ScuDo

Scuola di Dottorato ~ Doctoral School

WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation

Doctoral Program in Physics (30^{th}cycle)

Neuromorphic Systems based on Memristive Devices

From the material science perspective to bio-inspired learning hardware

By

Daniele Conti

******

Supervisor(s):

Prof. Fabrizio C. Pirri, Supervisor

Prof. Carlo Ricciardi, Co-Supervisor

Politecnico di Torino, Department of Applied Science and Technology, Torino, Italy

Doctoral Examination Committee:

Prof. Roberto Verucchi, Referee, CNR-IMEM, Institute of Materials for Electronics and Magnetism, Trento, Italy

Prof. Ettore Vittone, Referee, University of Torino, Department of Physics, Torino, Italy

Prof. Dario Daghero, Politecnico di Torino, Department of Applied Science and Technology, Torino, Italy

Prof. Daniele Ielmini, Politecnico di Milano and IU.NET, Department of Electronics, Informations and Bioengineering, Milano, Italy

Prof. Riccardo Zecchina, Bocconi University, Department of Decision Sciences, Milano, Italy; Politecnico di Torino, Institute of Condensed Matter Physics and Complex Systems, Torino, Italy

Politecnico di Torino
2018

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Daniele Conti
2018

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

I would like to dedicate this thesis to everyday life dreamers, to simple people that silently continue to explore ways of improving the world with their knowledge, while doing their job.
Don’t stop, find the way, build the opportunity and start with action.
Future belongs to day-dreamers.

Acknowledgements

I would like to acknowledge my supervisor, Prof. Fabrizio C. Pirri, and my co-supervisor, Prof. Carlo Ricciardi, for making me available their knowledge and precious experience, for the meaningful discussions we had in these three years of science and exploration of physics and technology. Thank you for having been my mentors along the way.
I would like to acknowledge Prof. Riccardo Zecchina, for having introduced me to such an exciting topic that stays at the frontier of physics, technology, computer science and biology.
I would like to acknowledge Prof. Giacomo Indiveri and his group, together with Doct. Sabina Spiga, Stefano Brivio and their colleagues at CNR-Agrate, for the excellent time I had in a top class environment such as Zuerich and the high quality research we did all together.
I would like to acknowledge my colleagues, Cecilia Giovinazzo, Gianluca Milano, Vittorio Fra and Davide Scaiola, for sharing with me both excitement and frustration typical of research. They helped me in growing both as a person and as a researcher. I am proud of the scientific exploration we did together.
I would like to acknowledge the people I collaborated with during my research activities: Samuele Porro, Krishna Rajan, Denis Perrone and many others, that made everyday life run nicely.
I would like to acknowledge the PhD in physics board, together with its coordinator, Prof. Arianna Montorsi, with which I collaborated so proficiently during my representative activities.
I would like to acknowledge the PhD students for all the good moments we shared together and, in particular, Lucrezia Ravera, with which we created AperiDISAT.
In conclusion, I would like to say thanks to my family that, following all of my choices, always found the way to give me good advice and stay close to me.
Thank you.

Abstract

Hardware computation is facing in the present age a deep transformation of its own paradigms. Silicon based computation is reaching its limit due to the physical constraints of transistor technology. As predicted by the Moore's law, downscaling of transistor dimensions doubled each year since the 60s, leading nowadays to the extreme of 16-nm channel width of the present state-of-the-art technology. No further improvement is possible, since laws of physics impose a different electrical behavior when lower dimensions are attempted. Multiple solutions are then envisaged, spanning the range from quantum computing to neuromorphic computing.
The present dissertation wants to be a preliminary study for understanding the opportunities enabled by neuromorphic computing based on resistive switching memories. In particular, brain inspires technology and architecture of new generation processors because of its unique properties: parallel and distributed computation, superposition of processing and memory unit, low power consumption, to cite only some of them. Such features make brain particularly efficient and robust against degraded data, further than particularly suitable to process and store in memory new information. Despite many research projects and some commercial products are already proposing brain-like computing processors, like spiNNaker or IBM's Bluenorth, they only mimic the brain functioning with standard Silicon technology, that is inherently serial and distinguish between processing and memory unit. Resistive switching technology on the other hand, would allow to overcome many of these issues, enabling a far better match between biological and artificial neuromorphic computation.
Resistive switching are, generally speaking, Metal-Insulator-Metal structures able to change their electrical conductance as a consequence of the history of applied electric signal. In such sense, they behave exactly as synapses do in a biological neural networks. For this reason, resistive switching when modeled as memristor, i.e. memory-resistor, can act as artificial synapses and, moreover, are particularly suitable to be interfaced with artificial Silicon neurons that are designed to replicate

the biological behavior when excited with electric pulses. Anyhow, from the technological standpoint, there is still no standard on the design and fabrication of resistive switching, so that multiple structure and materials are investigated.
In this dissertation, it is reported an analysis of multiple resistive switching devices, based on various materials, i.e. TiO_2, ZnO and HfO, and device architectures, i.e. thin film and nanostructured devices, with the scope of both characterizing and comprehending the physics behind resistive switching phenomena. Furthermore, numerical simulations of artificial spiking neural networks, embedding Silicon neurons and HfO-based resistive switching are designed and performed, in order to give a systematic analysis of the performances reached by this new kind of computing paradigm.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivations

Artificial intelligence, (AI), has became one of the most prominent topic in technology, able to agglomerate around itself wide segments of society that in the past usually remained disaggregated with respect to the development of science. Indeed, the scientific community devotes huge efforts in the understanding of basic principles of learning, both from the biologic and computational perspective, and its development in real-world application. Their results are constantly monitored by the industry world, that from applications into the production process up to management of new services, are progressively integrating autonomous robot and machine learning algorithms into their businesses. Interestingly, also the general public is enthusiastically following the success reached by scientists and engineers, developing in parallel to the technological progress a common thought interrogating itself about the ethical and moral implications of artificial intelligence. It may sound almost natural nowadays thinking about robots helping us side by side in our daily life, but still it is an astonishing result getting real day after day. AI signs milestones more and more frequently, winning world championship of chess (IBM's Deep Blue, May 1997), obtaining super-human performance in pattern recognition (IJCNN Competition, 2011, [1]), recently winning he strategy based Go championship with a self-teaching AI (DeepMind & Google's AlphaGo, March 2016). Tools that allow for direct interactions with machine through natural language are widespread, from mobile-phones (Apple's Siri), computers (Microsoft's Cortana), personal assistance

at home (Amazon's Alexa). The latest technology pops up from automotive industry and regards self-driving cars. This is a step further in the realization of the over-mentioned Internet of Things (IoT), constituted by network of devices able to sense stimuli from the real world, communicate among each-other and actuate in order to accomplish specific tasks. Devices in the internet of things are requested to operate autonomously in remote areas, where in principle, electric power could not be sufficiently available and remote control cannot be feasible. Nonetheless, device activity needs to be desirably efficient, avoiding failures in task performances. In order to accomplish such a demanding requirement, devices must be compact, power efficient, optimized from the algorithmic standpoint and, above all, *smart*.
Smart devices must be able to process information gathered from the environment and perform predictions about its dynamic. Such a prior information is then evaluated to take decisions leading to the successfully accomplished task. In principle, it is a particularly power demanding activity and it may be hardly implemented on remote devices, that nowadays are indeed constantly communicating with servers where computation is performed. This leads, just as examples, to delays in response to stimuli, approximated analysis due to incomplete data, suffer from connection failures... local computation would solve many of these limitations, struggling on the other hand with the limited computational power.
Such an important challenge brought the scientific community to investigate further machine learning algorithms and device architectures, looking for inspiration from nature. The most effective example of extreme computational power, combined with power efficiency, is undoubtedly the human brain.
The human brain has astonishing capabilities with a limited amount of resources. We are able to perform task characterized by a high level of complexity, such as recognizing faces in very corrupted environment, as well as sounds, communicate with multiple languages, define complex activity strategies based on external dynamic variables. All of this with a more efficient power consumption, [2]. Brain has developed in its evolution an highly parallelized architecture completely different from the Von Neumann one, that stays at the basis of computers. Von Neumann architectures requires the computation unit (ALU) and the memory unit to be distinguished inside the central processing unit (CPU), [3], see figure 1.1.

Within brain, such a distinction is not existing anymore and the computation unit is merged with the memory unit, through synapses. The neural network is indeed built out of about 10^{11} neurons, each of them generating an average of 10^4

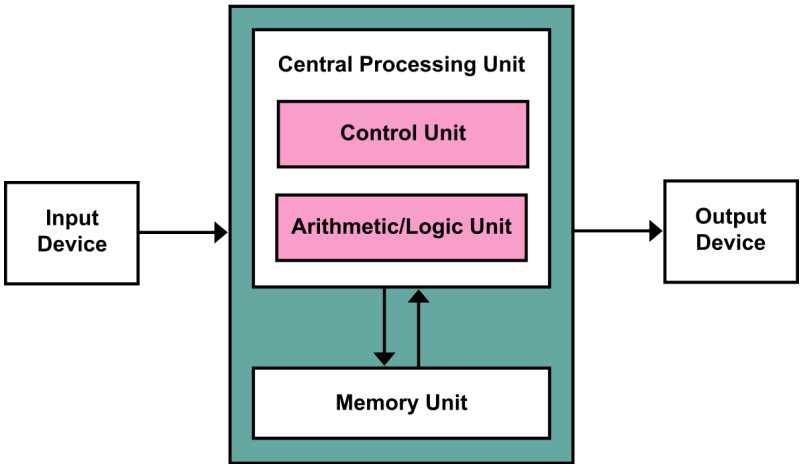


Fig. 1.1 Von Neumann Architecture

connections with other neurons in the cortex, [4]. These are dynamic connections, i.e. synapses, that emerge or disappear according to the network functioning. Moreover, their efficacy evolves in time according to the task performed, shaping what we call memory, [5]. From this perspective, synapses are both the processing unit of our brain, whose information is transmitted across neurons that acts as center of integration, and the center where memory is stored. It is the neurons activity itself that allows synaptic strength to change in time, according to a learning rule named spike-timing dependent plasticity, (STDP), [6–9]

According to STDP, the time-correlation between the firing activity of pre-synaptic neurons and post-synaptic neuron is responsible for potentiating or depressing the synapse efficacy. Specifically, if a post-synaptic spike follows in time a pre-synaptic spike, synapse strengthen its efficacy proportionally to their delay. Vice-versa, if post-synaptic spike comes earlier that the pre-synaptic spike, i.e. the spikes are anti-correlated in time, the synapse becomes depressed. In this way, information processing happens locally in time and space, i.e. only when spikes occur at the level of single synapse, and in parallel, thus for multiple synapses at the same time. Both features are particularly desirable for algorithms embedded on independent smart devices.

Because of all of these features expressed by the human brain functioning, the scientific community began to investigate it also from the robotics and computer science perspective. A new emergent technology seems to be particularly promising for the realization of artificial neural network in hardware, called resistive switchings.

1.2 Resistive Switching and Memristor

Resistive switching are devices able to perform a reversible transition in their resistivity according to the variation of an external variable, like temperature, pressure, magnetic field, voltage, [10]. Among the many physical mechanisms involved with different material, the resistive switching addressed in this dissertation are only those involving as control variable the applied voltage. Huge investigation in the field comes from micro-electronics companies. Indeed more-than-Moore technologies are of paramount importance in companies outlook, since the current transistor based technology is reaching its physical limit, [11]. The Moore's law, [12], empirically defined in 1965, by Gordon Moore, founder of Intel, stated that the micro-processors complexity doubles every 18 months. Taking in account that the chip board remains nearly constant, the Moore's law imply a constant doubling of chip density. Since its definition, Moore's law has been taken as growth target for all electronic components producers, pushing the technology process toward optimized Metal-Oxide-Semiconductor (MOS) fabrication with lower and lower channel size. From the 60s, the channel dimension of CMOS stepped from microns to the 32 nm currently implemented in standard commercial laptops. Further research scales down this reference dimension to 20 nm, touching the extreme limit of 16 nm, [13]. It is the limit imposed by physics to Moore's law, since below this dimension, the silicon doping allowing the functioning of MOS as logic gates cannot be guaranteed anymore to be equivalent for all the nodes, i.e. being the doping process intrinsically stochastic, the number of doping atoms in the channel cannot be tuned and lead to a higher failure rate of single nodes in operative conditions. From this scenario comes the pressure laying behind the investigation of resistive switching technology. Indeed, different physical mechanisms would allow a further scale down of chip components dimensions primarly for random access memories fabrication (RAM). Particularly attractive from this standpoint appear to be magnetoresistances (MRAM) [14], based on spintronics, phase change memories (PRAM), [15], and, more recently, resistive switching (ReRAM), [16, 17]. Resistive switching phenomenon is observed with a large varieties of devices characterized by different architectures, metal-insulator-metal (MIM) layers or nanostructures, different materials, like transition metal oxides [18], transitional metal dichalcogenides [19], graphene-based structures [20], organic compounds [21], colloids [22]. Among them, binary oxides, like TiO_2 [23], Ta_2O_5 [24], ZnO [25], HfO [26], WO_3 [27], NiO [28], appears to

be particularly promising for ReRAM applications. Mechanisms laying behind the resitive switching phenomenon in MIM architecture are still under deep investigation [29], but can classified in two major areas, i.e. valence change memories (VCM) and electro-chemical memories (ECM), [16]. The former mechanisms involves the presence of defects in the transition metal oxide matrix in the form of oxygen vacancies. The absence of one oxygen atom in the lattice (that often is amorphous or polycrystalline) creates a locally positive charged that is unstably occupied by electrons, in order to respect the neutrality principle. The positively charged oxygen vacancies thus are subject to external electric field imposed through the application of a voltage to the two electrodes of the MIM structure and drifts through the lattice. A distinction at this stage is done, identifying interface-type mechanism, [30, 31], and filamentary mechanisms [32], depending on the characteristics of the material adopted as insulator. With interface resistive switching, the insulator layer can be observed to contain a sublayer under-stoichiometric with respect to the whole lattice. Its resistance thus would be lower than the stoichiometric sub-layer and the whole structure would act as the series of two resistances. Under the effect of external electric field and with proper lattice properties, the interface between the two layers could move, reaching the opposite electrode from the one where voltage is applied. As soon as the whole lattice is homogeneously under-stoichiometric, the MIM device switches from a high resistance state (HRS) to a low resistance state (LRS). In filamentary RS on the other hand, rather than a moving interface, the movement of oxygen vacancies is favored withing a tiny region of the lattice that thus assumes locally a lower stoichiometry, [33]. The difference between the two mechanisms is strongly dependent on the lattice properties, as it is determined by the potential barriers height and distribution encountered by oxygen vacancies while drifting.
The second major subset of RS accounts for electro-chemical memories. ECM require the presence of an electro-migrating metal element as one of the two electrodes adopted for the MIM structure, like Ag or Cu, [34]. In ECM, when positive voltage is applied to the electro-migrating electrode, redox reaction happens at the interface between metal and insulator, leading to anodic dissolution of metallic cations subject to the electric field drift. Depending on the ion mobility within the insulator matrix, they will reach the opposite electrode creating a so called Conductive Filament (CF). CF is a metallic shortcut, cone-shaped, between the two electrode and through the insulating material, usually few tens of nm in diameter [35].
In both mechanism, the phenomenon is reversible when opposite polarity voltage is

applied and imposes drift toward the electrode of origin. When the LRS is reached in fact, the electronic current transport dissipates energy under form of heat, respecting the Joule law. Above a certain threshold thus, the CF can be interrupted and the newly generated ions (both vacancies in the case of VCM and metal ions in case of ECM) are drifted backward from the grounded electrode.
An optimized resistive switching reaches good repeatability in switching, cycle to cycle endurance and state retention. Companies' objective is the production of ReRAM featuring these quality measures comparable to the present transistor-based technology.
In 2008, the interest of the scientific community was awakened by results from HP labs. The group of Strukov and William in fact, published a description of TiO_2 based MIM resistive switching modeled as a memristor, [36].
Memristor, literally the contraction of the terms memory and resistor, was postulated by Leon Chua in 1971 as the fourth missing fundamental circuit element, together with resistor, capacitor and inductor, [37]. He proceeded from symmetrical reasoning, comparing the constitutive equations of the three elements,

$$\begin{aligned} dv &= R\, di \\ dq &= C\, dv \\ d\phi &= L\, di \end{aligned} \tag{1.1}$$

together with the definition of charge and magnetic flux as:

$$\begin{aligned} q &= \int i\, dt \\ \phi &= \int v\, dt \end{aligned} \tag{1.2}$$

Observing the previous set of equations, graphically depicted in figure 1.2, it emerges that all the constitutive equations derive from the relation between two of the fundamental quantities v, i, q and ϕ, but no relation exists between charge and magnetic flux. Chua in [37] demonstrate that such a relation can exists for a passive element, whose constitutive equation can be written as:

$$\begin{aligned} \phi &= f(q) \\ d\phi &= M(q)\, dq \end{aligned} \tag{1.3}$$

where $M(q)$ is the so called *memristance*.

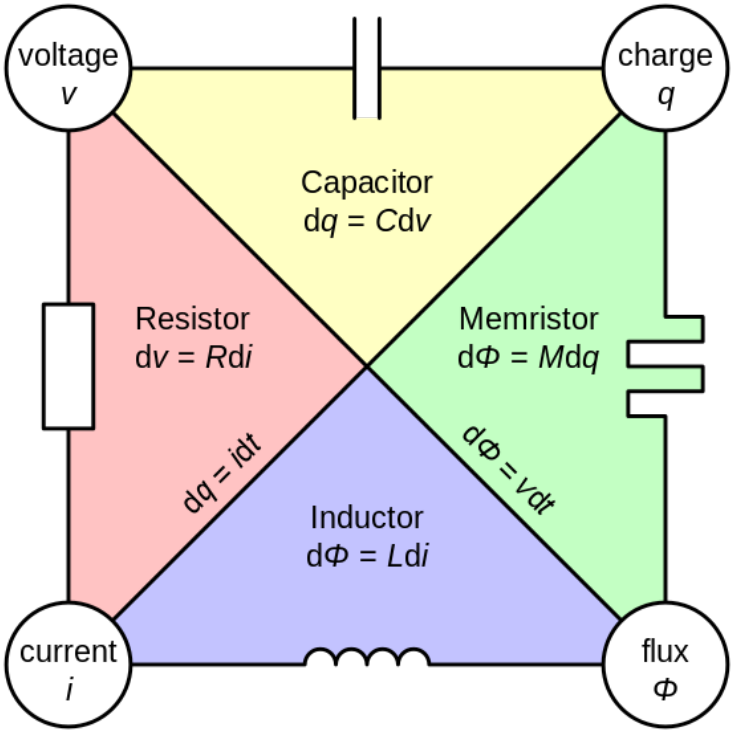


Fig. 1.2 Definition of the memristor from symmetric reasoning

Notice that one can always take the time derivative of such a constitutive equation, obtaining:

$$\begin{cases} v(t) = \dfrac{df(q)}{dq}\, i(t) = M(q)\, i(t) \\ \dfrac{dq}{dt} = i(t) \end{cases} \tag{1.4}$$

Some observation is important already at this stage about the memristor definition. First of all, the memristor is intrinsically a non-linear element, since if $M(q)$ would be constant, integrating the constitutive equation would lead to the case of

linear resistor. Moreover, the quantity ϕ involved in the constitutive equation comes from the Lenz's law, that states the existence of an electromotive force, i.e. a voltage, induced by the time variation of magnetic flux. This is certainly true in the case of an inductor, where a coil generates a magnetic field, but it is not necessarily implied in the case of a memristor, as it will be clarify in the next of this dissertation. Indeed, ϕ should be considered not as a magnetic flux, but only as the integral of voltage in time, and for this reason it will be called, as in the most recent literature, *voltage momentum*, [38].
Memristor is an element with memory, i.e. its memristance value depends on the quantity of charged that passed through the element under the effect of the applied voltage in time. Thus, the history of the applied signal determines the resistance state of the device.
As last consideration at this stage, from the equation 1.4 it can obtained that the memristor fingerprint is an hysteresis loop pinched at the origin of axis, as the one shown in figure 1.3.
In 1976, the memristor definition was generalized to non ideal cases by Leon Chua and Sung M. Kang, introducing a so called *state variable* collecting all other dependencies of the memristance different from charge, [39]. The new obtained set of equation becomes then, in terms of voltage momentum and charge:

$$\begin{cases} \phi = f(q, x(t)) \\ \dfrac{dx(t)}{dt} = g(x, i) \\ \dfrac{dq(t)}{dt} = i \end{cases} \tag{1.5}$$

and in terms of voltage and current:

$$\begin{cases} v(t) = M(q, x(t))\, i(t) \\ \dfrac{dx(t)}{dt} = g(x, i) \end{cases} \tag{1.6}$$

where $g(x,i)$ is the dynamic functional relation that specifies the evolution in time of the state variables controlling the memristor and specifies the whole physics involved. Examples of such variables could be the internal temperature, the applied electric field, defect concentration etc., depending on the system under test. As it can be noticed, the memristor is described as a non-linear dynamic systems.

In 2008, HP labs scientists, Dmitri B. Strukov and Stanley R. Williams published in Nature for the first time the description of a resistive switching, based on TiO_2 and with MIM architecture, as a memristor, titling the paper "The missing memristor found", [36]. In that case, they provided a simple model of interface-type resistive switching, with a separation between region rich in positively charged dopant, i.e. oxygen vacancies, and region poor of dopant, i.e. stoichiometric metal oxide, see figure 1.3. Application of the external bias imposes a drift of the region boundary, modeled as state variable of the system. Thus, considering the total thickness of the device D and the dopant mobility μ_V, they write the dynamic equation of the system as the series of two resistances:

$$\begin{aligned} v(t) &= \left(R_{ON}\frac{x(t)}{D} + R_{OFF}\left(1 - \frac{x(t)}{D}\right)\right) i(t) \\ \frac{dx(t)}{dt} &= \mu_V \frac{R_{ON}}{D} i(t) \end{aligned} \tag{1.7}$$

That yields for what concerns the state variable dynamics:

$$x(t) = \mu_V \frac{R_{ON}}{D} q(t) \tag{1.8}$$

The impact of such a paper on the scientific community was incredibly high, since boosted renovated research on a field that since 1976 was stuck to purely theoretical results, without any experimental counterpart. On the contrary, from that moment on, huge scientific effort is devoted to the investigation of the physical mechanism behind the resistive switching devices, together with their theoretical modeling in the framework of Chua's formalism.

From the perspective of neuromorphic systems, memristors offer particularly favorable features to mimic synapse functioning. In fact, synapse is a junction between two nodes of a circuit and it is in charge to transfer a weighted information depending on its relevance for accomplishing a task. Relevance then is gathered via *experience*, if we are talking from a biology perspective, or through what the network has learned during its training phase, talking from the computer science standpoint. Indeed, while learning synapses change their synaptic weight, i.e. their efficacy, according to the signal they observe and then preserve that change. At the

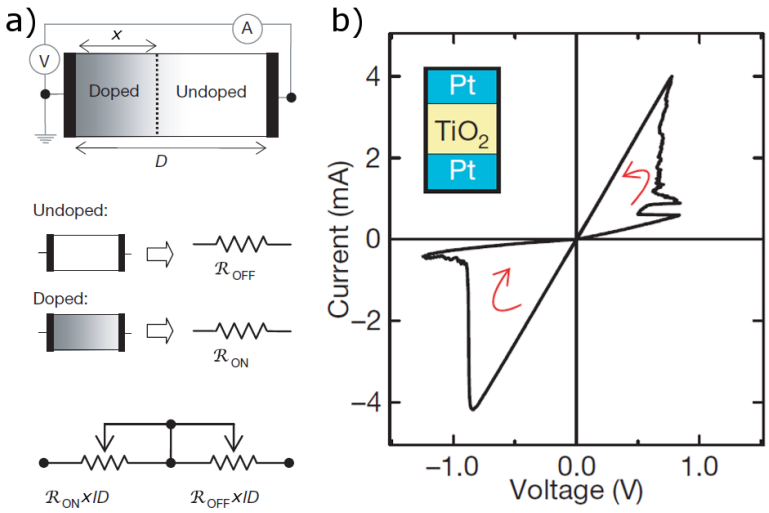


Fig. 1.3 HP Labs model of TiO_2-based memristor.

very same level, memristor reaches a resistance state, mathematically formalized by its state variable, depending on the history of signals applied and then preserves memory of it. It means that, within a network, signals in the form of electric current can be weighted by the internal resistance of memristors while being transferred to other nodes. In the extreme cases, a highly insulating memristor is equivalent to an open circuit and the information is not transferred, while a memristor in conductive state lets the information be transferred without attenuation.
Memristors, in the experimental form of resistive switching, can be reasonably investigated as synaptic element of hardware neural network to be embedded in future smart devices.

1.3 Spiking Neural Network

Hardware devices designed specifically for machine learning purposes are usually referred to as hardware neural networks (HNN). Since the end of the past century, together with the faster development of deep neural networks in software simulations, HNN received great attention because of their inherent advantages with respect to the traditional computing technologies. Indeed, state of the art neural networks, such as convolutional neural networks (CNN) and recurrent neural networks (RNN) are particularly computational demanding. Moreover, NN are intrinsically parallel in the computing architecture, but higher complexity of electronic systems is required

when transferred on a serial computing architecture.
HNN then are particularly desirable for several properties: *speed*, specialized hardware are optimized for learning tasks, outperforming standard technologies; intrinsically hardware-embedded *parallel* computing allows for improved computational power and speed on the one side and, on the other, it provides a particularly robust infrastructure against system failures and *degradation*. Indeed, serial architectures are particularly vulnerable to *fail-stop* operations, mainly because of lack of redundancy. On the contrary, neural networks are extremely tolerant to single connection failures as well as noise, and rather than stopping, they continue to operate with reduced performances. All of these characteristics make HNN of considerable advantage for the development of smart and remote technologies, [40].
If the number of advantages for HNN is considerable, a still large number of challenges must be faced by HNN designers. The most prominent regards the network topology, whose high connection degree could result in a complex pattern of connectivity to be transferred in silicon. Moreover, hardware technology could potentially introduce further variability to the one accounted in the normal function of learning. Another important fact is the non-linearity of activation function required by learning algorithm, that for digital technology could require complex and power-angry control circuits.
The complete landscape of HNN different implementations thus is still variegated and accounts for multiple solution dealing with the aforementioned problems. They include digital, analog, hybrid solutions, optical solutions, FPGA (Field Programmable Gate Array) based solutions, VLSI (Very Large Scale Integration) solutions.
Among those, neuromorphic Spiking Neural Networks (SNN) are gathering more and more interest for many reasons. First they are *event-based* networks, where computation happens only when an event, i.e. an electric (often differential) signal is inputed to the network sensors. It allows to save power and speed up performances. Moreover, they allow for bio-inspired learning algorithms implementation, such as the one currently recognized by the neuroscience community as at the basis of brain functioning, i.e. STDP.
Spiking Neural Networks convert the input stimulus into a train of Poisson spikes, with frequency directly proportional to input intensity, e.g. in the case of image recognition, pixel brightness is converted with direct proportionality into spiking trains of certain frequency. Nodes of the SNN are then Integrate&Fire (I&F) neurons, which receive input in form of current and accumulate it until a pre-defined threshold

is reached. When the threshold is reached, the I&F neuron emits an output spike that is forwarded in the network.
The state-of-the-art I&F neuron was developed in 2009 by Indiveri et al. with analogue VLSI technology embedding 21 metal-oxide-semiconductor field effect transistors (MOSFET) working in the subthreshold domain and capacitors, [41, 42]. Its schematics is reported in figure 1.4.

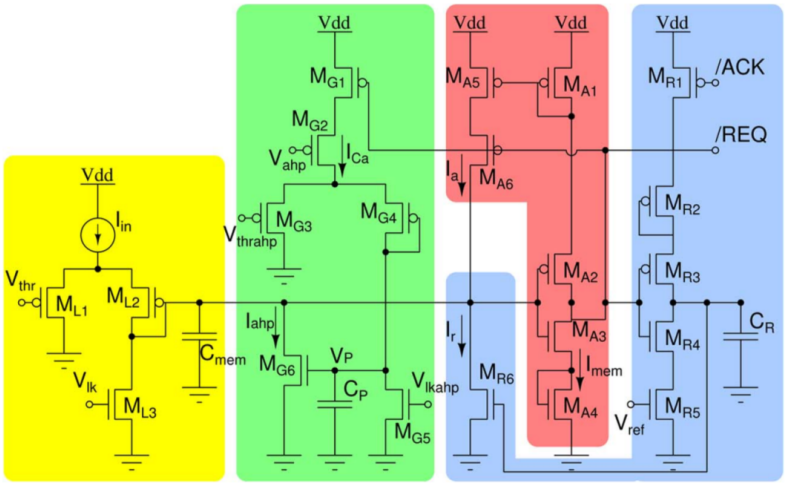


Fig. 1.4 Schematics of Integrate and Fire Neuron.

The exponential adaptive I&F neuron accounts for four different blocks, here represented in different colors. The yellow block, named Differential Pair Integrator (DPI) models the neuron's leak conductance returning exponential sub-threshold dynamics in response to constant input; this block contains the capacitor named C_{mem} that mimics the neuron's membrane capacitance. The green-block implements spike frequency adaptation mechanism, mimicking the calcium current adaptation mechanisms present in real neurons. The red block is responsible for generation of amplified spike events. To conclude, the blue block resets the state of the neuron to its resting state (set with V_{ref}), providing a refractory period during which spike generation is impeded.
The dynamics of the VLSI I&F neuron can be expressed in a single equation as follow:

$$\begin{aligned}\left(1+\frac{I_{th}}{I_{mem}}\right)\tau\frac{d}{dt}I_{mem}+I_{mem}\left(1+\frac{I_{ahp}}{I_{\tau}}\right)&=I_{mem_{\inf}}+f(I_{mem})\\ \tau_{ahp}\frac{d}{dt}I_{ahp}+I_{ahp}&=I_{ahp_{\infty}}u(t)\end{aligned}\tag{1.9}$$

that, with linearization (approximable for $I_{in} << I_{\tau}$) and no adaptation, can be reduced to:

$$\tau\frac{d}{dt}I_{mem}+I_{mem}=\frac{I_{th}}{I_{\tau}}I_{mem}+f(I_{mem})\tag{1.10}$$

with,

$$f(I_{mem})=\frac{I_a}{I_{\tau}}(I_{mem}+I_{th})\approx\frac{I_a}{I_{\tau}}I_{mem}\tag{1.11}$$

where I_{mem} is the neuron's membrane current, I_{ahp} is the current involved in spiking-frequency adaptation, I_a is the positive-feedback current and $u(t)$ is a step function responsible for the refractory period of the neuron. All other parameters are derived from MOSFET parameters and can be written as:

$$\begin{aligned}\tau&=\frac{C_{mem}U_T}{\kappa I_{tau}}\\ \tau_{ahp}&=\frac{C_pU_T}{\kappa I_{\tau_{ahp}}}\\ I_{\tau}&=I_0e^{\frac{\kappa}{U_T}V_{lk}}\\ I_{\tau_{ahp}}&=I_0e^{\frac{\kappa}{U_T}V_{lkahp}}\\ I_{mem_{\infty}}&=\frac{I_{th}}{I_{\tau}}(I_{in}-I_{ahp}-I_{tau})\\ I_{ahp_{\infty}}&=\frac{I_{th_{ahp}}}{I_{\tau_{ahp}}}I_{Ca}\\ I_{th}&=I_0e^{\frac{\kappa}{U_T}V_{thr}}\\ I_{th_{ahp}}&=I_0e^{\frac{\kappa}{U_T}V_{thrahp}}\end{aligned}\tag{1.12}$$

with κ and U_T the transistor permittivity and thermal voltage, I_0 the transistor dark current, I_{in} the input current, I_{th} and $I_{th_{ahp}}$ the current through n-type MOSFETs not present in figure 1.4.
As it can be noticed, the I&F neuron here presented is designed to faithfully replicate the functioning of biological neuron, including hyper-polarization, refractory period and adaptation. Moreover, such a design is low-power, compact and compatible with asynchronous logic. Similar design can be adopted for VLSI artificial synapses as well, embedding other tens of transistors. By the way, here it is where memristors can contribute to the design of neuromorphic system and their simplification. Indeed the memristor is particularly suitable to synaptic plasticity induced by input spikes as the one produced by the I&F neuron, so it could be set as connection between neurons in more complex network. Further control circuits are required for the learning algorithm implementation, regardless of the synaptic device adopted.

1.4 Dissertation Outlook

The present dissertation is devoted to the investigation of resistive switchings and their synaptic plasticity properties in the context of neuromorphic systems. Its structure will proceed as follow:

TiO_2 based resistive switching. In this first chapter, nanostructured devices and thin film devices are investigated with different purposes, i.e. to study the physics governing the switching mechanisms on the one hand and the device needs for technology application and engineered materials on the other.
The discussion will start from TiO_2 based nanotubes grown via anodic oxidation. Titania nanotubes offer an intrinsic multilayer structure, with multiple stoichiometry, that can potentially be exploited as VCM type resistive switching. In order to deepen the role of oxygen vacancies in the switching mechanism, multiple nanotubes arrays are produced in different conditions, including also a top coating of polyacrylic acid responsible for the catalysis of surface chemical reactions able to change the oxygen vacancies and the charge carrier densities available to conduction.
Consecutively, TiO_2 based thin layer will be briefly discussed. TiO_2 thin film are deposited via Atomic Layer Deposition (ALD). It is a technique particularly favorable for large scale production of nano-sized devices because of

its peculiar features like self-limiting deposition and conformal deposition without line-of-sight requirements. Deposition at different temperatures, i.e. lower than the normally used in TiO_2 processes are investigated, in order to simplify the device fabrication process. Consequently, the variation on materials properties and resistive switching response are analyzed.

ZnO based resistive switching. The research group focused particularly its attention on ZnO based resistive switching because of the numerous techniques available for material deposition and device fabrication. Indeed, both thin film devices as well as nanostructured devices are fabricated. Thin film devices were fabricated using either ALD deposition or Sputtering deposition. Nanostructured devices in the form of ZnO nanowires were fabricated via Chemical Vapor Deposition (CVD) and then characterized either as a forest of nanowires or as a single nanowire. All the variety of techniques available within the group allowed to compare performances and properties of devices among each-other, with the aim of investigate the different physical mechanisms implied in the resistive switching of ZnO, both in VCM and ECM configuration.
This chapter will be focused on ZnO thin film deposited via Sputtering technique. This work aims to compare ECM and VCM resistive switching responses with thin layers of different thicknesses, implying different crystallinity conditions. Insights about the switching mechanisms are required in order to obtain stable and repeatable electric response.

Simulation of learning network. The activity described in this chapter originated from an international collaboration among the PoliTO group, CNR division of material science, in Agrate Brianza, Italy, and the institute of Neuroinformatics - INI, University of Zurich and Polytechnic University of Zurich, Zurich, Switzerland. It is devoted to the simulation of a fully hardware spiking neural network embedding silicon I&F neurons, as the one described in the previous section, and experimentally characterized resistive switching, complete of device variability distributions. CNR provided the HfO based resistive switching characterized with voltage pulse trains, while INI is an internationally recognized institute for neuromorphic bio-inspired systems. The first section of the chapter will be devoted to the description of the HfO based resistive switching and the method of characterization and empirical modeling adopted. The second part of the chapter will be instead dedicated to the description of

STDP in the form of Fusi's learning rule. To conclude, it will be described the network architecture, reported and discussed the learning performances of the network while trained with a database of hand-written digits known as MNIST, currently used as benchmark for learning tasks.

Chapter 2

Experimental

The present chapter collects all the information about device fabrication processes and material characterization setups, since most of them remain the same from device to device. The first section reports the recipe for nanotubes and thin film fabrication. The second one reports parameters adopted for material characterization techniques. If special conditions occurs for specific device characterizations, they will be specified directly in the referring chapters.
All fabrication processes discussed here are optimized in previous research performed by collaborators. References to scientific literature reporting the optimization process are reported in the corresponding section.

