Synaptic realizations based on memristive devices

V. Milo[1], T. Dalgaty[2], D. Ielmini[1], E. Vianello[2]

[1] Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, p. Leonardo da Vinci 32, 20133 Milano, Italy.
[2]Univ. Grenoble Alpes, CEA, LETI, 38000 Grenoble, France. Email: Elisa.VIANELLO@cea.fr

Abstract

In the last 10 years, neuromorphic computing has emerged as a novel approach to tackle the challenges of the end of Moore's law. In this frame, memristive devices are very promising due to their unique properties, such as high compactness, high switching speed, low power consumption, and tunable resistance. In particular, memristive devices can be used as suitable synaptic connections that can replicate the local plasticity rules in biological networks, thus making the memristor an essential element to develop cognitive systems with the same capabilities as the human brain.
In this chapter, we present an overview of the current status on the synaptic circuits based on memristive devices. We review various implementations, including single-memristor synapse adopting resistive switching memory (RRAM) and phase change memory (PCM), hybrid structures combining complementary metal-oxide semiconductor (CMOS) transistors and memristive devices, and materials-based approaches aiming at reproducing biological learning rules by the physical properties of the device. Learning rules such as the spike-timing dependent plasticity (STDP), the spike-rate dependent plasticity (SRDP) and the short-term plasticity (STP) are described. We finally present few examples of learning circuits combining synaptic networks, thus supporting the promising prospect of memristive circuits capable of bio-realistic brain-inspired cognitive computing.

1. Introduction

Since the seminal works of Rosenblatt [1] and Minsky [2], the neural network has been recognized as the most powerful circuit to describe the human brain and achieve a certain level of 'intelligence' in hardware. Among the neural networks, the deep neural network (DNN) has been shown to achieve a high accuracy in learning objects, images and speech [3,4]. DNN requires however supervised learning with an extensive dataset, to train the system by iterative schemes such as the backpropagation and other gradient descent techniques. Such a learning scheme can be seen as a mere mathematical method to improve the fitting of existing data by iteratively updating the synaptic weights, which lacks any specific similarity with the human brain. On the other hand, the spiking neural network (SNN) aims at reproducing the cognitive processes in the human brain, which largely rely on the exchange of spikes among neurons to process information [5]. This is the so-called 'neuromorphic approach', where the circuit design aims at replicating the exact architecture, the information coding, and the learning methodology of the human brain. In neuromorphic SNNs, spikes also control learning via Hebbian rules such as the spike-timing dependence plasticity (STDP) and the spike-rate dependent plasticity (SRDP).
To implement DNNs and SNNs in hardware circuits and systems, the CMOS technology has been traditionally adopted in both digital and analogue (or mixed) circuits [6,7]. CMOS circuits combine a large flexibility of design, a good scaling and the possibility to operate transistors in the subthreshold regime, which is useful to minimize the power consumption and achieve a high energy efficiency, as in the human brain. On the other hand, CMOS circuits generally lack a synaptic device technology capable of storing a synaptic weight in a nonvolatile stability and analogue accuracy. Emerging nonvolatile memories, such as the resistance switching memory (RRAM) [8]

and the phase change memory (PCM) [9], instead, naturally provide the synaptic element which is needed for DNN and SNN. These types of memories are relatively small and scalable, since they have a 2-terminal resistive structure, where the resistance can be suitably changed by the application of electrical pulses. Thanks to the material storage concept in RRAM and PCM, the memory device can be miniaturized to the range of about 10 nm [10]. RRAM and PCM also display analogue switching, where the resistance can be increased or decreased gradually by the application of suitable voltage pulses [11]. Emerging memories can be easily implemented in CMOS circuits, thanks to the back-end-of-line (BEOL) integration [12]. Finally, RRAM and PCM has been shown to enable fast and energy-efficient in-memory computing [13], thanks to the implementation of physical matrix-vector multiplication (MVM) within a crosspoint array accelerating data processing [14-16] and the non-iterative solution of linear algebra problems [17]. Given these multiple advantages from the physical, architectural and scaling perspectives, the nonvolatile resistive memories have been recognized as a promising technology to implement synaptic elements within high density neuromorphic systems [18].

This chapter presents the hardware implementation of synapses with bio-realistic plasticity, relying on RRAM and PCM. First, the plasticity rules for biological synapses, such as STDP and SRDP, are reviewed with reference to in vivo and in vitro experiments. Implementing such bio-inspired plasticity rules in hardware is essential for designing SNNs which emulate some of the cognitive functions of the human brain, such as unsupervised learning, pattern recognition, association, attention, and planning. Synaptic implementations are then discussed by describing RRAM synapses, PCM synapses, and various hybrid implementations combining one or more transistors with resistive devices to enable higher functionality and flexibility of the synaptic circuit. Non-overlap synapses, differential synapses, 3D synapses, and 3-terminal synaptic transistor concepts are also presented to provide a comprehensive overview of various architectural approaches to STDP synapses. Triplet and SRDP learning synapses are also introduced with their applications in learning and filtering of spiking information. Finally, the chapter will provide an overview of full-hardware implementations of SNN for learning of patterns, thus further supporting the relevance of biological learning rules for enabling brain-inspired functions in silico.

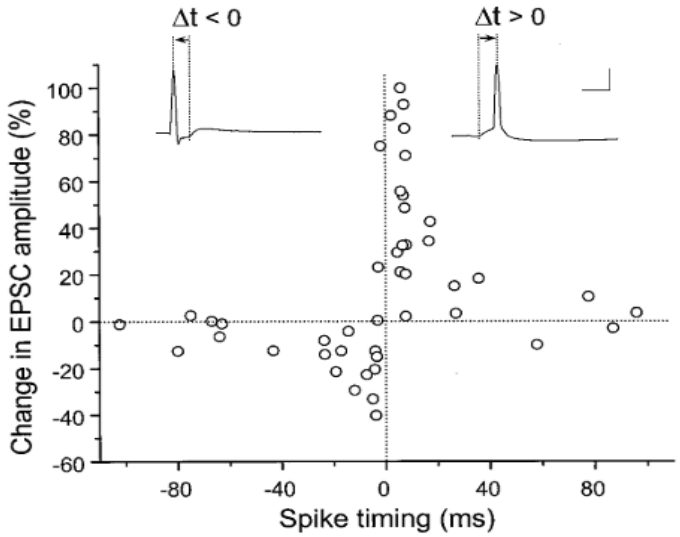

Figure 1. Experimentally observed pair based STDP characteristics. Reprinted by [19].

2. Biological synaptic plasticity rules

The computational elements of nervous systems, neurons and synapses, continuously adapt their properties for the purposes of homeostasis, short-term adaptation and long-term changes for learning and memory formation. This adaptation takes place by modifying the properties and number of ion-channels on their cell membrane. These modifications result in changes of ion-channel efficacy and temporal dynamics of ion exchange. For the case of synapses these modifications are usually abstracted to the idea of a change in a synaptic weight, which can be expressed as a function of the spike timing or spiking rate of the pre- and post- synaptic neurons. A body of literature work uses this organizational perspective to derive "learning rules" which govern the synaptic weight modification on the basis of data derived from biological experiments. Well-known learning algorithms are the STDP rule, which induces changes triggered by pairs of pre- and post- synaptic spikes, and the SRDP, where synaptic potentiation and depression are controlled by high and low presynaptic spike rates, respectively. The changes can be persistent for long-term plasticity or non-persistent for short-term plasticity (STP). The following sections summarize the most common models of plasticity.

2.1 Long-term STDP and SRDP

Changes in the synaptic weight are believed to encode the memory behavior and serve as the principal mechanism for learning in nervous systems. The most known STDP rule is a long-term plasticity induced by pairs of presynaptic and postsynaptic spikes, which was first experimentally observed in 1998 [19]. The changes in synaptic weight depend on the difference in spike timing between a pre- and a post-synaptic neuron and is persistent. The direction of the weight change depends on the polarity of this timing difference. The synaptic weight between two neurons increases for the case of the pre-synaptic neuron firing before the post-synaptic neuron, leading to the so-called long-term potentiation (LTP). On the other hand, the synaptic weight between two neurons decreases for the case of the pre-synaptic neuron firing after the post-synaptic neuron, leading to the so-called long-term depression (LTD) (Fig. 1). The weight change is higher when the spike-time interval is short and it tends to zero for increasing spiking interval, which is consistent with the Hebb's postulate [20]. The dependence between the spike time interval and the weight change can be modelled as a piecewise function of two exponentials. Other shapes for the STDP characteristic have been observed, such as a symmetric dependence or anti-Hebbian plasticity where the time dependence is reversed compared to the classical time dependence [21-24]. In all these cases, the change in the synaptic weight depends on the relative timing of the pre- and post-synaptic spikes, which is the core principle of the pair-based STDP rule. However, pair-based STDP fails to replicate the results of richer experimentally observed biological features. In particular, it has been demonstrated that a triplet rule (i.e., a rule that considers sets of three spikes, two pre and one post or two post and one pre) is more biologically realistic [25].

SRDP is another paradigm for implementing the Hebbian synaptic plasticity. The SRDP induction protocol is predominantly based on the neuronal firing rate to vary the sign and magnitude of synaptic plasticity [26-28]. As observed in the hippocampus/neocortex, the post-synaptic terminations underwent LTP when the pre-synaptic neuron fired with a high frequency (20–100 Hz), while LTD was observed instead for low-frequency spiking (1-5 Hz). A simple and effective learning rule to implement SRDP, often called the Fusi rule, relies on the post-synaptic firing rate instead [26]. After a pre-synaptic pulse, the synapse can be depressed or potentiated depending on whether the post-synaptic membrane potential is low or high gated by an additional variable called the calcium variable, which is determined by the neurons firing rate. Synapse potentiation is inhibited when the calcium variable is above a certain threshold, while synapse depression is inhibited for the calcium variable being below another threshold.

2.2 Short-term plasticity

Long term STDP and SRDP induce persistent synaptic weight changes. On the other hand, short term non-persistent synaptic weight changes can also take place after the synapse has propagated a spike. Following a pre-synaptic spike, the weight of the synapse can either transiently decrease (depression) or increase (facilitation), followed by a decay in time of the synaptic weight toward its baseline level. As is the case for long-term plasticity, short-term plasticity has been observed in biological experiments [29-31]. Short-term plasticity may result in either a depression when each presynaptic spike induces a decrease of the synaptic weight (Fig. 2a), or a facilitation when each presynaptic spike induces an increase of the synaptic weight (Fig. 2b). As changes induced by short-term plasticity only take effect during a short period and rapidly fade with time, it is not sufficient to cause a stable learning. However, it has interesting properties that contribute to the efficiency of the neural network. One notion is that the short-term change behaves as a temporal filter of spiking trains. For instance, a synapse exhibiting short-term depression acts as a low-pass filter since high frequency pre-synaptic activity is attenuated in the synapse before it can excite the post-synaptic neuron. The contrary is true for short-term facilitation whereby only a high rate of pre-synaptic activity is sufficient to achieve a synapse strong enough to significantly excite a post-synaptic cell.

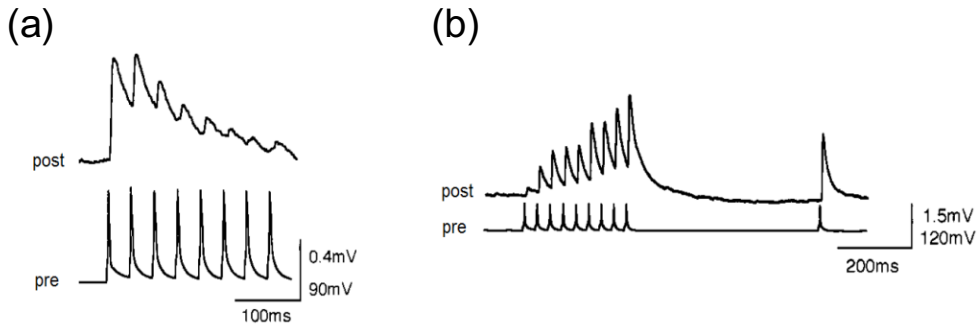

Figure 2. Experimentally observed short term plasticity. (a) Example of short- term depression. (b) Example of short-term facilitation. Bottom traces show the presynaptic spikes, top traces show the postsynaptic potential. Adapted from [29].

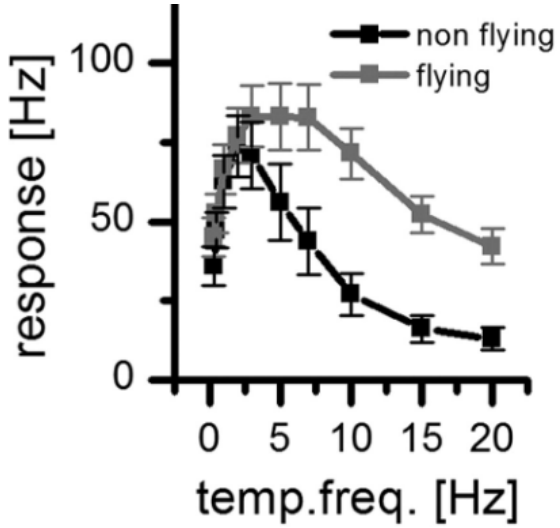

Figure 3. Temporal frequency sensitivity tuning curve of the mean response of Lobula Plate, a neuron dedicated to the processing of optic flow, in drosophila resting and flying states. Adapted from [32].

2.3 State-dependent synaptic modulation

Further synaptic temporary modulation can be induced by signals from neuromodulatory neurons dependent on the state of the animal [32]. An example can be found in the elementary motion detection system of drosophila where the neuromodulator octopamine tunes neuronal properties in the visual system as a function of whether the insect is resting or flying. This allows the insect to adapt its sensitivity to different velocities of stimulus as well as reduce power consumption while in a resting state. In Fig. 3 the response of Lobula Plate tangential cells, well-characterized neurons dedicated to the processing of optical flow, is reported for Drosophila stimulated with a moving grating when it is in resting and flying states. The area under the curve for the insect in its resting state is greatly reduced relative to that of its flying state which is thought to be an evolutionary adaptation to optimize the energy consumption.

3. Memristive implementations

To develop bio-inspired neuromorphic hardware, the implementation of the biological synaptic plasticity rules, such as STDP and SRDP, is essential. In fact, a key enabling feature of neuromorphic circuits is their ability for learning and adaptation, which requires synaptic plasticity as in the human brain. As a result, there has been a significant effort in the exploration of novel devices that could replicate bio-inspired learning rules with simple algorithms, low energy consumption, and high density of synaptic connections. To this purpose, memristive devices appear as a promising technology to emulate the synaptic behavior in artificial neural networks. In particular, strong interest was gained by a class of memristors including RRAM and PCM, also called first-order memristors [33], depicted in the sketch of Fig. 4a. In this type of devices, STDP can be achieved solely by the application of overlapping spikes at device terminals as schematically depicted in Fig. 4b [34]. In addition to first-order memristors, another class of memristors, called second-order memristors (Fig. 4c), has been recently proposed [33], evidencing that resistive switching phenomena can be induced by non-overlapping spikes applied across memristor device with variable positive/negative relative delay Δt (Fig. 4d). The non-overlap resistance switching can be explained by the occurrence of short-term conductance changes controlled by second-order internal variables such as the internal temperature [34]. This feature is extremely important to implement at device level significant processes such as the Ca^{2+} short-term dynamics [31], thus enabling the gradual weight update shown by biological STDP [19] and SRDP [27,28] with higher detail than the synaptic implementations with first-order memristors. Taking inspiration from these schemes, several hardware implementations of nanoscale synapses based on memristive materials capable of replicating synaptic plasticity have been developed and the most significant prototypes are discussed in the following.

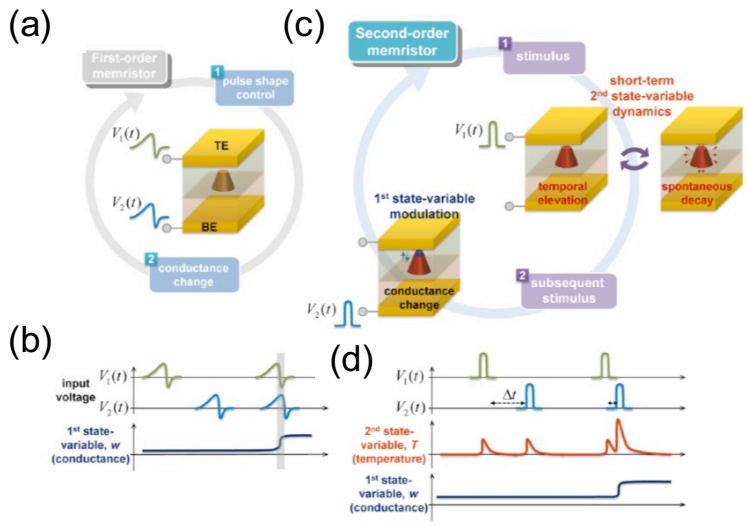

Figure 4. Comparison between (a) a first-order memristor where (b) only overlap of spikes applied at terminals can induce a conductance modification and (c) a second-order memristor where (d) conductance can be changed depending on the sign and magnitude of relative timing of applied spikes thanks to second-order variables (e.g. temperature) displaying a short-term dynamics. Reprinted with permission from [34]. Copyright (2015) American Chemical Society

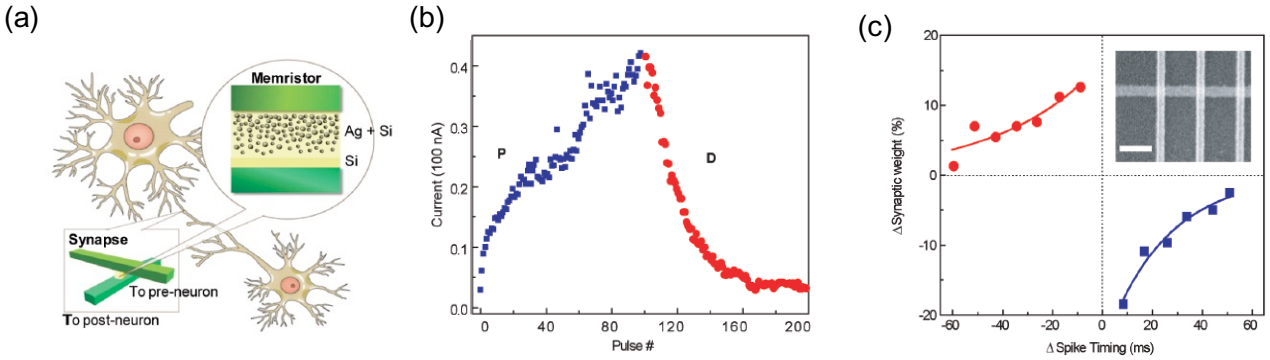

Figure 5. (a) Sketch of a synapse connection between a PRE and a POST neuron implemented by a memristor element. (b) Current response of Ag-Si RRAM device as a function of number of applied pulses for both potentiation (current increase) and depression (current decrease). (c) STDP implementation for Ag-Si memristor at experimental and simulation level by application of PRE/POST spikes with variable time delay. Adapted with permission from [36]. Copyright 2010 American Chemical Society

3.1 RRAM synapses

In last decade, RRAM technology has been intensively investigated to design memristive synapses capable of STDP for bio-realistic neuromorphic systems [34-42]. RRAM combines in fact low-voltage operation, large window, analogue-type multilevel operation, good cycling endurance and strong reliability [8].

