EMERGING RESISTIVE SWITCHING MEMORIES AND NEUROMORPHIC APPLICATIONS

SAPIENZA
UNIVERSITÀ DI ROMA

Paolo Lorenzi

Department of Information Engineering, Electronics and Telecommunications

Sapienza, University of Rome

Supervisor: Prof. Fernanda Irrera

Sempre e comunque ai miei genitori…

CONTENTS

Alla fine, in qualche modo si fa… [cit.]

1 Resistive Switching Memories

Resistive Switching (RS) Memories, or Resistive Random Access Memory (RRAM), are one of the most promising candidates for Non-Volatile Memory applications. In this chapter, after an introduction to the NAND Flash memory main characteristics and limits, Resistive Switching Memories are presented. Electrical behavior, materials and integration schemes will be described. An overview of the physical mechanisms governing the switching phenomena and the electrical conduction will be given. In the following paragraphs the state of the art of released prototypes and open issues will be addressed. A comparison between RRAM and memristors closes the chapter.

1.1 Non-Volatile Memories: NAND Flash

1.1.1 Introduction

Two key categories of electronic memory are volatile and nonvolatile. In general, volatile memories lose their bit state when the power is removed and nonvolatile memories retain this state for an extended period of time. This period is known as the retention time. Typical nonvolatile memories are expected to have a retention period of 10 years [1], whereas volatile memories like dynamic random access memory (DRAM) retain the bit state for less than 1 s [1]. Sequential and random access is another important distinction between modern electronic memories technologies. A random access memory (RAM) can access all bits with equal effort. The time needed to access a bit of data is known as latency. In general a RAM has a low bit access latency. There are two kinds of RAM, as shown in Figure 1.1: DRAM, where data is stored on capacitors and requires a periodic refreshment, and Static RAM (SRAM) where data is retained as long as there is power supply on. Some memory technologies like NAND flash are written in blocks, and thus are not considered a true RAM. However, because each block can be addressed individually, NAND flash is considered a hybrid.

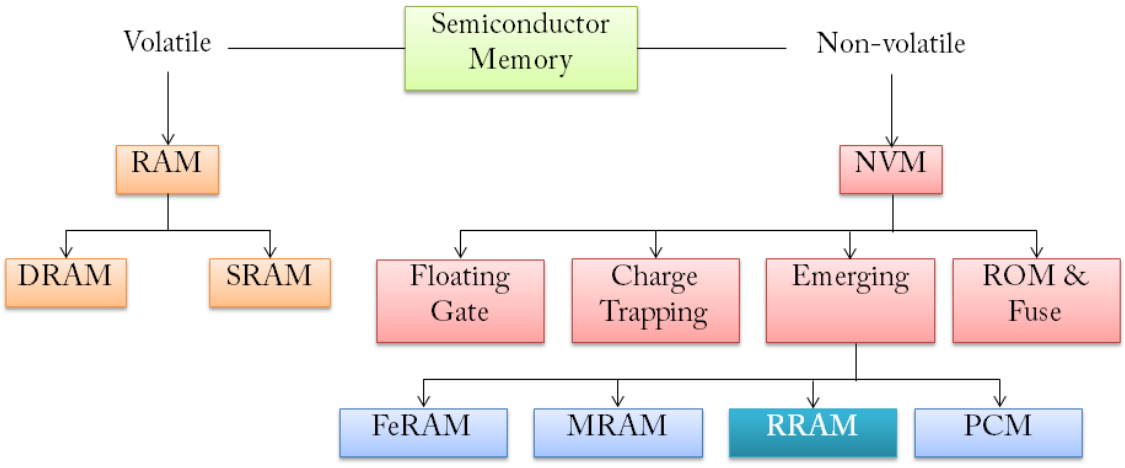


Figure 1.1 Semiconductor memories overview

The terms set, write, and program are often used to refer to the operation of switching a memory cell to the "1" state. This "1" state is often referred to as set, programmed, or on. Conversely, the terms reset and erase usually refer to the cell switching to a "0" state. Hence, the "0" state is referred to as reset, erased or off.

1.1.2 NAND Flash

With the recent strong increase in the demand for data storage, NAND Flash has solidified its leading position among different storage devices for the smallest chip size and cheapest bit cost [2]. The scaling in NAND Flash technology has progressed aggressively and successfully with trend shown in Figure 1.2.

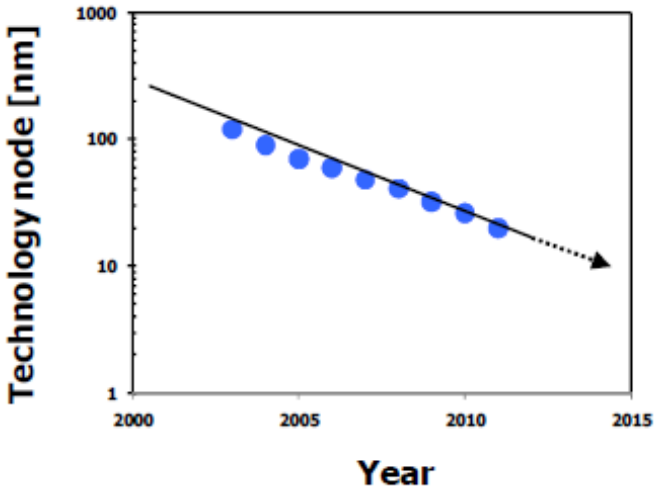


Figure 1.2 The scaling trend of floating gate NAND Flash cell over 10 years [2]

During the past years, the dimension of NAND Flash cell has shrunk dramatically. Due to this aggressive scaling, the chip production cost has become cheaper year-to-year by ~40% and consequently the market demand for NAND Flash memory has rapidly increased in many new consumer electronic applications such as MP3, SSD, USB pen drives, Tablets, Smart Phones and Memory Cards as depicted in Figure 1.3.

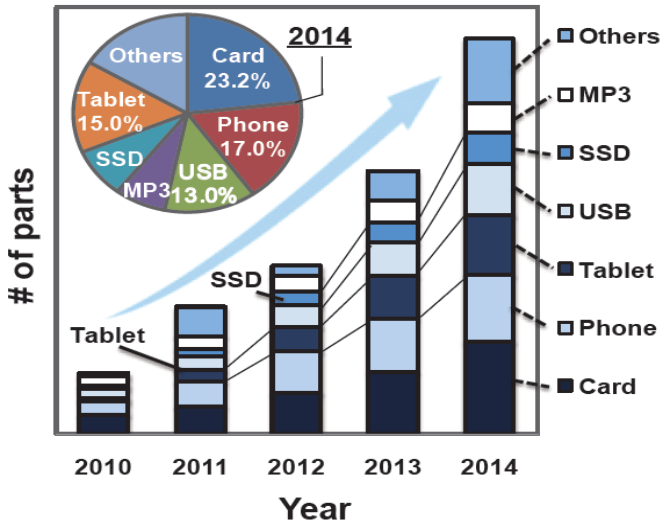


Figure 1.3 The growth of NAND Flash market [3]

The competition of scaling down technology is getting harder in the NAND industry. The current NAND Flash cell, the Floating Gate (FG) structure, is facing new technological challenges approaching the 10nm dimension. Major concerns are related to [4]:

- Physical dimension
- Electrical isolation
- Read window margin.

1.1.2.1 NAND Flash scaling issues

In Figure 1.4 a) is sketched the concept structure of the Floating Gate structure. Basically it is a transistor MOSFET with two gates: the accessible Control Gate (CG) and the charge storage gate called Floating Gate capacitively coupled with the CG. Since the FG is completely surrounded by insulating material, the charge stored in it is retained for long periods of time. Usually Fowler-Nordheim (FN) tunneling and hot-carrier injection mechanisms are used to modify the amount of charge stored in the FG.

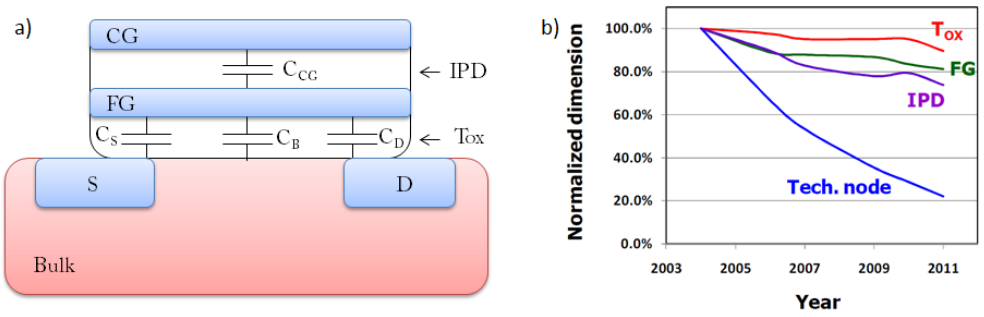


Figure 1.4 a) Floating Gate basic structure where CG is the control gate, FG is the floating gate, C_{CG}, C_S, C_D and C_B are the control gate, source, drain and bulk capacitances; in b) the scaling trend of the physical parameters

The operations of NAND flash memory are based on the storage of electric charge for writing/erasing and the sensing of the charge (current) for reading out the data. Thus, the intrinsic nature imposes limitations on the thickness of the dielectrics (the tunnel oxide and the inter-poly dielectric) which should be thicker than a minimum thickness required to reduce the charge loss as much as possible. This limitation is the origin of most of the scaling issues in NAND flash memory [5].

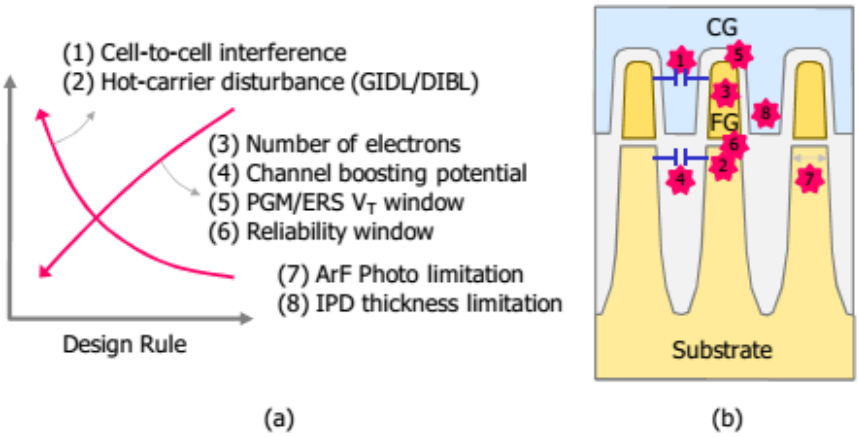


Figure 1.5 Scaling limitations and trends in NAND flash memory.

As shown in Figure 1.5, with the scaling of the device, most of the electrical properties have become worse. The cell-to-cell interference and the hot-carrier disturbance have

been increasing as the cells are scaling down continuously. The number of electrons in stored in the FG, the channel boosting potential, Program/Erase (PGR/ERS) threshold voltage (V_T) window, and reliability window are decreasing continuously. Finally, the thickness of IPD occupies a significant portion in the cell dimension because other dimensions have reached a comparable thickness with IPD.

The first critical issue of scaling down the NAND flash memory is the cell-to-cell interference. The cell-to-cell interference is a V_T shift of a monitored cell when V_T shift of a neighbor cell takes place due to the fringing electric field. The interference is inversely proportional with the cell dimension and it depends to the fixed thickness of the tunnel oxide and the IPD. Some solutions have been developed to reduce the interference by the change of the gate spacer from nitride to oxide and an introduction of air-gap both in the gate space and the active space [6].

Regarding hot-carrier disturbance by the boosted channel potential, as shown in Figure 1.6, most of the electrons are supplied by Drain Induced Barrier Lowering (DIBL) Gate Induced Drain Leakage (GIDL) at the Ground Select Line (GDL) transistor. Electrons are accelerated by the lateral electric field and injected into the GDL transistors [7]. The channel potential may reaches 5~10 V by a self-boosting of channel to suppress the disturbance in not selected lines. The channel boosting potential cannot be scaled down with the PGR/ERS voltages because it depends on the electric field. Thus, the hot-carrier disturbance becomes more serious as the thickness of the tunnel oxide and the IPD shrinks with the same operating voltage.

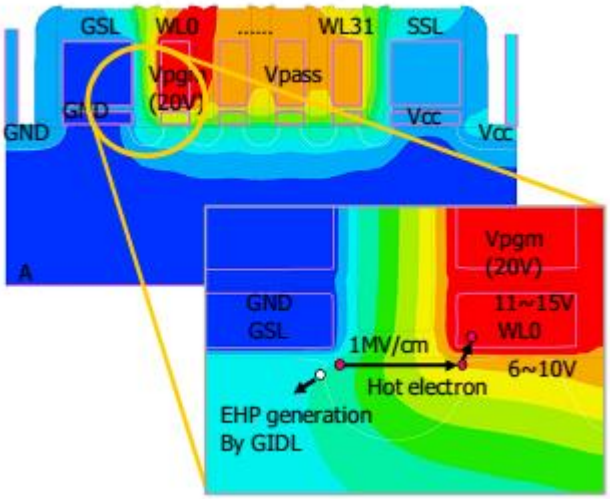


Figure 1.6 A hot-carrier disturbance mechanism. The potential diagram explains hot-electron injection generated by GIDL current at GSL transistor

The reduction in FG capacitance, proportional to the ratio between the FG area and the tunnel oxide thickness, caused by the technological shrink continuously reduces the number of electrons stored in the FG per unit voltage and the reliability strongly deteriorates as the number of electrons available for data storage becomes less than 10. Then, the amount of V_T shift by one electron increases and becomes a range of 0.1 V, hence it increases the dispersions of the cell V_T during operation [5].

The electrical coupling ratio, defined as the ratio between the control gate capacitance and the total gate stack capacitance, should be kept higher than 0.6 in order to achieve good control. To get this coupling value, scaling of tunnel oxide and IPD thickness is very critical and has to be managed very conservatively, as shown in Figure 1.4 b), for satisfying reliability constraints. Erase operation occurs either through electron de-trapping from the FG or hole injection from the substrate into the storage layer; at the same time, such operation causes an electron injection from the control gate to the storage layer through Fowler Nordheim (FN) tunneling, and this is the reason for the

erase saturation problem [5]. The InterPoly Dielectric (IPD), the dielectric between the control and the floating gate, becoming thinner with technology shrinkage will increase this effect. This is caused by the increased electric field in the IPD and it will deteriorate the saturation level of the threshold voltage Vt.

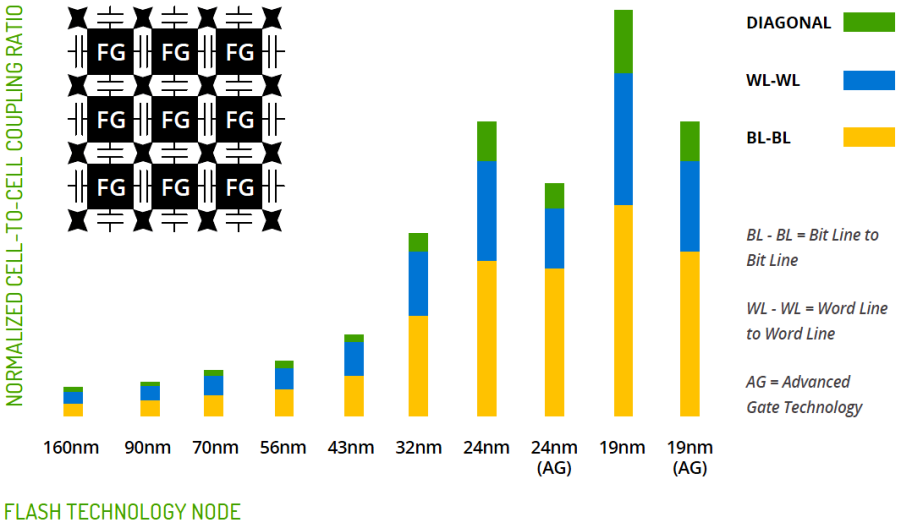


Figure 1.7 Cell to Cell coupling ratio trend versus technology node [8]. In the inset are sketched the parasitic capacitances surrounding the floating gates.

In the basic conventional concept, the FG voltage was determined by the control gate voltage through a coupling effect. As the design rule of NAND Flash memory is scaled down, however, parasitic capacitors surrounding the floating gate, as shown in the inset of Figure 1.7, should be involved to explain the secondary effects occurring in the cell operation. The FG voltage is determined by not only the corresponding control-gate voltage but also by the voltages of the surrounding floating gates and the control gates. Hence, a phenomenon called "floating-gate interference," occurs, in which change of a cell accompanies threshold voltage shift of the adjacent cells by floating-gate voltage shift [9]. In other words, the floating-gate voltage is coupled by the floating-gate

voltage changes of the adjacent cells via parasitic capacitors in the same manner as the control-gate voltage. Then a floating-gate interference coupling ratio can be defined. In Figure 1.7 is shown the influence of the parasitic effects due the increase of the coupling factor with the progress in the dimension limit for the technological processes. As the scaling of floating gate cell is proceeded, the Word Line (WL) to- WL and Bit Line (BL)-to-BL spaces decrease and the capacitive coupling between WL's and BL's become stronger.

The interface trap generation and the oxide bulk trap dominantly determine the reliability of NAND flash memory [10]. Due to trap assisted leakage, the thickness of the oxides cannot shrink too much. If the oxide is too thin, even the presence of few traps can be sufficient to discharge the FG in not acceptable time range. The trap generation mechanisms are much faster at the corner of the cell transistor, hence an increased trap generation at the corner of the active and the gate makes the reliability window decrease as the cell dimensions shrink [10, 11]. Furthermore, an increased lateral E-field between the WLs generates more traps in the cell [5].

Many researches are being conducted in extending the FG technology and resolve the mentioned issues. 3D cell structure has been studied as an alternative solution to overcome the scaling limitation and increase the bit density. The 3D NAND Flash memories can be classified largely into two groups according to the gate and channel directions. In a vertical channel (VC) scheme, WLs are stacked on Si substrate [12]. By digging holes through the stacked WLs, vertical tube-shaped strings are formed and poly-Si channels are plugged-in. On the other hand, a vertical gate (VG) scheme has stacked channels with subsequent deposition of the storage layer and gate materials

[13]. Through the advantages of the gate-all-around or dual-gate structures combined with the extended DR, cell current increases by ~ 2 times. In addition, subthreshold swing decreases by ~ 30 % despite the carrier mobility degradation by ~ 85 % due to the use of poly-Si channel, the high aspect ratio-induced process limits and complexities in decoding circuit.

In order to keep pursuing the scaling, there are two guide lines to follow: the optimization of the FG basic cell and array architecture and the proposal of new device structures with innovative operating concepts. Both the guide lines are currently experienced worldwide. So, in the last years, innovative concepts have been proposed, alternative to the conventional FG technology, as shown in Figure 1.1: silicon nano-crystals [14], SONOS [15-17] and TANOS [18-20], Magnetic RAM [21,22], Ferroelectric RAM [23] and Resistive Switching Memories, often referred as Resistive RAM (RRAM). In this thesis, we will focus on the Resistive Switching Memories.

1.2 Resistive Switching

The existence of the abrupt electrical switching event in oxides (insulator materials), that brings the material into a conductive state has been known for over 40 years. As shown in Figure 1.8, the first observations on these resistive switching phenomena were reported in the 1960s [24-27].

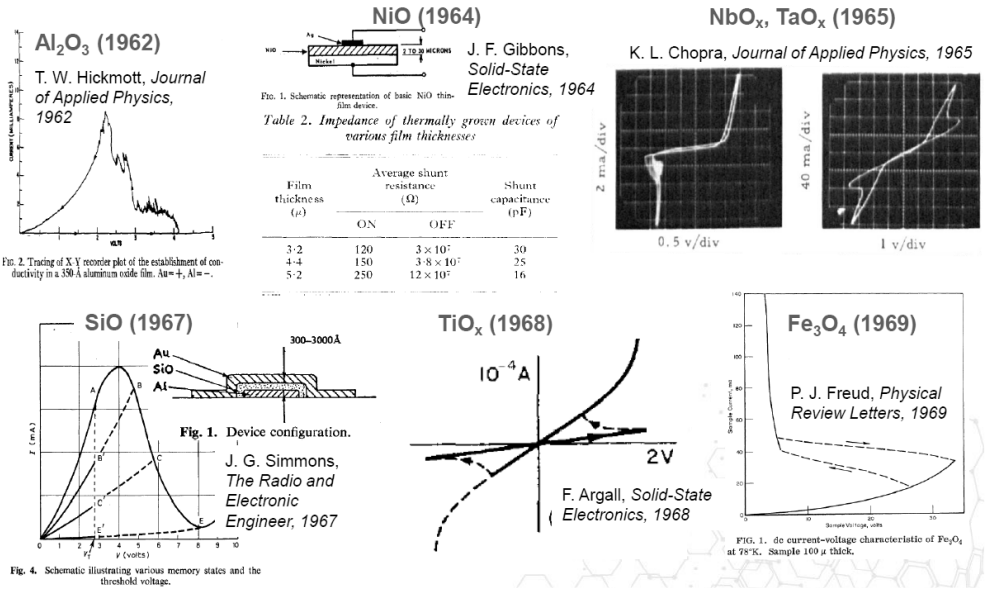


Figure 1.8 Some of memory/threshold switching in various oxide films already reported in 1960s

In its basic form, the device structure is an oxide material sandwiched between two metal electrodes, called bottom and top electrodes (BE and TE). The stack is a well-known metal–insulator–metal (MIM) structure. After the first works those reports remain in the domain of scientific studies. The recent revival of interest in resistive switching began in the late 1990s, first with complex metal oxides such as the pervoskite oxides of SrTiO3 [28], SrZrO3 [29], but later a strong interest has grown around binary metal oxides such as NiO [30] and TiO2 [31]. The research activities had a sensible increase after the International Electron Devices Meeting of 2004, where Samsung presented NiO-based memory cells integrated in a CMOS compatible process with a one-transistor–one-resistor (1T1R) device structure [32]. In that work a full electrical characterization data were presented including data retention, endurance, and programming/erasing characteristics suggesting that a memory technology based on resistive switching may be feasible. During these years, various acronyms such as

OxRAM, ReRAM, and RRAM, have been used in the literature for these devices that exhibit RS between a high-resistance state (HRS) and a low-resistance state (LRS). The expectation for RRAMs is that it will be a memory technology that can be easily integrated with conventional CMOS technology, using a materials compatible with the conventional CMOS fabrication environment and process temperatures that allow it to be fabricated in the Back End Of Line (BEOL), so on the metal layers or within the contact vias to the source and drain of a metal– oxide–semiconductor field-effect transistor (MOSFET) of a CMOS chip [33].

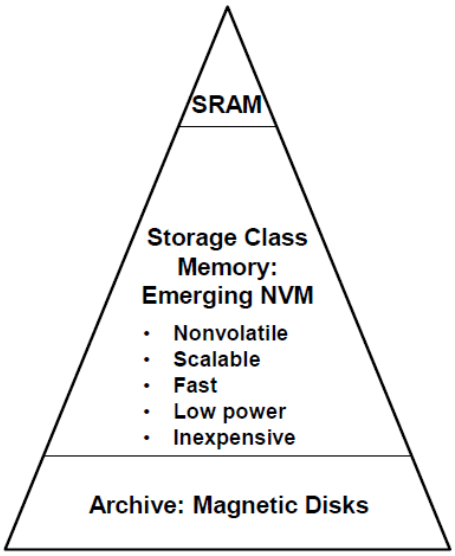


Figure 1.9 Simplified memory hierarchy [34]

In a computing system architecture to obtain the quantities of memory needed to perform useful computations and store large numbers of variables, a combination of high performance/low capacity (SRAM, DRAM) and low performance/high capacity memory must be used (Magnetic Disk). The modern memory hierarchy is often represented as a triangle (Figure 1.9) where the tip indicates a very small amount of very high performance memory, and ranges through the massive, slow archival storage at the

base. A new memory class, called Storage Class Memory (SCM) [34] with specific characteristics in between the ones of the extremes of the pyramid represented in Figure 1.9. is becoming important in order to fill the gap in the performances of fast but expensive SRAM and non-volatile cheap magnetic hard disk. Due to their characteristics, Resistive Switching memories are excellent candidates to become the next SCM.

1.3 Structures and materials

Basically, the RRAM element consists of the bottom electrode (BE), top electrode (TE), and the insulating oxide layer between those electrodes. In Figure 1.10 (a) is shown a widely used MIM structure where the memory cell is fabricated on the metallic via structure [32]. In this case, the material for BE can be different from that for the via, while in the case of the structure shown in Figure 1.10(b), the metallic via acts as BE [35]. When the area of the via under the oxide layer is smaller than that of TE, the via diameter can be regarded as the effective memory element size. Thus, the active size can be shrunk by reducing the via diameter and the oxide layers can be deposited on the via structure.

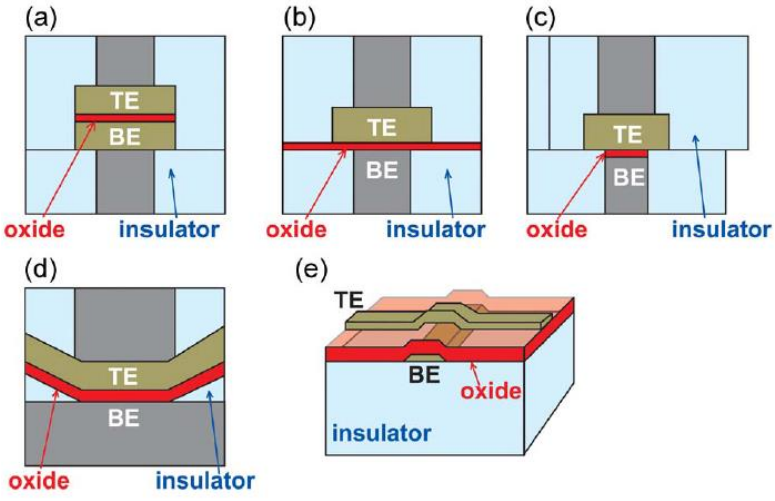


Figure 1.10(a) Typical MIM basic stacking structure of RRAM. (b) The memory element on the metallic via as a BE with oxide and TE layers. (c) The oxidized via material for the resistance switching oxide layer. (d) The concave structure and (e) the cross-bar structure consisting of the bottom and top electrode wires and blanket oxide film.

In order to simplify the fabrication process, it is possible to oxidize the metal in the via, as shown in Figure 1.10(c). Although the choice of the oxide material is strictly linked to the metal in the via, very good resistance switching has been reported in this structure with CuOx [36, 37] and WOx [38, 39] resistance switching layer obtained through the oxidation of Cu and W plugs. The cell in Figure 1.10(d) has a concave structure of the insulating layer. By decreasing the concave area, the memory element size can be scaled down. The resistance switching in this concave structure has been reported with NiO [40-42], TiOx [43] and HfOx [44-46] resistance switching layers. The cross-point structure is shown in Figure 1.10(e) [47,48]. The cells are sandwiched between word-lines and bit-lines and the memory element size is defined as the width of BE and TE wires. In any structures, stacking of memory elements is available, which can increase the effective memory element density and further consolidate the good scalability.

A wide range of binary metal oxides have been found to show resistive switching phenomena. Most of them are transition metal oxides. In Figure 1.11 are summarized the materials used for the resistive switching oxide layer and for the electrodes. In some cases, conductive nitrides, like TiN and TaN, are also used as electrode materials. As shown in Figure 1.12, where two graphs are sketched that resume the number of publications taken from Google Scholar regarding the materials used as oxides in

RRAMs, among the oxides, the most that have drawn attention are: HfOx, AlOx, NiOx, TiOx, and TaOx [33], because, in the past, they and have been extensively studied or integrated in CMOS compatible processes. The deposition of these kind of oxides usually can be done in different ways: oxidation of a corresponding metal, reactive sputtering and atomic layer deposition (ALD). Among these methods, ALD is widely used due to the ability to control the thickness of the thin film in a very accurate way.

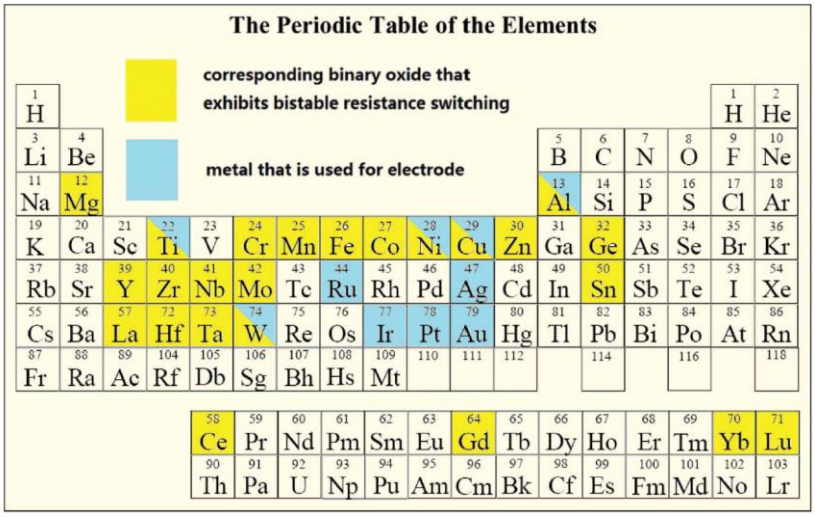


Figure 1.11 Summary of the materials used for binary metal–oxide RS memories.

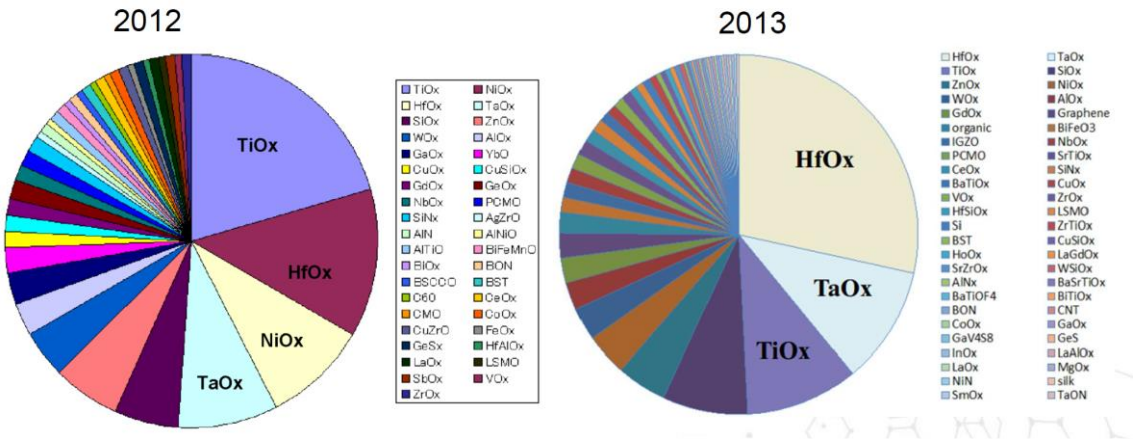


Figure 1.12 Oxides used for RRAM memories (from Google Scholar)

Although the preferred material for RS memories technology has not been determined yet, it is clear that the proper choice of the combination of electrodes and oxide material is crucial for the correct operation of the cell. Moreover, the fabrication processes should be carefully controlled in order to prevent the oxide layers from suffering process damage, such as plasma damages from the etching step and the deposition process.

1.4 Electrical Characteristics

According to the current-voltage (I-V) characteristics, it is possible to classify the RS devices in two general categories: unipolar (nonpolar) and bipolar. On the basis of I-V curves shown in Figure 1.13, in unipolar resistive switching mode (Figure 1.13a), the switching direction does not depend on the polarity of the applied voltage and generally occurs at higher voltage amplitude that of bipolar switching, where the memory operations are polarity dependent (Figure 1.13 b). A fresh memory device with high initial resistance state can be switched in to a low-resistance state (LRS) by applying a proper voltage. This process occurs just one time and it is called the 'electroforming' or simply 'forming' and it alters the resistance of the pristine device irreversibly [49-51]. After the forming process, the memory cell can be switched to a high-resistance state (HRS), generally lower than the initial resistance, by the application of a particular voltage called reset voltage. This process is called RESET. Switching from a HRS to a LRS is called 'SET'. In the SET process, generally, the current is limited by an external current compliance (CC) in order to avoid irreversible device damage. In bipolar resistive switching mode, the SET and RESET occur in the opposite polarity. For example, as shown in Figure 1.13 b), the SET operation can be done by applying

positive voltage on TE, while negative voltages reset the device. For bipolar switching to occur, the MIM stack should be asymmetric generally, such as different electrodes.

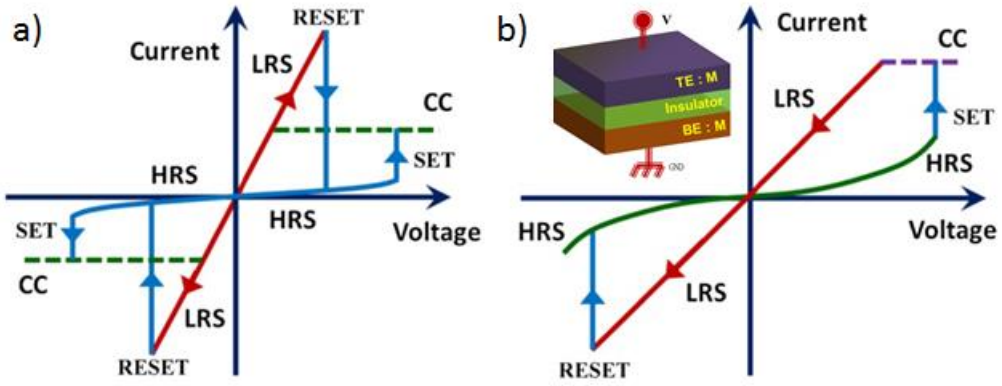


Figure 1.13 (a) I-V curves for unipolar (nonpolar) switching where the switching is independent on the polarity of the applied voltage. In (b) a typical bipolar I-V is shown: SET and RESET occur at opposite polarities. In the inset a schematic diagram of the RS memory structure.

In Figure 1.14 typical (not best or worst) device characteristics of the emerging memory technologies and the FLASH memories are compared [52]. Regarding the emerging memory technologies, in this thesis, we are interested in the RRAM properties. The first big advantage of the RRAM, compared to the FLASH technology, is the low voltage needed to program and erase (SET/RESET) the cells. While for FLASH, more than 10V are needed, for RRAMs volages lower than 3V are enough to perform the SET and RESET operations. The second important advantage of RS devices is the high speed of data reading but expecially the very high speed of writing/erasing. Compared to FLASH, they can be 1000 times faster. As one can see in Figure 1.14, also the cell area of RRAM is competitive with the FLASH (F in figure 1.14 is the

minimum dimension of the technological process). Before taking into account the reliability characteristcs, it is better to give some definitions:

- Retention: the time that a memory holds its bit state when the power is removed. This determine whether the memory is considered volatile or nonvolatile.
- Write Endurance: the number of times a memory can be programmed and erased before it fails (i.e. cannot be further programmed or read). Another parameter is the Read Endurance: the number of times a memory can be read before it fails, but this is typically greater than or equal to the write endurance, and hence the term "endurance" usually refers to write endurance.

While the data retention values are comparable for FLASH and emerging memories, the endurance for RRAMs can be even 8 orders of magnitude greater than the NAND. 10^{12} SET/RESET cycles have been demonstrated [53]. This is one of the reasons why RRAM devices are very promising candidates as Storage Class Memory [34]

	MAINSTREAM MEMORIES		EMERGING MEMORIES		
	FLASH				
	NOR	NAND	STT-MRAM	PCRAM	RRAM
Cell area	10 F^2	<4F^2 (3D)	6~50F^2	4~30F^2	4~12F^2
Multibit	2	3	1	2	2
Voltage	>10 V	>10 V	<1.5 V	<3 V	<3 V
Read time	~50 ns	~10 µs	<10 ns	<10 ns	<10 ns
Write time	10 µs–1 ms	100 µs–1 ms	<10 ns	~50 ns	<10 ns
Retention	>10 y	>10 y	>10 y	>10 y	>10 y
Endurance	>1E5	>1E4	>1E15	>1E9	>1E6~1E12

Figure 1.14 Device characteristics of mainstream and emerging memory technologies. Adapted from [52]

1.5 Architecture

One of the common RS memories array architectures is the 1 Transistor 1 Resistor (1T1R) array. In this sheme, each RS cell is connected in series to a selection transistor, as shown in Figure 1.13. With the addition of a selection transistor it is possible to isolate the selected cell from all other unselected cells. The WL controls the gate of the selector; thus, tuning the WL voltage it can control the write current that is delivered to the cell, e.g. the current compliance during SET.

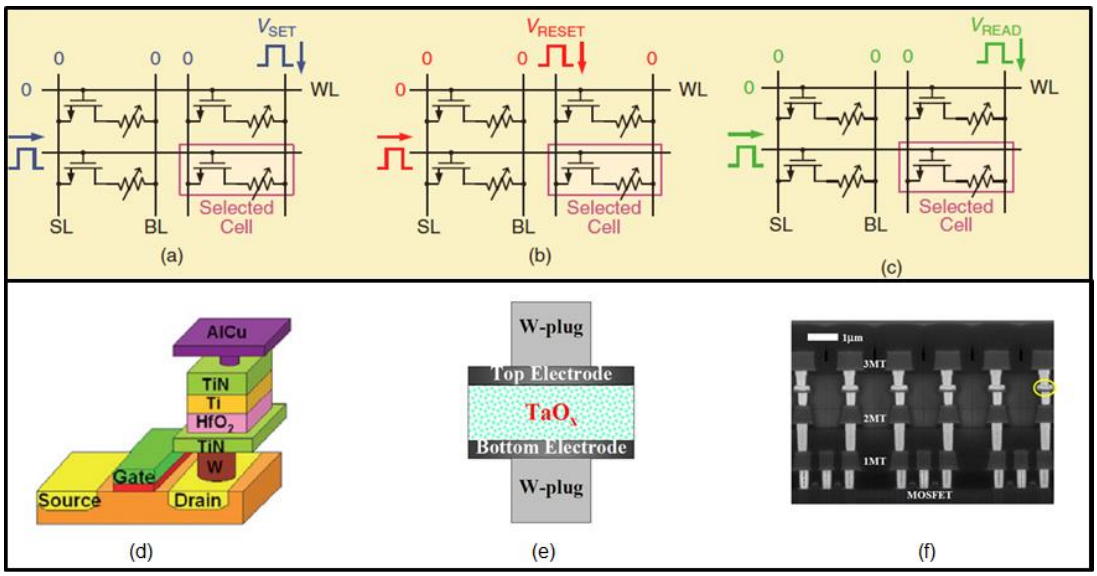


Figure 1.15 A scheme of the 1T1R array during (a) set, (b) reset and (c) read. In (d) a sketch of the direct contact between the memory cell and the drain of selector transistor is shown. In (e) a different cell structure [wei2008] is inserted between the metal layers M2 and M3 (f).

The memory cell top electrode connects to the BL while its bottom electrode connects to the contact via to the drain of the transistor, as shown in Figure 1.15 (d). It is also possible to insert the memory cell between two metal layers (Figure 1.15 (e)-(f)) in the back end of line (BEOL). The source line (SL) connects to the source of the transistor.

Figure 1.15(a)–(c) shows the typical write/read scheme for the 1T1R array. For the SET operation, WL voltage is applied to turn on the transistor of the selected cell and it imposes the current compliance, and a SET pulse is applied to the BL of the selected cell while SL is grounded. For the RESET operation, WL voltage is applied to turn on the selection transistor of the selected cell without imposing a current compliance. As RRAMs usually needs bipolar switching, a RESET pulse is applied to the SL of the selected cell while BL is grounded to reverse the polarity. For all the unselected cells, the WL, BL, and SL are grounded. A read voltage, usually lower than the SET/RESET voltages, is applied to the BL in order to read the data from the 1T1R array and a WL voltage is applied to turn on the selection transistor of the selected cell, while SL is grounded. A sense amplifier based circuitry can sense the difference in the read-out current for HRS and LRS through the BL comparing it to a reference current. Due to the off state for the unselected transistors, this writing/reading scheme is widely adopted in order to prevent crosstalk or interference issues, and each cell can be independently and randomly accessed. Multiple bits can be written/read in parallel into/from the same WL by activating multiple columns.

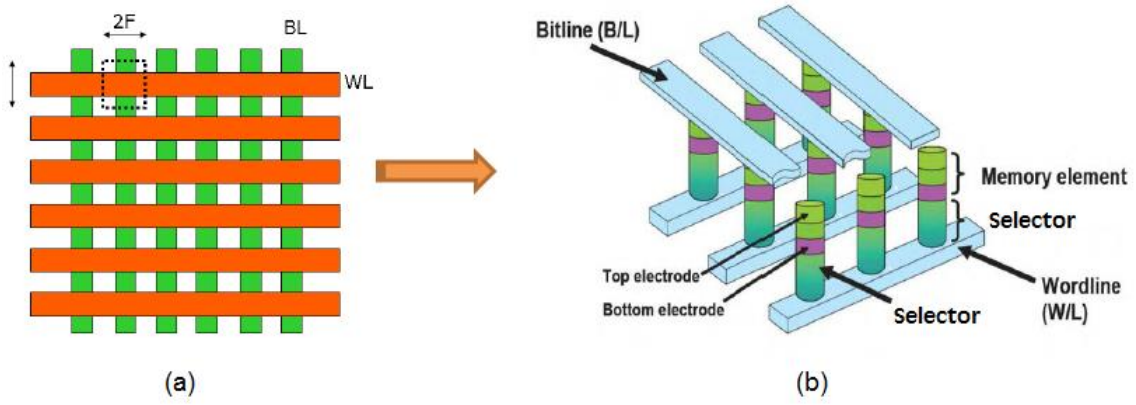


Figure 1.16 Cross Point array scheme

The other common array architecture is the cross-point array. It consists of perpendicular rows and columns with RRAM cells sandwiched in between, as shown in Figure 1.16. The cross-point array, in principle, can achieve a $4F^2$ cell area, where F is the minimun dimension allowable of the technological process; thus, it can reach higher integration density than the 1T1R scheme. In the oneselector and one-resistor (1S1R) architecture, the selectors are added in series with the memory element (Figure 1.16 (b)) to prevent cross talk or sneak current paths between cells.

In order to enable high density 1S1R cross-point arrays, an ideal selector should respect several specifications derived from circuit performances, device and process compatibility. One of the requirements for the selector is that a two-terminal device is needed to not cause extra memory array area overhead. Furthermore, a selector should be able to provide enough current for SET and RESET operations. For example, to enable resistive memory with ~10 A switching current, this translates to current density of ~10 MA/cm2 for a selector, where targeting 10x10 nm2 cell size. The maximum achievable cross-point array block size depends on the circuit performance like read margin, read/write power, etc. All these aspects are strongly affected by the leakage currents from the unselected memory elements. The leakage currents need to be as low as possible for improving the overall memory operation [54]. Considering that an ideal selector should have high current at high voltage as well as very small current at low voltages simultaneously, this translates into a highly nonlinear characteristic [55]. As most of the reported resistive memory cells exhibited better performance in bipolar operation mode, a bidirectional selector, which could provide symmetrical I-V, such as high drive current and highly nonlinearity at both polarities is required.