2.1 Device Fabrication

All the fabricated devices present a general trilayer structure, composed of bottom electrode (BE), active metal oxide, and top electrode (TE) deposited onto a substrate.

2.1.1 TiO_2 Nanotubes Array

Referring to TiO_2 nanotubes array (NTA), the detailed structure accounts a Titanium foil as both substrate and bottom electrode, TiO_2 NTA as active metal oxide with possibly Poly Acrylic Acid (PAA) coating and Platinum top electrode.
The Titanium foil (99.6% purity, thickness 200 μm, Goodfellow) was used as substrate. After cleaning via ultra-sonication in acetone, rinsing in ethanol and

nitrogen stream drying, the Ti foil was etched into 1 wt% HF solution in order to obtain a fresh metal surface for NT nucleation. The anodization process was carried out at room temperature in an ethylene glycol based electrolyte containing 0.5 wt% NH_4F and 2.5 vol% deionized water, in a two electrode electrochemical cell using a platinum sheet as counter electrode. The NTs growth was performed under continuous stirring applying a constant voltage of 60 V by a DC power supply (GW Instek SPD-3606). Optimized recipe at [43]. An anodization time of 30 s allowed the formation of an ordered array of nanotubes with a thickness of about 250 nm, well adhered to Ti foil substrate. The underlying Ti foil itself was used as bottom electrode for all the resistive switching devices. Some of the as-grown samples were thermally treated after anodic oxidation, being annealed in air at 450 °C for 1 h, in order to crystallize TiO_2 into anatase phase.
The thin film deposition of PAA (nominally 40 nm thick, if measured on a flat surface), was performed in a low pressure plasma enhanced CVD reactor (IONVAC PROCESS S.r.l., Pomezia, Italy), whose chamber base pressure was 37 mbar, ($3.7 \cdot 10^{+6}$ Pa) and the excitation radio frequency 13.56 MHz. The liquid monomeric precursor (acrylic acid, 99%) was stored in a quartz reservoir and the produced organic vapors were carried into the deposition chamber by bubbling argon in the reservoir. Acrylic acid vapors (3 sccm flow rate) were diluted in argon carrier gas (20 sccm flow rate), needed to sustain the plasma discharge. Polymerisation was performed by a pulsed plasma discharge with 200 W RF power, with 10% duty cycle ($t_{on} = 10ms, t_{off} = 90ms$) and 5 minutes deposition time. For further details and reference to the optimization process, refer to [44].
Circular Pt top electrodes (200 nm thick with diameters of 1 and 2 mm obtained through a shadow mask) were deposited at room temperature by DC sputtering in an Ar atmosphere of 10^{-4} bar, (10 Pa), (metal-coater Q150T-ES, Quorun Technologies) using a sputtering current of 50 mA for 180 s.

2.1.2 TiO_2 thin film devices

Metal-Insulator-Metal TiO_2 thin film devices account a Si wafer as substrate, Ti/Cu bottom electrode, ALD-deposited TiO_2 layer as active metal oxide and Pt top electrode. The fabrication process proceeds as follow, using standard ultraviolet (UV) photolithography and liftoff wet etching.

Silicon <100> p-type wafers with 4" diameters were used as substrates and washed in piranha solution followed by deionized water rinse. A thin Cu layer (200 nm thick) was deposited on a Si wafer by electron beam (e-beam) evaporation and used as a common bottom electrode for all devices without patterning. Because of the poor adhesion of evaporated Cu onto Si, a Ti adhesion interlayer (20 nm thick) was deposited during the same process before Cu evaporation.
The definition in the ALD oxide active layer was obtained via a liftoff procedure. First, a layer of photoresist was deposited on the Cu electrode and patterned by standard UV photolithography into circular pads with 1.2mm diameters. Then, 30 nm-thick TiO_2 thin films were deposited on the patterned photoresist via ALD in a Beneq TFS-200 reactor, using $TiCl_4$ as the precursor, H_2O as the coreactant and Ar as the carrier gas. Finally, wet liftoff etching of the TiO_2-coated photoresist was achieved via a mild sonication in dimethyl sulfoxide (DMSO) solvent at 50 °C, a subsequent DMSO rinse at room temperature, and a deionized water rinse and drying in N_2 gas.
The Pt top electrode (100 nm thick) was deposited via sputtering following a liftoff procedure similar to the one described above, to obtain circular electrodes 1 mm in diameter. To optimize the liftoff procedure for the patterning of the TiO_2 layers, several ALD tests at relatively low temperatures were achieved by varying the substrate temperature between 80 and 150 °C, while fixing the carrier gas flow rate (250 sccm), precursor pulse (100 ms), H_2O pulse (250 ms), purging times (2000 ms), and number of ALD cycles (400). A reference sample was deposited using the same growth conditions at a temperature of 230 °C. For the optimization process please refer to [20].

2.1.3 ZnO thin film devices

Metal-Insulator-Metal ZnO thin film devices account a Si/SiO_2 wafer as substrate, Ta/Pt bottom electrode, sputter-deposited ZnO layer as active metal oxide and either Pt or Cu top electrode. The fabrication process proceeds as follow.

Silicon <100> p-type wafers with 4” diameters were used as substrates and washed in piranha solution followed by deionized water rinse. Then, Ta(10 nm)/Pt(100 nm) bottom electrode was deposited at room temperature (base vacuum pressure 10^{-4}), in pure Ar atmosphere (gas pressure 1 Pa), by DC sputtering ($7.4 W cm^{-2}$ for Ta $5.9 W cm^{-2}$ for Pt) for 80 s and about 7 min, respectively. The thin Ta layer was

used to improve adhesion of BE to the underlying Si wafer, common to all devices present on the sample.
Deposition of ZnO thin films were performed by RF magnetron sputtering, starting from a 4" diameter ceramic ZnO target. Suitable vacuum conditions (base pressures ranging between $8.3{\cdot}10^{-6}$ and $2.5{\cdot}10^{-5}$ Pa) were obtained with a two-stage pumping system. Each deposition process was carried out by heating the Si/Ta/Pt substrates at 50°C and with a target-to-substrate distance of around 8 cm. All depositions were performed in a pure Ar atmosphere, with a fixed gas pressure of 0.66 Pa and a RF power density of 1.2 Watt/cm^2. An average deposition rate of 100 Å/min was achieved in such conditions. Prior to each ZnO deposition, pre-sputtering of the ZnO target was performed for 15 min to avoid any incorporation of contaminants in the grown films. For the optimization process please refer to [45, 46].
Finally, deposition of Pt and Cu circular top electrodes accomplished the fabrication of the MIM devices. To this purpose, Pt TE (200 nm thick, with diameters of 1 and 2 mm) were obtained deposited at room temperature by DC sputtering (metal-coater Q150T-ES, Quorun Technologies) in an Ar atmosphere of 10^{-4} bar, (10 Pa), using a sputtering current of 50 mA for 180 s. Cu TE (200 nm thick, with diameters of 1 and 2 mm) were deposited by thermal evaporation using a Cu target on Tungsten crucible, using an evaporation current of 125 A for 40 minutes. In both the cases, the desired geometry of TE was obtained by using a shadow mask.

2.1.4 HfO devices

Onto highly doped Si (0.001-0.005 Ω ·cm) substrate cleaned in HF from native oxide, 10 nm Ti and 40 nm TiN are grown by sputtering in 40 sccm Ar and mixed 4 sccm N_2 / 40 sccm Ar fluxes, respectively. After vacuum breaking, 5.5 nm HfO_2 is grown by atomic layer deposition in a Savannah 200 (Cambridge Nanotech) reactor with bis(methylcyclopentadienyl) methoxymethylhafnium(IV) (HfD-O4, by Sigma Aldrich) and water as Hf and O precursors respectively according to optimized recipes described at references [26, 47]. Top electrodes are defined by optical lithography and lift-off process after 50 nm Pt sputtering in 40 sccm Ar flux. Device lateral sizes used for this work are $40{\times}40\ \mu m^2$. The back of the substrates are coated with 100 nm Al by e-beam evaporation for lowering the contact resistance with the chuck of the electrical tests.

2.2 Material Characterization

The morphology of nanostructures and thin films was assessed by field-emission scanning electron microscopy (FESEM, ZEISS Auriga), equipped with an INCA Oxford Energy dispersive x-ray spectroscopy detector (EDX) for elemental analysis. X-Ray Photoelectron Spectroscopy (XPS) was performed to investigate the stoichiometry and chemical composition of materials, by using a PHI 5000 VersaProbe (Physical Electronics) system. The X-ray source was a monochromatic Al Kα radiation (1486.6 eV). XPS depth profile was obtained by means of an Ar+ flux at 2 kV accelerating voltage.
The crystal structure and orientation of the resistive switching devices was investigated by X-Ray Diffraction (XRD) measurements, using a Panalytical X'Pert Pro Diffractometer in Bragg-Brentano configuration (Cu Kα radiation, λ = 1.54059 Å). Structural characterization was assessed by Raman spectroscopy with a Renishaw inVia Reflex micro-Raman spectrophotometer, equipped with a cooled CCD camera. Samples were excited with a 514.5 nm wavelength solid state laser source.
Electrical characterization was performed using a Keithley 4200 Semiconductor Characterization System. I–V electrical measurements were performed in a two-point contact probe station at room temperature, maintaining the bottom electrode electrically grounded and the top electrode under DC voltage sweep. Pulsed electrical characterization is performed in a standard probe station equipped with Keysight B1500A instrument. Pulses are sent through a B1525A Semiconductor Pulse Generator Unit and current is read through a B1511B Source Measuring Unit both interfaced with the device through a custom board. Voltage is applied to top electrode while bottom electrode is kept grounded. Devices are forced with trains of identical pulses and the resistance is read after each pulse at 200 mV. Pulses are 30 μs-long with rise and fall times of 1 μs with applied positive voltage of 1.1 V and negative voltage of -0.8 V.

Chapter 3

TiO_2 based resistive switching

3.1 TiO_2 Nanotubes Array

Reference paper: Conti et al., ***Memristive Behavior in poly-acrylic acid coated*** *TiO_2* ***nanotubes arrays****, Nanotechnology 27 (2016) 485208 (10pp)*

TiO_2 has been extensively studied for its resistive switching properties since its modelling within the memristor formalism in 2008, [36]. A pletora of papers appeared in literature to investigate different deposition techniques, such as ALD [23] and sputtering deposition [48, 49], and different device architectures, like single layer TiO_2 devices [48], multi-layers devices [48, 50], doped TiO_2 devices [51]. A considerable amount of nanostructured TiO_2 based devices appeared as well, proposing nanotubes [52], nanowires [53] and nanorods [54] as viable structures for installing resistive switching mechanisms. Two main advantages derive from these peculiar structures: the first relies on the high surface to volume ratio, that could enable physical mechanisms not exploitable in thin film devices. The second one instead is based on the high aspect ratio of nanostructured devices, that still affecting the surface to volume ratio, represents also an advantage for interfacing artificial neuromorphic systems with the biologic counterpart [55].
Structures as TiO_2 NTA offers an extremely high control on the device fabrication, such high nanotubes height, diameter thickness, together with the low-cost parallel production of self-ordering, self-standing large arrays, [56, 57]. TiO_2 NTA are already widely studied for the many properties they express for various technological applications, like in dye-sensitized solar cells [43], water splitting devices

[58], Li-ion batteries [59, 60], supercapacitors [61], sensing devices (gas sensing [62], molecular sensing [63]) etc.. Moreover, NTA grows amorphous and can be crystallized to anatase phase through thermal annealing. They also present an inner multilayer structure that is more porous for the layer closer to the tube cavity and more stoichiometric at the tube bottom [57]. All of these properties could be advantageous for the RS properties and their fine tuning.
Application of a PAA coating goes in the direction of tuning RS properties of the tube. Nearly all of the works found in literature address the RS mechanisms observed in nanostructured devices (including TiO_2 NTA) to filamentary creation across the layers or to ions movement on top of the structure surface itself. Creation of localized ionic CF at the tube bottom could be advantageous for the control of RS electrical behavior, but on the contrary it requires an extremely ordered NTA, reachable only after multiple repetitions of the growth process and, even more important, a very high control on the deposition of metal top electrode, required to reach the tube bottom with internal surfaces coating, without creating shortcuts [64]. An homogeneous, partially conformal PAA coating on the other hand, does not imply specific requirements on the NTA matrix and, most of all, enable catalytic surface reaction that could affect the electric properties of the material, providing a fine tuning of voltage-driven RS.
The present study regards the complete characterization of four samples:

- as-grown $Ti/TiO_2/Pt$
- air annealed $Ti/TiO_2 - ann/Pt$
- as-grown, polymer coated $Ti/TiO_2/PAA/Pt$
- air-annealed, polymer coated $Ti/TiO_2 - ann/PAA/Pt$

A schematic of which is reported in figure 3.1, highlighting the NTA multilayer structures, panels a) and c), the PAA coating, panels b) and d) and a schematics of the equivalent electrical circuit created, panels e).

3.1.1 TiO_2 NTA structural characterization

The FESEM characterization of samples shows a regular array oriented orthogonally to the Ti foil, with an average height of 250 nm, external diameter of 120 nm and

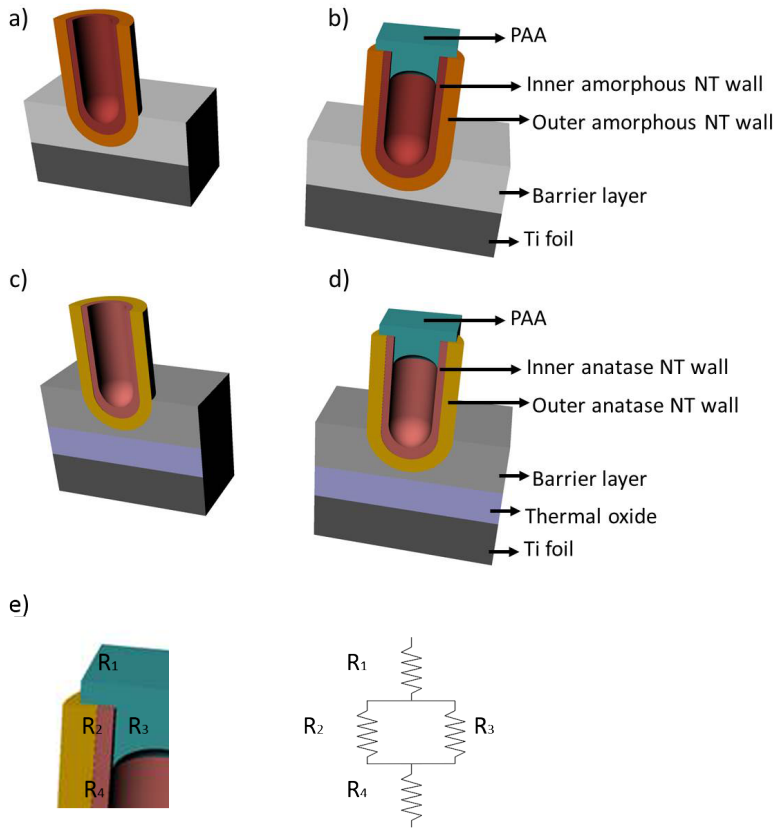


Fig. 3.1 3D drawing of the TiO_2 NTA samples, representing the multilayer structure of the nanotubes. Panels a) and b) show as-grown material without and with partially conformal PAA coating. Panels c) and d) show corresponding annealed NTA structure, highlighting insurgence of thermal oxide as intermediate layer. Panel e) shows the equivalent electric circuit schematics for the device.

internal diameter of the tube cavity that measured about 50 nm, see figure 3.2a). These dimensions are selected by time and voltage adopted for the anodic growth, since a linear relation exists between these quantities and the geometrical features expressed by the final tube, [57]. In particular, the application of 60 V for 30 s produces a bottom thickness of the tube of about 40 nm, so comparable with most of the thin layer studied found in literature for resistive switching investigation. During the nanotubes growth, anodic reactions induce the formation of a few nanometers-thick barrier layer at the nanotube base, see figure 3.1. The barrier layer presents a high concentration of impurities coming from the electrolytic solution that make the layer highly resistive.

A Pt top electrode was deposited onto the first set of as-grown NTA devices, in order to create electric contact. As it can be observed from the FESEM images,

figure 3.2b), Platinum nucleates on top of the NTA structure in a columnar grain configuration, feature possibly due to the particular pattern created by the top NTA openings. For this reason, Pt does not penetrate into the tube cavity and thus, direct metallic filamentary connection to the tube bottom are very unlikely.
On the second set of NTA devices, PAA was deposited on top of the array, creating an homogeneous and partially conformal coating of the nanotubes. As it can be seen in figure 3.2c), PAA fills the tube cavity for a depth comparable to the tube cavity diameter itself, creating a completely different equivalent electric circuit with respect to the one without polymer. Indeed, if the top layer of thin PAA creates a series resistance to the titania nanotube (R_1 in figure 3.1e)), within the first 50 nm close to the tube opening a parallel of two resistances is created, (R_2 and R_3), connected then with the tube resistance, R_4. Thanks to the functional carboxylic groups, PAA can exert polar and H-bond interactions with an oxygen-terminated surface (such as the TiO_2 one) and create a very stable surface interaction.
Generally, during a plasma polymer deposition, the plasma discharge activates the surface promoting adhesion and fragments the monomer molecules, [44, 65], that simultaneously recombine at the substrate surface. Therefore, a minimum thickness of the polymer is needed to ensure the presence of a crosslinked (polymerised) matrix also in z direction with respect to the surface plane. As reported in [66], the lowest thickness that allows to get a stable polymer is 40 nm. On the other hand a thicker film is not desirable as the 40 nm thick one already tends to hinder the NTA pores and thus produces a resistivity increase only.

In order to assess the chemical composition of the samples and the possible presence of impurities due to the anodic growth process, EDX analysis was conducted on equivalent samples $1\mu m$-thick, where only Ti NTA was present on Ti substrate, adopting an accelerating voltage of 15 keV. Data collected reported a relevant inclusions of Carbon and Fluorine, see figure 3.3, of 7% and 25% respectively, consistently with what reported in [43]. Thermal annealing of the sample lets outgas the impurities and a relative concentration of Ti:O elements of about 33% and 66% is obtained, i.e. returning a good stoichiometry for the nanostructured TiO_2.

Micro-Raman spectroscopy is performed in order to complete the structural analysis and assess the crystallinity of as-grown and annealed NTA. The collected data are compared with commercial TiO_2 crystalline powder with grain size in the range 15-20 nm, (Solaronix, Aubonne, Switzerland). As-grown sample returns a diffused photoluminescence, without the presence of clear peaks, as typical for

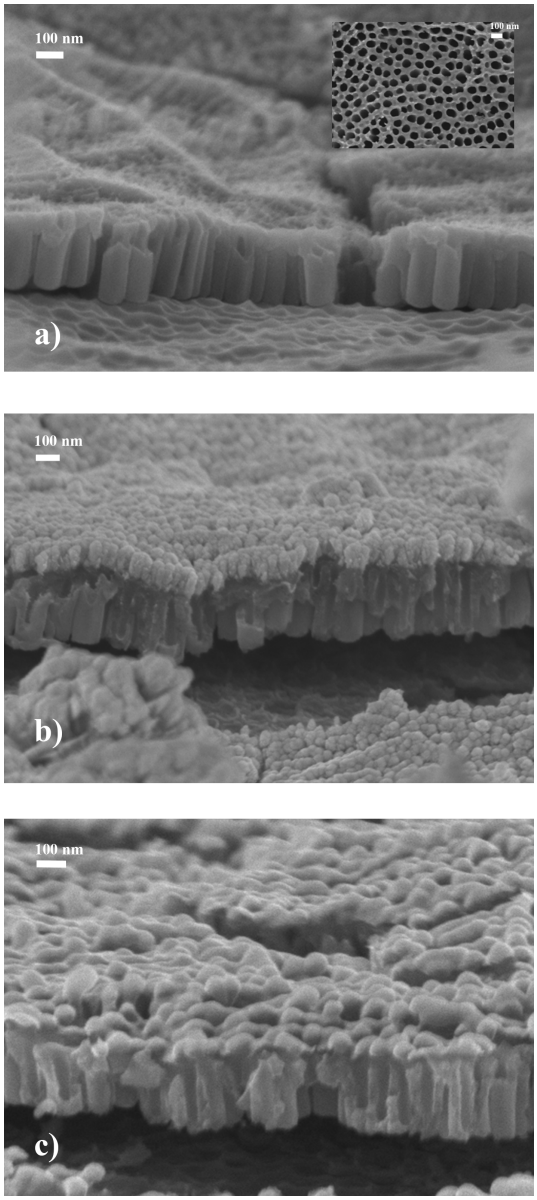


Fig. 3.2 FESEM characterization of NTA samples in top view, inset of panel a), and cross-section. Panel a), as grown material. Panel b), Pt coated NTA. Panel c), PAA coated NTA. PAA coating creates a homogeneous layer partially filling the tube cavities.

amorphous materials, see figure 3.4. Annealed NTA, on the other hands, showed Raman peaks typical of the anatase phase, in particular the Eg modes at about 144, 197, and 639 cm^{-1}, as well as two B_{1g} modes at about 399 and 519 cm^{-1}, showing high relative intensity. The latter result is comparable to the Raman spectrum

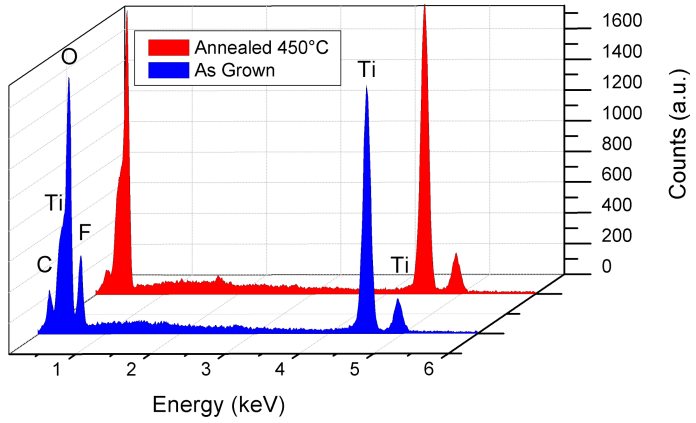


Fig. 3.3 EDX characterization of NTA samples. As-grown material shows typical peaks of Carbon and Fluorine (blue trace), while only Titanium and Oxygen are present in annealed sample (red trace).

obtained from commercial TiO_2 powder, assessing the anatase nanocrystalline nature of the annealed NTA.

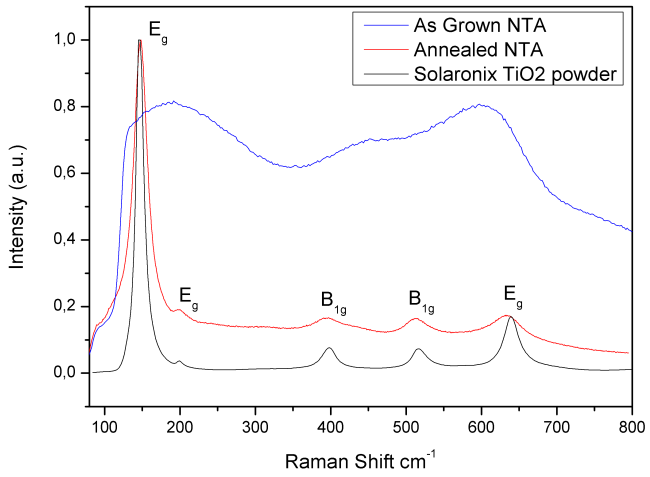


Fig. 3.4 Raman spectroscopy of as-grown (blue trace), and annealed (red trace) NTA devices. For comparison, it is reported Raman spectroscopy for commercial TiO_2 powder.

3.1.2 TiO_2 NTA electrical characterization

Electrical characterization of the NTA devices was performed in voltage sweep mode, applying voltage ramps up to $\pm 3V$. A current compliance, here named I_{cc} was also imposed in order to limit the maximum amount of current able to cross the device and prevent possible breakdown of the device itself. Since the quite large conductivity of the NTA samples, I_{cc} was kept in the range $1-10mA$.
In analyzing results obtained for RS behavior of TiO_2 NTA, the layered structure and the presence of defects should be considered, since the anodic titania act as a n-doped material. The best results found in literature for such a characterization are obtained by Schmuki et al., [52], for amorphous TiO_2 nanotubes, returning about 20 cycles with Off/On ratio of about 14. On the contrary, the annealed samples by that study returned only mild diode-like rectifying behavior.
In the following, uncertainty is considered as the cycle-to-cycle variation on the measured quantity and it is computed as standard deviation. Since results obtained from TiO_2 NTA electrical characterizations are not particularly numerous, in terms of number of data acquired, standard deviation may be slightly inaccurate and some other dispersion measures would be preferable, such as mean absolute deviation for example. It is here indicated as an approximate measure of data dispersion, but without the fully corrected statistical significance.

As-grown TiO_2 NTA was characterized applying voltage ramps in the range $[-1.5V, +1.5V]$, with imposed current compliance of 10 mA. The overall result obtained is an unstable resistive switching behavior, with very few (less than 10) cycles scarcely reproducible, see figure 3.5. The pristine resistance state of the device was equivalent to its low resistance state (LRS) and measured $R_{on} = 92 \pm 25\Omega$. When positive voltage was applied to the Pt electrode, the device exhibited a transition to a high resistance state (HRS) $R_{off} = 213 \pm 153\Omega$, with an average ratio of about $r \approx 2$. In the present section, ratio is computed considering measurements on few voltage ramp cycles with large cycle-to-cycle variability. Thus, ratio is meant here as a qualitative point of reference for understanding the device electrical behavior. As-grown device after few switching cycles came to a non-linear characteristic. It is evident then that such a devices cannot be adopted for any application since they produce unreliable electrical behavior.
When PAA coating is introduced on top of as-grown NTA, the measured initial resistance is $R_{on} = 970 \pm 292\Omega$ and a transition to HRS occurs in the positive

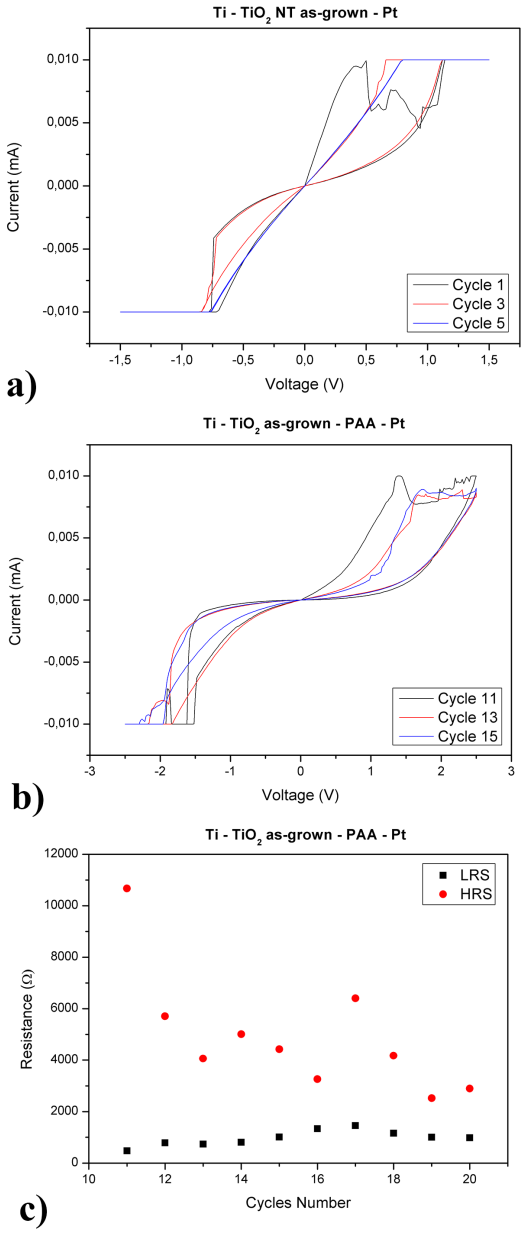


Fig. 3.5 Electrical characterization of as-grown NTA devices. Panel a), three qualitatively representative cycles for $Ti/TiO_2/Pt$ device. Panel b), three qualitatively representative cycles for $Ti/TiO_2/PAA/Pt$ device showing resistive switching behavior. Panel c), resistance states for the $Ti/TiO_2/PAA/Pt$ device, measured at $V = 0.1V$, for the first ten cycles.

polarity applied to Pt electrode, reaching a $R_{off} = (4.9 \pm 2.3)k\Omega$. Thus, the measured resistance Off/On ratio is $r \approx 5$, preserved for 20 cycles.

The first implication of PAA coating is an overall increased resistance, due to the series resistance introduced as in figure 3.1e). On the other hand, the presence of

the polymer tends to stabilize, in terms of endurance, the electrical behavior of NTA device and increase its Off/On ratio. A possible explanation of these facts can be found in the surface reactions induced by the polymer at the TiO_2 surface. Deprotonation of PAA carboxyl groups induces the adsorption of carboxylate onto the titanium dioxide surface, in bidentate configuration, [67]. Indeed, carboxylate provides negative charge to the oxide, allowing to increase conductivity by means of charge carriers' increment. A second and simultaneous reaction occurring at the NTA surface is due to the creation of hydroxyl groups. Hydrogen protons coming from deprotonated PAA react with TiO_2, by hydroxylation reaction, thus trapping an oxygen atom onto the substrate surface [67–69]. This process results in the generation of an oxygen vacancy and thus of a conductive region close to the surface that contributes to enhance the conductivity in the ON state.

Electrical characterization of annealed devices returned a highly resistive and non-linear behavior, coherently with what reported in [52]. Figure 3.6 reports qualitatively representative cycles for the device laying within cycle-to-cycle variability. The measured resistance state is $R_{on} = (26.3 \pm 1.1)k\Omega$ with a Off/On ratio between the two branches of hysteresis essentially equal to 1.
The main driver for electrical conductivity in a nanocrystalline material, like annealed TiO_2, are defects present in grain boundaries. It is established in literature how anodization conditions and post-processing of such a material could strongly affect the defect-based conduction of the specimens, altering surface states. In particular, temperature and annealing atmosphere could alter the material stoichiometry, reducing the number of oxygen vacancies and the related band gap trap states. Indeed, in literature it is reported how annealing process increases TiO_2 based devices conductivity. Conversely, in this case it is empirically observed an increment in the device resistivity, that is related to the low thickness of the NTA adopted. Indeed, thermal annealing produce also a thermal oxide layer at the interface between NTA and Ti substrate, see figure 3.1c). Being its thickness comparable with the thickness of NTA, it could strongly increase the overall device resistivity.
Annealed TiO_2 NTA, with PAA coating is characterized with voltage sweep within $[-1.5V, +1.5V]$ and $I_{cc} = 10mA$. The pristine device returned a LRS $R_{on} = (3.9 \pm 2.9)k\Omega$ and a transition with applied voltage in the positive polarity to HRS $R_{off} = (15.2 \pm 10.3)k\Omega$, with ratio $r \approx 4$ and endurance of 40 cycles. It should be highlighted that having such a large variation on the resistance values, the absolute values of series and parallel resistances due to PAA can be neglected and indeed,

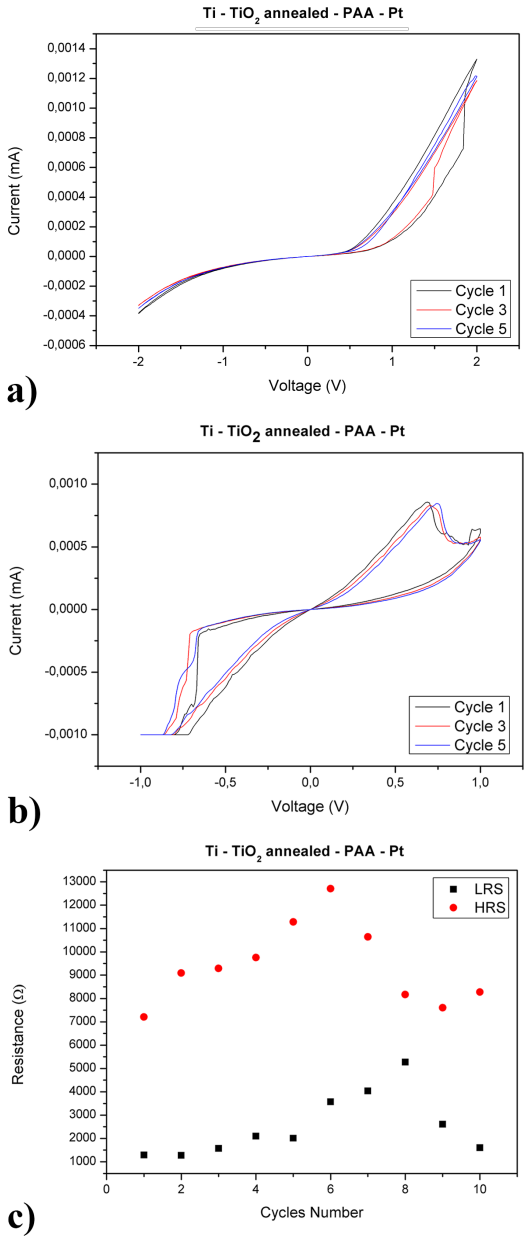


Fig. 3.6 Electrical characterization of annealed NTA devices. Panel a), three qualitatively representative cycles for $Ti/TiO_2 - ann/Pt$ device. Panel b), three qualitatively representative cycles for $Ti/TiO_2 - ann/PAA/Pt$ device showing resistive switching behavior. Panel c), resistance states for the $Ti/TiO_2/PAA/Pt$ device, measured at $V = 0.1V$, for the first ten cycles.

considering the large variability that characterize these samples, measured HRS of $TiO_2 - ann/PAA$ is comparable with resistance state measured for $TiO_2 - ann$ device.

Comparing all the obtained results, collected in table 3.1, at least two major features worth to be highlighted. First, PAA enable resistive switching with nanotubes, showing longer cycle endurance and higher Off/On ratio, likely due to the larger income of oxygen vacancies brought to the metal oxide by surface reactions.
Moreover, all devices show a pristine state equivalent to their LRS. This fact can be explained considering the growth dynamics of nanotubes. The self-aligned NTA was obtained with the application of an electric field able to activate redox reaction at the interface of the Ti substrate. Under the effect of the field, the Ti^{4+} ions at the metal surface interact with the OH^- ions present in the fluorine solution, producing oxidized Ti species. Dissolution of Ti ions moving toward the metallic cathode create fractures at the metal surface that acts as porous growth centers for the nanotubes. The weakened Ti-O bound allows for negative oxygen to move toward the bottom of the pore, establishing a positive feedback loop that generates the tube growth. When the anodic process is abruptly interrupted, excess of oxygen vacancies remain trapped at the tube bottom and result in an oxygen deficient, i.e. under-stoichiometric, TiO_2 layer resulting in the LRS of the NTA devices. As soon as RS is imposed, the oxygen vacancies are rearranged and the CF is interrupted.

Table 3.1 NTA devices electrical performances summary.

