Politecnico di Torino

GRENOBLE INP UGA Phelma

EPFL

DEPARTMENT OF ELECTRONICS AND
TELECOMMUNICATIONS ENGINEERING

MSc. in Nanotechnologies For ICTs

Master Degree Thesis

A CMOS-compatible bipolar analogue switching redox-based ReRAM with TaO_x/HfO_2 bilayer

Internal Supervisor
Prof. Carlo Ricciardi

Candidate
Roberto Guido

External Supervisors
IBM Research Europe, Zurich
Dr. Valeria Bragaglia
Research Staff Member in Neuromorphic Devices and Systems group
Youri Popoff and Tommaso Stecconi
PhD students

Academic year 2020-2021

A chi,
in questi anni,
è stato messo
in secondo piano
per raggiungere
questo obiettivo.

Alla mia famiglia.

Abstract

Complementary metal-oxide-semiconductor(CMOS)-based von Neumann architecture suffers from intrinsic power and speed inefficiencies, due to the data transfer between processor and the memory unit. This is commonly addressed as von Neumann bottleneck. The processing of the huge amount of information collected with big data requires the transition to novel computing paradigms and architectures, such as bioinspired computing. It relies on memristor technology which allows to co-locate the storage and the processing of data while ensuring low-cost fabrication and high-density of integration. Among the most promising memristive candidates for building future artificial neural networks (ANNs) there are redox-based valance-change mechanism (VCM) resistive random-access memories (ReRAMs). They gained interest for the large amount of distinguishable resistance levels achievable as well as for their scalability. Symmetry, linearity and stochastic nature of the weight update are the major challenges towards their adoption in neuromorphic systems. Stacking a conductive metal oxide (CMO) on top of the insulator layer represents a simple and cost-effective way to solve these issues. However, the switching mechanism in this novel stack configuration, as well as the role played by the CMO, is not fully understood, thus limiting the improvement of device performances.

In this master thesis work a CMOS-compatible ReRAM with substoichiometric tantalum oxide (TaO_x)/hafnium oxide (HfO_2) bilayer sandwiched between titanium nitride (TiN) electrodes is presented. To meet the challenging requirements of CMO materials, the parameters of the TaO_x sputtering deposition are carefully controlled. The chamber and target conditions revealed to be fundamental for the reproducibility of the TaO_x films. The electrical conductivity of the layer can be finely tuned by means of the pressure, while keeping constant and low the DC power and the oxygen (O_2)/argon(Ar) flow ratio. Material properties are investigated through Grazing Incidence X-ray Diffraction (GIXRD), X-ray Reflectivity (XRR) and Circular Transfer Length Method (CTLM).

Extensive quasi-static and pulsed electrical characterization have been conducted to assess the performances of the bilayer ReRAM devices, as well as to understand the impact of the CMO conductivity and thickness on the device switching. ReRAM devices embedding a 20 nm thick CMO with sheet resistance $R_{sh} \approx 90$ k$\Omega/\square$ showed

that a remarkable linear, symmetric and analogue weight update can be reached by sacrificing the dynamic range to roughly a factor 3.
Impedance Spectroscopy technique was employed to model the switching mechanism in the bilayer ReRAM devices. By analyzing the equivalent circuit used to describe the device in the different states, the role of the filament and of the defect induced leakage resistance in the TaO_x were decoupled.

Acknowledgements

Before starting to discuss the content of this master thesis, I would like to take some lines to express all my gratitude to the people that, even indirectly, contributed to it.

First, I would like to thank all the people of the Neuromorphic Devices and Systems group in IBM Research Zurich with who I worked closely during the past six months for the realization of this master thesis. In particular, I am deeply grateful to my supervisor Dr. Valeria Bragaglia for her scientific support during the whole project and for her patience in reviewing this thesis. Special thanks go also to Youri Popoff and Tommaso Stecconi, with who I shared ideas, laughs and great moments in the past months. I learned a lot from this experience on both scientific and human sides. I will be forever grateful for that.

I would like also to thank Prof. Carlo Ricciardi, Prof. Prejbeanu and Prof. Morfouli for their efforts in the supervision of the Master Nanotech in these hard times. This master was an amazing journey around three countries and an uncountable number of experiences. Thanks for this life-changing opportunity.

Vorrei ringraziare mia madre Lea, mio padre Vincenzo e i miei fratelli, Raffaele e Riccardo. Non dimostro mai quanto tenga a voi e quanto il vostro sostegno sia fondamentale per ogni mio piccolo passo. Vi ringrazio per avermi trasmesso quanto impegno e sacrificio siano fondamentali per raggiungere i propri obiettivi, anche se spesso questo mi ha portato a trascurarvi.

Ringrazio tutti i miei cari. Un ringraziamento particolare ai nonni, Antonio, Maria e Lina, e alle zie, Lucia, Rosaria e Ada, perché da sempre mi supportano e stimolano a raggiungere obiettivi sempre più grandi.

In fine, vorrei ringraziare i miei amici Luca, Pierfrancesco, Giorgio e Giulia che in questi anni non hanno mai perso la pazienza di aspettarmi e di dimostrarmi come l'amicizia sopravvive alla distanza. Siete dei veri amici.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Motivation

This chapter is intended to explain the reader the main motivations behind this master thesis work. First, an overview of the main memristive technologies for implementation in future computing paradigms is presented. Then, among these technologies, we will focus on resistive switching random-access memories (ReRAMs) that are our chosen candidates for this development and optimization work. The strengths, as well as the limitations of this technology, will here be explained. Finally, we give an insight into bilayer structure devices already presented in literature, to identify the major path pursued to improve ReRAM issues and we will propose our adopted strategy.

1.1 Memristive devices for novel computing architectures

Nowadays computing systems rely on the complementary metal-oxide-semiconductor (CMOS)-based von Neumann architecture, in which data buses are used for the connection between the processor and the memory part of the system. However, this kind of architecture suffers from intrinsic power and speed inefficiencies because of the different rate at which the performances of the two parts of the system (the processor and the memory) have improved. The delay caused by this performance difference is called von Neumann bottleneck and it gets even worse when the huge amount of information collected with big data must be processed [1][2][4]. In 2018, the International Roadmap for Devices and Systems (IRDS) stated that the conventional computing architecture is expected to reach is performance limit by 2024 [1]. This is mainly attributed to three "walls": memory wall, related to reliability issues in the non-volatility at extremely small technologically nodes; Moore's law, since scaling is encountering its lowest bound; and heat wall, which is related to the tremendous amount of energy required by edge computing products [1][3]. A transition to novel computing paradigms and architectures based on in-memory

computing, bioinspired computing and better memory storage is hence strongly required to feed the needs of big-data analysis and low-power intelligent systems.

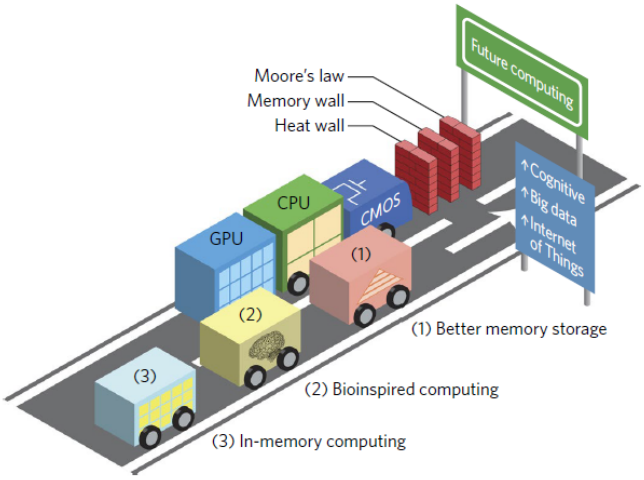


Figure 1.1: Alternative computing solutions towards future computing in order to overcome the limitations of conventional computing systems [3].

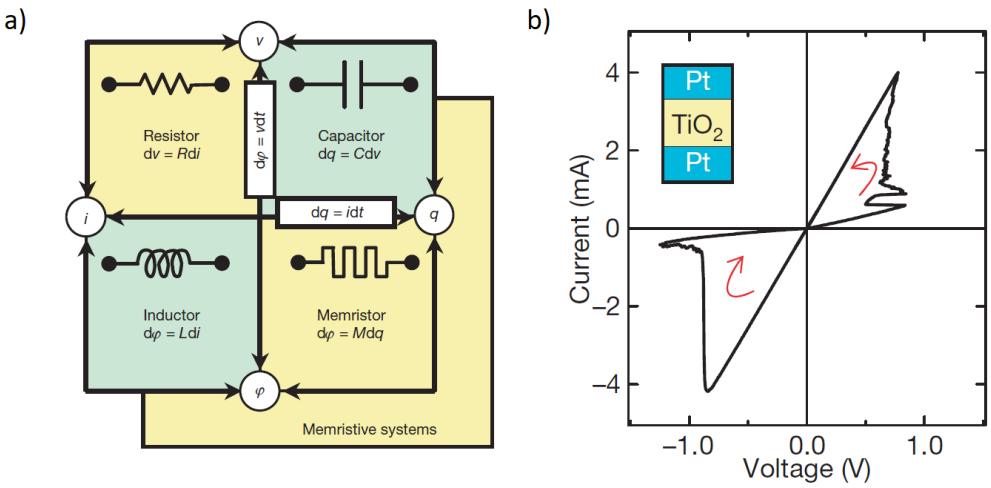


Figure 1.2: a) Schematic representation of the four fundamental two-terminal circuit elements: resistor, capacitor, inductor and memristor. b) Pinched I-V characteristic of $Pt/TiO_2/Pt$ memristor [6].

In this direction, memristor (abbreviation for memory resistor) technology is considered a very promising candidate to drive the new computing era for its low-cost fabrication and high-density of integration [1][3][4]. This device was firstly proposed

by Chua [5] in 1971 as the fourth fundamental circuit element. In [5], the basic mathematical relationship between charge and the flux through the memristance M is derived ($d\phi = Mdq$) and the current versus voltage (I-V) pinched hysteresis is demonstrated. However, only in 2008 Strukov et al. [6] reported the experimental observation of a typical I-V characteristic of a memristor in a metal/oxide/metal device (see Figure 1.2(b)), in which the metal layer is platinum (Pt) and the oxide layer is a 5 nm thick titanium dioxide (TiO_2).
Memristors store data as a resistance state unlike in conventional memory technologies, such as flash, Dynamic Random-Access-Memories (DRAM) and Static Random-Access-Memories (SRAM), in which information are stored as amount of charge in a capacitor. In particular, memristors can be switched between a low resistance state (LRS) and a high resistance state (HRS) by applying an electrical stimulation. A variation of resistance from HRS to LRS is called SET transition, whereas a RESET one occurs if the device is switched from LRS to HRS. The ratio between HRS and LRS is usually referred as On/Off ratio, memory window or dynamic range. Other important device properties are the endurance, which is defined as the maximum number of cycles to reliably switch between LRS and HRS, and the retention, which is the time after which the stored data is lost [1].
The main types of two-terminal memristors are:

a) Resistive switching memristors are usually based on a dielectric layer sandwiched between two electrodes. The switching mechanism relies on redox-based ion-migration in the dielectric, which can be oxygen vacancies in the case of devices based on valance-change mechanism (VCM), or active metal cations belonging to the metal electrodes for devices based on electrochemical mechanism (ECM). In general, these devices are called ReRAM even if a more specific name can be associated to each technology. Indeed, anion-based oxide random-access-memory (OxRAM) can be used to refer to VCM-type devices, while those relying on ECM are often called cation-based conductive bridge random-access-memory (CBRAM) [2].

b) Phase change memristors present different resistance states due to amorphous to crystalline phase transitions induced by Joule heating. Initially the device is characterized by an amorphous state, which is associated to the HRS because of the lack of electron density. Letting a current passing through a metal heater, the amorphous material is locally heated up to a temperature above the crystallization one. Hence it is locally melted and when it cools down, it crystallizes thus resulting in a localization of charge carriers that define the LRS in the device. The phase change material can be amorphized again by increasing its temperature usually above 1000 K through Joule heating effect.

c) Spintronic devices can rely on different effects. In particular, spin-based devices relying on giant magnetoresistance are made by two ferromagnetic layers

separated by a non-magnetic one. The resistance state is changed by the application of an external magnetic field which orients the polarization in the ferromagnetic layers in a parallel (LRS) or anti-parallel (HRS) way, thus favouring or not the motion of electrons having the same polarity of the magnetization. Another effect that can be exploited in spintronic devices is the Spin-Transfer Torque (STT) effect, which is based on modifying the magnetization of the ferroelectric layers by a spin-polarized current. Figure 1.3(c) shows a device relying on the Spin-Orbit Torque (SOT) effect. In this case the device stack is made by a heavy metal layer, a ferromagnetic film and an antiferromagnetic one. The magnetization of the ferromagnetic layer, and hence the device state, is defined by the spin angular momentum exerted by the accumulation of electrons at the interface between the heavy metal and the ferromagnet due to the in-plane current flowing in the heavy metal layer.

d) Ferroelectric Tunnel Junction (FTJ) devices are based on an insulting layer exhibiting ferroelectric properties which is placed between two metal electrodes. The state variation is based on modulating the flux of electrons through the insulator ultrathin energy barrier by reversing the ferroelectric polarization. The main effect that governs these systems is the giant tunnel electroresistance one.

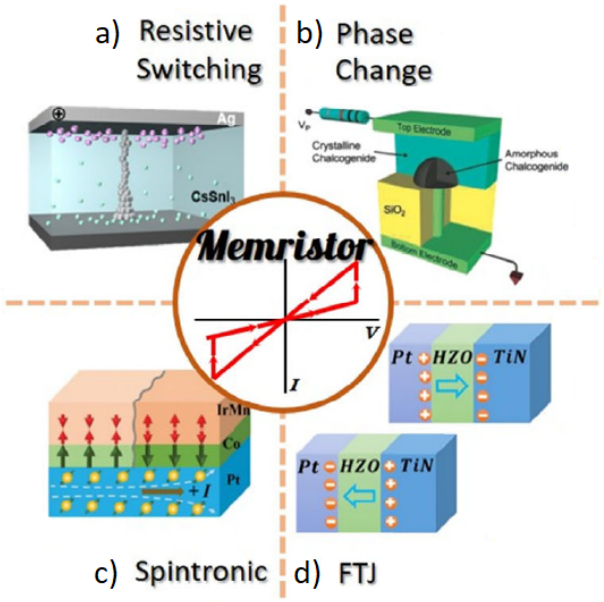


Figure 1.3: Main types of two-terminal memristors: resistive switching (a), phase change (b), spintronic (c) and FTJ (d). Adapted from [1].

Memristor technologies offer the possibility to co-locate the storage and the processing of data [2][4]. They can be densely packed in a crossbar array (CBA) architecture as it is shown in Figure 1.4. CBA architecture allows hence to overcome the CMOS scaling limit since emerging technologies do not show deterioration with size reduction [1]. Moreover, such architecture can be integrated on top of a CMOS chip because of its fully compatibility with backend-of-the-line (BEOL) process, hence enabling a further architecture enhancement. Although CBAs can be used for densely store huge amount of data or to perform logic-in-memory computations, neuromorphic computing systems represent surely the most fascinating perspective. It is from the biological inspiration of the brain connections that low-power and high-performance artificial neural networks (ANNs) can be built for satisfying the urgent demand for Artificial Intelligence (AI) hardware.

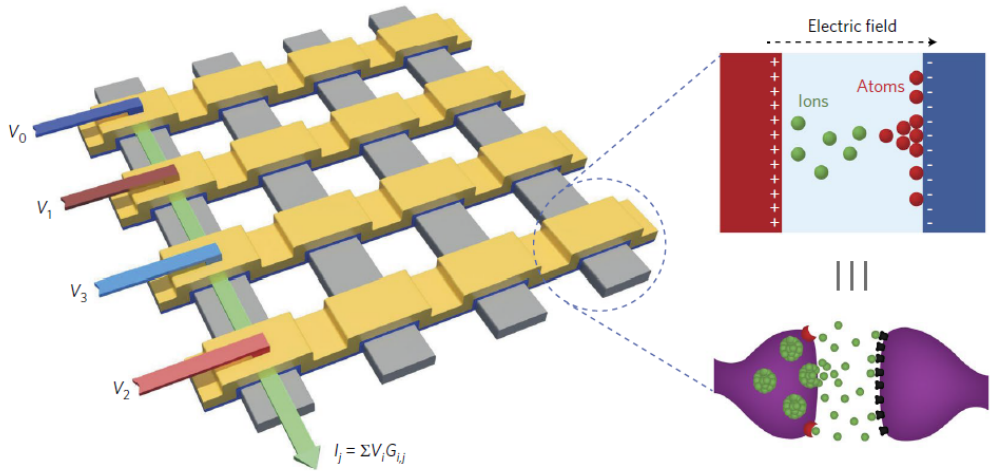


Figure 1.4: Schematic representation of a crossbar architecture [3].

In a biological neural network, the stored information is defined by the strength of the reconfigurable neural connections between the neurons, called synapses. In the same way, in ANNs, the memristors play the role of the synapses, since the synaptic strength corresponds to the resistive state of the devices. The memristor state is hence used to modulate the weighted synaptic connection between the neurons of the network. The network is then trained to perform a specific task by updating these synaptic weights through back-propagation operations. These operations are performed by applying voltage pulses with a proper amplitude and width to the devices, thus resulting in the weight increase (long-term potentiation or simply potentiation) or decrease (long-term depression or simply depression).
In this scenario, memristors represent an extremely promising candidate since they can store multiple states and not only two Boolean values [1]. Moreover, modern

networks can have multiple layers thus requiring large number of synaptic interconnections. This requirement is easier to be satisfied by the emerging memristor technologies with respect to conventional CMOS ones which are limited in their scaling.

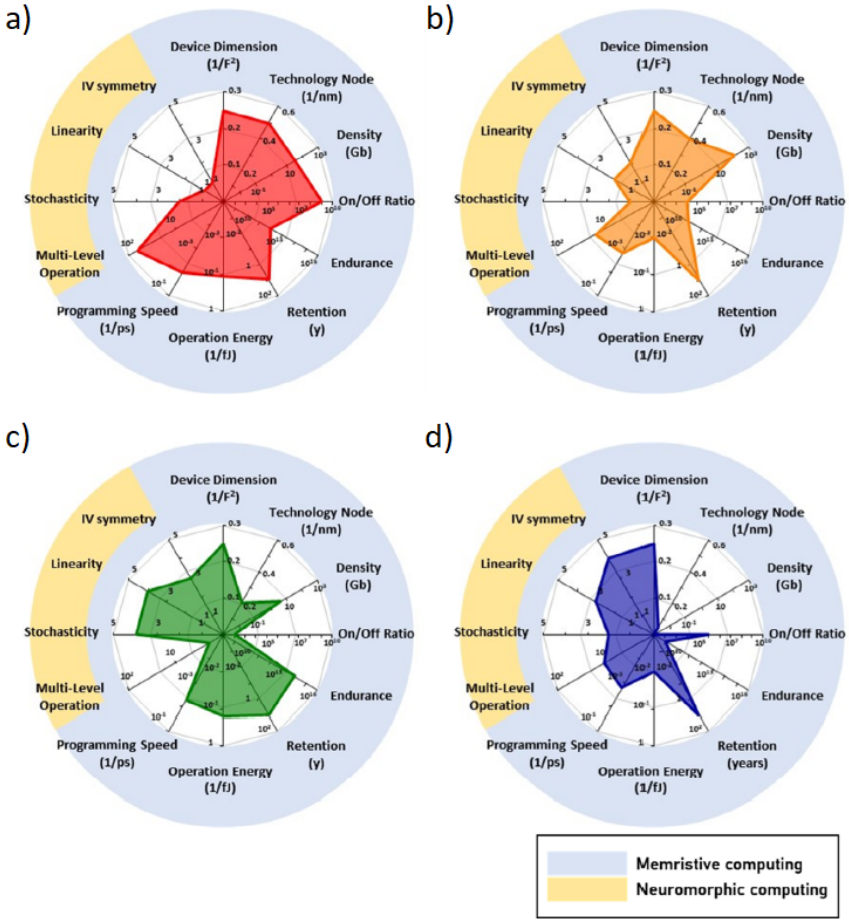


Figure 1.5: Comparison chart of the performances of the four types of two-terminal memristive devices: resistive switching (a), phase change (b), spintronic (c) and FTJ (d) [1].

In CBA, the computation of the signal sent between pre-synaptic neurons and post-synaptic neuron is based on multiplying the voltage level of the pre-synaptic neurons by the device conductance and summing these contributions to define the current level entering in the post-synaptic neuron, in a vector-matrix multiplication fashion, as it is depicted in Figure 1.4. Respect to implementation of ANNs based on loading the synaptic weights from the memory to the processor for the computation, CBA is a more efficient architecture capable to maximize the parallelism and

the distribution of the computation while overcoming the von Neumann bottleneck based on data transferring between memory and processor [3][2][4][7].
Between the enthusiastic discover of the potentiality of memristive devices and their complete adoption in future computing architectures, several challenges must still be faced. These challenges, as well as a summary of the main performances of the discussed two-terminal memristive technologies, are reported in Figure 1.5.
Focusing on the memristive computing performances in Figure 1.5, it is possible to observe that resistive switching devices are among the best candidates in terms of scalability and reliable operation, on its own based on the trade-off between memory window, endurance and retention. Indeed, ReRAM technology has demonstrated a scalability down to sub-10 nm regime without sacrificing its properties, since their chip footprint is determined only by the overlap between the metallic electrodes [2]. Moreover, the large On/Off ratio achievable with this technology translates in a better mapping capability of the synaptic weights in the algorithm to the resistive state of the devices [2]. This type of memristive devices show also outstanding programming speed and energy required per single operation if compared to the other technologies. For example, FTJ requires high voltages to induce the reversing of the polarization, while phase change materials rely on large currents and on long time to enable the crystallization, thus resulting non-adequate to fully solve the von Neumann bottleneck yet [1]. Instead, ReRAM requires energy around 100 fJ$\sim$10 pJ for the programming [2]. However, further device engineering is still required to improve the programming speed and the energy consumption to reach the brain capabilities [2].
In addition to the discussed properties, the adoption in neuromorphic computing systems of memristive devices passes through what Im et al. [1] define as neural plausibility. Neural plausibility describes how well the devices emulate the nervous system and it relies on four main properties:

- I-V symmetry, which indicates the symmetry in the resistance state modulation between a positive or negative voltage stimulation of the device. The more asymmetric the weight update in the device is, the more complex should result the driving circuitry.

- Linearity, which defines the resistance or conductance curve uniformity during the weight update. This parameter is fundamental for the accuracy of the learning process in ANNs, since an unpredictable state variation makes harder to reach a defined target value [1][3][7] or it creates an history dependence of the weight update [2].

- Stochasticity is the relative fluctuation of the resistance state in the range over which it fluctuates over a certain period of time. A low stochasticity value is required to improve the accuracy of ANNs [1].

- Multi-level operation refers to the number of distinguishable resistive states,

that on their own define the number of bits per device. Multilevel states are fundamental to resemble the biological synaptic plasticity, thus allowing for an improved and more robust learning capability of the ANN [2]. However, the weight precision requirement is strongly application dependent [2].

Resistive switching devices are one of the most favourable candidates for the implementation as artificial synapses in neuromorphic computing architectures mainly because of the large amount of distinguishable resistance levels that they can generate [1]. On the other hand, the I-V symmetry, linearity and stochastic nature of the switching mechanism seem to be the major challenges to address towards an actual system implementation based on this kind of devices. These aspects are discussed in more detail in Section 1.2.

1.2 VCM-ReRAM: working principle and technological limitations

ReRAM devices based on valance change mechanism are capacitor-like structures that typically comprise a transition metal oxide (TMO), such as hafnium oxide (HfO_2), titanium oxide (TiO_2) or tantalum oxide (Ta_2O_5), which is sandwiched between two metal electrodes, which are often referred as top (TE) and bottom (BE) electrodes.
The switching behaviour observed in these devices is mainly of two types: unipolar, in which the device changes its state between LRS and HRS, and viceversa, accordingly only to the amplitude of the applied voltage signal; bipolar, in which the device state transition has a defined voltage polarity dependence. The switching behaviour of these two sub-categories of devices are visually represented in Figure 1.6(a) and (b), respectively. However, this work is mainly focused on bipolar switching devices since they are known for their excellent endurance with respect to unipolar ones [4].
Another possible classification of VCM-based ReRAM regards the model that explains the switching behaviour of the device. As it is depicted in Figure 1.6(c) and (d), resistive switching originates from the migration of oxygen ions and subsequent valance changes. This can result in the formation and rupture of a locally confined conductive filament (CF) in the oxide or it can be a phenomenon that takes place homogeneously over the whole interface between the metal electrode and the TMO, thus resulting in a strong area dependence of the resistance. However, the interface-type devices are difficult to scale down and high voltages are required for switching the resistive state at high speed [9]. Moreover, when programmed through voltage pulses, interface-type ReRAM devices usually require long pulses (ms scale) for achieving a fast and reliable operation of the neuromorphic system [7]. For these reasons this work is focused on filamentary-based ReRAM.

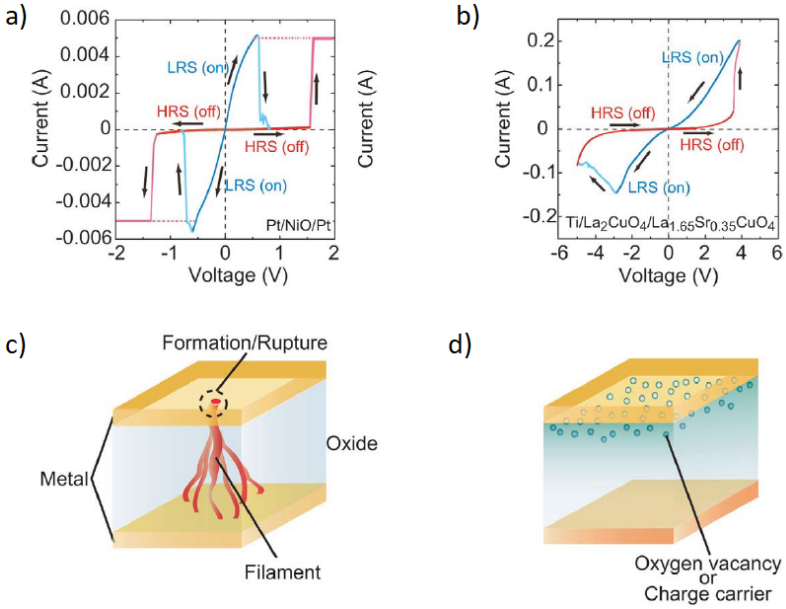


Figure 1.6: I-V curve for a unipolar switching ReRAM (a) and a bipolar one (b). The two main proposed models for resistive switching phenomenon are the filamentary (c) and interface-type one (d)[8].

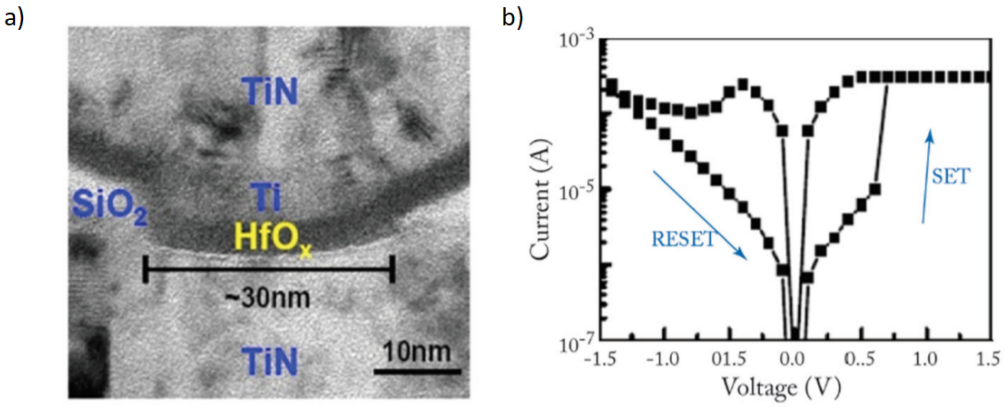


Figure 1.7: TEM image of a Ti/HfO_x/TiN ReRAM (a) and typical I-V characteristic of the device (b) [14].

The stability of the bipolar resistive switching in filamentary-based ReRAM is improved by employing an asymmetrical capacitor structure. In this work, to refer to this asymmetrical structure, the name "baseline" technology is used. In this technology, one electrode has a low work function, like in the case of titanium (Ti) or tantalum (Ta), while the other is typically platinum (Pt) or titanium nitride (TiN), since they have a very large work function [11]. The low work function metal is also

easier to oxidise with respect to the more inert Pt or TiN, and hence it is called oxygen exchange layer (OEL). An example Transmission Electron Microscope (TEM) image for a real device is shown Figure 1.7(a) and its corresponding I-V characteristic in (b). Instead, the baseline technology operation is schematically depicted in Figure 1.8.

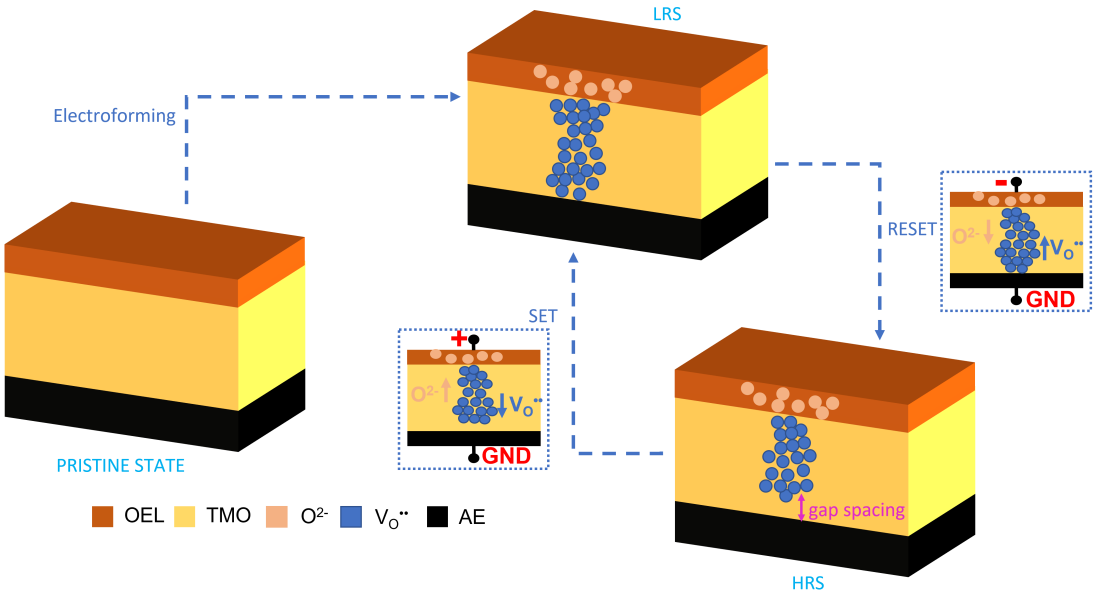


Figure 1.8: Schematic of baseline device operation.

In pristine state, filamentary-based devices are usually in an extremely high resistive state [4]. To enable the resistive switching phenomenon to occur, an initial electroforming step is required [10]. This process consists usually in the application of a positive or negative voltage at the OEL side, which can be thought as a distributed reservoir of defects at the metal-insulator interface [12]. The chemical reactions occurring at the OEL-TMO interface result hence in the gathering of oxygen ions (O^{-2}) by the OEL, with consequent injection of oxygen vacancy defects ($V_O^{\bullet\bullet}$ in the Kröger-Vink notation) forming a CF in the TMO through a soft-breakdown phenomenon [11][12][14]. After electroforming, the device is hence found in LRS and the switching to the HRS is reached by inducing the filament breakage. In particular, a negative voltage must be applied at the OEL to produce a migration of O^{-2} in the TMO, where they recombine through redox reactions with $V_O^{\bullet\bullet}$ till opening a gap at the oxide-inert electrode interface. On the other hand, applying a positive voltage at the OEL side results in the opposite migration of O^{-2} that leave behind $V_O^{\bullet\bullet}$. The filament breaking and restoring is thus the consequence of the oxygen vacancies redistribution within the TMO driven by a proper electric field and influenced by Joule heating effects [13]. It explains the modulation of the

local electronic conductivity due to the induced alteration in the valance state of transition metal cations [1][10][11]. A modulation of the width of the energy barrier at the interface with the high work function metal is the most accepted explanation of this resistive switching mechanism in literature [8][11]. According to this theory, being positively charged, oxygen vacancies act as donor-type centers [10] and their accumulation or depletion at the oxide-high work function metal interface respectively favour or not the carrier conduction [8]. The inert electrode plays hence a key role during the switching and for this reason it is also called active electrode (AE).

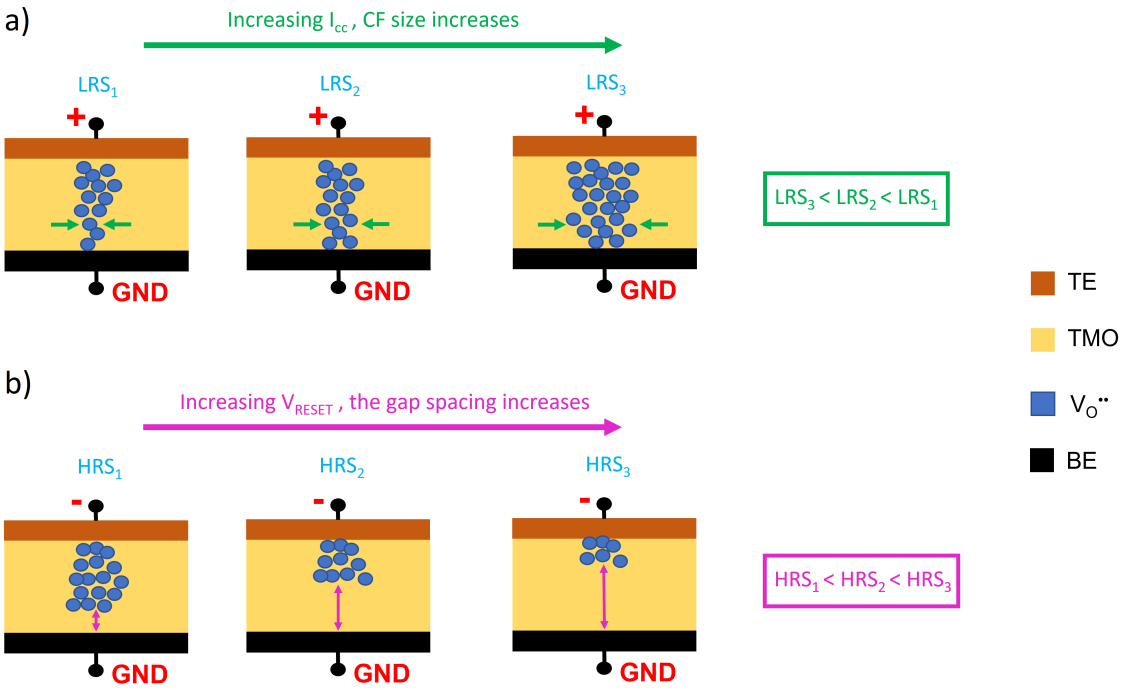


Figure 1.9: Schematic illustration on how distinguishable resistance levels can be achieved by varying the compliance current (I_{CC}) during SET transition (a) or the magnitude of the RESET voltage (V_{RESET}) (b). Redrawn from [14].

As already mentioned in Section 1.1, the large amount of distinguishable resistance levels achievable in resistive switching devices makes them suitable for the implementation as neuromorphic memristor-based artificial synapses. In baseline technology, multiple HRS and LRS can be achieved mainly in three ways:

- An upper bound to the current flowing in the device, which is called compliance current (I_{CC}), can be applied to limit the strength of the SET transition. As it is possible to observe from Figure 1.9(a), increasing the I_{CC} produces the widening of the CF due to the larger migration of oxygen vacancies [7] with

consequent reduction of the resistance value in LRS. In this way, multiple LRS levels can be achieved while the HRS remains constant. However, the wider it is the filament created and the larger should be the current required to flow in the device to break the filament during the RESET [14]. Practically, this method is implemented by controlling the gate voltage applied to a transistor in series to the ReRAM device (1T-1R configuration).

- The magnitude of the RESET voltage (V_{RESET}) can be controlled in order to obtain distinguishable HRS levels, while the LRS is maintained constant. Indeed, as it is observable in Figure 1.9(b), an increase of V_{RESET} produces an enlarging of the gap spacing with consequent increase of the resistance state in HRS. This approach is often employed since it reduces the complexity of the circuitry around the devices [14].

- The voltage pulse width (PW) during both potentiation and depression can be adjusted to define the resistance value in both LRS and HRS because of the more controlled electrical stimulation of the device. However, this method is poorly energy efficient due to the higher unwanted heat dissipation in the device [14].

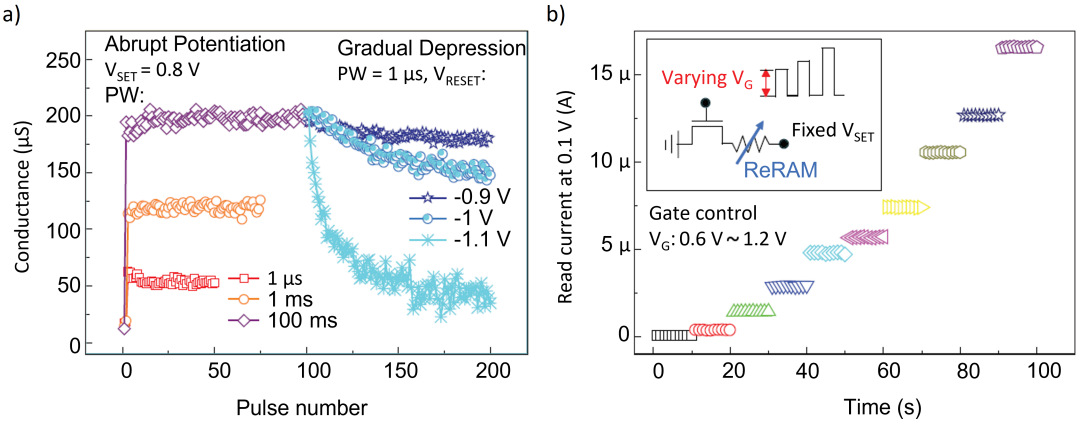


Figure 1.10: An example of pulsed programming of a Ti/HfO_2/TiN ReRAM device during both potentiation and depression (a). Multiple states can be achieved during potentiation by non-identical pulses that increase I_{CC} (b). Adapted from [7].

As motivated in Section 1.1, the accuracy in the learning process of ANNs strongly depends on the analogue tuning of the resistance state and it is actually in this regard that the baseline technology shows its limitations. Figure 1.10(a) shows how applying identical pulses for both potentiation and depression, the HRS in filamentary-based devices is abruptly switched to the LRS in a single pulse, whose width defines the value of the resistance state. The abruptness of the transition is

attributed to the positive feedback between the rise of the electric field in the gap region and the filament growth [2]. The LRS can be then gradually switched to the HRS by controlling the amplitude of the RESET pulse. The more gradual nature of the RESET transition is attributed to a negative thermal feedback and counteracting forces of drift and diffusion of $V_O^{\bullet\bullet}$ approaching equilibrium concentration [15].
The abrupt and stochastic behaviour during the SET transition is hence the main obstacle to a highly accurate learning process of the neural network. The problem can be mitigated by adjusting the resistance value in LRS through the amplitude of the voltage pulse applied to the gate of a transistor in series to the device, thus controlling the I_{CC}, as shown in Figure 1.10(b). However, the application of non-identical pulses results in a more complex peripheral circuitry, as well as in an increased latency and power consumption [7]. This challenges make the programming with non-identical pulses impractical for on-chip implementations [2]. In this regard, several optimized programming schemes for both potentiation and depression have been proposed [7]. They rely on combining voltage pulses in trains to obtain more gradual and linear state transitions by the partial connection or dissolution of the CF. This approach still poses some problems mainly regarding the time required to establish the best pulse pattern, even if the circuit complexity can be reduced since only one SET train and one RESET train must be applied.
Filamentary-based ReRAM are today the most advanced emerging technology in terms of scaling and integration with CMOS based architectures [9], but they are not yet ready to fulfil all the requirements of neuromorphic systems.

1.3 Bilayer ReRAM: improvement of the weight update

The intrinsic drawback of a poorly analogue and asymmetric weight update in baseline ReRAM devices lead to a further engineering of the memristive technology passing through several methodologies. In particular, the most successful methods were based on the manipulation of the properties of the insulator layer by modulating the parameters of its deposition process, or on employing various doping techniques, such as metal clusters or inorganic nanocrystals [4], that lower the formation energy of oxygen vacancies [7]. Although several advantages were introduced by these techniques, such procedures resulted in more complex manufacturing processes, and hence, also in more expensive ones [4]. To keep the fabrication process simple and cost-effective, in recent years bilayer ReRAM devices have been investigated to achieve the analogue, linear and symmetric resistive switching which is capable to mimic the learning process in biological synapses.
The phenomenon explaining the gradual switching is still open and different explanations appear in literature. In 2016, Woo et al. [16] reported the improvement

of a Ti/HfO_2/TiN ReRAM by depositing the HfO_2 on top of an aluminium (Al) layer. The ALD deposition of HfO_2 induces the oxidation of Al, with the consequent formation of an AlO_x/HfO_2 bilayer. According to the authors, the CF is formed through the entire thickness of the oxide stack but since the constriction is placed at the interface between AlO_x and HfO_2, an analogue switching is obtained. Indeed, the lower mobility of $V_O^{\bullet\bullet}$ in AlO_x than in HfO_2 limits their dissolution from the CF while the lateral modulation of the CF width is preferred. This incomplete annihilation of the filament and modulation of its width result in an extremely gradual, linear and symmetric weight update but the On/Off ratio is reduced, as it is possible to observe from Figure 1.11 (a).
A neuromorphic system based on a multilayer perceptron model was then used in [16] to test the accuracy of pattern recognition on the Mixed National Institute of Standards and Technology (MNIST) dataset. Figure 1.11(b) clearly shows that the training accuracy is drastically improved by linearly and symmetrical updating the weights of the network, even if the resistance window is narrower than in the case of baseline technology.

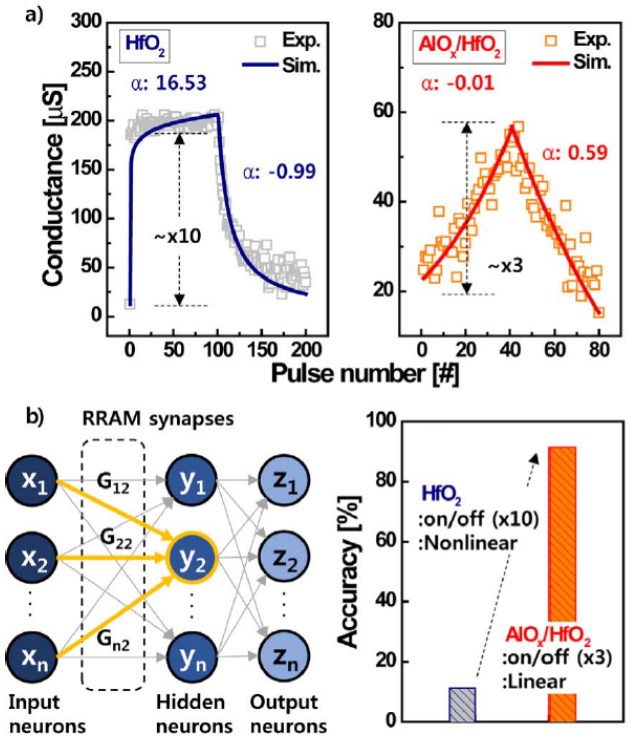


Figure 1.11: a) Comparison of the synaptic behaviour obtained from the HfO_2 and AlO_x/HfO_2 devices. b) Accuracy obtained with the two technologies in a pattern recognition with a neural network using a multilayer perceptron algorithm [16].

The idea that the bilayer ReRAM device works as a stack of two switching layers is proposed in a slightly different way by Kim et al. in [17], where the concentration of $V_O^{\bullet\bullet}$ and the CF are modulated by stacking a sputtered TaO_x on top of an ALD deposited HfO_2. Since the sputtering technique causes more physical defects than the ALD one, the TaO_x should have a larger amount of oxygen vacancies than HfO_2, thus leading to a two stages CF formation process. First a thicker filament is formed in the oxygen vacancies-rich TaO_x layer and then a weaker second one along the already formed CF is generated in the HfO_2. According to the authors, the control of the filament formation or dissolution in correspondence of its thinnest part at the interlayer between the two oxides determines the gradual resistive switching in both static and pulsed programming, as it is possible to observe in Figure 1.12.

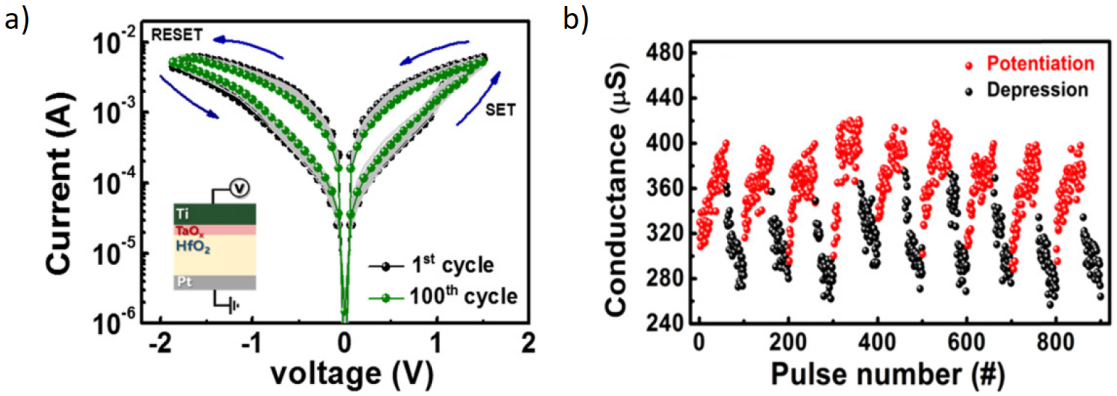


Figure 1.12: a) I-V characteristic of the ReRAM device based on $Ti/TaO_x/HfO_2/Pt$ stack. b) Pulsed programming of the bilayer ReRAM device with identical pulses [17].

Cüppers et al. [15] attribute the gradual transition on either SET or RESET to the inherent current limitation provided by the additional insulating layer stacked on top of the TMO, but the effectiveness of this mechanism depends on the resistance state of the device. In Figure 1.13(a) it is shown the device stack studied in [15] and it is possible to notice that the combination of the TiO_x interface layer and the metallic interconnection contributions acts as inherent conduction limiter (ICL) in the equivalent circuit model. The bilayer ReRAM device studied in [15] shows two modes of operation. In the first operation mode, the device starts from an extremely large or low resistance state, hence the ICL can limit the current through the device only initially, thus resulting in an abrupt transition to the opposite state. On the other hand, a gradual and self-compliant state update on both polarities is obtained by changing the applied voltage in steps of 10 mV so that the resistance difference between initial and final state is reduced, as it is represented by the coloured lines in Figure 1.13(b).

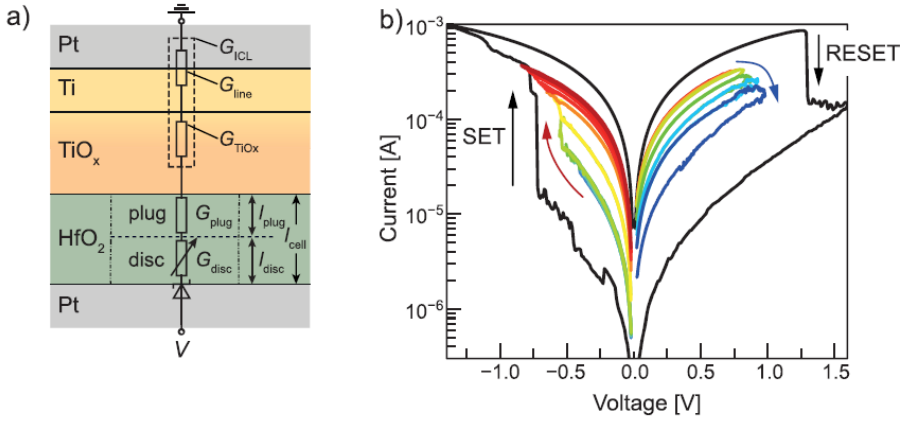


Figure 1.13: a) Stack of the TiO_x/HfO_2 ReRAM and equivalent circuit. b) I-V sweeps of the proposed device that can show an abrupt behaviour (black curves) or a gradual one (colored curves) [15].

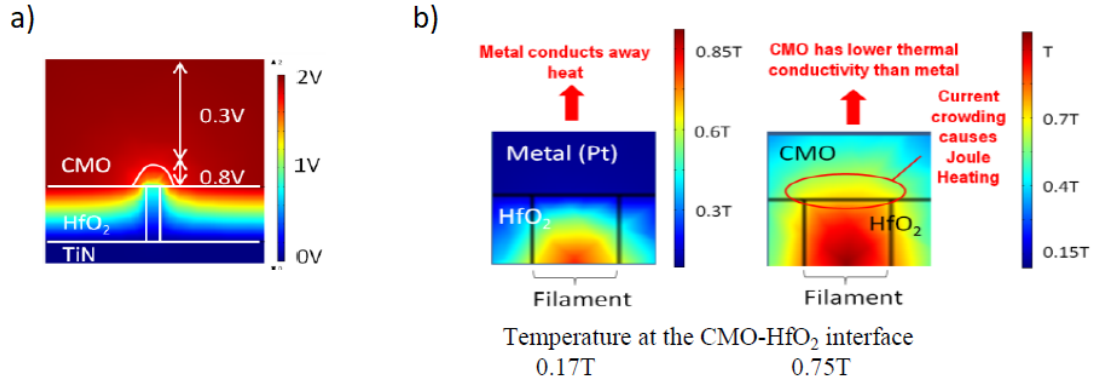


Figure 1.14: a) Simulation of the voltage drop in the TaO_x (CMO)/HfO_2/TiN device stack during SET. b) Comparison of the temperature distribution at the interface with the HfO_2 between having Pt or CMO as top electrode [19].

The role of the TiO_x in the device stack is, however, much more complicated. Indeed, in a previous article [18], the authors state that the TiO_x affects also the temperature distribution during switching (by modifying the heat dissipation) and the current conduction in the HRS, rather than acting as a simple series resistance. This idea is supported also by a previous work of Sekar et al. [19] on TaO_x/HfO_2 bilayer ReRAM device. In [19], the TaO_x is not a purely insulator, hence it is referred as conductive metal oxide (CMO). Figure 1.14(a) shows a significant voltage

drop occurring on the CMO during SET and the consequent current crowding at its interface with HfO_2. This suggests that the CMO resistivity has an impact on both switching and retention properties. Moreover, since the current crowding results in the temperature increase because of Joule heating effect, the thermal conductivity of the CMO layer must be carefully considered. The simulation in Figure 1.14(b) demonstrates how a lower thermal conductivity of CMO with respect to a metal electrode confines the heat at the TaO_x-HfO_2 interface, thus enhancing the ion motion.

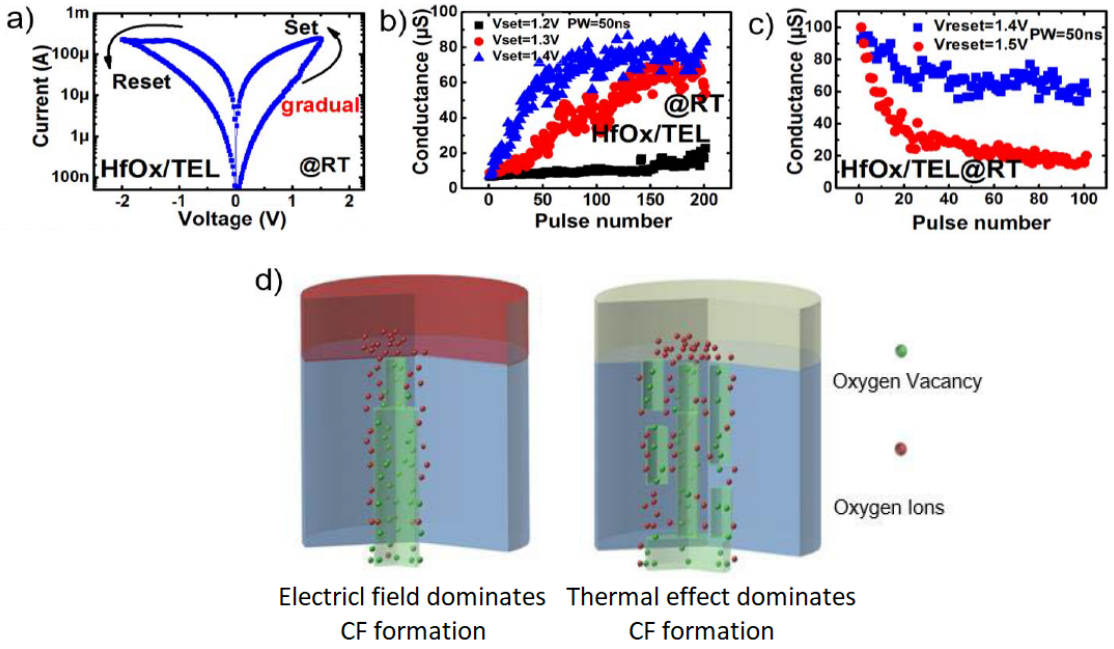


Figure 1.15: a) I-V sweep on TaO_x (TEL)/HfO_2 bilayer ReRAM at room temperature. b) Analogue conductance change using identical SET pulses. c) Analogue conductance change using identical RESET pulses. d) Schematic representation of the single strong CF due to the dominance of electric field and the multiple weak ones obtained when thermal effect is predominant [20].

Following the reasoning in [19] and observing a transition from abrupt to analogue behaviour in ReRAM devices by switching the device at high temperatures, Wu et al. [20] proposed to develop a thermal enhanced layer (TEL) that stacked on top of HfO_2 realizes an analogue weight update of the ReRAM. In [20] the TEL is still an oxygen deficient tantalum oxide, with a low thermal conductivity but a high electrical one. In agreement with [19], this layer can confine the heat during switching and thus allowing for the achievement of a gradual state transition at temperatures compatible with practical applications, as it is observable in Figure 1.15(a), (b) and

(c). According to the authors, a strong single CF should be formed if the positive feedback induced by the electric field dominates, thus resulting in a resistance state that depends on the CF formation and rupture. In contrast, the TEL enables the growth of multiple weaker CFs because of the induced more uniform distribution of oxygen vacancies. The resistive state results in the sum of the contribution of each CF, whose annihilation has a lower impact on the total resistance. A schematic representation of both scenarios is reported in Figure 1.15(d).
Although the resistive switching mechanism that enables the biological synaptic emulation is not completely clear, the potentialities of bilayer ReRAM devices as building blocks for future neuromorphic computing architectures are huge. This is especially true if a co-integration of crossbar array with CMOS transistors is realized. For example, Yao et al. [21] demonstrated the face recognition capabilities of a one-layer perceptron of 128×8 $TaO_x/HfAl_yO_x$ ReRAM trained with the Yale Face Database and tested on unseen faces. Using a write-verify weight update rule, the network is capable to achieve 91.7% of accuracy, while this value lowers to 87.5% if a simplified scheme without program-verify is employed.

1.4 Final goal of the proposed work

The practical realization of novel computing architectures based on bioinspired neuromorphic computing requires memristors to have an analogue multilevel behaviour as well as a low intra-device and inter-device variability. In this regard, the realization of bilayer redox-based VCM ReRAM devices seems to be a simple as well as cost-effective method to optimize these characteristics with respect to the baseline technology. However, several challenges are still present. To increase the success rate of this goal, a deeper understanding of how the analogue switching mechanism occurs is of paramount importance.
In this thesis work the deposition of the substoichiometric tantalum oxide layer is optimized to obtain a CMO, and the impact of its conductivity on the ReRAM performances is studied. To further improve the device switching, we analyzed the influence of the thickness of this CMO layer, which is related to both electrical and thermal effects. Eventually, we gave a physical interpretation of the mechanism behind the gradual resistive switching of the proposed ReRAM, by modelling the results obtained within a dedicated Impedance Spectroscopy experiment.
The contribution of this work will hopefully help to identify and to tune the main knobs that will make bilayer redox-based ReRAMs the artificial synapses of future ANNs.

Chapter 2

Processing methods

2.1 Thin Film Deposition

In this section a theoretical overview on the techniques used for the deposition of the thin-film materials is provided. These techniques can be divided in:

- Physical Vapor Deposition (PVD), such as sputtering and evaporation, in which material is physically removed from a source, or target, and it accumulates onto a cooler substrate;
- Chemical Vapor Deposition (CVD), in which the thin-film is formed by the reactions among gaseous phase precursors;
- Atomic Layer Deposition (ALD), in which the film is grown by the chemical reactions among gaseous precursors introduced sequentially into the process chamber.

2.1.1 Physical Vapor Deposition - Sputtering and Evaporation

Sputtering deposition

Sputtering is a thin-film deposition technique used for depositing both metallic and insulating films and it is based on the physical removal of material by ion bombardment from a source, usually called target. At the very beginning of the deposition, a gas consisting of neutral atoms, frequently argon (Ar), is inserted in the chamber and then a plasma is generated by the application of a Direct-Current (DC) voltage or a radio frequency (RF) power, typically at 13.56 MHz. As a consequence of the voltage application, free electrons are accelerated in the plasma and elastically collide with the neutral argon atoms, thus generating ions Ar^+. The ions are then accelerated towards the cathode that hosts the target by the voltage applied, thus

resulting in a physical ablation of material and the consequent deposition of the sputtered atoms, which mostly arrive as neutral ones on the substrate surface. The substrate is positioned usually in the bottom part of the chamber onto an holder. The holder acts as the anode and it is electrically connected with the chamber walls at a ground potential. Sometimes an heater is present at the anode in order to increase the temperature of the substrate during the deposition allowing for a better control on the crystallographic growth of the thin-film, on its deposition rate as well as on the material stoichiometry. An outlet and a vacuum pump allow to maintain a constant value of pressure in the chamber during the deposition. A schematic representation of DC and RF sputtering processes is provided in Figure 2.1.

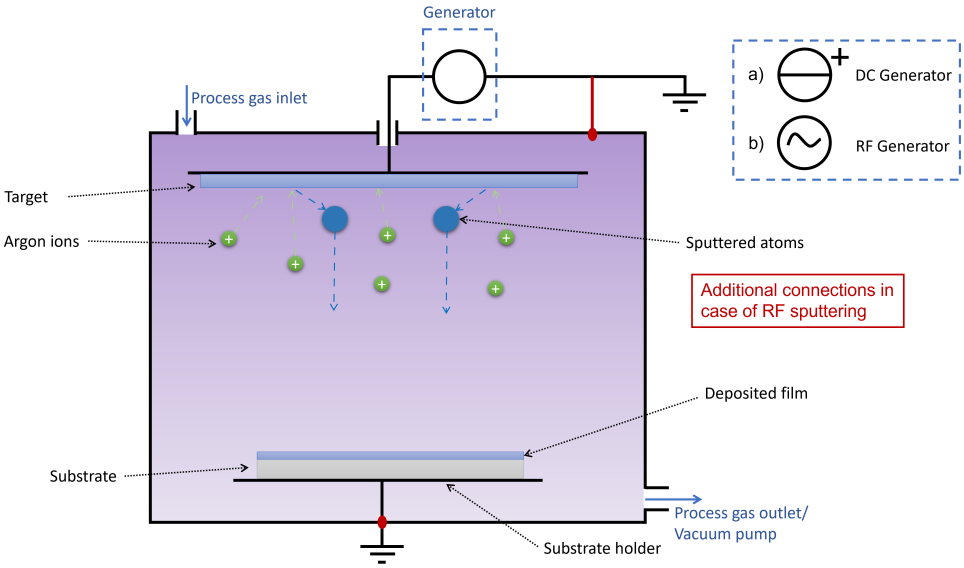


Figure 2.1: Schematic representation of DC sputtering (a) and RF sputtering (b). Red features stresses the connection of the substrate holder with the chamber walls, which is required to obtain a RF sputtering process mostly occurring at the target. Partially redrawn from [22].

Instead of DC sputtering, when a thin-film is deposited starting from a non-conductive target RF sputtering should be used in order to avoid the accumulation of the electric load [22]. In DC sputtering, the acceleration of the sputtering ions occurs directly in front of the target, since far from it the Ar^+ ions undergo additional collisions thus loosing their energy on the long path to the target. Moreover, no physical removal of material at the anode occurs [22]. This is not true in the case of RF sputtering, in which a similar re-sputtering rate is obtained at both electrodes, due to the slower mobility of ions versus electrons that results in a positively biased plasma with respect to both electrodes. To obtain a sputtering process mostly

occurring at the target, the anode is usually connected to the chamber walls, thus resulting in a smaller area of the cathode with respect to the anode [22]. Indeed, in this way the electric field is larger close to the smaller electrode (the target) since the overall current continuity must remain constant [22].
RF sputtering is used mostly for the deposition of insulating materials, while for conductive ones DC sputtering is preferred since it leads to a higher deposition rate.
Magnetic fields perpendicular to the electric one can be used to sustain the plasma in close proximity to the target thus allowing an improvement of the ionization rate and, as a consequence, of the sputtering yield at lower sputtering gas pressure [22]. The improvement in the ionization rate is explained by the higher ionization probability resulting from forcing electrons to move on helical path in close proximity of the target. The magnetic fields are generated by an array of magnets placed at the cathode and in this case the sputtering technique takes the name of magnetron sputtering. The magnetron sputtering configuration for both DC and RF sputtering is shown in Figure 2.2. The biggest advantages of magnetron sputtering with respect to the common configuration are the larger deposition rate and the lower incorporation of argon into the film working at lower gas pressures. On the other hand, a more inhomogeneous target erosion is obtained [22].

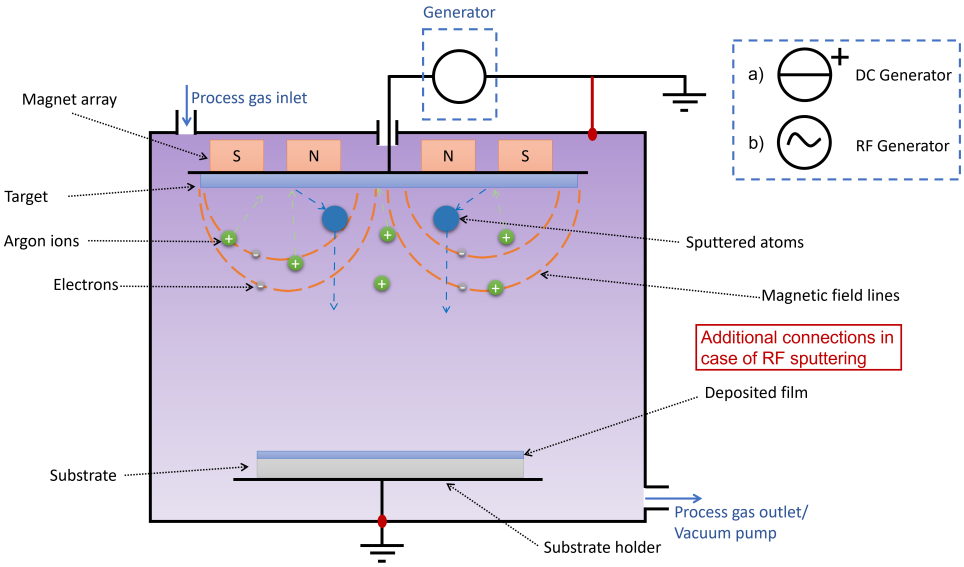


Figure 2.2: Schematic representation of DC magnetron sputtering (a) and RF magnetron sputtering (b). Red features stresses the connection of the substrate holder with the chamber walls, which is required to obtain a RF sputtering process mostly occurring at the target. Partially redrawn from [22].