Figure 5a illustrates the ideal concept of the RRAM-based synapse, where a memory element within a high-density crosspoint array can serve as synaptic connection between artificial neurons, similar to the biological synapse in the brain [36]. Interestingly, both the biological synapse and the memristive RRAM rely on the ionic diffusion for the plasticity mechanism [42]. One of the earliest implementations of RRAM-based synapses addressed a programmable metallization cell with an Ag-a/Si active layer where two regions with high and low Ag ion concentration, respectively, are formed by suitably setting the gradient of the Ag/Si mixture ratio [36]. Unlike memristors such as

HfO_x or TiO_x-based RRAM, which sometimes exhibits abrupt resistive transitions due to the formation and rupture of a conductive filament, the resistance of this device can be tuned with analogue precision by controlling the motion of Ag ions between Ag-rich and Ag-poor regions by application of an external voltage. To test the synaptic behavior of this device, a DC characterization study consisting of the application of two consecutive series of 100 300-µs-long pulses of amplitude 3.2 V and -2.8 V, respectively, was performed. As a result, Fig. 5b shows the incremental increase of the current during a first series of positive voltage pulses and the incremental decrease of current during the following series of negative voltage pulses, thus supporting the memristor capability of analogue potentiation/depression at positive/negative bias. Figure 5b also shows another feature of potentiation and depression processes for this type of synaptic device consisting of an increasing extent of weight variation in response to voltage pulses with longer width. Based on the characterization study at device level, STDP measurements were carried out. To capture STDP characteristics by Ag-Si RRAM device, a CMOS circuit was realized with two integrate-and-fire neurons connected through a RRAM memristor capable of mapping the relative time delay between occurrence times of PRE and POST spikes ($\Delta t = t_{PRE}$-t_{POST}) into the width of a pulse to be applied to synaptic device via a time-division multiplexing (TDM) scheme with globally synchronized time frames. According to this scheme, if the PRE spike anticipates the POST spike, a potentiation pulse with exponentially decreasing pulsewidth at increasing Δt is applied to the synapse. Otherwise, if the PRE spike follows the POST spike, a depression negative pulse with an exponentially decreasing pulsewidth at increasing $|\Delta t|$ is applied to the device. Figure 5c shows the resulting STDP characteristics obtained by measuring the percentage of synaptic weight update as a function of Δt which evidences an exponential decay of potentiation and depression in agreement with in vivo experimental data.

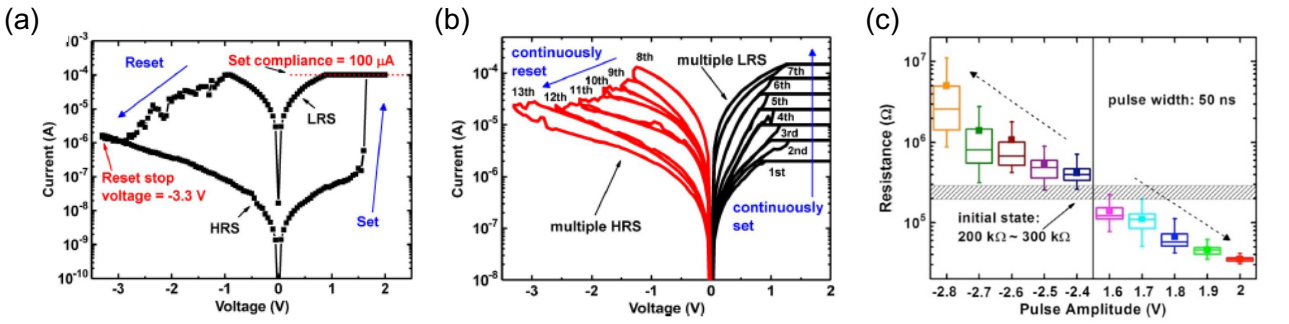

Figure 6. (a) Current -Voltage (I-V) characteristics of the HfO_x/AlO_x RRAM device with compliance current I_C = 100 µA and V_{stop} = -3.3 V. (b) I-V characteristics for increasing I_C, which results in multiple LRS, and increasing V_{stop}, which leads to multiple HRS. (c) Resistance response for HfO_x/AlO_x RRAM device evidencing a gradual resistance decrease/increase for positive/negative pulses of increasing amplitude and fixed 50 ns duration. Adapted from [37].

Although the results in Fig. 5 demonstrated the possibility to achieve STDP in silico for the first time, the TDM approach might require some additional circuit complexity. To reduce the complexity of the STDP scheme, a direct overlap scheme was adopted in a one-resistor (1R) structure of a bipolar RRAM device based on $TiN/HfO_x/AlO_x/Pt$ stack [37]. Figure 6a shows the I-V characteristics of the RRAM device with a relatively abrupt set transition and a more gradual reset transition whereas Fig. 6b shows the I-V curves obtained by a continuous increase of the compliance current I_C from 1 to 200 µA, which allows to set the device at increasingly high conductance. Also, the application of a reset sweep with incremental maximum voltage $|V_{stop}|$ from -1.3 V to -3.3 V allows to reset the device at increasing resistance. Therefore, the controllable set/reset operations support the multiple resistance states of the RRAM [43-46], enabling analogue synaptic potentiation/depression via continuous set/reset processes. Figure 6c further highlights the

multilevel operation capability of the RRAM, showing the measured resistance of synaptic RRAM in response to the application of individual 50-ns-long positive/negative pulses with increasing amplitudes. Starting from an intermediate initial state between 200 kΩ and 300 kΩ, the device resistance can be gradually increased up for pulse amplitudes varying from -2.4 V to -2.8 V, or the resistance can be gradually decreased for pulse amplitudes varying from 1.6 V to 2 V. The figure thus supports the ability to modulate the synaptic weight by applying short pulses of variable amplitude. The multilevel operation controlled by pulse amplitude was thus used as a basis to demonstrate STDP learning rule at device level. To achieve this objective, PRE and POST spikes were properly designed via a sequence of single pulses in consecutive timeslots, namely a negative pulse of period 1 µs followed by 5 positive pulses with identical period and decreasing amplitudes, such that only their overlap can effectively induce a synaptic weight modulation.

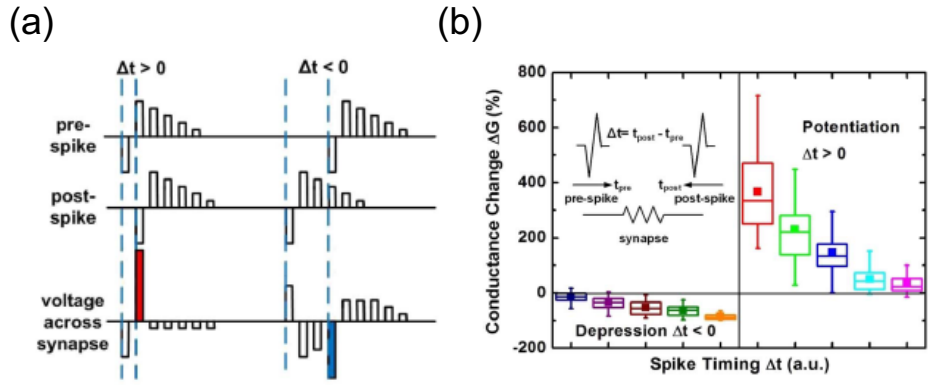

Figure 7. (a) Programming scheme based on overlap of PRE and POST spikes to capture synaptic potentiation and depression according to STDP rule. (b) Calculated relative change in conductance as a function of relative time delay between PRE and POST spikes suggesting the capability of HfO_x/AlO_x RRAM device of mimicking biological STDP rule. Adapted from [37].

Figure 7a shows the waveforms of two spikes that were devised such that if the relative timing between PRE and POST spikes, which is defined as $\Delta t = t_{post} - t_{pre}$ in this report, is positive, a single positive voltage pulse capable of triggering the set process is applied across the device causing potentiation whereas if Δt is negative, a single negative voltage pulse capable of triggering the reset process is applied across the device causing depression. As a result of the application of this overlap approach, an analogue STDP behavior approaching the biological one was captured in simulation. The resulting STDP characteristic is shown in Fig. 7b, which supports the HfO_x/AlO_x RRAM and the overlap scheme as a promising approach for hardware neuromorphic systems able to learn. The engineering of pulse shape/width of PRE and POST spikes applied to memristor terminals plays a crucial role to achieve the memristor conductance modulation, hence synaptic weight update. This is because conductance changes at a given time in memristors used in such synaptic structures, that are first-order memristors [33], is solely governed by the voltage/current input applied to the device and conductance state at that time. However, there is another class of memristors, referred to as second-order memristors [33] where the conductance is also controlled by one or more second-order state variables, which provide an additional degree of freedom to achieve the implementation of synaptic mechanisms increasingly similar to bio-realistic processes.

In this regard, Kim et al presented in [34] a second-order Ta_2O_{5-x}-based memristor device capable of replicating STDP rule with non-overlapping spikes exploiting the short-term dynamics of internal temperature, which thus serves as 2nd order state variable making weight modulation timing dependent. To capture STDP, memristor device was subjected to the application of non-overlapping PRE and POST spikes at two terminals (Fig. 8a) which, as evidenced in Fig. 8b, consist of two consecutive pulses with different amplitude and duration. In detail, PRE spike includes the

sequence of a 20-ns-long programming pulse of amplitude 1.6 V followed, after a time interval of 1 µs, by a longer pulse of amplitude of 0.7 V and width 1 µs for heat generation, whereas the POST spike coincides with PRE spike except for the amplitude of first pulse which is 1.1 V. The application of PRE and POST spike at top electrode (TE) and bottom electrode (BE), respectively, causes an overall voltage across device given by V_{pre}-V_{post} which changes as shown in Fig. 8c depending on whether PRE spike precedes the POST spike (left) or PRE spike follows POST spike (right). In the first case, which is featured by a positive time delay Δt between two spikes, the application of first spike (PRE spike) induces a temperature increase that affects the following spike (POST spike). Upon arrival of second spike, the heat generated by second spike is added to the decreasing heating previously activated by the first spike causing a memristor conductance increase, due to the short negative set pulse within POST spike, higher than conductance decrease induced by the positive reset pulse within PRE spike. This thus results in an overall conductance increase, hence the potentiation of memristive synapse. On the contrary, for negative Δt, an identical mechanism based on short-term dynamics of internal temperature leads memristor device to undergo a conductance decrease activated by second spike (PRE spike) higher than the conductance increase due to the first spike (POST spike), which results in an overall conductance decrease within memristor or synaptic depression. Importantly, note that in both cases the shorter/longer is Δt, the more/less pronounced is the impact of Joule heating summation effect on memristor conductance upon occurrence of the second spike, which thus results in an increasing/decreasing update of synaptic weight. As shown in Fig. 8d, this internal mechanism based on heat summation enables a second-order memristor to achieve at device level a very faithful replication of STDP characteristics observed in biological experiments where relative change in conductance is a function of both Δt sign and magnitude [19].

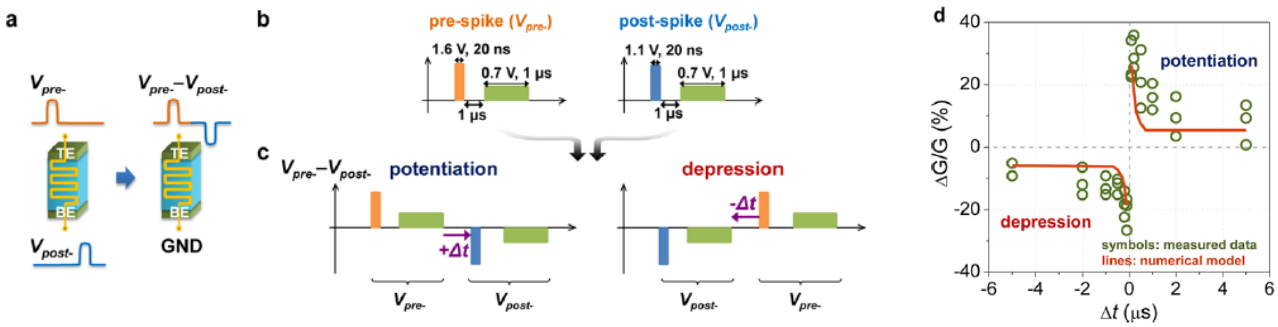

Figure 8. Sketch of memristive device whose terminals are applied two non-overlapping voltage pulses. (b) The application of PRE and POST spikes which consist of sequence of two positive pulses with different amplitude and width, at TE and BE, respectively, results in (c) a voltage across memristive element V_{PRE}-V_{POST} evidencing two consecutive spikes with no overlaps able to induce a conductance change depending on the order of presentation (sign of Δt) and short-term dynamics of internal temperature after pulse application (magnitude of Δt). (d) Experimental STDP characteristics achieved in a second-order Ta_2O_{5-x}-based memristor compared with a characteristics calculated by a numerical model**.** Adapted with permission from [34]. Copyright 2015 American Chemical Society

3.2 PCM synapses

In addition to RRAM technology, other novel non-volatile memory devices have been investigated as potential candidates to build electronic synapses. Among various types of memristors, PCM devices have received a strong interest mainly for their high resistance controllability via the gradual crystallization dynamics of chalcogenide-based active layer and the large resistance window (~10^3) which is ideal for efficient multilevel operation [47].

Similar to the approach described in [37] for RRAM synapses, a scheme based on the overlap between a PRE and POST pulsed voltage signals at device terminals was designed in [48] to demonstrate STDP in single element PCM-based synapses. As shown in Fig. 9a, POST signal consists of 8-ms-long negative pulse whereas PRE signal includes two sequences of 6 consecutive pulses of high and low positive voltages, respectively, separated by a zero period of 8 ms. In the first series, the pulses were designed with width of 50 ns, period of 10 ms and linearly increasing amplitudes to achieve synaptic depression. On the other hand, the following series includes pulses which were designed with width of 1 µs, period of 10 ms and linearly decreasing amplitudes to achieve synaptic potentiation. To validate such an overlap scheme, relative time delays Δt of opposite signs between PRE and POST signals were applied by keeping PRE spike and shifting the POST spike. Whereas Fig. 9a depicts the case for Δt = 0, Fig. 9b shows the overlapping spikes for a positive delay (Δt = 20 ms) evidencing that the net voltage across synaptic device given by V_{pre}-V_{post} crosses the minimum voltage threshold, thus leading to the increase of synaptic weight. Otherwise, if relative delay is negative (Δt = -40 ms), the voltage subtraction across PCM results in a single pulse of amplitude higher than the minimum voltage threshold, thus activating depression process (Fig. 9c). Based on these particular cases, the application of variable delay values ranging from -40 ms to 40 ms allowed Kuzum et al to achieve STDP capability at device level. This is confirmed by STDP measurements shown in Fig. 9d where the resulting STDP curve exhibits a nice agreement with biological data presented in [19]. This approach also offers great flexibility enabling to tune time constant of measured STDP characteristics by changing amplitude and separation of pulses within PRE spike. Specifically, gradually decreasing the spacing between consecutive pulses such that the highest pulses within each PRE sequence are the closest ones allows to reduce the time constants of STDP exponential curves, which are significant biological parameters marking synapses in the brain. The application of this scheme thus leads to the implementation of measured STDP characteristics for variable time constants |τ| between 10 ms - 30 ms shown in Fig. 9e, which supports the capability of PCM synapses of emulating various types of synapses with different biological functions. Finally, as shown Fig. 9f, the modulation of the order of pulses for potentiation and depression within PRE spike was also tested enabling to demonstrate two asymmetric and two symmetric different STDP kernel, thus paving the way to the possibility to build neuromorphic systems based on nanoscale memristive synapses increasingly approaching to the complex operation of human brain.