Sample	R_{on} (Ω)	R_{off} (Ω)	Ratio	Endurance
$Ti/TiO_2/Pt$	92 ± 25	213 ± 153	≈ 2	10
$Ti/TiO_2/PAA/Pt$	970 ± 292	$(4.9 \pm 2.3)10^3$	≈ 5	20
$Ti/TiO_2 - ann/Pt$	-	$(26.3 \pm 1.1)10^3$	≈ 1	-
$Ti/TiO_2 - ann/PAA/Pt$	$(3.9 \pm 2.9)10^3$	$(15.2 \pm 10.3)10^3$	≈ 4	40

3.1.3 Conclusions

In this study, a complete characterization of TiO_2 nanotubes array devices was given, investigating the electrical RS behavior of as-grown and annealed devices as well as both devices decorated with PAA coating. Structural characterization highlighted a regularly organized structure of nanotubes, orthogonally oriented with respect to the Ti substrate. The Pt top electrode nucleated on top of the nanotubes opening in a columnar configuration, preventing the metal from complete filling of the tube cavity. On the other hand, PAA coating provides a partially conformal coating of the tubes, with a penetration within the tube cavity to a depth equivalent to the

tube internal diameter. This fact implies a different equivalent electric circuit and, more important, the presence of surface reactions between the carboxylate groups of PAA and defective TiO_2 present at the tube surface. Raman spectroscopy and EDX characterizations show an amorphous and defective material, with considerable amount of Carbon and Fluorine absorbed, when as grown. On the other hand, thermal annealing in air produces a nanocrystalline material with good stoichiometry. Annealing causes also the creation of a thermal oxide layer at the interface between NTA and metallic electrode, responsible for a relevant increase in the overall device resistance.
Electrical characterization reported an unstable RS behavior for as-grown material, slightly improved when PAA coating is introduced. Annealed device, on the contrary, returns a diode-like behavior with a higher resistance expressed. Such a behavior is drastically changed by the introduction of PAA coating, which separates the two RS branches up to a ratio $r \approx 4$ with an endurance of 40 cycles. These major changes are presumably due to the surface reaction happening at the tube cavity surface, where carboxylate groups react at the TiO_2 surface and become adsorbed onto the metal oxide surface, providing negative charge available to transport that lower the value of LRS. Moreover, a second reaction happens at the same time between the hydrogen proton coming from carboxylate groups deprotonation and reacts with oxygen at the oxide interface, creating OH groups. Consecutively, an oxygen vacancy available for the CF during RS is created and contributes to the enhancement of conductivity in the ON state.
This analysis manifests the primary role of oxygen vacancies in RS and shows how RS performances can be tuned by controlling their concentration within the metal oxide, in order to obtain longer endurance and higher Off/On ratio. On the other hand, nanostructured TiO_2 devices offer poor performances for real devices applications and should abandoned in favor of thin film devices.

3.2 TiO_2 Thin Film

Reference paper: Porro et al., ***Low-temperature Atomic Layer Deposition of*** TiO_2 ***thin layers for the processing of memristive devices****, J. Vac. Sci. Technol. A 34(1), Jan/Feb 2016*

Thin film devices are favorable for integrated chip accounting for very large number of devices, because of the control they offer in the fabrication process and the opportunity to design very dense circuits on the same wafer. Indeed, progress in optical lithography sustained the Moore's law toward the scaling dawn of 2D technologies, allowing to embed on a single Silicon wafer billions of basic logic units in a single fabrication process.
Atomic Layer Deposition (ALD) is a key technological resource in such a fabrication process, thanks to its self-limited and self-organized deposition mechanism, its atomic thickness control and the relaxed constraint of line-of-sight deposition, [70]. It is a technique widely used also for the resistive switching technology, receiving strong impulse by microprocessors company as well that see the memristive technology as natural substitute of the Silicon technology in RAM fabrications, [10]. For this reason, starting from the HP lab first study on TiO_2 resistive switching in 2008, a pletora of scientific articles appeared in literature discussing properties of this material, most of which treated TiO_2 deposited via ALD. ALD allows the fabrication of a certain variety of TiO_2 based RS, such as single layer RS [36, 48], multilayer RS [50], doped RS [51] etc.
The present section regards the study of titanium dioxide based RS deposited via ALD at lower deposition temperature (i.e. < 150°C) with respect to the usual temperature window normally adopted (i.e. 200 - 250°C), [71]. The scope of such analysis is to enable a simple fabrication process for thin film devices based on TiO_2 that allow to skip complicated and inaccurate dry etching steps in the patterning process, [72]. Indeed, lower temperature deposition allows the metal oxide to be deposited directly onto pre-patterned photoresist and thus, to obtain the RS device simply via the easier lift-off technique. The study shows how with this simplified fabrication recipe, similar resistive switching electrical characteristics can be obtained.
The fabrication process proceeds as described in the previous chapter, following the steps depicted in figure 3.7.

First, Cu bottom electrode 200 nm thick was deposited as bottom electrode onto the substrate. A thin layer, 20 nm, of Ti was deposited in between to enhance adhesion of the metal layer onto the substrate. A subsequent step of standard optical lithography was performed in order to deposit and pattern the photoresist. ALD was then performed in order to deposit a metal oxide coating onto the complete substrate. Metal oxide was deposited at different temperatures, in between 80°C and 140°C, in order to assess the material quality for device production and RS properties. Lift-off

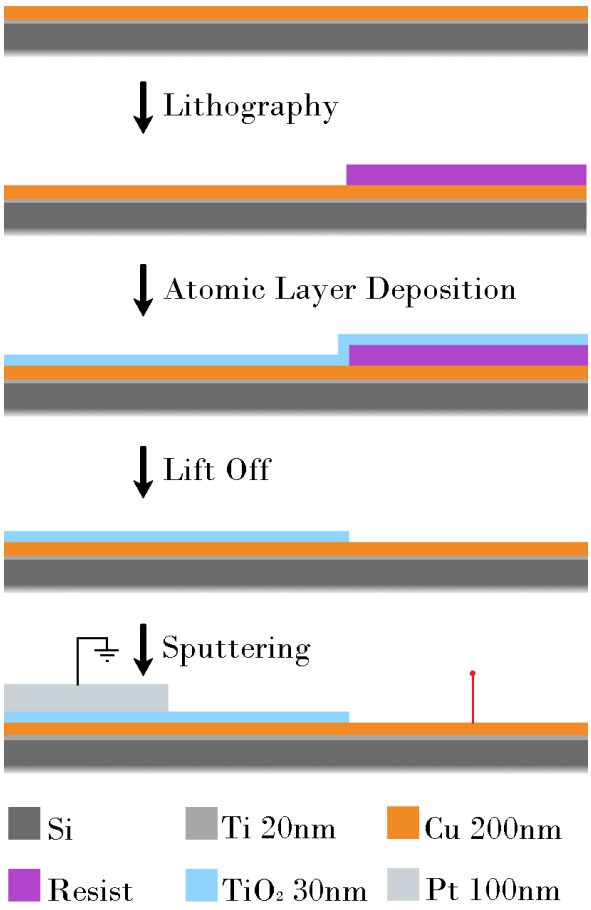


Fig. 3.7 Schematics of the process flow for fabrication of Si/ Ti(20nm)/ Cu(200nm)/ TiO_2(30nm)/ Pt(100nm) multilayer structure.

technique removed the metal oxide layer where it was not required by patterning. In conclusion, top Pt electrode was deposited through a shadow mask.

3.2.1 Material Characterization

FESEM analysis of the deposited layers show a smooth and continuous thin film successfully deposited in all the temperature range considered. See figure 3.8a) for a reference top view picture of the sample deposited at 120°C. Anyhow, inset of the same figure shows the formation of inclusions sparingly grown onto the surface. Such inclusions are much less present in higher temperature deposited samples, see figure 3.8b), and can be considered as by-product of the deposition parameters chosen, that will be discussed later in this section. FESEM investigation confirmed

that lift-off process was performed correctly, without any degradation of both metal oxide film and photoresist during the deposition process, see figure 3.8c).

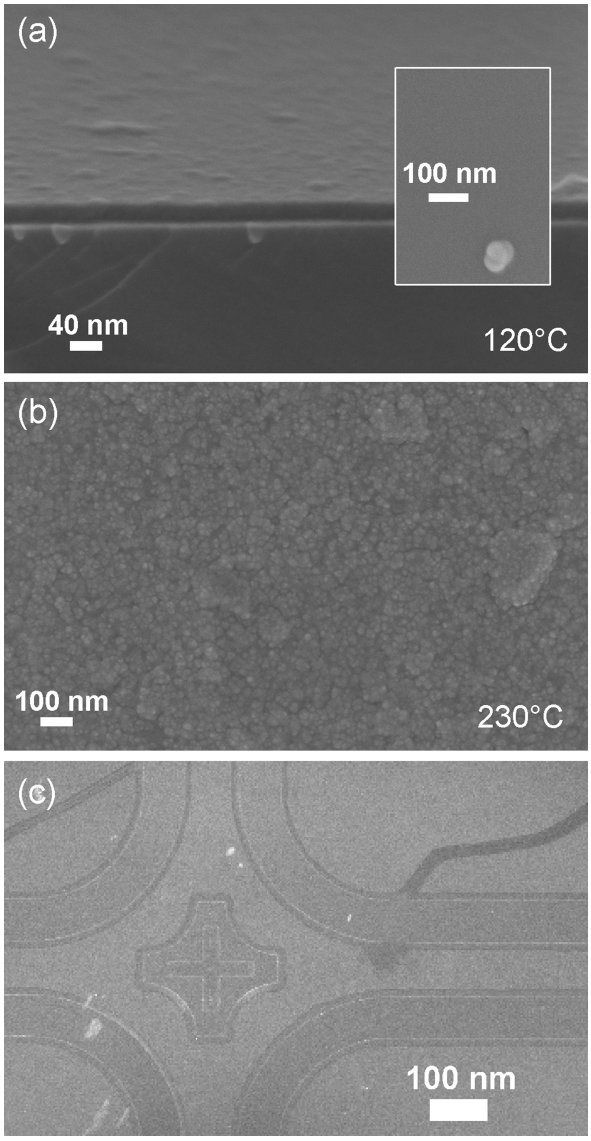


Fig. 3.8 FESEM characterization of TiO_2 samples. Panel a), cross section and top view (inset) of material deposited at 120°C. Panel b), top view of material deposited at temperature of 230°C. Panel c), top view of sample at 120°C patterned via DMSO lift off.

Figure 3.9 shows the growth per cycle (GPC) measured for the deposition processes at lower temperature. It is observed an inversed proportionality between GPC and deposition temperature, with as a reference about $\approx 0.5\text{Å}/cycle$ at 140°C and $\approx 1.1\text{Å}/cycle$ at 80°C.

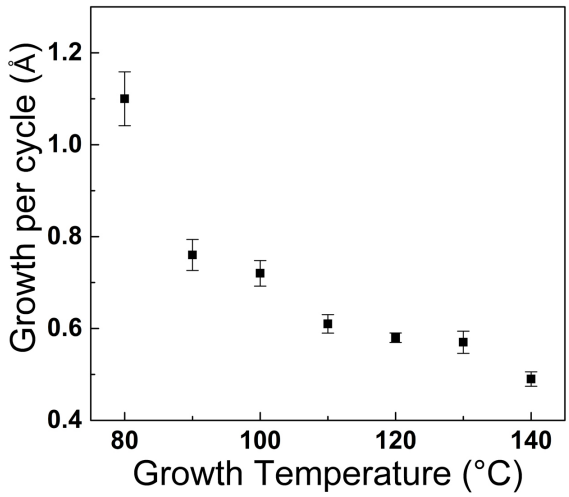


Fig. 3.9 Growth per cycle for 400 cycles of TiO_2 as a function of ALD deposition temperature.

High growth rate for metal oxide thin film could possibly lead to a less dense film reacher in localized defects. Indeed, two principal processes could occur as consequence of the low temperature, i.e. condensation and partial surface reaction during the ALD process. H-bonded OH groups adsorbed onto the substrate surface dominate the ALD reaction during the deposition process and condensation of these groups contributed to the generation of active sites for growth of both precursors and by-products of the deposition, contributing to the generation of deposition defects. For this reason, the temperature windows identified for the RS devices fabrication was limited to 120-140°C.

Raman spectroscopy was then performed for the samples deposited at 120°C, 130°C, 140°C and 230°C, taking a well crystallized Solaronix TiO_2 commercial powder as reference; data collected are reported in figure 3.10. Raman peaks typical for the crystalline TiO_2 in the anatase phase are observed. Well-crystallized TiO_2 shows three E_g vibration modes at ≈144, 197 and 639 cm^{-1} and two B_{1g} vibration mode at ≈ 399 and 519 cm^{-1}, that are comparable with what observed and reported in table 3.2. In order to assess the amorphous to crystalline ratio of the samples deposited at different temperature, peak fit was performed. Peaks center and full width at medium hight (FWMH) were compared. As can be observed from data reported in table 3.2 and 3.3, the three samples deposited within 120-140°C show similar features that, regarding the FWMH, can be collectively considered higher than the reference Solaronix powder, indicating a larger amorphous to crystalline ratio. On the other hand, peak centers show a shift that is symptomatic of a residual

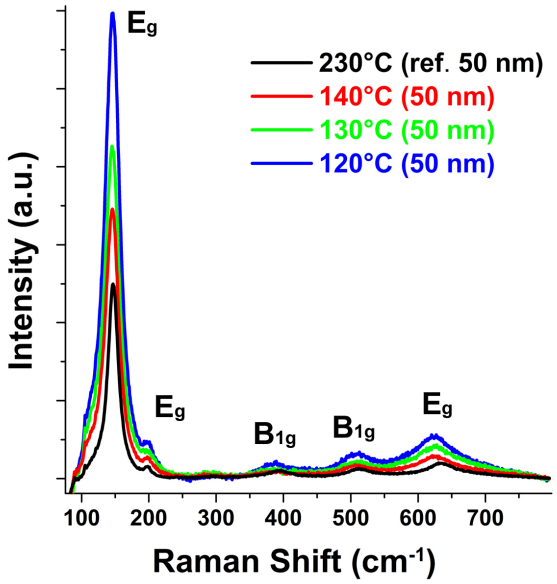


Fig. 3.10 Raman spectra fo TiO_2 deposited on Si wafer with Pt interlayer at 120-230°C. Measured peaks, as reported in table 3.2, are typical of TiO_2 in the anatase phase.

tensile stress in the substrate during deposition.

Table 3.2 TiO_2 Raman peaks center comparison.

Sample	Peak E_g	Peak B_{1g}	Peak B_{1g}	Peak E_g
	$144cm^{-1}$	$399cm^{-1}$	$519cm^{-1}$	$639cm^{-1}$
TiO_2-120C	145.7	384	507	626
TiO_2-130C	145	381	na	627
TiO_2-140C	145	381	509	628
TiO_2-230C	147	392	512	635
Solaronix	146.9	397.7	518	639.6

In conclusion, XPS was performed to assess the chemical composition of the samples and comparable and good stoichiometry was measured for all the low temperature devices, with $Ti \approx 33\%$ and $O \approx 66\%$, as shown in 3.11 for sample at $120°C$ for reference. Main peaks identified during XPS analysis are reported instead in figure 3.12, showing that no relevant concentration of impurities is present.

Table 3.3 TiO_2 Raman peaks FWMH comparison.

Sample	Peak E_g	Peak B_{1g}	Peak B_{1g}	Peak E_g
	$144cm^{-1}$	$399cm^{-1}$	$519cm^{-1}$	$639cm^{-1}$
TiO_2-120C	27.8	42.3	41.3	55.4
TiO_2-130C	27.7	na	na	55.4
TiO_2-140C	26.6	43.1	40.3	52.9
TiO_2-230C	19	32.9	35	45.2
Solaronix	13.8	23	24	25

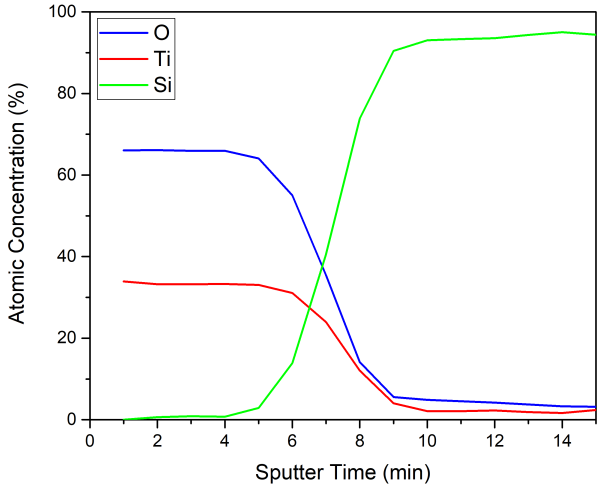


Fig. 3.11 XPS depth profile of TiO_2 thin film device deposited at $120°C$.

3.2.2 Electrical Characterization

The produced devices are electrically characterized in voltage sweep mode, with applied voltage ranging in between $[-1,+1]V$ and applied current compliance of $100\mu A$. The current compliance is set in order to reduce the probability of irreversible breakdown of the device during the soft breakdown imposed by the applied voltage. The measured pristine state of TiO_2 thin films is in the order of $10^{+7}\Omega$ and requires a higher applied voltage in order to create the first conductive filament bridging the two electrodes. Such a process is called *electroforming* and it is needed in particular with highly resistive thin films, such as TiO_2, HfO and Ta_2O_5. During this process, with devices in ECM configuration like these, positive voltage is applied to the Cu electrode, forcing the anodic dissolution of Cu^{2+} ions into the metal oxide

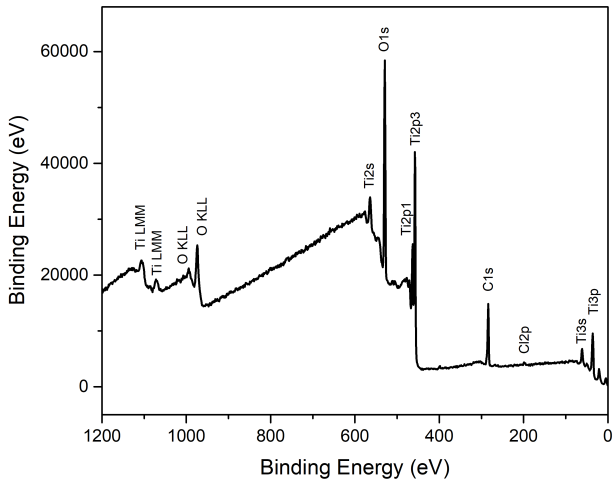


Fig. 3.12 XPS core levels of TiO_2 thin film device deposited at 120°C.

matrix. Positive ions are drifted toward the counter electrode until they create a metallic CF that brings the device in its LRS. A sufficiently high applied voltage in opposite polarity then is able to interrupt the CF because of the generated Joule heating. The metallic CF obtained in ECM are shown to be in conical shape and thus are likely interrupted close to the opposite electrode with respect from the one they originated, due to the thinner extreme of the CF itself and the reduced thermal conductance toward the electrode, where heat is easily dissipated.

Figure 3.13 shows representative cycles obtained for samples at 120°C, 140°C and 230°C, respectively panel a), b) and c). Table 3.4 reports a summary of results obtained for measured electrical behavior, highlighting resistance states and ratio with their respective cycle-to-cycle standard deviation. As it can be observed, performance of the three devices are comparable, with pristine state in the order of hundreds of $M\Omega$ and similar LRS, HRS and ratio. Standard deviations manifest a high variation on RS behavior quite typical for RS based on titanium dioxide and could easily make negligible possible variations due to the lower deposition temperature.

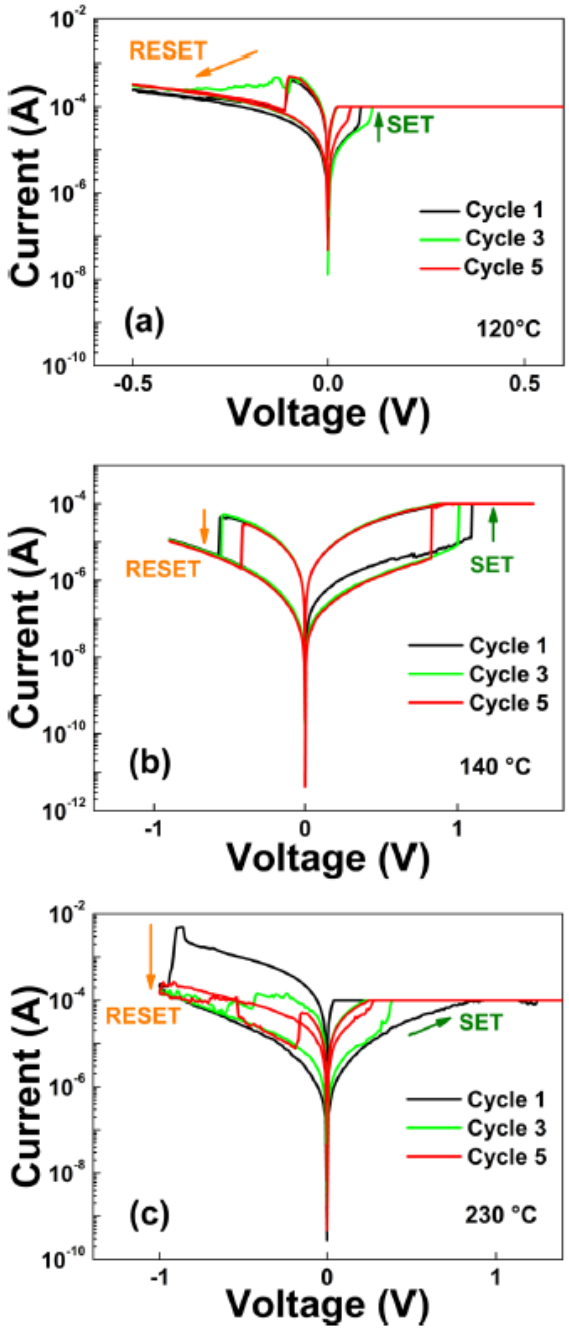


Fig. 3.13 Semilogarithmic plots of electrical characterization for TiO_2 devices deposited at different temperatures. Three representative cycles show comparable bipolar resistive switching behavior for devices deposited at 120°C, panel a), 140°C, panel b), and 230°C, panel c).

Table 3.4 TiO_2 thin film electrical performances summary.

Sample	$R_{Pristine}$ $(10^8\Omega)$	R_{on} $(10^2\Omega)$	R_{off} $(10^4\Omega)$	Ratio
TiO_2-120C	32 ± 13	0.35 ± 0.06	4.5 ± 2.0	9 ± 3
TiO_2-140C	0.66 ± 0.08	60 ± 20	260 ± 3	23 ± 8
TiO_2-230C	(0.60 ± 0.38)	2.1 ± 1.2	15 ± 10	17 ± 15

3.2.3 Conclusion

In this study, characterization of TiO_2 thin layer deposited via Atomic Layer Deposition at low temperature was given. ALD offers great advantages from the technological standpoint in terms of control of material deposition, conformal deposition without line-of-sight requirements and density of nodes producible on single wafer. On the other hand, TiO_2 patterning suffers of poorly controllable and inaccurate dry-etching process. Lower temperature deposition then becomes fundamental in enabling accurate patterning of the metal oxide through standard optical lithography, with ALD deposition of titanium dioxide directly onto the photoresist layer needed for the process.
In the present analysis, a temperature window between $[80°C, 140°C]$ was explored, well below the standard process temperature of 230°C. FESEM analysis shows a well patterned, compact and adhered layer to the substrate with sparingly grown inclusions. Such defects are likely to be produced as a consequence of the high growth per cycle observed at low temperature depositions. Indeed, high GPC allows inclusions due to condensation and partial surface reaction during the deposition process. An inverse proportionality between deposition temperature and GPC is observed. Raman spectroscopy shows TiO_2 crystallized in the anatase phase and fitting of peaks compared to commercial Solaronix powder highlights a higher amorphous to crystalline ratio for the deposited material. In conclusion, XPS analysis shows good stoichiometry of the layer, with 66% Oxygen and 33% Titanium relative concentration and no relevant impurities concentrations.
The electrical characterization of devices is performed in voltage sweep mode in the maximum range $[-1V, +1.5V]$. Current compliance is set to $I_{cc} = 10^{-4}A$ only in the positive polarity, in order to prevent damages to the dielectric because of too high current densities. When positive polarity is applied to the Cu electrode, SET transition is observed for all the tested devices, i.e. devices at 120°C, 140°C and 230°C deposition temperature; corresponding RESET transition is observed in the

opposite polarity. Large variability is observed in all the electrical characteristics, as typical for RS devices yet, but surely they can be considered comparable.
Low temperature deposition of ALD TiO_2 layer has demonstrated the successful patterning of devices via standard lithography, without impeding or relevantly affecting the resistive switching behavior of the produced devices.

Chapter 4

ZnO based resistive switching

4.1 ZnO Thin Films

Reference paper: currently undergoing writing

A different material was studied for the realization and characterization of thin film resistive switching, i.e. ZnO, zinc oxide. In particular, ZnO thin films are widely studied for the many properties they express, like piezo-electricity [73, 74], energy nanogeneration [75] and negative capacitance [45], optical properties [76], semiconductor properties [77, 78] etc. ZnO attracted also great interest due to its easiness of preparation, both as thin layer and nanostructure [79], device integration and electrical behavior. A plethora of methods is available for ZnO thin films preparation, like pulsed laser deposition [80], atomic layer deposition [23, 81], sputtering [45], chemical vapour deposition and hydrothermal synthesis [82], etc.. Currently, sputtering represents one of the most promising synthetic strategies for growing ZnO thin films with well-defined unipolar/bipolar RS behaviors, [83, 84]. Specifically for the research group from which this dissertation originates, ZnO offers the opportunity to compare resistive switching behavior originated from great variety of deposition and growth techniques. In particular, within the group are fabricated and investigated resistive switching produced via sputtering deposition, atomic layer deposition and chemical vapor deposition, in configuration of both single ZnO nanowire and nanowires forest grown onto substrate. Such a great opportunity allows to deepen the knowledge about ZnO properties due to different deposition methods, further understand the mechanisms that lead to resistive switching and engineer

devices toward the desired RS properties. The present dissertation, with this chapter, will focus on sputter-deposited ZnO RS.
In particular, Zno thin film RS are produced in both ECM and VCM architecture, adopting either Copper or Platinum as top electrode. Moreover, multiple device heights are produced in order to investigate the influence of thickness and consequent crystallinity on the RS properties. The studied devices are reported in the following:

- $Ta/Pt/ZnO-50nm/Pt$
- $Ta/Pt/ZnO-50nm/Cu$
- $Ta/Pt/ZnO-100nm/Pt$
- $Ta/Pt/ZnO-100nm/Cu$
- $Ta/Pt/ZnO-250nm/Cu$

With bottom 10nm-thick Tantalum as adhesion layer and Platinum as bottom electrode.

4.1.1 Material Characterization

Before analyzing electrical RS properties, morphology, chemical composition and crystal structure of the corresponding ZnO thin films are investigated.
Morphology of ZnO thin films is studied by FESEM analyses, both in top view and cross section. Figure 4.1a) shows the top view image of sample ZnO - 50 nm. The surface shows a fine-grained structure, with round-shaped grains and a typical size of 20 nm.The cross section view is shown in figure 4.1b) and highlights the presence of columnar grains oriented perpendicularly with respect to the Si/Ta/Pt substrate. Such a structure is typical for sputtered ZnO thin films, [85], and it is even more remarkable for the thicker ZnO samples, as shown in the next.

Figure 4.2 shows the cross sections FESEM images obtained for samples ZnO - 100 nm and ZnO - 250 nm. The corresponding top-views are shown in insets of the respective figures. Similarly to sample ZnO - 50 nm, also in this case the columnar grain structure of the material is observed. However, some differences can be appreciated. At first, the lateral size of the columns increases for higher ZnO thicknesses. Then columnar grain assumes a more defined geometrical shape, that

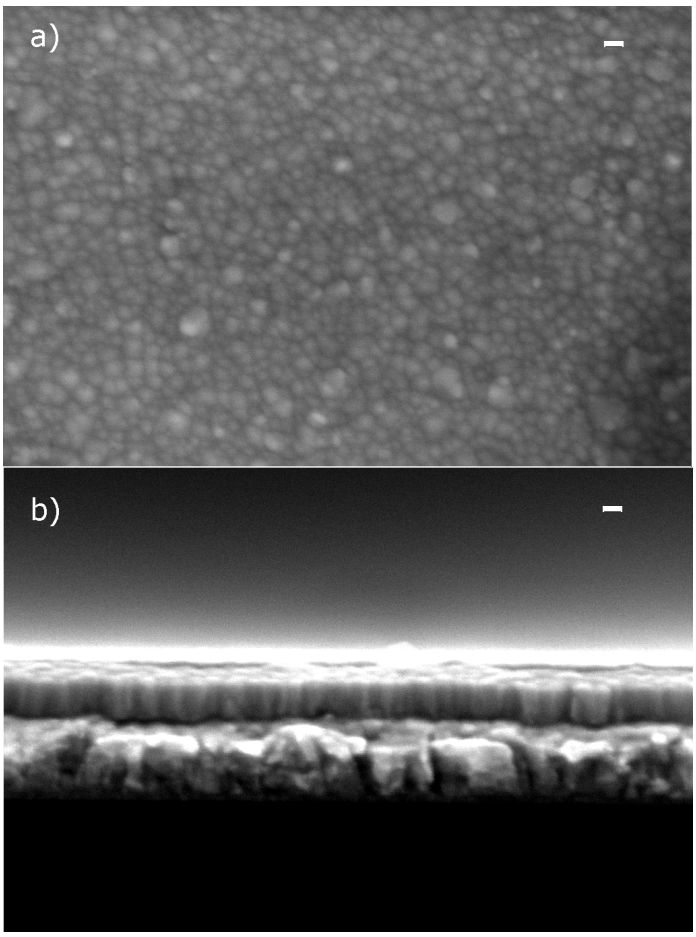


Fig. 4.1 FESEM characterization of ZnO - 50 nm sample deposited on Si/Ta/Pt substrate. Panel a), top-view. Panel b), cross-section. In both the cases, images are acquired at 500k magnification while the scale bar is 20 nm.

as obtained by XRD analysis reported in the following, approaches the hexagonal structure characterizing the wurtzite phase typical of thickest ZnO film (250 nm).

The compositional analysis of ZnO thin films is performed on sample ZnO - 100 nm as reference. Since all deposition parameters except for the deposition time were kept fixed for all the investigated ZnO thin films, no relevant changes in the oxide stoichiometry and composition are expected for the remaining ZnO samples. Figure 4.3 shows the XPS depth-profile analysis obtained for sample ZnO - 100 nm deposited on Si wafer. A quite good stoichiometry ratio between Zn and O is observed within the oxide layer. This is due to the mutual combination of several effects arising from the particular choice of the deposition parameters. Firstly, the use of a stoichiometric ceramic ZnO target, secondly the presence of inert Ar atmosphere,

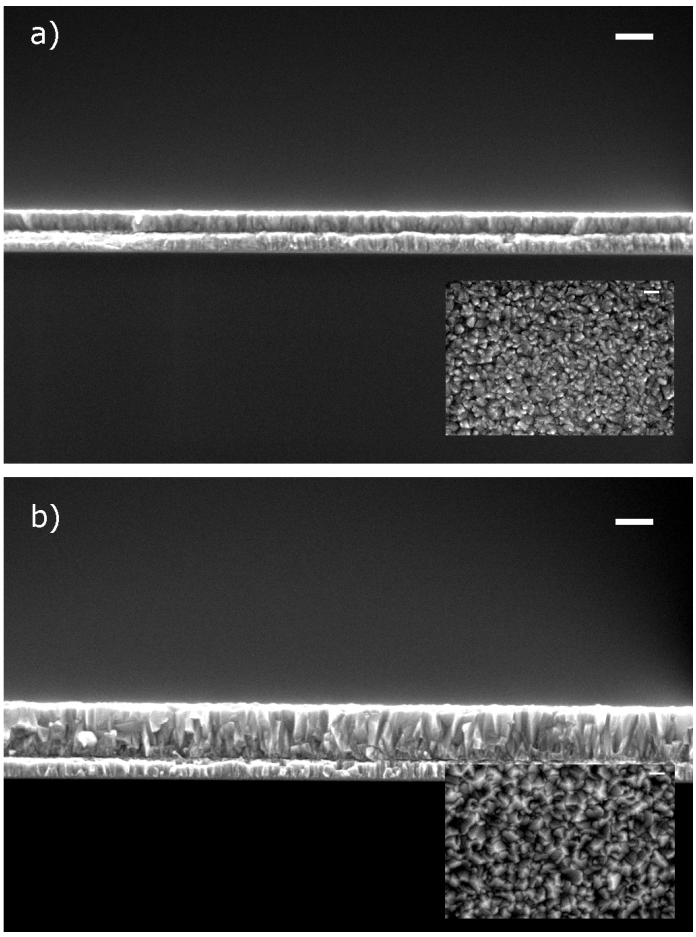


Fig. 4.2 FESEM characterization of ZnO samples. Panel a), ZnO - 100 nm layer as observed in cross section with SEM at 100k magnification, the shown scale bar is 200 nm. Panel a) inset, ZnO - 100 nm layer as observed in top view with SEM at 250k magnification, the shown scale bar is 100 nm. Panel b), ZnO - 250 nm layer as observed in cross section with SEM at 100k magnification, the shown scale bar is 200 nm. Panel b) inset, ZnO - 250 nm layer as observed in top view with SEM at 250k magnification, the shown scale bar is 100 nm.

thirdly the use of an optimal RF power density as defined in section 2, allowing for the incorporation of oxygen within the growing films, avoiding the undesirable oxygen deficiency that generally affects sputtered ZnO thin films.

The XPS core levels data reported in figure 4.4 for the same sample as above shows that no impurities in relevant atomic concentration are present in the thin layer produced.

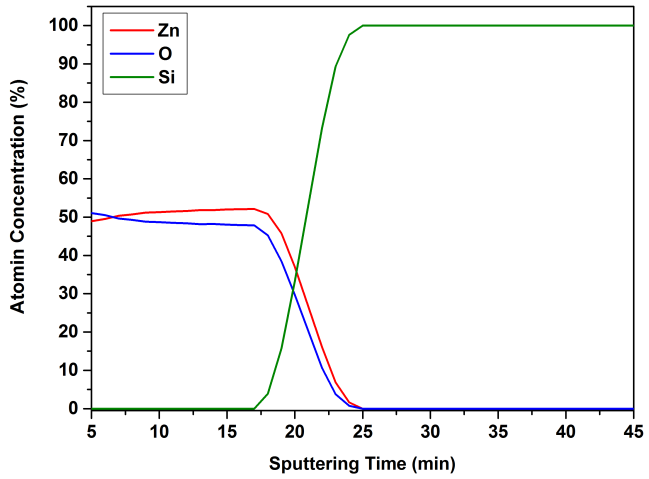


Fig. 4.3 XPS depth profile of ZnO-based MIM memristive devices, at ZnO film thickness of 100 nm as reference.

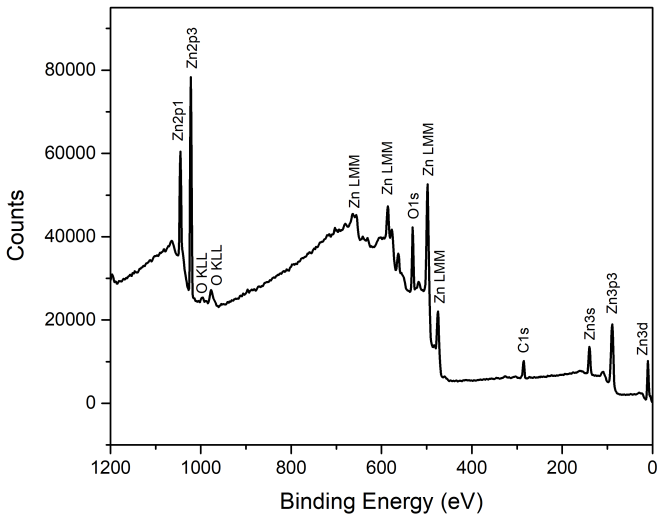


Fig. 4.4 XPS core levels spectrum of ZnO-based MIM memristive devices, at ZnO film thickness of 100 nm as reference.

Raman spectroscopy shows the typical peaks of nanocrystalline ZnO in their direct modes. As it could be expected, crystallinity of the material is preserved with the film thickness and peaks in Raman analysis becomes more and more intense, see figure 4.5.

