compatible STDP fully asynchronous learning system

output of this accumulated integral V_{int} is compared against a reference V_{REF}. If this reference is reached, the comparator output will trigger a spike generation circuit, which provides the output spike of the neuron. During spike generation, the input opamp is configured as a voltage buffer, thus copying the spike waveform generated by the spiking block at output node V_{out} to the neuron input node. An attenuated version of the spike is fed forward to the output of the neuron $V_{pre}(t) = \alpha_{pre} V_{out}(t)$. During the whole time of the spike (typically in the order of 20-100*ms*) the neuron is not integrating (computationally inactive). This time is also called "*refractory time*". During the absence of spike output, the spike generation circuit provides a constant voltage V_{rest}.

For the *Spike Circuit* in Fig. 5, an analog circuit can be devised that generates a specific action potential shape with some tunable parameters. Our proposed CMOS implementation of a spike circuit generating programmable spikes is shown in Fig. 6. This block generates a waveform whose shape is depicted in Fig. 4(a). In the proposed implementation, the main characteristics of the shape (duration of the positive pulse T_{high}, duration of the tail pulse T_{tail}, voltage level of the positive pulse V_{high}, lowest voltage level of the tail pulse V_{low}, reference voltage level of the resting neuron V_{rest}) are fully programmable through current and voltage biases.

The input to the spiking block (shown in Fig. 6) V_{comp} is the output of the comparator block of the I&F neuron shown in Fig. 5. The spiking block generates two outputs: an analog voltage at terminal V_{out}, and a digital control signal V_{sw}. When the input voltage V_{comp} gets activated (logic high) the spiking circuit generates an output signal at terminal V_{out} which has the shape shown in Fig. 4(a). The additional digital output voltage V_{sw} generates a pulse that activates the reset switch in Fig. 5 during the whole duration of the spike [15], as well as latches the input signal during the spiking time. As can be observed in Fig. 6, the designed spiking block is composed of 5 sub blocks: an *input latch* that latches the output of the input comparator during the whole duration of the generated spike, *two monostables*, the *tail voltage ramp generator* and the *switch control pulse generator* to generate the digital control output V_{sw}. The monostable blocks are able to generate two matched pulses of high voltage V_{high} and controllable duration. The *first monostable* has been modified so that its low voltage level is controlled by the *tail voltage ramp generator*. The

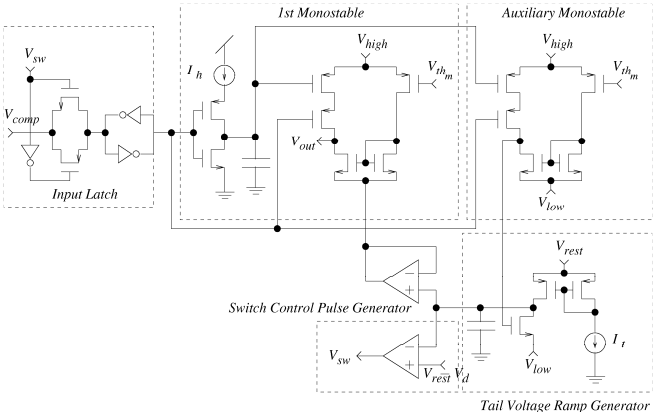

Fig. 6. Schematic of the Spiking Block

second or *auxiliary monostable* is used to know the time to initiate the generation of the pulse tail. The switch control pulse generator initiates an active pulse V_{sw} to control the reset switch when a high pulse is produced in the *auxiliary monostable* and uses the end of the tail voltage ramp to terminate the pulse V_{sw}. The pulse at V_{sw} is used also to control the input latch keeping it opaque during the whole duration of the output pulse. Its functionality is demonstrated through electrical simulations.

The correct operation of the spiking block has been verified through simulations. As an illustration, we show in Fig. 7 simulation results of the spikes generated at output V_{out} with different values of the current biases I_h and I_t. In these simulations, the voltage values are set to V_{REF}=1.65*V*, V_{rest}=1.65*V*, V_{high}=2.5*V*, V_{low}=1*V*. In Fig. 7 (a) the values of I_h are 10*pA*, 100*pA*, and 1*nA* and current I_t=1*pA*. For Fig. 7(b) I_h=5*pA* and current I_t is set to 1*pA*, 10*pA*, and 100*pA*.

