Extended Abstracts of the 2017 International Conference on Solid State Devices and Materials, Sendai, 2017, pp585-586

M-2-02 (Invited)

Emulating Synaptic Plasticity in Neuromorphic Systems with Resistive Memories

S. La Barbera, E. Vianello, T. Werner, B. De Salvo, L. Perniola
CEA, LETI, Minatec Campus, Grenoble, France E-mail: selina.labarbera@cea.fr

Abstract

Resistive memories (RRAM) are today promising candidates to implement artificial synapses in neuromorphic hardware systems. This paper focuses on RRAM to implement *synaptic plasticity*, the key feature allowing learning and memory in the brain. First, we will provide a comprehensive overview of current research on RRAM technologies to implement plasticity mechanisms, such as Spike-Timing-Dependent Plasticity (STDP), Short-Term Plasticity (STP) and Long-Term Plasticity (LTP). Second, we will present how different forms of plasticity can be implemented in a real application case: i.e. a spike-based neuromorphic system for real-time decoding of neural signals. We will show that STDP (a type of Synaptic Learning or LTP) allows the neural networks to learn patterns, while the Short-Term Plasticity (a type of Synaptic Adaptation) improves accuracy in the presence of background noise in the input data.

1. Introduction

Extracting useful information from unstructured data (e.g. raw sensor data) is a key requirement for future mobile and Internet of Things applications. Neuromorphic computing aims to reach this objective by miming the brain performances in terms of low-power and versatility [1]. Nevertheless, the industrial implementation of neuromorphic hardware still presents several major challenges, ranging from material science to innovative devices and circuits engineering. Impressive milestones have been reached recently, but the demonstrated brain-inspired systems mainly rely on purely-CMOS solutions and lack of scalable implementation for the synapses connecting the neurons [2]. Additionally, with pure CMOS technologies, it is quite expensive to provide synapses with 'plasticity' features, due to the large Silicon area needed. In this context, RRAM devices with attractive bio-inspired functionalities and related low-power performance provide an optimal solution for compact implementation of synapses and are, therefore, a key element to allow future implementation of neuromorphic hardware.

2. Synaptic Plasticity Features in RRAM

The *synaptic plasticity* can be defined as the modification of the synaptic conductance as a function of the neuron activity [3]. Different plasticity rules exist and several approaches can be used to classify the *synaptic plasticity* mechanisms [4]. The first approach is the *"causal description"*, based on the origin of the synaptic conductance modification (Fig.1). A first form of plasticity is the so-called *Synaptic Learning* (or Hebbian-type Learning), the synaptic weight depends on the correlation between the pre- and post-neuron spiking activity. Depending on the signal representation, i.e. rate- or temporal-coding, the *Synaptic Learning* can be formulated such as Spike-Rate-Dependent-Plasticity (SRDP) or Spike-Timing-Dependent Plasticity (STDP) with neuronal activity defined as the mean firing rate or the spike timing, respectively [5]. In the latter plasticity form, the pre-synaptic spike is required to shortly precede the post-synaptic one to induce potentiation (conductance increase), whereas the reverse timing of pre- and post-synaptic spike elicits depression (conductance decrease) (Fig1). RRAM devices have demonstrated to be good candidates to implement STDP [6-8]. A second form of *synaptic plasticity* can be referred to as *Synaptic Adaptation*. In this case, synaptic weight modification depends on the activity of the pre- or post-neuron activity only. The *Synaptic Adaptation* can lead to an increase (facilitation) or decrease (depression) of the synaptic conductance.

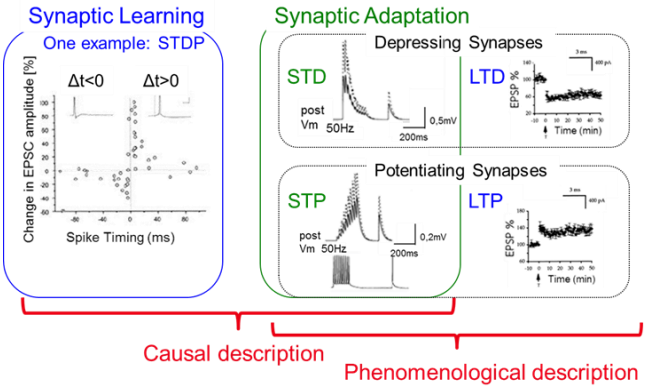

Fig.1 Synaptic Plasticity classification observed in biological synapses

The distinction between *Synaptic Learning* and *Synaptic Adaptation* seems very useful to classify the different synaptic processes and to evaluate their contribution to the learning process. One major difficulty is that both *Synaptic Learning* and *Synaptic Adaptation* can manifest simultaneously and it is complicated to make a clear distinction between them. By means of system level simulations of a real application, we tried to shed new light on the role of these two different types of plasticity [9-10].

A second approach, used to describe the *synaptic plasticity*, can be defined as *"phenomenological description"*: synaptic weight modifications can be either permanent (i.e., lasting for month or years) or temporary (i.e., recovering quickly their initial state) [11]. This observation leads to the definition of Long-Term Plasticity (LTP) and Short-Term Plasticity (STP), respectively (Fig1). The boundary classification into Long-Term (LT) and Short-Term (ST) effects is not well defined and should be considered with respect to the task to be realized. From a phenomenological point of view, the Spike-Timing-Dependent Plasticity (STDP) is

verified on long time scale (it is a form of LTP), the *Synaptic Adaptation*, on the contrary, is most often observed on short time scale (it is a form of STP). The first proposition of STP was realized in a nanoparticles/organic memory transistor (NOMFET) [12]. Equivalently, STP in two terminal devices can be implemented by taking advantage of the volatility of the different memory technologies, i.e. low retention of the state that is often a drawback for pure memory applications [13-14]. Moreover, STP has been implemented in non-volatile RRAM compounds using a specific programming methodology [9-10].

