Master's Degree in Nanotechnologies for ICTs

BEOL CMOS-Compatible Ferroelectric Fin-FET for Neuromorphic computing.

Falcone Donato Francesco

IBM Research, Zurich

Master Thesis

Supervisors :

Dr. Laura Bégon-Lours	IBM Research, Zurich
MSc. Mattia Halter	IBM Research, Zurich
Prof. C. Ricciardi	Politecnico di Torino

A.A. 2020/2021

Acknowledgements

I would like to thank:

- Dr. Laura Bégon-Lours and Mattia Halter as my master thesis supervisors at IBM Zurich Research Lab.
- Prof. Carlo Ricciardi as my master thesis supervisor at Politecnico di Torino.
- The whole Neuromorphic devices and systems group of IBM Zurich Research Lab for the scientific support.
- The Binnig and Rohrer Nanotechnology Center (BRNC).

Finally, I would like to thank my family and my girlfriend, who have supported me throughout the entire master program, and all the people having a special role in my life.

Abstract

The way in which hardware components are organized into a functional computer, namely the von Neumann architecture, has barely changed since its inception in 1945 [1]. The bottleneck of this architecture consists in the huge data transferring between the processor and the memory. Nevertheless, with the advent of the Internet of Things (IOT) and the Artificial Intelligence (AI), an exponential growth in the amount of processed data, has imposed critical requirements in terms of energy efficiency and processing speed. Neuromorphic hardware allows to perform computing at the site where data is stored, offering an attractive solution for these issues. Neuromorphic architecture can be based on a memristor, known as a programmable resistor, which is a circuit element that changes its resistance depending on how much charge flowed through it. Ferroelectric based memristors are a promising candidate to build energy efficient neuromorphic hardware.

In this work, a ferroelectric $Hf_{0.57}Zr_{0.43}O_2$ (HZO) field-effect transistor (FeFET) memristor has been electrically characterized in order to study the conduction mechanisms, as well as the physics behind the resistive switching, governed by the polarization screening charge in a WO_x channel. In particular, to find out the conduction nature both along the channel and through the gate stack, temperature-dependent electrical measurements have been carried out. In addition to the physical understanding of FeFET devices, to overcome the performance of standard planar devices, a new generation of devices has been processed and characterized, the Fin-FeFETs. They allow to increase the electrostatic gate control of the channel, and thereby getting a better resistive switching. The whole process flow has been studied and optimized, to achieve a fin's resolution of less than 10 nm width.

These devices can be used in crossbar array configuration to allow energy efficient vector-matrix multiplication, as well as Hebbian learning and Spike-timing-dependent plasticity (STDP) for spiking neural network tasks.

Contents

List of Figures

Chapter 1

Introduction and Background

1.1 Neuromorphic Computing

Computing capability of classical digital computers, based on complementary metal oxide semiconductor (CMOS) transistors, has advanced considerably in the past decades, mainly due to the shrinking down of transistor's dimensions, as predicted by Moore's law [2]. However, the architecture of classical computer, hence the way in which hardware components are organized, namely the von Neumann architecture, has barely changed since its inception in 1945 [1]. The main advantage of this architecture is the modularity of the engineering design, which allows to build extremely complex system without the need to understand all the components.

However, in the last years, with the advent of the Artificial Intelligence (AI) and the Internet of Things (IOT), an exponential growth in the amount of processed data, thus computing power, has imposed critical requirements in terms of energy efficiency and processing speed. In fact, AI methods, such as Artificial Neural Networks (ANNs) and Deep Neural Networks (DNNs), require tremendous computing resources during the learning to address ambitious problems such as speech and image recognition [3]. Digital computers based on von Neumann architecture, are not well suited to efficiently solve AI tasks. In fact, in this architecture, memory and processing unit are physically separated, hence data need to travel back and forth between memory and process unit. This operation takes a lot of time and energy, and is known as the *von Neumann bottleneck* [4]. In addition, performance mismatch between memory and process unit can lead to considerable latency, known as *memory wall* in the von Neumann architecture [2].

To overcome these limitations and speed up significantly specific computations used in ANNs and DNNs, a new computing paradigm is necessary. A first improvement in this direction, has been reached by the use of massively parallel and specialised architectures, like Graphical Processing Units (GPUs). In fact, the highly parallel structure of GPUs, makes them more efficient than conventional general-purpose processing unit (CPUs) for algorithms that process large amount of data in parallel, which is the case for ANNs and DNNs. However, GPUs, even being parallel com-

puting structures, are still based on the von Neumann architecture, hence they do not represent the ultimate energy efficient architecture for AI tasks.

A more radical alternative to classical von Neumann computing scheme is represented by the so-called *in-memory computing*. This architecture allows to perform computing at the site where data is stored, hence in-memory, avoiding time and energy wasting for data-travelling back and forth between the memory and the process unit. In-memory computing paradigm was proposed for the first time in the 1960s [5], and successively proved in digital domain [2]. However, probably due to the satisfying improvement of classical CMOS based computing capability, not enough attention was paid to this novel architecture.

Nowadays, in-memory computing paradigm is one of the most promising approaches to overcome the von Neumann bottleneck, offering an attractive solution to the energy consumption and speed issues of AI tasks.

An artificial neural network is a computing system, inspired by biological neural networks, that tries to imitate the way the human brain works. It is based on a collection of nodes, known as artificial neurons, connected to each other through elements that define the strength of each interconnection, known as artificial synapses. This mathematical abstraction can be used to process information and learn recurrent features from data, adapting the synaptic weights to better fit the data.

In order to be able to properly work, an artificial neural network, as its biological equivalent, has to be trained on a known data set, and interrogated on an unknown one, to infer information from it. Therefore, two main phases can be identified in ANNs employment, the training, where the ANN is trained by processing examples in a supervised or unsupervised approach, and the inference where a new data set is provided to the network, and information can be inferred from that. During the training, a probability-weighted associations between input and result of each training example, is formed. During this phase, the ANN synaptic weight can be adapted so that the actual output matches the desired output for the set of reference training data. During the inference phase, the ANN is actually used to predict the output for unknown new data.

A deep neural network (DNN) is a particular type of artificial neural network, composed by a succession of interconnected layers, each one containing a given number of neurons [6]. In figure 1.1 a schematic representation of a fully connected deep neural network is shown.

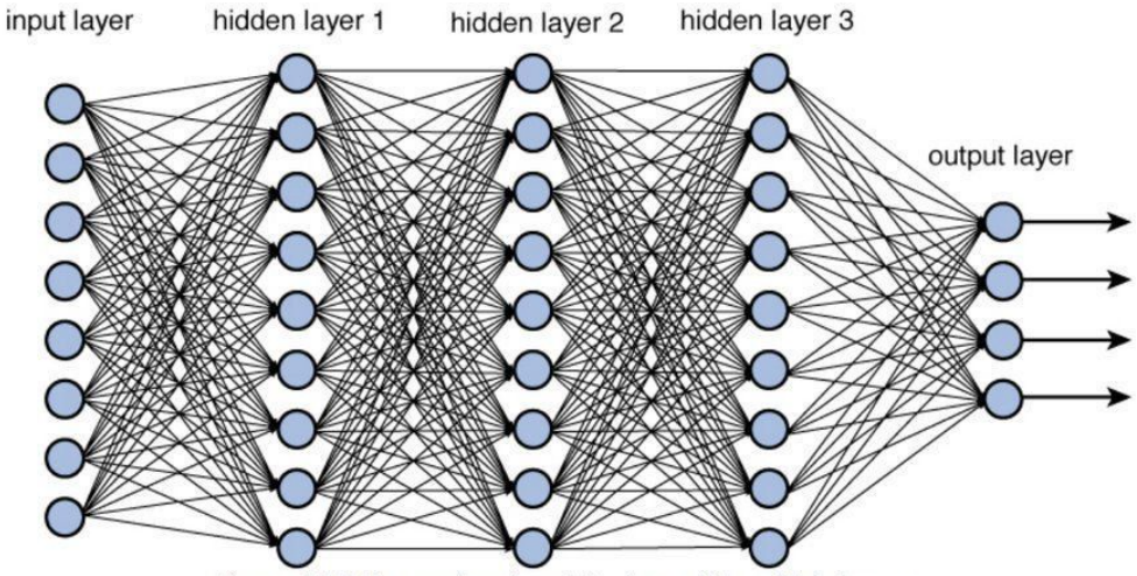


Figure 1.1: Schematic representation of a fully connected deep neural network with three hidden layers.

Neurons forming the input layer receive signals as input and propagate them to the neurons of the hidden layers, named in this way since not visible during the computation. Each neuron of the hidden layer gets as input the weighted sum of the outputs of all the neurons from the previous layer. The sum is processed through a non linear activation function, usually a sigmoid, and then propagated until it reaches the output layer. The synapses, represented as arrows in figure 1.1, have a strength which can be tuned during training. From a mathematical point of view, the synaptic weights can be represented by a matrix and the transmission of the information from one layer to the next occurs by matrix-vector multiplication. This operation is the most resource-demanding step in DNNs on a classical computer architecture [3], [7].

The in-memory computing paradigm can be applied to AI tasks and more specifically to DNNs, to perform matrix-vector multiplication in a much more efficient way. In fact, during the inference for instance, the synaptic weights contained in the mapping matrices are constant and only the input of the DNN changes. However, in a classical computer, for each evaluation, the actual layer and the corresponding synaptic matrix have to be brought from the memory to the processor.

This process may be made more efficient by exploiting a cross-bar architecture, where computing and storing take place in the same location, based on resistors. A crossbar architecture is composed by two superimposed perpendicular electrode lines, one for the input and one for the output. At each crosspoint between input and output electrodes, a resistor is placed. When voltages are applied on the input electrode, thus a vector $\vec{V} = V_i$, the corresponding vector $\vec{I} = I_m$ of the currents coming out from the output electrode is the multiplication of the synaptic matrix G $= G_{mi}$ of resistors and the voltage input vector.

$$I_m = \sum_{i=1}^{\#_{Voltages}} G_{mi} \cdot V_i \tag{1.1}$$

$$\vec{I} = G \cdot \vec{V} \tag{1.2}$$

In figure 1.2 a schematic representation of a crossbar array architecture, is shown.

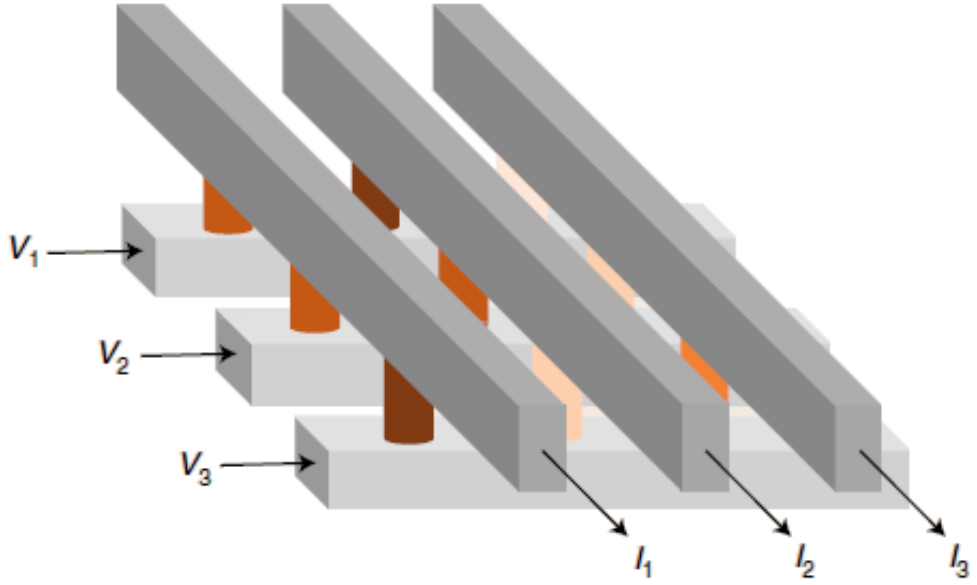


Figure 1.2: Schematic representation of a crossbar array based on resistors. Taken from [2].