V. SIMULATIONS OF HYBRID NANO-CMOS ARRAYS

In this section we demonstrate that a neural array composed of a matrix of 2-terminal memristive synaptic nanodevices interconnected to the spiking CMOS neurons proposed in section IV, exhibit STDP learning behaviour. The STDP learning of a synaptic array of memristive devices and CMOS spiking neurons has been verified through simulations. The schematic of the simulated array is shown in Fig. 8(a). An array of 4x4 synaptic elements is connected to 4 CMOS spiking neurons. The input spikes that would be generated by spiking neurons of a previous layer are distributed through the vertical lines, while the output spikes generated by the current neurons are distributed through horizontal lines. The post-

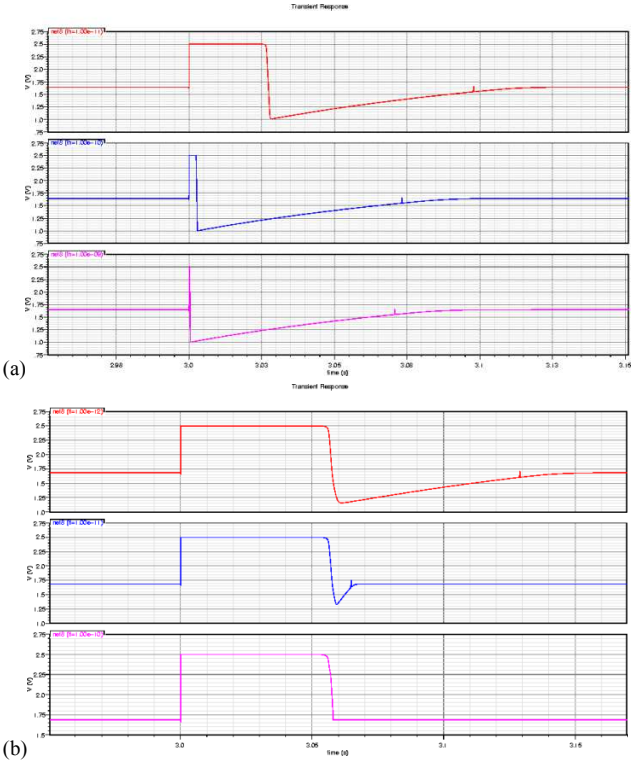

Fig. 7. Programmable Spike Voltage generated by the proposed CMOS Spiking Block (a) for different values of current I_h, and (b) for different values of current I_t.

synaptic voltages V_{pos} are kept at the resting voltage V_{rest} through the virtual ground effect of the high-gain opamp (see Fig. 5) when no spike is generated by a neuron. Each time a presynaptic pulse V_{pre} arrives through a vertical line, a current is integrated in each capacitor during the duration of the presynaptic pulse. As a result of each presynaptic pulse V_{pre}, the voltage at the input of each spiking block increases/decreases an amount that depends on the current value of the coupling memristance and the particular shape of the presynaptic pulse. During this time no weight updating should occur, thus the voltage difference V_{pre}-V_{rest} must be kept below the learning threshold voltage v_{th} (see Fig. 1(b)) of the memristors. When the integrated voltage of any of the neuron capacitors reaches its threshold V_{REF} (see Fig. 5) a postsynaptic spike is generated by the corresponding neuron. The voltage difference of the memristive device should go above the learning threshold v_{th} only when a pre- and post-synaptic spike overlap in time. In order to implement an STDP learning rule, depending on the delay between the presynaptic spike and the postsynaptic spike, the corresponding weight of the coupling memristance should be updated, producing an increase of the resistance if the presynaptic spike goes after the postsynaptic spike, and a decrease of the resistance if the presynaptic spike comes right before the postsynaptic spike.

We have verified through Spectre simulations that the operation of the CMOS spiking blocks connected to the synaptic memristors in the 4x4 array shown in Fig. 8(a) produces a correct STDP learning. The threshold considered for the memristor device is v_{th} =1V. In order to produce correct STDP learning for these memristors we set the voltage levels of the spikes generated by the spiking block to the following values: V_{REF}=2V, V_{rest}=1V, V_{high}=2V, V_{low}=0.5V.

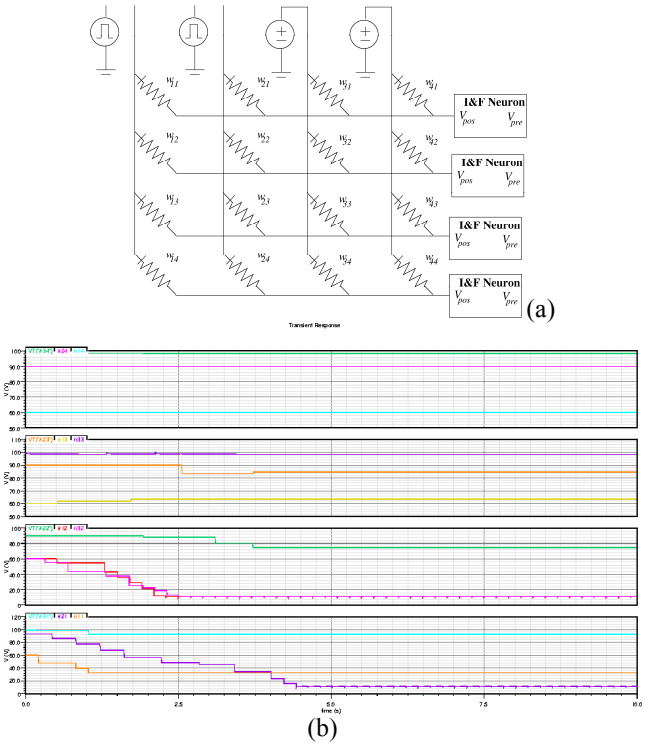

Fig. 8. (a) Simulated array of 4x4 memristors connected to 4 CMOS spiking neurons, and (b) evolution of the memristive weights