A solution to deposit dielectric materials starting from a conductive target or to

simply refine the stoichiometry of the deposited layer is reactive sputtering. In this case, a reactive gas is introduced into the chamber and it is typically oxygen (O_2) or nitrogen (N_2). The reactive gas molecules can react with the material sputtered from the target resulting in the deposition of a material different from the target one. A schematic representation of DC magnetron reactive sputtering is provided in Figure 2.3.

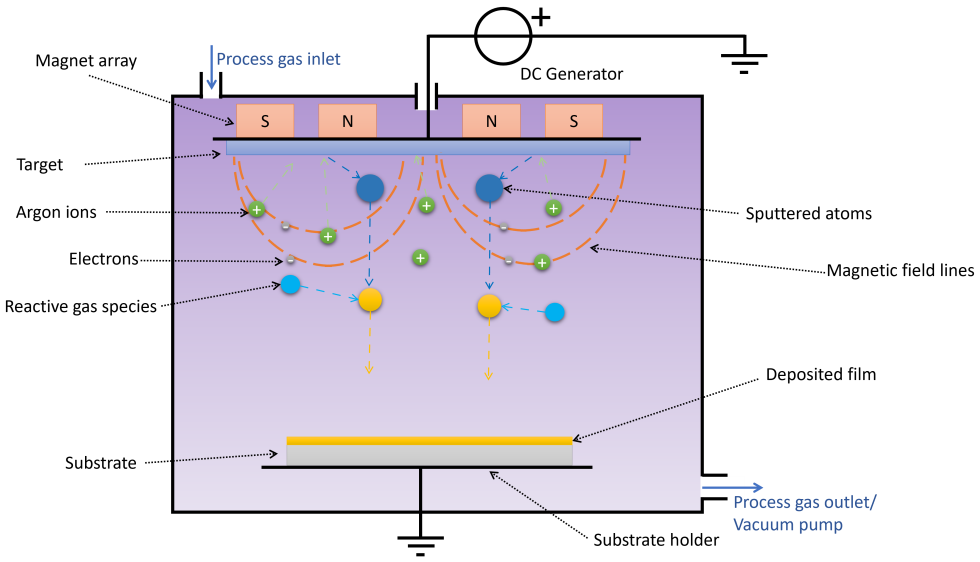


Figure 2.3: Schematic representation of DC magnetron reactive sputtering. Partially redrawn from [22].

The sputtering tool used in this work is the *CS 320 S* from *VON ARDENNE*.

Thermal and E-beam Evaporation

Evaporation is a PVD technique to deposit thin-film which can be performed in two ways: thermal evaporation and E-beam evaporation. In both cases, a source material hosted in a crucible is heated up in vacuum and the evaporated atoms condense on the surface of the sample, which is mounted upside down onto an holder to reduce the particulate absorption, and hence contamination [22].
The main difference between the two techniques consists in how the heating to generate the material to be deposited is obtained. In thermal evaporation, it is accomplished by using an heater like a tungsten filament in which a current typically of the order of 100 A flows. On the other hand, E-beam evaporation is based on the thermionic emission of a filament with consequent acceleration and focusing by a magnetic field of a high-energy electron beam onto the source material to locally melt it [22]. The schematic of the two evaporation processes is depicted in Figure

2.4.
With respect to thermal evaporation, higher temperatures can be reached in the E-beam process, thus allowing for the deposition of a wider range of materials. Moreover, being melted only the top part of the source material, the E-beam evaporation results in a reduced amount of contamination from the crucible [22].

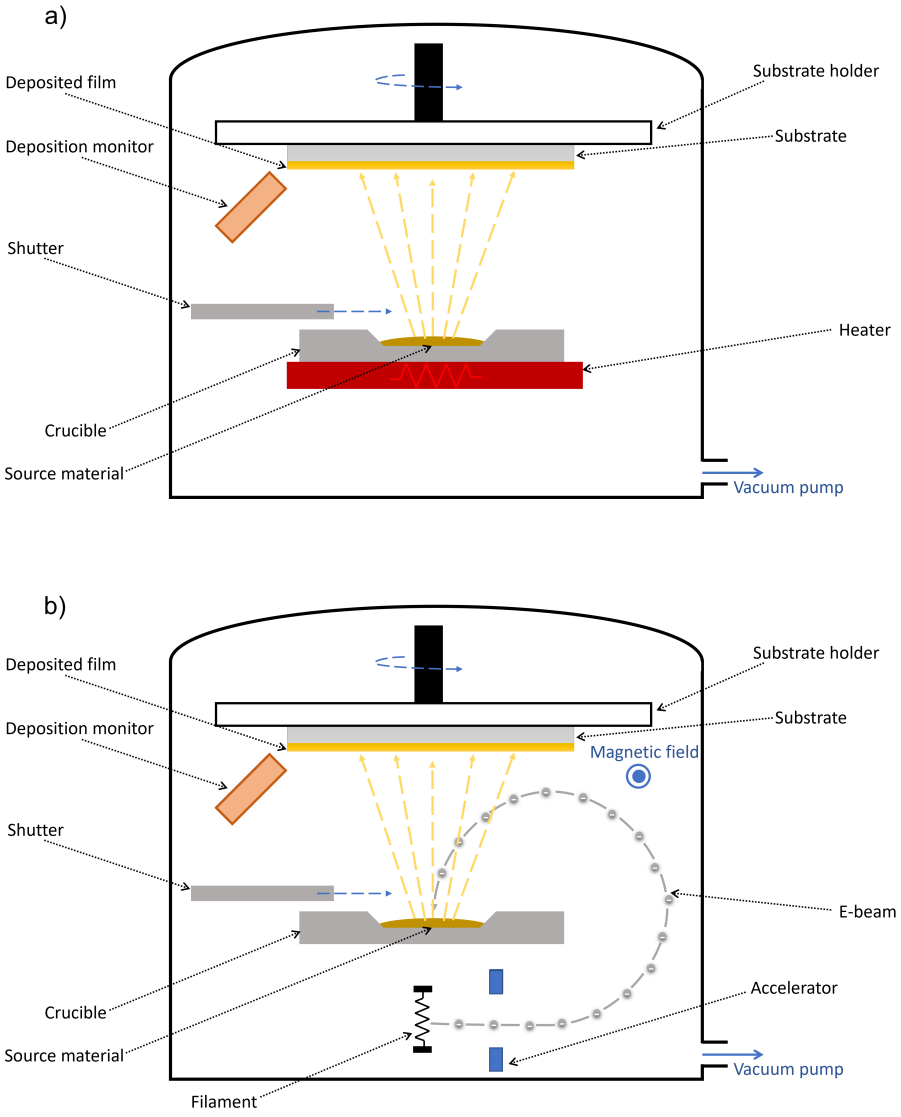


Figure 2.4: Schematic representation of thermal evaporation (a) and E-beam evaporation (b). Redrawn from [22] and [23].

The evaporation process is performed at even lower pressures than in sputtering. The high vacuum condition in which the evaporated atoms are transported results

in a reduction of contaminants in the chamber, as well as a strong reduction of gas-phase scattering phenomena. In particular, the reduction of scattering produces a longer mean free path for the atoms which travel mainly in essentially straight lines till reaching the surface of the sample [22]. The main consequence of this straight path of the evaporated atoms is the non-conformality of the deposition: the maximum of deposited material is on horizontal surfaces, while a minimal thickness is reached on the vertical sidewalls of the features, with no deposited material on features that result shadowed by others.
The deposition is usually controlled by monitoring tools like piezoelectric quartz disk and once the target thickness is reached, the material flux is abruptly stopped by a closing shutter. Moreover, another feature usually present in evaporation systems is the possibility to rotate the substrate holder in order to reduce the non-ideal and non-uniform behaviour of the deposition resulting from the high density of evaporated atoms right above the source [22].
The major advantages of the evaporation process are the little demage of the sample, the high purity of the deposited film and also the higher deposition rate if compared with other thin-film deposition techniques like sputtering. However, the evaporation technique presents also several limits such as the difficulty to deposit alloys or thin-film with precisely controlled stoichiometry, poor step coverage and the absence of any in situ pre-cleaning method [22].
The Evaporator tool used in this work is the *BAK501 LL* from *Evatec*.

2.1.2 Chemical Vapor Deposition - PECVD

Chemical Vapor Deposition (CVD) is a class of deposition techniques in which a thin-film is grown starting from gaseous reactants which are broken into reactive species by an high energy source [22]. Among the CVD deposition techniques can be cited for example Atmospheric Pressure CVD (APCVD), Low Pressure CVD (LPCVD) and Plasma Enhanced CVD (PECVD). The latter is here described in details, since it has been used in this work for the deposition of slicon dioxide (SiO_2) and silicon nitride (Si_3N_4). PECVD is usually the preferred technique to deposit these layers because it allows for low deposition temperatures but still maintaining an high deposition rate with respect for example to LPCVD. This makes PECVD also compatible with back-end-of-the-line (BEOL) processing. On the other hand, PECVD is a much more complex tool with respect to the other CVD techniques. Indeed, a plasma is generated by an RF bias in addition to the thermal source required to provide the energy for the occurrence of chemical reactions. The principle of the plasma generation is exactly the same of reactive sputtering, with the difference that now electrons generate free radicals by the collision with neutral gas molecules. The free radicals are extremely reactive species and thus the surface processes and deposition can occur at much lower temperatures than non-plasma systems, and even at room temperature [22]. The properties of the deposited layer,

such as the density and the stress, can be tailored acting on additional degrees of freedom with respect to the other CVD techniques, such as the RF power, the frequency and the bias. On the other hand, incorporation of byproducts into the layer should be carefully controlled [22]. An example of equipment configuration is shown in Figure 2.5, in which the reactive gas are fed in by inlets and an RF generator at a frequency usually of 13.56 MHz is used to generate the plasma. Byproducts are exhausted through the outlets.

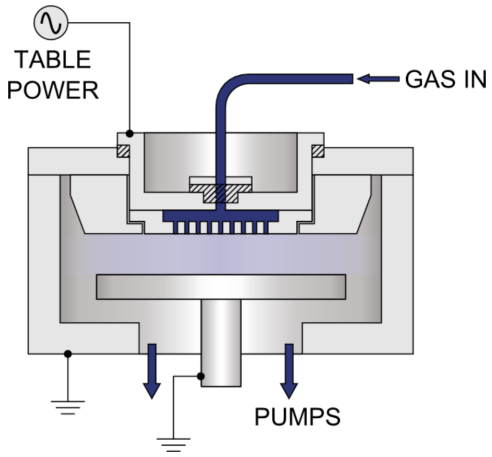


Figure 2.5: Schematic representation of PECVD tool [24].

The PECVD tool used in this work is the *PlasmaPro 100* from *Oxford Instruments*.

2.1.3 Atomic Layer Deposition - PEALD

Atomic Layer Deposition (ALD) is a thin-film deposition technique that allows to precisely deposit even few nanometers of material really layer by layer with excellent control, film quality and uniformity. Even 3D structures with high aspect-ratio can be covered in a conformal way at low process temperatures [25]. The technique relies on repeating for the time required to obtain a target thickness of deposited material a cyclic process of four steps, in which precursors are inserted one at the time in the chamber and residuals are purged after that self-limiting reactions take place. However, even though this deposition technique looks quite straightforward, it presents several challenges such as high thermal budget for actual device application and low throughput [26]. The material deposition can hence be performed at lower temperatures by introducing into the ALD chamber a plasma of reactant molecules that dissociates in reactive species during the process, thus the name of Plasma Enhanced ALD (PEALD) of this technique. In particular, the main steps

of the PEALD process, which are also illustrated in Figure 2.6(a), are the following:

1 Precursor adsorption step: the precursor is introduced into the chamber and it is chemisorbed on the surface through self-limiting reactions.

2 Purge step: A carrier gas is introduced into the chamber to remove the excess of the precursor and reaction byproducts from the chamber.

3 Plasma exposure step: a plasma containing strongly reactive molecules, like free radicals, is generated and these reactive species react with the already adsorbed precursor on the surface of the sample.

4 Purge step: reaction residues from the previous step are purged away from the chamber.

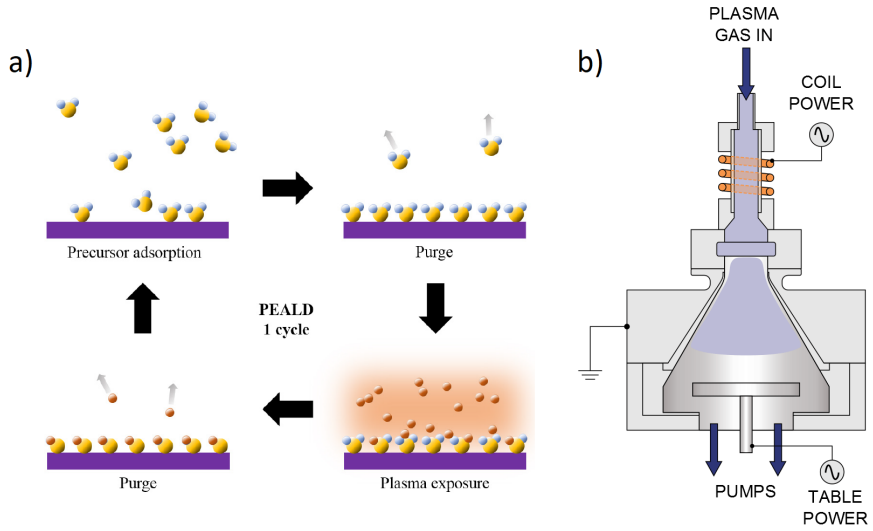


Figure 2.6: Schematic representation of a PEALD cycle (a) [26] and the PEALD tool (b) [25].

The plasma is generated using a capacitively coupled plasma technique, which is the the one also employed in standard sputtering deposition techniques, or inductively coupled plasma, whose example of reactor is shown in Figure 2.6(b). This kind of plasma generation is described in more details in Section 2.2.4 since it is also used for plasma etching.
Controlling the plasma conditions, such as the RF power and the frequency, is fundamental in this technique since the plasma characteristics, like density of radicals, energy and density of electrons and ions have a considerable influence on the deposited film [26]. For example an high plasma power could result in a larger surface

damage or unwanted incorporation of impurities in the deposited film [26].
The PEALD tool used in this work is the *FlexAL* from *Oxford Instruments*.

2.2 Patterning techniques

This section is intended to give the reader an overview about the main techniques used in this work for transferring a pattern onto a deposited film. These techniques comprehend lithography techniques, such as photolithography and laser lithography, to pattern a photosensitive material called photoresist, but also techniques used to remove material, such as lift-off and etching.

2.2.1 Photolithography

Photolithography is a technological process that allows to transfer a desired pattern onto a photosensitive material called photoresist, or simply resist, by exposing it with a light source. The goal of this process it to define the areas on the chip where following processes of material deposition or removal will take place. In order to define the pattern to be transferred on the chip, a physical light filter called photolithographic mask must be prepared to selectively expose the photoresist. It is designed as a .gdsII file with Computer-Aided Design (CAD) software, such as *KLayout* and *L-Edit* used in this work, and it can contain more than one lithographic layer according to the number of lithographic steps that must be performed. The information of each mask level is then transferred using a laser or an electron beam which exposes a photoresist deposited on a silica plate covered with a thin layer of chromium. Then the resist is developed and the chromium is etched away using usually a wet etching process. Finally, the remaining resist is stripped away thus leaving on the photolithographic mask transparent regions (where the silica is) and absorbent regions (where the chromium remained in place).
Once the mask has been physically realized, the photolithgraphic step can take place. It consists in printing the mask pattern in a layer of phototresist, which has been previously spun and baked on a wafer surface after the deposition of Hexamethyldisilane (HDMS) as adhesion promoter. The printing is performed in this work shining through the mask a high-intensity light, which is generated from a mercury (Hg) lamp, using a contact exposure system (*Mask Aligner MA6* from *Süss Microtec*). In this kind of system the chip is aligned with respect to the mask taking them apart at 30 µm of distance and then the resist on the chip is placed in direct contact with the chromium part of the mask for the required exposure time. This photolithographic system allows for high-resolution (≈1 µm of minimum feature dimensions transferable), since diffraction effects are minimized by the contact between the chip and the mask. However, the hard contact results also in high defect densities thus limiting the application of this exposure system to low volumes and small chip sizes [22].

Once the photoresist has been exposed to light, it undergoes a chemical alteration due to bond-breakage for then restructuring itself into a new stable form [22]. Two main polarities of the resist can be identified: positive photoresist, which becomes more soluble in the developer solution when exposed; whereas negative photoresist does the opposite. This different behaviour of photoresists results in the generation of a pattern transfer on the chip once the resist has been developed in the proper solvent, with the same pattern on the mask being transferred onto the chip in the case of a positive resist, while the inverse one in the case of a negative photoresist, as it is possible to observe in Figure 2.7.

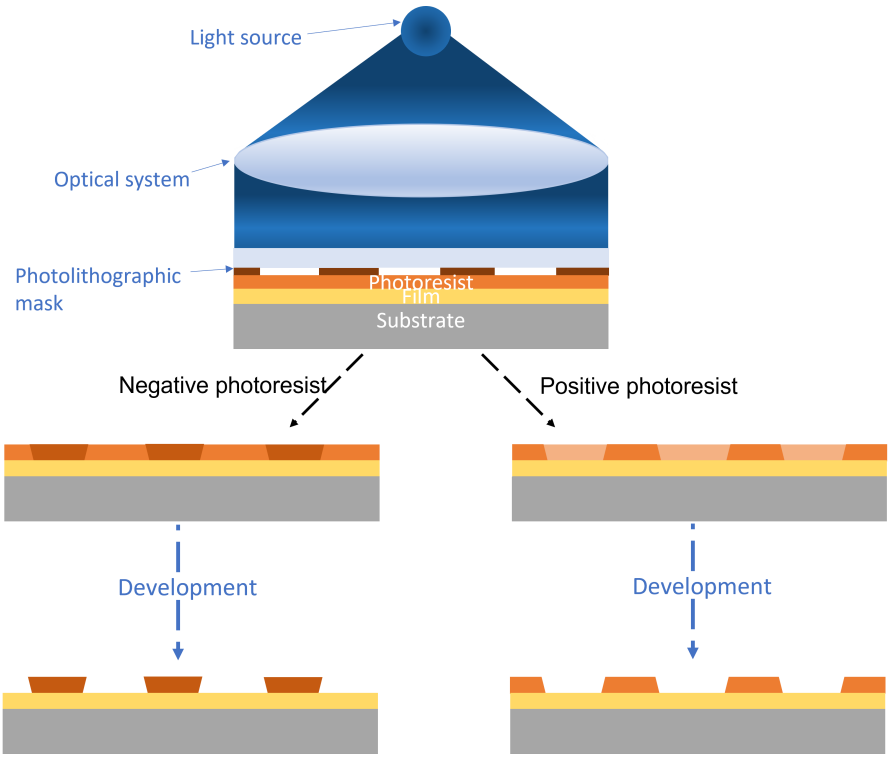


Figure 2.7: Schematic representation of photolithography in the case of a contact exposure system. The pattern transfer in the case of positive and negative polarity of the resist is represented. Partially redrawn from [22].

For more details about the discussed technique as well as other more advanced photolithographic exposure systems, the reader is advised to literature [22].

2.2.2 Maskless Laser Lithography

Laser lithography is a lithographic process that, as photolithography, allows to transfer a desired pattern onto a photoresist. The difference with respect to photolithoraphy is that the photosensitive material is exposed through a laser beam

and no more the light generated by an high-intensity Hg lamp. Moreover, the laser lithography is a maskless technique, in the sense that no physical mask must be fabricated to selectively define the region where the resist is exposed. Indeed, in laser lithography the pattern is exposed directly onto the chip using a spatial light modulator, which essentially works as a programmable mask of the design file [27]. A laser beam is hence deflected by a mirror and focused by an optical system on the photoresist surface in order to define on it the designed pattern. The principle of operation of maskless laser lithography is reported in Figure 2.8.

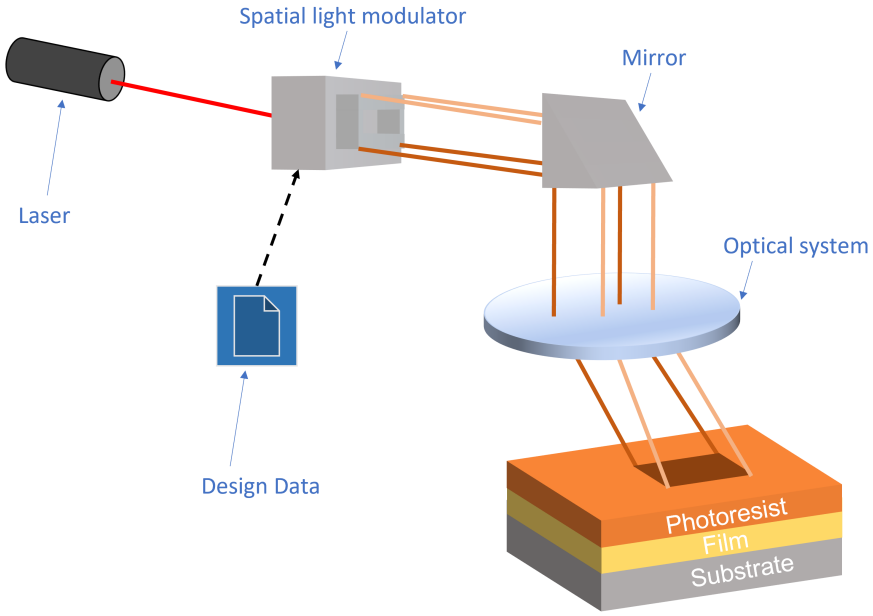


Figure 2.8: Schematic representation of maskless laser lithography process. Partially redrawn from [27].

The direct-writing approach followed by laser lithography allows to be much more time-efficient in the lithographic step, since the alignment is performed automatically by the tool, and also to have much more versatility in changing the design without the need to fabricate again a lithographic mask [27]. On the other hand, it is a much more expensive tool than a standard mask aligner used for contact photolithography. The tool that is used in this work is the *Heidelberg DWL 2000*, which can reach a sub-1 µm resolution using a 413 nm laser wavelength and 500 mW of maximum power.

2.2.3 Lift-off

Lithography usually precedes a processing step in which a material is deposited or removed using etching techniques, such as the ones described in the Section 2.2.4.

In particular, when the next step after the resist patterning through lithography is the deposition of new material, the resist acts as a mask to protect the regions where the new material should not be deposited. Once the deposition is completed, the unwanted material and the resist itself should be removed. This patterning approach is called lift-off and it is a very well known microfabrication technique especially for its simplicity, since no chemical selectivity of the layers should be considered as it is instead the case for etching approaches. Indeed, the removal of the resist with the material deposited on its surface is accomplished by submerging the chip into a solvent (typically acetone) and agitating it with ultrasonic in order to improve the efficiency of the stripping process. Sometimes an oxygen plasma may be required to further remove unwanted resist residues by chemical reaction with the organic material [28].

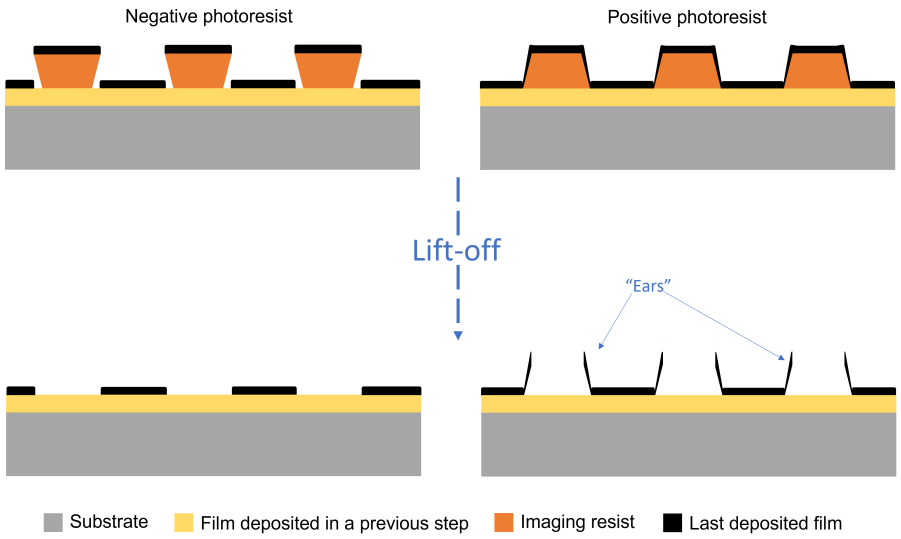


Figure 2.9: Schematic representation of lift-off process using a photoresist with negative or positive polarity. Partially redrawn from [28].

Positive resists are usually favoured for their better resolution with respect to photoresist having a negative polarity [22] but they are not well-suited for lift-off process. In this process, negative photoresists are usually preferred for the undercut profile of the resist sidewalls. Indeed, the absence of undercut profiles when using a positive photoresist leads to a partial deposition of material on the sidewalls with consequent reduction of the solvent access points to the resist and the appearance of "ears" at the edges of the defined features, as it is possible to observe in Figure 2.9. In order to avoid the appearance of the "ears" and still using a positive photoresist, common alternatives can be to increase the thickness of the resist, thus to reduce the amount of deposited material on the sidewalls, or to use positive photoresists with image reversal. In this work another approach is explored and it consists in

using a bi-layer photoresist by combining the *LOR B* resist with standard positive phototresist.

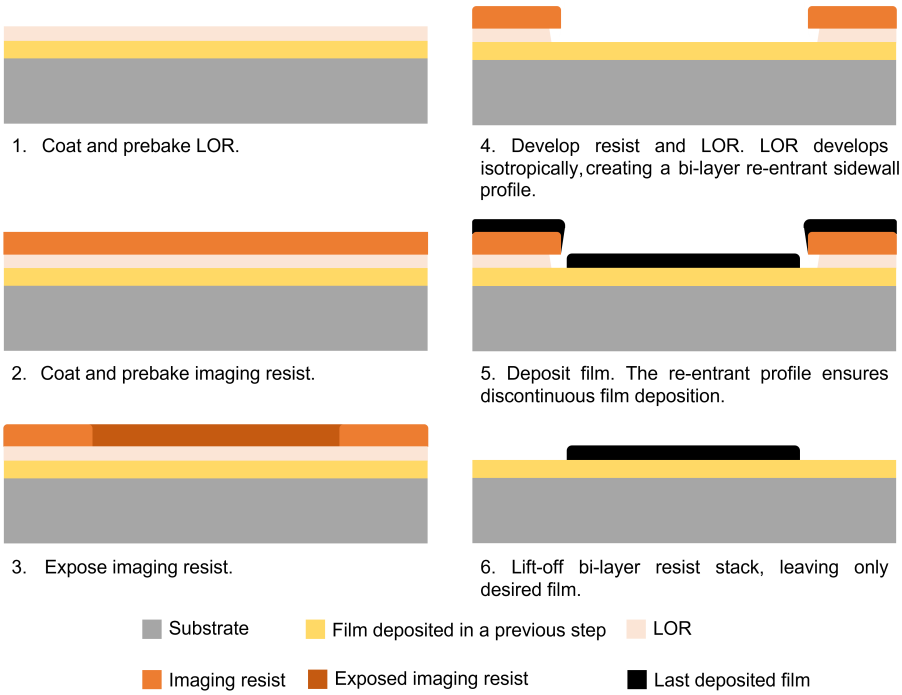


Figure 2.10: Schematic representation of lift-off process using bi-layer photoresist obtained from the combination of *LOR B* resist with standard positive phototresist. Redrawn from [29].

The main steps followed during the bi-layer resist approach are depicted in Figure 2.10, where it is possible to notice that no adhesion promoter is required for the already great adhesion of LOR resist on the most used substrates. Moreover, it is possible to observe the finely defined re-entrant sidewall profile that is obtained with this technique. The bi-layer approach hence inherits the simplicity of a standard one-resist process, even if an extreme control on the resist prebake is required to fix the development and undercut rate [29].

2.2.4 Wet and Dry Etching

Subtractive technological processes are based on the selective removal of deposited material from the sample, which is achieved using etching techniques. The speed at which the material is removed is called etch rate, and it depends on the composition

of the etchant as well as the material properties, such as its density [22].
Etching techniques can be classified in wet etching and dry etching. Wet etching is based on immersing the sample into a solution containing the etchant chemistry and hence removing the material by mostly chemical reactions. A common wet etching process used in this work is represented in Figure 2.11 and it consists in the removal of native SiO_2 which grows on both sides of a silicon (Si) substrate by using hydrofluoric acid (HF), with consequent generation of H_2SiF_6 as water-soluble byproduct [22]:

$$SiO_2 + 6HF \longrightarrow H_2SiF_6 + H_2O$$

In order to keep the etchant maximum strength over use and time, buffering agents are often added to the etch solutions. In particular, ammonium fluorine (NH_4F) is added to HF to help preventing the depletion of fluoride ions in the etchant and to reduce the etching of the photoresist, if this is used as masking layer during the etching process [22]. In this case, the etchant solution takes the name of "Buffered HF" or BHF, or more commonly "Buffered Oxide Etch" or BOE.

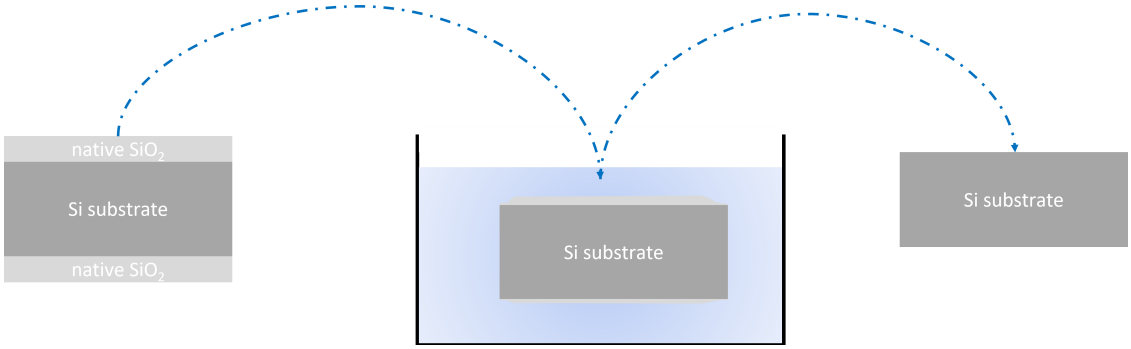


Figure 2.11: Schematic representation of the wet etching of native SiO_2 by HF.

An important parameter in the etching process is the selectivity, which is defined as the ratio of the etch rates of the different materials in an etch process [22]. Indeed, the ideal situation is to have a very fast etching of the layer that has to be removed compared to the masking layer and the layers underneath. In general, wet etchants are very selective because the process relies on chemical reactions. On the other hand, relying on a purely chemical processes, wet etching results in an isotropical removal of material, hence occurring even underneath the masking layer. Usually, a significant undercut called "etch bias" is obtained in wet etching, as it is possible to observe from Figure 2.12(a).
The directionality of the etching process is improved using dry etching, also called plasma etching since it involves the use of gas-phase etchants in the form of a plasma. In this case, the etching takes place as a combination of both chemical and physical processes. In particular, the more physical is the process and the less

lateral etching is obtained, thus resulting in extremely vertical profiles, like it is depicted in Figure 2.12(c). However, the more physical it is the process and the less selective it is, hence the tradeoff between directionality and selectivity should be evaluated accordingly to the processing requirements. Nevertheless, in plasma etch system the chemical and the physical components can work in synergy to obtain directionality while maintaining an acceptable degree of selectivity [22].

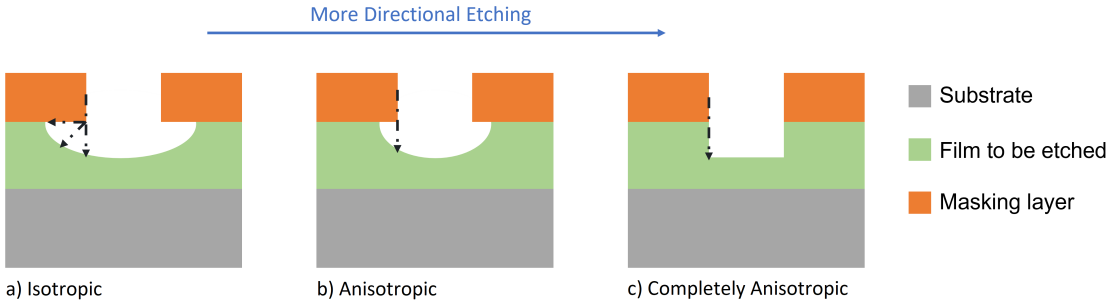


Figure 2.12: Schematic representation of the etching directionality in the case of isotropic (a), anisotropic (b) and completely anisotropic (c) etching. Redrawn from [22].

The plasma etch system is very similar to an RF sputtering one shown in Figure 2.1(b) but etchant species are introduced in gaseous phase in the chamber. The plasma is generated by applying an high electric field between the two electrodes of the system thus resulting in the generation of ionized molecules, ionized fragments of broken-up molecules, free electrons and free radicals [22]. The density of the plasma and the ion energy increase by increasing the RF power.
The chemical component in plasma etch process is due to free radicals, which are electrically neutral species that are extremely reactive because of their incomplete bonding. On the other hand, the physical component is due to ions which are accelerated in a very directional way by the generated electric field between the two electrodes. When the reactive neutral species and the ions act in synergistic fashion, the process is called ion-enhanced etching [22].
There exist different types of etching systems and here only the three used in this work are presented. The simplest plasma system is the barrel etcher, in which a plasma is generated by RF power, but it is kept away from the sample by the use of a perforated metal shield between the sample and the electrodes [22]. This separation results in a selective but purely isotropical chemical etching. Moreover, since the wafer is not sitting to any of the electrodes, non-uniformities are usually observed in the etching. For these reasons, barrel etcher has been used only for photoresist removal using an oxygen plasma. This process is more commonly called photoresist ashing. In this work the barrel etcher used is the *GIGAbatch 310 M* from *PVA TePla*.

A more anisotropical and uniform plasma etching is obtained using an etching system in parallel plate configuration. In this case the top electrode is electrically connected to the chamber walls and it is grounded, while the RF power is supplied to the bottom one. This kind of electrical connections allow to increase the size of the top electrode, resulting in an ion bombarding enhancement of the sample sitting on the bottom one [22]. For this reason, it is the configuration used in Reactive Ion Etching (RIE) and Inductively Coupled Plasma (ICP) etching, whose reactor is depicted in Figure 2.13.

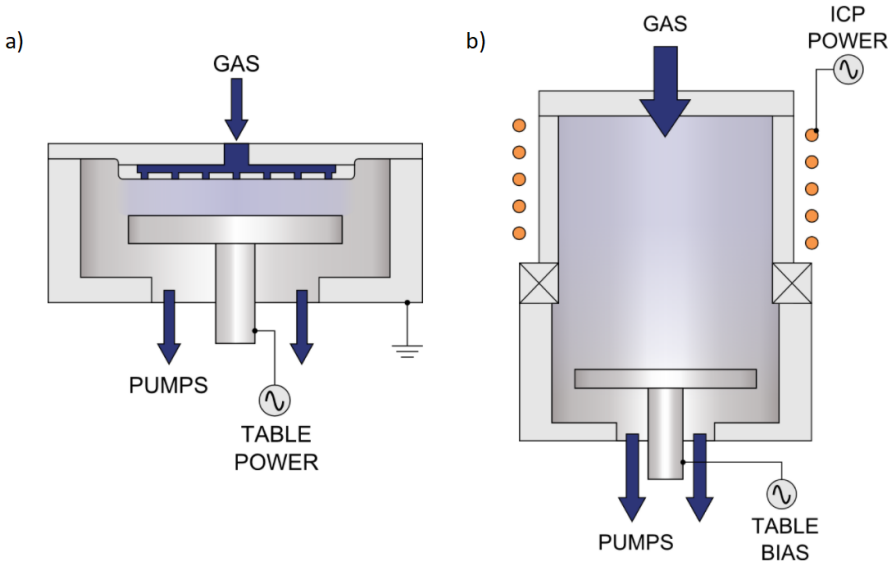


Figure 2.13: Schematic representation of RIE (a) [30] and ICP (b) [31] tool.

In RIE, both ions and reactive species are involved in the etching process and the system usually works at a pressure between 10 and 100 mtorr in order to achieve a good directionality and plasma density. However, the enhancement of the physical component could result in charging effects as well as lattice damage [22]. The RIE tool used in this work is the *PlasmaPro NPG 80* from *Oxford Instruments*.
The ICP etching is instead based on an high density plasma whose generation is decoupled from the ion acceleration. Indeed, Figure 2.13(b) shows as the RF power provided to the bottom electrode to accelerate the ions is separated from the inductively coupled one used for the plasma generation, thus allowing a larger process versatility. Moreover, the plasma has a higher density than the one produced by the RIE system, thus lower pressures can be exploited for having excellent profile control while keeping the etch rates and the selectivity high as well as ensuring a low damage of the sample surface [22]. The ICP tool used in this work is the *Plasmalab 100* from *Oxford Instruments*.

Chapter 3

Characterization methods

3.1 Material characterization

This section is intended to provide a theoretical overview on the characterization techniques that are exploited for understanding the material properties of the developed conductive tantalum oxide. Structural properties such as crystalline phase, thickness and density are investigated through techniques based on X-rays. The resistivity is then computed from the Circular Transfer Length Method (CTLM).

3.1.1 X-ray Diffraction (XRD)

X-ray diffraction (XRD) is a non-destructive technique used for the structural analysis of materials. In particular, this characterization technique is used to identify the crystallographic phase of the sample, as well as the crystallinity fraction if the material is polycrystalline and the average grain size.
XRD relies on the the diffraction of the X-ray electromagnetic waves by the crystal. The sample is indeed hit by a monochromatic X-ray beam at a certain incidence angle with respect to the sample surface. As a consequence, the atoms in the crystal are excited and, relaxing, they give rise to re-emitted radiation that does not excite other atoms, accordingly to the kinematic approximation. The re-emitted radiation can be approximated as constituted by spherical radiation waves but which can be considered plane waves at the detector, assuming a very large distance from the sample. Thus an interference pattern is generated, with a constructive interference occuring in only few directions due to the periodicity of the atoms in the crystal. The condition for constructive interference can be expressed by the Bragg's law:

$$2 \cdot d \cdot sin(\theta) = n \cdot \lambda \tag{3.1}$$

where d is the relative distance between two crystallographic planes which belongs to the same set of equivalent lattice planes (identified by the Miller index ($h\ k\ l$));

θ is the angle of incidence of the X-rays with respect to the crystal planes; λ is the wavelength of the X-ray radiation and n is an integer number representing the order of diffraction.
A graphical representation of the Bragg's law in the case of two parallel X-rays hitting the sample is depicted in Figure 3.1.

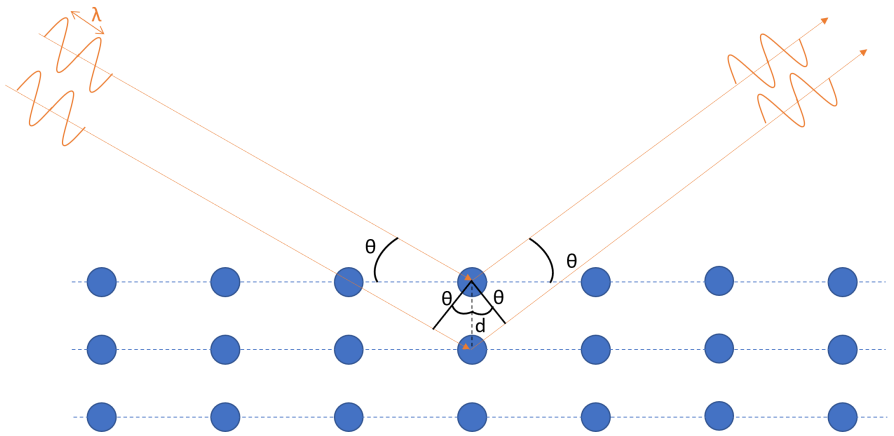


Figure 3.1: Schematic representation of the X-ray diffraction process in the direct space.

The energy of the X-ray radiation that hits the sample is usually about 5-20 keV, since the wavelength of the radiation λ must be of the same order of magnitude of the interatomic distance, usually around few angstroms.
Alternatively, the constructive interference can be expressed in the reciprocal lattice space. In particular, given the primitive vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$ of the Bravais lattice of the crystal in the direct space, the reciprocal space is defined by the primitive vectors $\vec{a^*}$, $\vec{b^*}$ and $\vec{c^*}$:

$$\vec{a^*} = 2\pi \frac{\vec{b} \times \vec{c}}{\vec{a} \cdot (\vec{b} \times \vec{c})}, \vec{b^*} = 2\pi \frac{\vec{c} \times \vec{a}}{\vec{a} \cdot (\vec{b} \times \vec{c})}, \vec{c^*} = 2\pi \frac{\vec{a} \times \vec{b}}{\vec{a} \cdot (\vec{b} \times \vec{c})} \tag{3.2}$$

Hence, considering the Miller index $(h\ k\ l)$ of a family of crystallographic planes, the reciprocal lattice is defined as the space of all the vectors $\vec{G}$ satisfying the relationship:

$$\vec{G} = \vec{a^*} \cdot h + \vec{b^*} \cdot k + \vec{c^*} \cdot l \tag{3.3}$$

The constructive interference condition is expressed by means of the Laue condition:

$$\vec{k'} - \vec{k} = \vec{G} \tag{3.4}$$

where $\vec{k'}$, $\vec{k}$ are respectively the wavevectors of the incident and scattered X-ray, and $\vec{G}$ is the reciprocal lattice vector.

The Laue condition can be visualized geometrically using the Ewald sphere, obtained rotating the wavevector $\vec{k}$ in the reciprocal space, as it is shown in Figure 3.2.

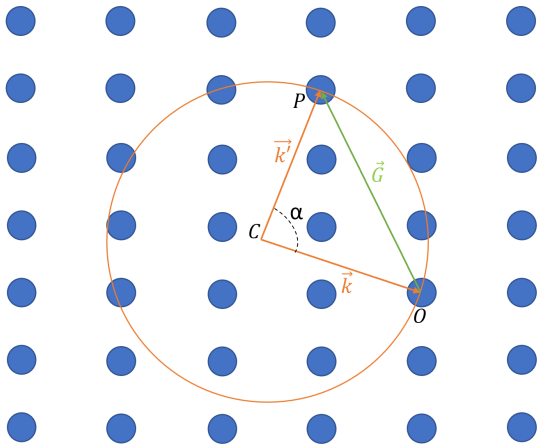


Figure 3.2: Schematic representation of the X-ray diffraction process in the reciprocal space.

As it is possible to observe from Figure 3.2, each reciprocal lattice point P that touches the sphere identifies a wavevector $\vec{k'}$ for which the Laue condition is satisfied. In other words, the condition of constructive interference is satisfied only for those angles α between the incident and scattered beam for which the Laue condition is true.
The intensity of the diffracted radiation is collected by a detector and it is analysed at different detection angles. However, multiple geometrical configurations can be used, accordingly to the information that has to be extracted and the kind of sample considered. The most common one is the Bragg-Brentano configuration, also known as θ-2θ configuration. This configuration is also called symmetric configuration because the angle of incidence ω and that of detection θ are always kept equal, as it is depicted in figure Figure 3.3(a). In this case, the deflection angle of the incident X-ray beam is 2θ. In the Bragg-Brentano configuration, the constructive interference is obtained only for set of equivalent planes parallel to the sample surface. A very interesting configuration is also the Grazing Incidence X-ray Diffraction (GIXRD) one, which is also called Thin-Film configuration since it allows to have a larger contribution in the signal from the crystallographic planes closer to the sample surface. In particular, this is achieved by keeping an angle of incidence ω fixed and at very low angles during the whole ω-2θ scan, as it is shown in Figure 3.3(b). Obviously, the incidence angle must be larger than the angle that leads to total reflection and the detection angle is varied keeping it always larger than the one of incidence. The result is that only crystal planes which are not

parallel to the sample surface give a contribution to the diffraction signal.

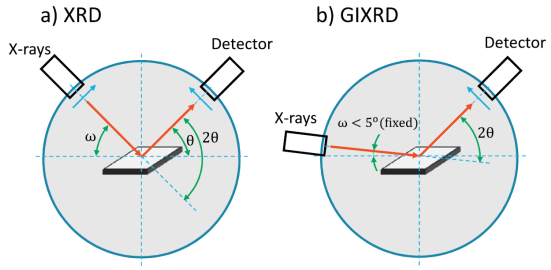


Figure 3.3: Schematic representation of the XRD(a) and GIXRD(b) measurement configuration [32].

As already stated, the contribution to the signal at the detector comes mainly from the crystallographic planes closer to the sample surface. In particular, considering a thin-film sample deposited on a silicon substrate hit by an X-ray radiation at an angle ω, as depicted in figure Figure 3.4, the depth of material contributing to the signal is given by:

$$d = \delta \cdot sin(\omega) \tag{3.5}$$

where δ is the penetration depth of the X-rays in the material, which can be defined as the distance travelled by the X-ray beam in the sample before seeing its intensity reduced by a factor $1/e$ [32].

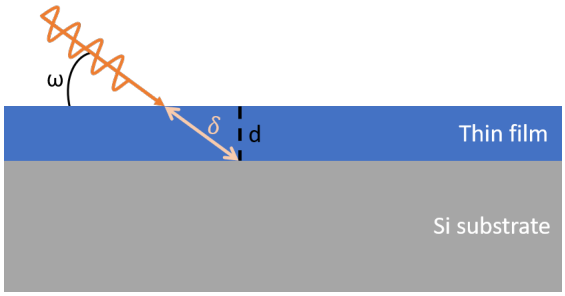


Figure 3.4: Schematic representation of the X-ray penetration in a thin-film sample during a GIXRD measurement.

The result of the XRD or GIXRD measurement is a diffraction pattern showing the intensity of the detected beam as a function of the angle of detection θ, or 2θ. The presence of a peak in the pattern is related to a particular crystallographic plane present in the sample, whose identification is reached through the comparison with the angular positions of diffraction peaks reported in databases. Examples of diffraction patterns are provided in Figure 3.5, from which the crystallographic

phase of the sample can be determined.

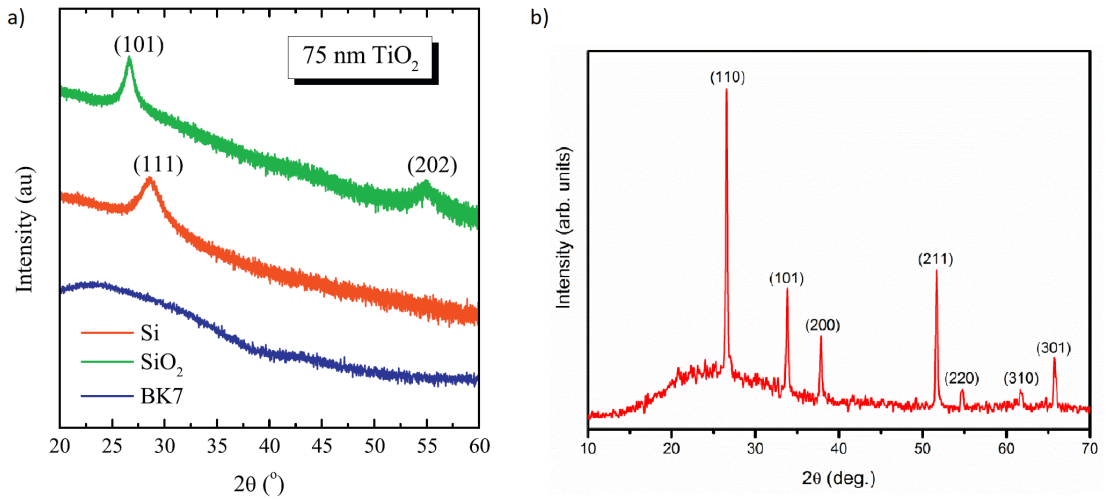


Figure 3.5: a) GIXRD pattern of a 75 nm titanium oxide sample deposited on various substrates [32]. b) XRD pattern of tin dioxide (SnO_2) thin-film deposited at 420 °C [33].

In Figure 3.5(a), the blue curve is typical of an amorphous sample, whereas the green and red ones identify monocrystalline samples. On the other hand, Figure 3.5(b) shows an XRD pattern that identifies a polycrystalline sample since peaks that belongs to different crystallographic orientations are observed. Moreover, Figure 3.5(a) shows that with XRD and GIXRD it is possible to observe the influence of the material substrate on the layer grown on top.
From the diffraction peak width the average grain size D can be also determined by means of the Sherrer formula:

$$D = \frac{0.9 \cdot \lambda}{\Delta \cdot 2\theta \cdot cos(\theta)} \tag{3.6}$$

where λ is the wavelength of the incident X-ray beam; Δ is the full width at half maximum (FWTH) of the peak and θ is the diffraction angle.
The diffractometer used is a *Bruker D8 Discover* with a $Cu\ K_\alpha$ radiation source at a wavelength λ=1.54184 Å and a scintillator as detector. All the measurements were performed with operating parameters of 50 kV and 50 mA at the X-ray generator. Once the XRD data are acquired, the background noise is removed with the software *DIFFRAC.EVA V5.2* and then a deconvolution of the XRD profile in the characteristic peaks is necessary. This is performed with *Origin 2018* software which allows the deconvolution of the XRD profile into Gaussian functions accordingly to the theoretical sample curve and formula of the Gaussian fit reported in

Figure 3.6. Finally, the peaks are indexed using the open-access databse *Materials Project* [34].

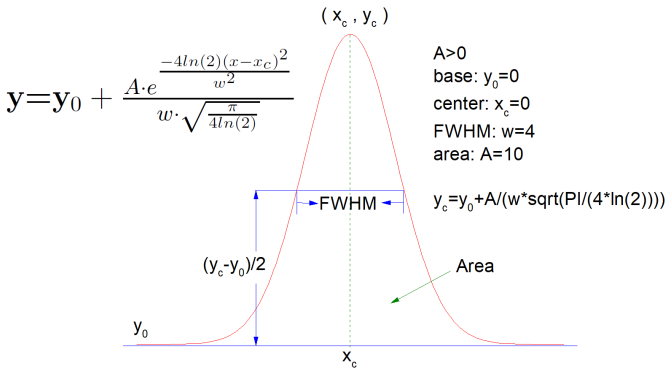


Figure 3.6: Sample curve and formula of the Gaussian fit used in the Origin software.

3.1.2 X-ray Reflectivity (XRR)

X-ray reflectivity (XRR) is a non-destructive technique for investigating structural parameters such as density, thickness and roughness of a multilayer film.
The diffractometer is operating in the symmetric θ-2θ configuration and the measurement is based on an incident X-ray at a grazing angle θ. The X-ray beam is partially reflected off the sample and partially transmitted, as it is shown in Figure 3.7(b). Since the transmitted electromagnetic waves undergo reflection at each interface in the sample, the reflectivity signal is constituted by oscillations, called Kiessing Fringes, which result from the interference of X-rays reflected at the different interfaces between the layers in the sample. Finally, the reflected X-ray beam is collected by a detector and its intensity is measured.

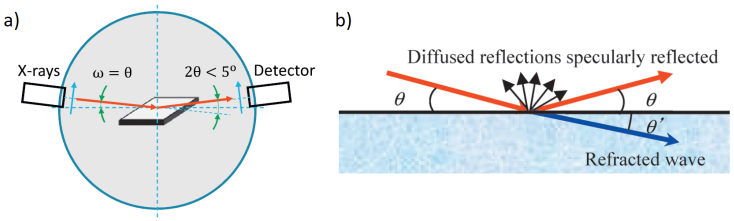


Figure 3.7: a) Schematic representation of the XRR measurement configuration [32]. b) Refraction and reflection of an incident X-ray beam on the material surface [35].

Total reflection is obtained when the X-ray beam is incident on the sample surface

at an angle smaller than the critical angle θ_c. This corresponds to the maximum of the X-ray reflectivity signal; instead, the signal decreases rapidly increasing the incidence angle θ above the critical one, as it is illustrated in Figure 3.8.

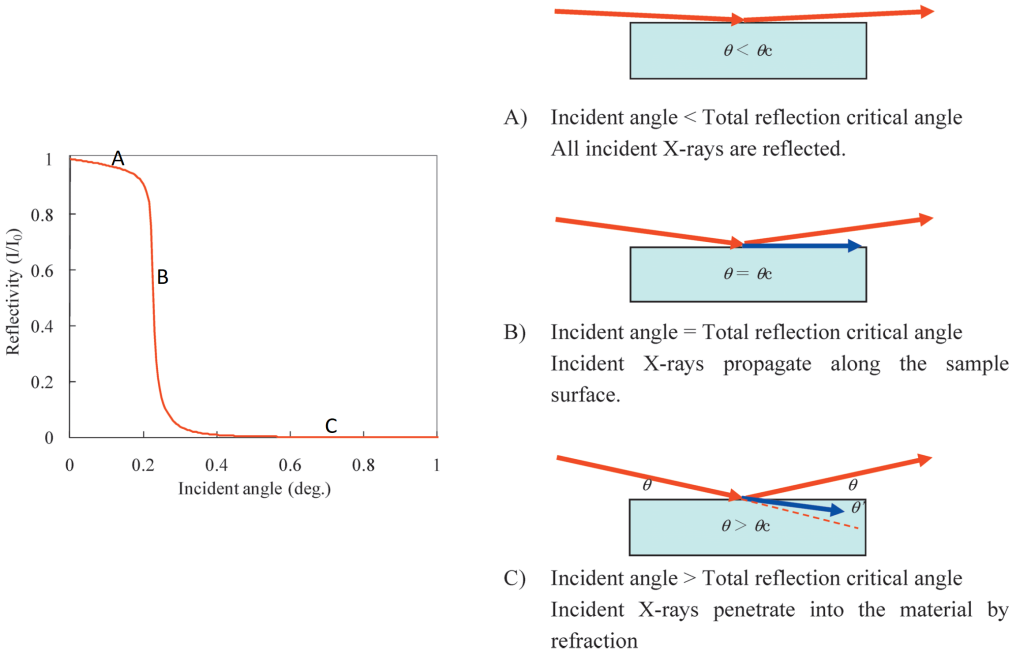


Figure 3.8: Reflectivity curve of a Si sample [35].

The X-ray reflectivity curve is used in order to acquire information on different structural parameters of the sample. In the example provided in Figure 3.9(a), the X-rays reflected from a gold (Au) film surface interfere with those reflected at the interface between Au and the silicon (Si) substrate, resulting in the typical oscillations in the X-ray reflectivity curve. The period of these oscillations is reduced by increasing the thickness of the film.
Figure 3.9(b) describes the effect of the thin-film density on the X-ray reflectivity curve. In particular, it is possible to observe that, considering films having the same thickness, the amplitude of the Kiessing Fringes and the critical angle for total reflection are influenced by the density of the film. The larger it is the density of the thin-film, the greater it results the total angle of reflection. On the other hand, the amplitude of the oscillation is larger, the bigger it is the density difference between the film and the substrate.
The surface roughness is another structural parameter that can be extracted from an X-ray reflectivity curve. In Figure 3.9(c), it is indeed possible to observe that the decay rate of the curve is much faster, the more rough it is the surface.

Information on interface roughness can be also obtained from the curve. In the example provided in Figure 3.9(d), it is clear that the amplitude of the oscillations increases with decreasing interface roughness.

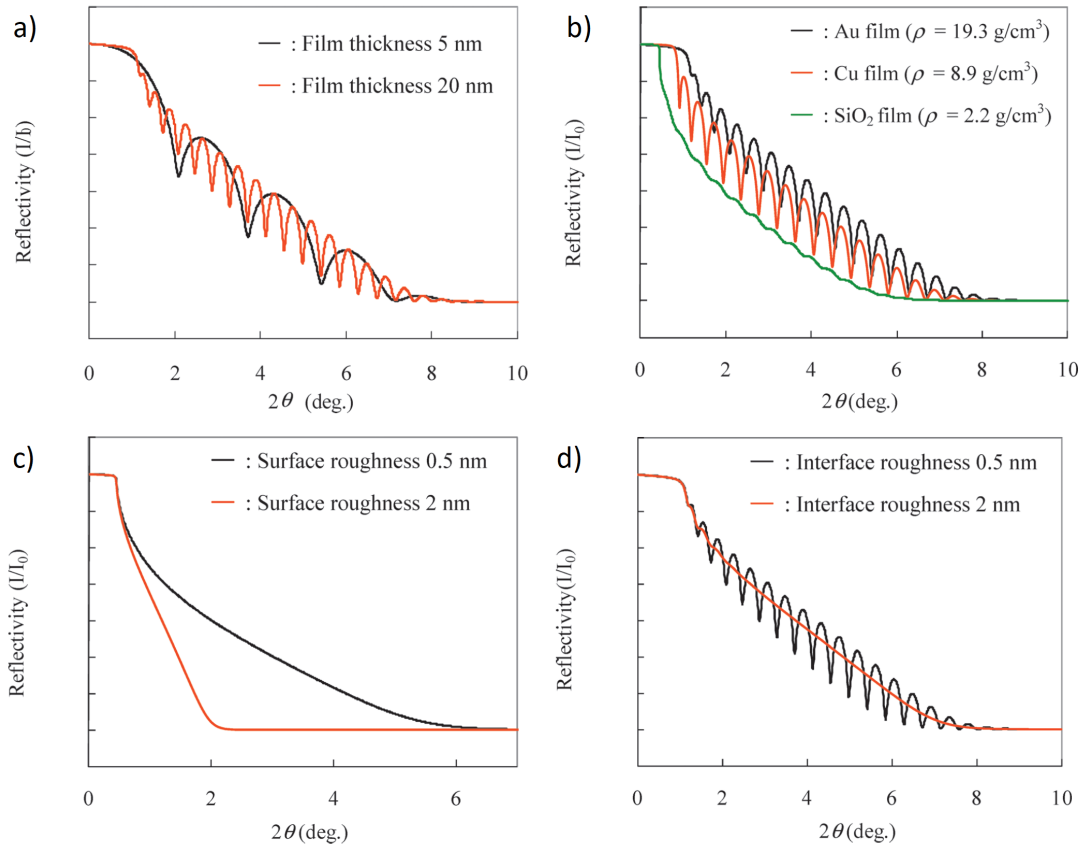


Figure 3.9: a) X-ray reflectivity curve of Au film grown on Si substrate [35].
b) X-ray reflectivity curves of 20 nm Au, Cu and SiO_2 film grown on Si substrate [35].
c) X-ray reflectivity curves of Si substrates with two different values of surface roughness [35].
d) X-ray reflectivity curve of Si substrates with different interface roughness [35].

In order to finely extract all these information on the layer or multilayer structure under investigation, the X-ray reflectivity curve obtained experimentally is compared with a theoretical one. The theoretical curve is built through the software *Leptos 7* on a model of the layered structure of the sample. The optimal structural parameters are hence identified like those which minimizes the residuals between the measured and calculated reflectivity curves.

3.1.3 Circular Transfer Length Method (CTLM)

Linear Transfer Length Method (TLM) is a characterization technique based on two contacts test structures defined on a layer under investigation. It is used to estimate the contact resistance and the specific contact resistance between the metal contact and layer underneath. It exploits the linear relationship between the measured resistance value and the gap spacing of the contacts. However, it is also particularly useful to estimate the sheet resistance (R_{sh}) of the layer under test, from which the resistivity (ρ) of the material can be extracted knowing the thickness t of the layer:

$$\rho = R_{sh} \cdot t \tag{3.7}$$

The technique is very powerful also because of its simplicity, since the test structures can be fabricated with only one lithographic step and lift-off. On the other hand, TLM tends to suffer from the spreading of the current between the contacts due to current crowding [36]. In order to overcome this issue, each test structure is made as a circular contact, separated from an external one by a ring-shaped gap, as it is depicted in Figure 3.10.

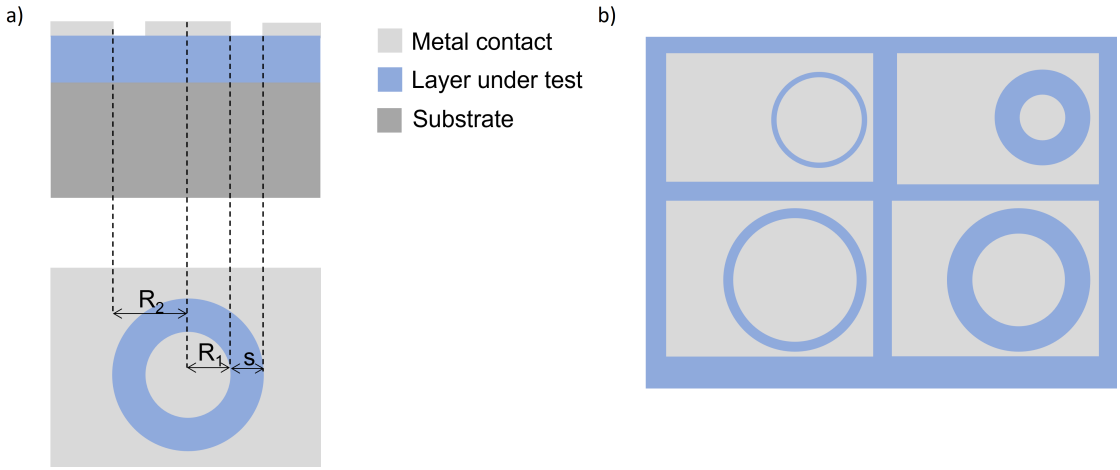


Figure 3.10: a) Cross-section and top view of the CTLM structure. R_1 and R_2 are the radii of the inner and outer disc contact, respectively, and S is the gap spacing, which is variable. b) Top view of four CTLM test structures with different gap spacing. Partially redrawn from [36].

A four-point-probe measurement is performed in order to compensate for losses due to probe contacting [36]. As it is possible to observe in Figure 3.11, a current is forced between the two metal contacts all the way through the intermediate layer, and a voltmeter is used to measure the voltage drop between the two contacts.

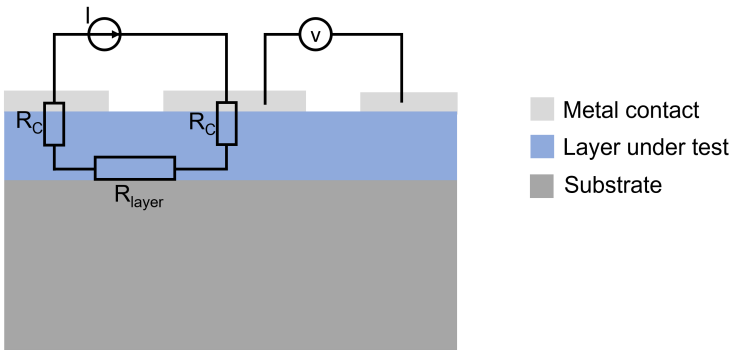


Figure 3.11: Schematic of CTLM four-point-probe measurement principle.

The value of the measured resistance R_T is hence given by the sum of the resistance of the intermediate layer R_{layer} and that of each contact R_C:

$$R_T = 2R_C + R_{layer} \tag{3.8}$$

The resistance of the layer under investigation R_{layer} depends on the gap spacing between the contacts, hence different test structures are measured and the curve of the total resistance R_T is plotted as function of the gap spacing (see Figure 3.12).

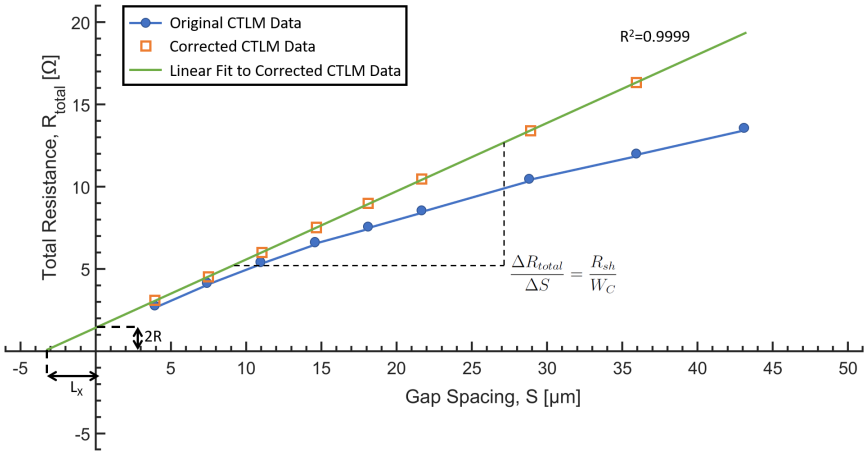


Figure 3.12: Total resistance plotted versus gap spacing before (with circles) and after (with squares) applying the correction factors [36]. Redrawn from [36].