Supposing that the resistors located at crosspoints can store information, hence their resistive values are programmable and stable in time, the crossbar array can memorize a matrix of synaptic weights. This intrinsically parallel architecture, allows programmable matrix-vector multiplication in one time step, independently of the array size, whereas in a classical computer the computing time of a matrix-vector multiplication scales with the size of the matrix.

An example of a programmable non-volatile resistor is the memristor, and a in-memory computing architecture based on memristors is known as neuromorphic architecture.

1.2 Memristor

A memristor is a non-volatile resistor which exhibits resistive switching [8]. In particular, the resistive state of a memristor can be programmed by external stimulus, such as voltage or current. However, since it is a non volatile element, the programmed value of resistance is retained even when the external stimulus disappears.

In the following, the main figures of merit of any memristor, independently from the physical principle behind the resistive switching, to be actually used in fast and energy efficient crossbar arrays, are listed [3], [9], [2].

- **Dynamic range** :
  Dynamic range is defined as the ratio between the maximum and minimum resistance states, namely the on/off ratio. A large dynamic range allows a proper mapping capability of the weights in the AI algorithm used. Usually, memristor's dynamic range larger than 2 is required.

- **Resistance** :
  The value of resistance is required to be high in order to reduce the total power consumption. In addition, considering integration in large cross-bar array, the resistance of the metal lines may be non negligible, thus a high resistance is required to preserve the switching capability of the memristor. In fact, if the high and low resistive values were comparable with metal lines resistance, since the latter is equivalent to an additional series resistance, the overall switching capability, hence the on/off, would be deteriorated. Usually, to address this issue, memristor's resistance larger than few hundreds of $k\Omega$ is required.

- **Multilevel states** :
  In order to get a more precise weight-conductance mapping, a high number of resolvable conductance states within the dynamic window, is required. This entails that the transition between the high and low resistive states has to be smooth, with several intermediate states, known also as analog states, in between. However, the number of resolvable conductance states remains strongly application-dependent, with the training that usually requires more analog states than the inference.

- **Size** :
  The memristor's footprint is particularly important to deal with large scale integration of modern neural networks. In addition, as previously mentioned, the advantage of cross-bar architecture with respect to the classical one, becomes more and more evident by increasing the array's size, hence neural network's size. This figure of merit, for which the criteria the smaller the better always holds, is strongly related to the number of resolvable analog states. However, since usually a trade-off between the memristor's size and its multilevel nature, exists, these figures of merit may be merged to define the number of bits stored per unit of area.

- **Weight-update linearity and symmetry** :
  The relationship between the device conductance and the number of identical programming pulses, defines the weight update process. The synaptic weight increase process is known as long term potentiation (LTP), or just potentiation, while the synaptic weight decrease is named long term depression (LTD), or simply depression. Ideally, for a direct mapping of the synaptic weights into actual conductance values, both potentiation and depression processes are required to be linear and symmetric. In fact, the non-linearity/asymmetry

makes the relative change of the synaptic weight ΔW dependent on the current weight W, causing a history dependence of the weight update process. It has been shown that both non linearity and asymmetry in the weight update process, negatively affects the learning accuracy in neural networks.

- **Retention** :
  Retention is defined as the ability to retain information over a period of time. During the inference, the memristor should behave as a long-term memory element, ideally storing the learned conductance value for a period of the order of 10 years at the maximum chip working temperature of 85 °C. These requirements are less critical for memristors used in neural networks with online training, where the weight update is continuous.

- **Endurance** :
  Endurance is defined as the number of set/reset cycles that can be applied to a memory element before it becomes unreliable. Endurance requirement is strongly application-dependent, since it depends on how many weight updates are required during the training process.

- **Yield and variability** :
  A low device-to-device variability is advisable for memristors integration in crossbar array. In addition, also the processing yield, defined as the ratio between the number of properly working devices and that of processed devices, may strongly impact crossbar's performances.

- **I-V linearity** :
  The matrix-vector multiplication, other than with binary inputs, may be properly emulated in hardware only if the memristors exhibit purely ohmic behaviors. In fact, if some non-linearity is present, the result of the matrix-vector multiplication may be distorted, hence the performances of the implemented neural network could be negatively affected.

There are several physical phenomena which can lead to a memristive behavior. In the following, the four major memristive technology candidates for crossbar-array integration are presented [9]:

- **Phase Change Memory (PCM)** :
  Phase change memory devices are based on specific materials that can switch between at least two structurally distinct solid phases, an amorphous and a crystalline one. In particular, the switching from amorphous to the energetically favorable crystalline phase occurs by heating the phase change material above its crystallization temperature for a time long enough for crystallization to occur [10]. On the other side, the switching from crystalline to amorphous phase is performed by melting and quickly quenching the phase change material. In both cases, the heat is provided by the Joule effect, however since the material shows a different resistivity according to its phase, it is possible to store a non-volatile information, getting a typical memristive behavior.

- **Resistive Random Access Memory (ReRAM)** :
  The working principle of ReRAM devices, composed by a dielectric material surrounded by two electrodes, relies on the formation and modulation of a conductive filament through the dielectric itself. An electroforming process, similar to a dielectric breakdown, is performed to electrically induce ion migration within the dielectric, forming a conducting filament. After that, the formed conductive filament may be strengthened or weakened by external stimulus, leading to a resistance modulation of the dielectric. However, single layer ReRAM usually exhibit an abrupt set (weight increase) process, which is difficult to control. This effect may be mitigated by material stack engineering, in particular using an oxide bi-layer, which allow to make weak or multiple weak filaments, as proved in TaO_x/HfO_2 [11] and AlO_x/HfO_2 [12] devices.

- **Ferroelectric Tunnel Junctions (FTJ)** :
  Ferroelectric tunnel junction devices, for example based on Metal-Ferroelectric-Insulator-Metal (MFIM) structure, exploit the partial polarization switching, present in multidomain ferroelectric materials, to modulate the tunneling barrier energy profile between the two electrodes. Polarization dependent tunneling current allows to define Tunnelling Electro-Resistance (TER), as the change in the electrical resistance associated with the polarization reversal in FTJ [13]. Therefore, the higher the tunnelling electro-resistance, the better is the resistive switching.

- **Ferroelectric Field Effect Transistor (FeFET)** :
  The ferroelectric field-effect-transistor synaptic device is a type of field-effect transistor, in which a ferroelectric material is placed between the gate electrode and the channel material. This memristor exploits the partial polarization switching, present in multidomain ferroelectric materials, to gradually tune the channel conductance. In addition, FeFET memristor, being a three-terminal device, allows to decouple the weight writing and reading paths [9]. In particular, the weight programming relies on the writing voltage applied to the gate, while the weight read-out on the reading voltage applied to the channel. Due to the three-terminal nature, FeFET memristors are organized into a pseudo-crossbar array architecture, which allows to mitigate several issues of the true crossbar array, such as the cross-talk effect due to sneak paths [9], [14].

This thesis focuses on ferroelectric field effect transistor synaptic devices, based on hafnium zirconium oxide HZO and tungsten oxide WO_3 as ferroelectric and channel material respectively. In the following section, a detailed overview of FeFET memristors used as starting point in this thesis, is provided.

1.3 Ferroelectric Field-Effect-Transistor

In the last decades, thin film HfO_2 attracted scientific interest into the research community, as best candidate to replace SiO_2 as gate dielectric for MOSFET applications. The interest in hafnium oxide material, basically motivated by its high dielectric constant ($\epsilon_r \simeq 25$) and moderate band gap ($E_g \simeq 5.8\,\text{eV}$) [15], brought several advancements in material knowledge. This allowed Böscke et al. in 2011 to report the discovery of ferroelectricity in silicon doped hafnium oxide capped with a TiN metal gate and subsequently annealed [16]. Silicon incorporation into the hafnium oxide lattice was discovered to induce a metastable tetragonal crystal phase. By further annealing the doped HfO_2 with a top electrode, the mechanical confinement suppressed the transition of the tetragonal to the monoclinic phase, and Böscke et al. postulated the crystallization into a polar orthorhombic phase (Pbc_{21}) responsible for the observed ferroelectricity [17].

Afterwards, the ferroelectric properties of hafnium oxide were explored using several dopants, such as Al, Gd, La, Zr. In this thesis work an ALD Zr-doped HfO_2 high-k dielectric film, which has been proved both to partially stabilize the orthorhombic structure of hafnium oxide, and to improve the surface morphology of the latter [18], is used to process ferroelectric field effect transistor synaptic devices.

In particular, in this thesis a $Hf_{0.57}Zr_{0.43}O_2$ (HZO)-based FeFET employing a tungsten oxide (WO_x) channel is reported. Since the memristor's process flow is CMOS compatible and exploits only abundant and CMOS friendly materials, the $HZO - WO_x$ stack is promising for large-scale integrated neuromorphic hardware based on ferroelectrics [19].

The FeFET structure used as starting point in this thesis is the one proposed by M. Halter et al. in IBM Zurich Research Lab in 2020 [19], and consists in a bottom-gate junction-less device, where the gate contact is accessed through the Si n+ substrate. On top of the gate, using a plasma-enhanced atomic layer deposition (PEALD) system, first 10 nm thick TiN layer is deposited using a tetrakis-(dimethylamino)titanium (TDMAT) precursor and N_2/H_2 plasma, then 10 nm thick HZO layer is grown using alternating cycles of tetrakis- (ethylmethylamino)hafnium (TEMAH) and ZrCMMM ((MeCp)- 2Zr(OMe)(Me)) at T = 300 °C. In addition, a W layer is sputtered and the crystallization of HZO using a millisecond flash lamp anneal system with a pre-heating temperature of 375 °C, is performed. After crystallization, tungsten is crystallized and oxidized to WO_3 in a rapid thermal annealer (RTA) at T = 350 °C and 50 sccm oxygen flow. In conclusion, an oxygen reduction from WO_3 to WO_x is performed in a Rapid Thermal Annealer by H_2 annealing [19]. After source and drain patterning by lift-off and channel etching by reactive ion etching (RIE), a passivation made of SiO_2, and exploiting Al_2O_3 as etch stop layer, is deposited, and VIAs contacts using W as metal, are defined.

In figure 1.3, a schematic illustration of bottom gate FeFET memristor developed by M. Halter et al. in IBM Zurich Research Lab in 2020 [19], is provided.

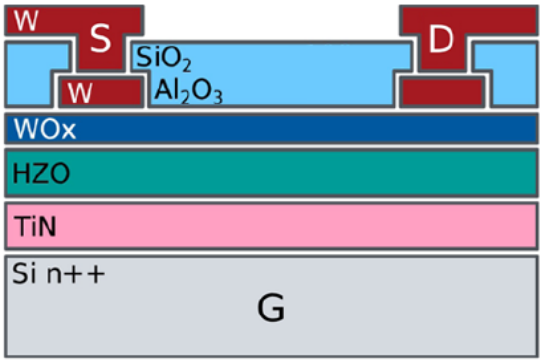


Figure 1.3: The schematic illustration of a bottom gate FeFET, indicating source (S), drain (D), gate (G), WO_x channel and ferroelectric HZO gate dielectric. Taken from [19].

The proposed bottom gate approach, which enables to crystallize hafnium zirconium oxide in its ferroelectric phase before the actual deposition of the channel material, allows to arbitrarily tune the resistivity of the latter by oxygen reduction or oxidation steps, since WO_x resistivity exponentially depends on the oxygen concentration, increasing with the increase of x stoichiometry [20]. By contrary, using a conventional MOSFET-like process flow, based on the deposition and patterning of channel material first, and the structuring and crystallization of the gate stack later, the resistivity of resulting channel can not be finely tuned and it has been noticed that tungsten oxide is strongly reduced by the millisecond flash lamp annealling during HZO crystallization.

