
Politecnico di Torino

# Resistive switching in ferroelectric synaptic weights

Tesi di Laurea Magistrale in Nanotechnologies for ICTs

**Supervisors**:

Dr. Laura Bégon-Lours (IBM Research Zurich)
Prof. Carlo Ricciardi (Politecnico di Torino)

**Author**:

Elisabetta Morabito
Student ID: s292468

Academic Year: 2022-23

# Abstract

The Von Neumann architecture has determined the characteristics and working principles of classical computing paradigms since 1945. CMOS technology had a major role in the improvement of modern computers thanks to advancements in device scaling. However, recent years saw the advancement in deep learning algorithms built on brain-inspired networks of artificial neurons, also referred to as artificial neural networks, concurrently with an exponential increase of processed data, representing a big challenge for conventional computer hardware. The "bottleneck" of the Von Neumann architecture is due to the unavoidable movement of data between the CPU and memory, causing latency and high energy consumption. Neuromorphic computing allows the hardware implementation of the "in-memory" computing paradigm taking advantage of cross-point technologies where the data can be processed and stored within the same location. This is possible thanks to beyond-CMOS devices known as memristors, whose programmable conductance states can be used to encode synaptic weights for implementing artificial neural networks algorithms such as deep neural networks and spiking neural networks. Among these, ferroelectric Tunnel Junctions are a great candidate for the hardware realization of spiking neural networks.

In this work, FTJ synaptic weights built with two different ferroelectric materials - $HfZrO_4$ and $BiFeO_3$ - are discussed. First, the fabrication processes of crossbar arrays based on the two distinct FTJs are presented. Then, particular attention is given to the $BiFeO_3$ crossbar array: electrical characterizations of the fully processed devices show a lower On/Off ratio than what is observed in the simplified test structures on the blanket film. Thus, temperature-dependent measurements have been analyzed in order to study the conduction mechanisms across the device, and an investigation of the resistive switching mechanism is reported. The oxidation of Ni during the processing affects the polarity of the FTJ and the On/Off ratio, which becomes comparable to that of CMOS-compatible $HfZrO_4$ junctions. Finally, the comparison between the two materials doesn't show any relevant advantage in using one technology over the other.

# Acknowledgements

This Master's thesis work was carried out in the IBM Research Zurich laboratories, with the supervision of Dr. Laura Bégon-Lours in the Neuromorphic Devices and Systems group.

First of all, I express my most sincere gratitude to Dr. Laura Bégon-Lours for the opportunity to join the group and for the trust you gave me in collaborating with your projects. Thank you for all the help and support over these months, and for contributing in my professional growth with valuable knowledge and expertise. You created a stimulating and collaborative environment thanks to your professionalism, but still leaving space for friendly chats and laughs.

I am very grateful to my office mate Donato Francesco Falcone for always helping me, giving me precious pieces of advice and supporting me as a friend. I really appreciated all the deep conversations but also the small talks we had, especially during the long travels in the car.

I am thankful to all the members of the Neuromorphic Devices and Systems group, who contributed to creating an inclusive and friendly environment, where to work with professionality and motivation. Thank you for the knowledge you shared and for all the conversations during lunch or in front of a coffee. Any of you had an important role in building my overall experience.

I would also like to show my gratitude to the OpTeam of the Binnig and Rohrer Nanotechnology Center (BRNC), who always gave me a useful hand during the cleanroom work.

I want to thank Prof. Carlo Ricciardi as my Master's thesis supervisor at Politecnico di Torino.

Last, but not for importance, I want to mention my family for the best support I could ever ask for, and all the important people in my life who contributed, and still contribute, to make my world brighter.

# Contents

**Abstract** **i**

**Acknowledgements** **iii**

**Contents** **v**

**1 Introduction** **1**

**2 Theoretical background** **5**

2.1 Neuromorphic computing . . . . . 5
2.2 Artificial Neural networks . . . . . 7
2.3 Memristor crossbar arrays . . . . . 11
2.4 Ferroelectricity . . . . . 15
2.5 Ferroelectric memories . . . . . 16
2.5.1 Ferroelectric Tunnel Junctions (FTJ) . . . . . 18

**3 Methods** **23**

3.1 Temperature dependent DC electrical measurements . . . . . 23
3.1.1 Conduction mechanisms in Dielectric films . . . . . 24
3.1.2 Electrode-limited conduction mechanisms . . . . . 24
3.1.3 Bulk-limited conduction mechanisms . . . . . 30
3.2 Processing . . . . . 37
3.2.1 Photolithography . . . . . 37
3.2.2 Deposition methods . . . . . 40
3.2.3 Etching methods . . . . . 43
3.2.4 Annealing methods . . . . . 44

**4 Passive crossbar arrays** **47**

4.1 HZO crossbar array . . . . . 47
4.1.1 Ferroelectric HZO . . . . . 47

4.1.2 Nanofabrication . . . . . 49
4.2 BFO crossbar array . . . . . 53
4.2.1 Ferroelectric BFO . . . . . 53
4.2.2 Nanofabrication . . . . . 56
4.2.3 Impact of the metal contact processing on the ferroelectric resistive switching . . . . . 58

**5 Analysis of the resistive switching mechanisms in the BFO FTJs 61**
5.1 Conduction mechanisms in BFO FTJ . . . . . 61
5.2 Analysis from low to medium fields . . . . . 62
5.2.1 The SCLC model . . . . . 62
5.2.2 Analysis in the Ohmic regime . . . . . 64
5.2.3 Analysis of the trap energy distribution in the EDT-SCLC regime . 66
5.2.4 Analysis of the trap density in the EDT-SCLC regime . . . . . 68
5.3 Analysis in the high field region . . . . . 70
5.3.1 Analysis in the TAT regime . . . . . 70
5.3.2 Analysis in the FNT regime . . . . . 71
5.4 Equivalent circuit model . . . . . 72
5.5 Resistive switching mechanism . . . . . 75

**6 Conclusions and future developements 79**

**Bibliography 81**

**List of Figures 91**

**List of Tables 99**

# 1 | Introduction

The past few decades have been characterized by an incomparable exponential growth of electronics and computing capacity which radically changed human habits. [1]. Thanks to the advancement in complementary metal-oxide-semiconductor (CMOS) technology, due to the reduction in device size following Moore's law and Dennard scaling, computer performance has improved year by year [2]. However, the past few years witnessed the limits of the classical computer architecture. At the device level, performance gains are no more straightforward to obtain due to increasing leakage current as the size of the transistor approaches physical limits [3]. From an architectural point of view, the constant movement of data between the central processing unit (CPU) and memory in the so-called "von Neumann" architecture, causes a significant decrease in the speed and energy efficiency [4].

The modern computing architecture was founded in 1945 by J. Von Neumann, inspired by a conceptual calculator structure suggested by Turing in 1936 [5]. In the Von Neumann architecture, the processor and memory regions are physically separated, and data are transferred via bridged buses [2]. The high cost of the Von Neumann architecture has been defined as the "Von Neumann bottleneck" or "memory wall" problem. This problem is characterized by three main aspects: (1) Latency deriving from data movement between memory arrays and the CPU; (2) Limited bandwidth of data movement in memory hierarchies; (3) high energy consumption [6].

With the advent of Artificial Intelligence (AI), Machine learning (ML) methods like Artificial neural networks (ANNs) and deep neural networks (DNNs) have become increasingly popular, concurrently with the exponential increase of data [3]. Predictions show that the amount of digital data to be analyzed and processed will be one million times greater than the current amount by 2040 [7]. The need for those networks to be trained on huge datasets constitutes a big challenge for conventional computer hardware based on the sequential Von Neumann architecture, which has limits in performance and energy efficiency [8].

The von Neumann bottleneck can be mitigated by increasing the number of parallel computing elements in the system [9]. However, real improvements couldn't be obtained keeping the same limited memory bandwidth [10]. State-of-the-art deep learning algorithms are trained with graphic processing units (GPUs) accelerators and several application-specific integrated circuits (ASIC) based on CMOS technology have been developed. However, they are not well suited for high-energy demanding tasks such as NN training [9, 11, 12]. In this context, a new computing paradigm called processing-in-memory (PIM) technology can be a promising solution to overcome the Von Neumann bottleneck. The idea consists of bringing memory and computing physically close, thus improving the efficiency of data movement [6].

Taking advantage of this "in-memory" concept, new architectures developed around emerging non-volatile memory devices, allowing for on-chip high-density storage and fast low-power parallel analog computing. Among these, some are able to achieve multilevel resistance states, emulating biological synapse functionality. Those emerging devices lie under the concept of "memristor", a beyond-CMOS electronic device whose controllable conductance states can be used to encode synaptic weights. In fact, this change in conductivity follows similar mechanisms that characterize the biological synapses at the base of the neuronal activity in the human brain. Thus, it has been regarded as an excellent technology suited for brain-inspired computing [3, 12]. Memristor crossbar arrays can also perform PIM computation by utilizing Ohm's law for multiplication and Kirchoff's current law for addition [6]. Those mathematical operations correspond to the fundamental multiply-accumulate (MAC) operation, or dot-product, which is the core of Vector-Matrix Multiplication (VMM). VMM is crucial in the operation of ANNs, but consumes a lot of energy if done using traditional architectures. However, in memristor crossbar arrays VMM is executed in the analog domain with less energy consumption and high parallelism [12–14]. Memristors are used to implement synaptic weight functionality in both non-spiking ANNs and Spiking-Neural-Networks (SNNs) [4]. The former aims at maximizing the parallel computation, and neuron values are encoded by binary bits or by the number of voltage levels; data in SNNs are encoded based on the timing and frequency of pre and post-synaptic spikes, making it more similar to actual biological nervous systems and more energy-efficient than other neural networks [1, 12].

The Ferroelectric Tunnel Junction is a structure comprising of a ferroelectric barrier that divides two conductive electrodes. It is a potential solution for creating non-volatile memristors, and several different ferroelectric materials have been shown to be suitable for FTJs to mimic brain-like operations [3].

This thesis will first discuss the fabrication of ferroelectric synaptic weights based on two different technologies, $HfZrO_4$ and $BiFeO_3$ ferroelectric tunnel junctions, and in particular a comprehensive analysis of the resistive switching mechanism in $BiFeO_3$ synaptic weights is reported. A brief overview of the theory behind neuromorphic computing, memristors and ferroelectric memories will be discussed in Chapter 2. In Chapter 3 we introduce the relevant methods employed to conduct research for this thesis. Then, the fabrication of memristor crossbar arrays based on $HfZrO_4$ and $BiFeO_3$ is reported in Chapter 4, with a discussion on possible applications of the two technologies and the material-related differences. The electrical characterization of $BiFeO_3$ crossbar arrays shows that the resistive switching is characterized by an On/Off ratio (the ratio between the currents in the High and Low Resistive States (HRS, LRS)) which is orders of magnitude smaller than what is observed in the simplified test structures on the blanket film. Chapter 5 will investigate the origin of this difference through the analysis of temperature-dependent measurements. As a last point, final remarks will be given in Chapter 6.

# 2 | Theoretical background

## 2.1. Neuromorphic computing

Traditional computing architectures experience a major performance bottleneck due to memory-related limitations, as outlined in Chapter 1. In comparison, biological nervous systems demonstrate exceptional capabilities and performance by utilizing stochastic networks of neurons with unreliable computational elements to carry out robust computations. [15]. In this context, neuromorphic computing paradigms are becoming attractive solutions since they allow the implementation of alternative non-Von Neumann architectures, using cutting-edge technologies.

Neuromorphic systems, emulating the biological ones, take advantage of energy-efficient asynchronous and event-driven methods to process information. While in traditional information processing systems CPUs and memory are two distinct and physically separated areas on the chip, both biological and artificial neural processing systems operate as 'in-memory' computing systems, where the neural network synapses simultaneously serve as memory storage and non-linear operators [9].

The research community has been facing a challenge in creating brain-like machines since the emergence of digital computers [16]. Already in 1956, J. Von Neumann was inspired by the brain in one of his works [17]. The first model of a perceptron was proposed in 1958 by Rosenblatt [18] and the retinas in 1970 by Fukushima [19]. The term Neuromorphic can be traced back to the works of C. Mead [20]. Mead was inspired by the way synaptic transmission works in the retina. He realized that transistors, which were only used as digital switches before, could be used in their analog form. This led to the development of neuromorphic circuits, which could simulate neural systems using fewer transistors than digital methods [21].

The concept of neuromorphic computing is tied to its hardware architecture, which initially relied on using analog circuits made of silicon as the fundamental components for constructing SNNs. In modern times, the concept has expanded to include hybrid analog/digital as well as purely digital implementations. Hardware-based neuromorphic ar-

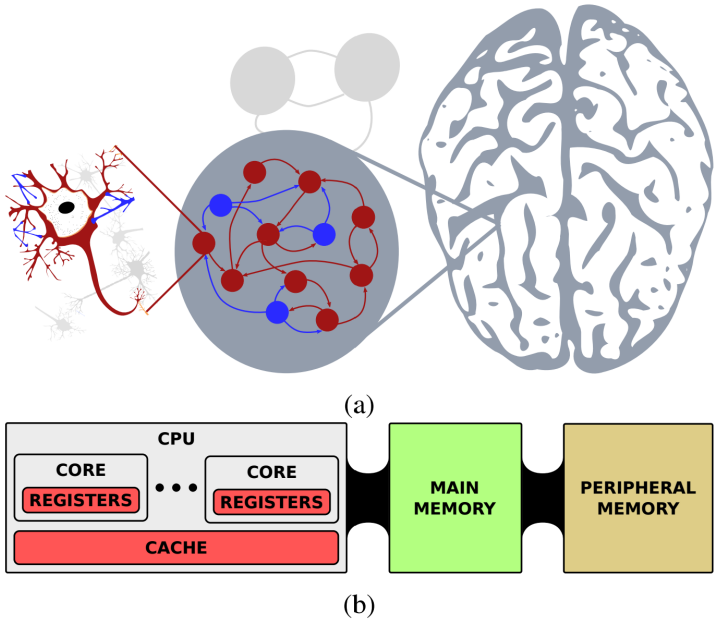



Figure 2.1: Illustration of the brain memory hierarchy with co-localized memory and computation **(a)** vs. **(b)** the classical computing architecture: Central Processing Units (CPUs) containing multiple cores, having each a micro-processor and local memory. This block is connected to the main memory block, the primary storage area, but requires longer access times than the memory blocks in the core. Worst is for peripheral memory block but with higher storage capabilities. Adapted from [9].

chitecture should be differentiated from software-based approaches to classification or recognition tasks, such as deep learning, which is developed within the Von Neumann framework [7].

Deep learning methods are built around the computational model of ANNs having more than two processing layers (also known as deep neural networks), in addition to revolutionary architecture, processing functions, and regularization techniques. ANNs and specifically DNNs models, have shown remarkable abilities to handle complex problems such as speech recognition, visual object recognition, image analysis, language translation, and even self-driving vehicles [22]. The goal of the hardware implementations of data-inspired computing is to supplement the Von-Neumann architecture for those application-specific intelligent tasks, without completely substituting it when it comes to conventional computing [12].

## 2.2. Artificial Neural networks

Artificial neural networks, commonly referred to as ANNs, are designed to realize artificial intelligence by replicating the functionality of biological neural networks in the brain [3]. The brain of humans consists of a vast network of interconnected neurons. Neurons are individual cells that carry out basic functions, such as responding to incoming signals, but when arranged in a network, they have the ability to perform complex tasks like recognizing speech and images. Electrochemical signals between neighboring neurons are transmitted across the so-called synapse (Figure 2.2). This has the effect of raising or lowering the electrical potential inside the body of the receiving cell. If the potential reaches a threshold, a pulse is sent down the axon and the cell is 'fired' [23, 24]. The inherent memory of the brain derives from the overall distribution of the synaptic connection strengths in the network, while learning is achieved through the synapses' reconfiguration known as *synaptic plasticity* [25].

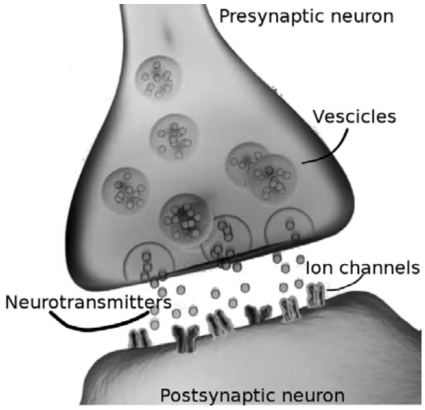



Figure 2.2: Transmission of electrochemical signals through neurotransmitters in a biological synapse. Adapted from [20].

Artificial neural networks (ANN) have been developed as generalizations of mathematical models of biological nervous systems. The basic processing elements of neural networks are called artificial neurons or nodes. The simplified mathematical model for a node in the ANN is shown in Fig. 2.3: the synapses are represented by connection weights that modulate the effect of the associated input signals, while the nonlinear characteristic exhibited by neurons is represented by a transfer function. The neuron output is then computed as the weighted sum of the input signals, transformed by the transfer function [24]. This process can be represented using a mathematical model [23]:

$$y = g\left(\sum_{i=0}^{n} \mathrm{w}_i \mathrm{x}_i - t\right)$$

where y is the output of the node, g is the transfer function, $w_i$ is the weight of input $x_i$. When the weighted sum of the inputs exceeds the threshold value of the node, the neuron is activated and the signal is transferred to the neighboring neurons. The transfer function g has many forms: the step function is the simplest one, however, non-linear transfer functions such as the sigmoid, are more appropriate since only a few problems are linearly separable [23].

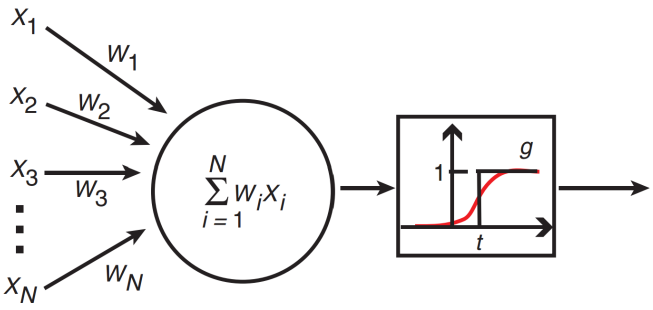



Figure 2.3: Graphical representation of the artificial neuron model. The neuron receives inputs $x_i$ from N other units or external sources, with associated weights $w_i$. The total input to a unit is the weighted sum over all inputs, $\sum_{i=1}^{N} w_i x_i = w_1 x_1 + w_2 x_2 + ... + w_N x_N$. If the total input is below a threshold t, the output of the unit would be 1 and 0 otherwise. The output is determined by the model $y = g\left(\sum_{i=0}^{n} w_i x_i - t\right)$ where the transfer function g is the step function. It can also be, for example, a continuous sigmoid as shown by the red curve. Adapted from [26].

When combining two or more artificial neurons we are getting an artificial neural network, where the nodes are organized into linear arrays, called layers. Figure 2.4 represents an ANN consisting of input, output, and hidden neuron layers [24]. Historically, a neural network with more than one hidden layer is referred to as a deep neural network (DNN), which is the computational framework to implement deep learning algorithms [4]. The basic architecture shown in Figure 2.4 is a feed-forward network, where the signal flow is from input to output units, contrary to recurrent networks which contain feedback connections. However, there are several other neural network architectures depending on the application.

ANNs can be trained to perform a specific task by teaching them how to adjust their weights according to a learning algorithm [24]. When the optimized synaptic weights are loaded into the neural network from an external site, it is called offline training. Online training, on the other hand, involves adjusting the synaptic weights during training in the same network that is used for inference. In addition, if the training samples are labeled according to expected known outputs, it is called supervised learning; otherwise, it is

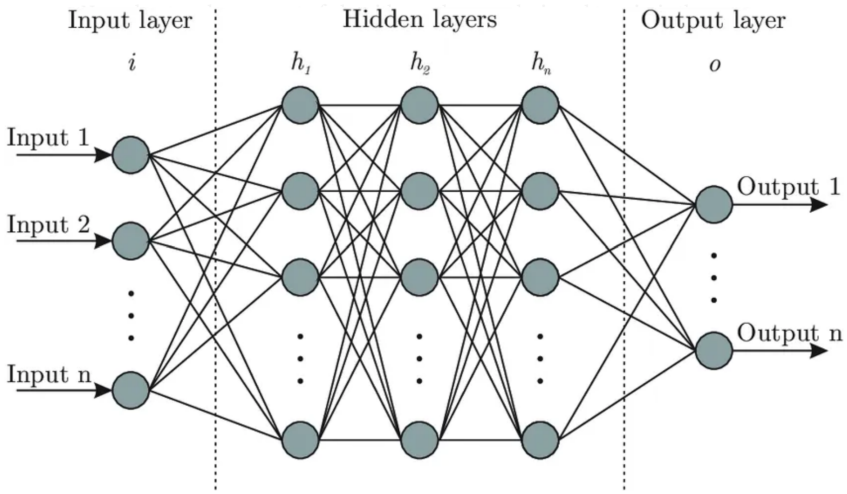



Figure 2.4: Schematic diagram of a feed-forward neural network with a number of hidden layers between input and output neuron layers. Adapted from [27].

unsupervised learning. In addition to these two, there is reinforcement learning where the decision-making process is self-adjusted based on early experiences, in order to maximize rewards for a specific problem [4].

Most of the learning techniques can be defined as variants of the *Hebbian learning* rule. This states that the simultaneous action of two neurons will strengthen their interconnection or weight; however, no bounds are given. The *Perceptron learning rule* is very similar to the Hebbian learning but weights are modified only when the input vectors, initialized with random weights, do not correspond to the output. This technique is only able to handle linearly separable problems. Deep learning, instead, is able to deal with non-linear problems [24]. It is based on the so-called "backpropagation algorithm", in which the difference between the expected (desired) and actual output values is minimized, and the weight is consequently updated but in a sequence that goes from the output to the input layer [7]. Particularly promising is the spike-timing-dependent plasticity (STDP), relying on the evolution of the synaptic strength depending on the timing and causality between the electrical signal of pre- and post-synaptic neurons, a rule observed in mammal's cortex [28]. Since no previous knowledge of the output weights is needed, STDP allows for unsupervised learning [2, 25]

When classifying ANNs it is important to look at the synaptic plasticity and the connectivity structure: neuromorphic hardware is based on the digital non-spiking implementation of DNNs or on the analog spiking behavior of Spiking Neural Networks (SNNs) [12]. Unlike other neural networks, SNNs encode data in the pulse timing and frequency between pre- and post-synaptic spikes. Like biological systems, neurons are only activated when

they detect an input signal, unless they are in a resting state. For this reason, SNNs are extremely energy efficient. SNNs are able to implement spike-timing-dependent plasticity (STDP) learning [2], thus conferring to the network unsupervised learning ability. The weights in DNNs are instead updated according to back-propagation, thus in a supervised training method. In this kind of network, the aim is to emulate the parallel computing feature of the brain to maximize the throughput, and the digital neuron values are encoded by binary bits or by the number of voltage pulses or levels. Nowadays, deep learning is trained in hardware platforms using GPUs accelerators, a combination of hundreds of parallel cores with high memory bandwidths able to increase speed and improve energy efficiency. But parallelization alone cannot solve the challenges related to the data movement between on-chip and off-chip memory and to the digital logic operations still on CPU. Several application-specific integrated circuits (ASIC) based on CMOS technology have been developed to further improve energy efficiency, using static random access memory (SRAM) as synaptic memory that however still needs to communicate with the off-chip memory. SRAM revealed also insufficient storage capacity for the increasing number of parameters of deep learning algorithms, and parallelism is limited by its characteristic row-by-row operation. CMOS-based hardware can be considered to be able to emulate the learning process of the brain. However, the brain efficiency is still unreachable, and the 'in-memory computing' feature is not realizable, partially because most CMOS neuromorphic hardware has been built upon the traditional von Neumann architecture [4]. As an example, the IBM TrueNorth neuromorphic chip based on SNNs shows comparable efficiencies to standard CPUs. Here, the weights need to be pre-trained offline and loaded to SRAM synaptic arrays. On the contrary, SNNs implemented for example in Intel Loihi, show important gains in latency and energy compared to TrueNorth neuromorphic chip [29]. Neuromorphic chips based on SNN implementations combine analog synaptic functionality with analog or digital neurons [9, 11, 12]. Both DNNs and SNNs have been realized with CMOS circuits, but the computational capability of SNNs has not been demonstrated as much as DNNs, mainly because of the lack of efficient algorithms and dynamic devices. The learning accuracy of deep learning based on back-propagation is higher than SNNs when solving traditional classification problems [2, 12].

Beyond digital memory technologies, non-volatile memory emerging devices like memristors (i.e. memory resistors) are attractive for hardware implementation of high-density storage neuromorphic hardware that can be monolithically integrated on CMOS [12, 14, 30]. Organized in a cross-point architecture or crossbar arrays, resistive memories have the intrinsic property to perform vector-matrix multiplication (VMM) locally, realizing the in-memory computing function. Vector matrix multiplication, which is at the core of

deep learning algorithms, is very energy-demanding for traditional digital computing systems. Memristors are able to locally store the weights, encoded in conductance values of each non-volatile element. Furthermore, memristors can be used as the physical weights not only for DNNs but also SNNs, sharing with biological neurons fundamentally similar functioning mechanisms [4]. Thus, the goal is to store and update the weights in a more parallel fashion and locally, replacing the SRAM arrays with the resistive crossbar arrays [12].

## 2.3. Memristor crossbar arrays

In 1971, Leon Chua theorized the existence of a new element, predicted as the fourth hidden fundamental element, in addition to resistor, capacitor, and inductor [2]. This element he called "memristor" was theorized to be a two-terminal device satisfying a non-linear relationship between charge and flux [31]:

$$d\phi = Mdq \tag{2.1}$$

The relationship can be equivalently expressed in terms of voltage and current:

$$M = \frac{d\phi/dt}{dq/dt} = \frac{V}{I} \tag{2.2}$$

According to Eq. 2.2, $M$ is measured in ohm, thus, the device should exhibit a resistance $M$ dependent on the charge flow. The revolutionary concept of the memristor lies in its inherent memory function: whenever current and voltage are zero it does not lose the values of magnetic flux and electric charge, but it will maintain memory of its most recent state. Thus "memristor" derives from a concept of memory resistor, and the same holds for the parameter $M$ defined as "memristance". The distinctive feature of any memristor is the pinched I-V hysteresis loop showed in Figure 2.5: the resistance of the device can be reversibly switched between a higher and a lower value, the so-called *Resistive switching* [3, 31].

A memristor is typically a three-layer system with two terminals, the electrodes, and a 'storage' layer in between. The device has inherent resistor properties but the storage layer can be dynamically reconfigured when stimulated by electrical inputs. This leads to memory effects since the change in resistance can be used to store data and also directly process them [1]. Unlike existing CMOS-based memory technology, which reads volatile capacitance states, memristor technology is a nonvolatile technology. Devices exhibit an

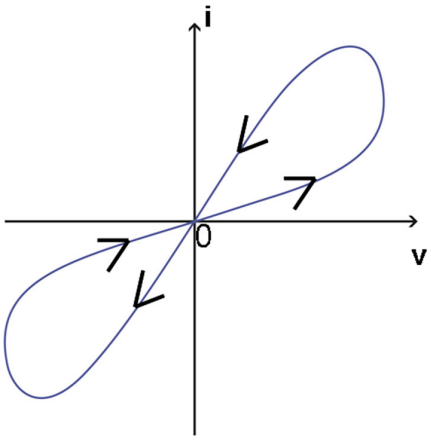



Figure 2.5: Representation of a generic current-voltage characteristic of a memristor, showing pinched hysteresis loop. Adapted from [31].

I-V hysteresis where changes in voltage lead to device switching from a low resistance state (LRS) to a high resistance state (HRS), and vice versa. Moreover, the conductance value can be gradually changed within the area surrounded by the I-V hysteresis; this property can be used to implement synaptic plasticity [2].

In 2008 the first demonstrator was built in Hewlett Packard laboratories, showing resistive switching, as predicted by Chua [2]. The growing interest in memristors brought several ideas for applications for those devices, such as reconfigurable logic and new memory concepts. However, most of the interest led to the development of memristors' capability to emulate the synaptic behavior [32]. Indeed, memristor research is particularly focused on utilizing them in neuromorphic hardware. This derives from the ability of neural networks to handle device non-idealities which include device-to-device variabilities due to fabrication non-uniformity, and cycle-to-cycle variabilities due to the stochastic switching process. Interestingly, this stochasticity is actually a beneficial property that mimics the behavior of biological synapses, acting as a regularizer during training.
Memristors can implement an ultrahigh-density memory layer that can be directly integrated on the processor through dense local interconnects. This eliminates the off-chip connections between memory and processor, thus significantly reducing the memory bottleneck and improving the energy efficiency and speed of the system. In fact, neural networks have been implemented on conventional computing hardware with the synaptic weights being stored in off-chip memory. In contrast, memristor-based implementations possess inherent co-location of memory and computing functions in the same device and a high level of parallelism which significantly improves system efficiency [1].

A memristor neural network can be realized in hardware in a crossbar form where inputs are fed into the rows and the outputs are connected to the columns (Figure 2.6 (c)). Rows and columns in the so-called crossbar array are perpendicular to each other, and the synaptic devices are sandwiched at each crosspoint. Their weights can be encoded by the conductance of the resistive synaptic devices. This architecture has been proposed for the implementation of the MAC operation, the most time-consuming step in neuromorphic algorithms, in a parallel fashion: read voltages are applied to all the rows, and then multiplied by the conductance $G$ of the synaptic devices (through Ohm's law $I = VG$) at each crosspoint, resulting in a weighted sum current in each column (trough Kirchoff's law $I_j = \sum_i V_i G_i$) (Figure 2.6) [1, 12]. In digital logic the multiplication and addition, which are the fundamental operations required by VMM, are realized instead by pipelining multiple adders and multiplier digital blocks [14]. (Figure 2.6 (b)). Thus, the parallelism of a crossbar architecture allows the mapping of the VMM in a single-read operation. Nonetheless, the presence of sneak paths affects the accurate reading of each resistive element individually. On the other hand, memristor-based architectures can read the VMM obtained through physics, which is not affected by sneak paths thanks to the simultaneous polarization of bit and word lines as well as simultaneous access to each cell. Thus, unlike RAM operations, they don't need selectors to access the individual memory cells but rather allow for selector-less passive crossbar operation. Programming the individual cells is however still challenging for sneak paths issues [14].
Communication between arrays is still in the digital domain; moreover, input voltages are better represented by the digital number of pulses or by the pulse width, to avoid non-linearity issues. Large-scale arrays can still face problems with long interconnect resistance and huge power demand from peripheral circuits [12].