The measured curve is clearly nonlinear, thus correction factors must be applied in order to linearise it and to extract the useful information. In particular, the following parameters can be extracted from the linear fit:

- The contact resistance $2R_C$ is extracted from the intercept with the vertical axis.

- The normalized contact resistance is obtained multiplying the contact resistance by the contact area.
- The transfer length L_T is obtained as half of the intercept with the x-axis L_X. It is expressed in micrometers and the smaller it is, the better the ohmic quality of the contact is [36].
- The sheet resistance R_{sh} of the layer below the metal contacts is obtained by multiplying the slope of the linear fit by the area of the inner circular contact.
- The resistivity ρ of the layer underneath the metal contacts is computed multiplying the sheet resistance R_{sh} by its thickness.
- The specific contact resistance is calculated from the multiplication of the sheet resistance R_{sh} and the transfer length L_T.

The correction factors that allow the linearisation of the fitting curve are derived by the formulation of the resistance between the two circular contacts [37]:

$$R_T = \frac{R_{sh}}{2\pi}[ln\frac{R_2}{R_2 - s} + L_T(\frac{1}{R_2 - s} + \frac{1}{R_2})] \tag{3.9}$$

where R_2 is the radius of the outer circular contact and s is the gap spacing.
Considering a radius of the inner contact R_1 much larger than the gap spacing s and $R_2 = R_1 + s$, the Taylor expansion can be applied:

$$R_T = \frac{R_{sh}}{2\pi R_1}[s + 2L_T] \cdot c \tag{3.10}$$

where the correction factor c is expressed as:

$$c = \frac{R_1}{s} ln(\frac{R_1 + s}{R_1}) \tag{3.11}$$

The correction factors c are hence defined by the sizing parameters of the test structures.
The measurements have been performed using a *Cascade* semi-automated probe station with *B1500A* semiconductor parameter analyzer.

3.2 Electrical characterization

In this section the electrical characterization techniques used for assessing the performances of the developed ReRAM are presented. In this regard, quasi-static measurement and pulsed measurement are described. Finally, Impedance Spectroscopy technique, which allows for a modelling of the ReRAM, and the setup used during this kind of measurements are presented.

3.2.1 Quasi-static Measurement

A quasi-static or direct-current (DC) measurement of a ReRAM device consists of I-V sweeps in which voltage stimuli are applied to the two-terminal device and the current passing through it is measured. The current value measured by a parameter analyzer reflects the change in the resistance state, since a larger current is measured when the resistance of the device is low, whereas a small value is observed if the device is in high resistance state.
Two measurement configurations are considered for the I-V sweeps, and they are shown in Figure 3.13 for a vertical device. These two measurement configurations allow to perform the three main I-V sweeps required for the quasi-static programming of the memory cell: electroforming, reading of the resistance state and switching cycles. In all the three cases the voltage is applied between the TE, which is contacted by a tip moved through a positioner, and the BE, which is connected to ground with the chuck contact that holds the device with a strong vacuum. In each case, the quasi-static voltage sweep is performed by fixing a value for the stop voltage of the sweep and, then, starting from 0 V a voltage level is applied to the device for 100 ms of hold time till reaching the stop value. The number of voltage levels during the sweep has been kept equal to one hundred points per sweep. Moreover, the sweep is bidirectional which means that the stop voltage is reached in one hundred voltage steps and then the same steps are retraced going backwards till 0 V. All these settings are defined in a *LabVIEW* command window and they are applied to the semiconductor parameter analyzer (*Agilent 4155C*) through a General Purpose Interface Bus (GPIB) connection. The *LabVIEW* interface is also used to visualize in real time the measurement.
The main difference between the two configurations is the presence of a 10 kΩ series resistor which is directly soldered on the tip for the electroforming case, as it is shown in Figure 3.13. This resistor allows to limit the voltage drop on the device during the electroforming step, thus preventing it from an hard breakdown condition. The working principle is the same of a current compliance set by the parameter analyzer, but it is much more efficient than the latter because it acts faster being very close to the device under test. A mechanism to limit the voltage drop on the device in this step is fundamental since the maximum voltage applied during electroforming is quite high if compared to the usual voltage levels used for the device programming, reaching even some volts. Obviously, a current compliance can be additionally set by the parameter analyzer if required.
The configuration shown in Figure 3.13 without any resistor soldered on the tip, instead, is used for performing the SET and RESET transitions and the reading of the resistance value of the vertical ReRAM device. In the case of the I-V sweep for programming the devices, the stop voltage is set in order to achieve a certain resistance value, but avoiding to break the device itself. Instead, the reading I-V sweep is always performed up to 0.25 V and the resistance value is read as the

ratio between 0.2 V and the current measured at this voltage. The choice for a stop voltage equal to 0.25 V is due to the fact that it is sufficiently small to do not modify the device state during the reading, but it is also enough to obtain an acceptable current level measurement from the parameter analyzer. To access the endurance of the devices, several cycling steps are also performed. In this case the device is forced to change its state between HRS and LRS, and viceversa, by simply alternating for a defined number of times the RESET and SET programming sweeps.

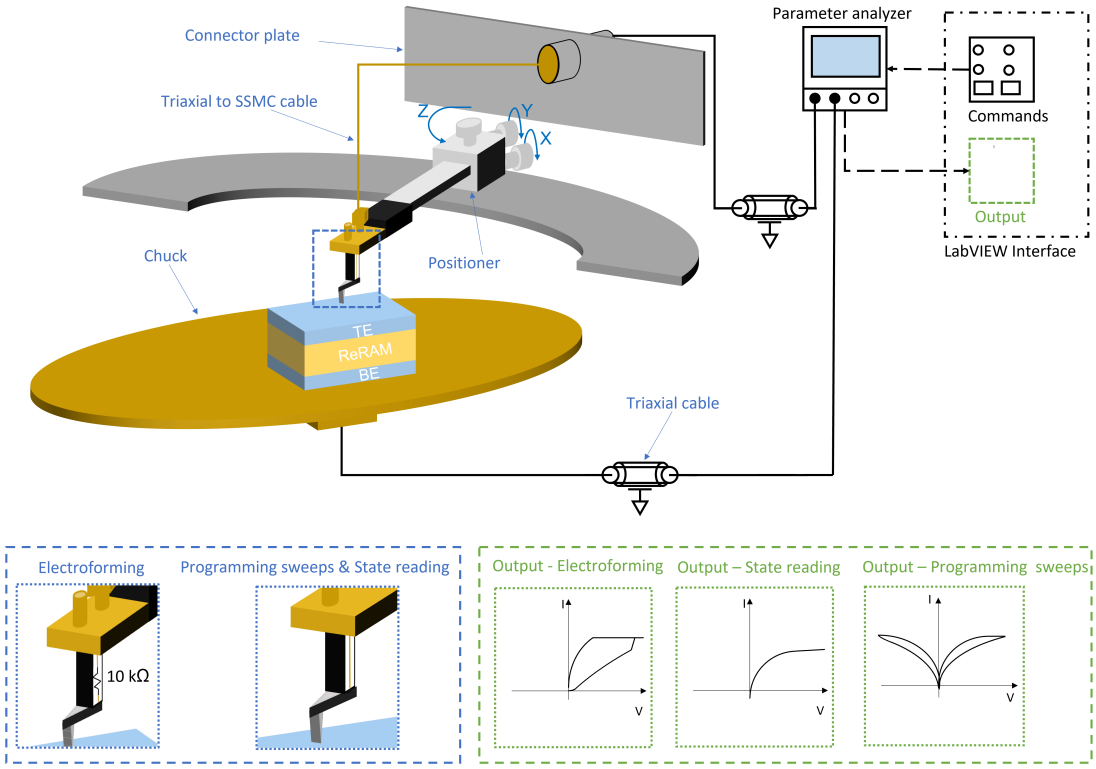


Figure 3.13: Schematic representation of quasi-static electrical characterization configuration in the case of electroforming, reading of the resistance state and switching cycles for a vertical device.

Figure 3.14 shows the two possible setup configurations for quasi-static programming of planar devices. In this case, the chuck is left floating during all the I-V sweeps and it is used just for keeping the chip in place through a strong vacuum. Instead, the device is contacted at both TE and BE with tips. However, because of the current overshoot due to the inductive contribution of the cables, as well as the absence of the parasitic resistive contribution of the chuck, additional protections to the device are required with respect to the case of vertical ones. During

elctroforming this consists in using a tip with a soldered 100 kΩ series resistor for contacting both TE and BE. On the other hand, during standard programming and reading of the state only a 10 kΩ series resistor at the BE is required to guarantee the device protection because of the smaller voltage applied to the device.
For completeness, it is mentioned that for the planar devices measurement a different quasi-automated probe station is employed with respect to the one used for measuring the vertical devices. The parameter analyzer used for the setup depicted in Figure 3.14 is an *Agilent B1500A*.

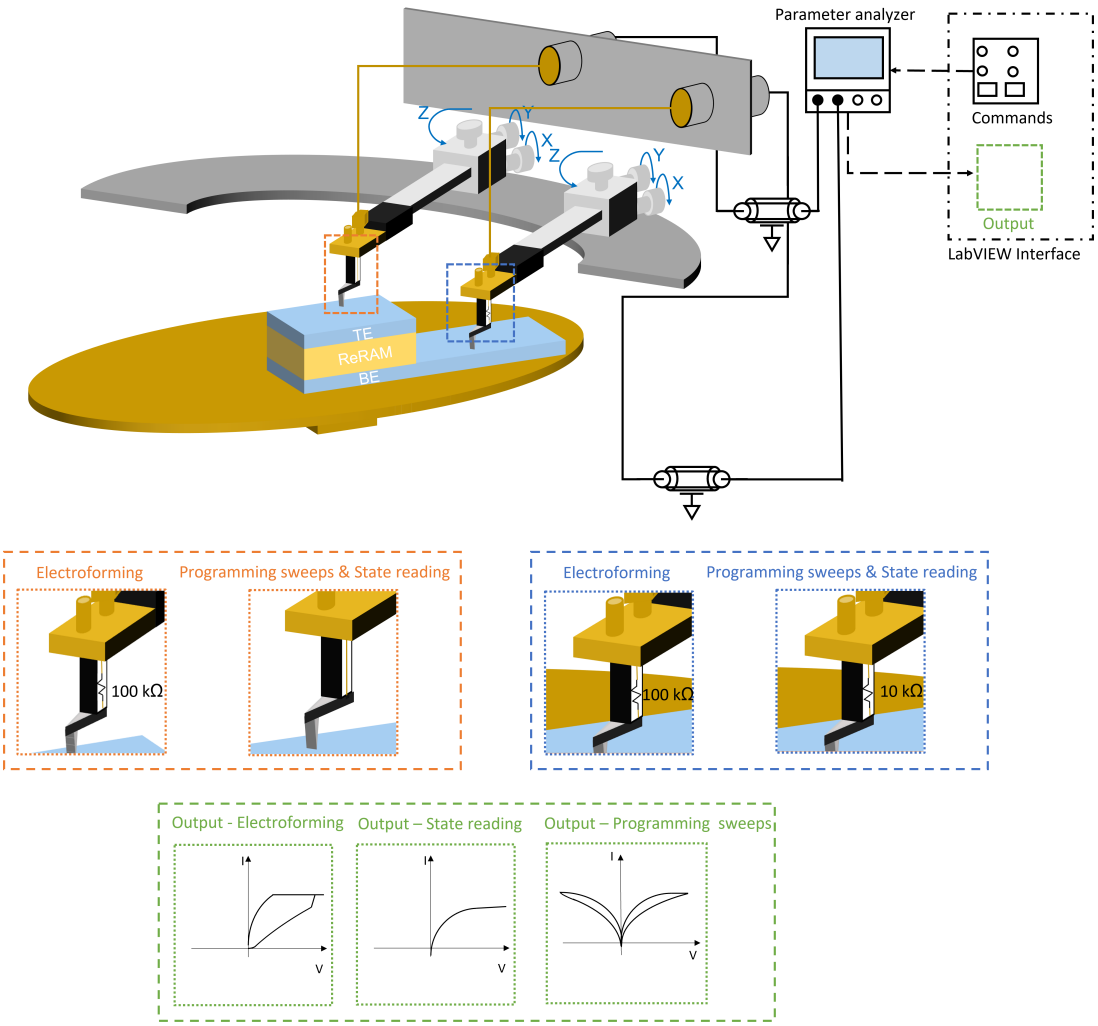


Figure 3.14: Schematic representation of quasi-static electrical characterization configuration in the case of electroforming, reading of the resistance state and switching cycles for a planar device.

3.2.2 Pulsed Measurement

ReRAM devices can emulate the behaviour of biological synapses if they are programmed by sending pulses with a defined amplitude and time characteristics, as already mentioned in Section 1.1. This kind of programming is the objective of pulsed measurements, also called Alternating-Current (AC) measurements. The schematic of the setup used for this kind of electrical characterization technique is shown in Figure 3.15.

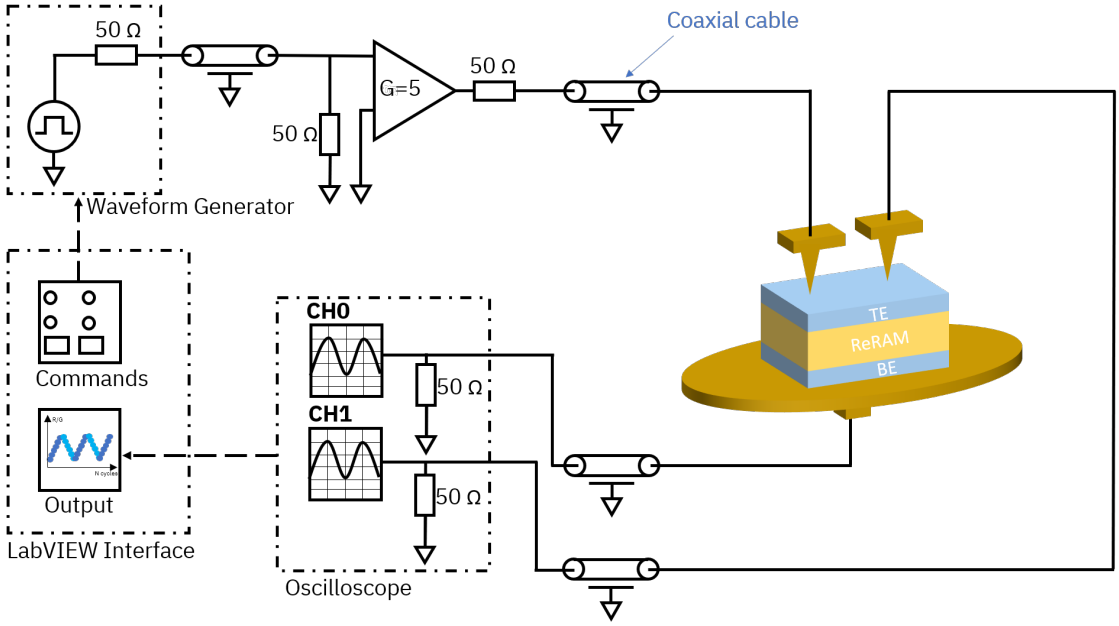


Figure 3.15: Schematic representation of the pulsed measurement characterization setup.

A software interface in *LabVIEW* is used to both send the commands to the arbitrary waveform generator (16-Bit 400 MS/s *NI PXIe-5451*), as well as for displaying the result of the measurement performed with an oscilloscope (400 MHz *NI PXIe-5164*). Coaxial cables (*RG-316*) are used to apply the signal to the device. In particular, the device is contacted from the top by two probe tips while the BE is contacted by the chuck to which it is stuck by a strong vacuum.
Before performing the measurement, in order to check that the tips are well in contact with the device, a signal is applied by the waveform generator and the channel 1 of the oscilloscope is used to read. The read signal should be almost the same as the applied one if the contact is good. The signal applied by the waveform generator is a square wave with a peak to peak amplitude A_{pp} of 200 mV in order to do not modify the device state during this step.
Once the contact is checked, the actual measurement can start by applying squared

voltage pulses at the TE. In particular, Figure 3.16(a) and (b) displays the theoretical shape of the positive and negative voltage pulses, respectively. The parameters that can be controlled in these pulses are the voltage amplitude $V_{amplitude}$ and the time window t_{window}. In the time window the values of the delay time t_{delay} of the pulse start and the duration of the pulse t_{width} can be chosen.
After the application of each pulse, the resistance state of the device must be determined. This is done applying to the device the sequence of four pulses depicted in Figure 3.16(c) at a fixed voltage of ±200 mV and averaging the resistance measurement done through each of them. This read measurement is performed by dividing the voltage drop on the device by the current that flows into it.

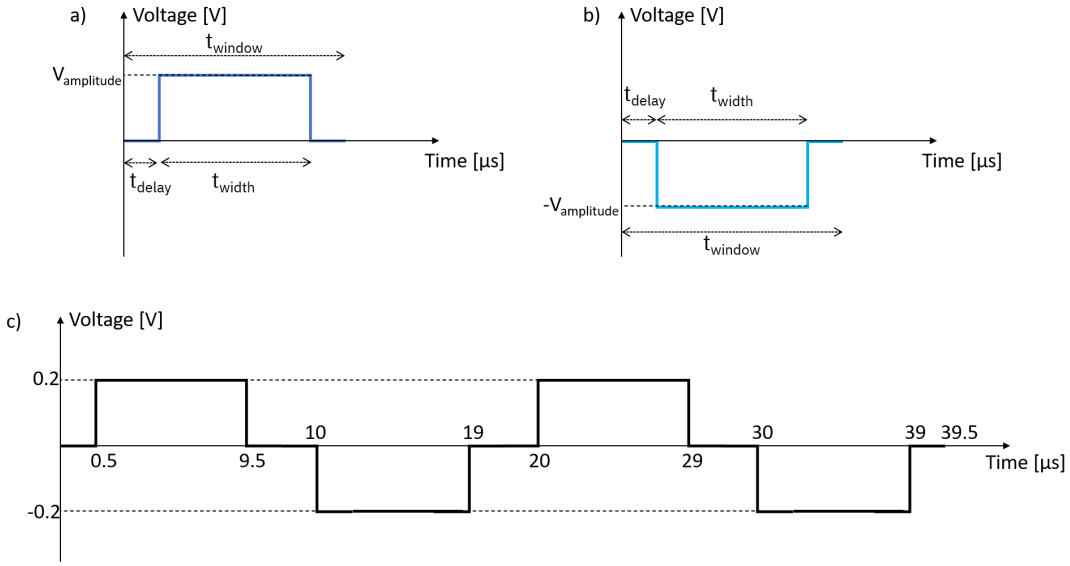


Figure 3.16: Schematic representation of the positive (a) and negative (b) voltage pulses, as well as the reading sequence of a pulsed measurement (c).

3.2.3 Impedance Spectroscopy

The impedance (Z) is an electrical parameter usually used to characterize electrical components. It can be thought as the opposition provided by the electrical component to an AC current with a defined frequency. The impedance is a complex number and thus it can be graphically represented by a vector in the plane of its real (R or Z') and imaginary (X or Z") part. A polar form representation (|Z| and Θ) is hence also possible, as it is shown in Figure 3.17.

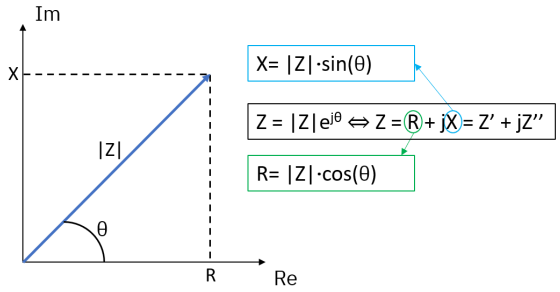


Figure 3.17: Schematic representation and mathematical expressions of the impedance as a vector plane.

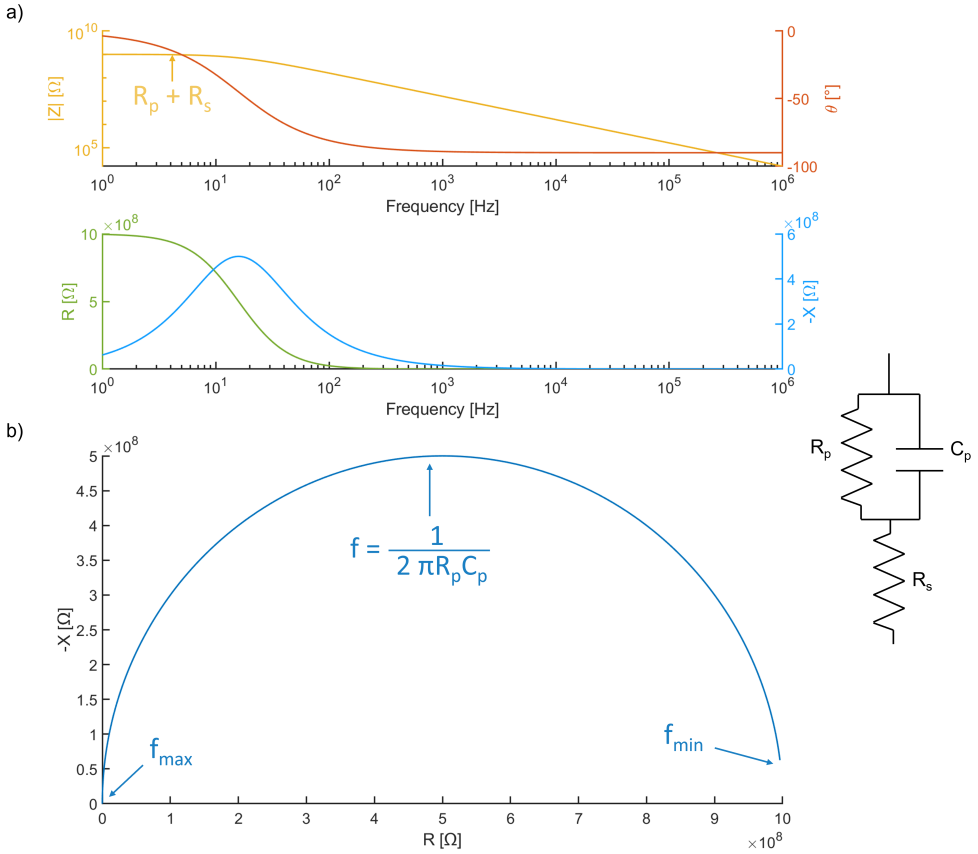


Figure 3.18: Bode plots of the planar and polar components (a) of the impedance of a capacitor C_p which is modelled with the parasitics contributions R_p and R_s. The Nyquist plot of the component is shown in (b).

A standard way to present the impedance data is to use a Bode plot of its polar and planar components (see Figure 3.18(a)) or a Nyquist plot in which the imaginary part is plotted as function of the real one (see Figure 3.18(b)). In the latter,

the larger it is the frequency at which the data are recorded and the closer they are to the origin of the graph. An example of this two representations is provided in Figure 3.18, where a capacitor C_p is modelled with the contribution of both a leakage resistance R_p and a series one R_s.
Based on the concept of complex impedance, Impedance Spectroscopy is a powerful technique used for investigating the electrical properties of materials and devices over a frequency range. Historically, it has been largely applied for studying electrochemical batteries, solar cells and even gas sensors [38][39]. Only recently, the huge potential of the technique has been applied to the study of ReRAM devices [40][42]. Indeed, Impedance Spectroscopy is a powerful as well as simple tool that can be used to measure in a cheap and non-destructive way devices if compared to other techniques, such as Transmission Electron Microscopy (TEM) combined with Energy-Dispersive X-ray Spectroscopy (EDS) [40]. In particular an equivalent electrical circuit (EEC) of the ReRAM device in each different states can be built and the contributions of the different regions of the sample can be decoupled. An example of the EEC of a Ti/HfO_2/Pt ReRAM in LRS and HRS is provided in Figure 3.19.

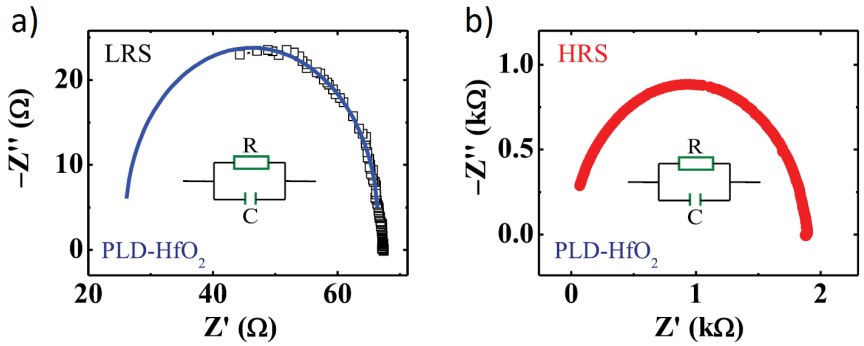


Figure 3.19: Example of the Nyquist plot and EEC of the LRS and HRS in a Ti/HfO_2/Pt ReRAM, in which HfO_2 is grown by Pulsed Laser Deposition (PLD) technique [40].

The Impedance Spectroscopy measurement consists in the application of an AC test signal (to which a DC bias may be superimposed) generated by an oscillator to a device-under-test (DUT) and in measuring the voltage drop on it V_x with a voltmeter, as well as the current that flows into it I_x with an ammeter. The ratio between the two gives the complex impedance data. In this work, a four-terminal configuration is used to get rid of the unwanted parasitics that can have a huge impact on the measurement, especially at high frequency. For this reason, four coaxial connections through *RG-316* cables are defined between the DUT and the tool used for the impedance measurement (a precision LCR meter *HP 4284A* connected to a

LabVIEW interface with GPIB connection):

- H_C: High current terminal;
- H_P: High potential terminal;
- L_C: Low current terminal;
- L_P: Low potential terminal.

As it is possible to notice from Figure 3.20, the outer shield conductors of the coaxial cables allow for a return path from the current signal without generating any inductive magnetic field, thus ensuring a more accurate measurement of both low and high impedances. Moreover, an Auto Balance Bridge configuration is used to compensate for the capacitive contribution of the cables by holding the L_P terminal to virtual ground, since the current that flows into the feedback of the operational amplifier is equal to the input current, which is the one flowing into the DUT.

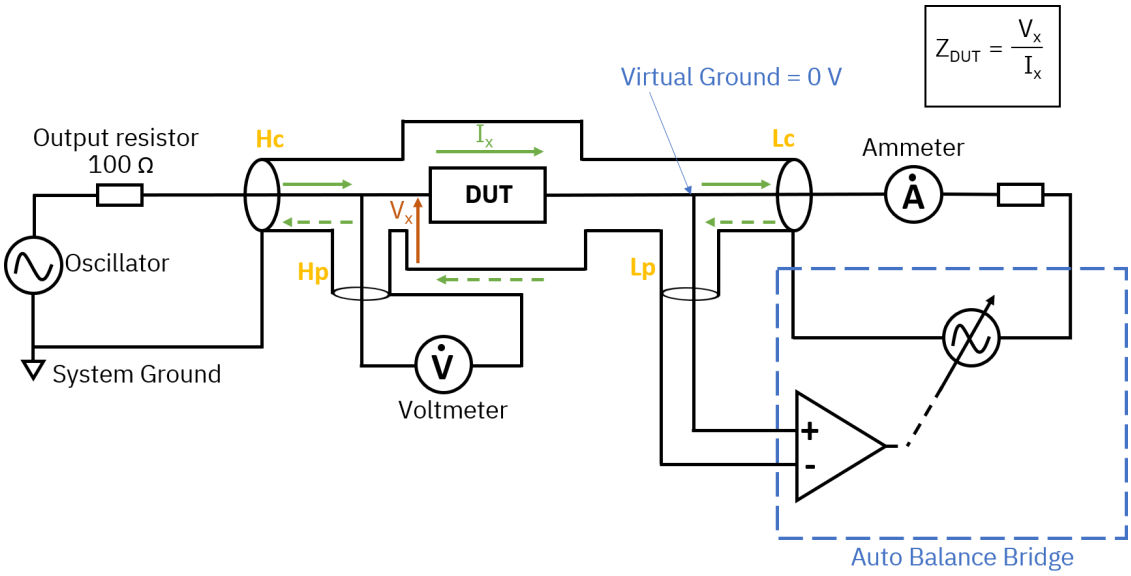


Figure 3.20: Schematic of the four-terminal measurement principle with auto balance bridge. Redrawn from [41].

In order to further improve the measurement accuracy the OPEN and SHORT corrections, as well as a Cable length one, are used to get rid of the contact resistance and of the crosstalk contribution due to the small distance between the tips used to contact the device.
A Printed Circuit Board (PCB) has been designed in order to perform a detailed

study of the setup with the final goal of determining the best device configuration, as well as optimal parameters to measure it. This study is presented in Section B.1. In Section B.2 instead, a deep explanation on how the measurement data of the complex impedance are fit to obtain the EEC model is presented.

Chapter 4

Results and discussion

4.1 Conductive Tantalum Oxide development

This section is dedicated to the development of conductive TaO_x with tunable resistivity using DC reactive magnetron sputtering technique. First, a description of the most critical parameters of the thin-film deposition technique is provided, as well as the strategy adopted to control them in order to obtain a TaO_x with a proper conductivity. A second part of the section is then dedicated to the material characterization. The material properties that have been assessed are the crystalline phase by means of GIXRD, the density and material deposition rate through XRR, as well as the sheet resistance and resistivity of the films obtained with CTLM measurements.

4.1.1 Material deposition

The final goal of the deposition process was to obtain a conductive TaO_x layer with tunable conductivity in order to substitute the metallic scavenging layer in baseline technology. As already mentioned in Section 1.3, this approach allows to further control the oxygen vacancies distribution thus enabling for a more gradual resistive switching behaviour. However, the deposition of a substoichiometric material needs an extreme control over the parameters of the sputtering tool to obtain the desired properties [4]. In this regard, the reproducibility of the proposed deposition process is highly affected by the conditions of the chamber and of the target, with the pressure of the sputter tool that can be used as a knob for the fine tuning of the material resistivity.
The base pressure of the sputtering chamber was kept lower than 5 µbar in order to ensure a good vacuum and to reduce the concentration of impurities in the chamber before the deposition of the layers. Since the TaO_x layer can be deposited starting from a metallic tantalum target, DC reactive magnetron sputtering has been chosen for the deposition. Before depositing the layer, a target cleaning was required

to remove any possible contamination or thin oxide layer from the target surface, since for our tool targets are usually stored in regular ambient and therefore they are exposed to air before being mounted in the chamber. This is a crucial step in order to ensure the reproducibility of the deposition. It consists of two minutes of pre-sputtering and then thirty minutes of sputtering of the target at a DC power of 200 W, a pressure of 6 µbar and a plasma generated with a constant flow rate of 20 sccm of Ar. The pre-sputtering step is performed with a shutter covering the sample and the reason for this step is to let the tool reaching the stable parameters for a deposition, or a cleaning in this case. After this step, the base pressure in the chamber tends to reach smaller values on the order of tenths of µbar, thus confirming the importance of the step to define the clean environment in which the deposition will occur.
All the conductive TaO_x layers were deposited keeping the substrate at room temperature and fixing the value of the DC power to 50 W. This very low value of the power results in small deposition rate of the material because the Ar ions are accelerated with less energy towards the target, sputtering away a smaller number of atoms. On the other hand, it is crucial to give time to the oxygen to react with the sputtered atoms thus allowing for the deposition of tantalum suboxides. Moreover, if the deposition power is increased, the energy of the species arriving to the substrate can be high enough to remove weakly bounded oxygen atoms favouring the deposition of a too conductive sample [43].
During the deposition process, the oxygen to argon flow ratio was kept fixed to a low value, with 1 sccm of O_2 and 40 sccm of Ar respectively. If the concentration of oxygen in the chamber would be too high, the target surface could oxidise leading to the deposition of an insulating material with an associated drop down of the deposition rate. This effect is known as target poisoning and it is accompanied by the sudden drop in discharge current related to the increase in the partial pressure of oxygen [44].
The pressure of the chamber during the deposition is then used as a knob to finely tune the resistivity of the deposited TaO_x. The importance of this parameter is extensively discussed in literature. In [45] it is explained how the deposition of substoichiometric TaO_x films in a controlled manner can only pass through the monitoring of the partial pressure of the gaseous species in the chamber rather than their flow rate. Since the tool used for our experiments does not allow for the control of the partial pressures, but only of the flow rate of the gasses, not all stoichiometries could be obtained within the explored deposition recipes. On the other hand, the conductive TaO_x layers can be obtained considering the range in which the reactive gas flow is proportional to the partial pressure of the reactive gas reported in [45]. In this region, keeping a constant flow rate of O_2, the increase of the total pressure in the chamber results in an increase in the concentration of the reactive gas, and a consequent larger probability of interaction between the sputtered atoms and the reactive species. The material characterization that follows

clearly shows that a more resistive film is obtained by increasing the pressure in the chamber.
The deposition of the TaO_x layers with different resistivity presented in this section were performed all in the same day. Prior all depositions, a target cleaning of five minutes and two minutes of pre-sputtering were performed. Additionally, to further ensure the interdependency of the depositions, between two sputtering processes a target cleaning of only five minutes at the conditions of the first longer one was performed.

4.1.2 GIXRD: crystallographic phase diagram

In this subsection it is shown the impact of the chamber pressure during growth on the crystallographic phase of TaO_x films. Six TaO_x samples have been deposited keeping the DC power (50 W) and the Ar and O_2 flow rates (40 sccm and 1 sccm respectively) constant, but varying the pressure(P_{dep}) between 6 µbar and 11 µbar, in steps of 1 µbar. The deposition time was fixed to the value of 200 s in order to have a comparable nominal thickness of 30 nm.

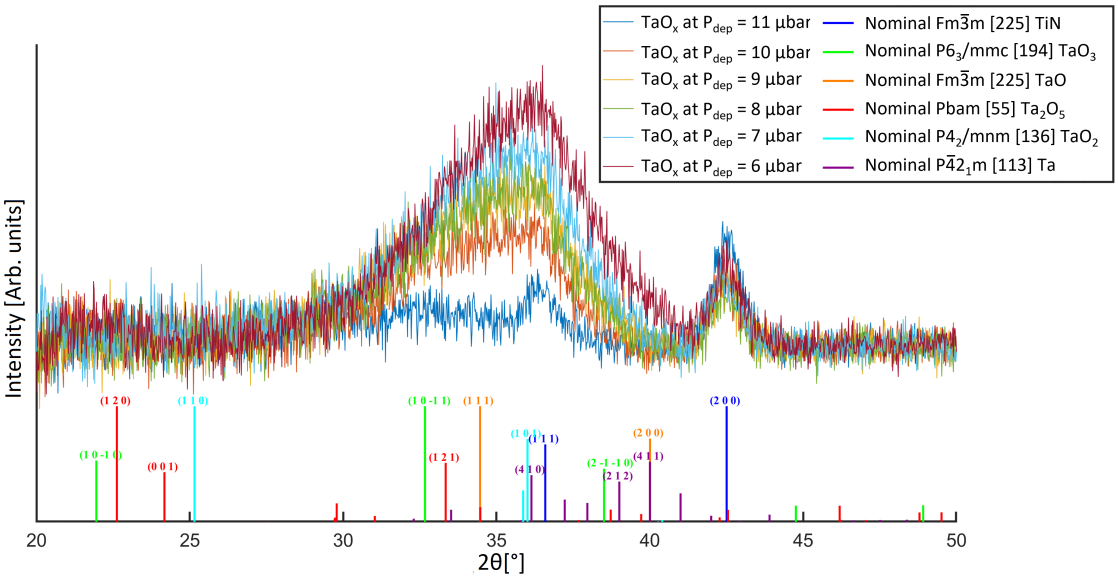


Figure 4.1: Background subtracted GIXRD scans (ω=0.65°) of TaO_x layers deposited at pressures (P_{dep}) from 6 µbar up to 11 µbar on HfO_2/TiN/Si substrates. The nominal peak positions of the various TaO_x stoichiometries, Ta and TiN are reported from the reference database [34], with the correspondent Hermann-Mauguin space group. Only peaks with an intensity larger than 30% are indexed for clarity of the representation.

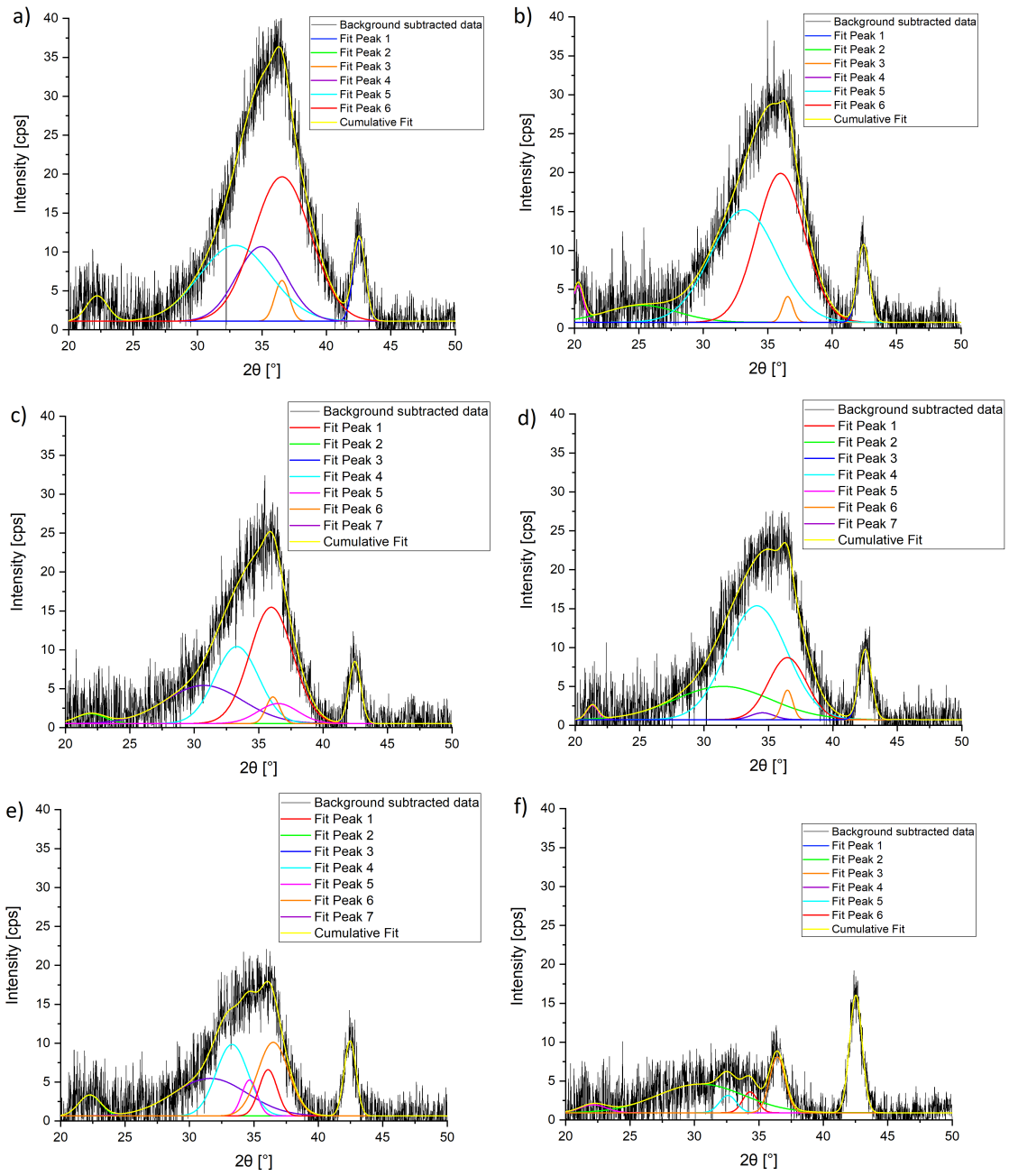


Figure 4.2: Gaussian deconvolution of the GIXRD peaks in the pattern of the TaO_x sample deposited at different pressures (6 µbar (a), 7 µbar (b), 8 µbar (c), 9 µbar (d), 10 µbar (e), 11 µbar (f)) on HfO_2/TiN/Si substrate.

In order to have a material growth as close as possible to the one of the actual layers that will be then embedded into the bilayer ReRAM devices, the deposition

of the TaO_x was performed on 6 nm of HfO_2 and 20 nm of TiN grown by PEALD on top of a standard 500 µm thick Si substrate. The impact of the substrate material during the growth is analysed in Section A.4, where the growth of one of the samples on top of a PECVD deposited SiO_2 is reported.
After the deposition of each layer, an ex situ GIXRD scan ($\omega - 2\Theta$) was performed fixing ω=0.65° and varying 2Θ between 20° and 50° in steps of 0.02°. The time per step was fixed to 10 s. The background subtracted raw data of the GIXRD pattern is shown in Figure 4.1.
Several peaks are detected. The peak at 42.5° is attributed to the (2 0 0) reflection on the TiN layer underneath [34]. Instead, the broader peak between 28° and 41° is attributed to the upper layer of TaO_x, which is crystalline. The typical amorphous bump between 20° and 30° is not detected in the set of samples. The nominal peak positions of the various TaO_x stoichiometries, Ta and TiN, accounting form more than 30% of the intensity, are reported in Figure 4.1 from the reference database [34], with the correspondent Hermann-Mauguin space group.
The broad peak originates from the superposition of multiple contributions that are deconvoluted with Gaussian functions, as shown in Figure 4.2.
Two main contributions are identified and they can be attributed to the (1 2 1) peak of orthorhombic Ta_2O_5 (light blue) and to the (4 1 0) peak of tetragonal Ta (red). The peak deconvolution highlights that the Ta peak dominates in the sample deposited at 6 µbar, but increasing the oxygen content in the chamber, the Ta_2O_5 peak intensity grows till overcoming the Ta one, thus indicating that more oxygen is incorporated in the Ta lattice when sputtering at high pressures, in agreement with a previous study [46]. This result is further corroborated by the reduction of conductivity in the samples grown with increasing chamber pressure, as reported in Section 4.1.4.
The presence of additional peaks to the main ones suggests the coexistance of multiple Ta-O sub-stoichiometries in the deposited layer. Therefore, the precise composition of the TaO_x samples cannot be immediately determined by solely looking to the XRD pattern. A more detailed material analysis through alternative techniques such as X-ray Photoelectron Spectroscopy (XPS) would be required.
The deconvolution of the diffraction peaks shown in Figure 4.2 underlines also the presence of two peaks associated to the (1 1 1) and (2 0 0) planes of cubic TiN phase which is hence sampled during the measurement. No peaks attributed to crystalline HfO_2 are found, confirming the expected amorphous nature of this layer.
The agreement of the showed results with those obtained in the reference [46] is further stressed by the tetragonal β phase of the tantalum sample deposited with the used DC magnetron sputtering tool, which is reported in Section A.1.

4.1.3 XRR: density and deposition rate as function of the chamber pressure

In order to assess the thickness and the density of the deposited TaO_x films, an XRR scan has been performed for each sample after deposition. The scans were performed varying 2Θ in the range 0° to 10° in steps of 0.01°. The integration time in each step was fixed to 1 s. The raw data and the simulated XRR curves are shown in Figure 4.3. The values of the fitting parameters are summarized in Table 4.1 for all the samples.

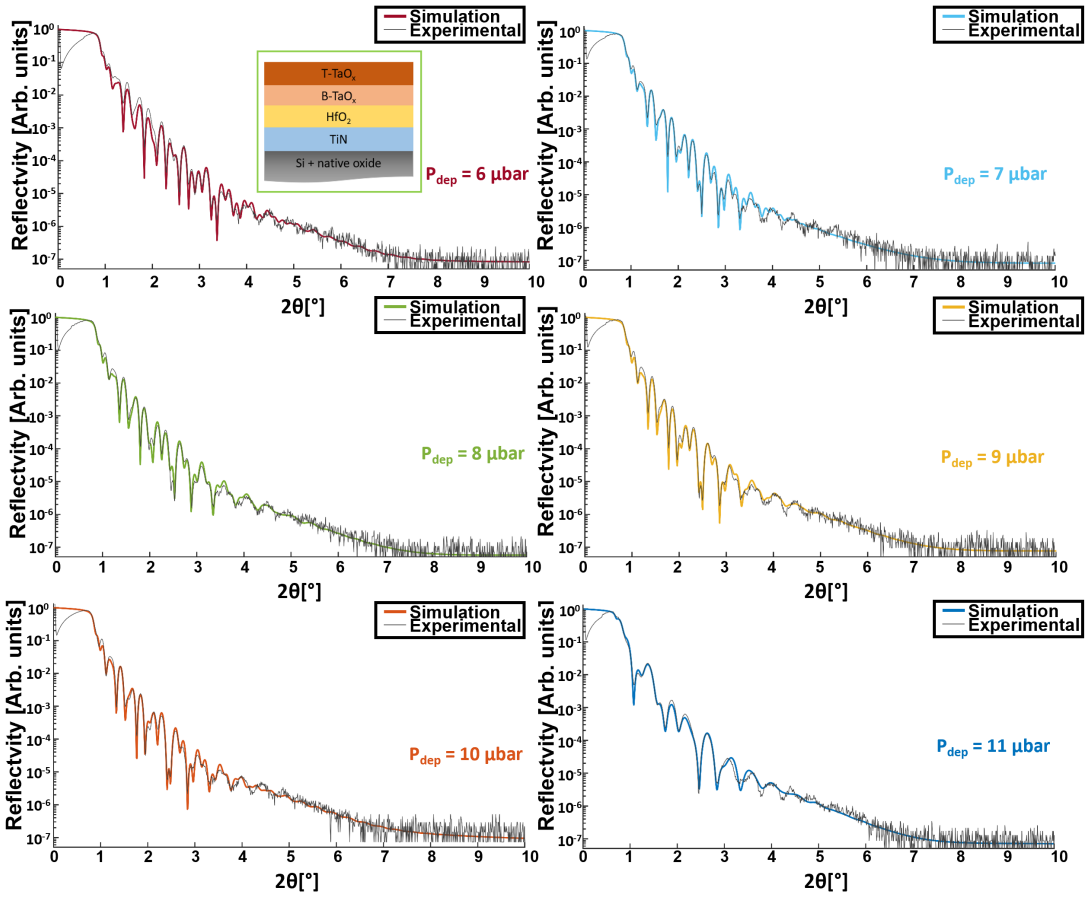


Figure 4.3: XRR experimental patterns (black line) and corresponding simulated curves (coloured line) for the six TaO_x samples deposited at P_{dep} from 6 µbar to 11 µbar. The inset shows the simulated stack.

	P_{dep} = 6 µbar			P_{dep} = 7 µbar			P_{dep} = 8 µbar		
Material	Thickness [nm]	Roughness [nm]	Density [g/cm³]	Thickness [nm]	Roughness [nm]	Density [g/cm³]	Thickness [nm]	Roughness [nm]	Density [g/cm³]
T-TaO_x	25.2	1.0	11.6	26.8	1.1	11.0	26.5	1.1	10.9
B-TaO_x	2.2	1.2	9.3	1.8	1.1	9.3	1.5	1.1	9.3
HfO_2	5.7	0.4	9.0	5.7	0.8	9.0	5.7	0.5	9.0
TiN_xO_Y	1.7	1.0	1.8	1.7	1.0	1.8	1.7	1.1	1.8
TiN	18.8	0.8	5.3	18.8	0.9	5.3	18.8	0.8	5.3
SiO_2	1.6	0.3	2.4	1.3	0.4	2.4	1.5	0.4	2.4

	P_{dep} = 9 µbar			P_{dep} = 10 µbar			P_{dep} = 11 µbar		
Material	Thickness [nm]	Roughness [nm]	Density [g/cm³]	Thickness [nm]	Roughness [nm]	Density [g/cm³]	Thickness [nm]	Roughness [nm]	Density [g/cm³]
T-TaO_x	26.4	1.2	10.7	26.6	1.1	10.7	4.1	1.2	9.9
B-TaO_x	2.2	1.0	9.0	2.3	1.2	8.6	2.2	1.0	8.5
HfO_2	5.7	0.4	9.0	5.7	0.5	9.0	5.7	0.7	9.0
TiN_xO_Y	1.7	1.1	1.8	1.7	0.9	1.8	1.7	1.1	1.8
TiN	18.8	0.8	5.3	18.8	0.9	5.3	18.8	0.8	5.3
SiO_2	1.5	0.3	2.4	1.8	0.3	2.4	1.6	0.4	2.4

Table 4.1: XRR fitting parameters of the TaO_x layers deposited on HfO_2/TiN/Si substrate at P_{dep} from 6 µbar to 11 µbar.

The simulated curves in Figure 4.3 follow very well the experimental reflectivity curves. The reliability of the analysis is also strengthened by the agreement between the measured and nominal thickness of the HfO_2 and TiN layers, which are 6 nm and 20 nm, respectively. Moreover, the density of the TiN agrees with the one of the cubic phase reported in the database [34], while for HfO_2 a good agreement is found with the value reported in literature for the material grown by a similar PEALD process [47].
The simulation of the multilayer stack highlights the formation of a thin layer of about 1.7 nm at the interface between HfO_2 and TiN. This layer is attributed to a partial oxidation of the TiN interface in the ALD tool when the deposition of HfO_2 starts [48]. The presence of this interfacial layer is also observed in the Transmission Electron Microscopy (TEM) image reported in Figure A.4. Moreover, the TaO_x layer cannot be fitted as a single layer but a bottom layer (B-TaO_x) of 2 nm with lower density needs to be added (see Table 4.1), as also confirmed by the TEM micrograph in Figure A.4. The micrograph also shows that the top layer (T-TaO_x) is characterized by columnar grains while the bottom one shows an amorphous nature. The presence of an amorphous onset layer when growing a crystalline one could be explained by the fact that the top layer needs a transition one to accommodate its crystalline growth on an amorphous oxide. This hypothesis is supported by the growth of TaO_x on other amorphous oxides, such as SiO_2, as it is reported

in Section A.4.

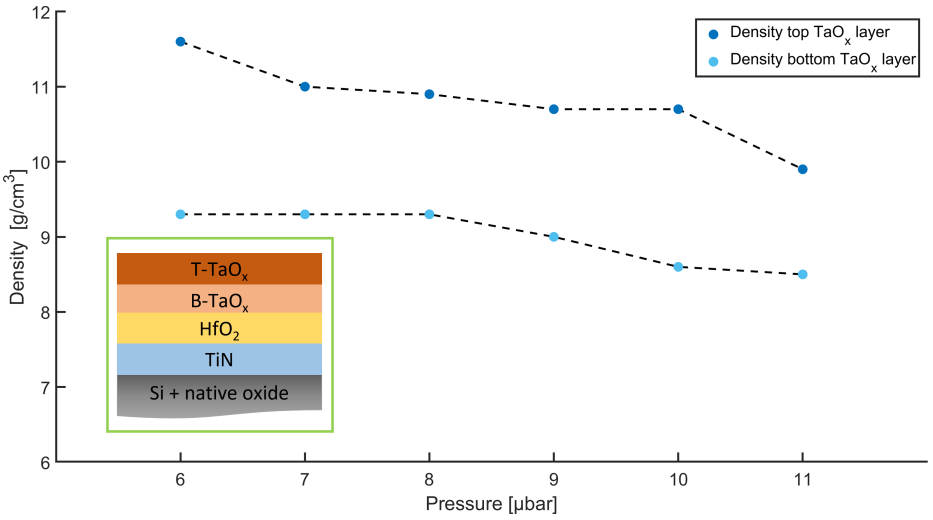


Figure 4.4: Density of both T-TaO$_x$ and B-TaO$_x$ layers as function of the deposition pressure. In the inset the layer stack is shown.

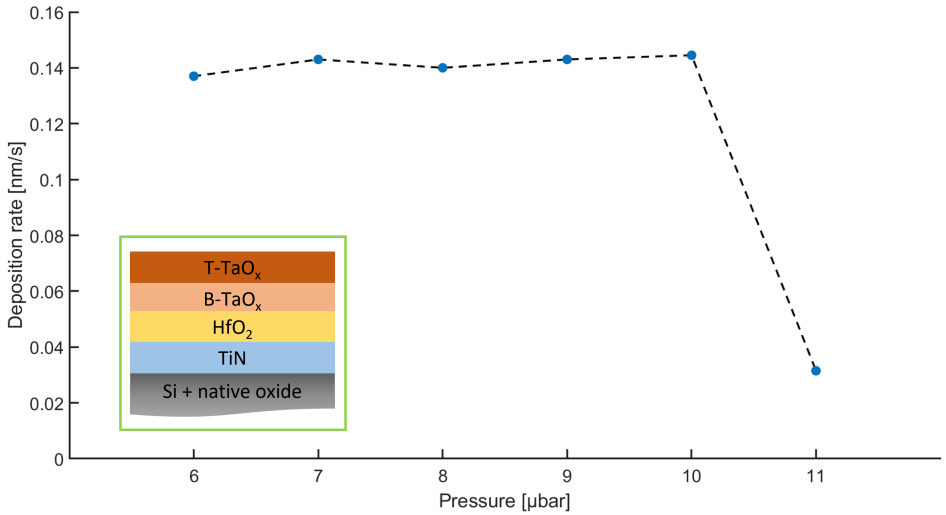


Figure 4.5: Deposition rate as function of the deposition pressure for TaO$_x$ samples. In the inset the layer stack is shown.

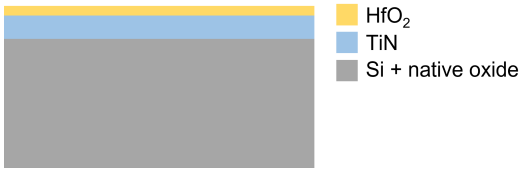
Figure 4.4 shows the evolution of the density of both T-TaO$_x$ and B-TaO$_x$ layers as function of the deposition pressure. In both cases, the density decreases by increasing the deposition pressure. This can be explained considering that increasing

the chamber pressure generally implies also an increase of the partial pressure of the reactive species. Thus, more oxygen is incorporated in the layers during the deposition making the material more oxidized and less dense.
The thickness of the layer is also influenced by the value of the pressure. Supposing a linear deposition rate for the material and considering the thickness of the TaO_x layer as the sum of those of top and bottom layer, the deposition rate as function of the pressure can be represented in Figure 4.5. The hypothesis of the linear deposition rate is supported by the cross-sections in Figure A.4 and Figure A.5, showing the measured thickness of about 20 nm and 30 nm for a deposition time respectively 167 s and 200 s in the TaO_x deposited at 9 µbar.
Figure 4.5 shows that the deposition rate is almost constant at the value of 0.14 nm/s for the samples that are deposited at a pressure up to 10 µbar; whereas at 11 µbar the deposition rate drops drastically down to about 0.03 nm/s. At this higher pressure, a drop in the discharge current could also be observed on the tool and it is due to the too high oxygen partial pressure, which explains the insulating nature of this last sample with respect to all the others (see Section 4.1.4).

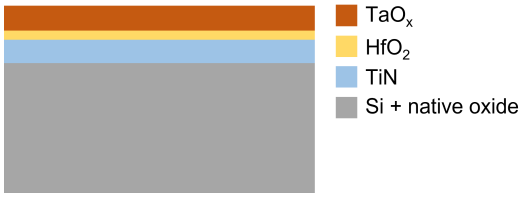
4.1.4 CTLM: sheet resistance and resistivity as function of the chamber pressure

CTLM is a very powerful technique also because of the simplicity in defining the structure of the contacts on the sample, which can be done with only one lithographic step and lift-off, as already mentioned in Section 3.1.3. In particular, the processing of CTLM structures is the following.

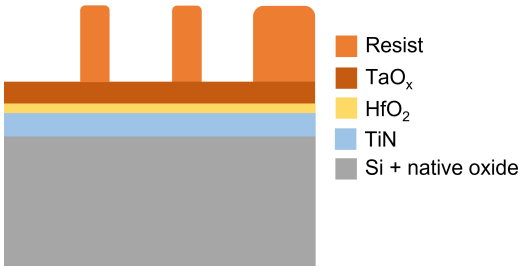
Cross-section of the sample	**Explanation of the processing step**
	The starting substrate is a standard 500 µm thick Si on top of which one or two nanometers of native SiO_2 are usually grown. There is no need to remove this SiO_2 layer since the CTLM structures are contacted from the top.



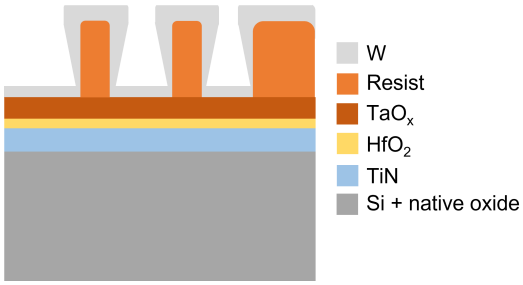
PEALD is used to deposit first 20 nm of TiN and then 6 nm of HfO_2, nominally. In particular, TiN is deposited at 300 °C alternating Tetrakis-(dimethylamido)-titan (TDMAT) as metal precursor and a nitrogen plasma for 260 cycles. After, HfO_2 is deposited at 290 °C using Tetrakis-(ethylmethtlamino)-hafnium (TEMAH) as metal precursor and an oxygen plasma alternated for 50 cycles.



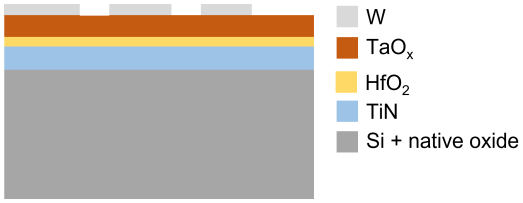
The deposition of the TaO_x layer is done by DC reactive magnetron sputtering as already explained in Section 4.1.1.



1.2 µm of positive polarity resist are spun on the chip after the deposition of HDMS as adhesion promoter. A baking step is then performed at 110 °C for one minute. Then the resist is exposed by contact photolithography for 3.6 s using Hg lamp with an intensity of 11 mW/cm^2. After the exposure, the development is performed dipping the chip for 25 s in a solution containing one part of developer and four parts of de-ionized water, and then the chip is rinsed with de-ionized water for one minute and dried with nitrogen. In this case it is not strictly required to spin a bi-layer resist because the structures to fabricate are simply contacts and, moreover they are quite large (in the µm range). The typical "ears" effect of Tungsten (W) expected after lift-off is minimized using a thick resist, hence they should not impact the sheet resistance measurement.



50 nm of W are deposited by DC magnetron sputtering. The W target is sputtered with 120 W of DC power and in a plasma formed with 20 sccm of argon. The chamber pressure is kept law to 6 µbar in order to reduce the incorporation of impurities in the deposited layer.



A lift-off step is performed leaving the chip for almost 5 minutes into a baker filled with pure acetone and agitating it with ultrasound. In order to remove possible acetone residues, a rinsing with isopropanol (IPA) is performed and then the chip is dried with nitrogen.

To have a much relevant value of the extracted parameters, the above described CTLM processing considers TaO_x deposited on HfO_2/TiN/Si substrate. In this way the material growth of TaO_x is the same of the devices. Nevertheless, CTLM structures were fabricated also on top of 100 nm thick SiO_2 layer, in order to remove any possible contribution from the bottom chuck during the measurement. The result of these measurements is provided in Section A.6.
Four groups of CTLM structures are defined on the chip to have a larger statistic on the extracted parameters. Each group includes four columns, which are on their own split in two sub- ones containing structures with a different radius of the inner contact (100 µm or 200 µm). For each structure then the gap spacing is varied from 15 µm to 5 µm in steps of 2.5 µm. An example of one CTLM block is depicted in Figure 4.6.
A current from 1 µA to 10 µA in steps of 1 µA is applied to each structure and the voltage drop is measured. This value is then divided by the associated current one to obtain the total resistance. An entire column must be measured to have a full measurement with the total resistance as function of the gap spacing. Then, the correction factors are applied and the corrected data are fitted, like in the example in Figure 4.7. For each sample at least five different columns were measured in order to have a reliable expression of the sheet resistance (R_{sh}) and the resistivity (ρ) in terms of mean value and standard deviation.

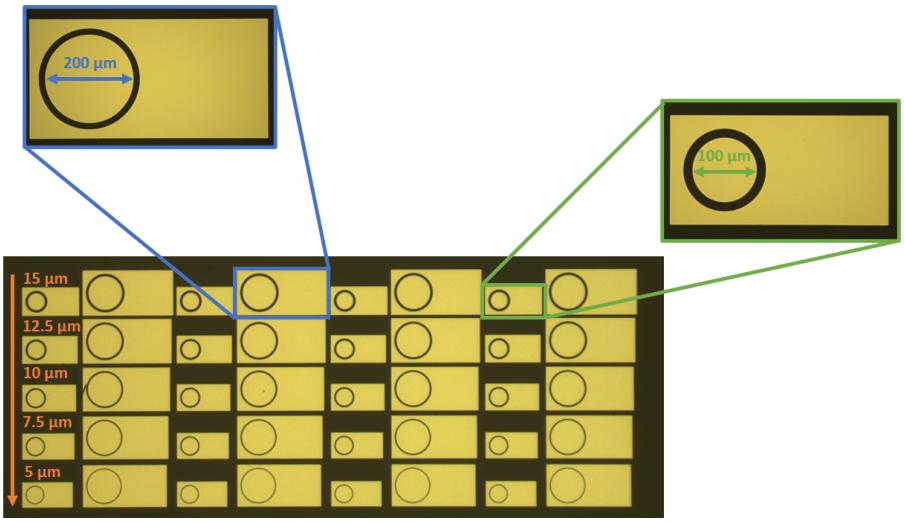


Figure 4.6: Microscope image of one group of CTLM test structures. In the insets a zoom on two structures with a different radius of the inner metal contact is shown, while the different gap spacings are reported in orange.

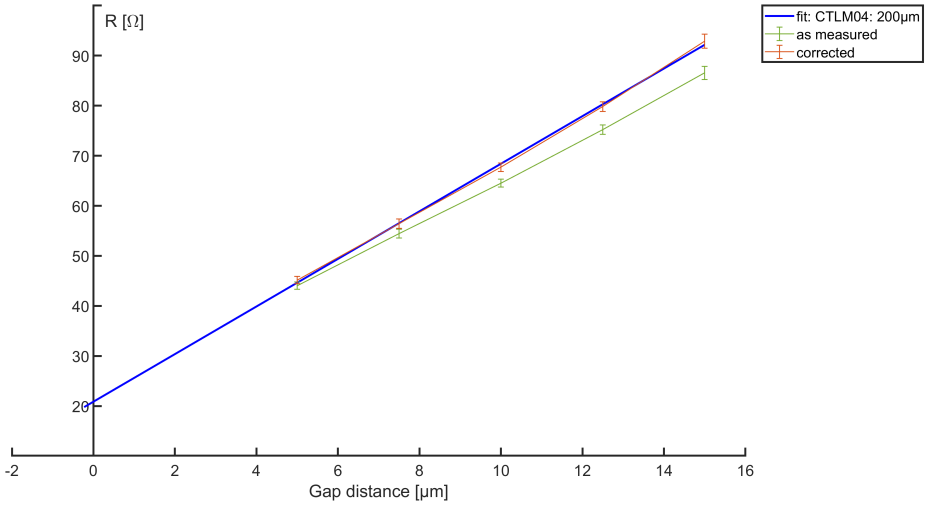


Figure 4.7: Example of fitting of CTLM data taken on a TaO_x deposited at 6 µbar on HfO_2/TiN/Si substrate. In this case, the inner radius of the measured structures is 200 µm.