The switching mechanism of $HZO - WO_x$ FeFETs relies on the voltage induced partial polarization switching, used to electrostatically deplete or accumulate free carriers in the thin WO_x channel. When the HZO ferroelectric remanent polarization points toward the interface with WO_x, free carriers accumulate at the interface to screen the depolarization field in HZO, thus the channel resistance R_{SD} decreases, and the memristor is in its low resistive state. By contrary, when remanent polarization points toward the $HZO - TiN$ interface, carrier depletion occurs in tungsten oxide at the interface with HZO, causing an increase of the channel resistance R_{SD}, resulting in a high resistive state. Since in both the HRS and LRS the polarization is screened in WO_x at the interface with HZO, it is possible to define a screening length x_d, representing the thickness of the ferroelectric modulated resistance, which increases as the carrier density N_D decreases [19]. The overall channel resistance R_{SD} in this junction-less FeFET can be thought as the resistance of two channels in parallel: one of thickness x_d, in which the resistivity is modulated upon polarization switching, and one of thickness $d_{WO_x} - x_d$ with a constant resistivity, where d_{WO_x} is the physical thickness of deposited tungsten oxide layer [19]. By decreasing the channel thickness d_{WO_x}, the dynamic range is enhanced, since the resistivity of an increasing proportion of the channel is modulated by ferroelectric polarization.

In addition, the higher the FeFET channel resistivity, the better the dynamic range is, however, if the channel is excessively oxidized, hence WO_x with x ≥ 3, stabiliza-

tion issue of remanent polarization can occur since there are not enough free-carriers in WO_x to screen the depolarization field.

However, since the FeFETs developed by M. Halter et al. in IBM Zurich Research Lab [19] are planar devices with micrometric sizes, a further improvement can be achieved exploiting a tri-gate architecture, as the one demonstrated by *Intel* in 2011 [21] for MOSFET applications. In this architecture, the gate surrounds the channel on three sides, creating a multigate device known as FinFET, with better gate control, thus performances, with respect to planar FET.

In figure 1.4, a graphical illustration of planar and fin based FET architecture is provided.

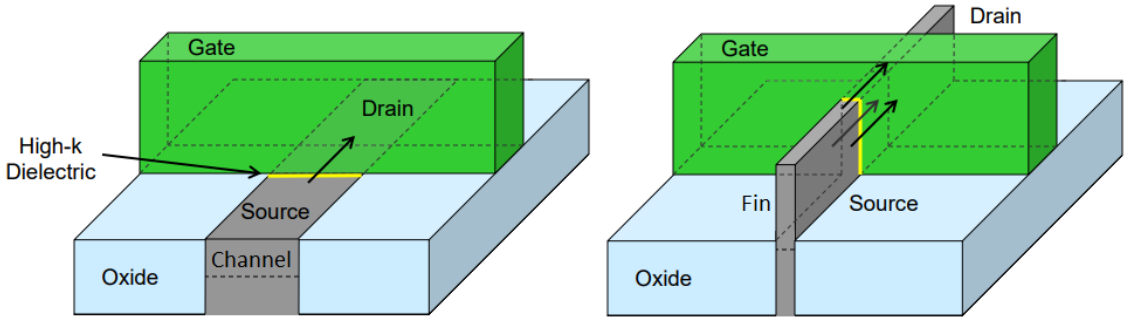


Figure 1.4: A graphical illustration of planar and fin based FET architecture. Redrawn from [21].

The main goal of this thesis is first to characterize planar FeFET memristors, as the one proposed by M. Halter et al. in IBM Zurich Research Lab [19], to explore and physically model the electrical transport both in the channel and along the gate stack, then to transfer the FeFET planar technology into a multigate FinFeFET one. A new process flow for $HZO - WO_x$ FinFeFETs has been studied and optimized to allow a dramatic scaling down of FeFET footprint, from micrometric planar devices to sub-10 nm fin based FeFETs, keeping the desired figures of merit of planar FeFETs. After fabrication and process flow optimization, FinFeFET electrical characterization showed promising results and paved the way for further device optimizations.

Chapter 2

Characterization Methods

This chapter details the different techniques used to characterize three terminal ferroelectric memristors. Since a memristor is a non-volatile electronic element, the electrical characterization, both DC and pulse based, is the way through which it is possible to estimate and quantify its overall performances. In addition, to understand the physics behind the switching mechanism, a characterization of physical conduction mechanisms both along the gate stack and between source and drain has been carried out with temperature dependent DC-electrical measurements. Material characterization has been performed through Circular Transfer Length Method, namely CTLM, which allows to extract parameters such as channel resistivity, as well as sheet and contact resistances. Finally, to study the crystalline structure of the materials involved, grazing incidence X-Ray diffraction, namely GIXRD, has been performed, primarily to check that $Hf_{0.57}Zr_{0.43}O_2$ has crystallized in its orthorombic phase.
Characterization tools such as Focused Ion Beam (FIB) and Scanning Electron Microscope (SEM) have been intensively employed after each critical fabrication step, to check the result and optimize the overall process flow.
The details of all the previously mentioned characterization methods are reported in the following sections.

2.1 Electrical Characterization

The DC and pulsed electrical characterization of memristors is performed using the Agilent B1500A Semiconductor Device Analyzer. In particular, the Agilent EasyEXPERT software has been used as operating environment to perform current-voltage sweep measurements. The signals are always applied through Source Measurement Unit cables, known as SMUs. In figure 2.1 an SMU terminal is shown. It has three leads, a central conductor to force or sense the signal, an encapsulating conductor that shields the center signal by employing the same voltage thus decreasing a possible leakage current, called Guard, and an outer conductor that serves as Common.

Figure 2.1: On the left a picture of SMU triax connector, while on the right a schematic representation of the cable. Taken from [22].

In addition to the Agilent B1500A, Cascade Summit 12000B-AP probe station is used for all the electrical measurements.
Electrical characterization of planar and fin based FeFET memristors consists in the following procedures:

- **Reset Operation**:
  To implement the Reset operation, the memristor has to be forced in its High Resistive State (HRS). This is done putting both the source and the drain as common terminal and applying a negative voltage sweep on the gate from $0\,\mathrm{V}$ to V_{reset} back and forth ($V_{reset} < 0$), to write the polarization towards the HZO-TiN interface, depleting electrons from WO_x channel.

- **Read Operation**:
  During the Read operation, a voltage sweep from $-V_{read}$ to $+V_{read}$ back and forth, is applied between source and drain. Usually the reading operation is done using $V_{read} = 0.2\,\mathrm{V}$, however for channel conduction exploration an higher V_{read} may be required. In the latter case, for $|V_{read}| > 1\,\mathrm{V}$, the reading operation is performed lifting up the gate tip from the corresponding pad, to avoid a potential drop between channel and the gate itself, which perturbs the measurements causing an hysteresis in I-V measurements.

- **Set Operation**:
  To implement the Set operation, the memristor has to be forced in its Low Resistive State (LRS). This is done putting both the source and the drain as common terminal and applying a positive voltage sweep on the gate from $0\,\mathrm{V}$ to V_{set} back and forth ($V_{set} > 0$), to write the polarization towards the HZO-WO_x interface, accumulating electrons in WO_x channel.

The previously described procedures are performed for both DC and pulsed characterization. However, the main difference is that in DC sweep, the voltage profile is a step-like function, with the amplitude kept constant for a defined amount of time, while in pulsed sweep, a single short pulse of varying amplitude is sent to program the device, and after each pulse the memristor state is read. In figure 2.2, a graphical representation of DC and pulsed sweep measurements is provided.

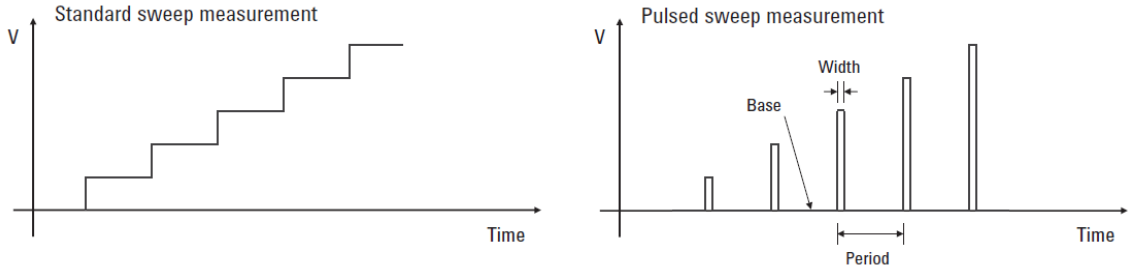


Figure 2.2: On the left a standard DC sweep profile, while on the right a pulsed measurement scheme. Taken from [22].

DC characterization allows to quantify memristor figures of merit such as the resistive range and the ON/OFF ratio, while from pulsed characterization, long term potentiation and depression cycles, as well as the linearity of weight update, are determined.

In the following, two additional electrical methods, based on DC electrical characterization, used to investigate the nature of conduction and the material properties, are detailed.

2.1.1 T-Dependent DC-Electrical Measurements

To investigate the physics and the conduction mechanisms of memristors, temperature dependent DC electrical measurements can be exploited. In fact since several conduction mechanisms have different temperature dependence, measuring the conduction currents at different temperatures, may allow to determine the different physical mechanisms describing the conduction.

The conduction mechanism in a dielectric can be of two types, electrode-limited or bulk-limited. The electrode-limited conduction mechanism depends on the electrical properties at the electrode-dielectric interface, while the bulk-limited one depends on the electrical properties of the dielectric itself [23]. This classification is essential since several conduction mechanisms may take part simultaneously.

In figure 2.3 a classification of all the conduction mechanisms taken into account in the following analysis, is provided.

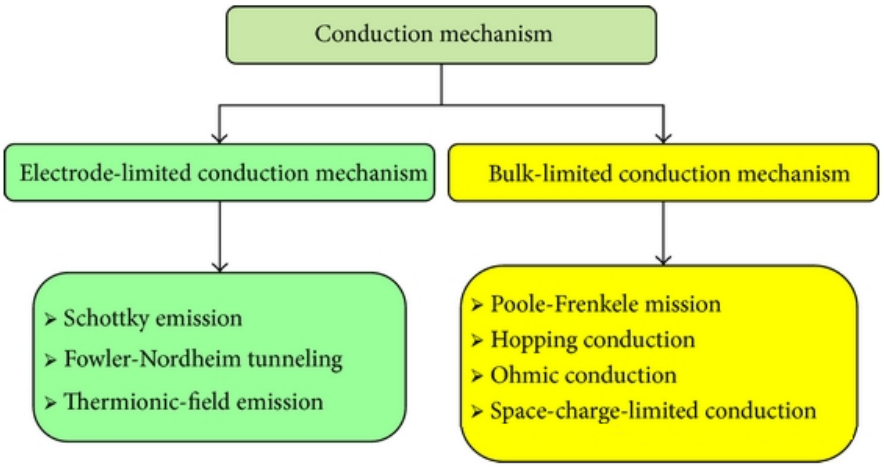


Figure 2.3: Classification of conduction mechanisms in dielectric films. Adapted from [23].

In the following, for each conduction mechanism taken into account, the analytical model used to fit the experimental data is reported.

- **Schottky Emission** :
  Schottky emission, also known as thermionic emission, consists in the liberation of electrons from an electrode by increasing its temperature. In a Metal-Insulator-Semiconductor (MIS) structure, if the electrons in the metal are thermally excited, they can obtain enough energy to overcome the energy barrier at the metal-dielectric interface, going directly in the dielectric. This conduction mechanism is usually relevant at high temperature.
  The implemented analytical model to describe this mechanism is the following [23]:

$$J = A^* \cdot T^2 \cdot \exp\left\{ \frac{-q(\phi_B - \sqrt{\frac{qE}{4\pi\epsilon_r\epsilon_0}})}{k_B T} \right\} \tag{2.1}$$

$$A^* = \frac{4\pi q k^2 m^*}{h^3} = \frac{120 m^*}{m_0} \tag{2.2}$$

  where J is the current density, A^* the effective Richardson constant, T the absolute temperature, $q\phi_B$ the Schottky barrier height, hence the conduction band offset, E the electric field across the dielectric, k_B the Boltzmann's constant, h the Planck's constant, ϵ_0 the vacuum permittivity, ϵ_r the dynamic dielectric constant, q the elementary electronic charge, m_0 the free electron mass and m^* the effective electron mass in the dielectric. Rearranging the previous equation, it is possible to show that the relation of $\log\left(\frac{J}{T^2}\right)$ as a function of $\sqrt{E}$, is linear.

$$\log\left(\frac{J}{T^2}\right) = \log(A^*) - \frac{q\phi_B}{k_B T} + \frac{\sqrt{\frac{q^3}{4\pi\epsilon_r\epsilon_0}}}{k_B T} \cdot \sqrt{E} \tag{2.3}$$

$$intercept = \log(A^*) - \frac{q\phi_B}{k_B} \cdot \frac{1}{T} \tag{2.4}$$

$$slope = \frac{\sqrt{\frac{q^3}{4\pi\epsilon_r\epsilon_0}}}{k_B} \cdot \frac{1}{T} \tag{2.5}$$

Plotting both the intercept and the slope of equation (2.3) as a function of $\frac{1}{T}$, known as Arrhenius plot, as in equations (2.4) and (2.5), it is possible to get further linear trends that allow to extract the A^* constant and $q\phi_B$ from the former, and the dynamic dielectric constant ϵ_r from the latter.