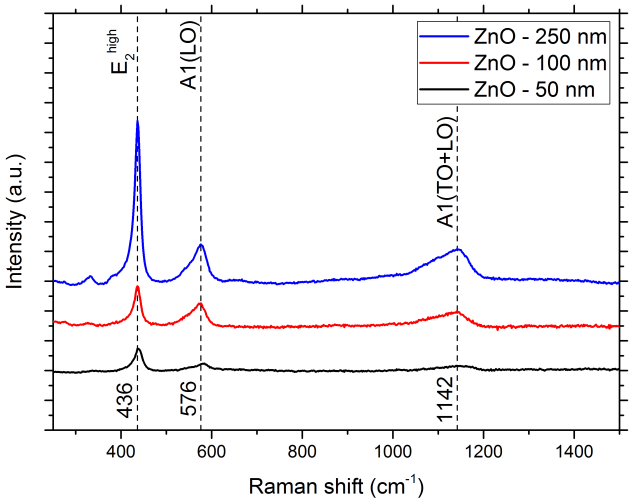


Fig. 4.5 Raman spectroscopy of ZnO samples at 50 nm thickness (black trace), 100 nm thickness (red trace), 250 nm thickness (blue trace)

Figure 4.6 shows the XRD spectrum of the MIM devices fabricated with differently thick ZnO films, i.e., 50 nm, 100 nm, and 250 nm. Apart from the strong diffraction contribution due to Si wafer and Platinum electrode, only reflections coming from ZnO wurtzite phase were detected. This is represented by the presence of (002) and (101) diffraction peaks, located at around 34.4° and 36° 2θ positions. The intensity of the corresponding peaks increases in line with the ZnO film thickness. In particular, very broad and weak peaks are detected for the thinnest ZnO film (50 nm), see figure 4.7. Both the ZnO diffraction contributions become more intense as the film thickness is increased up to 100 nm and, more noticeably, up to 250 nm. Differences in the crystal orientation of the analyzed ZnO samples may be noticed considering the corresponding $I(002)/I(101)$ intensity ratio; it changes from 0.7 to 1.1, when the ZnO thickness increases from 100 nm to 250 nm. This is due to the different intensities of the corresponding (002) peaks. Indeed, while the ZnO (101) peak does not remarkably change, the intensity of the ZnO (002) peak, due to diffraction of crystal planes oriented perpendicularly with respect to the thin-film growth direction, becomes the higher the more film thickness increases. This also stands for the existence of a preferential orientation within the ZnO structure along the c-axis crystal direction, consistently with the columnar grain structure previously shown in figure 4.2.

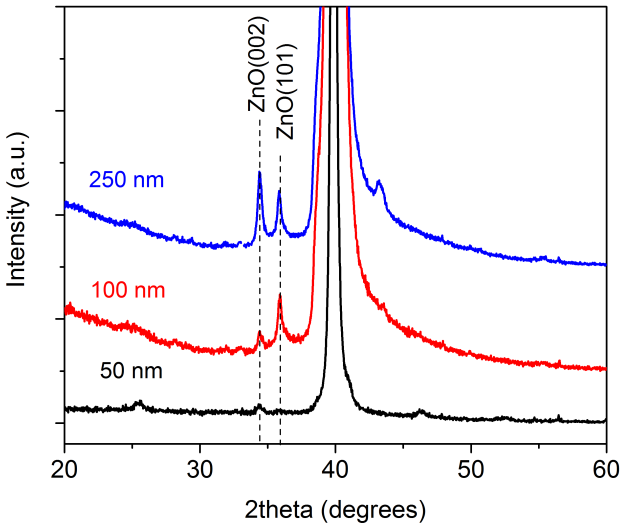


Fig. 4.6 XRD spectra of ZnO-based MIM memristive devices, at the different ZnO film thickness of 50 nm, 100 nm, and 250 nm.

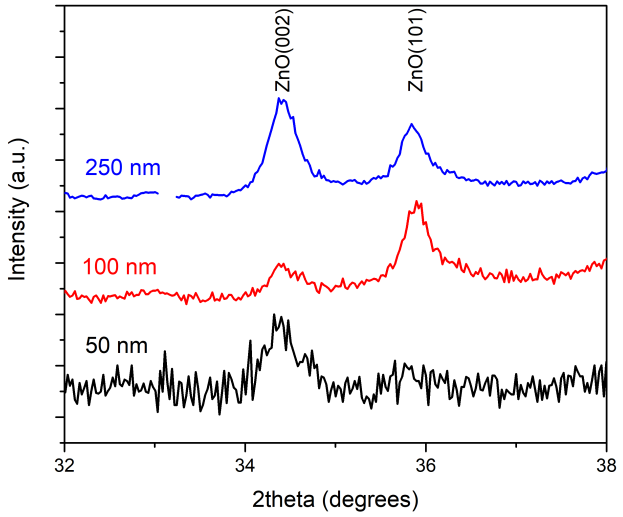


Fig. 4.7 XRD spectra detail of ZnO-based MIM memristive devices, at the different ZnO film thickness of 50 nm, 100 nm, and 250 nm.

4.1.2 Electrical Characterization

Devices are characterized in voltage sweep mode, applying consecutive positive and negative voltage ramps, not necessarily symmetric, while limiting the maximum amount of current transported across the device. The device is thus voltage controlled, up to the moment in which the current compliance is reached and becomes current controlled. This precaution is adopted in order to avoid irreversible breakdown of the device under test during transitions to lower resistance states.
In all the measures reported in the following, variability is computed as standard deviation of cycle-to-cycle measures.

The first device characterized here reported is ZnO – 50 nm VCM. Zinc oxide is particularly conductive and the measured resistance stays in the order of hundreds of ohms. For this reason, also the imposed current compliance is kept in the order of mA. Ten representative electric characterization cycles for the present device are reported in figure 4.8a), with one of them shown in red as reference. Electric characterization starts at 0-voltage with positive ramp up to 0.5V applied and consecutive negative ramp with absolute maximum negative voltage of -0.8V. Before the positive maximum voltage is reached, quick transition to LRS is observed. In the specific case represented, the current compliance is not required because of a low maximum voltage set. A quick and opposite transition is then observed in the negative voltage polarity, with a decrement to the original HRS. Already in the few cycles reported here, it appears evident how the characterization does not reach a stable electrical profile and observables like LRS, HRS, SET and RESET voltages are extremely variable. The resistance states during the voltage sweep themselves are unstable and largely noisy. Such characterization is consecutively repeated with more equivalent devices, spanning different range of maximum voltages as well as of current compliance values, without being able to stabilize the electrical response. On the other hand, irreversible electrical behavior is obtained for ZnO – 50 nm ECM. SET transition is obtained whenever positive voltage is applied to the copper electrode, but despite of the presence of current limitation, RESET transition is not obtained. Figure 4.8b) shows a representative characteristic of the ECM device once the breakdown is obtained. Many trials are attempted, reducing the current compliance as well as leaving the device forced only in the negative polarity up to 100 mA, but no reversible SET transition is observed. Copper is a highly electromigrating metal, that eventually under the application of electric field easily cover the whole 50 nm of zinc oxide

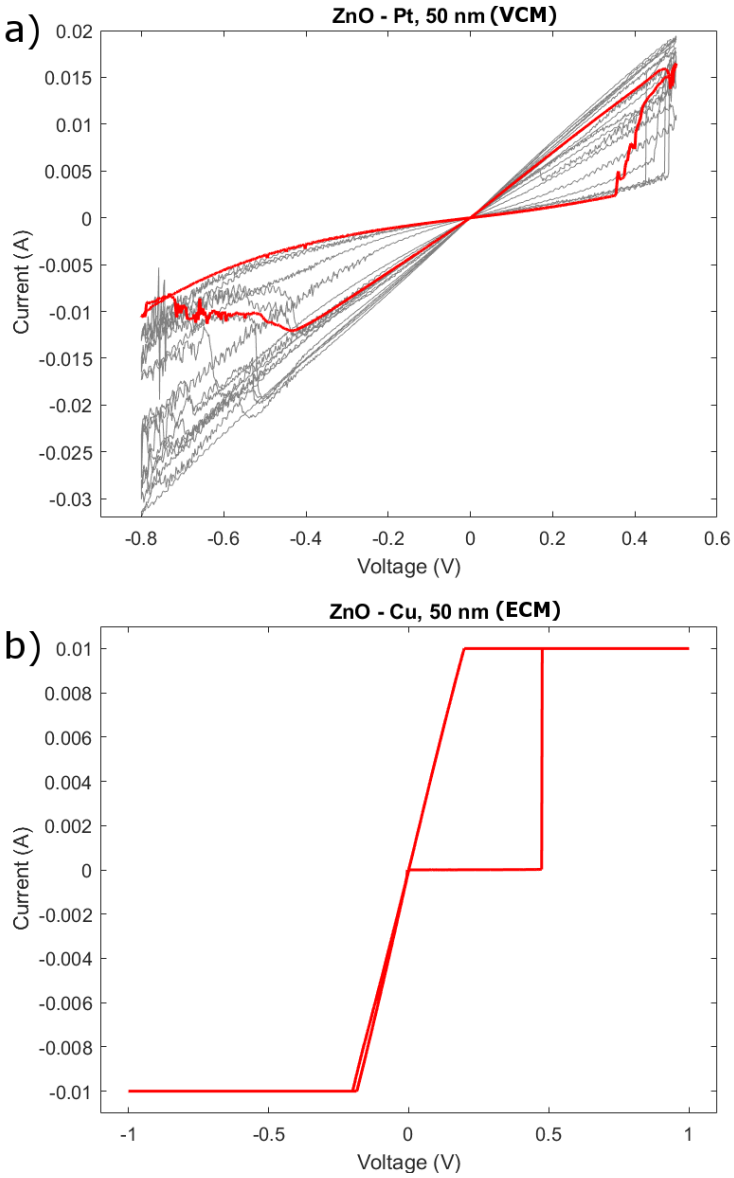


Fig. 4.8 Electrical characterization for ZnO - 50 nm samples. Panel a), voltage sweep cycles for ZnO/Pt sample showing bipolar resistive switching behavior; representative cycle is reported in red. SET transition is observed in positive voltage polarity, RESET transition is observed in negative voltage polarity. Panel b), voltage sweep cycle for ZnO/Cu sample. Dielectric breakdown is observed in positive polarity and irreversibly shortcut the device.

interposed between the two electrodes. On the other hand, VCM relies only on the presence of oxygen vacancies within the oxide matrix. The switching characteristics observed suggest the presence of a sufficient concentration of vacancies to start the resistive switching, but not for obtaining stable electrical behavior. Observing

the good stoichiometry of the material, one can suppose that the main source of vacancies within the oxide is then probably laying on the grain boundaries of ZnO and alongside the grain surfaces of the columnar structure observed, rather than in the polycrystalline and partially amorphous material.
In order to increase the amount of oxygen vacancies available for the resistance switching, a thicker device is thus realized, depositing 100 nm zinc oxide in both VCM and ECM configuration, with aim of stabilizing performances of VCM and further investigate those of ECM. In measurements performed with both devices, it is not observed a stable and sharp transition from pristine state to a higher conductance state, i.e. electroforming, but rather a smooth transition. Indeed in both measurements with ZnO - 100 nm, the pristine state is found to be in the same order of magnitude of high resistance state measured for the first few cycles.

Figure 4.9a) reports the observed behavior found for ZnO – 100 nm in VCM configuration. Among 100 cycles performed, two variable resistance states are observable, see inset in figure 4.9a). The device shows resistive switching behavior with SET transition dispersed around 2 V and RESET transition at about -2 V. As reported in table 4.1, standard deviations for these quantities are wide and the measured resistance in both LRS and HRS is about one order of magnitude higher than what previously measured. This is obviously a consequence of the thicker insulating material interposed between the two electrode, that is likely richer, in absolute terms, in oxygen vacancies available for the creation of a conductive filament, but still not enough to show a stable LRS. Indeed, observing the inset of figure 4.9a), it appears evident how in about 50% of measured cases, the positive voltage ramp is not enough to induce the SET transition. The LRS is not stabilized also by the relatively high current compliance, set to 35 mA. From this facts, it emerges that increasing further the thickness of a VCM device would likely bring to a scarcely relevant improvement in performances, while the increment of resistivity would lead toward a highly resistive device, that in this configuration would demand a particularly high applied voltage in order to induce sufficiently high electric field for the creation of the conductive filament.
On the contrary, ZnO – 100 nm ECM shows a much more reliable performance in comparison to the VCM counterpart. Figure 4.9b) reports the last 100 cycles out of 200 I-V curves measured. All 200 cycles corresponding resistance states are reported in the inset of the same figure. As can be observed, HRS undergoes a drift from higher resistance toward the nearly stable state observed in last 100 cycles.

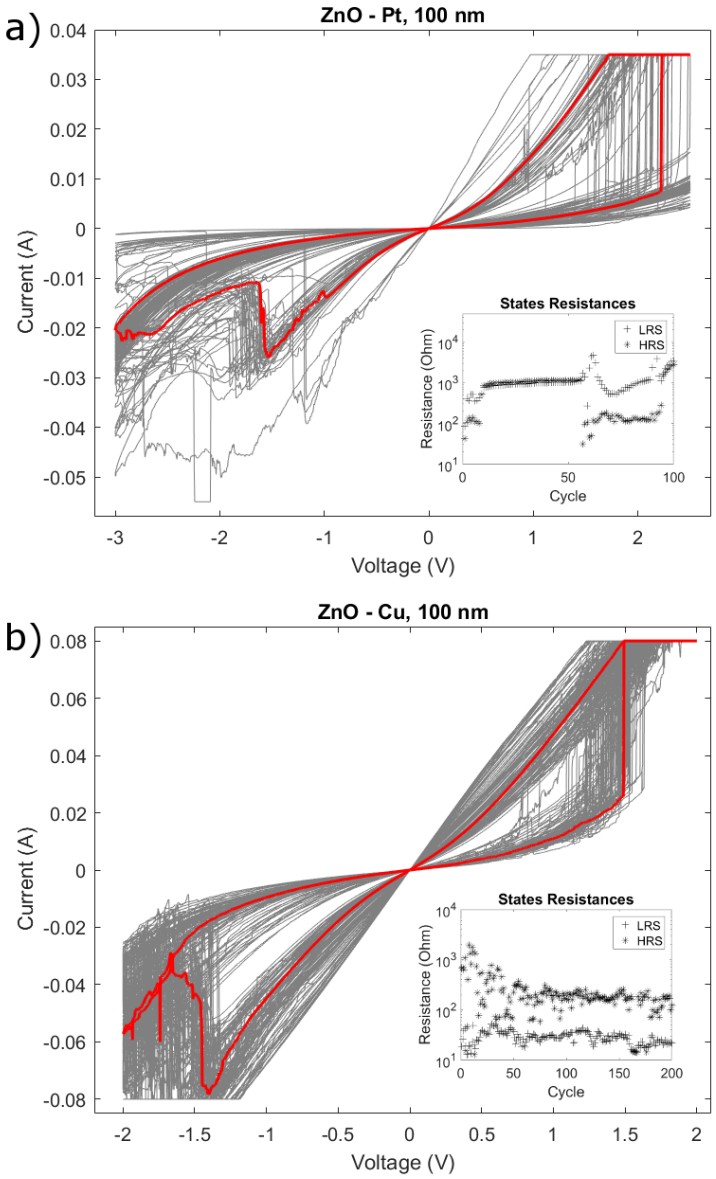


Fig. 4.9 Electrical characterization for ZnO - 100 nm samples. Panel a), voltage sweep cycles for ZnO/Pt sample showing bipolar resistive switching behavior; representative cycle is reported in red. High failure rate is observed. SET transition is observed in positive voltage polarity, RESET transition is observed in negative voltage polarity. Panel a) inset, resistance states measured at $V = 0.1V$ for 100 cycles. Panel b), voltage sweep cycles for ZnO/Cu sample showing bipolar resistive switching behavior; representative cycle is reported in red. SET transition is observed in positive voltage polarity, RESET transition is observed in negative voltage polarity. Panel b) inset, resistance states measured at $V = 0.1V$ for 100 cycles.

This is likely to be a consequence of ECM mechanism, in which the conductive filament composed of metal atoms is get enlarged by continuous operation of the device, that progressively reduces its resistance. The standard deviation of LRS, HRS, SET and RESET voltages is narrower than the previous case and indeed, the failure of transition happens only for a few cycles too. From this standpoint, performances of ECM appears to be largely greater than VCM and promising for applications. The peculiar conformation of ZnO polycrystalline matrix appears to play a stabilization role for copper ions moving orthogonally to the substrate surface without becoming a limit. In this perspective, device thickness does not seem to be a deterrent for operative devices in the ECM configuration and on the contrary, could favor better performances. In order to investigate further this insight and understand the feasibility of extremely thick resistive switching in ECM configuration, a ZnO – 250 nm device is fabricated, at this stage only with asymmetric Platinum-Copper electrodes.

Table 4.1 ZnO devices electrical performances summary: exhibited LRS, HRS and ratios.

Sample	R_{on} (Ω)	R_{off} (Ω)	Ratio
$ZnO-100-Pt$	124 ± 33	$(1.2 \pm 1.0)10^3$	14 ± 19
$ZnO-100-Cu$	27 ± 7	160 ± 57	6 ± 3
$ZnO-250-Cu$	24 ± 2	180 ± 17	7 ± 1

Table 4.2 ZnO devices electrical performances summary: voltage ranges for SET and RESET transitions.

Sample	SET	RESET
	Voltage Range (V)	Voltage Range (V)
$ZnO-100-Pt$	[1.5, 2.5]	[-1.5, -2]
$ZnO-100-Cu$	[1.0, 1.7]	[-1.3, -2]
$ZnO-250-Cu$	[1.0, 1.6]	[-0.5, -0.8]

The same electrical characterization is performed for ZnO - 250 nm. In this case, on the contrary to what observed with thinner devices, a sharp electroforming is observed, inducing transition from pristine resistance state, measured in the order of $7k\Omega$ at $V = 0.1V$, to resistance states one order of magnitude lower.
Figure 4.10a) shows 400 I-V curves measured with voltage ramps between +2V and -1.7V, with current compliance set to 20 mA only in the positive polarity. 400

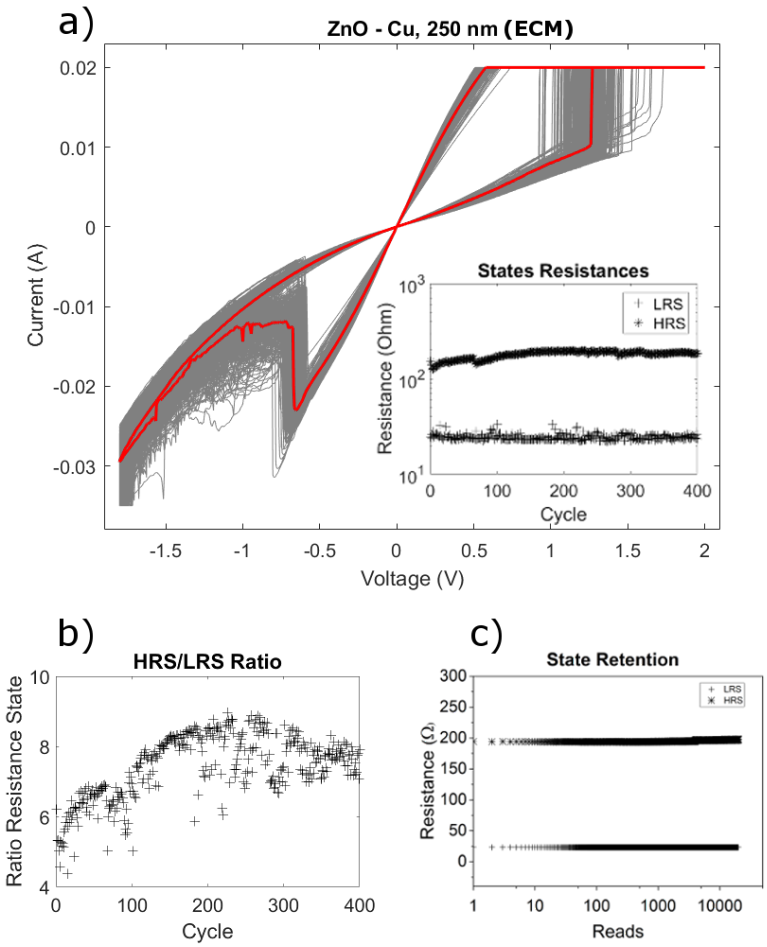


Fig. 4.10 Electrical characterization for ZnO - 250 nm samples. Panel a), voltage sweep cycles for ZnO/Cu sample showing stable and repeatable bipolar resistive switching behavior; representative cycle is reported in red. SET transition is observed in positive voltage polarity, RESET transition is observed in negative voltage polarity. Panel a) inset, resistance states measured at $V = 0.1V$ for 400 cycles. Panel b), resistance states ratio computed for 400 cycles. Panel c), retention measurements for LRS and HRS, measured with $V = 0.1V$, $T_{high} = 1ms$, $T = 10ms$

consecutive I-V cycles is the best endurance obtained for the produced devices and its stability is remarkable. Already from the plot reported, it appears evident the overlap between consecutive cycles. A more quantitative analysis then shows standard deviation for all the relevant observables narrower than the previous cases, as shown in table 4.1. The resistance states themselves are shown in inset of figure 4.10a), while panel b) of the same figure shows their ratio. It is observed a slightly increasing Off/On resistance ratio dispersed within the range [5,9]. Good stability of ZnO - 250 nm allowed for retention measures, performed in both LRS and HRS,

forcing with constant signal of 0.1V for with 10% duty cycle and 10 ms period. The performed retention measurement showed extremely stable resistance for $2.2 \cdot 10^{+4}$ reads both in HRS and LRS, see figure 4.10c), after which the measurement was interrupted.
ZnO - 250 nm shows far better performances in ECM configuration than all the previous devices characterized. Its peculiar features from the material standpoint relies mainly in the crystalline structure of the layer. Already from the FESEM characterization, the prevalent orientation of crystal grains orthogonally to the $SiO_2/Ta/Pt$ substrate is remarkable. The XRD analysis then confirms an emerged favoured crystallographic orientation along the c-axis, coexisting with direction (101). The particular crystallinity of the ZnO layer is direct consequence of the thicker layer and could stand at the basis of the observed electrical performance. Thick devices in VCM configuration already at 100 nm shows a high resistance and unstable electrical response, while when electromigrating metal atoms are involved, stability is reached. Moreover, in the latter case, values for LRS and HRS are comparable to the one measured with thinner MIM devices in ECM configuration and supports the hypothesis of the role of Cu atoms involved in the switching rather than Oxygen vacancies. Thus movement of ions and crystal orientation seems to be positively correlated and responsible for reliable performance of ZnO based resistive switching. On the other hand, the layer thickness is not impeding the formation of a conductive filament, that because of the thickness involved and the typical CF diameters reported in literature [35], seems to be unlikely continuous. Other works in literature shows that with relevant, still different, ZnO based devices the CF continuity is not required for observing resistive switching [86].

4.1.3 Conclusions

The present study investigated RS properties of ZnO thin film devices in both ECM and VCM configuration, at different thicknesses. The FESEM analysis of the samples highlighted a progressively more columnar crystallization of the ZnO oriented orthogonally to the $Si/Ta/Pt$ substrate. Samples at 50 nm thickness express barely visible columns, while wurtzite phase becomes predominant when the sample at 250 nm thickness is observed. Raman spectroscopy and XRD spectroscopy highlights also more and more intense peaks of nano-crystalline ZnO as the layer thickness increases, with predominance of peak (002), with respect to peak (101), as the

thickness reaches its maximum. Thus, grains are oriented alongside the c-axis, consistently with what observed with FESEM analysis. Finally, XPS spectroscopy shows a nearly stoichiometric film, due to the specific sputtering deposition conditions adopted. In particular, the choice of ZnO ceramic target, the inert Ar atmosphere and the accurate sputtering power calibration.
The electrical characterization of the devices shows an unstable and unreliable RS behavior for ZnO - 50 nm in VCM configuration. On the other hand, The same device in ECM configuration is easily shortcut within the first cycles of voltage sweeps. In order to increase the amount of oxygen vacancies available for the CF creation, the thickness of both devices is increased to 100 nm. Indeed, because of the particular structure of the ZnO film, oxygen vacancies available are mainly concentrated alongside the grain boundaries of the columnar grains present. ZnO - 100 nm shows an improved RS characteristic but, still, with very high failure rate, due to difficulties in performing the SET transition. On the contrary, the same device in ECM configuration performs far better, with only about 5% of failure rates and endurance shown up to 200 cycles. If thickness represents a problem for the creation of an oxygen vacancies based CF, Cu ions are able to electromigrate and stability create a more conductive area within the ZnO grains alongside the complete length of the film.
Finally, the largest thickness sample in ECM configuration exhibits the best performances of the series, with the narrower standard deviation on all measured quantities, 400 cycles of endurance reached and 10^4 reads of state retention.
In conclusion, ZnO thin films resistive switching devices shows reliable performances in voltage sweep mode preferably in ECM configuration and with a higher crystallinity, induced by a larger film thickness. Such a performance would be already exploitable for learning networks implementation. On the other hand, spiking neural networks, required a voltage pulse characterization, not yet available in the present study for these samples. Moreover, if the ZnO samples are particularly interesting for the study of mechanisms at the basis of resistive switching, because of their relatively large conductivity they are not the most favored for large network implementations, since they would induce a huge power dissipation. Other materials thus are adopted for the synapse-like implementation of RS in SNN, like Ta_2O_5 and HfO.

Chapter 5

Memristive Neuromorphic Systems

Reference paper: Brivio et al., ***Dynamics of spike-based learning in hybrid analog-digital CMOS-memristive neuromorphic systems****. Currently undergoing submission.*

Neuromorphic systems acquired more and more attention in the past decades because of their advantages with respect to the present computing technology, [40]. Indeed, they are particularly suitable for managing the noisy and unstructured amount of data constantly produced by the numerous device connected to the internet. An interpenetrated architecture merging computation unit and memory unit allows to overcome the limitation of Von Neumann's architecture, mostly in terms of communication bottleneck and chip size, boosting computational speed and reducing power consumption. Distinctive features of brain-like architecture are parallel and asynchronous computation, together with learning and adaptation to average activity mechanisms. Their potential have already been demonstrated by technologies like IBM's TrueNorth, [87], and SpiNNaker, [88], despite they are still based on conventional CMOS-based technology that hardly adapt to bio-inspired hardware.
Resistive switching, adopted together with refined chip architecture for emulating neurons functioning are particularly promising for the realization of bio-inspired devices. Analog electronics and event-driven computation indeed can be exploited to match the energy efficiency of brain, [89].
Most of the research activity in the field then is devoted to the understanding and modeling of resistive switching functioning in neuromorphic chips, dealing with the large device variability that such a new technology presents. As it already emerges

from the previous discussion of this dissertation, there is no standard yet for the memristor technology useful to be embedded in state of the art network. Numerous device architectures, like thin films or nanostructures, are explored in order to enable their successful implementation and control the switching mechanisms required for implementing learning algorithm in hardware. Variability is one of the main issues related to resistive switching, since both the final value of resistance state and of the control variable (electric field, internal temperature or others) needed for the switch could show a large variability. Many works in literature are devoted to demonstrate the robustness of neural networks against such variability, [90, 91] or to demonstrate its usability for the implementation of stochastic learning algorithms, [92].
Resistive switchings are also particularly suitable for neuromorphic systems because of their interoperability with spiking networks. Voltage spikes are of prominent importance in information transmission in brain, since they are generated by neurons and transmitted across the network through synapses. Artificial spiking neural networks are designed to behave similarly, implementing CMOS-based neurons able to integrate the input signals coming from the network and generate a voltage spike, i.e. a pulse, as similar as possible to the biological one. Such a pulse is then transmitted as the neuron output to the resistive switching and possibly imposes a transition in the resistance state. For this reason, it becomes important to have a complete characterization of the resistive switching in case of forcing pulse train, other than with voltage sweeps. As a second aspect, spiking computation can be described as a computation on demand, in which power is used only when an event, i.e. a spike generated by the input, is transmitted across the circuit. This fact contributes sensibly to the reduction of power consumption by the entire chip and favor its adoption for compact independent devices that cannot rely on constant accessible source of energy.
In the present chapter it is described the design of a spiking neural network embedding I&F CMOS-based neurons, as described in the Introduction of the present dissertation, and pulses-characterized resistive switching. Devices adopted to this purposes are realized by the Italian Council of Research, CNR - Agrate, with which a scientific collaboration as been established. The third side of the scientific collaboration comes from the Institute of NeuroInformatics of Zuerich, INI - ETH and UZH, that brings into the project competences for the circuit design and its implementation. Description of this second phase of the research project proceeds as follow: at first, the resistive switching devices will be described in their fabrication and character-

ization process. Empirical modeling of their functioning will be discussed, since it will be then employed in the spiking neural network learning algorithm. Then, Spike Timing Dependent Plasticity, together with its variation proposed by Fusi et al. [93], will be over-viewed. A brief description of the hardware designed is given. In conclusion, the behavior of the network is extensively analyzed with numerical simulations while performing the task of hand-written digits classification from the MNIST dataset, [94].

5.1 STDP and Fusi Learning

Spike Timing Dependent Plasticity is chosen as the learning rule implemented in the memristive network, because of its similarity to the biological processes believed to happen in brain functioning and because of its suitability to hardware [6–9, 89].
Human brain is thought to realize complex cognitive tasks and store memory thanks to the synapses composing its neural network. Indeed synaptic plasticity allows for a dynamic change of network configuration due to experience, realized in the form of variations in synaptic efficacy (also named synaptic weight, w) in transmitting information. Subset of synapses, during brain activity, becomes more responsive to specific stimuli rather then other; on the contrary, different subsets of synapses could become inhibited for the same reason. At its extreme, this behavior leads to the formation and pruning of synapses connected to dendrited of cells, physically changing the neuronal network itself, [5]. Synapses configuration becomes distinctive for the reaction to particular events experienced by the subject and are believed to be at the basis of our memory. The two types of dynamics leading to synapse efficacy strengthening and weakening are Long Term Potentiation (LTP) and Long Term Depression (LTD) respectively [9].
STDP models LTP and LTD as phenomena related to the firing activity of pre-synaptic neuron and post-synaptic neuron, considering their time-difference. Synaptic weight variation is defined as a function $\xi = f(\Delta T)$, with $\Delta T = t_{post} - t_{pre}$, being t_{post} and t_{pre} the time instant of post-synaptic spike and pre-synaptic spike respectively. STDP introduces a concept of causality in synaptic potentiation and depression, favoring the potentiation of connections between those neurons whose activity is positively correlated in time and reducing the relevance of those anti-

correlated. Moreover, as seen in figure 5.1, the relative weight change is proportional to the time delay happened between spikes.

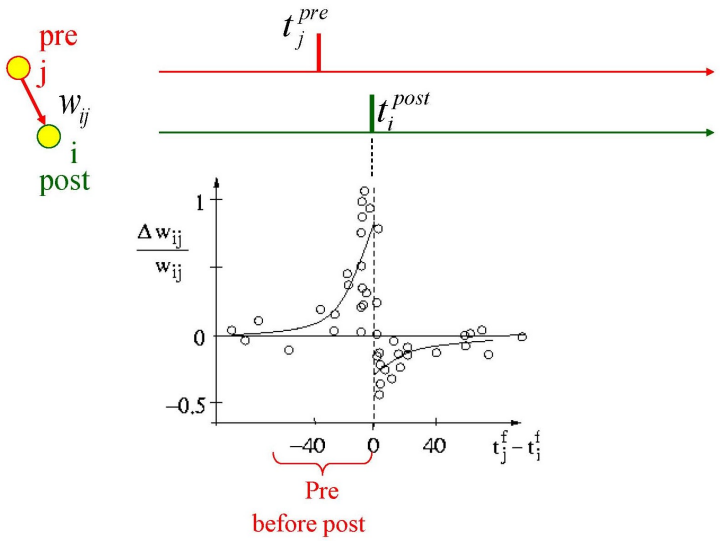


Fig. 5.1 Schematics of STDP functional behavior. Pre-synaptic neuron activity and post-synaptic neuron activity can induce synaptic weight update, following behavior depicted in the bottom panel, depending on their time correlation. Bottom panel, dot represents experimental data, solid lines represent mathematical inferred model. *Reproduced from Scholarpedia, 5(2):1362*

The scheme proposed by Fusi and coworkers, [93], corresponds to a generalization of the STDP learning rule. Conditions for the synaptic weight change to happen do not depend on the time of pre- and post-synaptic spike, but rather on the membrane current present at the post-synaptic neuron when stimulated by an input spike. In such a way, variation of synaptic weight is potentially always enabled, still depending on the activity of the post-synaptic neuron through the membrane current. In order to preserve a measure of time correlation, it is introduced a further variable inspired by biological feedback loop present in the neuron, i.e. Calcium ions current. It acts as a low pass filter of post-synaptic neuron activity and provides a measure of average firing rate on long time-scales. Setting opportunely thresholds to both membrane current, i.e. I_{mem}, and Calcium current, i.e. I_{Ca}, neuron activity could enable alternatively LTD or LTP depending on values assumed by both variables.
Looking more into the details of the circuit activity, a synapse connects couple of neurons. Neurons integrate current incoming at their input according to the following equation:

$$\begin{aligned} &\frac{dI_{mem}}{dt} = -\lambda + I(t), \\ &I_{mem} \geqslant 0, \end{aligned} \tag{5.1}$$

where $I(t)$ is the overall incoming current and λ is a constant leak. Whenever the membrane current overcomes a pre-determined threshold I_{spk_θ}, the post-neuron fires in its output and the membrane current is reset to a fixed value. This behavior emulates the refractory period experienced by biological neuron in which hyper-polarization occurs and silent the neuron activity until equilibrium is restored. Dynamics of the second internal variable of the neuron is governed as follow:

$$\frac{dI_{Ca}}{dt} = -\frac{1}{\tau_C} + J_C \sum_i \delta(t - t_i), \tag{5.2}$$

where contributions due to spiking $J_C \cdot \delta(t - t_i)$ with amplitude J_C occurring at time t_i are summed up and integrated with a time constant τ_C. The state variables I_{mem} and I_{Ca} define the neuron operative modes among LTD, LTP and neutral, according to the learning rule [93]:

$$\begin{aligned} LTD: &\quad I_{mem} < I_\theta \quad \wedge \quad \Theta_{D_{low}} < I_{Ca} < \Theta_{D_{up}} \\ LTP: &\quad I_{mem} > I_\theta \quad \wedge \quad \Theta_{P_{low}} < I_{Ca} < \Theta_{P_{up}} \quad , \\ neutral: &\quad \text{otherwise} \end{aligned} \tag{5.3}$$

where I_θ specifies thresholds on membrane current, $\Theta_{D_{up,low}}$ and $\Theta_{P_{up,low}}$ specify bands thresholds for LTD and LTP. Depending on the chosen thresholds, different regimes of learning can be observed, as it will further explained in the following sections.

5.2 HfO Resistive Switching

5.2.1 Characterization

Pt/HfO_2/TiN resistive switching are adopted as the weighting elements for synapses in the neuromorphic architecture. The inverse of the device resistance, i.e. the conductance, can be easily exploited as the measure of the synaptic weight. Devices

have been already demonstrated to undergo resistance switching operation through accumulation and dissolution of oxygen vacancies according to the picture of conductive filament formation and dissolution with good performances in terms of state retention, endurance and device scalability down to 28 nm [26, 95–97]. In this work, devices have been characterized and simulated through a simple compact model that both allows the simulation of a large amount of synaptic connection and at the same time captures the salient features in term of resistance levels and boundness of the synaptic weight that impact the final neuromorphic network.
Moreover, devices have been characterized with multiple set of temporal and electrical parameter, among which two are chosen for device simulations in neural networks, in order to observe the different electrical response in terms of conductance change. Indeed, what is usually reported as the voltage-time dilemma puts in evidence the still un-modeled variation in conductance dynamics observed when devices are characterized with electrical pulses of different amplitude and period. Intuitively, longer pulses of higher voltage, apply a stronger electric field within the oxide that induce a faster transition toward a lower resistance state. Conversely, short pulses of lower voltage require more numerous pulse trains to induce the same transition. Which the boundaries are for pulses parameter and how to model intermediate conditions is still object of research. In this study, two conditions are examined and are referred to as Pulse Parameter Set 1 and 2, i.e. PPS1 and PPS2. Please, refer to table 5.1 for details. Proceeding with the simulations, it has been found that PPS2 are more suitable for neural network with higher classification performances, as it will be clarified in the following of the present chapter. For such a reason, it is reported here characterization and modeling of devices with PPS2, while analysis performed with PPS1 is left for a more complete overview in appendix A

Pulse Parameter	$V_{LTP}(V)$	$V_{LTD}(V)$	$T(\mu s)$
PPS1	1.1	-0.8	30
PPS2	0.5	-0.45	30

Table 5.1 Pulse Parameter sets adopted for electrical characterization of RS devices.