An example of plot considering all the measured CTLM for one sample is provided in Figure 4.8, while in Section A.5 it is possible to find the plot for each TaO_x deposited on a HfO_2/TiN/Si substrate, with the exclusion of the one deposited at 11 µbar. The CTLM structures defined on the TaO_x deposited at the largest pressure

were not possible to be measured, since the layer did not show an ohmic behaviour even reducing the current in the nA range. Moreover, increasing the current in the structures to the µA range, a saturation in the voltage drop to the maximum value of the parameter analyzer (10 V) was observed. These observations suggest that the layer has an insulating nature with a sheet resistance in the GΩ range.

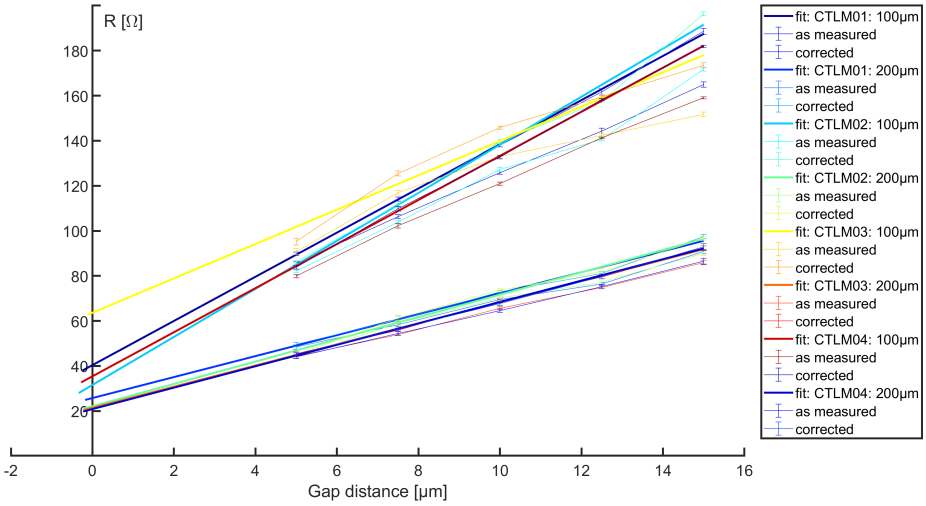


Figure 4.8: Example of fitting of all the CTLM data for the TaO_x deposited at 6 µbar on HfO_2/TiN/Si substrate.

The computed value of R_{sh} and ρ for each layer with the associated standard deviation are reported in Table 4.2, while a graphical representation of the evolution of these values is depicted in Figure 4.9.

Deposition pressure [µbar]	R_{sh}[kΩ/□]	ρ [Ωµm]
6	2.98 ± 0.27	81.76 ± 7.46
7	10.95 ± 0.28	313.16 ± 7.98
8	31.43 ± 4.73	880.00 ± 132.4
9	89.55 ± 0.52	2561.13 ± 14.92
10	172.23 ± 2.02	4977.44 ± 58.31
11	—	—

Table 4.2: Sheet resistance R_{sh} and resistivity ρ as function of the deposition pressure for TaO_x samples deposited on HfO_2/TiN/Si substrate. These values were not measurable for the sample deposited at 11 µbar.

Figure 4.9 clearly shows that increasing the pressure in the sputtering chamber,

the R_{sh} and ρ increase both because of a larger oxygen incorporation in the layer. Interestingly, the TaO_x film deposited at 10 µbar has a double resistivity with respect to the one deposited at 9 µbar, even though they are characterized by the same T-TaO_x density. This could mean that also the first nanometers of the TaO_x (B-TaO_x) play a role in the definition of the value of resistivity in the sample or that there is a different distribution of the oxygen content between these two samples. In both cases the electrical properties of the grown film are impacted in a non-negligible way. The values of resistivity found for the TaO_x samples deposited on HfO_2/TiN/Si substrate are in good agreement also with those of the same thin-films grown on SiO_2/Si substrate, as it is reported in Section A.6. This confirms that the layer that is actually sampled is the TaO_x.

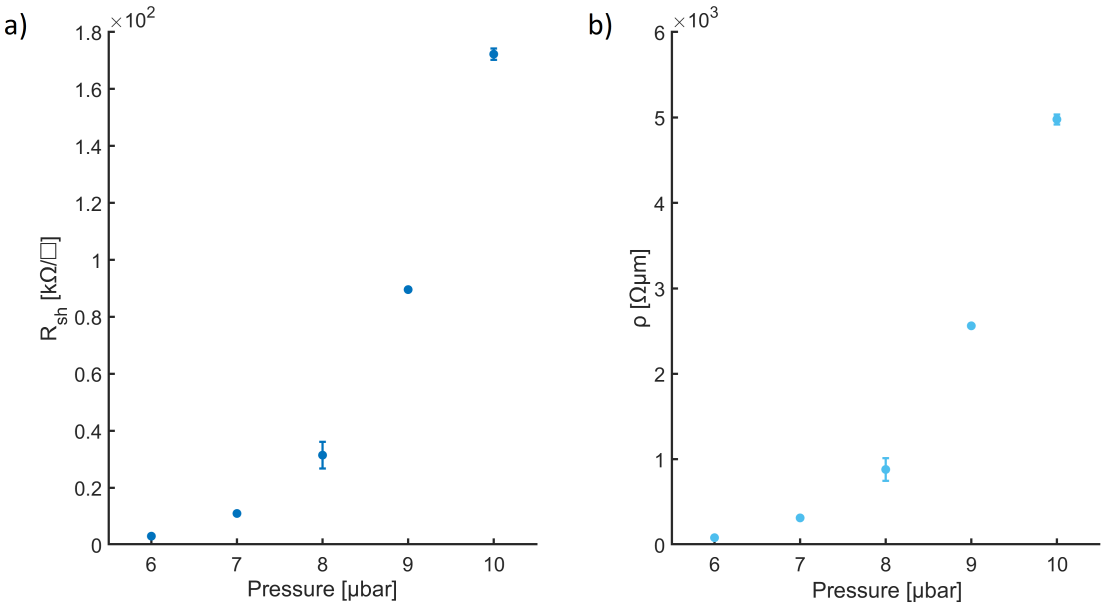


Figure 4.9: Plot of the sheet resistance R_{sh} (a) and the resistivity ρ (b) as function of the deposition pressure for the TaO_x samples deposited on HfO_2/TiN/Si substrate. Not visible error bars are hidden within the mean value dot.

4.2 Vertical ReRAM devices: processing and device optimization

In this section, the processing as well as the electrical measurements on the vertical ReRAM devices are presented, with the goal to optimize the bilayer structure towards the analogue behaviour. For this reason, a first comparison is made considering the quasi-static measurements on ReRAM embedding a CMO with a different sheet resistance, and hence resistivity. A further optimization is then obtained by

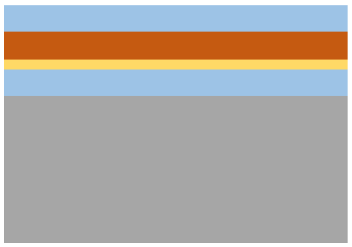
varying the thickness of the CMO layer giving to ReRAM the best DC performances. The analogue ReRAM technologies among the ones proposed are tested with pulsed measurements and the linear state variation is demonstrated. Finally, this section is intended to clarify the impact of the addition of the TaO_x in the stack, hence a comparison with devices having platinum as bottom electrode is also provided.

4.2.1 Processing of vertical ReRAM devices

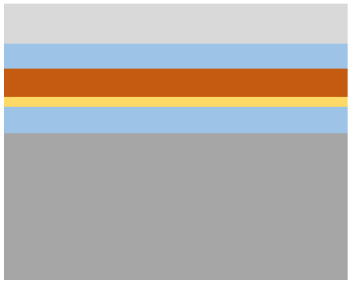
ReRAM devices are usually fabricated vertically because of an easier and more compact integration with MOS transistors with respect to planar ones. The main steps in the fabrication of vertical ReRAM devices are reported below. It is important to notice that the material used as well as the temperature of each different step are fully compatible with CMOS technology.

Cross-section of the sample	**Explanation of the processing step**
	The starting substrate is a 500 µm thick Si one doped with arsenic in order to be conductive. On top of it, one or two nanometers of native SiO_2 are usually grown.
	A 10 s dip in a solution containing one part of BOE and seven parts of de-ionized water is performed in order to get rid of the native SiO_2 on top of the substrate. This step is fundamental for the good electrical connection of the device to the probe station chuck.
	PEALD is used to deposit first nominally 20 nm of TiN and then 6 nm of HfO_2. In particular, TiN is deposited at 300 °C alternating Tetrakis-(dimethylamido)-titan (TDMAT) as metal precursor and a nitrogen plasma for 260 cycles. After, HfO_2 is deposited at 290 °C using Tetrakis-(ethylmethtlamino)-hafnium (TEMAH) as metal precursor and an oxygen plasma alternated for 50 cycles.

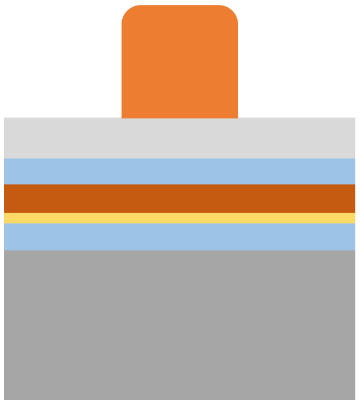
The TaO_x layer is deposited as already explained in Section 4.1.1.



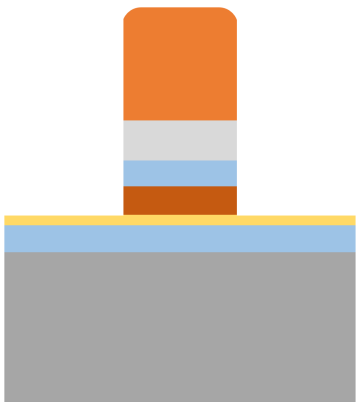
20 nm of TiN are deposited by RF reactive magnetron sputtering. The titanium target is sputtered with 400 W of DC power and 100 V of AC bias in a plasma formed starting from 2 sccm of nitrogen and 18 sccm of argon. The chamber pressure is kept low to 6 µbar in order to improve the layer purity.



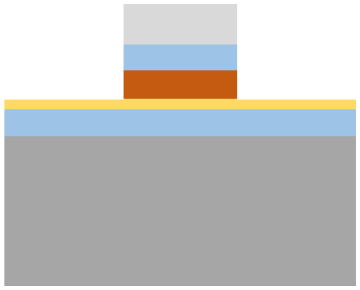
50 nm of W are deposited by DC magnetron sputtering. The W target is sputtered with 120 W of DC power and in a plasma formed with 10 sccm of argon. Also in this case, the chamber pressure is kept low to 6 µbar in order to reduce the incorporation of impurities.



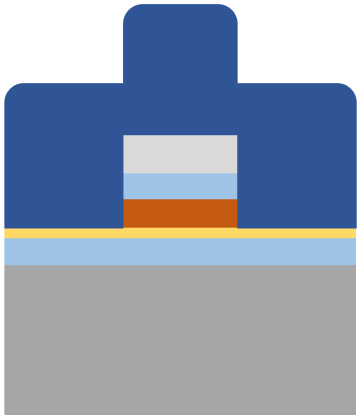
500 nm of positive polarity resist are spun on the chip after the deposition of HDMS as adhesion promoter. A baking step is then performed at 110 °C for one minute. Then the resist is exposed by laser writing at 45 W and developed. The development is done dipping the chip for 25 s in a solution containing one part of developer and four parts of de-ionized water, and then the chip is rinsed with de-ionized water for one minute and dried with nitrogen.



An etching step with ICP is performed in order to define the structure of the device, called MESA, which is protected by the resist that remained in place after the previous step. The etching is performed in a plasma generated with an RF power of 1200 W, 18 mbar of pressure and gas mixture constituted by 10 sccm of N_2, 15 sccm of SF_6 and 15 sccm of CHF_3. The etching recipe was calibrated by a colleague in order to have a high selectivity towards HfO_2, thus stopping at this layer.



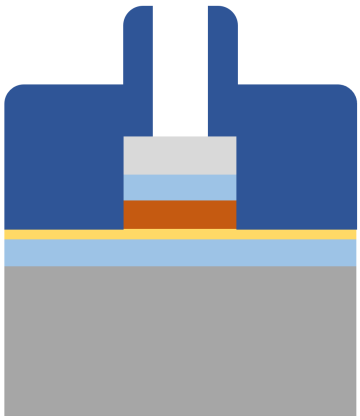
The resist is stripped dipping the chip in acetone and leaving it in agitation by ultrasound for a minute. In order to remove possible residues of resist on the chip, it is first rinsed with IPA and then further cleaned with an oxygen plasma for one and half minute. The oxygen plasma is generated with 50 sccm of O_2, 100 W of RF power and high pressure (150 mTorr) in the RIE tool in order to fix the temperature to 20 °C. The pressure in the chamber is kept high in order to reduce as much as possible unwanted etching of the sample. This oxygen step was tested to do not have any impact on the devices by quasi-static electrical characterization.



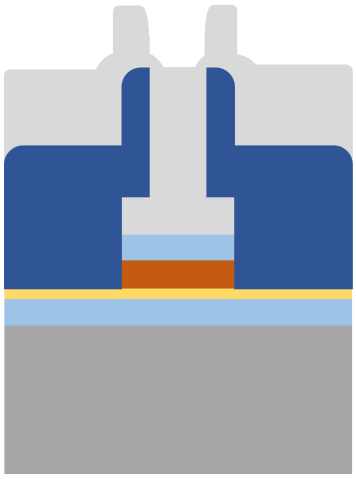
100 nm of Si_3N_4 are deposited as passivation layer using PECVD at a temperature of 300 °C. The Si_3N_4 is deposited with a deposition rate of 15.6 nm/min from a plasma of 2% of SiH_4 (120 sccm) and N_2 (1000 sccm), which is generated with an RF power of 20 W and a pressure of 1000 mTorr. Si_3N_4 was preferred to SiO_2 for the absence of oxygen which could be critical for the resistivity of the TaO_x.

500 nm of positive polarity resist are spun on the chip after the deposition of HDMS as adhesion promoter. The baking step is then performed at 110 °C for one minute. Then the resist is exposed by laser writing at 45 W and developed. The development is done dipping the chip for 25 s in a solution containing one part of developer and four parts of de-ionized water, and then the chip is rinsed with de-ionized water for one minute and dried with nitrogen.

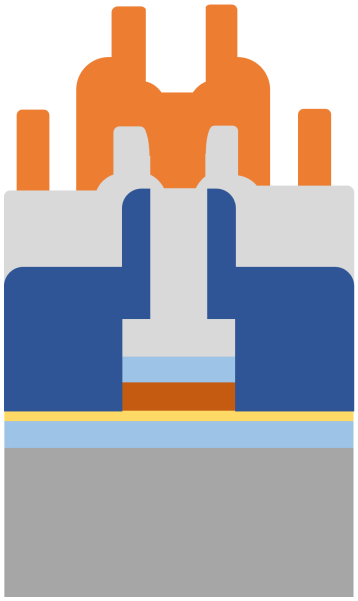
In order to open a VIA for the connection to the TE, an etching step with RIE is performed. The plasma is generated by applying an RF power of 100 W and a pressure of 55 mTorr at room temperature to a gas mixture composed by O_2 (5 sccm) and CHF_3 (50 sccm). The resulting etch rate of Si_3N_4 is 37.7 nm/min.



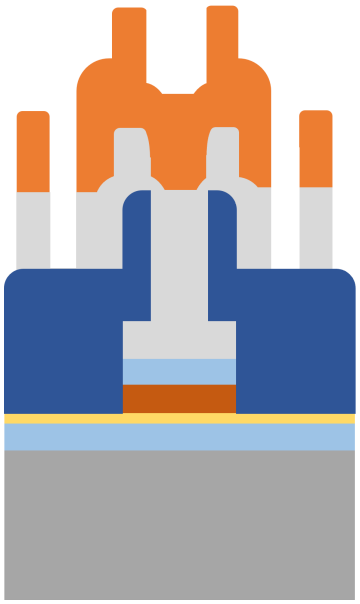
The resist is stripped dipping the chip in acetone and leaving it in agitation by ultrasound for a minute. In order to remove possible residues of resist on the chip, it is then rinsed with IPA. No oxygen treatment is required at this step since the RIE does not produce usually a large resist burning such the one obtained after the ICP process.



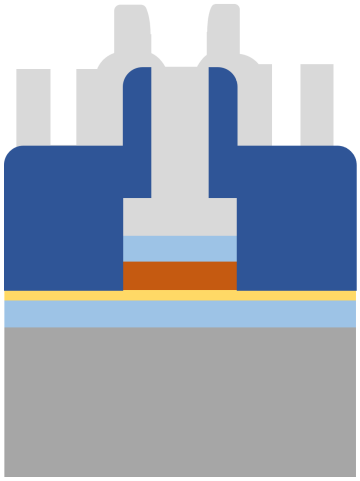
100 nm of W are deposited in order to contact the TE of the device. The deposition is performed using DC magnetron sputtering with a power of 120 W, 6 µbar of pressure and 20 sccm of argon. The larger argon flow rate with respect to the previous W deposition is due to the need to increase the deposition rate since two times the previous thickness is now deposited.



500 nm of positive polarity resist are spun on the chip after the deposition of HDMS as adhesion promoter. The baking step is then performed at 110 °C for one minute. Then the resist is exposed by laser writing at 45 W and developed. The development is done dipping the chip for 25 s in a solution containing one part of developer and four parts of de-ionized water, and then the chip is rinsed with de-ionized water for one minute and dried with nitrogen.



The contact pads and the outer W shielding is defined with a RIE etching of W. This step is also important to isolate the devices between them. In particular, the W is etched at a rate of 15.4 nm/min using plasma generated from N_2 (30 sccm) and SF_6 (30 sccm) generated with an RF power of 25 W, a pressure of 35 mTorr and at room temperature.



The resist is stripped dipping the chip in acetone and leaving it in agitation by ultrasound for a minute. In order to remove possible residues of resist on the chip, it is then rinsed with IPA. No oxygen treatment is required at this step since a RIE process was performed for the ectching.

Devices with different sizes have been fabricated: (60 × 60) µm^2, (30 × 30) µm^2, (12 × 12) µm^2 and (6 × 6) µm^2. In particular, a microscope image of (6 × 6) µm^2 devices is provided in Figure 4.10.

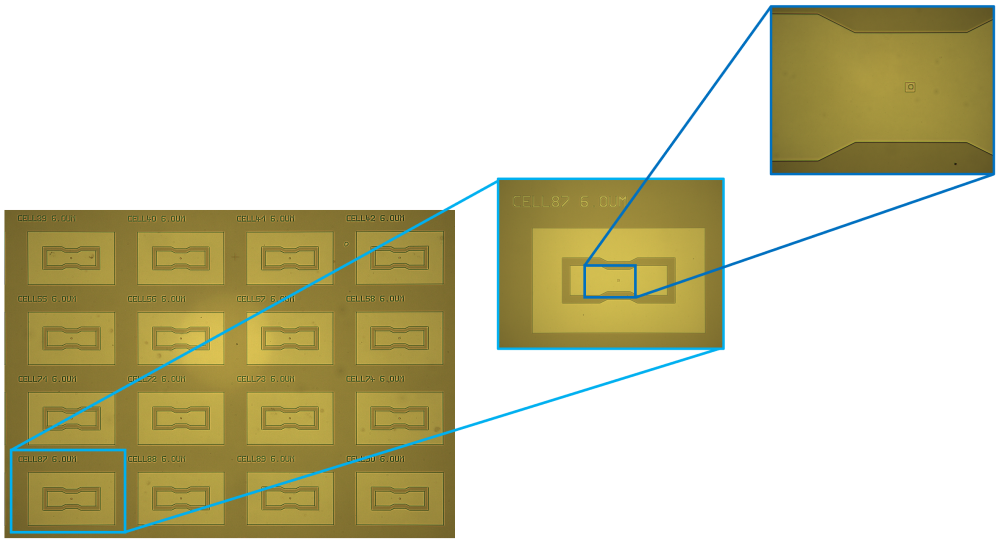


Figure 4.10: Microscope image of (6 × 6) µm^2 devices. In the insets a zoom on a single device and one on the overlap between the VIA and the MESA are shown.

4.2.2 Impact of the conductivity of TaO_x layer on ReRAM quasi-static performances

In Section 4.1 it has been shown how it is possible to obtain TaO_x layers with a different sheet resistance, and hence resistivity, by changing the pressure in the sputtering chamber. In this section we will discuss the impact of the TaO_x layer conductivity on the ReRAM performances. Bilayer ReRAM devices have been fabricated with two different kinds of substoichiometric tantalum oxide but having the same thickness (30 nm) for comparison purposes. In particular, the TaO_x layers embedded in the compared device stacks are deposited at a pressure of 6 µbar and 9 µbar, resulting in a sheet resistance R_{sh}≈3 kΩ/□ and ≈90 kΩ/□, respectively. A quasi-static analysis has been performed on twenty devices of each ReRAM stack. The considered devices have the same dimensions ((6 × 6) µm^2) to have a statistically relevant analysis.
The forming behaviour of the two ReRAM stacks is shown in Figure 4.11. As shown in the I-V curves of Figure 4.11(c) and Figure 4.11(d), a compliance current of 50 µA is applied during the electroforming step for devices embedding the more conductive TaO_x layer. The application of this upper bound for the current resulted to be critical to obtain stable working devices after electroforming for this specific

layer stack. It is not required for the ReRAM with a less conductive CMO.
We explain this finding as follow: the more resistive it is the TaO_x layer, the less current flows into the device, thus resulting in a softer breakdown of the HfO_2. Having a more resistive TaO_x provides an in-built compliance during electroforming. The TaO_x conductivity clearly impacts the voltage and the current at which the forming event occurs. As it is shown in Figure 4.11(a) and Figure 4.11(b), the lower it is the resistivity of the CMO, the lower it is the voltage required for the occurrence of electroforming. In accordance with this observation, the current at which this process occurs is instead lower the more resistive it is the TaO_x.

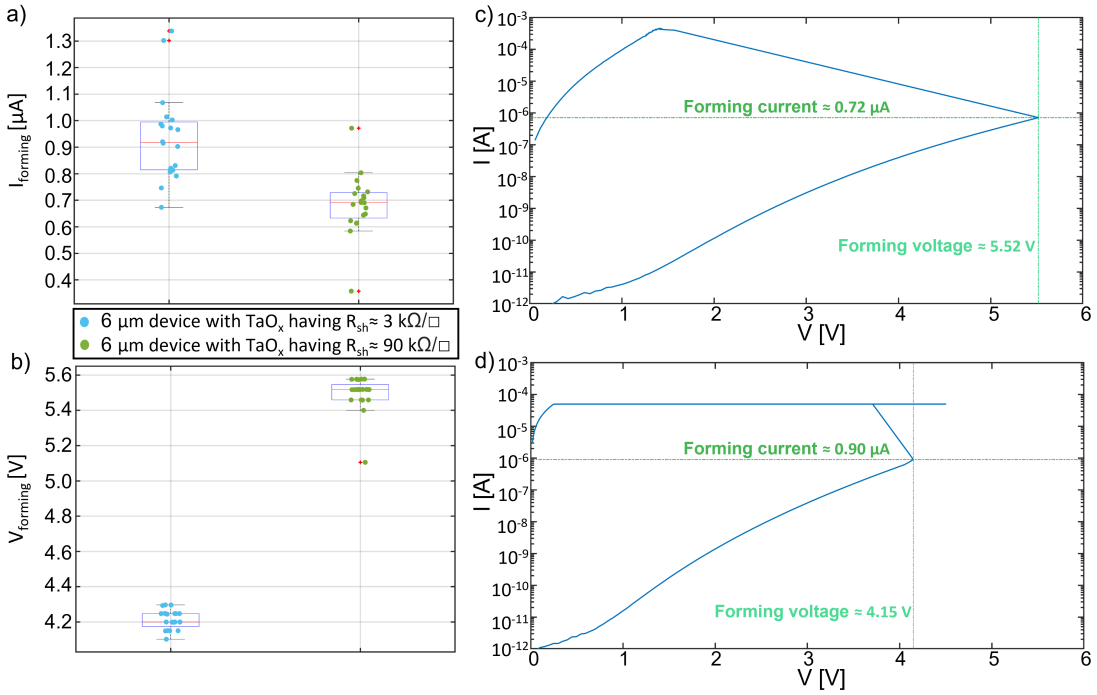


Figure 4.11: Boxplot representation of the forming current (a) and of the forming voltage (b) for ReRAM embedding TaO_x with different conductivity. An example of the typical electroforming I-V sweep is shown for ReRAM embedding a TaO_x with R_{sh}≈90 kΩ/□ (c) as well as for a device embedding a TaO_x with R_{sh}≈3 kΩ/□ (d).

An additional important parameter to consider is the value of the post-forming resistance $R_{post\text{-}forming}$ in the two ReRAM stacks. This value is usually associated to the size of the oxygen vacancy filament in the HfO_2. The statistical analysis shown in Figure 4.12 supports our considerations. The larger it is the resistivity of the TaO_x, the less aggressive it results the electroforming in the HfO_2, probably leading to a thinner filament. Indeed, if compared to the baseline ReRAMs, the devices seem to be in a HRS after the forming process. The filament size can be

further increased, but in a controlled manner, in the first SET operation.

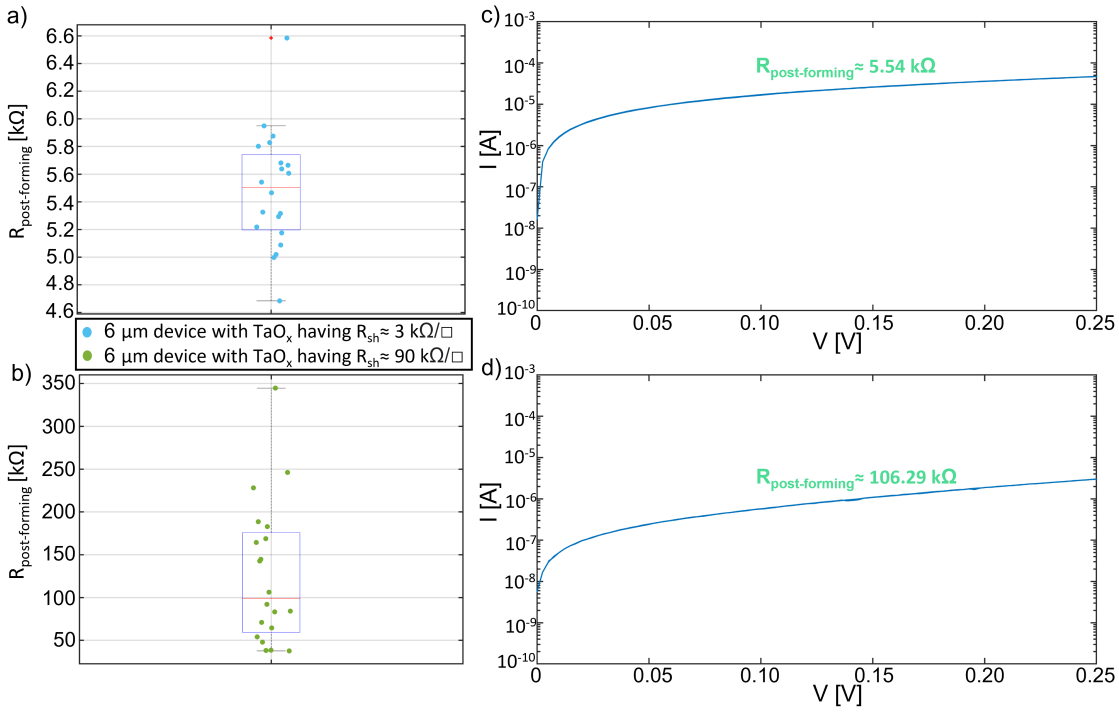


Figure 4.12: Boxplot representation of the $R_{post\text{-}forming}$ for ReRAM embedding TaO_x with different conductivity (a)(b). An example of the typical reading I-V sweep after forming is shown for ReRAM embedding a TaO_x with R_{sh}≈3 kΩ/□ (c), as well as for a device embedding a TaO_x with R_{sh}≈90 kΩ/□ (d).

After forming, the devices can be programmed in multiple resistive states between LRS and HRS. Figure 4.13 shows an example of the typical switching behaviour over 20 consecutive cycles between these two memory states. The stop voltages (V_{STOP}) during the programming are fixed to the same values for both devices in order to have a more relevant comparison. These voltage values are chosen high enough to have a good On/Off ratio but below the threshold value that induces complementary switching, a phenomenon that can harm the device behaviour upon cycling [49].
As shown in Figure 4.13(a) and (b), both ReRAM stacks show a gradual SET and RESET transitions and a self-compliance during the SET one since the process that leads to the LRS is limited without the application of any compliance current. The graduality of the switching, obtained introducing these two TaO_x layers in the ReRAM stacks, leads to remarkably improved analogue memristors. In standard ReRAMs with metallic scavenging layer (see Figure 1.7), the abruptness of the

SET transition is one of the main limiting aspects to obtain a continuous of analogue states, therefore hindering their implementation in crossbars. In addition, the CMO-based ReRAMs have opposite polarity respect to a standard baseline technology. This phenomenon is specific to the use of the CMO and it will be discussed more in details in Section 4.2.3 and 4.3.6.

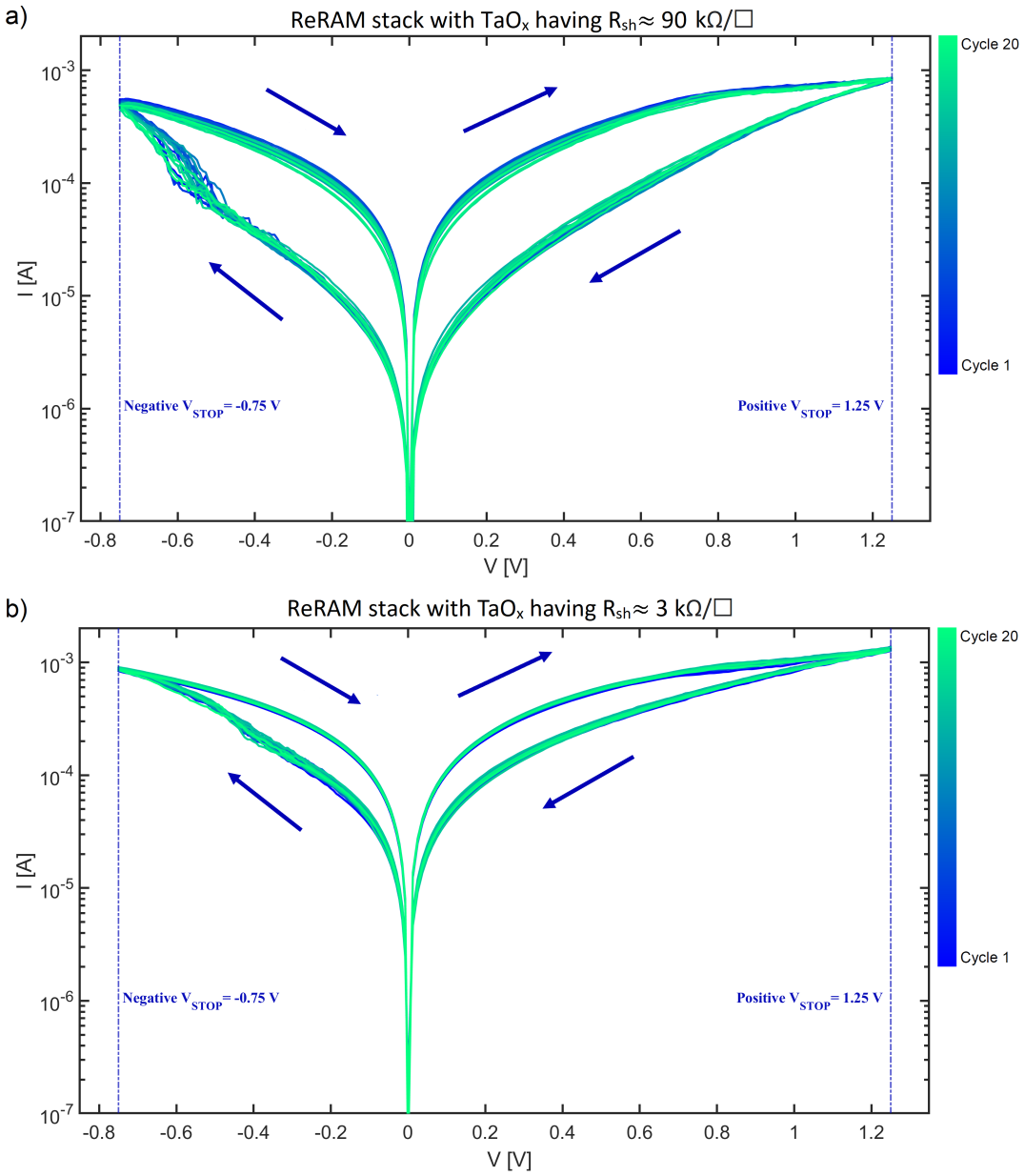


Figure 4.13: An example of the typical cycling behaviour over 20 consecutive cycles of ReRAM devices embedding a TaO_x with R_{sh}≈90 kΩ/□ (a) or a TaO_x with R_{sh}≈3 kΩ/□ (b).

From a further comparison of the two ReRAM stacks in Figure 4.13, it emerges that the On/Off ratio is different. From the statistical analysis on the mean On/Off ratio over the 20 cycles (see Figure 4.14), the devices that embed a more resistive TaO_x show a much higher mean On/Off ratio.

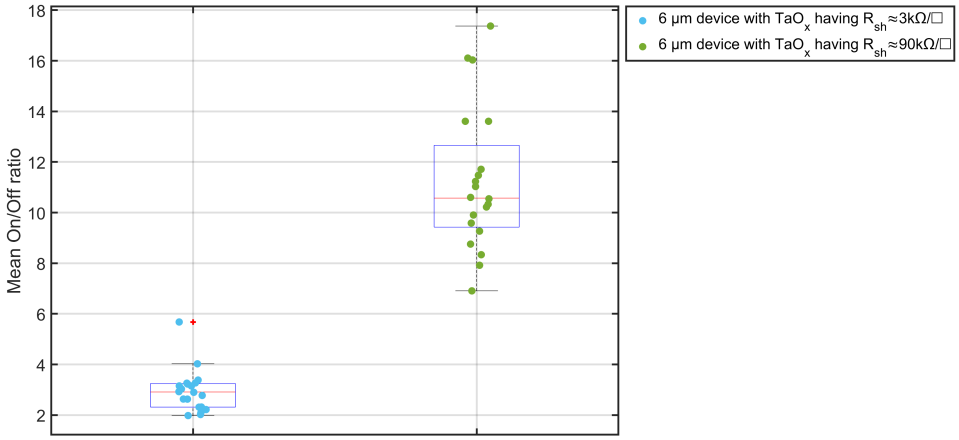


Figure 4.14: Boxplot representation of the On/Off ratio for ReRAM embedding TaO_x with different conductivity.

In particular, Figure 4.15(a) and (b) show that the ReRAM with a more resistive TaO_x presents a larger HRS with respect to the one embedding the more conductive CMO, while the LRS is almost the same in both cases. This result is related to the occurrence of a more efficient RESET process when a more resistive CMO is present in the stack and it can be explained by the role of the thermal conductivity of the TaO_x layer. As already reported in a previous study [50], the thermal conductivity of TaO_x is reduced by increasing the oxygen content, and thus by increasing its resistivity. The reduction of the thermal conductivity in the more resistive TaO_x layer could result in a slower heat dissipation in the layer stack, hence more energy is provided for the recombination of the oxygen vacancies and the oxygen ions. This can result in a deeper RESET with respect to the devices that embed a more conductive TaO_x. The LRS presents a tight distribution around its median value (see Figure 4.15(b)) while the HRS value shows a higher spread (see Figure 4.15(a)), reflecting a larger device-to-device variability. Figure 4.15(c) and (d) show the stability of HRS and LRS for both ReRAM devices. In both cases, the values of the LRS and HRS are stable and do not drift for at least 20 cycles.

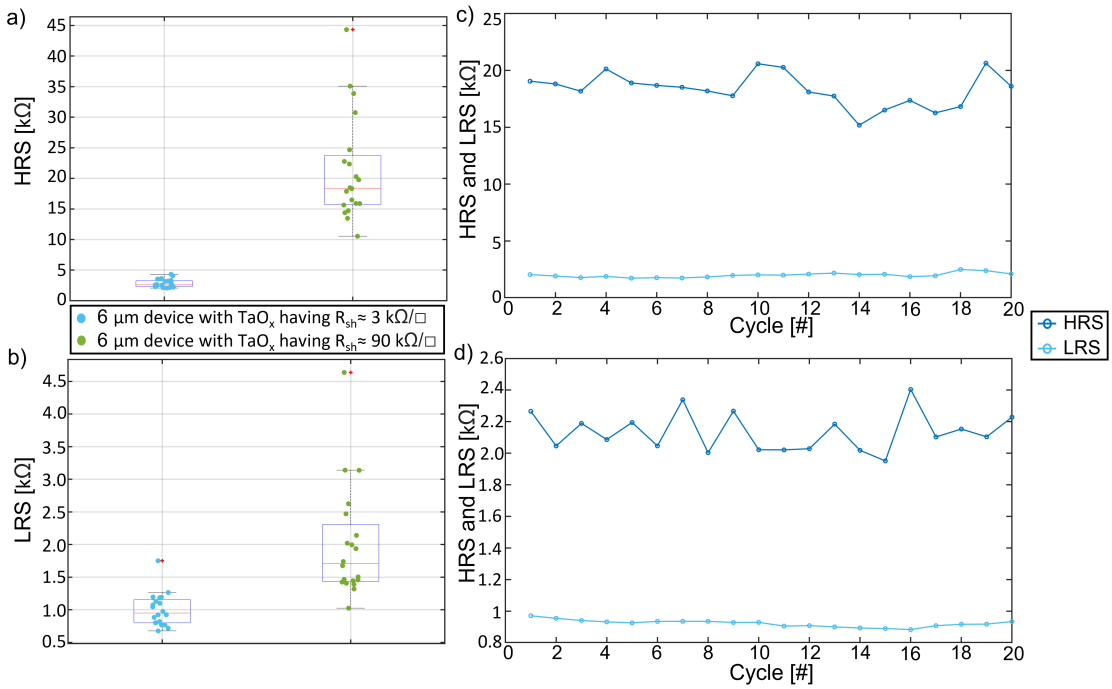


Figure 4.15: Boxplot representation of the HRS (a) and LRS (b) for ReRAM embedding TaO_x with different conductivity. An example of the endurance is provided for ReRAM embedding a TaO_x with R_{sh}≈90 kΩ/□ (c), as well as for a device embedding a TaO_x with R_{sh}≈3 kΩ/□ (d).

We now compare the mean HRS over 20 cycles with the devices post-forming resistance. As mentioned, after the electroforming step both TaO_x based ReRAMs are in a high resistance state. It is interesting to notice that the mean HRS is smaller in respect to the $R_{post\text{-}forming}$ (see Figure 4.16). This observation could be explained by a redistribution of the oxygen vacancies that results in a more stable HRS condition for the device at smaller resistance value than that right after post-forming. A possible motivation of this observation is provided in Section 4.3.6.

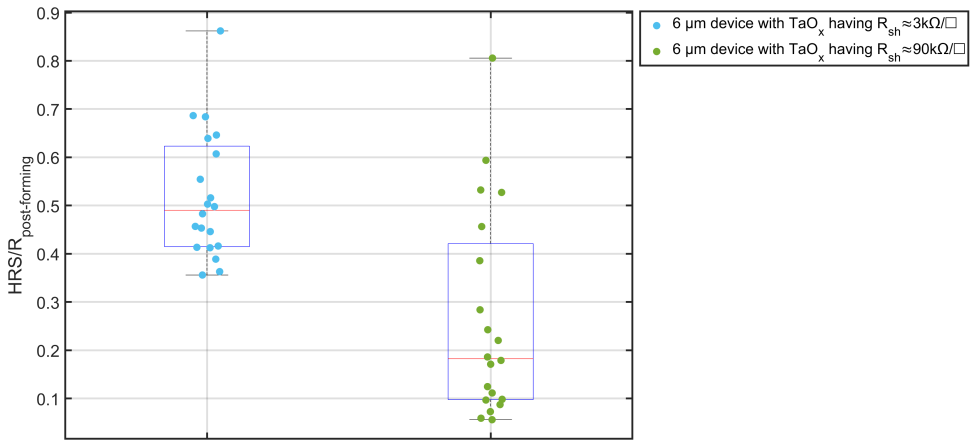


Figure 4.16: Boxplot representation of the ratio between the mean HRS over 20 cycles and $R_{post\text{-}forming}$ for ReRAM embedding TaO_x with different conductivity.

4.2.3 Explanation of the switching polarity in vertical ReRAM with platinum (Pt) as bottom electrode

The ReRAM devices presented in the Section 4.2.2 are characterized by the SET transition occurring at negative voltages, while the RESET transition is occurring at positive ones. This kind of polarity is opposite with respect to what is usually observed in literature for baseline ReRAM devices, even though the coexistence of two opposite switching modes in VCM-type ReRAM has been already demonstrated [51]. Typically, the switching event is located close to the metal electrode with the largest work function and the lowest oxidation enthalpy [51]. In the proposed ReRAM stack, hence a SET event is expected by applying a positive voltage to the TE because the electrode with the highest work function and the lowest oxidation enthalpy should be the bottom TiN. To understand if the observed opposite polarity in the device is somewhat due to the insertion of the the TaO_x layer in the stack and its interaction with the other layers, we have modified the layer stack to have a clear and strong asymmetry of the electrodes. Vertical bilayer ReRAM devices (with 30 nm TaO_x deposited at 9 µbar) have been processed with 20 nm of platinum (Pt) as BE, deposited by evaporation. Pt is extensively used in literature as inert electrode for ReRAM, since it is the metal with the largest work function. The behaviour of a (6 × 6) μm^2 device with Pt as BE is reported in Figure 4.17. The electroforming I-V sweep (see Figure 4.17(a)) can be performed without any current compliance since it is already provided by the TaO_x and it results in a high $R_{post\text{-}fomrming}$ (see Figure 4.17(b)). However, the forming voltage is more than 1 V smaller than the case with TiN as BE. This can be explained by the fact that

growing HfO_2 on top of the Pt electrode avoids the formation of the oxidized layer as for the case of the ReRAM with TiN BE (see Section 4.1.3). This thin oxidized layer resulted in an additional voltage drop during the electroforming step for the devices. The absence of this unwanted oxide layer for the ReRAM with Pt BE could also explain the lower forming current.

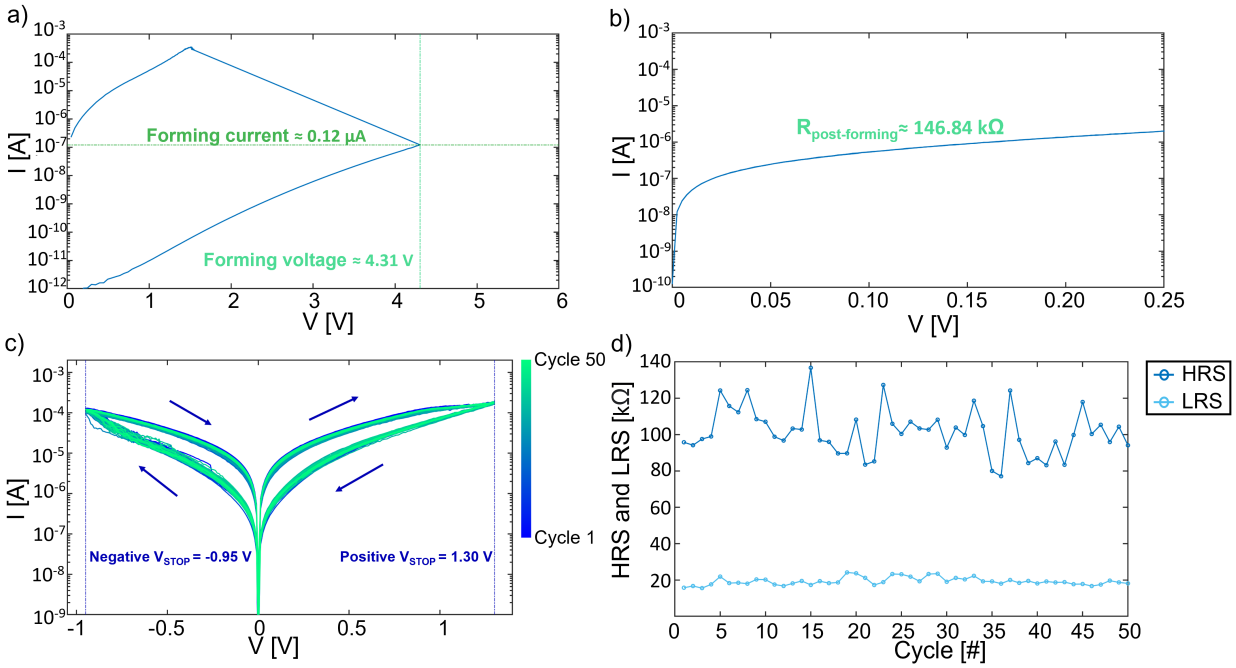


Figure 4.17: The behaviour of a vertical ReRAM device with Pt as BE is shown during electroforming (a), post-forming reading (b) and programming over 50 cycles (c). The device retention over 50 cycles is also reported (d).

The device has been cycled over 50 consecutive times showing no switching degradation. Figure 4.17(c) displays how the polarity remains the same of the CMOS-compatible ReRAM with TiN BE. Also, the analogue behaviour of the device in both transitions is not impacted by the change of the BE material. This observation suggests that the interface actually playing a role in the resistance state change of the ReRAM is the TaO_x-HfO_2 one. Indeed, the TaO_x behaves as an OEL and, as already stated by Celano et al. [12], it induces a distributed reservoir of defects at the interface with HfO_2 resulting in a large amount of building blocks available for the creation of one or more CFs in the oxide. However, the resistance value at both LRS and HRS is much higher than what is shown in Figure 4.15(c). The explanation for this could be found in the difference of the BE material in the two ReRAM stacks. Indeed, the higher work function of Pt with respect to TiN could result in an increased difficulty for the carriers to travel across the potential barrier at the BE. Moreover, the larger Pt work function could also be the reason for the

slightly larger absolute value of the V_{STOP} during the programming for obtaining a stable switching between LRS and HRS.

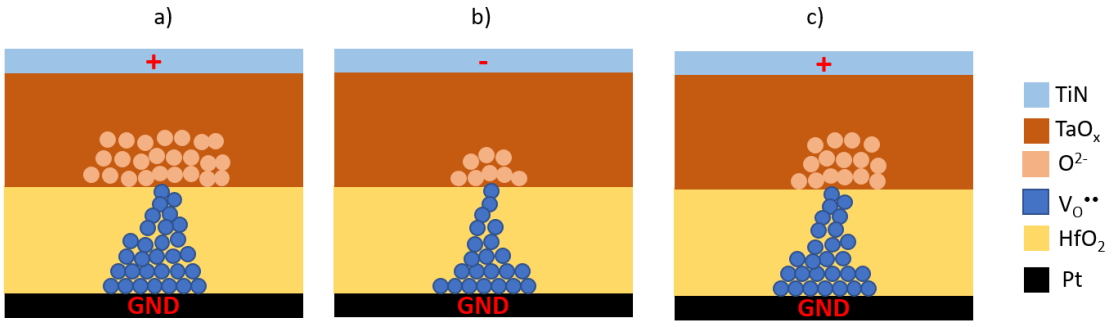


Figure 4.18: Microscopic description of oxygen ions (O^{-2}) and oxygen vacancies ($V_O^{\bullet\bullet}$) distribution in the device with Pt BE after electroforming (a), SET event (b) and RESET one (c). The voltage polarity applied to each electrode in each step is reported in red on the images.

A schematic representation of O^{-2} and $V_O^{\bullet\bullet}$ distribution in the ReRAM with Pt BE during the different transitions is given in Figure 4.18. Assuming the formation of a single filament, during the electroforming event, this one is created in the HfO_2 due to its dielectric nature, but since O^{-2} are accepted by the TaO_x, the conductivity of this layer is reduced possibly due to the appearance of an oxygen reach region at the interface with the filament. The appearance of this oxygen reach region explains the high value of the post-forming resistance. During the SET process, the negative voltage applied to the TE with respect to the grounded BE induces the migration of the O^{-2} into the HfO_2 layer, thus resulting in a thinning of the filament due to the recombination with the $V_O^{\bullet\bullet}$. This recombination is supposed to not destroy the filament and to reduce the extension of the oxygen reach region in the TaO_x, thus resulting in a more conductive state of the device (LRS), as it is shown in Figure 4.18(b). On the other hand, the application of a proper positive voltage on the TE results in the migration of the O^{-2} again in the TaO_x. Therefore, the filament is widening but an overall high resistance state is observed in the device (HRS) because of a larger oxygen reach region in the CMO, as depicted in Figure 4.18(c). The extension of this oxygen reach region never reaches the one present immediately after the occurrence of the electroforming event, thus explaining why the HRS is always lower than the post-forming resistance. A corroboration and refinement of the proposed microscopic description of the switching are provided through the Impedance Spectroscopy experiment illustrated in Section 4.3.6.

4.2.4 Impact of the thickness of TaO_x layer on ReRAM quasi-static performances

The results shown in Section 4.2.2 clearly demonstrate that for our proposed TaO_x-based ReRAM, the best trade-off between On/Off ratio and analogue behaviour is achieved embedding a more resistive CMO in the bilayer stack. The impact of thickness variations of such layer is now analysed comparing ReRAM devices with a thickness of the TaO_x layer equal to 20, 30, 40 or 60 nm.

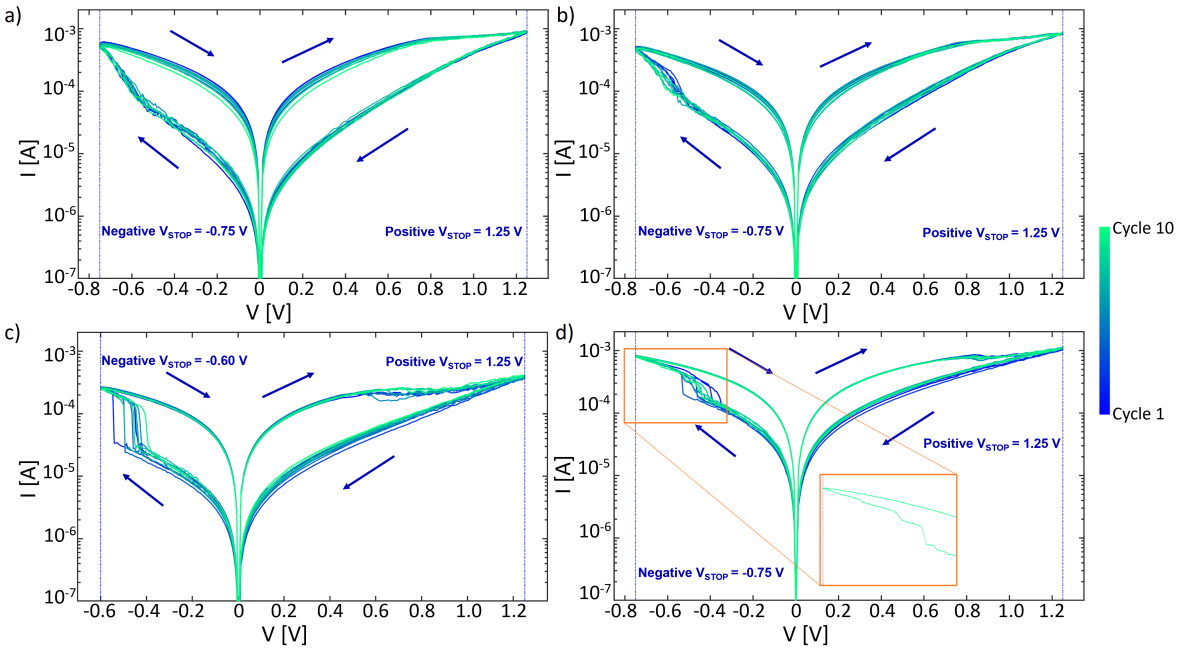


Figure 4.19: 10 switching cycles of vertical ReRAM devices with different thickness of the TaO_x layer: 20 nm (a), 30 nm (b), 40 nm (c) and 60 nm (d).

In particular, the switching behaviour over 10 cycles of devices with the four layer stacks variations is reported in Figure 4.19. To have a meaningful comparison all measured samples have the same size: (6×6) µm^2. Moreover, the programming voltages are kept the same for all the tested ReRAM with the only exception of the negative V_{STOP} in the layer stack with 40 nm thick TaO_x, which showed complementary switching increasing the absolute value of negative V_{STOP} beyond the used one.
The overview of I-V curves in Figure 4.19 shows that the devices with a 20 nm and a 30 nm thick TaO_x have comparable SET and RESET transitions. The analogue SET transition is instead degraded in the case of devices with 40 nm thick CMO. Indeed, in this case the typical abrupt and stochastic SET transition of baseline technology is observed. This change of behaviour in the device with the increased TaO_x thickness could be explained by the heat confinement produced by the CMO.

A thicker layer of TaO_x could result in a slower dissipation of the heat produced when the current is flowing into the ReRAM due to Joule heating effect. As a result, the SET transition is accelerated and the conductive filament is created abruptly. Counter intuitively, increasing the TaO_xthickness to 60 nm, the abruptness of the SET transition is reduced, and as shown by the inset in Figure 4.19(d), this is further amplified upon cycling. On the other hand, looking in more detail at the curve in the inset, we observe that in the tenth cycle the SET is characterized by multiple events and no more just one abrupt SET transition. The reason for the multiple steps during the SET transition could be found in the formation of multiple weak conductive filaments as consequence of the homogeneous heat distribution at the interface between the TaO_x and the HfO_2 [19][20]. Although the thickness of the layer is exactly the same of the ReRAM stack presented in [20], the observed state transitions are not analogue anymore. The different device behaviour obtained with respect to [20] highlights that not only the thermal properties but also the electrical conductivity of the CMO layer plays an important role in the definition of the device performances. Both thermal and electrical properties are influenced by the device thickness, but the proposed analysis clearly demonstrate that an optimal thickness for improving the device performances exists and it is between 20 and 30 nm, since they allow to have a good analogue fingerprint and a reasonable On/Off ratio. Figure 4.20 shows how the ReRAM devices with these two TaO_x layer thickness can be programmed in a different LRS and HRS by gradually changing the value of the V_{STOP}.

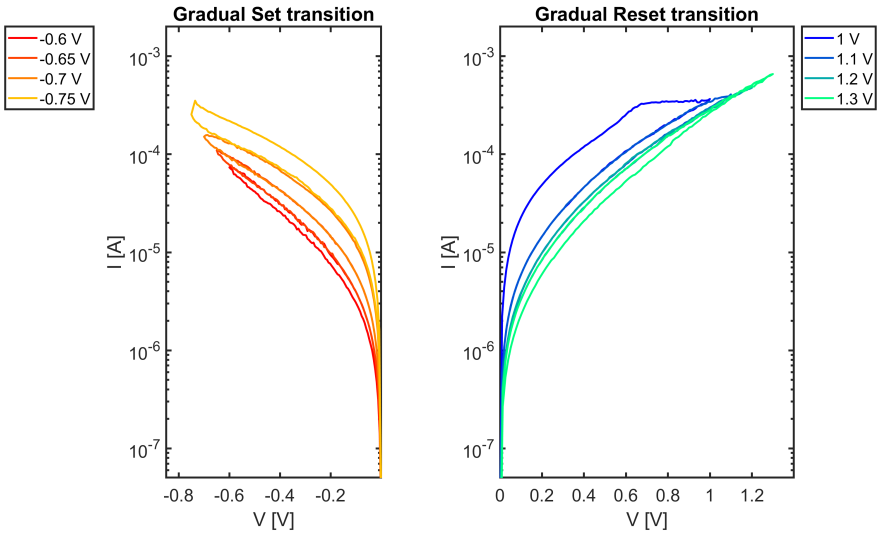


Figure 4.20: Example of gradual SET and RESET transition in a (12 × 12) µm^2 ReRAM device embedding a 20 nm thick TaO_x by changing the V_{STOP}.

To highlight the differences among these two ReRAM stacks a statistical analysis

is provided in Figure 4.21 considering ten devices for each stack and for each of the following different sizes: (6 × 6) µm^2, (12 × 12) µm^2 and (30 × 30) µm^2. All the devices underwent the same electroforming voltage sweep up to 6 V and the values of the forming voltage and of the forming current are plotted as function of the device area in Figure 4.21(a) and (b), respectively. The devices with the thinner TaO_x layer are shown to have a slightly smaller forming voltage and a higher forming current with respect to those with 30 nm thick TaO_x. The reason for this could be attributed to the smaller resistance value of the CMO. However, for both ReRAMs, clear trends for both quantities are observed as function of the device area. In particular, the forming current increases linearly with the device area (see Figure 4.21(b)), thus resulting in a constant value of the current density. On the other hand, the forming voltage follows an opposite trend, decreasing by increasing the device area (see Figure 4.21(a)).

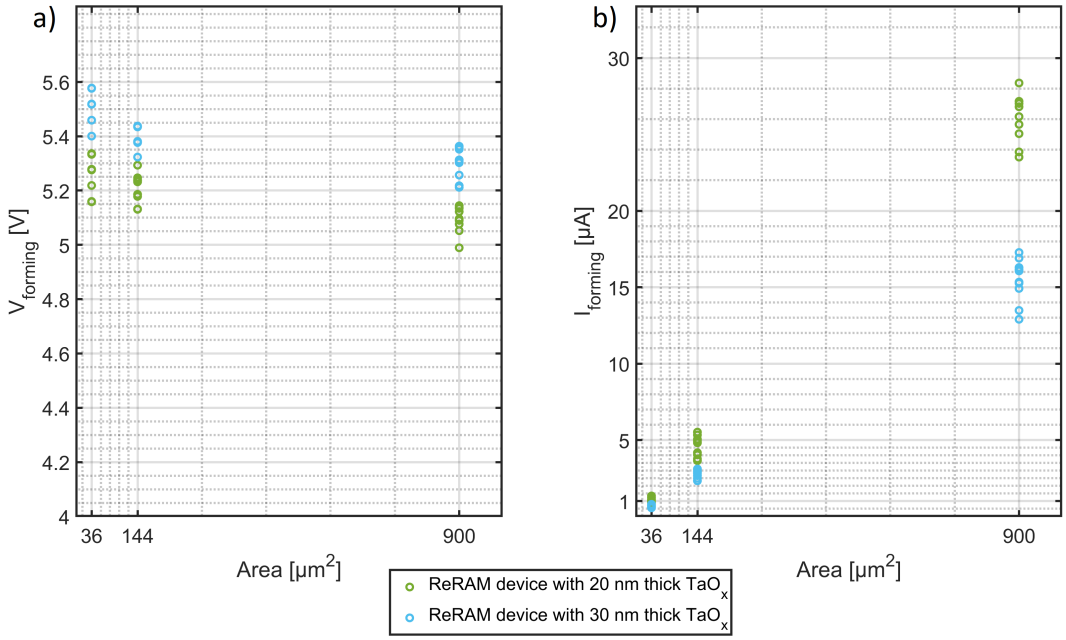


Figure 4.21: Forming voltage (a) and forming current(b) for the ReRAM devices embedding a 20 nm or 30 nm thick TaO_x.

These two effects can be correlated considering the probability of defects generation per unit of volume. Indeed, increasing the area of the device corresponds to an increase of the device volume, and hence a larger probability of defect generation. The bigger the device, the smaller it is the voltage required during the electroforming for the soft-breakdown of the HfO_2. The current passing into larger structures is higher since more defects acting as leakage paths are present.
A clear trend with respect to the device area is not observed for the post-forming

resistance shown in Figure 4.22. In general, both ReRAM stacks are characterized by high value of resistance after the forming step, even in the order of few hundreds of kΩ.

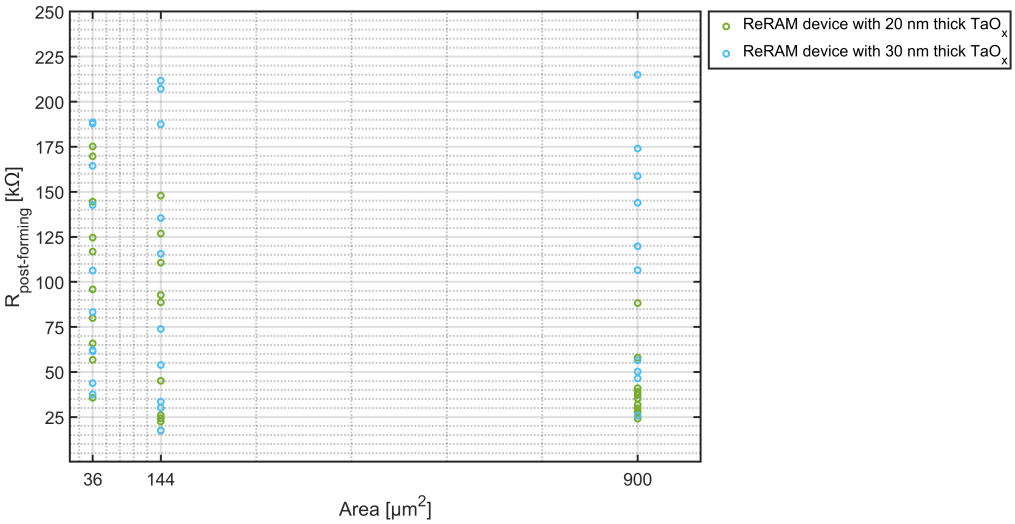


Figure 4.22: Post-forming resistance for the ReRAM devices embedding a 20 nm or 30 nm thick TaO_x.

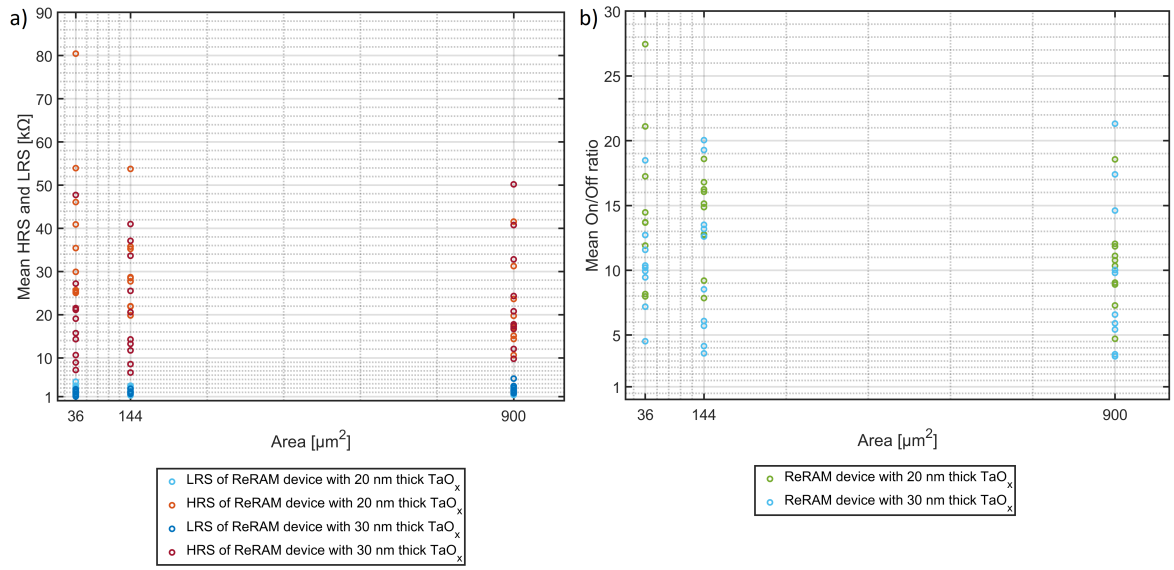


Figure 4.23: Mean value of resistive states (a) and On/Off ratio (b) for the ReRAM devices embedding a 20 nm or 30 nm thick TaO_x.