- **Modified Schottky Emission** :
  If the electronic mean free path l in the dielectric is less than the dielectric thickness t_d, standard thermionic emission must be modified as follows [24].

$$J = \alpha T^{\frac{3}{2}} E\mu(\frac{m^*}{m_0})^{\frac{3}{2}} \cdot \exp\left\{\frac{-q(\phi_B - \sqrt{\frac{qE}{4\pi\epsilon_r\epsilon_0}})}{k_B T}\right\} \tag{2.6}$$

where $\alpha = 3 \cdot 10^{-4} As/cm^3 K^{\frac{3}{2}}$ and the other notations are the same as defined before [23].
As for standard thermionic emission, rearranging the previous equation, it is possible to show that the plot of $\log\left(\frac{J}{T^{3/2}.E}\right)$ as a function of $\sqrt{E}$, is linear if this mechanism takes part to the conduction.

$$\log\left(\frac{J}{T^{3/2} \cdot E}\right) = \log\left(\alpha\mu(\frac{m^*}{m_0})^{\frac{3}{2}}\right) - \frac{q\phi_B}{k_B T} + \frac{\sqrt{\frac{q^3}{4\pi\epsilon_r\epsilon_0}}}{k_B T} \cdot \sqrt{E} \tag{2.7}$$

$$intercept = \log\left(\alpha\mu(\frac{m^*}{m_0})^{\frac{3}{2}}\right) - \frac{q\phi_B}{k_B} \cdot \frac{1}{T} \tag{2.8}$$

$$slope = \frac{\sqrt{\frac{q^3}{4\pi\epsilon_r\epsilon_0}}}{k_B} \cdot \frac{1}{T} \tag{2.9}$$

From Arrhenius plot of the intercept and the slope, it is possible to get $\mu(\frac{m^*}{m_0})^{\frac{3}{2}}$ and $q\phi_B$ from the former, and the dynamic dielectric constant ϵ_r from the latter.

- **Fowler-Nordheim Tunneling** :
  Fowler-Nordheim Tunneling, also known as field-emission, is a quantum mechanical effect, that may occur in a thin dielectric layer [23]. In particular, considering a metal-insulator-semiconductor structure, if the applied electric field is large enough, the electron wave function may penetrate through the triangular potential barrier into the conduction band of the dielectric, causing a tunneling current.
  The implemented analytical model to describe this mechanism is the following [23]:

$$J = \frac{q^3 E^2}{8\pi h q \phi_B} \cdot \exp\left\{ \frac{-8\pi (2qm_T^*)^{\frac{1}{2}}}{3hE} \phi_B^{\frac{3}{2}} \right\} \tag{2.10}$$

  where m_T^* is the tunneling effective mass in the dielectric and the other notations are the same as defined before.
  In Fowler-Nordheim tunneling, the relation between $\log\left(\frac{J}{E^2}\right)$ and $-\frac{1}{E}$ is linear.

$$\log\left(\frac{J}{E^2}\right) = \log\left(\frac{q^3}{8\pi h q \phi_B}\right) + \frac{8\pi (2qm_T^*)^{\frac{1}{2}}}{3h} \cdot \phi_B^{\frac{3}{2}} \cdot (-\frac{1}{E}) \tag{2.11}$$

$$intercept = \log\left(\frac{q^3}{8\pi h q \phi_B}\right) \tag{2.12}$$

$$slope = \frac{8\pi (2qm_T^*)^{\frac{1}{2}}}{3h} \cdot \phi_B^{\frac{3}{2}} \tag{2.13}$$

  Both the intercept and the slope of equation (2.11) are temperature independent. Assuming to know the tunneling effective mass in dielectric m_T^*, it is possible to extract the Schottky barrier height $q\phi_B$ from equation (2.12), and compare it with the one extracted from equation (2.13).

- **Thermionic-Field Emission** :
  Thermionic-field emission takes place intermediately between field-emission and thermionic emission. In this condition, tunneling electrons have an energy between the Fermi level of metal and the conduction band edge of dielectric. The difference between Thermionic-Emission, Thermionic-Field Emission and Field-Emission is shown in figure 2.4 [23].

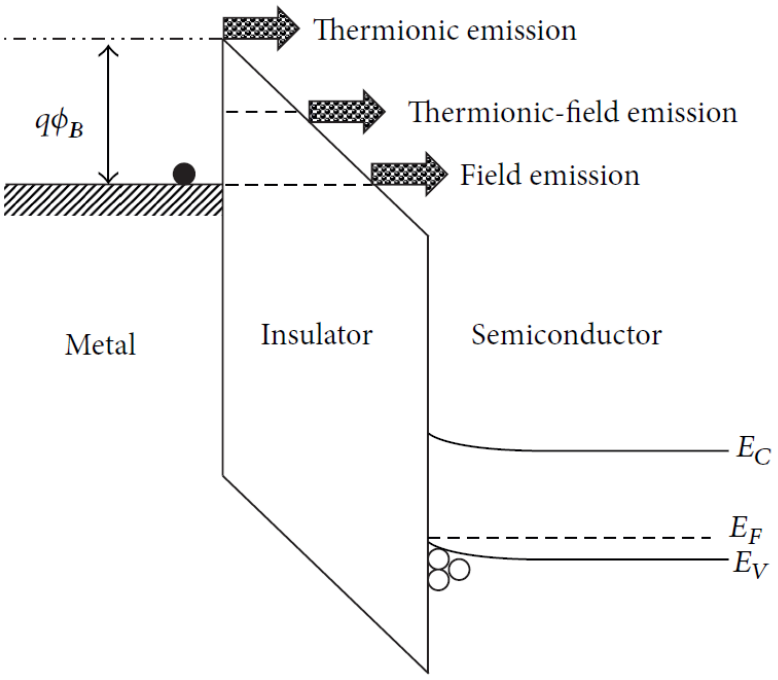


Figure 2.4: Thermionic-Emission is fully thermally activated, the Field-Emission is fully field activated, while the Thermionic-Field is activated by both the temperature and the field. Taken from [23].

The analytical model used to describe this conduction mechanism is [23]:

$$J = \frac{q^2\sqrt{m}(k_BT)^{\frac{1}{2}}E}{8\hbar^2\pi^{\frac{5}{2}}} \cdot \exp\left\{\frac{-q\phi_B}{k_BT}\right\} \cdot \exp\left\{\frac{\hbar^2q^2E^2}{24m(k_BT)^3}\right\} \tag{2.14}$$

In this case, the relation between $\log\left(\frac{J}{E\cdot T^{\frac{1}{2}}}\right)$ and $\frac{E^2}{T^3}$ is linear if this mechanism takes part to the conduction.

$$\log\left(\frac{J}{ET^{\frac{1}{2}}}\right) = \log\left(\frac{q^2\sqrt{m}k_B^{\frac{1}{2}}}{8\hbar^2\pi^{\frac{5}{2}}}\right) - \frac{q\phi_B}{k_BT} + \frac{\hbar^2q^2}{24m(k_B)^3} \cdot \frac{E^2}{T^3} \tag{2.15}$$

$$intercept = \log\left(\frac{q^2\sqrt{m}k_B^{\frac{1}{2}}}{8\hbar^2\pi^{\frac{5}{2}}}\right) - \frac{q\phi_B}{k_B} \cdot \frac{1}{T} \tag{2.16}$$

$$slope = \frac{\hbar^2q^2}{24m(k_B)^3} \tag{2.17}$$

From the Arrhenius plot of the intercept, the electron mass m and the conduction band offset $q\phi_B$ can be obtained. While from equation 2.17, it is possible to extract again the electron mass, that should be equal to the one previously extracted if the model is consistent.

- **Poole-Frenkel Emission** :
  Poole-Frenkel emission consists in the thermal excitation of electrons in dielectric traps, which may be emitted directly into the conduction band of the dielectric. In fact, considering an electron in a trap state, since its Coulomb potential energy can be decreased by an applied electric field, the probability that it can be thermally excited out up to the conduction band, is proportional to the applied electric field [23]. In figure 2.5 the schematic band diagram of Poole-Frenkel emission in metal-insulator-semiconductor structure, is shown.

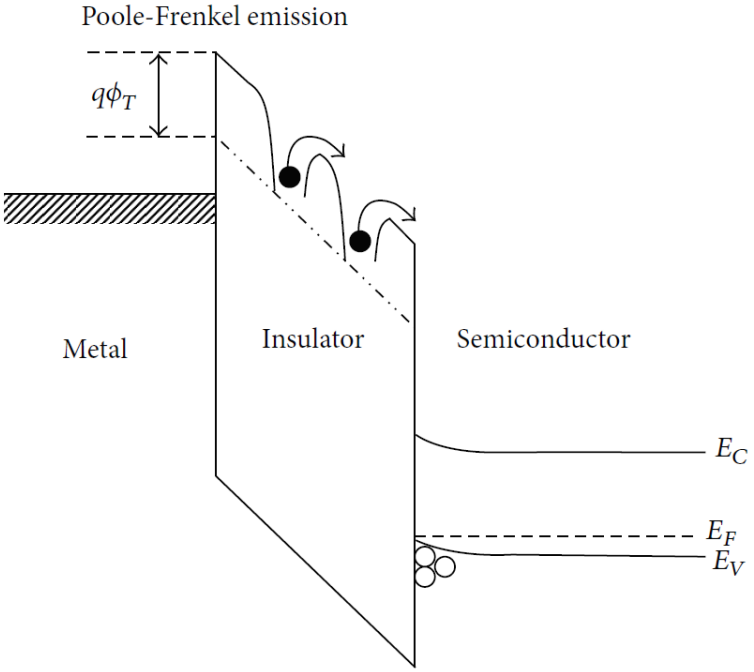


Figure 2.5: Schematic energy band diagram of Poole-Frenkel emission in MIS structure. Taken from [23].

This mechanism is analytically described by the following model [23]:

$$J = q\mu N_C E \exp\left\{ \frac{-q(\phi_T - \sqrt{\frac{qE}{\pi\epsilon_r\epsilon_0}})}{k_B T} \right\} \tag{2.18}$$

where N_C is the effective density of states in the conduction band, $q\phi_T$ is the energy barrier between the trap states and the bottom of the conduction band in the dielectric and the other notations are the same as defined before. In Poole-Frenkel emission, the relation between $\log\left(\frac{J}{E}\right)$ and $\sqrt{E}$ is linear.

$$\log\left(\frac{J}{E}\right) = \log\left(q\mu N_C\right) - \frac{q\phi_T}{k_B T} + \frac{\sqrt{\frac{q^3}{\pi\epsilon_r\epsilon_0}}}{k_B T}\sqrt{E} \tag{2.19}$$

$$intercept = \log\left(q\mu N_C\right) - \frac{q\phi_T}{k_B} \cdot \frac{1}{T} \tag{2.20}$$

$$slope = \frac{\sqrt{\frac{q^3}{\pi\epsilon_r\epsilon_0}}}{k_B} \cdot \frac{1}{T} \tag{2.21}$$

From intercept Arrhenius plot, hence equation 2.20, the product $\mu \cdot N_C$ and the height of the trap's barrier can be extracted, while from equation 2.21 the dynamic dielectric constant can be predicted.