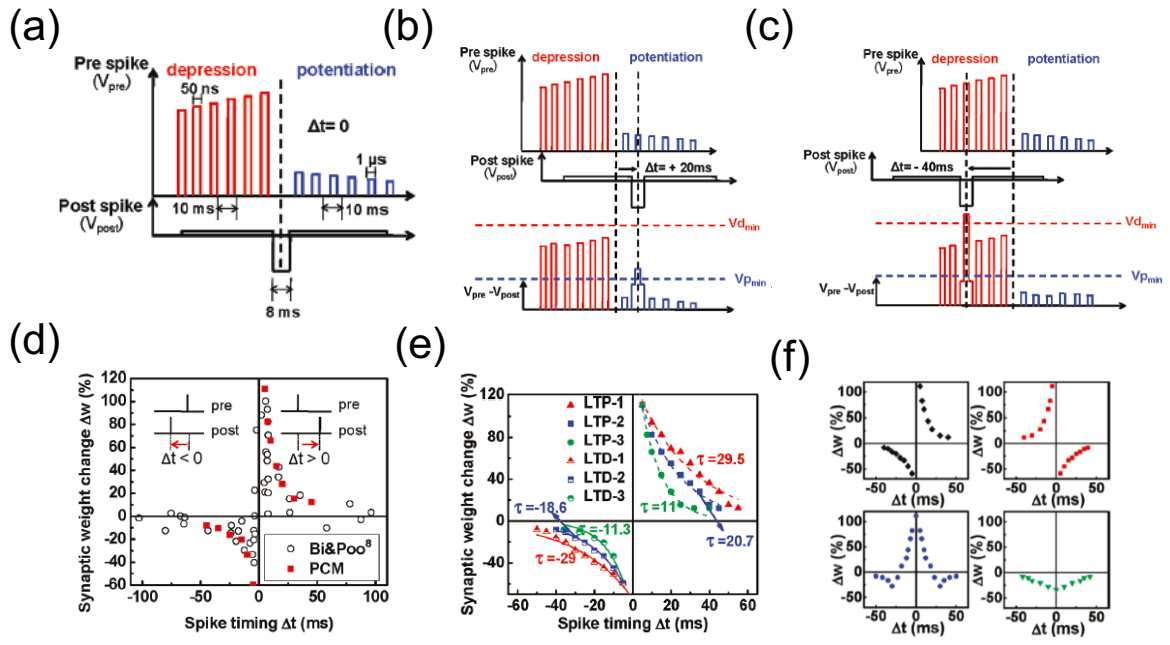

Figure 9. (a) Programming scheme based on overlap between pulses within PRE and POST spikes that is adopted to implement STDP in PCM synaptic device. (b) If relative delay is positive ($\Delta t = 20$ ms), spike overlap results in a voltage drop V_{PRE}-V_{POST} across PCM cell where a single 1-µs-long pulse can cross set threshold, thus inducing potentiation. (c) Conversely, if relative delay is negative ($\Delta t = -40$ ms), spike overlap results in a voltage drop V_{PRE}-V_{POST} across PCM cell where a single 50-ns-long pulse can overcome reset threshold, thus leading to synaptic depression. (d) STDP characteristics achieved by application of programming scheme on PCM cell against experimental data collected by Bi and Poo in [19]. (e) Measured STDP curves for variable time constants τ obtained tuning pulse amplitude/width within programming scheme. (f) Various asymmetric and symmetric STDP characteristics that can be implemented at device level changing the order of pulse sequences. Adapted with permission from [48]. Copyright 2012 American Chemical Society

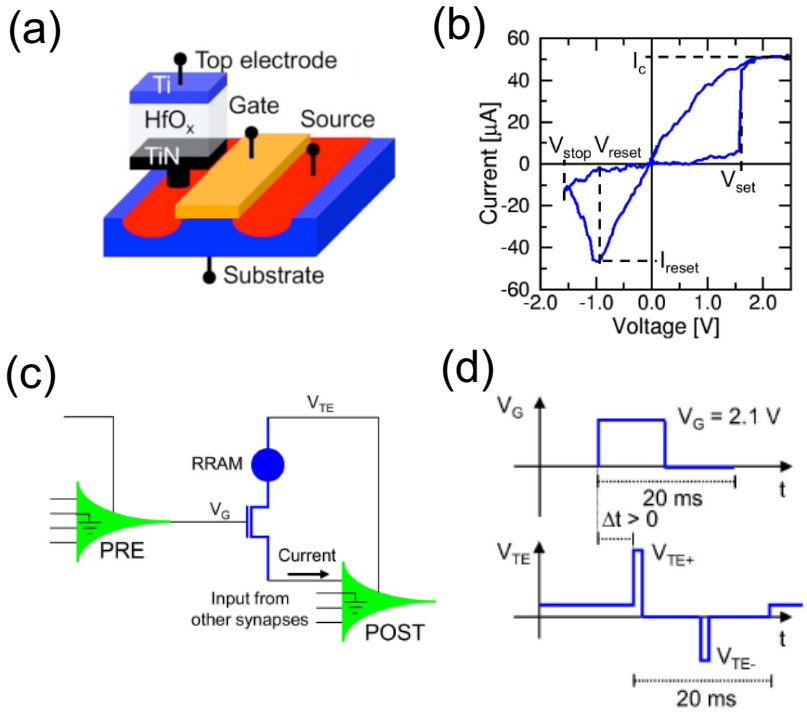

Figure 10. (a) Sketch of a 1T1R cell based on Ti/HfO$_x$/TiN RRAM device. (b) I-V characteristics of 1T1R RRAM structure. (c) Fundamental block using 1T1R cell as synaptic element connecting PRE neuron with POST neuron. (d) Programming strategy used to capture potentiation in 1T1R synapse: as Δt is positive, only positive pulse within POST spike applied to TE can overlap with PRE spike applied to the gate, thus activating a set transition, hence a weight change from HRS to LRS. Adapted from [55].

4. Hybrid CMOS/memristive synapses

4.1 1T1R synapses

Although single-element memristive synapses offer the prospect to build extremely dense neuromorphic circuits, their use in crossbar arrays however can lead to significant concerns such as leakage currents due to sneak paths and high-power consumption caused by the lack of current limiters. To bypass these issues while keeping relatively high integration density, a technological solution extensively adopted in recent years has been the use of a field effect transistor (FET) in series to the memristor device, which led to the development of hybrid CMOS/memristive synaptic structures such as the one-transistor/one-resistor (1T1R) [49-55].

Figure 10 shows a 1T1R structure based on serial connection of a FET to a Ti/HfO_x/TiN RRAM cell (a) and its I-V characteristic (b), which clearly evidences the current limitation to I_C = 50 µA during set transition achieved by FET. To operate such structure as an electronic synapse, the circuit scheme of Fig. 10c can be adopted [55]. According to this implementation, the PRE drives the gate terminal of FET, thus enabling synapse activation only as PRE spike occurs, whereas the POST controls the TE voltage V_{TE} which is generally set at low constant voltage to allow for communication between PRE and POST via the synapse. In this phase, the application of a PRE spike at FET gate when TE is biased at communication voltage induces a current proportional to the synaptic conductance across device being collected along with all the currents triggered by other activated PREs at the input of POST. Then, the sum of these currents is integrated by POST causing an increase of its internal potential until it exceeds a threshold eventually leading to the emission of a fire spike by POST which is delivered at TE to update the synaptic weight according to STDP rule. If the relative delay Δt between the PRE spike, which was designed as a 10-ms-long pulse of amplitude 2.1 V followed by a zero period of 10 ms, and the POST spike, which was designed as a 1-ms-long positive pulse followed by 1-ms-long negative pulse after a zero period of 10 ms, is positive, only the short positive pulse of amplitude $V_{TE+} > V_{set}$ within POST spike overlaps with PRE spike, thus inducing a set transition in RRAM cell resulting in the potentiation of synaptic weight (Fig. 10d). Conversely, if Δt is negative, only short negative pulse of amplitude $V_{TE-} < V_{reset}$ in the POST spike takes place at TE during PRE spike, thus causing a reset transition in RRAM cell leading to depression of synaptic weight.

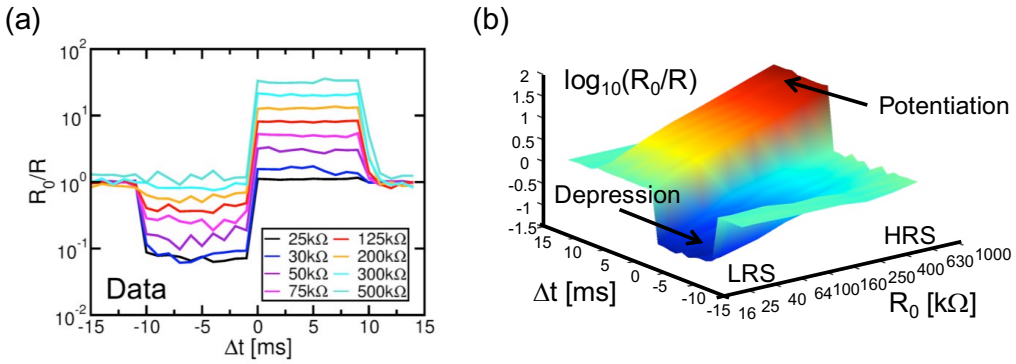

Figure 11. (a) Measured STDP characteristics achieved in LRS 1T1R RRAM device for variable initial state from HRS to LRS. (b) Color plot of experimental STDP implemented in 1T1R RRAM cell. Adapted from [55].

This synaptic operation scheme was validated by the measurements shown in Fig. 11a evidencing relative change of conductance in a single 1T1R synapse as a function of Δt for variable initial state from the full LRS (R_0 = 25 kΩ) and full HRS (R_0 = 500 kΩ). These characteristics first show that the more resistive is the initial state, the higher is the weight change via potentiation event, and the less resistive is the initial state, the higher is the weight change via depression event. Also, note that although the measured STDP characteristics show the synaptic potentiation/depression for positive/negative delays as expected by STDP biological protocol, their behavior is however uniform within overlap window of $|\Delta t|<10$ ms for any initialization because of binary nature of RRAM device which makes that the positive pulse at TE leads always device in full LRS set by I_C via V_G whereas negative pulse leads always device in full HRS, irrespective of Δt. This is also confirmed by color plot of measured STDP characteristics shown in Fig. 11b where the maximum potentiation for positive Δt is achieved starting from HRS whereas the maximum depression for negative Δt is obtained as the initial state is programmed in LRS. In addition to 1T1R RRAM synapses, 1T1R synaptic structures including PCM cell as memristive element have also been

investigated [49, 51, 54]. In this frame, Bichler et al devised the so-called 2-PCM synapse shown in Fig. 12a which is capable of implementing potentiation and depression by two 1T1R PCM structures referred to as LTP cell and LTD cell, respectively, using in both cases chalcogenide crystallization process [51]. In this way, a significant power saving due to the non-use of reset pulses at high current (hundreds of μA) for depression phase can be achieved. Also, since the progressive crystallization of chalcogenide active layer is carried out by application of sequences of voltage pulses with the same amplitude, pulse generation is easier than scheme adopted in [48]. In terms of functionality, this synaptic structure was used to capture a simplified STDP characteristics shown in Fig. 12b, according to which synaptic potentiation can occur only for a specific range of positive time delays between PRE and POST spikes of length T_{LTP}. In particular, to demonstrate this weight update rule, the pulse scheme for write operations schematically described in Fig. 12c was designed. Based on this scheme, as the total current integrated by an output neuron hits the threshold, the output neuron emits a POST spike being sent to all the input neurons by triggering write mode. During this phase, if an input neuron applies a single positive pulse called LTP pulse of amplitude V_{WR} at source of FET within 1T1R PCM structures means that the relative delay between PRE and POST falls in T_{LTP}, otherwise no signal is applied. In addition to this, the output neuron delivers at the same time voltage pulses of amplitude $-V_{WR}$ and $2V_{WR}$ at BEs of LTP PCM cells and LTD PCM cells, respectively, knowing that $V_{WR} < V_{set} < 2\ V_{WR}$. As a result, the conductance of LTP cells between firing input and output neurons increases since the total voltage across these devices is $2V_{WR} > V_{set}$, while the conductance of corresponding LTD cells remains unchanged because the total voltage across them is $V_{WR} < V_{set}$. Therefore, this involves that such 2-PCM synapses undergo synaptic potentiation in that the effective synaptic weight G_{LTP}-G_{LTD} increases. On the other hand, for synapses with no signal at source of FET, namely in all the cases with time delays outside LTP window, V_{WR} in absolute value drops on LTP cell and $2V_{WR}$ on LTD cell, which leads to conductance increase for LTD cell with unchanged conductance of LTP cell and consequently to the depression of those 2-PCM synapses. The application of this plasticity scheme however requires the execution of an additional refresh operation whenever the conductance of one of 2 cells within 2-PCM synapses saturates to the full LRS, which consists of a re-initialization in HRS of both devices followed by application of a series of set pulses to the LTP cell to restore the effective synaptic weight.

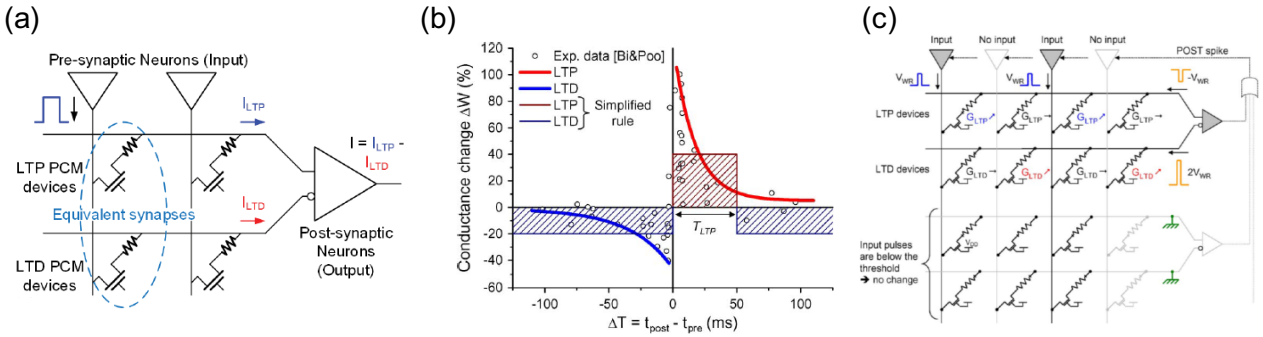

Figure 12. (a) Schematic representation of 2-PCM synapse whose weight is given by conductance difference between LTP device and LTD device. (b) STDP learning rule captured by 2-PCM synapse against biological STDP. (c) Programming algorithm used to implement potentiation and depression in 2-PCM synapses according to simplified STDP rule shown in (b). Adapted from [51].

4.2 2T1R synapses

Although very compact 1T1R synapses have been demonstrated to be capable of achieving neuromorphic applications such as visual pattern recognition via simplified STDP learning rules [49-55], more complex architectures are needed to gain higher flexibility and more detail in the emulation of biological processes. To this end, hybrid CMOS/memristive synaptic structure called two-transistor/one-resistor (2T1R) has recently been proposed using both RRAM device [56] and PCM device [57].

Figure 13a shows a 2T1R synapse with a TiN/HfO_x/TiN RRAM device which is serially connected to 2 transistors arranged with a parallel configuration [56]. To operate as electronic synapse, the communication gate (CG) of left transistor and RRAM TE are controlled by PRE, while the gate terminal of the right transistor, called fire gate (FG), and the RRAM BE are driven by POST integrate-and-fire circuit, which integrates all the synaptic currents activated by PREs via a brief pulse applied to CG during communication phase as long as a threshold is crossed, thus marking the generation of a fire spike. After the communication phase, which is performed by left path, this synaptic structure implements the plasticity phase, namely the weight update process, which is instead implemented separately exploiting the right path. Specifically, potentiation is achieved if the PRE voltage spike applied to the TE, which consists of the sequence of a negative 150-ms-long exponential pulse and very short (1 ms) positive pulse (top), anticipates ($\Delta t > 0$) the truncated positive exponential POST pulse applied to FG (center), in that their superposition results in very sharp current increase (bottom) inducing set transition of RRAM device (Fig. 13b). Conversely, as described by Fig. 13c, if POST spike precedes the PRE spike ($\Delta t < 0$), their overlap causes a reset transition within RRAM device leading to depression of 2T1R synapse. Applying the PRE and POST spikes at the 2T1R synapse with continuous change of Δt from -100 ms to 100 ms, its ability to capture bio-realistic analog behavior of potentiation and depression according to STDP was experimentally validated as evidenced by measured characteristics shown in Figs. 13d and Fig. 13e, respectively. In particular, note that a weak synaptic depression can also be obtained for very large positive Δt as a result of competition between the two synaptic processes. Importantly, this structure also offers an additional degree of freedom compared to 1T1R configuration namely the opportunity to change both potentiation characteristics (Fig. 13f) and depression characteristics (not shown) by proper tuning of time constant τ of FG voltage spike, which can serve as useful tool to replicate further biological phenomena.