Considering 10 cycles for the devices of each ReRAM with 20 and 30 nm TaO_x

layer, the mean value of the HRS and LRS for each device are plotted in Figure 4.23(a). In order to find a balance between having a statistically relevant analysis and the need to try to obtain from each device size the best possible performances, the V_{STOP} used for the programming are (-0.75 ± 0.05) V and (1.25 ± 0.05) V. Figure 4.23(a) shows how the LRS is practically the same for both ReRAMs while a slight device-to-device variability is observable for the HRS, even considering devices corresponding to the same ReRAM stack. Besides this variability in the HRS, no area dependence is observed for the switching behaviour of the devices thus indicating a filamentary nature of the switching. The absence of area dependence in the mean value of the On/Off ratio shown in Figure 4.23(b) further supports the observation of filamentary behaviour in the both TaO_x-based memories. Moreover, considering the On/Off ratio, the two kind of bilayer ReRAM devices analysed are quite similar.

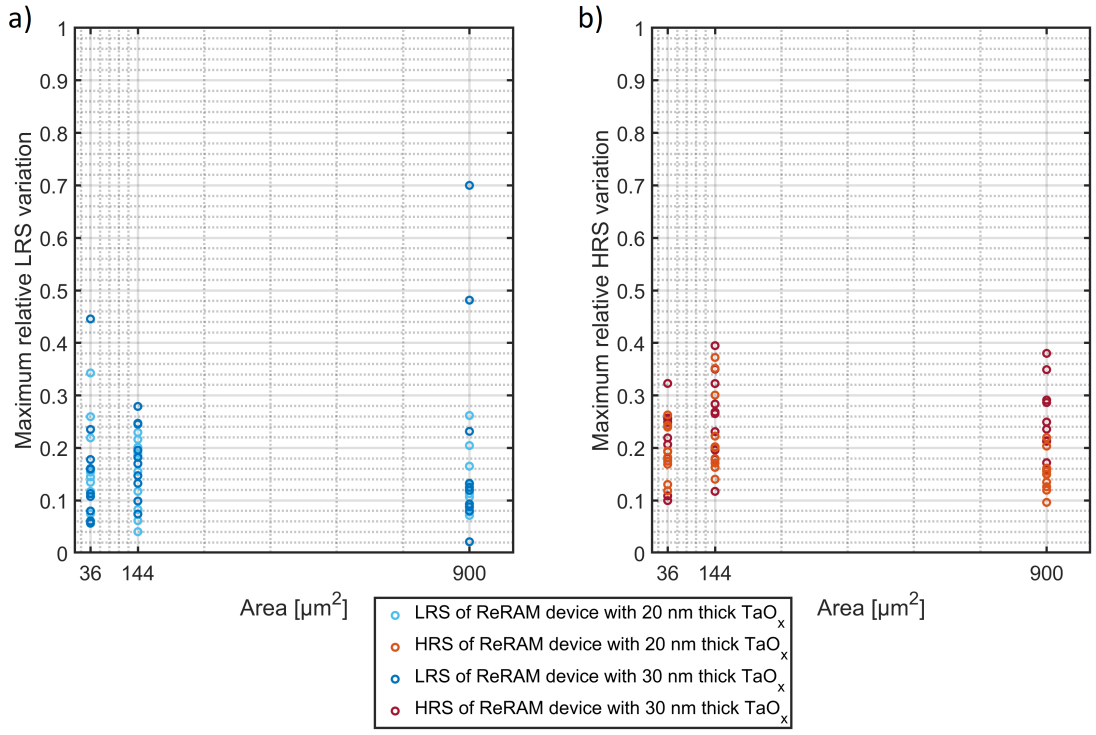


Figure 4.24: Maximum relative variation around the mean value of the LRS (a) and HRS (b) while cycling for the ReRAM devices embedding a 20 nm or 30 nm thick TaO_x.

Finally, an analysis of the stochasticity in the LRS and HRS for the two kinds of device stack has been performed. This information is evaluated by considering the maximum relative variation around the mean resistance value for both HRS and LRS while switching on 10 cycles each device. The correspondent mathematical

expression of this parameter is:

$$max(\frac{|R - R_{mean}|}{R_{mean}})$$

where R is the HRS or LRS accordingly to the parameter to compute.
As it is depicted in Figure 4.24, the two studied ReRAM stacks show a similar variation of the LRS and HRS around their mean value with no area dependence. In general, the variability in the resistance state is slightly smaller in the LRS than in the HRS and this can be attributed to the filamentary nature of the switching mechanism.
In standard baseline technology, the LRS is less stochastic than the HRS because the carriers follow the well-defined current path represented by the defect-rich filament in the HfO_2. On the other hand, the RESET transition leads to a larger variability since this event originates from the balance of drift and diffusion of the oxygen vacancies [11]. The origin of the stochasticity in HRS can be also found in the statistical fluctuation of the read current [49]. Since in the proposed devices, the two extreme resistance states have not a very different variability during cycling, this result could also suggest the existence of the oxygen reach region in the CMO at the interface with the HfO_2, which actually gives a contribution in the switching.

4.2.5 Pulsed measurements on analogue switching bilayer ReRAM devices

The quasi-static characterization demonstrated that the ReRAM stacks which are characterized by an analogue switching behaviour in both SET and RESET transitions are the one with a 30 nm thick TaO_x deposited at P_{dep}=6 µbar, and those embedding a TaO_x deposited at P_{dep}=9 µbar with a thickness of 20 or 30 nm. The performances of these devices are now presented from the point of view of pulsed measurements. The objective is to understand which among them can provide the best update of the resistance state in terms of linearity and symmetry, which are fundamental characteristics for training algorithms of neural networks. In literature, potentiation and depression curves that constitute the pulsed measurement are represented as the increase and decrease of the device conductance, respectively. This definition is kept also in this master thesis work but the potentiation and depression curves are shown on the resistance of the device. This choice is simply due to an hopefully easier visualization of the dynamic range of the proposed devices.
Figure 4.25 shows the pulsed measurement of a representative (6×6) µm^2 ReRAM device, in which the TaO_x has a thickness of 30 nm and a $R_{sh}\approx$3 kΩ/□. The presence of intermediate states in both transitions between the LRS and HRS in Figure 4.25 highlights the analogue behaviour of the device. Moreover, the device requires almost the same voltages as in the quasi-static I-V sweeps for the switching. The device is characterized by pretty linear potentiation and depression curves till reaching a saturation of the resistance state, where a more noisy behaviour is observed.

The noise in the resistance state is larger for the HRS than the LRS. This result is expected, since also in the quasi-static characterization, the HRS was characterized by a slightly larger stochasticity than the LRS (see Figure 4.15(d)). However, the more urgent problem to address is the low resistance value in both LRS and HRS, thus resulting in a poor robustness against the parasitics of the CBA in which this kind of devices should be embedded.

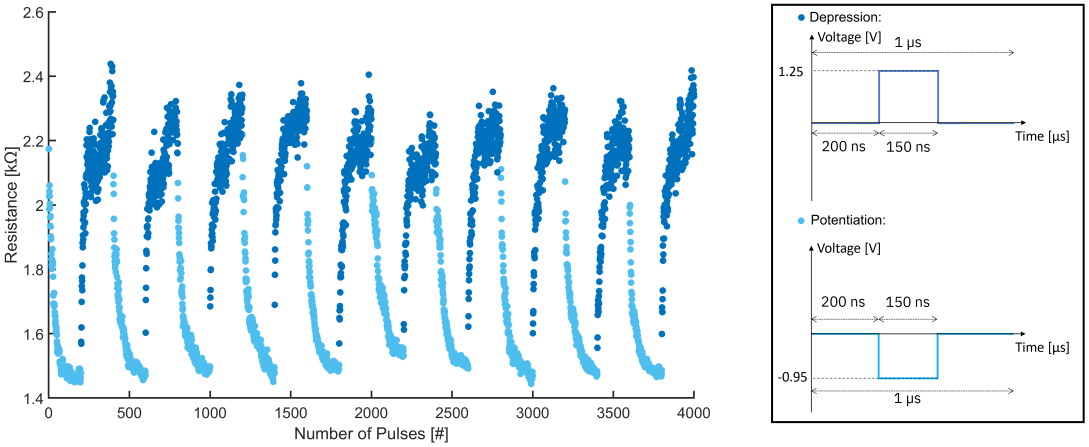


Figure 4.25: Pulsed measurement on 10 cycles and relative pulse scheme of a representative (6 × 6) µm^2 ReRAM device, in which the TaO_x has a thickness of 30 nm and a R_{sh}≈3 kΩ/□.

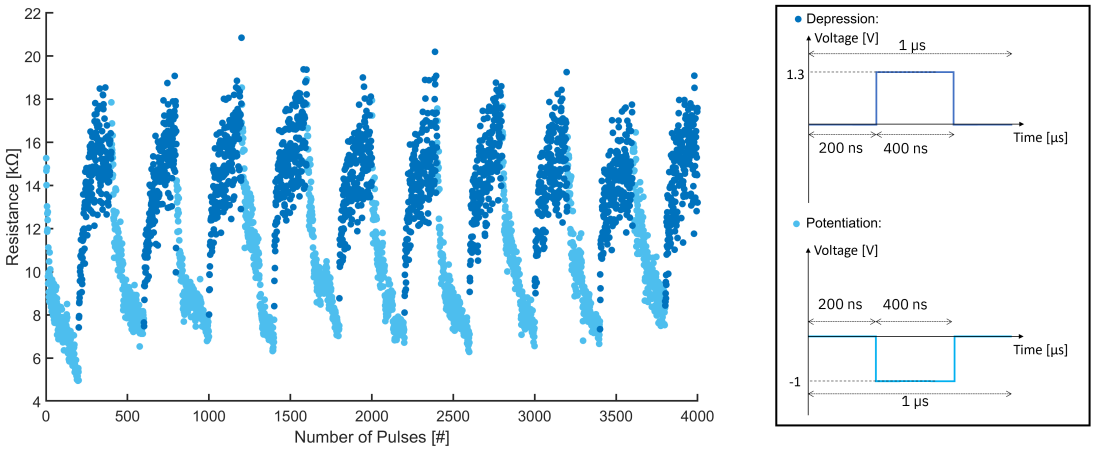


Figure 4.26: Pulsed measurement on 10 cycles and relative pulse scheme of a representative (6 × 6) µm^2 ReRAM device, in which the TaO_x has a thickness of 30 nm and a R_{sh}≈90 kΩ/□.

The resistance values can be increased by substituting the TaO_x with the more

resistive one (R_{sh}≈90 kΩ/□), and keeping the same thickness. Figure 4.26 clearly shows the increase of the resistance values with respect to the device shown in Figure 4.25. In this representative device the programming voltage amplitudes are still very low but the pulse duration for achieving a linear and stable potentiation and depression is longer. On the other hand, the noise in both potentiation and depression is larger with respect to Figure 4.25. This result is explained by the higher resistance values. Comparing the values of the LRS and HRS between Figure 4.26 and Figure 4.15, Figure 4.26 shows that the dynamic range is reducing, since even if the HRS comparable with what obtained in the quasi-static case (see Figure 4.15), the LRS is much larger.
The last device stack analysed is the one with the thinner TaO_x layer, but still characterized by an high value of resistivity. A first programming with still 400 ns pulses but lower voltages than Figure 4.26 is represented in Figure 4.27.

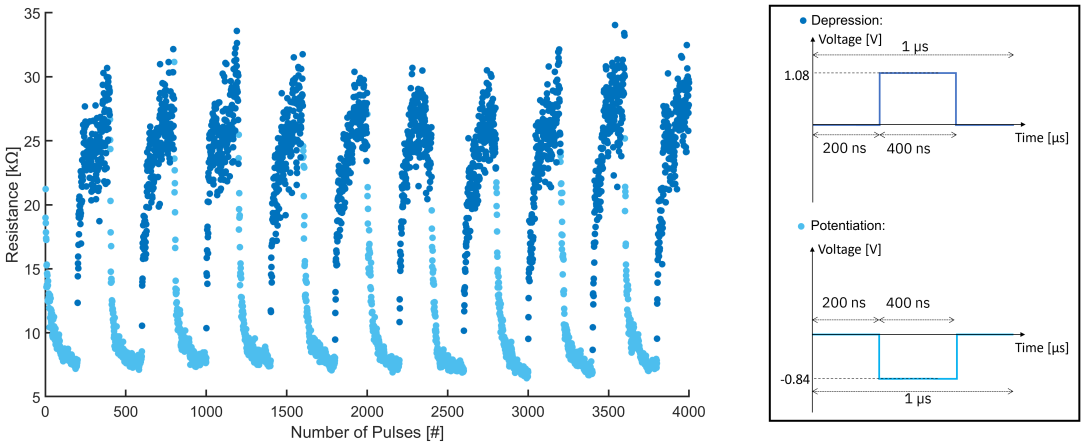


Figure 4.27: First pulsed measurement on 10 cycles and relative pulse scheme of a representative (6 × 6) µm^2 ReRAM device, in which the TaO_x has a thickness of 20 nm and a R_{sh}≈90 kΩ/□.

In Figure 4.27 shows that large memory window with less noise during both potentiation and depression can be achieved with respect to the devices with the TaO_x having the same resistivity but larger thickness. However, the number of intermediate states in Figure 4.27 is smaller than in Figure 4.26. If the same device is subjected to pulses with smaller duration and slightly larger programming voltage amplitudes, the result obtained is the one represented in Figure 4.28. As expected, Figure 4.28 shows that the larger increase of the programming voltage amplitude for potentiation with respect to the depression results in a smaller dynamic range with both LRS and HRS reduced, comparing them with Figure 4.27. On the other hand, the linearity and the number of intermediate states have improved in both transitions. These results clearly demonstrate the importance of the choice of the programming pulse parameters and that our bilayer ReRAM device can achieve a

linear and symmetry update of the resistance state.

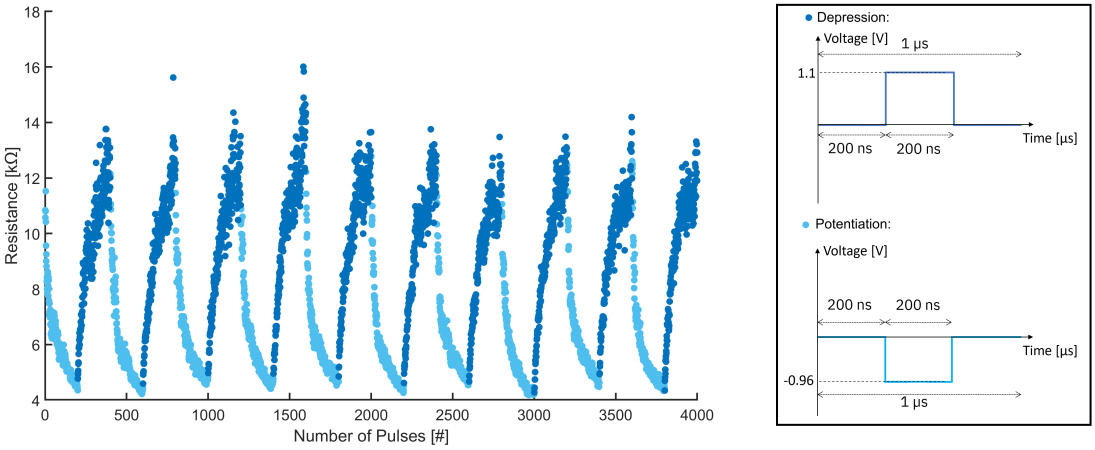


Figure 4.28: Second pulsed measurement on 10 cycles and relative pulse scheme of a representative (6 × 6) µm^2 ReRAM device, in which the TaO_x has a thickness of 20 nm and a R_{sh}≈90 kΩ/□.

4.3 Equivalent circuit model by Impedance Spectroscopy

In this section the EEC model of the proposed bilayer ReRAM technology is provided. In the attempt to decouple the contribution of the different layers, the analysis is based on modelling first reference structures in which only one oxide (TaO_x or HfO_2) is sandwiched between two metal electrodes. Then, real devices are measured, and the electrical model is built for the main resistive states in which they can be found: pristine state, post-forming, LRS and HRS. The model is built on planar devices to reduce the chuck contribution. The processing of these devices, as well as a comparison of the switching behaviour with respect to the vertical ones analysed in Section 4.2, is also provided.

4.3.1 Comparison between vertical and planar ReRAM devices

The chuck parasitics are detrimental for the Impedance Spectroscopy measurement, as it is explained in Section B.1. The need to remove this parasitic contribution leads to the processing of planar devices, whose fabrication is extensively explained in Section 4.3.2. To show that the proposed Impedance Spectroscopy model is valid also for vertical devices, a quasi-static characterization of a (40 × 40) µm^2 planar

device is provided here as an example. This device has Pt as BE and it embeds a 30 nm thick TaO_x with R_{sh}≈90 kΩ/□, even if the best performances in terms of linear and symmetric state update have been observed in the 20 nm thick one. The reason for these choices is to satisfy at the same time the requirement of Pt as BE for planar devices and the need to compare exactly the same ReRAM stack among vertical and planar configurations for a relevant modelling of the technology. The processing of planar devices is more challenging than the one of vertical devices. The reason is that the BE must be preserved during the etching of the MESA. This was not possible with the standard etching recipes already calibrated in the group, since the etching recipes are not selective towards TiN. Exploiting the fact that similar properties have been observed also in vertical devices (with 30 nm thick TaO_x) employing Pt as BE (see Section 4.2.3), and the fact that it is well known that this material is a very good etch stop because of its high inertness [52], we opted for the MESA stack of Section 4.2.3 to realize the planar devices. This choice allows for an easier processing and a meaningful modelling of the bilayer ReRAM technology.
The electroforming characteristic of the (40 × 40) µm^2 planar device and the reading of the post-forming resistance are shown in Figure 4.29.

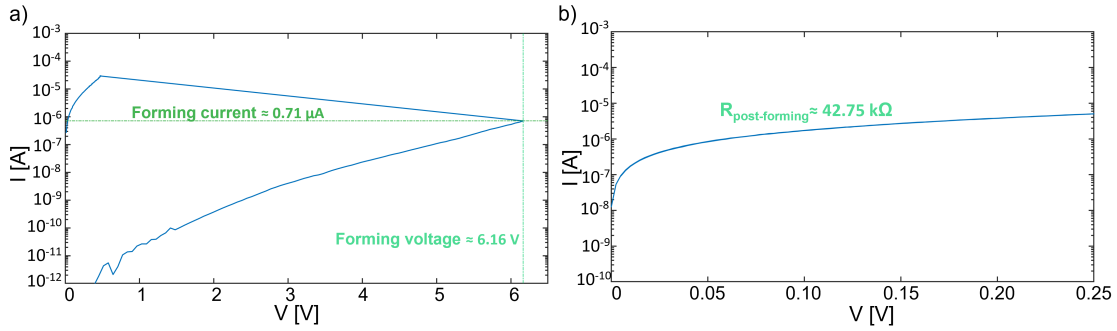


Figure 4.29: Electroforming I-V sweep for a (40 × 40) µm^2 planar device (a) and post-forming I-V sweep to read the $R_{post\text{-}forming}$ of the device (b).

Comparing the electroforming I-V sweep shown in Figure 4.29(a) with the one presented in Figure 4.17(a) for the vertical device, the current and the voltage at which the electroforming occurs are found to be larger in the planar device. This result can be attributed to the different configuration of the device and to the different electroforming conditions with which the planar devices are treated in respect to the vertical ones. Indeed, the BE in planar devices is not connected to the chuck but to another tip, as explained in Section 3.2.1. The larger current flowing into the device when the forming event occurs is a valid explanation for the smaller $R_{post\text{-}forming}$ with respect to the vertical devices. Moreover, the position of the VIA with respect to the MESA is different in the two device configurations. Figure 4.34

shows that several of the planar devices, like the one analysed, are fabricated with a VIA in the corner of the MESA to reduce the parasitics due to the overlap of the W metal pad connection and the MESA. As it is explained in Section 4.3.3, this not centred position of the VIA has a non-negligible impact on the electric field lines through which the current reaches the active part of the device, thus explaining the different electroforming parameters.
In order to protect the device from the current overshoots, it was cycled with a 10 kΩ resistor connected to the BE. The cycling behaviour of the device with and without the contribution of this series resistance is shown in Figure 4.30.

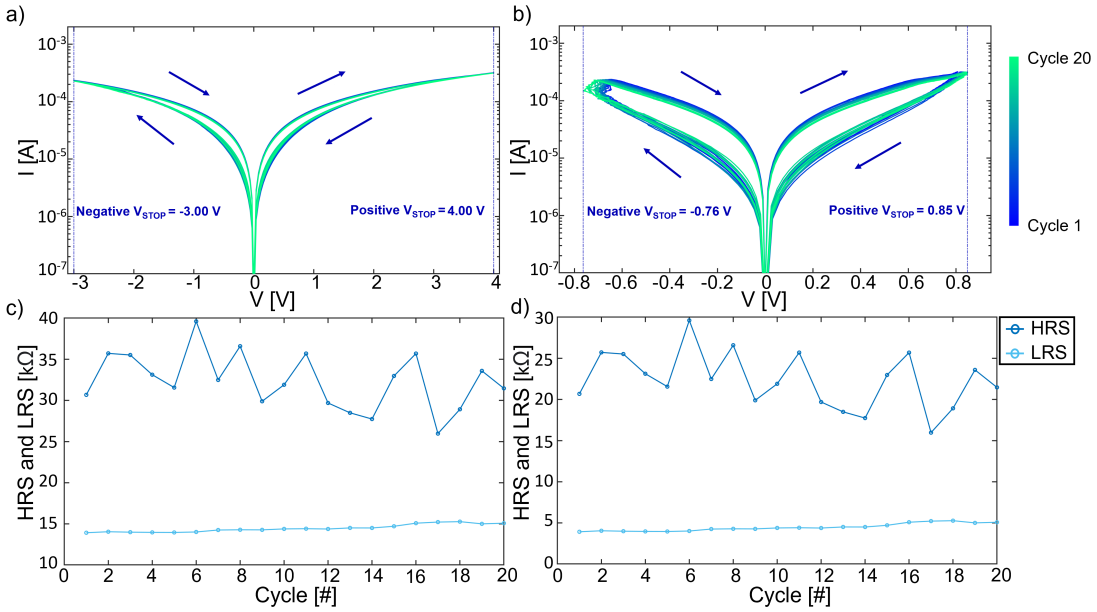


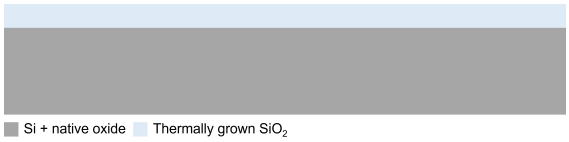
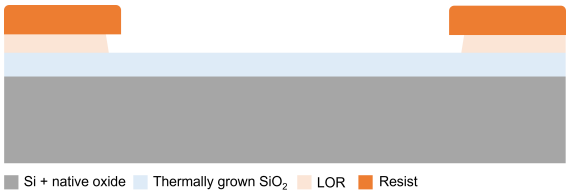
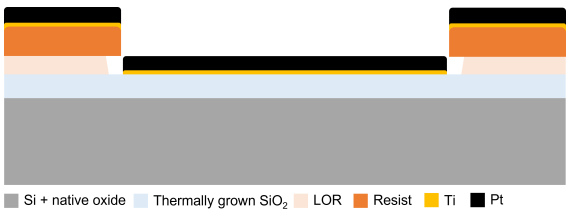
Figure 4.30: The cycling behaviour over 20 cycles for a (40 × 40) µm^2 planar device is shown with the 10 kΩ resistor connected in series to the BE (a) and without (b). Respectively, the evolution of HRS and LRS during cycling is shown in (c) and (d).

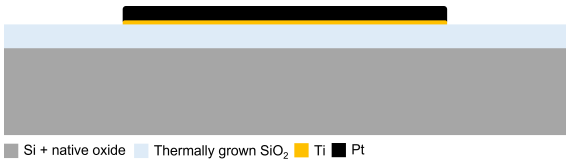
Figure 4.30(a) shows that the addition of the series resistance results in a more analogue state transition on both SET and RESET, even if this is obtained at much larger V_{STOP} due to the voltage drop on the 10 kΩ resistor. Removing the contribution of this resistor, still an analogue programming of the device is observed, as depicted in Figure 4.30(b). The cycling I-V curves show a good agreement with those of the vertical device (see Figure 4.17(c)). However, planar devices present a smaller value of both LRS and HRS than vertical ones. This result is attributed to the different $R_{post\text{-}forming}$ and V_{STOP} during programming of the two representative devices.
The differences between the two device configurations surely need a deeper understanding. Nevertheless, the analogue behaviour observed also in the case of planar

devices demonstrates that the Impedance Spectroscopy modelling of planar bilayer ReRAM can be extended to vertical ones with enough confidence.

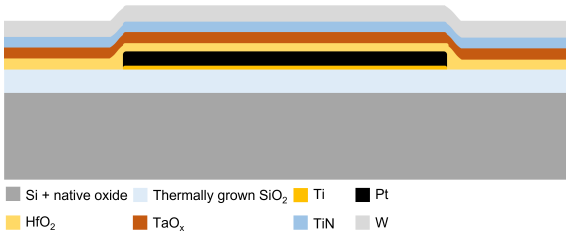
4.3.2 Design and processing of planar ReRAM devices for Impedance Spectroscopy

Planar devices allow for an easier modelling with respect to vertical ones because of the removal of the chuck contribution, as it is explained in Section B.1 with the PCB measurements. Their main fabrication steps are the following.

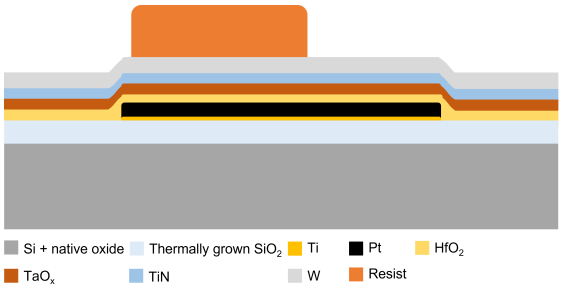
Cross-section of the sample	Explanation of the processing step
	The starting substrate is a 500 µm thick Si one on top of which 500 nm of SiO_2 are thermally grown.
	In order to have a precise definition with lift-off of the BE in the next step, a double resist is used. After a dehydration bake of five minutes at 180 °C, the two resists are spun and baked one after the other. In particular, LOR is baked at 175 °C for five minutes while the standard positive optical resist requires 110 °C for a minute. Then the resists are exposed for 1.8 s by contact photolithography using Hg lamp with an intensity of 11 mW/cm^2. After the exposure, the development is performed dipping the chip for 25 s in a solution containing one part of developer and four parts of de-ionized water, and then the chip is rinsed with de-ionized water for one minute and dried with nitrogen.
	50 nm of Pt with 3 nm of Ti as adhesion layer are deposited by e-beam evaporation. The deposition rate for the two materials is 10 nm/min for Pt and 6 nm/min for Ti.



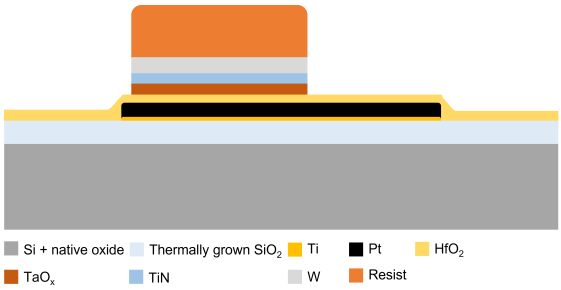
The lift-off is performed in two steps. First the top optical resist is removed leaving the chip in acetone for five minutes and in agitation with ultrasound. After rinsing with IPA and drying with nitrogen, the LOR is removed by immersing the chip in pure developer for two minutes. Then the chip is rinsed with de-ionized water and dried. Since Pt is a well-known inert material, the chip is cleaned with an oxygen plasma generated with plasma etcher at 600 W for two minutes.



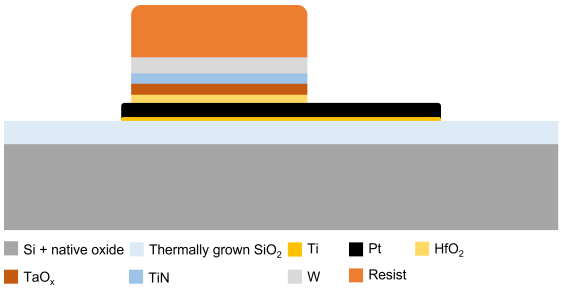
A PEALD deposition of HfO_2, as well as the sputtering of TaO_x, TiN and W are performed as already explained in the processing for vertical devices and in Section 4.1.1.



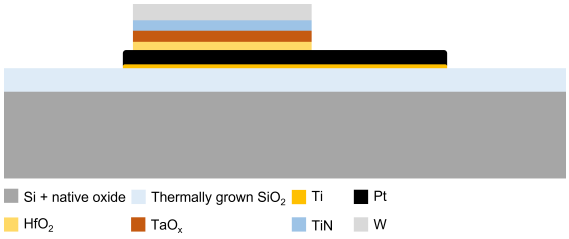
500 nm of positive polarity resist are spun on the chip after the deposition of HDMS as adhesion promoter. A baking step is then performed at 110 °C for one minute. Then the resist is exposed for 1.8 s by contact photolithography using Hg lamp with an intensity of 11 mW/cm^2. The development is done dipping the chip for 25 s in a solution containing one part of developer and four parts of de-ionized water. The chip is then rinsed with de-ionized water for one minute and dried with nitrogen.



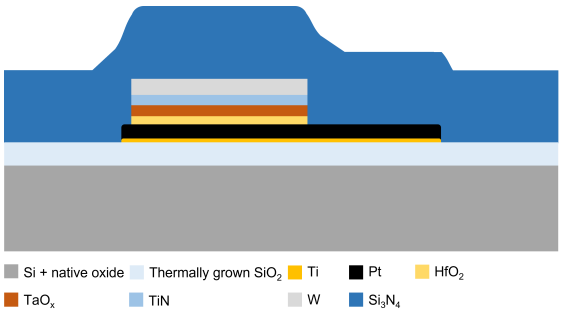
A first ICP etching step is performed in order to define the MESA in a plasma generated with an RF power of 1200 W, 18 mbar of pressure and gas mixture constituted by 10 sccm of N_2, 15 sccm of SF_6 and 15 sccm of CHF_3. The etching time is four minutes.



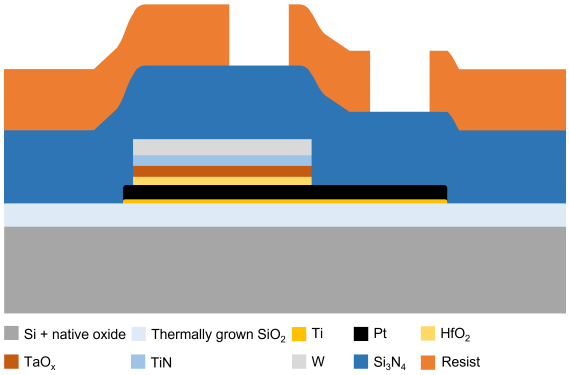
Since the previous etch step stops once the HfO_2 is encountered, a second one is required. Still it is performed with ICP but the plasma is different. The gaseous sources are Ar (5 sccm) and CF_4 (10 sccm) and the plasma is genereted with an RF power of 300 W and at a pressure of 2 mTorr. 30 s are required to get rid of the HfO_2 and due to the selectivity of the recipe towards Pt, the BE is not damaged during this process.



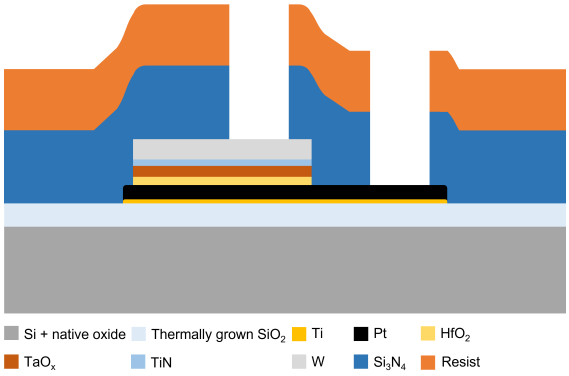
The resist is stripped dipping the chip in acetone and leaving it in agitation by ultrasound for a minute. In order to remove possible residues of resist on the chip, it is first rinsed with IPA and then further cleaned with an oxygen plasma for one and half minute. The latter cleaning is performed in the RIE as explained in Section 4.2.1.



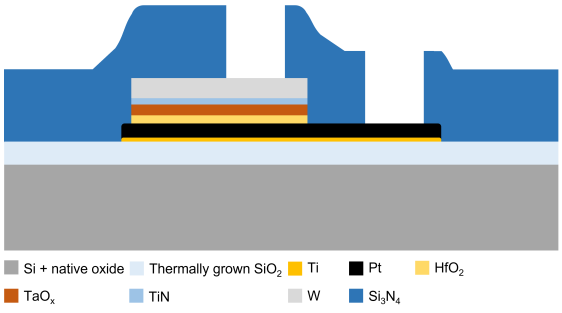
100 nm of Si_3N_4 are deposited as passivation layer using PECVD at a temperature of 300 °C. The Si_3N_4 is deposited with a deposition rate of 15.6 nm/min from a plasma of 2% of SiH_4 (120 sccm) and N_2 (1000 sccm), which is generated with an RF power of 20 W and a pressure of 1000 mTorr.



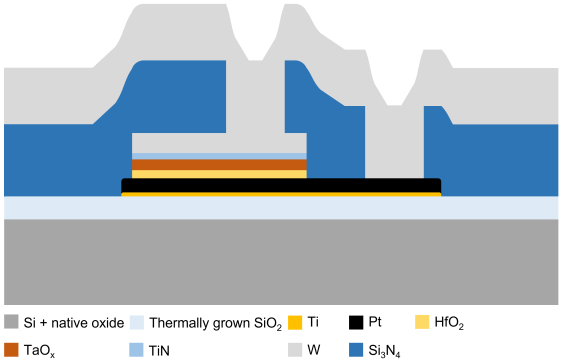
500 nm of positive polarity resist are spun on the chip after the deposition of HDMS as adhesion promoter. The baking step is then performed at 110 °C for one minute. Then the resist is exposed for 1.8 s by contact photolithography using Hg lamp with an intensity of 11 mW/cm^2. The development is done dipping the chip for 25 s in a solution containing one part of developer and four parts of de-ionized water, and then the chip is rinsed with de-ionized water for one minute and dried with nitrogen.



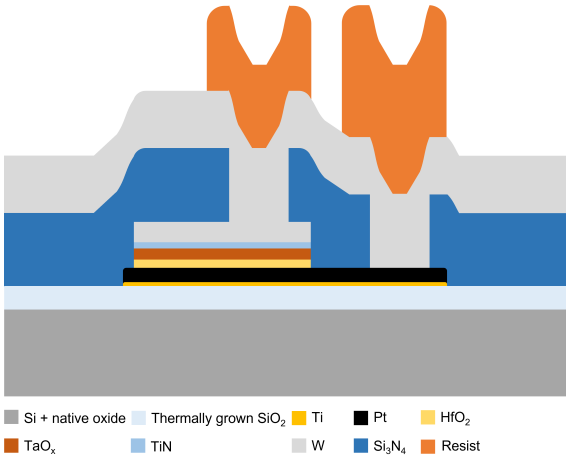
In order to open the VIAs for the connection to the electrodes an etching step with RIE is performed. The plasma is generated by applying an RF power of 100 W and a pressure of 55 mTorr at room temperature to a gas mixture composed by O_2 (5 sccm) and CHF_3 (50 sccm). The resulting etch rate of Si_3N_4 is 37.7 nm/min.



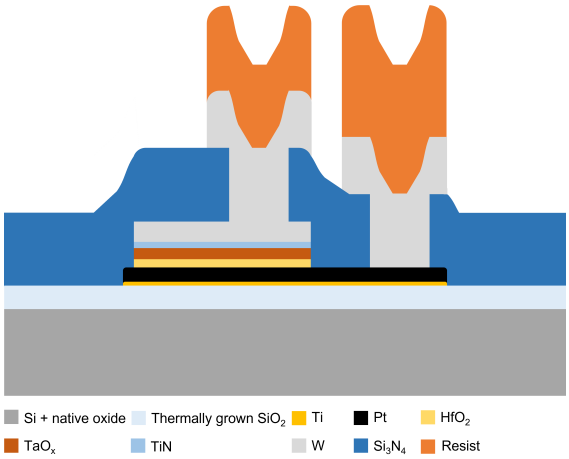
The resist is stripped dipping the chip in acetone and leaving it in agitation by ultrasound for a minute. In order to remove possible residues of resist on the chip, it is then rinsed with IPA. No oxygen treatment is required at this step since the RIE does not produce usually a large resist burning such the one obtained after the ICP process.



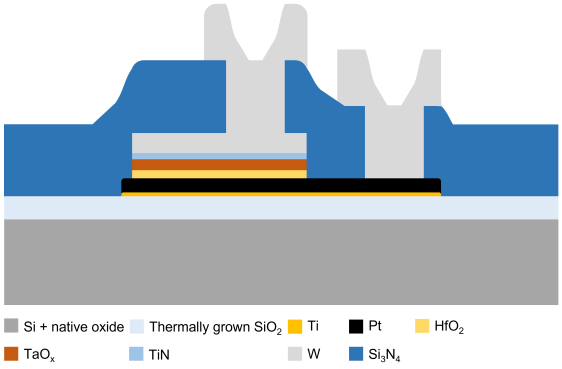
150 nm of W are deposited in order to contact the two electrodes of the device. The deposition is performed using DC magnetron sputtering with a power of 120 W, 6 µbar of pressure and 20 sccm of argon.



500 nm of positive polarity resist are spun on the chip after the deposition of HDMS as adhesion promoter. The baking step is then performed at 110 °C for one minute. Then the resist is exposed for 1.8 s by contact photolithography using Hg lamp with an intensity of 11 mW/cm^2. The development is done dipping the chip for 25 s in a solution containing one part of developer and four parts of de-ionized water, and then the chip is rinsed with de-ionized water for one minute and dried with nitrogen.

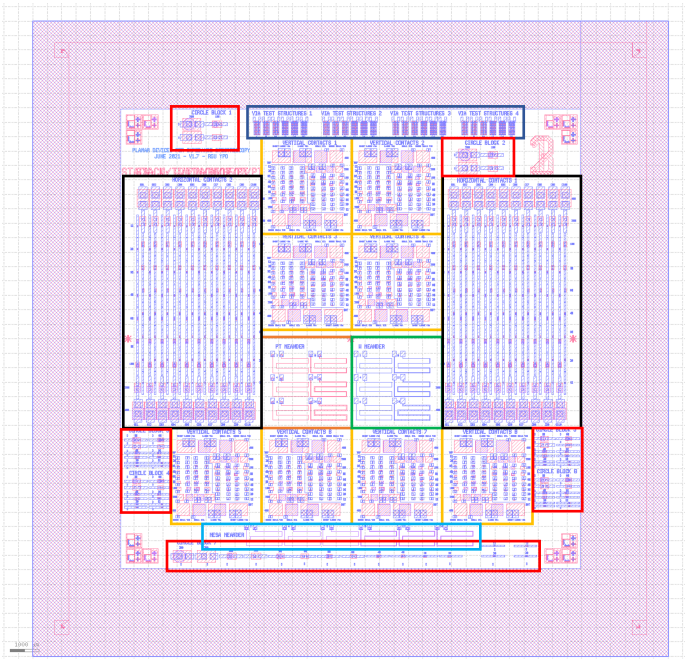


The contact pads are defined with a RIE etching of W. This step is also important to isolate the devices between them. In particular, the W is etched at a rate of 15.4 nm/min using plasma generated from N_2 (30 sccm) and SF_6 (30 sccm) generated with an RF power of 25 W, a pressure of 35 mTorr and at room temperature.



The resist is stripped dipping the chip in acetone and leaving it in agitation by ultrasound for a minute. In order to remove possible residues of resist on the chip, it is then rinsed with IPA. Still no oxygen treatment is required at this step since a RIE process was performed for the etching.

Actually the process flow reported above simply refers to one kind of planar devices (square devices with VIA in the corner) and not to the full set of test structures and devices fabricated for the Impedance Spectroscopy experiment. Indeed, the designed chip is much more complex and it consists of several different geometries.



- Circular devices and circular SHORT test structures (radius from 12 µm to 200 µm).
- Square devices with VIA in the middle of the MESA and square SHORT test structures (side of the mesa from 12 µm to 400 µm).
- Square devices with VIA in the corner of the MESA and square SHORT test structures (side of the mesa from 12 µm to 400 µm). Two dimensions of the VIA are designed for each structure.
- VIA test structures with VIA side varying from 4 µm to 32 µm in steps of 4 µm.
- Platinum meander with length of 5500 µm and width varying from 2 µm to 24 µm.
- Tungsten meander with length of 5500 µm and width varying from 2 µm to 24 µm.
- MESA meander with length of 5500 µm and width varying from 2 µm to 24 µm.

Figure 4.31: Schematic representation of the .gdsII design of devices for Impedance Spectroscopy with a summary of the different structures that it includes.

The complexity of the design is due to the need to decouple different contributions, as well as to try to reduce the effect of parasitics on the measurement. For example the pads were made at least 80 µm longer since the over-travel of the tip must be at least of 50 µm to make the contribution of the probe resistance (probe tip resistance and contact one) smaller than 1 Ω [53]. Moreover, the pads are placed at a distance of at least 16 µm between them in order to make the generated in-plane capacitance negligible during the measurement. A representation of the .gdsII with a summary of the different structures and devices is given in Figure 4.31. In particular, the proposed design allows to probe the impact on the device performance of the VIAs, through devices having different VIAs dimensions, as well as position with respect to the MESA. The impact of the device geometry could be understood by comparing the behaviour of square devices versus circular ones. Additionally, test structures in the form of meander were added to probe the R_{sh} of the metallic materials as well as the layers in the MESA. Even test structures to estimate the VIA resistance have been added. Figure 4.31 shows that the structures are repeated in several blocks, and the space on the chip is filled as much as possible by placing some structures in an interdigitated way within the same blocks.

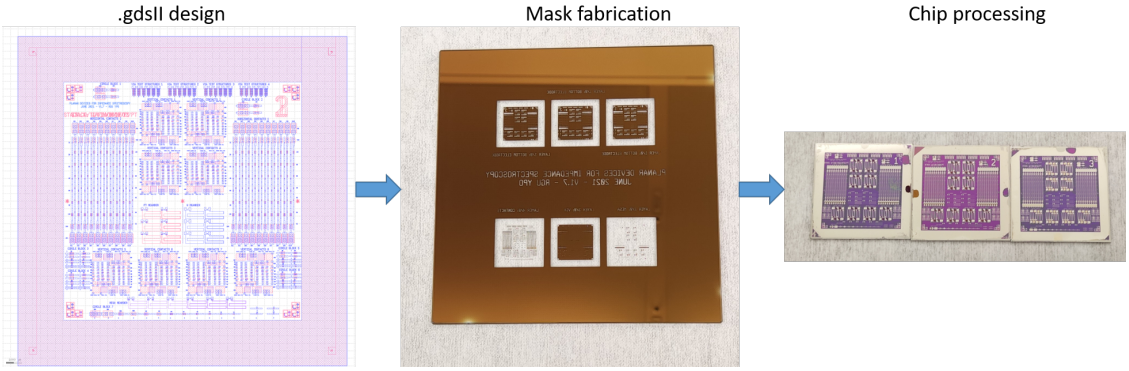


Figure 4.32: Schematic representation of the process flow of the planar devices fabrication, passing first from the design and the mask realization for then arriving to the actual processing of the chips.

Once the design was completed, the chromium mask for the contact photolithography was realized by a colleague and the chip was fabricated. Figure 4.32 shows that three chips were realized. The number one contains the planar devices, while the number two and three are made respectively with a W/TiN/TaO_x/Pt stack and W/TiN/HfO_2/Pt one in order to decouple the contribution of the different layers in the ReRAM stack. A cross-section of the stack present in each chip, as well as of the structures required for performing the SHORT correction, is depicted in Figure 4.33.

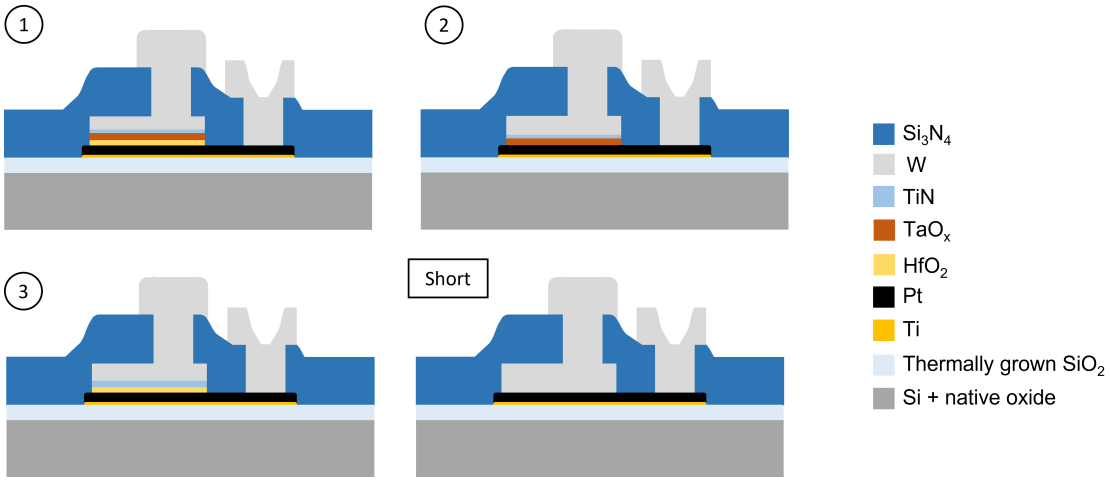


Figure 4.33: Schematic representation of the cross-section of square structures stack. They are labelled with the number of the corresponding chip in Figure 4.32, as well as of the test structure for the SHORT correction.

A microscope image of some of the blocks containing small and scaling VIAs square structures is shown in Figure 4.34, while Figure 4.35 shows how the structures look like when the tip for the measurement is approaching the pads.

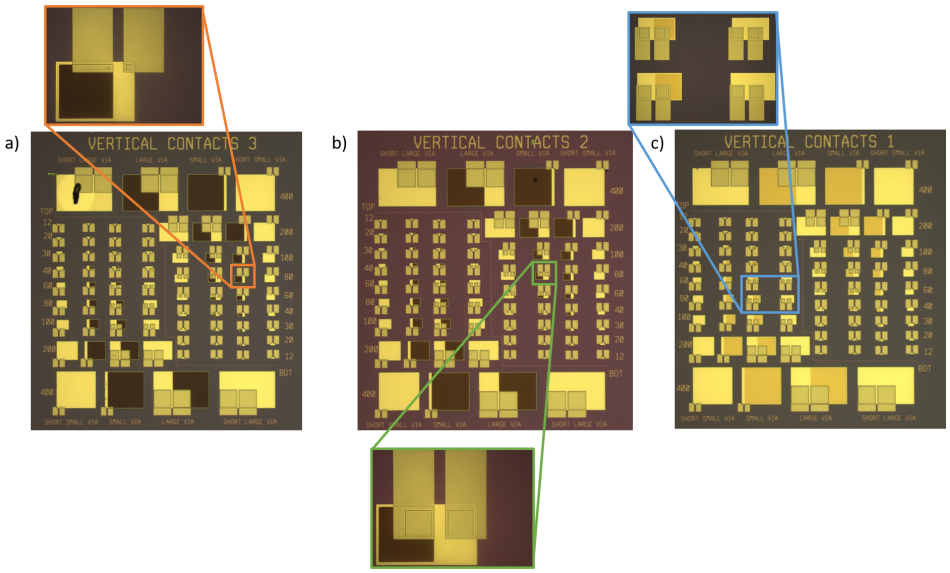


Figure 4.34: Microscope image of examples of some blocks containing square structures with VIA in the corner of the MESA for the chip having a W/TiN/TaO_x/HfO_2/Pt (a), W/TiN/TaO_x/Pt (b) and W/TiN/HfO_2/Pt (c) stack.

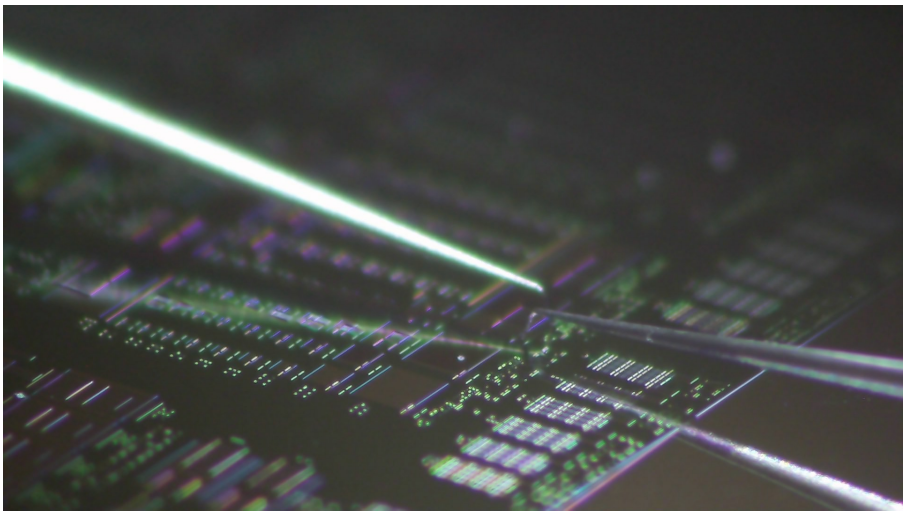


Figure 4.35: Camera view of the tested blocks right before performing the measurement.

4.3.3 Electrical equivalent circuit of TiN/TaO_x/Pt stack

The sample consisting of a nominally 30 nm thick TaO_x sandwiched between TiN and Pt can be modelled from an electrical point of view as a capacitor C in parallel to the leakage contribution R_{leak}. An equivalent series resistance R_{esr} must be added in the model for considering the contribution of the W-TiN and TaO_x-Pt interfaces. Other contributions, such as the contact resistance, the W connections and the W-TiN interface contribution are removed by the SHORT correction. This approach leads to two main overcorrections in the R_{esr} term:

- The W connection is much longer in the SHORT structure with respect to the actual device;
- The W-Pt interface contribution is counted twice in the SHORT structure while just one interface is present in the actual sample.

The SHORT structures measure a resistance of approximately 10 Ω, thus the W length overcorrection is not considered to have a huge impact on the proposed model. Moreover, being the W-Pt and W-TiN interfaces junctions between two metals, they are considered to bring the same electrical contribution to the R_{esr}. The EEC model for the TiN/TaO_x/Pt stack is hence the one shown in Figure 4.36. Figure 4.37 shows that the fit of the experimental data is not affected by the overcorrections.

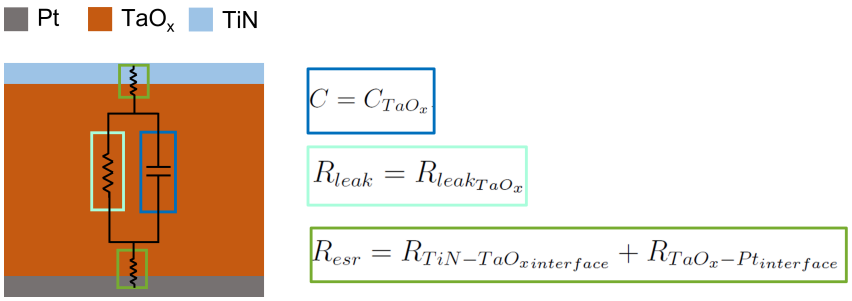


Figure 4.36: Representation of the EEC model for the TiN/TaO$_x$/Pt stack.

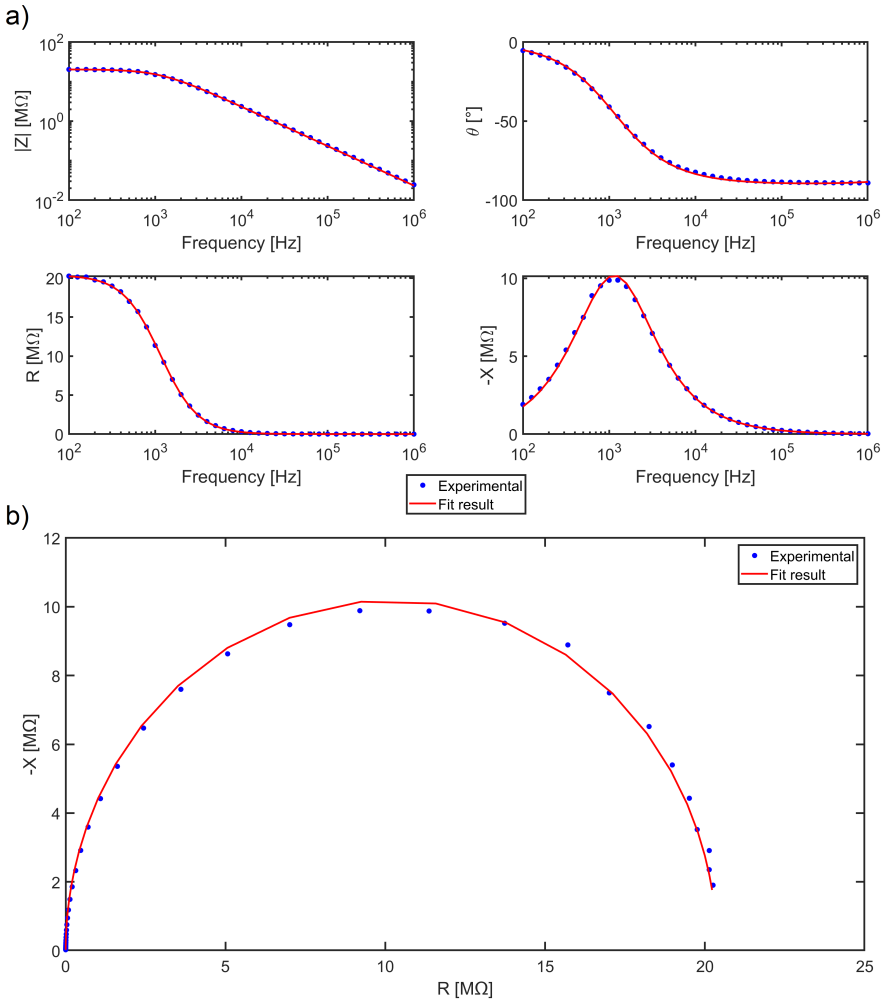


Figure 4.37: Fit result for a 144 µm^2 TiN/TaO$_x$/Pt structure with scaling VIA in the corner. It is represented as bode plots (a) and Nyquist plot (b).

The electrical parameters of the model have been studied as function of the area of the microstructure for different shapes of the MESA, as well as VIA position and dimensions. Figure 4.38 shows that the capacitance of the structure scales linearly with the area of the MESA, independently from the VIA position or the shape of both MESA and VIA. The linear scaling of the capacitive contribution with the area of the structure is expected if the layer stack is considered as a standard parallel plate capacitor, for which:

$$C = \frac{\epsilon_0 \epsilon_r A}{d} \tag{4.1}$$

where: ϵ_0 is the dielectric constant of free space, ϵ_r is the static relative permittivity of the dielectric medium; A is the area of the plates; and d is the distance between the plates of the capacitor.
Considering a thickness of the TaO_x of 28.6 nm (reported value in Table 4.1 for the layer deposited at 9 µbar) the static relative permittivity of the substoichiometric layer was computed. Figure 4.39 shows that it is constant to approximately 150, independently from the area of the structure or the VIA position and dimensions.

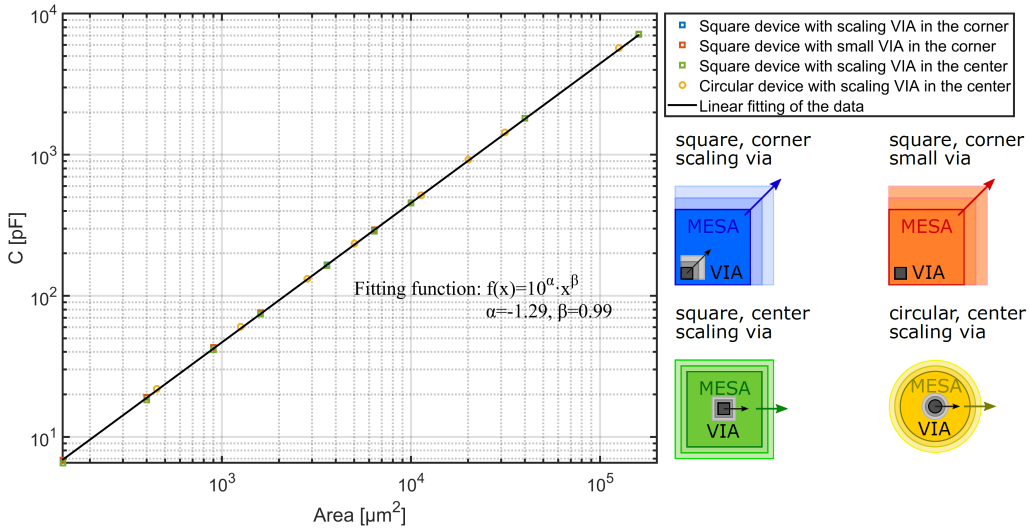


Figure 4.38: Plot of the capacitive contribution C in the model as function of the area of TiN/TaO_x/Pt microstucture.

The obtained value for the static dielectric constant is much higher than the one reported in literature, commonly between 25 and 35 [54]. Rather than considering the TaO_x as an high-κ dielectric, this high value of static relative permittivity is attributed to its more conductive behaviour with respect to a pure stoichiometric oxide. Indeed, if a pure metal is considered, its capacitance should be infinity, because of a null reactance is expected in the impedance; thus, also the ϵ_r in equation

(4.1) should be infinity. This explains the larger value of the static relativity permittivity of TaO_x with respect to the typical one for Ta_2O_5.

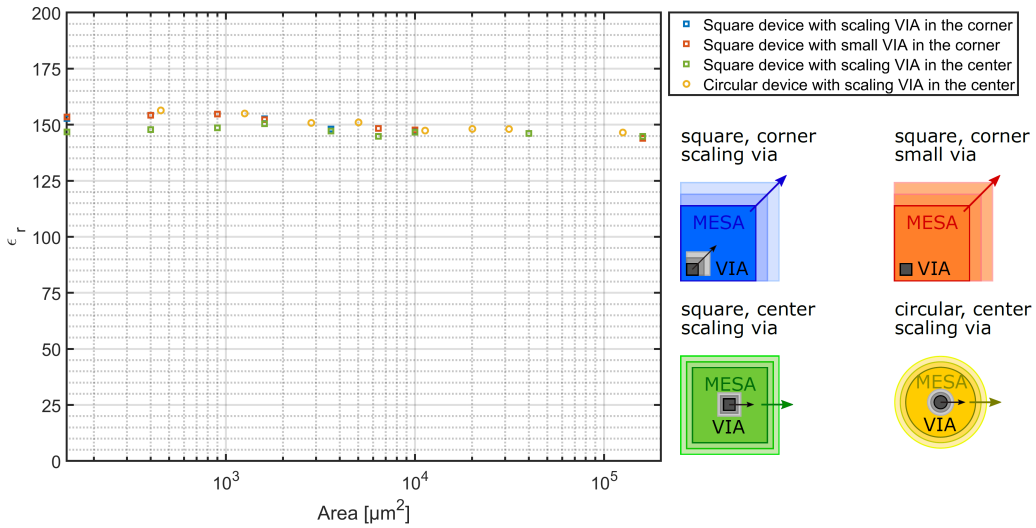


Figure 4.39: Plot of the static relative permittivity ϵ_r as function of the area of TiN/TaO_x/Pt microstucture.

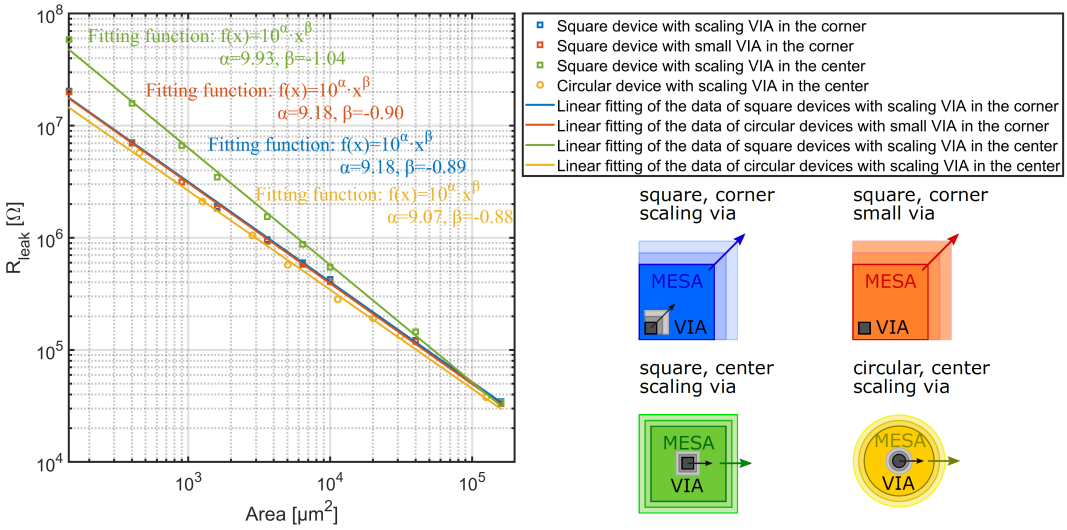


Figure 4.40: Plot of the leakage resistance R_{leak} as function of the area of TiN/TaO_x/Pt microstucture.

This hypothesis of more conductive behaviour in the TaO_x layer with respect to a stoichiometric oxide is confirmed by the value assumed by the leakage resistance

R_{leak} reported in Figure 4.40. It is two orders of magnitude lower than the value for HfO_2 structures, which is shown in Figure 4.47.
A linear inverse proportionality with the area is also observed in all the four device configurations. Regardless of the slightly different value of α in the fitting functions in Figure 4.40, the square devices with scaling VIA in the middle are characterized by a more linear scaling of R_{leak} with respect to the other device configurations. This difference could be attributed to processing variability regarding the region of the chip in which the devices are placed, as well as the distance between the microstructures. Figure 4.31 shows that the square devices with scaling VIA in the middle are placed in regions were processing steps could be not optimal (the left and right edges of the chip). Moreover, they are much closer than in the case of other device configurations. These placement conditions could result in a different etching of the MESA with respect to the circular devices or square ones with VIA in the corner. This could also explain why the change in the fitting function slope is much more evident considering smaller structures.
Impedance Spectroscopy has shown that the TaO_x layer cannot be considered as a pure resistor, which however was the objective of our material development and the result obtained from CTLM measurement shown in Section 4.1.4. This different conclusion of the two experiments can be attributed to a different regime of operation in which the TaO_x layer is measured, as well as a possible anisotropy in the current conduction if this takes place in-plane or cross-plane in the thin-film.