- **Hopping** :
  Hopping conduction consists of the tunneling of trapped electrons, from one trap site to another in a dielectric film. Figure 2.6 shows a schematic representation of band diagram where hopping conduction occurs.

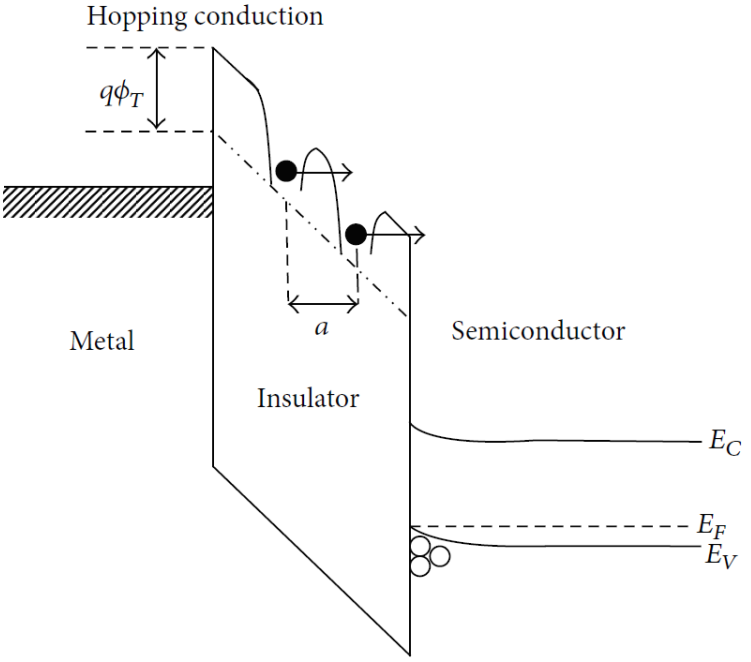


Figure 2.6: Schematic energy band diagram of hopping emission in MIS structure. Taken from [23].

This mechanism is modeled by the following analytical expression [23]:

$$J = qan\nu \exp\left\{\frac{qaE}{k_BT} - \frac{E_a}{k_BT}\right\} \tag{2.22}$$

where a is the mean hopping distance, ν is the frequency of thermal vibration of electrons at trap sites, and E_a is the activation energy, which means the energy level from the trap states to the bottom of conduction band, while all the other terms are as defined above.
The relation between $\log(J)$ and E is linear when hopping conduction occurs.

$$\log(J) = \log(qan\nu) - \frac{E_a}{k_B T} + \frac{qa}{k_B T} \cdot E \tag{2.23}$$

$$intercept = \log(qan\nu) - \frac{E_a}{k_B} \cdot \frac{1}{T} \tag{2.24}$$

$$slope = \frac{qa}{k_B} \cdot \frac{1}{T} \tag{2.25}$$

The activation energy E_a and the product $n \cdot \nu$ can be deduced from equation 2.24, while the hopping distance from equation 2.25.

- **Ohmic Conduction** :
  Ohmic conduction is due to the drift of mobile electrons in the conduction band and holes in the valence band. Although the energy band gap of dielectrics is usually large, at $T \neq 0\,\mathrm{K}$, there is a small amount of electrons in the conduction band and holes in the valence band, usually generated by thermal excitation, that may be drifted by the application of an external electric field, resulting in a net conduction current.
  The analytical model used to model this conduction mechanism is [23]:

$$J = \sigma \cdot E = nq\mu \cdot E = q\mu N_C \exp\left\{-\frac{E_C - E_F}{k_B T}\right\} \cdot E \tag{2.26}$$

$$n = N_C \exp\left\{-\frac{E_C - E_F}{k_B T}\right\} \tag{2.27}$$

  where σ is the electrical conductivity, n is the electron density in conduction band and $E_C - E_F$ is the energy distance between the bottom of the conduction band and the Fermi level.
  The relation between $\log(J)$ and E is linear when ohmic conduction occurs.

$$\log(J) = \log(q\mu N_C) - \frac{(E_C - E_F)}{k_B T} + \ln E \tag{2.28}$$

$$intercept = \log(q\mu N_C) - \frac{(E_C - E_F)}{k_B} \cdot \frac{1}{T} \tag{2.29}$$

$$slope = 1 \tag{2.30}$$

  From intercept Arrhenius plot, hence equation 2.29, the product μN_C and the energy distance $E_C - E_F$ can be extracted, while the slope is constant to 1.

- **Space Charge Limited Conduction** :
  Space-charge-limited conduction mechanism is similar to the transport conduction of electrons in a vacuum diode, but occurs in solid material. To explain this mechanism, a structure with a dielectric solid thin film surrounded by two electrodes, can be considered. By applying an external electric field, electrons are injected from the metal into conduction band of the dielectric, forming a space-charge region at metal-dielectric interface, which influences the transport.
  In low voltage regime, when the injected carrier density is lower than the thermally generated free carrier density, Ohm's law is fulfilled. On the contrary, in strong injection regime, hence for V $\geq$V$_{tr}$, the injected carrier density is higher than the thermally generated one, hence the transition from Ohm's law to space-charge limited region, takes place. During this regime, the traps in the dielectric are filled up, and a space charge appears. For V = V_{TFL}, all traps are filled up and the subsequently injected carriers are free to move in the dielectric film, hence the conduction is ruled by Child law [23].
  This conduction mechanism can be analytically described by the following model [23]:

$$J_{Ohm} = qn_0\mu E \tag{2.31}$$

$$J_{TFL} = \frac{9}{8}\mu\epsilon\theta\frac{V^2}{d^3} \tag{2.32}$$

$$J_{Child} = \frac{9}{8}\mu\epsilon\frac{V^2}{d^3} \tag{2.33}$$

$$V_{tr} = \frac{9}{8}\frac{qn_0d^2}{\epsilon\theta} \tag{2.34}$$

$$V_{TFL} = \frac{qN_td^2}{2\epsilon} \tag{2.35}$$

$$\theta = \frac{N_C}{g_nN_t}\exp\left\{\frac{E_t - E_C}{k_BT}\right\} \tag{2.36}$$

  where n_0 is the concentration of the free charge carriers at thermal equilibrium, V is the applied voltage, d is the thickness of thin film, θ is the ratio of the free carrier density to total carrier (free and trapped) density, g_n is the degeneracy of the energy state in the conduction band and E_t and N_t are respectively the trap energy level and trap density.
  In figure 2.7 the expected relation between $\log(J)$ and $\log(V)$ when space charge limited conduction occurs, is shown.

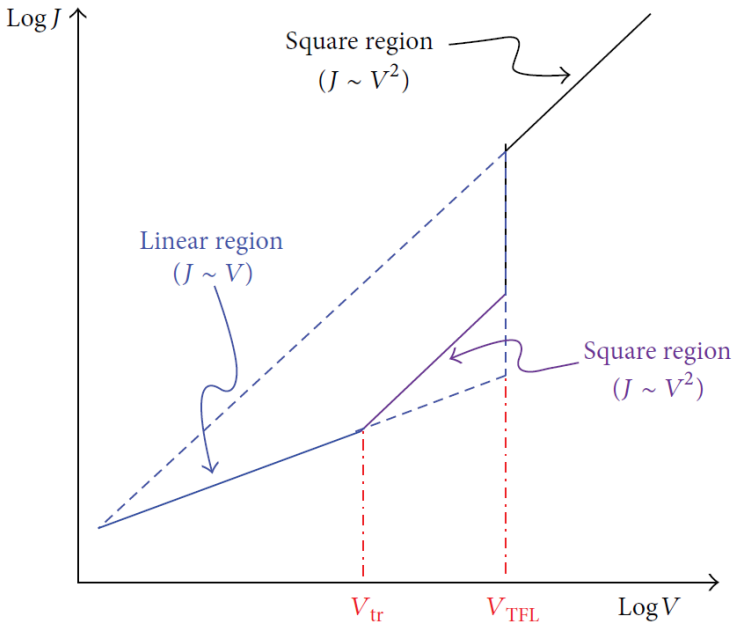


Figure 2.7: $\log(J)$ versus $\log(V)$ when SCLC mechanism occurs. Taken from [23].

When this mechanism occurs, parameters such as the trap density N_t, the energy distance $E_t - E_C$ as well as the degeneracy factor g_n, can be extracted.

All the previously detailed transport mechanisms are used to fit experimental data and extract physical parameters. If a specific conduction mechanism takes part to the overall conduction, the expected linear trends well describe the experimental data and physical parameters extracted by the fit are meaningful compared to what expected in literature.

2.1.2 Circular Transfer Length Method

Circular Transfer Length Method (CTLM) is a characterization technique that allows to determine the contact resistance R_c and specific contact resistance ρ_c through the linear relationship between the resistance and the gap spacing between multiple contacts [25]. The determination of the contact resistance R_c allows to get information about the interface resistance between the metal and the semiconductor. In addition to R_c, a useful quantity independent from the contact area is the specific contact resistivity $\rho_c = \frac{\partial V}{\partial J}\Big|_{V=0}$ [26].

In figure 2.8, a typical TLM test structure is reported. It consists of several metallic contacts with an increasing gap in between. If it is assumed that all the contacts have the same R_c, the increasing of the total resistance with gap distance d can be attributed to the increasing resistance through the semiconductor. However, considering a real contact, when the current flows from semiconductor into the metal, the highest current density is reached at the contacts edge, since the current flows through the less resistive path possible. This causes an exponential drop of the

current density J from the contact edge on, resulting in a reduction of the effective contact area. The transfer length L_T is defined as the distance from the edge at which the current density J has dropped by a factor e^{-1}.

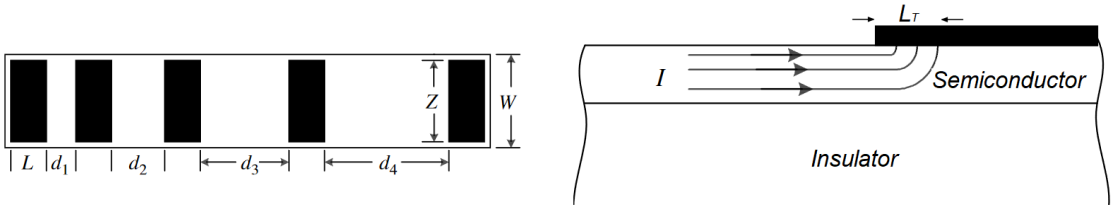


Figure 2.8: On the left, a typical TLM layout with a series of square contacts, while on the right a graphical explanation of transfer length L_T is shown. Redrawn from [26].

However, linear TLM technique tends to suffer from the fact that currents from one contact to another may spread due to current crowding [25]. For this reason, Circular TLM structures, whose typical layout is represented in figure 2.9, are used. They consist of an inner contact of radius r, a ring shaped gap of width d and a surrounding contact.

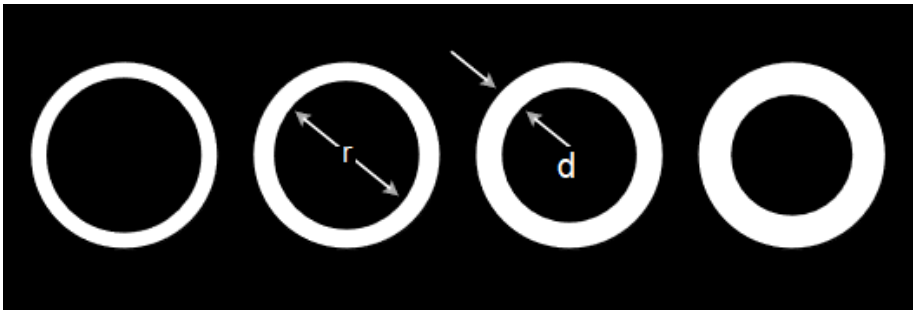


Figure 2.9: Illustration of a circular TLM layout. Redrawn from [26].