The switching characteristics of HfO_2-based devices, characterized with PPS2, are summarized in figure 5.2. When stimulated by sequences of identical positive spikes, devices undergo transition to LRS and LTP, see panel a), (increase of conductance). On the contrary, when stimulated by sequences of identical negative spikes,

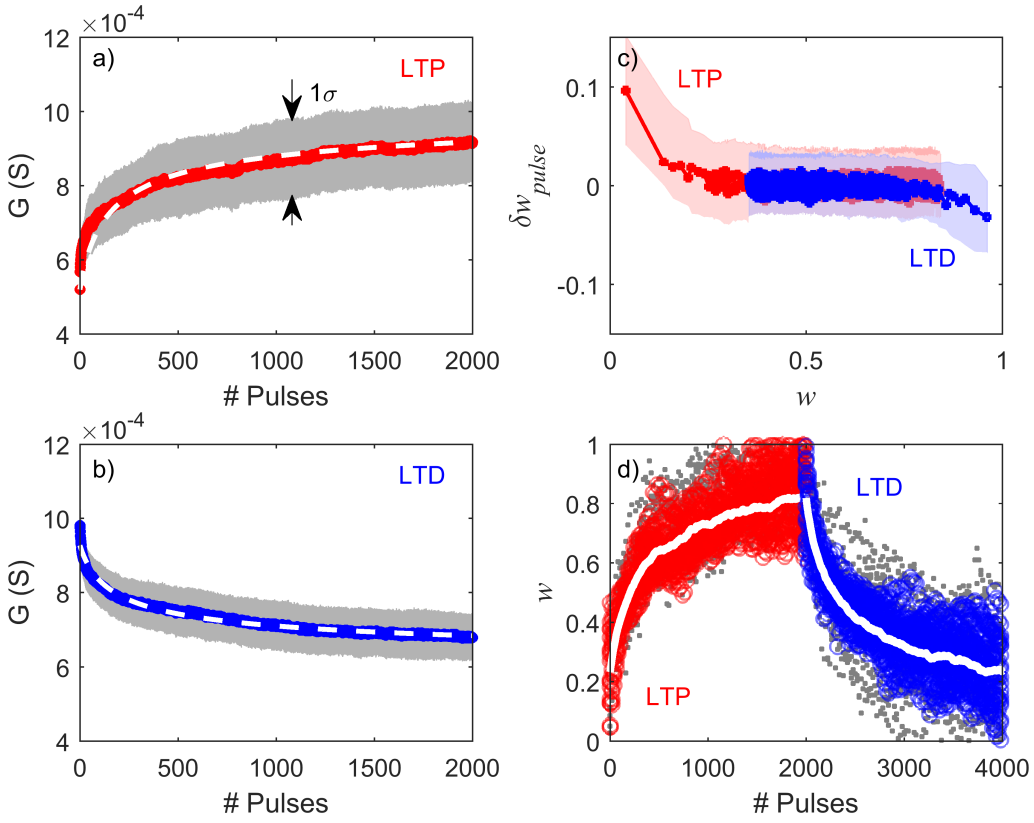


Fig. 5.2 RRAM representative conductance evolution as a function of the number of pulses from 1 to 2k pulses for LTP (a) and LTD (b): symbols correspond to the average of 50 experimental curves; dashed lines are the fitting curves. Grey shaded regions correspond to 1σ variability. (c) Conductance change driven by one pulse as a function of the initial resistance for both LTP and LTD, in red and blue respectively. (d) Comparison of tens of weight evolution as obtained from device and simulations including pulse-to-pulse variability for LTP (from pulse 1 to 2000) and LTD (from pulse 2000 to 4000). Grey small filled symbols are experimental data and large empty symbols corresponds to the simulated data. Thick white lines are the average of the simulated data.

devices undergo transition to HRS and LTD, see panel b), (decrease of conductance). Conductance is easily converted into synaptic weight through direct proportionality and mapped in the range $[0,1]$. In this way it becomes evident how the change in conductance, i.e. synaptic weight δw, depends on the current value of the weight itself w, see figure 5.2c). Observing the dynamics of LTD (blue dots) and LTP (red dots) process, asymmetries between the two processes emerge. Indeed, weight change is much faster for LTP and LTD, when the weight is close to the extreme value, and then saturates as the dynamics proceeds. The complete dynamic is shown

in 5.2d). The evolution of the conductance in LTD and LTP reported in figure 5.2a-d) evidences the major non-idealities of memristors for synaptic applications as already emphasized in several publications [98, 99]. The resistance change is non-linear as a function of number of pulses and it is faster for low pulse numbers (i.e. at low resistance for LTD and high resistances for LTP, respectively). The second aspect is that the LTP is observed to have a faster dynamics with respect to LTD, i.e. it presents an increased non-linearity than LTD.

5.2.2 Modeling

The gradual conductance change featured by the device in response to sequences of identical spikes is found to be thoroughly described by two separate equations easily expressed in terms of synaptic weight for both LTD and LTP processes, starting form LRS and HRS respectively. First of all, conductance is converted into synaptic weight according to equation 5.4.

$$w_{ex_i} = f * (\frac{G}{G_{max}} - \frac{G_{min}}{G_{max}}),\ w_{ex} \in [0,1] \tag{5.4}$$

with $f = 2$. In terms of w, LTD and LTP processes can be written as in equation 5.5.

$$\begin{aligned} LTD &: \delta w = -\alpha_D w^{\gamma_D} \\ LTP &: \delta w = \alpha_P (1-w)^{\gamma_p}\ , \\ & w_{t+1} = w_t + \delta w \end{aligned} \tag{5.5}$$

where parameters $\alpha_{D,P}$ and $\gamma_{D,P}$ are evaluated by fitting experimental data with equations 5.5. Fitting of experimental data are reported in figure 5.2a) and b) for both LTD and LTP with dashed white lines. Fitting curves follow nicely the same dynamics of experimental data and overlap data average over 50 experimental curves, figure 5.2a) and b). Pulse-to-pulse variability is captured as well, producing simulated conductance distributions that overlap the experimental data ones, figure 5.2d).
Variability is evaluated as the standard deviation value of the conductance change per unit pulse $\Delta G(n) = G(n) - G(n-1)$ taken over tens of cycles for both LTD and LTP. The experimental standard deviations $\sigma_{\Delta G}$ are almost constant for both LTD and LTP and they are used to faithfully reproduce device behavior. During

numerical simulations, when specified, variability is added to the new computed synaptic weight value as Gaussian noise with $\sigma_{\Delta G}$ standard deviation.

$$\begin{aligned} LTD &: \sigma_{\Delta G} = \delta_D \\ LTP &: \sigma_{\Delta G} = \delta_P \end{aligned} . \tag{5.6}$$

Parameter	α_D	γ_D	α_P	γ_P	δ_D	δ_P
value	0.0053	3.4	0.0064	3.2	0.005	0.005

Table 5.2 Values of the fitting parameters for the general synaptic weight evolution and for the modeling of the device variability for both LTD and LTP.

5.2.3 Memristor implications to Neural Networks

The saturation of synaptic weight is the most important issue that distinguish software neural networks from hardware realizations. Indeed, from very general considerations, it has been demonstrated that the existence of boundaries for synaptic weights definitely limits the storage capacity and the precision of neural networks [93, 100] As discussed by Fusi et al.[93], this issue is only mitigated by the increase of the number of accessible synaptic weights values. On the other side, boundaries of the resistance value spectrum are soft, according to the terminology of Fusi et al., in the sense that they are defined as a slowing down of the LTD and LTP processes at high and low resistances as shown in figure 5.2c). The figure, indeed, shows that the normalized conductance change, i.e. the synaptic weight change, depends on the actual conductance value and, in particular, that the conductance change is negligible at low conductances for LTD and at high conductances for LTP, respectively.
The existence of intermediate resistance levels between a minimum and a maximum and the softness of the boundaries allows to improve the memory capacity of the network [93].
The boundness of the accessible weight range of hardware synapses requires a slow learning operation for consolidating past experiences and memory lifetimes according to the slow learning - slow forgetting principle. To some extent, the gradual resistance evolution of memristive devices as shown in figure 5.2 is useful to implement a robust progressive learning and helps in keeping track of a certain amount of past events experienced by the synapse itself. On the other hand, the employment

of a generalized implementation of STDP learning rule allows slowing the learning process down.

5.3 Neuromorphic hardware architecture

In this section, it is proposed a neuromorphic system that implements the learning protocol described before. While the exact architecture proposed in this section has not been implemented in Silicon, the building blocks have been implemented and tested in Silicon [42, 101, 102].

Synapse architecture

Figure 5.3 illustrates the design of a memristor-based synapse circuit, that consists of a memristor, switching transistors, and a digital control circuit that controls the connectivity of the memristor to various nodes. The digital controller consists of two pulse extender circuits for generating extended pulses that are used for reading and programming the memristor. A synapse is addressed by row and column select signals labeled as *AERx* and *AERy* in figure 5.3. The pulse extender circuits are essential because the *AERx* and *AERy* signals are generated by the AER interfacing circuitry. The AER is an asynchronous communication protocol that uses very-short duration pulses, of the order of *ns*, to maximize through-put. These pulses are extended by tunable pulse extenders to interface with neuromorphic circuits operating in biologically plausible time-scales. Each synapse is provided with an independent digital control module which allows multiples synapses to be read from or written to simultaneously.
In the synapse shown in figure 5.3, the arrival of AER events triggers a read event by setting the *Read* and $\overline{Read}$ signals to logical high and low levels respectively. This turns on the transistors, S1 and S2, connecting the terminals of the memristor, D_{top} and D_{bot}, to V_{rd} and *bot*, respectively. The node marked *bot* is connected to a sensing circuit that feeds the sensed state to a neuron. The read phase ends with the falling edge of the *Read* pulse, which initiates the programming phase by triggering the second pulse extender that sets $Write$ and $\overline{Write}$ to logical high and low, respectively. The synaptic digital control block accepts UP and DN as inputs from learning controller of the neuron connected to the synapse. Depending on

the UP and DN signals, this either increases or decreases the conductance of the memristor. Note that the learning controller ensures that both UP and DN are not simultaneously high. The *Set* (*Reset*) signal enables the switching transistors $S3$ and $S4$ ($S5$ and $S6$) for the duration of the $Write$ pulse to increase (decrease) conductance of the memristor.

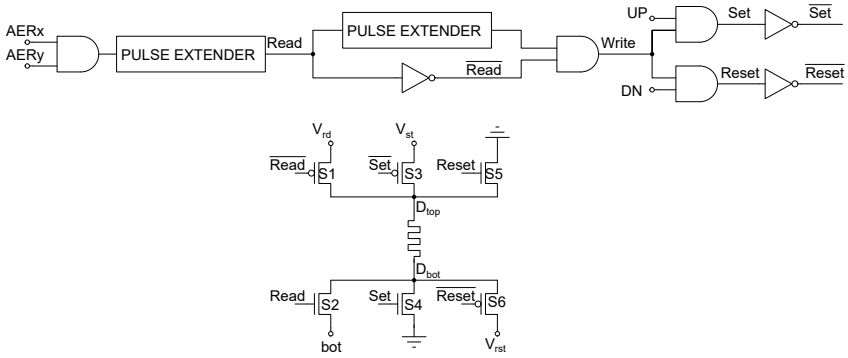


Fig. 5.3 A memristor synapse. *AERx* and *AERy* are the select signals used to access the synapse. The node marked *bot* will be connected to a sensing circuit that sets the node voltage to $Vrdb$. UP and DN are inputs to the synapse that determine if the conductance of the memristor should be increased or decreased. V_{st} and V_{rst} are the voltages applied across the memristor when increasing and decreasing its conductance, respectively.

System level architecture

Figure 5.4 illustrates how the synapse described in section 5.3 can be assembled into a neuromorphic system. The read and write voltage lines Vrd, Vst, and $Vrst$ are shared by all the memristors in the array. The *bot* node shown in figure 5.3 of all the synapses in a row are connected together. All the synapses in a row also share the resistance sense circuitry. The sense circuitry is implemented with an Op-Amp circuit. The sensing mechanism is described in the following paragraph using the top-most row of figure 5.4 as reference.

The Op-Amp, $Op1$, is connected in a negative feedback loop with a MOSFET that ensures that the *bot* node of the first synaptic row is set to voltage $Vrdb$. This ensures that the voltage across the memristors during read operation is $Vrd - Vrdb$. Therefore, the current into the current mirror pair $M1 - M2$ is a linear function of the conductance of the memristor. A sensing circuit using an op-amp in a similar feedback mechanism was used in an earlier work [102]. The output current from the current mirror is driven into a DPI circuit which implements a linear first-order

low-pass filter with tunable gain and bandwidth [42]. The output from the DPI is sent to the neuron $N1$. The neuron circuit is similar to the design from an earlier work [101]. Recall from Section 5.1, that each synaptic input to the neuron is scaled and convolved with a decaying exponential function. The arrangement described in this section implements this operation because the gain and bandwidth of the DPI circuit can be programmed to the desired value. The only constraint is that the circuit should be operated in sub-threshold. This can be ensured by making $Vrd - Vrdb$ small to make the currents small. Further note that the filtering operation can be shared by all the synapses because of the principle of linear superposition.

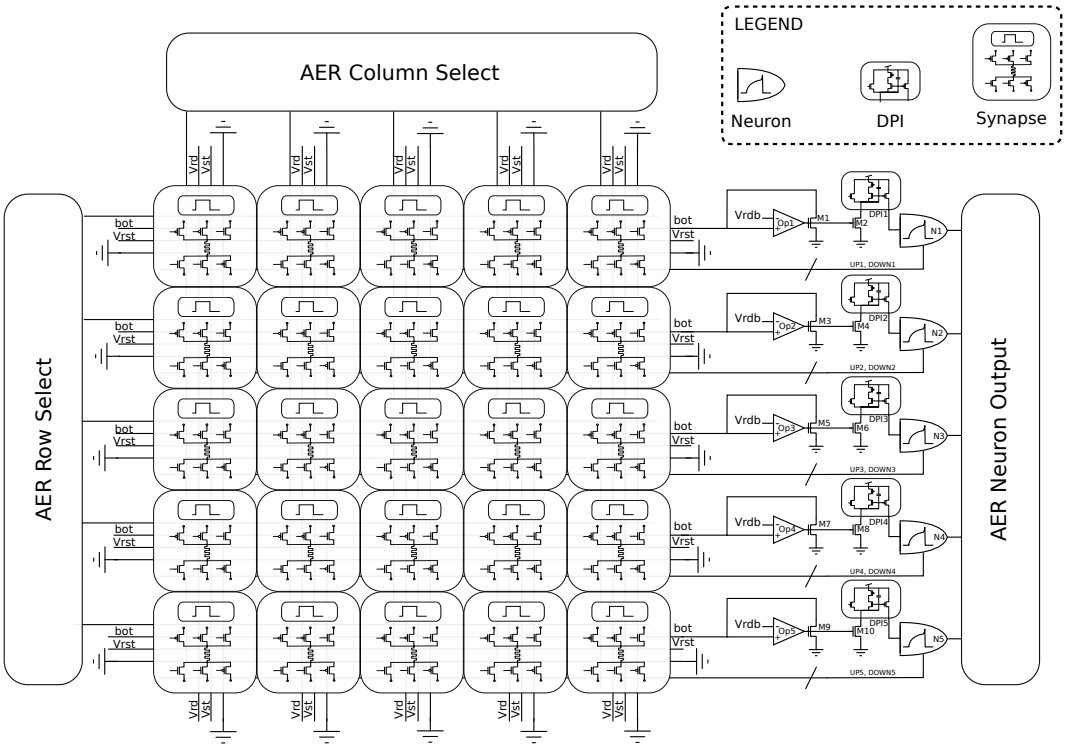


Fig. 5.4 A neuromorphic system implementing the proposed learning architecture. The AER Row and Column select modules generate the *AERx* and *AERy* signals, shown in Figure 5.3, to access each synapse. The AER Neuron output module communicates the spikes generated by the neurons to the configured target.

Each neuron is associated with its own learning block that implements the learning rule described in Section 5.1. This circuit has been implemented in an earlier work [101]. The learning block generates a 2-bit signal UP and DN that is used by the synapses connected to the neuron to perform the programming operation. The mechanism for programming the synaptic weights has already been described in

Section 5.3. This approach avoids use of pulses with complicated shapes [103–105] or long overlapping pulses [90, 104, 106, 107], making the system more power and area efficient.

5.4 Network

5.4.1 Neurons Populations

In the present study, the spiking neural network demanded to classify digits in MNIST database, is composed by four different neurons populations as described in the following and depicted in figure 5.5. The *input population*, (IP), receives the external stimulus, i.e. the MNIST digit, as pixel intensities and maps it to a Poisson process with spike train frequency directly proportional to the pixel intensity itself. The Poisson neurons population counts 784 nodes, corresponding to the total number of pixels presents in the 28x28 image size, and inputs to the output population. The *inhibitory population*, (InhP), accounts a limited number of Silicon neurons and receives input signal from the IP. In this network it is chosen an inhibitory population counting 392 neurons, thus one half of the excitatory population. Inhibitory neurons act as well as leaky integrate and fire neurons, but contribute to the wired neurons in the network with a negative input current. This is meant to modulate the total amount of current coming from the excitatory population and thus balance the firing rate of output neurons. The *teacher population*, (TP), is composed of the same number of neurons counted in the output population. Teacher neurons provide the signal for implementing a semi-supervised learning, [93], and bringing the membrane current of the respective output neuron above threshold only when the targeted digit is presented. Thus teacher neurons are Poisson neurons with fixed and high spike train frequencies. The *output population*, (OP), when referring to the simplest network architecture, defined as *perceptron* and composed by one input layer addressing one single output node, is composed of a single Silicon neuron and receives input signal from the IP, InhP and TP. Its activity is determined by the input current driving its output firing rate. In case of the complete network designed for the recognition of all digits, OP accounts for 10 times an integer number of neurons, i.e. 100 in case of 10 neurons addressed by each digit. It introduces a voting rule for the digits classification that improves considerably the recognition performances. Indeed,

being the process intrinsically stochastic, included also the spiking frequency of output neurons, the average firing rate of the neuron population computed on a larger number of nodes provides a more confident evaluation of performances.

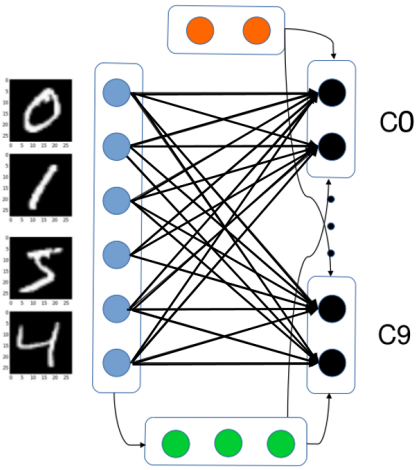


Fig. 5.5 Architecture of the complete network adopted in the learning task. Four different neurons population are represented in different colors. IP is represented in blue and inputed with MNIST digits as represented in the picture. InhP is represented in green, TP in orange, OP in black. IP and OP, IP and InhP as well as InhP and OP are fully connected, but for sake of clearness, only the first set of edges is shown in the picture. Only synaptic connections represented with thicker lines are plastic. TP and OP are connected one to one when a digit belonging to the specific class is presented.

5.4.2 Synaptic Connections

IP and OP are fully connected through plastic synapses adapting their weight according to the output neuron membrane current and calcium current values, only when an input spike is coming from IP. Resistive switchings conductances are directly proportional to synaptic weights, being multiplied by an integer factor f and rescaled within 0 and 1. IP and InhP are fully connected with non-plastic synapses, as well as InhP to OP. Pairs of neurons in TP and OP are connected. No further synaptic connections are present in the network.
Each synapse in the network, including the memristive synapses, is governed by an equation as described in 5.7:

$$\frac{dI_{syn}}{dt} = \frac{-I_{syn}}{\tau} \tag{5.7}$$

where τ is specific for any synaptic population. Moreover, the contribution to current coming through each synapses can be written as $I_{syn_i} = I_w * w_i$, with w_i synaptic weight and I_w is a current representing the synaptic efficacy. In these terms, the total current inputed to the OP through synapses can be written as 5.8:

$$I_{in} = \sum_{i}^{S_{ex}} I_{w_{ex}} w_{ex_i} + \sum_{i}^{S_{inh}} I_{w_{inh}} w_{inh_i} + \sum_{i}^{S_t} I_{w_t} w_{t_i} \tag{5.8}$$

where again I_w represents the synaptic efficacy for IP, InhP and TP, while w represents the synaptic weight for IP, InhP and TP and S represents the total number of synapses wired to the output neuron from IP, InhP and TP. The weights evolving in time are w_{ex}, because of synaptic plasticity, while all the other quantities in equation 5.8 are constants and fixed at the beginning of the simulation. In particular they are initialized as follows: $w_{ex} = U(0,1)$, thus representing resistances dispersed in the whole accessible states, $w_{IP-inh} = U(0,0.8)$, $w_{inh-OP} = G(\mu = 1.1, \sigma = 0.05)$ and $w_t = G(\mu = 1.3, \sigma = 0.05)$ respectively, where $U(\bullet)$ is uniformly distributed and $G(\bullet)$ is Gaussian distributed. All other custom parameters used into the simulation are specified in tables 5.3 and 5.4. Whenever an input spike reaches the synapses, the present membrane current and calcium current of the output neurons inputed by the synapse are checked against the respective thresholds. If conditions for LTP or LTD are satisfied, then a new resistance value is computed according to equation 5.5 and the synaptic weight is updated.

5.4.3 Perceptron

Preliminary tests with a perceptron network demanded to acquire features discriminative for 0s and 1s are performed. Only in this specific section, the presented analysis refer to devices characterized and simulated with pulse parameter set PPS1; such a choice is related to differences between the electrical characteristics observed in the two different cases, confront with appendix A. Indeed PPS1 induces more abrupt transitions in the devices conductance and thus, put in evidence their dynamics also in short representative simulation, better than what it would be observed with simulation involving PPS2.
Perceptron is a single layer neural network acting as linear discriminator. It accounts for a single input layer of nodes, connected through plastic synapses to a single

output neuron. The network is instructed to respond positively (output neuron active) to the input of a specific digit. Perceptron network, still composed of all the populations enumerated before, is the building block that allows to monitor correct functioning of all elements in the network. To this purpose, it is exploited to first check the correct behavior of simulated resistive switchings inside the network when LTD and LTP happens.
Observing figure 5.6, the perceptron is run for 900*ms* while being presented a single digit and requested to acquire features of digit 1. In both preliminary simulations, the resistive switchings are simulated excluding device variability. In the first case, a single digit 0 is presented. Figure 5.6a) shows the total weight map, i.e. the value of each synaptic weight connecting one input pixel of the digit to the output neuron, in false colors. Weights generally preserves the initial Gaussian distribution around mean 0.2, but it starts being evident a depression area of zero shape. This is the consequence of LTD for those synapses receiving high frequency of bright pixels defining the digit zero. Observing panel c), dynamics of two specific synaptic weight, i.e. one from the center of the weight map (in red) and one from the bottom left corner (in blue), is presented. Red line is constant for the whole duration of the simulation, since being in the background of the image, no input spike is presented. Blue line on the contrary shows a slow decay to lower synaptic weights, equivalent to higher resistances. The relative variation of the weight is low since it is already close to the asymptotic value of the resistance, reaching the bound of the allowed resistance distribution without variability.
Figure 5.6b) and d) show the same information for the case of digit 1 presented to the network. The specific input activates the teacher neuron that starts contributing to current income of the OP. Thus, output firing activity has higher frequency and enable LTP. Features of digit 1 appears on the weight map has a subset of weights with very high value, close to the maximum. The single weights dynamics show a consistent increase of synaptic weight for both values monitored, respecting the empirical functional behavior modeled from the real device characterization.

Figure 5.7 reports the same analysis described above in the case of device variability accounted in the simulation. Few considerations worth mentioning. First, depressed features of digit 0 in panel a) manifest a lower homogeneity with respect to the previous case. Looking at panel c), the weight dynamic represented in blue is not monotonic anymore, despite the same simulation condition. Thus, the dominant contribution of LTD remains, but variability allows for mixed increment

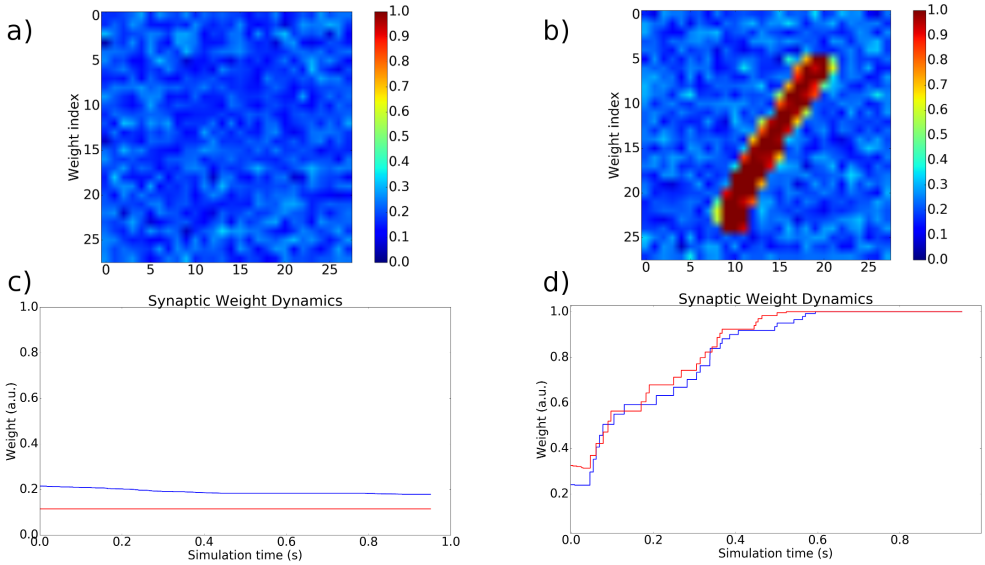


Fig. 5.6 Panel a), synaptic weights after a single 0 digit presented for 900 ms. Panel b), synaptic weights after a single 1 digit presented together with teacher signal. Panels c) and d), dynamics of weights in the former two cases respectively. Blue traces correspond to weight excited by both digits, while red traces correspond to weight excited only by digit 1. Traces make evident LTD process, when the only input is shown, and LTP process when teacher signal is applied. The narrow relative variation in weight for LTD is due to an already high value of initialized resistance, that brings dynamics toward saturation.

and decrement of synaptic weight. The red curve remains still constant since its variation depends on the pre-synaptic spike input. Looking at results for digit 1 as input, features are still acquired and evidently recognizable, but differently from the previous case, dynamics of two distinct synaptic weights are much less overlapping. In conclusion resistive switching behavior during network learning is perfectly comparable with the experimental characterization of real device; variability as well affects the dynamics of synaptic weight during both LTD and LTP, without impairing the features acquisition at the basis of digit learning.

Longer simulations with more than one input digit are then performed in order to gain insights about the dynamics of Silicon neurons variables, i.e. I_{mem} and I_{Ca}. Indeed, the two are of prominent importance in specifying the learning dynamics and the way in which current thresholds are specified affect deeply the network dynamics. In particular, at first the network dynamics is analyzed for two different Ca band configurations:

1. *thresholded* LTD/LTP modes;

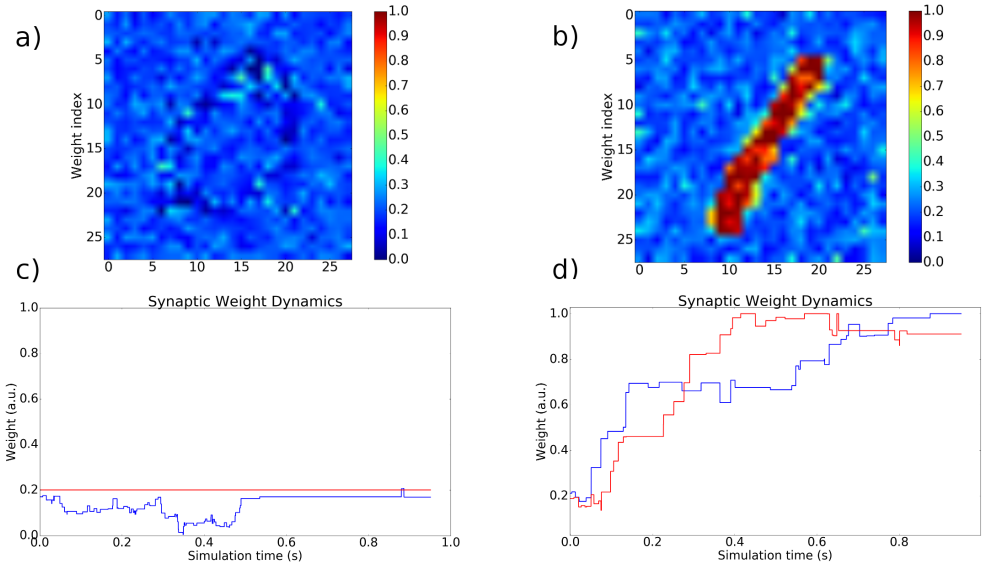


Fig. 5.7 Panel a), synaptic weights, including device variability, after a single 0 digit presented for 900 ms. Panel b), synaptic weights, including device variability, after a single 1 digit presented together with teacher signal. Panels c) and d), dynamics of weights in the former two cases respectively. Blue traces correspond to weight excited by both digits, while red traces correspond to weight excited only by digit 1. Traces make evident LTD process, when the only input is shown, and LTP process when teacher signal is applied. The device variability is evident in the noisy and non-monotonic dynamics.

2. *confined* LTD/LTP bands;

Let us consider the thresholded LTD/LTP case first, in which LTD and LTP bands do not overlap and together they cover the entire spectrum of values accessible by I_{Ca} (see top of figure 5.8). The evolution of the variable I_{mem} is a measure of the short time-scale activity at the post-synaptic neuron input terminal and presents in figure 5.8a) slowly varying or rapidly oscillating I_{mem} behavior whether a 0 or a 1 is shown to the network, respectively. The incoming stimulation provokes the output activity of the post-synaptic neuron that can be appreciated as a raising of the I_{Ca} variable in figure 5.8b). For all the time the Calcium current variable remains in the LTD (LTP) band the synaptic weight could undergo depression (potentiation) depending on the relative value of I_{mem} with respect to the chosen threshold, giving rise to the oscillating behavior shown in figure 5.8c), for two representative synaptic weights corresponding to pixels that are either shared by digits 0s and 1s or specific of digit 1s. It is evident from the previous description that a mean temporal correlation between the firing of the pre-synaptic and the post-synaptic neurons exists as required by a

STDP algorithms. In summary, in case of thresholded LTD/LTP modes, the large synaptic weights oscillation lead to fast learning and fast forgetting dynamics which prevents the storage of the information of the six presentation of the digit 1, as evident by the final configuration of the weight matrix in figure 5.8d).

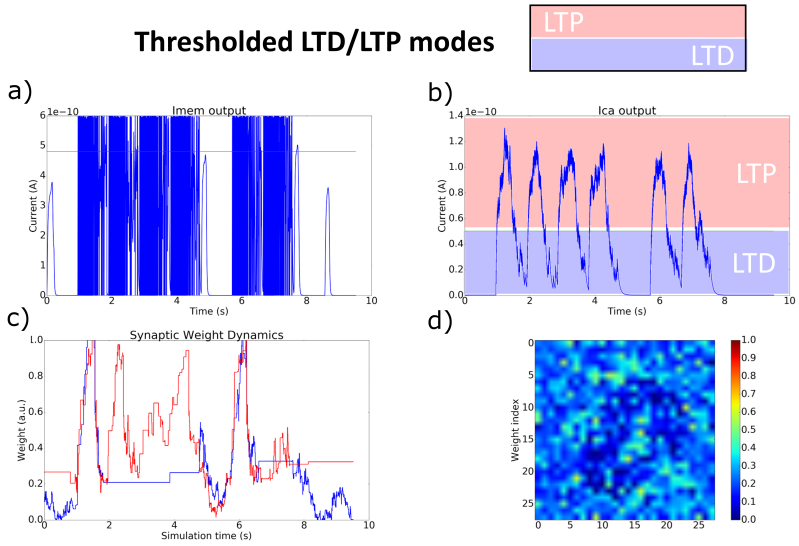


Fig. 5.8 Example of perceptron learning dynamics with thresholded I_{Ca} bands when 10 input digits are presented. Panel a), output neuron membrane current. Low frequency corresponds to digit 0 presented. High frequency corresponds to digit 1 presented together with teacher signal applied. Panel b), output neuron Calcium current acts as a low pass filter of membrane current. Regions for LTD and LTP are highlighted. Panel c), dynamics of weights during learning. Blue trace correspons to weight excited by both digits, red trace corresponds to weight excited only by digit 1. Large I_{Ca} bands allow for fast dynamics of weights that do not preserve features. Panel d), weight map after 10 digits presented.

In order to reduce the impact of this behavior, one needs to slow down the learning process, i.e. to reduce the transition probabilities both in LTD and LTP cases. Such a purpose is pursued limiting the current thresholds involved in the learning process. In particular, I_{Ca} thresholds are particularly suitable to this purpose, because narrowing the delimited bands for transitions act directly on the transition probabilities themselves. Furthermore, the image training time has been reduced to $250ms$.

Thus, in the second analyzed case, LTD and LTP bands are confined and exclude low and high values of I_{Ca}, as shown on top of figure 5.9, which correspond to high and low post-synaptic output firing rates, respectively. In this case, though I_{mem} and I_{Ca} evolutions reported in figure 5.9 are qualitatively similar to those of

figure 5.8, because the stimulation scheme is the same, the change in the LTD/LTP bands results in slowly varying synaptic weights that progressively retain more and more information about the presentation of the digit 1, as evident also by the weight map obtained at the end of the learning process. From this simple analysis is already evident how the choice of thresholds determines the network dynamics and consequently its performances. Moreover, at this stage becomes evident the reason behind the choice of pulse parameter set PPS2 for the device characterization and simulations. Devices characterized with less intense voltage pulses perform more gradual transitions toward the alternative conductance state. In such a way, they intrinsically reduce the impact of quickly varying synaptic weight described above. Slower resistive switching dynamics thus implies a better retention of information acquired during simulation.

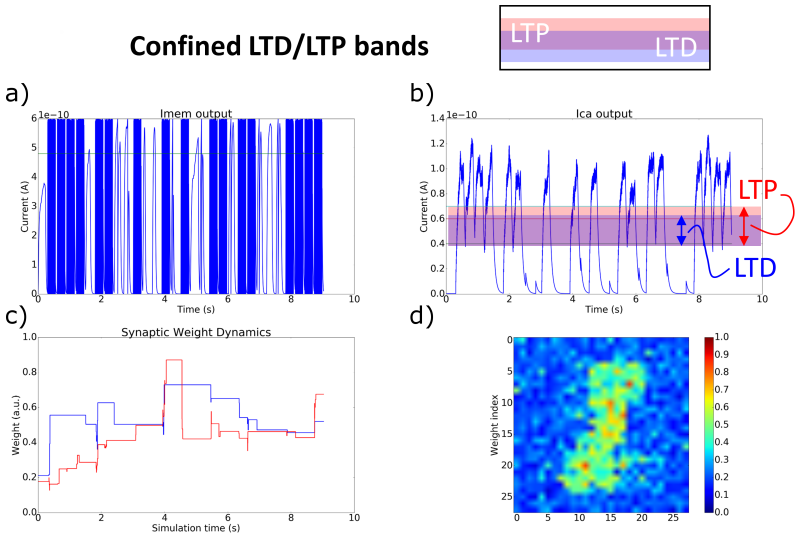


Fig. 5.9 Example of perceptron learning dynamics with confined I_{Ca} bands when 30 input digits are presented. Panel a), output neuron membrane current. Low frequency corresponds to digit 0 presented. High frequency corresponds to digit 1 presented together with teacher signal applied. Panel b), output neuron Calcium current acts as a low pass filter of membrane current. Regions for LTD and LTP are highlighted. Panel c), dynamics of weights during learning. Blue trace corresponds to weight excited by both digits, red trace corresponds to weight excited only by digit 1. Confined I_{Ca} bands allow for slow dynamics of weights that do preserve acquired features. Panel d), weight map after 30 digits presented.