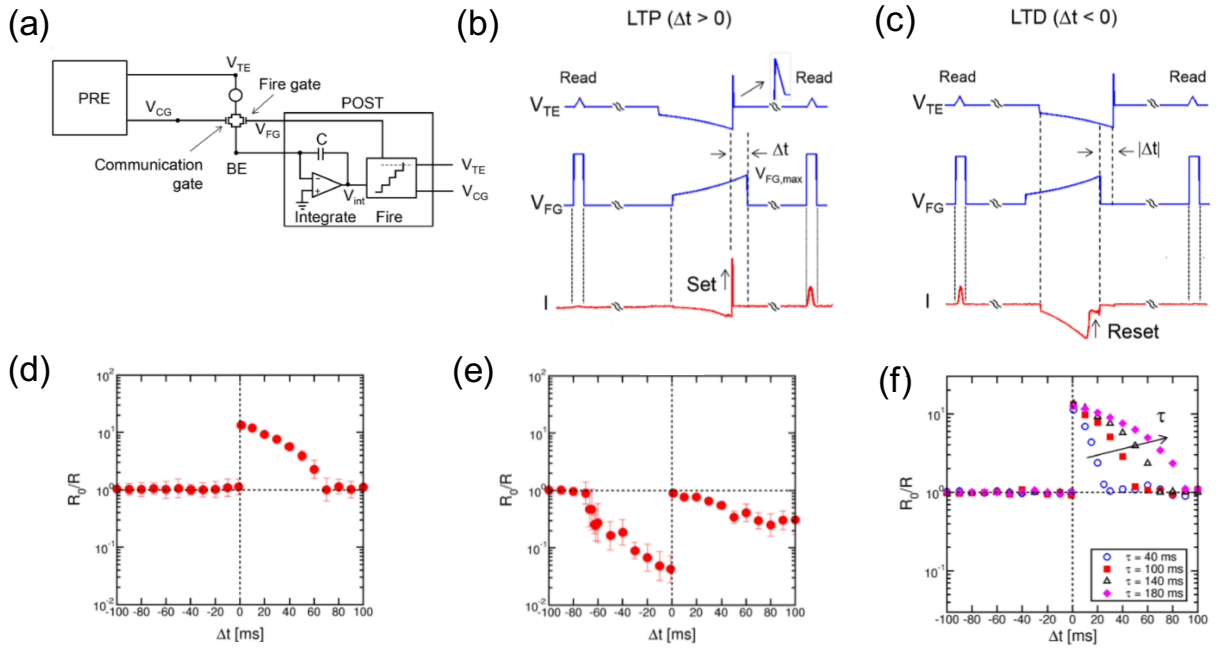

Figure 13. (a) Schematic representation of 2T1R RRAM synapse in PRE-synapse-POST circuit. Overlap between TE voltage and FG voltage triggering (b) set transition for RRAM device, hence potentiation for 2T1R synapse, in case of positive Δt and (c) reset transition for RRAM device, hence depression for 2T1R synapse, in case of negative Δt. STDP characteristics achieved by 2T1R RRAM structure for (d) potentiation and (e) depression, which can also occur for high positive Δt. (f) STDP characteristics under potentiation mode for variable time constant τ of FG pulse. Adapted from [56], which is distributed under CCBY.

Figure 14a shows an alternative 2T1R synapse using a PCM cell as memristive element [57]. Here, PCM cell is connected to the intermediate node between 2 transistors, called LIF transistor (top) and STDP transistor (bottom), respectively. This structure is connected to the PRE by the gate terminals of the LIF and STDP transistors, and to the POST by the LIF drain and the BE of the PCM device. Similar to the RRAM 2T1R synapse [56], two distinct paths were designed to achieve communication (LIF) and plasticity (STDP) operation modes, respectively. During LIF phase, which is explained in Fig. 14b, upon PRE spike, the LIF WL pulse generator included in the PRE circuit enables LIF transistor with STDP transistor turned off leading to the discharge of the capacitor of leaky-integrate-and-fire POST circuit as long as the voltage across the capacitor V_{cap} decreases below V_{th}. At that point, POST fires, activating after a time delay t_{delay} the STDP BL pulse generator which delivers a short positive pulse at the to BE of PCM cell. After LIF mode, the PRE circuit disables LIF transistor and activates the STDP transistor via a slowly-varying voltage signal emitted by STDP WL pulse generator, thus leading 2T1R synapse in STDP mode (Fig. 14c). In STDP mode, 2T1R synapse can update its weight according to STDP rule plasticity through the overlap of STDP BL pulse and STDP WL pulse. As shown in Fig. 14d, depression ($\Delta R > 0$) can be achieved for $t_{PRE} > t_{POST}$, namely as STDP BL pulse overlaps with increasing part of STDP WL signal since it induces high current programming PCM cell in HRS. Otherwise, potentiation ($\Delta R < 0$) can be achieved for $t_{PRE} < t_{POST}$, since in this case the overlap of STDP BL pulse and decreasing part of STDP WL signal results in a lower current leading PCM in LRS. Most importantly, this 2T1R synaptic implementation allows to capture the gradual nature of potentiation and depression dynamics via the properly designed STDP WL signal. This is confirmed by measured relative weight change as a function of Δt shown in Fig. 14e, which supports 2T1R synapse as valuable electronic synapse for neuromorphic applications.

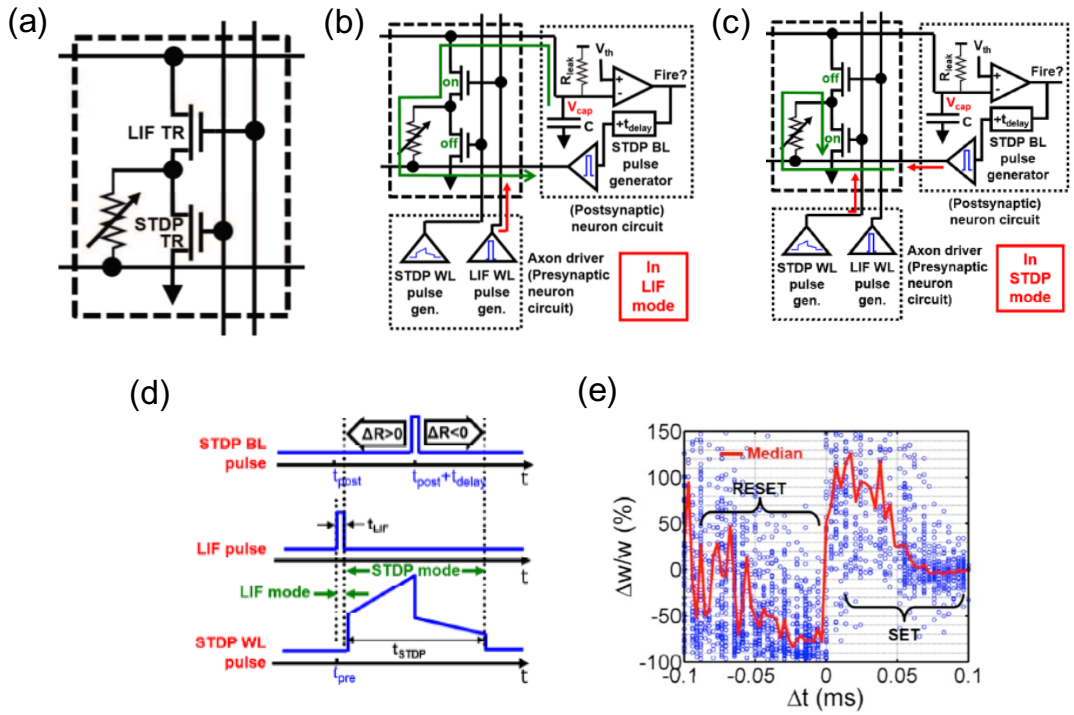

Figure 14. (a) Scheme of 2T1R PCM synapse where a transistor is used for leaky-integrate and fire phase (LIF TR) whereas the other one for weight update phase (STDP TR). (b) Schematic representation of 2T1R synapse operation during (b) LIF mode and (c) STDP mode. (d) Programming strategy used in 2T1R PCM synapse circuit to achieve potentiation and depression depending on timing of overlapping STDP BL pulse and STDP WL pulse. (e) Measured STDP characteristics demonstrated via 2T1R PCM synapse. Adapted from [57].

4.3 Differential synapses

As already discussed in Sections 4.1 and 4.2, the use of memristive devices such as RRAM and PCM devices in hybrid synaptic architectures involves a certain overhead in terms of complexity of structure and algorithm to capture biological behavior. First, these circuits need for the use of long overlapping spikes at PRE and POST terminals to trigger weight updates via atom configuration modifications, which results in significant reduction of data throughput in large-scale neuromorphic networks. Also, write operation of memristive devices governed by spike-based algorithms can require high programming currents, which has detrimental impact on power consumption and circuit size [58]. To tackle these severe issues featuring the majority of recently developed hybrid CMOS/memristive synapses, a novel memristive-based synaptic circuit was proposed in [58]. Such a synapse circuit exhibits a differential architecture based on 20 transistors and 2 HfO_2-based memristors, called D_{pos} and D_{neg}, respectively, being designed to store the synaptic weight in the conductance difference of two memristive devices. The operation scheme of this differential memristive synapse is divided into read and write mode phases. During read phase, synaptic weight at a given time can be tested measuring the currents flowing through memristive devices, and consequently the output currents, switching on only selectors controlled by read voltage V_{READ} while all the other transistors are turned off. Since one of the crucial goals for this novel circuit is to significantly reduce power consumption, all the switches in on-state were designed to work in sub-threshold regime. As a result, Nair et al demonstrate that, under certain bias conditions of transistors, the output currents measured during read phase, called I_{pos} and I_{neg}, are scaled versions of currents flowing through D_{pos} and D_{neg}, respectively. This positively affects not only power dissipation, but also on the area consumption in that a lower current entering POST allows to build POST circuits based on smaller capacitors and ultra-low power circuit elements. In addition, the reduced impact of memristor variability and the possibility to activate both excitatory and inhibitory currents represent additional advantages reachable through this differential synaptic structure based on no overlapping spikes at terminals. After read mode, write mode is sequentially turned on. This means that read signal is disabled ($V_{READ} = 0$) whereas write voltages V_{SET} and V_{RESET} are alternatively enabled according to whether the synaptic weight should be increased (potentiation) or decreased (depression), respectively. In case of high V_{SET}, D_{pos} undergoes a set transition and, simultaneously, D_{neg} undergoes a reset transition, thus leading to the increase of synaptic weight. Conversely, as V_{RESET} is high, D_{pos} is reset while D_{neg} is set, thus inducing a decrease of synaptic weight. To validate on-line learning ability of differential memristive synapses, learning simulations at network level have been implemented achieving significant performance in single pattern binary classification and multi-pattern classification.

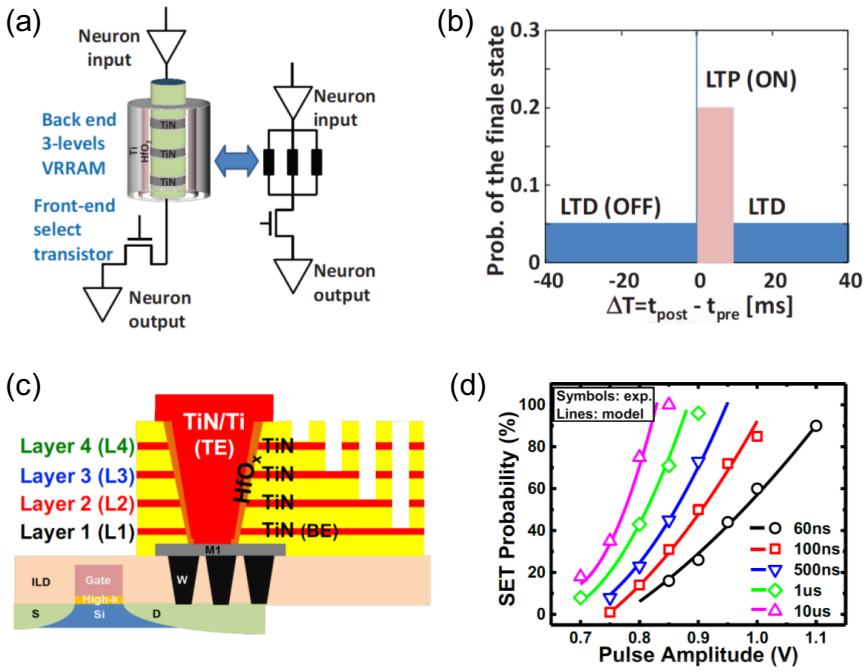

Figure 15. (a) Schematic representation of 3D 1T-nR synapse and (b) probabilistic STDP learning rule implemented at synaptic level**.** (c) Sketch of 4-layered 3D TiN/Ti/HfOx/TiN VRRAM synapse. Experimental and calculated behavior of set probability as a function of amplitude of applied pulse for increasing pulse width evidencing that the longer is the pulse, the lower can be the pulse amplitude to achieve set with high probability. (a) and (b) adapted from [60]. (c) and (d) adapted from [61].

4.4 1TnR synapses

Achieving complex cognitive functionalities performed by human brain is extremely challenging due to many reasons such as very low-power operation and unrivaled parallelism resulting from huge synaptic density. To emulate this latter feature in hardware, 2D crossbar arrays using very compact electronic synapses based on single-element or hybrid architectures have been extensively proposed without reaching, however, that of the human brain up to now. One of crucial reasons enabling the brain to host a huge number of synapses (~ 10^{15}) within an area lower than that of a shoebox [59], is the tridimensional architecture of the brain. Therefore, some novel hardware implementations of memristive synapses equipped with 3D architecture have been developed [40,60,61].

Figure 15a shows the vertical RRAM (VRRAM) structure presented in [60]. It consists of a stacked VRRAM, which includes a TiN/SiO_2 double layer with a TiN liner operating as BE surrounded by cylindrical-shaped HfO_2 switching layer and Ti-based TE, serially connected to a FET serving as selector and current limiter during set operation. This architecture allows to build a 1T-nR structure which, thanks to the multiple binary RRAM devices connected in parallel configuration, exhibits a conductance changing with gradual dynamics. In particular, it evidenced a strong potential as electronic synapse in auditory pattern extraction applications enabling to implement a simplified stochastic STDP-based learning rule similar to that proposed in [50], which is shown in Fig. 15b, via intrinsic variability of set and reset processes in RRAM elements. Another hardware implementation of 3D hybrid CMOS/memristive synapse was proposed in [61]. Its architecture, which is shown in Fig. 15c, evidences a four-layer 3D VRRAM, which includes a TiN/Ti layer as common TE, a HfO_x film as switching layer and 4 TiN layer as BEs, integrated with a p-channel FinFET operating as 3D selector. To implement synapses capable of stochastic learning, the intrinsic switching variability within RRAM was exploited. Specifically, as shown in Fig. 15d, set probability was characterized in experiments and simulation evidencing that the proper design of

applied pulses in terms of duration and amplitude can play a key role to optimize learning performance depending on the type of neuromorphic application.

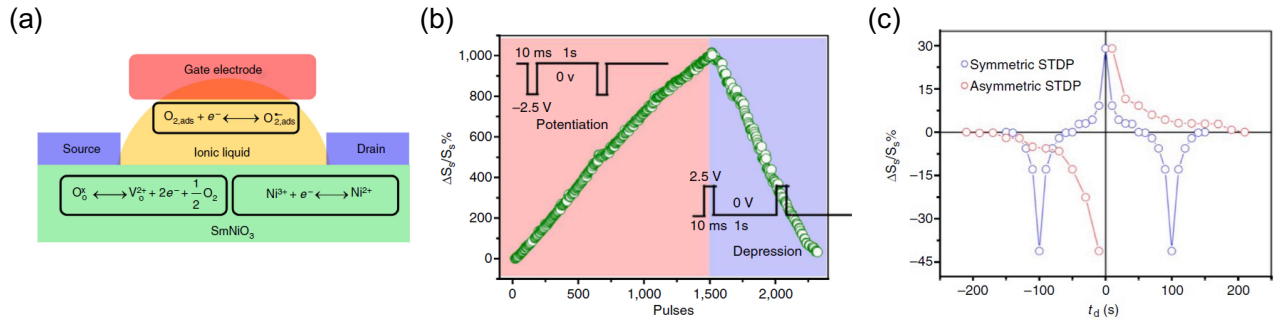

Figure 16. (a) Sketch of a nickelate $SmNiO_3$-based synaptic transistor capable of resistive switching from metal to insulator state and vice versa by electrochemical reactions (Ni reduction/oxidation) induced by application of positive/negative voltage pulses gating ionic liquid. (b) Calculated conductance response of three-terminal synapse as a result of application of an increasing number of negative pulses (potentiation) and positive pulses (depression). (c) Calculated symmetric and asymmetric STDP characteristics achieved implementing a PRE-POST delay-voltage conversion by a multiplexer. Adapted from [62].

5. Synaptic transistors (3-terminal synapses)

An important limitation for 2-terminal memristive synapses is that their operation relies on separation between communication and learning phases. However, it could be solved by adoption of new attractive transistor-based electronic synapses referred to as three-terminal synapses [62-64]. In [62], a three-terminal transistor device with a $SmNiO_3$ (SNO)-based channel (Fig. 16a) has been investigated in simulation demonstrating to be able to emulate STDP rule. The operation of this correlated nickelate synaptic transistor is based on resistive switching mechanism due to modulation of SNO perovskite stoichiometry which is achieved via application of gate pulses to ionic liquid (IL). As positive voltage pulses are provided to IL, the electric field drives the oxygen outside SNO layer inducing the generation of oxygen vacancies within SNO channel, which are responsible for electrochemical reduction of Ni^{3+} to Ni^{2+}. As a result, a resistance increase within device is obtained. On the contrary, under negative gate pulses, a field-driven motion of oxygen ions toward SNO is triggered, thus leading to the annihilation of oxygen vacancies with consequent oxidation of Ni^{2+} in Ni^{3+} that results in a resistance decrease. Based on this operation, synaptic potentiation and depression transitions were tested in simulation evidencing, as shown in Fig. 16b, that a linear increase of relative change in sheet conductance up to a factor 10 is achieved as a sequence of gate pulses of amplitude -2.5 V and width 10 ms with 1-s-long interval is applied to device, whereas a linear decrease of sheet conductance change up to initial state is achieved as gate voltage pulses with opposite (positive) polarity are provided. This result is very interesting since it highlights the strong linearity of both transitions for synaptic weight update, thus making this three-terminal SnO synaptic transistor more suitable than other widely reported memristor devices such as PCMO-based RRAM [65], TiO_x/TiO_2 RRAM [66], Ag:a-Si RRAM [36] and AlO_x/HfO_2 RRAM [67] for neuromorphic applications, i.e. speech recognition and image classification, implemented by deep neural networks using supervised learning schemes as backpropagation algorithm. Importantly, Fig. 16c shows the ability of this synaptic transistor to capture a very bio-realistic replication of both asymmetric and symmetric STDP characteristics, which were achieved connecting drain and source terminals to a multiplexer capable of converting the relative delay t_D between POST and PRE spikes applied to drain and source ($t_D = t_{POST}-t_{PRE}$) into a 10-s-long voltage pulse of proportional amplitude for gating operation.