By forcing a constant current through the structure, a voltage drop is induced over the two metal contacts and the intermediate semiconductor layer. In order to have more accurate measurements, 4 probes are used, two to force the current and two to sense the voltage. The total resistance R_T can be expressed as follows [25]

$$R_T = \frac{R_{sh}}{2\pi r} \cdot (d + 2L_T)) \cdot C \tag{2.37}$$

$$C = \frac{r}{d} \cdot \ln\left(1 + \frac{d}{r}\right) \tag{2.38}$$

where C is the correction factor to compensate for the difference between the linear and circular TLM layouts. Plotting the total resistance as a function of gap spacing yields a typical result as the one shown in figure 2.10.

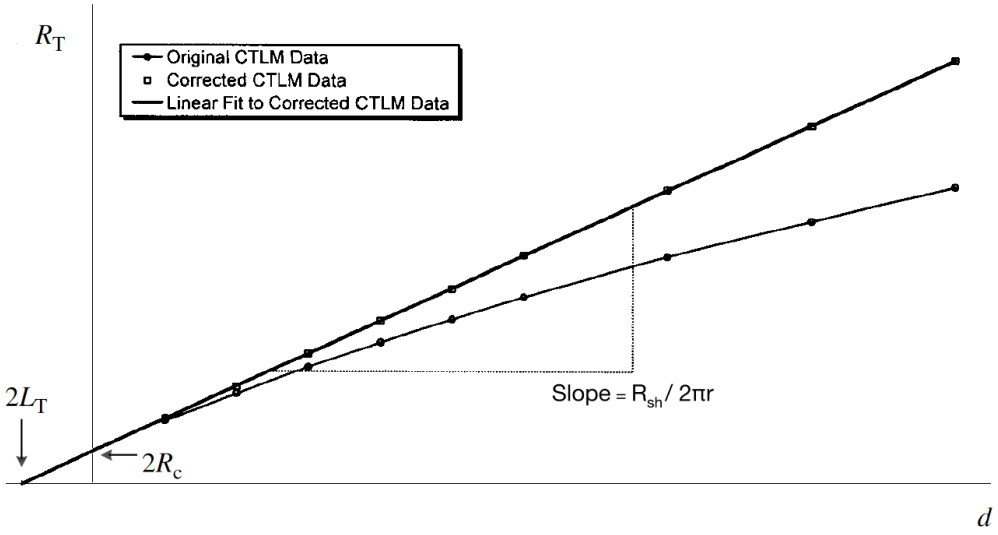


Figure 2.10: Total resistance plotted as a function of gap spacing before and after applying the correction factors. Adapted from [25].

Considering the linear fit of corrected CTLM data, the intercept with y-axes provides $2R_c$, which is the contact resistance, while d = -$2L_T$ is defined as the point at which the total resistance is zero. From equation 2.37, the sheet resistance R_{sh} of the semiconductor can be computed as follows:

$$R_{sh} = \frac{\partial R_T}{\partial d} \cdot 2\pi r \tag{2.39}$$

Finally, knowing the transfer length L_T and the contact resistance R_c, the specific contact resistivity can be computed as follows [26]:

$$\rho_c = R_c \cdot A_{c,eff} \simeq R_c \cdot 2\pi r L_T \tag{2.40}$$

2.2 Grazing Incidence X-Ray Diffraction

The structural analysis of crucial materials for both planar and fin-based FeFET, such as HZO and WO_x, has been performed using Grazing Incidence X-Ray Diffraction, known as GIXRD. This characterization method has been preferred to standard XRD in Bragg-Brentano configuration, since the investigated materials are in the form of thin films.

The working principle of GIXRD consists in irradiating the sample with an X-ray beam, impinging at a low incidence angle with respect to the sample surface to prevent the contribution of the thin film to be hidden by the contribution of the substrate. As consequence of this irradiation, the crystal atoms are excited and relax emitting spherical radiation waves, that at distance very long with respect to X-ray wavelength, can be approximated as plane waves. These plane waves interfere each other, and give rise to constructive interference according to Bragg law:

$$2 \cdot d \cdot \sin\theta = n\lambda \tag{2.41}$$

where θ is the angle of incidence of X-rays respect to the crystal planes, λ the wavelength of X-ray radiation, d the distance between crystalline planes and n the diffraction order.

The X-ray diffractometer system used in this work is the Bruker D8 Discover, equipped with a rotating Cu anode generator, and controlled by the measurement software *diffrac.measurement center*.

2.3 Scanning Electron Microscope

Scanning Electron Microscope (SEM) is a versatile characterization tool that exploits a focused electron beam rather than light, to perform high resolution imaging. The limit of resolution in microscopy is mostly determined by the wavelength λ of the electromagnetic radiation used to scan the surface of the sample. For this reason, accelerated electrons are suitable for high resolution imaging, since their relative tunable wavelength is of the order of picometer [27].

In an SEM, the electron beam, generated by an electron gun, is modulated and focused, by electromagnetic lenses and apertures, onto the specimen surface, as depicted in figure 2.11. Both the generation and the shaping of electron beam occurs in high-vacuum environment, to allow electron travelling without scattering. Incident electron beam penetrates into the sample for some distance before interacting with specimen atoms. The region of interaction between primary electron and the atoms of the sample is called volume of interaction, shown in figure 2.11.
From this region a variety of signals bringing several kind of information, are produced.

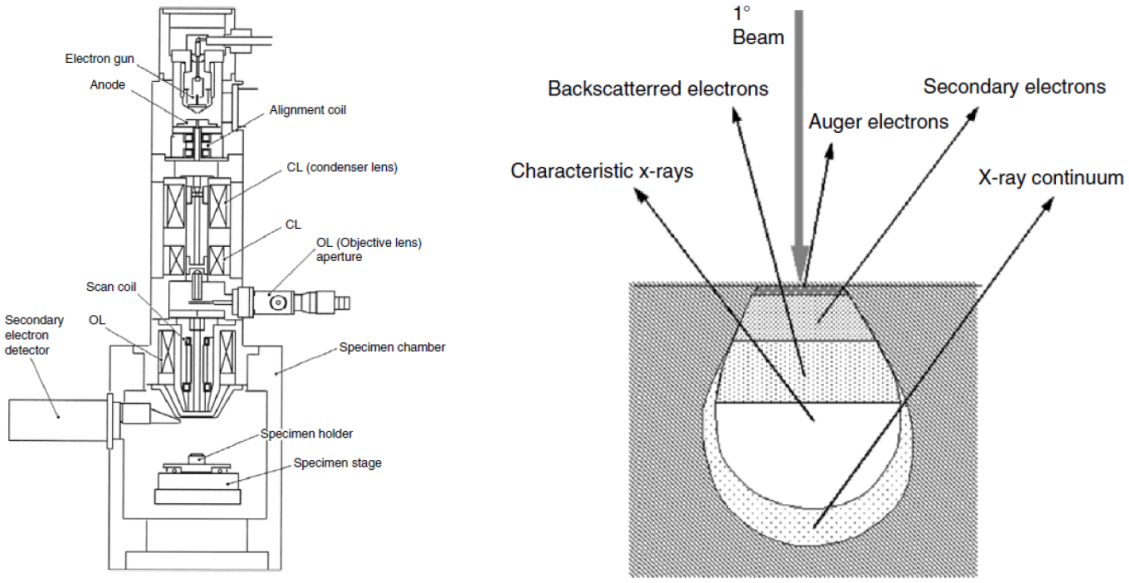


Figure 2.11: On the left the schematic of an SEM setup, while on the right the volume of interaction and the different types of signals for imaging. Taken from [27].

The incident primary electron beam mostly produces low-energy ($< 50\,\text{eV}$) secondary electrons (SE) from surface, which give topographic information and high surface resolution, and back-scattered electrons (BSE) from a deeper region, which bring information about sample composition. Other signals, such as X-rays and Auger electrons, are also generated during the interaction between specimen and primary electrons and may be used for microstructure analysis.

Scanning Electron Microscope can resolve features down to few nm size [28], however, since the imaging is performed with an electron beam, the sample outer layer is required to be conductive. If this is not the case, a charge density accumulates at surface of the specimen, changing the trajectory of electrons escaping from sample and causing distorted images. This phenomenon is known as charge-up effect.

The Scanning Electron Microscope system used in this work is the FEI Helios NanoLab 450S.

2.4 Focused Ion Beam

The working principle of a Focused Ion Beam (FIB) is almost identical to that of a Scanning Electron Microscope. However, in FIB, the accelerated and focused charged particles are not electrons, but ions. Since the ions are much more massive than electrons, FIB imaging is usually destructive, which allows the nanofabrication of specimen cross-sections and lamellaes.
In most commercially available FIB instrument, *Ga* ions are used to form the beam. The primary ion beam, after being properly shaped, is focused onto the specimen surface, allowing precise and localized physical etching. In addition to this, it is also possible to inject gas into the chamber to enhance the etching process, or perform an ion beam activated deposition [29].

Most modern FIB tools supplement the FIB column with an additional SEM column so that the tool becomes a "dual-beam" platform, named FIB–SEM [29]. This versatile setup allows not only the imaging of specimen, with the electron beam, but also the deposition and etching of material on a length scale of few nm. In addition, this tool becomes necessary when the area of investigation is not a surface but a cross-section. In this case, first the ion beam is used to etch the desired region, uncovering the cross-section, then the sample is tilted and with the electron beam the cross-section is imaged.

The Focused Ion Beam system used in this work is the FEI Helios NanoLab 450S.

Chapter 3

Processing Methods

In this section, all the processing tools used for the fabrication of fin based FeFETs are reported. In particular, since the process-flow of FinFeFETs consists of about 60 steps, several microelectronic fabrication techniques have been performed, such as Chemical and Physical Vapor Depositions, known as CVDs and PVDs, for material growth, as well as physical and chemical dry etching processes to faithfully transfer lithographically defined photoresist patterns into underlying layers. In addition, thermally activated processes, such as rapid thermal annealing and millisecond flash lamp annealing, have been used to crystallize the materials into the desired phase. For the transferring of the desired pattern directly in the substrate, both optical and e-beam lithography, and both additive and subtractive approaches, are used.
All these methods are described in the following sections.

3.1 Photolithography

The photolithography is the process through which device patterns are transferred from layout, such as GDSII file, to real substrates. To do so, a photosensitive chemical liquid polymer, called photoresist, is spinned onto the material to be patterned, and then specific regions of that are exposed by an electromagnetic radiation. According to the polarity of the photoresist, the polymer chains of the exposed parts are broken in positive resist, hence become more soluble in the developing solution, or cross linked in negative resist, becoming less soluble in the developing solution.

After the exposure of the photo-sensitive resist, the latter is developed in specific chemical solution. This allows to transfer the desired pattern into the resist. At this step, to pattern the desired material, the two following strategies can be exploited:

- **Subtractive photolithography**
  In this case, resist development is followed by a subtracting technique, thus etching, of the desired material, already deposited before resist spinning. During the etching, the patterned resist is used as protecting mask, and just after, it is stripped using a wet or dry subtractive process.

- **Additive photolithography (Lift-off)**
  In this approach, resist development is followed by an additive technique, thus a deposition, of the desired material on top of the patterned resist. After the deposition, the resist, with on top the deposited target material, is lifted-off using chemical solvent such as DMSO or acetone. After the lift-off, the target material remains only in the regions where it had a direct contact with the substrate. This approach is usually used for metal patterning, for which direct etching would be particularly challenging and would have undesirable effects on the layer below.

In figure 3.1, a schematic view of both subtractive photolithography and lift off process, is provided.