Parameter	Ideal value
ON current I_{on}	$>10MA/cm^2$
Threshold Voltage, V_{th}	~ 0V
ON/OFF ratio	$>10^6$
Processing temperature	<400 °C
Operating temperature	85 °C
Switching speed	<50 ns
Operation polarity	Compatible with memory element
Scalability	Comparable with memory element

Figure 1.17 Selector Device requirements

In order to transfer selector nonlinearity to the 1S1R full cell, it is important that the selector element is compatible with the memory cell, to guarantee enough current during writing operations and limited leakage current from the unselected memory elements. This means that the selector characteristics as: currents, speed, reliability, array yield and variability should be as good as or even better than the resistive memory cell [55]. Moreover, the materials composing the selector should be CMOS compatible, as for the memory element, thus limiting the usage of materials such as Pt, Ag, Au, etc. To enable 3D stacking of the memory arrays, the thermal budget of selector device fabrication should be BEOL compatible. It is also desired that a selector has a simple structure and low aspect ratio, to reduce the process complexity. Some of the parameters discussed are summarized in Figure 1.17.

1.6 Physical mechanisms of operation

It is widely accepted, that the physical mechanisms governing the switching phenomena in the oxide RRAM are based on the migration of oxygen vacancies and ions with related electrochemical reactions [56]. This ion drift is responsible for the formation and modification of a conducting filament (CF) between electrodes. This physical mechanism is also referred as redox (reduction/oxidation) effect [57]. The forming operation, corresponding to the first set process in the fresh sample, is similar to a dielectric soft breakdown. Initially, the oxygen vacancies (Vo^{2+}) density, namely double positive charged defects, in the metal oxide is relatively low. Under an appropriate electric field, the pairs of Vo^{2+} and oxygen ions (O^{2-}), double negative charged, are generated. The potential barrier for O^{2-}migration is lowered due to applied electric field and the oxygen ions drift toward the anode, where they form an oxygen reservoir, as shown in Figure 1.18.

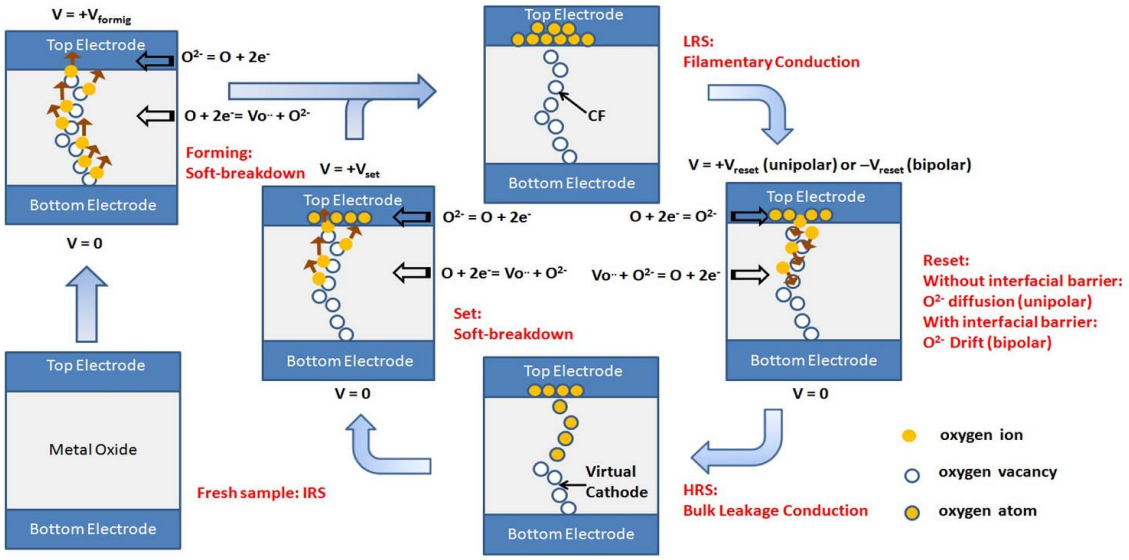


Figure 1.18 Picture of the switching mechanism in RRAMs

At the same moment, in the bulk, oxygen vacancies are generated and they lead to the formation of the conductive filament. Usually the as-deposited RRAM oxide films are amorphous or poly-crystalline, thus the CF is preferentially created along the grain boundaries [58]. The forming operation brings the RRAM cell in the LRS. During the RESET process, due the reversed electric field, O^{2-} are injected again in the insulating layer and they recombine with the Vo^{2+} composing the CF. In this way a gap region is created and the conductive filament is partially ruptured, thus originating the HRS. The residual CF with oxygen vacancies rich region is referred as the "virtual electrode" and it is responsible of the fact that the current in HRS is higher than the current in the pristine material. In the following SET process, the migration of oxygen ions occurs in the gap region and the CF reconnects both electrodes. This picture gives a phenomenological description of the experimental observations for many binary oxide RRAM devices.

1.6.1 Resistive Switching

As discussed in the previous paragraph, the oxygen ion migration is responsible of the switching phenomena in the RRAMs memory cell. This migration can be described through the ionic drift equation [59] and it can be derived from the simple rigid point-ion model illustrated schematically in Figure 1.19 [60, 61]. In solids, an ion hops in the net potential of the lattice ions. The solid parameters playing a role are the periodicity a and the activation energy U_A. The application of an electric field E, through an external voltage applied, modifies the potential shape and results in an effective barrier lowering by a factor $\sim \pm qEa/2$. Therefore, the overall effect on the average drift velocity can be described by the formula:

$$v \approx fae^{-\frac{U_A}{kT}} \sinh\left(\frac{qaE}{2kT}\right) \qquad 1.1$$

Where f is the frequency of escape attempts, q is the electron charge, K the Boltzman constant and T is the temperature.

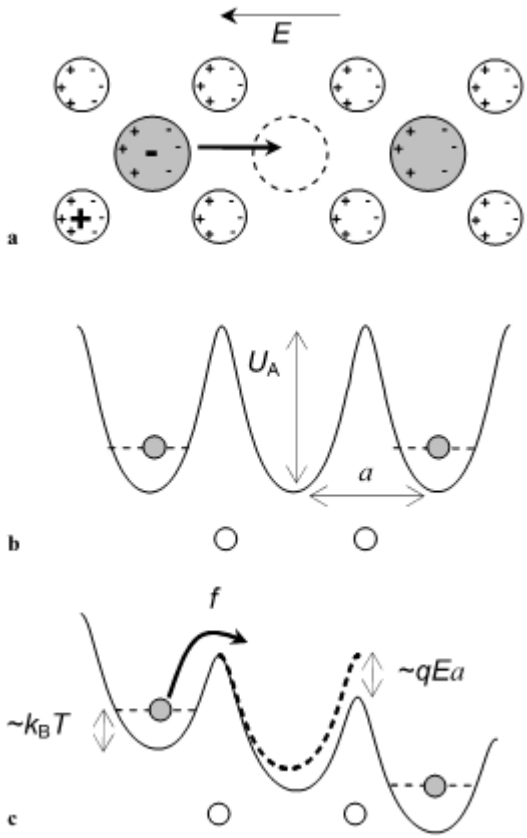


Figure 1.19 Illustration of vacancy drift. In a) the rigid point-ion model. In b) the corresponding potential energy profile and in c) the barrier lowering due to an applied electric field.

We should notice that the electric field in 1.1 is the local electric field and it can be much higher than the average field E_{mean} that is determined from the applied bias voltage $V= \int E_{mean}(x)dx$. In addition, the temperature expressed in 1.1 is the local temperature and it takes in to account Joule heating, which contribution is important

due to high current densities and electric field. The local temperature can be described by the equation:

$$T = T_0 + \frac{R_{th}}{R} V^2, \qquad 1.2$$

where T_0 is the room temperature and R_{th} is the effective thermal resistance [62].

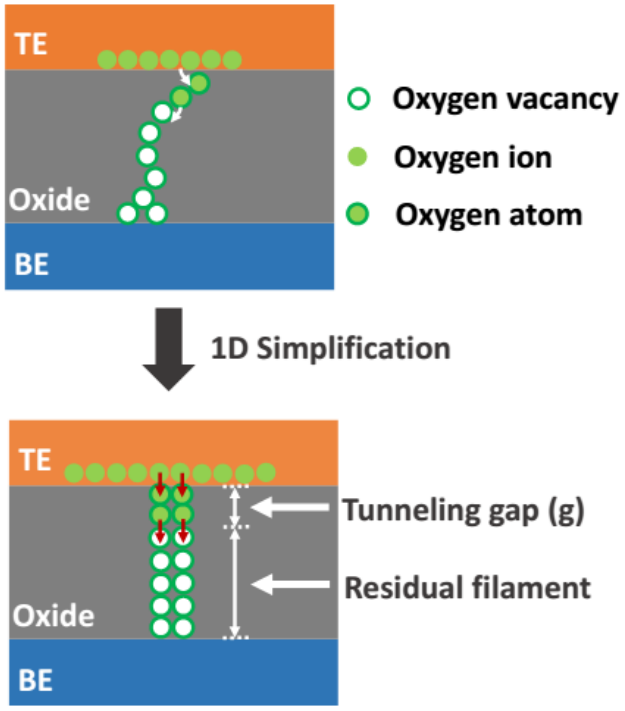


Figure 1.20 Conceptual picture of the conductive filament growth

In Figure 1.20, all the complex phenomena involving oxygen ions and vacancies migration are simplified into the growth of a single dominant filament that governs the switching process [63]. The size of the gap (g) between the tip of the filament and the electrode is the state variable determining the device resistance value. The kinetic of the state variable g can be described by the equation 1.1 in combination with 1.2 and taking in to account that the electric field E in 1.1 is the local electric field. This mean that this is the electric field in the gap region and it is much higher than the mean electric field simply calculated by $E=V/t_{ox}$., where V is the external applied voltage

and t_{ox} is the oxide thickness. A local enhancement factor that takes into account the non-uniform potential distribution in the cell is needed. The electric field in the gap region generally increases when the gap size decreases. In this way, field decrease as the tip of the filaments gets farther away from the electrode. A simplified expression of the dependence of the enhancement factor with the gap length is:

$$\gamma = \gamma_0 - c_0 g^b , \tag{1.3}$$

where γ is the enhancement factor and γ_0, c_0 and b are fitting parameters [63-651].

An alternative approach is the description of resistance switching via the numerical solution of differential equations that include continuity equations for charge carriers (Poisson equation), heat transport (Fourier equation), and ionized defects based on a drift/diffusion model [61]. The set of equations is given by:

$$\begin{aligned} \nabla\sigma\nabla\Psi &= 0 \\ -\nabla k_{th}\nabla T &= \left|\sigma\nabla\Psi\right|^2 \\ \frac{\partial n_D}{\partial t} &= \nabla\left(D\nabla n_D - \mu F n_D\right) \end{aligned} \tag{1.4}$$

σ in the first equation of 1.4 is the electrical conductivity and Ψ is the local potential. In the heat transport equation k_{th} is the thermal conductivity. The term $\left|\sigma\nabla\Psi\right|^2$ represents the local dissipated power density given by the product of the field by the current density, while $-\nabla k_{th}\nabla T$ term is the corresponding space variation of the heat flow. In the drift/diffusion ionic current equation n_D is the ions concentration, D is the ionic diffusivity, and μ is the ionic mobility. Diffusivity and mobility were assumed to be described by the Einstein relation. Finally, the electric conductivity σ and thermal conductivity k_{th} were assumed to depend on n_D.

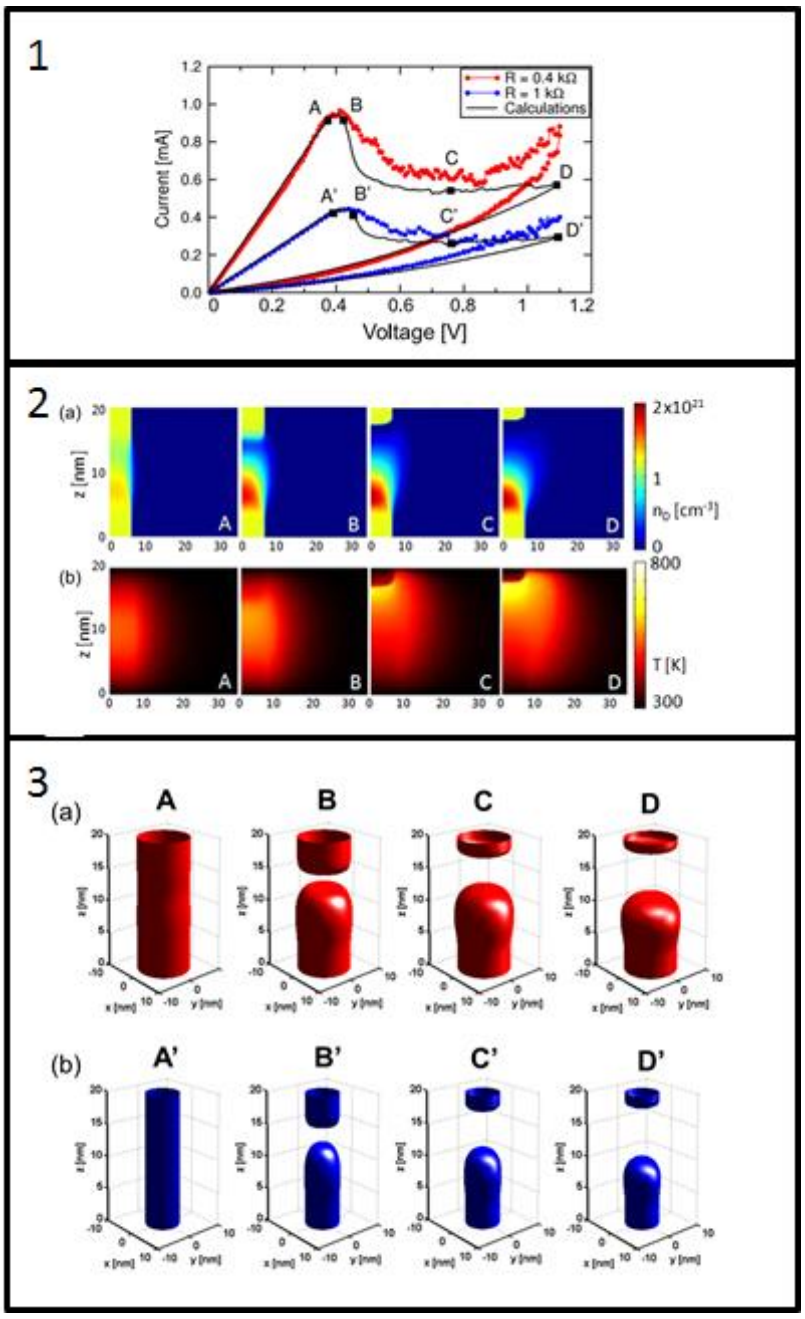


Figure 1.21 1) Experimental data for RESET and model [66]. In 2 a) the defect density and b) the local temperature from the points A to D marked in the IV curve in 1. In 3 the contour plot of the density concentration from A to D and from A' to D'.

In Figure 1.21 1) Experimental RESET data, with points A,B,C and D highlighted and the curves calculated with equation 1.4 are shown [66]. In 2) the profile of calculated defect concentration n_D (a) and temperature (b) for points A-D are depicted.

At the RESET voltage, corresponding to point B, the temperature has a maximum value around 500 K thus the ion migration started in the direction of the electric field (towards the BE). Due to this process, the regions above the middle of the CF strongly decreased the concentration of defects, while the region below the center of the CF shows the increase of defects concentration. Increasing the voltage, the depleted gap length increases, thus increasing the resistance as observed in the I–V curves in 1). As the depleted gap extends, the local temperature in 2) (b) changes, since the voltage drop mostly occurs in the gap region. As a result, the portions of the CF above and below the depleted gap stay at lower temperature and electric field, decreasing the migration flux of ionized defects. To sustain more defect migration and the corresponding resistance increase. This is why we observe a gradual RESET in Figure 1.21 1). In 3) it is shown the contour plot of the defect concentration for the points A-D (a) and A'-D' (b) corresponding to two different CF diameters. As one can see, the initial diameter does not affect the evolution of gap increasing with the voltage. This is because the local field and temperature are not dependent on the lateral CF dimension. In conclusion:

- Gap region extension is driven by the directional drift of ionized defects along the electric field direction
- Diffusion effects can be seen as a fattening of the CF at increasing voltage.

1.6.2 Current Conduction Models

While most of the scientific community observes a linear or ohmic behavior in the LRS, on the HRS different electron transport mechanisms interpretations are reported: Poole–Frenkel emission [67,68], Schottky emission [69,53], the space charge limited current (SCLC) characteristic [70,71] .

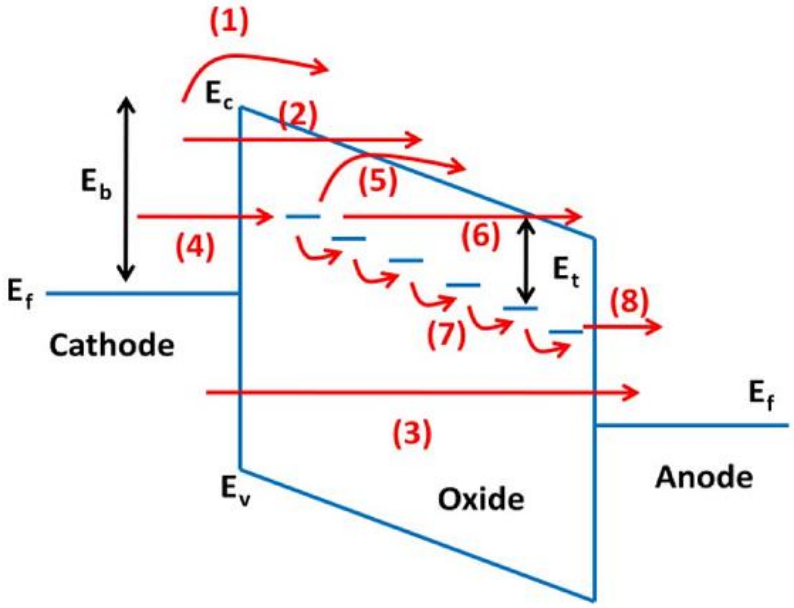


Figure 1.22 Picture of possible conduction mechanism involved in a MIM stack [33]

Various metal–oxide RRAMs may have different dominant conduction mechanism depending on different structure characteristics: material properties like the dielectric gap or the defect trap proprieties; the fabrication process conditions like the depositions of both metals and dielectric and the interface between the oxides and the electrodes proprieties. In Figure 1.22 the possible conduction mechanisms through a MIM stack are resumed:

1. Schottky emission:
2. Fowler–Nordheim (F–N) tunnelling
3. Direct tunnelling
4. Tunneling from cathode to traps
5. Poole–Frenkel emission
6. F–N-like tunneling from trap to conduction band
7. Trap assisted hopping or tunnelling

8. Tunneling from traps to anode

As it is clear from this summary of the transport mechanism that can be involved in the conduction of RRAM devices, the complete and accurate description of the I-V characteristic may consider the contemporary presence of different transport processes. But, in most of the cases there is a predominant mechanism that can describe the cell current voltage characteristic.

Considering the description given in paragraph 1.6.1, the parameter governing the switching phenomena of RRAMs is the gap between the tip of the conductive filament inside the oxide and the opposite electrode. In this scenario, since most of the tunneling mechanisms have an exponential dependence on the tunneling distance and field strength, the amount of current flowing through the cell can be generalized to be: [63,64]:

$$I = I_0 e^{\left(-\frac{g}{g_0}\right)} \sinh\left(\frac{V}{V_0}\right), \qquad 1.5$$

where I_0, g_0 and V_0 are fitting parameters. With equation 1.5 it is possible to catch both the LRS and HRS currents. Considering a fixed gap (g) in LRS, at low voltages the *sinh* approximates to linear function and the combination of the parameter I_0 and V_0 determines the slope of the IV. During the RESET increasing the gap distance due to ions migration according to kinetic expressed in equations 1.1-1.3 the current decreases with voltage.

Alternative approaches [72,73] regarding the conduction mechanism in the memory cell assume that, regardless of the conduction state, if the current flows through a narrow

constriction (filamentary path) between two electron reservoirs, called quantum point contact (QPC), the Landauer theory predicts that the current I flowing under the application of a finite bias V can be calculated as [74]:

$$I = \frac{2q}{h}\int T(E)\left[f\left(E-\beta qV\right) - f\left(E+(1-\beta)qV\right)\right]dE, \qquad 1.6$$

where E is the energy, T is the transmission probability, f is the Fermi distribution, q is the electron charge, and h is the Planck constant, β is the fraction of the applied bias that drops on the source side of the constriction. Assuming an inverted parabolic potential barrier for the first quantized sub-band inside the constriction, we obtain an expression for I:

$$I = \frac{2q}{h}\left\{ qV + \frac{1}{\alpha}\ln\left[\frac{1+\exp\left[\alpha\left(\Phi-\beta qV\right)\right]}{1+\exp\left[\alpha\left(\Phi+\left(1-\beta\right)qV\right)\right]}\right]\right\}, \qquad 1.7$$

where α is a constant related to the sub-band spatial curvature, Φ is the barrier height. For the sake of simplicity, assuming that each component Φ,α,n follows a time dependent smoothing hysteron. With this description, varying the parameters values, is possible to obtain different I-V characteristics, covering a multitude of I-V shapes, as shown in Figure 1.23

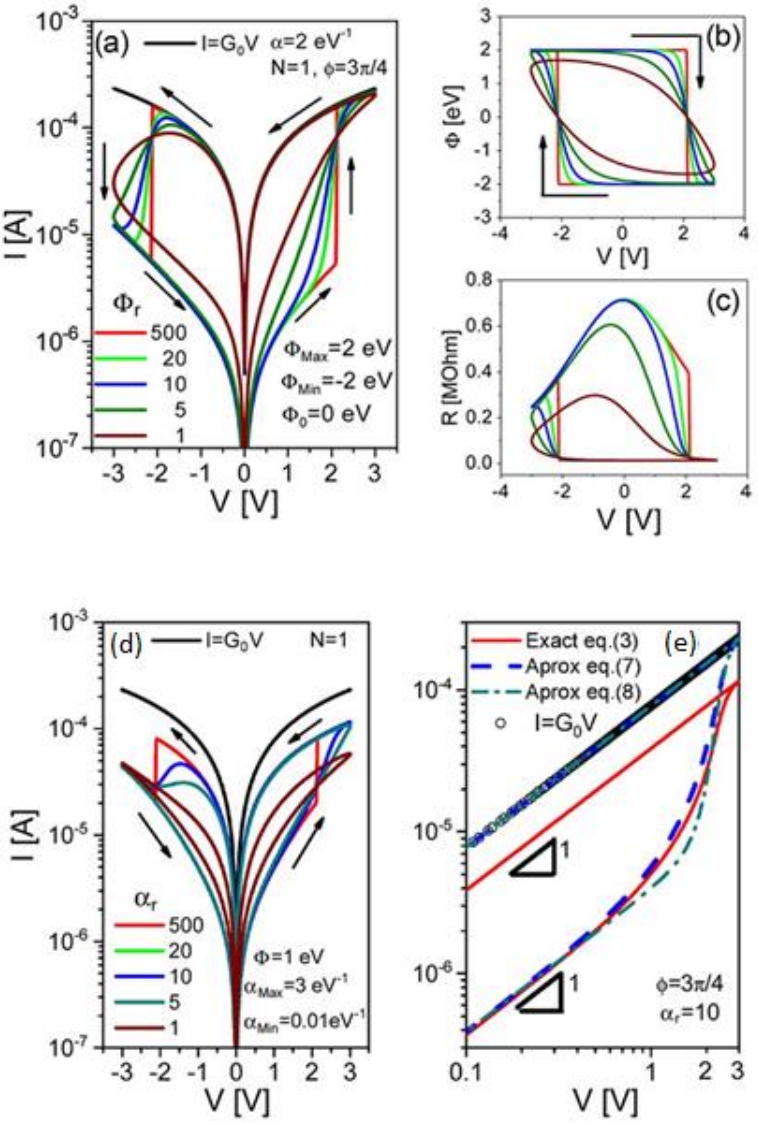


Figure 1.23 (a) I–V curves, (b) Φ–V loops and (c) R–V for different Φ values. (d) I–V curves for different α values. (e) Exact and approximate expressions for the current [72].

1.7 Open issues

The key issue of RRAM cells is the variability of the switching parameters. Due to the stochastic nature of ionic (oxygen vacancies or metal ions) migration, defects characteristics and material proprieties like the grain boundaries in polycristalline materials [58, 75, 76], the filament shape varies from device to device and also from

cycle to cycle in the same device. The fluctuation in the number of particles defining, for example, the radius of the conductive filament or the fluctuations in constriction geometry can be the reasons behind the resistances distributions shown in Figure 1.24 [77] where both the variabilities intracell (a,b) and intercell (c) are represented. The HRS resistance variation comes from the variation of the ruptured CFs length, thus any small variation of the tunneling gap distance may be magnified to be an exponential dependence of the tunneling current on the tunneling distance.

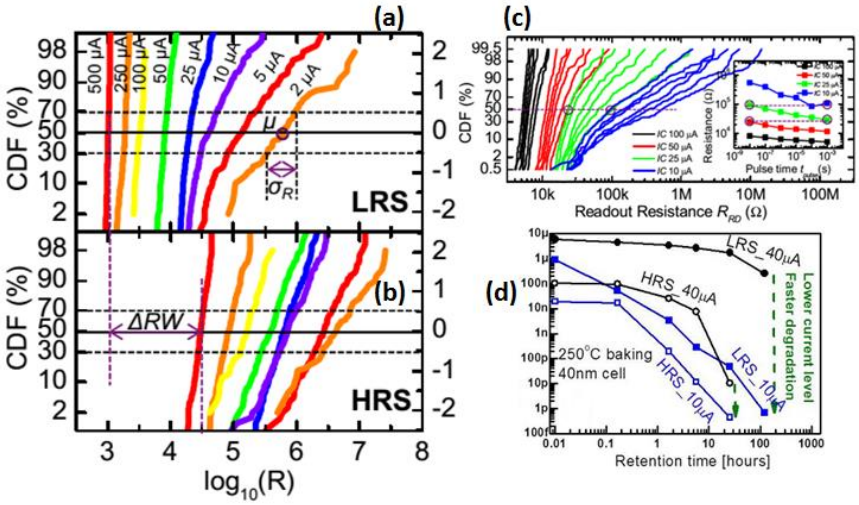


Figure 1.24 (a) LRS and (b) HRS distribution for 200 cycles in the same cell for different current compliances+ values. Distribution are approximately lognormal. In (c) LRS distribution in 100 different cells for various current compliances and pulse time. In (d) data retention measurements in HRS and LRS for two different current compliances.

Important variations in resistance distribution (even bigger than one order of magnitude) mean a more complex sensing circuit design and requires the program and verify techniques to acheive the target states, which could be latency consuming for the MLC operations. Although RRAM could require small write current (e.g., ~10 μA)

due to the filamentary switching mechanism, reducing too much the programming current may lead to increasing instabilities like a wider resistance distribution as shown in Figure 1.24 or it could reduce data retention performances when filament is too thin [78] as shown in Figure 1.24 (d) where data retention tests are reported for two different current compliances.

Altough reducing the the RRAM currents is critical for the reliability of the device, it is important for the 1T1R array architecture described in paragraph 1.5, because with new technological nodes the current of the low power transistor for logic accordingly scales too.

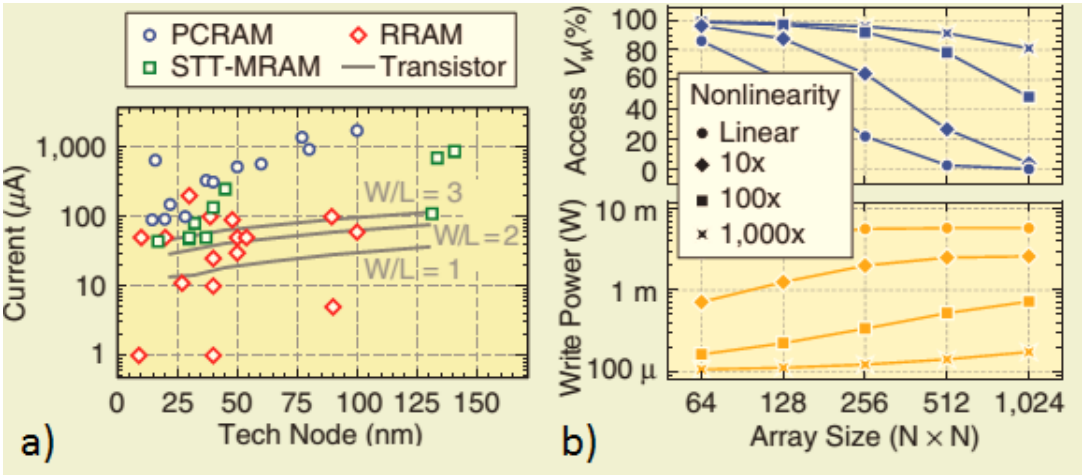


Figure 1.25 a) Silicon CMOS low-power logic transistor's drive current with the scaling from 130 nm down to 20 nm for different W/l simulated with the Predictive Technological Model [79] and Resistive Switching write currents. B) SPICE simulation of write margin and write power as a function of cross-point array size. increasing the nvm cell's i-v nonlinearity (N) by adding selectors is helpful to minimize the ir drop problem along the interconnect wire and the sneak path problem

Figure 1.25(a) shows the silicon CMOS low-power logic transistor's drive current with different technology nodes from 130 nm to 20 nm for different ratio of the channel Width and Length simulated with Predictive Technology Models (PTM) [79]. The typical write current of Resistive Switching Devices, including RRAM, from the literature data is also marked. As one can sse, the RRAM's write current typically ranges from 10 to 100 μA and it does not depend on the device area due to the filamentary conduction mechanism, so it does not scale with the technology node. Considering the transistor characteristics, for example, the drive current at V_{GS} = 5V can approach 40 μA for a W/L =1 transistor at 27-nm node [80]. So, for this reason, reducing the write current down to sub-10 μA is of great importance for continuing the scaling of the 1T1R array. In addition, reducing the write voltage down to sub-1 V is also necessary for embedded applications if using a logic-compatible process. The alternative array architecture to the 1T1R is the the cross-point array that suffers of other challenges: 1) the voltage drop problem along the interconnections and 2) the sneak path problem through the unselected cells. The voltage drop (usually referred as IR drop) issue becomes not negligible when the WL and BL wire associated resistance increases. This happens scaling the technological node in the sub-50-nm regime where the interconnect resistivity drastically due to the increased electron surface scattering. For example, at 20-nm node, the copper interconnect resistance between two neighboring cells is ~2.93 Ω; thus, the IR drop along the wire for a large array is no longer negligible [52]. The farthest cell from the driver sees an interconnect resistance that can be comparable to the typical LRS resistances, thus a part of the write voltage will drop on

the wire instead of the RRAM cell. A compensation voltage could be necessary to guarantee a successful write operation.

Figure 1.25(b) shows the SPICE simulation of the write margin and write power as a function of the cross-point array size for different I–V nonlinearity (N). The nonlinearity is defined as the current ratio between the current at the writing Voltage and the current at the half of the writing voltage. The cell resistance is fixed and the wire width is fixed at 20 nm. It is seen that at least , N 1000 > is needed for maintaining sufficient write margin and minimizing write power for a large array (e.g., a 1,024 × 1,024 array).

1.8 Prototypes

In recent years, several RRAM prototypes have been reported in literature. A wide range of switching materials were adopted including metal oxide based RRAM cells and conductive bridge RAM. As shown in Table 1 [62], most of prototypes were fabricated with technological nodes above 100 nm, in relatively small array size (typically below 10 MB). The exploitation of the capabilities of RRAMs devices in relaxed technology node and small array size demonstrates the strong interest for non-volatile embedded applications like microcontroller, where RRAM offers the advantages of faster read operation and lower power consumption.

Table 1 RRAM prototypes, adapted from [62]

Year	Tech.Node	Capacity	Company/Institute	Switching Material	Ref.
2010	130 nm	64 M	Unity	CMOX	[81]
2011	180 nm	4 M	ITRI	Ti/HfO2/TiN	[82]
2011	130 nm	384 k	Adesto	Ag/GeS2	[83]
2011	180 nm	4 M	Sony	CuTe/GdOx	[84]
2012	180 nm	8 M	Panasonic	TaN/TaO2/Ta2O5/Ir	[85]
2013	110 nm	512 k	Panasonic	TaN/TaO2/Ta2O5/Ir	[86]
2013	24 nm	32 G	Sandisk/Toshiba	MO	[87]
2014	27 nm	16 G	Micron/Sony	CuO	[88]
2015	90 nm	2 M	Renesas	/MetalTa2O5/Ru	[89]

In 2013 Panasonic released in the market the first microcontroller with 64KB of RRAM (Figure 1.26). The Panasonic website describes it as "*MN101L Resistive RAM (ReRAM) Embedded 8-bit MCUs are industry-leading, low-power microcomputers delivering enhanced processing performance, 50% lower power consumption, and over 5 times faster rewriting than flash memory or EEPROM. […]. With the low-power consumption ReRAM, shorter processing time and voltage control by high-performance CPU, and leakage current reduction of new fabrication process, MN101L reduces power consumption by 50% compared to the existing Flash microcomputer. Unlike flash memory or EEPROM, MN101L does not require a data erase to provide over 5 times faster rewriting rate. This design makes the Panasonic's MN101L ReRAM Embedded 8-bit MCUs ideally suited for portable healthcare, security, and sensor equipment.*" The key points for the adoption of RRAM devices instead the classical FLASH in the Panasonic microcontroller are the reduced

power consumption due to the very low voltages needed to write and read the memory cells and the low leakage, the high speed of operations and the higher cyclability.

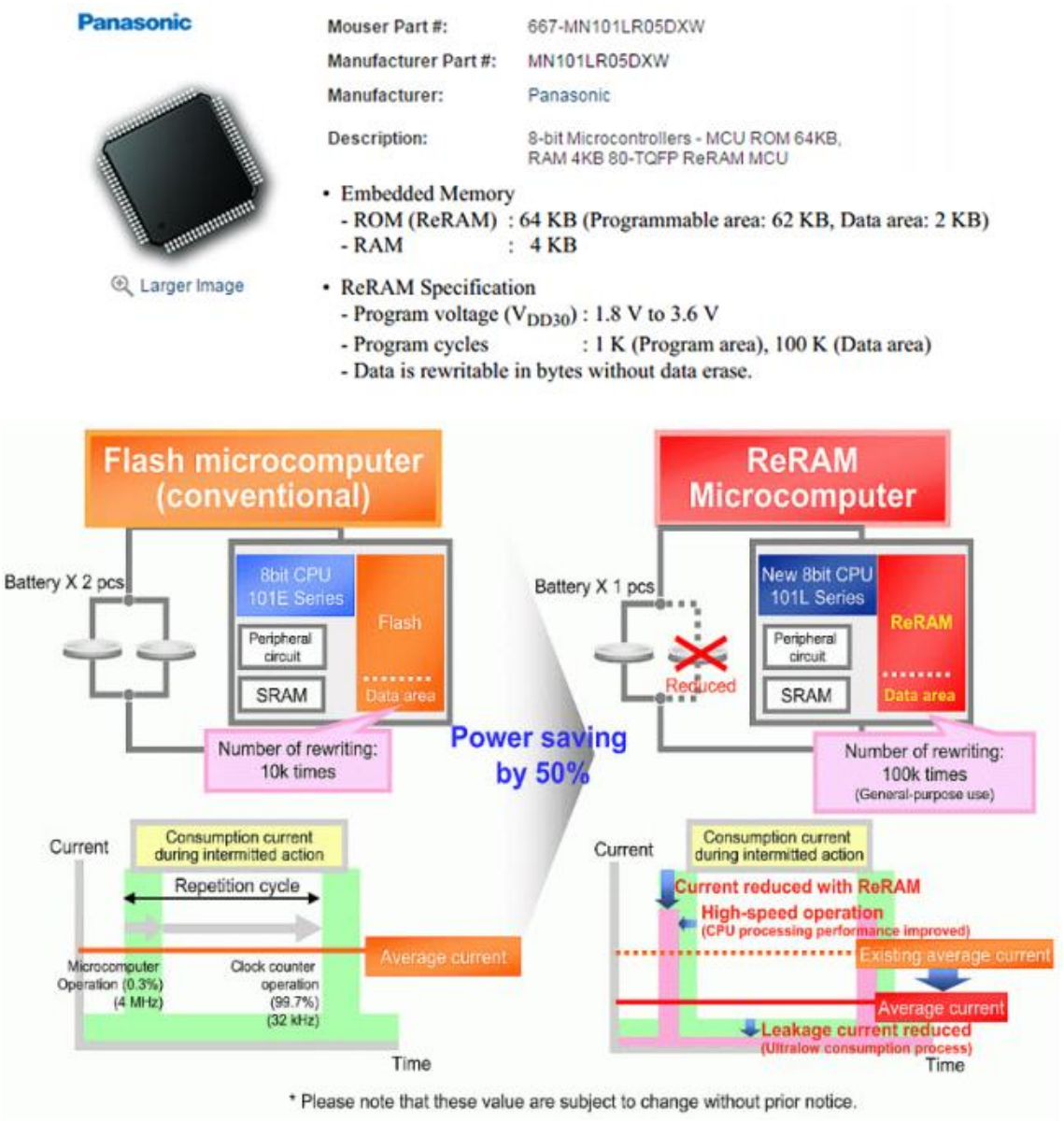


Figure 1.26 Panasonic 8-bit microcontroller MN101LR05DXW with 64KB ReRAM. In the bottom part of the figure, the advantages of using ReRAM respect to FLASH conventional microcontroller are highlighted.

Considering Table 1, in two cases, a more aggressive technology node were adopted in order to obtain device size below 30 nm, thus obtaining an array size of several GB

[87,88]. In [88] a 16Gb ReRAM designed in a 27nm node is presented. The 1T1R memory cell (Figure 1.27) is dual-layered with a CuTe conductive material and a thin insulator. The resistive element with top and bottom electrodes (TE, BE) operates through bipolar switching. A 1GB/s DDR interface and an 8-bank concurrent DRAM-like core architecture. High parallelism, a pipelined data-path architecture and innovations such as concurrent set/reset verify. The selector employs a buried-wordline architecture and is also used for current control during set (transition to LRS).

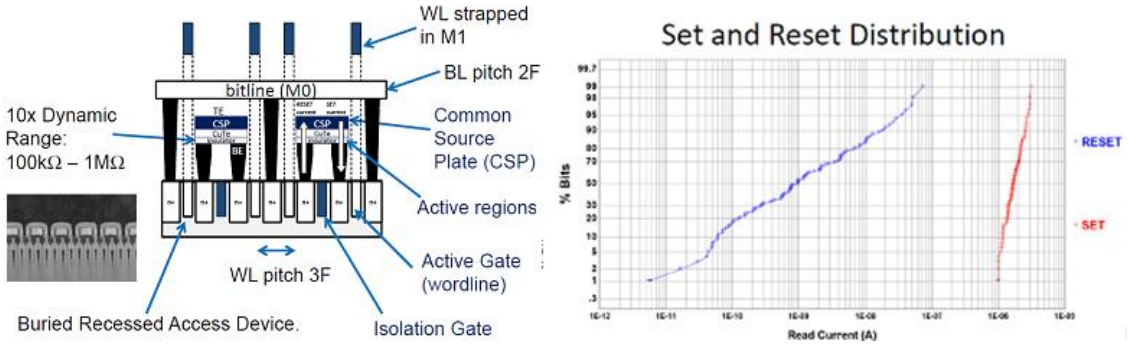


Figure 1.27 16GB RRAM chip memory cell and LRS and HRS distributions

1.9 RRAM as Memristor

In the last paragraph of this chapter, the author wants to give a mention on the Memristor. This is motivated by the fact that, during last years, the RRAM community and the Memristor community are merging together and it is not unusual in literature to refer to Resistive Switching devices as Memristors.

In a 2008 Nature Article entitled "The missing memristor found" [90], the authors Dmitri B. Strukov, Gregory S. Snider, Duncan R. Stewart and R. Stanley Williams of the HP Laboratories claimed that they found the fourth passive element theorized by Leon Chua in 1971 [91]. The memristive propriety arises naturally in nanoscale systems

in which resistive switching phenomena and ionic transport are driven by external applied voltage.

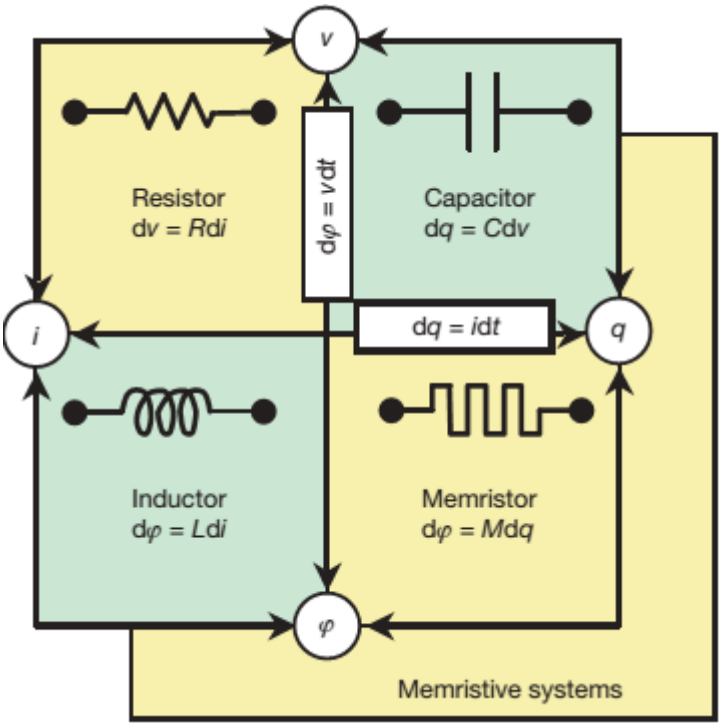


Figure 1.28 The four fundamental two-terminal circuit elements: resistor, capacitor, inductor and memristor. Resistors and memristors are subsets of a more general class of dynamical devices, memristive systems. Note that R, C, L and M can be functions of the independent variable in their defining equations, yielding nonlinear elements. For example, a charge-controlled memristor is defined by a single-valued function M(q).