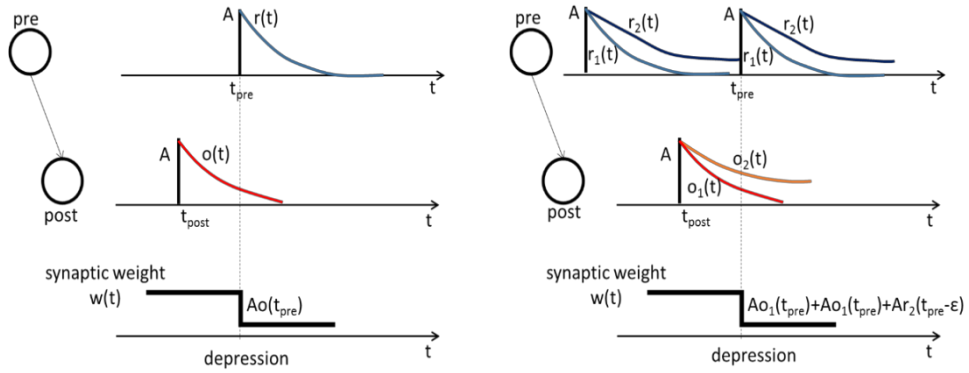

Figure 17. Schematic representation of the pair-based (left) and of the triplet based STDP rules (right). Synaptic weight change (depression event) is evidenced. Adapted from [68].

6. Triplet-based synapses

Pair-based synaptic modulation has been a staple in the implementation of neuromorphic computing systems capable of learning. This is owed to the algorithm's simplicity in comprehension and realization. However, beyond experiments where synaptic efficacy is measured after pairs of pre- and post-synaptic spikes, as a function of their relative timing, the plasticity rule fails to replicate the results of more complicated experiments. In particular the relationship between the frequency of spike pairing, where synaptic efficacies change more for higher pairing frequency, and for sequences of three (triplets) or four (quadruplets) spikes. This is believed to result from an asymmetry in the impacts of the spike timings of the pre- and post-synaptic cells in favour of the post-synaptic one. In order to break this symmetry, extensions of the pair-based algorithms have been proposed and are often termed triplet (or quadruplet) rules harking back to the experiments which motivated their development [25]. Typical pair-based STDP rules make use of one local variable each at the pre- and post-synapse which exponentially decay in time with the weight change being a function of the two states: 'o' represents the exponentially decaying post-synaptic variable, while 'r' denotes the pre-synaptic variable in Fig. 17. The values of these local variables can be thought of as being 'stamped' in time giving the famous form of the synaptic weight change expression (Δw), as a function of the spike times:

$$\Delta\omega(t_{pre}, t_{post}) = \begin{cases} Ae^{\left(\frac{t_{pre}-t_{post}}{\tau}\right)}, & t_{post} > t_{pre} \\ -Ae^{\left(\frac{t_{post}-t_{pre}}{\tau}\right)}, & t_{pre} > t_{post} \end{cases} \quad (1)$$

where, A is amplitude of the maximum synaptic efficacy change, t_{pre} is the timestamp of the last pre-synaptic spike, t_{post} is the timestamp of the last post-synaptic spike, τ is time constant of the decay from maximum synaptic change to zero change, w is the synaptic efficacy. As an extension triplet-based rules make use of an extra exponentially decaying variable per pre- and post-synapse and explicitly use their value in time to update the synaptic weight. These synaptic variables are stepped by a constant value when their respective neuron fires taking. This can be written as follows:

$$\frac{dx(t)}{dt} = \frac{-x(t)}{\tau}, at\ the\ moment\ of\ spike\ arrival\ (t = t_{pre}\ or\ t = t_{post}), x \rightarrow x + 1 \quad (2)$$

In the formulation for the triplet rule, each presynaptic spike t_{pre} induces an increase of two pre-synaptic variables, r_1 and r_2, and each postsynaptic spike t_{post} induces an increase of other two post-synaptic variables, o_1 and o_2. All these variables, o_1, o_2, r_1 and r_2 follows Eq. (2) where the time constant for each variable is independent as in Fig. 17. Using these four synaptic time dependent variables, Eq. (3) describes the triplet rule synaptic updates where the ratios between the time constants of $o_{2/1}$ and $r_{1/2}$ introduce the asymmetry in favor of the post-synapse.

$$\Delta\omega(t) = \begin{cases} -o_1(t)(A_2^- + A_3^- r_2(t-\varepsilon)), if\ t = t_{pre} \\ r_1(t)(A_2^+ + A_3^+ o_2(t-\varepsilon)), if\ t = t_{post} \end{cases} \quad (3)$$

where o_1 and o_2 are the post-synaptic variables which vary in time as described in Eq. (2), r_1 and r_2 are the pre-synaptic variables which vary in time as described in Eq. (2), A_2 is the maximum amplitude of change resulting from pairing of two spikes as in standard STDP, A_3 is the maximum amplitude of change resulting from pairing of three spikes extending the original STDP update to triplet STDP. Note that for the case of setting the constants A_3 to zero equation (3) assumes an alternate form of Eq. (2) where the local variables are explicitly written instead of the spike time. It is therefore important to realize that triplet STDP is not a novel rule but a higher order extension of pair-based STDP – analogous to using a higher order function to better fit data. Like higher order fitting, value should only come from developing an understanding of how to correspond these pre- and post-synaptic variables to real chemical variables inside the cells like somatic calcium concentration or that of synaptic glutamate. This work has motivated the development of synapses capable of implementing triplet learning algorithms for neuromorphic computing systems [68]. The work is based on the assumption that a resistive memory follows a behavioral model: the resistance of the device decreases exponentially if the applied voltage to the two terminal of the device (Δv) is higher than a given threshold (v_{th}), while it increases exponentially if the applied voltage is lower than $-v_{th}$, as described by Eq. (4)

$$f(\Delta v) = \begin{cases} I_0 \times \Delta v \left(e^{\frac{\Delta v - v_{th}}{v_0}} \right), |\Delta v| > |v_{th}| \\ 0, |\Delta v| < |v_{th}| \end{cases} \quad (4)$$

where $f(\Delta v)$ is a function which returns a change in the current passing through the resistive memory given an applied voltage Δv, I_0 and v_0 are two fitting parameters. Since the synaptic variables are exponential functions of time, a parallel exists with the exponential dependence on applied voltage of the resistance. It is then possible to use two resistive memories per triplet rule synapse whose superposition encodes the total synaptic weight (Fig. 18). One memory codes for the base-pair change as in standard STDP and the other for the extra change that results from the triplet rule. It is possible to simplify the triplet algorithm by removing the higher order change during presynaptic events, at the expense of slightly less biological correspondence, as in the spike-time dependent form written in Eq. (5).

$$\Delta\omega(t_{pre}, t_{post}) = \begin{cases} -Ae^{\left(\frac{t_{post} - t_{pre}}{\tau_1}\right)}, & if\ t = t_{pre} \\ Ae^{\left(\frac{t_{pre} - t_{post}}{\tau_2}\right)} + Ae^{\left(\frac{t_{pre} - t_{post}}{\tau_2}\right)} \times e^{\left(\frac{t_{post(n)} - t_{post(n-1)}}{\tau_3}\right)}, & if\ t = t_{post} \end{cases} \quad (5)$$

where A is the amplitude of maximum synaptic efficacy change, t_{pre} is the timestamp of the last pre-synaptic spike, t_{post} is the timestamp of the last post-synaptic spike, τ is the time constant of the decay from maximum synaptic change to zero change, w is the synaptic efficacy, $t_{post}(n)$ is the most recent

post-synaptic spike time, $t_{post}(n-1)$ is the second most recent post-synaptic spike time. With suitably generated voltages, which are a function of spike events, their combination over the terminals of the simple circuit of Fig. 18 can result in changes to the two devices such that their superimposed weight changes in the manner of a triplet rule.

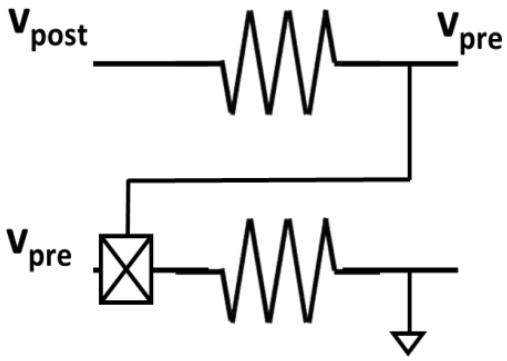

Figure 18. Two resistive memories synapse proposed in [68] to implement the triplet rule. The circuit is composed of two resistive memories and a multiplier/rectifier circuit shown as a crossed square. Adapted from [68].

7. SRDP synapses

7.1 1R synapses

In the human brain, crucial cognitive functionalities such as memory and learning are governed by complex synaptic mechanisms that are not yet fully understood. Some experimental studies such as the ones reported in [27,28] have revealed that, in addition to the timing of spikes underlying the well-known STDP learning rule, repetition rate of spikes also plays a key role in such processes. For this reason, the bio-realistic SRDP phenomenon taking into account the effect of spike rate on synaptic plasticity has attracted much attention to achieve a more faithful reproduction of synaptic behavior in hardware. Because of limitations due to the abrupt nature of resistive switching process in RRAM materials as [69], the implementation of SRDP at device level has required the exploration of alternative devices/structures such as single-element Ag_2S inorganic synapses [70], one-selector/1-resistor (1S1R) structures equipped with SiO_xN_y:Ag diffusive memristors [42] and second-order memristors as in [34].

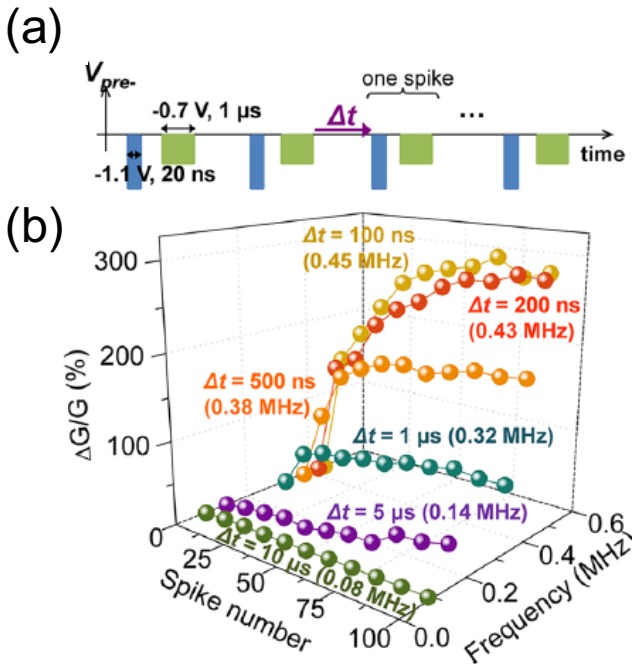

Figure 19. (a) SRDP implementation in a second-order Ta_2O_{5-x}-based memristor by application of series of set/heating pulses for variable time interval Δt. (b) Measured SRDP characteristics as a function of number of applied spikes with decreasing Δt from 10 µs to 100 ns. Adapted with permission from [34]. Copyright 2015 American Chemical Society

In [34], rate-based potentiation process was experimentally studied applying to the TE of Ta_2O_{5-x} RRAM device with grounded BE a sequence of PRE spikes, which consist of a negative 20-ns-long set pulse of amplitude -1.1 V followed by a 1-µs-long pulse of amplitude -0.7 V for heat generation, separated by time Δt as shown in Fig. 19a. In this manner, the shorter/longer is Δt, the stronger/weaker is the temporal heat accumulation effect on memristor conductance change already discussed in Section 3.1, which results in an increasing/decreasing synaptic potentiation. This is supported by SRDP characteristics for synaptic potentiation shown in Fig. 19b which evidence both an increase in conductance change for increasing number of applied spikes and a higher final weight for increasing stimulation frequency. Similar results were also obtained in rate-based synaptic depression experiments evidencing a stronger/weaker conductance decrease for high/low frequency stimulation of second-order memristor by programming pulses within PRE spikes with positive voltage polarity to reach reset transition (positive or negative polarity of heating pulses is unimportant). Therefore, these experimental results corroborate the ability of second-order memristors to implement another long-term plasticity bio-realistic rule as SRDP. Although the key role played by long-term plasticity in fundamental brain functionalities such as memory and learning has been supported by several biological experiments, the number of processes controlling real synaptic behavior is much wider and not yet totally understood. Among these additional effects, STP is one the most important since it enables to explain a crucial process at synaptic level as Ca^{2+} ion dynamics [31]. Motivated by experimental observations [29-31], significant solutions aiming at capturing STP by various memristive devices have been proposed in recent years [42, 70-72]. An interesting approach is the one presented by Werner et al in [71] where STP was implemented using non-volatile RRAM devices. To achieve STP, 10 Ti/HfO_2 RRAM cells were used in parallel to realize a single synapse and the programming scheme described in Fig. 20a was implemented. According to this scheme, every PRE spike applied to all RRAM TEs causes abrupt reset transitions within resistive synapse (weight decrease) which are followed by weak set transitions at each period ΔT with no input, thus gradually restoring the initial synaptic state.

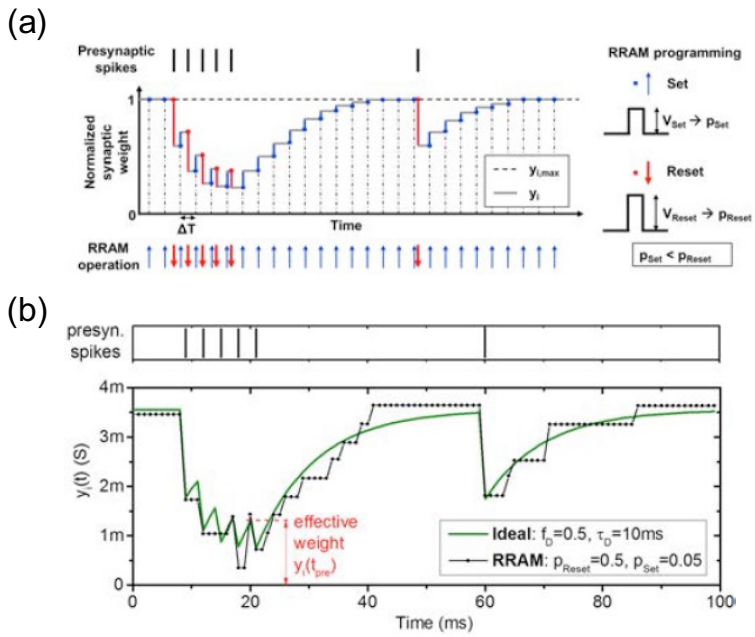

Figure 20. (a) Programming strategy used to achieve short-term plasticity (STP) in stochastic synapse based on 10 RRAM devices in parallel according to which each incoming PRE spike leads to abrupt depression and probabilistic set events can occur at each time slot ΔT with no external input, thus enabling to recover initial high conductance state. (b) STP implementation at experimental and simulation level based on pulse scheme shown in (a). Adapted from [71].

Based on this strategy, Fig. 20b shows the experimental and calculated evolution of synaptic weight y(t) as a function of time evidencing short-term changes which can be tuned controlling set/reset probabilities (p_{set} = 0.05 and p_{reset} = 0.5 in this case). Other significant approaches, however, enabled to achieve STP simply exploiting physical mechanisms underlying memristor operation, thus avoiding complex synaptic structures and programming schemes. In this context, particular focus should be attributed to Ag_2S-based inorganic synapses presented in [70] where STP is captured by spontaneous rupture of the metallic filament induced by low frequency spiking stimulation, and to diffusive SiO_xN_y:Ag memristor [42] which is capable of implementing short-term PPD and PPF, similarly to [72], thanks to diffusive dynamics of Ag ions in response to low frequency spike trains.

7.2 4T1R synapses

Because of the abrupt nature of resistive switching mechanism in many RRAM materials, most of RRAM devices do not enable to reproduce SRDP protocol unless complex synaptic structures and programming schemes are implemented. In this frame, a synapse circuit based on a hybrid CMOS/RRAM structure capable of SRDP functionality was presented in [73]. As shown in Fig. 21a, PRE and POST blocks are connected by a synaptic hybrid structure which is called 4-transistors/one-resistor (4T1R) synapse since it comprises one HfO_x RRAM device and two parallel branches each of which including a pair of FETs, M_1/M_2 for left branch and M_3/M_4 for right branch, in serial configuration. PRE block includes two spike generators emitting Poisson distributed asynchronous PRE spikes, which are applied to the gate of M_1 and, after being shifted by a delay Δt_D, to the gate of M_2, and PRE noise spikes driving the gate of M_3. In addition to PRE block,

POST block consists of an integrate-and-fire stage followed by a multiplexer (MUX) and an inverter. When external stimulation rate (f_{PRE}) is higher Δt_D^{-1}, the probability that M_1 and M_2 are simultaneously enabled by PRE spikes and their delayed copies is high, thus leading a current to flow across left branch. This current is integrated by POST and induces the emission of a fire pulse which is backward applied to the TE inducing a set transition, hence synaptic potentiation, as a result of a PRE-PRE-POST modified triplet-based weight scheme [25]. Also, note that the fire pulse, after being inverted by the inverter gate, is applied to the gate of M_4 disabling the right branch during potentiation mode. This means that the M_1/M_2 branch is the branch designed to capture synaptic potentiation. Instead, as f_{PRE} is much lower than Δt_D^{-1}, there is no chance that spike coincidences at inputs of potentiation branch occur. Therefore, a second branch based on M_3/M_4 pair was necessarily added in parallel to capture weight decrease at low f_{PRE}. To this aim, PRE block drives M_3 via PRE noise spikes at frequency $f_3 < f_{PRE}$ while POST, in addition to fire pulses, also emits random noise spikes at frequency $f_4 < f_{PRE}$ activating M_4 and TE. As these 3 random pulses overlap, the M_3/M_4 branch is enabled and a stochastic reset transition is triggered in RRAM device leading to a weight decrease, given the negative polarity of voltage pulse at TE. As a result, 4T1R synapse operation allows for SRDP algorithm by a selective synaptic potentiation for high-frequency spiking stimulation and a stochastic synaptic depression for low-frequency spiking stimulation using biologically inspired stochastic noise spikes emitted by PRE and POST [74]. The ability of 4T1R synapse circuit to implement high-frequency potentiation and low-frequency depression was validated in experiments separately studying potentiation and depression operation modes via 2T1R integrated structures.