A very interesting feature of memristor-based hardware is its compatibility with online learning since the weights can be gradually changed by applied voltage pulses, according to the learning rule. In addition, raw signal processing can be made more efficient by computing in the analog domain, and further reduce latency and chip area by eliminating the need for digital conversion.
Improvements to the device's performance include increasing speed, ON/OFF ratio, cycling endurance, and data retention time. Additional efforts involve the reduction of the operating voltage and current, as well as addressing device variability challenges. A small device footprint for large-scale neural network integration is needed, with a trade-off between scalability and analog synaptic properties to be faced. The realization of 3D crossbar arrays with two-terminal devices is a future goal [1, 12].

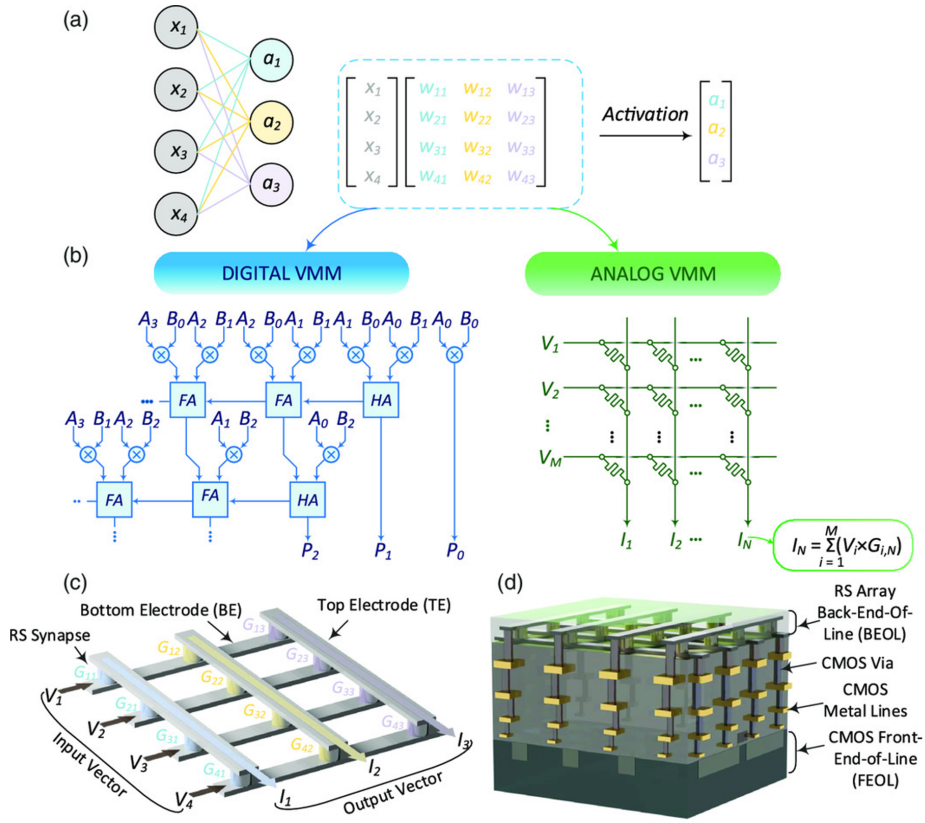



Figure 2.6: **(a)** Basic NN structure and related VMM operation, where the inputs $x_i$ are organized in a vector, the weights $W_i$ in a matrix, and the output of the operation is also represented as a vector. **(b)** Schematic of the digital VMM, obtained by pipelining digital blocks, versus the analog VMM, obtained by summing the currents obtained in each of the N columns. **(c)** Crossbar array where a memristor is formed at each crosspoint and can be used to simultaneously store data and process information ( $V_i$, voltage applied at each row i; $I_j$, current trough column j; $G_{i,j}$, conductance of the memristor at the intersection of row i and column j. **(d)** 3D illustration of how a memristor-based crossbar can be monolithically integrated in the Back-End-Of-Line (BEOL) of a CMOS. Adapted from [14]

The resistive switching (RS) mechanism can be related to various physical reasons, such as growth/fracture of conductive filaments, ferroelectric polarization, electron spin, charge trapping, Schottky barrier, Pool–Frenkel emission and space charge limited current (SCLC) [3]. Ferroelectric non-volatile memories have been recognized with great potentiality in neuromorphic applications. Other concepts have been proposed, such as Resistive random-access memory (RRAM), phase change material (PCM), and magnetic random-access memory (MRAM). The following section will provide an in-depth analysis of the ferroelectric technology.

## 2.4. Ferroelectricity

Discovered in 1921 in Seignette or Rochelle salt, ferroelectricity was known as "Seignette electricity". To understand how ferroelectricity can arise it is reasonable to take a look at its crystallographic structure. There exists a sub-set of dielectric materials whose crystal structure lacks a center of symmetry, i.e. they are non-centrosymmetric [33].
Asymmetric molecules have a dipole moment provided by the following definition [34] :

$$\vec{p} = \int dV \rho(\vec{r})\vec{r} \tag{2.3}$$

where $\rho(\vec{r})$ is the charge density in the molecule. The total dipole moment in a solid is the sum of all single dipole moments. The polarization $\vec{\mathrm{P}}$ is defined as the total dipole moment per unit volume [34].

Ferroelectrics belong to a set of non-centrosymmetric crystals with a dipole moment pointing along a special axis aligned with the crystal, called *polar axis*. For most ferroelectrics, the polar state only exists over a limited range of temperatures below the so-called *Curie temperature* $T_c$. Above this point, a transition occurs from a polar ferroelectric phase to a non-polar paraelectric phase, having higher crystal symmetry, and thus not showing ferroelectricity. Ferroelectrics possess a dielectric polarization in the unit cell of the crystal structure known as spontaneous polarization $\mathrm{P}_s$. The application of a field of sufficient magnitude will cause the spontaneous polarization to switch to a different stable direction, and, upon removal of the field, the polarisation will not spontaneously return to its original direction and magnitude. The polarization is not only dependent on the electric field but also on its history, yielding to the characteristic hysteretic behaviour of those materials. It must be noted that a ferroelectric material can be subdivided, in a simplistic picture, into many microscopic regions having each a homogeneous polarization. Those regions are called domains. There are many reasons for the existence of domains, including non-uniform strain, microscopic defects, and the thermal and electrical history of the sample. But even in an ideal crystal, domains are expected for energetic reasons. In normal conditions domains are randomly oriented, leading to a zero macroscopic polarization of the material, but when an electric field, higher than the so-called coercive field $\mathrm{E}_c$ is applied, they tend to align along its direction. The material will thus show a non-zero macroscopic polarization. In order to change the polarization direction, the ferroelectric domains must overcome an energy barrier equal to the coercive field [33, 34].

In Figure 2.7 we can see the typical hysteresis of the polarization under an applied electric field. Starting from zero, an increase of the field in the positive direction corresponds to

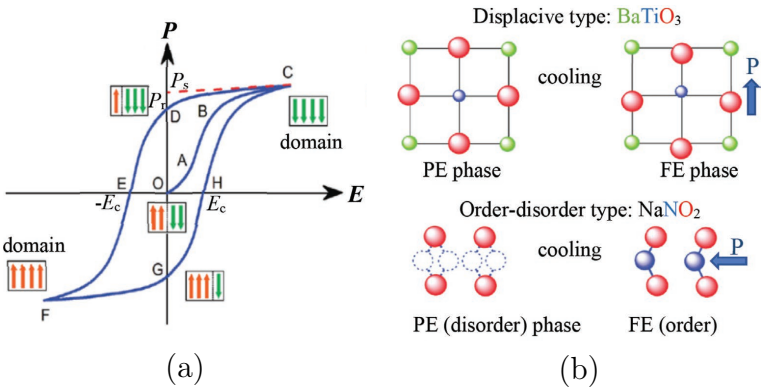



Figure 2.7: **(a)** The blue curve depicts the P-E ferroelectric hysteresis loop. The insets with the arrows represent the domain evolution during polarization switching. **(b)** Schematic representation of paraelectric-ferroelectric phase transition upon cooling below $T_c$, in both displacive type and order-disorder type materials. Adapted from [33].

an increase in the polarisation of the material. This is due to the gradual alignment of the random-oriented dipoles of ferroelectric domains (curve OC). As the field is decreased some domains are depolarized, but the polarization at null field is non-zero. This is called remanent polarization $P_r$ (slightly less than $P_s$) because it remains fixed also upon full removal of the external field (curve CD). Extrapolation of the CD curve back to the abscissa gives the saturation polarization value. A negative field will cause instead the polarization to reduce until it reaches zero at the negative coercive field ($-E_c$). Further decreasing the field will eventually switch all the domains along the negative field direction, yielding a polarisation reversal. Also in this case, the polarization will remain stable when the field is removed. A macroscopic model to describe ferroelectrics has been proposed by Landau and Devonshire. To understand the microscopic nature of the ferroelectric (and paraelectric) phase, we can refer to the so-called *order-disorder model* and *displacive model*. According to the order–disorder model, the paraelectric phase is characterized by thermally disordered electric dipoles, oriented along two or more random directions so that the average polarization is zero. Reorientation of these dipoles below $T_c$ generates ferroelectricity. In displacive ferroelectrics, there are no dipoles in the crystal above $T_c$, but appear only after a relative displacement of ions (Figure 2.7) [33, 35].

## 2.5. Ferroelectric memories

Ferroelectric materials can be regarded being those dielectrics with spontaneous two-way remanent polarization even in the absence of an external electric field. The resistive switching behaviour of the ferroelectric memristor is associated with the reversal of the

dielectric polarization in the ferroelectric. As a consequence, those materials (typically perovskites) attracted great interest in non-volatile memory applications, since they are able to encode the binary states in the two opposite polarizations. In addition to that, they can gradually change their polarization state using voltage pulses, allowing for multilevel resistance states that are ideal for neuromorphic applications. However, those materials were not compatible with the CMOS process. In 2011, it was reported that $SiO_2$-doped $HfO_2$ ultra-thin film showed ferroelectricity. Since then, Zr doping has shown the advantage of requiring a lower heat treatment temperature than other dopants, in addition to having a similar atomic radius and lattice parameter to Hf. So HZO ($Hf_{1-x}Zr_xO_4$) has become increasingly popular [36], mostly due to its CMOS compatibility, scalability, and easy integration already demonstrated with $HfO_2$ high-k technology for logic transistors [36, 37].

While the writing operation is very similar for all ferroelectric devices, readout depends strongly on the device concept. Ferroelectric memories can be classified into three main types: ferroelectric tunnel junction (FTJ), Ferroelectric Field Effect Transistor (FeFET) and Ferroelectric Random-Access Memory (FeRAM). FeRAM has a 1T–1C structure (T stands for transistor, C for capacitor) that utilizes a ferroelectric as a capacitor, thus it has a larger cell size than other memories. The binary information is stored in the ferroelectric capacitor, with non-volatile memory (NVM) properties, and the transistor allows random access for read/write operations. It has destructive characteristics in the read process, but it can read and write faster than other NVM devices. FeFET, a very promising NVM for AI computing, has a similar structure of a MOSFET where the oxide is substituted by a ferroelectric layer, and the gate voltage can be used to change the threshold voltage by polarization reversal, so two different conductance levels can be written in the semiconductor. It can achieve higher integration than FeRAM since it can be implemented without a selector, and a non-destructive readout is possible. The limited endurance at the moment might originate from extrinsic factors, which can be improved by process optimization. FTJ is a ferroelectric memory based on a change in resistance due to a change in the tunneling barrier when switching the polarization state in a metal-ferroelectric-metal (MFM) or metal-ferroelectric-semiconductor (MFS) structure. Due to its non-destructive characteristics and simple structure, a high-density memory can be implemented, but there is a problem with endurance due to charge trapping in the oxide. FTJ applications are more focused towards neuromorphic computing, being used as a multi-level brain-like device. Instead, FeRAM is expected to show RAM functionality with endurance of at least $10^{12}$ cycles or higher, and FeFETs would rather be a replacement for traditional non-volatile memories like Flash or EEPROM with $10^4$ – $10^6$ cycles, which

seems to be possible with the current status of the technology [36, 37, 39].
FTJs will be discussed in more detail in section 2.5.1.

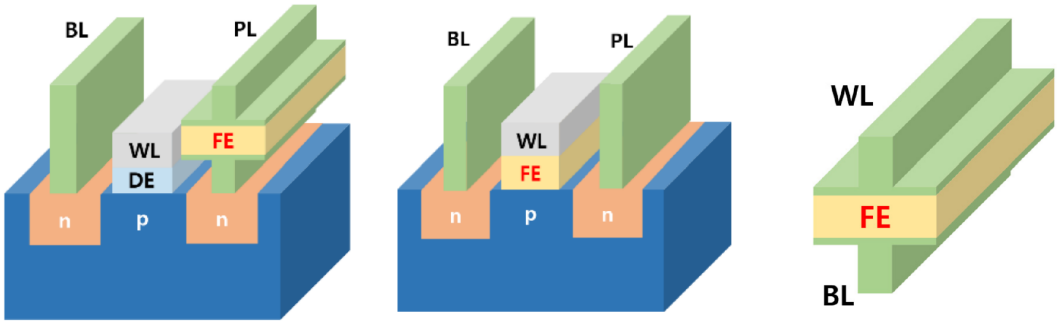



Figure 2.8: ferroelectric NVM devices: FeRAM **(a)** MOSFET plus ferroelectric capacitor 1T-1C structure; FeFET **(b)**, MOSFET structure with ferroelectric gate oxide; FTJ **(c)** parallel plate capacitor with ferroelectric layer. (BL=Bitline; WL=Wordline; PL=Plateline; DE=Dieletric; FE=Ferroelectric). Adapted from [37].

### 2.5.1. Ferroelectric Tunnel Junctions (FTJ)

The original concept of a FTJ consists of a MFM structure in which the ultra-thin ferroelectric film is sandwiched between two electrodes (Figure 2.10). Suggested by Esaki et al. in the year 1970 [36], it only attracted a lot of interest when recent experiments proved the existence of ferroelectricity down to nanometer scale. During the initial stages of development, ferroelectric devices were developed using perovskite materials such as $Pb(Zr,Ti)O_3$ (PZT), $BaTiO_3$ (BTO). However, ferroelectricity was only shown in thick films and their small bandgap (3 - 4 eV) was a problem for leakage current and electrical breakdown. In 2011, a discovery created new interest in FTJs: $SiO_2$-doped $HfO_2$ thin films showed excellent ferroelectricity even at thicknesses below 10 nm; moreover, the application of $HfO_2$ in high-k metal gate technology for logic transistors has already been successful. Compatibility and scalability with CMOS technology made $HfO_2$ very popular for ferroelectric memory applications [37].

The electrical resistance change in MFM junctions with ultrathin barriers is associated to the change in the orientation of the electric polarization [40]. The so-called tunneling electro-resistance ratio (TER) is measured as the ratio between the resistance of the device in low-resistive state (LRS) and high-resistive state (HRS), respectively, with typical values ranging up to 100 [41]. Many factors could explain the RS in ferroelectric heterostructures. For example, barrier height and depletion layer width modulation due to accumulation or depletion of carriers at interfaces or Schottky barrier modulation, with

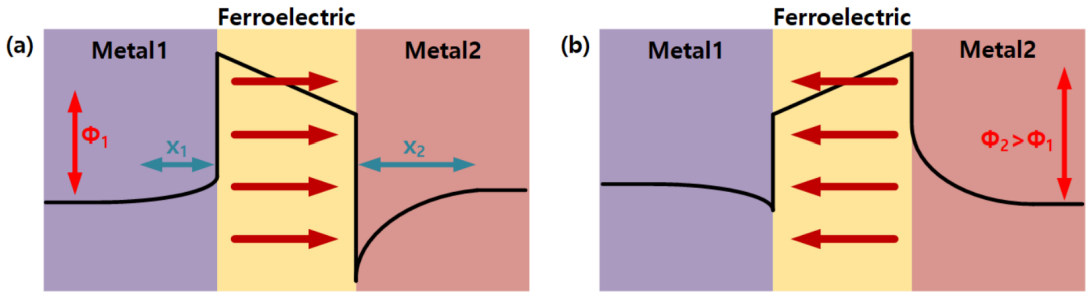



Figure 2.9: Generic MFM structure of a FTJ device. The energy band profile is asymmetric at the two metal/ferroelectric interfaces due to the different screening lengths ($x_1$, $x_2$). Red arrows show the ferroelectric polarization modulating the barrier height. Adapted from [36].

polarization switching [36, 42]. The physical mechanism typically used to account for TER in ferroelectric tunnel junctions is the change of the electrostatic potential profile induced by the reversal of the polarization in the ferroelectric. Indeed, surface charges in the thin ferroelectric film are not completely screened by the adjacent metals, giving rise to non-zero depolarizing field in the dielectric. When the polarization charges are fully compensated, the screening length of the electrodes $\lambda$ is ideally zero and so is the depolarization field. However, the screening length is always finite, with values related to the electric properties of the electrodes, causing incomplete screening. The electrostatic potential associated with the depolarization field depends on the direction of the electric polarization. The incomplete screening of polarization charges has been the limiting factor for the fabrication of ultra-thin ferroelectric films where ferroelectric polarization is destabilized by the depolarization field. Larger polarizability of the bound charges in the electrodes corresponds to a reduced repolarization field. Nowadays, the thickness limit in perovskite ferroelectrics is no longer a problem: the size effect is related to film microstructure, impurities, interface chemistry, and electrical/mechanical boundary conditions. With current technological advancement, the critical thickness for ferroelectric thin films is continuously reducing, reaching 2 nm in BFO. Unconventional ferroelectrics like $HfO_2$-based oxides have no thickness limits, with thin films realized up to 1 nm, but other problems arise with thickness reduction, such as the increase of the coercive field [38].

Different metal electrodes give rise to asymmetric energy band profiles due to the different screening lengths at the two interfaces, so carriers see a different potential barrier with the polarization reversal [40]. If one of the electrodes is a doped semiconductor the asymmetry increases. The point is that semiconductors will screen one of the carriers more efficiently and within a wider region. In that case, the increased screening length

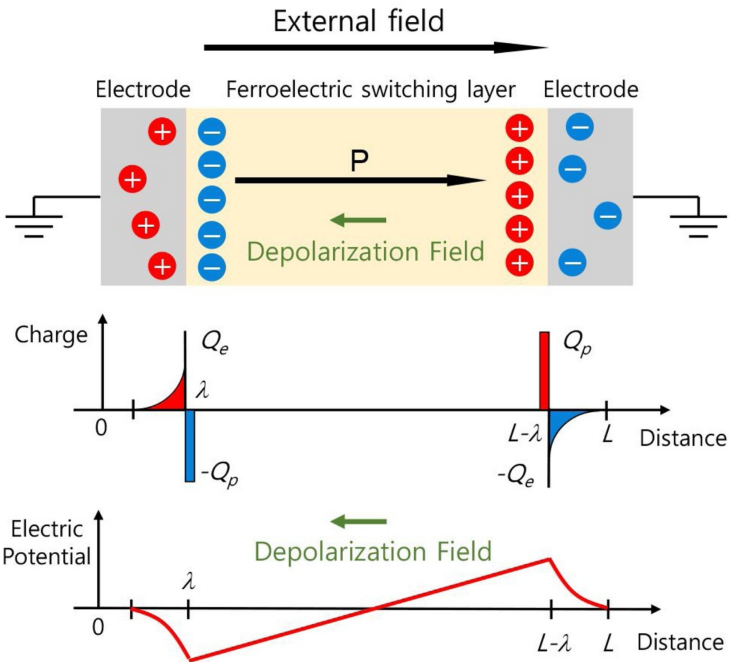



Figure 2.10: Schematic representation of bound charges developed at ferroelectric/metal interfaces in a ferroelectric tunnel junction due to polarization, and the consequent change and depolarization field distributions versus distance. Adapted from [38].

adds to the tunneling barrier width, leading to a change in tunneling probability and thus electrical resistance. This effect is sometimes referred to as Giant TER [43]. The TER ratio is defined as follows [2]:

$$TER(\%) = \frac{\mathrm{R}_{p^+} - \mathrm{R}_{p^-}}{\mathrm{R}_{p^-}} \times 100 \tag{2.4}$$

However, many works show that the defect crystal structure at the metal/ferroelectric interface plays a very essential role in the effect of the resistive switching. It is clear that the RS behavior of FTJs is mainly controlled by interface effects through mechanisms related either to the ferroelectric polarization or to defect states, which sometimes contribute in parallel [31, 42].

The tunneling current mainly arises thanks to conduction mechanisms like Direct tunneling (electrons tunnel across a rectangular barrier), "Fowler-Nordheim tunneling" (electrons tunneling across a rectangular barrier), and "thermionic emission" (when carriers receive thermal energy and exceed potential barrier) [36]. Details about these mechanisms will be given in Chapter 3. In order to ensure large tunneling currents a very thin ferroelectric layer ($<$ 3nm) is required [41]. As mentioned above, ferroelectric $HfO_2$ is more suitable

for FTJs because it can easily form a thinner ferroelectric layer, compared to perovskites. However, when a very thin polycrystalline layer is used, the memory window is reduced by the presence of parasitic currents [36]. In polycrystalline ferroelectrics like HZO, the existence of boundaries between the ferroelectric grains and the presence of defects can provide additional leakage current paths [41] or cause charge trapping effects, which lowers the On/Off ratio. The insertion of a dielectric layer between one of the electrodes and the ferroelectric (MFIM structure) has been demonstrated to be a solution, in fact, most work on FTJ using HZO is on double-layer FTJs [39]. MFM or MFIM double layer structures have demonstrated crossbar operation, but the inherent linearity of FTJs can be a problem for high current density. Solutions to implement non-linear FTJs have been developed, reducing the leakage currents and improving integration in crossbar arrays. Furthermore, the non-linear characteristic is advantageous for synaptic devices in neuromorphic applications [36]. For applications involving FTJs as synaptic weights in the analog VMM, gradual switching is needed. FTJs show synaptic plasticity with multiple intermediate resistance levels obtained through gradual switch of more ferroelectric domains by applying incremental voltage [44]. Conductance can indeed be modulated continuously by varying the number of voltage pulses and the pulse amplitude. This requires however proper selection of the programming pulse amplitudes and widths that are constrained by the CMOS technology [41, 45].

For readout operations, an electrical field smaller than the ferroelectric coercive field is applied, thus leading to a non-destructive readout. Furthermore, FTJs feature a low read current compared to other memristive devices, thus being the most power-efficient resistive switching devices, and due to their non-destructive characteristics and simple structure, a high-density memory can be implemented [41]. High endurance $> 10^{11}$ have been reported for double layer structures, retention up to 10 years, and reading times in the order of few nanoseconds [36, 37, 41].

# 3 | Methods

The following chapter will present the methods employed to conduct research for this thesis. In particular, we carried out a comprehensive analysis of the conduction mechanisms of $BiFeO_3$ (BFO) Ferroelectric Tunnel Junction through DC electrical measurements, to better understand the physics behind the resistive switching mechanism. Additionally, the text will outline the detailed microelectronic fabrication techniques used to build passive crossbar arrays.

## 3.1. Temperature dependent DC electrical measurements

The conduction currents in dielectric materials are negligible in normal conditions of applied field, due to their insulating nature. However, the application of large electric fields on those materials yields measurable leakage currents. The knowledge of those leakage paths is crucial in the improvement of the devices. Furthermore, the knowledge of the conduction mechanisms allows to gain insights about the resistive switching behaviour in FTJs. typical measurements are conducted on metal-insulator-metal (MIM) or metal-insulator-semiconductor (MIS) capacitors [46].
A distinguishing feature of those mechanisms is the temperature dependence, so measuring the temperature-dependent conduction currents is a simple way to determine which are the mechanisms governing the conduction in the sample. The idea is that each conduction mechanism is characterized by a specific current-voltage analytical expression, and a linear relationship can be highlighted in each analytical formulation, typically representing the data in the log(J) - log (V) domain. Then, a linear regression is performed on the represented data to verify if the expected linearity is satisfied by the analyzed mechanisms or not.

### 3.1.1. Conduction mechanisms in Dielectric films

The origin of conduction paths in dielectric films is of different nature: some depend only on the electrical properties of the dielectric film and so are called "*Bulk-limited*" conduction mechanisms, others derive from the electrical properties of the interface between the dielectric film and the electrode, thus the name "*electrode-limited*" conduction mechanisms [46]. Let's see more in detail the mechanisms that have been tested in this work, which are summarized in Figure 3.1, in order to understand their physical derivation and their linear representation.

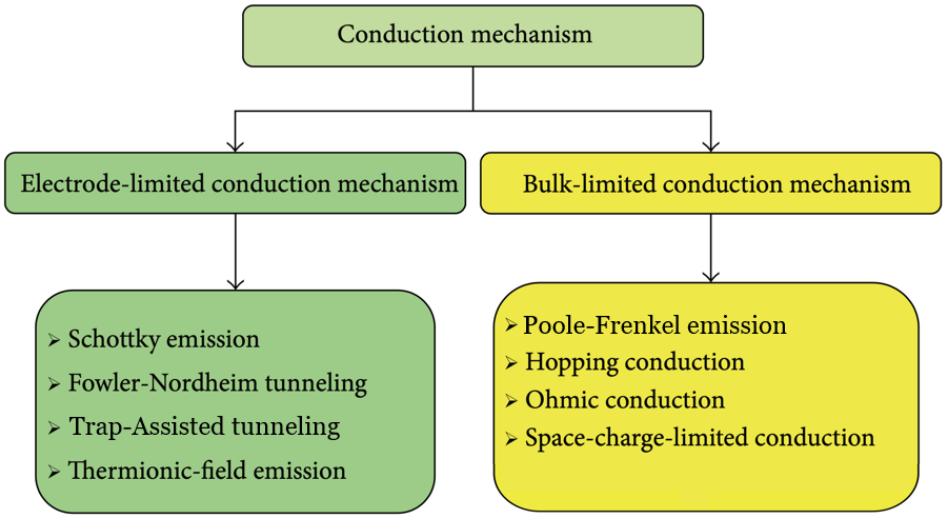



Figure 3.1: Summary of the conduction mechanisms in dielectrics that have been analyzed in this work. Adapted from [46]

### 3.1.2. Electrode-limited conduction mechanisms

The electrode-limited conduction mechanisms include Schottky emission, thermionic field emission, Fowler-Nordheim tunneling, and Trap-assisted tunneling.

- **Schottky emission**:
  Schottky emission occurs when electrons in the metal overcome the energy barrier at the interface with the oxide thanks to the thermal activation energy, as shown in Figure 3.2. It takes its name from the Schottky effect, i.e. the barrier-lowering effect due to image force. The expression of Standard Schottky emission is [46]:

$$J = AT^2 \exp\left[-\frac{q(\phi_B - \sqrt{qE/4\pi\epsilon_r\epsilon_0})}{kT}\right] \tag{3.1}$$

Where $A$ is the Richardson constant, dependent on the effective electron mass, $T$ is the absolute temperature, $q$ is the electronic charge, and $E$ is the electric field. $\epsilon_0$ and $\epsilon_r$ are the vacuum and optical permittivities, while $q\phi_B$ is the Schottky barrier height, measured as the offset between the metal and the insulator energy band edge. To be selected as a mechanism in the analyzed structure, the Schottky plot of $\log(J/T^2)$ versus $E^{1/2}$ should be linear. From the Arrhenius of the intercept of the fit is possible to extract the value of the Schottky barrier and the optical dielectric constant from the Arrhenius of the slope. In particular, not only linearity must be satisfied, but also the dielectric constant should be close to the square of the optical refractive index.

$$\log\left(\frac{J}{T^2}\right) = \log(A^*) - \frac{q\phi_B}{k_B T} + \frac{\sqrt{\frac{q^3}{4\pi\varepsilon_0}}}{k_B T}\sqrt{E} \tag{3.2}$$

$$\text{intercept} = \log(A^*) - \frac{q\phi_B}{k_B} \cdot \frac{1}{T} \tag{3.3}$$

$$\text{slope} = \frac{\sqrt{\frac{q^3}{4\pi\varepsilon_0}}}{k_B} \cdot \frac{1}{T} \tag{3.4}$$

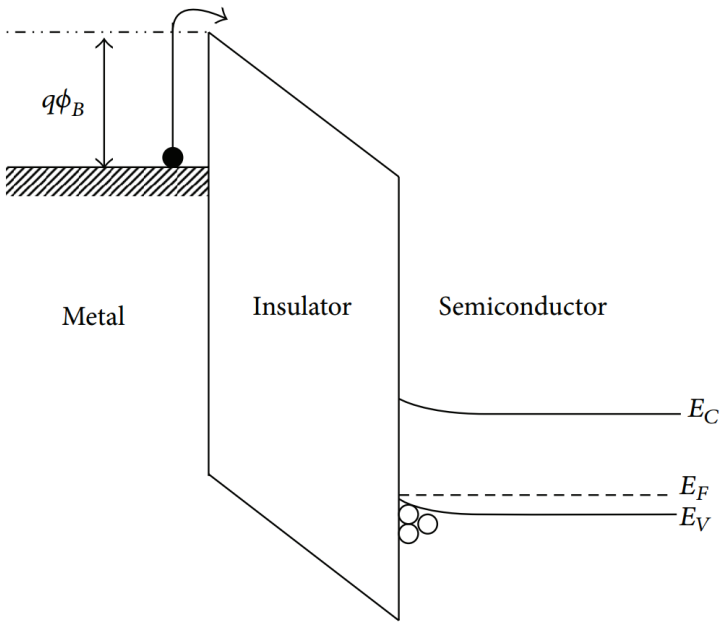



Figure 3.2: Schematic of Schottky emission in MIS structure. Adapted from [46]

When traps and interface defect states affect the conduction, the thermionic emission process becomes a trap-limited mechanism: this is the case of **modified Schottky emission**, which occurs by definition when the electronic mean free path is less

than the dielectric thickness. The modified expression is [46]:

$$J = \alpha T^{3/2} E \mu \left(\frac{m^*}{m_0}\right)^{3/2} \exp\left[-\frac{q(\phi_B - \sqrt{qE/4\pi\epsilon_r\epsilon_0})}{kT}\right] \tag{3.5}$$

where $\alpha$ is a constant and $\mu$ the carrier mobility. In this case the plot of $\log(J/(T^{3/2}E))$ versus $E^{1/2}$ should show linearity.

$$\log\left(\frac{J}{T^{3/2}\cdot E}\right) = \log\left(\alpha\mu\left(\frac{m^*}{m_0}\right)^{3/2}\right) - \frac{q\phi_B}{k_B T} + \frac{\sqrt{\frac{q^3}{4\pi\varepsilon_0}}}{k_B T}\cdot\sqrt{E} \tag{3.6}$$

$$\text{intercept} = \log\left(\alpha\mu\left(\frac{m^*}{m_0}\right)^{3/2}\right) - \frac{q\phi_B}{k_B}\cdot\frac{1}{T} \tag{3.7}$$

$$\text{slope} = \frac{\sqrt{\frac{q^3}{4\pi\varepsilon_0}}}{k_B}\cdot\frac{1}{T} \tag{3.8}$$

From the Arrhenius plot of the intercept it is possible to get $\mu\left(\frac{m^*}{m_0}\right)$, while from the Arrhenius of the slope the dynamic dielectric constant $\epsilon_r$.