Prof. Leon Chua noted that six different mathematical relations connect pairs of the four fundamental circuit variables: electric current *I*, voltage *V*, charge *Q* and magnetic flux *Φ*. One of these relations (the charge is the time integral of the current) is determined from the definitions of two of the variables, and another (the flux is the time integral of the voltage) is determined from Faraday's law of induction. Thus, there should be four basic circuit elements described by the remaining relations between the

variables. The 'missing' element—the memristor, with memristance M—provides a functional relation between charge and flux, $d\Phi=Mdq$.

If M is a constant, the memristor is identical to the resistance. The interesting case is when the memristance is not constant but depends on the history of current that had previously flowed through the device. This means that the present state depends on how much electric charge has flowed and in what direction through it in the past. In [90] the HP Labs team realized a memristor composed by a thin film of titanium dioxide and Platinum metal electrodes, thus connecting the operation of RRAM devices to the memristor concept. After the publication of the HP Labs, the interest in memristors grew up dramatically. The discovery of the missing element had great attention by the media and newspapers.

Following this, Prof. Leon Chua has argued [92] that the memristor definition could be generalized to cover all forms of two-terminal non-volatile memory devices based on resistance switching effects. All the RRAM devices exhibit a distinctive "fingerprint" characterized by a pinched hysteresis loop confined to the first and the third quadrants of the v–i plane whose contour shape in general changes with both the amplitude and frequency of any periodic "sinewave-like" input voltage source, or current source. In particular, the pinched hysteresis loop shrinks and tends to a straight line as frequency increases. So RRAMs and Memristors can be considered as the same device.

2 KEY ASPECTS OF RESISTIVE SWITCHING MEMORIES OPERATIONS: EXPERIMENTAL ANALYSIS AND MODELING

In this chapter, a description of the tested devices will be given in terms of device structure, fabrication and electrical characteristics. Key aspects of the behavior of the cell will be addressed. Experimental characterization and modeling of the influence of the applied voltage waveform, device geometry, external temperature, and electrical stress will be described.

2.1 Devices

Choice of proper materials composing the MIM stack is a key issue for fabrication and electrical operation of RRAM devices. In this Ph.D. project different MIM structure were used. Most of them were fabricated by the "Commissariat à l'énergie atomique et aux énergies alternatives" (CEA), Grenoble, France.

In the tested devices HfO2 was used as insulating layer, while different combination of metal electrodes, including Platinum (Pt), Titanium (Ti) and Titanium Nitride (TiN) were adopted. Two different set of samples fabricated in two phases were studied. In the first set, resistive memory cells, with dimensions ranging from 1.8µm down to 300 nm, were fabricated above a metal 1 level (see Figure 2.1)

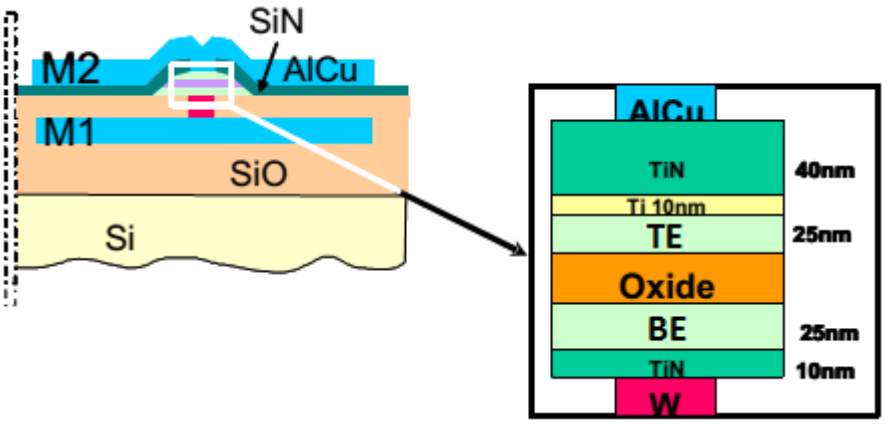


Figure 2.1 Schematic of the integrated memory cells (cell diameter ranges from 1.8µm to 300 nm)

A TiN layer was deposited before the bottom electrode in order to ensure its adhesion. A TiN/Ti layer was used at the top of the structure as antireflective layer for the lithography. To fabricate samples showed in Figure 2.2 HfO2 was deposited through Atomic Layer Deposition (ALD) at 350 °C on Physical Vapor Deposition (PVD) TiN or Pt. Top electrode (Ti or Pt) was then deposited by PVD and patterned via Ion Beam

Etching (IBE). X-Ray diffraction data revealed coexistence of both orthorhombic and monoclinic phases in HfO2.

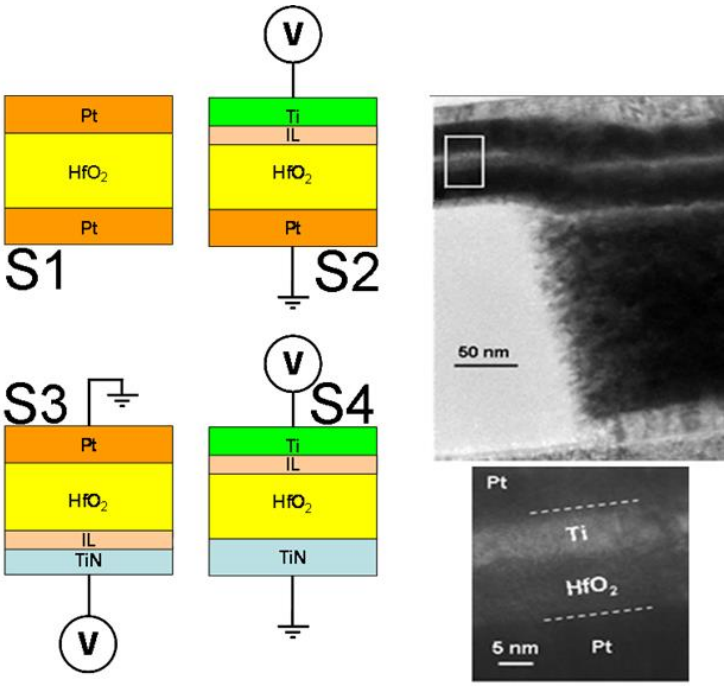


Figure 2.2 Studied device splits. Interfacial layer (IL) schematically represented according to physico-chemical analysis and high resolution TEM image of sample (Pt-Ti). Inset: closer view on the active stack [45].

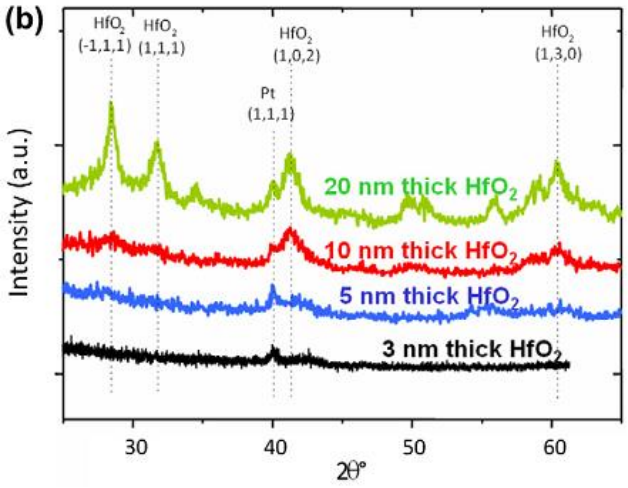


Figure 2.3 X-ray diffraction HfO_2thin films (with varying thickness), deposited in all cases on Pt [46].

As shown in Figure 2.3 a 10nm thick HfO2 on Pt is mainly monoclinic [46], while it is found mainly orthorhombic on TiN.

Atomic concentration profiles measured by Energy-dispersive X-ray spectroscopy (EDX: an analytical technique used for the elemental analysis or chemical characterization of a sample. It relies on an interaction of some source of X-ray excitation and a sample) upon Pt/HfO2/Ti samples (see High Resolution TEM, Figure 2.2 and EDX, Figure 2.3 a)) and TiN/HfO2/Pt (Figure 2.3 b))) devices. Oxygen profiles were found shifted either towards Ti (a) or TiN (b) electrodes, indicating the formation of HfOx/TiOx or HfOx/TiOxN interfacial layers (IL) induced by the O-getter character of Ti [93,94].

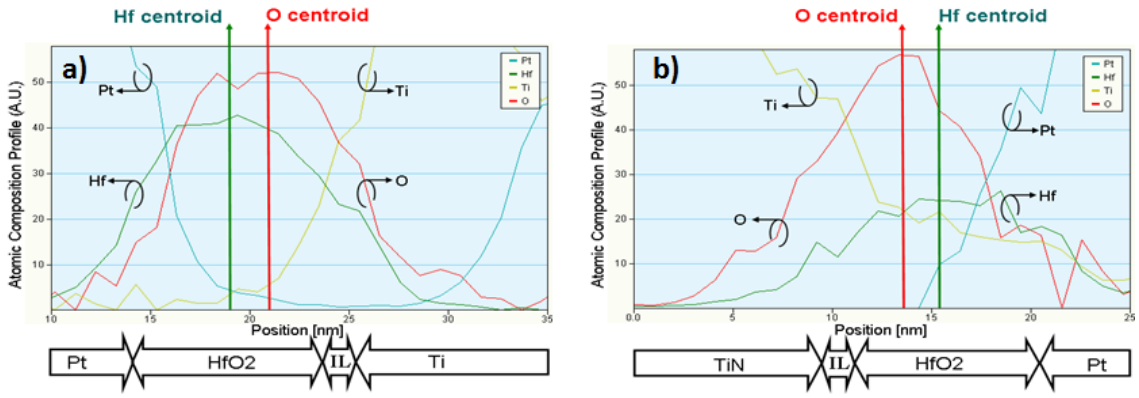


Figure 2.4 (a) Atomic composition of the active layers in as fabricated (Pt-Ti) device measured by EDX. Distribution of O from HfO2 is found to be shifted towards Ti electrode indicating the presence of an interfacial oxide layer. In TiN-Pt samples (b) distribution of O from HfO2 is found to be shifted towards TiN electrode.

The second set of samples has a TiN/HfOx/Ti stack. The devices were fabricated in a in mesa process. TiN bottom electrodes (BE) were deposited by PVD and patterned by DUV (Deep Ultra Violet) lithography to obtain device areas from 600nm×600nm to

3μm×3μm. Following this step, HfOx was deposited by ALD at 300°C and annealed at 400°C in N2 ambient for 30min. The Ti top electrode was deposited by PVD. A final anneal was performed at 450 °C for 30 m.

2.2 Quasi-Static electrical Characteristics

Quasi-static characterization of Forming, SET, RESET of the different kind of samples was performed using the Agilent B1500 Semiconductor Parameter Analyzer (SPA). The voltage sweep speed was set at $1 V/s$ and the current compliance was externally imposed to the devices by the SPA. Forming and SET/RESET current-voltage characteristics are shown in Figure 2.5. To initiate switching a preliminary forming operation is required. It is a one-time writing process at voltage higher than regular operating bias. When the four samples are compared (Table 2, Figure 2.5) it appears that the Ti presence strongly decreases the forming voltage (under 3V). We identified device polarity behaviors, i.e. unipolar (non-polar) vs. bipolar. To this aim, we performed several SET/RESET I-V sweeps on the devices. The results are summarized in Table 2, which reports SET and RESET voltages, and LRS and HRS resistances. Pt-Pt samples are found to be unipolar (non-polar), while samples TiN-Pt, TiN-TiN and TiN-Ti are bipolar. They require positive voltage applied on TiN electrode except in case with Ti where devices need positive voltage applied to Ti to achieve SET. This behavior suggests that the presence of an interlayer rich of oxide vacancies plays a fundamental role in determine the polarity of the switching of the devices [45]. For the unipolar samples the LRS was found to be lower than the others, while the HRS higher, thus giving the highest ratio between the two resistive state. The bipolar samples have LRS value around 500 Ω and HRS between 10kΩ and rarely 1MΩ.

Furthermore, bipolar samples show symmetric behaviour between SET and RESET in terms of switching voltages.

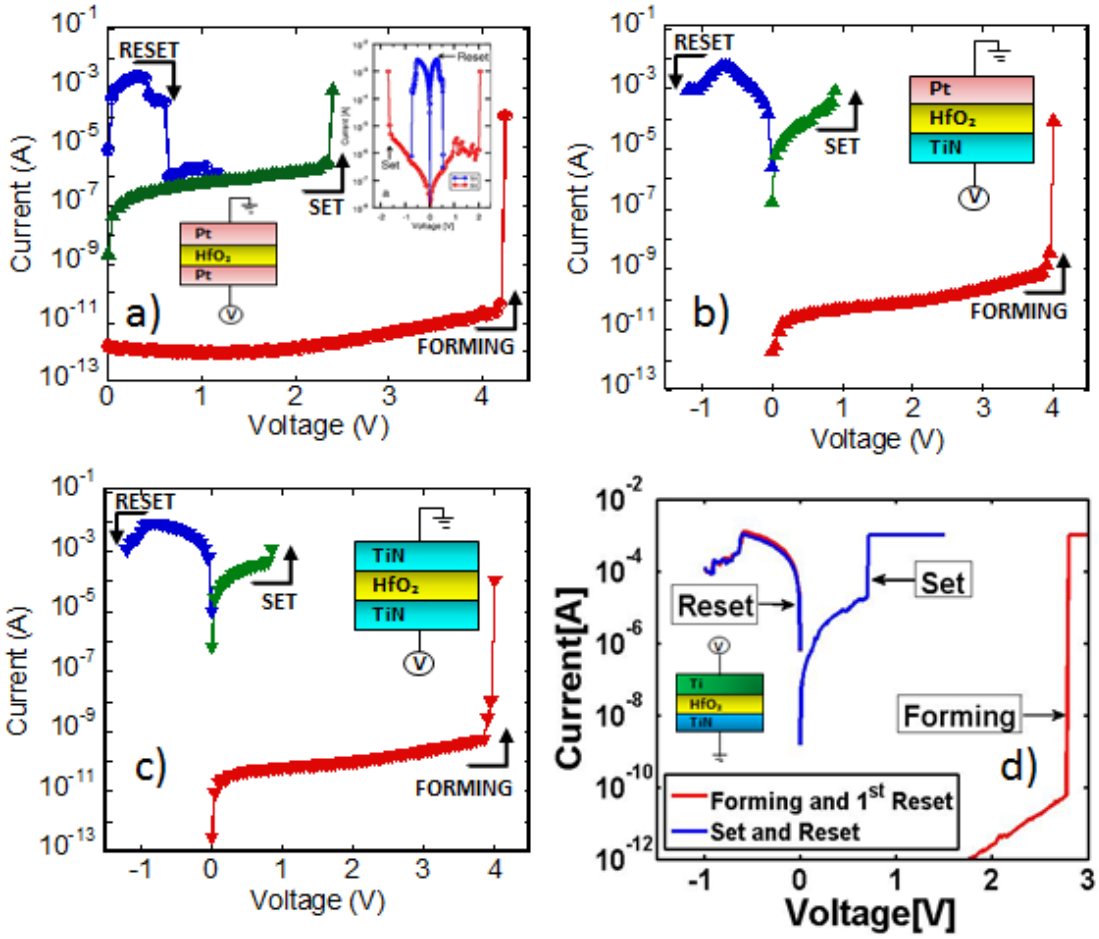


Figure 2.5 Current – Voltage (IV) curves for Forming SET and RESET for a) Pt-Pt b) TiN-Pt, c) TiN-TiN and d) TiN-Ti samples. In the inset of a) the unipolar characteristics is shown. The MIM stack of different samples are also sketched in the insets of (a-d).

Table 2 Electrical parameters measured on the four splits under investigation

	$V_{forming}$[V]	V_{SET}[V]	V_{RESET}[V]	R_{SET}[Ω]	R_{RESET}[Ω]
UNIPOLAR BEHAVIOUR					
Pt-Pt	4.4 ± 0.3	3 ± 1.5	0.4 ± 0.15	<100	~1M
BIPOLAR BEHAVIOUR					
Pt-Pt	4.7 ± 0.3	3 ± 1.5	-0.3 ± 0.1	~500	1-100M
TiN-TiN	3.70 ± 0.58	0.8 ± 0.2	-0.8 ± 0.1	~600	10k-100k
TiN-Pt	3.75 ± 0.82	0.7±0.2	-0.7±0.1	~600	10k-1M
TiN-Ti	2.7 ± 0.52	0.51 ± 0.15	-0.48 ± 0.1	~800	6k-20k

2.3 Waveform

As discussed in the previous paragraph, all the results were obtained in Quasi-Static conditions. This means that the applied voltage sweeps were so slow to be considered as steady state, thus considering only the influence of the voltage variable and neglecting the other one: time. In this paragraph, we analyze the effects of different voltage waveforms on the switching characteristics of the memory cells. Before discussing the results, an overview on the setup used to this aim is needed.

2.3.1 Experimental Setup

The experimental set-up is sketched in Figure 2.6. Voltage waveforms are applied to the BE of device under test (DUT) through the MOSFET M1 and displayed on channel 1 of an oscilloscope (LeCroy WaveRunner 44Xi). Current (I_{DUT}) flows through the DUT and a reference resistance (R_{REF}) that is in series with the memory cell. The voltage drop on R_{REF} is amplified by a factor G and acquired on channel 2 (V_{CH2}), thus being:

$I_{DUT}=V_{CH2}/(G\cdot R_{REF})$. In order to limit the current during SET, the voltage drop on R_{REF} is monitored: when it reaches the critical value V_{REF}, externally set, the MOSFET M1 limits current to the value: $I_{COMPL}=V_{REF}/R_{REF}$.

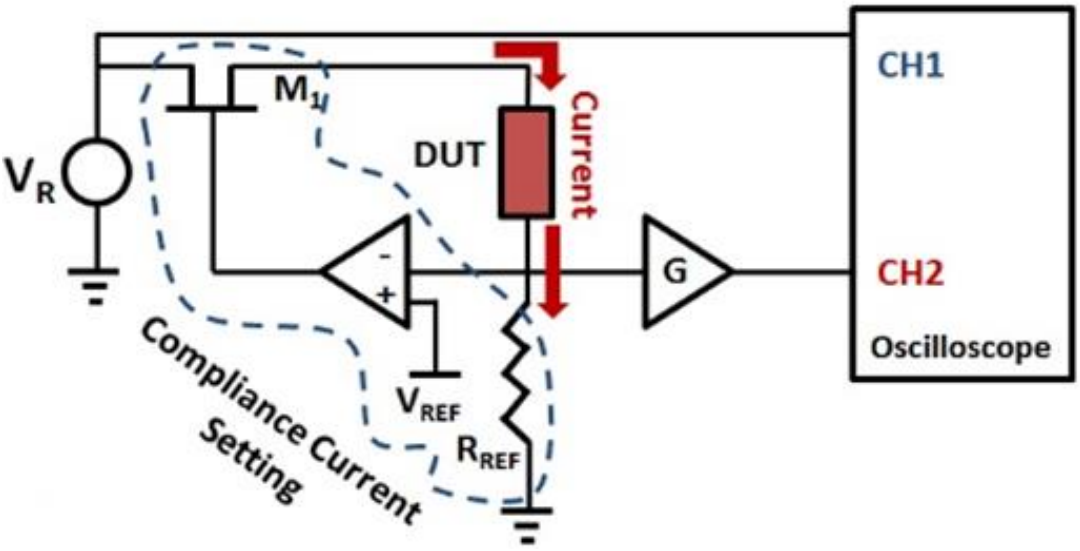


Figure 2.6 Experimental setup. Highlighted the current compliance limitation part

This kind of external current limitation was preferred to the insertion of just an external series resistance because it prevents voltage partition during the SET ramp between the external resistance itself and the device resistance which changes during the ramp. Using $R_{REF}= 24\Omega$, the partition is always negligible compared to the RRAM LRS values, as shown in table 2 for bipolar samples, and the current measurement is reliable.

Examples of waveform traces are shown in Figure 2.7 where blue marks are the applied voltage waveform and the current is in red. We can observe the current limitation imposed by the circuit during the SET operation. In particular, the current compliance was set to 1 mA in this case.

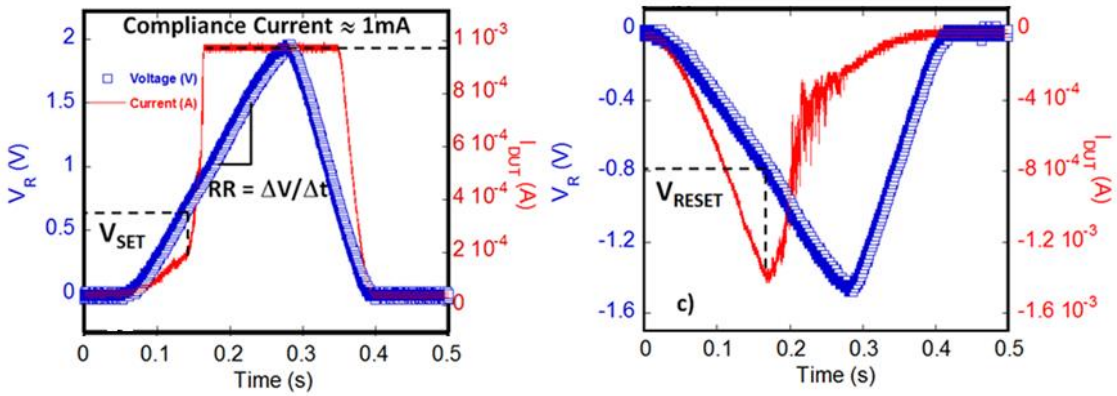


Figure 2.7 Example of applied voltage (thick blue line with open squares) and RRAM current (thin red line) during SET (left) and RESET (right)

2.3.2 Ramp Rate influence on SET and RESET

The influence of the voltage ramp speed (rate) defined as the ratio $RR=\Delta V/\Delta t$ on the switching voltage was investigated [95]. In these experiments, V_{SET} is defined as the voltage corresponding to the increase of the current slope and V_{RESET} is defined as the voltage corresponding to the maximum current (Figure 2.7). Data of V_{SET} and V_{RESET} vs RR were collected over a set of 15 samples of two different kind of stacks: TiN/HfO_2/Pt (denoted with C1) and TiN/HfO_2/Ti (denoted with C2). The averages values are displayed in Figure 2.8. As expected, both $|V_{RESET}|$ and V_{SET} increase with the front speed following a logarithmic behaviour [96]:

$$V_{SET/RESET}(RR) = a_{S/R} + b_{S/R} \cdot \log(RR/c), \qquad 2.1$$

where $a_{S/R}$, $b_{S/R}$ and c are constants whose values are determined after interpolation of experimental data (dashed lines in Figure 2.8). It is worth noticing that the two curves are parallel (with $b_{S/R}$~$0.025\ V$ and $c=4.925\ V/s$). Actually, the switching process is due to the concurrence of more than one mechanism, first of all the ion drift diffusion both for

SET and RESET [59], but also the oxygen bond breaking for SET, oxygen vacancies recovery with the oxygen ion release from the interface for RESET [64,97]. The fact that in Figure 2.8 the dynamics for SET and RESET is the same can be attributed to the predominance of the ion drift diffusion over other mechanisms, for both the closure and dissolution of the conductive filament.

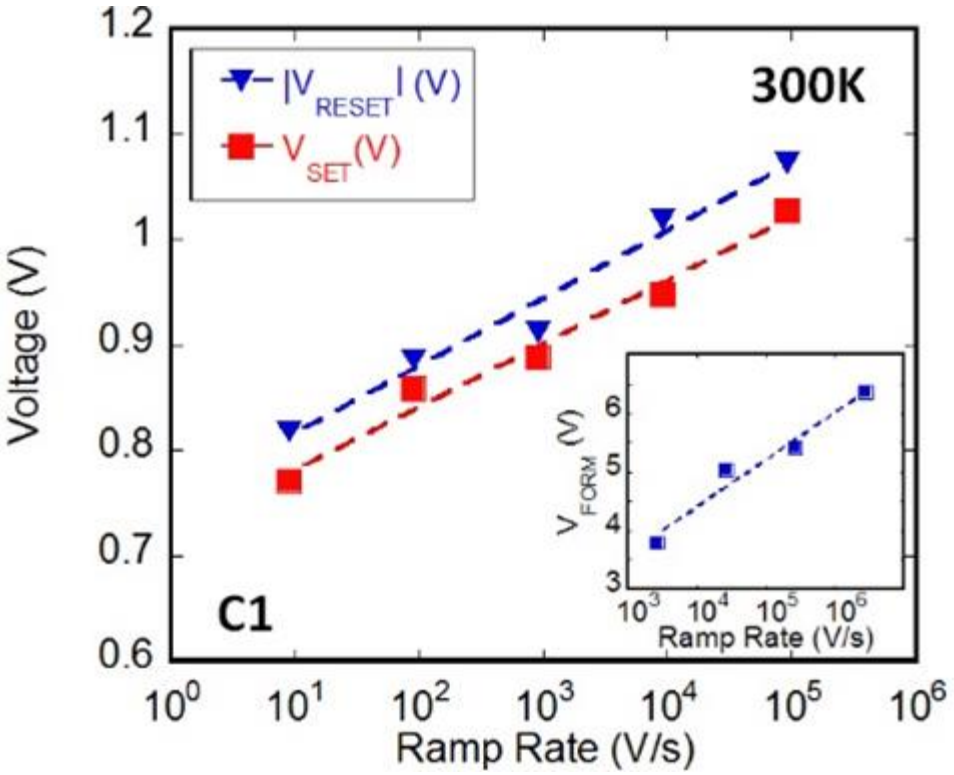


Figure 2.8 (Color online) $|V_{RESET}|$ and V_{SET} vs RR (symbols) for C1 samples. Inset: V_{FORM} vs. RR. Lines are logarithmic interpolations.

As expected, also V_{FORM} behaves logarithmically with the ramp rate, but with a different dynamics, as shown in the inset of Figure 2.8, since in this case the whole filament from one electrode to the other needs to be formed, while in SET just a portion of it must be completed.

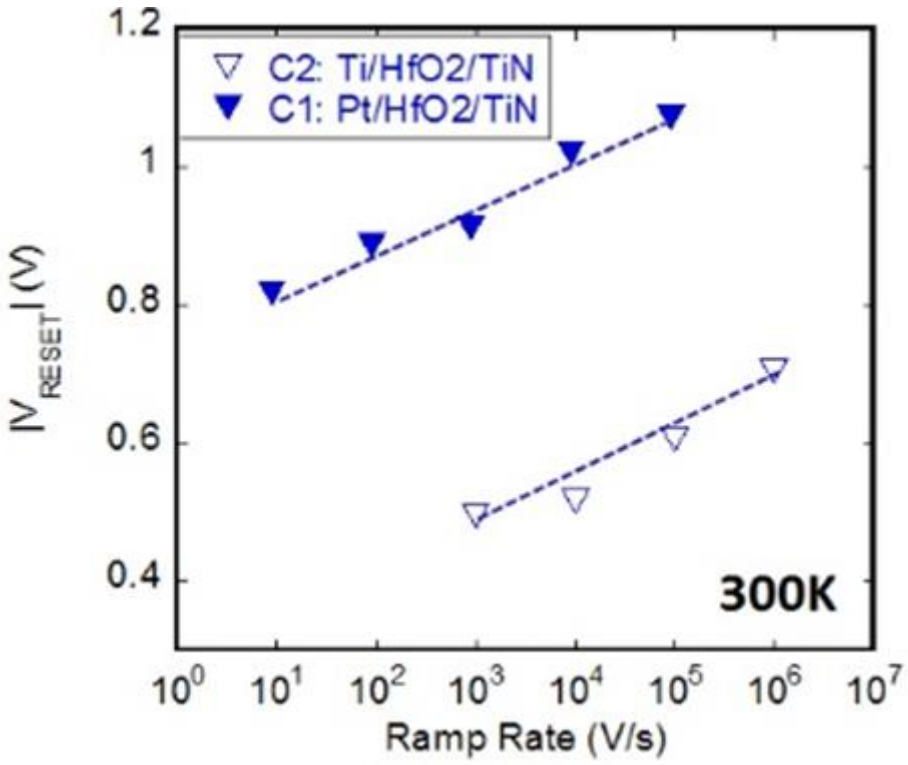


Figure 2.9 (Color online) V_{RESET} as a function of RR for samples C1 (closed symbols) and C2 (open symbols). Lines are logarithmic interpolations

In Figure 2.9 the curves of $|V_{RESET}|$ for the C1 (with Pt as TE) and C2 (with Ti as TE) samples are compared as function of the RR, at room temperature. A few considerations can be done: first, the C1 curve stands at higher values than the C2 curve; second, both curves increase with the ramp rate and, third, they are parallel. Recalling the probability P_R of release of oxygen O^{2-} ions from the electrode defined in [97] and substituting the electrode oxygen release time dt with the expression holding in our experiment $dt=|dV|/RR$ it can be written:

$$P_R = f \cdot \frac{|dV|}{RR} \cdot e^{-\frac{E_i - \gamma Ze|V|}{K_B T}}, \qquad 2.2$$

where f is the vibration frequency of oxygen atoms, $Z = 2$ is the charge number of oxygen ions, e is the electron charge, T is the local temperature, K_B is the Boltzmann constant, E_i is the energy barrier between the electrode and oxide and γ is the enhancement factor of the external voltage during the O^{2-} release process. The quantity $(E_i - \gamma Ze|V|)$ is positive and decreases increasing voltage, with the consequent exponential increase of the release probability. On the other hand, the probability reduces linearly with increasing RR. Therefore, increasing RR the voltage should be slightly increased in order to get the same probability, as evidenced in Figure 2.9. The barrier energy E_i to extract oxygen ions from the TiON layer is higher than from the TiO_x layer. This affects the exponential term in the release probability P_R, favoring C2 samples. For this reason, and also recalling that the effective dielectric thickness of C2 samples is lower than in C1 samples, a greater voltage is needed in C1 samples to switch. Finally, the two curves of RESET voltage in Figure 2.9 are logarithmic and parallel, indicating that the dynamics of filament interruption is the same, governed by the same hopping barrier [59] of ion diffusion, and is not influenced by the electrode.

2.3.3 Pulsed Forming

Pulsed measurements at very short time scale were performed to get deeper insight in the Forming dynamics [51]. Voltage pulses with 10 ns rise time (T_{RISE}) and amplitude (V_P) were used. The circuit used in this experiment is sketched in Figure 2.14 (a). A simplified version of the circuit was adopted because in Forming characterization an external resistor is sufficient to obtain current compliance limitation without sensible partition of the applied voltage. This is due to the very high resistance value of pre-formed devices. Pulses were displayed on channel 1 (CH1) of an oscilloscope. The

voltage drop on a resistor in series to the device under test (DUT) was acquired on channel 2 (CH2) of the oscilloscope. R_L=10 KΩ fixed the current compliance to V_P/10 KΩ.

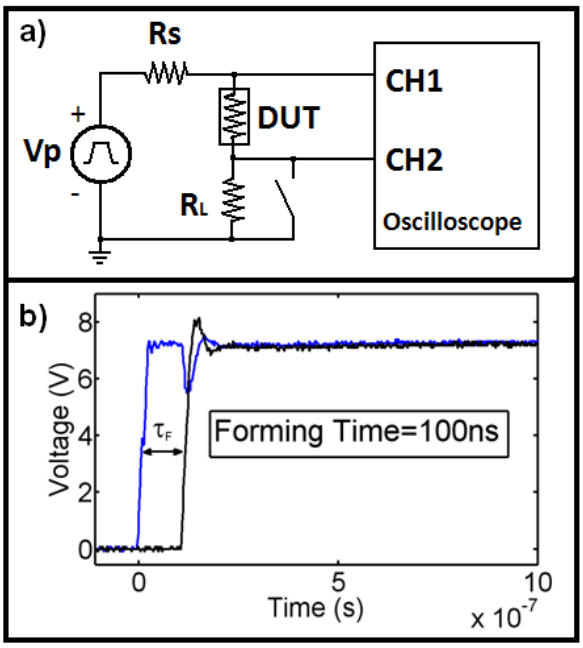


Figure 2.10 (a) Sketch of the simplified setup used for measuring forming time and resistance; (b) waveforms detected on CH1 and CH2 of the scope.

The typical waveforms acquired on CH1 and CH2 are sketched in Figure 2.10(b). The signal displayed on CH2 allowed measuring the Forming time τ_F. Data of τ_F for Pt-Pt (S1), TiN-Pt (S2) and TiN-TiN (S3) samples were collected at room temperature. In the inset is shown the distributions of forming voltage for S2. The average value τ_F was then used to plot $V_P \cdot \tau_F$ against V_P in Figure 2.11. An exponential dependence was found for all the samples (dashed lines in Figure 2.11).

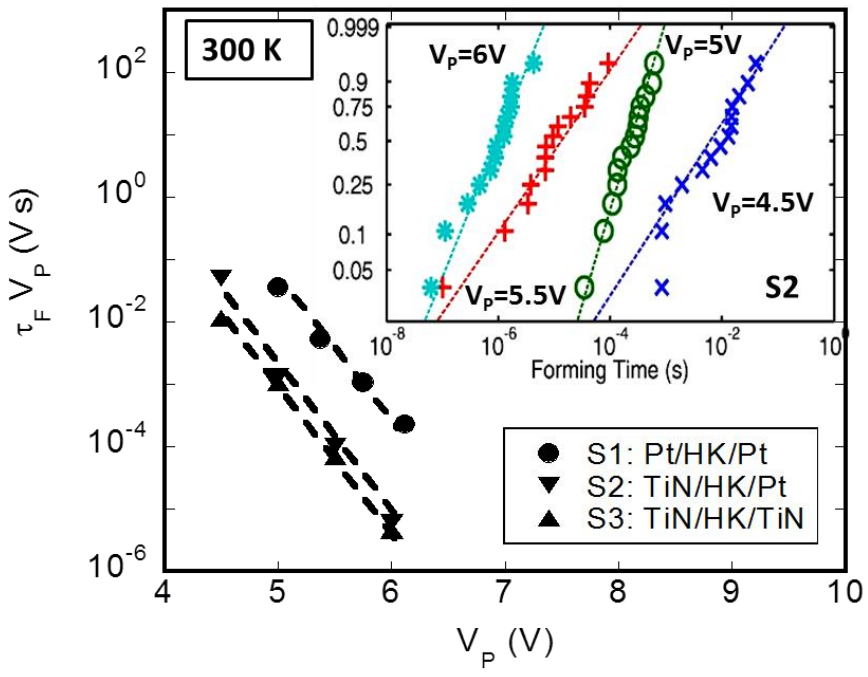


Figure 2.11 Data of forming time (average value) varying the voltage pulse amplitude for S1, S2, S3 samples. Markers are experimental data, line is interpolation. Inset: Weibull distributions of one set of samples (S2).

As a result, we say that voltage scaling is limited by the tradeoff with the forming time. As an example, in samples S1 characteristic time of forming is 200 μs at 6.12 V, and grows up of two orders of magnitude with a voltage decrease of only 18% (30ms at 5V). The parameter relating the forming time with voltage in the exponential term is the most important for the fast programming with low voltage. Forming times for S2 samples (featuring TiN as BE and Pt as TE) and S3 samples (featuring TiN as BE and also as TE) is approximately the same. At the same time, the S2 and S3 curves lie at times shorter respect to the S1 curve (featuring Pt as BE and TE). This indicates that the chemical reactivity of the growth electrode (rather than the TE) plays a main role in the forming kinetics. The measured shorter τ_F in the case of TiN as BE can be

explained with the formation of a TiON layer with consequent enrichment of oxygen vacancies at the bottom interface, which accelerates creation of percolative paths.

2.4 RESET Dynamics

Alternatively a different approach can be considered to model the influence of the applied voltage waveform on the I-V characteristic, starting from the consideration that the filamentary electron transport in resistive switching HfO2-based metal-insulator-metal structures can be modeled using a diode-like conduction mechanism with a series resistance [98,99]. A number of conduction models have been invoked to explain the *I–V* curves after the initial forming step, such as Schottky emission, Poole-Frenkel emission, tunneling and point-contact conduction, among others [73, 100-102]. All these conduction mechanisms have associated to an exponential *I-V* curve for the HRS compatible with a diode-like expression of the type: $I=I_0*[exp(\alpha V)-1]$, where I_0 and α are model constants related to the particular conduction mechanism considered. The inclusion of a series resistance allows simulating the linear *I-V* curve associated with the LRS, under the appropriate limits. It is possible to show that this simple equivalent circuit composed by a diode and a series resistance can also describe the progressive reset dynamics if the pre-exponential diode current factor is properly modulated. The modulation is related to the sequential deactivation of conductive channels, which is mathematically expressed in terms of a generalized logistic growth model.

According to previous reports [103], the filamentary current in RS devices can be modeled using an equivalent electric circuit formed by a diode and a series resistance *R*. In this case, the current is:

$$I(V,\lambda) = I_0(\lambda)\{\exp[\alpha(V - IR)] - 1\}, \quad 2.3$$

where I_0 is the pre-exponential diode amplitude factor and α a parameter related to the slope of the curve. Assuming that I_0 can be expressed as a function of the normalized number of active channels $0 \leq \lambda \leq 1$ spanning the dielectric film it is:

$$I_0(\lambda) = I_{0\min}(1-\lambda) + I_{0\max}\lambda, \quad 2.4$$

where I_{0max} and I_{0min} are the maximum and minimum values of I_0, respectively. Combining equations 2.3 and 2.4 it is possible to extract λ from the experimental *I-V* curves as:

$$\lambda(V) = \frac{I(V,\lambda) - I(V,\lambda=0)}{I(V,\lambda=1) - I(V,\lambda=0)}, \quad 2.5$$

The values of I_{0max}, I_{0min}, α and R are chosen so as to fit simultaneously the LRS and HRS *I-V* curves.

Reset I-V curves averaged over 30 cycles and over 4 different device of the type Ti/HfO2/Ti for 3 different ramp rates (100V/s, 1000 V/s, 10000 V/s) are shown with marks in Figure 2.12. Each curve was obtained averaging data from 30 reset curves. The black solid lines represent the stationary HRS/LRS currents used to compute with equation 2.5. In order to correlate λ with the number of active conductive channels at a voltage *V*, a simplified approach is considered first. Within this context, the idea of a filamentary path is nothing but a bunch of conductive channels of atomic dimensions. We assume that during the reset process the channels are sequentially deactivated following a Gaussian dependence with the applied voltage:

$$f(V) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(V - V_R)^2}{2\sigma^2}\right], \qquad 2.6$$

where V_R is the average reset voltage and σ^2 the standard deviation. Notice that equation 2.6 should not be confused with the reset voltage distribution obtained from cycle to cycle.

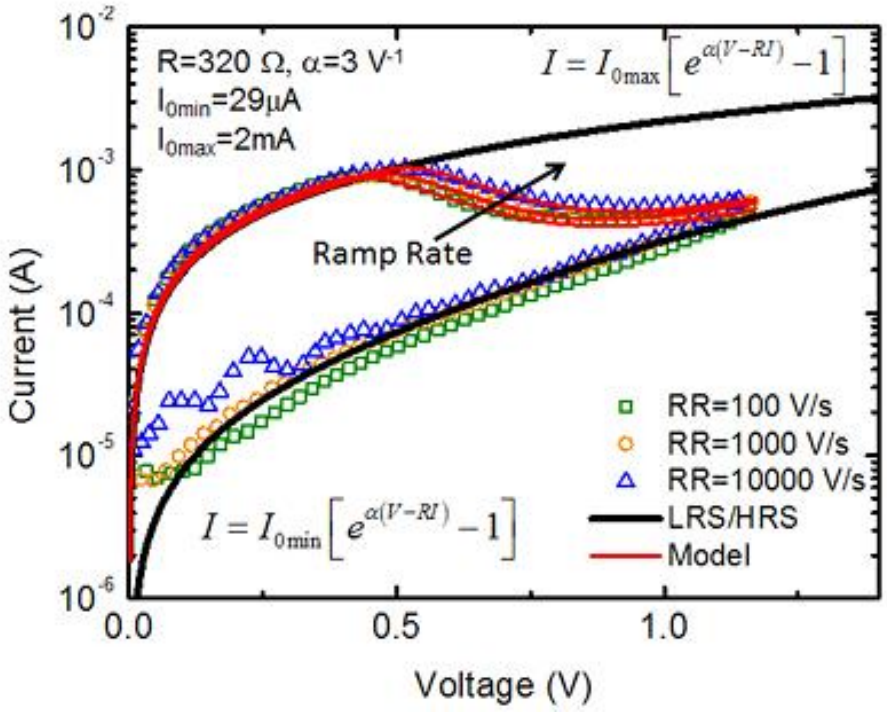


Figure 2.12 Reset I-V curves averaged over 30 cycles and over 4 different device for each ramp rate. Black lines are the calculated current in the case of $I_0=I_{0max}$ and $I_0=I_{0min}$. Red lines are calculations during the transition for the 3 different ramp rates.

The convolution of *f(V)* with the Heaviside step function *H* gives the normalized number of active conductive channel as a function of *V*:

$$\lambda(V)=\int_{V}^{\infty}H(V-\xi)f(\xi)d\xi=\frac{1}{2}\left[1-erf\left(\frac{V-V_R}{\sqrt{2}\sigma}\right)\right]\approx\{1+\exp[\eta(V-V_R)]\}^{-1}, \qquad 2.7$$

where *erf* is the error function and $\eta=1.702/\sigma$ [104]. The integral in equation 2.7 represents the area under the Gaussian curve from a voltage V to ∞, which can be approximated by a simple logistic function.

Experimental values for $\lambda(V)$ are calculated with equation 2.5 and shown in Figure. 2.13 with symbols. Dashed lines are the calculated curves using the logistic expression in equation 2.7. As shown in the inset of Figure 2.13 the logistic curves seem to fit very well the experimental results, but a closer inspection reveals that the calculated curves largely depart from the experimental data for a low number of conductive channels. This is outlined by the logarithmic scale of Figure 2.13. In order to improve the description of the conductive channels deactivation process, a more complex dynamics is considered next.

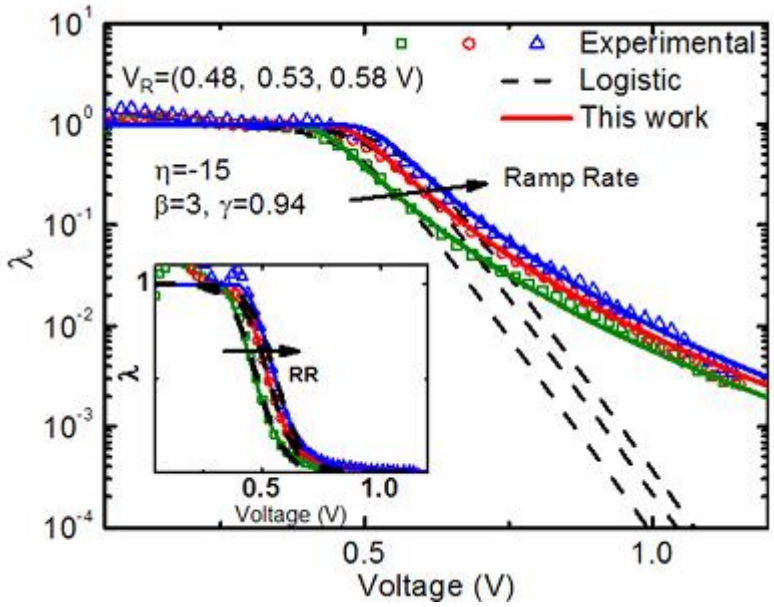


Figure 2.13 Cumulative distribution of normalized conductive channels as a function of the applied voltage in logarithmic scale. Markers are values extracted trough equation

2.5. Dashed lines are the logistic curves. Solid lines are the generalized logistic curves calculated with equation 2.9. In the inset data and calculations are shown in linear scale.