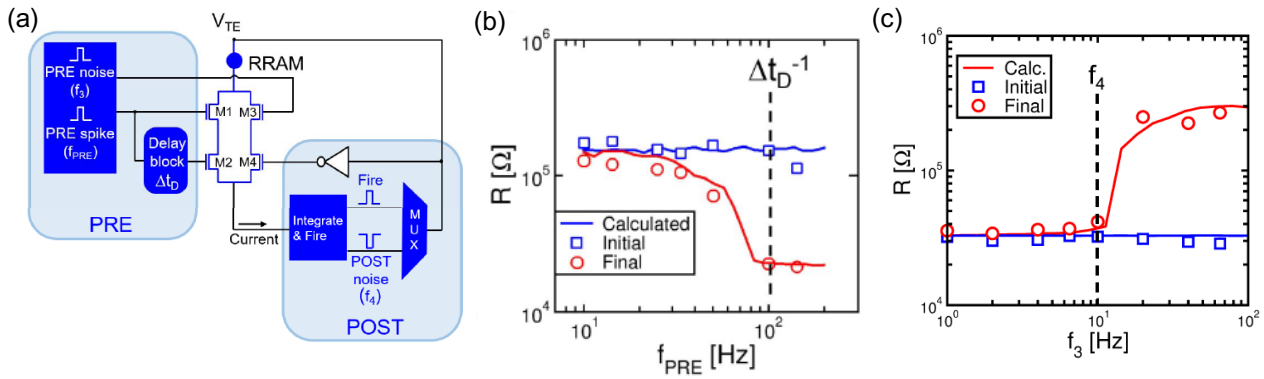

Figure 21. Schematic representation of hybrid 4T1R RRAM synapse capable of replicating SRDP bio-realistic rule. Experimental demonstration of (b) synaptic potentiation for $f_{PRE} > \Delta t^{-1}$ and (c) synaptic depression for $f_3 > f_4$ in case of $f_{PRE} << \Delta t^{-1}$. Adapted from [73].

As shown in Fig. 21b, given a delay Δt_D = 10 ms, resistance change from HRS to LRS in RRAM device, hence potentiation, can be achieved only for f_{PRE} >= 100 Hz, that is Δt_D^{-1}, thus supporting high frequency potentiation. On the other hand, a resistance transition from LRS to HRS in RRAM device can be triggered by PRE and POST noise spikes provided that $f_3 > f_4$, as supported by Fig. 21c where f_4 was set to 10 Hz. This result also confirms the feasibility of stochastic depression, and consequently SRDP, in 4T1R RRAM synapses.

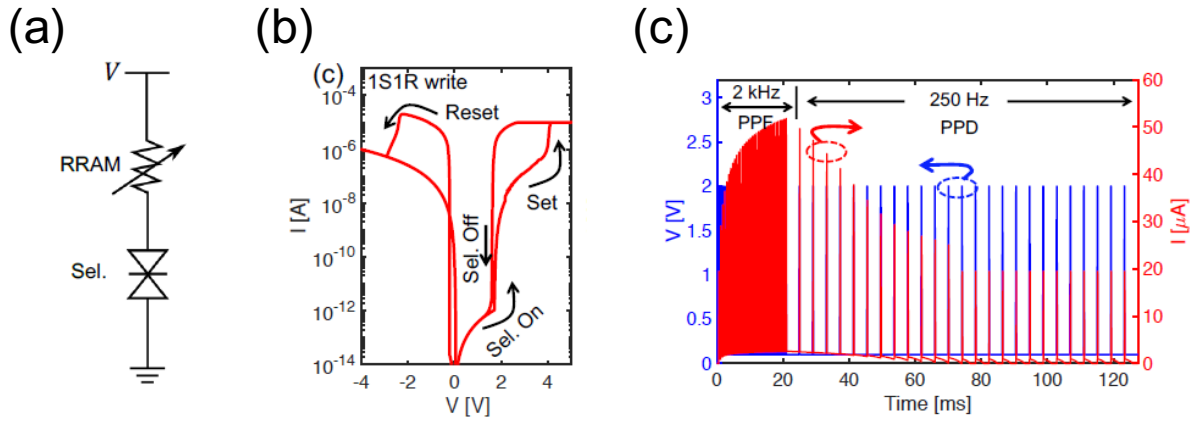

Figure 22. (a) Scheme a 1S1R structure obtained combining a non-volatile RRAM device with a volatile RRAM select device and (b) its I-V characteristics. (c) Current response of 1S1R structure evidencing SRDP capability via paired-pulse facilitation (PPF) for high-frequency spiking stimulation and paired-pulse depression (PPD) for low-frequency spiking stimulation. Adapted from [72].

7.3 1S1R synapses

In parallel to hybrid CMOS/RRAM structures capable of mimicking synaptic behavior using non-volatile resistive switching phenomenon in various RRAM devices such as 1T1R cell (Section 4.1) and 2T1R cell (Section 4.2), other attractive hybrid structures based on memristor devices were also intensively explored to further approach a more detailed replication of biological dynamics. Among them, strong interest was gained by 1S1R structure using RRAM devices based on material stacks such as SiO_xN_y:Ag [42], Ag/SiO_x [75] and Cu/SiO_x [75] showing volatile resistive switching as a result of spontaneous retraction of metallic filaments within a short retention time in the range from few μs to few ms. In [72], volatile switching of Ag/SiO_x RRAM within 1S1R structure was investigated by extensive simulations to reproduce SRDP at synaptic device level.
Figure 22 show the scheme of a 1S1R cell based on a non-volatile RRAM device serially connected with a volatile RRAM selector (a), and its corresponding I-V characteristics calculated by combined use of two physics-based analytical models for non-volatile RRAM [76] and volatile RRAM [72] (b). Figure 22c shows the current for a 1S1R device in response to different spiking stimulation regimes. In particular, it should be noted that the application of a spike train at high frequency (f_{spike} = 2 kHz) leads to a gradual current (conductance) increase thanks to the gradual growth of filament induced by spikes, which results in the so-called paired-pulse facilitation (PPF). On the contrary, under a low frequency spiking stimulation (f_{spike} = 250 Hz), conductance gradually decreases because the filament dissolution dominates on its growth, leading to another regime known as paired-pulse depression (PPD). The implementation of these two processes thus suggests the ability of volatile RRAM devices in 1S1R cell to capture biologically inspired SRDP algorithm with the added value, compared to the 4T1R RRAM synaptic structure proposed in [73], to gain a significant area saving making it very promising for building of dense crosspoint synaptic networks capable of brain-inspired cognitive functionalities.

8. Self-learning networks with memristive synapses

In recent years, we have seen a boost in the performance and applications of machine learning (ML), driven by several factors: (i) the availability of large data sets for training and models; (ii) the increased computational power of modern computers (GPUs are an excellent match for ML thanks to the high degree of parallelization). Among the many fields of ML, Deep Learning (DL) is the most

popular. Deep neural networks fall into three classes of architectures: fully connected neural networks (FCNN), convolutional neural networks (CNN) and recurrent neural networks (RNN).

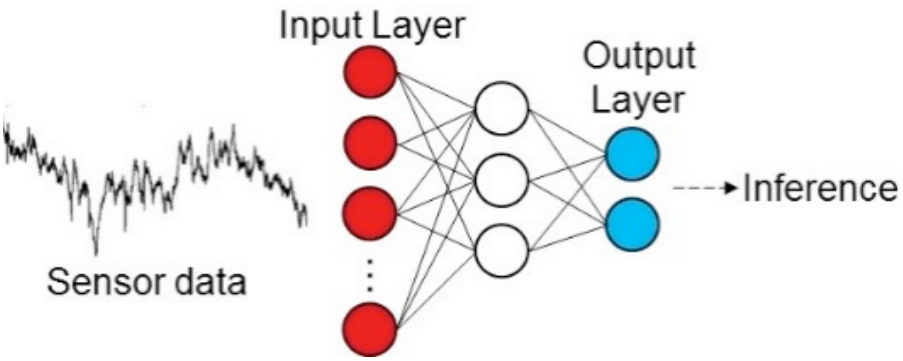

Figure 23. Example of two-layer Fully Connected Neural Network (FCNN).

As shown in Fig. 23, a FCNN is composed of fully-connected layers, each of which contain a collection of processing units (neurons) and weights (synapses). The neurons of a given layer are connected to every neuron of the previous layer by a large number of synapses. Raw data (e.g. video, audio, biological data…) initialize the values of the first layer (the input layer). The output layer corresponds to the inference classes (each output neuron is associated to a class of objects, e.g. dog, cat, car…). The number of weights and operations is directly proportional to the dimensions of the layers. On the other hand, CNN is composed of one or more convolutional layers, pooling or sub-sampling layers, and fully connected output layers (Fig. 24). In a convolutional layer a small set of synapses (constituting a kernel) allows subsequent network layers to extract spatially localized features before the information is subsampled and pooled and often used to drive further convolutional layers. The output of the convolutional layers (feature maps) contain information about the locations where features extracted by learned kernels are present in the input. The fully connected layer (classification module) is applied to complete the classification. Inference in CNN is identical to that of FCNN. The input data initializes the processing units of the first layer and the algorithm moves forward layer by layer. The activity of the processing units in the output layer correspond to the inferred classes as for the FCNN. CNN can achieve superb classification accuracy for image processing at much lower weight count than FCNN. Unlike FCNNs and CNNs, RNNs have loops enabling information to persist since the input at each step is composed of the data at that step in conjunction with the network output obtained at the previous step (Fig. 25). They are the natural architecture to use for sequential or temporal data. In the last few years there have been incredible success applying RNN to a variety of problems such as speech recognition, language modeling, translation. In particular, the well-known long-short-term memory (LSTM) RNN has recently found extensive application in text and speech recognition tasks. The pattern detection and classification in neural networks are the result of a training phase, by the repeated presentation of a training set and application of the learning rule, networks can learn to produce the correct responses to a set of inputs. In the last decades, new class of learning frameworks (such as supervised, unsupervised, reinforcement), with almost no resemblance to biological systems, have been developed in order to implement them in neural networks. After the training phase, the neural network infers things about new data (inference operation). During the inference operation, neural networks carry out enormous calculations of multiply accumulate (MAC) operation between weights and input data, and thus it needs high-performance hardware such as graphics processing unit (GPU). RRAM arrays are ideal to implement the MAC operation: the multiply operation is performed at every cross-point by Ohm's law, with current summation along rows or columns. Moreover, since are fabricated in the BEOL are increasingly attractive for high density, as they inherently lead to a benefit with respect to equivalent

SRAM macros. In addition, there is interest to use RRAM in more biologically inspired architectures and learning rules as presented in Section 3.

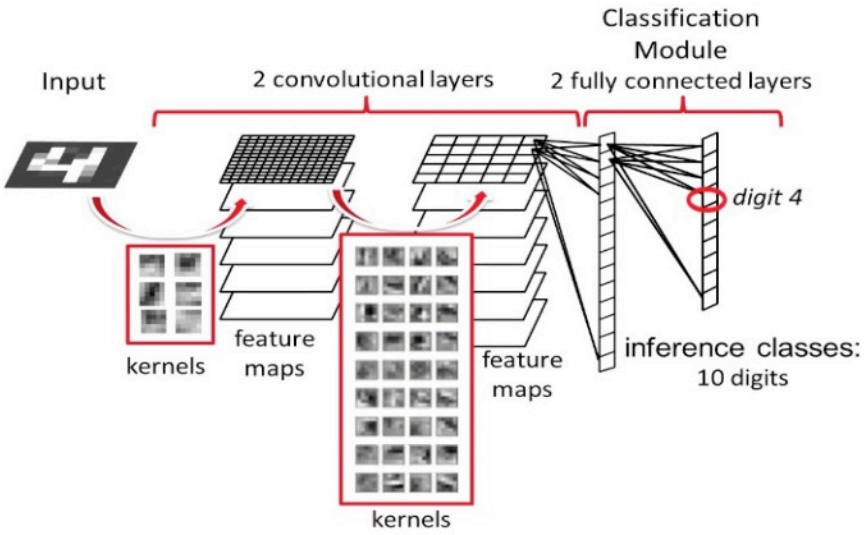

Figure 24. Schematic of Convolutional Neural Network (CNN) used for handwritten digits recognition (MNIST database).

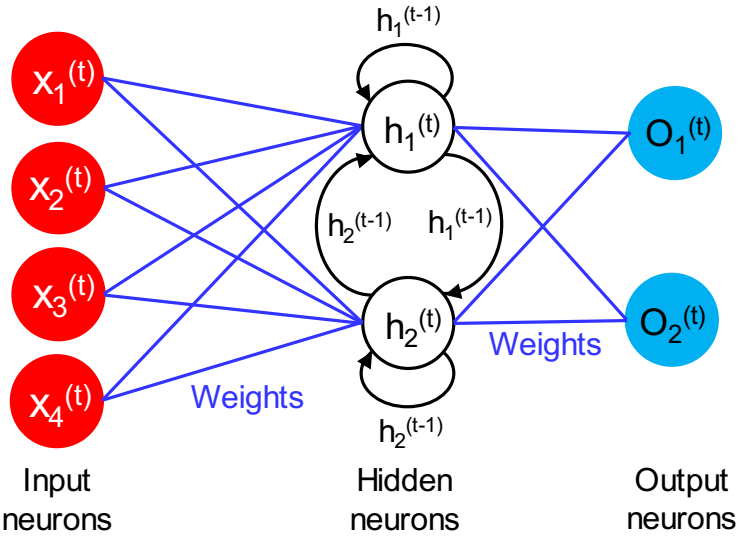

Figure 25. Sketch of a multilayer Recurrent Neural Network (RNN).

Hardware implementations of the inference operation in neuromorphic hardware have been presented in the literature [77-83]. A RRAM perceptron classifier implemented entirely in integrated hardware is presented in [81]. Multivalued resistance levels are stored in the RRAM cells. The test chip, 2M synapses integrated into 130nm CMOS, results in 90.8% MNIST recognition rate (ex-situ training). A small-scale perceptron classifier based on RRAM crossbar array board integrated with discrete CMOS components is presented in [82]. The network was trained both in-situ and ex-situ to perform classification of 4x4 pixel images.

Brain-inspired learning in spiking neural networks with RRAM synapses has been widely explored in recent years [84-96]. A perceptron-like neuromorphic hardware capable of STDP was presented in

[90]. This hardware network consists of a fully connected perceptron neural network (16 PREs and 2 POSTs) where all the PREs were connected with each POST by individual 1T1R RRAM synapses identical to the ones described in Section 4.1. Inhibitory synapses between the two POST neurons enable implementation of the well-known winner-take-all scheme [97] according to which the POSTs are not allowed to fire together in order to maximize storage capability of multiple visual patterns. The system was implemented on a PCB connecting an Arduino Due μC and synaptic elements with 1T1R integrated structure. The learning of two patterns has been experimentally demonstrated: two patterns and random noise were stochastically submitted to the 1st layer of the network. Noise submission induces depression within background synapses, thus allowing to 'forget' the previously learnt pattern when a new one is submitted. Noise is shown to decrease learning time and reduce the probability of 'false firing'. However, excessive noise results in unstable learning increasing the probability of 'false firing' [91,92].

In addition to hardware demonstration of ability to learn static visual patterns via STDP, the 1T1R RRAM synapses adopted in [90,91] were also used to connect 16 PREs with a single POST in a perceptron network in order to implement learning of spatiotemporal sequences [94]. To this end, PREs were subjected to the presentation of spatiotemporal patterns consisting of sequences of 4 spikes which were labeled as true/false patterns according to a teacher signal. Fig. 26a shows experimental demonstration of learning of spatiotemporal patterns in the same perceptron network with 1T1R RRAM synapses evidencing (top) the supervision signal and V_{int} measured in response to the sequence submission during training, (center) true fire, false fire and false silence spikes generated during the experiment, and (bottom) the color plot of potentiation/depression behavior of all the synaptic weights at increasing training cycle which suggests that 1-4-9-16 sequence was chosen as true spatiotemporal pattern. Fig. 26b shows some experimental results for recognition phase following training phase. Fig. 26b (top) shows that submission of true pattern allows V_{int} to cross voltage threshold, thus supporting the network ability to capture the true sequence learnt during training. In addition to this, as shown in Fig. 26b (bottom), the network is able to recognize false patterns submitted at input layer, as for instance 16-7-4-1 sequence, since V_{int} cannot hit the voltage threshold in these cases.