The results of the simulations are shown in Fig. 8(b). In this simulation we apply presynaptic spikes to the first and second columns of synapses while the third and fourth columns are tied to voltage V_{rest} so that no learning occurs in the memristors located in those columns. The presynaptic spikes are applied with a period of 200ms. The evolution of the memristor weights is plotted in Fig. 8(b). As can be observed, the weights of the memristors located in the first and second column do not change, as no presynaptic pulse is applied to those columns. However, a change in the weights of the memristors located in the third and fourth columns ocurrs when pre- and post- synaptic spikes overlap.

VI. CONCLUSIONS

A CMOS spiking neuron generating tunable spike waveforms has been designed. The STDP learning of memristive devices connected to that neuron has been verified through simulations of a 4x4 array of memristors connected to 4 spiking neurons. The next step will be to fabricate the proposed neuron and the experimental interaction with memristive devices.

REFERENCES

[1] M. Khan, D. Lester, L. Plana, A. Rast, X. Jin, E. Painkras, and S. Furber et al., "SpiNNaker: Mapping Neural Networks onto a Massively-Parallel Chip Multiprocessor," IEEE Int. Joint Conf. on Neural Networks (IJCNN), pp. 2849–2856, June 2008.

[2] C. Zamarreño-Ramos, et al. "Multi-Casting Mesh AER: A Scalable Assembly Approach fo Reconfigurable Neurmorphic Structured AER Systems. Application to ConvNets," *IEEE Trans. BioCAS*, in Press.

[3] A. Cassidy, A. Andreou, and J. Georgiou, "Design of a one million neuron single fpga neuromorphic system for real-time multimodal scene analysis," 45th Annual Conf. on Inf. Sciences and Systems (CISS), pp. 1–6, 23-25 March 2011.

[4] S. Joshi, et al. "Scalable Event Routing in Hierarchical Neural Array Architecture with Global Synaptic Connectivity," CNNA, Feb. 2010.

[5] J. Fieres et al., "Realizing Biological Spiking Network Models in a Configurable Wafer-Scale Hardware System," IEEE Int. Joint Conf. on Neural Networks (IJCNN), pp. 969–976, June 2008.

[6] W. Gerstner, et al. "Why Spikes? Hebbian learning and retrieval of time-resolved excitation patterns," *Biological Cybernetics,* 69, 503-515, 1993.

[7] T. Masquellier et al., "Spike timing dependent plasticity finds the start of repeating patterns in continous spike trains," PLOS ONE, 3(1), e1377.

[8] T. Masquellier et al., "Competitive STDP-based spike pattern learning," Neural Comp. 21, 1-18, 2008.

[9] G. Bi and M. Poo, "Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type," *J. Neurosci.,* 18(24), 10464-10472, 1998.

[10] Strukov, et al. (2008). The missing memristor found. Nature 453, 80-83.

[11] J. Borghetti, et al. A hybrid nanomemristor/transistor logic circuit capable of self-programming. Proc. Natl. Acad. Sci. U.S.A. 106, 1699–1703.

[12] Jo, S. H., Kim, K.-H., and Lu, W. (2009). High-density crossbar arrays based on a Si memristive system. *NANO Lett.* 9, 870–874.

[13] L. O. Chua, "Memristor - the missing circuit element," *IEEE Trans. Circ. Theory*, vol. 18, pp. 507-519, 1971.

[14] B. Linares-Barranco and T. Serrano-Gotarredona, "Memristance can explain spike-time-dependent-plasticity in neural synapses," in Nature precedings, http://hdl.handle.net/10101/npre.2009.3010.1

[15] B. Linares-Barranco and T. Serrano-Gotarredona, "Exploiting Memristance in Adaptive Asynchronous Spiking Neuromorphic Nanotechnology Systems," *Proc. IEEE NANO*, July 2009.

[16] C. Zamarreño-Ramos, et al., "On Spike-timing-dependent-plasticity, memristive devices, and building a self-learning visual cortex," Frontiers in Neuroscience, http:www.frontiersin.org, March 2011, Vol. 5, Article 26.

[17] C. Mayr, P. Staerke, J. Partzsch, R. Schueffny, "Waveforms driven plasticity in BiFeO3 memristive devices: model and implementation," Workshop on Neural Information Processing Systems, NIPS 2009.

[18] Snider, G. S. (2008). Spike- timingdependent learning in memristive nanodevices. *IEEE Int. Symp. on Nano Architectures 2008*, 85–9.