A variety of models have been proposed in literature to interpret specific growth dynamics with limited resources[105-107]. Most successful predictive models are based on extended forms of the classical Verhulst growth equation, a standard tool of population dynamics [108 - 110]. A generalized form of the logistic growth model was introduced by A. Tsoularis [111] as:

$$\frac{d\lambda}{dV} = \eta\lambda^{1+\beta(1-\gamma)}\left(1-\lambda^{\beta}\right)^{\gamma}, \qquad 2.8$$

where η, β and γ are fitting constants related to the asymmetry of the λ-V curve. β and γ are positive numbers which satisfy the condition [23_APL_2015] $\gamma < 1+1/\beta$. We use this model to represent the CC rupture dynamics in our RS devices. First of all, we notice that $\beta = \gamma = 1$ in equation 2.8 corresponds to the logistic curve obtained in equation 2.7. Then, $\gamma = 1$ was considered by other authors [112] in order to represent the window function in the state equation of nonlinear memristors. The differential equation 2.8 has the analytic solution:

$$\lambda(V) = \left\{1+\left[\eta\beta(\gamma-1)(V-V_R)\right]^{\frac{1}{1-\gamma}}\right\}^{-\frac{1}{\beta}}, \qquad 2.9$$

V_R is introduced here to account for the shift of the reset voltages obtained with the three different ramp rates. For negative values of the argument (i.e., for $V<V_R$) $\lambda=1$ is posed. The solid lines in Figure 2.13 were calculated using equation 2.9. They exhibit an excellent agreement with the experimental data in the whole investigate range. Looking at the experimental curves in Figure 2.13, it can be observed that higher sweep

rates imply larger reset voltages. In this connection, a logarithmic dependence of the reset voltage shift with the ramp rate was found [95]. Remarkably, the transition dynamics remains essentially the same as illustrated in Figure 2.14. Following Cagli *et al.* [113], we assume that the reset probability per unit time dP_R/dt scales exponentially with the applied voltage according to:

$$\frac{dP_R}{dt} = \frac{1}{\tau}\exp(\theta V), \tag{2.10}$$

where τ and θ are constants. Since $RR=dV/dt$, integrating eq. (8) leads to:

$$V_R = a + b\ln(RR). \tag{2.11}$$

where V_R is the reset voltage corresponding to the maximum current and a and b are fitting constants. It is worth noticing that if the translation expressed in equation 2.11 is used the three curves with different RR are superimposed. This is shown in Figure 2.14. This result highlights once again that the dynamics ruling the conductive channels deactivation process is not depending on the waveform features. Keeping in mind that λ is associated with the activated paths, $d\lambda/dV$ given by equation 2.8 expresses the distribution of the channel deactivation voltages. It is clear that this distribution is more complex than the Gaussian distribution initially assumed in 2.6. Figure 2.12 suggests an asymmetric distribution with a longer tail towards the higher voltages range.

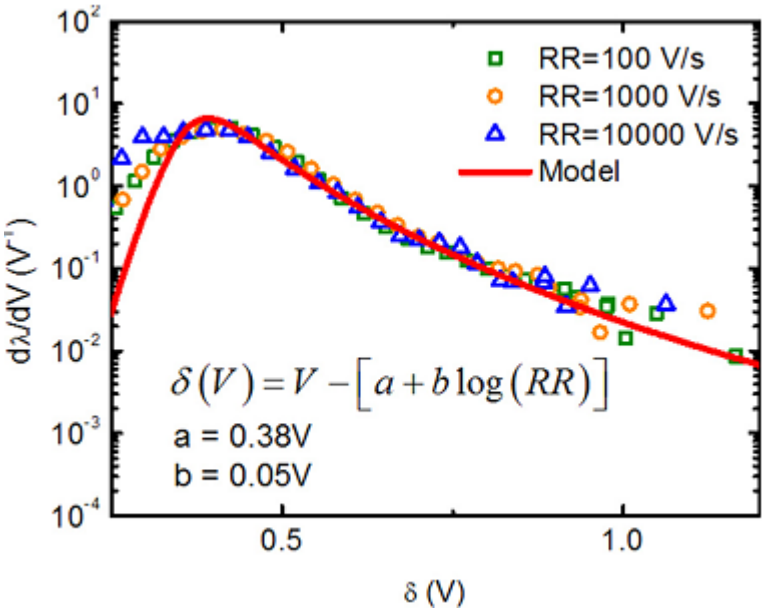


Figure 2.14 Derivative of λ as a function of δ(V). Markers are experimental data, and lines are calculated values.

As a final demonstration of the proposed approach, partial reset *I-V* curves were investigated. In this case the measurements correspond to a sequence of triangular voltage ramps with ramp rate RR=10^3 V/s. The maximum voltage (V_{STOP}) was increased in steps of 100 mV starting from 0.6 V and ending at 1.2 V. The measurement was repeated 30 times and the average value of the current is shown with symbols in Figure 2.13 a) as a function of the sweep time. Equations 2.3, 2.4 and 2.9 were used to calculate the current during the partial reset sweeps. In order to account for the hysteretic behavior, the last value of λ obtained at the end of each triangular waveform was used as the starting value for the subsequent sweep.

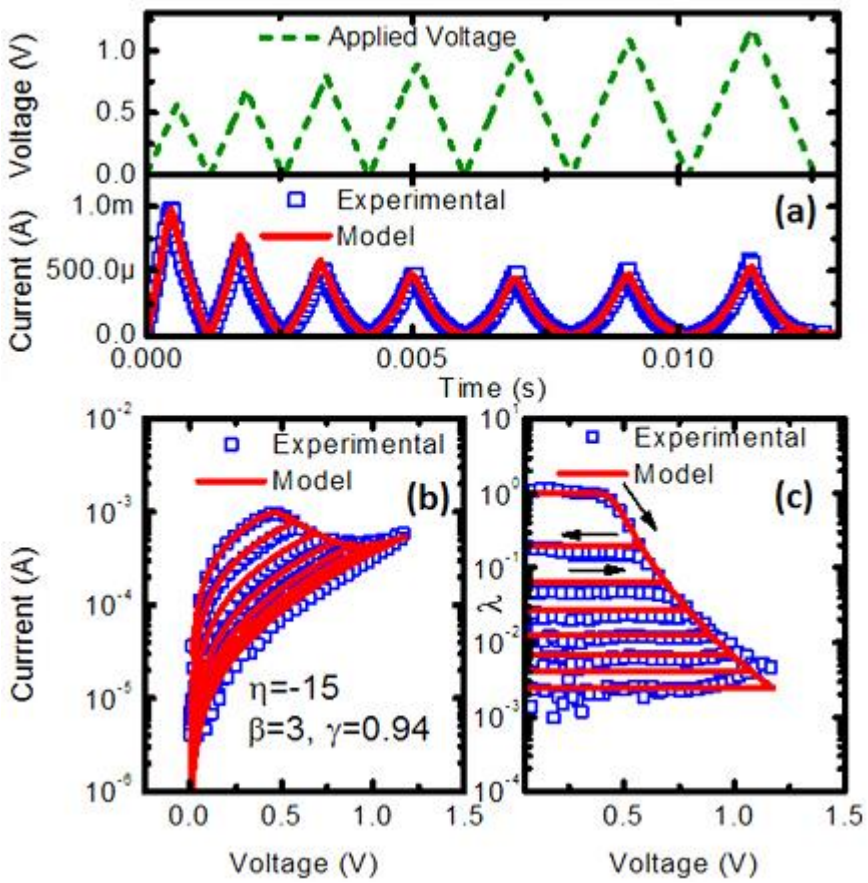


Figure 2.15 (a) Partial Reset voltage waveform (dashed line), measured current (markers), and the calculated current (solid) are shown. Data were obtained averaging 30 cycles; I-V characteristic (b) and normalized number of activated channels λ(c) are also shown.

In Figure 2.15 b) the complete *I-V* characteristic was also reconstructed using this approach. During partial reset at the end of every single triangular sweep a fraction of the total normalized number of active channels was achieved following the generalized logistic curve (Figure 2.15 c). The change of the state of the RS device is permanent in the sense that after each successive sweep the initial current of the cell is the same measured at the end of the previous one. These results point out that our RS devices exhibit multiple memory states.

The progressive reset transition between the low and high resistance states in HfO_2-based RS devices can be modeled using a diode-like conduction mechanism with a series resistance combined with a generalized logistic model for the pre-exponential diode amplitude factor. The dynamics has been interpreted in terms of the sequential deactivation of multiple conduction channels spanning the dielectric film. Fitting results indicate that the switching behavior dynamics can be described with the same equation regardless of the voltage sweep rate. Finally, partial reset curves were characterized and modeled using the same approach.

2.5 Device Geometry

The influence of the device geometry on switching behavior was investigated. In Figure 2.16, RESET/SET voltages (black dots and lines) and LRS/HRS (red dots and lines) data were reported as a function of cell dimension. As one can see, no influence of cell scaling is observed in both the kind of data. This clearly confirms switching being driven by generation-disruption of a conductive filament.

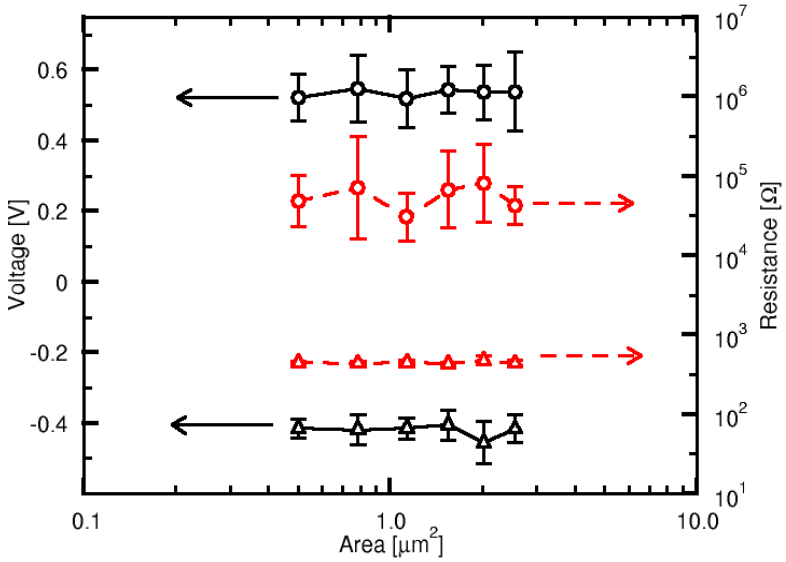


Figure 2.16 Area dependence of Set and Reset voltages (left coordinate axis) and set and reset resistances (right coordinate axis) for TiN-Ti

While the device area size has no influence on the SET and RESET characteristics, it has a strong influence on the Forming operation.

The forming time was monitored, with respect to device area (A) in the range 0.125 μm^2-2.544 μm^2. This investigation was performed only on Pt-Pt (S1) samples. Results (each point is an average on 20 samples) are shown in Figure 2.17 for different values of the pulse amplitude.

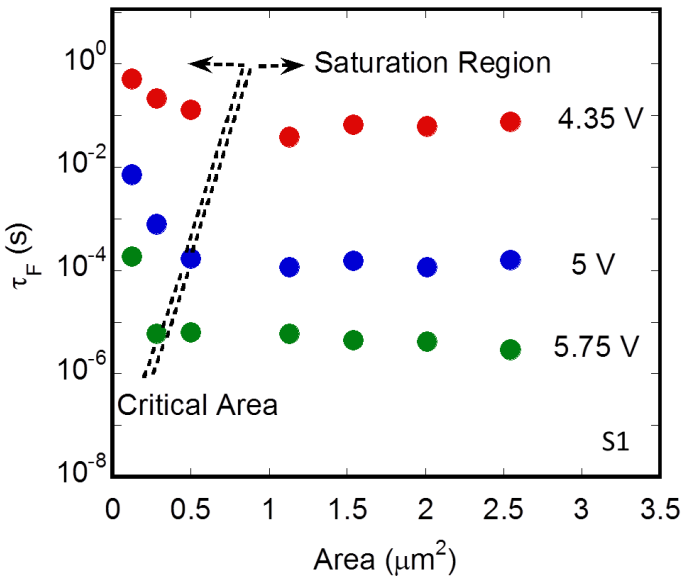


Figure 2.17 Forming time vs area for three pulse amplitudes (samples S1).

As a result, it was found that above a critical area (A_C) τ_F tends to saturate. The asymptotic value of τ_F is higher at lower values of V_P, in agreement with previous results. Below A_C, we find the expected behavior: the smaller the area, the lower the number of traps, the higher the time required for the filament forming. A_C is inversely related to V_P. This fact is a concern for RRAM device scaling, since the contemporary

scaling of area and voltage implies a strong increase of forming time. To explain the behavior of τ_F at large areas, we can consider that after increasing the area the probability to form a filament tends to unity. Probably more filaments add in parallel, but this has no impact on the time for forming. In the investigated case (Pt electrodes) values of A_C are greater than what is foreseen by the ITRS for the next node of nonvolatile memory, which implies that the region of interest for practical applications is in the left side of Figure 2.17, where the forming time increases with area scaling. The dependence of forming voltage on the cell size in 5nm thick Ti/HfO_2/TiN MIM device was reported in [114]. As a result, in that work a drastic increase of breakdown voltage was found below a characteristic size. This is in complete agreement with the results shown.

2.6 Dependence of Forming and RESET on Temperature

The temperature dependences of switching parameters are important in order to highlight the physical mechanisms governing the behaviour of cells. First of all, the forming dynamics vs temperature was investigated. Data of τ_F were collected over the Pt-Pt (S1), Pt-TiN (S2) and TiN-TiN (S3) samples. The devices temperature (T) was increased from room temperature (300K) up to 473 K. During the experiment the forming pulse amplitude V_P was varied.

Here we assume that forming time is thermally activated in an Arrhenius dependence:

$$\tau_F \propto e^{-\frac{E_A}{kT}}, \qquad \textbf{2.12}$$

where E_A is the activation energy, k the Boltzmann constant and T the temperature.

Data of τ_F/T versus 1000/T (and V_P) are graphed in Figure 2.18 for samples S1, S2 and S3. Dashed lines are the interpolations obtained with equation 2.12.

The only parameter which varied between the different samples was the hopping activation energy. Average values of E_A extracted from fits are 0.72 eV for S1 samples and 0.56 eV for S2 and S3 samples.

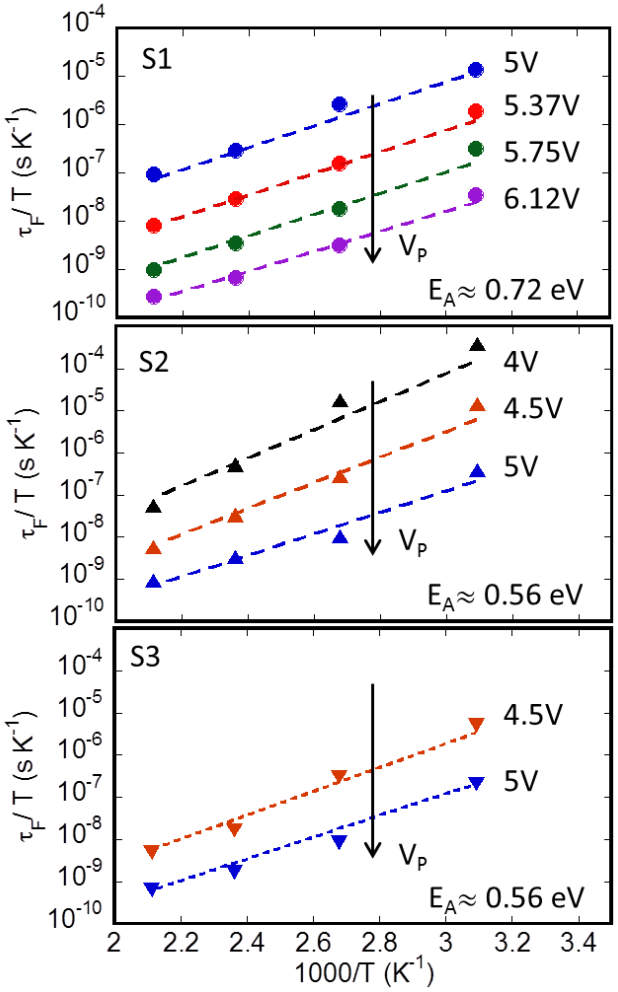


Figure 2.18 Data of forming time varying the temperature for S1, S2 and S3 samples. Markers are experimental data, line is interpolation.

The hopping energy for samples with a bottom interface layer rich of oxygen vacancies is lower than that for samples with Pt electrode, highlighting again the different

physical mechanism governing the switching of the unipolar (S1) and bipolar (S2-S3) devices.

Considering the RESET operation, voltage and current at the RESET switch were monitored increasing the measurement temperature and the conductance just before the RESET switch was calculated as $G=I_{RESET}/V_{RESET}$. Data of G are plotted as function of the inverse temperature in Figure 2.19 for 3 Ramp Rate: 100 *V/s* 1000 *V/s* and 10000 *V/s*. As one can see, the conductance of HfO_2 films in C1 (Pt-Pt) samples is lower than in C2 (TiN-Ti) ones and increases with temperature, while the conductance of C2 samples does not change with temperature. The higher conductance of C2 samples can be explained recalling that the effective thickness of the dielectric film is reduced respect to C1 samples and considering the different microscopic phase of the HfO_2 in the filament for the two kind of samples. The different sensitivity of conductance to the measurement temperature can be ascribed to the microscopic phase of the HfO_2 too. Interpolating data of conductance with the usual expression of conduction holding for semiconductors one finds out a much lower value of the activation energy in C2 samples than in C1 ones (shown in Figure 2.19). This picture allows to figure out the fact that conduction is so favored in C2 that increasing temperature does not yield any appreciable enhancement. (As a reference, the same model of LRS conductance was applied in ZnO films extracting an activation energy of 0.4 eV [115]). On the other hand, the presence of temperature activated mechanism of conduction could be also taken into account to explain the sensitivity of conductance to temperature. As an example, the LRS conductance in SiO_x films was recently interpreted in terms of Poole-Frenkel (PF) emission from traps [116]. In principle, an emission model could be

applied to the case of C1 samples, but not in the case of C2 samples, where the step of HT annealing could have reduced the trap density. However, three temperature values are not enough to allow a meaningful extraction of the trap features and for this reason any emission modeling is not carried on.

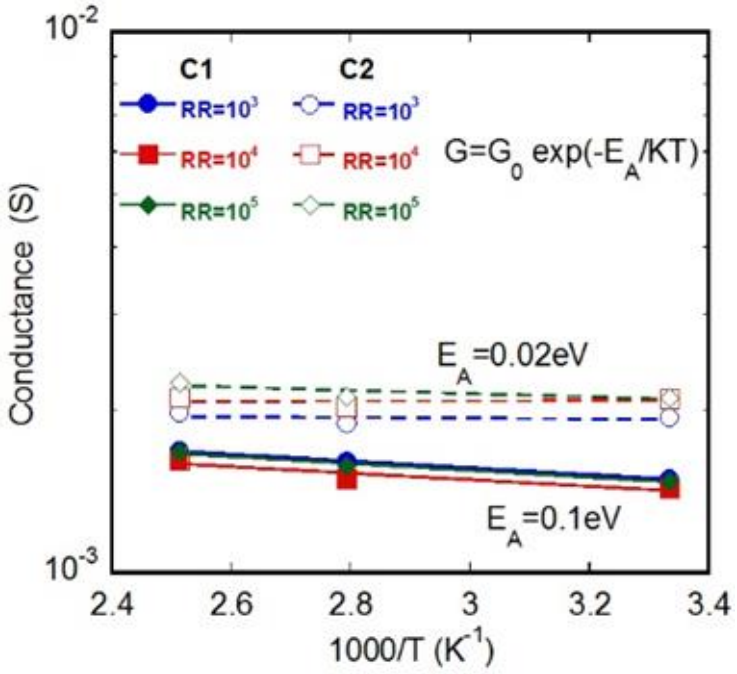


Figure 2.19 G in the LRS state vs temperature in C1 (closed symbols) and C2 (open symbols) samples, for three different ramp rates.

2.7 Constant Voltage Stress

In this paragraph aspects related to the field of reliability are examined. In particular the case of electrical stress. An experiment of read disturb is performed on HfO_2 based resistive memories [117]. The experiment elucidates the role of the forming conditions and the electrode materials on the robustness of the low resistance state against electrical disturb. It is performed in three steps: 1) two sets of samples with different electrodes (TiN and Pt) are formed using voltage pulses with different amplitudes; 2) formed

samples are subject to constant voltage stress of different entities; 3) the current flowing through the MIM during stress is monitored and processed.

The complete setup used is sketched in Figure 2.20. In Figure 2.20(a) the four pads of a device are drawn (TE and BE), but it should be noticed that only two terminals are connected during measurements: during forming time measurements only the two probes on the left are connected, while for the stress measurements the probes on the right are connected. In b) the circuit used for measuring the forming time (τ_F) is shown. In this case, forming was achieved in pulsed condition as discussed in paragraph 2.3.3. Typical waveforms are displayed in Figure 2.20 c). The right side of Figure 2.20 regards the stress measurements.

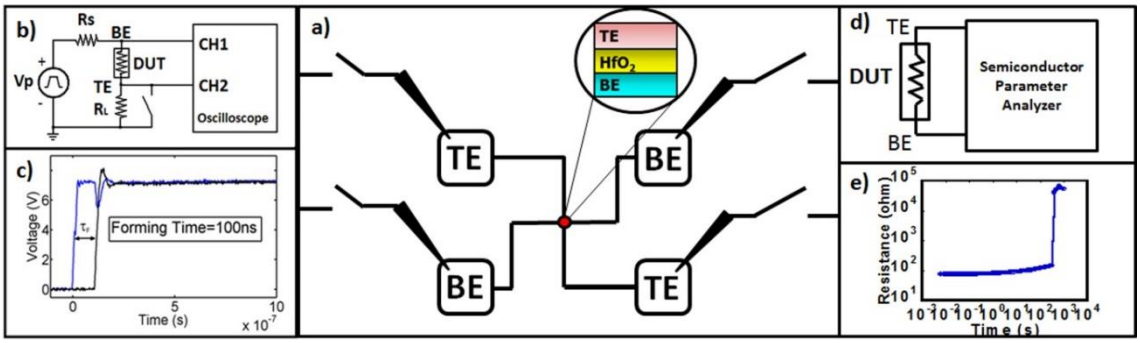


Figure 2.20 Experimental set-up: (a) Sketch of the pads of the device under test (DUT); b) circuit used for measuring forming time; (c) waveforms detected on CH1 and CH2 of the oscilloscope; d) circuit used for the extraction of the resistance during the stress; e) an example of the evolution of the resistance during the stress

We investigated the stability of the LRS when an electrical read disturb is applied. To the aim, all the samples did undergo a Constant Voltage Stress (CVS) featuring the same polarity of the RESET, but lower amplitude (V_{STRESS}). In Figure 2.20(d) it is displayed the electrical connection between the semiconductor parameter analyzer and

the device for monitoring the current flowing during the stress. Current data are then processed to extract the resistance curve versus stress time, as shown in Figure 2.20(e).

Results obtained on TiN-Pt and Pt-Pt samples using a forming pulse featuring V_P= 6V are displayed in Figure 2.21. In that experiment TiN/Pt samples were stressed for 1000 s applying a CVS at -0.3V (since they are bipolar), while Pt/Pt samples were stressed with a CVS at +0.3V (since they are unipolar). This comparison is meaningful since the RESET voltage had the same absolute average value (+0.8 V in the Pt-Pt case and -0.8 V in the Pt-TiN case). As one can see resistance of the Pt-Pt samples underwent appreciable and progressive variations during the stress, whereas the TiN-Pt did not at all.

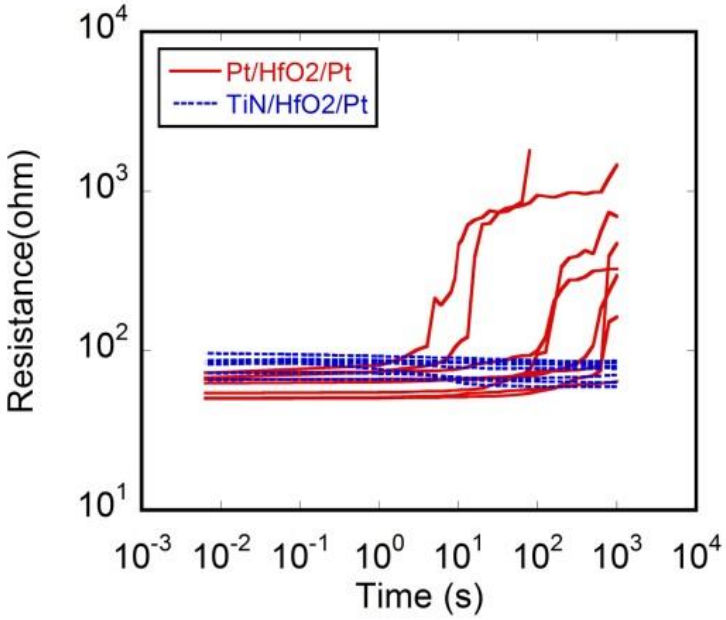


Figure 2.21 Comparison between TiN-Pt (blue dashed lines) and Pt-Pt (red solid lines) samples: resistance curves during stress. V_P = 6V, $|V_{STRESS}|$ = 0.3 V

Influence of forming in combination with the V_{STRESS} entity was analysed. Forming was performed using trapezoidal voltage pulses with the same duration (T_P = 500 ms) and

the same rise/fall time (T_{RISE}=50 ns). The pulse amplitude V_P was varied: V_P = 4 V, 5 V, 6 V. On the basis of the preliminary characterization reported in paragraph 2.3.3, that value of T_P is longer than the values of τ_F in TiN-Pt and Pt-Pt samples distributions, and we are confident that 1) the conductive filament close in stationary condition; 2) the traps in the conductive path have time enough to settle and re-arrange; 3) the current flows through the filament for a certain time, from its formation to the end of the 500 ms lasting pulse; 4) the temperature locally grows and the filament has time to eventually vary its size. For TiN-Pt samples, the entity of the stress was further increased, and results are shown in Figure 2.22 (V_{STRESS}= -0.6 V, -0.7 V).

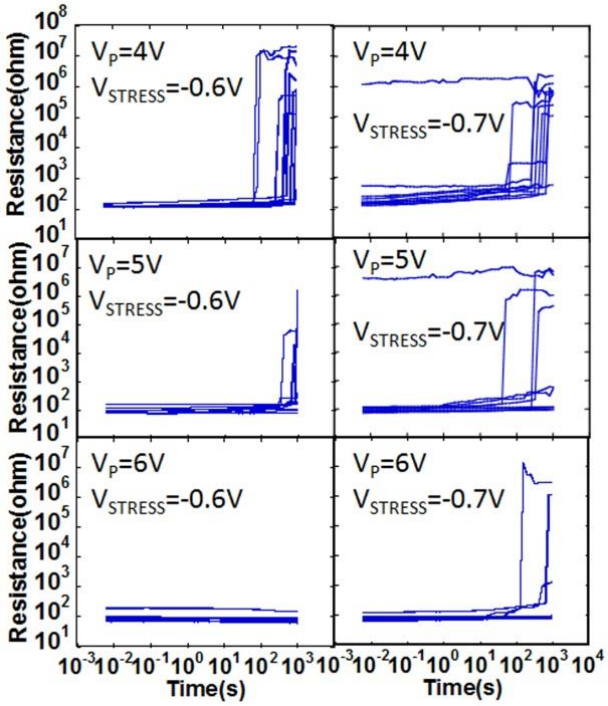


Figure 2.22 TiN-Pt samples: resistance curves during stress. V_P and V_{STRESS} are indicated inside the plots.

First of all, it is possible to notice that still with a 1000 s long CVS at -0.6 V the LRS looks rather robust. The number of samples showing an increase of resistance after a

certain stress time becomes lower increasing V_P (scroll down the columns in Figure 2.22). On the contrary, that number increases whit V_{STRESS} (scroll horizontally the rows). Devices exhibiting an increase of resistance by a factor 10x are addressed as failing ones. Stress of lower entities were also applied, but the percentage of samples exhibiting an appreciable increase of resistance was extremely low in any forming condition, and results are not reported.

Experimental data for TiN-Pt have been simply processed and results are resumed in the graph of Figure 2.23, where the percentage of TiN/Pt samples which has failed is drawn as a function of the forming voltage, for the two stress conditions.

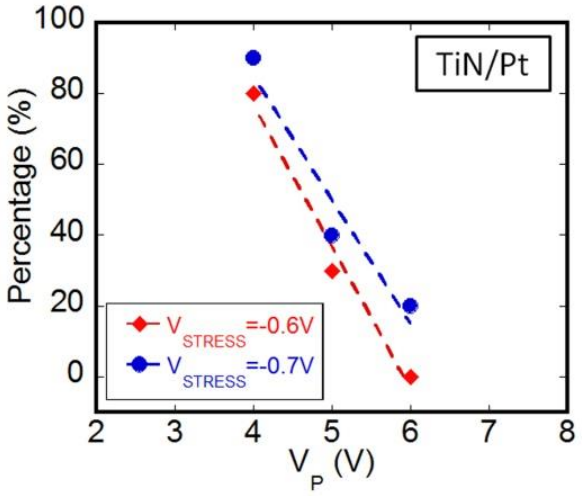


Figure 2.23 TiN-Pt samples: percentage of samples which failed during stress as a function of forming and stress voltages

The plot clearly outlines the higher robustness of the LRS to electrical disturb when forming is achieved with the higher pulse voltage. This can be due to the fact that using higher pulse amplitude the forming time is shorter as previously described. However, the pulse duration was fixed to 500 ms, independently of the voltage amplitude.

Therefore, in the case of higher V_P the percolation current flows between the two electrodes for a longer time, during which the filament size can eventually increase.

Looking at Figure 2.22 it can be noted that the forming voltage amplitude has an impact also on the value of the (high) resistance reached by devices at the end of the CVS. This is particularly evident for V_{STRESS}=-0.6 V.

In Figure 2.24 the box distributions of resistance at the end of the stress experiment (called R_{1000}) are drawn for the three forming conditions (stress at -0.6 V). The value of the initial resistance (R_0) is also indicated with dashed line, as a reference. Each box contains 50% of the occurrences (from 25% to 75%), the vertical bars indicate the tails, the horizontal line is the median value. As one can see, both the distributions and the medians tend to the value R_{1000}=R_0 with increasing V_P

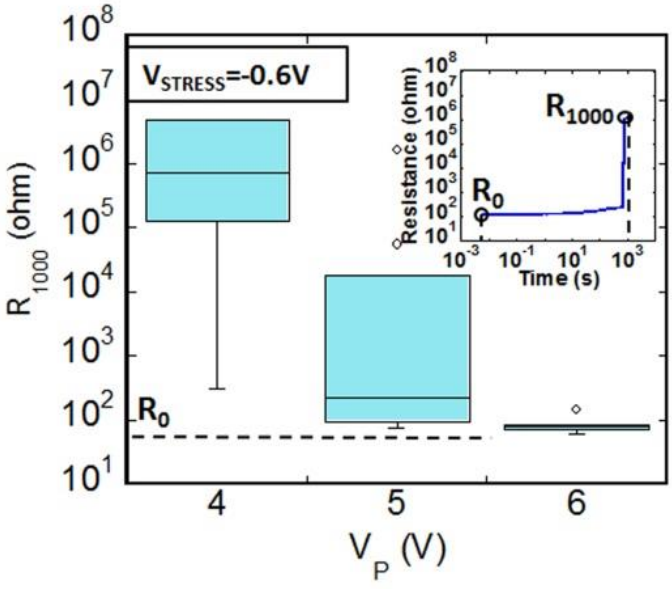


Figure 2.24 TiN-Pt samples: box distribution of the resistance measured after a 1000 s long-stress performed at -0.6 V.

All the data collected so far are coherent with two different pictures of the mechanism involved in the RESET operation in Pt-Pt and TiN-Pt samples. In Pt-Pt samples, the great and progressive sensitivity of the LRS samples to electrical disturb even at low entity can be related to a local heating of the filament during the stress with consequent migration of oxygen vacancies out of the filament [118,119]. This progressive mechanism is sketched in Figure 2.25a where, independently of the voltage polarity, the current flowing in the conductive filament may cause vacancy migration where the local temperature is higher. On the contrary, in TiN/Pt samples the voltage stress does not affect the LRS for stress of low entity. It seems that the stress voltage should be greater than a certain critical value (which depends on the forming condition), after which the LRS abruptly switches into the HRS.

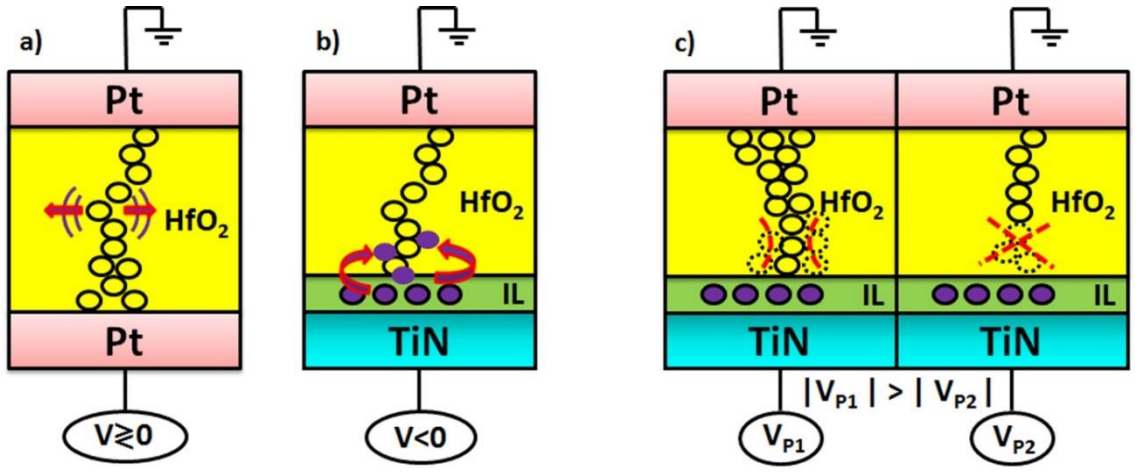


Figure 2.25 Sketch of filament rupture in Pt-Pt cells (empty circles are oxygen vacancies); b) recombination of oxygen vacancies with oxygen ions from the O-rich interface layer (blue circles) in TiN-Pt cells; c) different forming conditions give rise to different sizes of the filament: thicker filaments are more robust against rupture through vacancy-ion recombination.

This is coherent with a picture of the mechanism of RESET as related to the presence of a barrier at the IL/HfO_2 interface that the oxygen ions injected from the IL must overcome to recombine with vacancies in the filament [45,97,120,121] (O/V recombination). This threshold mechanism is elucidated in Figure 2.25b, where there is sketched the injection of oxygen ions and their recombination with vacancies in the filament, nearby the injection side. In this case, V_P determines how much effective is the mechanism of O/V recombination in the filament rupture, since the forming conditions determine the size of the filament. This is elucidated in Figure 2.25c, where two filaments with different size are drawn: in the case of the thicker filament (formed with higher values of V_P), the filament interruption via O/V recombination is more difficult, given the same time and voltage stress [122]. Obviously, also the increase of temperature has a role in favoring the filament rupture through O/V recombination, since temperature increases the probability that oxygen ions overcome the IL/HfO_2 barrier. This reflects in the fact that there is a distribution of the time-to-rupture in Figure 2.22.

A similar experiment was performed also on the TiN-Ti samples. The resistance evolution under constant voltage stress of both the low resistive state and the high resistive state of the TiN-Ti cells is studied from an experimental and theoretical point of view. A filamentary model based on ions hopping and oxygen vacancies generation phenomena is used to interpret the behavior of the cells [123]. The gap between the tip of the filament and the metal electrode is the parameter governing the device resistance. The evolution of the cell resistance in both the HRS and the LRS is crucial for the correct state reading with time and it is one of the key aspects of the memory reliability.

Usual retention tests consist in accelerating resistance changes by applying proper electrical stress [103, 117].

The memory cells were subject to electrical stress using positive constant voltages at the bottom gate in the HRS, and using negative voltages in the LRS. The resistance was systematically monitored at the same voltage of the stress. The CVS amplitude (V_{STRESS}) was varied in a wide range (0.3V-0.5V for the HRS and from -0.6V to -0.3V for the LRS). In Figure 2.26 there are reported data collected on distinct samples at various V_{STRESS}.

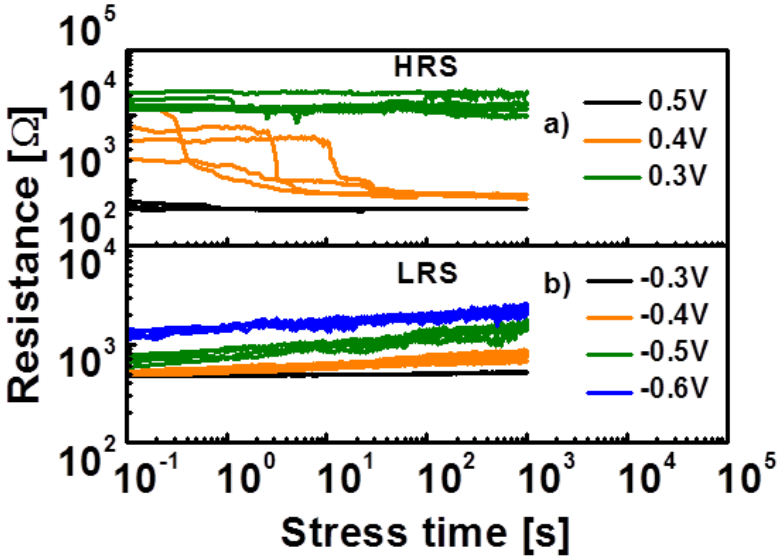


Figure 2.26 Resistance of the HRS (up) and LRS (down) measured during a CVS performed at the voltage values listed in the figure.

As for the HRS, in the studied timescale the monitored resistance did not change with V_{STRESS}= 0.3 V. Using V_{STRESS}= 0.4 V the resistance changed abruptly from HRS to LRS, with a wide time distribution. Finally, using V_{STRESS}=0.5 V the switch could not be monitored since it had already occurred at the time of the first measurement (100 ms). On the contrary, the stress in LRS made the resistance to vary progressively,

showing a trend of the variation rate with the stress voltage (Figure 2.26 b). In consequence of the stress, the LRS approached the resistance value of the HRS after a considerable time, depending on the stress voltage amplitude.

As described in the previous chapter the switching mechanisms can be simplified into the creation of a single dominant filament responsible of the conduction change in the cell. The length of a tunnelling gap is the parameter governing the device resistance. In this frame, the I-V current behaviour may be interpreted in terms of a mixed conduction model which takes into account the gap formation and evolution. The current equation can be written as follows:

$$I = I_0 \exp\left(-\frac{g}{g_{\min}}\right)\sinh\left(\frac{V}{V_0}\right), \qquad \textbf{2.13}$$

where the pre-factor I_0 is related to the slope of the ON current, V_0 is a constant voltage scaling term, g_{min} is the minimum gap length which is finite and it is assumed to be equal to the atomic distance a (0.25 nm, corresponding to the ON condition). The gap (g) kinetics is calculated taking into account the electric field and temperature enhanced oxygen ion hopping during RESET and oxygen vacancies generation for SET [97].:

$$\frac{dg}{dt} = -fa\exp\left(-\frac{E_{h,A}}{K_B T}\right)\sinh\left(\frac{qa\gamma V}{K_B TL}\right), \qquad \textbf{2.14}$$

where E_h is the hopping barrier, E_A is the activation energy for oxygen vacancy generation, f is the frequency attempt to escape, a is the atomic distance, K_B the Boltzmann constant and q is the electron charge. We fixed: E_h =1.5 eV, E_A = 1.3 eV and f = 10^{13} s^{-1}[97]. The local electric field is approximated with the formula $F = \gamma V/L$, where L is the film thickness (10 nm) and γ is a field enhancement factor taking into

account the local field modifications caused by the conductive filament and the strong polarizability in high-k dielectrics. Such a field enhancement factor depends on the gap length in the sense that it decreases with increasing *g*:

$$\gamma = \gamma_0 - c_0 g^3 , \qquad \textbf{2.15}$$

where γ_0 and c_0 are fit parameters. The cubic dependence with the gap length was obtained empirically by other authors [63]. The temperature T is the local temperature in the filament which is greater than the room temperature T_0 as described in paragraph 1.6.1:

$$T = T_0 + \xi_J \cdot P_J = T_0 + \xi_J \cdot V \cdot I , \qquad \textbf{2.16}$$

where ξ_J is an equivalent thermal resistance.

Application of an electrical stress can imply a degradation of the memory state, in the sense that the cell can vary its resistance state during the stress. In fact, when a negative CVS is applied in the LRS, the resistance increases its value and when a positive CVS is applied in the HRS, the resistance decreases. Referring to the conductive filament picture, this can be regarded as whether the gap increases or decreases its length, respectively. The conduction model and the gap dynamics expressed in equations 2.13-2.16 are now used to simulate the resistance value evolution under the effect of the electrical stress. First, we focus on the LRS and try to interpolate the curves sketched in Figure 2.26. Calculation of the LRS resistance was obtained with the equations 2.13-2.16. The simulated curves are displayed in Figure 2.27 with solid lines, superimposed to experimental data (symbols).

Table 3

γ_0	55
ξ_J	$1x10^5$ W/K
c_0	$(4.2\ nm)^3$
V_0	0.35 V
I_0	1.8 mA

All the curves were obtained using the parameter values listed in Table 3 and varying only the value of the applied voltage.