The role of synaptic variability (due to the intrinsic cell to cell and cycle to cycle variability) during unsupervised learning by STDP is investigated in [96] by means of system level simulations calibrated on the characterization of a 4kbit RRAM array. A fully connected feed-forward neural network topology with leaky integrate and fire neurons and RRAM-based synapses is adopted. A detection task in dynamic input data is investigated. The network is composed of one-layer fully connected network topology. The input layer is an image sensor composed of 128x128 spiking pixels, fully connected to an input layer of 60 neurons. The results are based on system level simulations, calibrated on the experimental data (measurements have been performed on a 4 kbit 1T1R array). The results demonstrate that, similarly to biology, SNNs are not only robust to variability but a certain amount of it can improve the network performance. More precisely the performance of the proposed application for real measured RRAM conductance distributions and an artificial device with zero variability are studied. For a given memory window at three standard deviations (3σ) in the cumulative conductance distribution (ratio between the high and the low conductance values at 3σ) of 2.25 the detection score is 0.63 for the artificial synapse with no variability and 0.952 for the real RRAM. Another way to improve the network performance is to increase the memory window. The increase of both conductance variability and memory window allows for an increase of the ratio between the conductance values of potentiated and depressed synapses, thus improving the learning accuracy.

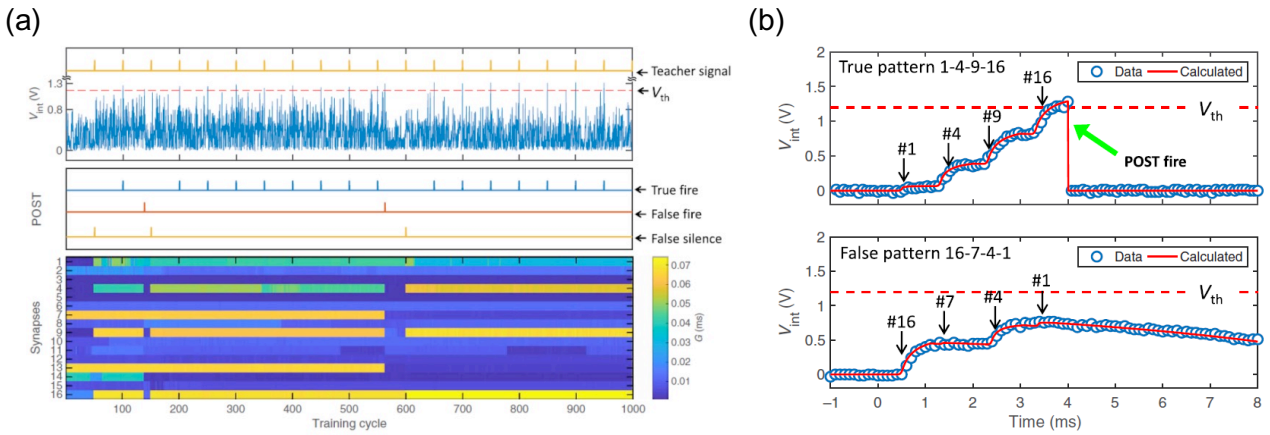

Figure 26. (a) Experimental training of spatiotemporal patterns captured by a perceptron network with 16 RRAM synapses. (top) Sequence of teaching spikes used as labels for true pattern and measured evolution of V_{int} during training phase. (center) True fire, false fire and false silence spikes occurring during training and (bottom) evolution of measured conductance for each synapse during training which evidences potentiation of weights associated to true pattern 1-4-9-16. (b) Experimental results for recognition phase evidencing that crossing of voltage threshold by V_{int} marks if (top) true sequence is effectively recognized or (bottom) if submitted sequence is a false pattern. Adapted from [94].

9. Conclusions

This chapter reviews the implementation of synaptic elements within neuromorphic hardware by using memory and memristive devices. RRAM and PCM synapses show analogue switching, scalable size, low voltage/power, thus offering a promising technology for both spiking and non-spiking neural networks for cognitive computing. To emulate the learning processes in the human brain, bio-inspired STDP and SRDP processes can be realized by using either overlap or non-overlap algorithms. The physics of RRAM devices can be used to naturally implement STDP and SRDP, e.g., by thermal effects or ionic diffusion at the nanoscale. By combining neuron and synapse elements within a neuromorphic circuit, learning and recognition functions can be achieved, thus allowing to benchmark CMOS and memristive technologies for cognitive computing.

10. Acknowledgments

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 648635).

11. References

[1] F. Rosenblatt, “The Perceptron: A perceiving and recognizing automaton,” Report 85-460-1, Cornell Aeronautical Laboratory, Buffalo, New York (1957).

[2] M.L. Minsky and S.A. Papert, “Perceptrons: An introduction to computational geometry,” The MIT Press, Cambridge MA (1972).

[3] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning," *Nature* **521**, pp. 436-444 (2015). DOI: 10.1038/nature14539

[4] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of IEEE* **86**, pp. 2278-2324 (1998). DOI: 10.1109/5.726791

[5] W. Maass, "Networks of spiking neurons: The third generation of neural network models," *Neural Networks* **10**(9), pp. 1659-1671 (1997). DOI: 10.1016/S0893-6080(97)00011-7

[6] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, and D. S. Modha, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science* **345**(6197), pp. 668-673 (2014). DOI: 10.1126/science.1254642

[7] E. Chicca, F. Stefanini, C. Bartolozzi, G. Indiveri, "Neuromorphic electronic circuits for building autonomous cognitive systems," *Proceedings of the IEEE* **102**(9), pp. 1367-1388 (2014). DOI: 10.1109/JPROC.2014.2313954

[8] D. Ielmini, "Resistive switching memories based on metal oxides: mechanisms, reliability and scaling," *Semicond. Sci. Technol.* **31**(6), 063002 (2016). DOI: 10.1088/0268-1242/31/6/063002

[9] S. Raoux, W. Welnic and D. Ielmini, "Phase change materials and their application to non-volatile memories," *Chem. Rev.* **110**, 1, pp. 240-267 (2010). DOI: 10.1021/cr900040x

[10] B. Govoreanu, G.S. Kar, Y-Y. Chen, V. Paraschiv, S. Kubicek, A. Fantini, I.P. Radu, L. Goux, S. Clima, R. Degraeve, N. Jossart, O. Richard, T. Vandeweyer, K. Seo, P. Hendrickx, G. Pourtois, H. Bender, L. Altimime, D.J. Wouters, J.A. Kittl, M. Jurczak , "10x10 nm^2 Hf/HfO_x crossbar resistive RAM with excellent performance, reliability and low-energy operation", in *IEEE International Electron Devices Meeting (IEDM)*, pp. 729-732 (2011). DOI: 10.1109/IEDM.2011.6131652

[11] F. Nardi, S. Larentis, S. Balatti, D. C. Gilmer and D. Ielmini, "Resistive switching by voltage-driven ion migration in bipolar RRAM – Part I: Experimental study," *IEEE Trans. Electron Devices* **59**(9), pp. 2461-2467 (2012). DOI: 10.1109/TED.2012.2202319

[12] F. Arnaud, *et al.*, "Truly Innovative 28nm FDSOI Technology for Automotive Micro-Controller Applications embedding 16MB Phase Change Memory," in *IEEE International Electron Devices Meeting (IEDM)*, pp. 424-427 (2018). DOI: 10.1109/IEDM.2018.8614595

[13] D. Ielmini and H.-S.P. Wong "In-memory computing with resistive switching devices", *Nature Electronics* **1**, pp. 333-343 (2018). DOI: 10.1038/s41928-018-0092-2

[14] C. Li, M. Hu, Y. Li, H. Jiang, N. Ge, E. Montgomery, J. Zhang, W. Song, N. Dávila, C.E. Graves, Z. Li, J.P. Strachan, P. Lin, Z. Wang, M. Barnell, Q. Wu, R.S. Williams, J.J. Yang, and Q. Xia, "Analogue signal and image processing with large memristor crossbars," *Nature Electronics* **1**, pp. 52-59 (2018). DOI: 10.1038/s41928-017-0002-z

[15] P.M. Sheridan, F. Cai, C. Du, W. Ma, Z. Zhang and W.D. Lu, "Sparse coding with memristor networks," *Nat. Nanotechnol.* **12**, pp. 784-789 (2017). DOI: 10.1038/nnano.2017.83

[16] M. Le Gallo, A. Sebastian, R. Mathis, M. Manica, H. Giefers, T. Tuma, C. Bekas, A. Curioni, E. Eleftheriou, "Mixed-precision in-memory computing," *Nature Electronics* **1**, pp. 246-253 (2018). DOI: 10.1038/s41928-018-0054-8

[17] Z. Sun, G. Pedretti, E. Ambrosi, A. Bricalli, W. Wang, and D. Ielmini, "Solving matrix equations in one step with crosspoint resistive arrays," *PNAS* **116**(10), pp. 4123-4128 (2019). DOI: 10.1073/pnas.1815682116

[18] D. Ielmini, "Brain-inspired computing with resistive switching memory (RRAM): Devices, synapses and neural networks," *Microelectron. Eng.* **190**, pp. 44-53 (2018). DOI: 10.1016/j.mee.2018.01.009

[19] G.-Q. Bi and M.-M. Poo, "Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and post synaptic cell type," *J. Neurosci.* **18**(24), pp. 10464-10472 (1998). DOI: 10.1523/JNEUROSCI.18-24-10464.1998

[20] D.O. Hebb, "The Organization of Behavior," New York: JohnWiley & Sons (1949).

[21] M. A. Woodin, K. Ganguly, M.M. Poo, "Coincident pre- and postsynaptic activity modifies GABAergic synapses by postsynaptic changes in Cl- transporter activity," *Neuron* **39**(5), pp.807-820 (2003). DOI: 10.1016/S0896-6273(03)00507-5

[22] Y. Luz and M. Shamir, "Balancing feed-forward excitation and inhibition via hebbian inhibitory synaptic plasticity," *PLoS Computational Biology* **8**(1), pp.1-12 (2012). DOI: 10.1371/journal.pcbi.1002334

[23] L. F. Abbott and S. B. Nelson, Synaptic plasticity: taming the beast', *Nat. Neurosci*. **3,** pp. 1178-1183 (2000). DOI:10.1038/81453

[24] T.P. Vogels, *et al.*, "Inhibitory synaptic plasticity: spike timing-dependence and putative network function," *Frontiers in Neural Circuits*, **7**:119 (2013). DOI: 10.3389/fncir.2013.00119

[25] J. Pfister and W. Gerstner, "Triplets of spikes in a model of spike timing-dependent plasticity," *Journal of Neuroscience* **26**(38), pp. 9673-9682 (2006). DOI: 10.1523/JNEUROSCI.1425-06.2006

[26] J. M. Brader, W. Senn, S. Fusi, "Learning real-world stimuli in a neural network with spike-driven synaptic dynamics," *Neural Computation* **19**(11), pp. 2881-2912 (2007). DOI: 10.1162/neco.2007.19.11.2881

[27] M. F. Bear, "A synaptic basis for memory storage in the cerebral cortex," *Proc. Natl. Acad. Sci. USA* **93**, pp. 13453-13459 (1996). DOI: 10.1073/pnas.93.24.13453

[28] P.J. Sjöström, G.G. Turrigiano and S.B. Nelson, "Rate, timing, and cooperativity jointly determine cortical synaptic plasticity," *Neuron* **32**, pp. 1149-1164 (2001). DOI: 10.1016/S0896-6273(01)00542-6

[29] M. F. Bear and R. C. Malenka. "Synaptic plasticity: LTP and LTD," *Curr Opin Neurobiol.* **4**(3), pp. 389-399 (1994). DOI:10.1016/0959-4388(94)90101-5

[30] H. Markram, Y. Wang and M. Tsodyks, "Differential signaling via the same axon of neocortical pyramidal neurons," *Proc Natl Acad Sci U S A*, **95**(9), pp. 5323-5328 (1998). DOI: 10.1073/pnas.95.9.5323

[31] R.S. Zucker and W.G. Regehr, "Short-term synaptic plasticity," *Annual Review of Physiology*, **64**(1), pp. 355-405 (2002). DOI: 10.1146/annurev.physiol.64.092501.114547

[32] S. N. Jung, A. Borst, and J. Haag, "Flight activity alters velocity tuning of fly motion-sensitive neurons," *Journal of Neuroscience*, **31**(25), pp. 9231-9237 (2011). DOI: 10.1523/JNEUROSCI.1138-11.2011

[33] Y.V. Pershin and M. di Ventra, "Neuromorphic, digital, and quantum computation with memory circuit elements", *Proc. IEEE* **100**(6), pp. 2071-2080 (2012). DOI: 10.1109/JPROC.2011.2166369

[34] S. Kim, C. Du, P. Sheridan, W. Ma, S.H. Choi and W. D. Lu, "Experimental demonstration of a second-order memristor and its ability to biorealistically implement synaptic plasticity," *Nano Lett.* **15**(3), pp. 2203-2211 (2015). DOI: 10.1021/acs.nanolett.5b00697

[35] G.S. Snider, "Spike-Timing-Dependent Learning in memristive devices," in *IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH 2008)*, pp. 85-92 (2008). DOI: 10.1109/NANOARCH.2008.4585796

[36] S.H. Jo, T. Chang, I. Ebong, B.B. Bhadviya, P. Mazumder, and W. Lu, "Nanoscale memristor device as synapse in neuromorphic systems," *Nano Lett.* **10**(4), pp. 1297-1301 (2010). DOI: 10.1021/nl904092h

[37] S. Yu, Y. Wu, R. Jeyasingh, D. Kuzum, and H.-S. P. Wong, "An electronic synapse device based on metal oxide resistive switching memory for neuromorphic computation," *IEEE Trans. Electron Devices* **58**(8), pp. 2729-2737 (2011). DOI: 10.1109/TED.2011.2147791

[38] K. Seo, I. Kim, S. Jung, M. Jo, S. Park, J. Park, J. Shin, K. P. Biju, J. Kong, K. Lee, B. Lee, and H. Hwang, "Analog memory and spike-timing-dependent plasticity characteristics of a nanoscale titanium oxide bilayer resistive switching device," *Nanotechnology* **22**(25), 254023 (2011). DOI: 10.1088/0957-4484/22/25/254023

[39] T. Serrano-Gotarredona, T. Masquelier, T. Prodromakis, G. Indiveri and B. Linares-Barranco, "STDP and STDP variations with memristors for spiking neuromorphic learning systems," *Front. Neurosci.* **7**:2 (2013). DOI: 10.3389/fnins.2013.00002

[40] I.-T. Wang, Y.-C. Lin, Y.-F. Wang, C.-W. Hsu, and T.-H. Hou, "3D synaptic architecture with ultralow sub-10 fJ energy per spike for neuromorphic computation," in *IEEE International Electron Devices Meeting (IEDM)*, pp. 665-668 (2014). DOI: 10.1109/IEDM.2014.7047127

[41] M. Prezioso, F. Merrikh Bayat, B. Hoskins, K. Likharev, and D. Strukov, "Self-adaptive spike-time-dependent plasticity of metal-oxide memristors," *Sci. Rep.* **6**:21331 (2016). DOI: 10.1038/srep21331

[42] Z. Wang, S. Joshi, S.E. Savel'ev, H. Jiang, R. Midya, P. Lin, M. Hu, N. Ge, J.P. Strachan, Z. Li, Q. Wu, M. Barnell, G.-L. Li, H.L. Xin, R.S. Williams, Q. Xia, and J.J. Yang, "Memristors with

diffusive dynamics as synaptic emulators for neuromorphic computing," *Nat. Mater.* **16**, pp. 101-108 (2017). DOI: 10.1038/nmat4756

[43] S. Yu, Y. Wu, and H.-S. P Wong, "Investigating the switching dynamics and multilevel capability of bipolar metal oxide resistive switching memory," *Appl. Phys. Lett.* **98**(10), 103514 (2011). DOI:10.1063/1.3564883

[44] S. Balatti, S. Larentis, D.C Gilmer, and D. Ielmini, "Multiple memory states in resistive switching devices through controlled size and orientation of the conductive filament," *Adv. Mater.* **25**(10), pp. 1474-1478 (2013). DOI: 10.1002/adma.201204097

[45] L. Zhao, H.-Y. Chen, S.-C. Wu, Z. Jiang, S. Yu, T.-H. Hou, H.-S. P. Wong, and Y. Nishi, "Multi-level control of conductive nano-filament evolution in HfO_2 ReRAM by pulse-train operations," *Nanoscale* **6**(11), pp. 5698-5702 (2014). DOI: 10.1039/C4NR00500G

[46] A. Prakash, J. Park, J. Song, J. Woo, E.-J. Cha, and H. Hwang, "Demonstration of low power 3-bit multilevel cell characteristics in a TaO_x-based RRAM by stack engineering," *IEEE Electron Device Lett*. **36**(1), pp. 32-34 (2015). DOI: 10.1109/LED.2014.2375200

[47] A. Athmanathan, M. Stanisavljevic, N. Papandreou, H. Pozidis, E. Eleftheriou, "Multilevel-cell Phase-Change Memory: A viable technology," *IEEE J. Emerging and Selected Topics in Circuits and Systems (JETCAS)* **6**(1), pp. 87-100 (2016). DOI: 10.1109/JETCAS.2016.2528598

[48] D. Kuzum, R.G.D. Jeyasingh, B. Lee, and H.-S.P. Wong, "Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing," *Nano Lett.* **12**(5), pp. 2179-2186 (2012). DOI: 10.1021/nl201040y

[49] M. Suri, O. Bichler, D. Querlioz, O. Cueto, L. Perniola, V. Sousa, D. Vuillaume, C. Gamrat, B. DeSalvo, "Phase change memory as synapse for ultra-dense neuromorphic systems: Application to complex visual pattern extraction," in *IEEE International Electron Devices Meeting (IEDM),* pp. 79-82 (2011). DOI: 10.1109/IEDM.2011.6131488