- **Fowler-Nordheim Tunneling (FNT)**:
  Quanto-mechanical tunneling across a potential barrier allows electrons from the metal to penetrate into the dielectric conduction band. FNT accounts for carriers tunneling through a triangular barrier of height $q\Phi_B$ (Figure 3.3), which generates a measurable tunneling current density only at high fields [46]:

$$J = \frac{q^2 E}{8\pi\hbar q\Phi_B}\exp\left[-\frac{8\pi(2qm_T^*)^{1/2}}{3\hbar E}\phi_B^{3/2}\right] \tag{3.9}$$

  $m_T^*$ is the tunneling effective mass of carriers in the dielectric, and it is generally assumed to be equal to the carrier effective mass. If FNT is an active conduction mechanism in the structure, the plot of $\log(J/E^2)$ versus 1/E should be linear, and from the slope, the electron effective mass and the barrier height can be extracted:

$$\log\left(\frac{J}{E^2}\right) = \log\left(\frac{q^3}{8\pi\hbar q\phi_B}\right) + \frac{8\pi\left(2qm_T^*\right)^{1/2}}{3\hbar}\cdot\phi_B^{3/2}\left(-\frac{1}{E}\right) \tag{3.10}$$

$$\text{intercept} = \log\left(\frac{q^3}{8\pi\hbar q\phi_B}\right) \tag{3.11}$$

$$\text{slope} = \frac{8\pi\left(2qm_T^*\right)^{1/2}}{3\hbar}\cdot\phi_B^{3/2} \tag{3.12}$$

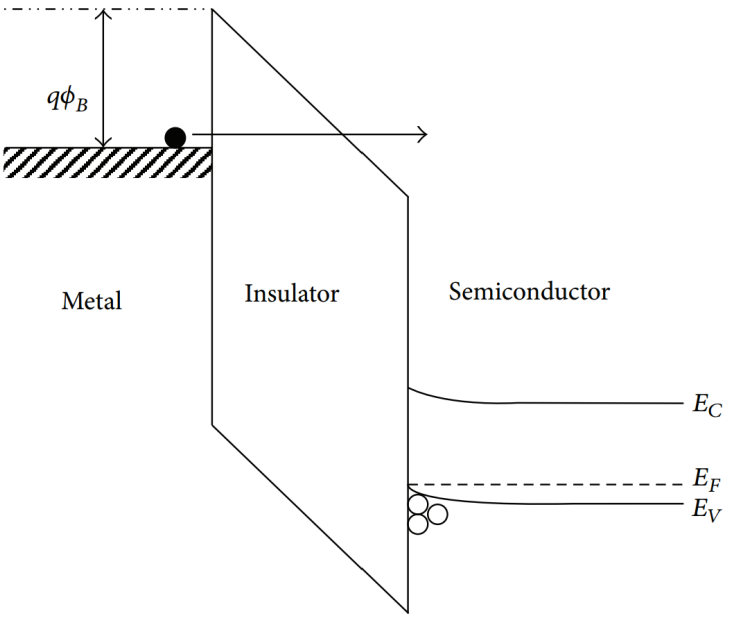



Figure 3.3: Schematic of Fowler-Nordheim tunneling process in MIS structure. Adapted from [46]

- **Trap-Assisted tunneling (TAT)**:
  The physical model of TAT considers a triangular barrier in the insulator and the presence of some trapping centers near the conduction band edge (n-type case). When moderate electric fields, lower than required for FNT, are applied across the insulator, the electrons from the electrodes can be injected by tunneling into the trap centers. This type of transport can be divided into three processes: (i) tunneling from the electrode into the trap, (ii) tunneling from trap to trap, and (iii) tunneling from the trap into the electrode (Figure 3.4) The tunneling levels could be different, so tunneling for example from the trap to the next electrode may require a thermal excitation. If thermal excitation is rapid the conductance is limited by the tunneling rate to the traps [47, 48]. The TAT current density has a simplified expression derived by [49]:

$$J = A \exp\left[ -\frac{4\sqrt{2qm_{ox}}}{3\hbar}\phi_t^{3/2}\frac{1}{E}\right] \tag{3.13}$$

with the trap level represented by $\phi_t$. In this case, A represents a pre-factor. In this theory, the transport of electrons from one electrode to the other electrode is supposed to go over one trap only. The tunneling from trap to trap, however, is neglected. Also, the thickness of the oxide layer and the trap distance are not considered, which makes the model oversimplified [47]. However, this is a generalized model which considers the more general case of both triangular and trapezoidal barrier tunneling, to accurately model J –E curves for a wider range of electric fields than simple TAT [49]. TAT fit should be linear in the representation $\log J$ versus $-1/|V|$, and the slope of the fit yields the trap energy level.

$$\log(J) = -\frac{1}{|V|}\cdot\frac{4\sqrt{2qm_{ox}}}{3\hbar}\phi_t^{3/2} + \log(A) \tag{3.14}$$

$$\text{intercept} = \log(A) \tag{3.15}$$

$$\text{slope} = \frac{4\sqrt{2qm_{ox}}}{3\hbar}\phi_t^{3/2} \tag{3.16}$$

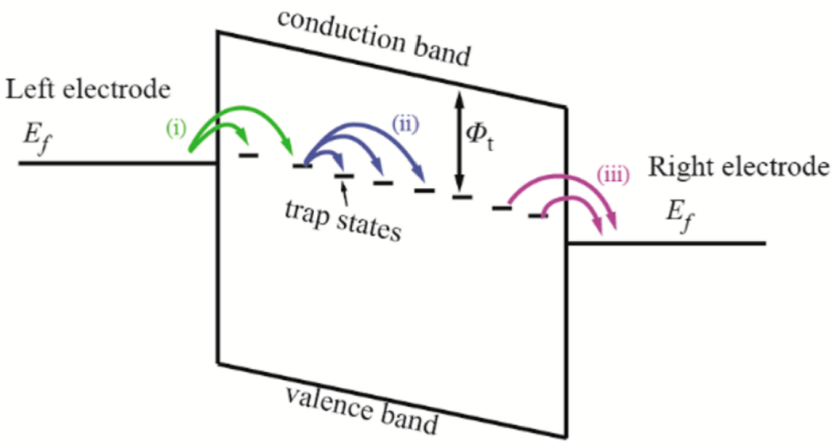



Figure 3.4: Schematic of Trap-Assisted tunneling process in MIM structure. The induced field induces a tilted rectangular band structure. The trap level $\phi_t$ lies in the oxide band gap below the conduction band. The phenomenon occurs in three steps:(i) tunneling into the trap from one electrode, (ii) tunneling from trap to trap, and (iii) tunneling from the trap into the other electrode. Adapted from [48]

- **Thermionic-field emission**:
  The conditions for this conduction mechanism are intermediate between thermionic (Schottky) emission and field emission (tunneling). This situation is schematized in Figure 3.8. The current density is expressed as [46]:

$$J = q^2 \frac{\sqrt{m}(kT)^{1/2}E}{8\hbar^2\pi^{5/2}} exp\left(-q\frac{q\phi_B}{kT}\right) exp\left[\frac{\hbar^2 q^2 E^2}{24m(kT)^3}\right] \tag{3.17}$$

  In this case, one should verify if the plot of $\log(J/(T^{1/2}E))$ versus $E^2/T^3$ is linear.

$$\log\left(\frac{J}{E\sqrt{T}}\right) = \log\left(\frac{q^2\sqrt{mk_B}}{8\hbar\pi^{5/2}}\right) - \frac{q\phi_B}{k_B T} + \frac{\hbar^2 q^2}{24m(k_B T)^3} \cdot \frac{E^2}{T^3} \tag{3.18}$$

$$\text{intercept} = \log\left(\frac{q^2\sqrt{m}k_B}{8\hbar\pi^{\frac{5}{2}}}\right) - \frac{q\phi_B}{k_B} \cdot \frac{1}{T} \tag{3.19}$$

$$\text{slope} = \frac{\hbar^2 q^2}{24m(k_B T)^3} \tag{3.20}$$

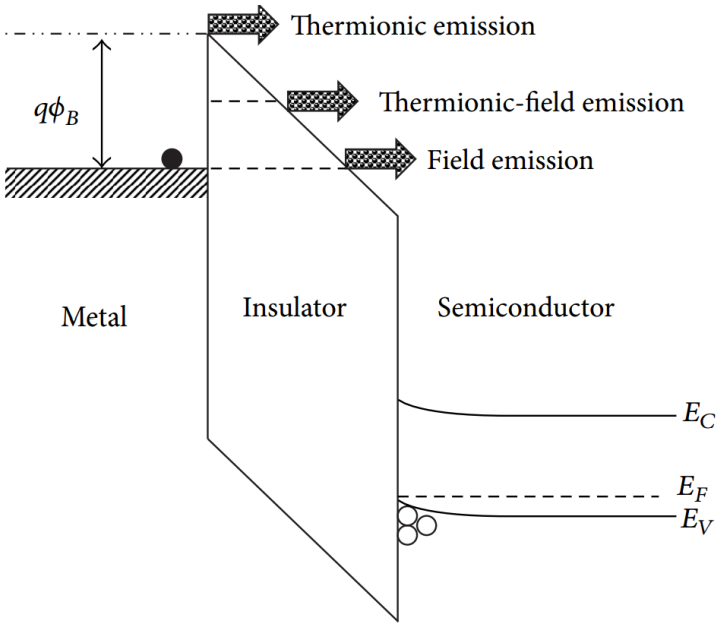



Figure 3.5: Schematic of the energy band diagram of thermionic-field emission in MIS structure, compared to thermionic emission and field emission. Adapted from [46]

### 3.1.3. Bulk-limited conduction mechanisms

Bulk-limited mechanisms include Poole-Frenkel (PF) emission, variable-range hopping (VRH), Ohmic conduction, and Space-charge limited conduction (SCLC). Ionic and grain-boundary limited conduction have not been taken into account in this work. From the analysis of those mechanisms, we can gain insights into the electrical properties of the dielectric film through the extraction of parameters such as the trap energy level and density, the trap spacing, the electron mobility, and the effective density of states in the conduction/valence band.

- **Poole-Frenkel (PF) emission**:
  This mechanism involves thermionic emission from bulk traps. The bulk of the dielectric is characterized by a certain number of traps that confine electrons inside high potential barriers. For electrons to escape from the trapping centers a lot of energy is needed, but it can be reduced by an applied field across the dielectric film. Thus, thermionic emission of electrons from the traps can become easier (Figure 3.6) and a measurable current is generated. The PF emission current density is [46]:

$$J = q\mu N_c E \exp\left[\frac{-q(\phi_T - \sqrt{qE/\pi\epsilon_i\epsilon_0})}{kT}\right] \tag{3.21}$$

  Here, the parameter $\mu$ is the carrier drift mobility, $N_c$ is the density of states in the conduction band (for an n-type material) and $q\phi_T$ represents the trap energy level. In this case, the plot of $\ln(J/E)$ versus $E^{1/2}$ must be linear. From the Arrhenius of the intercept of the fit, the trap level can be extracted, while the slope Arrhenius yields the optical dielectric constant. Also in this case, it is important to verify that the square of $\epsilon_r$ matches with the optically measured refractive index. This mechanism is typically expected to occur at high fields and temperatures.

$$\log\left(\frac{J}{E}\right) = \log(q\mu N_c) - \frac{q\phi_T}{k_B T} + \frac{\sqrt{\frac{q^3}{\pi\varepsilon_0\varepsilon_r}}}{k_B T}\sqrt{E} \tag{3.22}$$

$$\text{intercept} = \log(q\mu N_c) - \frac{q\phi_T}{k_B} \cdot \frac{1}{T} \tag{3.23}$$

$$\text{slope} = \frac{\sqrt{\frac{q^3}{\pi\varepsilon_0\varepsilon_r}}}{k_B} \cdot \frac{1}{T} \tag{3.24}$$

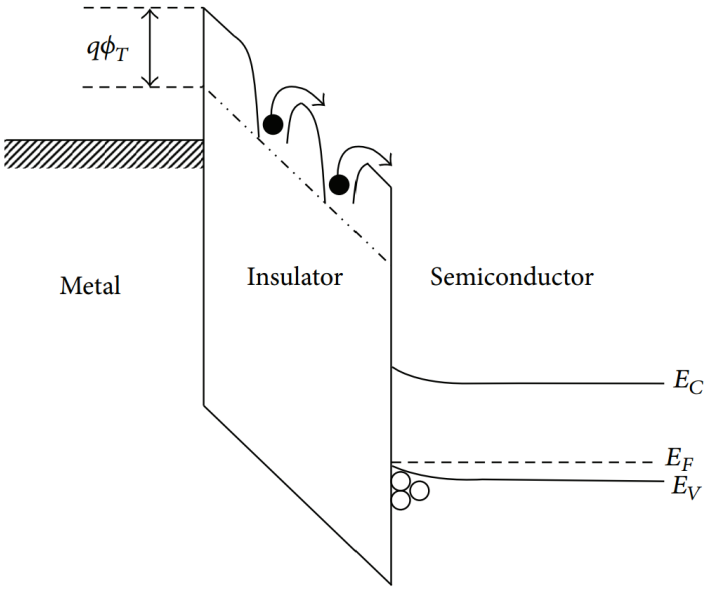



Figure 3.6: Schematic of the energy band diagram of Poole-Frenkel emission process in a MIS structure. Adapted from [46]

- **Variable Range hopping (VRH)**:
  Conduction by hopping is due to the tunneling of trapped electrons, "hopping" from one trap site to another. This situation is better exemplified in Figure 3.7. The consequent analytical current density is [46]:

$$J = qan\nu exp\left[\frac{-qaE}{kT} - \frac{E_a}{kT}\right] \tag{3.25}$$

  Information about the traps is included in the parameter $a$, the mean hopping distance, and $E_a$, the trap activation energy. Then, $n$ is the electron concentration in the conduction band of the dielectric and $v$ is the thermal frequency of vibration of the trapped electrons. For VRH the plot $\log(J)$ versus $E$ should yield a linear fit. From the slope, we can obtain the hopping distance, while the activation energy from the intercept.

$$\log(J) = \log(qan\nu) - \frac{E_a}{k_B T} + \frac{qa}{k_B T} \cdot E \tag{3.26}$$

$$\text{intercept} = \log(qan\nu) - \frac{E_a}{k_B} \cdot \frac{1}{T} \tag{3.27}$$

$$\text{slope} = \frac{qa}{k_B} \cdot \frac{1}{T} \tag{3.28}$$

  This effect is expected at relatively low electric fields, with an exponential decrease

of the current density at high temperatures

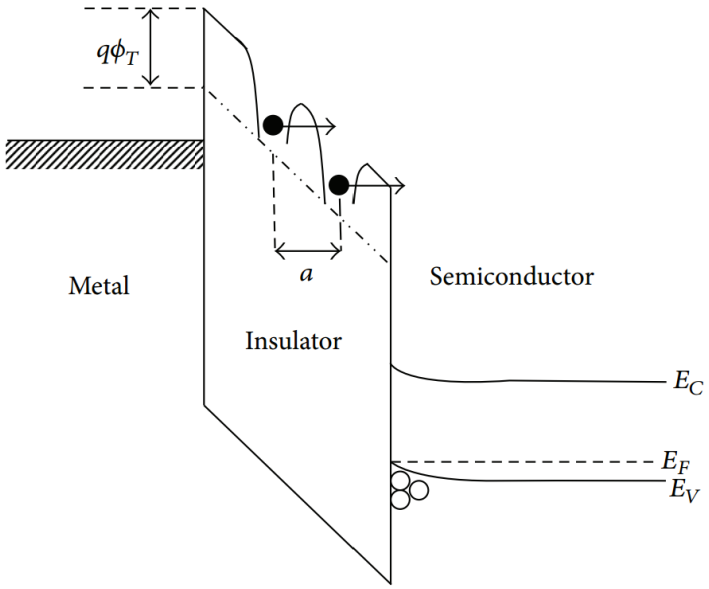



Figure 3.7: Schematic of the energy band diagram of variable-range hopping process in MIS structure. Adapted from [46]

- **Ohmic conduction**:
  At very small electric fields it is possible to measure a current of carriers moving in the conduction and valence bands of the insulator, generated by thermal excitation. The current density, which is very small, is expressed as [46]:

$$J = \sigma E = nq\mu E = q\mu N_c E \exp\left[-\frac{(E_c - E_F)}{kT}\right] \tag{3.29}$$

  Where $\sigma$ is the electrical conductivity, $\mu$ is the electron mobility, and $N_C$ is the effective density of states in the conduction band. $E_C$ is the conduction band edge and $E_F$ the oxide Fermi level. Note that this is valid for a n-type material. For Ohmic conduction to be observed, the $\log(J)$ - $\log(V)$ plot must be linear.

$$\log(J) = \log(q\mu N_c) - \frac{(E_C - E_F)}{k_B T} + \log(E) \tag{3.30}$$

$$\text{intercept} = \log(q\mu N_c) - \frac{(E_C - E_F)}{k_B} \cdot \frac{1}{T} \tag{3.31}$$

$$\text{slope} = 1 \tag{3.32}$$

From the intercept, we can extract the conductivity, which has an exponential re-

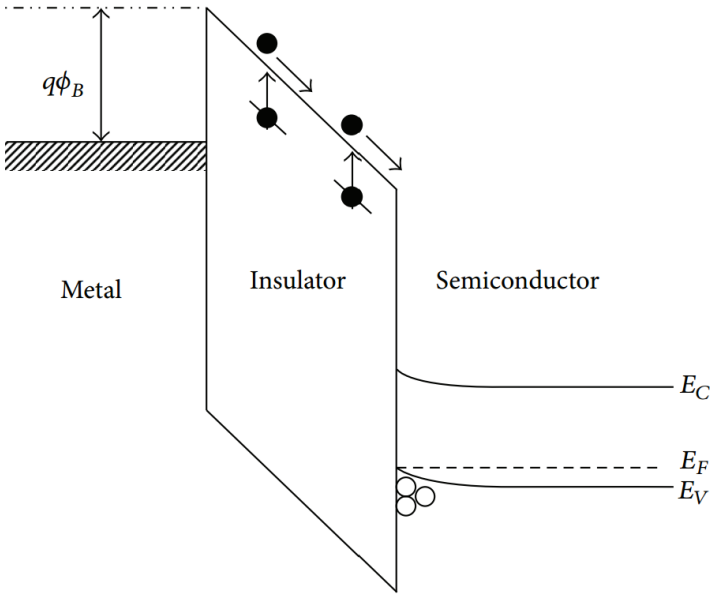



Figure 3.8: Schematic of the energy band diagram of Ohmic conduction in a MIS structure. Adapted from [46]

lation with the temperature i.e. $\sigma \propto exp\left[\frac{(E_c-E_F)}{kT}\right]$ and the parameters such as $E_c - E_F$ and the product $\mu N_c$ can be obtained using the Arrhenius plot of the intercept.

- **Space-charge-limited conduction (SCLC)**:
  The SCLC characteristics in the log $J$ - log $V$ plot are represented in Figure 3.9. The log-log plane is bounded by three curves: Ohm's law a smaller voltages ($J \propto V$), traps-filled limit (TFL) current ($J \propto V^2$), and Child's law at higher voltages ($J \propto V^2$). The voltages $V_{tr}$ and $V_{TFL}$ set the transition from a region to the neighboring one. Let's consider a trap-free insulator sandwiched between two electrodes (MIM or MIS structure). When the system reaches a dynamic equilibrium, a space-charge region emerges in the dielectric giving rise to a space-charge limited current of a certain magnitude. In order to observe a significant amount of current at least one of the two electrodes must make ohmic contact to the insulator, and the insulator must be relatively free from trapping defects. The expression for SCL currents is ruled by the so-called Child's law for solids [46]:

$$J_{Child} = \frac{9}{8}\mu\epsilon\frac{V^2}{d^3} \tag{3.33}$$

  Where $\mu$, $\epsilon$ and $d$ are the carrier mobility, dielectric constant and thickness of the

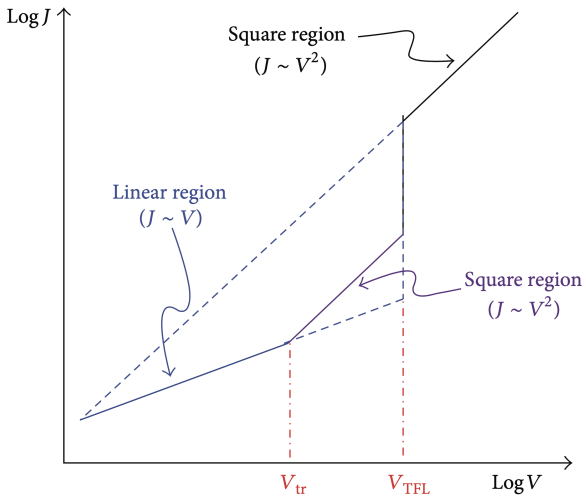



Figure 3.9: Typical current density-voltage (J-V) characteristic in the log $J$-log $V$ plane for space-charge-limited conduction current for an insulator with a discrete shallow trap level. The three bounding curves are the Ohm's law ($J \propto V$), traps-filled limit (TFL) current ($J_{TFL} \propto V^2$), and Child's law ($J_{Child} \propto V^2$). $V_{tr}$ and $V_{TFL}$ are the transition voltage at the departure from ohm's law and TFL curve. Adapted from [46]

dielectric. Let now the insulator have shallow traps, that is, traps lying close enough to the conduction band to be in thermal equilibrium with the electrons. The same expression of Child's law will be obtained, with the difference that the drift mobility for free carriers is multiplied by the factor $\theta$, indicating the fraction of the total charge that is free. The value of this fraction is determined by the number and depth ($E_t$) of traps and is independent of the applied voltage. Accordingly, the space-charge-limited current has the same square-law dependence on voltage as in the simple trap-free model, but it decreases in magnitude as observed in the TFL region [52]. The TFL current is modeled as [46]:

$$J_{\mathrm{TFL}} = \frac{9}{8}\epsilon\boldsymbol{\mu\theta}\frac{V^2}{d^3} \tag{3.34}$$

where the free-to-trapped carrier density ratio is:

$$\theta = \frac{N_t}{N_c}e^{-\frac{E_t}{kT}} \tag{3.35}$$

Therefore, the SCLC mechanisms can be explained as follows: at low applied voltages ($\mathrm{V} < \mathrm{V_{tr}}$) the J-V characteristic follows the ohm's law, which implies that the density of thermally generated free carriers ($\mathrm{n_0}$) inside the films is larger than the

injected carriers. The ohmic mode takes place when trap centers are partially filled at weak injection. The transition at $V_{tr}$ implies that the carrier's transit time across the insulator $\tau_c$ is equal to the dielectric relaxation time $\tau_d$. When instead $V < V_{tr}$, $\tau_c > \tau_d$, so the injected carriers will not be able to travel across the insulator but will rather redistribute themselves in order to maintain charge neutrality. This phenomenon is known as dielectric relaxation. This situation is schematized in Figure 3.10.

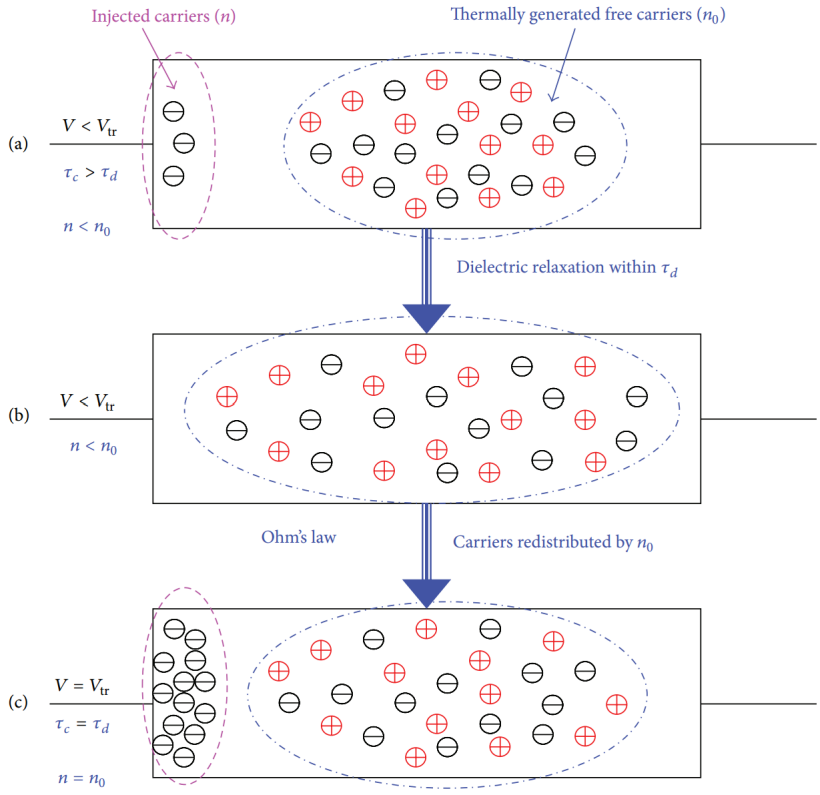



Figure 3.10: Schematic of the physical mechanisms occurring in SCLC during weak injection. For voltages lower than $V_{tr}$ conduction is Ohmic **(b)**, until $V = V_{tr}$ where SCL conduction starts. Adapted from [46]

As voltage increases we have strong injection, so the injected carriers will dominate over the thermally generated ones. Since their transit time decreases with increasing V, they will be able to travel across the insulator, and a space-charge region builds up. While the voltage reaches $V_{TFL}$, the increase of the density of free carriers will reach such a value that the Fermi level moves up above the electron trapping level, and the traps get gradually saturated. This results in a strong increase in the number of free electrons, thus explaining the increase of the current at $V_{TFL}$. This voltage is defined as the voltage required to fill the traps or, in other words, as the

voltage at which the Fermi level passes through the trap level. At the voltages V $>$ $V_{TFL}$, all traps are filled up and the subsequently injected carriers will be free to move in the dielectric films, so the current will rapidly jump from its low trap-limited value to a high trap-free SCL current. This strong injection regime is represented in Figure 3.11.

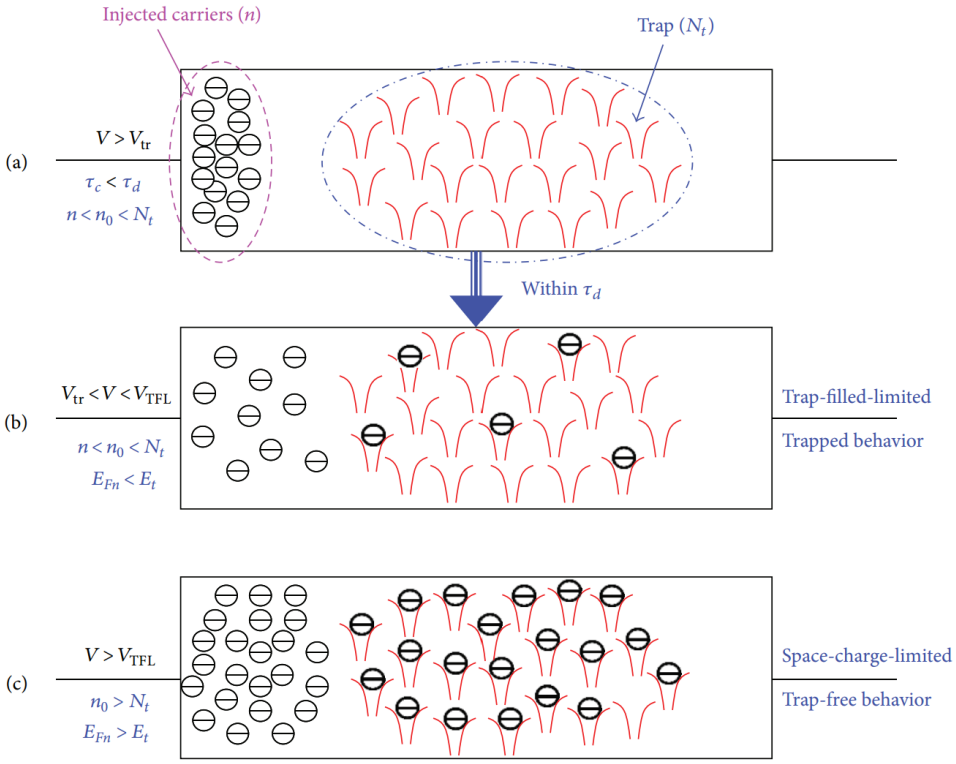



Figure 3.11: Schematic of the physical mechanisms occurring in SCLC during weak injection. For $V_{tr} < V < V_{TFL}$ conduction is trap-filled limited **(b)**, and for $V > V_{TFL}$ it becomes trap-free.Adapted from [46]

- **Space-charge-limited conduction with exponentially-distributed traps (EDT-SCLC)**:

  A very common situation is that traps are not associated with a single discrete energy level, but rather with a distribution of energies. Let's consider an exponential distribution, and let the steepness of the trap distribution be approximated by a characteristic temperature $T_C = E_T/k$ (k is the Boltzmann constant) . Equivalently, one can characterize the exponential tail of defect density of states in terms of energy with the characteristic energy $E_T = T_C k$. The trap density $N_t$ is thus a function of exponentially distributed traps with parameter $T_C$ (or $E_T$) [50] :

$$N_t(E) = \frac{N_t}{kT_C} \exp -\frac{E}{kT_V} \tag{3.36}$$

Small values of $T_C$, lead to trap distributions varying rapidly with energy, while large values of $T_C$ approximate a slowly varying trap distribution [51].
The voltage dependence of the SCL current density is given by [52] :

$$J \propto V^{\frac{T}{T_C}+1} \tag{3.37}$$

Thus, V has an exponent $> 2$ if $T_C > T$. For $T_C < T$ it reduces to the case of discrete shallow traps [52]. For SCLC with shallow traps the plot $\log(J)$ versus $\log(V)$ should be linear, and a slope of 2 is expected. From the Arrhenius plot of the intercept of the fit one can extract information about the characteristic energy $E_T$ and the trap density $N_t$. In the case of SCLC with exponentially distributed traps one must seek linearity in the $\log(J)$ versus $\log(V)$ with a slope higher than 2, where the slope is the parameter $m = \frac{T}{T_C} + 1$. The Arrhenius plot of the slope yields the value of $T_C$, while information about the trap density can be obtained with advanced techniques which are discussed in Chapter 5.