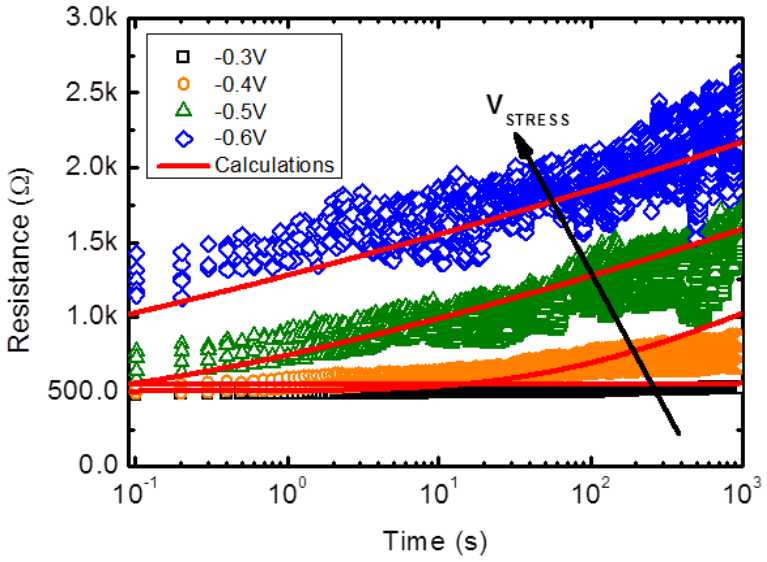


Figure 2.27 Calculated curves of LRS resistance obtained with the equations 2.13-2.16 (lines) and experimental data collected on samples for different V_{STRESS} (symbols).

The gap calculated integrating equation 2.14 is plotted in Figure 2.28a. At time zero (ON state) an initial (minimum) gap length *g=0.25nm* is assumed, to which the minimum resistance corresponds. The stress progressively widens the gap as shown in the figure. The higher the stress voltage, the faster gap widening. The local temperature in the filament is calculated using equation 2.16 of the model. Results are shown in Figure 2.28b. As expected, the maximum value of temperature is in the ON state and

rapidly decreases with the gap widening, since the flowing current decreases, as well as the Joule dissipated power.

Turning to the HRS evolution during stress, it can be noted that it strongly depends on the stress voltage. In fact, in one case (V_{STRESS}=0.3V) the resistance does not change in the time range monitored, while in another case (V_{STRESS}=0.5V) the device switches its state at the first measured point (100ms), thus denoting a threshold behavior of the cell. When V_{STRESS}=0.4V the devices exhibit an abrupt change of the resistance in a wide time range(10^{-1}s – 10^{1}s).

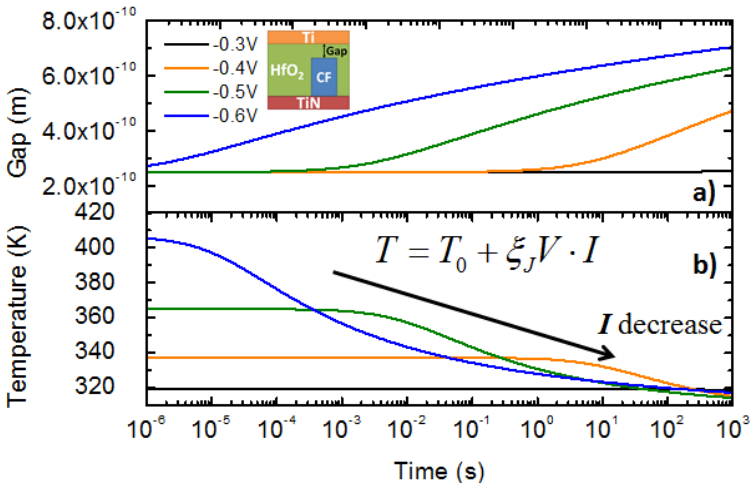


Figure 2.28 (a) Curves of the gap and (b) local temperature in the filament during the CVS calculated with the model for different values of V_{STRESS}. Inset: dependence of the gap on V_{STRESS}.

We focused on this case in order to understand this variability in the cell behavior. Equations 2.13-2.16 were used to describe the HRS evolution during the stress. A parameter g_0 representing the starting gap length was adopted in order to consider the different resistance starting value at the first measured point. This different starting configuration of the devices influences the parameter γ responsible of the enhancement

of local electric the field in the oxide. γ defined in equation 2.15 is a global parameter that approximates in a simple dependence with the gap length all the possible conductive filament configurations that modify the local electric field. In the case of HRS it is well known that there is a wider dispersion of the resistance value [124, 125] respect to the LRS. In Figure 2.29 are reported the HRS evolution during the CVS at 0.4V for four different samples. Markers are experimental data and lines are the calculated values using equations 2.13-2.16 and parameter listed in table 3 excepting c_0. The starting gap was varied between g_0=*1.17nm* and g_0=*0.7nm* and, in order to take into account the wide time range of the abrupt resistance change, also the parameter c_0 was varied. An empirical dependence of the type $c_0=a+bg_0$ (*a=0.622* and *b=-5.37*) was found (inset of Figure 2.29)

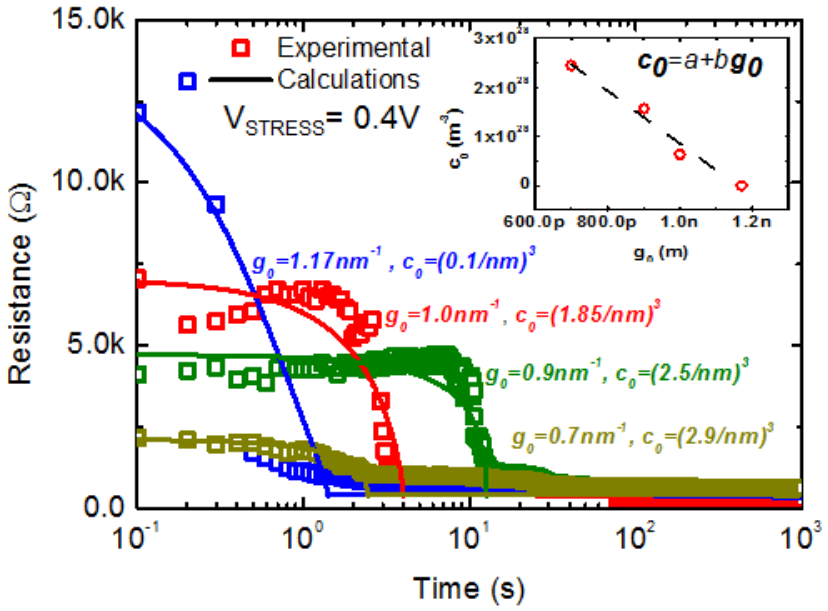


Figure 2.29 Solid lines are calculated HRS resistance evolution during the stress obtained with the model for V_{STRESS}=0.4V. Experimental data collected on 4 samples are reported with symbols. In the inset linear relation of c_0 with the initial gap is shown.

Calculations of the gap evolution during the positive CVS for the HRS are shown in Figure 2.30 a. Comparing results shown in Figure 2.28 a, it is possible to see the differences in the dynamic of the processes for the two resistive states. In Figure 2.28 a the gap increasing is gradual, while the gap decreasing in Figure 2.30 a is abrupt. In fact, looking to equation 2.15, the relation of the enhancement factor γ with the gap expresses a decreasing of its value increasing g. Also the temperature decreases due to the increasing of the gap length and the consequent reduction of the current.

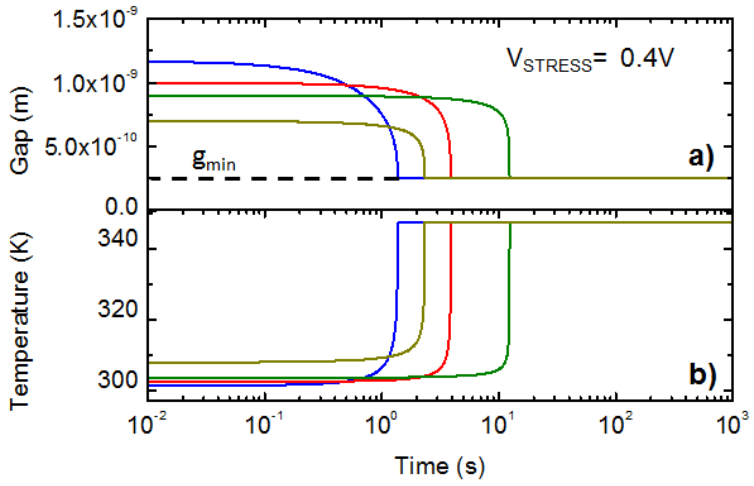


Figure 2.30 (a) Curves of the gap and (b) local temperature in the filament during the CVS calculated with the model for V_{STRESS}=0.4V.

This can be seen like a negative feedback: increasing the gap decreases the enhancement factor and, as a consequence, the electric field and the temperature. On the contrary, when the gap is reducing its value γ increases and so the electric field, the current and the temperature. In this case the phenomenon can be seen as a positive feedback self-accelerated process [66].

To clarify how the parameter variability influences the behavior of the cell in the HRS e LRS a Gaussian distribution of c_0 was adopted:

$$c_0 = c_{0\_mean} + \delta_{c0}\,, \tag{2.17}$$

where δ_{t0} is a zero mean value normal distribution with a standard deviation σ_{c0} of 10% of the mean value c_{0_mean}. Results are shown in Figure 2.31 a and 2.31 b. For clarity, only one starting gap g_0 was considered in the HRS. As one can see, the same variability of the same parameter gives a wider time range dispersion in the HRS state, while in the LRS the variability of c_0 reflects in a slightly slower or faster increase of the gap length during the electrical stress. The HRS case is a typical behavior of positive feedback systems, where small variations in parameters values give strong variations in the output [126].

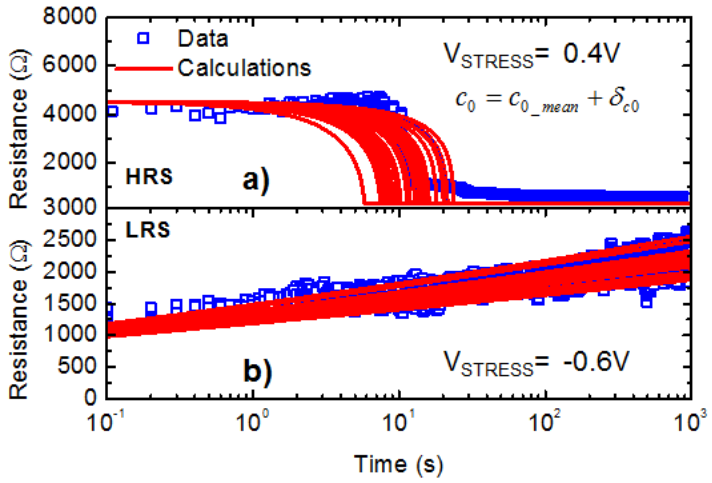


Figure 2.31 30 Curves of the gap evolution calculated with a Gaussian distribution of c0 for the HRS (a) and the LRS (b).

3 Neuromorphic Application

In this chapter, an introduction of the neuromorphic bio-inspired architectures will be given in terms of biological basics, neurons and synapses electrical models and circuits. Bio-inspired learning paradigm Spike Time Dependent Plasticity (STDP) will be described and adopted for simulations. Different approaches in RRAM based networks architectures and simulations will be described and results on visual pattern recognition will be discussed.

3.1 Introduction

Neuromorphic computing refers to an emerging interdisciplinary field that takes its inspiration from biological neural architectures and computations occurring inside the brain or the cerebral cortex. It comprises principles and knowledge from neurobiology, computational neuroscience, computer science, machine learning, VLSI circuit design, and more recently nanotechnology. According to Vonn Neumann architecture a computing machine is composed by different units: a unit that performs calculations, a unit responsible of controlling the sequence of the operations, a storage unit that holds data and program and inputs and outputs devices (Figure 3.1). In non-Von Neumann neuromorphic architectures, computing hardware and processing are not completely isolated tasks. Memory is intelligent and participates in processing of information. Neuromorphic computing may also referred to as Cognitive computing.

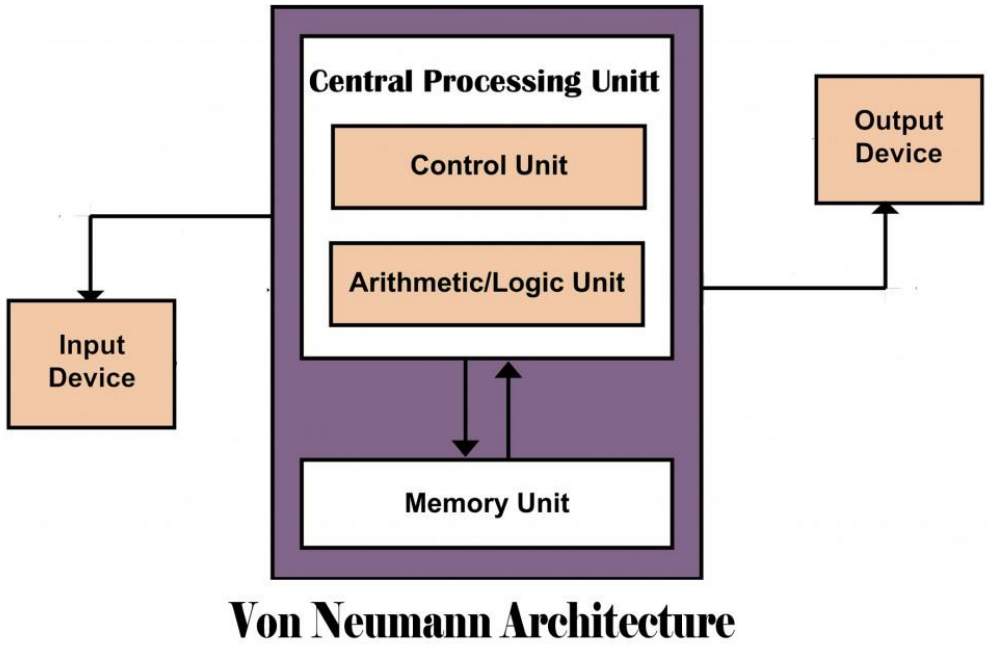


Figure 3.1 Von Neumann Architecture

Neuromorphic and bio-inspired computing paradigms have been proposed as the third generation of computing or the future successors of von-neumann machines as described in Figure 3.2.

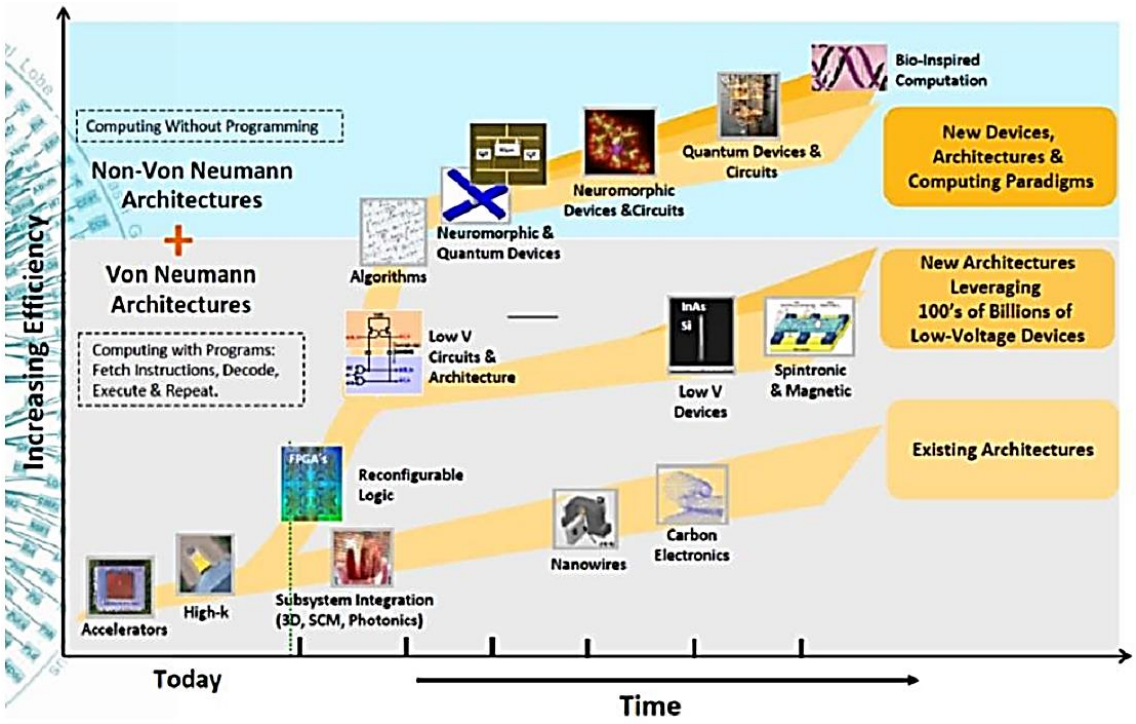


Figure 3.2 Computing architectures, circuits and devices

Architectures abstracted from neural systems provide a promising solution to different cognitive tasks like the object detection and the speech recognition, inspiring new paradigms, which can drastically improve the performance and efficiency of computing systems [127]. Due to the configurable and multiparallel architecture constituted by a complex network of ~10^{12} neurons and 10^{15} synapses, the human brain is able to perform a wide range of cognitive tasks with a power consumption of less than 20 W [128]. This value is much lower respect to the modern multi-core based computers that require 10,000 times more power [129]. The superior efficiency of the brain in performing fuzzy and fault-tolerant computation has motivated engineers to mimic the

key algorithmic and computational features of the brain in software and silicon-based hardware [130]. So, low power consumption and fast calculations are main goals in implementation of Artificial Neural Networks (ANNs). Dedicated hardware ANN architectures, compared to software implementations, can offer very high computational speed [131] at very low cost with high reliability, due to the redundancy of the architecture [132]. In the past years, pre-trained ANNs have been implemented in CMOS technology, with flexible design and a variety of architectures featuring good computing performances [133]. However, CMOS networks still have limitations in density and power consumption [134]. For example, in SRAM based ANNs, the synaptic weights are stored in volatile memories, which implies a considerable power leakage due to the subthreshold current [135].

In this scenario, Non-Volatile Memories synapses stand out for their excellent power management and integration due to their small sizes that allow high density structures. Resistive switching memories are most attractive for very large-scale system demonstration [136]. Today, RRAMs are considered among the most promising devices to realize neuromorphic systems capable of implement learning algorithms based on pulses [137]. This is due to several advantages offered by this technology: the simplicity of operation and low energy consumption (~0.1 pJ/bit) [138], the compatibility with CMOS technology and the possibility of 3D crossbar integration in back-end-of-line [139], the good scaling properties of the elemental cell (minimum size below 10 nm [33]), the excellent reliability (~10^{12} SET/RESET cycles [33]), the short programming time (~ns [33]), the online learning algorithms in which appropriate pulses update the synaptic weights.

3.1.1 Biological basics

Neurons are the basic functional units of the nervous system, and they generate electrical signals called action potentials, which allow them to quickly transmit information over long distances. Neurons, like other cells, have a cell body (called the soma). The nucleus of the neuron is found in the soma. Various processes (appendages or protrusions) extend from the cell body. These include many short, branching processes, known as dendrites, and a separate process that is typically longer than the dendrites, known as the axon (Figure 3.3).

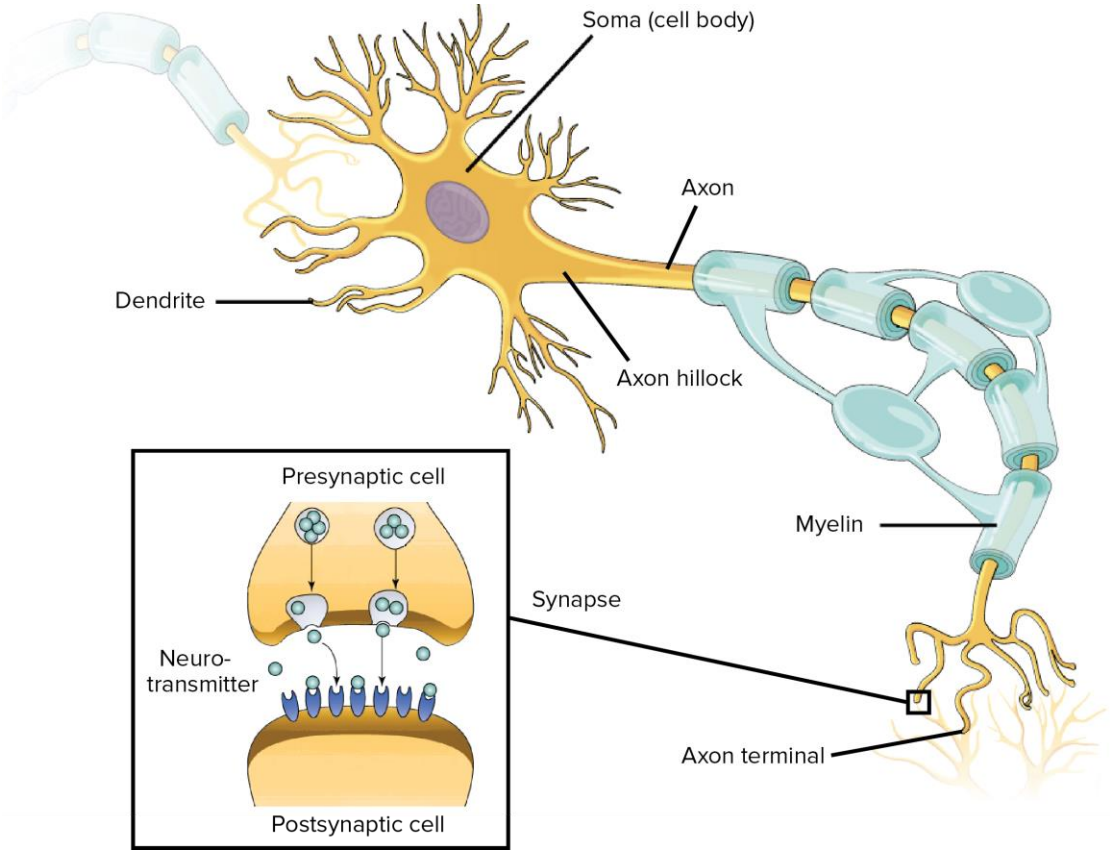


Figure 3.3 Neurons structure. In the inset a zoom of the synapse

The first two neuronal functions, receiving and processing incoming information, generally take place in the dendrites and cell body. Incoming signals can be either

excitatory – which means they tend to make the neuron fire (generate an electrical impulse) – or inhibitory – which means that they tend to keep the neuron from firing.

Most neurons receive many input signals throughout their dendritic trees. A single neuron may have more than one set of dendrites, and may receive many thousands of input signals. Whether or not a neuron is excited into firing an impulse depends on the sum of all of the excitatory and inhibitory signals it receives. If the neuron does end up firing, the nerve impulse, or action potential, is conducted down the axon. An axon is a special cellular extension (process) that arises from the cell body at a site called the axon hillock and travels for a distance, as far as 1 meter in humans or even more in other species. We can consider the dendrites as the inputs of the neuron and the axon as the output.

Connections between the neurons are made onto the dendrites and cell bodies of other neurons. These connections are known as synapses. They are the sites at which information is carried from the presynaptic neuron, to the target neuron uusally called postsynaptic neuron (Inset of Figure 3.3).

The information is transmitted in the form of chemical messengers called neurotransmitters. When an action potential travels down an axon and reaches the axon terminal, it triggers the release of neurotransmitter from the presynaptic cell. Neurotransmitter molecules cross the synapse and bind to membrane receptors on the postsynaptic cell, conveying an excitatory or inhibitory signal.

Thus, the axon and the axon terminals carry out the function of communicating information to target cells.

When a neurotransmitter binds to its receptor on a receiving cell, it causes ion channels to open or close. This can produce a localized change in the membrane potential—voltage across the membrane—of the receiving cell. In some cases, the change makes the target cell more likely to fire its own action potential. In this case, the shift in membrane potential is called an excitatory postsynaptic potential, or EPSP. In other cases, the change makes the target cell less likely to fire an action potential and is called an inhibitory post-synaptic potential, or IPSP. An EPSP is depolarizing: it makes the inside of the cell more positive, bringing the membrane potential closer to its threshold for firing an action potential. Sometimes, a single EPSP isn't large enough bring the neuron to threshold, but it can sum together with other EPSPs to trigger an action potential. IPSPs have the opposite effect. That is, they tend to keep the membrane potential of the postsynaptic neuron below threshold for firing an action potential. IPSPs are important because they can counteract, or cancel out, the excitatory effect of EPSPs.

A neuron constantly integrates or sums all the incoming PSPs, that it receives at its dendrites, from several pre-synaptic neurons. The incoming EPSPs and IPSPs lead to a change in the resting potential of the membrane. When the membrane potential depolarizes beyond a certain threshold, it leads to spiking or action potential generation inside the post-synaptic neuron, as shown in Figure 3.4.

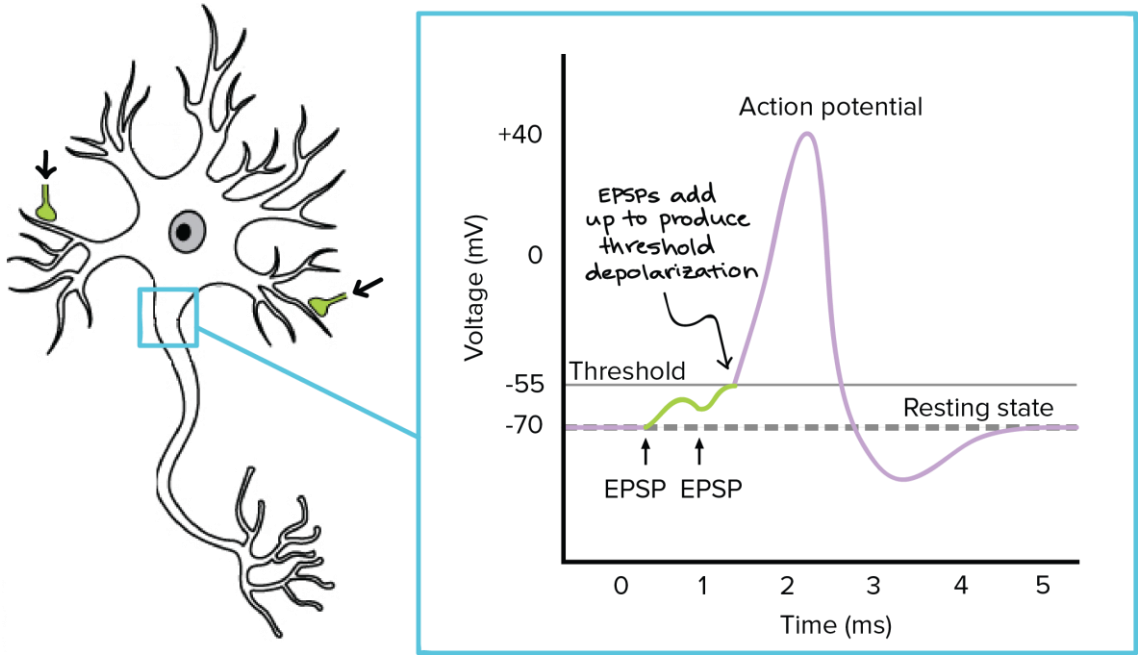


Figure 3.4 Action Potential

The weight of a synapse is defined by the strength of change that it can induce in the membrane potential of a post-synaptic neuron when an action potential is generated. The ability of a synapse to change its strength, in response to neuronal stimuli is defined as synaptic plasticity.

One of the most diffused and widely accepted model of the synaptic plasticity is the so called: Spike Timing Dependent Plasticity (STDP) [140,141]. According to this learning paradigm, the synaptic weights update according to relative spiking time of pre and post-synaptic neurons. According to STDP, when repeated pre-synaptic spikes arrive just before post-synaptic spikes the synapses weights increases following a long-term potentiation (LTP) (right part of Figure 3.5) . On the contrary, repeated spikes arriving after post-synaptic spikes leads to long-term depression (LTD) of the same synapses weights (left part of Figure 3.5). Thus, inputs that might be the cause of the post-synaptic neuron's excitation are made even more likely to contribute in the future,

whereas inputs that are not the cause of the post-synaptic spikes are made less likely to contribute in the future. Note that the relative change of synaptic strength is more profound if the time difference (Δt) between the spikes is smaller. As Δt increases, the effect of LTD and LTP becomes less profound like an exponential decay.

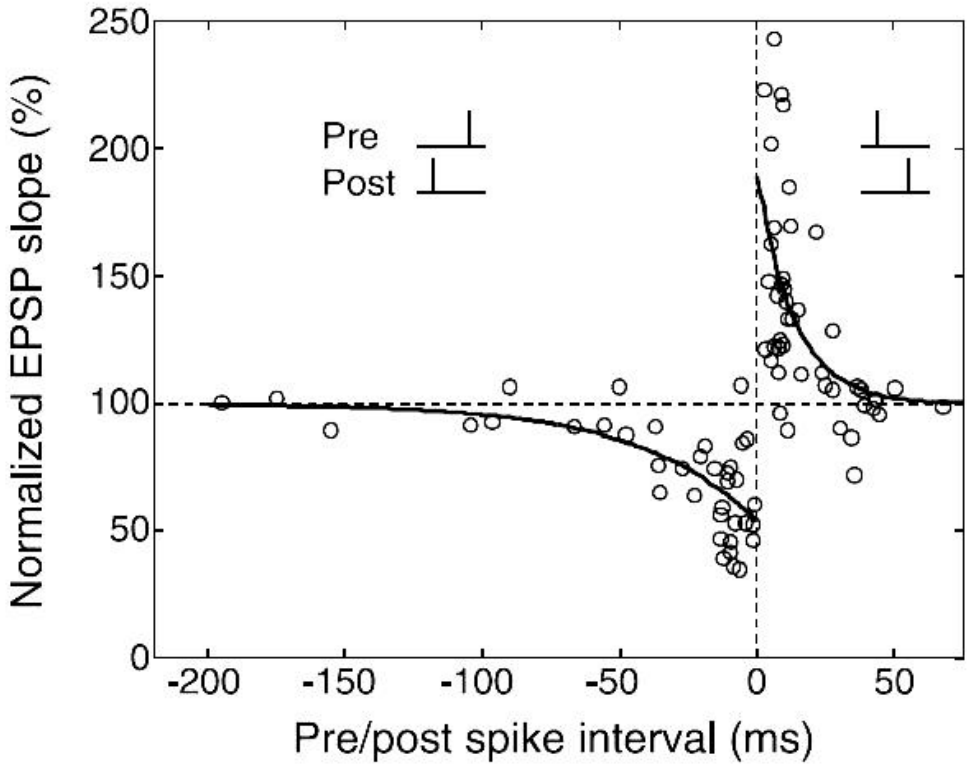


Figure 3.5 Illustration of spike timing dependent plasticity time windows, taken from [142]. Depending on the precise time difference between a post- and a pre-synaptic spike, the synaptic weight can be either depressed or potentiated.

3.2 Neuron electrical models and circuits

It is possible to find various models of biological neurons (and synapses), with different degrees of complexity and abstraction in literature. The complexity and the proper choice of a model depends on the context and on the application. In medical or biological fields, having a detailed model that takes in to account the dynamics at the

level of individual ion-channels and underlying biophysical mechanisms is fundamental to understand the working of the biological neurons. In neuromorphic computing, simple behavioral models are sufficient to describe, with a certain accuracy, the neuron features that are exploited in such applications.

One of the earliest and simplest neuron models is the Leaky Integrate-and-Fire (LIF) [143]. In this model, a neuron is represented by a simple capacitive and resistive differential equation, where *Cm* denotes the neuron membrane capacitance and *Rm* denotes the membrane resistance:

$$I(t) - \frac{V_m(t)}{R_m} = C_m \cdot \frac{dV_m(t)}{dt}, \tag{3.1}$$

The LIF model takes into account the leakage-effect of the neuron membrane potential by drift of some ions, assuming that the neuron membrane is not a perfect insulator. The spiking events are not explicitly modelled in the LIF model. Instead, the neuron constantly sums or integrates the incoming pre-synaptic currents and when the membrane potential reaches a certain threshold voltage (Vth) it instantaneously fire, generating an action potential.

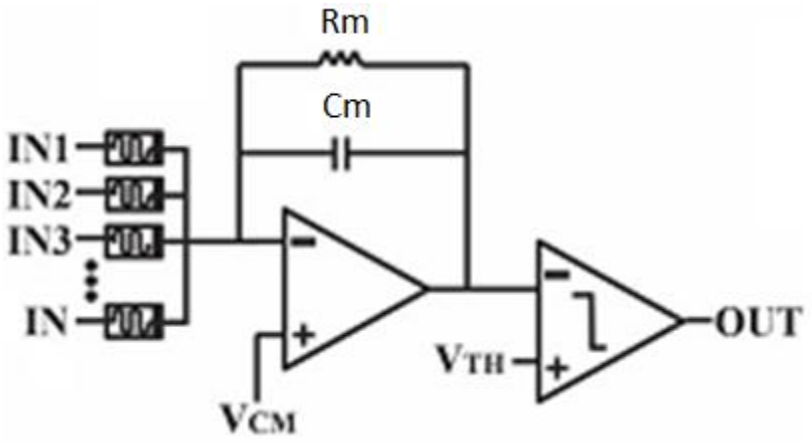


Figure 3.6 Schematic of LIF neuron

From a circuital point of view the LIF neuron can be implemented with an integrator, a comparator, switches, and eventually control logic. In the example in Figure 3.6 [144] the integrator has an inverting output; hence, the output decreases as the current flowing through the memristor is accumulated on the capacitance *Cm*. As soon as the integrator output goes down below a certain threshold voltage *Vth*, the comparator output goes high (logic "1"), and it can be recognized by the control logic as the neuron has fired.

In application where we want to include more biological aspects in the neuron model a deeper description is needed. In neuromorphic architectures, time represents itself and so the neuron (and synapse) circuits must process input data in the moment when they are available, producing output responses as much as possible in real time. In order to interact with the real-world sensory signals efficiently, neuromorphic systems must use synchronized with the real-world events circuits that have biologically plausible time constants (tens of milliseconds) [145]. This constraint is not easy to satisfy using analog VLSI technology [146-148]. A way to overcome this problem is to use current mode design techniques [149] and subthreshold circuits [150-154]. When MOSFETs operate in the subthreshold domain, the carrier diffusion is the main physical mechanism governing the behavior of the transistor. Therefore, MOSFETs have an exponential relationship between drain current I_D and gate voltage V_{GS}, at very low currents regime. The time constants of MOS subthreshold circuits are inversely proportional to the reference currents and directly proportional to the circuit capacitance, thus allowing the integration of relatively small capacitors in integrated circuits in order to implement temporal circuits that are both compact and have

biologically realistic time constants, ranging from tens to hundreds of milliseconds. A sub-threshold neuron circuit is shown in Figure 3.7. It is composed by:

- input differential pair integrator (DPI) circuit used as a low pass filter (LPF) (ML1-ML3)
- spike-event generating amplifier with current-based positive feedback (MA1-MA6)
- spike reset circuit with refractory period functionality (MR1-MR6)
- spike-frequency adaptation mechanism implemented by an additional DPI LPF (MG1-MG6).

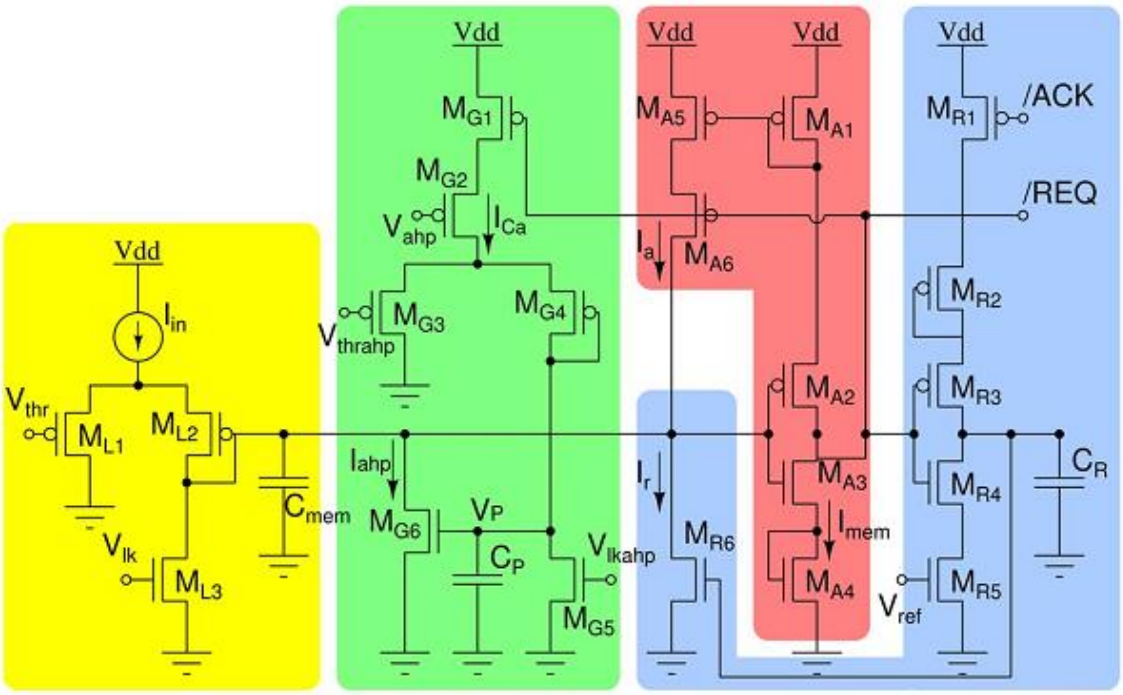


Figure 3.7 Adaptive exponential neuron circuit schematic. The input DPI circuit (ML1-3) models the neuron's leak conductance. A spike event generation amplifier (MA1-6) implements current-based positive feedback (modeling both sodium activation and inactivation conductances) and produces address–events at extremely low-power operation. The reset block (MR1-6) resets the neuron and keeps it in a

resting state for a refractory period, set by the Vref bias voltage. An additional LPF (MG1-6) integrates the spikes and produces a slow after-hyperpolarizing current Iahp responsible for spike-frequency adaptation.

The DPI block ML1-3 models the neuron's leak conductance; it produces exponential subthreshold dynamics in response to constant input currents. The neuron's membrane capacitance is represented by the capacitor C_{mem}, while sodium channel activation and inactivation dynamics are modeled by the positive-feedback circuits in the spike-generation amplifier MA1-6. The reset MR1-6 block models the potassium conductance and refractory period functionality. The spike-frequency adaptation block MG1-6 models the neuron's calcium conductance that produces the after-hyperpolarizing current *Iahp*, which is proportional to the neuron's mean firing rate. A first order set of differential equations describes the dynamics of the current:

$$\begin{aligned} &\left(1+\frac{I_{th}}{I_{mem}}\right)\tau\frac{d}{dt}I_{mem}+I_{mem}\left(1+\frac{I_{ahp}}{I_{\tau}}\right)=\mathrm{I}_{mem_{\infty}}+f\left(I_{mem}\right) \\ &\tau_{ahp}\frac{d}{dt}I_{ahp}+I_{ahp}=I_{ahp_{\infty}}u(t) \\ &f\left(I_{mem}\right)=\frac{I_a}{I_{\tau}}\left(I_{mem}+I_{th}\right) \end{aligned} \quad , \qquad 3.2$$

Where I_{mem} is the subthreshold current that represents the real neuron's membrane potential variable, I_{ahp} is the slow variable responsible for the spike-frequency adaptation mechanisms, and *u(t)* is a step function that is unity for the period in which the neuron spikes and null in other periods. Term *f(Imem)* is a function that depends on both membrane potential variable and positive-feedback current *Ia*. The other parameters are defined as (referring to Figure 3.7 for the symbols):

$$\tau = \frac{C_{mem} V_T}{k I_\tau}; \tau_{ahp} = \frac{C_p V_T}{k I_{\tau_{ahp}}}$$
$$\mathrm{I}_{mem_\infty} = \frac{I_{th}}{I_\tau}(I_{in} - I_{ahp} - I_\tau); I_{ahp_\infty} = \frac{I_{th_{ahp}}}{I_{\tau_{ahp}}} I_{Ca} \quad , \qquad \textbf{3.3}$$

Where I_{th} and I_{thahp} represent currents through n-type MOSFETs not present in Figure 3.7. For values of $I_{in} >> I_\tau$ it is possible to simplify equation 3.3 in:

$$\tau \frac{d}{dt} I_{mem} + I_{mem} = \frac{I_{th}}{I_\tau} I_{in} + \frac{I_a}{I_\tau} I_{mem}, \qquad \textbf{3.4}$$

Where $f(I_{mem}) \approx (I_a/I_t) I_{mem}$. So, under these conditions, the circuit of Figure 3.7 implements a generalized Integrate and Fire neuron model [155]. Learning and long-term memory of information in biological neurons is due to calcium flux controlled by N-Methyl-D-Aspartate (NMDA) channel induced synaptic plasticity [156, 157]. Biological NMDA receptors have been shown to demonstrate spike timing dependent plasticity. Synaptic weights can learn to associate positive or negative correlations with the differential timing patterns of pre- and post-synaptic spikes. Since the conception of this theory, biological neuron activity has been shown to exhibit behavior closely - modeling such calcium-based Hebbian learning [157-160]. Since calcium concentration decays exponentially, this behavior can be easily implemented on hardware using subthreshold transistors. The calcium signal integrates the postsynaptic spike sequence and accumulates according to:

$$\frac{d}{dt} I_{Ca} = \frac{I_a - I_{Ca}}{\tau_{Ca}}, \qquad \textbf{3.5}$$

where I_{Ca} is the postsynaptic calcium variable and is a function of postsynaptic spiking activity, with a long time constant τ_{Ca}.

3.3 Synapses

Several different hardware implementation of artificial synapses exist in literature. A complete summary on this topic is out of the scope of this thesis. In this paragraph we consider just the CMOS synapse implementation described in [145] that was the starting point for the results described in the following paragraph 3.4.3 and the RRAM based synapses characteristics.

3.3.1 CMOS implementation

Synapses are the ports in which both biological and silicon neurons receive and compile information from downstream neurons, thus playing a crucial role in neural learning. This silicon synapse is based on an integrator that is capable of emulating short-term plasticity, well as NMDA receptor behavior, and produces biologically plausible Excitatory-Post-Synaptic-Currents (EPSCs) [145]. An example of a full excitatory synapse circuit is shown in Figure 3.8. The input spike (the voltage pulse *Vin*) is applied to both MD3 and MS3. The output current *Isyn*, sourced from MD6 and through MG2, rises and decays exponentially with time. The temporal dynamics are implemented by the DPI block MD1-6. The circuit time constant is set by $V\tau$ while the synaptic efficacy, which determines the EPSC amplitude, depends on both *Vw0* and *Vth*. The circuit dynamic is described by the equation:

$$\frac{d}{dt} I_{syn} = I_{syn}\left(\frac{I_w}{I_\tau} + 1\right), \qquad 3.6$$

Equation 3.6 says that the change in circuit response increases with every spike, by an amount greater than one, for as long as condition *Isyn*<<*Ith* is satisfied. As *Isyn* increases, this condition starts to fail, and eventually the opposite condition (*Isyn* >>I*th*)

is reached. This is the condition for linearity, under which the circuit starts to behave as a first-order LPF.