[50] M. Suri, O. Bichler, D. Querlioz, G. Palma, E. Vianello, D. Vuillaume, C. Gamrat, and B. De Salvo, "CBRAM devices as binary synapses for low-power stochastic neuromorphic systems: auditory (cochlea) and visual (retina) cognitive processing applications," in *IEEE International Electron Devices Meeting (IEDM),* pp. 235-238 (2012), DOI: 10.1109/IEDM.2012.6479017

[51] O. Bichler, M. Suri, D. Querlioz, D. Vuillaume, B. De Salvo and C. Gamrat, "Visual pattern extraction using energy-efficient '2-PCM synapse' neuromorphic architecture," *IEEE Trans. on Electron Devices*, **59**(8), pp. 2206-2214 (2012). DOI: 10.1109/TED.2012.2197951

[52] S. Ambrogio, S. Balatti, F. Nardi, S. Facchinetti, and D. Ielmini, "Spike-timing dependent plasticity in a transistor-selected resistive switching memory," *Nanotechnology* **24**, 384012 (2013). DOI: 10.1088/0957-4484/24/38/384012

[53] D. Garbin, E. Vianello, O. Bichler, Q. Rafhay, C. Gamrat, G. Ghibaudo, B. De Salvo, and L. Perniola, "HfO_2-based OxRAM devices as synapses for convolutional neural networks," *IEEE Trans. Electron Devices* **62**(8), pp. 2494-2501 (2015). DOI: 10.1109/TED.2015.2440102

[54] S. Ambrogio, N. Ciocchini, M. Laudato, V. Milo, A. Pirovano, P. Fantini and D. Ielmini, "Unsupervised learning by spike timing dependent plasticity in phase change memory (PCM) synapses," *Front. Neurosci*. **10**:56 (2016). DOI: 10.3389/fnins.2016.00056

[55] S. Ambrogio, S. Balatti, V. Milo, R. Carboni, Z. Wang, A. Calderoni, N. Ramaswamy, and D. Ielmini, "Neuromorphic learning and recognition with one-transistor-one-resistor synapses and bistable metal oxide RRAM," *IEEE Trans. Electron Devices* **63**(4), pp. 1508-1515 (2016). DOI: 10.1109/TED.2016.2526647

[56] Z.-Q. Wang, S. Ambrogio, S. Balatti and D. Ielmini, "A 2-transistor/1-resistor artificial synapse capable of communication and stochastic learning for neuromorphic systems," *Front. Neurosci.* **8**:438 (2015). DOI: 10.3389/fnins.2014.00438

[57] S. Kim, M. Ishii, S. Lewis, T. Perri, M. BrightSky, W. Kim, R. Jordan, G.W. Burr, N. Sosa, A. Ray, J.-P. Han, C. Miller, K. Hosokawa, and C. Lam, "NVM neuromorphic core with 64k-cell (256-by-256) phase change memory synaptic array with on-chip neuron circuits for continuous in-situ learning," in *IEEE International Electron Devices Meeting (IEDM),* pp. 443-446 (2015). DOI: 10.1109/IEDM.2015.7409716

[58] M.V. Nair, L.K. Muller and G. Indiveri, "A differential memristive synapse circuit for on-line learning in neuromorphic computing systems," *Nano Futures* **1**, 035003 (2017). DOI: 10.1088/2399-1984/aa954a

[59] M.M. Waldrop, "Nanoelectronics: Smart connections," *Nature* **503**(7474), pp. 22-24 (2013). DOI: 10.1038/503022a

[60] G. Piccolboni, G. Molas, J. M. Portal, R. Coquand, M. Bocquet, D. Garbin, E. Vianello, C. Carabasse, V. Delaye, C. Pellissier, T. Magis, C. Cagli, M. Gely, O. Cueto, D. Deleruyelle, G. Ghibaudo, B. De Salvo, L. Perniola., "Investigation of the potentialities of Vertical Resistive RAM (VRRAM) for neuromorphic applications," in *IEEE International Electron Devices Meeting (IEDM),* pp. 447-450 (2015). DOI: 10.1109/IEDM.2015.7409717

[61] H. Li, K.-S. Li, C.-H. Lin, J.-L. Hsu, W.-C. Chiu, M.-C. Chen, T.-T. Wu, J. Sohn, S.B. Eryilmaz, J.-M. Shieh, W.-K. Yeh, and H.-S.P. Wong, "Four-layer 3D vertical RRAM integrated with FinFET as a versatile computing unit for brain-inspired cognitive information processing," *IEEE Symposium on VLSI Technology (VLSI Technology)*, pp. 1-2 (2016). DOI: 10.1109/VLSIT.2016.7573431

[62] J. Shi, S. D. Ha, Y. Zhou, F. Schoofs and S. Ramanathan, "A correlated nickelate synaptic transistor," *Nat. Commun.* **4**:2676 (2013). DOI: 10.1038/ncomms3676

[63] E. J. Fuller, F. El Gabaly, F. Léonard, S. Agarwal, S. J. Plimpton, R. B. Jacobs-Gedrim, C. D. James, M. J. Marinella, and A. A. Talin, "Li-ion synaptic transistor for low power analog computing," *Adv. Mater.* **29**, 1604310 (2017). DOI: 10.1002/adma.201604310

[64] J. Tang, D. Bishop, S. Kim, M. Copel, T. Gokmen, T. Todorov, S.H. Shin, K.-T. Lee, P. Solomon, K. Chan, W. Haensch, J. Rozen, "ECRAM as scalable synaptic cell for high-speed, low-power neuromorphic computing," in *IEEE International Electron Devices Meeting (IEDM)*, pp. 292-295 (2018). DOI: 10.1109/IEDM.2018.8614551

[65] S. Park, A. Sheri, J. Kim, J. Noh, J. Jang, M. Jeon, B. Lee, B. R. Lee, B. H. Lee, and H. Hwang, "Neuromorphic speech systems using advanced ReRAM-based synapse," in *IEEE International Electron Devices Meeting (IEDM)*, pp. 625-628 (2013). DOI: 10.1109/IEDM.2013.6724692

[66] L. Gao, I.T. Wang, P. Y. Chen, S. Vrudhula, J. S. Seo, Y. Cao, T. H. Hou, S. Yu, "Fully parallel write/read in resistive synaptic array for accelerating on-chip learning," *Nanotechnology*, **26**(45), p. 455204, 2015. DOI: 10.1088/0957-4484/26/45/455204

[67] J. Woo, K. Moon, J. Song, S. Lee, M. Kwak, J. Park, H. Hwang, "Improved synaptic behavior under identical pulses using AlO_x/HfO_2 bilayer RRAM array for neuromorphic systems," *IEEE Electron Device Lett.*, **37**(8), pp. 994-997 (2016). DOI: 10.1109/LED.2016.2582859

[68] M. R. Azghadi, B. Linares-Barranco, D. Abbott, P. H. W. Leong "A hybrid CMOS-memristor neuromorphic synapse," *IEEE Trans. on Biomedical Circuits and Systems*, **11**(2), pp. 434-445 (2017). DOI: 10.1109/TBCAS.2016.2618351

[69] W. He, K. Huang, N. Ning, K. Ramanathan, G. Li, Y. Jiang, J. Y. Sze, L. Shi, R. Zhao and J. Pei, "Enabling an integrated rate-temporal learning scheme on memristor" *Sci*. *Rep*. 4:4755 (2014). DOI: 10.1038/srep04755

[70] T. Ohno, T. Hasegawa, T. Tsuruoka, K. Terabe, J. K. Gimzewski and M. Aono, "Short-term plasticity and long-term potentiation mimicked in single inorganic synapses," *Nat. Mater.* **10**(8), pp. 591-595 (2011). DOI: 10.1038/nmat3054

[71] T. Werner, E. Vianello, O. Bichler, A. Grossi, E. Nowak, J.-F. Nodin, B. Yvert, B. DeSalvo, L. Perniola, "Experimental demonstration of short and long-term synaptic plasticity using OxRAM multi k-bit arrays for reliable detection in highly noisy input data," in *IEEE International Electron Devices Meeting (IEDM)*, pp. 432-435 (2016). DOI: 10.1109/IEDM.2016.7838433

[72] W. Wang, A. Bricalli, M. Laudato, E. Ambrosi, E. Covi, and D. Ielmini, "Physics-based modeling of volatile resistive switching memory (RRAM) for crosspoint selector and neuromorphic computing," in *IEEE International Electron Devices Meeting (IEDM)*, pp. 932-935 (2018). DOI: 10.1109/IEDM.2018.8614556

[73] V. Milo, G. Pedretti, R. Carboni, A. Calderoni, N. Ramaswamy, S. Ambrogio, and D. Ielmini, "A 4-transistors/1-resistor hybrid synapse based on resistive switching memory (RRAM) capable of Spike-Rate-Dependent Plasticity (SRDP)," *Trans. on Very Large Integration (VLSI) Systems*, **26**(12), pp. 2806-2815 (2018). DOI: 10.1109/TVLSI.2018.2818978

[74] A.A. Faisal, L.P.J. Selen, D.M. Wolpert, "Noise in the nervous system," *Nat. Rev. Neurosci.* **9**(4), pp. 292-303. DOI: 10.1038/nrn2258

[75] A. Bricalli, E. Ambrosi, M. Laudato, M. Maestro, R. Rodriguez, and D. Ielmini: "Resistive switching device technology based on silicon oxide for improved on-off ratio - Part II: Select devices," *IEEE Trans. Electron Devices* **65**(1), pp. 122-128 (2018). DOI: 10.1109/TED.2017.2776085

[76] S. Ambrogio, S. Balatti, D. C. Gilmer, and D. Ielmini, "Analytical modeling of oxide-based bipolar resistive memories and complementary resistive switches," *IEEE Trans. Electron Devices* **61**(7), pp. 2378-2386 (2014). DOI: 10.1109/TED.2014.2325531

[77] G.W. Burr, R.M. Shelby, C. di Nolfo, J.W. Jang, R.S. Shenoy, P. Narayanan, K. Virwani, E. U. Giacometti, B. Kurdi, and H. Hwang, "Experimental demonstration and tolerancing of a large-scale neural network (165,000 synapses) using phase-change memory as the synaptic weight element," in *IEEE International Electron Devices Meeting (IEDM)*, pp. 697-700 (2014). DOI: 10.1109/IEDM.2014.7047135

[78] M. Prezioso, F. Merrikh-Bayat, B.D. Hoskins, G.C. Adam, K K. Likharev and D.B. Strukov, "Training and operation of an integrated neuromorphic network based on metal-oxide memristors," *Nature* **521**(7550), pp. 61-64 (2015). DOI: 10.1038/nature14441

[79] S. Yu, Z. Li, P.-Y. Chen, H. Wu, B. Gao, D. Wang, W. Wu and H. Qian, "Binary neural network with 16 Mb RRAM macro chip for classification and online training," in *IEEE International Electron Devices Meeting (IEDM),* pp. 416-419 (2016). DOI: 10.1109/IEDM.2016.7838429

[80] P. Yao, H. Wu, B. Gao, S.B. Eryilmaz, X. Huang, W. Zhang, Q. Zhang, N. Deng, L. Shi, H.-S.P. Wong, and H. Qian, "Face classification using electronic synapses," *Nat. Commun.* 8:15199 (2017). DOI: 10.1038/ncomms15199

[81] R. Mochida, K. Kouno, Y. Hayata, M. Nakayama, T. Ono, H. Suwa, R. Yasuhara, K. Katayama, T. Mikawa, Y. Gohou, "A 4M synapses integrated analog ReRAM based 66.5 TOPS/W neural-network processor with cell current controlled writing and flexible network Architecture," in *IEEE Symposium on VLSI Technology (VLSI Technology),* pp. 175-176 (2018). DOI: 10.1109/VLSIT.2018.8510676

[82] F. Merrikh-Bayat, M. Prezioso, B. Chakrabarti, H. Nili, I. Kataeva, D. Strukov, "Implementation of multilayer perceptron network with highly uniform passive memristive crossbar circuits," *Nat. Commun.* **9**:2331 (2018). DOI: 10.1038/s41467-018-04482-4

[83] C. Li, D. Belkin, Y. Li, P. Yan, M. Hu, N. Ge, H. Jiang, E. Montgomery, P. Lin, Z. Wang, W. Song, J. P. Strachan, M. Barnell, Q. Wu, R. S. Williams, J.J. Yang, and Q. Xia, "Efficient and self-adaptive in-situ learning in multilayer memristor neural networks," *Nat. Commun.* **9**:2385 (2018). DOI: 10.1038/s41467-018-04484-2

[84] S. Yu, B. Gao, Z. Fang, H. Yu, J. Kang and H.-S.P. Wong, "Stochastic learning in oxide binary synaptic device for neuromorphic computing," *Front. Neurosci.* **7**:186 (2013). DOI: 10.3389/fnins.2013.00186

[85] D. Garbin, O. Bichler, E. Vianello, Q. Rafhay, C. Gamrat, L. Perniola, G. Ghibaudo and B. De Salvo, "Variability-tolerant Convolutional Neural Network for pattern recognition applications based on OxRAM synapses," in *IEEE International Electron Devices Meeting (IEDM)*, pp. 661-664 (2014). DOI: 10.1109/IEDM.2014.7047126

[86] E. Covi, S. Brivio, A. Serb, T. Prodromakis, M. Fanciulli, and S. Spiga, "Analog memristive synapse in spiking networks implementing unsupervised learning," *Front. Neurosci.* **10**:482 (2016). DOI: 10.3389/fnins.2016.00482

[87] V. Milo, G. Pedretti, R. Carboni, A. Calderoni, N. Ramaswamy, S. Ambrogio, and D. Ielmini, "Demonstration of hybrid CMOS/RRAM neural networks with spike time/rate-dependent plasticity," in *IEEE International Electron Devices Meeting (IEDM)*, pp. 440-443 (2016). DOI: 10.1109/IEDM.2016.7838435

[88] A. Serb, J. Bill, A. Khiat, R. Berdan, R. Legenstein, and T. Prodromakis, "Unsupervised learning in probabilistic neural networks with multi-state metal-oxide memristive synapses," *Nat. Commun.* **7**:12611 (2016). DOI: 10.1038/ncomms12611

[89] T. Werner, E. Vianello, O. Bichler, D. Garbin, D. Cattaert, B. Yvert, B. DeSalvo and L. Perniola, "Spiking neural networks based on OxRAM synapses for real-time unsupervised spike sorting," *Front. Neurosci.* **10**:474 (2016). DOI: 10.3389/fnins.2016.00474

[90] G. Pedretti, V. Milo, S. Ambrogio, R. Carboni, S. Bianchi, A. Calderoni, N. Ramaswamy, A. S. Spinelli, and D. Ielmini, "Memristive neural network for on-line learning and tracking with brain-inspired spike timing dependent plasticity," *Sci. Rep.* **7**:5288, (2017). DOI: 10.1038/s41598-017-05480-0

[91] G. Pedretti, V. Milo, S. Ambrogio, R. Carboni, S. Bianchi, A. Calderoni, N. Ramaswamy, A. S. Spinelli, D. Ielmini, "Stochastic learning in neuromorphic hardware via spike timing dependent plasticity with RRAM synapses," in *IEEE J. Emerging Topics in Circuits and Systems (JETCAS)* **8** (1), pp. 77-85 (2018). DOI: 10.1109/JETCAS.2017.2773124

[92] G. Pedretti, S. Bianchi, V. Milo, A. Calderoni, N. Ramaswamy, and D. Ielmini, "Modeling-based design of brain-inspired spiking neural networks with RRAM learning synapses," in *IEEE International Electron Devices Meeting (IEDM)*, pp. 653-656, 2017. DOI: 10.1109/IEDM.2017.8268467

[93] Z. Wang, S. Joshi, S. Savel'ev, W. Song, R. Midya, Y. Li, M. Rao, P. Yan, S. Asapu, Y. Zhuo, H. Jiang, P. Lin, C. Li, J.H. Yoon, N.K. Upadhyay, J. Zhang, M. Hu, J.P. Strachan, M. Barnell, Q. Wu, H. Wu, R.S. Williams, Q. Xia and J.J. Yang, "Fully memristive neural networks for pattern classification with unsupervised learning," *Nature Electronics* **1**, pp. 137-145 (2018). DOI: 10.1038/s41928-018-0023-2

[94] W. Wang, G. Pedretti, V. Milo, R. Carboni, A. Calderoni, N. Ramaswamy, A.S. Spinelli, D. Ielmini, "Learning of spatiotemporal patterns in a spiking neural network with resistive switching synapses," *Sci. Adv.* **4**(9):eaat4752 (2018). DOI: 10.1126/sciadv.aat4752

[95] T. Dalgaty, E. Vianello, D. Ly, G. Indiveri, B. De Salvo, E. Nowak, and J. Casas, "Insect-inspired elementary motion detection embracing resistive memory and spiking neural networks," in: *Vouloutsi V. et al. (eds) Biomimetic and Biohybrid Systems. Living Machines 2018. Lecture Notes in Computer Science*, vol 10928. Springer, Cham (2018). DOI: 10.1007/978-3-319-95972-6_13

[96] D. Ly, A. Grossi, C. Fenouillet-Beranger, E. Nowak, D. Querlioz, and E. Vianello, "Role of synaptic variability in resistive memory-based spiking neural networks with unsupervised learning," *Journal of Physics D: Applied Physics*, **51**:44 (2018). DOI: 10.1088/1361-6463/aad954

[97] T. Masquelier, R. Guyonneau, and S.J. Thorpe, "Competitive STDP-based spike pattern learning," *Neural Computation* **21**(5), pp. 1259-1276 (2009). DOI: 10.1162/neco.2008.06-08-804