## 3.2. Processing

In this section, all the processing techniques employed for the fabrication of FTJ crossbar arrays are described. Starting from conventional photolithography, the typical process flow for microelectronic fabrication is reported, from exposure to pattern transfer, even with advanced exposure tools. Patterning techniques that are used in this work include physical and chemical dry etching processes, as well as Chemical and Physical Vapor Deposition methods for thin films growth. In addition, thermally activated processes are employed to induce proper crystallization of the materials.

### 3.2.1. Photolithography

Photolithography is a microfabrication technique employed to pattern parts of a thin film or bulk of a substrate. The pattern is transferred from a GDSII file onto a light-sensitive material, the photoresist, which has been previously spun onto the substrate. Light can be exposed either directly (without mask), or with a projected image by using a customized photomask. The exposure to light changes the solubility of the photoresist such that some parts of it can be easily dissolved in a developer solution, depending on the polarity of the photoresist: in positive resists the exposed polymeric chains become more soluble in the developer, while for a negative resist, light induces cross-linking of the polymer, thus the unexposed parts will be removed during development (Figure 3.12). The resist

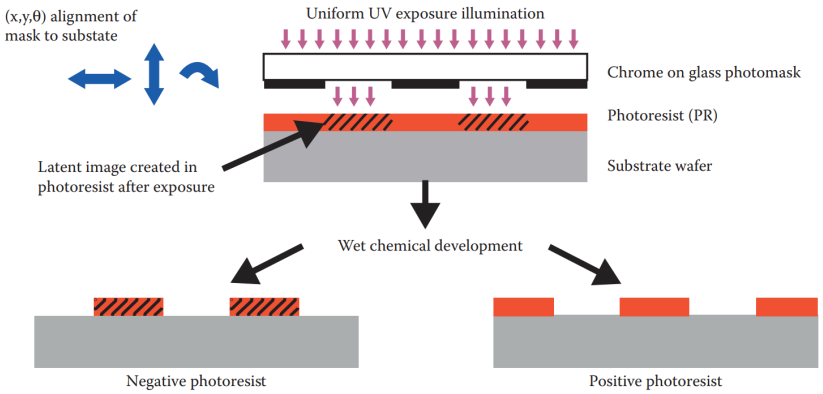



Figure 3.12: Schematic representation of the pattern transfer on a negative versus a positive photoresist after light illumination. Adapted from [53].

is then developed in an alkali solution, which removes either the exposed (for the positive photoresist) or the unexposed (for the negative photoresist) polymer. Thus, the final resist pattern is binary: parts of the substrate are covered with resist while other parts are completely uncovered. This binary pattern is needed to faithfully transfer the desired pattern since the parts of the substrate covered with resist will be protected during the subsequent patterning.

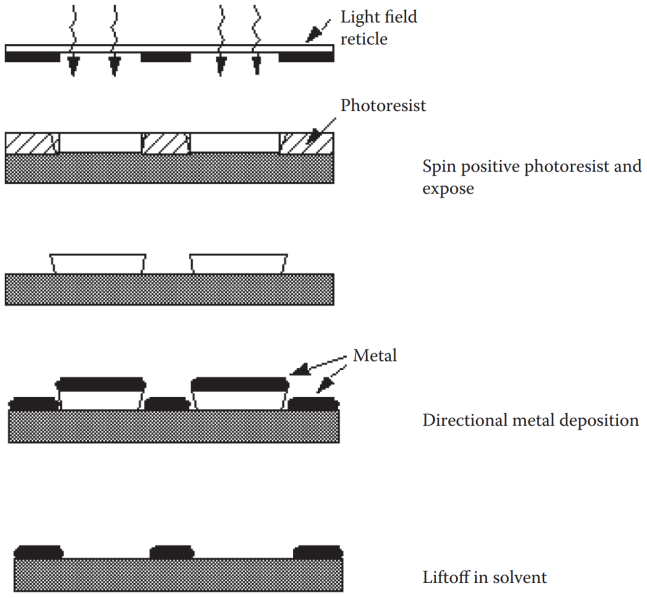



Figure 3.13: Schematic process flow during lift-off transfer. Adapted from [53].

There are two basic pattern transfer approaches: subtractive transfer (etching) is the most common approach, while additive transfer techniques (lift-off) are used whenever etching processes are not available, for example for copper interconnects. In this case,

the photoresist is patterned in such a way to open the areas where the new layer must be grown; then, the material is deposited over the whole area of the wafer, reaching the surface of the substrate in the unprotected regions and staying on the top of the photoresist. When stripping the photoresist, the material on the top is lifted off and washed together with it, while the material remains deposited on the uncovered areas. Lift-off is schematized in Figure 3.13.

Resolution, the smallest printable feature, is ruled by the Rayleigh criterion:

$$R = 0.6\frac{\lambda}{NA} \tag{3.38}$$

R is the minimum feature size, $\lambda$ is the light source wavelength and NA is the objective numerical aperture of the projection system ($\simeq 0.6$ for standard systems). However, this is valid for a particular case of a "point source". A generalized expression defines the resolving power of the photolithography as:

$$R = k_1\frac{\lambda}{NA} \tag{3.39}$$

where $k_1$ is the process factor ($\simeq 0.4$ for mass production), an experimentally determined dimensionless parameter that depends on processing factors. From the above relation, it is clear that reducing the wavelength of the light is an effective alternative to decrease the feature size, as it happens for example in the electron-beam lithography described next. Present photolithography equipment employs deep ultraviolet (DUV) light from excimer lasers having a wavelength of 248 and 193 nm. Thus the minimum feature sizes can be reached up to 50 nm [53–57].

- **Electron-beam lithography (EBL)** :
  Traditional photolithography needs the use of a photomask with the customized CAD pattern, transferred with a mask aligner. On the contrary, with e-beam lithography, it is possible to directly expose the resist by scanning a focused electron beam onto the substrate. The beam reaches a very small spot size that can be less than 10 nm, using a series of condenser lenses and beam deflection coils that deflect the beam so that it follows the desired pattern. The process flow is similar to conventional photolithography, that is exposure of the sample surface, followed by the development of the exposed area, and finally the pattern transfer. Also in EBL, it is possible to have both polarities for the resist, which are now electron-sensitive materials like poly(methyl methacrylate) denoted as PMMA.
  Although the high resolution capability outperforms conventional DUV tools, it

has significantly slower throughput, which limits its use in high volume commercial manufacturing [58, 59].

- **Laser lithography**:
  Direct laser writing (DLW) is another maskless photolithography approach, in which a laser beam and a light modulator are employed to "write" the features directly onto the surface. A spatial Light Modulator works as a "dynamic mask", in the sense that it is used to project the CAD design directly onto the wafer. Applications include research and small-volume production in most of the areas where optical lithography is required. The main use for DLW is the fabrication of the photomasks used in conventional optical lithography, but it is also suitable for the direct exposure of chips and small wafers, bypassing the need for an expensive mask. In this case, the process flow is exactly comparable to classical photolithography with the only difference in the exposure tool. The mask design is loaded in a software and it is then written directly onto a conventional photosensitive resist. The features are designed in the form of closed polygons, so it is possible to use two exposure modes: in the dark field mode the polygons are not exposed, thus in a positive photoresist they will not be developed, while in the clear field mode, the exposed polygons become soluble in the developer solution.
  The laser beam has typical wavelengths of hundreds of nanometers which allows for micrometric resolution. Achieving a sub-micrometer resolution by conventional DLW techniques is more challenging than by EBL, but the latter is more expensive and requires a high-vacuum environment, while DLW is simpler and low-cost [60].

### 3.2.2. Deposition methods

In general, the deposition processes are divided into two main categories. If physical processes are involved in the film growth, we talk about Physical Vapor Deposition (PVD), but films can also be deposited as a consequence of chemical processes, referred to as Chemical Vapor Deposition (CVD). In a PVD process, the material to be deposited (target material) is first vaporized from a solid form into plasma or ions, then transferred to the substrate surface and allowed to condense. CVD, instead, usually involves a gaseous target material that reacts near the substrate to produce the desired thin film.

- **Atomic layer deposition (ALD)**:
  ALD (or thermal ALD) is a self-limiting deposition method that allows the growth of thin films with atomic scale deposition control. It is based on surface reactions between a precursor and the reactant molecules on the surface. Gas precursors

are released in a sequential way into a vacuum chamber, left to react with the surface in a self-limiting process, and then, the unreacted precursor is purged away with the by-products so that the next precursor can be introduced. In each ALD cycle, the precursor and the co-reactant are pulsed into the chamber to produce a single monolayer during the "half-reactions" processes. After the purge, a new cycle starts, and a new monolayer is deposited until the target film thickness is achieved. Figure 3.14 shows a generic ALD cycle. The growth can be therefore controlled with high precision by controlling the number of deposition cycles. The equipment includes a heated reactor (in thermal ALD) inside a chamber where a moderate vacuum is applied. Typical temperatures for conventional thermal ALD processes are $< 350°C$. Temperature affects the growth according to the "ALD temperature window", specific for each process: temperatures outside of the window generally result in poor growth rates and non-ALD type deposition due to slow reaction rates or precursor condensation (at low temperature), or thermal decomposition and rapid desorption of the precursor (at high temperature).

It is possible to increase the reactivity at lower temperatures involving a plasma, from which the name Plasma-enhanced (PE) ALD. The gaseous plasma generates radicals which are more reactive than co-reactants, thus improving the ALD process. More precursors are available since the ALD window for some materials is expanded, because of the lower activation energy needed, widening the range of possible films that can be deposited. The lower temperatures allow for faster deposition times and avoid precursor decomposition and contaminants, yielding better film properties. However, more expensive and complex equipment is needed. ALD-deposited films allow for unique conformality of high aspect ratio and three dimensionally-structured materials, thanks to its self-limiting characteristics. [61, 62]

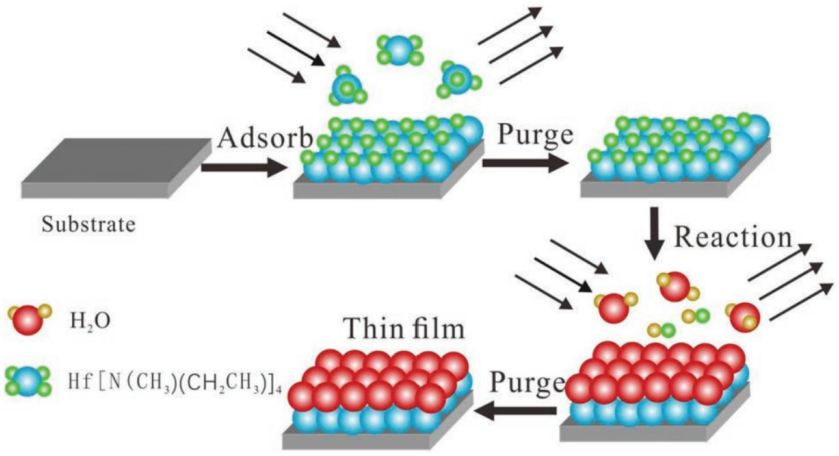



Figure 3.14: Typical ALD cycle. Adapted from [44]

- **Pulsed laser deposition (PLD)**:
  Thin films can also be deposited by PLD, where a pulsed laser beam is continuously focused onto a target of the material to be deposited. When the target spot reaches a sufficient temperature the material vaporizes, generating a plasma of high-energy ions. Inside the cavity under vacuum, the gas-phase ions collide with the background gas in the cavity and finally deposit in thin films on a substrate. In general, PLD uses a KrF excimer laser with a wavelength of 248 nm. It allows for a controlled epitaxial growth of thin films, with various crystallinity, with a relatively simple process. Thickness control is possible by tuning the number of times in which the laser hits the target in a continuous way. The deposition time is short and it ensures good film quality, but it suffers from poor adhesion and poor repeatability [37, 44].

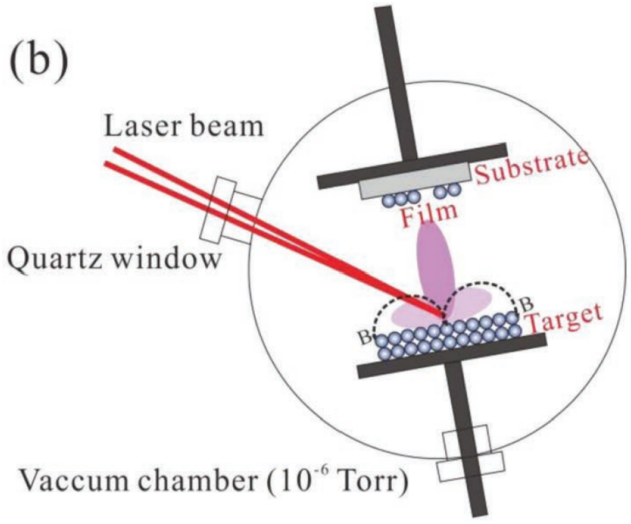



Figure 3.15: Schematic representation of a PLD process. Adapted from [44]

- **DC-sputtering**:
  DC sputtering is a PVD technique that uses a direct current as a power source. The basic configuration of DC Sputtering equipment consists of a vacuum chamber where the target material (cathode) is placed in front of the substrate to be coated (anode). The chamber is filled with ionized Ar gas, so when a DC electrical current (typically in the -2 to -5 kV range) is applied to the target the latter is bombarded by the gas atoms. These collisions will sputter the target atoms, which are vaporized in the gas plasma. Finally attracted by the positively biased anode, they condense in a thin film onto the substrate (Figure 3.16).
  DC sputtering is the simplest and most economical process among PVD metal deposition methods, and it has an impressive sputtering rate in comparison. This allows

to process large substrates quickly, and it is suitable for a wide range of materials, with problems when it comes to insulating materials [63, 64].

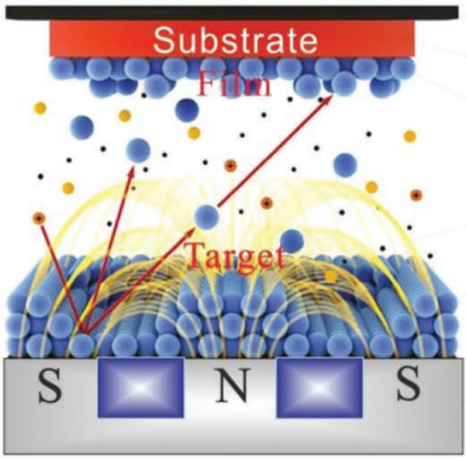



Figure 3.16: Schematic representation of a typical sputtering process. Adapted from [44]

- **Plasma-Enhanced Chemical Vapor Deposition (PECVD)**:
  Plasma-enhanced chemical vapor deposition is a CVD process for thin film growth. The reactants in the gas state (plasma) are involved in chemical reactions with the substrate, where they condense into a solid state, forming a thin film. The system comprises a vacuum chamber where the deposition process takes place, and the plasma is generated by a RF power supply. PECVD can be carried out at a relatively low temperature than CVD. Indeed, conventional CVD utilizes heat as an energy source to drive the chemical reactions, thus the substrate must be able to withstand relatively high temperatures. By utilizing plasma to provide some of the energy for the chemical reactions, PECVD enables lower temperatures for the substrates, thus coating occurs with less stress to the thin film interfaces which allows for stronger bonding. In addition, it shows higher deposition efficiency and higher tunability of the deposition conditions [65, 66].

### 3.2.3. Etching methods

During the lithographic process, the areas to be etched are left unprotected by the photoresist. The photoresist is generally not damaged by the etching and protects the material underneath. When the etching is complete, the resist is stripped leaving the desired pattern etched into the previously deposited layer. Etching is performed either using wet chemicals (Wet etching) or more commonly in a dry plasma environment (Dry etching).

- **Reactive ion etching (RIE)**:
  RIE is classified as a dry etching method since it employs a chemically reactive plasma to remove material deposited on wafers. The substrate, typically covered by a patterned photoresist, is placed inside a vacuum chamber where a plasma of reactive gases is generated by applying an electric field. The high-energy ions accelerated by the electric field towards the substrate are involved in both chemical reactions and physical sputtering at the surface. The combination of these two mechanisms removes the material selectively to create the desired pattern.
  The ion bombardment in RIE allows for high etch rates and better control of the etching direction, resulting in high anisotropy. Additionally, RIE is a versatile technique that can be used on materials including metals, semiconductors, and insulators in several applications. Furthermore, RIE etching is a dry etching method that avoids drawbacks associated with wet etching like contamination and the handling of hazardous chemicals. Critical parameters that influence the etching are the pressure, the temperature, and the gas flow rate [67].

- **Ion beam etching (IBE)**:
  The IBE tool utilizes a beam of ions focussed on the target to remove material from it. The ion beam, often made of inert gas ions like Argon, is directed towards the target surface, where the energetic ions collide with the surface atoms, causing them to be sputtered away. Unlike RIE, it doesn't rely on chemically reactive gases but only on the physical sputtering process.
  An ion source produces a collimated beam of ions with well-controlled energy and direction, which provides IBE with the ability to create anisotropic etch profile with excellent depth control. Furthermore, since the process does not involve chemically reactive species, the etching process is relatively clean and less prone to contamination. IBE is compatible with a wide range of materials, making it suitable for various applications in microelectronics, optics, and materials science. However, ion beam etching systems can be more expensive and complex compared to other ion milling techniques due to the requirements for high-vacuum environments and sophisticated ion beam generation and control systems. Moreover, it is characterized by a low etch rate, poor selectivity, and potential for device damage [68, 69].

### 3.2.4. Annealing methods

Annealing is a heat treatment used in semiconductor manufacturing to improve the crystal quality and surface roughness of wafers. It can also be used to remove defects and impurities, to activate ion-implanted dopants or to relieve stress in silicon. The normal

temperature range for annealing processes lies between 900 and 1100°C and it is normally performed in an inert ambient such as nitrogen. In this work, annealing methods have been used to induce the crystallization of oxide materials in the ferroelectric phase.

- **Rapid Thermal annealing (RTA)**:
  Rapid thermal annealing is a process used in semiconductor device fabrication that utilizes heat to affect the wafer's electrical properties. RTA was originally developed for ion implantation annealing but has broadened its application to oxidation, silicide formation, and advanced applications such as modifying the crystallographic phase of elements, enhancing lattice interface, or stress relaxation.
  During RTA, the wafer is rapidly heated from a low to a high processing temperature ($T > 900°C$) for a short time and then brought back rapidly to a low temperature. The heat source is typically an array of lamps in an optical system. In contrast to conventional furnaces, where a batch of wafers is introduced into the furnace and oxidized at the same time, RTA systems are single-wafer machines. However, due to the high processing temperature, the processing time required for oxidation is reduced, ranging from 1 s to 5 min. This makes RTA very suitable for growing thin oxide films ($< 40$ nm), where precise temperature control and short oxidation times are important.
  However, the temperature control of the wafer is challenging since a small support is used to heat the wafer rapidly. This makes it very difficult to use thermocouples for temperature measurement as it is done in a furnace. An infrared pyrometer can be used for back-side measurement, but only low precision is possible because the back-side surface conditions affect the reading. Furthermore, the wafer temperature can change by approximately 1000°C in a few seconds, which also complicates an accurate temperature measurement. Cooling must also be perfectly controlled to prevent dislocations and sample breakage [70, 71].

- **Flash Lamp Annealing (FLA)**:
  FLA is a short-time annealing method that can provide high temperatures on a time scale of milliseconds or microseconds, at the surface or the near-surface region of a substrate. During the thermal process, the surface of a wafer is treated with one or more pulses of a flash lamp. A reflector is put above the substrate to maximize the light intensity on the front side of it, while an optional halogen lamp preheats the substrate, up to a maximum of 1670 K, to reduce the energy required from the flash and consequently attenuate thermal gradients.
  The light pulse is absorbed by the substrate or wafer and converted into heat: light absorption occurs in the front side of the substrate, i.e. the surface, and the

generated heat is dissipated by thermal conduction towards the backside. As a result, the temperature on the substrate surface rises, reaches a maximum value, and then drops to an equilibrium value, while the temperature on the backside rises continuously to the low equilibrium value. The thermal stress on the backside is therefore significantly lower than on the front side. This enables higher maximum temperatures to be reached while reducing thermal stresses on the substrate below, which is not thermally loaded. Because of that, the use of temperature-sensitive substrates is possible. Furthermore, FLA allows to limit the thermally-induced dopant diffusion due to the short process duration.

FLA is suitable for a wider range of temperature-sensitive materials with respect to RTA. As a drawback, temperature is now much more difficult to estimate as the temperature profile within a sample depends on the material properties. Also, thermal stress has to be managed, and additional measures have to be taken in order to ensure process homogeneity. FLA has already been used in microelectronics, mostly to activate dopants or to recrystallize amorphous semiconductor layers, but also in the formation of silicide and germanide materials for contact fabrication [72–74].

# 4 | Passive crossbar arrays

In this chapter, we will discuss two types of ferroelectric materials, $BiFeO_3$ (BFO) and $HfZrO_4$ (HZO), that are used to build synaptic weights for two different passive crossbar arrays. These synaptic weights consist of FTJs fabricated using distinct material systems for each of the crossbars. We will then provide a comprehensive overview of how the crossbars are fabricated. In particular, it must be specified that the fabrication of the BFO crossbar took place prior to the beginning of this thesis, but we report it here for contextualizing, while the HZO crossbar was personally co-processed during this research project. A final section will be dedicated to the investigation of the effect of the metal contact on the resistive switching of BFO crossbars.

## 4.1. HZO crossbar array

### 4.1.1. Ferroelectric HZO

In the earlier stages, ferroelectric materials employed in non-volatile memory applications were developed using Perovskite materials. However, to reach thin films down to few nanometers, a price must be paid in terms of ferroelectricity suppression due to incomplete screening of polarization charges. Actually, with recent advancements in film deposition technologies, this is no longer an issue for most perovskite materials such as $BiFeO_3$ (treated in section 4.2), $BaTiO_3$ (BTO), $SrTiO_3$ (STO), $PbTiO_3$ [75]. Further issues regard CMOS process incompatibility due to mismatch with Silicon interfaces and high processing temperatures [76]. Hafnium Zirconium Oxide (HfZrO4 or HZO in this work) is a ternary metal oxide composed of two binary oxides with fluorite structure, $HfO_2$ and $ZrO_2$ [77]. Hafnium oxide ($HfO_2$) has been used for years in high-k field-effect transistor (FET) technology. Its high dielectric constant, wide bandgap and CMOS compatibility made it the most popular material for those applications. In 2011 $HfO_2$ was the protagonist of a discovery by Boscke et.al that set a breakthrough in the field of ferroelectric non-volatile memories, overcoming many of the issues related to the perovskites. He found the emergence of ferroelectricity in $SiO_2$-doped $HfO_2$ deposited in ultra-thin films.

$HfO_2$ exhibit a polycrystalline structure having multiple phases. At room temperature and pressure, the bulk crystal adopts in its stable form a monoclinic phase (space group $P2_1/c$, m-phase), or tetragonal phase (space group $P4_2/nmc$, t-phase) and cubic phase (space group $Fm\bar{3}m$, c-phase) stabilized via doping or nanostructuring. Those are centrosymmetric structures, which have dielectric properties but do not show ferroelectricity [44]. Surprisingly, experimental evidence shows that, under proper doping and annealing conditions, there exists actually a non-centrosymmetric orthorhombic phase (space group $Pca2_1$) in $HfO_2$, which can thus be ferroelectric [78]. In particular, the polar o-phase has been postulated as the transformation phase between the t- and m-phases, as shown in Figure 4.1 [80]. The doping concentration should be enough high to suppress monoclinic phase formation during the crystallization, but not too high to oppose to the transition to the orthorhombic phase [37].

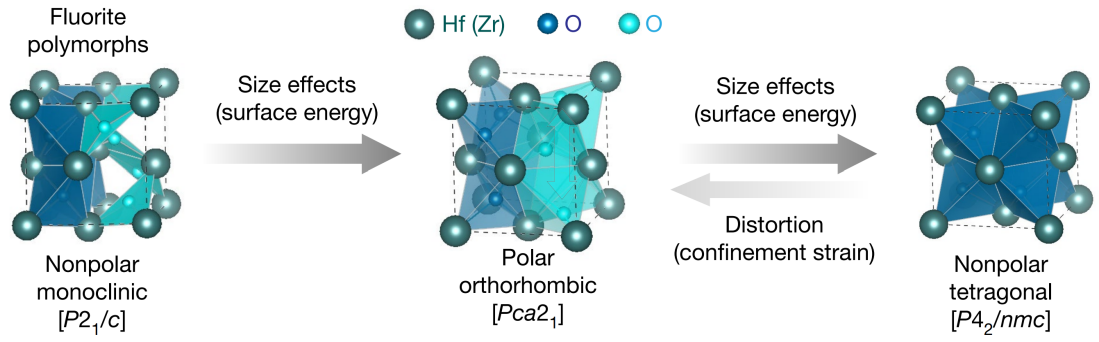



Figure 4.1: Illustration of the effect of surface energy and confinement strain in inducing the polar distorted phase. The cyan arrow represents the anions (cyan) displacement in the orthorhombic phase with respect to the cations (blue). Grey arrows indicate that surface energy always favors higher symmetry structures (monoclinic $\longrightarrow$ orthorhombic $\longrightarrow$ tetragonal) while distortion effects enhance lower symmetry (tetragonal $\longrightarrow$ orthorhombic). Adapted from [76]

Other recent studies demonstrated the same behaviour by doping with various elements such as Al, Zr, Y, Ga, and La, using different growth methods, different substrates and electrodes, and testing the most appropriate annealing method. Among all the dopants, the most promising element was Zirconium (Zr): the required doping content for maximum ferroelectric polarization can be stable at 50%, due to its very similar physical and chemical properties to those of Hafnium, while other dopants are stable at much lower doping concentrations ($< 20\%$) [78–80]. Ferroelectricity has been demonstrated in 1 nm thick ALD-grown HZO. The study has also highlighted size and confinement effects peculiar to fluorite oxides, favoring ultrathin inversion symmetry breaking. Thus, HZO has the most promising characteristics for the development of ultra-thin ferroelectric films [76].

A challenging factor is the crystallization temperature: after deposition, the annealing step is mandatory to induce ferroelectricity in the film, thus stabilizing the orthorhombic phase, but the thermal budget increases as we try to thin down the film. Ferroelectricity in HZO films can be achieved at relatively low temperatures because the crystallization temperature of Zirconium oxide ($ZrO_2$) is generally lower than that of other high-k dielectrics. The thermal budget varies in the range 400-500°C, which is the limit for compatibility with CMOS technology [77, 79]. Annealing occurs mostly under a Nitrogen atmosphere since it does not oxidize TiN, which is the typical material employed for the electrodes. The electrode material choice is fundamental in determining the ferroelectric properties of HZO; experimental results agree on the fact that HZO thin films sandwiched between TiN top and bottom electrodes show the most excellent ferroelectric properties, reporting the higher $P_r$ value [77, 79]. Another critical point to consider is the deposition process to grow HZO thin films: in addition to doping, the stable ferroelectric phase is indeed influenced by dimensional confinement, mechanical stress, ferroelectric film thickness, or surface energy induced by electrode capping layers [79, 80]. Deposition methods induce notable variations in those parameters, thus influencing the film's ferroelectric properties. Most works suggest the use of ALD, and ferroelectricity is induced by rapid thermal annealing. Other deposition methods include PVD, PLD, and chemical solution deposition (CSD). ALD allows for the growth of polycrystalline HZO films with high control of the deposition conditions. The conformal character of the growth allows to cope with surface topography problems and gives promising perspectives for 3D integration [81]. Nonetheless, it is not straightforward to obtain uniform ferroelectricity in ultra-thin films [37].
The fabrication process of HZO ferroelectric devices is CMOS compatible, and due to the low thermal budget, they can be integrated in the BEOL [81, 82]. Resistive switching behaviour was demonstrated in BEOL-compatible HZO synaptic weights scaled down to 3.5 nm. The electroresistance loop shows that very defined voltage pulses can switch the device into stable intermediate states. However, the coercive field distribution is very broad. These results are related to the polycrystalline nature of HZO, where grains have different orientations and, therefore require different field amplitudes or duration to switch. As a consequence, those devices are more suited for non-volatile memory applications and inference DNNs applications [82].

### 4.1.2. Nanofabrication

The material system of the single FTJ device for the HZO crossbar consists of a 5.5 nm thick active layer, sandwiched between 20 nm and 10 nm of TiN bottom and top electrodes. The stack is grown on 200 nm thick $SiO_2$ substrate by PE-ALD. The active

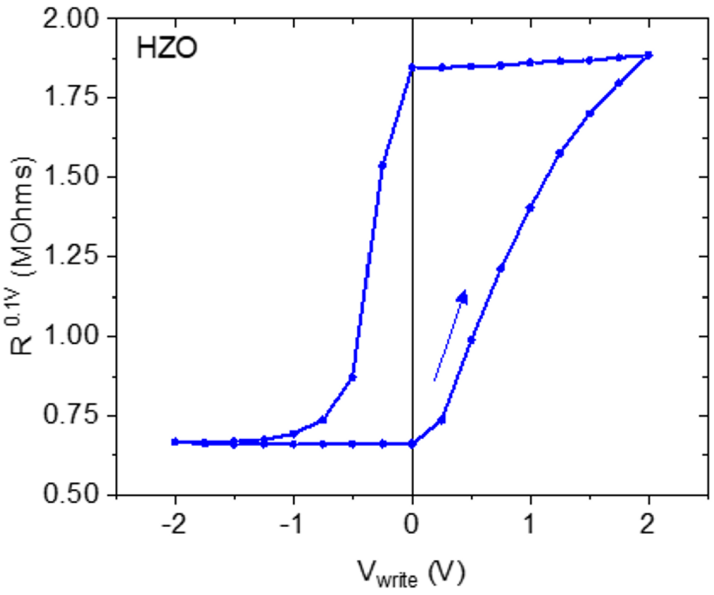



Figure 4.2: Electroresistance loop measured at 100 mV after the application of voltage pulses of variable amplitude, in TiN/HZO/WOx/TiN structure. The device has a broad coercive field distribution and well-defined intermediate resistance states. Adapted from [82]

layer consists of 2 nm thick WOx and 3.5 nm of HZO ferroelectric layer. Tungsten (W) is used as a material for the contacts for the top and bottom electrodes.

As specified above, in order to provide HZO with asymmetric interfaces, MFIM or MFS (metal-ferroelectric-semiconductor) structures are the common choice for HZO ferroelectric devices. TiOx was first studied as a metal oxide interlayer, substituted here by $WO_x$ interlayer, a metal oxide with n-type semi-conducting properties due to the presence of oxygen vacancies [83]. Here, the use of the $WO_x$ interlayer allows to induce an asymmetry in the energy profile of the memristor and an increased On/Off ratio. The bottom and top metallic electrodes are made of TiN, to favor the crystallization of HZO in the ferroelectric phase during annealing [81, 82].