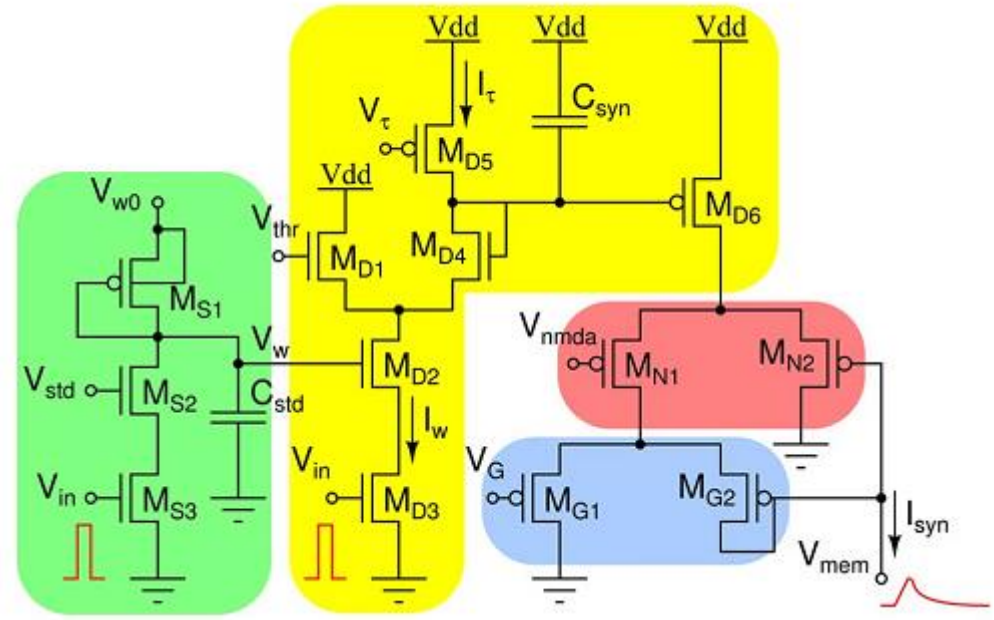


Figure 3.8 Complete DPI synapse circuit, including short-term plasticity, NMDA voltage gating, and conductance-based functional blocks. The short-term depression block is implemented by MOSFETs MS1-3; the basic DPI dynamics are implemented by the block MD1-6; the NMDA voltage-gated channels are implemented byMN1-2, and conductance-based voltage dependence is achieved withMG1-2

3.3.2 RRAM based synapses

Two-terminal RRAM devices have been proposed as artificial synapses in neuromorphic circuits thanks to the capability of analog gradual resistance modulation, strong device area scalability, low power consumption, compatibility with CMOS technology combined with a 2-terminal structure. Due to these characteristics, RRAMs naturally satisfy the requirements to act as a connection for communication between a pre-synaptic neuron and a post-synaptic neuron. The RRAM resistance state can be

considered inversely proportional to the synaptic weight w. Different resistive switching devices were used for practical implementations, such as: phase change [161,162], ferroelectric [163,164], and oxide-based RRAMs [165,161]. When RRAMs are employed in neuromorphic networks, two main operational modes are used, binary and analog. The binary operations is based on the two limits states HRS and LRS, and it is proved to be effective in specific applications [161,165,166]. At the same time, gradual analog resistance modulation is desirable to improve the performances of the network [167- 169].However, the difficulty of operating RRAM in an analog way, especially in transitions from HRS to a more conductive state, implies challenging hardware implementations of the networks [168].

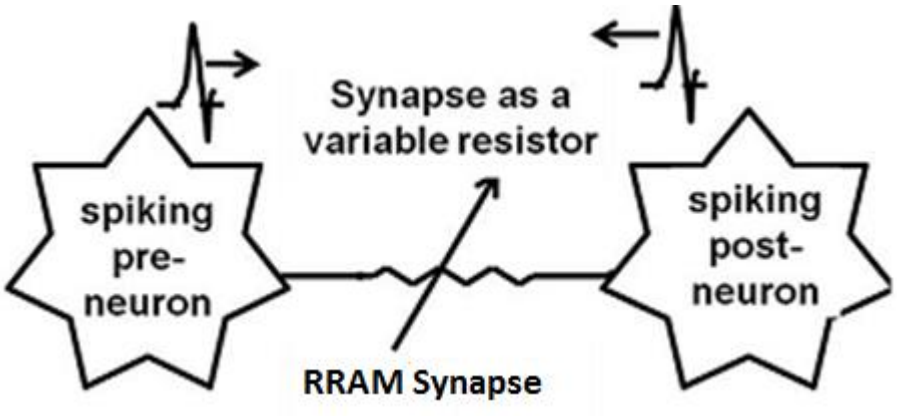


Figure 3.9 Basic RRAM based synapse scheme connecting a pre-synaptic neuron and a post-synaptic neuron

In Figure 3.9 is sketched the basic connection scheme of a pre-synaptic and a post-synaptic neuron through a RRAM based synapse. As previously said, the memory cell resistance value is linked to the weight of the synapse. In neuromorphic network, the evolution of the weight is responsible of the learning of the circuit. Essentially, during the training of the network, a defined learning paradigm is adopted in order to obtain

an evolution of the weights in a way that the network learned a certain task. So the use of a proper learning paradigm responsible of the evolution of the RRAM resistive state is a key aspect for the entire behavior of the network.

In the following paragraphs, different approaches will be considered in neuromorphic networks simulations with RRAM based synapses for visual pattern recognition applications:

- PSPICE simulation of a small network
- Matlab simulations of a large network based on STDP
- Brian (Python) simulations of a large array based on a modified version of the STDP

3.4 Networks

A neuromorphic network has the neurons divided into subgroups of fields and elements in each subgroup are usually placed in a row or a column. Each subgroup is called layer of neurons in the network. A neuromorphic network may have input layer that supply the input signals for the neurons in the next layer, output layer where output is generated and in between them hidden layer(s) that process information between input and output layers. Two neurons are connected with a weight (synapse) that can have various values. Basically, the raw output of a neuron in a simple network is a weighted sum of its inputs multiplied by weights connected to it.

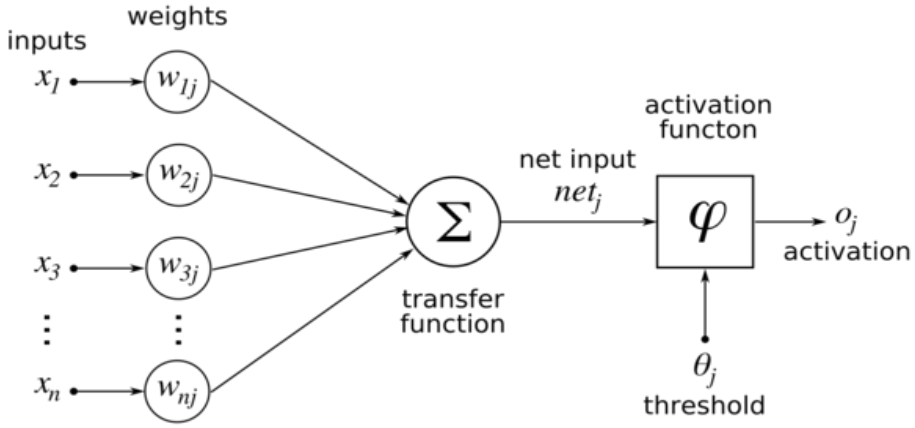


Figure 3.10 Basic scheme of a neuron with the weights and the inputs

A single neuron processes multiple inputs applying an activation function on a linear combination of the inputs:

$$y_i = \varphi_i\left(\sum_{j=1}^{N} w_{ij}x_j + b_i\right), \qquad 3.7$$

where $\{x_j\}$ is the set of inputs, w_{ij} is the synaptic weight connecting the j^{th} input to the i^{th} neuron, b_i is a bias, $\varphi i(\cdot)$ is the activation function, and y_i is the output of the i^{th} neuron considered. The activation function is usually strongly nonlinear.

3.4.1 PSPICE simulated network

The leaky integrate and fire (LIF) model, described in previous paragraphs, was adopted to implement the neurons. A neuromorphic network composed by a first 25 sensory neurons layer and a second layer made of 10 output neurons connected through 250 RRAM synapses was used (Figure 3.11) [170].

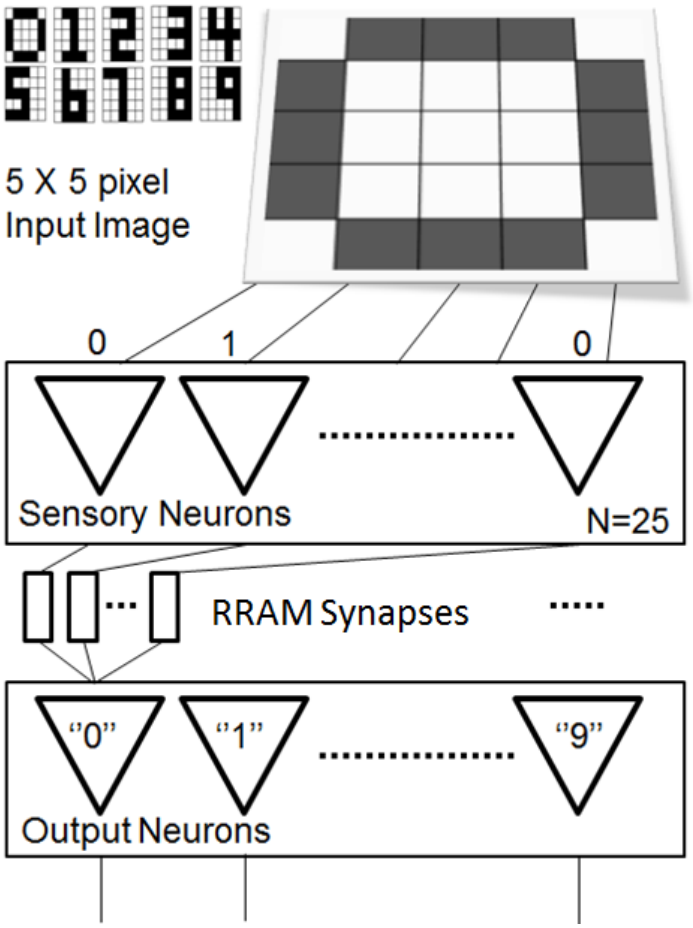


Figure 3.11 Circuit architecture: 5x5 pixel binary input image, 25 sensory neurons, 250 RRAMs and 10 output neurons

TiN/HfO2/Ti RRAMs were electrically characterized and modeled to be used as synapses. The steady-state current voltage (I-V) characteristics for SET and RESET is displayed in Figure 3.12. In order to simulate the I-V characteristic of the cells, a

physically based simplified model (as described in Chapters 1 and 2) was used to reproduce the I-V characteristics. The equation adopted are included in the figure.

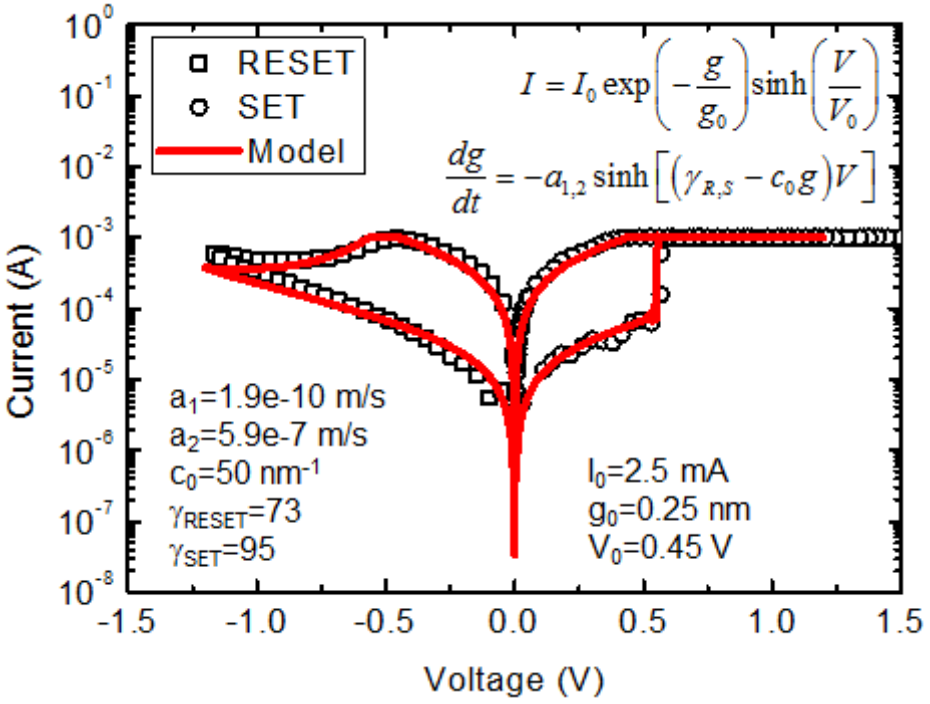


Figure 3.12 RRAM experimental (markers) current-voltage characteristic and model (solid line). Current and tunneling gap dynamic equations are also shown.

Starting from these equations a PSPICE behavioral model was implemented in order to perform circuital simulations. SPICE (Simulation Program with Integrated Circuit Emphasis) is a general-purpose, open source analog electronic circuit simulator. It is a program used in integrated circuit and board-level design to check the integrity of circuit designs and to predict circuit behavior. The proposed model is based on the equations in Figure 3.12. In the following page is listed the PSPICE code used. The RRAM device is described with two main blocks: the first one is an integrator composed by a capacitor and a source of current. This part is devoted to integrate the differential equation related to the gap dynamics. The second block is a voltage

controlled current source that represents the conduction current dependent on the actual value of the gap.

```
* SPICE model for RRAM devices, Created by Paolo Lorenzi
* Connections: TE - top electrode BE - bottom electrode
.SUBCKT MEMRISTOR-LORENZI TE BE
.PARAM a1=-1.9199e-10 a1_2=-5.9199e-07 a2=0.0025 c0=50
gamma0=73 gini=0.26gmin=0.25 g0=0.25 tens0=0.45 gmax=10
gamma0_set=95
.func Arg(V1,V2) =  {LIMIT(0.9667*(gamma0_set-
c0*V2)*V1,1000,-5800)}
*function F(V1, V2, V3) - Describes the SV motion (V2 is
the gap), V1 is the applied voltage, V2 * is the limited
sinh argument and V3 is the gap; 0.9667 in the sinh
considers a/(l*ke*T)
.func
F(V1,V2,V3)={IF(V1<=0,a1*sinh(V2),IF(V3<=gmin,0,a1_2*sinh(V
2)))}
* IV Response - Hyperbolic sine due to MIM structure
*a2=I0, V2=gap, tens0=V0
.func IVRel(V1,V2) =  {a2*exp(-V2/g0)*sinh(V1/tens0)}
* Circuit to determine state variable
* dx/dt = F(V(t),x(t))
Cx XSV 0 1
Rx XSV 0 1g
.IC V(XSV) = 0.26
Ex 2 0 value={ Arg(V(TE,BE),V(XSV,0))}
Gx 0 XSV value={ F(V(TE,BE),V(2,0),V(XSV,0))}
* Current source for memristor IV response
Gm TE BE value = {IVRel(V(TE,BE),V(XSV,0))}
.ENDS MEMRISTOR-LORENZI
```

The entire system was adopted in order to recognize 5X5 pixel binary images representing numbers between “0” and “9” (Figure 3.11). The network operations are divided in two distinct phases: the training and the testing ones. During the training phase, all the memory cells started in LRS and voltage pulses of -1.2V were applied to the resistive switching devices. The output spike of the sensory neurons were used to modify the resistance values of RRAMs. The simulated resistance behavior, with the

PSPICE model described above, as a function of the number of sensory neuron pulses is shown in Figure 3.13 for three different pulse widths. For this simulations the same parameters used to fit the I-V experimental characteristic shown in Figure 3.12 were adopted. Unfortunately it was not possible to compare the pulsed simulated values with experimental data on our samples. However, it is clear that the trend is in agreement with already published works in literature [63]. For the training, a simple protocol was adopted: every neuron was trained separately to recognize one of the 10 numbers just applying training pulses of -1.1 V in correspondence of black pixels in the number image.

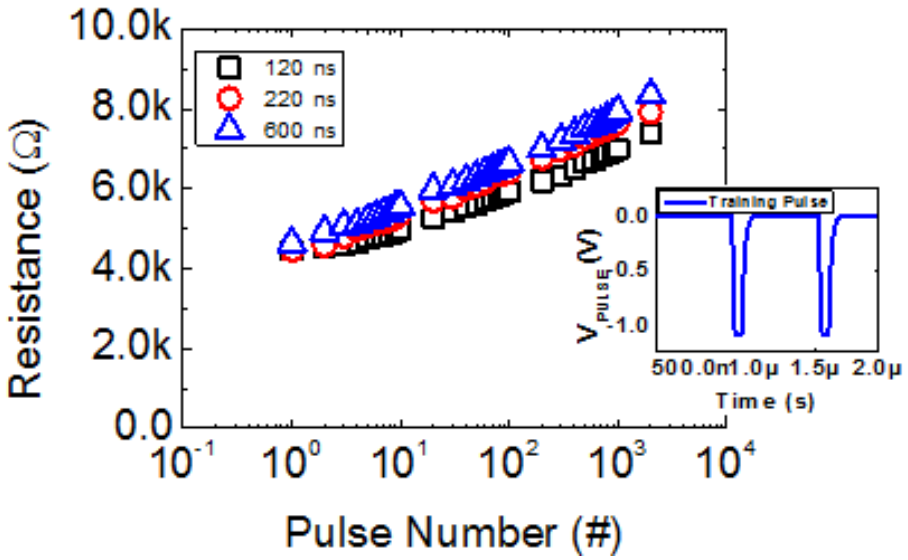


Figure 3.13 PSPICE RRAM resistance evolution as a function of training pulse number for three different pulse width. In the inset an example of the sensory neuron training pulse

In the following, binary images represented by just low and high resistive state will be considered. During the testing phase an input image is applied (numbers between "0" and "9"). Low voltage pulses (0.1V), corresponding to black pixels, were applied to

RRAM previously trained and integrated by output neurons. When the voltage across the capacitor reaches the threshold voltage (Figure 3.14), the winner neuron stops the others through a feedback network (not shown here) and the recognition is completed. In Figure 3.14 an example is shown: during testing an image corresponding to a "0" is applied to the network. All the output nodes start the integration of the spikes coming from the synapses. The output neuron trained to recognize the number "0" was the first one reaching the threshold voltage.

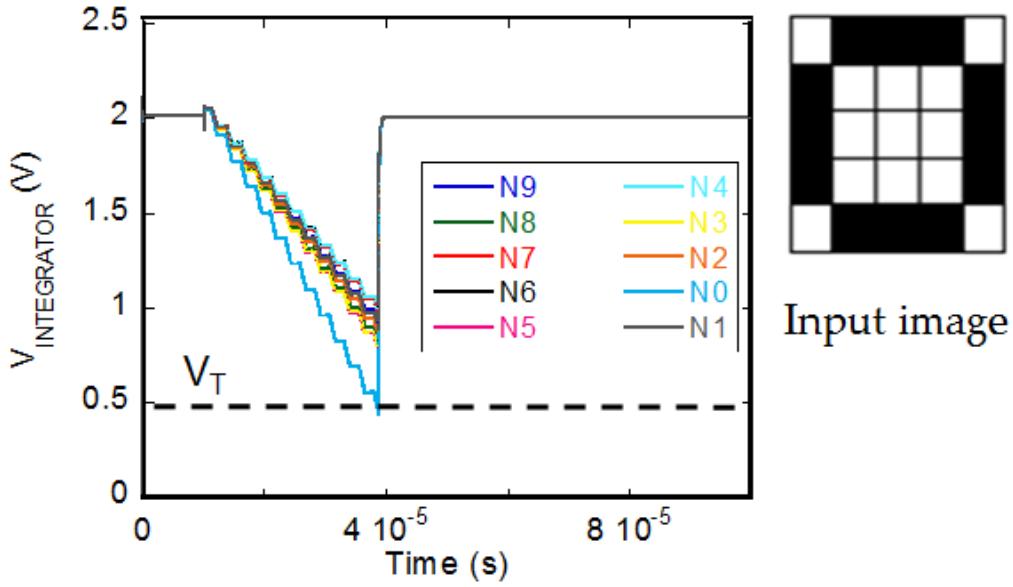
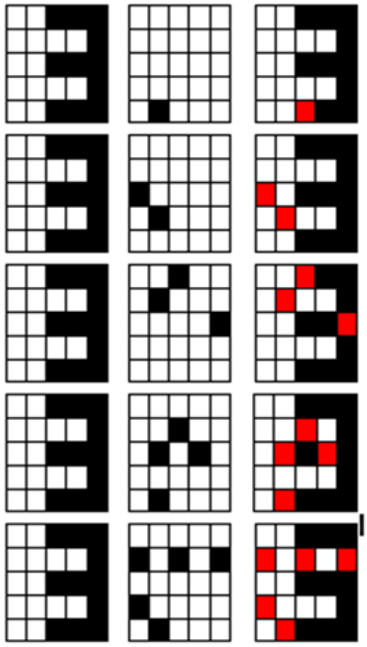


Figure 3.14 Voltage at the integrator node of all the output neurons during the testing phase with "0" as input image.

The performances of the circuit were tested in two ways. The first one was to apply the same inputs image adding a noise (Figure 3.15). In particular, noisy pixel with random timing features were added to the input images with different percentage (4% - 20%). It was found that if the noise level was under the 10% (corresponding to 2 pixels) the network recognized correctly all the patterns. The second way to evaluate the performance of the network was to build an home-made test set as shown in Figure

3.16 and monitoring the success rate of recognition. A test set of 100 images composed by 10 different representations of numbers between "0" and "9". A success rate of the 44% was achieved with the test set used. Obviously, in order to increase the performances of the network, it is mandatory to improve the learning paradigm during the training phase. Furthermore, the adoption of a meaningful data set is fundamental to give robustness to the network. These points will be addressed in the following paragraph.



Noisy Pixels	Recognition
0%	All patterns were recognized successfully.
4%	All patterns were recognized successfully.
8%	All patterns were recognized successfully.
12%	Failed to recognize pattern 3. Failed to recognize pattern 2.
16%	Failed to recognize pattern 3. Failed to recognize pattern 2.
20%	Failed to recognize pattern 7.

Figure 3.15 The effect of noisy pixels on the recognition of the input image "3" is shown on the left. Network recognition performance as function of the noisy pixels is resumed.

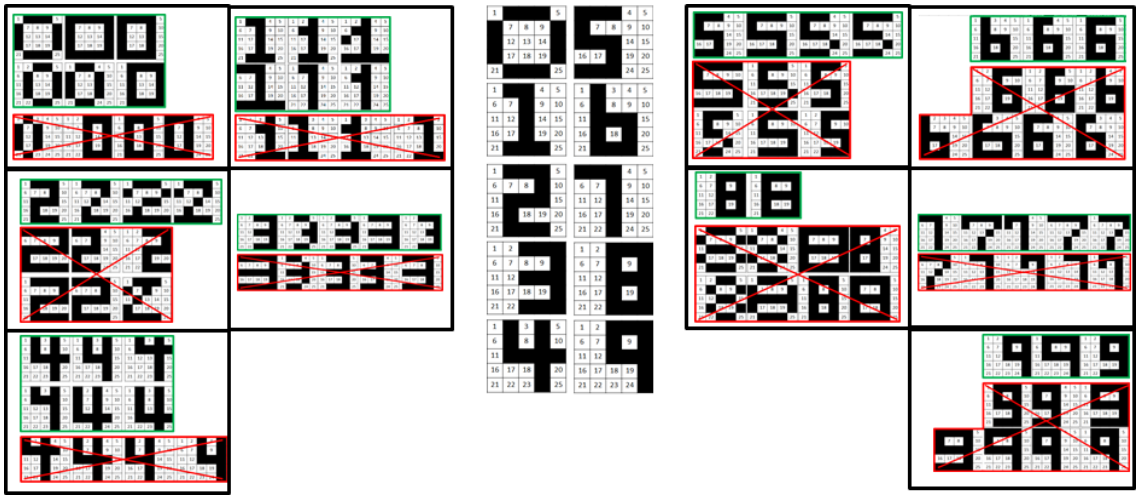


Figure 3.16 Test set used to evaluate the success rate

3.4.2 STDP-based network for handwritten digits recognition

A fully connected brain inspired neuromorphic network composed by 784 input neurons, M output neurons (with M ranging between 100 and 2000) and 784xM TiN/HfO_2/Ti Resistive RAM synapses was simulated. The proposed network has been trained with the biologically inspired learning paradigm Spike Time Dependent Plasticity in order to recognize handwritten digits between “0” and “9” from the Mixed National Institute of Standards and Technology (MNIST) database. The MNIST database is composed by 60000 training and 10000 testing images. Every image is made by 28x28 pixels (784 in total) coded in a grey scale of 256 levels. The MNIST database represents a classic data set used as a benchmark for pattern recognition algorithms. In Figure 3.17 an example of the images composing the MNIST database in shown.

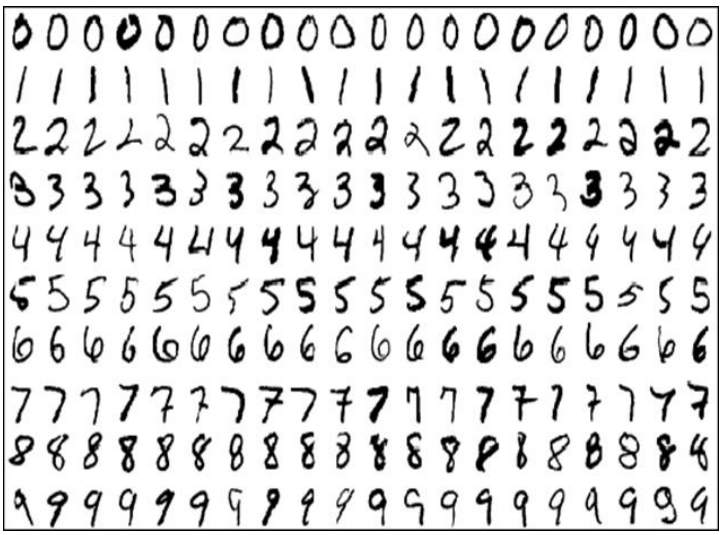


Figure 3.17 Subset of the handwritten digits taken from the MNIST database

A network composed by 784 input neurons and an output layer made of M neurons connected through 784xM synapses was considered [171]. The measured high and low resistance values of the TiN/HfO2/Ti Resistive RAMs fix the boundaries of the synapses weights.

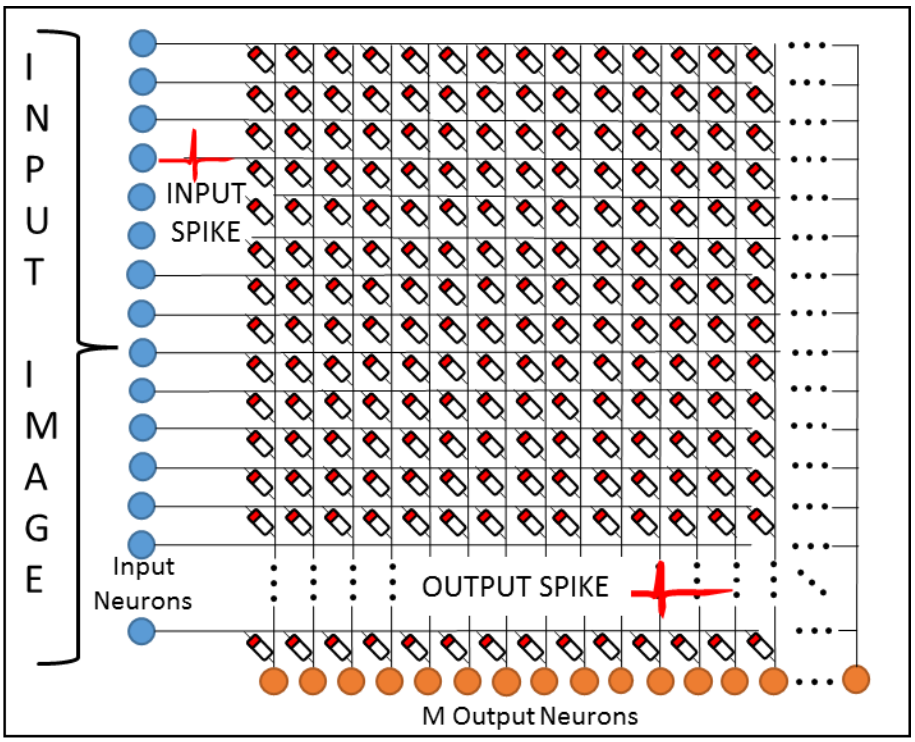


Figure 3.18 Sketch of the matrix made of 784 input sensory neurons and M output neurons. Synapses connect input and output neurons

The neurons were modeled as leaky integrators inspired by the leaky integrate and fire (LIF) model. In the training phase, input neurons emit spikes corresponding to black pixels of the input training image. When the input neuron spikes, a positive voltage is applied to the RRAM connected to that neuron. This voltage induces a synaptic current small enough to maintain unchanged the state of the resistive memory. The synaptic current is added to the other synaptic currents coming from the RRAMs belonging to the same column. The output neuron spikes when its voltage drop reaches an arbitrary predefined threshold that is the same for all neurons. Then, the synapse weight changes according to the STDP paradigm, as shown in Figure 3.19. Unfortunately, besides the quasi static I-V characteristics of the RRAM, no experimental data about the trained and recognized characteristics of the neuromorphic network were collected. Although,

the simulated results seems to lack the link with the real device, the curve represented in Figure 3.19 has a shape comparable with the same curve that is possible to find in various works in literature [165, 172] where RRAMs are used to implement STDP learning rule. Different methods are described in literature to obtain the STDP curve using RRAM devices. Most of them are based on the overlapping of waveforms coming both from the pre-synaptic and the post-synaptic neuron [172, 173]. In these work the curve in Figure 3.19 was used as a starting point for the simulation of the network and it was not derived from experimental data. Only the limits of the range of resistances that the RRAM could explore were fixed by experimental data. So the high limit was fixed at 10kohms, while the low limit was fix at 500ohms.

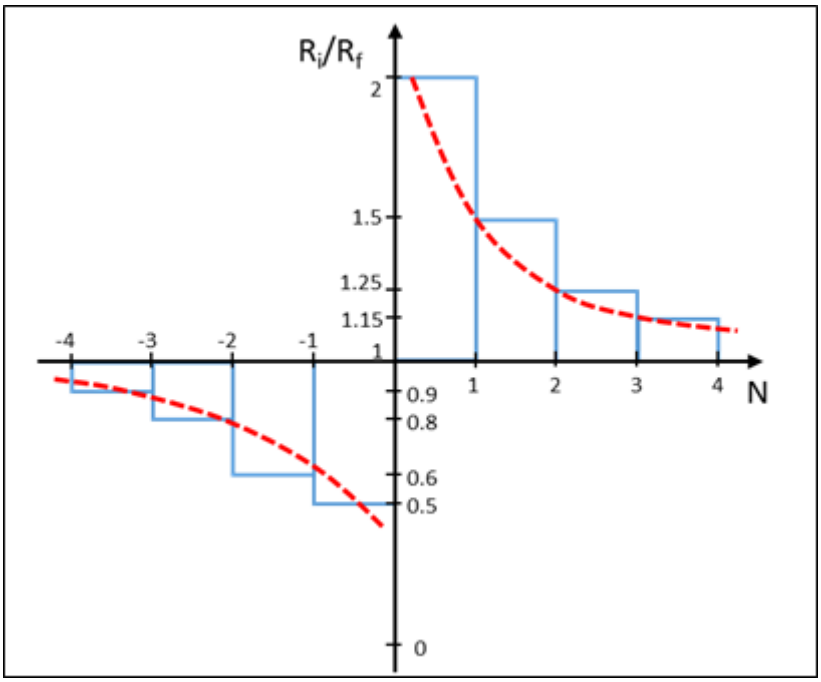


Figure 3.19 Ratio between initial and final RRAM resistance vs simulation step. The dashed curve represents the STDP paradigm

To train the network, the entire MNIST training set (60,000 examples) was used. A new image was presented to the input layer when an output neuron emitted a spike.

After training, a class was assigned to each neuron, based on its highest response to the ten classes of digits between "0" and "9" over the presentation of the training set. The recognized class was determined by averaging the responses of each neuron per class and then choosing the class with the highest number of spikes. The response of the neurons with the assigned class was then used to measure the classification accuracy of the network on the MNIST test set. After the classification, the network was tested with 10000 testing images without updating the RRAM resistance values. All the simulations were performed in MATLAB. In Table 4 it is shown the overall success rate in recognizing actual numbers for different values of the output neuron number (M between 100 and 2000). As a result, increasing M up to 500, the overall success rate increases and then it assess around 83%. Beyond this limit, the success rate becomes independent on M.

Table 4

Test / Num. / M	0	1	2	3	4	5	6	7	8	9	TOT
100	83.54	92.16	79.12	91.45	45.49	42.57	96.48	93.04	32.14	55.26	73.75
500	99.99	95.25	93.74	96.04	82.51	64.13	84.69	88.44	59.52	56.3	82.062
1000	96.12	99.53	97.64	85.2	82.83	73.75	95.7	87.42	76.19	51.09	84.594
2000	96.63	95.56	95.41	89.27	74.03	79.33	89.64	84.25	64.82	53.59	83.02

The confusion table of the 784x1000 network is displayed in Table 5. It reports the percentage of recognized numbers as function of the actual numbers over the 10000 test

images. The diagonal gives the correct recognitions. The most failing tests regards number "9", which exhibits around 40% failures in total, and number "8", which exhibits more than 20% failures. These failures are strictly related to the intrinsic similarity between symbols in the MNIST database. Recognition rates greater than 80% are achieved for seven of the ten numbers, while the most unsatisfying results refer to confusion between numbers "9", "4" and "7".

Table 5

Recognized \ Actual	0	1	2	3	4	5	6	7	8	9
0	**96.12**	0	0.50	1.97	0.75	10.45	2.97	0.30	8.76	2.60
1	0	**99.53**	0.20	0.20	0.64	1.42	0	0.71	0.54	1.56
2	0.53	0	**96.94**	2.60	0.21	0.59	0.55	2.35	3.67	0.52
3	0.53	0	0.40	**85.20**	0.10	7.83	0	0.10	4.43	1.14
4	0	0	0.50	0.72	**82.83**	0.83	0.22	5.52	1.62	27.42
5	0.21	0	0.20	5.62	0.53	**73.75**	0.22	0	4.00	0.31
6	1.39	0	0.20	0.20	1.28	2.13	**95.70**	0	0.10	0.10
7	0.10	0.46	0.20	0.62	0.42	0.11	0	**87.42**	0.10	12.30
8	1.07	0	0.81	2.81	1.28	2.61	0.22	1.53	**76.19**	2.91
9	0	0	0	0	11.90	0.23	0.11	2.04	0.54	**51.09**

Some neural networks proposed in literature in the last few years and used for MNIST classification are reported in Table 6. They should be compared with the present work, keeping in mind that the network proposed here is moved by the feasibility of an easy

hardware realization. As one can see, the best performance in terms of success rate is achieved using feedforward multilayer perceptron topology with backpropagation algorithm [174]. However, due to the complexity of the network and the use of a supervised and computationally demanding algorithm, this architecture does not represent the best option for an easy hardware implementation. In [175], a very good success rate is obtained with an array architecture using a STDP algorithm. That result was achieved with the additional use of two techniques known as Homeostasis, that dynamically adapts the spiking thresholds of the output neurons in order to uniform the firing activities, and Lateral Inhibition, that is the capability of an excited neuron to reduce the activity of its neighbors. Unfortunately, Homeostasis and Lateral Inhibition increase the complexity and the network is therefore definitely not suitable for easy hardware implementation. Finally, the architecture proposed in [176], features 784 input neurons and 300 output neurons, is based on RRAMs and uses a STDP unsupervised algorithm. It is very similar to the one proposed here in terms of topology and learning paradigm and achieves very good success rate (93%), but, again, it takes advantage of the Homeostasis technique which allows a better distribution of the neurons activity. The use of STDP learning paradigm in conjunction with Homeostasis makes the hard implementation likely more difficult than using STDP only. Anyway, in order to obtain an higher success rate for the MNIST recognition a more complex scheme is needed.

Table 6

	Present Work	[174]	[175]	[176]
Total Success Rate	85%	93%	95%	99.7%
Architecture	784x1000	784x300	784x6400	784x2500x2000x 1500x1000x500x10
Supervised/Unsupervised	Unsupervised	Unsupervised	Unsupervised	Supervised
Learning paradigm	STDP	STDP + *Homeostasis*	STDP + *Homeostasis* + *Lat. Inhibition*	Back Propagation
ReRAM Based	Yes	Yes	No	No
Suitable for Hardware	Yes	Yes?	No	No

3.4.3 Spiking Neural network with adapted STDP learining rule

A spiking neural network (SNN) has been simulated on a typical machine learning task: the recognition of the handwritten digits of the MNIST database [177]. In addition to neuronal and synaptic state, SNNs also incorporate the concept of time into their operating model. The idea is that neurons in the SNN do not fire at each propagation cycle, but rather fire only when a membrane potential reaches a specific threshold. When a neuron fires, it generates a signal which travels to other neurons which, in turn, increase or decrease their potentials in accordance with this signal. In the context of spiking neural networks, the current activation level (modeled as some differential equation) is normally considered to be the neuron's state, with incoming spikes pushing this value higher, and then either firing or decaying over time. Various coding methods

exist for interpreting the outgoing spike train as a real-value number, either relying on the frequency of spikes, or the timing between spikes, to encode information.

The proposed network has been implemented in the BrianSimulator software [178]. BrianSimulator is a simulating environment implemented in Python and it is an open-source library for simulating spiking neural networks. The Brian toolbox provides a flexible platform to customize parameters from network architectures, down to the differential equations that govern individual neuron and synapse dynamics.

The considered network architecture (Figure 3.20) consists of a single feedforward layer composed of 784 input neurons fully connected by plastic synapses to 100 outputs neurons (a total of 78400 synapses). The MNIST images are coded by Poisson input spikes with a spike frequency proportional to the intensity of the pixel. These Poisson-distributed spike trains are assumed to adequately provide the stochasticity needed to emulate noisy biological and hardware systems. In particular, the maximum spiking frequency has been set at 25Hz and the minimum at 0Hz. Every training image was presented for 250ms to the network. Neurons in the output layer have no lateral connection and are subdivided into pools of size 10, each selective to a particular digit. In addition to the signal from the input layer, the output neurons receive additional signals from inhibitory and teacher populations. The inhibitory population is composed by 20 neurons and it provides a signal proportional to the coding level of the stimulus and serves to balance the excitation coming from the input layer. A stimulus-dependent inhibitory signal is important, as it can compensate for large variations in the coding level of the stimuli. The non-plastic synapses shown in Figure 3.20 were used to weight

the contribution of the inhibitory population. They did not change their value during the training.

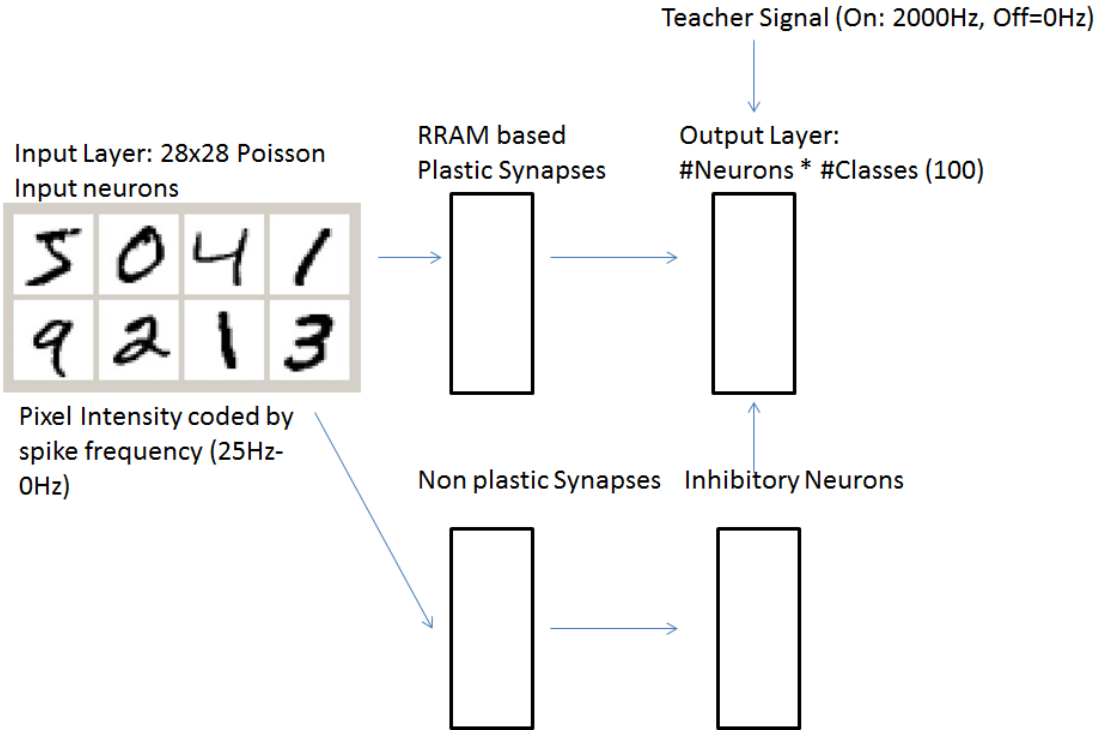


Figure 3.20 Sketch of the network architecture

Teacher signals are commonly used in machine learning tasks to train outputs by selectively applying a gain to the training neuron. The teacher population is active during training and imposes the selectivity of the output pools with an additional excitatory signal forcing the selected neurons to spike in a determined frequency range. The excitatory teacher signal controls the output neuron to raise the spiking frequency to a value where the probability of weight potentiation is high. Following learning, the response of the output neurons to a given stimulus can be analyzed by selecting proper threshold frequencies to determine which neurons are considered active (express a vote). The classification result is determined using a majority rule decision between the selective pools of output neurons.

The learning rules used in this network are based on ones proposed by Brader et al. in 2007 [179] that have been shown to execute STDP algorithms to solve classification problems both in software and hardware simulations. In this model, the internal synaptic strength is represented by the weight variable w, and is restricted to the range ($w_{min} > w > w_{max}$). The weight can update in the presence of presynaptic spikes; and the updates can push the weight towards long-term potentiation (LTP) if there is a positive weight step $\Delta w+$, or long-term depression (LTD) if there is a negative weight step $\Delta w-$. Referring to the neuron equations (3.2-3.5) described in the previous paragraph 3.2, weight updates can only occur if I_{mem} and I_{Ca}, which are both dependent on the past average firing rate, are within the appropriate range. When a presynaptic spike arrives at the soma, the weight updates according to the following rules:

$$\begin{aligned} w &\to w + \Delta w_{+} \ \ if \ \ I_{mem}\left(t_{pre}\right) > \vartheta_{mem} \ \ and \ \ \vartheta_{up}^{l} < I_{Ca}\left(t_{pre}\right) < \vartheta_{up}^{h} \\ w &\to w - \Delta w_{-} \ \ if \ \ I_{mem}\left(t_{pre}\right) \le \vartheta_{mem} \ \ and \ \ \vartheta_{down}^{l} < I_{Ca}\left(t_{pre}\right) < \vartheta_{down}^{h} \end{aligned} \quad , \qquad 3.8$$

Where $\Delta w+$ and $\Delta w-$ are the positive weight step sizes, θ_{mem} is the membrane threshold and θ^{l}_{up} ,θ^{hl}_{up} , θ^{l}_{down} , θ^{h}_{down} are the calcium variable thresholds.