5.4.4 Linear Synapses Network

Before starting with simulations of the complete network embedding resistive switching dynamics, preliminary tests with linear dynamics synapses are led. Linear dynamics for synaptic change means a constant increase or decrease in synaptic weight for each LTP or LTD event, respectively. Such a behavior is defined additive weight dynamics, since it does not involve any dependency on the actual synaptic weight and a constant variation is imposed to the weight. Moreover, in this way the total number of available synaptic states is perfectly defined at the beginning of the simulation. Indeed, these two features are opposite respect to those expressed by the resistive switching dynamics, that on the contrary is multiplicative, i.e. the synaptic weight change depends on the actual synaptic weight, and do not allow to determine precisely the number of available states because of the large variability and the asymptotic behavior. To this extent, these preliminary tests with linear synapses allow first to understand the performance of the designed network and learning algorithm. Then, when resistive switching dynamics will be embedded in the simulation, their impact on performances will be clearly distinguishable.
The network simulated here is composed of the complete Input population and Inhibitory Population, while the Output Population counts 10 neurons, i.e. one neuron per class. Correspondigly, the Teacher Population counts 10 neurons, one active at a time. Three different set of simulations are led, imposing three different weight change amplitudes, as reported in the following:

1. $\delta w = 0.1$; 10 states available
2. $\delta w = 0.01$; 100 states available
3. $\delta w = 0.001$; 1000 states available

Parameters adopted in the network and related to the firing activity are reported in the table 5.3; parameters related to learning activity are instead reported in table 5.4. The main STDP learning rule drawback relies in its difficulty in defining network parameters, being them significantly detached from an optimization model that allows their inference during the learning process itself. More practically, parameters of the network should be empirically determined by the user from the network functioning in a greedy-like fashion. For this reason, and particularly for what concerns thresholds on output neurons variables, multiple tests are performed

with different simulations. Change in the membrane current thresholds as well as on Calcium current thresholds have major impact on the network classification power. Meaningful collection of these tests are reported and briefly discussed in appendix B.

Parameter	$I_{spkthr,ex}$	$I_{spkthr,inh}$	$I_{w_{ex}}$	$I_{w_{inh}}$	$I_{w_{inhout}}$	I_{w_t}
value	600 pA	600 pA	7 pA	5 pA	5 pA	1 nA

Table 5.3 Parameters adopted in the memristive network.

Parameter	$I_{mem,thr}$	$\theta_{D,l}$	$\theta_{D,u}$	$\theta_{D2,l2}$	$\theta_{D2,u}$	$\theta_{P,l}$	$\theta_{P,u}$
value	500 pA	125 pA	132 pA	15 pA	30 pA	120 pA	145 pA

Table 5.4 Thresholds adopted for I_{mem} and I_{Ca} in the doubled LTD band network.

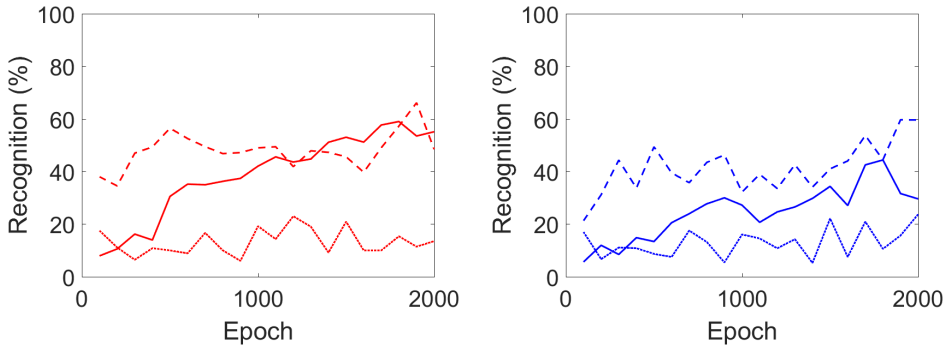
At first, simulations are run for 2000 training images in batch of 100 images at a time, alternated with test batch of 100 images during which classification rate is computed.
Recognition rate in literature is computed in many possible ways, all sharing the common feature of preserving the highest average firing frequency for the targeted output population i, here indicated as $F_i = \frac{1}{N}\sum_{k=1}^{N} f_k$, with N being the number of neurons in a class of the Output Populations and f_k the output neuron firing rate. In the present analysis, it is computed with two separated measures, imposing either a soft-constraint or a strong one: the first, here named r_{avg} computes the average firing rate for each population when a single input is presented to the network. An input digit is taken as correctly classified by the network if the corresponding subpopulation of output neurons fires with the highest average firing rate, i.e. when $F_{i^*} > F_j, \forall j \neq i and i^* = digit$. A stronger-constraint recognition measure is also evaluated with a different procedure. First, the average firing rate of a class is computed on all the test epochs in which a digit associated to that class is presented: $\overline{F_i} = \frac{1}{M_{dgt}}\sum_{j=1}^{M_{dgt}} F_{i,j}$, where M_{dgt} represents the total number of epochs in which a digit dgt belonging to class i is presented and F_i is defined as above. The winner pool, i^*, is the one that when test digit is presented has an average firing rate, F_{i^*}, that

overcomes $\overline{F_{i^*}}$ times an arbitrary threshold θ_{rec}, and in the meanwhile, all other classes have an average firing rate below the same quantity. In formulas:

$$F_{i^*} > \overline{F_{i^*}} \cdot \theta_{rec} \bigvee F_i < \overline{F_i} \cdot \theta_{rec} \forall i \neq i^* \tag{5.9}$$

The threshold is newly chosen at every iteration as the one maximizing the recognition rate and frequently falls in the range $[0.9, 1]$. MNIST digits pixel intensities are multiplied by an amplification factor and linearly converted to Poisson frequency. During training, each digit is applied to the input layer for 250ms, while during testing phase, each digit is applied for 800ms.

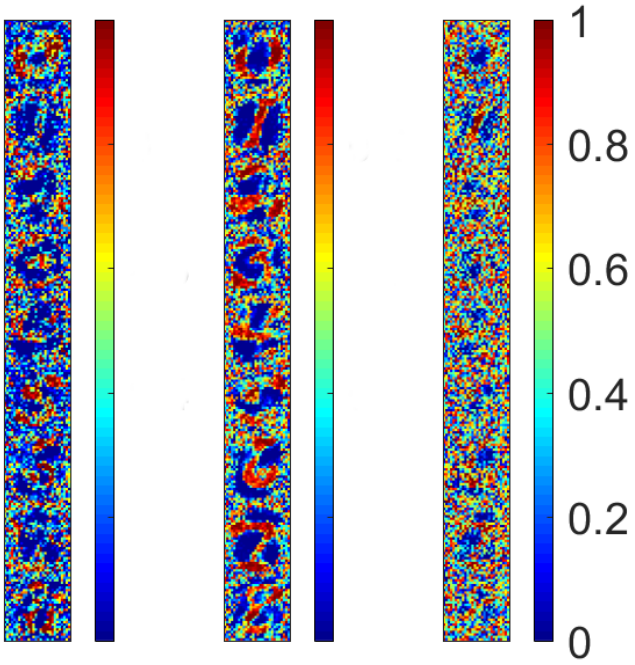
Looking at figure 5.10, it is reported the recognition rate r_{avg}, 5.10a) with red lines, and r_{avg}, 5.10b) with blue lines, for the three cases of linear synapses under analysis. Figure 5.11 reports instead the final weight maps reached after 2000 training images presented and 5.12 the corresponding distributions of synaptic weights.



(a) Recognition rate r_{avg} (b) Recognition rate r_{th}

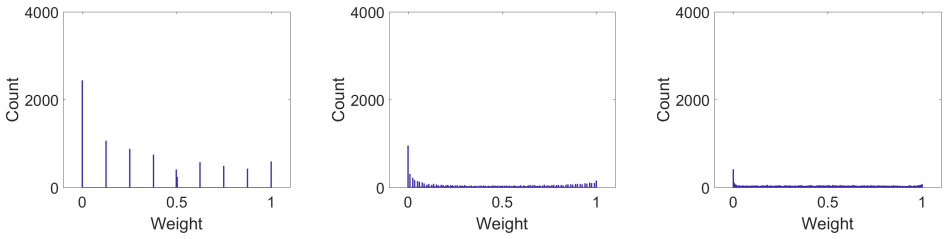
Fig. 5.10 Recognition rate r_{avg}, panel a) red lines, and r_{th}, panel b) blue lines, computed for 10 synaptic states available (dotted line), 100 states (dashed line) and 1000 states (solid line) in a 2000 training images simulation.

As it can be observed, simulation performed with 10 synaptic levels available does not reach a recognition rate above 20% with both measures, meaning that such a low number of available levels together with the simple network architecture are not sufficient to learn ten different patterns. Anyhow, looking at the corresponding weight map in figure 5.11a), it is evident that some features of digit presented start to appear. Eventually, their definition does not allow the network to perform correct



(a) $\delta w = 0.1$ (b) $\delta w = 0.01$ (c) $\delta w = 0.001$

Fig. 5.11 Weightmaps obtained after 2000 training images simulations for network embedding linear synapses. Panel a), 10 states available, Panel b) 100 states available, Panels c) 1000 states available.

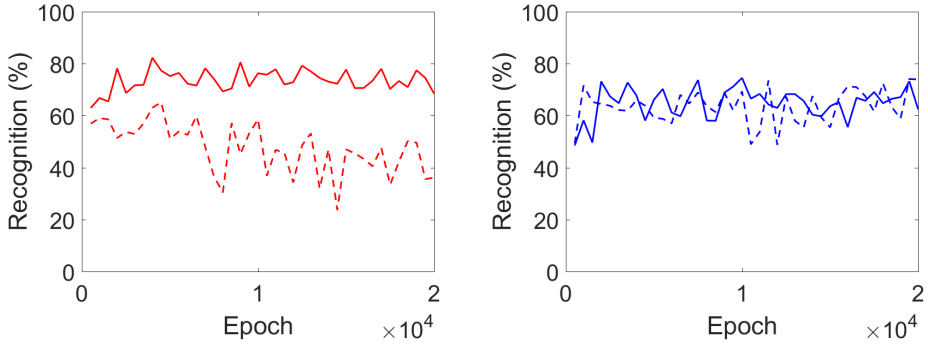


(a) $\delta w = 0.1$ (b) $\delta w = 0.01$ (c) $\delta w = 0.001$

Fig. 5.12 Weight distributions obtained after 2000 training images simulations for network embedding linear synapses. Panel a), 10 states available, Panel b) 100 states available, Panels c) 1000 states available.

classification of the query pattern during the test phase. On the contrary, simulations performed with $\delta w = 0.01$ and $\delta w = 0.001$, i.e. with 100 and 1000 levels available show better results. Indeed, both simulations lead to higher recognition rate with both measures r_{avg} and r_{th}. Focusing on r_{avg}, see 5.10a), the former simulation with

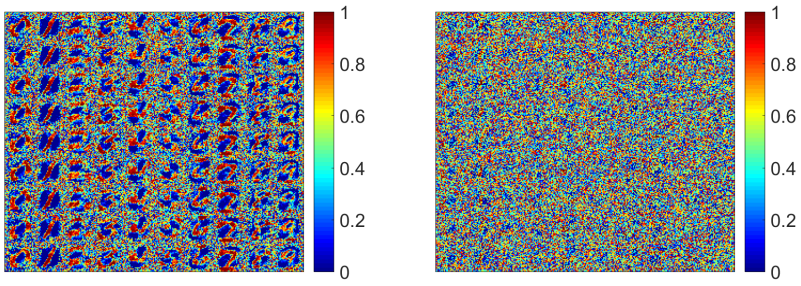
100 states available, dashed red line, oscillates around 50% of correctly classified pattern, while the latter simulation with 1000 states available, solid red line, tends to overcome 60% without reaching saturation. Same behavior is shown monitoring r_{th}, see 5.10b), with overall lower recognition rate with respect to r_{avg}. Indeed, also from the weightmap viewpoint, features in 5.11b) are much more defined that what they are in 5.11c), making emerge that in the latter case the process of feature extraction is still evolving. In the three cases, the distribution of synaptic weight, that was initialized uniformly in the range $[0,1]$, accumulates to the two extremes allowed, producing a bimodal distribution that in this case reduces to two sharp peaks, see figure 5.12. This is an immediate consequence of the additive learning rule, that do not prevent synaptic weights to diverge from their mean value.
Starting from these results, longer simulations are performed in the cases of $\delta w = 0.01$ and $\delta w = 0.001$, discarding the case with $\delta w = 0.1$ because no further improvement would be reached. Moreover, in these simulation are introduced 10 output neurons per class, so that the output population is composed of 100 neurons in total. This is helpful because introduces a voting rule in determining the class that correctly classifies the input by averaging the ouput firing activity among all neurons composing the class. Results are reported in figure 5.13, again reporting different classification rates measured with both r_{avg}, see 5.13a), and r_{th}, see 5.13b).



(a) Recognition rate r_{avg} (b) Recognition rate r_{th}

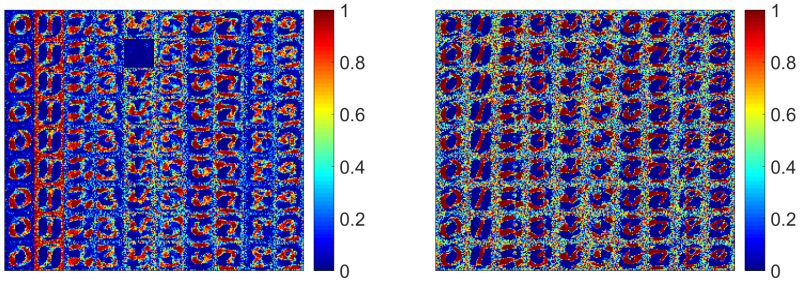
Fig. 5.13 Recognition rate r_{avg}, panel a) red lines, and r_{th}, panel b) blue lines, computed for 100 synaptic states available (dashed line) and 1000 states (solid line) in a 20000 training images simulation.

Two main observations can be done from results reported in figure 5.13. Simulation involving 100 states available soon reaches recognition rate close to 60%, dashed red line, but, after about 5000 train images presented, it begins to constantly decrease



(a) 100 states, 1k images

(b) 1000 states, 1k images

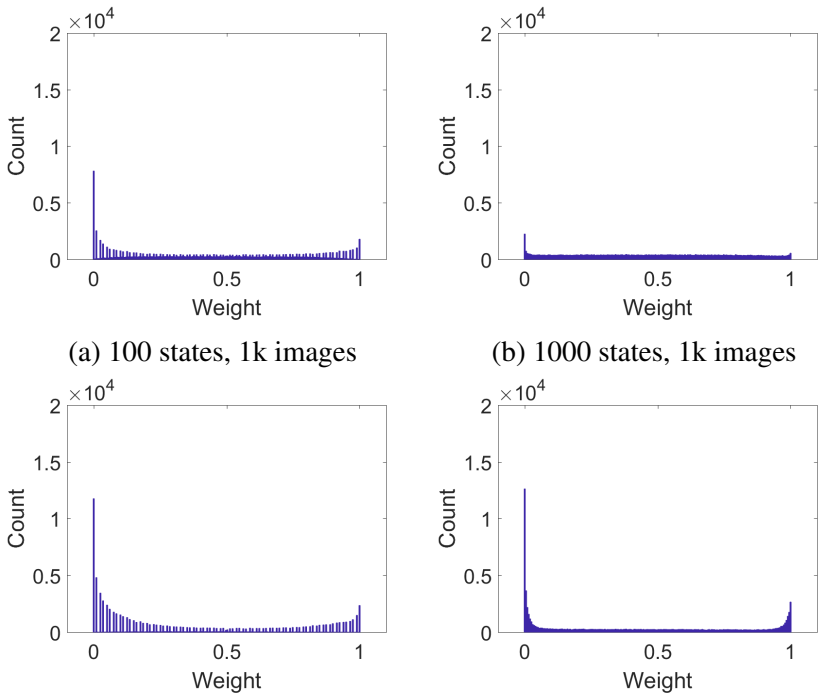


(c) 100 states, 20k images

(d) 1000 states, 20k images

Fig. 5.14 Weightmaps obtained after 1000, Panel a) and b), and 20000, Panel c) and d), training images simulations for network embedding linear synapses. Panel a) and c), 100 states available, Panel b) and d), 1000 states available

for the rest of the simulation. It is most likely a sign of the limited storage capacity of the network, as discussed in Brader et al. [93]. In his paper, Brader correlates the memory capacity of the network to the number of available states offered by synapses, describing the process of forgetting the information acquired as soon as the limit capacity is reached. Such a behavior is less evident in 5.13b), dashed blue line, where r_{th} is monitored, most likely because the strong-constraint recognition rate makes the performances measure safer against early weight variations. On the contrary, this phenomenon is not observed in simulations with $\delta w = 0.001$ and 1000 states available in which saturation is only reached after 5000 training images presented and express a recognition rate stable in between 70% and 80% for both recognition measures, see solid red and blue lines. Moreover in this simulation, defined features are evident from the weightmap reported in figure 5.14, as well as bimodal distribution peaked at the two extreme of the available range of synaptic weight distribution, see figure 5.15. In figure 5.14, comparison between the simulation with 100 states available, 5.14a) and 5.14c), and simulation with 1000 states



(a) 100 states, 1k images

(b) 1000 states, 1k images

(c) 100 states, 20k images

(d) 1000 states, 20k images

Fig. 5.15 Weight distributions obtained after 1000, Panel a) and b), and 20000, Panel c) and d), training images simulations for network embedding linear synapses. Panel a) and c), 100 states available, Panel b) and d), 1000 states available

available, see 5.14b) and 5.14d) are reported. It is evident the slower pace of feature extraction that lead the latter simulation to a more variable weightmap, while in the former simulation discrimination between background and feature extracted is already emerged. The same analysis is reported for weight distribution in figure 5.15. Results reported in this section are taken as reference for simulations performed in the following, since they enclose the classification capacities and limits related to network architecture and learning rule, rather than the effects due to resistive switching components.

5.4.5 Fast Dynamic Memristive Network

The first simulations including resistive switching dynamics performed refers to device characterized with pulse parameter set PPS1, thus with conductance change manifesting a faster dynamics as a function of number of applied voltage pulses. It must be noticed that for this set of simulations, excitatory weights have a differ-

ent initialization, i.e. $w_{ex} = G(\mu = 0.2, \sigma = 0.05)$. As mentioned above, resistive switching dynamics imposes a multiplicative update rule for synaptic weights, that means that the amplitude of weight change is the more important as the lower the weight is and vice-versa. In this way, the resistive switching behavior brings network dynamics toward weights stability, i.e. synaptic weights tends to remain confined close to a mean value rather than diverge during network activity. Stability, generally speaking, is a desirable property, in particular for networks embedding unbounded synaptic weights, while in cases like the one presented in this dissertation, resistive switching are already naturally imposing a soft-bound to the synaptic dynamics. For this reason, multiplicative networks often requires further algorithmic expedient to enable the strong competition among output neurons needed to acquire discriminant features from input pattern. Indeed, if we take as an example the potentiation process of a single synapse, in order to reach the highest reachable value of the synaptic weight a long pulse train of the same voltage polarity is required. However, if such a pulse train is interrupted by the realization of a depression event, LTP process is immediately interrupted and the synaptic weight is decreased with a sudden jump that prevent the asymptotic value to be preserved. Moreover, if LTP would be excessively prominent with respect to LTD, it would favor features blurring, allowing those pixels that stay close to the digit border to be potentiated with few input spikes. With the scope of reducing weight stability effectiveness, it is necessary to act on LTD and LTP events distributions, leveraging on membrane current threshold and calcium current thresholds. Since the former encloses the information about time correlation between pre- and post- spikes, Calcium thresholds are preferred for the further network tuning required.

In order to improve performances in terms of recognition rate of the network and reach satisfying performances with both the recognition measures, learning is slowed down reducing the transition probabilities for both LTD and LTP. Thus, thresholds for calcium current are increased. Moreover, it has been chosen to enforce competitive learning and depression with respect to potentiation introducing a second LTD calcium current band. In literature, equivalent expedients are usually adopted with multiplicative STDP learning networks, [91]. In the specific case reported, Calcium current bands are kept thresholded as in the previous simulations, then a second LTD band is added with thresholds $\theta_{D2,l} = 15pA, \theta_{D2,u} = 30pA$, see table 5.5. The joint effect on learning of teacher signal and doubled calcium band intuitively allow to acquire features of the presented digit within the targeted subpopulation and simulta-

neously to possibly depress correlated weights within different subpopulations. A greater separation between background and feature extracted is thus favored.

Parameter	$I_{mem,thr}$	$\theta_{D,l}$	$\theta_{D,u}$	$\theta_{D2,l2}$	$\theta_{D2,u}$	$\theta_{P,l}$	$\theta_{P,u}$
value	500 pA	125 pA	132 pA	15 pA	30 pA	120 pA	145 pA

Table 5.5 Thresholds adopted for I_{mem} and I_{Ca} in the doubled LTD band network.

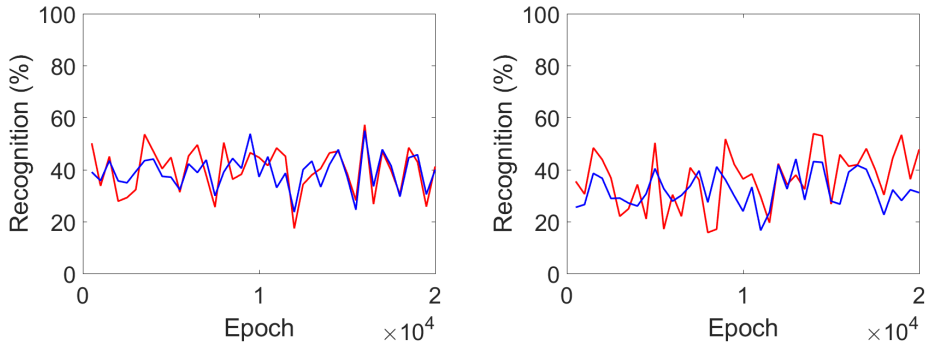


Fig. 5.16 Panel a), recognition rate achieved by PPS1 memristive network with parameters as specified in tables 5.3 and 5.5, without device variability. Performance are computed as r_{avg} (red line) and r_{th} (blue line) across 20k training epochs. Panel b), recognition rate achieved by memristive network with parameters as specified in 5.5 with device variability. Performance are computed as r_{avg} (red line) and r_{th} (blue line) across 20k training epochs.

Results obtained from the simulation of this new network are shown in figure 5.16, comparing performances for synapses not reporting, panel a), and reporting, panel b), resistive switching variability. Plots show recognition performances for both r_{avg} (red line) and r_{th} (blue line). As it can be observed, curves are nicely overlapping, both in terms of the two recognition measures and in terms of device variability. It signifies that device variability is not impeding network learning and performances are not relevantly affected.
Looking at the weight distribution, figure 5.17, as in the previous case histograms compare the weight evolution in case of variability accounted and not accounted at different steps of the simulations. Figure 5.17a) and 5.17c) report the distribution after 1000 and 20000 training images presented, respectively, in the case in which no device variability is embedded. Figure 5.17b) and 5.17d) report the distribution after 1000 and 20000 training images presented, respectively, in the case in which device variability is embedded. All distributions preserve partially the weight initialization,

centered around $w = 0.2$, that creates the mean peak of the distributions. LTP and LTD bring weights to evolve toward the extremes, creating two shoulders at higher weight values and lower weight values. The doubled LTD band is effective in allowing depression toward minimum weight. Nonetheless, in case of variability accounted, distributions highlight more dispersed weights in comparison to what reported in figure 5.17a) and 5.17c). Indeed variability allows to explore the high number of intermediate configuration of resistance state reachable with real devices and without limiting the classification performances.

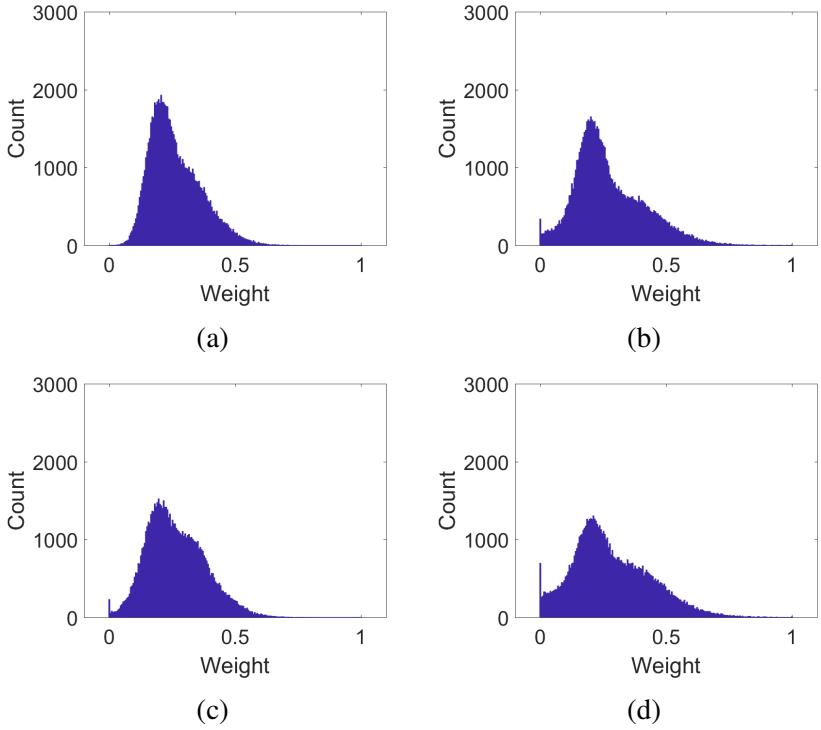


Fig. 5.17 Weight distributions obtained after 1000, Panel a) and b), and 20000, Panel c) and d), training images simulations for network embedding resistive switching synapses. Panel a) and c) report simulations not accounting for device variability, Panel b) and d) report simulations accounting for device variability.

The weight maps, see figure 5.18, reflects this behavior. Panels a) and c) reports the weight map for simulation without device variability at simulation epoch 1000 and 20000. For comparison, panels b) and d) reports weight maps for simulations with device variability at the same epochs.

Figure 5.18 highlights two main features: first, the chosen thresholds on Calcium current allow a consistent depression nearby digits (see for example digit 1) and

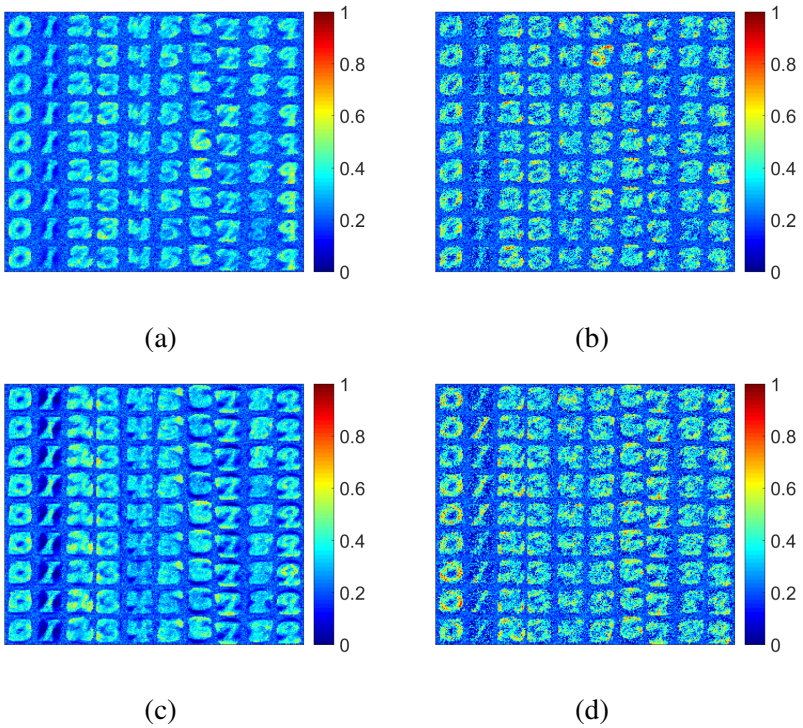


Fig. 5.18 Comparison between features maps of synaptic weights for network without device variability, panel a) and c), and network with device variability, panels b) and d). Maps are reported after 1000 epochs, panel a) and b), and after 20000 epochs, panel c) and d).

an enhancement of those subsets of features that are discriminative for the digit itself (see digit 2 enhanced at its extrema). Evidently, such a process does not reach a conclusion within the simulated training epochs and acquired features are too broad to accomplish better classification. Secondly, comparison between features without and with device variability do not show evident differences, but for a higher homogeneity in acquired features in the former case. Thus, from the perspective of recognition rate, weight distribution and acquired features, the network is robust against resistive switching variability.

These results show the importance of an extended LTD distribution that allows to reach the overlap between the two recognition measures. It also initiates the acquisition of discriminative features during learning, but still it is not sufficient to reach good classification performances.

5.4.6 Slow Dynamic Memristive Network

The set of simulation reported in this section are performed taking into account the resistive switching dynamics characterized with pulse parameter set PPS2. Accordingly to such electric characterization, the switching dynamics is slower and more gradual, tending toward the asymptotes in a larger number of applied pulses. Anyhow, switching dynamics is still strongly non-linear and gives rise to a multiplicative update rule for synaptic weights. On the other hand, the gradual process leading to extreme values of synaptic weights, allows to virtually explore a larger number of available synaptic states, with respect to those available when a fast dynamic resistive switching is accounted.
Network parameters adopted in the present simulations are as those reported in table 5.3 and 5.5, adopted in the previous simulations. In this network, excitatory synapses weights are randomly initialized with uniform distribution in the range $[0, 1]$.

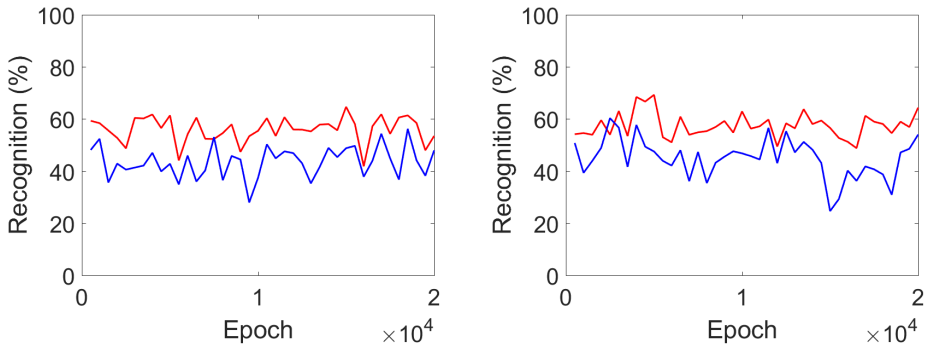


Fig. 5.19 Panel a), recognition rate achieved by PPS2 memristive network with parameters as specified in tables 5.3 and 5.5, without device variability. Performance are computed as r_{avg} (red line) and r_{th} (blue line) across 20k training epochs. Panel b), recognition rate achieved by memristive network with parameters as specified in 5.5 with device variability. Performance are computed as r_{avg} (red line) and r_{th} (blue line) across 20k training epochs.

Results found from simulations of this new network are shown in figure 5.19, in which performances are compared for synapses not accounting, panel a), and accounting, panel b), resistive switching variability. Plots show recognition performances for both r_{avg} (red line) and r_{th} (blue line). First of all, again comparing the two performances of the classification tasks, variability is not affecting the network functioning and classification rates are not sensibly reduced. Moreover, recognition

rates are higher than those obtained with devices characterized with pulse parameter set PPS1, reaching an average rate close to 60% for r_{avg} measure and slightly lower for r_{th}. Indeed, the lower synaptic weight change allows to explore the number of synaptic states available and improve learning abilities.
Such a recognition rate is similar to what obtained for linear synapses with 100 and 1000 synaptic states available. In particular, staying bounded between the two results observed in the previous case, classification abilities of the new network do not appear to be degraded as the simulation proceeds, like in the case of $\delta w = 0.01$ linear simulations. Moreover, the present network does not reach average classification rate as high as in $\delta w = 0.001$ linear weight change network. Thus, it can be deduced that the number of available synaptic states offered by resistive switching expressing a slow dynamics are included in between 100 and 1000. They should be higher than the former, since the network preserve the information learned without classification performance degradation, and lower than the latter, because of the slightly worse performances in terms of recognition rates.
Looking at the weight distribution, figure 5.20, as in the previous case histograms compare the weight evolution in case of variability not accounted and accounted at different steps of the simulations at different time steps of the simulation. Figure 5.20a) and 5.20c) report the distribution after 1000 and 20000 training images presented, respectively, in the case in which no device variability is embedded. Figure 5.20b) and 5.20d) report the distribution after 1000 and 20000 training images presented, respectively, in the case in which device variability is embedded. Again, the multiplicative weight update rule brings rapidly the synaptic weights toward a unimodal distribution centered close to the mid value of the synaptic weight range. All weights, including those that lay in the background of the input pattern, when evolving as a consequence of alternative LTP and LTD events, move toward an average value, leaving the weight map with a low contrast between what is background and what digit features, see figure 5.21a-c). Nonetheless, digit features are just visible on top of the weight map, and drive the recognition rate toward the observed value. Indeed, the weight distribution observed in figure 5.20 is not symmetric, but present a shoulder toward lower weights. Such a shoulder represents those weights belonging to pixels common to multiple digits, that when excited by digits from another class are depressed. This behavior manifests the tendency of the network toward a bimodal distribution, that encodes information belonging to the pattern background (synaptic depression leads toward a peak close to minimum weights)

and discriminative features of the pattern (synaptic potentiation leads toward a peak close to the maximum weights). On the other hand, multiplicative weight update rule imposed by resistive switching dynamics drastically limits the drift of the two peaks one apart from eachother, reducing also the classification performances of the network.

In order to put more in evidence these aspects of the network and the tendency to acquire discriminative features of input digit, in Figure 5.22 is reported the weightmap reached after 20000 images presented, without device variability, with color-map rescaled in a narrower range of weights. Here it becomes evident the difference between the enhanced digit features in red, encoded by synapses that become potentiated, and the digit background in blue, where synapses become suppressed.

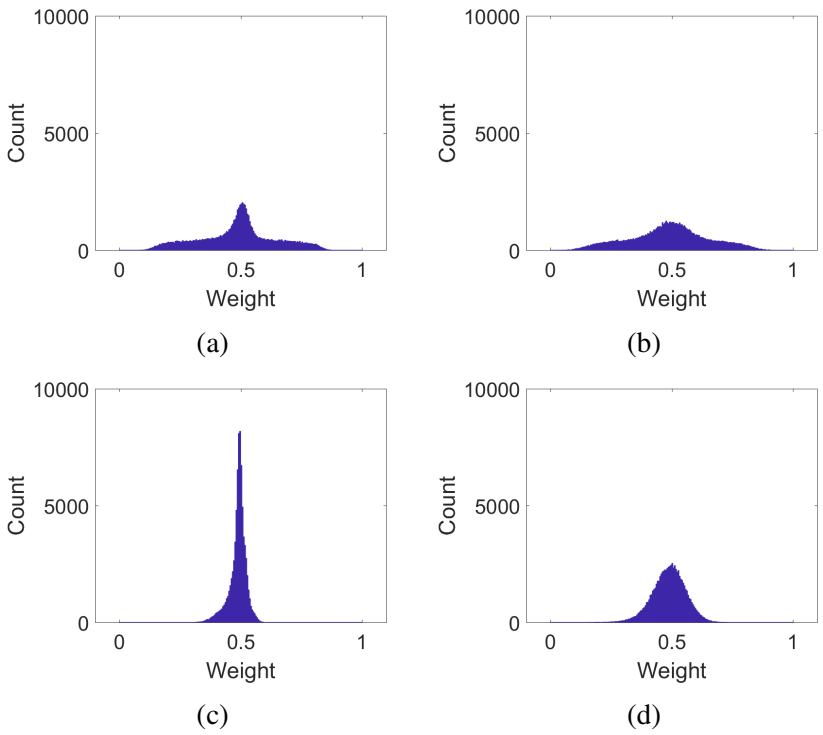


Fig. 5.20 Weight distributions obtained after 1000, Panel a) and b), and 20000, Panel c) and d), training images simulations for network embedding resistive switching synapses. Panel a) and c) report simulations not accounting for device variability, Panel b) and d) report simulations accounting for device variability.