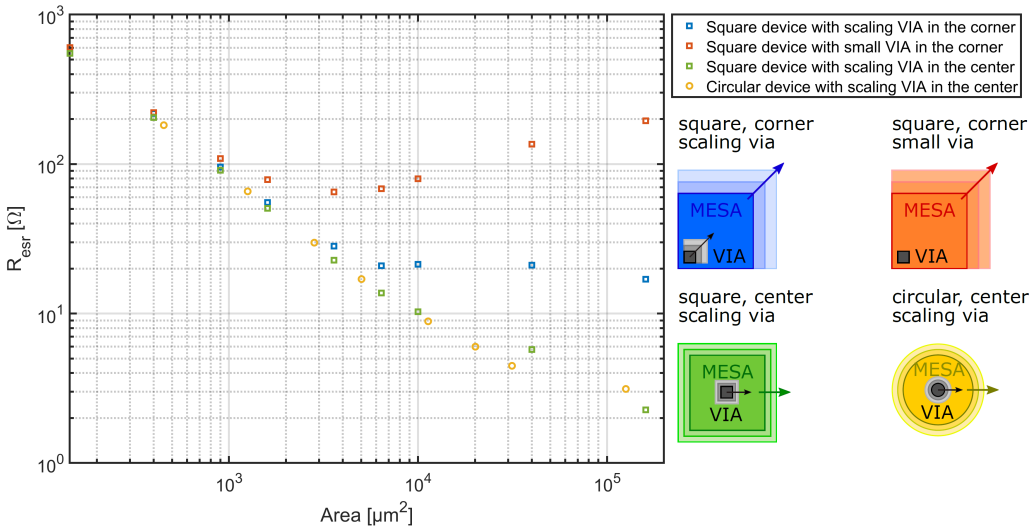


Figure 4.41: Plot of the equivalent series resistance R_{esr} as function of the area of $TiN/TaO_x/Pt$ microstucture.

Finally, the variation of the equivalent series resistance R_{esr} as function of the device area was studied. Figure 4.41 shows that the shape of the VIA and the MESA

have no influence on R_{esr}, while a clear relation with the VIA position can be noticed. In particular, having all the structures (except the smallest one for tackling lithography resolution issues) a VIA that is roughly a quarter of the MESA, a linear scaling with the area of the MESA is observed if the VIA is placed in its middle. However, if the VIA is placed in the corner, but still scaled, a saturation in the R_{esr} is observed at the device size of (80 × 80) μm^2. Not scaling the VIA dimensions leads to an inversion of the relationship and, in particular, to a linear increase of the R_{esr} with the device size starting from (60 × 60) μm^2. This phenomenon is observed in all the device stacks, as it is presented also in Section 4.3.4 and Section 4.3.5. Hence it is not attributed to any property of the TaO_x layer. Instead, it can be explained by much longer current paths for reaching the MESA edges resulting from the placement of the VIA in the corner with respect to the case in which the VIA is in the middle of the MESA. These paths become even longer if the VIA is not scaled, as it is schematically depicted in Figure 4.42.

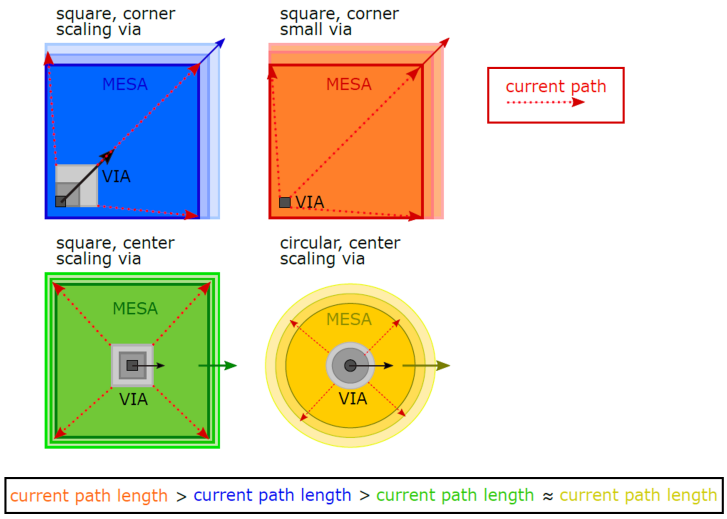


Figure 4.42: Schematic representation of the current paths in the four configurations of TiN/TaO_x/Pt to explain the R_{esr} dependence on the area of the MESA.

4.3.4 Electrical equivalent circuit of TiN/HfO_2/Pt stack

In order to decouple the HfO_2 contribution in the model of ReRAM devices, a nominally 6 nm layer has been deposited between TiN and Pt in a planar capacitor structure. In this case, like for the sample described in Section 4.3.3, the oxide layer can be modelled as a capacitive contribution C in parallel to a leakage resistor R_{leak}, while the contribution of the TiN-HfO_2 and HfO_2-Pt interfaces can be taken into account as an equivalent series resistance R_{esr} (after the application of a SHORT correction).

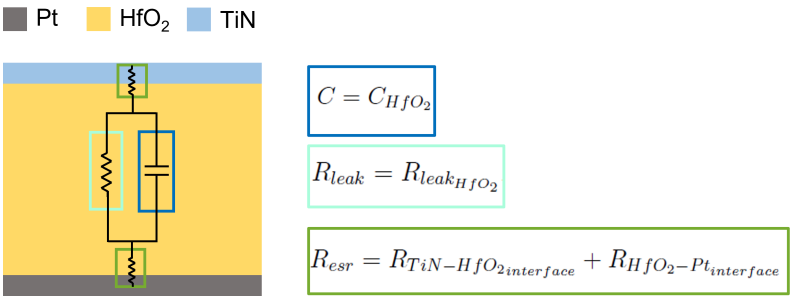


Figure 4.43: Representation of the EEC model for the TiN/HfO_2/Pt stack.

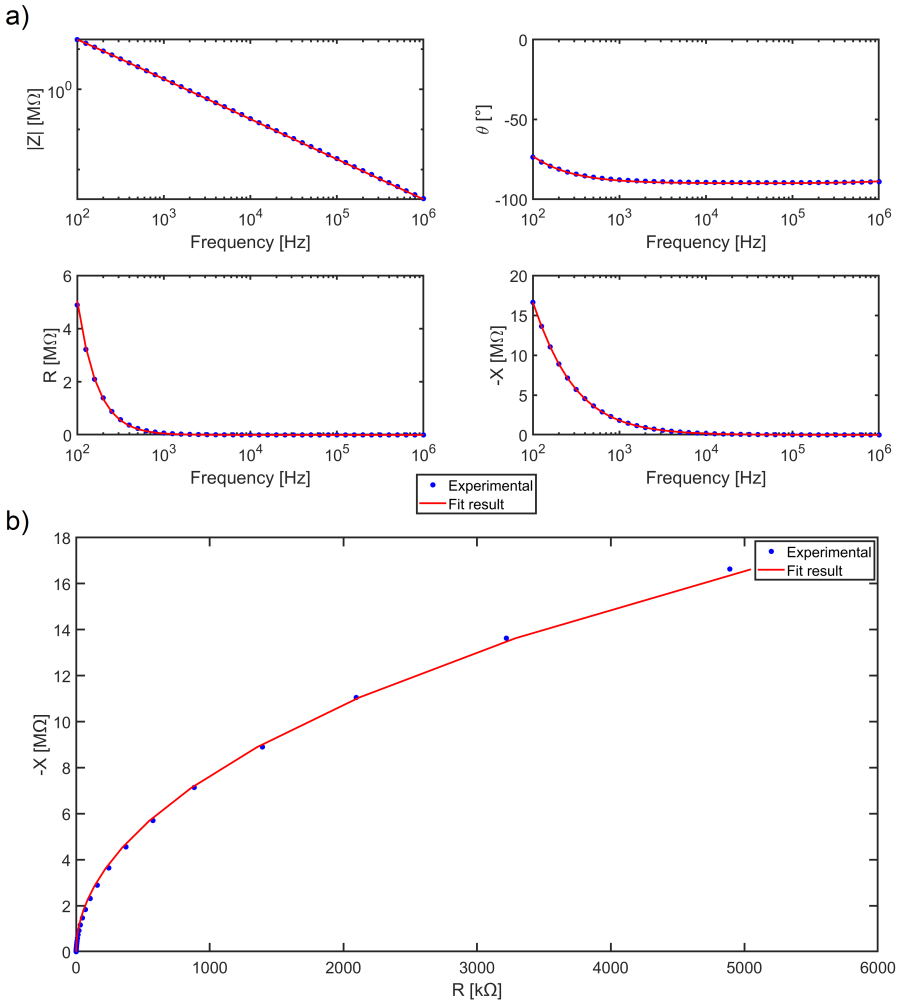


Figure 4.44: Fit result for a 3600 µm^2 TiN/HfO_2/Pt structure with small VIA in the corner. It is represented as bode plots (a) and Nyquist plot (b).

The proposed model is depicted in Figure 4.43 and it is in good agreement with the experimental data, as it is observable in Figure 4.44.
In this section an analysis of the three degrees of freedom of the model is provided while showing the reader a study of the block-to-block variability on the chip. A comparison between square devices with scaling VIA in the corner and those with a small VIA in the same position is provided in Section B.3 for completeness.

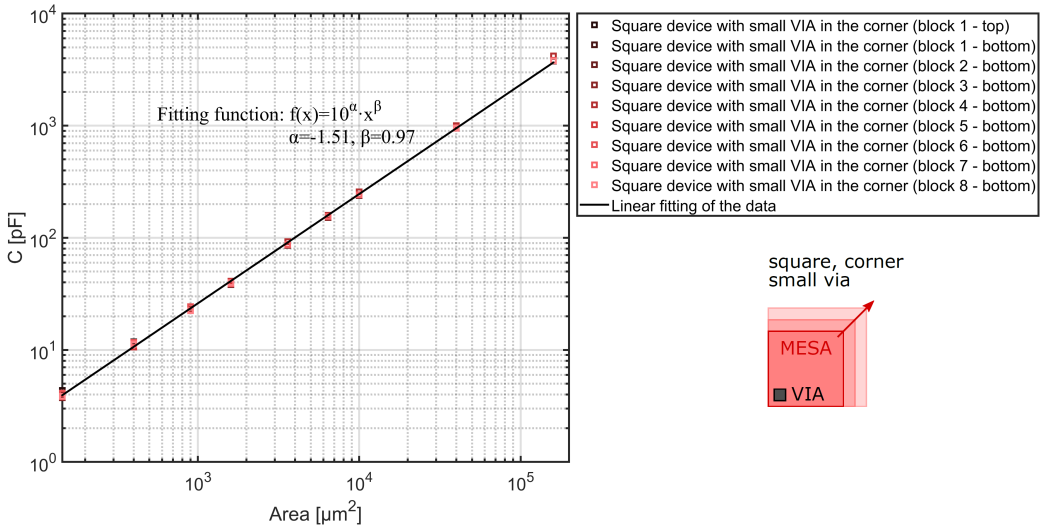


Figure 4.45: Plot of the capacitive contribution C in the model as function of the area of TiN/HfO_2/Pt microstucture with small VIA in the corner.

Figure 4.45 shows the expected linear decrease of the capacitive contribution in the sample by decreasing the device area, independently from the block considered. This linear relation between C and the device area allows to apply the general formula of a parallel plate capacitor to extract the static relative permittivity. For the calculation, a thickness of 5.7 nm has been considered for the oxide layer, in agreement with the XRR fitting values reported in Table 4.1.
Figure 4.46 shows that the static dielectric permittivity is constant to approximately 15.6 scaling down the device area to 3600 μm^2, while it starts to increase for the devices having an area between 900 μm^2 and 144 μm^2. All the computed values for ϵ_r agree with those reported in literature [55]. However, the observed increment in the static dielectric constant is unexpected since the HfO_2 is deposited on all the chip at the same time in a very uniform and conformal manner by PEALD. The possible explanation for the increment of ϵ_r in smaller structures could be found in a no more negligible contribution of fringing fields at the edges of the microstructures or in the reduced accuracy of the LCR meter, due to the smaller test current

entering in the tool. The latter hypothesis is supported by the larger spreading of the points going down to the area of 144 µm^2, as it is observable in Figure 4.46.

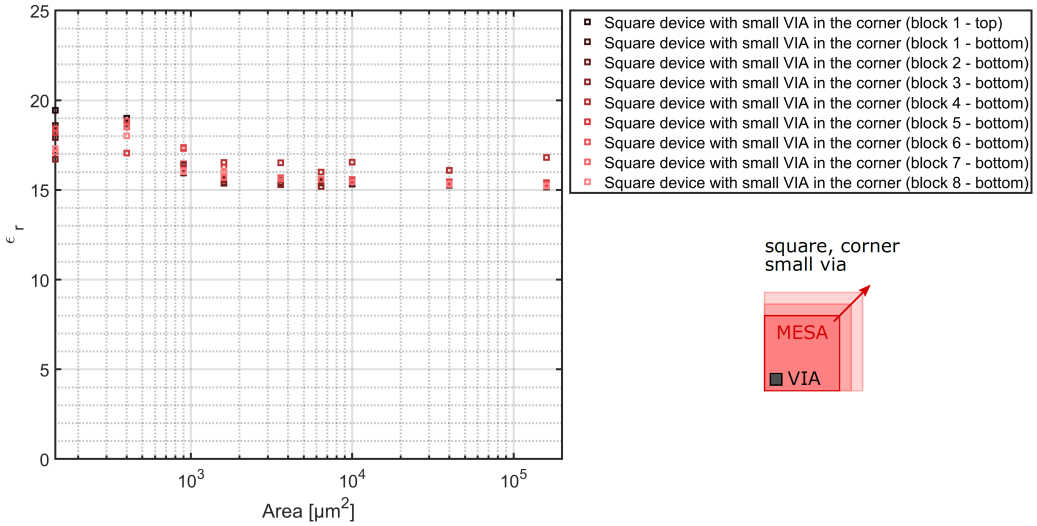


Figure 4.46: Plot of the static relative permittivity ϵ_r as function of the area of TiN/HfO_2/Pt microstucture with small VIA in the corner.

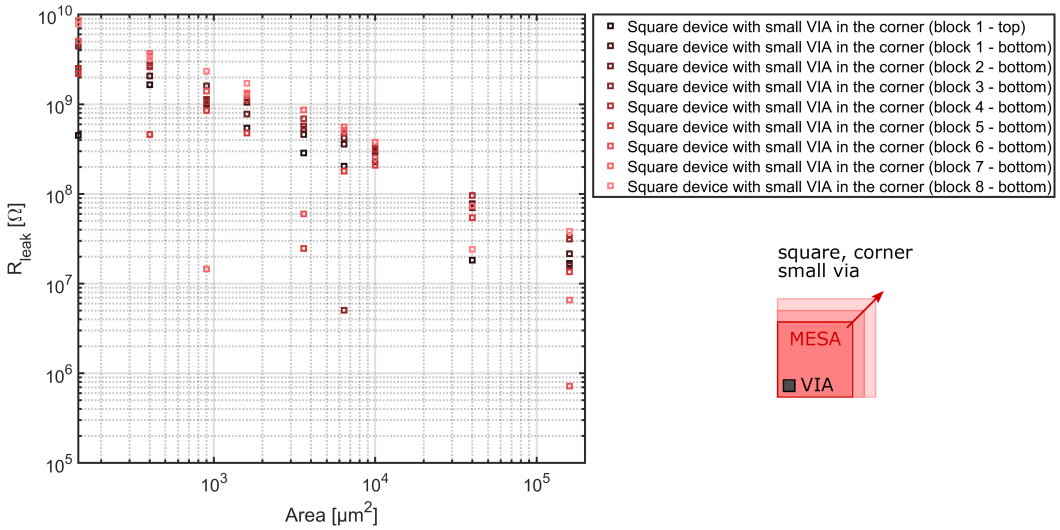


Figure 4.47: Plot of the leakage resistance R_{leak} as function of the area of TiN/HfO_2/Pt microstucture with small VIA in the corner.

As already mentioned in Section 4.3.3, the R_{leak} contribution is much larger for the

sample embedding the HfO_2 layer than the one containing the TaO_x, thus highlighting the insulating properties of HfO_2. Moreover, Figure 4.47 shows the linear scaling of this resistive parameter with the area of the device. The spreading of the points suggests a block-to-block variability with some structures that could be considered outliers with respect to the general linear trend.
Plotting the values assumed by R_{esr} as function of the MESA area (see Figure 4.48), a degradation of the linear resistance decrease with the area of the devices is observed. In particular, a linear increase of the resistive contribution, as function of the same parameter starting at 3600 µm^2, is observable, like already discussed for Figure 4.41.

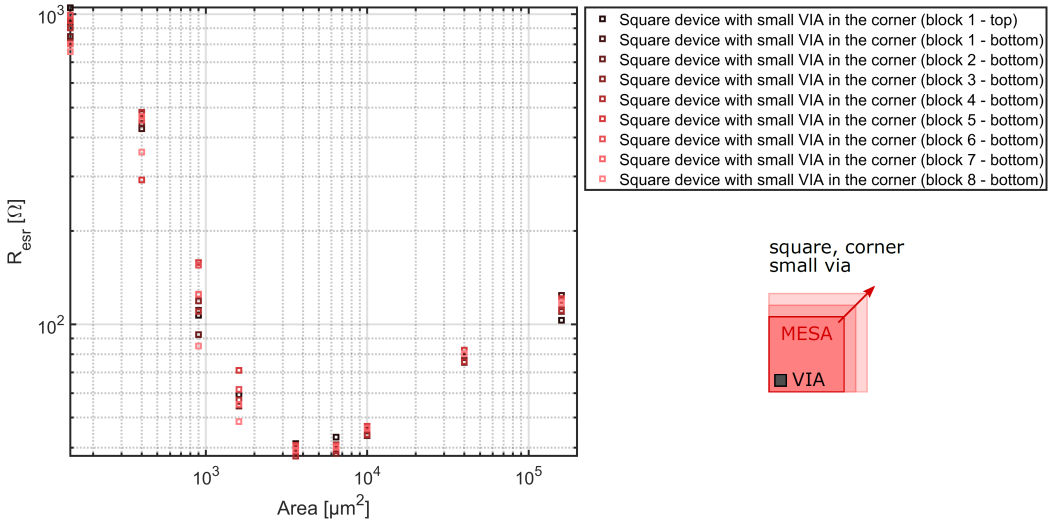


Figure 4.48: Plot of the equivalent series resistance resistance R_{esr} as function of the area of TiN/HfO_2/Pt microstucture with small VIA in the corner.

4.3.5 Electrical equivalent circuit of TaO_x/HfO_2 bilayer ReRAM devices in pristine state

We now evaluate if the two models obtained for the TaO_x and HfO_2 reference samples can be applied straightforwardly to the pristine state of the actual ReRAM stack or if a more complex model should be adopted.
Figure 4.49 shows the fitting of experimental data using a first-order model having only three variables (C, R_{leak} and R_{esr}). This model can be the one represented in Figure 4.50(a) or that in Figure 4.50(b). The only difference between the case (a) and (b) is the presence of an interface between the two oxide layers, whose role may be considered during the modelling. The other possible models explored are those in Figure 4.50(c) and (d). These higher-order models are discarded since they do

not produce a good fitting result due to the absence of more than one pole in the bode plots of the experimental data.

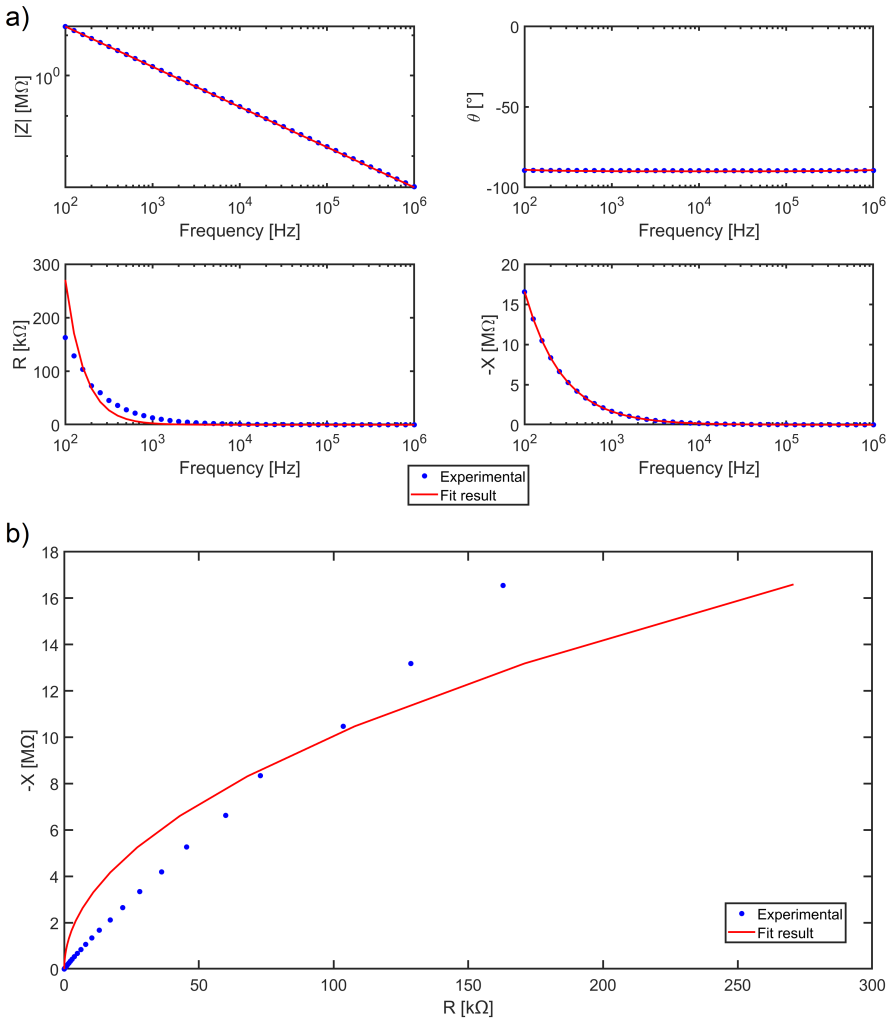


Figure 4.49: Fit result for a 6400 µm^2 TaO_x/HfO_2 bilayer ReRAM device in pristine state with scaling VIA in the corner. It is represented as bode plots (a) and Nyquist plot (b).

The observation of a first-order system is supported by the measurement performed down to the minimum frequency possible with the LCR meter (20 Hz), which is provided in Section B.5. Moreover, an equivalent surface potential at the interface between the different layers is not expected. Nevertheless, even if also in [56] a first-order system describes the pristine state of baseline ReRAM technology, the presence of other poles in the system can be masked by the dominant one. This reasoning could explain the non-perfect fit of the real part of the impedance. However, this non-perfect fit can be attributed also to the limited test current entering

in the tool due to the very high R_{leak} of the stack.
In this section the values assumed by the three parameters of the model are discussed referring to square devices with scaling VIA in the corner or small VIA in the same position.

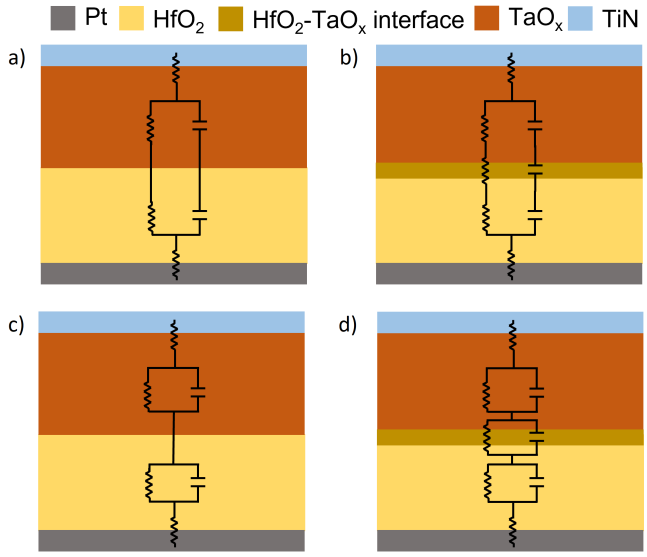


Figure 4.50: Graphical representation of the evaluated EEC models for the TaO_x/HfO_2 ReRAM device in pristine state: first-order systems (a) and (b), a second-order one (c) and a third-order one (d).

The capacitive contribution increases linearly with the area of the device, as in the case of the devices embedding only TaO_x or HfO_2. However, if the model represented in Figure 4.50(a) would be the right one, the expected capacitance value should be computed as in equation (4.2).

$$C_{expected} = (\frac{1}{C_{TaO_x}} + \frac{1}{C_{HfO_2}})^{-1} \tag{4.2}$$

Considering a single device for each area with only TaO_x or HfO_2 as oxide layers, the formula in equation (4.2) was applied and the result is represented by the orange and red squares in Figure 4.51. $C_{expected}$ is always a bit larger than the measured values. This suggests the presence of an interface layer between the TaO_x and the HfO_2. Indeed, being the interface capacitive contribution in series with the capacitance of the two bulk oxides, it reduces the overall capacitance of the system. The small capacitive impact of the interface layer on the overall capacitance is explained by its large value due to its low thickness (using equation (4.1)).

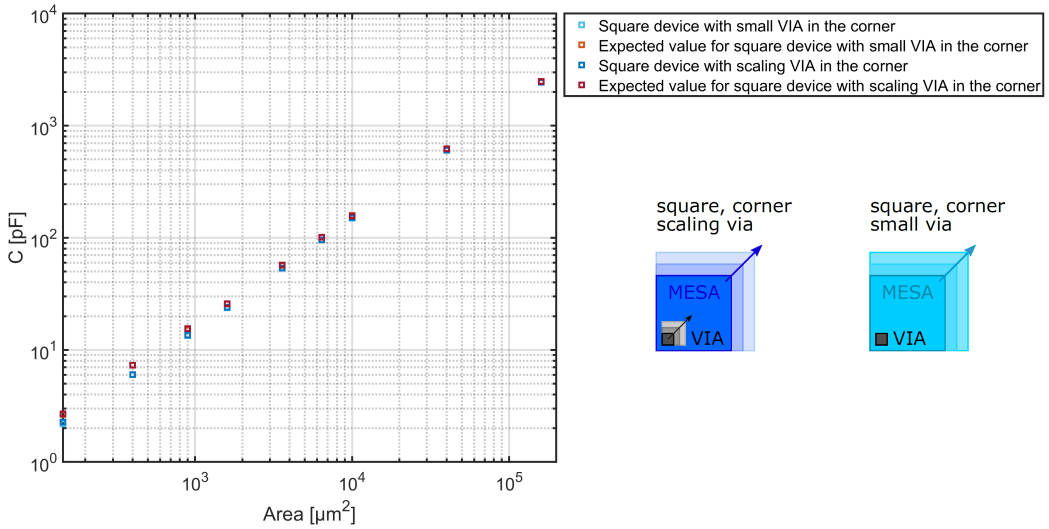


Figure 4.51: Plot of the capacitive contribution C in the model as function of the area of TaO_x/HfO_2 ReRAM device in pristine state.

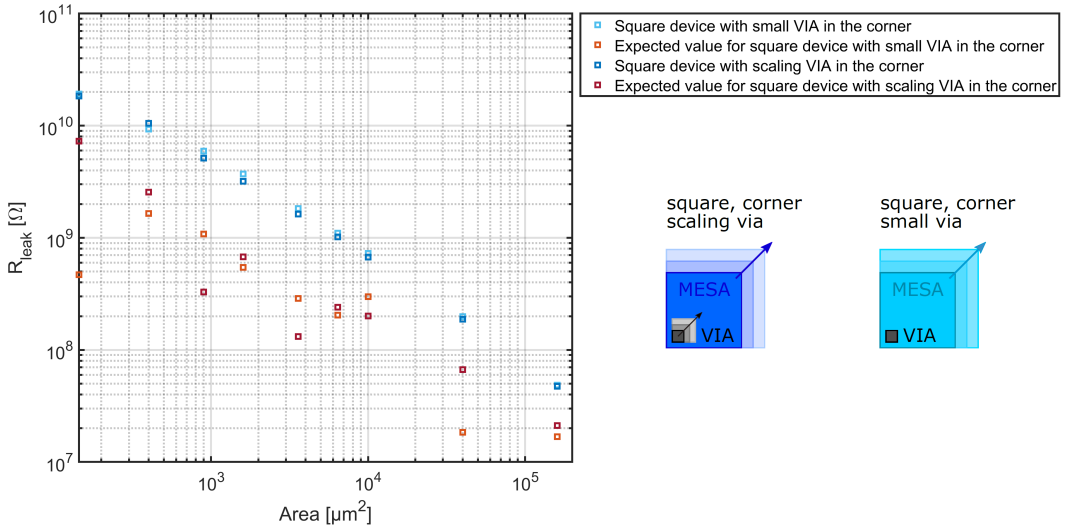


Figure 4.52: Plot of the leakage resistance R_{leak} as function of the area of TaO_x/HfO_2 ReRAM device in pristine state.

The presence of a non-negligible interface layer between TaO_x and the HfO_2 layers is confirmed by the value assumed by R_{leak}, as reported in Figure 4.52. The simple model of Figure 4.50(a) can be applied to get an insight on the expected value for the leakage resistance, which is calculated using equation (4.3) on data of devices embedding only TaO_x or HfO_2.

$$R_{leak_{expected}} = R_{leak_{TaO_x}} + R_{leak_{HfO_2}} \tag{4.3}$$

Regardless of the device-to-device variability that results in the more scattered $R_{leak_{expected}}$ points, the measured leakage resistance is always larger than the expected one. This suggests an additional resistive contribution coming from the interface layer.

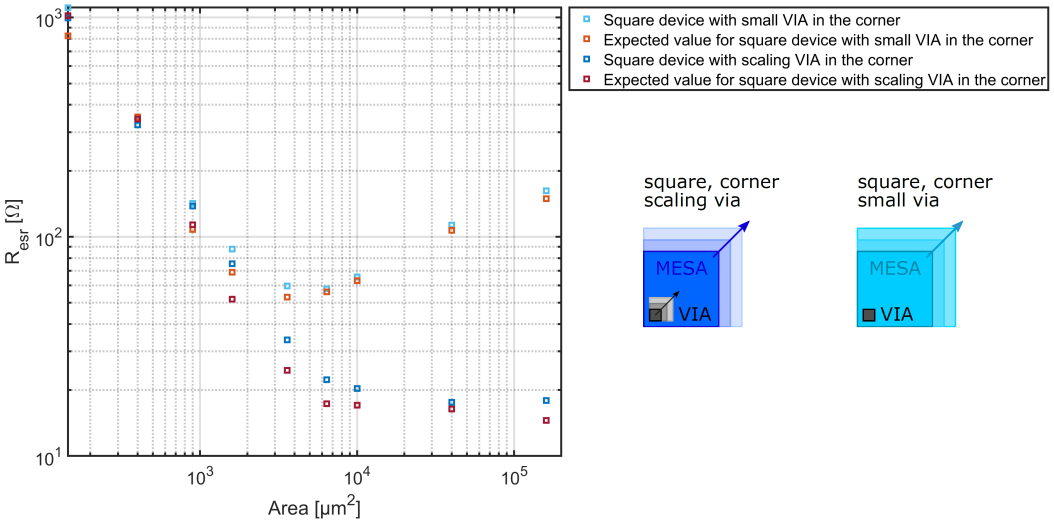


Figure 4.53: Plot of the equivalent series resistance resistance R_{esr} as function of the area of TaO_x/HfO_2 ReRAM device in pristine state.

Finally, the equivalent series resistance R_{esr} was evaluated. Figure 4.53 reports the obtained values as function of the device area. A good agreement in the R_{esr} trend is found with what already discussed in Section 4.3.3. Moreover, considering that the series resistive contribution is only due to the junctions with the metal layers and that they remain almost invariant, the fitting values resemble the expectations built using equation (4.4).

$$R_{esr_{expected}} = \frac{R_{esr_{TaO_x}} + R_{esr_{HfO_2}}}{2} \tag{4.4}$$

The comparison between the expected model proposed in Figure 4.50(a) and the fit result using the first-order system leads to the conclusion of the presence of an interface between the TaO_x and the HfO_2. Such interface adds both a resistive and a capacitive contribution to the model. In conclusion, the bilayer ReRAM devices in pristine state can be described in first approximation using a first-order system and the different contributions can be decoupled as it is presented in Figure 4.54.

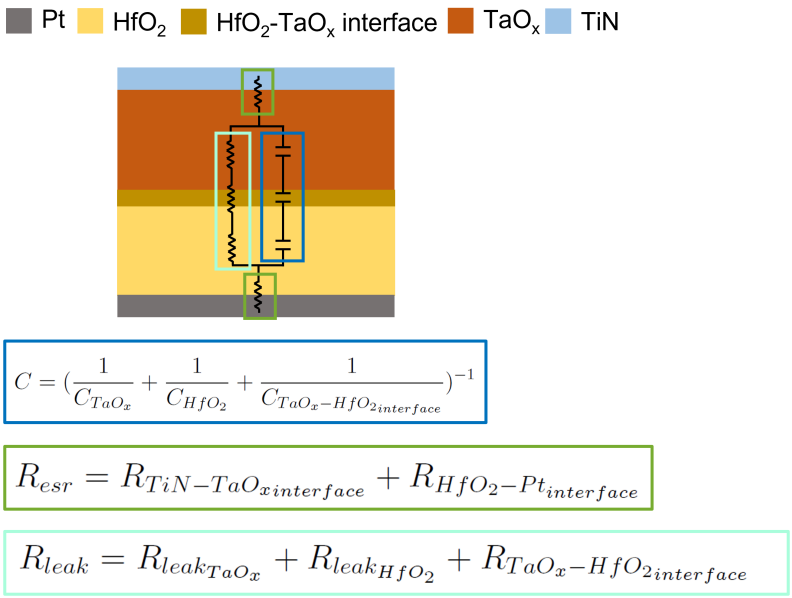


Figure 4.54: Graphical representation of the EEC model for the TaO_x/HfO_2 ReRAM device in pristine state.

4.3.6 Electrical equivalent circuit of TaO_x/HfO_2 bilayer ReRAM devices in post-forming, LRS and HRS

We now model the bilayer ReRAM devices after applying an electrical stimulus that sets them in post-forming, LRS and HRS. In this section the modelling of a (40 × 40) μm^2 device with small VIA in the corner is presented and it is exploited to justify the proposed filamentary-based model. The DC programming of the same device has been already discussed in Section 4.3.1.
In all the different resistive states, the simple first-order model with three degrees of freedom ensures a good fitting accuracy, thus well representing the system. Some examples of the fitting curves are provided in Section B.6. However, the parameters values vary depending on the device state, as it is observable in the plot reported in Figure 4.55. The values assumed by the three parameters of the model are also summarized in Table 4.3.
When the electroforming process occurs, a change in both resistive parts of the system (R_{leak} and R_{esr}) is observed while the capacitive contribution C remains almost unaltered. In particular, the R_{leak} is subjected to a strong reduction reaching a value that is even two orders of magnitude lower than the mean one (1.8 MΩ) for devices (in the same configuration and size) embedding only the TaO_x layer. This abrupt decrease can be explained by the formation of defects due to the high DC voltage applied during the electroforming step. These defects enable an easier carrier transport in the TaO_x layer and at the TaO_x-HfO_2 interface. Nevertheless, a

possible alteration of the conductivity of the layer could also be related to a different growth of the TaO_x on HfO_2 than on Pt. Additional analyses should be performed to identify if the defects are confined in the specific region close to the filament or if the difference in R_{leak} is due to changes in the material defects concentration due to the growth on a different substrate.

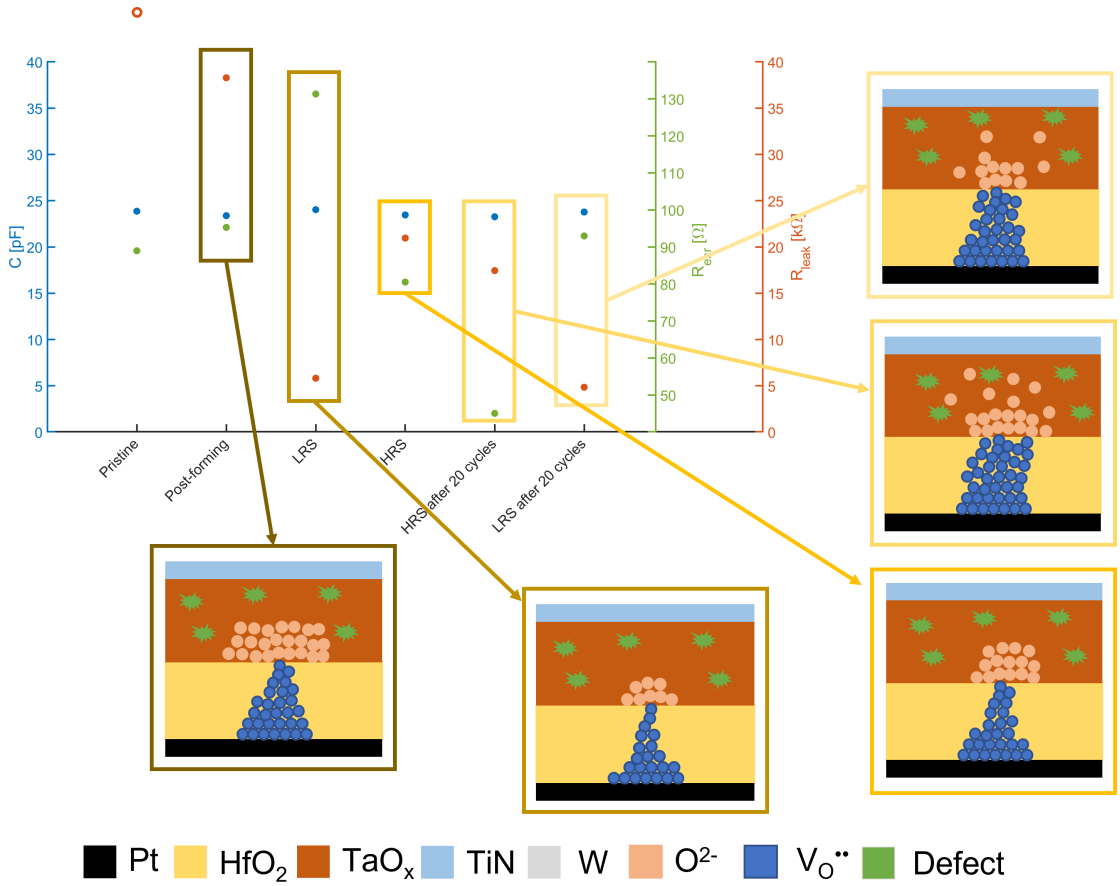


Figure 4.55: Plot of the value assumed by the EEC model parameters in the different resistive states for a (40 × 40) μm^2 TaO_x/HfO_2 ReRAM square device with small VIA in the corner. A graphical representation of the filamentary model for each state is also provided. The R_{leak} in pristine state is outside the range of the plotted values.

On the other hand, the slight increase in the R_{esr} can be associated to the formation of a filament of $V_O^{\bullet\bullet}$ in the HfO_2 layer because of the gathering of the O^{-2} by the CMO layer. Indeed, the filament formation results in an additional and local series resistive component that must be decoupled from the R_{leak}. The leakage term now counts only for the contributions of the TaO_x layer and its interface with HfO_2. The capacitive contribution brought by the HfO_2 layer cannot be removed from the

C term because of its bulk origin. The fact that C is a bulk contribution explains also why it remains unaltered after electroforming. A graphical explanation of the change of contributions to the EEC model after electroforming is provided in Figure 4.56.

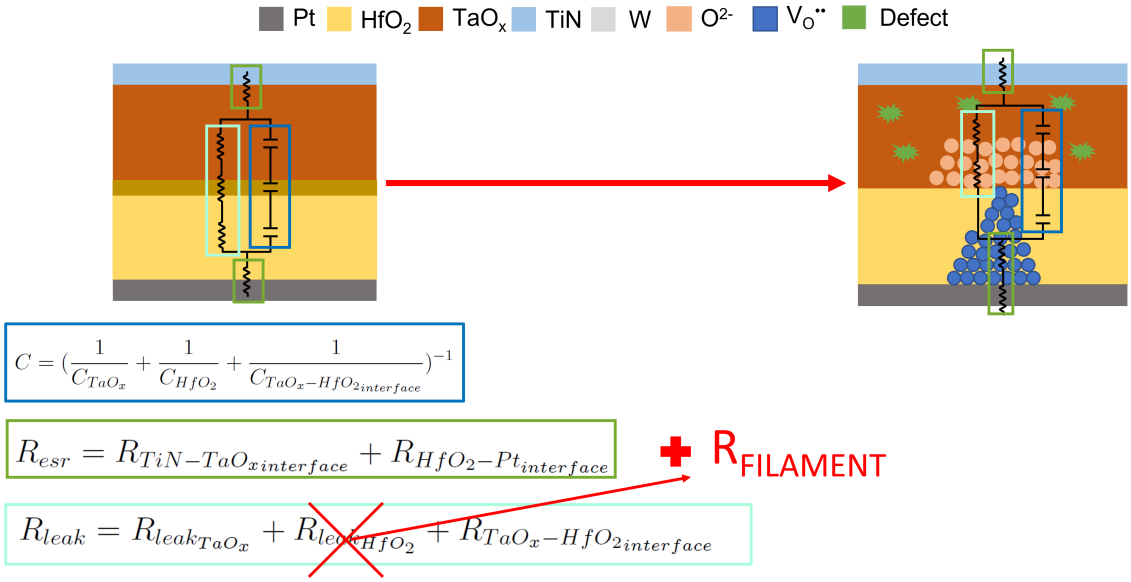


Figure 4.56: Graphical explanation of the change in the contributions to the EEC model parameters in a TaO_x/HfO_2 ReRAM device after electroforming.

Figure 4.55 shows that when the device is SET for the first time, O^{-2} are removed from the CMO layer because of the negative voltage applied at the TE, thus explaining the increment of conductivity in the whole system and reduction of R_{leak}. This process is accompanied by redox reactions occurring at the region of the filament that produce its thinning and the consequent increase of R_{esr}. However, the analogue behaviour of the memory suggests that the filament is not broken by these reactions otherwise the R_{leak} should be in the MΩ range, as it is explained in Section B.7.
The argument of the bulky nature of the capacitive term still holds upon cycling the device between LRS and HRS. However, the observation of the variation of the capacitive contribution associated to the TaO_x-HfO_2 interface could be hidden by the dominant contribution brought by the two oxide thin-films, as discussed in Section 4.3.5.
When the device is programmed in HRS, the positive voltage applied at the TE results in the re-incorporation of O^{-2} in the TaO_x layer and in the widening of the filament. Therefore, R_{leak} increases and R_{esr} is reduced. The observation of a lower value of R_{esr} in HRS with respect to the value assumed in pristine state could suggest a modulation of the barrier height at the filament-Pt interface because of

a broadening of the filament that takes place also at its root due to nucleation processes [57].
Figure 4.55 also shows how the parameters of the model in LRS and HRS evolve after 20 switching cycles. In particular, in HRS a slight decrease of the R_{leak} and a strong reduction of the R_{esr} are observed with respect to the first HRS. This results in a degradation of the switching mechanism. Indeed, the decrease in the leakage resistance can be attributed to a deeper migration of O^{-2} in the CMO layer because of the higher magnitude of the DC voltage applied during the RESET with respect to the SET. This phenomenon O^{-2} migration can make easier the carriers travelling through the current paths in the TaO_x layer, as graphically observable in Figure 4.55. This phenomenon of O^{-2} migration could also explain the strong reduction of the R_{esr}, because of a lower amount of O^{-2} available for taking part to the redox reactions. Hence, the filament gets broader while cycling. The very low value of R_{esr} suggests that the major contribution to this parameter is due to the filament.
The degradation of the switching mechanism can be noticed as well by switching the device in LRS and comparing the values assumed by the variables in the model with those in the first LRS. The reduction of the R_{esr} is due to the widening of the filament while the slight reduction of R_{leak} could be due to the opened current paths. Nevertheless, these easier current paths have a weaker impact on the resistance state if compared to what happens in HRS.
The hypothesis of O^{-2} migration in the TaO_x layer during the repetitive cycling could be the explanation of the impossibility to overcome the $R_{post\text{-}forming}$ by the resistance value in HRS, as discussed in Section 4.2.

State	C [pF]	R_{leak} [kΩ]	R_{esr} [Ω]
Pristine	23.9	$4.9 \cdot 10^6$	89.0
Post-forming	23.4	38.3	95.3
LRS	24.0	5.8	80.5
HRS	23.5	21.0	131.3
HRS after 20 cycles	23.3	17.5	45.1
LRS after 20 cycles	23.8	4.8	93.0

Table 4.3: Values assumed by the EEC model parameters in the different resistive states for a (40×40) µm^2 TaO_x/HfO_2 ReRAM square device with small VIA in the corner.

Chapter 5

Conclusions and Outlook

5.1 Conclusions

In this master thesis work substoichiometric TaO_x films have been developed to integrate them as CMO material into an HfO_2-based ReRAM stack. The resulting bilayer redox-based VCM ReRAM devices have been characterized and further optimized in a feedback loop.
A physical interpretation of the gradual switching behaviour was also proposed employing a dedicated setup for Impedance Spectroscopy characterization.
Several challenges were faced during both sputtering deposition and device characterization, leading to the following learnings.

I The reproducible DC reactive sputtering deposition of substoichiometric TaO_x films requires a dedicated Ta cleaning process and a tight control of the chamber pressure. Keeping constant and low the values of DC power and O_2/Ar flow rate ratio, the pressure can be used as a knob to finely tune the material resistivity while maintaining the fully CMOS-compatibility of the process. Increasing the pressure, a more resistive and less dense TaO_x layer is obtained. This effect is observable analyzing in the relative intensities of the (1 2 1) peak of orthorhombic Ta_2O_5 and the (4 1 0) peak of tetragonal Ta in the GIXRD profiles. All the deposited conductive TaO_x are characterized by the same deposition rate of 0.14 nm/s and a double-layer structure, with an amorphous B-TaO_x and a crystalline T-TaO_x.

II The conductivity of the TaO_x layer significantly affects the electrical properties of bilayer ReRAM devices. During electroforming, the $I_{forming}$ is lower and the $V_{forming}$ is larger, the more resistive it is the TaO_x layer, suggesting that it acts as an in-built compliance. Consequently, the $R_{post\text{-}forming}$ is larger for a bilayer ReRAM device embedding a more resistive TaO_x.
Bilayer ReRAMs show remarkably analogue SET and RESET transitions and a self-compliance during the SET in both cases of low and high conductivity

TaO_x embedded in the stack. The LRS and HRS values are stable over at least 20 cycles. However, devices with a more resistive TaO_x show a larger On/Off ratio. This observation could be explained by the lower thermal conductivity in the more resistive CMO which results in a slower heat dissipation in the layer stack and in a deeper RESET.

III The thickness of the CMO layer considerably impacts its electrical and thermal properties thus also affecting the switching behaviour of ReRAM devices. If ReRAMs embedding a 20 or 30 nm thick TaO_x show analogue switching, this is not true for devices having a 40 nm thick TaO_x. In this case, the typical abrupt and stochastic SET transition of baseline technology is observed. This finding can be associated to the slower heat dissipation in the stack that accelerates the CF formation. The abruptness of the SET transition gets reduced increasing the CMO thickness to 60 nm, even if multiple steps are observed. The reason for these multiple steps during the SET process can be found in the formation of multiple weak CFs.
According to quasi-static measurements, it is possible to conclude that the optimal thickness of TaO_x is between 20 and 30 nm.

IV Quasi-static measurements performed on bilayer ReRAM devices with different area show that the $I_{forming}$ increases linearly with the device area whereas an opposite trend is found for the $V_{forming}$. These two effects are correlated by the probability of defects generation that is higher in devices with bigger area, leading to the presence of more leakage paths. Regardless of the electroforming process, no area dependence in observed neither in $R_{post\text{-}forming}$ or On/Off ratio. The independence of the LRS and HRS from the device area indicates the filamentary nature of the switching.

V The insertion of TaO_x in the ReRAM stack results in an analogue behaviour of the memory cell during both potentiation and depression at almost the same V_{STOP} of the quasi-static measurements. However, ReRAM devices embedding the most conductive TaO_x developed results in low values for LRS and HRS, thus posing some problems in their crossbar integration. Embedding in the stack a more resistive CMO with the same thickness results in an increase of resistance values but also of the programming noise. The latter can be mitigated embedding in the stack a 20 nm thick TaO_x. In this case a remarkable linear, symmetric, and analogue weight update can be reached by sacrificing the dynamic range to roughly a factor 3.

VI The Impedance Spectroscopy model of TaO_x is described by the parallel between a leakage contribution R_{leak} and a capacitive one C. The higher value of the extracted static relative permittivity with respect to previous studies confirms the conductive nature of the TaO_x material. The capacitive contribution in TaO_x apparently contradicts the conclusion of a pure resistor coming from

CTLM measurements. On the other hand, it may suggest the anisotropy in the current conduction or a different regime of operation in the two experiments.

VII The substitution of TiN with Pt as BE in the bilayer ReRAMs, as well as the variation from vertical to planar configuration, does not impact the graduality and the switching polarity in the devices, thus suggesting the key role played by the TaO_x in the mechanism.

VIII In pristine state, bilayer ReRAM devices can be approximated by a first-order system C//R_{leak}+R_{esr}. The C and R_{leak} values suggest the presence of an interface layer between TaO_x and HfO_2, whose presence is also confirmed by a TEM cross-section. After electroforming, R_{leak} gets below the value observed in the reference TaO_x sample, thus suggesting a strong modification of this layer with an increased defect density or a different growth of TaO_x on HfO_2 or Pt. The increase of R_{esr} is attributed to the filament formation. The observed trends for R_{leak} and R_{esr} suggest that during SET O^{-2} are removed from the CMO and they recombine with $V_O^{\bullet\bullet}$ in the filament without breaking it. During RESET, O^{-2} are reincorporated in the CMO, leaving behind $V_O^{\bullet\bullet}$ that make the CF wider. The main contribution determining the resistance state is the oxygen content at the TaO_x-HfO_2 interface which modulates the R_{leak}. C remains unaltered during switching suggesting a bulk origin of this parameter.
Cycling the device 20 times, the values assumed by R_{leak} and R_{esr} display a degradation of the switching mechanism. It can be attributed to a deeper migration of O^{-2} in the CMO with reduction of the number of redox-reaction occurring at the filament region.

5.2 Outlook

The presented master thesis work highlights the outstanding potentialities of ReRAM devices based on TaO_x/HfO_2 bilayer for realizing novel computing architectures based on bioinspired neuromorphic computing. The promising results obtained in this master thesis work strongly suggest to further proceed in the research of this powerful, as well as simple, improvement methodology of baseline ReRAM.
The further improvement of the proposed CMO must consider the assessment of its stoichiometry with advanced techniques such as XPS, thus ensuring an even finer tuning of the oxygen content in the layer and hence of the switching behaviour in the devices [4]. The importance of the TaO_x thermal properties, mentioned in the present work as well as in [18], [19] and [20], clearly motivates for the determination of the thermal conductivity of this layer. This material property can be obtained through the techniques described in [62], thus allowing for an additional degree of freedom in the improvement of device performances. Moreover, the apparently contradictory electrical equivalent model obtained from CTLM experiment

and Impedance Spectroscopy one suggests the requirement of a deeper understanding of the conduction mechanisms in the substoichiometric layer. More electrical characterization and temperature-dependent measurements could be performed to acquire this knowledge [63]. The control of the CMO properties is fundamental for further reducing the intra-device and inter-device variability.
The achieved linear, symmetric, and analogue weight update during pulsed programming of the proposed bilayer ReRAM pushes for its integration into CBA. This step hides several challenges that must be addressed but that go beyond the scope of this work. They include processing challenges mainly regarding the scaling-down of the device area to the nm^2 range while still preserving the bilayer ReRAM properties and stability. Moreover, related to the array size, issues associated to device yield and array parasitics that can degrade the system performance may show up [2]. Finally, the power consumption of a CBA embedding the proposed bilayer ReRAM devices must be considered, with MΩ range resistances that would be desirable in future neuromorphic systems [9].
The interpretation of the switching behaviour in the proposed bilayer ReRAM by means of Impedance Spectroscopy is surely exciting. However, further investigations are required to establish the proposed physical model. A deeper insight in the complex nanoscale phenomena can be obtained thorough temperature-dependent Impedance Spectroscopy [64]. More advanced techniques, such as in situ Scanning TEM (STEM) and Electron Energy-Loss Spectroscopy (EELS)[65] could also be beneficial, allowing to assess the redox changes during the switching process. Scalpel Scanning Probe Microscopy (SPM) [12] could be used to locate and characterize the OEL region as well as the CF.

Appendix A

Supplementary information on material characterization

A.1 GIXRD and XRR analysis on a tatalum sample

A nominally 10 nm thick Ta sample was deposited to have a reference that can help during the identification of the crystalline phase of TaO_x. It was deposited by DC magnetron sputtering on a standard 500 µm Si substrate (with native SiO_2 always present on top). The sputtering process was performed after five minutes of target cleaning and three of pre-sputtering at the following conditions:

- DC power: 300 W.
- Chamber pressure: 4 µbar.
- Argon flow rate: 20 sccm.
- Sputtering time: 100 s.

A GIXRD scan ($\omega - 2\Theta$) was performed at ω=1.30° and varying 2Θ between 20° and 50° in steps of 0.02°. The time per step was fixed to 2 s. The background subtracted raw data of the GIXRD pattern is shown in Figure A.1.
In Figure A.1 is shown as the Ta sample is characterized by a tetragonal β-Ta crystallographic phase, in agreement with [46]. The observed peaks, which are labelled in Figure A.1, are determined by the Gaussian deconvolution of the GIXRD profile, as reported in Figure A.2.

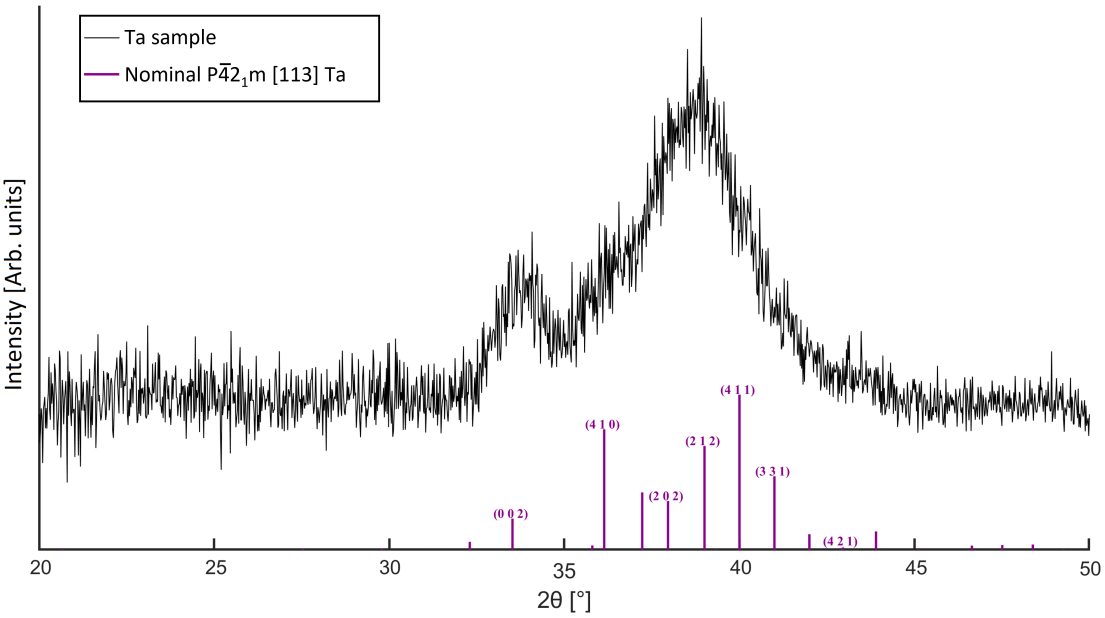


Figure A.1: Background subtracted GIXRD scan at ω=1.3° of a Ta sample deposited on a standard Si substrate. The nominal peak position of tetragonal β-Ta are reported from the reference database [34], with the correspondent Hermann-Mauguin space group. Only observed peaks are indexed.

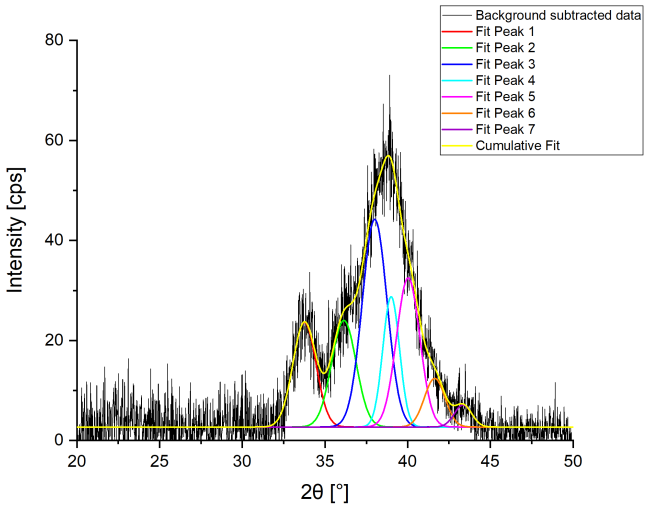


Figure A.2: Gaussian deconvolution of the GIXRD peaks in the pattern of the Ta sample deposited on Si substrate.

An XRR scan was also performed to check the thickness and the density of the deposited film. The scan was performed varying 2Θ in the range 0° to 10° in steps of 0.01°. The integration time in each step was fixed to 1 s. The raw data and the simulated XRR pattern are reported in Figure A.3, while the fitting parameters are summarized in Table A.1.

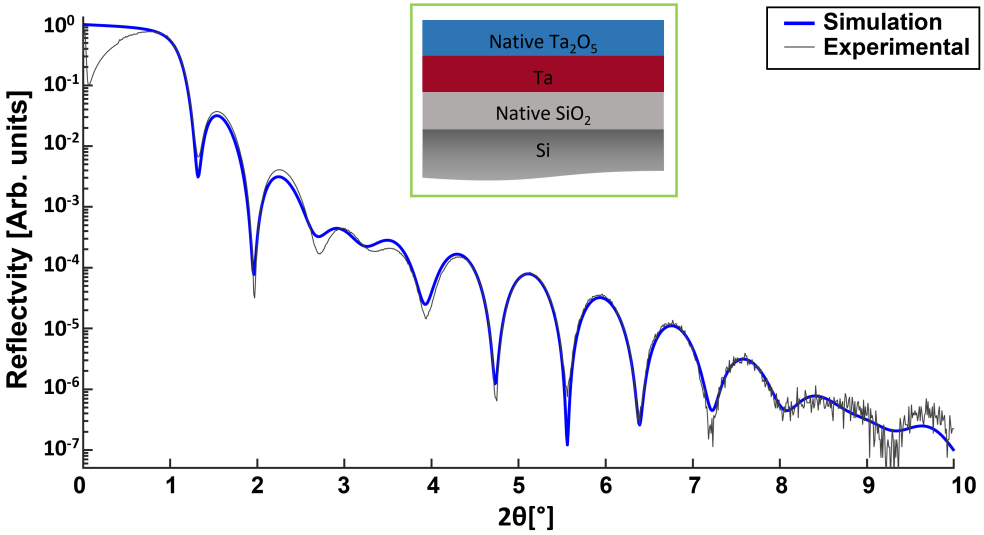


Figure A.3: XRR experimental pattern (black line) and corresponding simulated curve (blue line) for the Ta sample. The inset shows the simulated stack.

Material	Thickness [nm]	Roughness [nm]	Density [g/cm^3]
Native Ta_2O_5	1.5	0.4	8.4
Ta	9.8	0.3	16.4
Native SiO_2	2.6	0.3	2.4

Table A.1: XRR fitting parameters of the Ta sample deposited on Si substrate.

The expected thickness of the Ta layer is obtained and the high density is typical of a metallic material. A native oxide with a thickness of 1.5 nm tends to form on top of the metallic film because the layer was exposed to open air before performing the X-ray characterization.

A.2 Transmission Electron Microscopy (TEM) analysis of a ReRAM device in pristine state

A Transmission Electron Microscopy (TEM) analysis was performed by a colleague on a pristine ReRAM device embedding in the stack a nominally 20 nm thick TaO_x deposited at 9 µbar. A TEM cross-sectional image of this device is reported in Figure A.4.

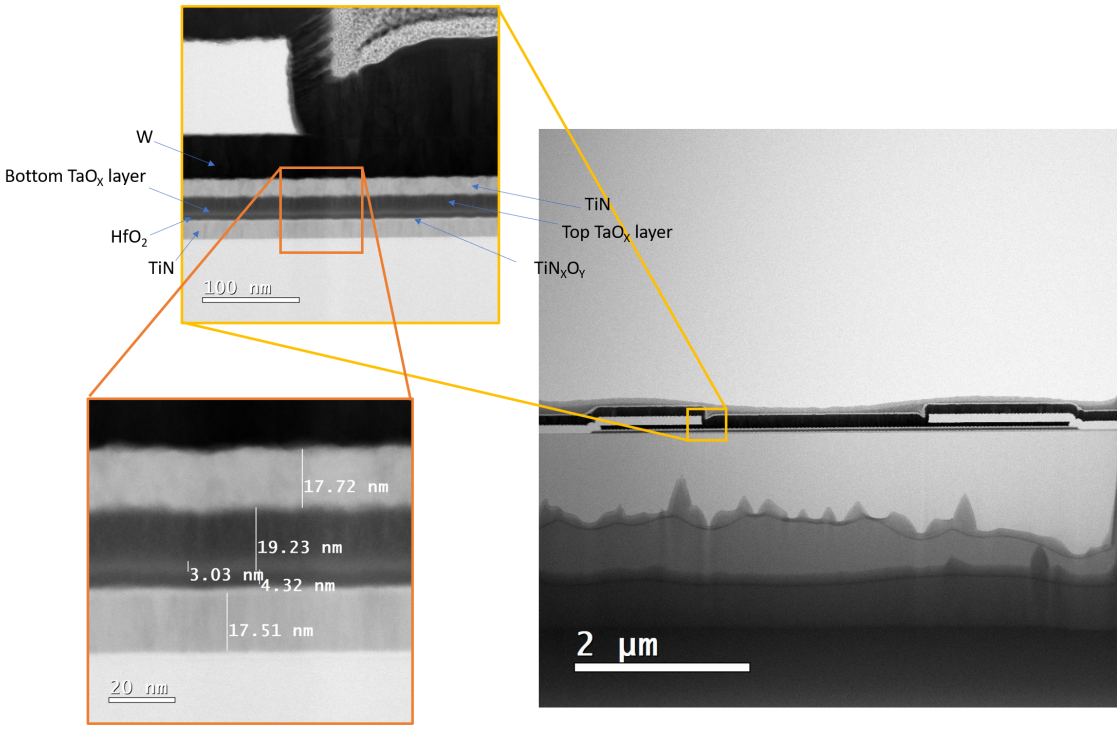


Figure A.4: TEM image of a (6×6) µm^2 pristine device embedding in the stack a nominally 20 nm thick TaO_x deposited at 9 µbar. In the insets a zoom on the left edge of the VIA and a further one on the device stack are shown.

The TEM micrograph supports the material analysis since the device has not been subjected to any electrical stimulus. Only processing steps were performed and, because of their CMOS-compatibility, they should not drastically modify the characteristics of the deposited layers. The measured TaO_x thickness from the TEM image of almost 20 nm is in agreement with the hypothesis of linear growth rate of the layer, since it was deposited in 167 s. However, the measurements reported in Figure A.4 could be not extremely accurate, since they are obtained trying to recognize colour contrast. This is true especially for very thin films like the HfO_2 or the interfacial layers.
The result shown in Figure A.4 clearly confirms the presence in the layer stack

of an interfacial layer between the dark TaO_x and the HfO_2, as well as between the HfO_2 and the TiN. Moreover, in the insets of Figure A.4 it is clearly visible a colour contrast in the dark TaO_x, which identifies vertically columnar grains, thus confirming the crystallinity of this layer.

A.3 Focused Ion Beam (FIB) cross-section of a ReRAM device in pristine state

A Focused Ion Beam (FIB) cross-section of a pristine device was performed by a colleague in order to validate all the processing steps. In Figure A.5, a Scanning Electron Microscope (SEM) image of the cross-section of a pristine device embedding a 30 nm thick TaO_x deposited at 9 µbar is shown. Figure A.5 supports the computed TaO_x thickness by the XRR fit presented in Section 4.1.3. The thickness value reported in Figure A.5 must be taken only as a reference, because it has been measured manually by looking to the colour contrast.