It follows the detailed process flow: a 200 nm thick $SiO_2$ oxide was grown on Si by thermal oxidation. The active stack was then deposited by PE-ALD: 20 nm of TiN was deposited at 300°C with Tetrakis(dimethylamino)titanium and $N_2$ as precursors. 2 nm of WOx was deposited at 375°C with $(BuN)_2W(NMe_2)_2$ and $O_2$, then 3.5 nm of HZO was deposited at 300°C alternating one cycle with Tetrakis (ethylmethylamino) hafnium (IV) and $O_2$, and two cycles with Bis (methylcyclopentadienyl) (methyl) (methoxy) zirconium (IV) and $O_2$. Ten additional nanometers of TiN were deposited. The crystallization was performed with the millisecond flash lamp annealing technique: the sample was preheated to 400°C,

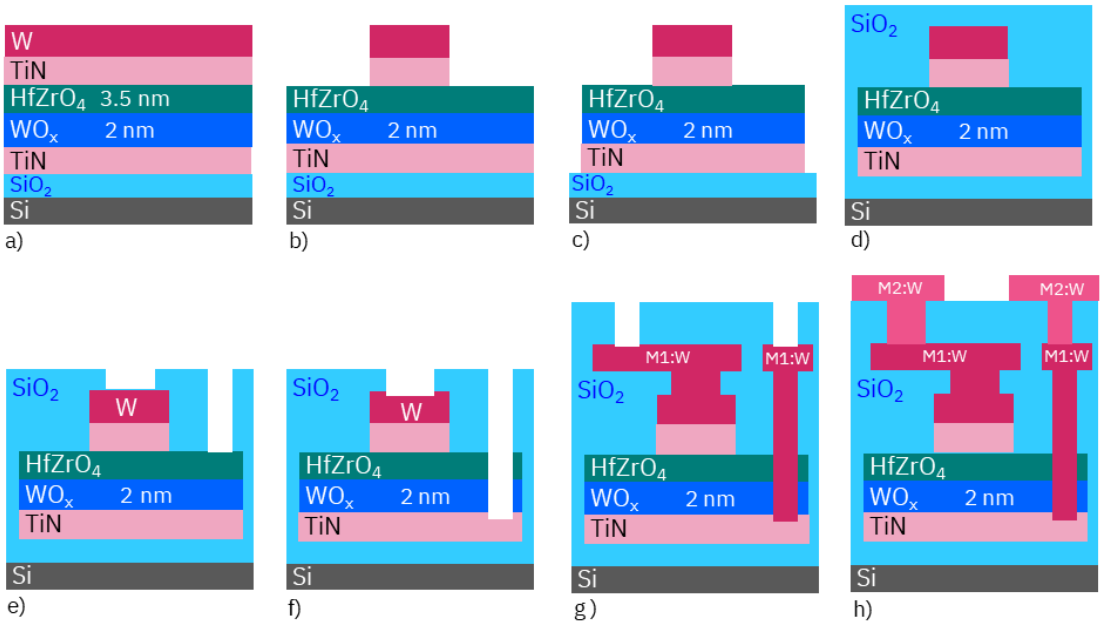
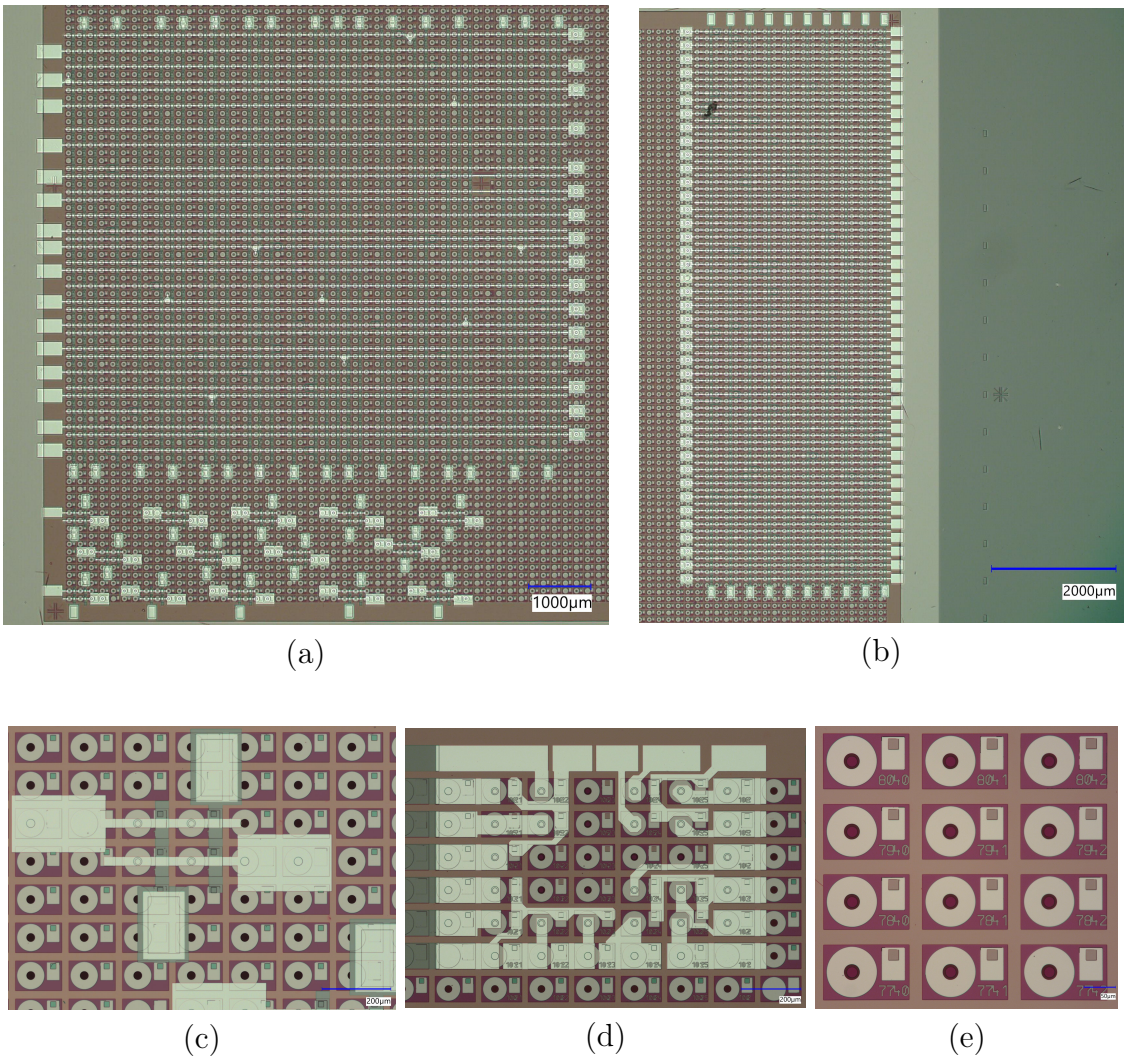



Figure 4.3: Process-flow for HZO crossbar: **(a)** Deposition of the full material stack after annealing. **(b)** Definition of the top electrode by optical lithography and RIE of top TiN and W layers. **(c)** Definition of the bottom electrode by optical lithography and IBE of the HZO, WOx and TiN layers. **(d)** The devices are isolated from each other by a $SiO_2$ passivation layer deposited by PECVD. **(e)** Vias to the device contacts are etched, $SiO_2$ layer by RIE, **(f)** then the HZO and the WOx by IBE. **(g)** M1 is deposited by sputtering and patterned by RIE. $SiO_2$ passivation layer is deposited by PECVD and vias to the bottom electrode are etched. **(h)** A second passivation layer of $SiO_2$ is deposited to isolate M2 from M1. Vias to M1 are etched. M2 is deposited by sputtering, then patterned by RIE.

then a 20 ms long energy pulse of 90 $J{\cdot}cm^{-2}$ was applied. A 100 nm thick W metal electrode was then deposited by sputtering (Figure 4.3 (a)). The top electrode, defining the area of the junction, was defined by optical lithography and RIE of the W and top TiN layers. Using this method, the HZO layer acted as an etch stop (Figure 4.3 (b)). The bottom electrode was then defined by optical lithography and IBE of the HZO, WOx and TiN layers, as shown in Figure 4.3 (c). A 100 nm thick $SiO_2$ passivation layer was deposited at 300°C by PECVD, represented in Figure 4.3 (d). Vias to the top and the bottom electrodes were defined by optical lithography. The $SiO_2$ layer was etched by RIE, then the HZO and the WOx were etched by IBE, exposing the TiN layer to air. The etch was immediately followed by the sputtering of 100 nm of W (Figure 4.3 (e) and (f)). The first metal lines (M1) were then defined by optical lithography and etched by RIE. A 100 nm thick $SiO_2$ passivation layer was deposited at 300°C by PECVD. Vias to the

bottom electrode contacts were defined by optical lithography (Figure 4.3 (g)). The $SiO_2$ layer was etched by RIE. 100 nm of W was sputtered. Figure 4.3 (h)) shows the final cross-section where also the second metal lines (M2) are defined by optical lithography and RIE.

Two crossbar arrays of sizes 32x32 and 81x10 have been fabricated on a single chip, plus some 2x2 and 4x4 arrays for device characterization (Figure 4.4).




(a) (b)

(c) (d) (e)

Figure 4.4: Optical microscope snapshots of HZO crossbar arrays. 32x32 array and 2x2 arrays for device characterization **(a)**, 81x10 array **(b)**, 2x2 array **(c)** and 4x4 array **(d)**. Snapshot of non-contacted FTJ devices (e).

The mask design of the chip, represented in Figure 4.5, has been designed with the aid of Klayout software. The design consists of several layers corresponding to different lithographic steps. The Klayout design is loaded into the DLW software as a GDSII file,

and for each lithographic step, the proper pattern is selected to be directly written on the substrate.

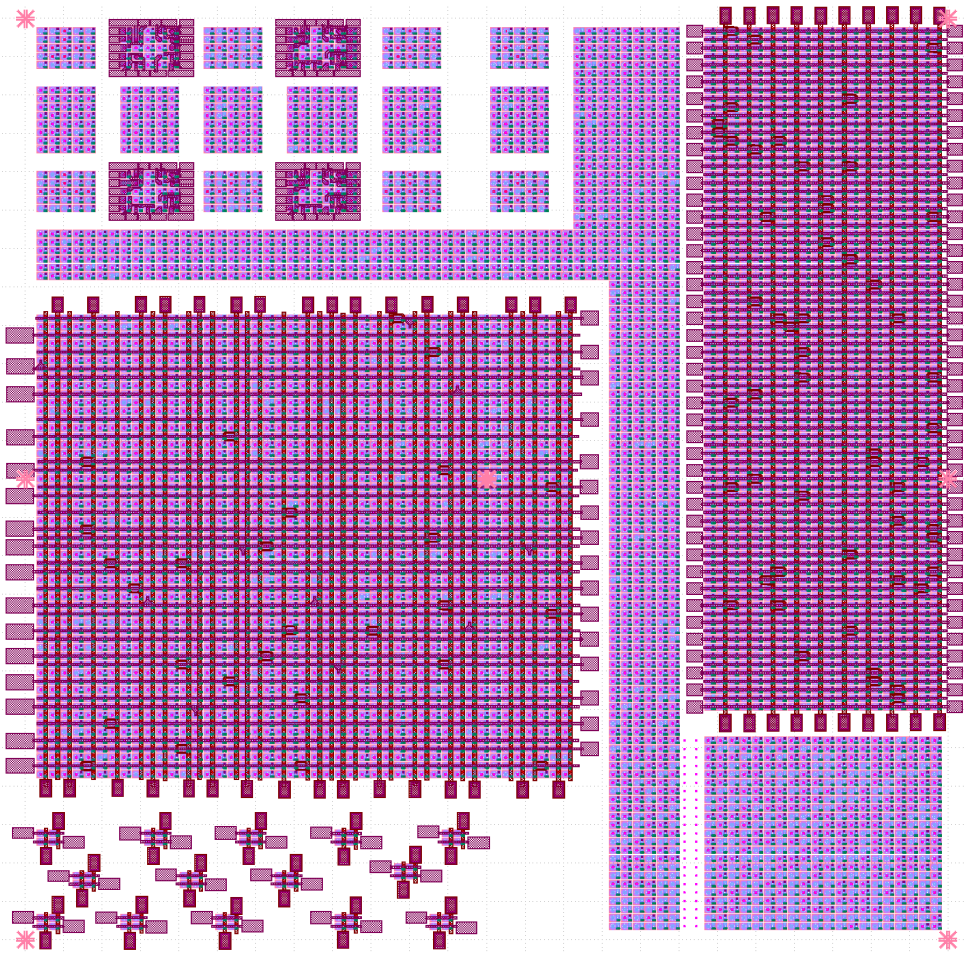


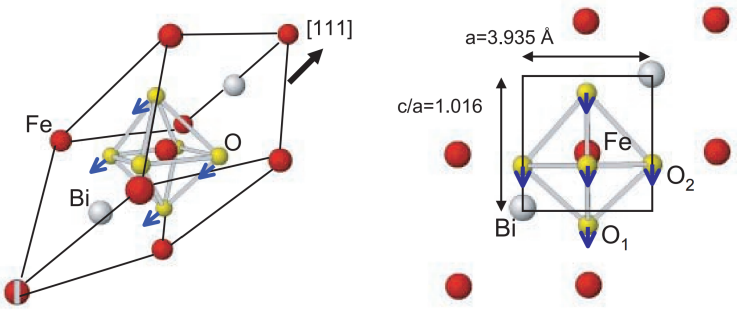
Figure 4.5: Mask design for DLW lithography, performed on Klayoyt software. Full chip with 32x32, 81x10, 2x2, and 4x4 HZO crossbars.

## 4.2. BFO crossbar array

### 4.2.1. Ferroelectric BFO

Bismuth ferrite ($BiFeO_3$ or BFO) is a perovskite oxide known for its excellent multiferroic properties. Multiferroic materials have gathered a lot of research interest due to their potential in storage and low-power spintronic applications. They have simultaneous ferroelectric, magnetic, and ferroelastic properties [84–86]. Bulk BFO was shown to have polarization values of $\simeq 100\mu C\ cm^{-2}$ along the [111] direction. This is the highest reported value for a bulk ferroelectric. It is ferroelectric below a very high Curie tem-

perature of $\simeq 1100K$ and antiferromagnetic with Néel temperature $T_N \simeq 640K$. First synthesized in the late 1950s, its bulk form crystallizes in the distorted rhombohedral structure (space group $R3c$) [84].




(a) (b)

Figure 4.6: Rhombohedral **(a)** and Tetragonal **(b)** BFO unit cells. Adapted from [86].

High-quality thin film BFO can be grown epitaxially on single-crystal substrates. Interestingly, it has been shown that epitaxial strain exerted by the substrate will eventually cause the film to assume a structure of reduced symmetry with respect to the bulk. In particular, the growth technique and the substrate can be studied to allow a desired percentage of epitaxial strain on the growing film by the so-called *strain engineering*. For large compressive strain of a percentage higher than 4.5, along the (001) orientation, a BFO highly-distorted tetragonal $P4mm$ phase (T-phase) can be stabilized (Figure 4.7). According to theoretical predictions, values of giant polarization of $150\mu C\ cm^{-2}$ can be obtained in thin films. A precise tailoring of growth parameters and substrate choice in terms of induced epitaxial strain is fundamental for the stabilization of the T-phase BFO. Most of the progress in BFO epitaxial thin films has been achieved using STO (0 0 1) substrates, which induce a compressive strain to the film. Evidence has shown pure tetragonal BFO ultrathin films on STO up to 10 nm. Transition to monoclinic BFO after 15 nm is usually obtained and over 30–90 nm the BFO reaches the bulk structure.

Heteroepitaxial growth of thin films has reached great improvements due to recent advancements in PLD technology. The growth of polar oxide heterostructures and superlattices with perovskite structures is achieved by PLD. However, the lack of Si compatibility has limited the practical application of the PLD-based polar oxides [87]. Most of the applications of BFO thin films are appealing for its large polarization. Very attractive for gate oxides in FeFETs where large carrier density modulation can be achieved in the channel, and in ferroelectric tunnel junctions, where giant electroresistance effect has been reported [84]. Previous studies have reported very large non-volatile modulation of the transport properties in FeFETs based on supertetragonal BFO as gate oxide coupled with

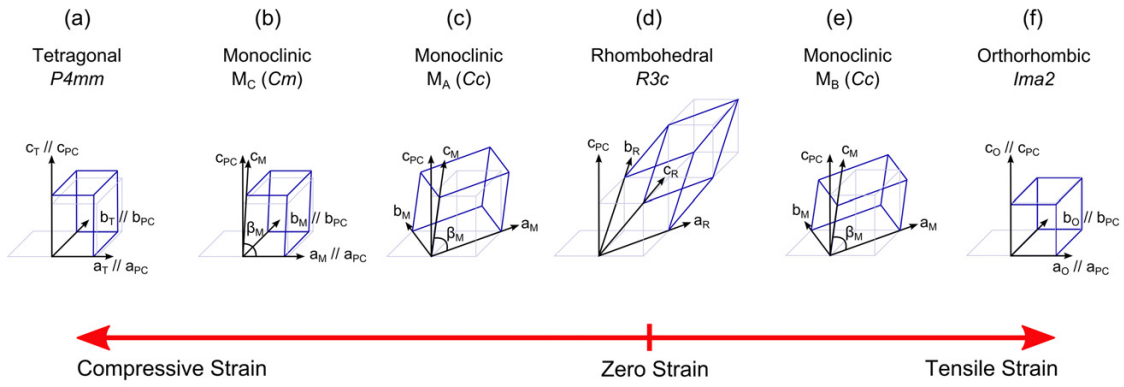



Figure 4.7: Representation of the various crystal structures that BFO thin film assumes under epitaxial strain. In blue are the unit cell of the relative structure, in grey the pseudocubic perovskite unit cell- Adapted from [84].

a $CaMnO_3$ (CMO) channel [88]. CMO is a Mott insulator, belonging to the class of strongly correlated oxides, that can become metallic upon slight electron doping. This is the case of $Ca_{1-x}Ce_xMnO_3$, where a concentration of x=0.04 of $Ce^{4+}$ induces a metallic transition in the parent material due to electron-doping [88]. Correlated oxides used as electrodes in FTJs can be subjected to field-induced electronic phase transitions upon polarization reversal, resulting in enhanced tunnel electroresistance. In a study of 2017 by *Boyn et al.* [25] synaptic functionality has been demonstrated on a FTJ consisting of supertetragonal BFO tunnel barrier combined with CCMO and Co electrodes. In order to stabilize the tetragonal BFO in ultrathin films, the selected substrate is $YAlO_3$ (YAO), since it can exert enough strain on the BFO thin film. In order to allow epitaxial growth of BFO, CCMO is chosen as a bottom electrode since it is also well lattice-matched with YAO [89].

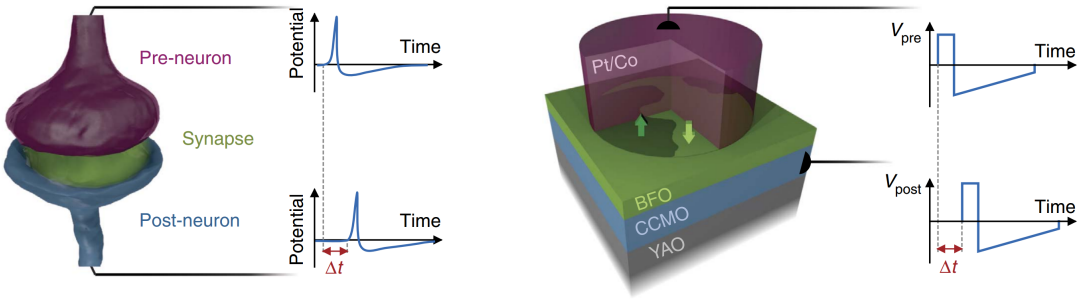



Figure 4.8: Sketch of pre- and post-neurons connected by a biological synapse and an artificial one, i.e. the Co/BFO/CCMO/YAO ferroelectric tunnel junction. The synaptic plasticity relies on the causality ($\Delta t$) of neuron spikes. Adapted from [25]

The study demonstrated synaptic functionality with giant electroresistance up to $10^4$, in addition to high endurance and operation speed. In particular, the electroresistance hysteresis loop (Figure 4.9) is characterized by well-defined voltage thresholds ($V_{th}^{+}$ and $V_{th}^{-}$), a feature that allowed to demonstrate STDP in the ferroelectric memristor (Figure 4.8). In fact, the synapse is strengthened (weakened) only if there is a causal (anticausal) relationship between pre- and post-neuron spikes; in order to achieve this functionality, only closed-time spikes can produce a conductance change, so it is important that a single spike never exceeds the threshold, leading to an unwanted resistance variation. The study demonstrated that the application of cumulative pulses of constant amplitude leads to a gradual reversal of the polarization by gradually expanding the existing domains and nucleating new ones. This cumulative switching feature can be ascribed to the single-crystal nature of epitaxial BFO, and it is the cause of the sharp coercive field distribution.

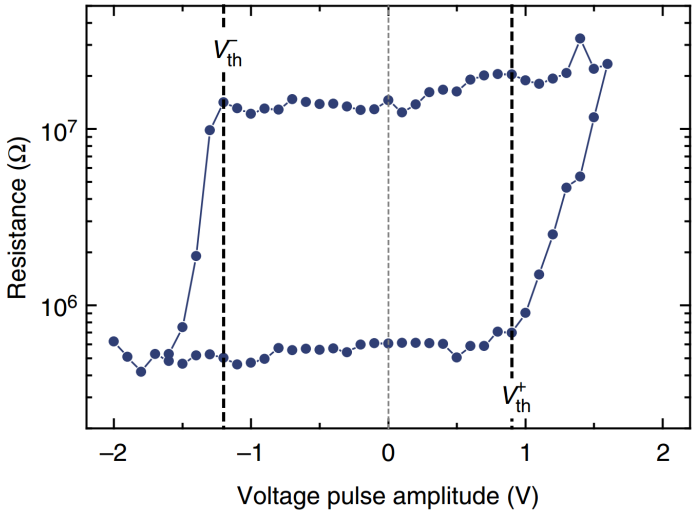



Figure 4.9: Electroresistance hysteresis loop of the ferroelectric memristor displaying clear voltage thresholds $V_{th}^{+}$ and $V_{th}^{-}$. A single voltage pulse is applied to switch the device into many intermediate states. Adapted from [25]

The next section will present the fabrication process of a crossbar array whose synaptic weights are the ones reported above and proposed in ref.[25], where instead of Co electrodes, Ni has been tested for the first time as a material for top and bottom electrodes.

### 4.2.2. Nanofabrication

The material system of the BFO crossbar array consists of a 4 nm thick ferroelectric layer (BFO) and a 17 nm thick oxide electrode ($Ca_{0.96}Ce_{0.04}MnO_3$ – CCMO) grown on $1 \times 1$ $cm^2$ single crystal substrates of $YAlO_3$ by pulsed laser deposition. CCMO was grown at

670°C under an oxygen pressure of 0.2 mbar, and BFO was subsequently grown at 580°C with an oxygen pressure of $6 \times 10^3$ mbar. The samples were then annealed for 30 min at 500°C under high oxygen pressure (300 mbar) [90].

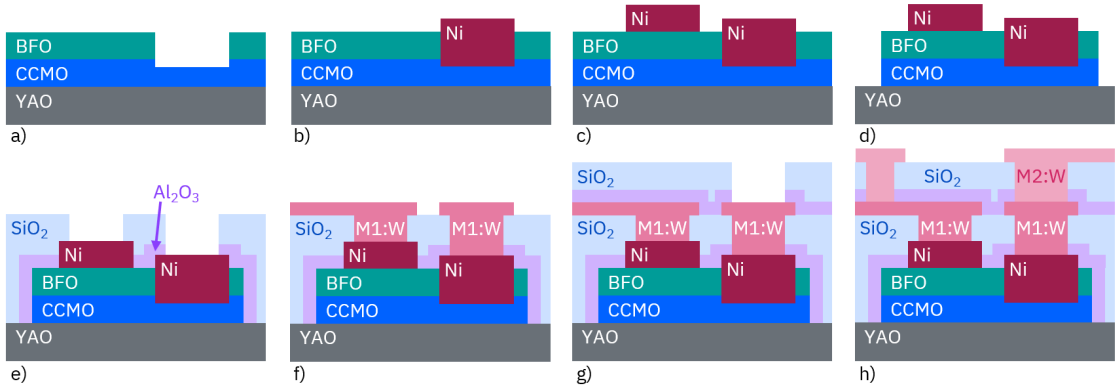



Figure 4.10: Process-flow for BFO crossbar: **(a)** BFO is etched by ion milling to access the CCMO back electrode. **(b)** Using the same e-beam resist from **(a)**, Ni contacts to the CCMO are realized by lift-off. **(c)** Using a lift-off process, Ni contacts are placed above BFO to form a FTJ device (Ni/BFO/CCMO). **(d)** Each device is isolated from each other by a surrounding trench that is etched by ion milling. **(e)** A thin passivation ($Al_2O_3$ (5 nm) / $SiO_2$ (100 nm)) is deposited to isolate M1 from the BFO. Vias to the device contacts are etched, and **(f)** M1 is deposited by sputtering and patterned by RIE. **(g)** A second passivation ($Al_2O_3$ (5 nm) / $SiO_2$ (300 nm)) is deposited to isolate M2 from M1. Vias to M1 are etched. **(h)** M2 is deposited by sputtering, then patterned by RIE.

To demonstrate the fabrication of crossbars based on FTJ devices with sub-micrometer size, an EBL process was established. The non-conducting and transparent properties of the $YAlO_3$ substrate require extra precaution. The original approach reported in this process-flow consists of using a sacrificial water-soluble polymer, coated on top of the e-beam resist. Then, chromium is deposited by sputtering to avoid any charging effect during the e-beam exposure of the resist. After the exposure, the metal is lifted off by dissolving the polymer in water.
The detailed process flow for the BFO FTJ devices with Ni contacts is the following: EBL is applied to open the back contact to the device by locally etching the BFO layer (ion milling tool with an Ar plasma and a beam current of 250 mA, Figure 4.10(a)). Then, by using the same patterned resist, a Ni lift-off is performed (Figure 4.10(b)). Simultaneously, e-beam markers are deposited. Next, the top contact (Ni) is realized by a lift-off process on the BFO to form the FTJ, as represented in Figure 4.10(c). Figure 4.10(d) shows the result of the device isolation step, necessary to electrically disconnect all devices from each other. A 100 nm thick $SiO_2$ passivation (with 5 nm $Al_2O_3$ as etch stop) is added on top

by PECVD at 300°C, vias to the device contacts are dry etched by RIE in a $CHF_3$ plasma (Figure 4.10(e)). Then, tungsten (100 nm) metal line and pads (M1) are formed on top of the passivation (Figure 4.10(f)). At this stage, a single test device can be measured. An additional passivation layer ($Al_2O_3$ (5 nm) /$SiO_2$ (300 nm), (Figure 4.10(g)) and metal layer (M2) is then patterned. The final cross-section is depicted in Figure 4.10(h) [91].

### 4.2.3. Impact of the metal contact processing on the ferroelectric resistive switching

In addition to the crossbar arrays, single devices with access pads at the first and second metal levels (M1 and M2) were fabricated for device characterization. In particular, resistance hysteresis measurements were taken from a test FTJ, having a diameter of 1 $\mu m$, after processing up to M1 and then, after processing up to M2. The setup is shown in Figure 4.11, where voltage pulses of varying amplitude are applied to the top electrode of the FTJ while the CCMO electrode is grounded.

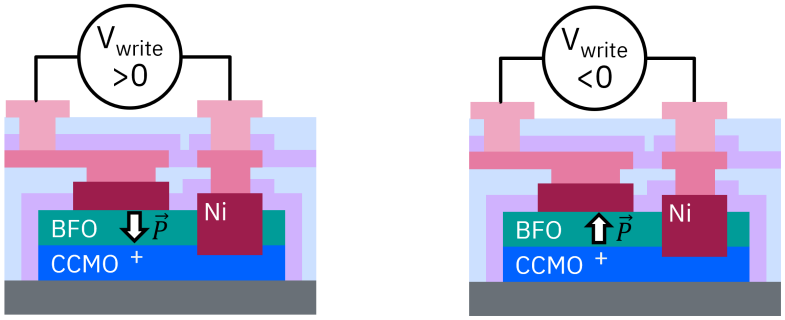



Figure 4.11: Experimental setup for single device characterization, with access pads to M1 and M2. The device is set to LRS after writing with a positive amplitude **(a)** and to HRS with a negative writing amplitude **(b)**.

The measurement process consists first in the application of a write signal, with an amplitude "$V_{write}$" and duration of 10 ms, on the top electrode. Then, the resistance of the junction "$R_{0.2V}$" is read by applying an I-V sweep from -0.2 V to 0.2 V.

After each reading step the write amplitude is increased or decreased, and the same procedure is repeated until a full resistance hysteresis loop is complete. Starting from 0 V, the write amplitude is increased towards positive values to switch the device into the Low Resistive State (LRS). From the maximum positive value (4 V), the write amplitude is then decreased towards the maximum negative value (-4 V) to switch the device into the HRS (Figure 4.11). In Figure 4.12 five hysteresis cycles are represented, the direction is indicated by the black arrow. At the M1 stage (blue curve), the low resistive state

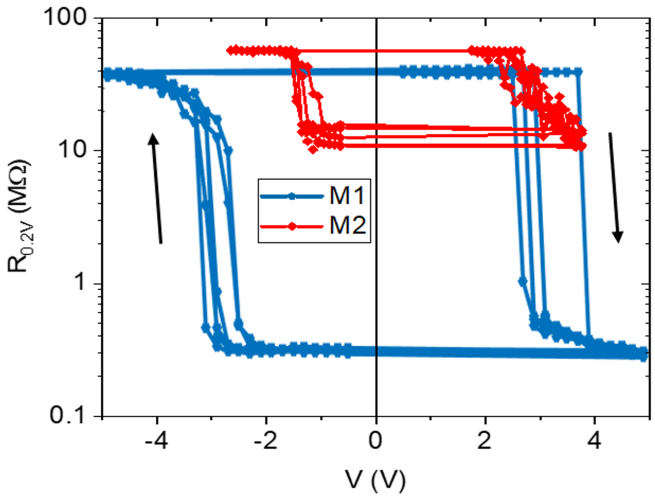



Figure 4.12: Resistance hysteresis measurements of a FTJ (diameter 1 $\mu$ m). Each measurement is repeated five times, in each the resistance is measured at $V_{read}$ = 0.2V after applying pulses of amplitude $V_{write}$. Blue curve: after defining the first metal level. Red curve: after defining the second metal level. The arrow indicates the direction of the hysteresis. Published in [91].

has a resistance of $R_{On}$ = 275 kΩ and the On/Off ratio was 152. After finishing the full process up to M2, the low resistive state has increased more than an order of magnitude to $R_{On} \simeq$ 10 MΩ (red curve) and the On/Off ratio decreased to 6, a value comparable to the On/Off of the HZO crossbar (Figure 4.2). Those devices are subjected to an aging of the contact to the BFO, leading to the formation of a NiOx layer at the Ni/BFO interface [91].
To understand the origin of the inverted polarity and of the change in the On/Off ratio, we investigate the resistive switching mechanisms in the aged system.