3. Phenomenological Implementation: co-existence of STP and LTP in the same RRAM device

If the contribution of STP and LTP processes to computing is not completely understood in biological systems, both STP and LTP effects in synaptic connections has been evidenced and should play a crucial role. A first approach to combine STP and LTP processes is to consider that repetition of STP effects should lead to a LTP modification in the same synaptic connection. Ohno et al. [11] reported for the first time the transition from STP to LTP in atomic bridge technology. In this case, the synaptic conductivity is increased due to the formation of a conductive filament across the insulating gap. While for low frequency, the filament tends to relax between pulses; higher frequencies lead to a strong bridge that maintains the device in the ON state. In this case, the transition from STP to LTP, is associated to the mean firing rate of the pre-neuron. Both STP and LTP follow the *Synaptic Adaptation* plasticity rule. A similar approach to combine both STP and LTP in the same electrochemical metallization cell has been proposed in [14]. The transition from the volatile to the non-volatile regime is tuned by the number and the size of the conductive filaments.

4. Causal Implementation: co-existence of STP and LTP in a neuromorphic system (using two RRAM devices)

We proposed a circuit strategy to implement Spike-Timing- Dependent Plasticity (Synaptic Learning) and depressing Short-Term Plasticity (Synaptic Adaptation) using two separate RRAM-based synapses [9-10]. The main advantage of this approach is that STPD (Long-Term Plasticity) and depressing Short-Term Plasticity can occur at the same time. The circuit, presented in Fig.2a, allows to connect the RRAM-based synaptic compounds $y_i(t)$ and w_{ij} that are able to reproduce the depressing STP and the STDP (LTP) as described in [9-10]. The strength of the synaptic connection (or synaptic conductance) between the input neuron, A_i, and the output neuron, B_j, is $w_{ij}*y_i(t)$, where w_{ij} is the strength modulated according to the STDP (LTP) and $y_i(t)$ is a scaling factor that describes the depressing STP. When an input neuron receives an incoming event, it generates a small voltage pulse (V_{in}) that propagates to the output neuron through $y_i(t)$, the buffer and w_{ij}. The resulting post-synaptic current (I_{out}) is proportional to $w_{ij}*y_i(t)$. In order to study the role of the STDP (LTP) and depressing STP, we

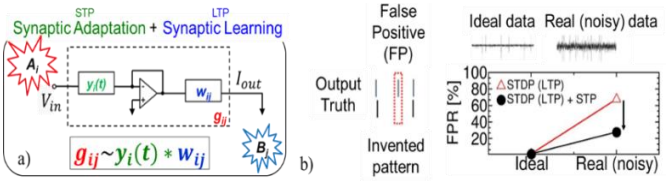

Fig.2 Co-existence of STP and LTP in a neuromorphic system

performed full system-level simulations of a real application. A Fully Connected Neural Network (FCNN) for real-time decoding of neural signals has been implemented using the N2D2 simulator [9]. The neurons are modeled with Leaky Integrate and Fire model and the synapses emulate the depressing STP and the Long-Term STDP as described in [9]. In Fig.2b the metrics to evaluate the network performances (i.e. the Detection Rate, DR, and the False Positive Rate, FPR) are shown. The network is designed to learn and distinguish action potentials (i.e. spikes) from background noise. If the signal-noise-ratio (SNR) is high (80 in the 'ideal' case), the network without STP achieves a DR and FPR around 97% and 1.6%, respectively. For a SNR=27, the network performance without depressing STP decreases significantly due to the high FPR. For a SNR=27, the introduction of depressing STP is mandatory to guarantee the network functionality. It allows to decrease the FPR by 35% while the DR decreases only 2.5%. Thanks to the introduction of the Synaptic Adaptation (depressing STP), the network learns to stop responding to noise stimuli, thus improving the performances in presence of highly noisy input data.

5. Conclusions

In this paper, we reviewed different plasticity mechanisms that can be implemented in RRAM devices. Then, we demonstrated that the combination of depressing Short-Term Plasticity (STP) and STDP (LTP) makes the neuromorphic systems highly robust against environmental noise in the input data. The STP implementation allows to reduce the False Positive events generated by the input noise, thus improving the network performances.

Acknowledgements

This work has been supported by the h2020 NeuRAM3 project.

References

[1] G. Indiveri, and S. C. Liu, *Proceedings of the IEEE*, (2015)
[2] P. A. Merolla et al., *Science*, (2014)
[3] T.V. Bliss et al., *Nature*, (1993)
[4] S. La Barbera and F. Alibart, *Springer India*, (2017)
[5] Bi and Poo, *Journal of neuroscience* (1998)
[6] D. Kuzum al., *Nano letters* (2011)
[7] Kim et al., *Nano letters* (2015)
[8] M. Prezioso et al., *Nature Scientific Reports* (2016)
[9] T. Werner et al., *IEDM* (2016)
[10] E. Vianello et al., *IMW* (2017)
[11] H. Markram et al., *Science* (1997)
[12] F. Alibart et al., *Advanced Functional Materials* (2010)
[13] T. Ohno et al., *Nature Materials* (2011)
[14] S. La Barbera et al., *ACS nano* (2015)