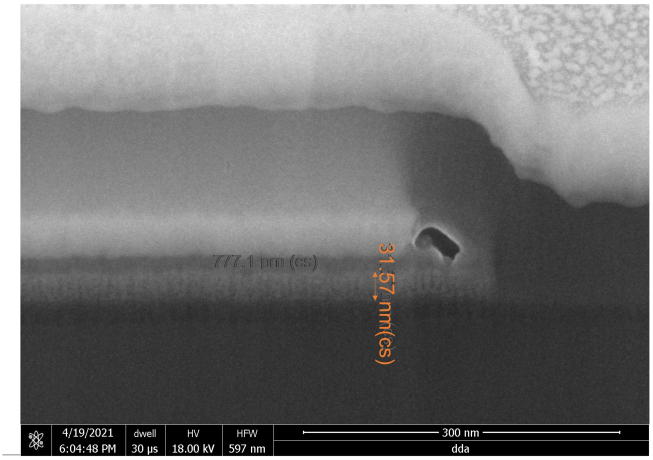


Figure A.5: SEM image of a (12×12) µm^2 pristine device embedding in the stack a nominally 30 nm thick TaO_x deposited at 9 µbar.

A.4 GIXRD and XRR analysis on TaO_x deposited on SiO_2/Si substrate

A TaO_x sample was deposited with the same sputtering conditions, and a chamber pressure of 9 µbar on a 100 nm thick SiO_2 layer, to understand the impact of the growth of the substoichiometric layer on a different substrate than the device stack one. The deposition of the SiO_2 layer was performed with PECVD at a temperature

of 300 °C and a pressure of 1000 mTorr. The plasma was generated with 2% of SiH_4(150 sccm) and N_2O (710 sccm) as gaseous sources, and an RF power of 20 W. A GIXRD scan ($\omega - 2\Theta$) was performed at ω=0.65° and varying 2Θ between 20° and 50° in steps of 0.02°. The time per step was fixed to 10 s. The background subtracted raw data of the XRD pattern is shown in Figure A.6.

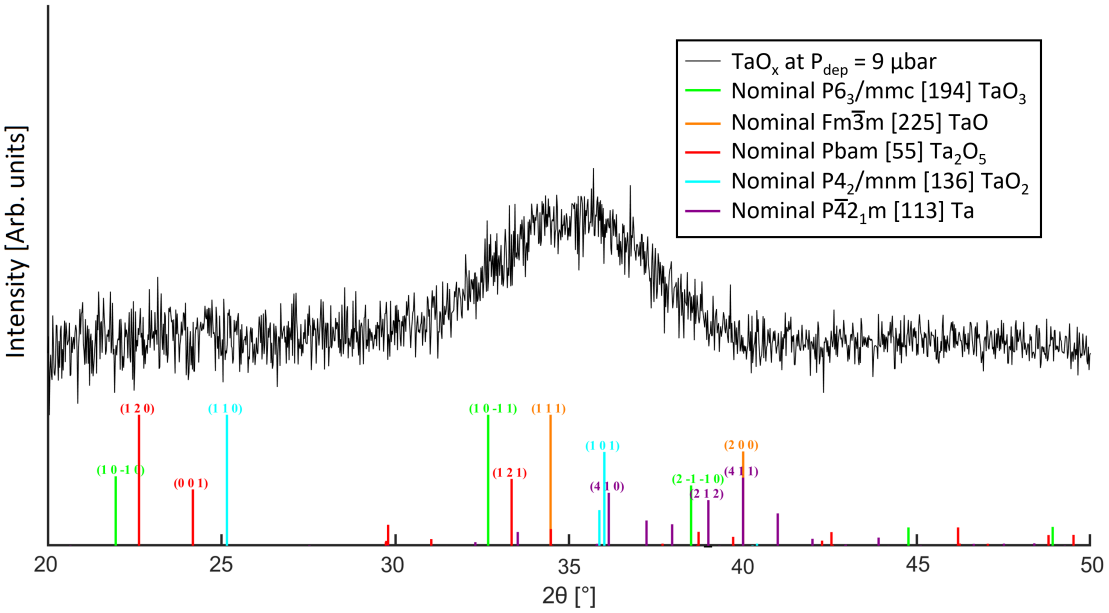


Figure A.6: Background subtracted GIXRD scan at ω=0.65° of a TaO_x sample deposited at 9 µbar on SiO_2/Si substrate. The nominal peak positions of the various TaO_x stoichiometries and Ta are reported from the reference database [34], with the correspondent Hermann-Mauguin space group. Only peaks with an intensity larger than 30% are indexed for clarity of the representation.

Figure A.6 shows an X-ray diffraction pattern that is slightly different from the one reported in Figure 4.1. The absence of the (2 0 0) peak of TiN is a further confirmation that this peak in Figure 4.1 is associated to the TiN layer in the stack. Moreover, it is observed a difference in the cumulative peak of the GIXRD pattern when the TaO_x is deposited on SiO_2/Si substrate rather than on HfO_2/TiN/Si substrate. In the latter stack the additional contribution of the (1 1 1) peak of TiN must be considered. However, the deconvolution of the cumulative peak shown in Figure A.7 highlights a slight difference with respect to Figure 4.2(d), underling a possible different growth of the material. As mentioned in Section 4.1.2, it is not possible to classify the sample in a defined crystalline phase, while the presence of different suboxide phases is more probable. Indeed, both the (1 1 1) peak of TaO and the (1 0 1) of TaO_2 are present in the GIXRD pattern shown in Figure A.6. They are in agreement with the fitted peaks 4 and 5 in A.7, respectively.

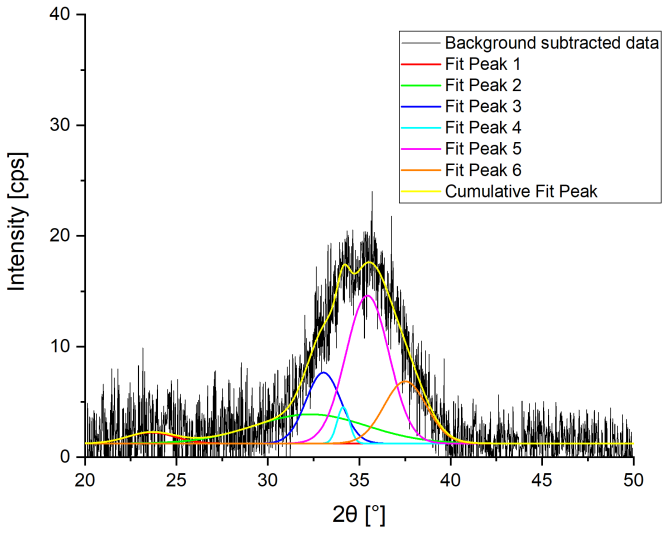


Figure A.7: Gaussian deconvolution of the GIXRD peaks in the pattern of a TaO_x sample deposited at 9 µbar on SiO_2/Si substrate.

An XRR scan was performed varying 2Θ in the range 0° to 10° in steps of 0.01°. The integration time in each step was fixed to 1 s. The raw data and the simulated XRR pattern are reported in Figure A.8, while the fitting parameters are summarized in Table A.2.

Material	Thickness [nm]	Roughness [nm]	Density [g/cm^3]
T-TaO_x	26.4	1.4	10.7
B-TaO_x	2.3	0.8	9.7
SiO_2	100.0	1.3	2.4

Table A.2: XRR fitting parameters for a nominally 30 nm TaO_x deposited at 9 µbar on SiO_2/Si substrate.

As it is possible to observe from the values reported in Table A.2, the TaO_x thickness is close to the expected 30 nm. The layer can be divided in a top layer (T-TaO_x) and bottom one (B-TaO_x). The T-TaO_x shows a thickness, roughness and density that are in agreement with the deposition on HfO_2/TiN/Si substrate (see Table 4.1). On the other hand, B-TaO_x is characterized by a larger density than the one reported in Table 4.1. This observation could hide a difference in the growth of the first nanometers of the TaO_x. This growth difference can be associated to the thickness of the amorphous oxide on top of which the substoichiometric layer is deposited or also to the stoichiometry of the oxide itself.

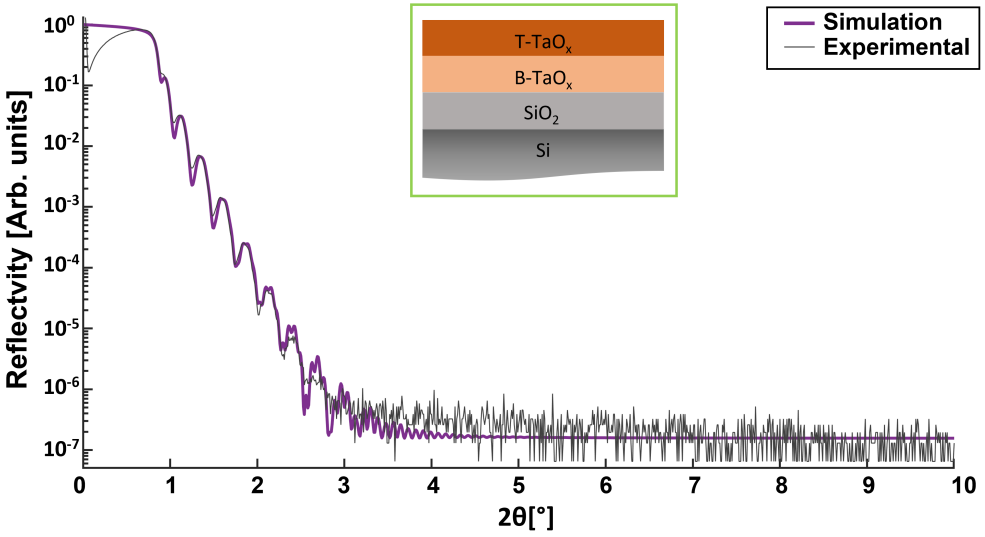


Figure A.8: XRR experimental pattern (black line) and corresponding simulated curve (violet line) for the TaO_x sample deposited at P_{dep} of 9 µbar. The inset shows the simulated stack.

A.5 CTLM fitting curves for TaO_x deposited on HfO_2/TiN/Si substrate

In this section the fitting curves for all the CTLM structures measured for each TaO_x deposited on HfO_2/TiN/Si substrate are reported. As already mentioned in Section 4.1.4, the sample deposited at 11 µbar was not measurable within the instrument limits because its insulating nature, thus it is not reported in Figure A.10.
In Figure A.10, it is possible to notice the presence of some curves that give rise to a negative value of contact resistance. This result is attributed to the difference in dimensions between the structures on the mask and the actual ones on the chip that gives rise to an horizontal shift in the fitting curves. Nevertheless, the large slope observed in all the samples highlights the very small contribution of the contact resistance and the reliability of the sheet resistance measurement. Some experimental data points were excluded during the fitting because the correspondent structures were short-circuited. However, each fitting is done on at least three measured data out of five in order to have a reliable measurement.

A.6 CTLM fitting curves for TaO_x deposited on SiO_2/Si substrate

CTLM structures were fabricated on a 100 nm thick SiO_2 film deposited by PECVD on a Si substrate. The fitting curves with the data points used during the fitting, as well as their correction, are reported in Figure A.11.
From these fits, the values of the sheet resistance and of the resistivity are computed. These values are reported in Table A.3, while they are also plotted as function of the deposition pressure in Figure A.9.

Deposition pressure [µbar]	R_{sh}[kΩ/□]	ρ [Ωµm]
6	3.96 ± 0.06	108.63 ± 1.74
7	14.74 ± 0.79	421.65 ± 22.56
8	33.46 ± 0.53	936.97 ± 14.77
9	98.32 ± 3.02	2811.95 ± 86.42
10	234.71 ± 1.58	6783.03 ± 45.52
11	—	—

Table A.3: Sheet resistance R_{sh} and resistivity ρ as function of the deposition pressure for TaO_x samples deposited on SiO_2/Si substrate. These values were not measurable for the sample deposited at 11 µbar.

Figure A.9 shows that both the sheet resistance and the resistivity increase as function of the deposition pressure, in agreement with the deposited samples on HfO_2/TiN/Si substrate. However, even if the observed trend is the same, a slightly difference between the values reported in Table A.3 and those in Table 4.2 is noticed. It could be associated to a slightly difference in the growth of the material accordingly to the substrate.

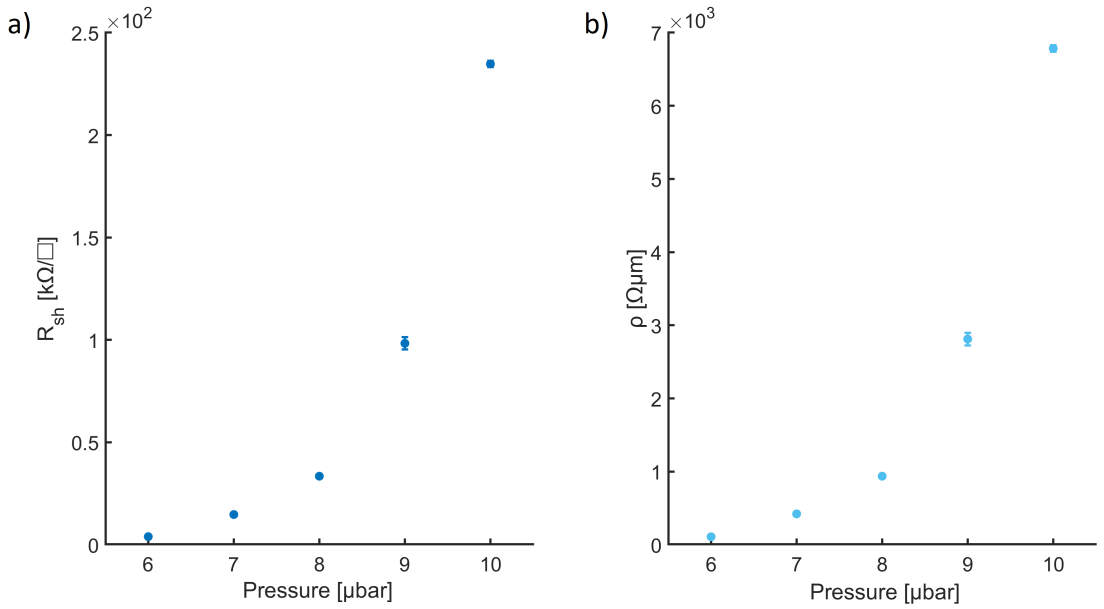
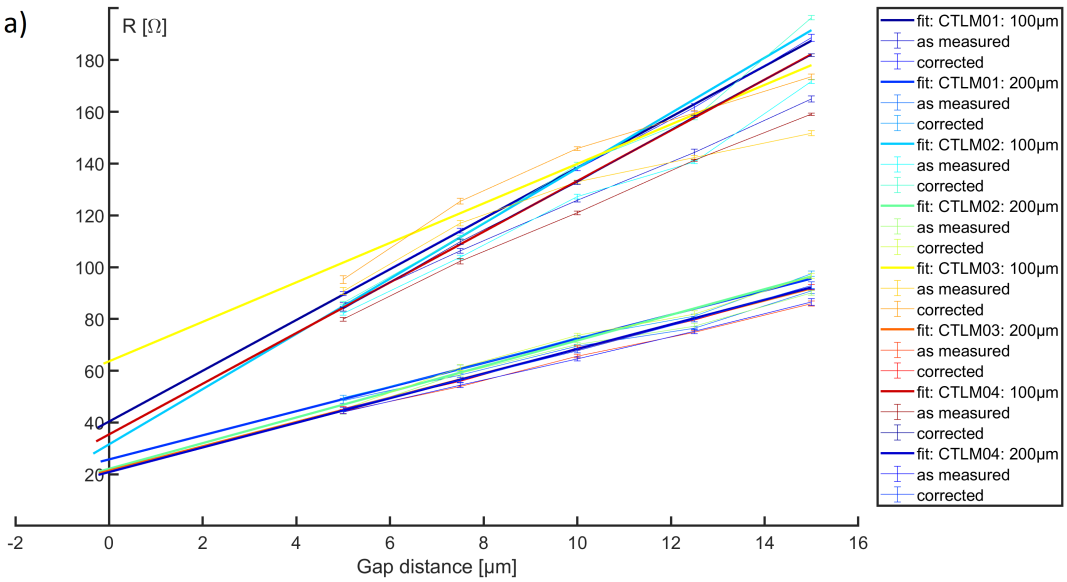
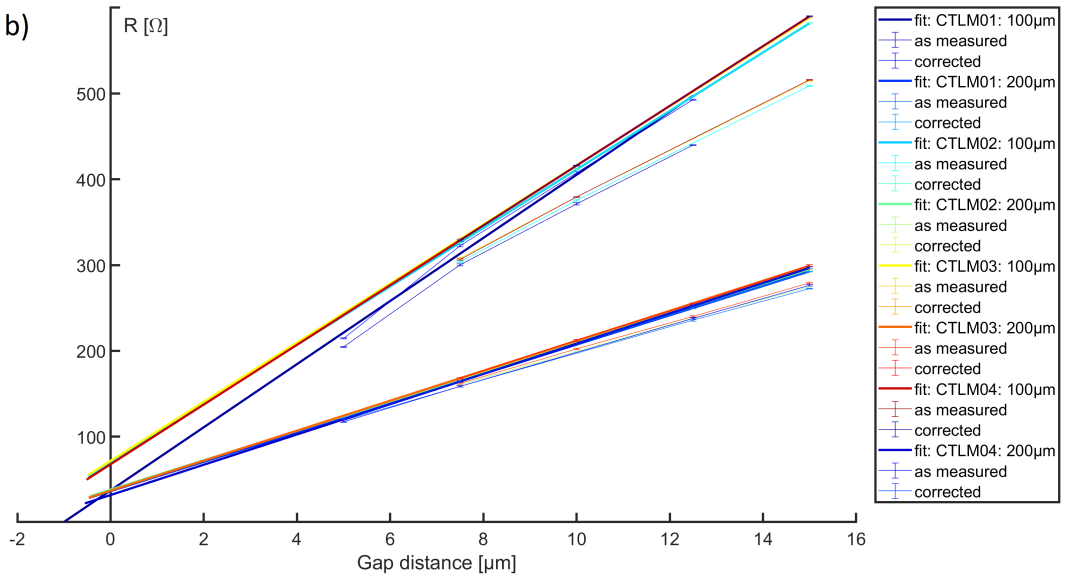
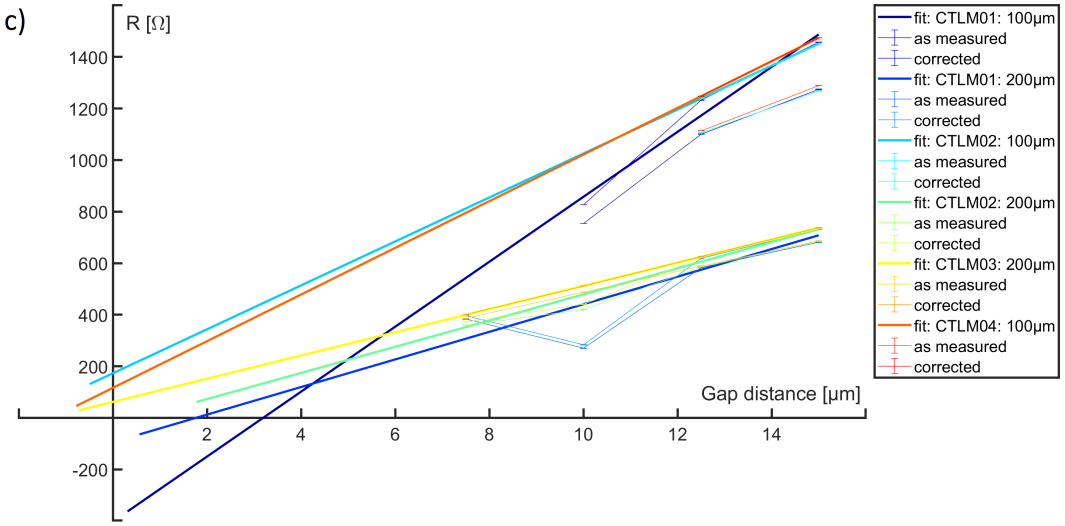


Figure A.9: Plot of the sheet resistance R_{sh} (a) and the resistivity ρ (b) as function of the deposition pressure for the TaO_x samples deposited on SiO_2/Si substrate. Not visible error bars are hidden within the mean value dot.




b)
R [Ω]
500
400
300
200
100
-2
0
2
4
6
8
10
12
14
16
Gap distance [µm]
fit: CTLM01: 100µm
as measured
corrected
fit: CTLM01: 200µm
as measured
corrected
fit: CTLM02: 100µm
as measured
corrected
fit: CTLM02: 200µm
as measured
corrected
fit: CTLM03: 100µm
as measured
corrected
fit: CTLM03: 200µm
as measured
corrected
fit: CTLM04: 100µm
as measured
corrected
fit: CTLM04: 200µm
as measured
corrected


c)
R [Ω]
1400
1200
1000
800
600
400
200
-200
2
4
6
8
10
12
14
Gap distance [µm]
fit: CTLM01: 100µm
as measured
corrected
fit: CTLM01: 200µm
as measured
corrected
fit: CTLM02: 100µm
as measured
corrected
fit: CTLM02: 200µm
as measured
corrected
fit: CTLM03: 200µm
as measured
corrected
fit: CTLM04: 100µm
as measured
corrected

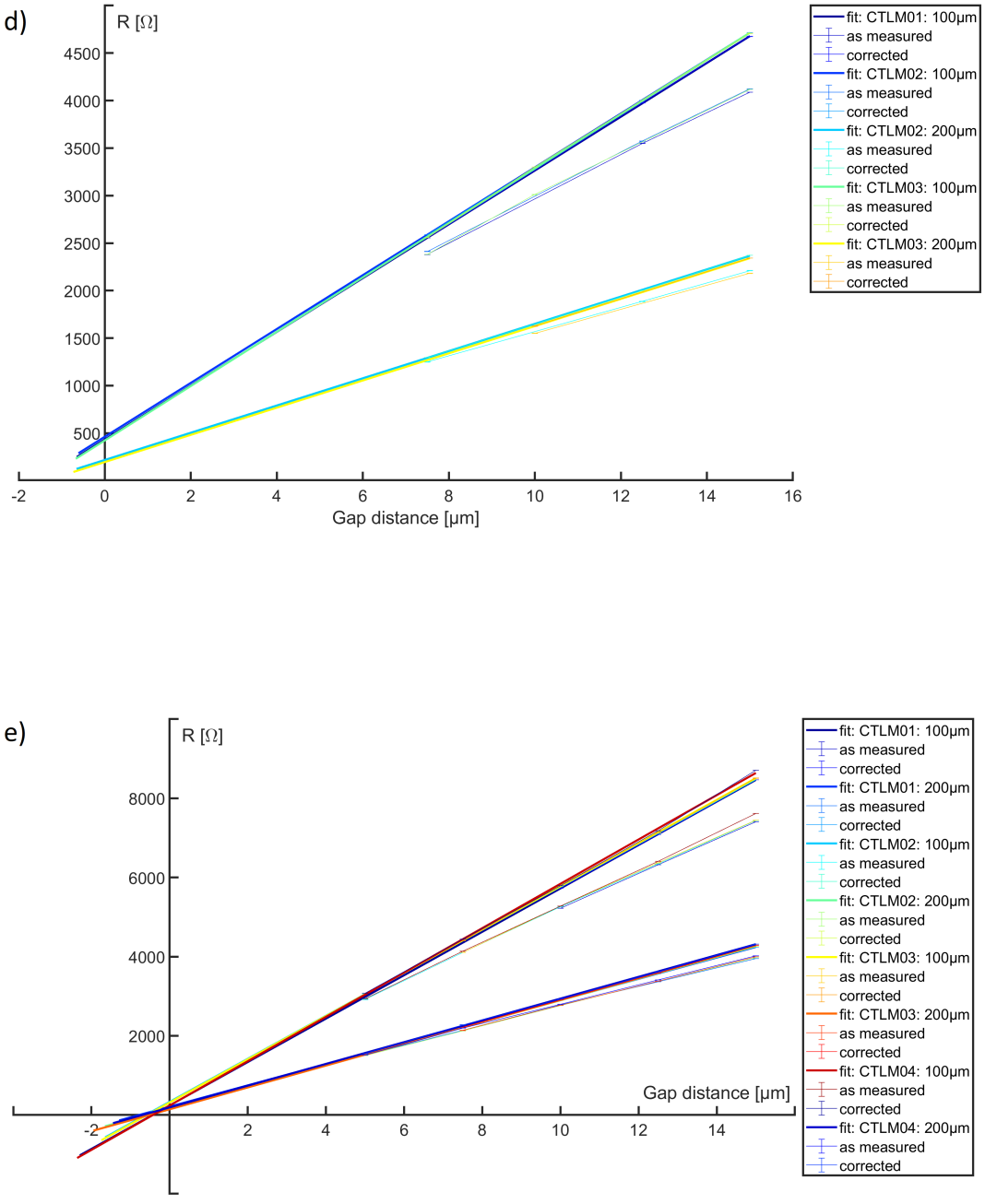
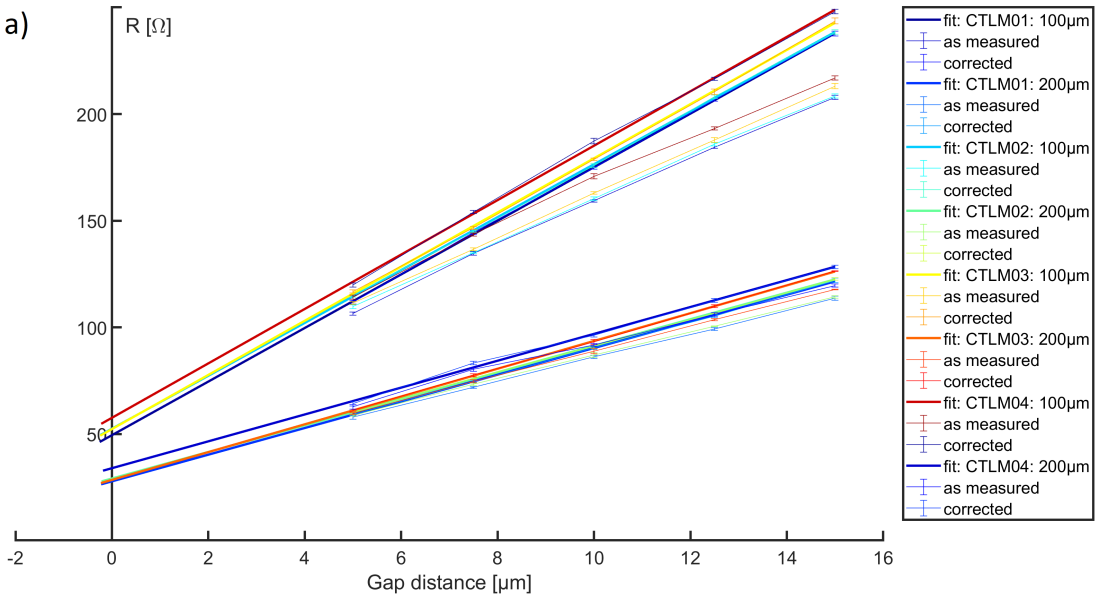
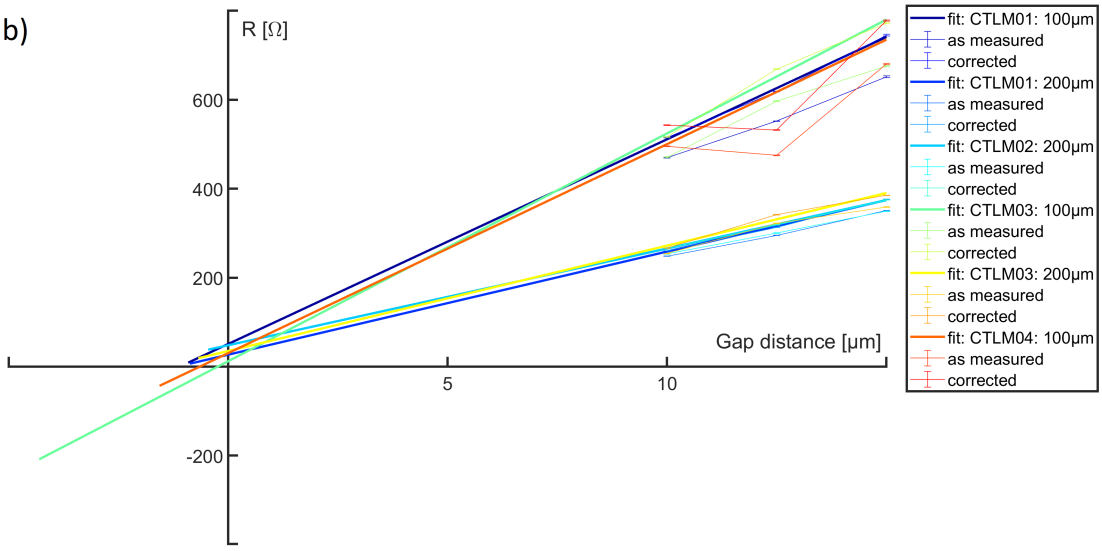
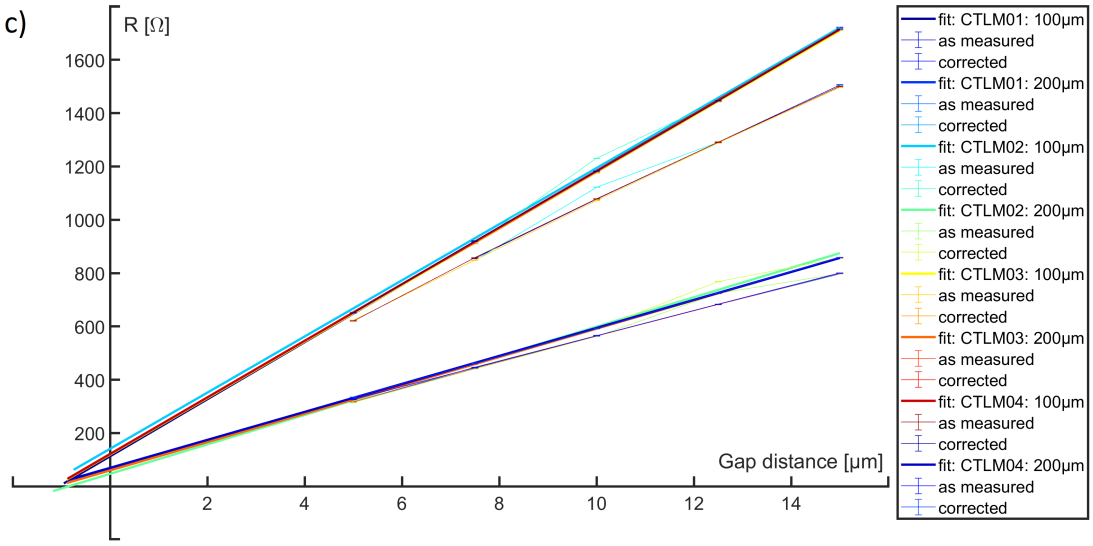
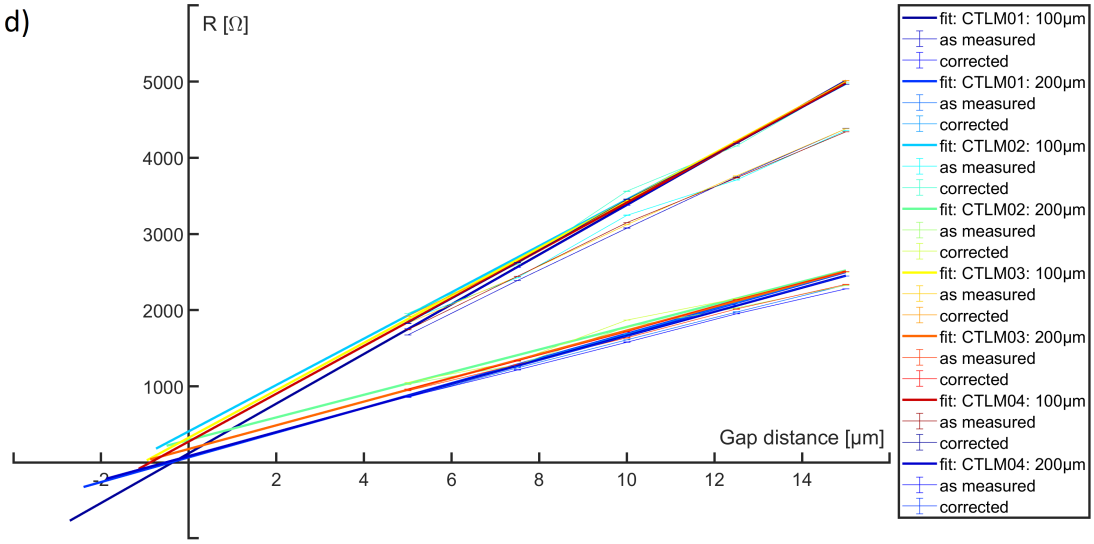


Figure A.10: Fitting curves of all the CTLM data for the TaO_x deposited on HfO_2/TiN/Si substrate at a different value of deposition pressure: 6 µbar (a), 7 µbar (b), 8 µbar (c), 9 µbar (d), 10 µbar (e).


a)
R [Ω]
200
150
100
50
-2
0
2
4
6
8
10
12
14
16
Gap distance [µm]
fit: CTLM01: 100µm
as measured
corrected
fit: CTLM01: 200µm
as measured
corrected
fit: CTLM02: 100µm
as measured
corrected
fit: CTLM02: 200µm
as measured
corrected
fit: CTLM03: 100µm
as measured
corrected
fit: CTLM03: 200µm
as measured
corrected
fit: CTLM04: 100µm
as measured
corrected
fit: CTLM04: 200µm
as measured
corrected


b)
R [Ω]
600
400
200
-200
5
10
Gap distance [µm]
fit: CTLM01: 100µm
as measured
corrected
fit: CTLM01: 200µm
as measured
corrected
fit: CTLM02: 200µm
as measured
corrected
fit: CTLM03: 100µm
as measured
corrected
fit: CTLM03: 200µm
as measured
corrected
fit: CTLM04: 100µm
as measured
corrected


c)
R [Ω]
1600
1400
1200
1000
800
600
400
200
2
4
6
8
10
12
14
Gap distance [µm]
fit: CTLM01: 100µm
as measured
corrected
fit: CTLM01: 200µm
as measured
corrected
fit: CTLM02: 100µm
as measured
corrected
fit: CTLM02: 200µm
as measured
corrected
fit: CTLM03: 100µm
as measured
corrected
fit: CTLM03: 200µm
as measured
corrected
fit: CTLM04: 100µm
as measured
corrected
fit: CTLM04: 200µm
as measured
corrected


d)
R [Ω]
5000
4000
3000
2000
1000
-2
2
4
6
8
10
12
14
Gap distance [µm]
fit: CTLM01: 100µm
as measured
corrected
fit: CTLM01: 200µm
as measured
corrected
fit: CTLM02: 100µm
as measured
corrected
fit: CTLM02: 200µm
as measured
corrected
fit: CTLM03: 100µm
as measured
corrected
fit: CTLM03: 200µm
as measured
corrected
fit: CTLM04: 100µm
as measured
corrected
fit: CTLM04: 200µm
as measured
corrected

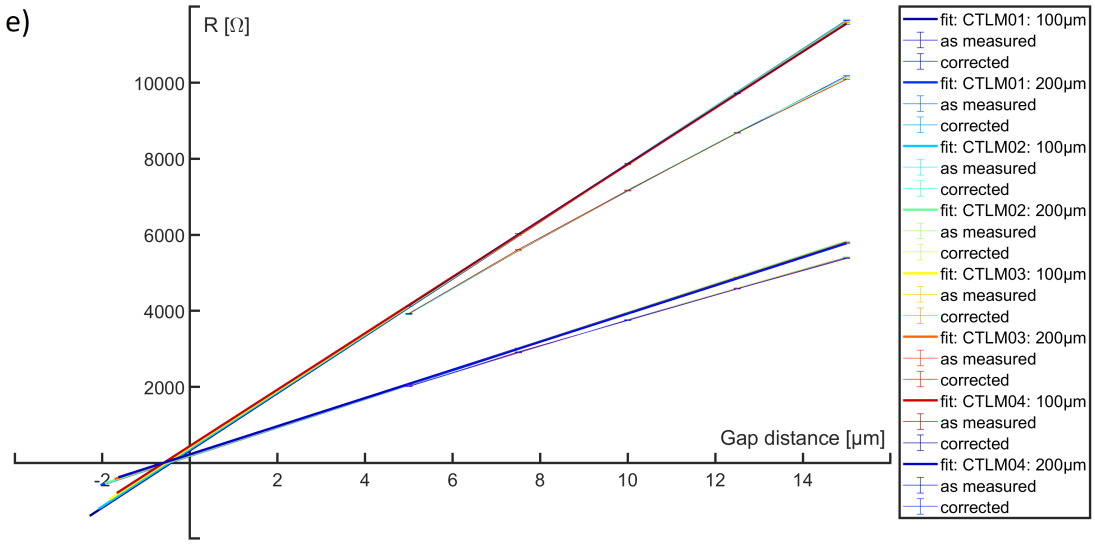


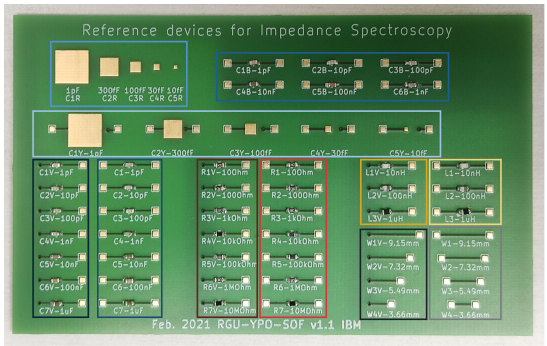
Figure A.11: Fitting curves of all the CTLM data for the TaO_x deposited on SiO_2/Si substrate at a different value of deposition pressure: 6 µbar (a), 7 µbar (b), 8 µbar (c), 9 µbar (d), 10 µbar (e).

Appendix B

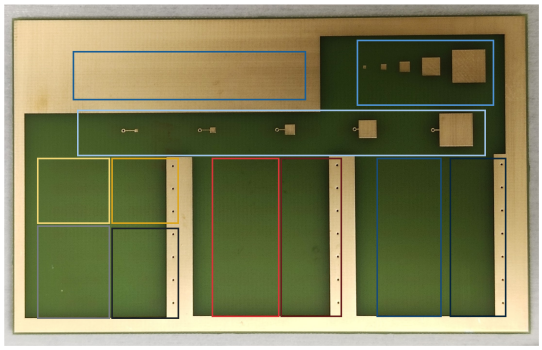
Supplementary information on Impedance Spectroscopy

B.1 Printed-Circuit-Board (PCB) design and measurements

A Printed-Circuit-Board (PCB) was designed and soldered to validate the Impedance Spectroscopy setup presented in Section 3.2.3. In this section, an analysis of the measurements on surface-mounted devices (SMD), such as capacitors and resistors, is presented.



- SMD capacitors in vertical configuration: C=1 pF to 1 µF.
- SMD capacitors in planar configuration without conductive backside: C=1 pF to 1 µF.
- SMD capacitors in planar configuration with conductive backside: C=1 pF to 100 nF.
- PCB capacitors in vertical configuration: C= 10 fF to 1 pF.
- PCB capacitors in planar configuration: C= 10 fF to 1 pF.



- SMD resistors in vertical configuration: R=10 Ω to 10 MΩ.
- SMD capacitors in planar configuration: R=10 Ω to 10 MΩ.
- SMD inductors in vertical configuration: L=10 nH to 1 µH.
- SMD inductors in planar configuration: L=10 nH to 1 µH.
- PCB wires in vertical configuration: Length= 3.66 mm to 9.15 mm.
- PCB wires in planar configuration: Length= 3.66 mm to 9.15 mm.

Figure B.1: PCB frontside (a) and backside (b) with a summary of the passive components integrated.

The proposed analysis justifies the measurement parameters and the device configuration used during the Impedance Spectroscopy experiments on ReRAM devices. Figure B.1 shows that the PCB includes all the possible passive components that can be found in an electronic circuit: capacitors, resistors and inductors. These components are in the form of 0603 SMD ones which allow for a reduced parasitic contribution and smaller dimensions if compared to devices soldered with through-holes technology. Moreover, additional capacitors are designed considering the static relative permittivity of the *FR4* insulator of the PCB board (4.4) and the thickness of the board itself (1.6 mm). Simple wire connections with different length are also present on the PCB for testing their contribution and to allow for a better SHORT correction.
Two configurations of connections to the components have been designed in order to discriminate the impact of measuring with and without the chuck connection. Figure B.2 shows the measured magnitude and phase of the impedance of the resistor components contacted using two pads (planar configuration) and those of the devices connected between a pad and the chuck through a VIA in the board (vertical configuration).

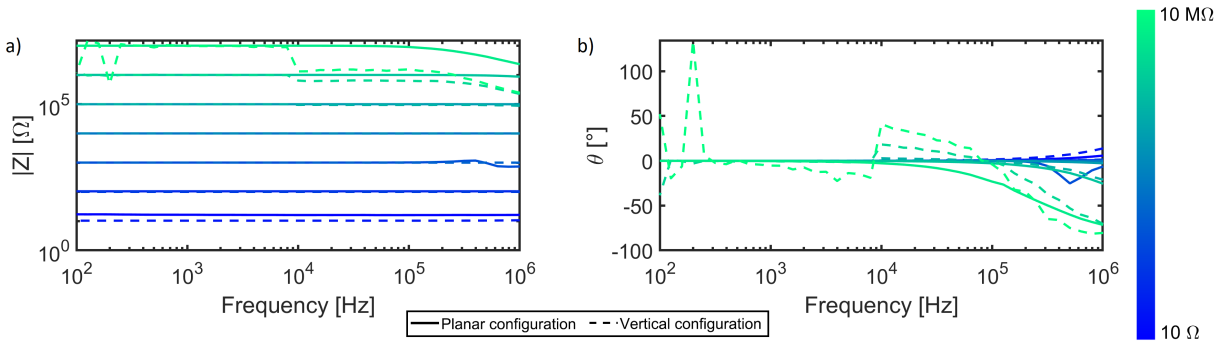


Figure B.2: Impedance magnitude (a) and phase (b) of the SMD resistors in planar and vertical configuration measured using an AC voltage V_{AC} in rms of 50 mV superimposed to a bias voltage V_{DC} of 0 V and averaging over 3 measurements each frequency point.

The two configurations of resistors are measured in the same conditions. After performing an OPEN correction with the tool, a SHORT correction is also performed by using W1 and W1V structures on the board. Then, the devices are measured using an alternating voltage V_{AC} in rms of 50 mV, which is superimposed to a bias voltage V_{DC} of 0 V, and each of the 41 frequency points is measured as the average of three measurements. Each frequency point is measured with an integration time set to LONG in the tool in order to obtain a better accuracy and stability of the measurement, while the delay time between two data points was set to 200 ms.
Apart from a capacitive contribution in the more resistive components, Figure B.2

displays a noisier behaviour in both phase and magnitude for the vertical connected resistors. The noise increases going towards larger values of resistance since the test current measured by the tool is reduced. This behaviour is observed also by measuring the capacitor devices in both configurations with the same measurement parameters, as it is represented in Figure B.3. Indeed, for the vertical connected capacitors, reducing the capacitance value towards 1 pF results in a noisier measurement due to the capacitive parasitic contribution of the chuck connection. Moreover, a larger inductive contribution at high frequency is observed for the vertical components with respect to planar ones. These results clearly highlight the need of planar devices for performing the modelling of bilayer ReRAM. In this way, it is possible to get rid of the parasitic effects created by the chuck connection at the BE of the devices.

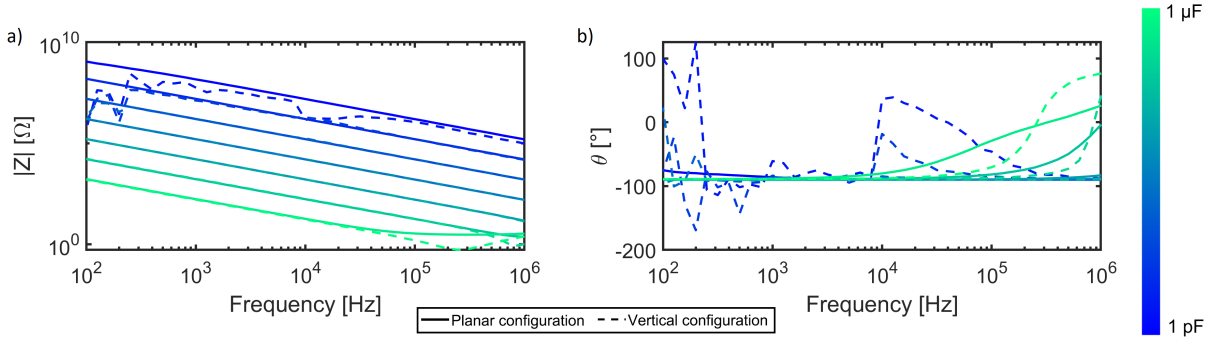


Figure B.3: Impedance magnitude (a) and phase (b) of the SMD capacitors in planar and vertical configuration measured using an AC voltage V_{AC} in rms of 50 mV superimposed to a bias voltage V_{DC} of 0 V and averaging over 3 measurements each frequency point.

We evaluated also the impact of the alternating voltage amplitude on the measurement of a 100 pF capacitor in planar configuration. The result shown in Figure B.4 highlights how increasing the alternating voltage magnitude, a better accuracy of the measurement is reached in the phase, since it gets closer to the theoretical constant value of 90°. Instead, the measurement of the magnitude of the impedance is not affected by the amplitude of the applied AC signal. As a consequence, the measurement on the ReRAM devices is performed at 200 mV. This voltage level allows to obtain a good accuracy in the measurement without altering the device itself, as it is demonstrated in section B.4.
The impact of the averaging was also studied considering both SMD planar capacitors and resistors. The measurements done averaging 32 times each frequency point or only 3 times are compared for this purpose. Figure B.5 and Figure B.6 show that increasing the averaging times allows to reach values closer to the theoretical ones in the components impedance, especially in the phase. Moreover, averaging 32 times the frequency points allows to reduce the parasitic inductive effect observed

at high frequency. On the other hand, the averaging does not play a crucial role in measuring SMD resistors. This is probably due to the larger test current that the tool measures for these components with respect to the SMD capacitors.

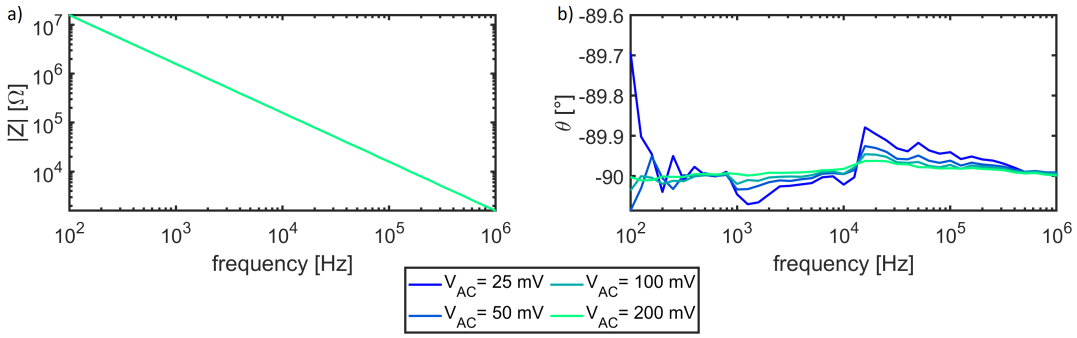


Figure B.4: Impedance magnitude (a) and phase (b) of the 100 pF SMD capacitors in planar configuration without conductive backside measured at different AC voltages V_{AC} superimposed to a bias voltage V_{DC} of 0 V and averaging over 3 measurements each frequency point.

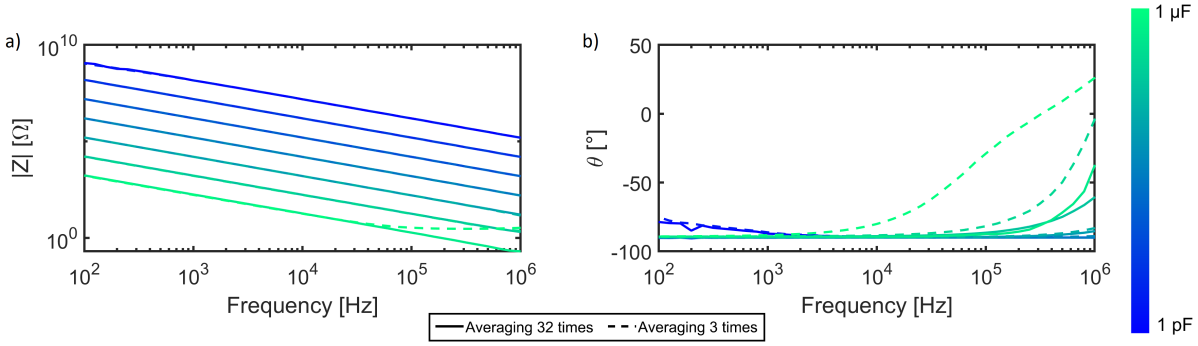


Figure B.5: Impedance magnitude (a) and phase (b) of the SMD capacitors in planar configuration without conductive backside measured at AC voltages V_{AC} in rms of 50 mV superimposed to a bias voltage V_{DC} of 0 V and averaging over 32 or 3 measurements each frequency point.

As already motivated, planar devices are the best configuration to measure since parasitic contributions are reduced. However, when planar devices are measured the chuck cannot be removed since it holds in place the chip by the application of vacuum; hence it must be left to a floating potential. The effect of having or not a conductive backside connection to the electrically floating chuck was studied using SMD capacitors. The result reported in Figure B.7 clearly shows that having or not a conductive backside connected to the electrically floating chuck does not play any role in determining the accuracy of the impedance measurement.

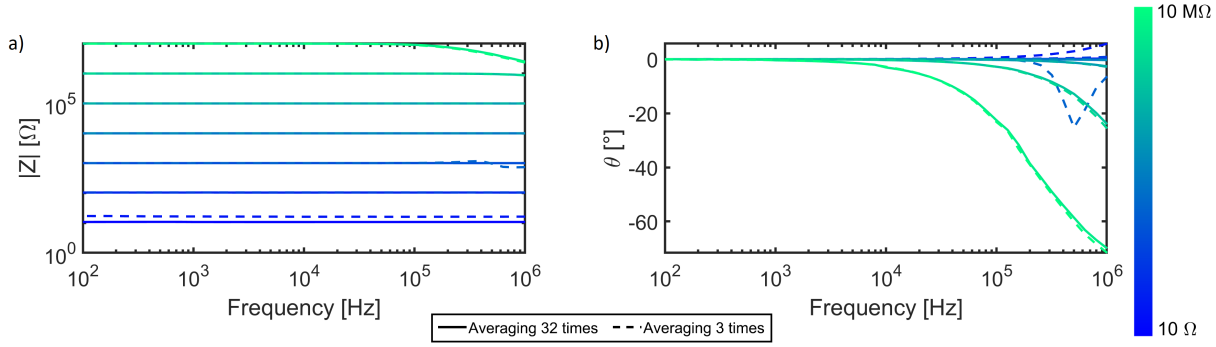


Figure B.6: Impedance magnitude (a) and phase (b) of the SMD resistors in planar configuration without conductive backside measured at AC voltages V_{AC} in rms of 50 mV superimposed to a bias voltage V_{DC} of 0 V and averaging over 32 or 3 measurements each frequency point.

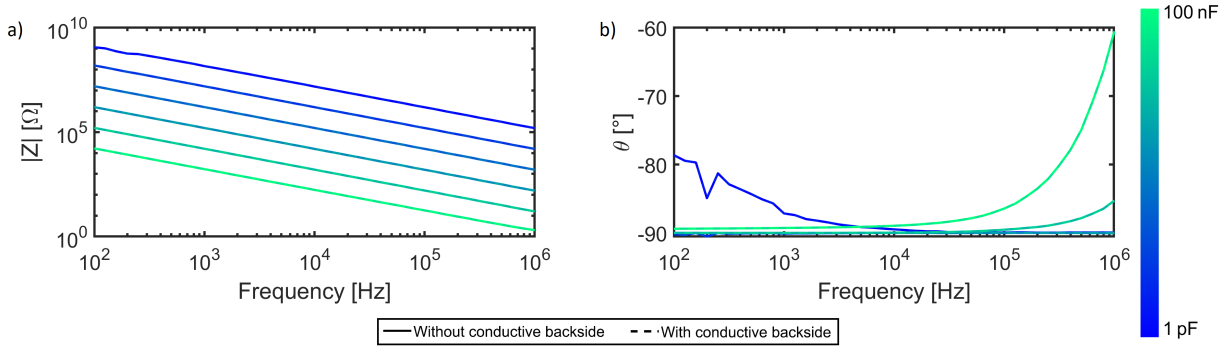


Figure B.7: Impedance magnitude (a) and phase (b) of the SMD capacitors in planar configuration with and without conductive backside. They are measured using an AC voltage V_{AC} in rms of 200 mV superimposed to a bias voltage V_{DC} of 0 V and averaging over 32 measurements each frequency point.

B.2 Mathematical description of how to fit complex impedance data

This section is intended to guide the reader through the post-processing of the Impedance Spectroscopy data, explaining how the raw data is mathematically elaborated and fitted to obtain the equivalent electrical model of the DUT.
Once the tool corrections are applied, the impedance of the device Z_{DUT} and the one of a "dummy" structure Z_{SHORT} are measured. The "dummy" structure is important to remove all additional series contribution to the device. Examples of "dummy" structures can be the W1 and W1V lines on the PCB. In the case of the

device, a "dummy" structure is simply the same device structure but W substitutes the active layers, thus accounting for the unwanted series contribution coming from the metal layers that connect the tip to the device. The correction for this series contribution can be obtained applying equation (B.1).

$$Z_{corrected} = Z_{DUT} - Z_{SHORT} \tag{B.1}$$

A further open correction could be also applied mathematically in this post-processing step. In this work, it was not applied since no further improvement in the measured data was observed and even an overcorrection was obtained in some cases.
Once the corrected data are ready, the fitting step can start. Here the fitting is explained by the example of a model that takes into account a capacitor C_p in parallel with a resistor R_p, then connected to a series component R_s. We have chosen this example since it is the one observed in all the states of the ReRAM devices studied in this master thesis. A graphical representation of the model used is drawn in Figure 3.18. In this case, the nonlinear function to fit the data depends on the three parameters of the model and it is reported in equation (B.2).

$$Z(C_p, R_p, R_s) = (j\omega C_p + \frac{1}{R_p})^{-1} + R_s \tag{B.2}$$

where ω is the angular frequency, which can be related to the standard frequency f, with the equation (B.3).

$$\omega = 2\pi f \tag{B.3}$$

The fitting approach followed consists to evaluate the model on the data and to adjust the model parameters in order to obtain a better electrical description of the acquired data. In order to define how much close the model is with respect to the data, a function which accepts the model parameters, as well as the n data, and that returns the sum of squared errors for the model $Z(C_p, R_p, R_s)$ was defined. The sum of squared error function is reported in expression (B.4).

$$\sum_{i=1}^{n} (\frac{|Z_{data}| - |Z|_i}{Z_{range}})^2 + (\frac{\Theta_{data} - \Theta_i}{\Theta_{range}})^2 \tag{B.4}$$

where $|Z|_i$ and Θ_i are the values of the magnitude and phase of the computed impedance by the model for the data i, while the ranges elements refer to:

$$Z_{range} = |Z_{data}|_{max} - |Z_{data}|_{min} \tag{B.5}$$

$$\Theta_{range} = \Theta_{data_{max}} - \Theta_{data_{min}} \tag{B.6}$$

The `fminsearch` function in *MATLAB* is then used to find the parameters that minimize the objective function (B.4). However, since `fminsearch` searches for a local minima in the objective function and not for an absolute one, a good set of starting values for the model degrees of freedom should be passed to the function. This is fundamental to achieve a good fit without reaching the maximum number of iterations allowed by the software.

B.3 Electrical equivalent circuit of TiN/HfO_2/Pt devices with VIA in the MESA corner

A comparative analysis for the two configurations of TiN/HfO_2/Pt devices having a VIA in the MESA corner which is scaling or not, is here provided. For both configurations the model used to fit the data is exactly the same as the one explained in Section 4.3.4 with three degrees of freedom: a capacitive element C; a leakage resistor R_{leak}; and an equivalent series resistance R_{esr}.

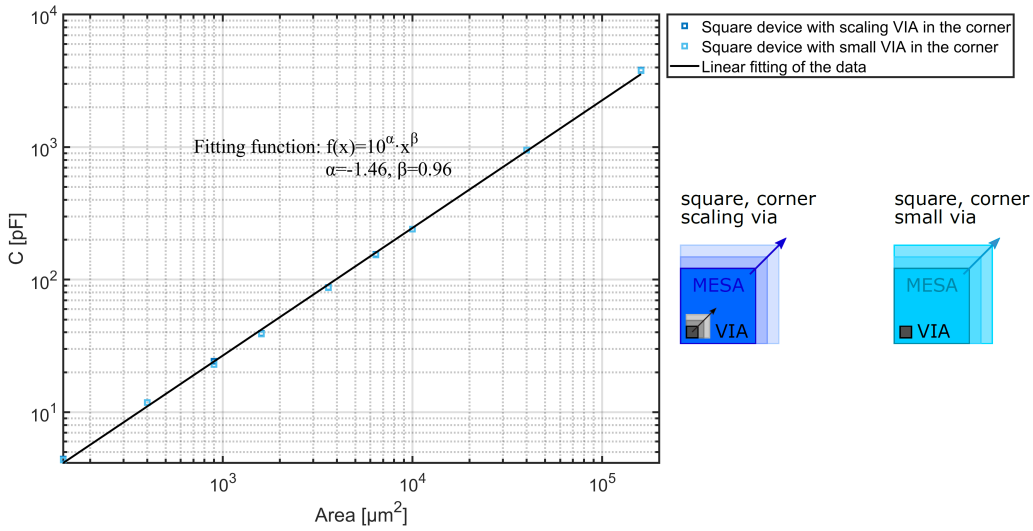


Figure B.8: Plot of the capacitive contribution C in the model as function of the area of TiN/HfO_2/Pt microstucture with scaling or small VIA in the corner.

The capacitive contribution increases linearly with the device area, as already shown in Figure 4.45. The standard formula of the capacitance in a parallel plate capacitor can be inverted to obtain the static relative permittivity. The result plotted as function of the device area in Figure B.9 shows that the static dielectric constant is unexpectedly increasing for the devices having an area between 900 µm^2 and 144 µm^2, like in Section 4.3.4. This result can be commented in the same way as in

Section 4.3.4 referring to the possible contribution of fringing fields or to the tool limitations.

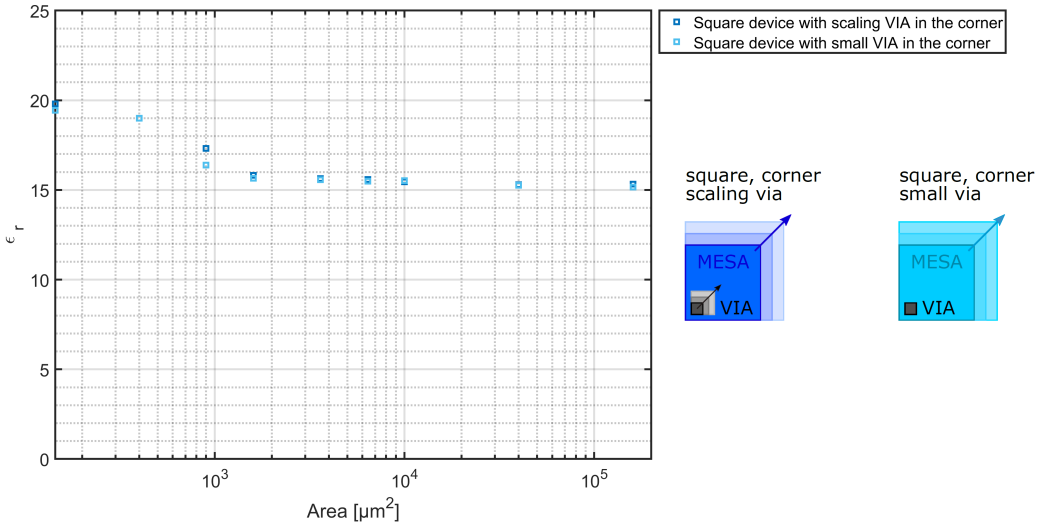


Figure B.9: Plot of the static relative permittivity ϵ_r as function of the area of TiN/HfO_2/Pt microstucture with scaling or small VIA in the corner.

Regardless of device-to-device variability, Figure B.10 shows as the R_{leak} parameter decreases by increasing the device area. This trend is independent from the VIA dimensions thus stressing the dependence of this parameter on the HfO_2 properties. The high values assumed by R_{leak} further confirm the more insulating behaviour of this sample with respect to the one embedding the TaO_x layer.
Finally, the result plotted in Figure B.11 shows the saturation of the R_{esr} for square devices with scaling VIA in the corner and larger areas than 3600 μm^2. Moreover, an increase of this resistive contribution for square devices with small VIA in the corner and areas larger than 3600 μm^2 is observed. This behaviour in the R_{esr} is exactly the same of Figure 4.41, thus it can be attributed in the same way to the different length of the current paths.

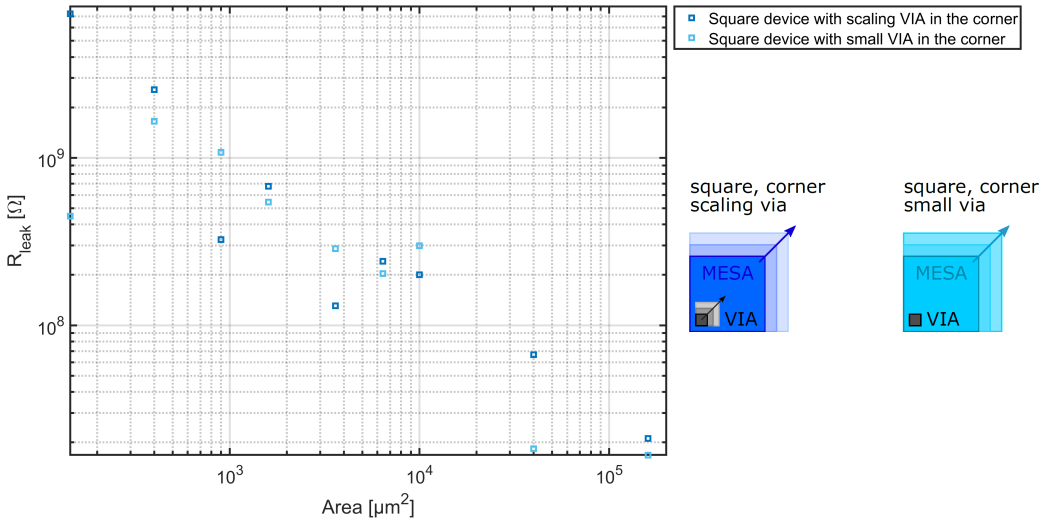


Figure B.10: Plot of the leakage resistance R_{leak} as function of the area of TiN/HfO_2/Pt microstucture with scaling or small VIA in the corner.