# 5 | Analysis of the resistive switching mechanisms in the BFO FTJs

## 5.1. Conduction mechanisms in BFO FTJ

To understand how the metal contacts affect the On/Off ratio of the aged devices, we investigated the dominant conduction mechanism for carriers inside the $BiFeO_3$ films. To this aim, we employed temperature-dependent I-V measurements represented in Figure 5.1. As indicated by the arrows, the applied voltage amplitude is increased towards positive values starting from 0 V, subsequently decreased towards the maximum negative voltage value, and increased again to reach 0 V.

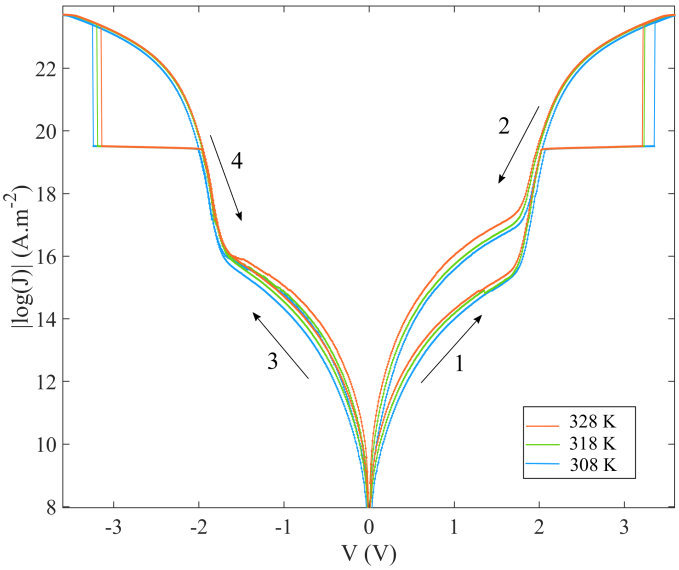



Figure 5.1: Temperature-dependent I-V measurements in $|log(J)| - V$ representation. The arrows are numbered according to the chronological order in which the applied voltage amplitude is increased or decreased.

To decorrelate the ferroelectric switching from the transport mechanisms, only the non-switching curves shown in Figure 5.2 of the full I-V sweep are analyzed. The I-V in HRS is analyzed for $V < 0$, while in the LRS, the I-V is analyzed for $V > 0$. For voltages smaller than 1.6 V in absolute value, the resistance decreases with increasing temperature, suggesting that the transport is thermally activated. The possible conduction mechanisms include interface-limited Schottky emission, space-charge-limited bulk conduction (SCLC), bulk-limited Poole-Frenkel emission (PF), and variable-range hopping (VRH). A the voltage exceeds 1.6 V, the current becomes temperature dependent, indicating that in this regime a tunneling mechanism (direct tunneling, Fowler-Nordheim tunneling [46] or trap-assisted tunneling [49]) is dominant.

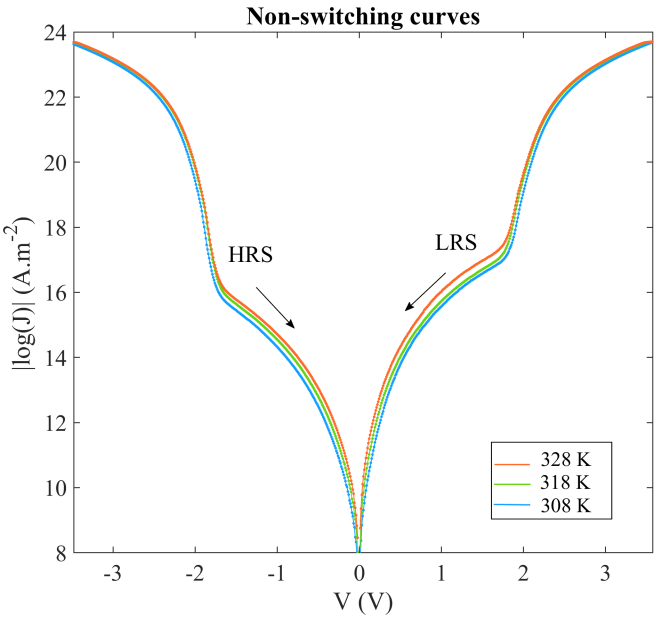



Figure 5.2: Non-switching branch of the temperature-dependent I-V measurements. The positive branch corresponds to the LRS, while the negative branch to the HRS.

To identify the possible conduction mechanisms, the data are represented according to the analytical expression relative to each process, which shows linear behavior, as proposed by [46]. Then, a linear regression is performed to fit the data.

## 5.2. Analysis from low to medium fields

### 5.2.1. The SCLC model

In the low/medium field range, Poole-Frenkel emission, Schottky emission, and variable-range hopping fittings were not satisfying. Although the PF emission fit showed linearity

in a limited range, an attempt to compute the dielectric constant $\epsilon_r$ yielded a value as high as 100, which is too high when compared to the reported value of $\approx 6.5$ [92, 93].

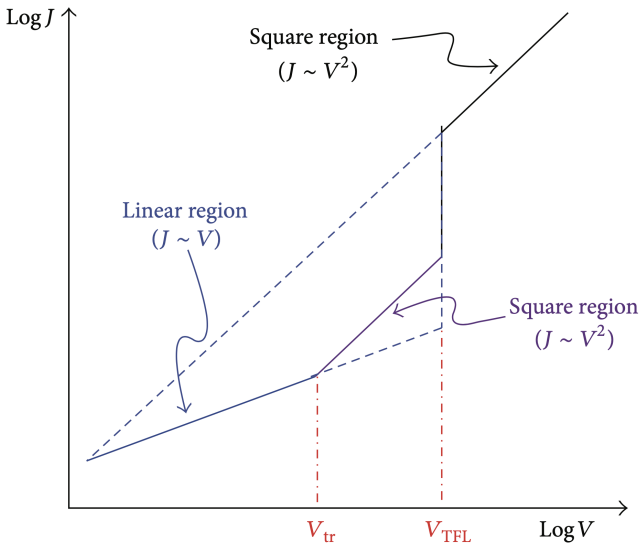



Figure 5.3: Typical current density-voltage (J-V) characteristic in the $\log J$-$\log V$ plane for space-charge-limited conduction current for an insulator with a discrete shallow trap level. The three bounding curves are the Ohm's law ($J \propto V$), traps-filled limit (TFL) current ($J_{TFL} \propto V^2$), and Child's law ($J_{Child} \propto V^2$). $V_{tr}$ and $V_{TFL}$ are the transition voltage at the departure from ohm's law and TFL curve. Adapted from [46].

For a SCLC model based on the assumption that the dielectric has a discrete shallow trap level, we expect behavior in accordance with the predictions shown in Figure 5.3 [46].

Figure 5.4 shows $\log_{10}(V)$ vs. $\log_{10}(J)$ curves in both HRS and LRS. At low voltages (below 70 mV), the slope is approximately 1, suggesting linear Ohmic behavior. In the intermediate bias region (1V < V < 1.58 V for LRS and −1.4 V < V < 0.5 V for HRS), a power law dependence ($J \propto kV^m$) yields a good fit, suggesting SCLC mechanism as the dominant one.

The parameter $m$(T) represents the slope of the $\log_{10} V - \log_{10} J$ characteristics in region II. However, the slope values are higher than 2 and they decrease with an increase in temperature. Furthermore, as shown in Figure 5.5, $m$(T) varies linearly with $T$. This analysis suggests a SCLC current governed by traps that are exponentially distributed in energy (EDT-SCLC) [52, 94–99].

BFO thin films seem to show high leakage current density due to the presence of acceptor or donor states in the material band gap. The former may originate from the presence

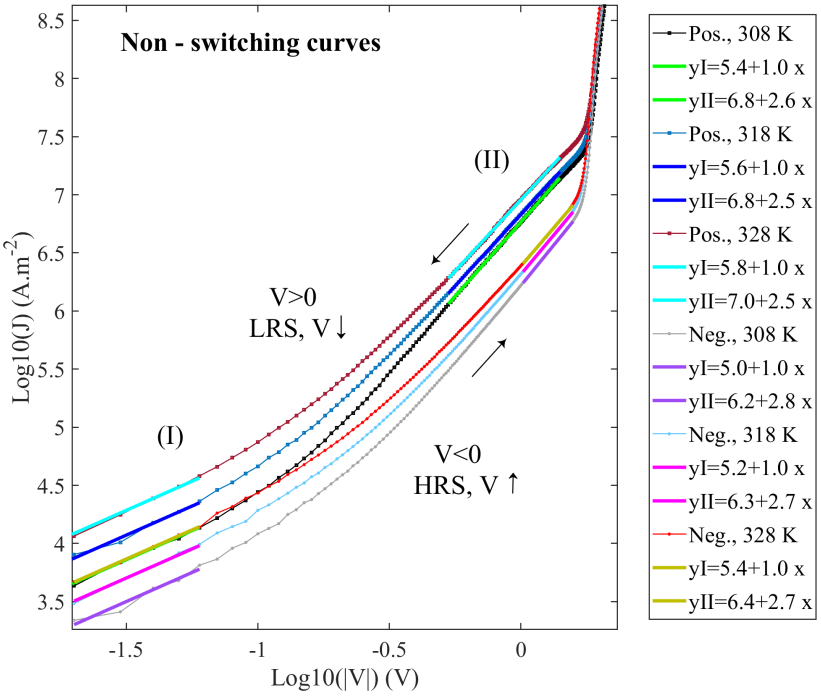



Figure 5.4: Current-voltage characteristics in the Ohmic and Space-Charge-limited conduction (SCLC) representations. At each temperature, a linear regression is performed on the negative, increasing branch (HRS, nonswitching) and on the positive, decreasing branch (LRS, nonswitching).

of Bi deficiencies in the lattice [100, 101], while the presence of oxygen vacancies or the valence fluctuation of Fe ions would result in n-type conduction [102, 103]. We assume that BFO films in the present study are p-type due to the high stoichiometry of the films, which excludes the dominance of oxygen vacancies.

### 5.2.2. Analysis in the Ohmic regime

The Ohmic current is described by:

$$J = \sigma E = \mu q N_\nu \exp\left[-\frac{E_F - E_\nu}{k_B T}\right] E \tag{5.1}$$

or equivalently:

$$\log_{10}(J) = \log_{10}\left(\frac{\mu q N_\nu}{d}\right) - \frac{E_F - E_\nu}{k_B} \times \frac{1}{T} + \log_{10}(V) \tag{5.2}$$

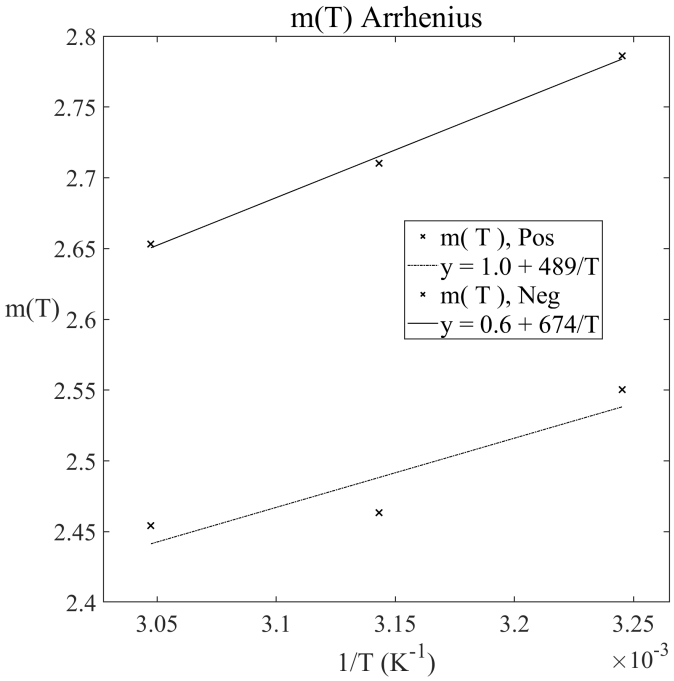



Figure 5.5: Arrhenius plot of the slopes ($m$) obtained in the SCLC region of I-V characteristics for **(a)** positive nonswitching branch (LRS) and **(b)** negative nonswitching branch (HRS). From the slope of the Arrhenius plot, the parameter $T_C$ is calculated in the LRS and HRS.

where $\mu$ is the hole mobility, $q$ is the electronic charge, $k_B$ is the Boltzmann constant, and $T$ is the absolute temperature. $E_F - E_v$ is the energy difference between the valence band and the Fermi level, and $N_v$ is the effective valence band density of states. The electric field $E$ was assumed to drop mainly across the ferroelectric layer, resulting in $E = V/d$, where the thickness $d$ is equal to 4.1 nm.

Figure 5.6 shows the Arrhenius plot of the intercepts of the linear regression in the Ohmic regime: the position of the Fermi level is then estimated to be at 0.43 eV above the valence band of BFO in the LRS, and decreases to 0.38 eV in the HRS, indicating that the barrier height change is not the dominant mechanism for the resistive switching. The carrier concentration-mobility product $\mu N_v$ can be considered to be temperature-independent because $\mu$ varies as $T^{-3/2}$ and $N_v$ as $T^{3/2}$. The extracted values are $\mu N_v^+ = 5.48 \times 10^{20}\,(\mathrm{cmVs})^{-1}$ in LRS, which decreases to $\mu N_v^- = 2.68 \times 10^{19}\,(\mathrm{cmVs})^{-1}$ in the HRS, consistently with the larger resistance observed.

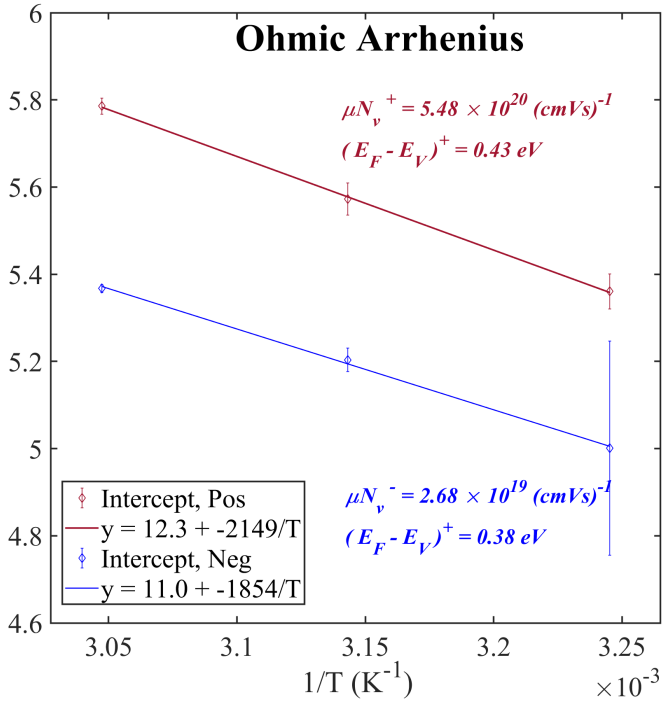



Figure 5.6: Arrhenius plots of the intercepts of the linear regressions in the Ohmic regime. The $\mu N_v$ and $E_F - E_v$ parameters are calculated from the intercept and the slope, respectively.

| | $\mu N_v$ | $E_F - E_v$ |
|---|---|---|
| LRS | $5.48 \times 10^{20}\,(\mathrm{cmVs})^{-1}$ | 0.43 eV |
| HRS | $2.68 \times 10^{19}\,(\mathrm{cmVs})^{-1}$ | 0.38 eV |

Table 5.1: $\mu N_v$ product and $E_F - E_v$ in the Ohmic regime for LRS and HRS.

### 5.2.3. Analysis of the trap energy distribution in the EDT-SCLC regime

For EDT-SCLC data are fitted with the following model:

$$\log_{10}(J) = \log_{10}(k) + m(T)\log_{10}(V) \tag{5.3}$$

According to the Mark and Helfrich model [50], $m(T) = l(T) + 1$ where $l(T) = T_C/T$. $T_C$ is the characteristic temperature, a parameter that indicates how many shallow traps exist and determines the profile of the energy distribution of the trap sites. A larger $T_C$ value indicates a smaller number of shallow traps [51] and a narrow profile, and vice versa, if $T_C$ is small. Its physical meaning could be related to the temperature in the

final cooling-down stage of the preparation of the material at which annealing effectively ceased [94]. Equivalently, one can characterize the exponential tail of defect density of states in terms of characteristic energy with the parameter $E_t = T_C \cdot k_B$.

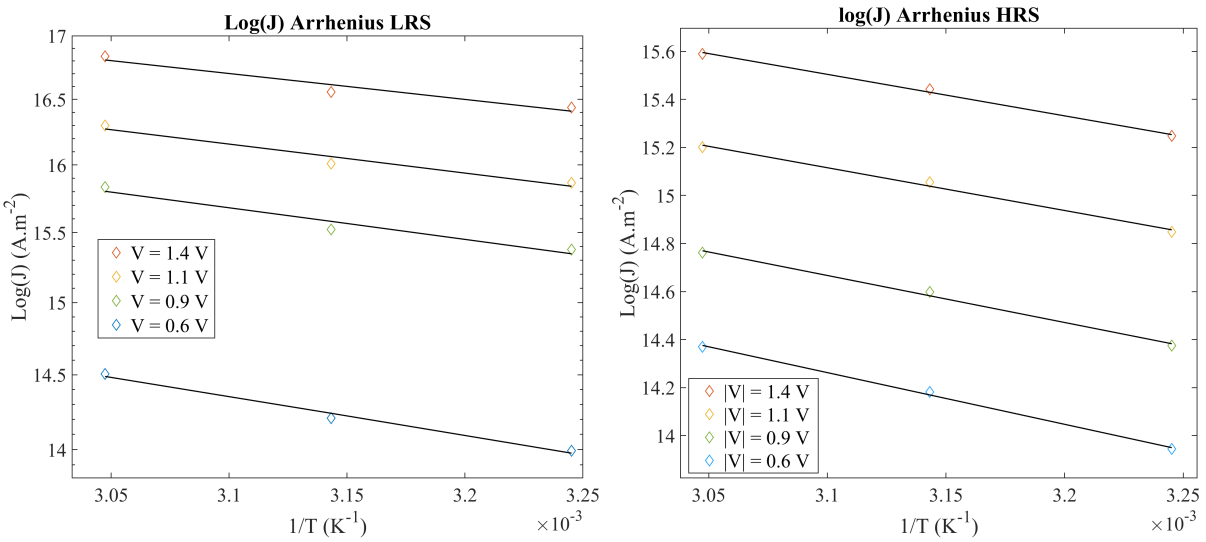



Figure 5.7: Arrhenius plots of $\log J$ for **(a)** positive nonswitching branch (LRS) and **(b)** negative nonswitching branch (HRS). The current density $J$ is computed at 4 different bias values using Eq. (5.4). The Arrhenius slopes are related to the activation energies.

The slope of the Arrhenius plot of $m(T)$ (Figure 5.5) directly determines the $T_C$ value, being 489 K for LRS and 546 K for HRS, for which $E_t = 0.047$ eV and $E_t = 0.042$ eV, respectively. These characteristic energies indicate that the exponential distribution has a relatively narrow profile, but they don't give information about the energy range around which traps are distributed.

| | $T_C$ | $E_t$ |
|---|---|---|
| LRS | 489 K | 0.047 eV |
| HRS | 546 K | 0.042 eV |

Table 5.2: Characteristic temperatures and energies in EDT-SCLC regime for LRS and HRS.

We expect that the current superlinear increase is due to a dominant shallow level which can be completely filled up while deep traps start to be filled only partially. These results lead to the following physical interpretation: Ohmic conduction at low biases indicates that carriers are injected from the electrode and partially fill the shallow traps, and with increasing voltage, the shallow traps become completely filled at the transition voltage $V_{tr}$, and then deep traps start to become partially filled.

Additional information can be recovered from the estimation of the traps' activation energies. In fact, the temperature dependence in this regime shows Arrhenius thermal activation, and the related slopes are directly linked to the traps' activation energy [51, 97, 98, 104, 105]. The Arrhenius dependence of $\log(J)$ is shown in Figure 5.7 for a few voltages. The fact that the slope changes with voltage is a further confirmation that traps are distributed exponentially in energy [98]. The activation energies are distributed around 0.17 eV and 0.23 eV for LRS and 0.15 eV and 0.23 eV for HRS. These are the trap levels involved in the conduction, which are already known to be distributed exponentially with parameter $T_C$ (Figure 5.8) [106].

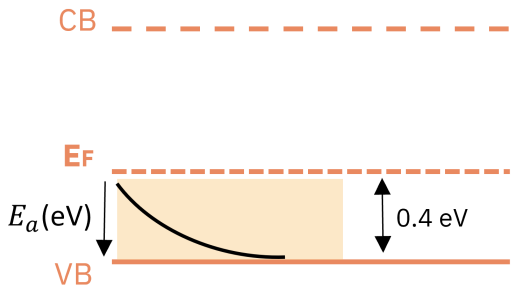



Figure 5.8: Schematic representation of shallow traps distributed exponentially with parameter $T_C$ in the BFO band gap. The black solid line denotes the variation of the trap density $N_t$ with the black arrow, which corresponds to the energy axis along which the activation energy varies.

### 5.2.4. Analysis of the trap density in the EDT-SCLC regime

It is worth showing that the Mark and Helfrich model current density can be written in the Arrhenius form to bring out the temperature dependence [107]:

$$J = \left(\frac{\mu N_\nu q V}{2d}\right) \exp\left[-\frac{E_t}{k_B T} \ln\left(\frac{q N_t d^2}{2\epsilon_0 \epsilon_r V}\right)\right] \tag{5.4}$$

where the slope at constant voltage is the activation energy $E_a$, expressed as:

$$E_a = E_t \ln\left(\frac{q N_t d^2}{2\epsilon_0 \epsilon_r V}\right) \tag{5.5}$$

Figure 5.9 shows the expected linear dependence of the activation energy on $\log(1/V)$. An examination of the temperature dependence in the EDT-SCLC regime shows that the current density decreases with decreasing temperature at low voltages, while at higher voltages, the trend is expected to be reversed. The log(J) - log(V) curves will converge

at a point corresponding to a critical voltage $V_C$, where the current in Eq. (5.4) is almost temperature independent, namely $E_a \approx 0$ (the variation in current at the crossover voltage for different values of $m$ is less than 1%). This voltage is actually the trap-filled limit voltage $V_{TFL}$. It can be of particular interest since it allows computation of the total trap density $N_t$ from the following expression [95, 97, 108**?** ]:

$$V_C = \frac{ed^2 N_t}{2\epsilon_0 \epsilon_r} \tag{5.6}$$

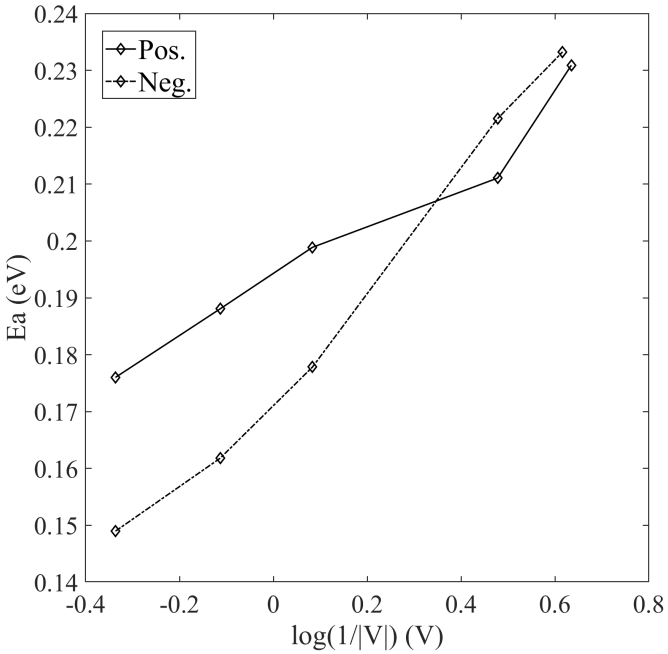



Figure 5.9: Dependence of the activation energy on $\log(|V|)$. The data show the linear dependence expected from Eq. 5.5

As shown in Figure 5.10, the extrapolation of the power-law fitted lines in the $\log_{10}(V) - \log_{10}(J)$ plot meet at a critical voltage. If all other parameters are known, an estimation of $N_t$ is straightforward to obtain. However, the value of $\varepsilon_r$ is not experimentally known. Assuming its value is constant with temperature, it is possible to estimate the ratio $N_t/\epsilon_r$. The results are $N_t/\epsilon_r \approx 4.6 \times 10^{19}\,\mathrm{cm}^{-3}$ for LRS data and $N_t/\epsilon_r \approx 1.2 \times 10^{20}\,\mathrm{cm}^{-3}$ for HRS data. It must be noted that we chose not to include the 328 K data in the positive LRS extrapolation as we observed some irreversible changes at high bias at this temperature.

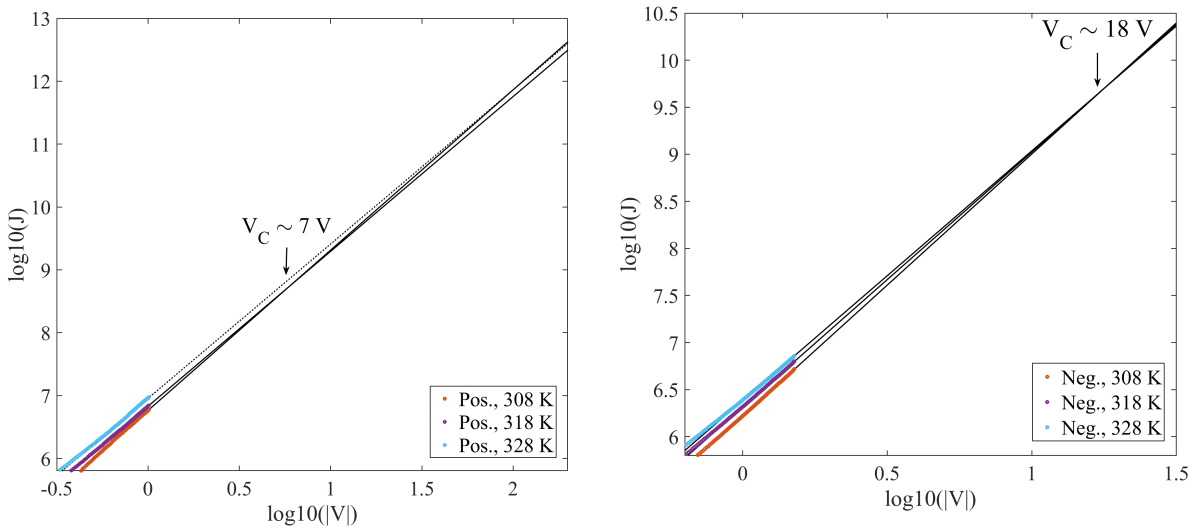



Figure 5.10: Extrapolation of $\log J - \log V$ curves in **(a)** LRS and **(b)** HRS. The fitting lines intersect at the critical voltage $V_C = 7\,\mathrm{V}$ in **(a)** and $V_C = 18\,\mathrm{V}$ in **(b)**.

## 5.3. Analysis in the high field region

In the high field region, the transport is clearly not thermally activated, indicating that this phenomenon is related to a kind of tunneling leakage mechanism which can be FNT or a form of TAT.

### 5.3.1. Analysis in the TAT regime

A simplified model of TAT shows linearity in the representation $\log(J)$ versus $-1/|V|$ in a reasonable range between 1.78 V and 2 V (Figure 5.11). A simplified expression for TAT current is [110]:

$$J = A \exp\left(-\frac{4d\sqrt{2qm^*}\Phi_t^{3/2}}{3\hbar V}\right) \tag{5.7}$$

which is equivalent to:

$$\log(J) = \log(A) - \frac{1}{V} \times \frac{4d\sqrt{2qm^*}\Phi_t^{3/2}}{3\hbar}. \tag{5.8}$$

$A$ is a constant, $d$ is the BFO thickness, and $m^*$ is the electron effective mass in the oxide. $\Phi_t$ is the level of the trap with respect to the valence band edge. The slope of the $\log(J)$ - $-1/|V|$ plots allows us to extract the value of $\Phi_t$: $\Phi_t \approx 1.2\,\mathrm{eV}$ for LRS and $\Phi_t \approx 1.5\,\mathrm{eV}$ for HRS.

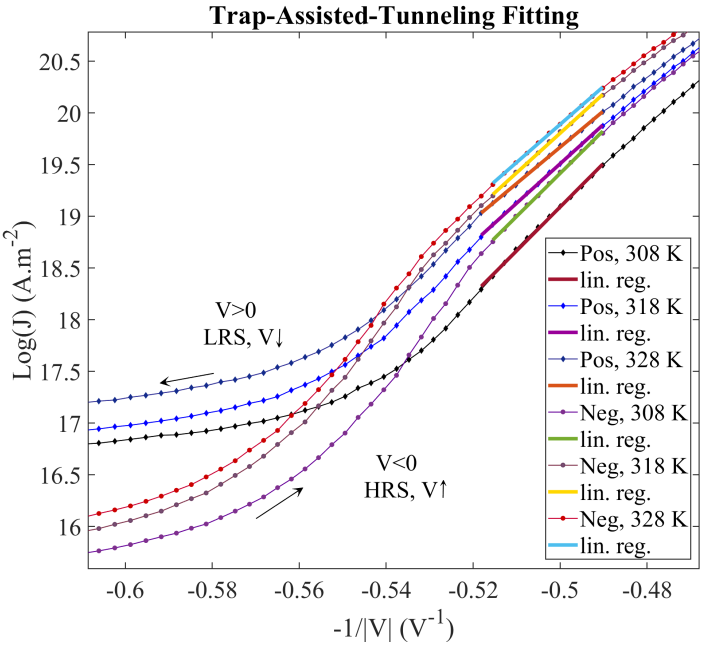



Figure 5.11: Current-voltage characteristics in the TAT representation. At each temperature, a linear regression is performed on the negative, increasing branch (HRS, nonswitching) and on the positive, decreasing branch (LRS, nonswitching).

### 5.3.2. Analysis in the FNT regime

In the highest field range ($2.8\,\mathrm{V} < \mathrm{V} < 3\,\mathrm{V}$), the $\log(J/V^2)$ versus $-1/|V|$ curves shown in Figure 5.12 exhibit linearity, implying that the conduction is governed by FNT.