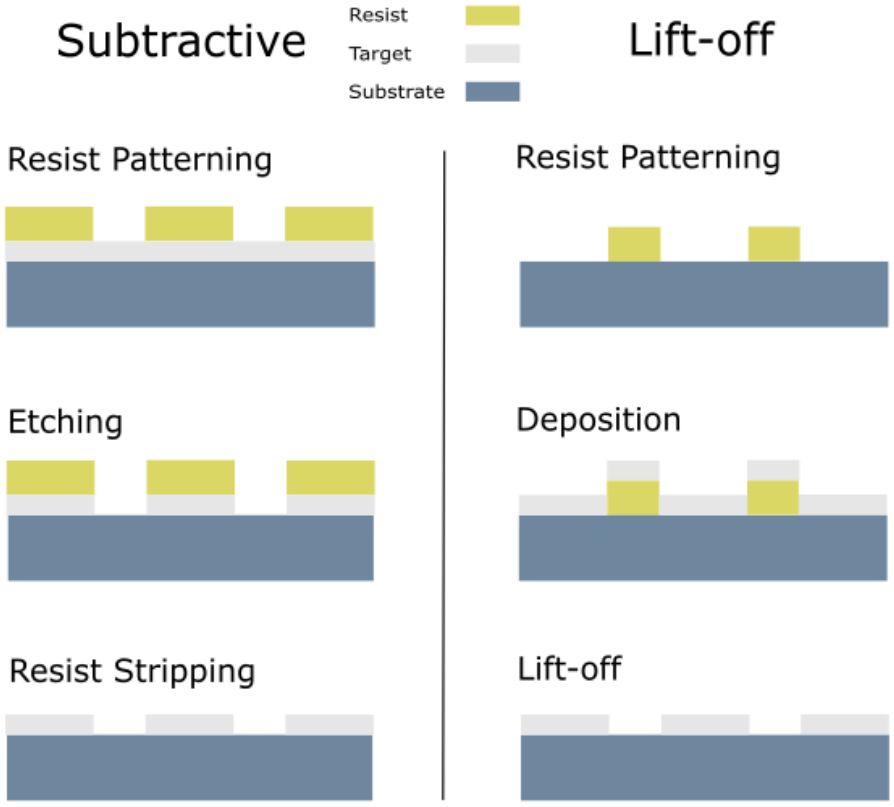


Figure 3.1: Illustration of steps in subtractive photolithography and lift-off patterning.

Both the previously described patterning techniques can achieve the same result. Subtractive photolithography can achieve higher resolution, thus it is preferred when critical resolution is needed. On the other hand, lift-off is a meant for complementary application, and it is usually used when direct etching is particularly challenging. The main limitations of lift-off are:

- **Retention**
  It occurs when unwanted parts of the target material, such as metal, remain on the sample's surface. This can be due to a non proper dissolution of the resist in the chemical solvent.

- **Ears**
  During the deposition, target material can cover also the sidewalls of the resist, thus vertical structures standing upwards from the surface, known as *lift-off ears*, can be formed.
- **Redeposition**
  During lift-off process, some particles of target material can redeposit onto the sample's surface, at a random location.

To mitigate these problems, especially the *lift-off ears* one, when lift-off is exploited for pattern transferring, a resist double layer is usually used instead of a single one. In particular, if the photo-sensitivity of the bottom resist is higher than that of the top layer, for the same dose it will develop on a larger surface, which will result in a undercut profile. When a directional metallization method is used, such as evaporation, this profile will limit the deposition of the metal on the edges of the features and will result in a more precise patterning.

3.1.1 Electron-beam Lithography

The resolution of a photolithographic step depends on both the specific resist features and the wavelength λ of the radiation used. In fact, the resolution limit of an optical system, defined as the minimum distance by which two structures can be separated and still appear as two distinct objects, is mostly limited by diffraction phenomenon. In particular, the minimum critical dimension CD_{min} in diffraction-limited optical system, is dictated by Abbe diffraction law:

$$CD_{min} = \frac{\lambda}{2 \cdot n \sin\theta} \tag{3.1}$$

where λ is the wavelength of the radiation used, n is the refractive index of the medium in which the radiation travel before reaching the resist and θ is half collecting angle. For this reason, decreasing the wavelength of the impinging radiation allows to increase the resolution of the process.

Electron-beam Lithography exploits a focused beam of electrons to directly expose e-beam sensitive resists, such as *Poly (methyl methacrylate)* (PMMA). Using an electron radiation allows to overcome the previously described diffraction limit, pushing it in sub-nm range. In fact, since electrons have wave-like properties with equivalent wavelengths λ between 0.2 Å and 0.5 Å, the diffraction becomes totally negligible, and the exposure system is not anymore limited by that, thus it is possible to transfer patterns with sub 10 nm resolution. However, since e-beam lithography is maskless, it has low throughput, limiting its usage to photomask fabrication, low-volume production of semiconductor devices, and research and development.

The Electron-beam Lithography system used in this work is the Vistec EBPG 5200+.

3.1.2 Laser Lithography

This exposure method exploits a laser radiation to directly expose a photo-sensitive resists. The primary usage of the direct laser writer is the fabrication of hard masks for optical lithography, suitable for high-volume production. Direct laser writing can also be used on individual chips, as e-beam lithography. The critical dimension, thus the resolution as well as the alignment precision, of this exposure technique is about 1 µm, thus lower if compared to e-beam lithography, since laser radiation has higher wavelength than electrons. However, this tool is meant for complementary applications with respect to e-beam lithography, since the exposure times, as well as the tool's cost, are considerably lower.

The laser writer system used in this work is the Heidelberg DWL 2000, with 413 nm diode laser as radiation source.

3.2 Deposition methods

In the following sub-sections, several tools to deposit thin film of semiconductors, oxides and metals are described. In general, the deposition techniques can be categorised into two main groups, according to the material growth. If the material synthesis occurs as consequence of chemical reaction of vapor phase reactants, it is called chemical vapour deposition CVD. Otherwise, if material growth is achieved by physical deposition of it onto the substrate, it is called physical vapour deposition PVD.

3.2.1 Atomic Layer Deposition

Atomic layer deposition (ALD) is a process that allows the deposition of very thin film with atomic scale precision. This technique, as all the Chemical Vapor Depositions (CVDs), exploits gaseous precursors. However, the peculiarity is that ALD breaks the CVD reaction into two half reactions keeping the precursor materials separated during the reaction. Atomic layer deposition can be used to deposit metallic nitrides, such as TiN as well as higk κ dielectric, such as Al_2O_3 and HfO_2.

The ALD process, performed in a sealed reactor, starts with the pulsing of the first processing gas into the reaction chamber, which reacts with substrate surface until it is totally coated. Since the precursor gas does not react with itself, the reaction terminates with the formation of a monolayer. After this step, the excess of processing gas, as well as the byproducts, are pumped away during the first purge step. The second processing gas then flows into the chamber and reacts with the previously adsorbed molecules to form a molecular layer of compound material. Finally, a second purge step ensures the removal of all volatile byproducts. In figure 3.2, a typical ALD cycle is shown.

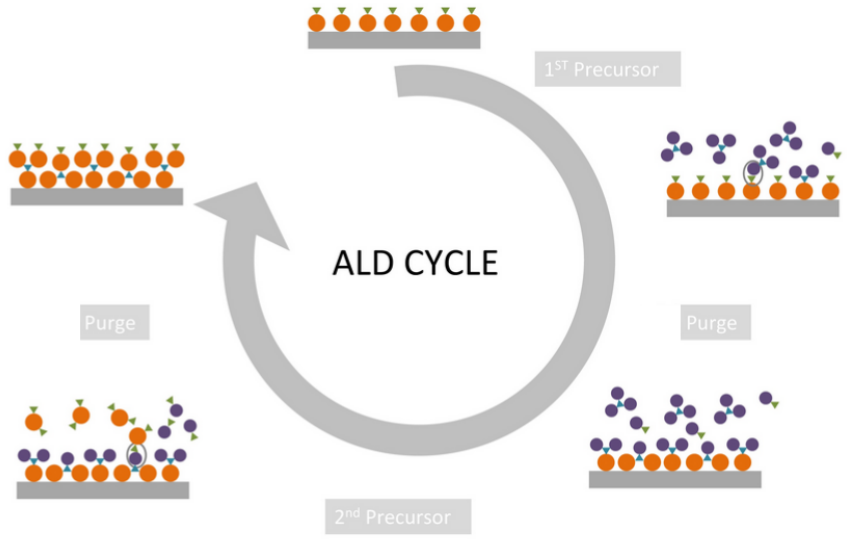


Figure 3.2: Illustration of an atomic layer deposition cycle. Redrawn from [30].

Since ALD film growth is self limited and produces exactly one monolayer, the thickness of the resulting film may be precisely controlled by tuning the number of deposition cycles.

Most of ALD processes occurs in a temperature window between 50 °C and 500 °C, depending on precursor volatility and reactivity [30]. In order to lower the processing temperature, as well as the deposition time and the contamination content, Plasma Enhanced ALD can be used. This technique, which is a further advancement of standard ALD, exploits a pulse step with O_2 or Ar RF-plasma, to create the necessary chemical reactions with lower activation energy.

The plasma enhanced atomic layer deposition system used in this work is the Oxford FlexAL.

3.2.2 Electron beam evaporation

Evaporation is a Physical Vapor Deposition (PVD) technique, commonly used to deposit thin films. The physical principle consists in heating the desired material above its melting point, so that it starts to evaporate. During the process, since the processing chamber is under high vacuum condition, the mean free path of the evaporated particles is such that they move in straight-line trajectories without scattering events, condensing onto the substrate. In electron beam evaporation, an electron beam, accelerated to high kinetic energy, is focused onto the target material, which is stored in proper crucible called liners. Since the heating is localized, only the target material is evaporated, reducing the contaminations from the crucible. In addition, due to the small area of the source, the emission profile follows a cosine law [31], which makes evaporation a technique characterized by poor conformality and high directionality. In fact, with this technique it is difficult to get a continuous

coverage on structures with a strong topography, but it is an ideal method for lift-off processes. A wide number of metals can be evaporated, depending on their melting point.

The electron beam evaporation system used in this work is the Evatec BAK501 LL.

3.2.3 DC-Sputtering

Sputtering is a Physical Vapor Deposition (PVD) technique, which exploits a plasma, usually Argon based, to bombard the material to be deposited, called target, and dislodge its atoms which are then deposited onto the sample to form thin film. Sputtering chamber is characterized by two electrodes, with the target material used as cathode and the sample as anode, and gas inlets to inject and pump the plasma gas. This technique is so versatile that can be used both to deposit material as well as for physical etching purpose.
If Sputtering is used as deposition technique, after the injection of the inert gas, usually Ar, the anode is grounded, and a negative voltage is applied at the cathode to generate a DC-plasma Argon based. The applied electric field accelerates the free electrons in the initial neutral Argon atmosphere, which elastically collide with Ar atoms, causing an ionization of the latters. Ionized Argon ions are used to sputter the target, whose atoms arrive at the sample mostly as neutral atoms. This technique is more versatile than evaporation, since metals, but also oxides or nitrides, can be sputtered, but it is characterized by a poor directionality.
Reversing the voltage sign between cathode and anode, it is possible to physically etch the sample's surface with Ar plasma. This mode of operation of Sputter, called Inverse Sputter Etching (ISE) allows to clean sample's surface, which may have been previously oxidized or covered by impurities, before performing the actual deposition.

The sputtering system used in this work is the vonArdenne CS 320 S.

3.2.4 Plasma Enhanced Chemical Vapor Deposition

Plasma Enhanced Chemical Vapor Deposition (PECVD) is a technique used to deposit thin film starting from gaseous precursors. The process consists in the transport of precursor molecules into the reactor chamber, the diffusion and adsorption of the latters on the surface of the sample, the decomposition as well as the incorporation of them into solid films and finally the desorption of volatile byproducts. In PECVD, the energy to break gaseous precursors into reactive species for deposition, is provided through a plasma. First the chamber is evacuated and filled with precursor gases, then an RF signal is applied to generate a plasma, which at steady-state, mainly consists of electrons, ionized molecules and free radicals. Since free radicals are electrically neutral species, characterized by incomplete bonding and high reactivity, the deposition process occurs at lower temperature with respect to non-plasma based CVDs, making PECVD suitable for back end of line (BEOL) processing. This technique allows high deposition rates, compared for example to ALD or sputtering, but can result in films with more defects related to mechanical stress. For this reason, it is usually used to deposit thick passivation layer.

The plasma enhanced chemical vapor deposition system used in this work is the STS Oxford PlasmaPro 100 PECVD System.

3.3 Annealing methods

Annealing is a heat treatment process that may change the physical, electrical and chemical properties of a material, altering its internal microstructure. In particular, in this work, annealing processes are used both to crystallize materials in the desired phase, and tune their electrical properties through oxidation processes. In the following sub-sections, the annealing techniques used are detailed.