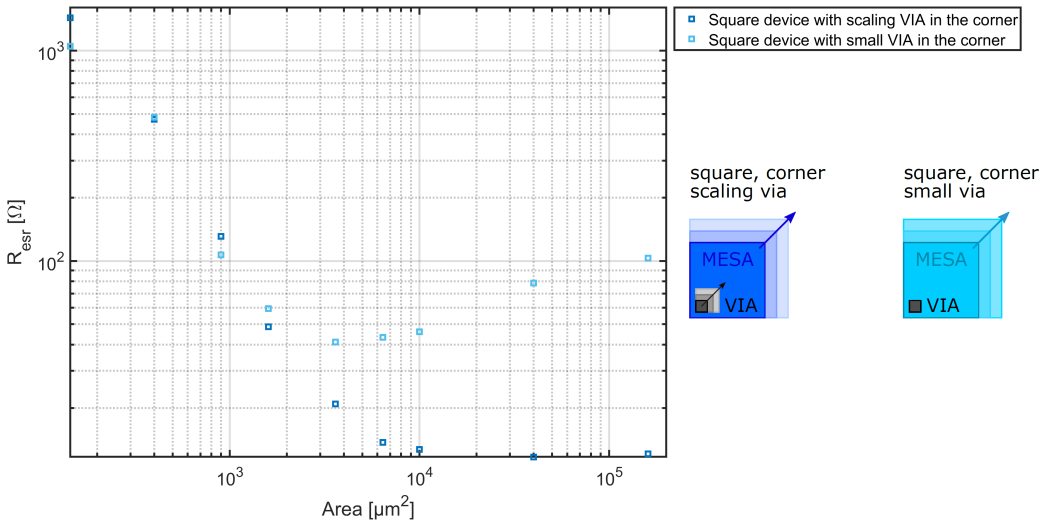


Figure B.11: Plot of the equivalent series resistance resistance R_{esr} as function of the area of TiN/HfO_2/Pt microstucture with scaling or small VIA in the corner.

B.4 AC voltage impact on Impedance Spectroscopy measurement of TaO_x/HfO_2 bilayer ReRAM device in pristine state

Increasing the alternating voltage amplitude used for the impedance measurement leads to a more accurate result, as already shown in Figure B.4. However, it is not yet clear if this measurement parameter has an impact on the ReRAM devices. For this reason, a (80 × 80) µm^2 square device with scaling VIA in the corner was subjected first to five consecutive measurements at an increasing voltage amplitude from 25 mV to 500 mV (in rms). Then, it was remeasured at a voltage of 200 mV to observe if the device has changed during the measurements or not. The bode plots of the experimental data, as well as the fitting result, are reported in Figure B.12.

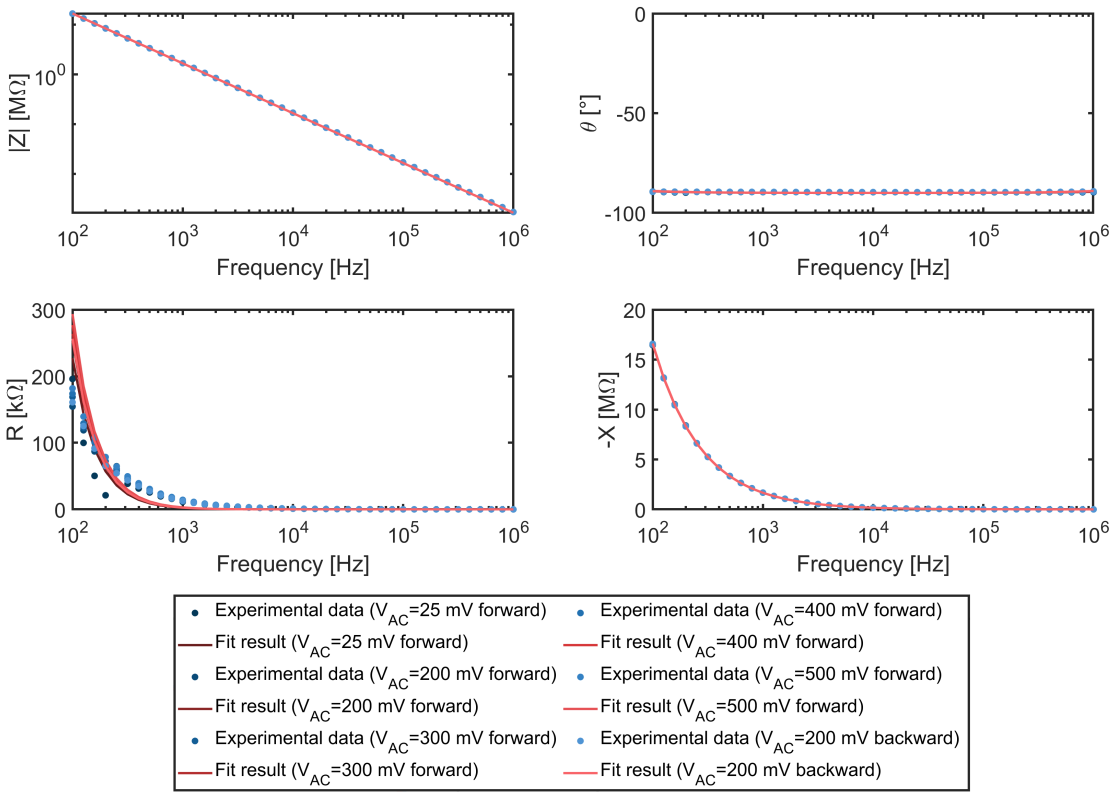


Figure B.12: Bode plots of the impedance expressed in magnitude and phase, as well as real and imaginary part, for a (80 × 80) µm^2 square device with scaling VIA in the corner subjected first to five consecutive measurements at an increasing voltage amplitude V_{AC} (in rms) from 25 mV to 500 mV and then again to 200 mV. The V_{AC} is superimposed to a bias voltage V_{DC} of 0 V and each frequency point is averaged 16 times.

Figure B.12 shows that the imaginary part is not affected by V_{AC} amplitude, being all the experimental points and the fitting curves perfectly overlapped. This result is extended to the magnitude and the phase of the impedance, since the imaginary part represents the major contribution. On the other hand, the increase of V_{AC} amplitude results in the reduction of the noise in the real part at low frequencies. Indeed, the experimental points below 500 Hz are scattered if the measurement is taken at 25 mV, whereas a smoother curve is already observed at 200 mV. However, increasing the amplitude of the AC measurement voltage does not imply a better fit in the real part of the impedance of the ReRAM device in pristine state. A larger frequency range may be beneficial for this purpose.
The good superposition between the data points recorded in the two measurements performed at 200 mV clearly shows that this V_{AC} amplitude can be used to measure the bilayer ReRAM devices without producing a major alteration in them.

B.5 Impedance Spectrocopy measurement of TaO_x/HfO_2 bilayer ReRAM device in pristine state down to 20 Hz

The observation of a first-order system in the Impedance Spectroscopy measurement described in Section 4.3.5 could be related to the limited frequency range in which the measurement is performed. The precision LCR meter *HP 4284A* used in this master thesis work allows to take measurements at most from 20 Hz to 1 MHz. In order to further support the proposed model for the ReRAM devices in pristine state, a 1600 μm^2 square device with small VIA in the corner was measured in both the typical frequency range used in this work (from 100 Hz to 1 MHz) and from 20 Hz to 1 MHz. The results are compared in Figure B.13, which shows that the two measurements are perfectly overlapped from 100 Hz to 1 MHz. However, below 100 Hz the measurement starts to be very noisy with the appearance of even some negative values in the real part R of the impedance. This result demonstrates that the measurement is affected by tool limitations which are probably related to a non-perfect calibration at low frequencies. These limitations prevent from performing the Impedance Spectroscopy measurements with the widest possible range.
The value assumed by the three model variables in both frequency ranges is reported in Table B.1. C and R_{esr} remain almost unaffected by the enlargement of the frequency range. On the other hand, R_{leak} presents one order of magnitude of difference between the two measurements. This observation suggests that limiting the measurement in the range 100 Hz-1 MHz leads to an underestimation of the R_{leak}. However, the poor tool accuracy at low frequencies does not allow to reach a solid conclusion.
The tool limitations are not observed in the imaginary part of the impedance and

hence neither in its magnitude and phase. This is sufficient to conclude as the first-order system proposed to describe the bilayer ReRAM devices in pristine state is still valid even reducing the lower measurement frequency bound to 20 Hz. Higher order contributions are outside from this range or hidden behind the dominant contribution observed.

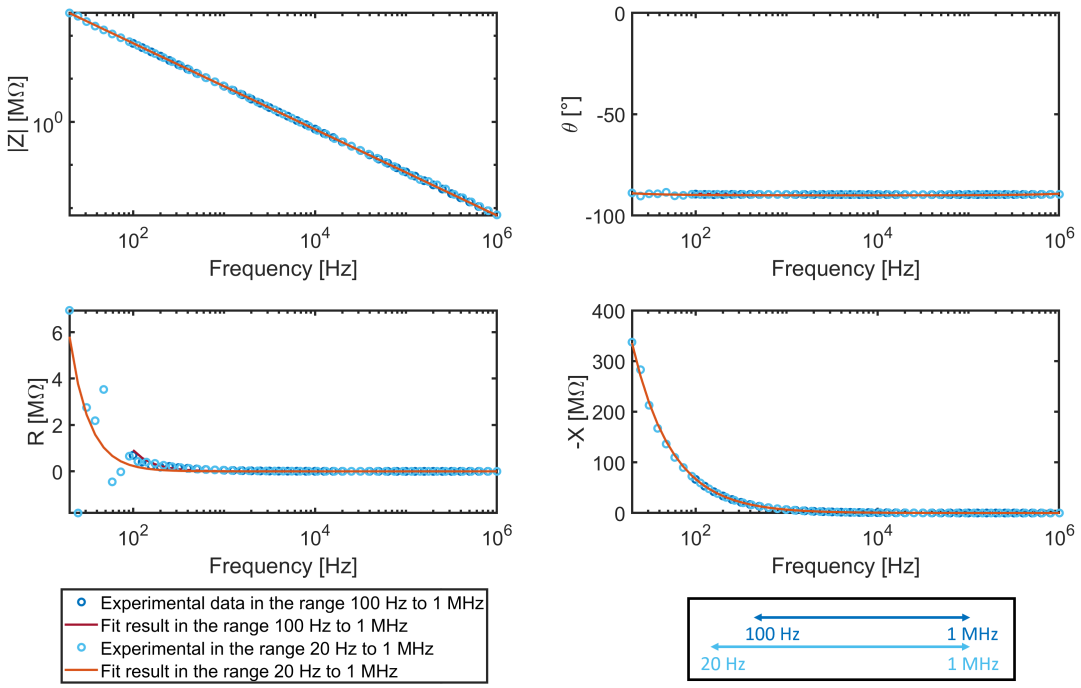


Figure B.13: Bode plots of the impedance expressed in magnitude and phase, as well as real and imaginary part, for a (40 × 40) µm^2 square device with small VIA in the corner measured in the frequency range from 20 Hz to 1 MHz and from 100 Hz to 1 MHz. A V_{AC} of 200 mV (in rms) is superimposed to a bias voltage V_{DC} of 0 V and each frequency point is averaged 32 times.

f [Hz]	C [pF]	R_{leak} [GΩ]	R_{esr} [Ω]
100 - 1·10^6	23.86	4.85	88.97
20 - 1·10^6	23.70	19.40	90.90

Table B.1: Value of the parameters of the model for a 1600 µm^2 square device with small VIA in the corner measured in the frequency ranges of 20 Hz-1 MHz and 100 Hz-1 MHz.

B.6 Fitting of Impedance Spectrocopy data for a TaO_x/HfO_2 bilayer ReRAM device in post-forming, HRS and LRS

The fitting curves of the experimental data after electroforming, in LRS and HRS for the TaO_x/HfO_2 bilayer ReRAM device analysed in Section 4.3.6 are reported in this section to stress the validity of the proposed first-order model also when the device is electrically stimulated.

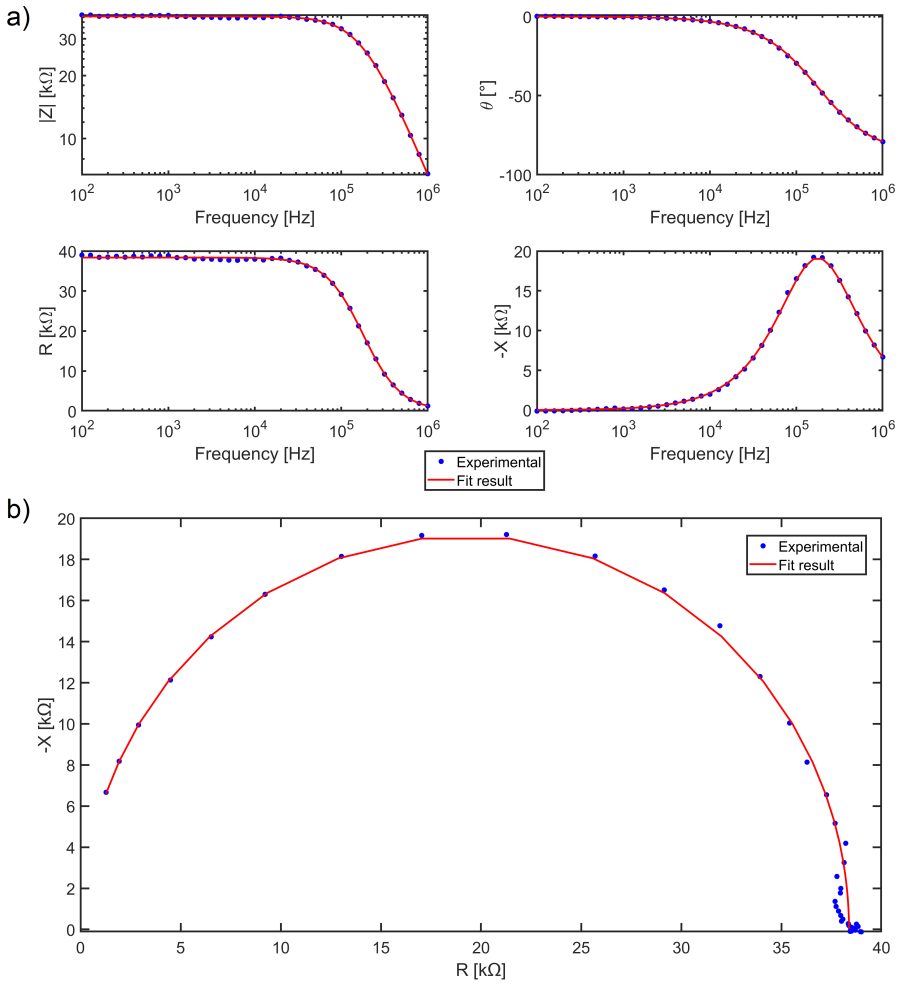


Figure B.14: Fit result for a 1600 µm^2 TaO_x/HfO_2 bilayer ReRAM device with small VIA in the corner in post-forming state. It is represented as bode plots (a) and Nyquist plot (b).

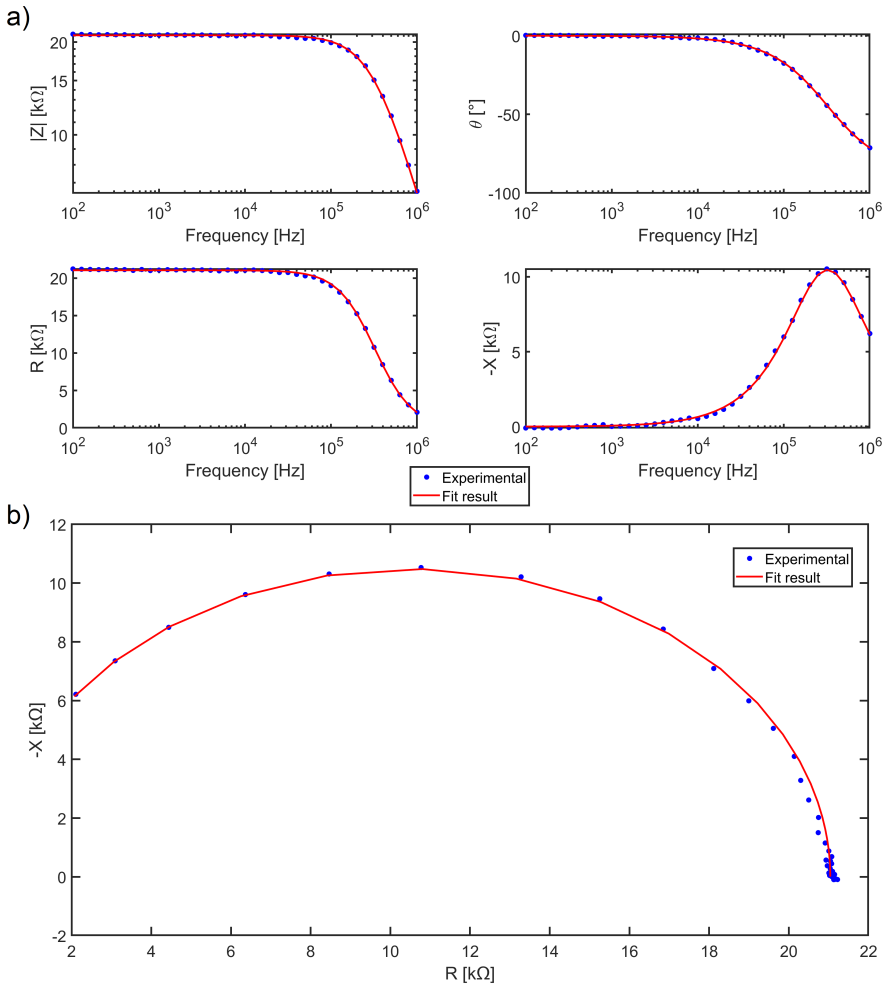


Figure B.15: Fit result for a 1600 µm^2 TaO_x/HfO_2 bilayer ReRAM device with small VIA in the corner in HRS. It is represented as bode plots (a) and Nyquist plot (b).

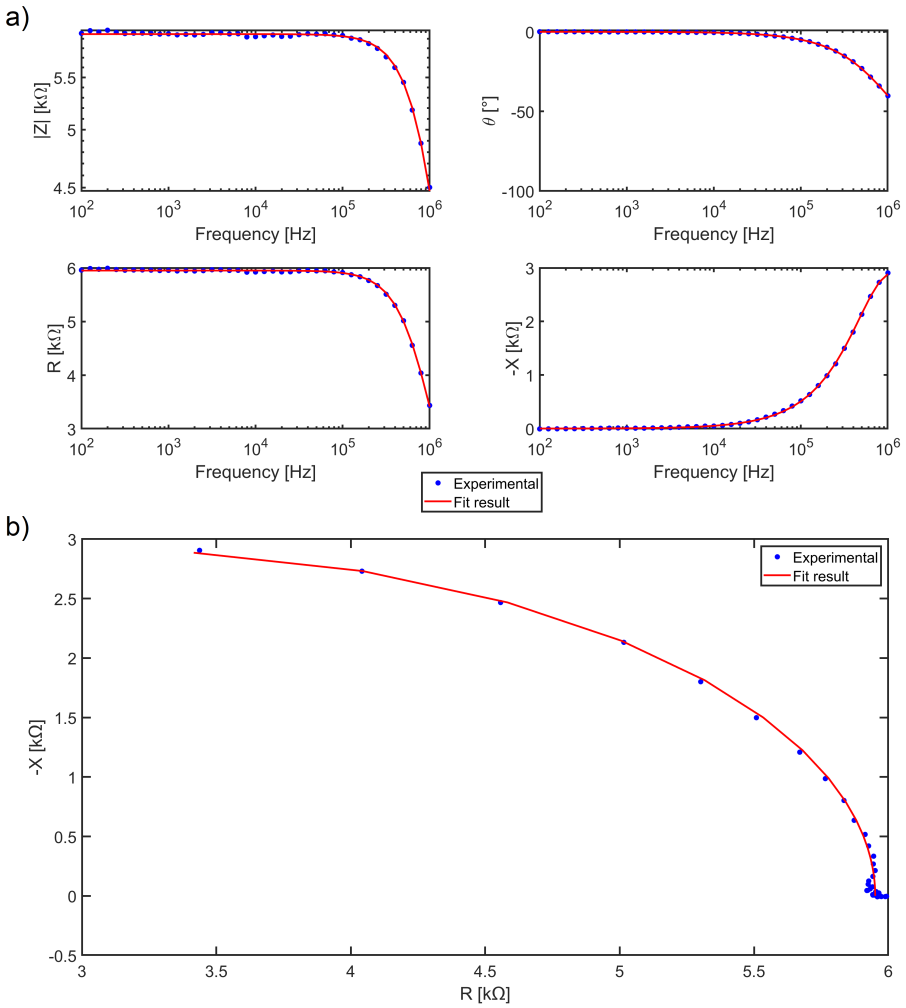


Figure B.16: Fit result for a 1600 µm^2 TaO_x/HfO_2 bilayer ReRAM device with small VIA in the corner in LRS. It is represented as bode plots (a) and Nyquist plot (b).

B.7 Evolution of the electrical equivalent circuit of TaO_x/HfO_2 bilayer ReRAM device during SET process

The modelling of the device in the different resistive states presented in Section 4.3.6 was performed on data acquired by measuring its impedance after the DC programming described in Section 4.3.1. In this section instead, a different approach is followed to measure the device impedance: the impedance is measured while performing the voltage sweep required for programming the device in LRS.

Figure B.17 depicts the two components of the voltage signal V_{SIGNAL} applied on the device: a DC bias V_{DC} for making the state transition occurring and an AC voltage V_{AC} of 200 mV (in rms) for measuring the impedance. All the measurements are performed averaging only 8 times each frequency point to reduce both the electrical stress on the device and the measurement time with respect to averaging 32 times.

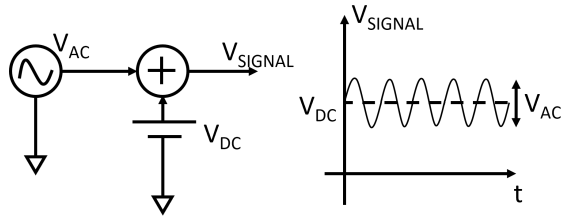


Figure B.17: Schematic representation of the voltage signal V_{SIGNAL} applied on the device for measuring its impedance while the resistive state is changing.

A (60 × 60) µm^2 square device with a small VIA in the corner was subjected to electroforming with standard DC method. Consequently, its resistance dropped to 248 kΩ (if it is read a 0.2 V). Then a first bidirectional voltage sweep on V_{DC} from 0 V to -1.4 V in steps of 0.2 V was applied to SET the device. The impedance was measured for each V_{DC}, and the evolution of the parameters of the model while the switching was occurring is reported in Figure B.18. The values assumed by the model parameters are also listed in Table B.2.
Figure B.18 shows that the capacitive contribution C of the system remained constant during both forward and backward sweep. This further supports the considerations made in Section 4.3.6 on its bulk origin or the impossibility to observe its possible change due to phenomena occurring at the interface between TaO_x and HfO_2. On the other hand, increasing the V_{DC} during the forward sweep, the leakage resistance R_{leak} decreased. This reduction is still attributed to the removal of oxygen ions from the CMO layer that on their own recombine with oxygen vacancies in the filament region. During the backward sweep, R_{leak} was always lower than its corresponding value on the forward sweep, thus highlighting the occurrence of a SET transition. A non-linear character in the LRS can be also observed in Figure B.18.
The thinning of the filament (without any breakage occurring) due to the SET event in the filamentary-based model proposed can be observed in the slight increase of the values assumed by the equivalent series resistance R_{esr} in the backward sweep with respect to the forward one. However, a negative value of R_{esr} was obtained from the fitting when the biggest change in R_{leak} was observed. This negative value does not have any electrical meaning from a model point of view, but it strongly suggests that the device was heavily modified when V_{DC}=-1.4 V was applied. This

strong modification occurring in the device results in very noisy data points which make the fitting much more difficult, if compared for example to the one of the precedent measurement at V_{DC}=-1.2 V, as it is shown in Figure B.19. Figure B.18 highlights also that an higher magnitude of V_{DC} was required for setting the device in LRS with respect to what is reported in Section 4.3.1. We attribute this result to the larger post-forming resistance of this device with respect to the one analyzed in Section 4.3.1, as well as to the limitation of the DC current provided by the precision LCR meter *HP 4284A*, as reported in [41].

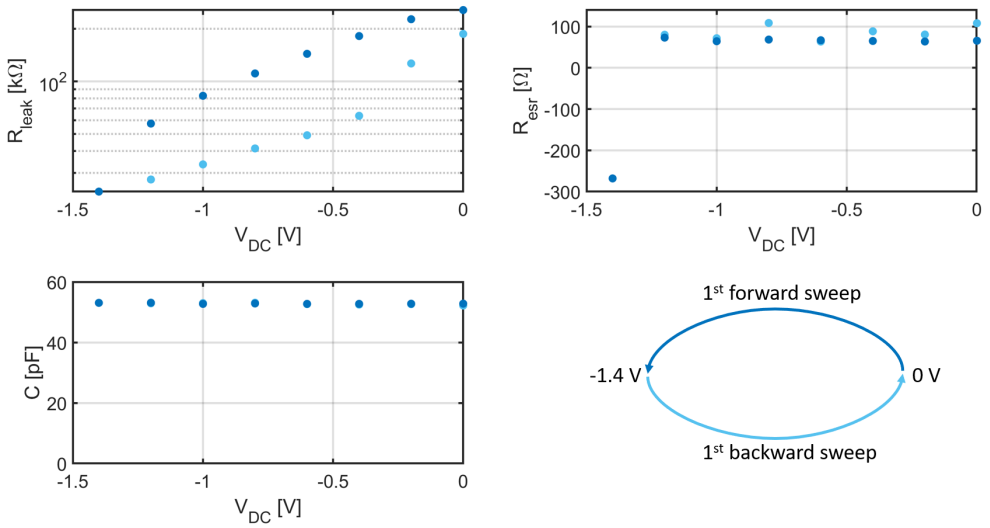


Figure B.18: Evolution of the three model parameters C, R_{leak} and R_{esr} during a bidirectional voltage sweep on V_{DC} from 0 V to -1.4 V for a (60 × 60) µm^2 square device with a small VIA in the corner.

FWD V_{DC} [V]	C [pF]	R_{leak} [kΩ]	R_{esr} [Ω]	BWD V_{DC} [V]	C [pF]	R_{leak} [kΩ]	R_{esr} [Ω]
0.0 V	52.8	256.0	65.7	0.0 V	52.2	187.0	108.0
-0.2 V	52.8	227.0	63.8	-0.2 V	52.7	127.0	80.7
-0.4 V	52.8	182.0	65.1	-0.4 V	52.5	63.6	88.5
-0.6 V	52.8	144.0	66.9	-0.6 V	52.7	49.2	63.6
-0.8 V	52.9	111.0	68.5	-0.8 V	53.1	41.4	109.0
-1.0 V	52.7	82.8	64.4	-1.0 V	53.0	33.6	71.7
-1.2 V	53.1	57.4	73.1	-1.2 V	52.9	27.5	80.1
-1.4 V	53.1	23.5	-268.0	---	---	---	---

Table B.2: Values assumed by the three model parameters C, R_{leak} and R_{esr} during a bidirectional (FWD=forward, BWD=backward) voltage sweep on V_{DC} from 0 V to -1.4 V for a (60 × 60) µm^2 square device with a small VIA in the corner.

A second voltage sweep on V_{DC} from -1.4 V to -2.4 V was then applied to the device to achieve a lower LRS. A positive value of R_{esr} was measured at a DC bias

of -1.4 V, thus suggesting the stabilization of the device state with respect to the previous voltage stimulation. However, even if coherently with the proposed model the R_{leak} was decreasing by reducing V_{DC} to -1.8 V, the R_{esr} was also decreasing. This can be explained by the still non-optimal fitting because of the noisy data points collected or also it can be attributed to the appearance of some competing mechanisms that tend to wider the filament. An example of this competing mechanism could be the incorporation of oxygen ions in the Pt grains (most probably due to physisorption or chemisorption within grain boundaries [58]) which results in the creation of additional oxygen vacancies that reduce the energy barrier at the BE interface.

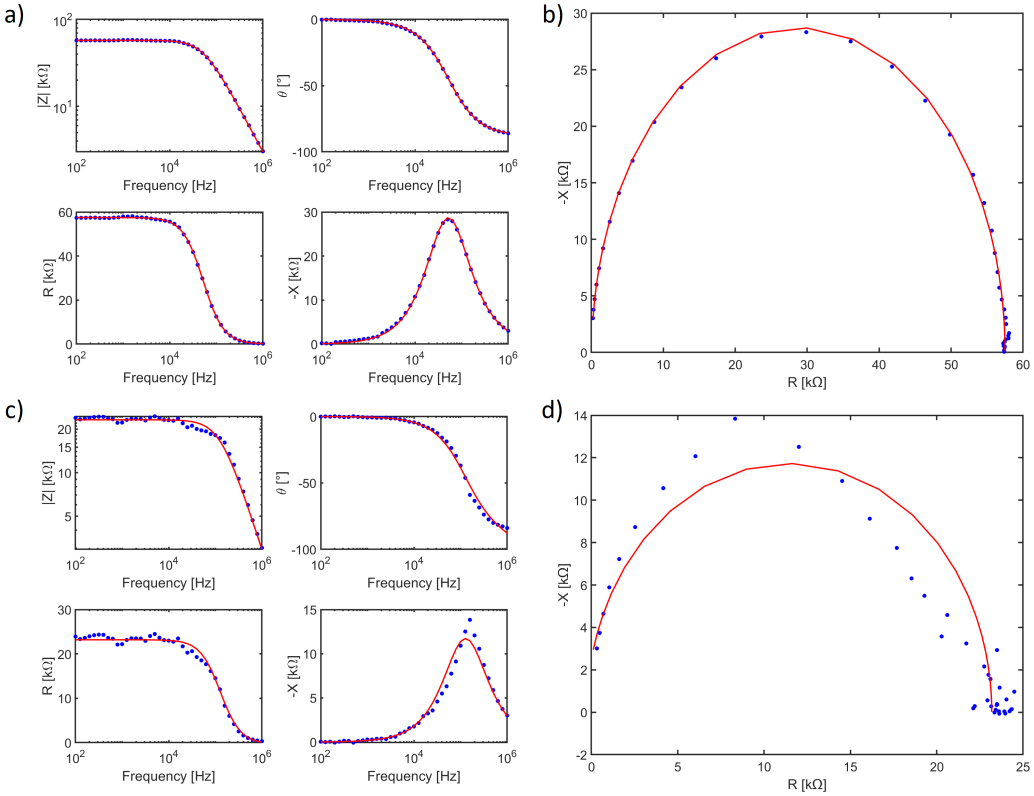


Figure B.19: Fit result for a (60 × 60) µm^2 square device with a small VIA in the corner when a DC bias voltage of -1.2 V(a-b) or -1.4 V(c-d) is applied. The fit result is represented as bode plots (a-c) and Nyquist plot (b-d).

The jump in the equivalent series resistance to 498.0 Ω when -2.0 V are applied may suggest the dominance among the competing processes of the filament thinning. Moreover, when the DC bias reached the value of -2.4 V, an abrupt increase

of the leakage resistance was observed while R_{esr} dropped down. The strongly correlated change in the two resistance contributions can be explained by the breaking of the filament as consequence of the huge oxygen vacancies recombination induced by the very large negative voltage applied. This hypothesis is supported by the values assumed by the R_{leak} during the second backward sweep, which were in the order of hundreds of kΩ or even MΩ. Indeed, the gap spacing between the active electrode and the filament can be modelled as the parallel between a resistor and a capacitor, as already reported in [40]. However, the total capacitance of the system was not subjected to major changes even when the filament breaks, thus further stressing its bulk origin.

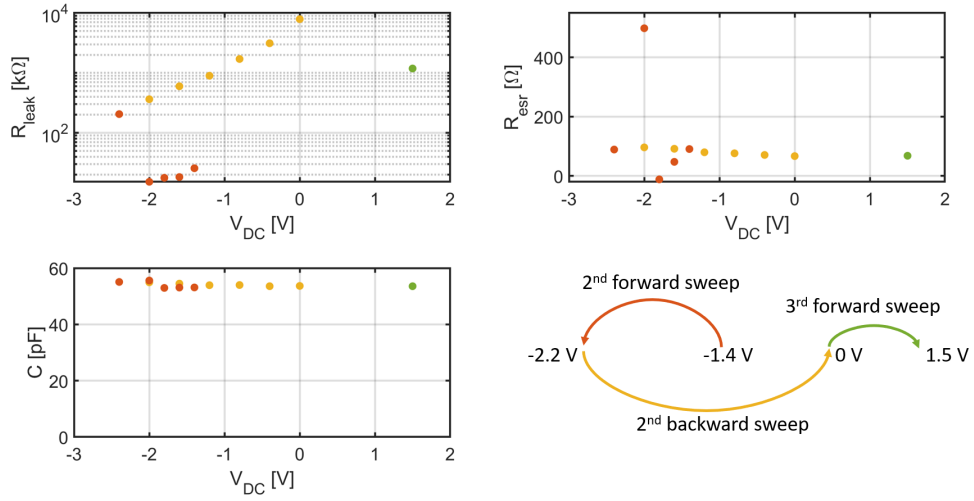


Figure B.20: Evolution of the three model parameters C, R_{leak} and R_{esr} during a voltage sweep on V_{DC} from -1.4 V to -2.4 V and then back to 0 V for a (60 × 60) µm^2 square device with a small VIA in the corner. The values of the model parameters are also reported in the plots when a V_{DC}=1.5 V is applied.

FWD V_{DC} [V]	C [pF]	R_{leak} [kΩ]	R_{esr} [Ω]	BWD V_{DC} [V]	C [pF]	R_{leak} [kΩ]	R_{esr} [Ω]	FWD V_{DC} [V]	C [pF]	R_{leak} [kΩ]	R_{esr} [Ω]
-1.4 V	53.1	25.6	90.6	0.0V	53.7	$7.9 \cdot 10^3$	66.7	1.5 V	53.6	$1.2 \cdot 10^3$	67.9
-1.6 V	53.1	18.3	47.4	-0.4 V	53.6	$3.1 \cdot 10^3$	70.7	---	---	---	---
-1.8 V	53.0	17.7	-11.8	-0.8 V	54.0	$1.7 \cdot 10^3$	76.2	---	---	---	---
-2.0 V	55.6	15.2	498.0	-1.2 V	53.9	896.0	79.7	---	---	---	---
-2.4 V	55.1	205.0	88.9	-1.6 V	54.5	598.0	91.5	---	---	---	---
---	---	---	---	-2.0 V	55.0	363.0	96.2	---	---	---	---

Table B.3: Values assumed by the three model parameters C, R_{leak} and R_{esr} during a voltage sweep on V_{DC} from -1.4 V to -2.4 V and then back to 0 V for a (60 × 60) µm^2 square device with a small VIA in the corner. The values of the model parameters at V_{DC}=1.5 V are also reported.

When the filament breaks, the model remains in first approximation the one for a first-order system with the contributions to the parameters that are those illustrated in Figure B.21.

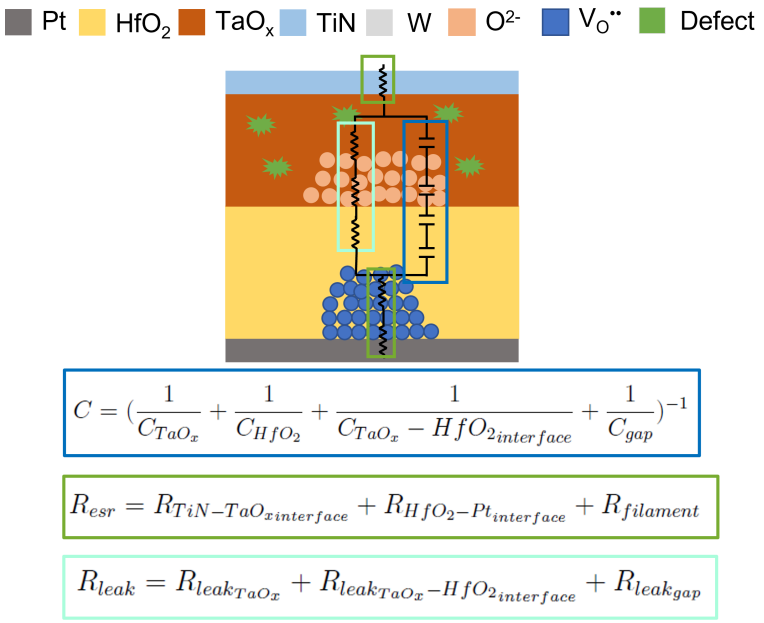


Figure B.21: Graphical representation of the contributions to the EEC model parameters in a TaO_x/HfO_2 bilayer ReRAM device when the filament breaks.

Finally, if a positive DC bias of 1.5 V was applied, the oxygen ions were gathered in the TaO_x layer, thus leaving behind additional oxygen vacancies that started to fill the gap, as it is observable from the reduction of the R_{leak} in Figure B.20. However, the still high value of R_{leak}, as well as the unchanged value of R_{esr}, suggests that the filament is not yet "plugged" to the CMO layer.
The experiment discussed in this section strongly supports the proposed filamentary-based model, since it demonstrates that the model can describe also a device of a different size and a different post-forming resistance with respect to the one analyzed in Section 4.3.6.

B.8 Analysis of the relaxation time constant in TaO_x/HfO_2 bilayer ReRAM device while switching

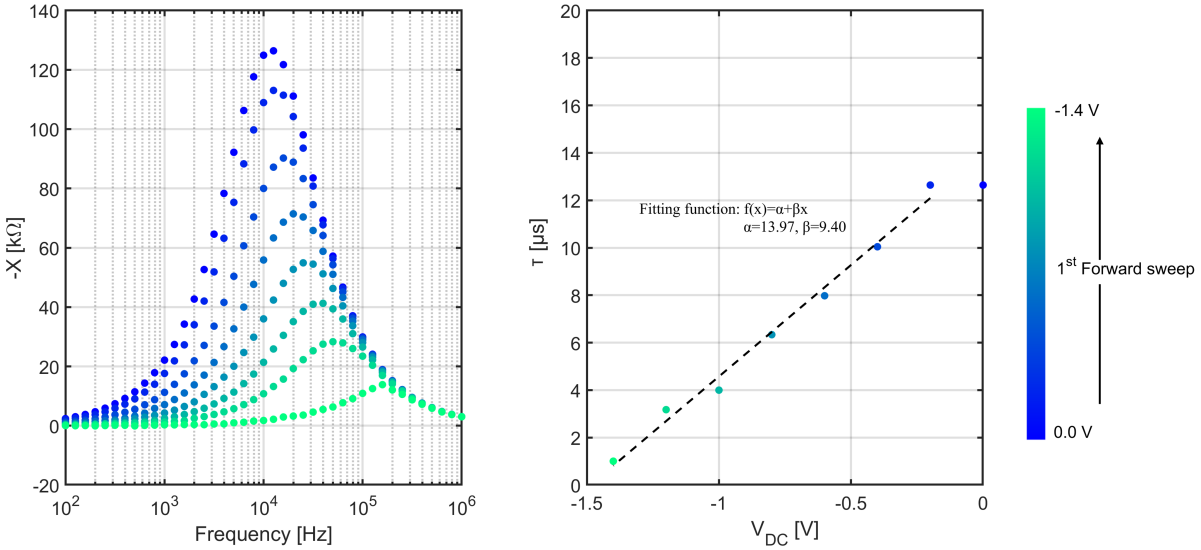
A deeper analysis of the data collected during the sweeps on the DC bias voltage V_{DC} reported in Section B.7 highlights a shift in the frequency f_{max} at which the maximum in the imaginary part of the impedance is observed. The shift can be

visualized for each of the DC bias sweeps in Figure B.22 and it indicates a single relaxation process in the device. A relaxation time constant τ can be computed for each measurement accordingly to equation (B.7) [59] and its evolution during each sweep is reported in Figure B.22.

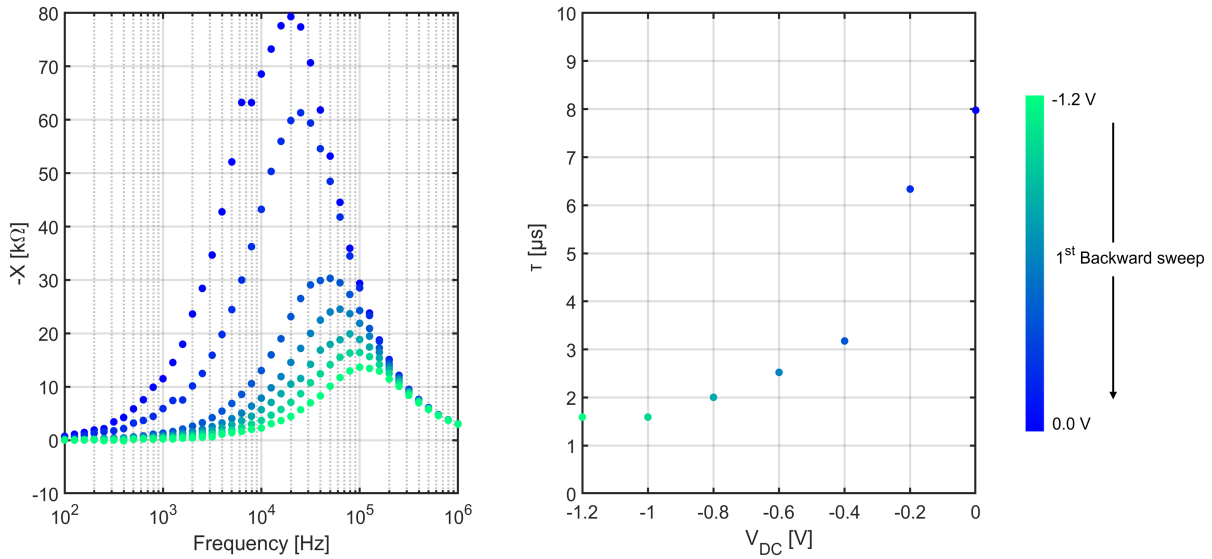
$$\tau = \frac{1}{2\pi f_{max}} \tag{B.7}$$

Figure B.22 shows that τ decreases by increasing the magnitude of the V_{DC} applied to the device, with an interesting linear trend for what concerns the first voltage sweep applied to the device. This behaviour of τ is associated to an enhancement of the electrical conductivity in the device, and it is perfectly in agreement with similar observations found in literature on Metal-Insulator-Semiconductor (MIS) structures [59]. In [59], the phenomena is explained as a further carrier injection in the capacitor. However, being the ReRAM device more complex than a simple capacitor, this explanation must be refined considering more advanced analyses. Indeed, the presence of the filament in the HfO_2 suggests to attribute the observed decrease of the relaxation time constant to solely a larger transition probability in the CMO, as already observed in other semiconducting transition metal oxide like TiO_2 [60]. This argumentation leads to classify the TaO_x as a relaxation semiconductor, which is defined by the condition that the dielectric relaxation time is longer than the free carrier lifetime [61]. Nevertheless, this also means that a rupture in the filament should not play any role on the time constant. However, Figure B.22 (c) and (d) show that when the filament is ruptured, τ increases. This result highlights a filament contribution and when it breaks, the carrier conduction becomes much more difficult, as it is expected.
Making a complete reasoning about the observed phenomena with this simple analysis is for sure premature but the proposed results can be a starting point for more advanced analyses which may lead to the full identification of the conduction mechanisms taking place in the bilayer ReRAM devices.

a)


-X [kΩ]
Frequency [Hz]
τ [μs]
Fitting function: f(x)=α+βx
α=13.97, β=9.40
V_{DC} [V]
-1.4 V
1st Forward sweep
0.0 V

b)


-X [kΩ]
Frequency [Hz]
τ [μs]
V_{DC} [V]
-1.2 V
1st Backward sweep
0.0 V

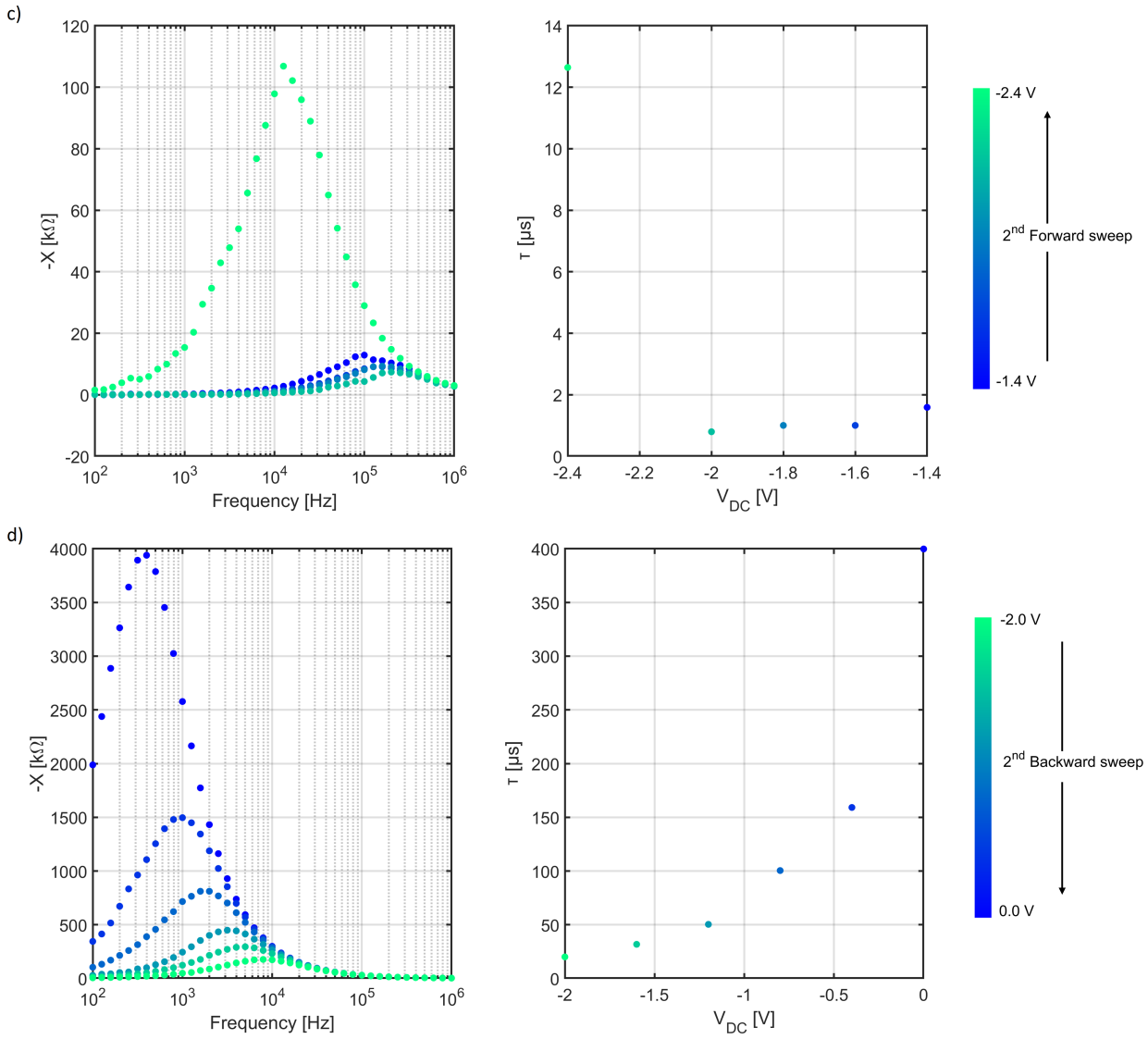


Figure B.22: Plot of the imaginary part X of the impedance for a (60 × 60) µm^2 square device with a small VIA in the corner as function of the measurement frequency and the computed relaxation time constant τ during a voltage sweep on V_{DC} from 0 V to -1.4 V (a), from -1.2 V to 0 V (b), from -1.4 V to -2.4 V (c) and from -2 V to 0 V (d).

Bibliography

[1] I.H. Im, S.J. Kim, H.W. Jang, *Memristive Devices for New Computing Paradigms*, Advanced Intelligent Systems, vol. 2, p. 2000105, 2020.

[2] S. Yu, *Neuro-inspired computing with emerging nonvolatile memorys*, Proceedings of the IEEE, vol. 106, pp. 260-285, 2018.

[3] M.A. Zidan, J. Strachan, W. Lu, *The future of electronics based on memristive systems*, Nature Electronics, vol. 1, pp. 22-29, 2008.

[4] A. Sokolov, H. Abbas, Y. Abbas, C. Choi, *Towards engineering in memristors for emerging memory and neuromorphic computing: A review*, Journal of Semiconductors, vol. 42, 2021.

[5] L.O. Chua, *Memristor-the missing circuit element*, IEEE Transactions on Circuit Theory, vol. 18, pp. 507-519, 1971.

[6] D. Strukov, G. Snider, D. Stewart, R. Williams, *The Missing Memristor Found*, Nature, vol. 453, pp. 80-83, 2008.

[7] J. Woo, S. Yu, *Resistive Memory-Based Analog Synapse: The Pursuit for Linear and Symmetric Weight Update*, IEEE Nanotechnology Magazine, vol. 12, pp. 36-44, 2018.

[8] A. Sawa, *Resistive Switching in Transition Metal Oxides*, Materials Today, vol. 11, pp. 28-36, 2008.

[9] D. Christensen, R. Dittmann, B. Linares-Barranco, A. Sebastian, M.L. Gallo, A. Redaelli, S. Slesazeck, T. Mikolajick, S. Spiga, S. Menzel, I. Valov, G. Milano, C. Ricciardi, S. Liang, F. Miao, M. Lanza, T. Quill, S. Keene, A. Salleo, J. Grollier, D. Marković, A. Mizrahi, P. Yao, J. Joshua Yang, G. Indiveri, J. Strachan, S. Datta, E. Vianello, A. Valentian, J. Feldmann, X. Li, W. Pernice, H. Bhaskaran, E. Neftci, S. Ramaswamy, J. Tapson,

F. Scherr, W. Maass, P. Panda, Y. Kim, G. Tanaka, S. Thorpe, C. Bartolozzi, T.A. Cleland, C. Posch, S. Liu, A.N. Mazumder, M. Hosseini, T. Mohsenin, E. Donati, S. Tolu, R. Galeazzi, M.E. Christensen, S. Holm, D. Ielmini, N. Pryds, *2021 Roadmap on Neuromorphic Computing and Engineering*, ArXiv, vol. abs/2105.05956, 2021.

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

[11] A. Schnöhals, C.M.M. Rosário, S. Hoffmann-Eifert, R. Waser, S. Menzel, D. Wouters, *Role of the Electrode Material on the RESET Limitation in Oxide ReRAM Devices*, Advanced Electronic Materials, vol. 4, p. 1700243, 2017.

[12] U. Celano, J. Op de Beeck, S. Clima, M. Luebben, P. Koenraad, L. Goux, I. Valov, W. Vandervorst, *Direct Probing of the Dielectric Scavenging-Layer Interface in Oxide Filamentary-Based Valence Change Memory*, ACS Applied Materials & Interfaces, vol. 9, pp. 10820-10824, 2017.

[13] H. Zhang, S. Yoo, S. Menzel, C. Funck, F. Cüppers, D. Wouters, C. Hwang, R. Waser, S. Hoffmann-Eifert, *Understanding the Coexistence of Two Bipolar Resistive Switching Modes with Opposite Polarity in Pt/TiO_2/Ti/Pt Nano-sized ReRAM Devices*, ACS Applied Materials & Interfaces, vol. 10, pp. 29766-29778, 2018.

[14] F. Zahoor, T.Z.A. Zulkifli, F.A.. Khanday, *Resistive Random Access Memory (RRAM): an Overview of Materials, Switching Mechanism, Performance, Multilevel Cell (mlc) Storage, Modeling, and Applications*, Nanoscale Research Letters, vol. 15, 2020.

[15] F. Cüppers, S. Menzel, C. Bengel, A. Hardtdegen, M.V. Witzleben, U. Böttger, R. Waser, S. Hoffmann-Eifert, *Exploiting the switching dynamics of HfO_2-based ReRAM devices for reliable analog memristive behavior*, APL Materials, vol. 7, p. 091105, 2019.

[16] J. Woo, K. Moon, J. Song, S. Lee,M. Kwak, J. Park, H. Hwang, *Improved Synaptic Behavior Under Identical Pulses Using AlO_x/HfO_2 Bilayer RRAM Array for Neuromorphic Systems*, IEEE Electron Device Letters, vol. 37, pp. 994-997, 2016.

[17] S. Kim, Y. Abbas, Y. Jeon, A. Sokolov, B. Ku, C. Choi, *Engineering synaptic*

characteristics of TaO_x/HfO_2 bi-layered resistive switching device, Nanotechnology, vol. 29, p. 415204, 2018.

[18] A. Hardtdegen, C. La Torre, F. Cüppers, S. Menzel, R. Waser, S. Hoffmann-Eifert, *Improved Switching Stability and the Effect of an Internal Series Resistor in HfO_2/TiO_x Bilayer ReRAM Cells*, IEEE Transactions On Electron Devices, vol. 65, pp. 3229-3236, 2018.

[19] D.C. Sekar, B. Bateman, U. Raghuram, S. Bowyer, Y. Bai, M. Calarrudo, P. Swab, J. Wu, S. Nguyen, N. Mishra, R. Meyer, M. Kellam, B. Haukness, C.J. Chevallier, H. Wu, H. Qian, F. Kreupl, G. Bronner, *Technology and circuit optimization of resistive RAM for low-power, reproducible operation*, IEEE International Electron Devices Meeting, pp. 28.3.1-28.3.4, 2014.

[20] W. Wu, H. Wu, B. Gao, N. Deng, S. Yu, H. Qian, *Improving Analog Switching in HfOx-Based Resistive Memory With a Thermal Enhanced Layer*, IEEE Electron Device Letters, vol. 38, pp. 1019-1022, 2017.

[21] P. Yao, H. Wu, B. Gao, S. Eryilmaz, X. Huang, W. Zhang, Q. Zhang, N. Deng, L. Shi, H. Wong, H. Qian, *Face classification using electronic synapses*, Nature Communications, vol. 8, 2017.

[22] J.D. Plummer, M.D. Deal, P.B. Griffin, *Silicon VLSI technology: Fundamentals, Practice, and Modeling*, Upper Saddle River, NJ: Prentice Hall, 2000.

[23] J.X.J. Zhang, K. Hoshino, *Fundamentals of nano/microfabrication and scale effect*, Molecular Sensors and Nanodevices, pp. 43-111, 2019.

[24] Oxford Instruments, *Plasma Enhanced Chemical Vapour Deposition (PECVD)*, accessed 24 May 2021, `https://plasma.oxinst.com/technology/pecvd`.

[25] Oxford Instruments, *Atomic Layer Deposition (ALD)*, accessed 4 July 2021, `https://plasma.oxinst.com/technology/atomic-layer-deposition`.

[26] Y.J. Ji, K.S. Kim, K. Kim, J. Byun, G. Yeom, *A Brief Review of Plasma Enhanced Atomic Layer Deposition of Si_3N_4*, Applied Science and Convergence Technology, vol. 28, pp. 142-147, 2019.

[27] Heidelberg Instruments, *Maskless Laser Lithography*, accessed 3 August 2021, `https://heidelberg-instruments.com/key-features/maskless-laser-lithography`.

[28] D.R. Hines, N.P. Siwak, L.A. Mosher, R. Ghodssi, *MEMS Lithography and Micromachining Techniques*, *MEMS Materials and Processes Handbook*, Springer, vol.1, 2011.

[29] MicroChem Corp., *LOR Lift-Off Resists*, Datasheet, 2002.

[30] Oxford Instruments, *Reactive Ion Etching (RIE)*, accessed 3 August 2021, `https://plasma.oxinst.com/technology/reactive-ion-etching`.

[31] Oxford Instruments, *Inductively Coupled Plasma Etching (ICP)*, accessed 3 August 2021, `https://plasma.oxinst.com/technology/icp-etching`.

[32] I. Strabrawa, A. Kubala-Kukuś, D. Banaś, G. Pepponi, J. Braziewicz, M. Pajek, M. Teodorczyk, *Characterization of the morphology of titanium and titanium (IV) oxide nanolayers deposited on different substrates by application of grazing incidence X-ray diffraction and X-ray reflectometry techniques*, Thin Solid Films, vol.671, pp. 103-110, 2019.

[33] Ö. Erken, C. Gümüş, *Determination of The Thickness and Optical Constants of Polycrystalline SnO_2 Thin Film by Envelope Method*, Adiyaman University Journal of Science, 2018.

[34] Materials Project, University of California at Berkeley, accessed 30 July 2021, `https://materialsproject.org`.

[35] M. Yasaka, *X-ray thin-film measurement techniques*, The Rigaku Journal, vol. 26, pp. 1-9, 2010.

[36] J.H. Klootwijk, C.E. Timmering, *Merits and Limitations of Circular TLM structures for contact resistance determination for novel III-V HBTs*, Proceedings of the 2004 IEEE International Conference on Microelectronic Test Strucures, pp. 247-252, 2004.

[37] B. Jacobs, M. Kramer, E.J. Geluk, F. Karouta, J. Mullin, *Optimisation of the Ti/Al/Ni/Au ohmic contact on AlGaN/GaN FET structures*, Journal of Crystal Growth, vol. 241, pp. 15-18, 2002.

[38] W. Choi, H. Shin, J.M. Kim, J. Choi, W. Yoon, *Modeling and Applications of Electrochemical Impedance Spectroscopy (EIS) for Lithium-ion Batteries*, Journal of Electrochemical Science and Technology, vol. 11, pp. 1-13, 2020.

[39] V. Balasubramani, S. Chandraleka, T. Rao, R. Sasikumar, M.R. Kuppusamy,

T. Sridhar, *Review-Recent Advances in Electrochemical Impedance Spectroscopy Based Toxic Gas Sensors Using Semiconducting Metal Oxides*, Journal of The Electrochemical Society, vol. 167, p. 037572, 2020.

[40] X. Jiang, Y. Zhao, Y. Chen, D. Li, Y. Luo, D. Zhao, Z. Sun, J. Sun, H.W. Zhao, *Characteristics of different types of filaments in resistive switching memories investigated by complex impedance spectroscopy*, Applied Physics Letters, vol. 102, p. 253507, 2013.

[41] Hewlett Packard, *Precision LCR meter 4284A*, Operation manual, 1998.

[42] A. Kindsmüller, A. Schönhals, S. Menzel, R. Dittmann, R. Waser, D. Wouters, *The influence of interfacial (sub)oxide layers on the properties of pristine resistive switching devices*, Non-Volatile Memory Technology Symposium (NVMTS), p. 1-4, 2018.

[43] Y.M. Zhou, Z. Xie, H. Xiao, P. Hu, J. He, *Effects of deposition parameters on tantalum films deposited by direct current magnetron sputtering*, Journal of Vacuum Science & Technology A, vol. 27, pp. 109-113, 2009.

[44] J. Sarkar, *Sputtering Materials for VLSI and Thin Film Devices*, William Andrew Publishing, 2014.

[45] J.E. Stevens, A. Lohn, S. Decker, B. Doyle, P.R. Mickel, M. Marinella, *Reactive sputtering of substoichiometric Ta_2O_x for resistive memory applications*, Journal of Vacuum Science & Technology A, vol. 32, p. 021501, 2014

[46] C.M.M. Rosário, B. Thöner, A. Schönhals, S. Menzel, A. Meledin, N. Barradas, E. Alves, J. Mayer, M. Wuttig, R. Waser, N. Sobolev, D. Wouters, *Metallic filamentary conduction in valance change-based resistive switching devices: the case of TaO_x thin film with $x \sim 1$*, Nanoscale-The Royal Society of Chemistry, vol. 11, 2019.

[47] A. Sharma, V. Longo, M. Verheijen, A. Bol, W. Kessels, *Atomic layer deposition of HfO_2 using $HfCp(NMe_2)_3$ and O_2 plasma*, Journal of Vacuum Science & Technology A, vol. 35, 2017.

[48] M. Halter, L. Bégon-Lours, V. Bragaglia, M. Sousa, B. Offrein, S. Abel, M. Luisier, J. Fompeyrine, *A back-end, CMOS compatible ferroelectric Field Effect Transistor for synaptic weights*, ACS Applied Materials & Interfaces, vol. 12, 2020.

[49] D. Ielmini, *Resistive switching memories based on metal oxides: mechanisms, reliability and scaling*, Semiconductor Science and Technology, vol. 31, p. 063002, 2016.

[50] C.D. Landon, R. Wilke, M. Brumbach, G. Brennecka, M.A. Blea-Kirby, J. Ihlefeld, M. Marinella, T. Beechem, *Thermal transport in tantalum oxide films for memristive applications*, Applied Physics Letters, vol. 107, p. 023108, 2015.

[51] H. Zhang, S. Yoo, S. Menzel, C. Funck, F. Cüppers, D. Wouters, C. Hwang, R. Waser, S. Hoffmann-Eifert, *Understanding the Coexistence of Two Bipolar Resistive Switching Modes with Opposite Polarity in Pt/TiO_2/Ti/Pt Nanosized ReRAM Devices*, ACS Applied Materials & Interfaces, vol. 10, p. 29766-29778, 2018.

[52] K.R. Williams, K. Gupta, M. Wasilik, *Etch rates for micromachining processing-Part II*, IEEE ASME Journal of Microelectromechanical Systems, vol. 12, pp. 761-768, 2003.

[53] Cascade Microtech, *DCP-HTR Parametric DC Probing of Small Pads*, accessed 28 July 2021, `https://www.cascademicrotech.com/files/DCPHTR-AN.pdf`.

[54] K. Rathee, B. Malik, *Structural and Electrical Properties of Tantalum Pentaoxide (Ta_2O_5) Thin Films - a Review*, International Journal of Modern Physics Conference Series, vol. 22, pp. 564-569, 2013.

[55] A. Gehring, S. Selberherr, *Modeling of Tunneling Current and Gate Dielectric Reliability for Nonvolatile Memory Devices*, Device and Materials Reliability, IEEE Transactions on Device and Materials Reliability, vol. 4, pp. 306-319, 2004.

[56] A. Kindsmuller, A. Schönhals, S. Menzel, R. Dittmann, R. Waser, D. Wouters, *The influence of interfacial (sub)oxide layers on the properties of pristine resistive switching devices*, Non-Volatile Memory Technology Symposium (NVMTS), pp. 1-4, 2018.

[57] S. Dirkmann, J. Kaiser, C. Wenger, T. Mussenbrock, *Filament Growth and Resistive Switching in Hafnium Oxide Memristive Devices*, ACS Applied Materials & Interfaces, vol. 10, pp. 14857-14868, 2018.

[58] D. Cooper, C. Baeumer, N. Bernier, A. Marchewka, C. La Torre, R. Dunin-Borkowski, S. Menzel, R. Waser, R. Dittmann, *Anomalous Resistance Hysteresis*

in Oxide ReRAM: Oxygen Evolution and Reincorporation Revealed by In Situ TEM, Advanced Materials, vol. 29, 2017.

[59] A. Buyukbas Uluşan, A. Tataroğlu, *Impedance spectroscopy of Au/TiO_2/n-Si metal-insulator-semiconductor (MIS) capacitor*, Physica B-Condensed Matter, vol. 580, p. 411945, 2020.

[60] M. Sriubas, K. Bočkutė, D. Virbukas, G. Laukaitis, *Dynamics of electrical charge carriers in Mg-doped TiO_2 thin films under reducing conditions*, Journal of Vibroengineering, vol. 17, pp. 3333-3340, 2015.

[61] N.M. Haegel, *Relaxation Semiconductors: In Theory and in Practice*, Applied Physics A, vol. 53, pp. 1-7, 1991.

[62] F. Völklein, H. Reith, A. Meier, *Measuring methods for the investigation of in-plane and cross-plane thermal conductivity of thin films*, Physica Status Solidi (a), vol. 210, pp. 106-118, 2013.

[63] F. Chiu, *A Review on Conduction Mechanisms in Dielectric Films*, Advances in Materials Science and Engineering, vol. 2014, pp. 1-18, 2014.

[64] Z. Wang, H. Yu, X. Tran, Z. Fang, J. Wang, H. Su, *Transport properties of HfO_{2-x} based resistive-switching memories*, Physical Review B, vol. 85, 2012.

[65] D. Cooper, C. Baeumer, N. Bernier, A. Marchewka, C. La Torre, R. Dunin-Borkowski, S. Menzel, R. Waser, R. Dittmann, *Anomalous Resistance Hysteresis in Oxide ReRAM: Oxygen Evolution and Reincorporation Revealed by In Situ TEM*, Advanced Materials, vol. 29, 2017.