A simplified expression for the current density in the FNT regime under an electric field E is given by [46] :

$$J = \frac{V^2 e^2}{d^2 8\pi h \Phi_B} \exp\left(-\frac{1}{V}\frac{d 8\pi\sqrt{2m^*}(\Phi_B e)^3}{3he}\right) \tag{5.9}$$

$\Phi_B$ is the potential barrier, and $m^*$ is the effective mass of the tunneling charge carriers. For completeness, some correction factors should be included, as suggested by the theory developed by [111]. Here, we considered the same approach of [112] where they introduce a free correction factor $C$ which allows to reproduce the experimental data quantitatively:

$$J = C\frac{V^2 e^2}{d^2 8\pi h \Phi_B} \exp\left(-\frac{1}{V}\frac{d 8\pi\sqrt{2m^*}(\Phi_B e)^3}{3he}\right) \tag{5.10}$$

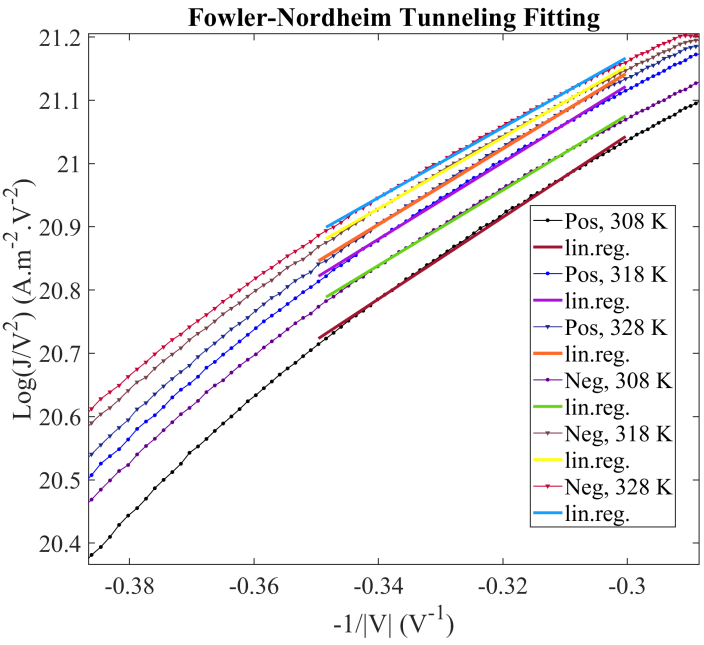



Figure 5.12: Current-voltage characteristics in the FNT representation. At each temperature, a linear regression is performed on the negative, increasing branch (HRS, nonswitching) and on the positive, decreasing branch (LRS, nonswitching).

The slope of the $\log(J/V^2)$ - $-1/|V|$ fit can be used to compute $\Phi_B$. We obtain values of 0.35 eV for LRS and 0.32 eV for HRS.

## 5.4. Equivalent circuit model

In Figure 5.13 we propose an equivalent circuit model to simulate the $J-V$ curves. Each conduction mechanism is modeled as a voltage-controlled current source since the electric field is assumed to be uniform across the insulator [114].

The current density is computed as:

$$J_{Ohmic} + J_{SCLC} + \left(\frac{1}{J_{TAT}} + \frac{1}{J_{FNT}}\right)^{-1} \tag{5.11}$$

where $J_{Ohmic}$ is computed with Eq. 5.1, $J_{FNT}$ with Eq. 5.10 and $J_{TAT}$ with Eq. 5.7. The EDT-SCLC contribution is [50]:

$$J_{SCLC} = q\mu N_c \left(\frac{\epsilon_0 \epsilon_r}{qN_t}\right)^l \frac{V^{l+1}}{d^{2l+1}} \tag{5.12}$$

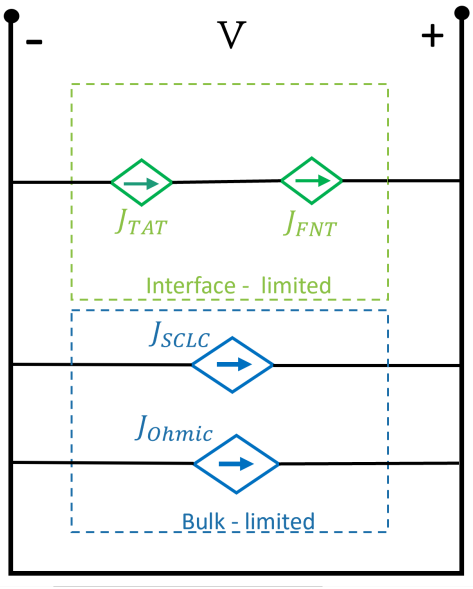



Figure 5.13: Equivalent circuit model describing the $J-V$ behavior of the dielectric. The bulk-limited and the interface-limited conduction paths contribute in parallel to the transport. The interface-limited current is modeled with two series-connected current sources ($J_{TAT}$ and $J_{FNT}$) to emphasize the dominance of the smallest contribution.

In this equivalent circuit model, all the separate conduction channels contribute in parallel to the current. However, the electrode-limited FNT and TAT mechanisms are only active in their respective electric field ranges, preventing them from contributing simultaneously to the current. So it is evident that the current in the high-field region is equal to the reciprocal sum of $J_{TAT}$ and $J_{FNT}$ , therefore the smallest J contribution will dominate [114]. Figure 5.14 depicts the experimental data and the analytical curves, obtained using the equivalent circuit model proposed above. In order to calculate the SCL current from the analytical expression proposed in Eq. 5.12, the value of $\mu N_v$ corresponding to the EDT-SCLC regime is computed from the fit of the experimental curves. The values are reported in table 5.3: as expected, a temperature-dependent value lower than the $\mu N_v$ for the Ohmic regime is obtained.

| Temperature (K) | $\mu N_c^+ \, (\mathrm{cmVs})^{-1}$ | $\mu N_c^- \, (\mathrm{cmVs})^{-1}$ |
|---|---|---|
| 308 | $0.9797 \times 10^{17}$ | $1.0709 \times 10^{18}$ |
| 318 | $1.0020 \times 10^{17}$ | $1.0453 \times 10^{18}$ |
| 328 | $1.1826 \times 10^{17}$ | $0.9745 \times 10^{18}$ |

Table 5.3: Values of $\mu N_c$ for the EDT-SCLC regime at different temperatures, for both LRS and HRS.

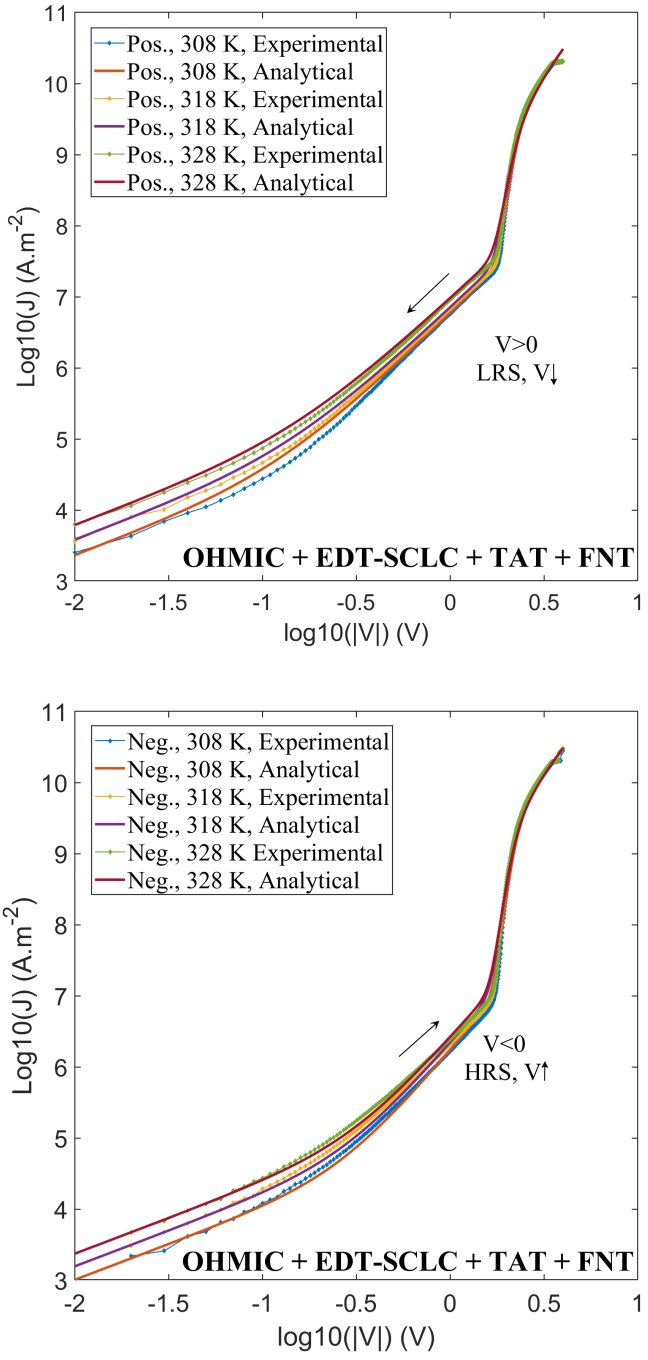



Figure 5.14: Reproduced experimental data using the equivalent circuit model proposed above.

From Figure 5.14 it is clear that the model successfully emulates the experimental data, thus it is reasonable to explain the conduction as the interplay between Ohmic, EDT-SCLC, TAT, and FNT.

## 5.5. Resistive switching mechanism

Resistive switching in BFO thin films has been observed to derive mainly from two mechanisms: the formation of a conductive filament or the modulation of the interface by ferroelectric polarization. Local conductive paths of oxygen vacancies or metallic particles can be reversibly formed inside the material, switching the device from the HRS to the LRS. The presence of grain boundaries facilitates the movement of defects, so this mechanism is usually observed in polycrystalline films or MFM structures. In the other case, a Schottky barrier at the metal/BFO interface can dominate the resistance due to polarization modulation of the barrier. It can also happen at the interface between BFO and a n-type semiconductive electrode, thus forming a p-n junction [42, 113]. In this kind of tunnel junctions, the screening of the polarization charges in the semiconductor creates an accumulation (resp. depletion) of electrons when the polarization points towards (resp. outwards) the semiconductor, leading to a LRS (rep. HRS) [91].

The results reported in the previous chapters are fundamental for the interpretation of the resistive switching mechanisms in the device and justify the moderate On/Off ratio. First, both Ohmic and SCLC are bulk-limited conduction regimes, contrary to the interface-limited one observed in nanocapacitors on blanket films [90]. Additional information can be recovered from the polarity of the switching: considering that CCMO is n-type (with a carrier density of typically around $1.55 \times 10^{21}\text{cm}^{-3}$) the stack consists of metal/ferroelectric/n-type semiconductor. The polarization pointing towards (resp. outwards) the semiconductor should lead to a Low Resistive State (LRS), (resp. to a High Resistive State (HRS)), and this is in agreement with Figure 4.12 for both the fresh and the aged devices with Ni electrodes. Instead, previous work with Co/BFO/CCMO system, showed that the HRS occurs when the polarization points toward the CCMO electrode. The inverted polarity was explained by the presence of an interfacial dielectric layer at the Co/BFO interface. The On-Off ratio, in this case caused by the TER, not only depends on tunneling barrier height modulation through incomplete screening but also on the electrostatic potential across the ferroelectric by different metal work functions, and tunneling barrier width modulation by enhanced depletion of one electrode or partial metallization of one of the interfaces to the ferroelectric [91].

In Chapter 4 we mentioned that Ni contacts oxidize at the surface of the BFO, forming a NiOx layer at the Ni/BFO interface. However, it shows an opposite polarity with respect to the Co/BFO/CCMO (plain film) system discussed in [90], even if both metal/ferroelectric interfaces are characterized by the formation of a metal oxide layer. This can be explained by taking into account the whole material system, namely the

stack Ni/BFO/CCMO/Ni, and in particular the Ni/CCMO interface. The interfacial oxide may form also at the Ni/CCMO interface, where oxygen is scavenged from the CCMO, increasing its resistance. In fact, CCMO becomes metallic when about 4% of the $Ca^{2+}$ ions are substituted by the smaller $Ce^{4+}$ ions. This yields a two-electron doping for the parent material ($CaMnO_3$), leading to mixed valent $Mn^{3+/4+}$ ions. Removing oxygen from the compound reduces Mn, which leads to an insulating layer. Such oxidation can be induced by the high processing temperatures (300°C) of the passivation deposition ($SiO_2$, PECVD), and occur over time at room temperature. This means that, because of oxygen scavenging by the Ni bottom electrode contact, the patterned CCMO layer has a smaller carrier density compared to the CCMO plain film. Therefore, the depletion length is large and the BFO/CCMO interface controls the FTJ polarity [91]. Other works reported metal-insulator phase transition in CCMO electrostatically induced by the BFO [115]. Here the material system is Pt/BFO/CCMO and a reversed polarity is also reported as Co/BFO/CCMO system. In this case, the phenomenon is mainly due to the induced insulating phase in the CCMO, creating wide space charge regions in both BFO and CCMO.

Taking in mind the above results, in the following discussion we propose the mechanisms governing resistive switching. Area-dependent measurements have excluded the presence of a conductive filament in the structure [91]. To gain insights about the On/Off ratio it is important to consider the low-field region, where the conduction is bulk-limited. This implies the presence of space charge regions at the BFO/CCMO and at the Ni/BFO (also described as Ni/NiOx/BFO) interfaces. Assuming no contribution from the polarization, a depletion region with a certain width forms across the p-n junction after reaching dynamic equilibrium. After a positive bias is applied, the polarization points toward the CCMO layer, so the positive bound charges aggregate at BFO/CCMO interface. Due to the ferroelectric field effect, the carriers in the n-type semiconductor are attracted toward the interface by the positive bound charges, resulting in a decrease in the depletion layer width. This results in a small resistance at the BFO/CCMO interface. On the other hand, the Ni/BFO interface is characterized by a NiOx interlayer preventing the direct tunneling of electrons across the BFO. The oxide is depleted from electrons, thus the Ni / BFO interface has a high resistance. In addition, according to the theory of polarization screening, the depolarization field developed inside the ferroelectric layer leads to an energy band bending at the interface which decreases the potential barrier when a positive pulse is applied. Therefore, when the polarization points downward, the carriers can easily pass through the depletion region having a narrow width and lower barrier height, but the former mechanisms is dominant (Figure 5.15 (a)). With the polarization reversal, upon

the application of a negative voltage, a complementary effect occurs so the resistance of the BFO/CCMO interface increases while that of the Ni/BFO interface decreases (Figure 5.15 (b)). The two effects cancel each other, so the currents measured in the two opposite polarities, namely in the LRS and HRS, do not show substantial differences, leading to the small On/Off ratio [91, 113].

In the high field region the situation is different, since we observed a tunneling-dominated behaviour. The I-V characteristics in the FNT regime are symmetric, and the tunneling barriers for carriers computed in section 5.3.2 are comparable in the two different resistance states. The formation of a NiOx interlayer at both interfaces, for the reasons explained above, yields a heterostructure of the type Ni/NiOx/BFO/CCMO/NiOx/Ni. Thus, carriers will see a symmetric potential barrier. The values computed in section 5.3.2 are in good agreement with the potential barrier at the Ni/NiOx interface. The work function value for Ni is $\simeq 5.15$ eV [116] while reported values for p-type NiOx are electron affinity $\chi \simeq -1.5$ eV and band gap of $\simeq 3.9$ eV [117, 118]. The potential barrier for carriers at the interface when the system reaches dynamic equilibrium is $\simeq 0.3$ eV.

The resistive switching in ferroelectrics is influenced by the deposition conditions and the microstructure of the films. A variety of techniques that can help in the improvement of the RS behaviour include doping of the ferroelectric layer or controlling the oxygen vacancy concentration, as well as optimizing the thickness of the ferroelectric and electrode layers or constructing an insertion layer between the ferroelectric material and the substrate. Doping the ferroelectric layer has been demonstrated to be very simple and effective, however appropriate element selection is needed. The oxygen vacancy concentration can be controlled by changing the stoichiometry of the film during the deposition process; a higher oxygen pressure leads to lower oxygen vacancy concentration in the film and better ferroelectricity. Defects can also induce RS effects related to trapping/detrapping or tunneling mediated by defects. They can also induce the creation of local conductive filament, which typically plays an auxiliary role to the interface-regulated mechanisms. These methods are devoted to the control of defect concentration, which is relatively easy to achieve, but they are limited by material type, preparation process, and related characterization methods. In contrast, regulating the interface is more feasible: a key factor is the thickness of the ferroelectric layer which affects the microstructure and the strain state, as well as the electrode thickness.
In the presented case study a more appropriate material for the electrodes must be chosen to recover a better performance in terms of On/Off ratio. As reported in [91], a new process flow has been developed, where W and Pt have been chosen as top and bottom electrodes, replacing the Ni. In this structure, the W/BFO interface is dominant over

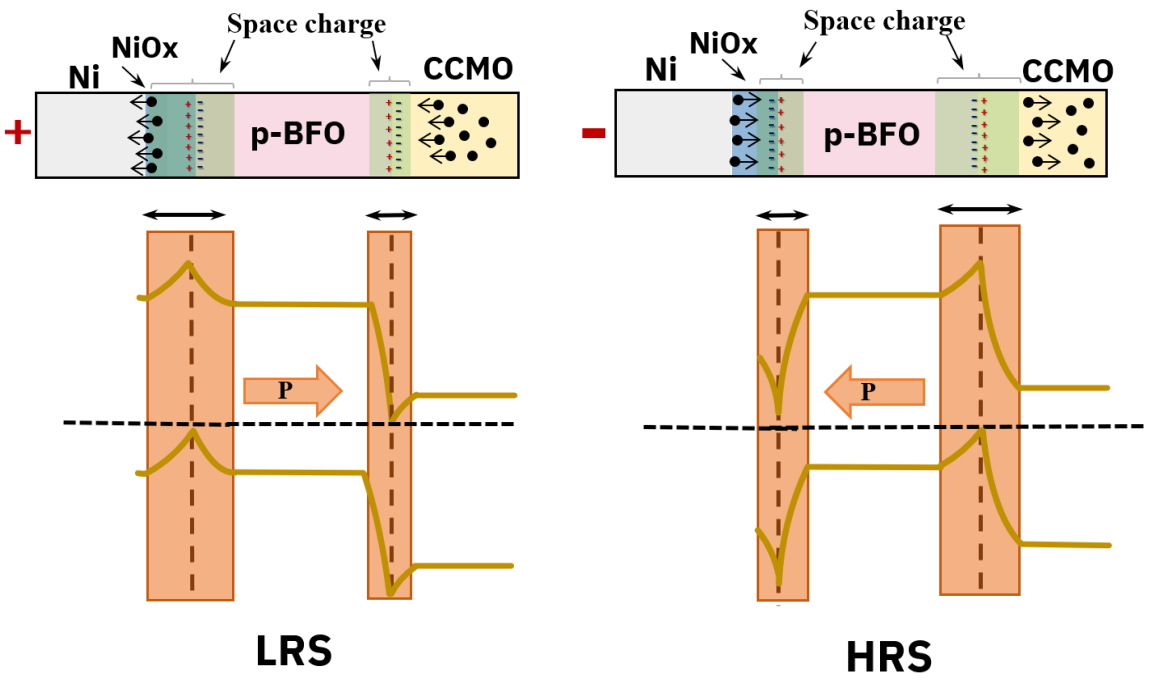



Figure 5.15: Resistive switching mechanisms with polarization reversal. Charge distributions and energy-band diagrams of Ni/NiOx/BFO/CCMO heterostructure for the **(a)** LRS state and **(b)** HRS state. The modulation of the barrier height and width by the polarization reversal at the BFO/CCMO interface determines the resistive switching. V. For BFO, the band gap is 2.8 eV and the electron affinity is 3.3 eV [119]. Black dots represent electrons, while positive red signs and blue minus signs are the positive and negative bound charges, respectively.

the BFO/CCMO interface, and the Pt contact to the CCMO does not scavenge anymore oxygen from the latter. The measured On/Off ratio is two orders of magnitude larger than the one measured in the fully processed BFO with Ni electrodes. The design of the M1 and M2 metal lines has been further optimized for the interconnection of a 784 $\times$ 100 crossbar array to a neuromorphic circuit using an advanced assembly technique, the controlled collapse chip connection or flip-chip bonding.

# 6 | Conclusions and future developements

In this thesis, the fabrication of ferroelectric crossbar arrays based on FTJ devices has been presented for two technologies differing because of the ferroelectric material used, namely HZO and BFO. The fluorite HZO is the most widely used material for the realization of ferroelectric memristors due to the well-developed CMOS-friendly fabrication process which allows co-integration. Its ferroelectric behaviour makes it more suited for non-volatile memory and inference DNNs applications. In contrast, the perovskite BFO has shown incredible performances in terms of advanced synaptic functionality such as STDP, leading to interesting applications such as online learning in DNNs and unsupervised learning in SNNs. Furthermore, the cumulative switching feature is convenient from an electronic point of view. Even if is it more performant than HZO, its deposition requires extra precaution, in addition to the high processing temperatures required during epitaxial growth of BFO and CCMO, which are not CMOS-compatible. Thus, there is no clear advantage in using one technology over the other.

The use of Ni contacts in BFO crossbars, motivated by the large On/Off ratio reported on Ni/BFO/CCMO/YAO capacitors, led to a degradation of the latter in the fully processed devices. The origin of the degradation was investigated, in particular using temperature-dependent measurements. The oxidation of the Ni contacts at the BFO and the CCMO interfaces is also responsible for a change in the device polarity. It blocks tunneling currents through BFO at a small bias and changes the conduction from interface (tunneling) to bulk (Ohmic, SCLC) limited. The use of a different metal electrode system has demonstrated improved crossbar operation in a larger crossbar array. As a next step, a demonstrator for the BEOL-integration of the crossbar in a neuromorphic circuit can be realized via flip-chip bonding.

# Bibliography

[1] Zidan, M.A., Strachan, J.P. & Lu, W.D. The future of electronics based on memristive systems. Nat Electron 1, 22–29 (2018). https://doi.org/10.1038/s41928-017-0006-8.

[2] Seung Ju Kim, Sangbum Kim, Ho Won Jang, Competing memristors for brain-inspired computing, iScience, Volume 24, Issue 1, 2021, 101889, ISSN 2589-0042, https://doi.org/10.1016/j.isci.2020.101889.

[3] Sun, K., Chen, J., Yan, X., The Future of Memristors: Materials Engineering and Neural Networks. Adv. Funct. Mater. 2021, 31, 2006773. https://doi.org/10.1002/adfm.202006773.

[4] Xia, Q., Yang, J.J. Memristive crossbar arrays for brain-inspired computing. Nat. Mater. 18, 309–323 (2019). https://doi.org/10.1038/s41563-019-0291-x.

[5] J. von Neumann. "First draft of a report on the EDVAC". In: IEEE Annals of the History of Computing 15.4 (1993), pp. 27–75. doi:10.1109/85.238389.

[6] Zou, X., Xu, S., Chen, X. et al. Breaking the von Neumann bottleneck: architecture-level processing-in-memory technology. Sci. China Inf. Sci. 64, 160404 (2021). https://doi.org/10.1007/s11432-020-3227-1.

[7] Jeong, D. S., Kim, K. M., Kim, S., Choi, B. J., & Hwang, C. S. (2016). Memristors for Energy-Efficient New Computing Paradigms. Advanced Electronic Materials, 2(9), 1600090. https://doi.org/10.1002/aelm.201600090.

[8] Beilliard, Yann & Alibart, Fabien. (2021). Multi-Terminal Memristive Devices Enabling Tunable Synaptic Plasticity in Neuromorphic Hardware: A Mini-Review. Frontiers in Nanotechnology. 3. 779070. doi:10.3389/fnano.2021.779070.

[9] G. Indiveri and S. -C. Liu, "Memory and Information Processing in Neuromorphic Systems," in Proceedings of the IEEE, vol. 103, no. 8, pp. 1379-1397, Aug. 2015, doi: 10.1109/JPROC.2015.2444094.

[10] S. Moore: “Multicore is bad news for supercomputers”, IEEE Spectrum 45, 15 (2008), doi: 10.1109/MSPEC.2008.4659375.

[11] T. N. Theis and H. . -S. P. Wong, "The End of Moore's Law: A New Beginning for Information Technology," in Computing in Science & Engineering, vol. 19, no. 2, pp. 41-50, Mar.-Apr. 2017, doi: 10.1109/MCSE.2017.29.

[12] S. Yu, "Neuro-inspired computing with emerging nonvolatile memorys," in Proceedings of the IEEE, vol. 106, no. 2, pp. 260-285, Feb. 2018, doi: 10.1109/JPROC.2018.2790840.

[13] Yang, R. In-memory computing with ferroelectrics. Nat Electron 3, 237–238 (2020). https://doi.org/10.1038/s41928-020-0411-2.

[14] Amirsoleimani, A., Alibart, F., Yon, V., Xu, J., Pazhouhandeh, M.R., Ecoffey, S., Beilliard, Y., Genov, R. and Drouin, D. (2020), In-Memory Vector-Matrix Multiplication in Monolithic Complementary Metal–Oxide–Semiconductor-Memristor Integrated Circuits: Design Choices, Challenges, and Perspectives. Adv. Intell. Syst., 2: 2000115. https://doi.org/10.1002/aisy.202000115.

[15] W. Maass, "Noise as a Resource for Computation and Learning in Networks of Spiking Neurons," in Proceedings of the IEEE, vol. 102, no. 5, pp. 860-880, May 2014, doi: 10.1109/JPROC.2014.2310593.

[16] E. Chicca, F. Stefanini, C. Bartolozzi and G. Indiveri, "Neuromorphic Electronic Circuits for Building Autonomous Cognitive Systems," in Proceedings of the IEEE, vol. 102, no. 9, pp. 1367-1388, Sept. 2014, doi: 10.1109/JPROC.2014.2313954.

[17] Neumann, J. von. "Probabilistic Logics and the Synthesis of Reliable Organisms From Unreliable Components". Automata Studies. (AM-34), Volume 34, edited by C. E. Shannon and J. McCarthy, Princeton: Princeton University Press, 1956, pp. 43-98. https://doi.org/10.1515/9781400882618-003.

[18] Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. Psychol. Rev. 65, 386–408.

[19] Fukushima, K., Yamaguchi, Y., Yasuda, M., and Nagata, S. (1970). An electronic model of the retina. Proc. IEEE 58, 1950–1951.

[20] Giacomo Indiveri et al 2013 Nanotechnology 24 384010 DOI 10.1088/0957-4484/24/38/384010.

[21] Indiveri G and Horiuchi TK (2011) Frontiers in neuromorphic engineering. Front. Neurosci. 5:118. doi: 10.3389/fnins.2011.00118-

[22] Grekousis, George. (2018). Artificial neural networks and deep learning in urban geography: A systematic review and meta-analysis. Computers, Environment and Urban Systems. 74. 10.1016/j.compenvurbsys.2018.10.008.

[23] Livingstone, David. (2009). Artificial Neural Networks: Methods and Applications. 10.1007/978-1-60327-101-1.

[24] Abraham, Ajith. "129 Artificial Neural Networks." (2005).

[25] Sören Boyn, Julie Grollier, Gwendal Lecerf, Bin Xu, Nicolas Locatelli, et al.. Learning through ferroelectric domain dynamics in solid-state synapses. Nature Communications, 2017, 8, pp.14736. ⟨10.1038/ncomms14736⟩. ⟨hal-02288726⟩

[26] Krogh, A. What are artificial neural networks?. Nat Biotechnol 26, 195–197 (2008). https://doi.org/10.1038/nbt1386.

[27] Alfayoumi, Bayan & Alshraideh, Mohammad & Leiner, Martin & Aldajani, Iyad. (2021). Machine Learning Predictions for the Advancement of the Online Education in the Higher Education Institutions in Jordan. Hunan Daxue Xuebao/Journal of Hunan University Natural Sciences. 48.

[28] Mu Y, Poo MM. Spike timing-dependent LTP/LTD mediates visual experience-dependent plasticity in a developing retinotectal system. Neuron. 2006;50:115–125.

[29] David, M., "Early Benchmarking Results for Neuromorphic Computing: What It Reveals About Future Adoption.", Intel Corporation.

[30] Zhong Sun et al. ,One-step regression and classification with cross-point resistive memory arrays.Sci. Adv.6,eaay2378(2020).DOI:10.1126/sciadv.aay2378.

[31] Wang, Lei & Yang, Cihui & Wen, Jing & Gai, Shan & Peng, YuanXiu. (2015). Overview of emerging memristor families from resistive memristor to spintronic memristor. Journal of Materials Science: Materials in Electronics. 26.10.1007/s10854-015-2848-z.

[32] Andy Thomas 2013 J. Phys. D: Appl. Phys. 46 093001

[33] Shahrokhi, S., Gao, W., Wang, Y., Anandan, P. R., Rahaman, Md Z., Singh, S., Wang, D., Cazorla, C., Yuan, G., Liu, J.-M., Wu, T., Emergence of Ferroelectricity in Halide Perovskites. Small Methods 2020, 4, 2000149. https://doi.org/10.1002/smtd.202000149.

[34] P.B. Littlewood, 2002. "Physics of ferroelectrics"

[35] R. Whatmore: “Ferroelectric Materials”, in: Springer Handbook of Electronic and Photonic Materials, ed. by S. Kasap & P. Capper, Boston, MA: Springer US, 2006, 597.

[36] Yoo et al. , Electronics 2023, 12(10), 2297; https://doi.org/10.3390/electronics12102297.

[37] Song, Chong-Myeong, and Hyuk-Jun Kwon. 2021. "Ferroelectrics Based on HfO2 Film" Electronics 10, no. 22: 2759. https://doi.org/10.3390/electronics10222759.

[38] Huimin Qiao, Chenxi Wang, Woo Seok Choi, Min Hyuk Park, Yunseok Kim, Ultra-thin ferroelectrics, Materials Science and Engineering: R: Reports, Volume 145, 2021, 100622, ISSN 0927-796X, https://doi.org/10.1016/j.mser.2021.100622.

[39] H. Mulaosmanovic, P. D. Lomenzo, U. Schroeder, S. Slesazeck, T. Mikolajick and B. Max, "Reliability aspects of ferroelectric hafnium oxide for application in non-volatile memories," 2021 IEEE International Reliability Physics Symposium (IRPS), Monterey, CA, USA, 2021, pp. 1-6, doi: 10.1109/IRPS46558.2021.9405215.

[40] M. Ye. Zhuravlev, R. F. Sabirianov, S. S. Jaswal, and E. Y. Tsymbal Phys. Rev. Lett. 94, 246802 – Published 20 June 2005; Erratum Phys. Rev. Lett. 102, 169901 (2009)

[41] E. Covi et al., "Ferroelectric Tunneling Junctions for Edge Computing," 2021 IEEE International Symposium on Circuits and Systems (ISCAS), Daegu, Korea, 2021, pp. 1-5, doi: 10.1109/ISCAS51556.2021.9401800.