Considering the synapses, the synaptic weights w can be linked to the RRAMs conductance values. The RRAM experimental evidences suggest that the device conductance can be incrementally adjusted by tuning the duration and sequence of the applied programming voltage. For example, the application of potentiating voltage pulses incrementally increases the resistive switching device conductance, and the application of depressing voltage pulses incrementally decreases the memory conductance. Among these interesting capabilities, some critical aspects of using RRAMs as synapse should be also taken in consideration. For example, these devices

suffer of variability of the states. In particular, applying the same programming or erasing (SET/RESET) voltage pulse, even at the same cell, it is possible to obtain different conductance values.

TiN/HfOx/Pt RRAMs devices were fabricated by the Institute of MicroElectronics and MicroSystems (IMM) of the Research National Council (CNR), Agrate, Brianza, Italy [180]. The devices were defined through the pattering of the top electrode into 40×40 μm^2 pads by optical lithography. A sketch of the device structure is reported in the inset of Figure 3.21 a). Devices were measured applying voltage or current to the top electrode and grounding the bottom electrode. The measurements were carried out mainly through a custom board interfacing the sample with the Source Pulse Generator Unit (SPGU) and the Source. Devices were subjected to trains of identical pulses and the resistance was read after each pulse at 200 mV. Pulses were 30 ms-long with time and fall times of 1 m s. Their maximum voltage was 1.1 V for LTD and -0.8 V for LTP. The state dependent switching has been verified on 10 different devices for 10 to 150 LTD/LTP cycles (not showed here).

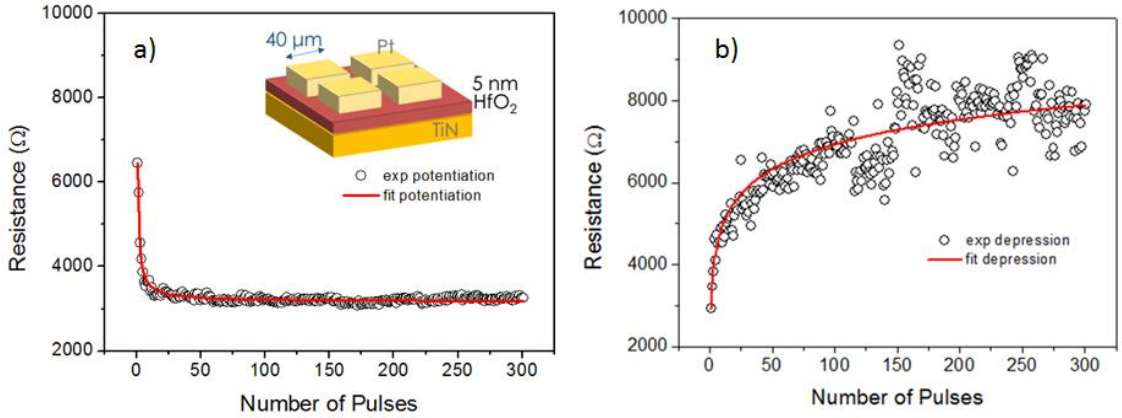


Figure 3.21 a) Potentiation curves obtained with 300 pulses at -0.8V; b) depression curves obtained with 300 pulses at 1.1V. Markers are experimental data and lines are the fitting.

In Figure 3.21 typical potentiation (LTP) and depression (LTD) curves are shown. To fit the two curves a fitting model was derived:

$$\begin{aligned} R &= R_{0,D} + \alpha_D \ln(n) \\ R &= R_{0,P} + \alpha_P \left(\frac{1}{n^{\beta}} - 1 \right), \end{aligned} \qquad 3.9$$

where the first equation is for the depression and the second is for potentiation. Referring to the symbols $R_{0,x}$, α_x and β are fitting parameters and n is the number of the applied pulses. To take into account the variability of the RRAM devices a stochastic terms was added in the simulations when the variability was included. In order to link the RRAM resistance value with the weight (w) of the synapse a simple inverse proportion relation was adopted.

Due to simulation time constrains, a reduced set of the complete MNIST database was used in order to evaluate the performances of network. 2000 Training images were used during the learning phase and the entire test set (10000 images) was used to test the recognition performances of the proposed architecture. After completing the training the teacher signals were switched off and a threshold per each pool was calculated based on the spiking frequency of the neurons during the presentation of a validation set of images (1000) taken from the training set. The test images were presented to the network for a duration of 350ms. The network was tested with and without the RRAM variability.

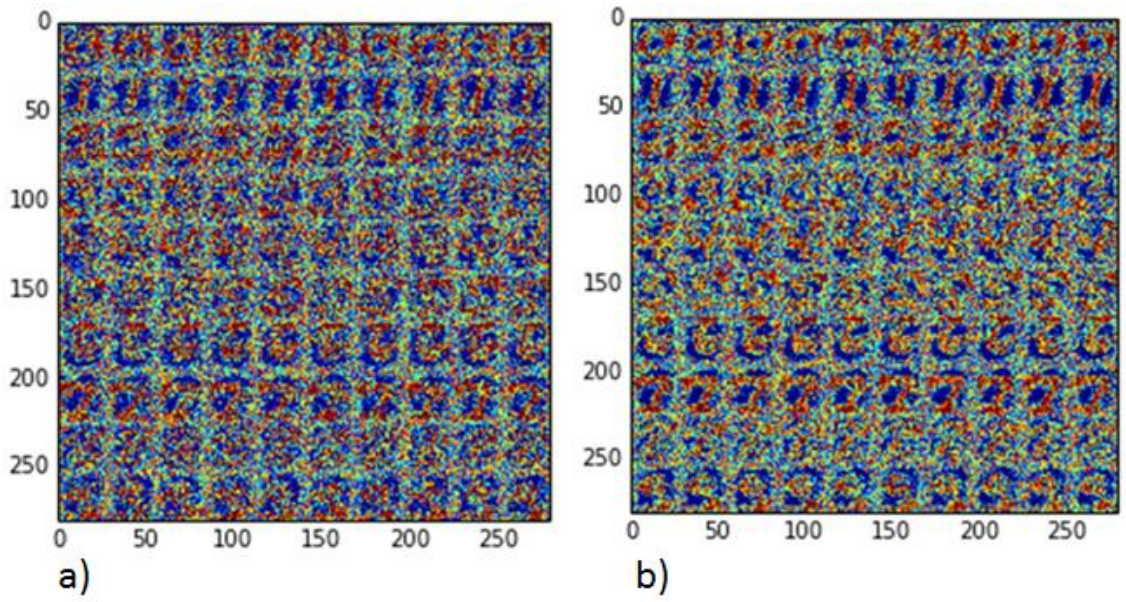


Figure 3.22 Features maps (weights of the synapses organized in a matrix form) of the trained network; a) with and b) without variability

In Figure 3.22 are depicted the features maps of the network (the weights of the synapses) for the two cases. On the left the case with variability (a) and on the right the case without (a). The first thing we notice is that some digits are clearly represented (numbers: 0, 1, 7), some other are possible to be distinguished (2,3,6 and 9) while numbers 8 and 5 are very confused. This is due to the large variability in these numbers of the MNIST database. As we can see, there is not an appreciable difference in the two features map. This suggests that the RRAM variability does not affect the behavior of the network. Unfortunately, the success rate for the recognition of the test set, averaged over 10 runs of the simulation, was 42%. This value is not comparable with success rate of networks that is possible to find in literature [168, 174-176]. However, we should keep in mind that these results are still remarkable considering the fact that we were simulating a network based on silicon neurons that emulate the biological response of real neurons and RRAM devices, including the experimental variability. Anyway, in

order to get better performances, further improvements are needed, in particular: an optimization of the learning algorithm, in order to fit better the behavior of the RRAM cells, a more sophisticated conversion of the of the RRAM conductance value and the synaptic weights and a more complex network architecture. Improvements on this study are still on going.

4 SUMMARY AND CONCLUSIONS

With the recent strong increase in the demand for data storage, NAND Flash has solidified its leading position among different storage devices for the smallest chip size and cheapest bit cost. The scaling in NAND Flash technology has progressed aggressively and successfully and during the past years, the dimension of the cell has shrunk dramatically. Due to this aggressive scaling, in the last decade the chip production cost has become cheaper year-to-year by ~40% and consequently the market demand for NAND Flash memory has rapidly increased in many new consumer electronic applications such as MP3, SSD, USB pen drives, Tablets, Smart Phones and Memory Cards. However, the competition of scaling down technology is getting harder in the NAND industry. The current NAND Flash cell, the Floating Gate (FG) structure, is facing new technological challenges approaching the 10nm dimension.

Major concerns are related to physical dimension, electrical isolation and read window margin. The key issue in scaling down the physical dimensions of the FG cell is that the scaling in horizontal direction cannot be followed by the scaling in the vertical dimension. The electrical gate coupling factor in FLASH memories should be kept higher than 0.6 in order to achieve good control. To get this coupling value, scaling of tunnel oxide and IPD thickness is very critical and has to be managed very conservatively. Furthermore, the scaling of the cell dimension and consequently the reduction of the floating gate capacitance reduces the number of electrons stored in the FG per unit voltage and the reliability strongly deteriorates as the number of electrons available for data storage becomes less than 10. As the design rule of NAND Flash memory is scaled down, parasitic capacitors surrounding the floating gate influence the FG voltage more and more. Hence, a phenomenon called "floating-gate interference," or cross-talk occurs, in which change of a cell accompanies threshold voltage shift of the adjacent cells by floating-gate voltage shift. In this scenario, in order to keep pursuing the scaling, there are two guide lines to follow: the optimization of the FG basic cell and array architecture and the proposal of new device structures with innovative operating concepts. Both the guide lines are currently experienced worldwide. So, in the last years, innovative concepts have been proposed alternative to the conventional FG technology. Among these, Resistive Switching Memories, often referred as RRAM are one of the most promising candidates for Non-Volatile Memory applications.

The expectation for RRAMs is that it will be a memory technology that can be easily integrated with conventional CMOS technology, using materials compatible with the conventional CMOS fabrication environment and process temperatures that allow it to

be fabricated in the Back End Of Line (BEOL). The ultimate target is integration on the metal layers or within the contact vias to the source and drain of a metal– oxide–semiconductor field-effect transistor (MOSFET) of a CMOS chip. In its basic form, the device structure is an oxide material sandwiched between two metal electrodes, called bottom and top electrodes (BE and TE). The stack is a well-known metal–insulator–metal (MIM) structure. A wide range of binary metal oxides have been found to show resistive switching phenomena. Most of them are transition metal oxides.

According to the current-voltage (I-V) characteristics, it is possible to classify the RS devices in two general categories: unipolar (nonpolar) and bipolar: in unipolar resistive switching mode the switching direction does not depend on the polarity of the applied voltage and generally occurs at higher voltage amplitude that of bipolar switching, where the memory operations are polarity dependent. A fresh memory device with high initial resistance state can be switched in to a low-resistance state (LRS) by applying a proper voltage. This process occurs just one time and it is called forming. After the forming process, the memory cell can be switched to a high-resistance state (HRS), generally lower than the initial resistance, by the application of a particular voltage called reset voltage. This process is called RESET. Switching again from a HRS to a LRS is called 'SET'. In the SET process, generally, the current is limited by an external current compliance in order to avoid irreversible device damage. In bipolar resistive switching mode, the SET and RESET occur in the opposite polarity.

The first big advantage of the RRAM, compared to the FLASH technology, is the low voltage needed to program and erase (SET/RESET) the cells. While for FLASH more than 15V are needed, for RRAMs voltages lower than 3V are enough to perform the

SET and RESET operations. This means that high voltage generation peripheral circuitry is not needed, thus reducing drastically the area occupied by the total memory chip. Moreover, this beneficial aspect is enhanced by the capability of RRAM to be stacked in the BEOL of the CMOS process. The second important advantage of RS devices is the high speed of data reading, but especially the very high speed of writing/erasing. Compared to FLASH, they can be 1000 times faster. While the data retention values are comparable for FLASH and Resistive Switching memories, the endurance for RRAMs can be even 8 orders of magnitude greater than the NAND: 10^{12} SET/RESET cycles have been demonstrated.

It is widely accepted that the physical mechanisms governing the switching phenomena in the oxide RRAM are based on the migration of oxygen vacancies and ions with related electrochemical reactions. This ion drift is responsible for the formation and modification of a conducting filament (CF) between electrodes. During the RESET a gap region is created and the conductive filament is partially ruptured. The size of the gap between the tip of the filament and the electrode is the state variable determining the device resistance value.

While most of the scientific community observes a linear or ohmic behavior in the LRS, on the HRS different electron transport mechanisms interpretations are reported: Poole–Frenkel emission, Schottky emission, the space charge limited current (SCLC) characteristic, quantum point contact depending on the material properties. The dielectric bandgap, the defect trap proprieties and the fabrication process conditions like the depositions of both metals and dielectric and the interface between the oxides and the electrodes proprieties.

Although RRAM devices show very promising aspects for the next future Non Volatile Memory, still there are some open issues. In particular, the key point of RRAM cells is the variability of the switching parameters. Due to the stochastic nature of ionic migration, defects characteristics and material properties like the grain boundaries in polycristalline materials, the filament shape varies from device to device and also from cycle to cycle in the same device.

In this scenario, the present Ph.D. thesis studies systematically RRAM operation features from both the experimental and modeling viewpoints and explores the potentialities of RRAM artificial synapsis in neuromorphic application. Key aspects of the behavior of the cell were addressed. Experimental characterization and modeling of the influence of the applied voltage waveform, device geometry, external temperature, and electrical stress were performed. An introduction of the neuromorphic bio-inspired architectures was given in terms of biological basics, neurons and synapses electrical models and circuits. Bio-inspired learning paradigm Spike Time Dependent Plasticity (STDP) was described and adopted for simulations. Different approaches in RRAM based networks architectures and simulations were described and results on visual pattern recognition were discussed.

In the tested devices HfO_2 was used as insulating layer, while different combination of metal electrodes, including Platinum (Pt), Titanium (Ti) and Titanium Nitride (TiN) were adopted. To initiate switching a preliminary forming operation is required and it appeared that the Ti presence strongly decreases the forming voltage. Pt-Pt samples were found to be unipolar (non-polar), while samples TiN-Pt, TiN-TiN and TiN-Ti were bipolar. They required positive voltage applied on TiN electrode except in case

with Ti where devices need positive voltage applied to Ti to achieve SET. This behavior suggested that the presence of an interlayer rich of oxide vacancies plays a fundamental role in determine the polarity of the switching of the devices. The influence of the voltage ramp speed (rate) defined as the ratio $RR=\Delta V/\Delta t$ on the switching voltage was investigated. It was found that both $|V_{RESET}|$ and V_{SET} increased with the front speed following a logarithmic behaviour with same slope. The fact that the dynamics for SET and RESET was the same can be attributed to the predominance of the ion drift diffusion over other mechanisms, for both the closure and dissolution of the conductive filament. $|V_{RESET}|$ data for the samples with Pt as top electronde (C1) and samples with Ti as top electrode (C2) (both with TiN as bottom electrode) were compared as function of the RR, at room temperature. The C1 curve stands at higher values than the C2 curve due to the different barrier energy to extract oxygen ions from the TiON and TiOx interlayers. Both the curves increased with the ramp rate and they were parallel, indicating that the dynamics of filament interruption is the same, governed by the same hopping barrier of ion diffusion, and is not influenced by the electrode.

Regarding the RESET dynamics, it was demonstrated that the progressive reset transition between the low and high resistance states in HfO_2-based RS devices could be modeled using a diode-like conduction mechanism with a series resistance combined with a generalized logistic model for the pre-exponential diode amplitude factor. It was shown that a Verhulst logistic model does not provide accurate results. The proposed dynamics was interpreted in terms of the sequential deactivation of multiple conduction channels spanning the dielectric film. Fitting results indicated that the switching behavior dynamics could be described with the same equation regardless of the voltage

sweep rate. Partial reset curves were characterized and modeled using the same approach, thus giving further support to the proposed picture. This part of the work was done in collaboration with Prof. E. Miranda at the Autonomous University of Barcelona, where I spent three months working on this topic.

The influence of the device geometry on switching behaviour was investigated. No dependence of RESET/SET voltages and LRS/HRS on the cell dimensions on was found. This clearly confirmed switching being driven by generation-disruption of a conductive filament. While the device area size had no influence on the SET and RESET characteristics, a strong influence on the Forming operation was found. For smaller device the forming was higher. The lower number of traps in devices with smaller area explained this behaviour. As a consequence, the time required for the filament forming is higher.

Temperature impact on the forming dynamics and current conduction during RESET was addressed. A thermally activation Arrhenius dependence for the forming time was found. The comparison between different samples highlighted the activation energy dependence on the MIM stack composition. In particular higher activation energy for Pt-Pt sample was found. The conductance of HfO_2 films in C1 (Pt-Pt) samples was found to be lower than in C2 (TiN-Ti) ones and increased with temperature, while the conductance of C2 samples did not change with temperature. The higher conductance of C2 samples was explained recalling that the effective thickness of the dielectric film is reduced respect to C1 samples and considering the different microscopic phase of the HfO2 in the filament for the two kind of samples.

Specific aspects related to the RRAM reliability were also examined, as the case of electrical constant voltage stress (CVS). An experiment of read disturb was performed and the role of the forming conditions and the electrode materials on the robustness of the low resistance state against electrical disturb was investigated. It was performed on two sets of samples with different bottom electrodes (TiN and Pt) and the same Pt TE. The stress included three steps: 1) The samples were formed using voltage pulses with different amplitudes; 2) formed samples were subject to CVS of different entities; 3) the current flowing through the MIM during stress was monitored and processed. As a result, it was found that: a) the electrode material has an impact on the stability of the low resistance state, since the mechanism involved in the interruption of the conductive path is different if the electrode is inert or reactive with HfO_2; b) samples with inert electrodes more easily switch into a high resistance state; c) samples formed with higher forming pulses are more robust (as if a stronger filament is formed) and it is more difficult to interrupt it; d) switching to a higher resistance state is favored by increasing the stress entity; e) the value of the (high) resistance achieved by the failing device depends on the forming condition in the sense that the portion (or portions) of filament undergoing rupture decreases for stronger forming, and the corresponding value of resistance is lower.

A similar experiment was performed also on the TiN (BE)-Ti (TE) samples. The resistance evolution under constant voltage stress of both the low resistive state and the high resistive state of the TiN-Ti cells was studied from an experimental and theoretical point of view. An analytical model was used to simulate the resistance degradation in the LRS and HRS during CVS. The degradation during stress was interpreted in terms

of an increase or decrease of the tunnelling gap, that is the distance between the tip of the conductive filament inside the oxide and the metal electrode. The gap evolution was ruled by oxygen ions hopping and oxygen vacancies generation. Differences in the dynamic of the processes for the two resistive states were observed and modeled. In the HRS stress the gap increasing was gradual, while the gap decreasing in the LRS stress was abrupt and it occurred in a wide time range. This phenomena were interpreted in terms of negative or positive feedback related to the enhancement factor of the electric field. The variability of parameters were also investigated and a stronger influence in the LRS stress was found.

Finally, possible neuromorphic applications were investigated. Neuromorphic computing refers to an emerging interdisciplinary field that takes its inspiration from biological neural architectures and computations occurring inside the brain or the cerebral cortex. It comprises principles and knowledge from neurobiology, computational neuroscience, computer science, machine learning, VLSI circuit design, and more recently nanotechnology. In non-Von Neumann neuromorphic architectures, computing hardware and processing are not completely isolated tasks. Memory is intelligent and participates in processing of information. Architectures abstracted from neural systems can provide a promising solution to different cognitive tasks like the object detection and the speech recognition, inspiring new paradigms, which can drastically improve the performance and efficiency of computing systems.

Due to the configurable and multiparallel architecture constituted by a complex network of ~10^{12} neurons and 10^{15} synapses, the human brain is able to perform a wide range of cognitive tasks with a power consumption of less than 20 W. This value is much lower

respect to the modern multi-core based computers that require 10,000 times more power. The superior efficiency of the brain in performing fuzzy and fault-tolerant computation has motivated engineers to mimic the key algorithmic and computational features of the brain in software and silicon-based hardware. Dedicated hardware Artificial Neural Networks (ANN) architectures, compared to software implementations, can offer very high computational speed at very low cost with high reliability, due to the redundancy of the architecture. In this scenario, Non-Volatile Memories synapses stand out for their excellent power management and integration due to their small sizes that allow high density structures. RRAMs are considered among the most promising devices to realize neuromorphic systems capable of implement learning algorithms based on pulses. This is due to several advantages offered by this technology: a) the simplicity of operation and low energy consumption (~0.1 pJ/bit), b) the compatibility with CMOS technology and the possibility of 3D crossbar integration in back-end-of-line, c) the good scaling properties of the elemental cell (minimum size below 10 nm), d) the excellent reliability (~10^{12} SET/RESET cycles), e) the short programming time (~ns) and the online learning algorithms in which appropriate pulses update the synaptic weights.

Neurons and synapses are the the basic functional units of the nervous system. Connections between the neurons are made through synapses. One of the most diffused and widely accepted model of the synaptic plasticity is the so called: Spike Timing Dependent Plasticity. According to this learning paradigm, the synaptic weights update according to relative spiking time of pre and post-synaptic neurons. According to STDP, when repeated pre-synaptic spikes arrive just before post-synaptic spikes the

synapses weights increases following a long-term potentiation (LTP). On the contrary, repeated spikes arriving after post-synaptic spikes leads to long-term depression (LTD) of the same synapses weights. Thus, inputs that might be the cause of the post-synaptic neuron's excitation are made even more likely to contribute in the future, whereas inputs that are not the cause of the post-synaptic spikes are made less likely to contribute in the future. The Leaky Integrate and Fire (LIF) model descripts the behavior of biological neurons. In this model, a neuron is represented by a simple capacitive and resistive differential equation. The spiking events are not explicitly modelled in the LIF model. Instead, the neuron constantly sums or integrates the incoming pre-synaptic currents and when the membrane potential reaches a certain threshold voltage (Vth) it instantaneously fire, generating an action potential. From a circuital point of view the LIF neuron can be implemented with an integrator, a comparator, switches, and eventually control logic. In application where we want to include more biological aspects in the neuron model a deeper description is needed. Subthreshold CMOS circuits are an option to emulate neurons and synapses behavior. MOSFETs have an exponential relationship between drain current I_D and gate voltage V_{GS}, at very low currents regime. The time constants of MOS subthreshold circuits are inversely proportional to the reference currents and directly proportional to the circuit capacitance, thus allowing the integration of relatively small capacitors in integrated circuits in order to implement temporal circuits that are both compact and have biologically realistic time constants, ranging from tens to hundreds of milliseconds.

Two-terminal RRAM devices have been proposed as artificial synapses in neuromorphic circuits thanks to the capability of analog gradual resistance modulation,

strong device area scalability, low power consumption, compatibility with CMOS technology combined with a 2-terminal structure. Due to these characteristics, RRAMs naturally satisfy the requirements to act as a connection for communication between a pre-synaptic neuron and a post-synaptic neuron. When RRAMs are employed in neuromorphic networks, two main operational modes are used, binary and analog. The binary operations is based on the two limits states HRS and LRS. At the same time, gradual analog resistance modulation is desirable to improve the performances of the network. The RRAM resistance state can be considered inversely proportional to the synaptic weight *w*.

In the Ph.D. thesis, different approaches were considered in neuromorphic networks simulations with RRAM based synapses for visual pattern recognition applications: a) PSPICE simulation of a small network, b) Matlab simulations of a large network based on STDP and c) Brian (Python) simulations of a large array based on a modified version of the STDP.

A neuromorphic network composed by a first 25 sensory neurons layer and a second layer made of 10 output neurons connected through 250 RRAM synapses was considered in PSPICE simulations. TiN/HfO2/Ti RRAMs were used as synapses. Starting from gap dynamics and current conduction equations a PSPICE behavioral model was implemented in order to perform circuital simulations. The RRAM device was described with two main blocks: the first one was an integrator composed by a capacitor and a source of current. This part was devoted to integrate the differential equation related to the gap dynamics. The second block was a voltage controlled current source that represents the conduction current dependent on the actual value of the gap.

The entire system was adopted in order to recognize 5X5 pixel binary images representing numbers between “0” and “9”. The network operations were divided in two distinct phases: the training and the testing ones. For the training, a simple protocol was adopted: every neuron was trained separately to recognize one of the 10 numbers just applying training pulses of -1.1 V in correspondence of black pixels in the number image. During the testing phase an input image is applied. Low voltage pulses (0.1V), corresponding to black pixels, were applied to RRAM previously trained and integrated by output neurons. When the voltage across the capacitor reached the threshold voltage the winner neuron stopped the others through a feedback network and the recognition was completed. The performances of the circuit were tested in two ways. The first one was to apply the same inputs image adding a noise. It was found that if the noise level was under the 10% (corresponding to 2 pixels) the network recognized correctly all the patterns. The second way to evaluate the performance of the network was to build an home-made test set. A success rate of the 44% was achieved with the test set used. Obviously, in order to increase the performances of the network, it was mandatory to improve the learning paradigm during the training phase. Furthermore, the adoption of a meaningful data set was alos fundamental to give robustness to the network. To this aim, a fully connected brain inspired neuromorphic network composed by 784 input neurons, M output neurons (with M ranging between 100 and 2000) and 784xM HfO_2 Resistive RAM synapses is presented here. The network was trained in Matlab system level simulations using an unsupervised algorithm of the class “winner take all” with 60000 images of the MNIST database in order to recognize handwritten digits between “0” and “9”. The neuromorphic architecture was tested with 10000 testing images

without updating the synapses. The measured high and low resistance values of the Resistive RAMs fix the boundaries of the synapses weights. A success rate of 85 % over 10^4 test images representing the entire test dataset was obtained at best.

In order to explore networks that were closer to the biological behavior, a spiking neural network (SNN) was simulated on the recognition of the handwritten digits of the MNIST database. This work was done in collaboration with the S.Spiga group of IMM group of the CNR, Agrate and the prof. G. Indiveri of the Institute of NeuroInformatics, ETH and UZH, Zurich, Switzerland where I spent three months working on this topic. In addition to neuronal and synaptic state, SNNs also incorporate the concept of time into their operating model. The network was implemented in the BrianSimulator software. The considered network architecture consisted of a single feedforward layer composed of 784 input neurons fully connected by plastic synapses to 100 outputs neurons (a total of 78400 synapses). The MNIST images were coded by Poisson input spikes with a spike frequency proportional to the intensity of the pixel. Every training image was presented for a certain time to the network. Neurons in the output layer have no lateral connection and are subdivided into pools of size 10, each selective to a particular digit. In addition to the signal from the input layer, the output neurons received additional signals from inhibitory and teacher populations. Following learning, the response of the output neurons to a given stimulus can be analyzed by selecting proper threshold frequencies to determine which neurons are considered active (express a vote). The classification result is determined using a majority rule decision between the selective pools of output neurons. The learning rules used in this network were based on ones proposed by Brader et al. in 2007 that is a modified version of the

STDP paradigm. In particular it is based on a set of thresholds to be compared with the neurons membrane potential and an auxiliary variable Calcium linked to the post-synaptic firing activity. Considering the synapses, the synaptic weights w can be linked to the RRAMs conductance values. The RRAM experimental evidences suggested that the device conductance could be incrementally adjusted by tuning the duration and sequence of the applied programming voltage obtaining potentiation (LTP) and depression (LTD) curves. At the same time an intrinisc variability is present, in particular, applying the same programming or erasing (SET/RESET) voltage pulse, even at the same cell, it is possible to obtain different conductance values. To fit the behavior of the RRAM synapses a fitting model was derived and a stochastic contribution was added to take in to account the variability. Due to simulation time constrains, a reduced set of the complete MNIST database was used in order to evaluate the performances of network. The test images were presented to the network for a duration of 350ms. The network was tested with and without the RRAM variability. The features maps of the network highlighted that a learning was performed after the training, since some digits were clearly represented, while some other were distinguishable. Considering or not the variability had no effect on the network performances. Unfortunately, the success rate for the recognition of the test set, averaged over 10 runs of the simulation, was only 42%. However, we should keep in mind that this result is still remarkable considering the fact that we were simulating a network based on silicon neurons that emulate the biological response of real neurons and RRAM devices, including the experimental variability. Anyway, in order to get better performances, further improvements are needed, in particular: an optimization of

the learning algorithm, in order to fit better the behavior of the RRAM cells, a more sophisticated conversion of the of the RRAM conductance value and the synaptic weights and a more complex network architecture.

5 PUBLICATIONS

Journals:

- P. Lorenzi, R. Rao and F. Irrera*; "Forming Kinetics in HfO2 -Based RRAM Cells"*, IEEE Transactions on Electron Devices, vol. 60, pp. 438-443, (2013)
- P. Lorenzi, R. Rao and F. Irrera*; "Role of the electrode metal, waveform geometry, temperature, and postdeposition treatment on SET and RESET of HfO2-based resistive random access memory 1R-cells: Experimental aspects"*, Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena, vol. 33, pp. 01A107, (2015)
- P. Lorenzi, R. Rao, F. Irrera, J. Suñé and E. Miranda*; "A thorough investigation of the progressive reset dynamics in HfO2-based resistive switching structures"*, Applied Physics Letters, vol. 107, pp. 113507, (2015)
- P. Lorenzi, R. Rao, T. Prifti and F. Irrera*; "Impact of the forming conditions and electrode metals on read disturb in HfO2-based "*, Microelectronics Reliability, vol. 53, pp. 1203 - 1207, (2013)

- P. Lorenzi, R. Rao and F. Irrera*; "Conductive filament evolution in HfO2 resistive* device during constant voltage stress", Microelectronics Reliability, vol. 55, pp. 1446 - 1449, (2015)

Conferences:

- *Lorenzi, P; Rao, Rosario; Irrera, F , European Symposium on Reliability of Electron Devices, Failure Physics and Analysis. ESREF 2013. September 30th - October 4th 2013. Arcachon – France; "Impact of the forming conditions and electrode metals on read disturb in HfO 2-based RRAM"*
- *Lorenzi, P.; Rao, R.; Irrera, F. , Workshop on Dielectrics in Microelectronics, WoDiM 2014, 09/06/2014-11/06/2014, Kinsale, Ireland; "Role of the electrode metal, waveform geometry, temperature, and postdeposition treatment on SET and RESET of HfO2-based resistive random access memory 1R-cells: Experimental aspects"*
- *Lorenzi, P; Sucre V., Rao, R.; Irrera, F. International Conference on Memristive Systems, MEMRISYS 2015, 08/11/2015 – 10/11/2015, Paphos, Cyprus, "Memristor based neuromorphic circuit for visual pattern recognition"*
- *Lorenzi, P; Rao, R.; Irrera, F. , European Symposium on Reliability of Electron Devices, Failure Physics and Analysis. ESREF 2015. /, 2015 – 05/10/2015 – 09/10/2015, Toulouse – France, "Conductive filament evolution in HfO 2 resistive RAM device during constant voltage stress"*

- *Romano, G.; Lorenzi P., Rao, R., F. Irrera,, Workshop on Dielectrics in Microelectronics, WoDiM 2016, 23/06/2016 – 26/06/2016, Catania, Italy, "HfO2-ReRAM based Neuromorphic Architecture for Digit Recognition"*

Conference (accepted):

- *J. Frascaroli, S. Brivio, E. Covi, P. Lorenzi, M. V. Nair, F. Irrera, G. Indiveri, S. Spiga, EMRS 2016, 22/05/2017 – 26/05/2017, Strasbourg, France, "HfO 2-Based Devices for Neuromorphic Network with Real-Time Operation"*

6 REFERENCES

[1] ITRS Roadmap of Semiconductors, http://www.semiconductors.org/main/2015_international_technology_roadmap_for_semiconductors_itrs/

[2] S. Lee; *"Scaling Challenges in NAND Flash Device toward 10nm Technology"*, 2012 4th IEEE International Memory Workshop, pp. 1-4, (2012)

[3] J. Choi and K. S. Seol; *"3D approaches for non-volatile memory"*, 2011 Symposium on VLSI Technology - Digest of Technical Papers, pp. 178-179, (2011)

[4] S. Hong; *"Memory technology trend and future challenges"*, 2010 International Electron Devices Meeting, pp. 12.4.1-12.4.4, (2010)

[5] Y. Park, J. Lee, S. S. Cho, G. Jin and E. Jung; *"Scaling and reliability of NAND flash devices"*, 2014 IEEE International Reliability Physics Symposium, pp. 2E.1.1-2E.1.4, (2014)

[6] J. Seo, K. Han, T. Youn, H. E. Heo, S. Jang, J. Kim, H. Yoo, J. Hwang, C. Yang, H. Lee, B. Kim, E. Choi, K. Noh, B. Lee, B. Lee, H. Chang, S. Park, K. Ahn, S. Lee, J. Kim and S. Lee; *"Highly reliable M1X MLC NAND flash memory cell with novel active air-gap and p #x002B; poly process integration technologies"*, 2013 IEEE International Electron Devices Meeting, pp. 3.6.1-3.6.4, (2013)

[7] J.-D. Lee, C.-K. Lee, M.-W. Lee, H.-S. Kim, K.-C. Park and W.-S. Lee; *"A New Programming Disturbance Phenomenon in NAND Flash Memory By Source/Drain Hot-Electrons*

Generated By GIDL Current", 2006 21st IEEE Non-Volatile Semiconductor Memory Workshop, pp. 31-33, (2006)

[8] C. Technologies; *"CTWP010: SLC vs MLC NAND and The Impact of Technology Scaling White Paper"*

[9] J.-D. Lee, S.-H. Hur and J.-D. Choi; *"Effects of floating-gate interference on NAND flash memory cell operation"*, IEEE Electron Device Letters, vol. 23, pp. 264-266, (2002)

[10] J.-D. Lee, J.-H. Choi, D. Park and K. Kim; *"Data retention characteristics of sub-100 nm NAND flash memory cells"*, IEEE Electron Device Letters, vol. 24, pp. 748-750, (2003)

[11] A. Fayrushin, K. Seol, J. Na, S. Hur, J. Choi and K. Kim; *"The new program/erase cycling degradation mechanism of NAND flash memory devices"*, 2009 IEEE International Electron Devices Meeting (IEDM), pp. 1-4, (2009)

[12] J. Jang, H. S. Kim, W. Cho, H. Cho, J. Kim, S. I. Shim, Younggoan, J. H. Jeong, B. K. Son, D. W. Kim, Kihyun, J. J. Shim, J. S. Lim, K. H. Kim, S. Y. Yi, J. Y. Lim, D. Chung, H. C. Moon, S. Hwang, J. W. Lee, Y. H. Son, U. I. Chung and W. S. Lee; *"Vertical cell array using TCAT(Terabit Cell Array Transistor) technology for ultra high density NAND flash memory"*, 2009 Symposium on VLSI Technology, pp. 192-193, (2009)

[13] W. Kim, S. Choi, J. Sung, T. Lee, C. Park, H. Ko, J. Jung, I. Yoo and Y. Park; *"Multi-layered Vertical Gate NAND Flash overcoming stacking limit for terabit density storage"*, 2009 Symposium on VLSI Technology, pp. 188-189, (2009)

[14] B. D. Salvo, C. Gerardi, R. van Schaijk, S. A. Lombardo, D. Corso, C. Plantamura, S. Serafino, G. Ammendola, M. van Duuren, P. Goarin, W. Y. Mei, K. van der Jeugd, T. Baron, M. Gely, P. Mur and S. Deleonibus; *"Performance and reliability features of advanced nonvolatile memories based on discrete traps (silicon nanocrystals, SONOS)"*, IEEE Transactions on Device and Materials Reliability, vol. 4, pp. 377-389, (2004)

[15] C. W. Oh, N. Y. Kim, S. H. Kim, Y. L. Choi, S. I. Hong, H. J. Bae, J. B. Kim, K. S. Lee, Y. S. Lee, N. M. Cho, D. W. Kim, D. Park and B. I. Ryu; *"4-Bit Double SONOS Memories (DSMs) Using Single-Level and Multi-Level Cell Schemes"*, 2006 International Electron Devices Meeting, pp. 1-4, (2006)

[16] M. K. Kim, S. D. Chae, H. S. Chae, J. H. Kim, Y. S. Jeong, J. W. Lee, H. Silva, S. Tiwari and C. W. Kim; *"Ultrashort SONOS memories"*, IEEE Transactions on Nanotechnology, vol. 3, pp. 417-424, (2004)

[17] A. Subirats, A. Arreghini, R. Degraeve, D. Linten, G. V. den bosch and A. Furnemont; *"In Depth Analysis of Post-Program VT Instability after Electrical Stress in 3D SONOS Memories"*, 2016 IEEE 8th International Memory Workshop (IMW), pp. 1-4, (2016)

[18] P. Lorenzi, R. Rao, F. Palma, F. Irrera and G. Ghidini; *"Charge trapping NanoElectronic memory"*, 2009 9th IEEE Conference on Nanotechnology (IEEE-NANO), pp. 214-217, (2009)

[19] S. M. Jung, J. Jang, W. Cho, H. Cho, J. Jeong, Y. Chang, J. Kim, Y. Rah, Y. Son, J. Park, M. S. Song, K. H. Kim, J. S. Lim and K. Kim; *"Three Dimensionally Stacked NAND Flash Memory Technology Using Stacking Single Crystal Si Layers on ILD and TANOS Structure for Beyond 30nm Node"*, 2006 International Electron Devices Meeting, pp. 1-4, (2006)

[20] A. Padovani, L. Larcher and P. Pavan; *"Compact modeling of* program/erase operations for SPICE-like circuit simulations", Microelectronics Journal , vol. 44, pp. 50 - 57, (2013)

[21] X. Fong, R. Venkatesan, D. Lee, A. Raghunathan and K. Roy; *"Embedding Read-Only Memory in Spin-Transfer Torque MRAM-Based On-Chip Caches"*, IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 24, pp. 992-1002, (2016)

[22] K. Narayanapillai, X. Qiu, Y. Wang, J. Kwon, J. Yu, L. M. Loong, W. Legrand, J. Yoon, K. Banerjee and H. Yang; *"Spin-transfer versus spin-orbit torque MRAM"*, 2016 IEEE International Nanoelectronics Conference (INEC), pp. 1-2, (2016)

[23] J. H. Park, H. Y. Kim, G. S. Jang, D. Ahn and S. K. Joo; *"Ultimate multibit 1T-FeRAM with selectively nucleated grown single-grain PbZr 0.52 Ti 0.48 O 3 for very-large-scale-integrated memory"*, Journal of Physics D: Applied Physics, vol. 49, pp. 075106, (2016)

[24] T. W. Hickmott; *"Low-Frequency Negative Resistance in Thin Anodic Oxide Films"*, Journal of Applied Physics, vol. 33, pp. 2669-2682, (1962)

[25] J. Gibbons and W. Beadle; *"Switching properties of thin Nio films"*, Solid-State Electronics, vol. 7, pp. 785 - 790, (1964)

[26] G. Dearnaley, A. M. Stoneham and D. V. Morgan; *"Electrical phenomena in amorphous oxide films"*, Reports on Progress in Physics, vol. 33, pp. 1129, (1970)

[27] J. G. Simmons; *"Conduction in thin dielectric films"*, Journal of Physics D: Applied Physics, vol. 4, pp. 613, (1971)

[28] Y. Watanabe, J. G. Bednorz, A. Bietsch, C. Gerber, D. Widmer, A. Beck and S. J. Wind; *"Current-driven insulator–conductor transition and nonvolatile memory in chromium-doped SrTiO3 single crystals"*, Applied Physics Letters, vol. 78, pp. 3738-3740, (2001)

[29] A. Beck, J. G. Bednorz, C. Gerber, C. Rossel and D. Widmer; *"Reproducible switching effect in thin oxide films for memory applications"*, Applied Physics Letters, vol. 77, pp. 139-141, (2000)

[30] S. Seo, M. J. Lee, D. H. Seo, E. J. Jeoung, D.-S. Suh, Y. S. Joung, I. K. Yoo, I. R. Hwang, S. H. Kim, I. S. Byun, J.-S. Kim, J. S. Choi and B. H. Park; *"Reproducible resistance switching in polycrystalline NiO films"*, Applied Physics Letters, vol. 85, pp. 5655-5657, (2004)

[31] C. Rohde, B. J. Choi, D. S. Jeong, S. Choi, J.-S. Zhao and C. S. Hwang; *"Identification of a determining parameter for resistive switching of TiO2 thin films"*, Applied Physics Letters, vol. 86, pp. 262907, (2005)

[32] I. G. Baek, M. S. Lee, S. Seo, M. J. Lee, D. H. Seo, D. S. Suh, J. C. Park, S. O. Park, H. S. Kim, I. K. Yoo, U. I. Chung and J. T. Moon; *"Highly scalable nonvolatile resistive memory using simple binary oxide driven by asymmetric unipolar voltage pulses"*, IEDM Technical Digest. IEEE International Electron Devices Meeting, 2004., pp. 587-590, (2004)

[33] H. S. P. Wong, H. Y. Lee, S. Yu, Y. S. Chen, Y. Wu, P. S. Chen, B. Lee, F. T. Chen and M. J. Tsai; *"Metal #x2013;Oxide RRAM"*, Proceedings of the IEEE, vol. 100, pp. 1951-1970, (2012)

[34] M. Marinella; *"The future of memory"*, 2013 IEEE Aerospace Conference, pp. 1-11, (2013)

[35] C. Kügeler, R. Weng, H. Schroeder, R. Symanczyk, P. Majewski, K.-D. Ufert, R. Waser and M. Kund; *"Study on the dynamic resistance switching properties of NiO thin films "*, Thin Solid Films , vol. 518, pp. 2258 - 2260, (2010)

[36] A. Chen, S. Haddad, Y. C. Wu, T. N. Fang, S. Kaza and Z. Lan; *"Erasing characteristics of Cu2O metal-insulator-metal resistive switching memory"*, Applied Physics Letters, vol. 92, pp. 013503, (2008)

[37] T. N. Fang, S. Kaza, S. Haddad, A. Chen, Y. C. Wu, Z. Lan, S. Avanzino, D. Liao, C. Gopalan, S. Choi, S. Mahdavi, M. Buynoski, Y. Lin, C. Marrian, C. Bill, M. VanBuskirk and

M. Taguchi; *"Erase mechanism for copper oxide resistive switching memory cells with nickel electrode"*, 2006 International Electron Devices Meeting, (2006)

[38] W. C. Chien, Y. C. Chen, K. P. Chang, E. K. Lai, Y. D. Yao, P. Lin, J. Gong, S. C. Tsai, S. H. Hsieh, C. F. Chen, K. Y. Hsieh, R. Liu and C. Y. Lu; *"Multi-Level Operation of Fully CMOS Compatible WOX Resistive Random Access Memory (RRAM)"*, 2009 IEEE International Memory Workshop, pp. 1-2, (2009)

[39] B. Jiao, N. Deng, J. Yu, Y. Bai, M. Wu, Y. Zhang, H. Qian and H. Wu; *"Resisitive switching variability study on 1T1R AlOx/WOx-based RRAM array"*, 2013 IEEE International Conference of Electron Devices and Solid-state Circuits, pp. 1-2, (2013)

[40] M.-J. Lee, S. I. Kim, C. B. Lee, H. Yin, S.-E. Ahn, B. S. Kang, K. H. Kim, J. C. Park, C. J. Kim, I. Song, S. W. Kim, G. Stefanovich, J. H. Lee, S. J. Chung, Y. H. Kim and Y. Park; *"Low-Temperature-Grown Transition Metal Oxide Based Storage Materials and Oxide Transistors for High-Density Non-volatile Memory"*, Advanced Functional Materials, vol. 19, pp. 1587–1593, (2009)

[41] U. Russo, D. Ielmini, C. Cagli and A. L. Lacaita; *"Filament Conduction and Reset Mechanism in NiO-Based Resistive-Switching Memory (RRAM) Devices"*, IEEE Transactions on Electron Devices, vol. 56, pp. 186-192, (2009)

[42] C. Cagli, D. Ielmini, F. Nardi and A. L. Lacaita; *"Evidence for threshold switching in the set process of NiO-based RRAM and physical modeling for set, reset, retention and disturb prediction"*, 2008 IEEE International Electron Devices Meeting, pp. 1-4, (2008)

[43] A. Regoutz, I. Gupta, A. Serb, A. Khiat, F. Borgatti, T.-L. Lee, C. Schlueter, P. Torelli, B. Gobaut, M. Light, D. Carta, S. Pearce, G. Panaccione and T. Prodromakis; *"Role and Optimization of the Active Oxide Layer in TiO2-Based RRAM"*, Advanced Functional Materials, vol. 26, pp. 507–513, (2016)

[44] Y. S. Chen, H. Y. Lee, P. S. Chen, P. Y. Gu, C. W. Chen, W. P. Lin, W. H. Liu, Y. Y. Hsu, S. S. Sheu, P. C. Chiang, W. S. Chen, F. T. Chen, C. H. Lien and M. J. Tsai; *"Highly scalable hafnium oxide memory with improvements of resistive distribution and read disturb immunity"*, 2009 IEEE International Electron Devices Meeting (IEDM), pp. 1-4, (2009)

[45] C. Cagli, J. Buckley, V. Jousseaume, T. Cabout, A. Salaun, H. Grampeix, J. F. Nodin, H. Feldis, A. Persico, J. Cluzel, P. Lorenzi, L. Massari, R. Rao, F. Irrera, F. Aussenac, C.