The same analysis is reported for the case with device variability considered in the simulation and showed in figure 5.20b) and 5.20d). As it can be observed, weights distribution preserves show the same unimodal peak centered ad mid-values,

but with a sensibly higher dispersion in comparison to what reported in figure 5.20a) and 5.20c). Indeed variability allows to explore the high number of intermediate configurations of resistance state reachable with real devices without limiting the classification performances.

The weight maps, see figure 5.21, reflect this behavior. Panels a) and c) reports the weight map for simulation without device variability at simulation epoch 1000 and 20000. For comparison, panels b) and d) reports weight maps for simulations with device variability at the same epochs. Similarities between the two conditions are particularly relevant. Thus, from the perspective of recognition rate and weight distribution, the network can be said robust against resistive switching variability.

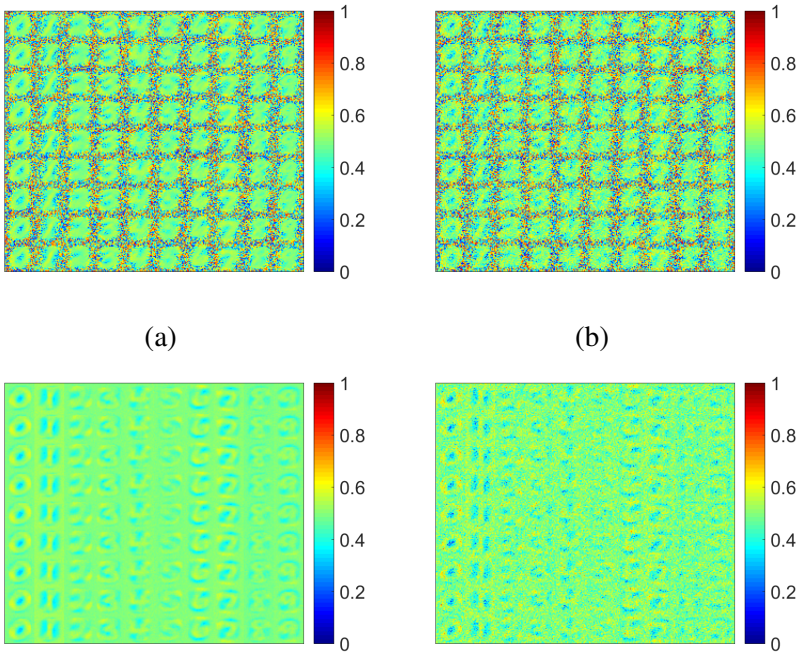


Fig. 5.21 Comparison between features maps of synaptic weights for network without device variability, panel a) and c), and network with device variability, panels b) and d). Maps are reported after 1000 epochs, panel a) and b), and after 20000 epochs, panel c) and d).

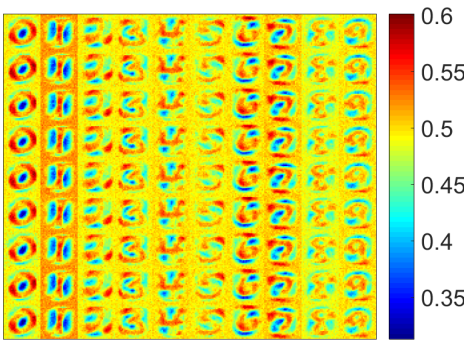


Fig. 5.22 Weightmap of synaptic weights for network without device variability after 20000 training images presented. False colors map has been rescaled in a narrow range in order to put in evidence discriminative features acquired.

5.5 Conclusions

In the present chapter it has been presented the complete study performed for the design of a memristor-based spiking neural network.
Resistive switching adopted for the task are produced and characterized by CNR - Agrate Brianza, partner of the scientific collaboration set up for the present study, together with the Institute of Neuroinformatics of Zuerich. RS are HfO-based MIM structures characterized with trains of pulses with equal polarity and period. Two sets of electrical parameters are adopted for device characterization. Empirical models for the resistance transition to HRS (LTD) and LRS (LTP) are then extrapolated from data, capturing the pulse dependent behavior of both processes. Cycle-to-cycle variability has been quantified as well. All these features are accounted in simulations of the designed spiking neural network, demanded to perform a learning task with handwritten digits coming from the MNIST database.
Spiking Neural Network accounts for four different populations: input population, converting pixel intensities into Poisson spike trains, output population, composed of Silicon I&F neurons mimicking the biological behavior and connected to IP through resistive switching plastic synapses. Inhibitory population, composed of Silicon I&F neurons and inputing the OP with negative current, and to conclude, teacher population, designed to perform semi-supervised learning.
The adopted learning rule is a generalization of STDP formalized by Fusi and Brader in 2007. It is particularly suitable for hardware application since it involves two slow varying variables, i.e. membrane current and Calcium current of the output neurons,

to discriminate between LTD and LTP events.
After preliminary tasks with a perceptron network, performed in order to verify the correct function of all elements of the network and the empirical behavior of resistive switching synapses, the study has proceeded with simple network embedding synapses with linear weight update. The different amount of available weight states during network training has been shown to deeply affect the network performances in terms of recognition rates, stating the better performances, the more synaptic states available. Results obtained at this stage are taken as reference for the network learning capabilities and compared with results obtained from networks embedding resistive switching synapses.
Simulation running the complete network with 10 neurons per digit output class are then performed, comparing performances reached with both fast and slow dynamics resistive switching synapses. First, a slow dynamics in conductance change has been shown to be preferable, since it allows performances closer to linear synapses with a sufficiently high number of synaptic level. In this way, slow dynamic resistive switching synapses offer a higher number of weight levels. Moreover, the study highlights the mandatory requirement of narrow I_{Ca} bands in order to limit the number of events for both LTD and LTP. I_{Ca} bands are placed at high spiking frequency, in order to capture variation during teacher imposed learning, and are associated with a second LTD band at lower frequency, allowing for depression of correlated digit features. This choice for thresholds allow to start the acquisition of discriminative features of input digits and to increase the network recognition rate with both measures defined, i.e. r_{avg} and r_{th}. By the way, it is not sufficient to break the weight stability due to multiplicative weight update rule intrinsically embedded in resistive switching. Indeed, weight update relies on the new value of resistance imposed by a new incoming pulse, that has been shown to be pulse-dependent, highly non-linear and asymmetric, e.g. high with low number of pulses and saturating with high number of pulses. Such a feature brings weight distribution to be unimodal and peaked at weight mean value, making it unable to discriminate between input digit features and input background.
The present study does not go into the change to network architecture demanded to break weight stability, that are postponed to further studies.

Chapter 6

Conclusions

The present dissertation wants to be a preparatory study for the new generation of smart devices based on hardware artificial intelligence. An always increasing number of publications has appeared in the recent years in literature discussing properties of neural networks implemented in various hardware architectures, both with standard Silicon technology and with new type of devices. The renown interest for brain-inspired hardware neural networks has been strongly pushed forward by the resistive switching technology and its description with the memristor formalism. Indeed, memristor could behave as artificial synapses able to tune their strength, i.e. conductance and synaptic efficacy depending on the field of application, as a consequence of the history of electric signals applied to the two terminals device. The applied voltage imposes a change in the device resistance that is then preserved and used to modulate the input information transferred across the network. A pletora of studies concern the various architectures, nanostructures and materials adopted for resistive switching fabrication and operation, without converging toward a standard for memristive devices exploitable in intelligent hardware design.
The present study, realized as the start of a new research line for the group in this university, is an investigation of the physical mechanisms involved in the resistive switching phenomenon as well as an engineering process devoted to the fabrication of devices suitable for synaptic application in hardware neural networks. The second part of the research activity is then focused on acquiring the fundamental knowledge about spiking neural network simulation tools, STDP learning rule and the functioning of hardware building blocks composing the designed circuit, such as I&F Silicon neurons. An extensive analysis of the learning behavior is lead, in

order to comprehend the dynamics of resistive switching devices during learning task, together with advantages and limitations imposed by the physics governing the device functioning.
At first, TiO_2 based resistive switching are investigated. Nanostructured TiO_2, in the form of nanotubes arrays, are grown via anodic oxidation on Ti foil. Both as-grown and air annealed material is studied and adopted for device fabrication with poly-acrylic acid coating and Platinum top electrode. The multi-wall structure generated during tube growth can be exploited for installing the resistive switching behavior based on Oxygen vacancies movement. Indeed, valence change memories relies on the creation of a conductive filament composed of Oxygen vacancies present in under-stoichiometric material. The PAA coating is then introduced in the device structure as a tuner for charge carriers density and Oxygen vacancies concentration, allowing to establish a more stable resistive switching phenomenon in terms of repeatability and endurance. More in detail, the study on PAA coated TiO_2 NTA allows to investigate the RS behavior of VCM exploiting chemical surface redox reactions, that are particularly advantageous in a high aspect-ratio device such as the one here presented. Despite the great interest in the physical phenomena involved in RS for nanostructures, such a kind of devices are of hard implementation in more complex electronic circuits. For this reason, TiO_2 thin film devices are realized.
TiO_2 is among the first materials studied for the realization of RS and formalized as memristor. Numerous laboratories, both in private companies as well as in universities, devoted great effort in the modeling and engineering of resistive switching based on this material, aiming to fabricate the next generation of dense, low-power ReRAM memories. Particular interest comes also from the opportunity to fabricate TiO_2 based devices via Atomic Layer Deposition, an extremely controllable and accurate techniques for thin layer depositions. On the contrary, patterning of TiO_2 is a problematic step in circuit fabrication because of the mechanical resistance of the material and usually results in imprecise process or even failure. The investigation reported in the present dissertation is thus devoted to validate a new recipe for TiO_2 device fabrication based on low temperature deposition and optical lithography.
Low temperature deposition of TiO_2 via ALD is performed in the range [80°C, 140°C] and an inverse proportionality between growth per cycle and temperature is observed. High GPC induces the formation of defects and inclusions on the thin layer, due to possible processes of condensation or other chemical reactions as by product of the deposition. On the contrary, correct patterning of the material

needed for device fabrication is successfully demonstrated. Moreover, no significant change in material properties is observed in all characterizations performed. From the electrical characterization standpoint, RS is observed with comparable features in all the measured devices, i.e. 120°C, 140°C and 230°C as reference. In conclusion, low deposition temperatures do not affect the RS switching properties of the material and enables a new and easier fabrication process with respect to the most adopted one.

TiO_2 based RS devices do not express particularly stable electrical behavior and, on the contrary, manifest a quite high variability in RS response. This issue is addressed in numerous studies present in literature and has lead the scientific community toward the consideration of different materials among metal oxides.

The last material investigated for the RS properties is ZnO. ZnO offers the opportunity to be deposited exploiting different techniques. i.e. ALD and sputtering, and grown in nanostructures such as nanowires. Object of the present dissertation is ZnO thin films RS deposited via sputtering. A number of devices has been realized and studied, both in VCM and ECM configuration, in order to obtain the most accurate and stable electrical response from the produced devices. Devices with Pt top electrode and Cu top electrode are fabricated at crescent thickness, i.e. 50 nm, 100 nm and 250 nm.

FESEM characterization shows a columnar structure of the layer that goes to be more evident at highest thicknesses. Raman spectroscopy shows material typical peaks of nanocrystalline phase and XRD analysis shows preferred orientation (002) and (101). Deposition of thicker ZnO layer returns more intense peaks and highlights the prevalence of (002) orientation in the film. XPS analysis shows nearly stoichiometric material.

Electrical characterization of thinner ZnO highlights unstable RS behavior for VCM type devices, while ECM type devices do not express any significant electrical response. Indeed, irreversible breakdown is reached for the characterized devices in ECM configuration. The increase in ZnO layer thickness aims to provide higher concentration of Oxygen vacancies available to RS in the VCM type cell, prevalently present on grain surfaces. Anyhow, electrical response of ZnO - 100 nm devices in VCM configuration, despite being improved, is not yet satisfactory because of high failure rate in installing the SET transition. On the contrary, ECM device shows much better RS response. Thicker ECM type ZnO device is the best performing and returns stable and repeatable RS response in terms of endurance and state retention,

showing the narrowest standard deviations on all measured quantities. The fact that such a thick MIM structure produce the stablest RS response let think that the formation of a CF does not require to be continuous and can instead exploit the preferred orientation of crystals within the ZnO layer.
The observed performances, despite not being sufficient for the production of commercial ReRAM, could be preliminary exploitable for the realization of neuromorphic circuits. Anyhow, ZnO would not be the preferred material for circuit realization because of its relatively high conductance, that would results in high power dissipation during the circuit functioning. For this reason, other insulating materials, like HfO and Ta_2O_5 are preferred to this purpose. Moreover, voltage pulse characterization is required for correct embedding of resistive switching in spiking neural network, and it is not available yet for the present devices. For this reasons, a collaboration between the National Council for Research, site of Agrate Brianza - Italy, and the Institute of Neuroinformatics in Zuerich - Switzerland has been established.
The last part of this dissertation is devoted to the simulation of spiking neural network embedding HfO based resistive switching and totally designed for its hardware implementation.
Pulse characterization of RS returns highly-non linear and asymmetric behavior for SET transition to LRS and RESET transition to HRS. Data collected by the CNR group are used to extrapolate an empirical models of the transitions, complete of cycle to cycle variability, that can be embedded into the neural network simulations. SET transition, when remapped into the learning context, is equivalent to synaptic long term potentiation, while RESET transition is equivalent to long term depression. Both LTP and LTD show a fundamental dependence on the pulse numbers for the conductance update, that results in a so called multiplicative weight update rule. Indeed, weight update, i.e. conductance update, is high with a low number of pulses and low with high number of pulses.
The network simulates the function of real hardware circuit demanded to perform the learning of handwritten digits present in the MNIST database exploiting a generalized and hardware suitable STDP learning rule. MNIST classification is a task considered as benchmark for the testing of neural networks as well as different learning rules. The network architecture considered in this study is a single layer network composed of four different neuron populations: input population, converting pixel intensities into Poisson spike trains, output population, composed of Silicon

I&F neurons mimicking the biological behavior and connected to IP through resistive switching plastic synapses, inhibitory population, composed of Silicon I&F neurons and inputing the OP with negative current, and, to conclude, teacher population, designed to perform semi-supervised learning.
After preliminary test of a simplified perceptron network functioning and simulations with linear synapses, complete analysis of the learning dynamics with resistive switching dynamics is performed. Indicative recognition performances are reached with opportunely chosen thresholds on neurons variables, but the main issue that requires to be addressed in order to reach state-of-the-art learning performances comes from multiplicative weight update due to RS physical behavior. Indeed dependence on pulse number for the weight update brings dynamics towards a stable configuration with weights distributed in a unimodal distribution peaked at weight mean value. This particular feature impedes to acquire discriminative features of the input and thus correctly classify the presented digits.
Future work will still see a long way before of real hardware implementation of memristive neuromorphic systems, that goes from the resistive switching physical phenomenon description, to the proper memristive modeling, concluding to a formalized STDP learning rule able to solve issues imposed by multiplicative weight updates without the introduction of complex network architectures. Nonetheless, the author of the present dissertations is faithful that in the nearest future, this technology will grow side by side and even further than the actual state-of-the-art technology.

References

[1] Dan Cireşan, Ueli Meier, Jonathan Masci, and Jürgen Schmidhuber. Multi-column deep neural network for traffic sign classification. *Neural Networks*, 32:333–338, aug 2012.

[2] Biswa Sengupta and Martin B. Stemmler. Power Consumption During Neuronal Computation. *Proceedings of the IEEE*, 102(5):738–750, may 2014.

[3] John von Neumann. First Draft of a Report on the EDAVAC. *IEEE Annals of the History of Computing*, 15(4):27–75, 1993.

[4] Suzana Herculano-Houzel. The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in Human Neuroscience*, 3(November):1–11, 2009.

[5] Joshua T. Trachtenberg, Brian E. Chen, Graham W. Knott, Guoping Feng, Joshua R. Sanes, Egbert Welker, and Karel Svoboda. Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. *Nature*, 420(6917):788–794, dec 2002.

[6] O. Steward W. B. Levy. Temporal Contiguity Requirements for Long-Term Associative Potentiation/Depression in the Hippocampus. *Neuroscience*, 8(4):791–797, 1983.

[7] H. Markram. Regulation of Synaptic Efficacy by Coincidence of Postsynaptic APs and EPSPs. *Science*, 275(5297):213–215, jan 1997.

[8] Henry Markram, Wulfram Gerstner, and Per Jesper Sjöström. A history of spike-timing-dependent plasticity. *Frontiers in Synaptic Neuroscience*, 3(AUG):1–24, 2011.

[9] H. Markram, W. Gerstner, and P. J. Sjöström. Spike-timing-dependent plasticity: A comprehensive overview. *Frontiers in Synaptic Neuroscience*, 4(JULY):2010–2012, 2012.

[10] Zheng Yang, Changhyun Ko, and Shriram Ramanathan. Oxide Electronics Utilizing Ultrafast Metal-Insulator Transitions. *Annual Review of Materials Research*, 41(1):337–367, 2011.

[11] Chris A. MacK. Fifty years of Moore's law. *IEEE Transactions on Semiconductor Manufacturing*, 24(2):202–207, 2011.

[12] G E Moore. Cramming more components onto integrated circuits (Reprinted from Electronics, pg 114-117, April 19, 1965). *Proceedings Of The Ieee*, 86(1):82–85, 1965.

[13] Shekhar Borkar. Designing reliable systems from unreliable components: The challenges of transistor variability and degradation. *IEEE Micro*, 25(6):10–16, 2005.

[14] Yiming Huai. Spin-Transfer Torque MRAM (STT-MRAM): Challenges and Prospects. *AAPPS bulletin*, 18(6):33–40, 2008.

[15] Matthias Wuttig and Noboru Yamada. Phase-change materials for rewriteable data storage. *Nature Materials*, 6(11):824–832, 2007.

[16] Rainer Waser, Regina Dittmann, Ceorgi Staikov, and Kristof Szot. Redox-based resistive switching memories nanoionic mechanisms, prospects, and challenges. *Advanced Materials*, 21(25-26):2632–2663, 2009.

[17] J. Joshua Yang, Dmitri B. Strukov, and Duncan R. Stewart. Memristive devices for computing. *Nature Nanotechnology*, 8(1):13–24, 2012.

[18] Kyung Min Kim, Gun Hwan Kim, Seul Ji Song, Jun Yeong Seok, Min Hwan Lee, Jeong Ho Yoon, and Cheol Seong Hwang. Electrically configurable electroforming and bipolar resistive switching in Pt/TiO2/Pt structures. *Nanotechnology*, 21(30):305203, 2010.

[19] Ruijing Ge, Xiaohan Wu, Myungsoo Kim, Jianping Shi, Sushant Sonde, Li Tao, Yanfeng Zhang, Jack C. Lee, and Deji Akinwande. Atomristor: Nonvolatile Resistance Switching in Atomic Sheets of Transition Metal Dichalcogenides. *Nano Letters*, 18(1):434–441, jan 2018.

[20] Samuele Porro, Eugenio Accornero, Candido Fabrizio Pirri, and Carlo Ricciardi. Memristive devices based on graphene oxide. *Carbon*, 85:383–396, apr 2015.

[21] Tatiana Berzina, Anteo Smerieri, Marco Bernab, Andrea Pucci, Giacomo Ruggeri, Victor Erokhin, and M. P. Fontana. Optimization of an organic memristor as an adaptive memory element. *Journal of Applied Physics*, 105(12), 2009.

[22] A. Chiolerio, I. Roppolo, K. Bejtka, A. Asvarov, and C. F. Pirri. Resistive hysteresis in flexible nanocomposites and colloidal suspensions: interfacial coupling mechanism unveiled. *RSC Adv.*, 6(61):56661–56667, 2016.

[23] Samuele Porro, Alladin Jasmin, Katarzyna Bejtka, Daniele Conti, Denis Perrone, Salvatore Guastella, Candido F. Pirri, Alessandro Chiolerio, and Carlo Ricciardi. Low-temperature atomic layer deposition of TiO2 thin layers for the processing of memristive devices. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 34(1):01A147, 2016.

[24] Tong Liu, Yuhong Kang, Sarah El-Helw, Tanmay Potnis, and Marius Orlowski. Physics of the voltage constant in multilevel switching of conductive bridge resistive memory. *Japanese Journal of Applied Physics*, 52(8), 2013.

[25] Marco Laurenti, Samuele Porro, Candido F. Pirri, Carlo Ricciardi, and Alessandro Chiolerio. Zinc Oxide Thin Films for Memristive Devices: A Review. *Critical Reviews in Solid State and Materials Sciences*, 42(2):153–172, 2017.

[26] S Brivio, G Tallarida, E Cianci, and S Spiga. Formation and disruption of conductive filaments in a HfO2/TiN structure. *Nanotechnology*, 25(38):385705, 2014.

[27] T. D. Dongale, S. V. Mohite, A. A. Bagade, P. K. Gaikwad, P. S. Patil, R. K. Kamat, and K. Y. Rajpure. Development of Ag/WO3/ITO thin film memristor using spray pyrolysis method. *Electronic Materials Letters*, 11(6):944–948, 2015.

[28] S. G. Hu, Y. Liu, T. P. Chen, Z. Liu, Q. Yu, L. J. Deng, Y. Yin, and Sumio Hosaka. Emulating the Ebbinghaus forgetting curve of the human brain with a NiO-based memristor. *Applied Physics Letters*, 103(13):1–5, 2013.

[29] Jae Sung Lee, Shinbuhm Lee, and Tae Won Noh. Resistive switching phenomena: A review of statistical physics approaches. *Applied Physics Reviews*, 2(3):031303, sep 2015.

[30] S. Asanuma, H. Akoh, H. Yamada, and A. Sawa. Relationship between resistive switching characteristics and band diagrams of Ti/Pr(1-x)CaxMnO3 junctions. *Physical Review B*, 80(23):235113, dec 2009.

[31] Akihito Sawa. Resistive switching in transition metal oxides. *Materials Today*, 11(6):28–36, 2008.

[32] Rainer Waser and Masakazu Aono. Nanoionics-based resistive switching memories. *Nature materials*, 6(11):833–40, 2007.

[33] Feng Miao, John Paul Strachan, J. Joshua Yang, Min Xian Zhang, Ilan Goldfarb, Antonio C. Torrezan, Peter Eschbach, Ronald D. Kelley, Gilberto Medeiros-Ribeiro, and R. Stanley Williams. Anatomy of a nanoscale conduction channel reveals the mechanism of a high-performance memristor. *Advanced Materials*, 23(47):5633–5640, 2011.

[34] Ilia Valov, Rainer Waser, John R Jameson, and Michael N Kozicki. Electrochemical metallization memories—fundamentals, applications, prospects. *Nanotechnology*, 22(28):289502, 2011.

[35] Umberto Celano, Ludovic Goux, Attilio Belmonte, Karl Opsomer, Alexis Franquet, Andreas Schulze, Christophe Detavernier, Olivier Richard, Hugo Bender, Malgorzata Jurczak, and Wilfried Vandervorst. Three-dimensional observation of the conductive filament in nanoscaled resistive memory devices. *Nano Letters*, 14(5):2401–2406, 2014.

[36] Dmitri B Strukov, Gregory S Snider, Duncan R Stewart, and R Stanley Williams. The missing memristor found. *Nature*, 453(7191):80–3, 2008.

[37] Leon Chua. Memristor-The missing circuit element. *IEEE Transactions on Circuit Theory*, 18(5):507–519, 1971.

[38] Fernando Corinto and Mauro Forti. Memristor Circuits : Flux — Charge Analysis Method. 63(11):1997–2009, 2016.

[39] L.O. Chua and Sung Mo Kang. Memristive devices and systems. *Proceedings of the IEEE*, 64(2):209–223, 1976.

[40] Janardan Misra and Indranil Saha. Artificial neural networks in hardware: A survey of two decades of progress. *Neurocomputing*, 74(1-3):239–255, 2010.

[41] Paolo Livi and Giacomo Indiveri. A current-mode conductance-based silicon neuron for Address-Event neuromorphic systems. *Proceedings - IEEE International Symposium on Circuits and Systems*, pages 2898–2901, 2009.

[42] Elisabetta Chicca, Fabio Stefanini, Chiara Bartolozzi, and Giacomo Indiveri. Neuromorphic electronic circuits for building autonomous cognitive systems. *Proceedings of the IEEE*, 102(9):1367–1388, 2014.

[43] Andrea Lamberti, Adriano Sacco, Stefano Bianco, Diego Manfredi, Marco Armandi, Marzia Quaglio, Elena Tresso, and Candido F. Pirri. An easy approach for the fabrication of TiO2 nanotube-based transparent photoanodes for Dye-sensitized Solar Cells. *Solar Energy*, 95:90–98, 2013.

[44] Paola Rivolo, Sara Maria Severino, Serena Ricciardi, Francesca Frascella, and Francesco Geobaldo. Protein immobilization on nanoporous silicon functionalized by RF activated plasma polymerization of Acrylic Acid. *Journal of Colloid and Interface Science*, 416:73–80, 2014.

[45] Marco Laurenti, Alessio Verna, and Alessandro Chiolerio. Evidence of Negative Capacitance in Piezoelectric ZnO Thin Films Sputtered on Interdigital Electrodes. *ACS Applied Materials and Interfaces*, 7(44):24470–24479, 2015.

[46] M. Laurenti, S. Stassi, M. Lorenzoni, M. Fontana, G. Canavese, V. Cauda, and C. F. Pirri. Evaluation of the piezoelectric properties and voltage generation of flexible zinc oxide thin films. *Nanotechnology*, 26(21):215704, may 2015.

[47] S. Brivio, J. Frascaroli, and S. Spiga. Role of metal-oxide interfaces in the multiple resistance switching regimes of Pt/HfO 2 /TiN devices. *Applied Physics Letters*, 107(2):023504, jul 2015.

[48] John Paul Strachan, J. Joshua Yang, L. A. Montoro, C. A. Ospina, A. J. Ramirez, A. L D Kilcoyne, Gilberto Medeiros-Ribeiro, and R. Stanley Williams. Characterization of electroforming-free titanium dioxide memristors. *Beilstein Journal of Nanotechnology*, 4(1):467–473, 2013.

[49] Antonio C Torrezan, John Paul Strachan, Gilberto Medeiros-Ribeiro, and R Stanley Williams. Sub-nanosecond switching of a tantalum oxide memristor. *Nanotechnology*, 22(48):485203, 2011.

[50] Pablo F Siles, Muriel de Pauli, Carlos Cesar Bof Bufon, Sukarno O Ferreira, Jefferson Bettini, Oliver G Schmidt, and Angelo Malachias. Tuning resistive switching on single-pulse doped multilayer memristors. *Nanotechnology*, 24:035702, 2013.

[51] Woojin Jeon, Sang Ho Rha, Woongkyu Lee, Yeon Woo Yoo, Cheol Hyun An, Kwang Hwan Jung, Seong Keun Kim, and Cheol Seong Hwang. Controlling the Al-doping profile and accompanying electrical properties of rutile-phased TiO2 thin films. *ACS Applied Materials and Interfaces*, 6(10):7910–7917, 2014.

[52] Jeong Eun Yoo, Kiyoung Lee, Alexei Tighineanu, and Patrik Schmuki. Highly ordered TiO2 nanotube-stumps with memristive response. *Electrochemistry Communications*, 34:177–180, 2013.

[53] Kazuki Nagashima, Takeshi Yanagida, Masaki Kanai, Keisuke Oka, Annop Klamchuen, and Sakon Rahong. Switching Properties of Titanium Dioxide Nanowire Memristor. *Japanese Journal of Applied Physics*, 51:9–12, 2012.

[54] Batu Ghosh and Amlan J. Pal. Conductance switching in TiO2 nanorods is a redox-driven process: Evidence from photovoltaic parameters. *Journal of Physical Chemistry C*, 113(42):18391–18395, 2009.

[55] Gregory S. Snider and R. Stanley Williams. Nano/CMOS architectures using a field-programmable nanowire interconnect. *Nanotechnology*, 18(3):035204:1–11, 2007.

[56] P. Roy, S. Berger, and P. Schmuki. TiO2 nanotubes: Synthesis and applications. *Angewandte Chemie - International Edition*, 50(13):2904–2939, 2011.

[57] J. M. Macak, H. Hildebrand, U. Marten-Jahns, and P. Schmuki. Mechanistic aspects and growth of large diameter self-organized TiO2 nanotubes. *Journal of Electroanalytical Chemistry*, 621(2):254–266, 2008.

[58] Simelys Herná, Diana Hidalgo, Adriano Sacco, Angelica Chiodoni, Andrea Lamberti, Valentina Cauda, Elena Tresso, and Guido Saracco. Comparison of photocatalytic and transport properties of TiO2 and ZnO nanostructures for solar-driven water splitting. *Phys. Chem. Chem. Phys.*, 17(ii):7775–7786, 2015.

[59] Gregorio F. Ortiz, Ilie Hanzu, Thierry Djenizian, Pedro Lavela, Jos L. Tirado, and Philippe Knauth. Alternative Li-ion battery electrode based on self-organized titania nanotubes. *Chemistry of Materials*, 21(1):63–67, 2009.

[60] Andrea Lamberti, Nadia Garino, Adriano Sacco, Stefano Bianco, Angelica Chiodoni, and Claudio Gerbaldi. As-grown vertically aligned amorphous TiO2 nanotube arrays as high-rate Li-based micro-battery anodes with improved long-term performance. *Electrochimica Acta*, 151:222–229, 2015.

[61] X Lu, G Wang, T Zhai, M Yu, J Gan, Y Tong, and Y Li. Hydrogenated TiO2 nanotube arrays for supercapacitors. *Nano Lett*, 12(3):1690–1696, 2012.

[62] Sorachon Yoriya, Haripriya E. Prakasam, Oomman K. Varghese, Karthik Shankar, Maggie Paulose, Gopal K. Mor, Thomas J. Latempa, and Craig A. Grimes. Initial Studies on the Hydrogen Gas Sensing Properties of Highly-Ordered High Aspect Ratio TiO2 Nanotube-Arrays 20 μm to 222 μm in Length. *Sensor Letters*, 4(3):334–339, 2006.

[63] Andrea Lamberti, Alessandro Virga, Alessandro Chiadò, Angelica Chiodoni, Katarzyna Bejtka, Paola Rivolo, and Fabrizio Giorgis. Ultrasensitive Ag-coated TiO2 nanotube arrays for flexible SERS-based optofluidic devices. *Journal of Materials Chemistry C*, 3(26):6868–6875, 2015.

[64] Ning Liu, Kiyoung Lee, and Patrik Schmuki. Reliable metal deposition into TiO2 nanotubes for leakage-free interdigitated electrode structures and use as a memristive electrode. *Angewandte Chemie - International Edition*, 52(47):12381–12384, 2013.

[65] Renate Forch, Zhihong Zhang, and Wolfgang Knoll. Soft plasma treated surfaces: Tailoring of structure and properties for biomaterial applications. *Plasma Processes and Polymers*, 2(5):351–372, 2005.

[66] P. Rivolo, M. Castellino, F. Frascella, and S Ricciardi. Ultra-Thin Plasma-Polymerized Functional Coatings for Biosensing: Polyacrylic Acid, Polystyrene and Their Co-Polymer. In Pietro Mandracci, editor, *Crystalline and Non-crystalline Solids*, chapter Chapter 5., page 182. InTech, 2016.

[67] Maria Buchholz, Mingchun Xu, Heshmat Noei, Peter Weidler, Alexei Nefedov, Karin Fink, Yuemin Wang, and Christof Wöll. Interaction of carboxylic acids with rutile TiO2(110): IR-investigations of terephthalic and benzoic acid adsorbed on a single crystal substrate. *Surface Science*, 643:117–123, 2016.

[68] Yan-Yan Yu and Xue-Qing Gong. Unique adsorption behaviors of carboxylic acids at rutile TiO2(110). *Susc*, 641(C):82–90, 2015.

[69] Xue Qing Gong, Annabella Selloni, and Andrea Vittadini. Density functional theory study of formic acid adsorption on anatase TiO2(001): Geometries, energetics, and effects of coverage, hydration, and reconstruction. *Journal of Physical Chemistry B*, 110(6):2804–2811, 2006.

[70] Markku Leskelä and Mikko Ritala. Atomic layer deposition (ALD): From precursors to thin film structures. *Thin Solid Films*, 409(1):138–146, 2002.

[71] Ville Miikkulainen, Markku Leskelä, Mikko Ritala, and Riikka L. Puurunen. Crystallinity of inorganic films grown by atomic layer deposition: Overview and general trends. *Journal of Applied Physics*, 113(2), 2013.

[72] Väino Sammelselg, Ivan Netšipailo, Aleks Aidla, Aivar Tarre, Lauri Aarik, Jelena Asari, Peeter Ritslaid, and Jaan Aarik. Chemical resistance of thin film materials based on metal oxides grown by atomic layer deposition. *Thin Solid Films*, 542:219–224, 2013.

[73] I.B. Kobiakov. Elastic, Piezoelectric and Dielectric Properties of ZnO and CdS Single Crystals in a Wide Range of Temperatures. *Solid State Communications*, 35:305–310, 1980.

[74] Z Wang and J Song. Piezoelectric nanogenerators based on zinc oxide nanowire arrays. *Science*, 312(5771):242, 2006.

[75] Stefano Stassi, Valentina Cauda, Carminna Ottone, Angelica Chiodoni, Candido Fabrizio Pirri, and Giancarlo Canavese. Flexible piezoelectric energy nanogenerator based on ZnO nanotubes hosted in a polycarbonate membrane. *Nano Energy*, 13:474–481, 2015.

[76] X. W. Sun and H. S. Kwok. Optical properties of epitaxially grown zinc oxide films on sapphire by pulsed laser deposition. *Journal of Applied Physics*, 86(1):408–411, 1999.

[77] Katsutoshi. Takeda and Masao. Isomura. United States Patent, Method for forming ZnO Film, Method for forming ZnO Semiconductor Layer, Method for fabricating Seminconductor Device, and Semiconductor Device, 2006.

[78] S Ong, Chensha Li, Yuning Li, Yiliang Wu, and Rafik Loutfy. Stable , Solution-Processed , High-Mobility ZnO Thin-Film Transistors. 129(10):2750–2751, 2007.

[79] Valentina Cauda, Rossana Gazia, Samuele Porro, Stefano Stassi, Giancarlo Canavese, Ignazio Roppolo, Alessandro Chiolerio, B Bhushan, D Luo, S Schricker, W Sigmund, and S Zauscher. Nanostructured ZnO Materials: Synthesis, Properties and Applications. In *Handbook of Nanomaterials Properties*, volume XXVII, pages 137–177. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[80] B. J. Jin, S. Im, and S. Y. Lee. Violet and UV luminescence emitted from ZnO thin films grown on sapphire by pulsed laser deposition. *Thin Solid Films*, 366(1-2):107–110, 2000.

[81] P. F. Carcia, R. S. McLean, and M. H. Reilly. High-performance ZnO thin-film transistors on gate dielectrics grown by atomic layer deposition. *Applied Physics Letters*, 88(12):2004–2007, 2006.