[42] Wang, Z. J., Bai, Y., Small 2019, 15, 1805088.https://doi.org/10.1002/smll.201805088.

[43] Mattia Halter: Ferroelectric Memristors for Neuromorphic Applications: Design, Fabrication, and Integration, © 2022

[44] Li, Z., Wang, T., Yu, J., Meng, J., Liu, Y., Zhu, H., Sun, Q., Zhang, D. W., Chen, L., Ferroelectric Hafnium Oxide Films for In-Memory Computing Applications. Adv. Electron. Mater. 2022, 8, 2200951. https://doi.org/10.1002/aelm.202200951.

[45] M. Saitoh et al., "HfO2-based FeFET and FTJ for Ferroelectric-Memory Centric 3D LSI towards Low-Power and High-Density Storage and AI Applications," 2020 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2020, pp. 18.1.1-18.1.4, doi: 10.1109/IEDM13553.2020.9372106.

[46] Fu-Chien Chiu, "A Review on Conduction Mechanisms in Dielectric Films", Advances in Materials Science and Engineering, vol. 2014, Article ID 578168, 18 pages, 2014. https://doi.org/10.1155/2014/578168.

[47] Funck, Carsten & Menzel, Stephan. (2021). Comprehensive Model of Electron Conduction in Oxide-Based Memristive Devices. ACS Applied Electronic Materials. XXXX. 10.1021/acsaelm.1c00398.

[48] T. P. Xiao, X. Zhao, S. Agarwal and E. Yablonovitch, "Impact of interface defects on tunneling FET turn-on steepness," 2015 Fourth Berkeley Symposium on Energy Efficient Electronic Systems (E3S), Berkeley, CA, USA, 2015, pp. 1-2, doi: 10.1109/E3S.2015.7336796.

[49] Mau Phon Houng, Yeong Her Wang, Wai Jyh Chang; Current transport mechanism in trapped oxides: A generalized trap-assisted tunneling model. J. Appl. Phys. 1 August 1999; 86 (3): 1488–1491. https://doi.org/10.1063/1.370918.

[50] Journal of Applied Physics 33, 205 (1962); https://doi.org/10.1063/1.1728487.

[51] Kim, JK., Cho, K., Kim, TY. et al. Trap-mediated electronic transport properties of gate-tunable pentacene/MoS2 p-n heterojunction diodes. Sci Rep 6, 36775 (2016). https://doi.org/10.1038/srep36775.

[52] A. Rose, (1955). Space-Charge-Limited Currents in Solids. Physical Review, 97, 1538-1544

[53] Hoang T.N., Photolithography, Academia, https://www.academia.edu/35007088/Photolithography.

[54] Inflibnet centre, "Photolithography",http://epgp.inflibnet.ac.in/epgpdata/uploads/epgp_content/S000831ME/P001852/M030127/ET/1525847027Module-7_Unit2_NSNT.pdf.

[55] M. Van Rossum, Integrated Circuits, Editor(s): Franco Bassani, Gerald L. Liedl, Peter Wyder, Encyclopedia of Condensed Matter Physics, Elsevier, 2005, https://doi.org/10.1016/B0-12-369401-9/00503-9.

[56] Mack C.A., Semiconductor Lithography (Photolithography) - The Basic Process, http://www.lithoguru.com/scientist/lithobasics.html.

[57] Wikipedia contributors, "Photolithography," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Photolithography&oldid=1175788937.

[58] Omer Nur, Magnus Willander, Chapter 3 - Conventional nanofabrication methods, William Andrew Publishing, 2020.

[59] D.M. Tennant, A.R. Bleier, 4.02 - Electron Beam Lithography of Nanostructures, Academic Press, 2011.

[60] Heidelberg Instruments, Maskless Laser Lithography- Direct Writing On The Microscale, https://heidelberg-instruments.com/core-technologies/maskless-laser-lithography/.

[61] Richard W. Johnson, Adam Hultqvist, Stacey F. Bent, A brief review of atomic layer deposition: from fundamentals to applications, Materials Today, Volume 17, Issue 5, 2014.

[62] The plasma enhanced atomic layer deposition database, Introduction to plasma enhanced atomic layer deposition, https://www.plasma-ald.com/.

[63] Matt Hughes, What is DC sputtering, Semicore, https://www.semicore.com/news/94-what-is-dc-sputtering.

[64] Stoner J, What is DC sputtering? An overview, Korvus Technology, 2022.

[65] V. Teixeira, J. Carneiro, P. Carvalho, E. Silva, S. Azevedo, C. Batista, 11 - High barrier plastics using nanoscale inorganic films, Editor(s): José-María Lagarón, Multifunctional and Nanoreinforced Polymers for Food Packaging, Woodhead Publishing, 2011.

[66] Fuzhen Wang, Junwei Wu, Chapter 10 - Plasma-enhanced chemical vapor deposition, Editor(s): Fuzhen Wang, Junwei Wu, Modern Ion Plating Technology, Elsevier, 2023.

[67] Kohli V., Reactive Ion Etching: A Comprehensive Guide, Wevolver, 2023, https://www.wevolver.com/article/reactive-ion-etching-a-comprehensive-guide.

[68] Ayodele A., Ion Milling: A Comprehensive Guide to Material Etching Techniques, Wevolver, 2023, https://www.wevolver.com/article/ion-milling-a-comprehensive-guide-to-material-etching-techniques.

[69] R.E. Lee, Chapter 11 - Ion-Beam Etching (Milling), Elsevier, Volume 8, 1984.

[70] Wikipedia contributors, "Rapid thermal processing," Wikipedia, The Free Encyclopedia,https://en.wikipedia.org/w/index.php?title=Rapid$_$thermal$_$processing&oldid=1117571044.

[71] Ch. Hollauer: Modeling of Thermal Oxidation and Stress Effects, TU Wien, https://www.iue.tuwien.ac.at/phd/hollauer/node13.html.

[72] Rebohle, L., Prucnal, S., Berencén, Y. et al. A snapshot review on flash lamp annealing of semiconductor materials. MRS Advances 7, 1301–1309 (2022). https://doi.org/10.1557/s43580-022-00425-w.

[73] R.A. McMahon, M.P. Smith, K.A. Seffen, M. Voelskow, W. Anwand, W. Skorupa, Flash-lamp annealing of semiconductor materials—Applications and process models, 2007.

[74] FLA-flash lamp annealing, Blitzlab, https://www.blitzlab.de/en/fla-flash-lamp-annealing/.

[75] Appl. Phys. Lett. 109, 242901 (2016); https://doi.org/10.1063/1.4971996.

[76] Cheema SS, Kwon D, Shanker N, Dos Reis R, Hsu SL, Xiao J, Zhang H, Wagner R, Datar A, McCarter MR, Serrao CR, Yadav AK, Karbasian G, Hsu CH, Tan AJ, Wang LC, Thakare V, Zhang X, Mehta A, Karapetrova E, Chopdekar RV, Shafer P, Arenholz E, Hu C, Proksch R, Ramesh R, Ciston J, Salahuddin S. Publisher Correction: Enhanced ferroelectricity in ultrathin films grown directly on silicon. Nature. 2020 May;581(7808):E5. doi: 10.1038/s41586-020-2297-6. Erratum for: Nature. 2020 Apr;580(7804):478-482. PMID: 32433606.

[77] Mohammed Bilal Hachemi. Contribution to the study of structural and ferroelectric properties of HZO thin films. Micro and nanotechnologies/Microelectronics. Université Grenoble Alpes [2020-..], 2022. English. ffNNT : 2022GRALT071ff. fftel-03946210f.

[78] Sang, Xiahan & Grimley, Everett & Schenk, Tony & Schroeder, Uwe & LeBeau, James. (2015). On the Structural Origin of Ferroelectricity in HfO2 Thin Films. Applied Physics Letters. 106. 162905. 10.1063/1.4919135.

[79] Kim, S.J., Mohan, J., Summerfelt, S.R. et al. Ferroelectric Hf0.5Zr0.5O2 Thin Films: A Review of Recent Advances. JOM 71, 246–255 (2019). https://doi.org/10.1007/s11837-018-3140-5.

[80] Wei, Y., Nukala, P., Salverda, M. et al. A rhombohedral ferroelectric phase in epitaxially strained Hf0.5Zr0.5O2 thin films. Nature Mater 17, 1095–1100 (2018). https://doi.org/10.1038/s41563-018-0196-0.

[81] Laura Bégon-Lours et al 2022 Neuromorph. Comput. Eng. 2 024001

[82] Bégon-Lours, L., Halter, M., Puglisi, F. M., Benatti, L., Falcone, D. F., Popoff, Y., Dávila, D., Sousa, M., Offrein, B. J., Scaled, Ferroelectric Memristive Synapse for

Back-End-of-Line Integration with Neuromorphic Hardware. Adv. Electron. Mater. 2022, 8, 2101395. https://doi.org/10.1002/aelm.202101395.

[83] L. Bégon-Lours et al., "Analog Resistive Switching in BEOL, Ferroelectric Synaptic Weights," in IEEE Journal of the Electron Devices Society, vol. 9, pp. 1275-1281, 2021, doi: 10.1109/JEDS.2021.3108523.

[84] D Sando et al 2014 J. Phys.: Condens. Matter 26 473201

[85] Kwi Young Yun, Minoru Noda, Masanori Okuyama, Hiromasa Saeki, Hitoshi Tabata, Keisuke Saito; Structural and multiferroic properties of BiFeO3 thin films at room temperature. J. Appl. Phys. 15 September 2004; 96 (6): 3399–3403. https://doi.org/10.1063/1.1775045.

[86] J. Wang et al. Multiferroic Thin Film Heterostructures. Science 299, 1719 (2003); DOI: 10.1126/science.1080615.

[87] Appl. Phys. Rev. 6, 041403 (2019); https://doi.org/10.1063/1.5118737.

[88] Yamada, H., Marinova, M., Altuntas, P. et al. Ferroelectric control of a Mott insulator. Sci Rep 3, 2834 (2013). https://doi.org/10.1038/srep02834.

[89] Appl. Phys. Lett. 109, 232902 (2016); https://doi.org/10.1063/1.4971311.

[90] H. Yamada et al., "Giant Electroresistance of Super-tetragonal BiFeO 3 -Based Ferroelectric Tunnel Junctions," ACS Nano, vol. 7, no. 6, pp. 5385–5390, Jun. 2013, doi: 10.1021/nn401378t.

[91] Halter, M., Morabito, E., Olziersky, A. et al. Crossbar operation of BiFeO3/Ce–CaMnO3 ferroelectric tunnel junctions: From materials to integration. Journal of Materials Research (2023). https://doi.org/10.1557/s43578-023-01158-8.

[92] S. Farokhipoor, B. Noheda; Local conductivity and the role of vacancies around twin walls of (001)BiFeO3 thin films. J. Appl. Phys. 1 September 2012; 112 (5): 052003. https://doi.org/10.1063/1.4746073

[93] Appl. Phys. Lett. 90, 072902 (2007) https://doi.org/10.1063/1.2535663.

[94] M. A. Lampert, Phys. Rev. 103, 1648 (1956)

[95] Arora, A., Mourya, S., Singh, N., Kumar, S., Chandra, R., Malik, V. K. (2023). Coexistence of Space Charge Limited and Variable Range Hopping Conduction Mechanism in Sputter- Deposited Au/SiC Metal-Semiconductor-Metal Device. IEEE Transac-

tions on Electron Devices, 70(2), 714-719. https://doi.org/10.1109/TED.2022.3232472.

[96] Appl. Phys. Lett. 97, 093105 (2010); https://doi.org/10.1063/1.3484956.

[97] Journal of Applied Physics 94, 1283 (2003); https://doi.org/10.1063/1.1582552.

[98] R. D. Gould; The interpretation of space-charge-limited currents in semiconductors and insulators. Journal of Applied Physics 1 April 1982; 53 (4): 3353–3355. https://doi.org/10.1063/1.331003.

[99] J G Simmons 1971 J. Phys. D: Appl. Phys. 4 613

[100] H. Khassaf, G. A. Ibanescu, I. Pintilie, I. B. Misirlioglu, L. Pintilie; Potential barrier increase due to Gd doping of BiFeO3 layers in Nb:SrTiO3-BiFeO3-Pt structures displaying diode-like behavior. Appl. Phys. Lett. 18 June 2012; 100 (25): 252903. https://doi.org/10.1063/1.4729816.

[101] Tsurumaki, A., Yamada, H. and Sawa, A. (2012), Impact of Bi Deficiencies on Ferroelectric Resistive Switching Characteristics Observed at p-Type Schottky-Like Pt/Bi1–$\delta$ FeO3 Interfaces. Adv. Funct. Mater., 22: 1040-1047. https://doi.org/10.1002/adfm.201102883.

[102] Yao Shuai et al 2011 Appl. Phys. Express 4 095802

[103] Journal of Applied Physics 99, 054104 (2006) https://doi.org/10.1063/1.2177430.

[104] Appl. Phys. 105, 074103 (2009); https://doi.org/10.1063/1..

[105] Journal of Applied Physics 95, 3120 (2004); https://doi.org/10.1063/1.1646441.

[106] Appl. Phys. Lett. 34, 226 (1979); https://doi.org/10.1063/1.90739.

[107] Journal of Applied Physics 94, 1283 (2003); https://doi.org/10.1063/1.1582552.

[108] Appl. Phys. 103, 122103 (2013); https://doi.org/10.1063/1.4821185.

[109] Appl. Phys. Lett. 103, 122103 (2013); https://doi.org/10.1063/1.4821185.

[110] X. Li et al., "Investigation on Carrier Transport Through AlN Nucleation Layer From Differently Doped Si(111) Substrates," in IEEE Transactions on Electron Devices, vol. 65, no. 5, pp. 1721-1727, May 2018, doi: 10.1109/TED.2018.2810886.

[111] R. G. Forbes: "Physics of generalized Fowler-Nordheim-type equations", J. Vac. Sci. Technol., B: Microelectron. Nanometer Struct.–Process., Meas., Phenom. 26, 788 (2008).

[112] Sören Boyn. Ferroelectric tunnel junctions : memristors for neuromorphic computing. Materials Science [cond-mat.mtrl-sci]. Université Paris-Saclay, 2016. English. <NNT : 2016SACLS089>. <tel-01382194.

[113] Zhongqiang Hu, Qian Li, Meiya Li, Qiangwen Wang, Yongdan Zhu, Xiaolian Liu, Xingzhong Zhao, Yun Liu, Shuxiang Dong; Ferroelectric memristor based on Pt/BiFeO3/Nb-doped SrTiO3 heterostructure. Appl. Phys. Lett. 11 March 2013; 102 (10): 102901. https://doi.org/10.1063/1.4795145.

[114] Chiang, Tsung-Han Wager, John. (2017). Electronic Conduction Mechanisms in Insulators. IEEE Transactions on Electron Devices. PP. 1-8. 10.1109/TED.2017.2776612.

[115] Wenhao Yu, Luqiu Chen, Yifei Liu, Bobo Tian, Qiuxiang Zhu, Chungang Duan; Resistive switching polarity reversal due to ferroelectrically induced phase transition at BiFeO3/Ca0.96Ce0.04MnO3 heterostructures. Appl. Phys. Lett. 9 January 2023; 122 (2): 022902. https://doi.org/10.1063/5.0132819.

[116] Ofuonye, Benedict & Lee, Jaesun & Yan, Minjun & Sun, Changwoo & Zuo, Jian-Min & Adesida, Ilesanmi. (2014). Electrical and microstructural properties of thermally annealed Ni/Au and Ni/Pt/Au Schottky contacts on AlGaN/GaN heterostructures. Semiconductor Science and Technology. 29. 095005. 10.1088/0268-1242/29/9/095005.

[117] Rajagopal Reddy, Varra & Reddy, P.R & REDDY, NEELAKANTA & Choi, ChelJong. (2016). Microstructural, electrical and carrier transport properties of Au/NiO/n-GaN heterojunction with a nickel oxide interlayer. RSC Adv.. 6. 10.1039/C6RA23476C.

[118] Vadim P Sirkeli et al 2015 Semicond. Sci. Technol. 30 065005

[119] Appl. Phys. Lett. 98, 192901 (2011) https://doi.org/10.1063/1.3589814

# List of Figures

2.1 Illustration of the brain memory hierarchy with co-localized memory and computation **(a)** vs. **(b)** the classical computing architecture: Central Processing Units (CPUs) containing multiple cores, having each a microprocessor and local memory. This block is connected to the main memory block, the primary storage area, but requires longer access times than the memory blocks in the core. Worst is for peripheral memory block but with higher storage capabilities. Adapted from [9]. . . . . 6

2.2 Transmission of electrochemical signals through neurotransmitters in a biological synapse. Adapted from [20]. . . . . 7

2.3 Graphical representation of the artificial neuron model. The neuron receives inputs $x_i$ from N other units or external sources, with associated weights $w_i$. The total input to a unit is the weighted sum over all inputs, $\sum_{i=1}^{N} w_i x_i = w_1 x_1 + w_2 x_2 + ... + w_N x_N$. If the total input is below a threshold t, the output of the unit would be 1 and 0 otherwise. The output is determined by the model $y = g\left(\sum_{i=0}^{n} w_i x_i - t\right)$ where the transfer function g is the step function. It can also be, for example, a continuous sigmoid as shown by the red curve. Adapted from [26]. . . . . 8

2.4 Schematic diagram of a feed-forward neural network with a number of hidden layers between input and output neuron layers. Adapted from [27]. 9

2.5 Representation of a generic current-voltage characteristic of a memristor, showing pinched hysteresis loop. Adapted from [31]. . . . . 12

2.6 **(a)** Basic NN structure and related VMM operation, where the inputs $x_i$ are organized in a vector, the weights $W_i$ in a matrix, and the output of the operation is also represented as a vector. **(b)** Schematic of the digital VMM, obtained by pipelining digital blocks, versus the analog VMM, obtained by summing the currents obtained in each of the N columns. **(c)** Crossbar array where a memristor is formed at each crosspoint and can be used to simultaneously store data and process information ( $V_i$, voltage applied at each row i; $I_j$, current trough column j; $G_{i,j}$, conductance of the memristor at the intersection of row i and column j. **(d)** 3D illustration of how a memristor-based crossbar can be monolithically integrated in the Back-End-Of-Line (BEOL) of a CMOS. Adapted from [14] . . . . . 14

2.7 **(a)** The blue curve depicts the P-E ferroelectric hysteresis loop. The insets with the arrows represent the domain evolution during polarization switching. **(b)** Schematic representation of paraelectric-ferroelectric phase transition upon cooling below $T_c$, in both displacive type and order-disorder type materials. Adapted from [33]. . . . . 16

2.8 ferroelectric NVM devices: FeRAM **(a)** MOSFET plus ferroelectric capacitor 1T-1C structure; FeFET **(b)**, MOSFET structure with ferroelectric gate oxide; FTJ **(c)** parallel plate capacitor with ferroelectric layer. (BL=Bitline; WL=Wordline; PL=Plateline; DE=Dieletric; FE=Ferroelectric). Adapted from [37]. . . . . 18

2.9 Generic MFM structure of a FTJ device. The energy band profile is asymmetric at the two metal/ferroelectric interfaces due to the different screening lengths ($x_1$, $x_2$). Red arrows show the ferroelectric polarization modulating the barrier height. Adapted from [36]. . . . . 19

2.10 Schematic representation of bound charges developed at ferroelectric/metal interfaces in a ferroelectric tunnel junction due to polarization, and the consequent change and depolarization field distributions versus distance. Adapted from [38]. . . . . 20

3.1 Summary of the conduction mechanisms in dielectrics that have been analyzed in this work. Adapted from [46] . . . . . 24

3.2 Schematic of Schottky emission in MIS structure. Adapted from [46] . . . 25

3.3 Schematic of Fowler-Nordheim tunneling process in MIS structure. Adapted from [46] . . . . . 27

3.4 Schematic of Trap-Assisted tunneling process in MIM structure. The induced field induces a tilted rectangular band structure. The trap level $\phi_t$ lies in the oxide band gap below the conduction band. The phenomenon occurs in three steps:(i) tunneling into the trap from one electrode, (ii) tunneling from trap to trap, and (iii) tunneling from the trap into the other electrode. Adapted from [48] . . . . . 28
3.5 Schematic of the energy band diagram of thermionic-field emission in MIS structure, compared to thermionic emission and field emission. Adapted from [46] . . . . . 29
3.6 Schematic of the energy band diagram of Poole-Frenkel emission process in a MIS structure. Adapted from [46] . . . . . 31
3.7 Schematic of the energy band diagram of variable-range hopping process in MIS structure. Adapted from [46] . . . . . 32
3.8 Schematic of the energy band diagram of Ohmic conduction in a MIS structure. Adapted from [46] . . . . . 33
3.9 Typical current density-voltage (J-V) characteristic in the log $J$-log $V$ plane for space-charge-limited conduction current for an insulator with a discrete shallow trap level. The three bounding curves are the Ohm's law ($J \propto V$), traps-filled limit (TFL) current ($J_{TFL} \propto V^2$), and Child's law ($J_{Child} \propto V^2$). $V_{tr}$ and $V_{TFL}$ are the transition voltage at the departure from ohm's law and TFL curve. Adapted from [46] . . . . . 34
3.10 Schematic of the physical mechanisms occurring in SCLC during weak injection. For voltages lower than $V_{\mathrm{tr}}$ conduction is Ohmic **(b)**, until V = $V_{\mathrm{tr}}$ where SCL conduction starts. Adapted from [46] . . . . . 35
3.11 Schematic of the physical mechanisms occurring in SCLC during weak injection. For $V_{\mathrm{tr}} <$ V $< V_{\mathrm{TFL}}$ conduction is trap-filled limited **(b)**, and for V $> V_{\mathrm{TFL}}$ it becomes trap-free.Adapted from [46] . . . . . 36
3.12 Schematic representation of the pattern transfer on a negative versus a positive photoresist after light illumination. Adapted from [53]. . . . . 38
3.13 Schematic process flow during lift-off transfer. Adapted from [53]. . . . . 38
3.14 Typical ALD cycle. Adapted from [44] . . . . . 41
3.15 Schematic representation of a PLD process. Adapted from [44] . . . . . 42
3.16 Schematic representation of a typical sputtering process. Adapted from [44] 43

4.1 Illustration of the effect of surface energy and confinement strain in inducing the polar distorted phase. The cyan arrow represents the anions (cyan) displacement in the orthorhombic phase with respect to the cations (blue). Grey arrows indicate that surface energy always favors higher symmetry structures (monoclinic ⟶ orthorhombic ⟶ tetragonal) while distortion effects enhance lower symmetry (tetragonal ⟶ orthorhombic). Adapted from [76] . . . . . 48
4.2 Electroresistance loop measured at 100 mV after the application of voltage pulses of variable amplitude, in TiN/HZO/WOx/TiN structure. The device has a broad coercive field distribution and well-defined intermediate resistance states. Adapted from [82] . . . . . 50
4.3 Process-flow for HZO crossbar: **(a)** Deposition of the full material stack after annealing. **(b)** Definition of the top electrode by optical lithography and RIE of top TiN and W layers. **(c)** Definition of the bottom electrode by optical lithography and IBE of the HZO, WOx and TiN layers. **(d)** The devices are isolated from each other by a $SiO_2$ passivation layer deposited by PECVD. **(e)** Vias to the device contacts are etched, $SiO_2$ layer by RIE, **(f)** then the HZO and the WOx by IBE. **(g)** M1 is deposited by sputtering and patterned by RIE. $SiO_2$ passivation layer is deposited by PECVD and vias to the bottom electrode are etched. **(h)** A second passivation layer of $SiO_2$ is deposited to isolate M2 from M1. Vias to M1 are etched. M2 is deposited by sputtering, then patterned by RIE. . . . . 51
4.4 Optical microscope snapshots of HZO crossbar arrays. 32x32 array and 2x2 arrays for device characterization **(a)**, 81x10 array **(b)**, 2x2 array **(c)** and 4x4 array **(d)**. Snapshot of non-contacted FTJ devices (e). . . . . 52
4.5 Mask design for DLW lithography, performed on Klayoyt software. Full chip with 32x32, 81x10, 2x2, and 4x4 HZO crossbars. . . . . 53
4.6 Rhombohedral **(a)** and Tetragonal **(b)** BFO unit cells. Adapted from [86]. 54
4.7 Representation of the various crystal structures that BFO thin film assumes under epitaxial strain. In blue are the unit cell of the relative structure, in grey the pseudocubic perovskite unit cell- Adapted from [84]. . . . . 55
4.8 Sketch of pre- and post-neurons connected by a biological synapse and an artificial one, i.e. the Co/BFO/CCMO/YAO ferroelectric tunnel junction. The synaptic plasticity relies on the causality ($\Delta t$) of neuron spikes. Adapted from [25] . . . . . 55

4.9 Electroresistance hysteresis loop of the ferroelectric memristor displaying clear voltage thresholds $V_{th}^{+}$ and $V_{th}^{-}$. A single voltage pulse is applied to switch the device into many intermediate states. Adapted from [25] . . . . 56
4.10 Process-flow for BFO crossbar: **(a)** BFO is etched by ion milling to access the CCMO back electrode. **(b)** Using the same e-beam resist from **(a)**, Ni contacts to the CCMO are realized by lift-off. **(c)** Using a lift-off process, Ni contacts are placed above BFO to form a FTJ device (Ni/BFO/CCMO). **(d)** Each device is isolated from each other by a surrounding trench that is etched by ion milling. **(e)** A thin passivation ($Al_2O_3$ (5 nm) / $SiO_2$ (100 nm)) is deposited to isolate M1 from the BFO. Vias to the device contacts are etched, and **(f)** M1 is deposited by sputtering and patterned by RIE. **(g)** A second passivation ($Al_2O_3$ (5 nm) / $SiO_2$ (300 nm)) is deposited to isolate M2 from M1. Vias to M1 are etched. **(h)** M2 is deposited by sputtering, then patterned by RIE. . . . . 57
4.11 Experimental setup for single device characterization, with access pads to M1 and M2. The device is set to LRS after writing with a positive amplitude **(a)** and to HRS with a negative writing amplitude **(b)**. . . . . 58
4.12 Resistance hysteresis measurements of a FTJ (diameter 1 $\mu$ m). Each measurement is repeated five times, in each the resistance is measured at $V_{read}$ = 0.2V after applying pulses of amplitude $V_{write}$. Blue curve: after defining the first metal level. Red curve: after defining the second metal level. The arrow indicates the direction of the hysteresis. Published in [91]. 59

5.1 Temperature-dependent I-V measurements in $|log(J)| - V$ representation. The arrows are numbered according to the chronological order in which the applied voltage amplitude is increased or decreased. . . . . 61
5.2 Non-switching branch of the temperature-dependent I-V measurements. The positive branch corresponds to the LRS, while the negative branch to the HRS. . . . . 62
5.3 Typical current density-voltage (J-V) characteristic in the log $J$-log $V$ plane for space-charge-limited conduction current for an insulator with a discrete shallow trap level. The three bounding curves are the Ohm's law ($J \propto V$), traps-filled limit (TFL) current ($J_{TFL} \propto V^2$), and Child's law ($J_{Child} \propto V^2$). $V_{tr}$ and $V_{TFL}$ are the transition voltage at the departure from ohm's law and TFL curve. Adapted from [46]. . . . . 63

5.4 Current-voltage characteristics in the Ohmic and Space-Charge-limited conduction (SCLC) representations. At each temperature, a linear regression is performed on the negative, increasing branch (HRS, nonswitching) and on the positive, decreasing branch (LRS, nonswitching). . . . . 64
5.5 Arrhenius plot of the slopes ($m$) obtained in the SCLC region of I-V characteristics for **(a)** positive nonswitching branch (LRS) and **(b)** negative nonswitching branch (HRS). From the slope of the Arrhenius plot, the parameter $T_C$ is calculated in the LRS and HRS. . . . . 65
5.6 Arrhenius plots of the intercepts of the linear regressions in the Ohmic regime. The $\mu N_v$ and $E_F - E_v$ parameters are calculated from the intercept and the slope, respectively. . . . . 66
5.7 Arrhenius plots of $\log J$ for **(a)** positive nonswitching branch (LRS) and **(b)** negative nonswitching branch (HRS). The current density $J$ is computed at 4 different bias values using Eq. (5.4). The Arrhenius slopes are related to the activation energies. . . . . 67
5.8 Schematic representation of shallow traps distributed exponentially with parameter $T_C$ in the BFO band gap. The black solid line denotes the variation of the trap density $N_t$ with the black arrow, which corresponds to the energy axis along which the activation energy varies. . . . . 68
5.9 Dependence of the activation energy on $\log(|V|)$. The data show the linear dependence expected from Eq. 5.5 . . . . . 69
5.10 Extrapolation of $\log J - \log V$ curves in **(a)** LRS and **(b)** HRS. The fitting lines intersect at the critical voltage $V_C = 7\,\mathrm{V}$ in **(a)** and $V_C = 18\,\mathrm{V}$ in **(b)**. 70
5.11 Current-voltage characteristics in the TAT representation. At each temperature, a linear regression is performed on the negative, increasing branch (HRS, nonswitching) and on the positive, decreasing branch (LRS, nonswitching). . . . . 71
5.12 Current-voltage characteristics in the FNT representation. At each temperature, a linear regression is performed on the negative, increasing branch (HRS, nonswitching) and on the positive, decreasing branch (LRS, nonswitching). . . . . 72
5.13 Equivalent circuit model describing the $J - V$ behavior of the dielectric. The bulk-limited and the interface-limited conduction paths contribute in parallel to the transport. The interface-limited current is modeled with two series-connected current sources ($\mathrm{J}_{TAT}$ and $\mathrm{J}_{FNT}$) to emphasize the dominance of the smallest contribution. . . . . 73

5.14 Reproduced experimental data using the equivalent circuit model proposed above. . . . . 74
5.15 Resistive switching mechanisms with polarization reversal. Charge distributions and energy-band diagrams of Ni/NiOx/BFO/CCMO heterostructure for the **(a)** LRS state and **(b)** HRS state. The modulation of the barrier height and width by the polarization reversal at the BFO/CCMO interface determines the resistive switching. V. For BFO, the band gap is 2.8 eV and the electron affinity is 3.3 eV [119]. Black dots represent electrons, while positive red signs and blue minus signs are the positive and negative bound charges, respectively. . . . . 78

# List of Tables

5.1 $\mu N_v$ product and $E_F - E_v$ in the Ohmic regime for LRS and HRS. . . . . 66
5.2 Characteristic temperatures and energies in EDT-SCLC regime for LRS and HRS. . . . . 67
5.3 Values of $\mu N_c$ for the EDT-SCLC regime at different temperatures, for both LRS and HRS. . . . . 73