Carabasse, M. Coue, P. Calka, E. Martinez, L. Perniola, P. Blaise, Z. Fang, Y. H. Yu, G. Ghibaudo, D. Deleruyelle, M. Bocquet, C. Müller, A. Padovani, O. Pirrotta, L. Vandelli, L. Larcher, G. Reimbold and B. de Salvo; *"Experimental and theoretical study of electrode effects in HfO2 based RRAM"*, 2011 International Electron Devices Meeting, pp. 28.7.1-28.7.4, (2011)

[46] V. Jousseaume, A. Fantini, J. Nodin, C. Guedj, A. Persico, J. Buckley, S. Tirano, P. Lorenzi, R. Vignon, H. Feldis, S. Minoret, H. Grampeix, A. Roule, S. Favier, E. Martinez, P. Calka, N. Rochat, G. Auvert, J. Barnes, P. Gonon, C. Vallée, L. Perniola and B. D. Salvo; *"Comparative study of non-polar switching behaviors of NiO- and HfO2-based oxide resistive-RAMs "*, Solid-State Electronics , vol. 58, pp. 62 - 67, (2011)

[47] L. Song, J. Zhang, A. Chen, H. Wu, H. Qian and Z. Yu; *"An efficient method for evaluating* crossbar array performance ", Solid-State Electronics , vol. 120, pp. 32 - 40, (2016)

[48] Y. Deng, P. Huang, B. Chen, X. Yang, B. Gao, J. Wang, L. Zeng, G. Du, J. Kang and X. Liu; *"RRAM Crossbar Array With Cell Selection Device: A Device and Circuit Interaction Study"*, IEEE Transactions on Electron Devices, vol. 60, pp. 719-726, (2013)

[49] A. Sawa; *"Resistive switching in transition metal oxides "*, Materials Today , vol. 11, pp. 28 - 36, (2008)

[50] T. Tsuruoka, K. Terabe, T. Hasegawa and M. Aono; *"Forming and switching mechanisms of a cation-migration-based oxide resistive memory"*, Nanotechnology, vol. 21, pp. 425205, (2010)

[51] P. Lorenzi, R. Rao and F. Irrera; *"Forming Kinetics in HfO_2 -Based RRAM Cells"*, IEEE Transactions on Electron Devices, vol. 60, pp. 438-443, (2013)

[52] S. Yu and P. Y. Chen; *"Emerging Memory Technologies: Recent Trends and Prospects"*, IEEE Solid-State Circuits Magazine, vol. 8, pp. 43-56, (2016)

[53] Z. Wei, Y. Kanzawa, K. Arita, Y. Katoh, K. Kawai, S. Muraoka, S. Mitani, S. Fujii, K. Katayama, M. Iijima, T. Mikawa, T. Ninomiya, R. Miyanaga, Y. Kawashima, K. Tsuji, A. Himeno, T. Okada, R. Azuma, K. Shimakawa, H. Sugaya, T. Takagi, R. Yasuhara, K. Horiba, H. Kumigashira and M. Oshima; *"Highly reliable TaOx ReRAM and direct evidence of redox reaction mechanism"*, 2008 IEEE International Electron Devices Meeting, pp. 1-4, (2008)

[54] R. Aluguri and T. Y. Tseng; *"Overview of Selector Devices for 3-D Stackable Cross Point RRAM Arrays"*, IEEE Journal of the Electron Devices Society, vol. 4, pp. 294-306, (2016)

[55] G. W. Burr, R. S. Shenoy, K. Virwani, P. Narayanan, A. Padilla, B. Kurdi and H. Hwang; *"Access devices for 3D crosspoint memory"*, Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena, vol. 55, pp. 040802, (2014)

[56] S. Yu; *"Overview of resistive switching memory (RRAM) switching mechanism and device modeling"*, 2014 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 2017-2020, (2014)

[57] R. Waser, R. Dittmann, G. Staikov and K. Szot; *"Redox-Based Resistive Switching Memories – Nanoionic Mechanisms, Prospects, and Challenges"*, Advanced Materials, vol. 21, pp. 2632–2663, (2009)

[58] M. Lanza, K. Zhang, M. Porti, M. Nafría, Z. Y. Shen, L. F. Liu, J. F. Kang, D. Gilmer and G. Bersuker; *"Grain boundaries as preferential sites for resistive switching in the HfO2 resistive random access memory structures"*, Applied Physics Letters, vol. 100, pp. 123508, (2012)

[59] D. B. Strukov and R. S. Williams; *"Exponential ionic drift: fast switching and low volatility of thin-film memristors"*, Applied Physics A, vol. 94, pp. 515–519, (2009)

[60] P. A. Leighton; *"Electronic Processes in Ionic Crystals (Mott, N. F.; Gurney, R. W.)"*, Journal of Chemical Education, vol. 18, pp. 249, (1941)

[61] M. Dignam; *"Ion transport in solids under conditions which include large electric fields"*, Journal of Physics and Chemistry of Solids, vol. 29, pp. 249 - 260, (1968)

[62] D. Ielmini; *"Resistive switching memories based on metal oxides: mechanisms, reliability and scaling"*, Semiconductor Science and Technology, vol. 31, pp. 063002, (2016)

[63] Z. Jiang, S. Yu, Y. Wu, J. H. Engel, X. Guan and H. S. P. Wong; *"Verilog-A compact model for oxide-based resistive random access memory (RRAM)"*, 2014 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), pp. 41-44, (2014)

[64] X. Guan, S. Yu and H. S. P. Wong; *"A SPICE Compact Model of Metal Oxide Resistive Switching Memory With Variations"*, IEEE Electron Device Letters, vol. 33, pp. 1405-1407, (2012)

[65] S. Yu, B. Gao, Z. Fang, H. Yu, J. Kang and H. S. P. Wong; *"A neuromorphic visual system using RRAM synaptic devices with Sub-pJ energy and tolerance to variability: Experimental*

characterization and large-scale modeling", 2012 International Electron Devices Meeting, pp. 10.4.1-10.4.4, (2012)

[66] S. Larentis, F. Nardi, S. Balatti, D. C. Gilmer and D. Ielmini*; "Resistive Switching by Voltage-Driven Ion Migration in Bipolar RRAM #x2014;Part II: Modeling"*, IEEE Transactions on Electron Devices, vol. 59, pp. 2468-2475, (2012)

[67] Y.-M. Kim and J.-S. Lee*; "Reproducible resistance switching characteristics of hafnium oxide-based nonvolatile memory devices"*, Journal of Applied Physics, vol. 104, pp. 114115, (2008)

[68] W.-Y. Chang, Y.-C. Lai, T.-B. Wu, S.-F. Wang, F. Chen and M.-J. Tsai*; "Unipolar resistive switching characteristics of ZnO thin films for nonvolatile memory applications"*, Applied Physics Letters, vol. 92, pp. 022110, (2008)

[69] C.-Y. Lin, S.-Y. Wang, D.-Y. Lee and T.-Y. Tseng*; "Electrical Properties and Fatigue Behaviors of ZrO2 Resistive Switching Thin Films"*, Journal of The Electrochemical Society, vol. 155, pp. H615-H619, (2008)

[70] Q. Liu, W. Guan, S. Long, R. Jia, M. Liu and J. Chen*; "Resistive switching memory effect of ZrO2 films with Zr+ implanted"*, Applied Physics Letters, vol. 92, pp. 012117, (2008)

[71] H. Y. Lee, P. S. Chen, T. Y. Wu, Y. S. Chen, F. Chen, C. C. Wang, P. J. Tzeng, C. H. Lin, M. J. Tsai and C. Lien*; " HfO_x Bipolar Resistive Memory With Robust Endurance Using AlCu as Buffer Electrode"*, IEEE Electron Device Letters, vol. 30, pp. 703-705, (2009)

[72] E. A. Miranda, C. Walczyk, C. Wenger and T. Schroeder*; "Model for the Resistive Switching Effect in HfO_2 MIM Structures Based on the Transmission Properties of Narrow Constrictions"*, IEEE Electron Device Letters, vol. 31, pp. 609-611, (2010)

[73] E. Miranda, D. Jimenez and J. Sune*; "The Quantum Point-Contact Memristor"*, IEEE Electron Device Letters, vol. 33, pp. 1474-1476, (2012)

[74] S. Datta*; "Electronic Transport in Mesoscopic Systems"*, (1997)

[75] H. Clarke, T. Brown, J. Hu, R. Ganguli, A. Reed, A. Voevodin and P. J. Shamberger*; "Microstructure dependent filament forming kinetics in HfO 2 programmable metallization cells"*, Nanotechnology, vol. 27, pp. 425709, (2016)

[76] G. Bersuker, D. C. Gilmer, D. Veksler, P. Kirsch, L. Vandelli, A. Padovani, L. Larcher, K. McKenna, A. Shluger, V. Iglesias, M. Porti and M. Nafría*; "Metal oxide resistive memory*

switching mechanism based on conductive filament properties", Journal of Applied Physics, vol. 110, pp. 124518, (2011)

[77] A. Fantini, L. Goux, R. Degraeve, D. J. Wouters, N. Raghavan, G. Kar, A. Belmonte, Y. Y. Chen, B. Govoreanu and M. Jurczak*; "Intrinsic switching variability in HfO2 RRAM"*, 2013 5th IEEE International Memory Workshop, pp. 30-33, (2013)

[78] Y. Y. Chen, M. Komura, R. Degraeve, B. Govoreanu, L. Goux, A. Fantini, N. Raghavan, S. Clima, L. Zhang, A. Belmonte, A. Redolfi, G. S. Kar, G. Groeseneken, D. J. Wouters and M. Jurczak*; "Improvement of data retention in HfO2/Hf 1T1R RRAM cell under low operating current"*, 2013 IEEE International Electron Devices Meeting, pp. 10.1.1-10.1.4, (2013)

[79] *; "Predictive Technology Model (PTM). [Online]. Available: http://ptm.asu.edu/"*, Arizona State University, (2011)

[80] J. Zahurak, K. Miyata, M. Fischer, M. Balakrishnan, S. Chhajed, D. Wells, H. Li, A. Torsi, J. Lim, M. Korber, K. Nakazawa, S. Mayuzumi, M. Honda, S. Sills, S. Yasuda, A. Calderoni, B. Cook, G. Damarla, H. Tran, B. Wang, C. Cardon, K. Karda, J. Okuno, A. Johnson, T. Kunihiro, J. Sumino, M. Tsukamoto, K. Aratani, N. Ramaswamy, W. Otsuka and K. Prall*; "Process integration of a 27nm, 16Gb Cu ReRAM"*, 2014 IEEE International Electron Devices Meeting, pp. 6.2.1-6.2.4, (2014)

[81] C. J. Chevallier, C. H. Siau, S. F. Lim, S. R. Namala, M. Matsuoka, B. L. Bateman and D. Rinerson*; "A 0.13 #x00B5;m 64Mb multi-layered conductive metal-oxide memory"*, 2010 IEEE International Solid-State Circuits Conference - (ISSCC), pp. 260-261, (2010)

[82] S. S. Sheu, M. F. Chang, K. F. Lin, C. W. Wu, Y. S. Chen, P. F. Chiu, C. C. Kuo, Y. S. Yang, P. C. Chiang, W. P. Lin, C. H. Lin, H. Y. Lee, P. Y. Gu, S. M. Wang, F. T. Chen, K. L. Su, C. H. Lien, K. H. Cheng, H. T. Wu, T. K. Ku, M. J. Kao and M. J. Tsai*; "A 4Mb embedded SLC resistive-RAM macro with 7.2ns read-write random-access time and 160ns MLC-access capability"*, 2011 IEEE International Solid-State Circuits Conference, pp. 200-202, (2011)

[83] C. Gopalan, Y. Ma, T. Gallo, J. Wang, E. Runnion, J. Saenz, F. Koushan, P. Blanchard and S. Hollmer*; "Demonstration of Conductive Bridging Random Access Memory (CBRAM) in logic* process ", Solid-State Electronics , vol. 58, pp. 54 - 61, (2011)

[84] W. Otsuka, K. Miyata, M. Kitagawa, K. Tsutsui, T. Tsushima, H. Yoshihara, T. Namise, Y. Terao and K. Ogata*; "A 4Mb conductive-bridge resistive memory with 2.3GB/s read-throughput*

and 216MB/s program-throughput", 2011 IEEE International Solid-State Circuits Conference, pp. 210-211, (2011)

[85] A. Kawahara, R. Azuma, Y. Ikeda, K. Kawai, Y. Katoh, Y. Hayakawa, K. Tsuji, S. Yoneda, A. Himeno, K. Shimakawa, T. Takagi, T. Mikawa and K. Aono*; "An 8 Mb Multi-Layered Cross-Point ReRAM Macro With 443 MB/s Write Throughput"*, IEEE Journal of Solid-State Circuits, vol. 48, pp. 178-185, (2013)

[86] A. Kawahara, K. Kawai, Y. Ikeda, Y. Katoh, R. Azuma, Y. Yoshimoto, K. Tanabe, Z. Wei, T. Ninomiya, K. Katayama, R. Yasuhara, S. Muraoka, A. Himeno, N. Yoshikawa, H. Murase, K. Shimakawa, T. Takagi, T. Mikawa and K. Aono*; "Filament scaling forming technique and level-verify-write scheme with endurance over 107 cycles in ReRAM"*, 2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers, pp. 220-221, (2013)

[87] T. Y. Liu, T. H. Yan, R. Scheuerlein, Y. Chen, J. K. Lee, G. Balakrishnan, G. Yee, H. Zhang, A. Yap, J. Ouyang, T. Sasaki, S. Addepalli, A. Al-Shamma, C. Y. Chen, M. Gupta, G. Hilton, S. Joshi, A. Kathuria, V. Lai, D. Masiwal, M. Matsumoto, A. Nigam, A. Pai, J. Pakhale, C. H. Siau, X. Wu, R. Yin, L. Peng, J. Y. Kang, S. Huynh, H. Wang, N. Nagel, Y. Tanaka, M. Higashitani, T. Minvielle, C. Gorla, T. Tsukamoto, T. Yamaguchi, M. Okajima, T. Okamura, S. Takase, T. Hara, H. Inoue, L. Fasoli, M. Mofidi, R. Shrivastava and K. Quader*; "A 130.7mm2 2-layer 32Gb ReRAM memory device in 24nm technology"*, 2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers, pp. 210-211, (2013)

[88] R. Fackenthal, M. Kitagawa, W. Otsuka, K. Prall, D. Mills, K. Tsutsui, J. Javanifard, K. Tedrow, T. Tsushima, Y. Shibahara and G. Hush*; "19.7 A 16Gb ReRAM with 200MB/s write and 1GB/s read in 27nm technology"*, 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), pp. 338-339, (2014)

[89] M. Ueki, K. Takeuchi, T. Yamamoto, A. Tanabe, N. Ikarashi, M. Saitoh, T. Nagumo, H. Sunamura, M. Narihiro, K. Uejima, K. Masuzaki, N. Furutake, S. Saito, Y. Yabe, A. Mitsuiki, K. Takeda, T. Hase and Y. Hayashi*; "Low-power embedded ReRAM technology for IoT applications"*, 2015 Symposium on VLSI Technology (VLSI Technology), pp. T108-T109, (2015)

[90] D. B. Strukov, G. S. Snider, D. R. Stewart and R. S. Williams*; "The missing memristor found"*, Nature, vol. 453, pp. 80–83, (2008)

[91] L. Chua; *"Memristor-The missing circuit element"*, IEEE Transactions on Circuit Theory, vol. 18, pp. 507-519, (1971)

[92] L. Chua; *"Resistance switching memories are memristors"*, Applied Physics A, vol. 102, pp. 765–783, (2011)

[93] H. Y. Lee, P. S. Chen, T. Y. Wu, Y. S. Chen, C. C. Wang, P. J. Tzeng, C. H. Lin, F. Chen, C. H. Lien and M. J. Tsai; *"Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO2 based RRAM"*, 2008 IEEE International Electron Devices Meeting, pp. 1-4, (2008)

[94] B. Traoré, K. H. Xue, E. Vianello, G. Molas, P. Blaise, B. D. Salvo, A. Padovani, O. Pirrotta, L. Larcher, L. R. C. Fonseca and Y. Nishi; *"Investigation of the role of electrodes on the retention performance of HfOx based RRAM cells by experiments, atomistic simulations and device physical modeling"*, 2013 IEEE International Reliability Physics Symposium (IRPS), pp. 5E.2.1-5E.2.6, (2013)

[95] P. Lorenzi, R. Rao and F. Irrera; *"Role of the electrode metal, waveform geometry, temperature, and postdeposition treatment on SET and RESET of HfO2-based resistive random access memory 1R-cells: Experimental aspects"*, Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena, vol. 33, pp. 01A107, (2015)

[96] P. Lorenzi, R. Rao and F. Irrera; *"Impact of Forming Pulse Geometry and Area Scaling on Forming Kinetics and Stability of the Low Resistance State in HfO2-Based RRAM Cells"*, 2012 4th IEEE International Memory Workshop, pp. 1-4, (2012)

[97] P. Huang, X. Y. Liu, B. Chen, H. T. Li, Y. J. Wang, Y. X. Deng, K. L. Wei, L. Zeng, B. Gao, G. Du, X. Zhang and J. F. Kang; *"A Physics-Based Compact Model of Metal-Oxide-Based RRAM DC and AC Operations"*, IEEE Transactions on Electron Devices, vol. 60, pp. 4090-4097, (2013)

[98] P. Lorenzi, R. Rao, F. Irrera, J. Suñé and E. Miranda; *"A thorough investigation of the progressive reset dynamics in HfO2-based resistive switching structures"*, Applied Physics Letters, vol. 107, pp. 113507, (2015)

[99] M. Morales-Masis, H.-D. Wiemhofer and J. M. van Ruitenbeek; *"Towards a quantitative description of solid electrolyte conductance switches"*, Nanoscale, vol. 2, pp. 2275-2280, (2010)

[100] E. Miranda, D. Jiménez, A. Tsurumaki-Fukuchi, J. Blasco, H. Yamada, J. Suñé and A. Sawa; *"Modeling of hysteretic Schottky diode-like conduction in Pt/BiFeO3/SrRuO3 switches"*, Applied Physics Letters, vol. 105, pp. 082904, (2014)

[101] D. Ielmini and Y. Zhang; *"Analytical model for subthreshold conduction and threshold switching in chalcogenide-based memory devices"*, Journal of Applied Physics, vol. 102, pp. 054517, (2007)

[102] J. Borghetti, D. B. Strukov, M. D. Pickett, J. J. Yang, D. R. Stewart and R. S. Williams; *"Electrical transport and thermometry of electroformed titanium dioxide memristive switches"*, Journal of Applied Physics, vol. 106, pp. 124504, (2009)

[103] B. Traoré, P. Blaise, E. Vianello, E. Jalaguier, G. Molas, J. F. Nodin, L. Perniola, B. D. Salvo and Y. Nishi; *"Impact of electrode nature on the filament formation and variability in HfO2 RRAM"*, 2014 IEEE International Reliability Physics Symposium, pp. 5E.2.1-5E.2.5, (2014)

[104] J. Blasco, N. Ghenzi, J. Su??, P. Levy and E. Miranda; *"Modeling of the Hysteretic I-V Characteristics of TiO_2 -Based Resistive Switches Using the Generalized Diode Equation"*, IEEE Electron Device Letters, vol. 35, pp. 390-392, (2014)

[105] S. Bowling, M. Khasawneh, S. Kaewkuekool and B. Cho; *"A logistic approximation to the cumulative normal distribution"*, ournal of Industrial Engineering and Management, (2009)

[106] W. Easterly and R. Levine; *"What have we learned from a decade of empirical research on growth? It's Not Factor Accumulation: Stylized Facts and Growth Models"*, The World Bank Economic Review, vol. 15, pp. 177-219, (2001)

[107] J. M. Kim and J. M. Kosterlitz; *"Growth in a restricted solid-on-solid model"*, Phys. Rev. Lett., vol. 62, pp. 2289–2292, (1989)

[108] R. Buchanan, R. Whiting and W. Damert; *"When is simple good enough: a comparison of the Gompertz, Baranyi, and three-phase linear models for fitting bacterial growth curves"*, Food Microbiology, vol. 14, pp. 313 - 326, (1997)

[109] M. Peleg, M. G. Corradini and M. D. Normand; *"The logistic (Verhulst) model for sigmoid microbial growth curves revisited "*, Food Research International , vol. 40, pp. 808 - 818, (2007)

[110] R. Law, D. J. Murrell and U. Dieckmann; *"POPULATION GROWTH IN SPACE AND TIME: SPATIAL LOGISTIC EQUATIONS"*, Ecology, vol. 84, pp. 252–262, (2003)

[111] A. Tsoularis and J. Wallace; *"Analysis of logistic growth models "*, Mathematical Biosciences , vol. 179, pp. 21 - 55, (2002)

[112] P. Georgiou; *"Ph.D. thesis"*, Imperial College London, (2013)

[113] C. Cagli, D. Ielmini, F. Nardi and A. L. Lacaita; *"Evidence for threshold switching in the set process of NiO-based RRAM and physical modeling for set, reset, retention and disturb prediction"*, 2008 IEEE International Electron Devices Meeting, pp. 1-4, (2008)

[114] P. Y. Gu, Y. S. Chen, H. Y. Lee, P. S. Chen, W. H. Liu, W. S. Chen, Y. Y. Hsu, F. Chen and M. J. Tsai; *"Scalability with silicon nitride encapsulation layer for Ti/HfOx pillar RRAM"*, Proceedings of 2010 International Symposium on VLSI Technology, System and Application, pp. 146-147, (2010)

[115] F. C. Chiu, W. Y. Chang, P. W. Li, C. C. Chen and W. P. Chiang; *"Trap exploration of ZnO-based resistance switching memory devices"*, 2013 IEEE 5th International Nanoelectronics Conference (INEC), pp. 234-235, (2013)

[116] Y. F. Chang, Y. C. Chen, J. Li, F. Xue, Y. Wang, F. Zhou, B. Fowler and J. C. Lee; *"Comprehensive trap-level study in SiOx-based resistive switching memory"*, 71st Device Research Conference, pp. 135-136, (2013)

[117] P. Lorenzi, R. Rao, T. Prifti and F. Irrera; *"Impact of the forming conditions and electrode metals on read disturb in HfO2-based "*, Microelectronics Reliability , vol. 53, pp. 1203 - 1207, (2013)

[118] U. Russo, D. Ielmini, C. Cagli and A. L. Lacaita; *"Self-Accelerated Thermal Dissolution Model for Reset Programming in Unipolar Resistive-Switching Memory (RRAM) Devices"*, IEEE Transactions on Electron Devices, vol. 56, pp. 193-200, (2009)

[119] C. Cagli, F. Nardi and D. Ielmini; *"Modeling of Set/Reset Operations in NiO-Based Resistive-Switching Memory Devices"*, IEEE Transactions on Electron Devices, vol. 56, pp. 1712-1720, (2009)

[120] P. Huang, X. Y. Liu, W. H. Li, Y. X. Deng, B. Chen, Y. Lu, B. Gao, L. Zeng, K. L. Wei, G. Du, X. Zhang and J. F. Kang; *"A physical based analytic model of RRAM operation for circuit simulation"*, 2012 International Electron Devices Meeting, pp. 26.6.1-26.6.4, (2012)

[121] B. Chen, B. Gao, Y. H. Fu, R. Liu, L. Ma, P. Huang, F. F. Zhang, L. F. Liu, X. Y. Liu, J. F. Kang and G. J. Lian; *"Co-existed unipolar and bipolar resistive switching effect of HfOx-based RRAM"*, 2012 IEEE Silicon Nanoelectronics Workshop (SNW), pp. 1-2, (2012)

[122] T. Ninomiya, Z. Wei, S. Muraoka, R. Yasuhara, K. Katayama and T. Takagi; *"Conductive Filament Scaling of TaO_x Bipolar ReRAM for Improving Data Retention Under Low Operation Current"*, IEEE Transactions on Electron Devices, vol. 60, pp. 1384-1389, (2013)

[123] P. Lorenzi, R. Rao and F. Irrera; *"Conductive filament evolution in HfO2 resistive* device during constant voltage stress "

[124] F. M. Puglisi, L. Larcher, P. Pavan, A. Padovani and G. Bersuker; *"Instability of HfO2 RRAM devices: Comparing RTN and cycling variability"*, 2014 IEEE International Reliability Physics Symposium, pp. MY.5.1-MY.5.5, (2014)

[125] S. Deora, G. Bersuker, M. G. Sung, D. C. Gilmer, P. D. Kirsch, H. F. Li, H. Chong and S. Gausepohl; *"Statistical assessment of endurance degradation in high and low resistive states of the HfO2-based RRAM"*, 2013 IEEE International Reliability Physics Symposium (IRPS), pp. MY.2.1-MY.2.5, (2013)

[126] P. R. Gray, P. J. Hurst, S. H. Lewis and R. G. Meyer; *"Analysis and design of analog integrated circuits"*, (5th ed.)Wiley, (2009)

[127] R. Lippmann; *"An introduction to computing with neural nets"*, IEEE ASSP Magazine, vol. 4, pp. 4-22, (1987)

[128] S. Thorpe, D. Fize and C. Marlot; *"Speed of processing in the human visual system"*, Nature, vol. 381, pp. 520–522, (1996)

[129] C. Teuscher, D. Mange, A. Stauffer and G. Tempesti; *"Bio-inspired computing tissues: towards machines that evolve, grow, and learn "*, Biosystems , vol. 68, pp. 235 - 244, (2003)

[130] C. Teuscher, D. Mange, A. Stauffer and G. Tempesti, "*Bio-inspired computing tissues: towards machines that evolve, grow, and learn*", Biosystems, vol. 68 pp. 235–244 (2003).

[131] M. Hänggi and G. Moschytz, "*Cellular neural networks: analysis, design and optimization.*" Springer Science & Business Media. (2013).

[132] C. Poon and K. Zhou, "*Neuromorphic silicon neurons and large-scale neural networks: challenges and opportunities*" Frontiers in neuroscience, vol. 5 pp. 108 (2011).

[133] P. Merolla, J. Arthur, R. Alvarez-Icaza, A. Cassidy, J. Sawada, F. Akopyan et al., "*A million spiking-neuron integrated circuit with a scalable communication network and interface*" Science, vol. 345 pp. 668–673 (2014).

[134] J.Burr. "*Digital neural network implementations.*" Neural networks, concepts, applications, and implementations 3 (1991).

[135] J. Seo, B. Brezzo, Y. Liu, B. Parker, S. Esser, R. Montoye, and D. Friedman. "*A 45nm CMOS neuromorphic chip with a scalable architecture for learning in networks of spiking neurons*". 2011 IEEE Custom Integrated Circuits Conference (CICC). (2011, September).

[136] L. Myoung-Jae, et al. "*2-stack 1D-1R cross-point structure with oxide diodes as switch elements for high density resistance RAM applications.*" 2007 IEEE International Electron Devices Meeting. IEEE, (2007).

[137] S.Jo, T. Chang, I. Ebong, B. Bhadviya, P. Mazumder, and W. Lu. "*Nanoscale memristor device as synapse in neuromorphic systems.*" Nano letters, 10(4), 1297-1301. (2010).

[138] S. Yu, B. Gao, Z. Fang, H. Yu, J. Kang and H. Wong, "*A neuromorphic visual system using RRAM synaptic devices with Sub-pJ energy and tolerance to variability: Experimental characterization and large-scale modeling*" Electron Devices Meeting (IEDM), (2012).

[139] B.Govoreanu, et al. "*10× 10nm 2 Hf/HfO x crossbar resistive RAM with excellent performance, reliability and low-energy operation.*" Electron Devices Meeting (IEDM), 2011 IEEE International. IEEE, (2011).

[140] S. Song, K. Miller and L. Abbott, "*Competitive Hebbian learning through spike-timing-dependent synaptic plasticity*" Nature neuroscience, vol. 3 pp. 919–926 (2000).

[141] R. Rao and T. J. Sejnowski. "*Spike-timing-dependent Hebbian plasticity as temporal difference learning.*" Neural computation, 13(10), 2221-2237, (2001).

[142] G.-q. Bi and M.-m. Poo*; "Synaptic Modifications in Cultured Hippocampal Neurons: Dependence on Spike Timing, Synaptic Strength, and Postsynaptic Cell Type"*, Journal of Neuroscience, vol. 18, pp. 10464–10472, (1998)

[143] K. Sasaki, T. Morie and A. Iwata*; "A spiking neural network with negative thresholding and its application to associative memory"*, Circuits and Systems, 2004. MWSCAS '04. The 2004 47th Midwest Symposium on, vol. 3, pp. iii-89-92 vol.3, (2004)

[144] M. Chu, B. Kim, S. Park, H. Hwang, M. Jeon, B. H. Lee and B. G. Lee*; "Neuromorphic Hardware System for Visual Pattern Recognition With Memristor Array and CMOS Neuron"*, IEEE Transactions on Industrial Electronics, vol. 62, pp. 2410-2419, (2015)

[145] E. Chicca, F. Stefanini, C. Bartolozzi and G. Indiveri*; "Neuromorphic Electronic Circuits for Building Autonomous Cognitive Systems"*, Proceedings of the IEEE, vol. 102, pp. 1367–1388, (2014)

[146] G. Rachmuth, Z. Shouval, M. Bear, and C.-S. Poon, "*A biophysically-based neuromorphic model of spike rate- and timing-dependent plasticity*,"Proc. Nat. Acad. Sci. USA, vol. 108, no. 49, pp. E1266–E1274, Dec. (2011).

[147] J. Wijekoon and P. Dudek, "*Compact silicon neuron circuit with spiking and bursting behaviour*,"Neural Netw., vol. 21, no. 2–3, pp. 524–534, Mar.–Apr. (2008).

[148] J. Schemmel, D. Bru ¨derle, K. Meier, and B. Ostendorf, "*Modeling synaptic plasticity within networks of highly accelerated I&F neurons*," inProc. IEEE Int. Symp. Circuits Syst., 2007, pp. 3367–3370, (2007)

[149] C. Tomazou, F. Lidgey, and D. Haigh, Eds., *Analogue IC Design: The Current-Mode Approach*. Stevenage, U.K.: Peregrinus, (1990)

[150] E. Drakakis, A. Payne, and C. Toumazou,*'Log-domain state-space': A systematic transistor-level approach for log-domain filtering*," IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process., vol. 46, no. 3, pp. 290–305, Mar. (1999).

[151] R. Edwards and G. Cauwenberghs, "*Synthesis of log-domain filters from first-order building blocks*," Int. J. Analog Integr. Circuits Signal Process., vol. 22, pp. 177–186, (2000).

[152] S.-C. Liu, J. Kramer, G. Indiveri, T. Delbruck, and R. Douglas,Analog *VLSI: Circuits and Principles*. Cambridge, MA, USA: MIT Press, (2002).

[153] T. Yu and G. Cauwenberghs, "*Log-domain time-multiplexed realization of dynamical conductance-based synapses*," inProc. IEEE Int. Symp. Circuits Syst.,Jun.2010, pp. 2558–2561., (2010)

[154] S. Mitra, G. Indiveri, and R. Etienne-Cummings, "*Synthesis of log-domain integrators for silicon synapses with global parametric control*," in Proc. IEEE Int. Symp. Circuits Syst., 2010, pp. 97–100., (2010)

[155] R. Jolivet, T. Lewis, and W. Gerstner, *"Generalized integrate-and-fire models of neuronal activity approximate spike trains of a detailed model to a high degree of accuracy,"*J. Neurophysiol., vol. 92, pp. 959–976, (2004)

[156] D. Hebb; "*The organization of behavior*", Wiley & Sons, (1949)

[157] H. Shouval, G. Castellani, B. Blais, L. Yeung, L. Cooper; "*Converging evidence for a simplified biophysical model of synaptic plasticity*", Biological Cybernetics, (2002)

[158] K. Giese, N. Fedorov, R. Filipkowski, A. Silva; "*Autophosphorylation at Thr286 of the a Calcium-Calmodulin Kinase II in LTP and learning*", Science, (1998)

[159] M. Gomez, E. DeCastro, E. Guarin, H. Sasakura, A. Kuhara, I. Mori, T.Bartfai, C. Bargmann and P. Nef; "*Ca2+ Signaling via the Neuronal Calcium Sensor-1 regulates associative learning and memory in C. Elegans*", Neuron, (2001)

[160] P. Koplas, R. Rosenberg, G. Oxford; "*The role of calcium in the densensitization of capsaisin responses in rat dorsal root ganglion neurons*", The Journal of Neuroscience, (1997)

[161] S. Ambrogio, S. Balatti, V. Milo, R. Carboni, Z. Q. Wang, A. Calderoni, N. Ramaswamy and D. Ielmini*; "Neuromorphic Learning and Recognition With One-Transistor-One-Resistor Synapses and Bistable Metal Oxide RRAM"*, IEEE Transactions on Electron Devices, vol. 63, pp. 1508-1515, (2016)

[162] D. Kuzum, R. G. D. Jeyasingh, B. Lee and H.-S. P. Wong*; "Nanoelectronic Programmable Synapses Based on Phase Change Materials for Brain-Inspired Computing"*, Nano Letters, vol. 12, pp. 2179-2186, (2012)

[163] N. Du, M. Kiani, C. Mayr, T. You, D. Bürger, I. Skorupa, O. Schmidt and H. Schmidt*; "Single pairing spike-timing dependent plasticity in BiFeO3 memristors with a time window of 25 ms to 125 μs"*, Frontiers in Neuroscience, vol. 9, pp. 227, (2015)

[164] Y. Nishitani, Y. Kaneko and M. Ueda*; "Supervised Learning Using Spike-Timing-Dependent Plasticity of Memristive Synapses"*, IEEE Transactions on Neural Networks and Learning Systems, vol. 26, pp. 2999-3008, (2015)

[165] Z. Wang, S. Ambrogio, S. Balatti and D. Ielmini*; "A 2-transistor/1-resistor artificial synapse capable of communication and stochastic learning in neuromorphic systems"*, Frontiers in Neuroscience, vol. 8, pp. 438, (2015)

[166] M. Suri, D. Querlioz, O. Bichler, G. Palma, E. Vianello, D. Vuillaume, C. Gamrat and B. DeSalvo; *"Bio-Inspired Stochastic Computing Using Binary CBRAM Synapses"*, IEEE Transactions on Electron Devices, vol. 60, pp. 2402-2409, (2013)

[167] J. Bill and R. Legenstein; *"A compound memristive synapse model for statistical learning through STDP in spiking neural networks"*, Frontiers in Neuroscience, vol. 8, (2014)

[168] D. Garbin, E. Vianello, O. Bichler, Q. Rafhay, C. Gamrat, G. Ghibaudo, B. DeSalvo and L. Perniola; *"HfO2-Based OxRAM Devices as Synapses for Convolutional Neural Networks"*, IEEE Transactions on Electron Devices, vol. 62, pp. 2494–2501, (2015)

[169] S. Park, M. Chu, J. Kim, J. Noh, M. Jeon, B. H. Lee, H. Hwang, B. Lee and B.-g. Lee; *"Electronic system with memristive synapses for pattern recognition"*, Scientific Reports, vol. 5, pp. 10123, (2015)

[170] P. Lorenzi, V. Sucre, G. Romano, R. Rao and F. Irrera; *"Memristor based neuromorphic circuit for visual pattern recognition"*, 2015 International Conference on Memristive Systems (MEMRISYS), pp. 1-2, (2015)

[171] G. Romano, P. L. ans Rosario Rao and F. Irrera; *"HfO2-ReRAM based Neuromorphic Architecture for Visual Recognition"*, Workshop on Dielectrics in Microelectronics, WoDiM 2016, (2016)

[172] T. Serrano-Gotarredona, T. Masquelier, T. Prodromakis, G. Indiveri and B. Linares-Barranco; *"STDP and STDP variations with memristors for spiking neuromorphic learning systems"*, Frontiers in Neuroscience, vol. 7, pp. 2, (2013)

[173] E. Covi, S. Brivio, A. Serb, T. Prodromakis, M. Fanciulli and S. Spiga; *"Analog Memristive Synapse in Spiking Networks Implementing Unsupervised Learning"*, Frontiers in Neuroscience, vol. 10, pp. 482, (2016)

[174] D.C. Ciresan, U. Meier, L.M. Gambardella, and J. Schmidhuber, "*Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition*," Neural computation, 22 (12), pp.3207-3220. (2010).

[175] P.Diehl, and M. Cook. "*Unsupervised learning of digit recognition using spike-timing-dependent plasticity.*" Frontiers in computational neuroscience 9, pp.99. (2015).

[176] D. Querlioz, O. Bichler, and C. Gamrat. "*Simulation of a memristor-based spiking neural network immune to device variations.*" Neural Networks (IJCNN), The 2011 International Joint Conference on, pp. 1775-1781.(2011).

[177] J. Frascaroli, S. Brivio, E. Covi, P. Lorenzi, M. V. Nair, F. Irrera, G. Indiveri and S. Spiga*; "HfO2-Based Devices for Neuromorphic Network with Real-Time Operation"*, Submitted to EMRS 2016, (2016)

[178] D. Goodman and R. Brette*; "Brian: a simulator for spiking neural networks in Python"*, Frontiers in Neuroinformatics, vol. 2, pp. 5, (2008)

[179] J. M. Brader, W. Senn and S. Fusi*; "Learning Real-World Stimuli in a Neural Network with Spike-Driven Synaptic Dynamics"*, Neural Comput., vol. 19, pp. 2881–2912, (2007)

[180] S. Spiga, B. Stefano, E. Covi, M. Fanciulli, A. Serb, T. Prodromakis, H. Mostafa and G. Indiveri*; "(Invited) Analog Hfox-Rram Switches for Neural Networks"*, Meeting Abstracts, vol. MA2016-02, pp. 1473, (2016)

7 ACKNOWLEDGMENTS

In realtà, i ringraziamenti preferisco scriverli in Italiano. Questa tesi di dottorato rappresenta la sintesi finale di un percorso iniziato ben prima del 2013. Un percorso iniziato già durante la tesi di laurea, nel lontano ormai 2008, e che termina in questi giorni con la scrittura di questa tesi. Un percorso lungo, faticoso e meraviglioso al tempo stesso, condiviso con persone altrettanto meravigliose. Un'esperienza bellissima che mi ha portato ad innamorarmi di questo lavoro.

Per questo motivo ci tengo a ringraziare tutte quelle persone che hanno fatto parte di questo viaggio.

Non posso non iniziare col ringraziare la Prof.ssa Irrera, mentore e guida scientifica e morale. In tutti questi anni ha rappresentato l'unico punto fermo da cui partire per intraprendere qualunque attività: dai progetti di ricerca fino alle cene di laboratorio.

Il Prof. Palma, presenza silenziosa, ma importante.

Il Prof. Miranda e il Prof. Indiveri che mi hanno accolto nei loro laboratori come fossi uno di casa.

Rosario, che definire collega è assolutamente riduttivo. Un amico, un fratello con cui ho condiviso più di tutti il bello e il brutto di questo lavoro.

Stefano e Olga con i quali, insieme a Rosario, ho avuto il piacere di formare il nucleo originario del MindLab. Colleghi e amici anche loro.

Giulio e Ardian che nell'ultimo periodo hanno condiviso con me l'onore e l'onere di portare avanti il lavoro.

Gli studenti che hanno avuto la sfortuna di assistere alle mie lezioni o avermi avuto come co-relatore di tesi.

Tutte le persone del dipartimento con le quali ho avuto il piacere di scambiare due parole o bere un caffè.

A Marta, per le nostre passeggiate e i nostri pranzetti a Monti.

E infine, più di tutti, il mio ringraziamento va sempre ai miei genitori e a mio fratello che, fino all'ultimo, non hanno mai capito che lavoro facessi.

A tutti voi,

Grazie.