[82] L. V. Podrezova, S. Porro, V. Cauda, M. Fontana, and G. Cicero. Comparison between ZnO nanowires grown by chemical vapor deposition and hydrothermal synthesis. *Applied Physics A*, 113(3):623–632, 2013.

[83] Seunghyup Lee, Heejin Kim, Jinjoo Park, and Kijung Yong. Coexistence of unipolar and bipolar resistive switching characteristics in ZnO thin films. *Journal of Applied Physics*, 108(7):076101, oct 2010.

[84] Wen-Yuan Chang, Yen-Chao Lai, Tai-Bor Wu, Sea-Fue Wang, Frederick Chen, and Ming-Jinn Tsai. Unipolar resistive switching characteristics of ZnO thin films for nonvolatile memory applications. *Applied Physics Letters*, 92(2):022110, 2008.

[85] E.M Bachari, G Baud, S Ben Amor, and M Jacquet. Structural and optical properties of sputtered ZnO films. *Thin Solid Films*, 348(1):165–172, 1999.

[86] Yuchao Yang, Xiaoxian Zhang, Min Gao, Fei Zeng, Weiya Zhou, Sishen Xie, and Feng Pan. Nonvolatile resistive switching in single crystalline ZnO nanowires. *Nanoscale*, 3(4):1917, 2011.

[87] 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):668–673, 2014.

[88] Steve B. Furber, David R. Lester, Luis A. Plana, Jim D. Garside, Eustace Painkras, Steve Temple, and Andrew D. Brown. Overview of the SpiNNaker system architecture. *IEEE Transactions on Computers*, 62(12):2454–2467, 2013.

[89] Giacomo Indiveri, Elisabetta Chicca, and Rodney Douglas. A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity. *IEEE Transactions on Neural Networks*, 17(1):211–221, 2006.

[90] Damien Querlioz, Olivier Bichler, and Christian Gamrat. Simulation of a memristor-based spiking neural network immune to device variations. *Proceedings of the International Joint Conference on Neural Networks*, pages 1775–1781, 2011.

[91] Damien Querlioz, Olivier Bichler, Philippe Dollfus, and Christian Gamrat. Immunity to device variations in a spiking neural network with memristive nanodevices. *IEEE Transactions on Nanotechnology*, 12(3):288–295, 2013.

[92] Hrika Manem, Jeyavijayan Rajendran, and Garrett S. Rose. Technique for a CMOS / Nano Memristive Trainable. *IEEE Transactions on Circuits and Systems*, 59(5):1051–1060, 2012.

[93] Joseph M. Brader, Walter Senn, and Stefano Fusi. Learning Real-World Stimuli in a Neural Network with Spike-Driven Synaptic Dynamics. *Neural Computation*, 19(11):2881–2912, 2007.

[94] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[95] S. Brivio, E. Covi, A. Serb, T. Prodromakis, M. Fanciulli, and S. Spiga. Gradual set dynamics in HfO 2 -based memristor driven by sub-threshold voltage pulses. In *2015 International Conference on Memristive Systems (MEMRISYS)*, pages 1–2. IEEE, nov 2015.

[96] Jacopo Frascaroli, Flavio Giovanni Volpe, Stefano Brivio, and Sabina Spiga. Effect of Al doping on the retention behavior of HfO2 resistive switching memories. *Microelectronic Engineering*, 147(April):104–107, nov 2015.

[97] Jacopo Frascaroli, Stefano Brivio, Federico Ferrarese Lupi, Gabriele Seguini, Luca Boarino, Michele Perego, and Sabina Spiga. Resistive Switching in High-Density Nanodevices Fabricated by Block Copolymer Self-Assembly. *ACS Nano*, 9(3):2518–2529, mar 2015.

[98] Zongwei Wang, Minghui Yin, Teng Zhang, Yimao Cai, Yangyuan Wang, Yuchao Yang, and Ru Huang. Engineering incremental resistive switching in TaOx based memristors for brain-inspired computing. *Nanoscale*, 8(29):14015–14022, 2016.

[99] Shimeng Yu, Pai-Yu Chen, Yu Cao, Lixue Xia, Yu Wang, and Huaqiang Wu. Scaling-up resistive synaptic arrays for neuro-inspired architecture: Challenges and prospect. In *2015 IEEE International Electron Devices Meeting (IEDM)*, volume 2016-Febru, pages 17.3.1–17.3.4. IEEE, dec 2015.

[100] G Parisi. Asymetric neural networks and the process of learning. *Journal of Physics A: Mathematical and General*, 19:L675–L680, 1986.

[101] Ning Qiao, Hesham Mostafa, Federico Corradi, Marc Osswald, Fabio Stefanini, Dora Sumislawska, and Giacomo Indiveri. A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128K synapses. *Frontiers in Neuroscience*, 9(APR):1–17, apr 2015.

[102] Hesham Mostafa, Christian Mayr, and Giacomo Indiveri. Beyond spike-timing dependent plasticity in memristor crossbar arrays. In *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, volume 2016-July, pages 926–929. IEEE, may 2016.

[103] Stefano Ambrogio, Nicola Ciocchini, Mario Laudato, Valerio Milo, Agostino Pirovano, Paolo Fantini, and Daniele Ielmini. Unsupervised learning by spike timing dependent plasticity in phase change memory (PCM) synapses. *Frontiers in Neuroscience*, 10(MAR):1–12, 2016.

[104] Erika Covi, Stefano Brivio, Alexander Serb, Themis Prodromakis, Marco Fanciulli, and Sabina Spiga. Analog memristive synapse in spiking networks implementing unsupervised learning. *Frontiers in Neuroscience*, 10(OCT):1–13, 2016.

[105] M. Prezioso, F. Merrikh Bayat, B. Hoskins, K. Likharev, and D. Strukov. Self-Adaptive Spike-Time-Dependent Plasticity of Metal-Oxide Memristors. *Scientific Reports*, 6(1):21331, 2016.

[106] Daniele Garbin, Elisa Vianello, Olivier Bichler, Quentin Rafhay, Christian Gamrat, Gérard Ghibaudo, Barbara Desalvo, and Luca Perniola. HfO 2 - Based OxRAM Devices as Synapses for Convolutional Neural Networks. 62(8):2494–2501, 2015.

[107] Adrien F. Vincent, Jérôme Larroque, Nicolas Locatelli, Nesrine Ben Romdhane, Olivier Bichler, Christian Gamrat, Wei Sheng Zhao, Jacques Olivier Klein, Sylvie Galdin-Retailleau, and Damien Querlioz. Spin-transfer torque magnetic memory as a stochastic memristive synapse for neuromorphic systems. *IEEE Transactions on Biomedical Circuits and Systems*, 9(2):166–174, 2015.

Appendix A

Fast Dynamics Resistive Switching

Pt/HfO_2/TiN resistive switching are adopted as the weighting elements for synapses in the neuromorphic architecture. The conductance can be easily exploited as the measure of the synaptic weight. Depending on the characteristics of applied voltage pulses used for the electrical characterization of the device, different conductance dynamics can be reached and then exploited in learning simulations.
In this appendix, it is reported the characterization of resistive switching devices with pulses as reported in table A.1.

Pulse Parameter	$V_{LTP}(V)$	$V_{LTD}(V)$	$T(\mu s)$
PPS1	1.1	-0.8	30

Table A.1 Pulse Parameter sets adopted for electrical characterization of RS devices.

The switching characteristics of HfO_2-based devices are summarized in figure A.1 and a representative characterization of device endurance is reported in figure A.2, where it emerges the repeatability of resistive switching characteristics for a number of pulses in the order of 10^4.

LTD process brings the device from LRS gradually to the HRS as shown in figure 5.2a) and b), from 1 to 50 and from 1 to 300 pulses, respectively. LTP brings the device from the HRS gradually to the LRS as shown in figure 5.2c) and d) from 1 to 50 and from 1 to 300 pulses, respectively. It can be noticed that LTP is more abrupt at low number of pulses with respect to LTD because the former reaches almost the saturation within 50 pulses (figure 5.2b) while LTD slowly proceeds beyond 50

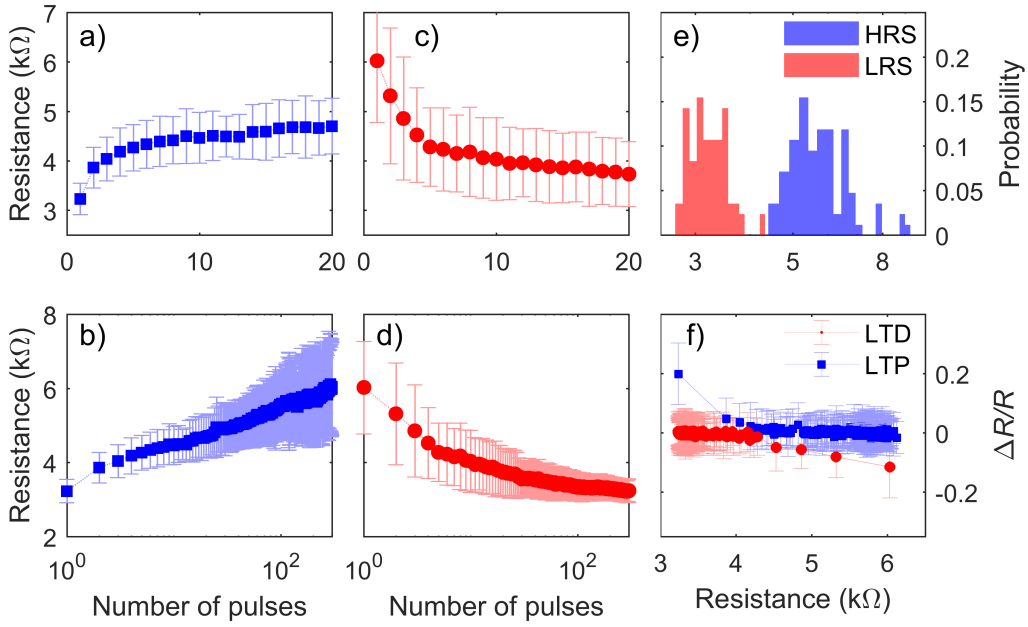


Fig. A.1 Representative resistance evolution as a function of the number of pulses from 1 to 50 pulses and from 1 to 300 pulses starting from LRS and HRS for LTD (blue, a and b) and LTD (red, c and d), respectively. Panel e), distribution of resistance values for HRS and LRS obtained after 300 LTD and LTP spikes. Panel f), dependence of the 1-spike normalized resistance change as a function of the pre-spike resistance value. Symbols reports the average among tens of experimental curves acquired on a single device and error bars corresponds to one sigma standard deviation.

pulses (figure 5.2a). The evolution of the resistance in LTD and LTP reported in figure 5.2a-d) evidences the major non-idealities of memristors for synaptic application. The resistance change is non-linear as a function of number of pulses and it is faster for low pulse numbers (i.e. at low resistance for LTD and high resistances for LTP, respectively). The second aspect is that the LTP is observed to have a faster dynamics with respect to LTD, i.e. it presents an increased non-linearity than LTD.
With set of pulses parameters PPS1, transitions toward LRS and HRS happen faster than what observed in chapter 5, in particular for what concerns the first set of applied pulses, cfr. figure 5.2 and A.1. Saturation of the potentiation and depression processes lead to bimodal resistance distribution of LRS and HRS as shown in figure 5.2e.

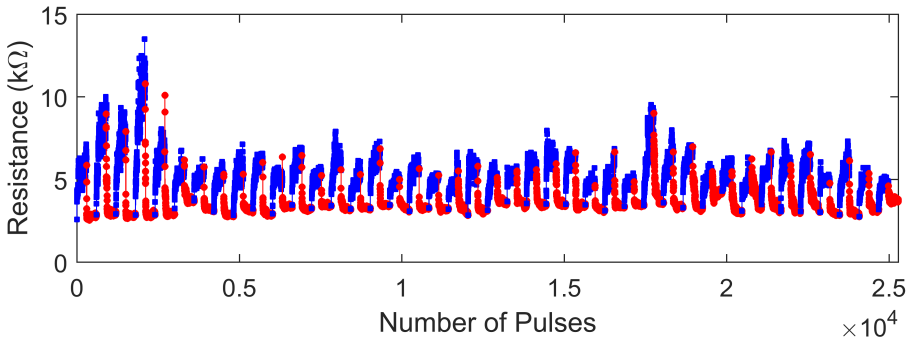


Fig. A.2 Representative endurance test over LTD (blue squares) and LTP (red circles)

A.1 Modeling

The gradual resistance change featured by the device in response to sequences of identical spikes is found to be thoroughly described by two separate equations for the resistance R for LTD and LTP processes starting form LRS and HRS respectively.

$$\begin{aligned} LTD &: R = R_{0,D} + \alpha_D \cdot ln(n) \\ LTP &: R = R_{0,P} - \alpha_P \cdot (1 - \frac{1}{\beta_P}) \end{aligned}, \tag{A.1}$$

where n is the number of pulses and parameters $\alpha_{D,P}$ and β_P are evaluated by fitting experimental data with equations A.1. Representative fitting of experimental data are reported in figure A.3a) and b) for LTD and LTP, respectively.

Variability is evaluated as the standard deviation value of the resistance change per unit pulse $\Delta R(n) = R(n) - R(n-1)$ taken over tens of cycles for both LTD and LTP and is reported in figure A.3c) and d). The experimental standard deviations $\sigma_{\Delta R}$ are almost constant for LTD and is decreasing as a function of the number of pulses (i.e. with decreasing resistances) for LTP. Therefore, beyond the asymmetry in the average resistance evolution between LTD and LTP processes, an asymmetry exists also in their variability.

$$\begin{aligned} LTD &: \sigma_{\Delta R} = \delta_D \\ LTP &: \sigma_{\Delta R} = \frac{\gamma_P}{n} + \delta_P \end{aligned}. \tag{A.2}$$

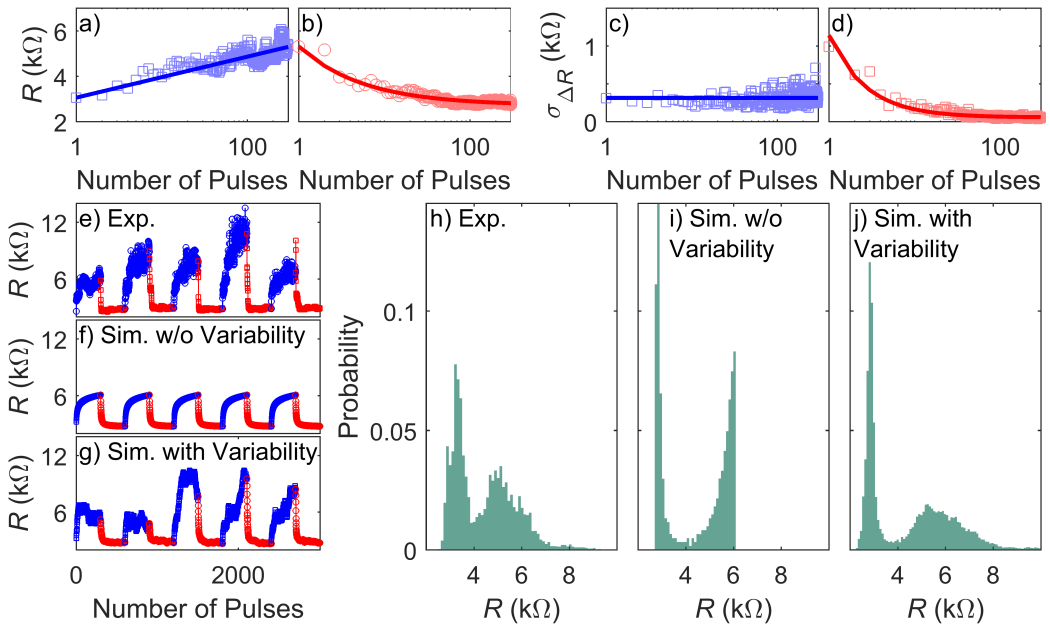


Fig. A.3 Fitting and simulation of the HfO_2 memristors. Panel a) and b), representative LTD and LTP processes respectively (symbols, the experimental results and lines, fitting according to equation A.1. Panels c) and d), standard deviation of resistance variation per pulse ΔR for LTD and LTP respectively. Representative switching cycles: experimental, panel e), simulated without variability, panel f), and simulated with pulse to pulse variability, panel g). Resistance values histograms obtained with tens of cycles: experimental, panel h), simulated without variability, panel i) and simulated with pulse to pulse variability, panel j).

The fitting parameters of equations A.1-A.2 are used to simulate the behavior of the devices according to the general laws that link the resistance variation per pulse to the actual device resistance, reported below:

$$\begin{aligned} LTD: \Delta R &= \alpha_D \cdot exp\left(-\frac{R-R_{0,D}}{\alpha_D}\right) + \Delta R_{var.} \\ LTP: \Delta R &= -\alpha_P \cdot \beta_P \cdot (\alpha + R_{0,P} - R)^{-\frac{1+\beta_P}{\beta_P}} + \Delta R_{var.} \end{aligned} \quad , \tag{A.3}$$

where ΔR_{var} is the pulse to pulse variability contribution simulated as a Gaussian noise with standard deviation according to equations A.2. Representative simulations without variability and with pulse to pulse variability are reported in figure A.3f) and g) and compared to the experimental results in figure A.3e). The comparison evidences that the average behavior without variability describes a very ideal case, while simulation with variability faithfully reproduce that device behavior. A statistical assessment of the same consideration can be obtained by comparing the

histograms of the resistance values reached among tens of LTD and LTP cycles, as reported in figure A.3h) and j) for experimental data and simulated without and with variability. Indeed, the bimodal distribution of experimental results is reproduced in the simulation. However, average simulations performed with LTD and LTP processes stopped at 300 pulses display an unrealistic cut off of the distribution at low and high resistances. Such an issue is solved in simulations taking into account device variability.

Parameter	$R_{0,D}$	α_D	$R_{0,P}$	α_P	β_P	δ_D	γ_P	δ_P
average	3234	502	6032	3350	0.612	139	286	17

Table A.2 Averages and standard deviations values of the fitting parameters for the general resistance evolution and for the modeling of the device variability for both LTD and LTP.

Appendix B

Parameters Investigation

Parameters adopted for the network simulations deeply affect the network classification performances. This is particularly evident in case of variations on learning thresholds applied to output neurons variables: membrane current and Calcium current. Membrane current embeds information related to events time-causality. Calcium current embeds informations about neurons average activity. Different choices of thresholds can be particularly restrictive on transition events that pre-synapses experience. Nonetheless, the whole set of parameters has relevance on the network dynamics and classification performances.
As already described in this dissertation, STDP learning rule lacks a rigorous method for cost function optimization that includes all the network parameters. For such a reason, only an empirical testing of parameters sets can give insights about the network functioning.
In this appendix are reported tables of subsets of parameters sets tested. Resistive switching embedded in the network tested here are characterized with pulse parameter set PPS1. Some tests include also slight variation of the network learning dynamics that have been tested, without bringing tangible improvement to the classification performances. All of them will be briefly explained in the following subsections.

B.1 Unbounded Calcium Current

Time correlation is encoded mainly in membrane current of the output neuron, since the post-synaptic spike is triggered by a sufficient income of pre-synaptic spikes. On the other hand, Calcium current encodes average activity of the output neuron, low-pass filtering the firing rate of the neurons itself, i.e. its membrane current dynamics.
In figure B.1 are reported parameters for a set of simulations in which no bounds (or very soft bound) are imposed to Calcium current, in order to investigate to which extent only membrane current can be exploited for successful STDP to take place.
From such simulations emerge that depression is overcoming completely potentiation, since Fusi's STDP does not focus on correlation between two spike time instants, but rather on a time window. Reducing the threshold on the membrane current enlarges the time window loosing correlation. On the contrary, increasing the threshold captures better correlation but sensibly increases probability of depression events.

B.2 Bounded Calcium Current

Time correlation is encoded mainly in membrane current of the output neuron, since the post-synaptic spike is triggered by a sufficient income of pre-synaptic spikes. On the other hand, Calcium current encodes average activity of the output neuron, low-pass filtering the firing rate of the neurons itself, i.e. its membrane current dynamics.
In figure B.2 are reported parameters for a set of simulations in which bounds are imposed to Calcium current, in order to limitate the probability of LTD and LTP events.
First trials report thresholds on neurons variables that are inspired by the work of Brader et al, that in the following trials are adapted to the present simulations. Despite the multiple combinations tried, only a subset of which is reported here, no condition on variable thresholds allows to limit the impact of multiplicative learning rule imposed by resistive switching on weight update. More clearly, no combination of thresholds is able to discriminate and sufficiently separate LTD and LTP events, so that discriminative features of input digit are acquired.

B.3 Dynamic Calcium Current

Starting from parameters identified in the previous set of simulations, it is imposed a priori a linear dynamic on Calcium current bands, see figure B.3. The ratio behind this subset of simulations emerges from observation of the early epochs of network functioning. Indeed, features extraction of input digits already happens in the first 1000 epochs of learning, but then they roughly disappear because of the general drift of synaptic weights toward their mean value.
This subset of simulation then imposes a linear drift on Calcium current lower thresholds that tend to make narrower Ca bands both for LTD and LTP events, resulting in a much lower probability of transitions after the first epochs of learning.
In figure B.3, four more lines reports angular coefficient and offset used for the threshold dynamics, following the simple equation: $\Theta_{Dl,Pl} = m_{LTD,P} * epoch + q_{LTD,P}$. Update happens in a number of limited epochs at the beginning of the simulations, that in table are referred as *Update window*.

B.4 Homeostatic Principle

Homeostatic principle is an expedient that is usually applied to biologically inspired neural network. It is usually embedded to reduce the over- and under-activity of network elements, avoiding saturations. Homeostatic principle comes from biological observation of specific operative range for cells in brain.
Within the present simulations, see figure B.4, homeostatic principle is imposed on teacher signal. It modulates the contribution coming from teacher neurons in such a way that the output neurons firing rate remains within specified targets. Moreover, this expedient is intended to allow a positive feedback on features extraction, since a reduction of teacher signal is expected to let emerge a higher contribution of potentiated (depressed) synaptic weights in enabling LTP (LTD). Unfortunately, no relevant contribution of this kind is observed during simulations. Dynamics on the teacher signal can happen in a number of limited epochs at the beginning of the simulations, that in table are referred as *Homeo window*. After such a number of epochs, if specified, the teacher contribution is imposed to a fixed value, reported in figure as *asymptotic* $I_{w,t}$.

B.5 Double Calcium Bands

The only expedient allowing a tangible improvement on learning performance, due to the enabling of discriminative features acquisition of input pattern, comes from the introduction of a second Calcium band for Long Term Depression, see figures B.5 and B.6. Indeed, in order to boost suppression of synaptic weights referring to input pattern background, a second LTD Ca band at low values exploits the weak activity of not excited input neurons. It sensibly increases the probability of LTD events for those synapses that are not primarily excited by the input digit.
Simulations reported in chapter 5 results from the parameter optimization process reported here.

Trial Name	**BRMem06**	**BRMem07**	**BRMem08**	**BRMem09**	**BRMem10**	**BRMem11**
$I_{spkthr,ex}$	400pA	400pA	400pA	600pA	600pA	600pA
$I_{spkthr,inh}$	600pA	600pA	600pA	600pA	600pA	600pA
$I_{mem,thr}$	350pA	350pA	350pA	580pA	550pA	550pA
$\Theta_{D,l}$	0pA	0pA	0pA	0pA	0pA	1pA
$\Theta_{D,u}$	200pA	200pA	200pA	200pA	200pA	200pA
$\Theta_{P,l}$	100pA	0pA	0pA	0pA	0pA	1pA
$\Theta_{P,u}$	200pA	200pA	200pA	200pA	200pA	200pA
$I_{w,ex}$	7pA	7pA	7pA	7pA	7pA	7pA
$I_{w,inh}$	5pA	5pA	5pA	5pA	5pA	5pA
$I_{w,inhout}$	5pA	5pA	5pA	5pA	5pA	5pA
$I_{w,t}$	1nA	1nA	1nA	0.5nA	1nA	0.5nA
$w_{in,out}$	0	0	0	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*
$w_{in,inh}$	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*
$w_{inh,out}$	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*
$w_{teach,out}$	*G(0.8,0.1)*	*G(0.8,0.1)*	*G(0.8,0.1)*	*G(0.8,0.1)*	*G(0.8,0.1)*	*G(0.8,0.1)*
F_{teach}	500Hz	500Hz	100Hz	500Hz	500Hz	500Hz
Refractory time	0.5ms	0.5ms	0.5ms	0.5ms	0.5ms	0.5ms
Training image time	350ms	20ms	100ms	250ms	250ms	250ms

Fig. B.1

Trial Name	**BRMem12**	**BRMem13**	**BRMem14**	**BRMem15**	**BRMem16**	**BRMem17**	**BRMem18**
$I_{spkthr,ex}$	600pA	600pA	600pA	600pA	600pA	600pA	600pA
$I_{spkthr,inh}$	600pA	600pA	600pA	600pA	600pA	600pA	600pA
$I_{mem,thr}$	550pA	550pA	550pA	550pA	550pA	580pA	550pA
$\Theta_{D,l}$	30pA	20pA	10pA	30pA	30pA	30pA	125pA
$\Theta_{D,u}$	40pA	40pA	40pA	60pA	130pA	40pA	130pA
$\Theta_{P,l}$	30pA	30pA	30pA	30pA	30pA	30pA	80pA
$\Theta_{P,u}$	130pA	130pA	130pA	130pA	130pA	130pA	130pA
$I_{w,ex}$	7pA	7pA	7pA	7pA	7pA	7pA	7pA
$I_{w,inh}$	5pA	5pA	5pA	5pA	5pA	5pA	5pA
$I_{w,inhout}$	5pA	5pA	5pA	5pA	5pA	5pA	5pA
$I_{w,t}$	0.5nA	0.5nA	0.5nA	0.5nA	0.5nA	0.5nA	0.5nA
$w_{in,out}$	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*
$w_{in,inh}$	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*
$w_{inh,out}$	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*
$w_{teach,out}$	*G(0.8,0.1)*	*G(0.8,0.1)*	*G(0.8,0.1)*	*G(0.8,0.1)*	*G(0.8,0.1)*	*G(0.8,0.1)*	*G(0.8,0.1)*
F_{teach}	500Hz	500Hz	500Hz	500Hz	500Hz	500Hz	500Hz
Refractory time	0.5ms	0.5ms	0.5ms	0.5ms	0.5ms	0.5ms	0.5ms
Training image time	250ms	250ms	250ms	250ms	250ms	250ms	250ms

Fig. B.2

Trial Name	TestIca01	TestIca02	TestIca03	TestIca04	TestIca05	TestIca06	TestIca07	TestIca08	TestIca08var
$I_{spkthr,ex}$	600pA	600pA	600pA	600pA	600pA	600pA	600pA	600pA	600pA
$I_{spkthr,inh}$	600pA	600pA	600pA	600pA	600pA	600pA	600pA	600pA	600pA
$I_{mem,thr}$	550pA	550pA	550pA	550pA	550pA	550pA	550pA	550pA	550pA
$\Theta_{D,l}$	127pA	127pA	127pA	127pA	127pA	127pA	127pA	135pA	135pA
$\Theta_{D,u}$	130pA	130pA	130pA	130pA	130pA	130pA	130pA	140pA	140pA
$\Theta_{P,l}$	120pA	120pA	120pA	120pA	120pA	120pA	120pA	120pA	120pA
$\Theta_{P,u}$	130pA	130pA	130pA	130pA	130pA	130pA	130pA	140pA	140pA
$I_{w,ex}$	7pA	7pA	7pA	7pA	7pA	7pA	7pA	7pA	7pA
$I_{w,inh}$	5pA	5pA	5pA	5pA	5pA	5pA	5pA	5pA	5pA
$I_{w,inhout}$	5pA	5pA	5pA	5pA	5pA	5pA	5pA	5pA	5pA
$I_{w,t}$	0.5nA	0.5nA	0.5nA	0.5nA	0.5nA	0.5nA	0.5nA	0.5nA	0.5nA
$w_{in,out}$	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*
$w_{in,inh}$	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*
$w_{inh,out}$	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.3,0.05)*	*G(1.3,0.05)*	*G(1.3,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*
$w_{teach,out}$	*G(0.8,0.1)*	*G(0.8,0.1)*	*G(0.8,0.1)*	*G(0.8,0.1)*	*G(0.8,0.1)*	*G(0.8,0.1)*	*G(0.8,0.1)*	*G(0.8,0.1)*	*G(0.8,0.1)*
F_{teach}	500Hz	500Hz	500Hz	500Hz	500Hz	500Hz	500Hz	500Hz	500Hz
Refractory time	0.5ms	0.5ms	0.5ms	0.5ms	0.5ms	0.5ms	0.5ms	0.5ms	0.5ms
Training image time	250ms	250ms	250ms	250ms	250ms	250ms	250ms	250ms	250ms
m_{LTD}	0.0125pA	0.0140pA	0.05pA	0.05pA	0.05pA	0.05pA	0.05pA	0.05pA	0.05pA
q_{LTD}	120pA	120pA	120pA	120pA	120pA	120pA	120pA	120pA	120pA
m_{LTP}	0.100pA	0.140pA	0.166pA	0.166pA	0.166pA	0.166pA	0.166pA	0.166pA	0.166pA
q_{LTP}	60pA	50pA	70pA	70pA	60pA	60pA	70pA	70pA	70pA
Update Window	600 img	500 img	300 img	300 img	300 img	300 img	300 img	300 img	300 img

Fig. B.3

Trial	**NH04**	**NH05**	**NH06**	**NH09**	**NH10**	**NH07_r02**	**NH12**
$I_{spkthr,ex}$	600pA	600pA	600pA	600pA	600pA	600pA	600pA
$I_{spkthr,inh}$	600pA	600pA	600pA	600pA	600pA	600pA	600pA
$I_{mem,thr}$	590pA	590pA	590pA	590pA	580pA	580pA	580pA
$\Theta_{D,l}$	125pA	125pA	125pA	125pA	125pA	125pA	125pA
$\Theta_{D,u}$	132pA	132pA	132pA	132pA	132pA	132pA	132pA
$\Theta_{P,l}$	111pA	111pA	111pA	111pA	111pA	111pA	111pA
$\Theta_{P,u}$	145pA	145pA	145pA	145pA	145pA	145pA	145pA
$I_{w,ex}$	7pA	7pA	7pA	7pA	7pA	7pA	7pA
$I_{w,inh}$	5pA	5pA	5pA	5pA	5pA	5pA	5pA
$I_{w,inhout}$	5pA	5pA	5pA	5pA	5pA	5pA	5pA
$I_{w,t}$	1nA	1nA	1.4nA	1nA	1nA	1nA	1nA
$w_{in,out}$	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*
$w_{in,inh}$	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*
$w_{inh,out}$	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*
$w_{teach,out}$	*G(1.3,0.2)*	*G(1.3,0.2)*	*G(1.3,0.2)*	*G(1.3,0.2)*	*G(1.3,0.2)*	*G(1.3,0.2)*	*G(1.3,0.2)*
F_{teach}	500Hz	500Hz	500Hz	500Hz	500Hz	500Hz	500Hz
Refractory time	0.5ms	0.5ms	0.5ms	0.5ms	0.5ms	0.5ms	0.5ms
Training image time	250ms	250ms	250ms	250ms	250ms	250ms	250ms
$F_{target,up}$	250Hz	275Hz	150Hz	120Hz	$I_{w,t}$ *const*	120Hz	120Hz
$F_{target,up}$	250Hz	225Hz	100Hz	80Hz	$I_{w,t}$ *const*	80Hz	80Hz
F_{th}	90Hz	90Hz	90Hz	80Hz	$I_{w,t}$ *const*	80Hz	80Hz
$\Delta I_{w,t}$	200pA	200pA	200pA	50pA	$I_{w,t}$ *const*	50pA	50pA
Homeo Window	-	-	-	-	1000img	300img	300img
asymptotic $I_{w,t}$	-	-	-	-	250pA	50pA	300pA

Fig. B.4

Trial Name	**NC06**	**NC07**	**NC08**	**NC09**	**NC10**	**NC11**	**NC13**
$I_{spkthr,ex}$	600pA	600pA	600pA	600pA	600pA	600pA	600pA
$I_{spkthr,inh}$	600pA	600pA	600pA	600pA	600pA	600pA	600pA
$I_{mem,thr}$	480pA	480pA	480pA	480pA	520pA	590pA	580pA
$\Theta_{D,l}$	120pA	120pA	111pA	125pA	125pA	125pA	125pA
$\Theta_{D,u}$	132pA	132pA	145pA	130pA	130pA	130pA	130pA
$\Theta_{D2,l}$	20pA	20pA	15pA	15pA	15pA	15pA	15pA
$\Theta_{D2,u}$	40pA	40pA	40pA	40pA	40pA	40pA	40pA
$\Theta_{P,l}$	111pA	111pA	120pA	120pA	120pA	120pA	120pA
$\Theta_{P,u}$	145pA	145pA	132pA	145pA	145pA	145pA	145pA
$I_{w,ex}$	7pA	7pA	7pA	7pA	7pA	7pA	7pA
$I_{w,inh}$	5pA	5pA	5pA	5pA	5pA	5pA	5pA
$I_{w,inhout}$	5pA	5pA	5pA	5pA	5pA	5pA	5pA
$I_{w,t}$	1nA	1nA	1nA	1nA	1nA	1nA	1nA
$w_{in,out}$	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*
$w_{in,inh}$	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*
$w_{inh,out}$	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*
$w_{teach,out}$	*G(1.3,0.2)*	*G(1.3,0.2)*	*G(1.3,0.2)*	*G(1.3,0.2)*	*G(1.3,0.2)*	*G(1.3,0.2)*	*G(1.3,0.2)*
F_{teach}	500Hz	500Hz	500Hz	500Hz	500Hz	500Hz	500Hz
Refractory time	0.5ms	0.5ms	0.5ms	0.5ms	0.5ms	0.5ms	0.5ms
Training image time	250ms	250ms	250ms	250ms	250ms	250ms	250ms

Fig. B.5

Trial Name	**tmp_NC13**	**tmp_NC14**	**tmp_NC14_r01**	**tmp_NC14_r02**	**NC15**	**NC16b**
$I_{spkthr,ex}$	600pA	600pA	550pA	550pA	600pA	550pA
$I_{spkthr,inh}$	600pA	600pA	550pA	550pA	600pA	550pA
$I_{mem,thr}$	580pA	580pA	500pA	500pA	500pA	500pA
$\Theta_{D,l}$	125pA	125pA	125pA	125pA	125pA	125pA
$\Theta_{D,u}$	130pA	130pA	130pA	130pA	130pA	130pA
$\Theta_{D2,l}$	15pA	15pA	15pA	15pA	15pA	15pA
$\Theta_{D2,u}$	30pA	30pA	30pA	30pA	30pA	30pA
$\Theta_{P,l}$	120pA	120pA	120pA	120pA	120pA	120pA
$\Theta_{P,u}$	145pA	145pA	145pA	145pA	145pA	145pA
$I_{w,ex}$	7pA	7pA	7pA	7pA	7pA	7pA
$I_{w,inh}$	5pA	5pA	5pA	5pA	5pA	5pA
$I_{w,inhout}$	5pA	5pA	5pA	5pA	5pA	5pA
$I_{w,t}$	1nA	1nA	1nA	1nA	1nA	1nA
$w_{in,out}$	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*	*G(0.2,0.05)*
$w_{in,inh}$	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*	*U([0,0.8])*
$w_{inh,out}$	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*	*G(1.1,0.05)*
$w_{teach,out}$	*G(1.3,0.2)*	*G(1.3,0.2)*	*G(1.3,0.2)*	*G(1.3,0.2)*	*G(1.3,0.2)*	*G(1.3,0.2)*
F_{teach}	500Hz	500Hz	500Hz	500Hz	500Hz	500Hz
Refractory time	0.5ms	0.5ms	0.5ms	0.5ms	0.5ms	0.5ms
Training image time	250ms	250ms	250ms	250ms	250ms	250ms

Fig. B.6