3.3.1 Rapid Thermal Annealing

Rapid Thermal Annealing (RTA) is a process in semiconductor device fabrication, which consists in heating a sample up to a specific temperature, to affect its crystalline structure and electrical properties. In addition to the heating, a flow of one or several process-gases can be inserted into the processing chamber, to induce further reactions and modifications of the sample, such as oxidation or reduction. During the process, sample's temperature may be monitored by temperature sensors such as thermocouple and pyrometer, where the first is preferred for $T_{setpoint} < 400\,^{\circ}\mathrm{C}$, and the second otherwise. The heating ramp up to $T_{setpoint}$ occurred on a timescale of few minutes maximum, since the heating is performed by high intensity lamps, usually halogen based. Finally, the sample's temperature is controlled and kept constant through power adjustment, for a set amount of time before cooling.

The rapid thermal annealing system used in this work is the Annealsys AS-one 150.

3.3.2 Flash Lamp Annealing

Similar to rapid thermal annealing, flash lamp annealing (FLA) is a thermal process which allows to modify the crystalline structure of materials. However, in this process, the energy needed to alter the internal microstructure of the material is provided both through heating and flashing. In fact, first a pre-heating step is performed through conventional halogen lamp heater located under the sample's holder, as for RTA, then a millisecond flash, with a maximum energy density of $110\,\mathrm{J/cm^2}$ and a duration variable between 0.3 ms and 20 ms, is generated by xenon flash lamp array located at the top of the chamber.
The actual temperature reached during the millisecond flash lamp anneal is not uniform across the sample and depends on the initial pre-heat temperature, and the energy and duration of the flash pulse. During the pre-heating step, temperature is monitored via thermocouple, while there is no facility to monitor the temperature spike during the flash. In addition, the actual energy absorbed by the sample during flashing process depends on the specific material structure of the sample.

The flash lamp annealing system used in this work is the DTF FLA50AS/150 PH.

3.4 Etching Methods

In semiconductor device manufacturing, etching refers to any technology that selectively removes material from a wafer surface to achieve a predefined pattern. Usually, the desired pattern is first transferred in a photoresist, which then is used to define areas that will be etched. Etching techniques can be categorised into dry and wet, according to the nature of etchants. Wet etching exploits liquid etchants to chemically remove materials from a wafer, while dry etching involves a plasma or gaseous etchants to remove the substrate material.

Every etching process is characterized by a specific etch rate, which defines how fast a material is removed and is particularly important for non-selective etch processes since the etch time directly determines the etch depth. Knowing the etch rate, another important figure of merit of an etching process is its selectivity, defined as the ratio between the etch rate of the material required to be etched and the one of the material that should not be etched. In fact, if the process is not selective enough, not only the desired material will be etched, but also the one underneath, causing unwanted overetching. To avoid this condition, high selectivity between the material to be etched and the one underneath, which should work as etch stop layer, is required.

In addition, according to the space uniformity of material removal, an etching processes may be isotropic, if the etch rates are comparable along the different crystallographic directions, or anisotropic, if the etch rate in one crystallographic orientation is much larger than the others.
In the following subsections, the tools used to etch thin film of various compounds are detailed.

3.4.1 Reactive Ion Etching

Reactive ion etching (RIE) is a dry etching process, which being based on a combination of chemical and physical etch, allows to exploit the advantages of both. Chemical etching process, which exploits a selective chemical reaction between etchant gases and the target material, to remove the latter, is usually characterized by fast etch rate and isotropic etch profile. On the contrary, physical etching, which requires high kinetic energy ion beam to physically bombard and etch off the substrate atoms, similarly to Inverse Sputter Etching (ISE), is purely directional, allowing anisotropic etch profile. Reactive ion etching allows to merge these advantages in one technique, characterized by high etch rate and anisotropic patterning capability.

The process starts with the introduction of gaseous chemical etchant precursors, usually fluorine or chlorine based, into the reaction chamber, then an RF energy is applied to a pair of parallel plate electrodes to generate a plasma, whose radicals are used to chemically react and isotropically etch the surface of the target material. In addition to this plasma based chemical etch, an inert gas, usually Ar, is first introduced in the process chamber, then ionized by the plasma, and finally accelerated by an electric field to directional bombard and etch the surface of the target material. The ion beam directional bombardment breaks the chemical bonds of the

surface to be etched, creating dangling bonds, which quickly react with the plasma free radicals, increasing the overall etch rate.

The combination of isotropic chemical and anisotropic physical etching makes this process a versatile tool, where the chemical and physical etch rates can be tuned to get the desired etch profile, keeping a high etch rate.

The reactive ion etching system used in this work is the Oxford Instruments PlasmaPro NPG 80.

3.4.2 Inductively Coupled Plasma Reactive Ion Etching

Inductively Coupled Plasma Reactive Ion Etching (ICP-RIE) is a further optimization of conventional Reactive Ion Etching, which allows to better control the etching profile, overcoming the limitation of conventional RIE.

In conventional Reactive Ion Etching process, a single RF power supply is used to sustain the plasma, used both to generate the radicals necessary for chemical etch, and to ionize inert gas, such as Ar, which is then accelerated by a DC power to physically etch the target material. This makes difficult to independently control the radical etching and the ion flux, hence the amount of isotropic and anisotropic etching. In addition, another limitation of conventional RIE based on parallel plate electrodes, is its low plasma density [32].

To overcome both these limitations, inductively coupled plasma reactive ion etching tool exploits two independent RF power supplies, one controlling the radical generation and the other the Ar ionization, to be able to span from purely physical etching to purely chemical etching. In addition, to generate a high density radical plasma, a time-varying electric current is passed through an inductive coil, to create oscillating magnetic field around it, inducing, according to Maxwell's equations, an electric field. This field forces the electrons to move in spiral paths generating a high density plasma of radicals [32]. As for conventional RIE, the radicals diffuse to the sample to do chemical etching, while the ions, accelerated by a DC signal, provide a physical component to etching. The larger the bias, the greater the energy of the impinging ions.

The inductively coupled plasma reactive ion etching system used in this work is the Oxford Instruments Plasmalab 100.

Chapter 4

Results

4.1 FeFET

Planar FeFET, processed with bottom gate approach as described in section 1.3, are electrically characterized to provide insights on the role of defects and oxygen vacancies in the transport and resistive switching mechanisms. In FeFET planar memristor, the accumulation and depletion depth x_d of the WO_x channel is expected to extend for few nm from WO_x-HZO interface [19], depending on the carrier density thus the stoichiometry. For this reason, to improve the memristor performances and the switching control, thin WO_x channel, just 4 nm thick, is used. Keeping constant thickness, several geometries, having channel width and length of 5 µm, 10 µm, 20 µm and 30 µm, are characterized.

To investigate the conduction mechanisms both in WO_x channel and along the gate stack of planar FeFETs, temperature dependent transport measurements are performed. In particular the temperature is swept from 298 K to 348 K with a step of 10 K. Before performing the electrical measurement, the sample is kept at the target temperature for $\simeq$ 10 min to allow the establishment of thermodynamic equilibrium between the sample and the heated gold chuck of the probe station. At each temperature, the characterization procedure is the one described in section 2.1. In particular first a reset cycle is performed with a gate voltage sweep from 0 V to $V_{reset} = -5$ V with a step of 50 mV back and forth, while grounding source and drain terminals. Then, the channel in the HRS is explored by a voltage sweep between source and drain from -2 V to 2 V with a step of 25 mV, keeping the gate tip disconnected from the corresponding pad, hence floating, to avoid a potential drop between the resistive channel and the gate itself, which could perturb the measurements. Finally, exactly the same couple of measurements but with $V_{set} =$ 5 V, are carried out to investigate the set cycle of the gate and the channel in the LRS.

Several devices have been characterized, however in the following sub-sections, the result of these measurements are provided both for the channel and the gate of one representative device having both channel length and width of 10 µm, since the

considerations about the nature of conduction are not affected by device's geometry.

4.1.1 Channel conduction

To study the conduction along WO_x channel, all the physical mechanisms detailed in sub-section 2.1.1 have been used to fit the experimental data. However, the only mechanism able to properly fit the data is the Ohmic one, while all the others lead to non reasonable parameters. The Ohmic fit for both HRS and LRS of the representative device are reported in figure 4.1 and 4.2.

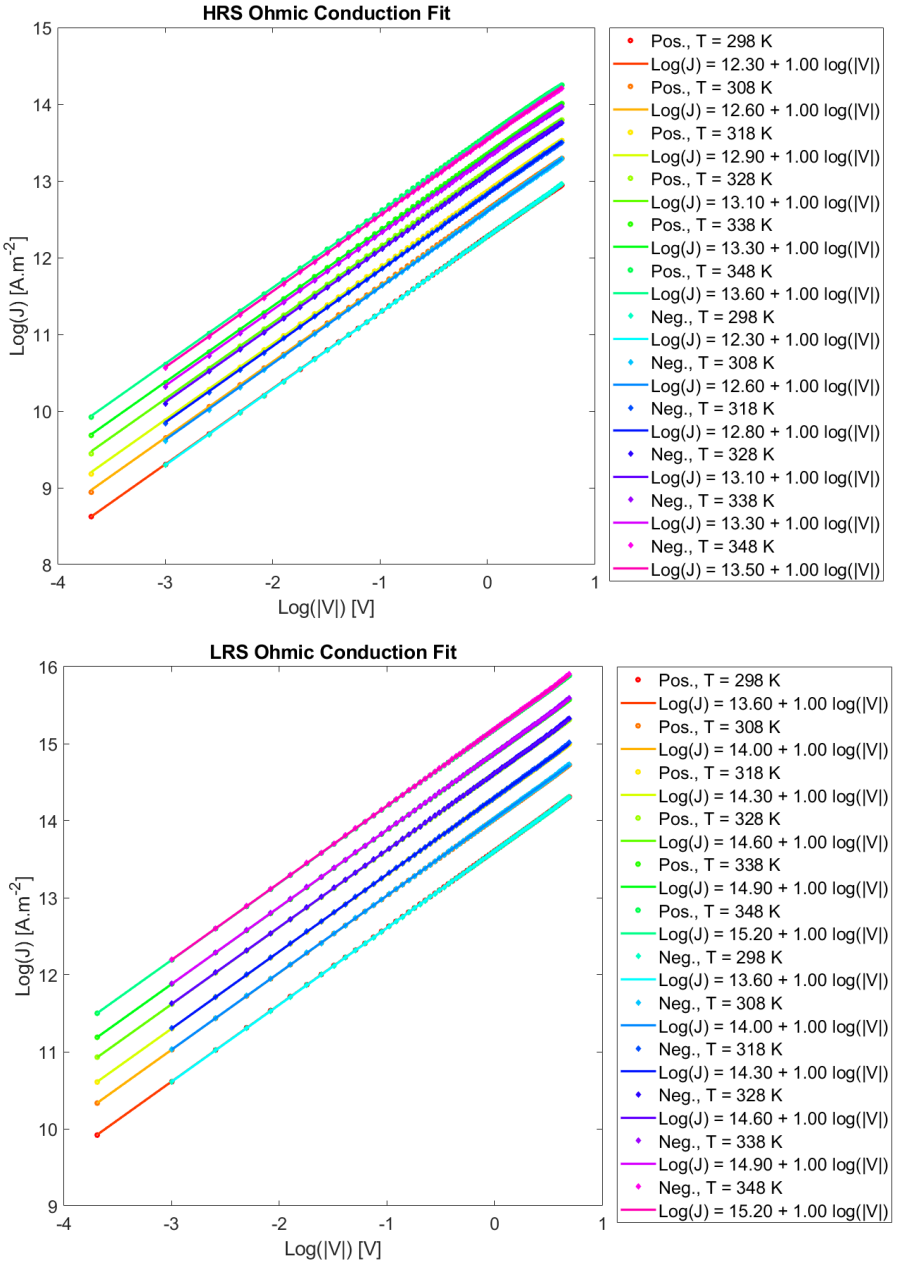


Figure 4.1: Experimental data and ohmic conduction fit at different temperatures, for both HRS and LRS.

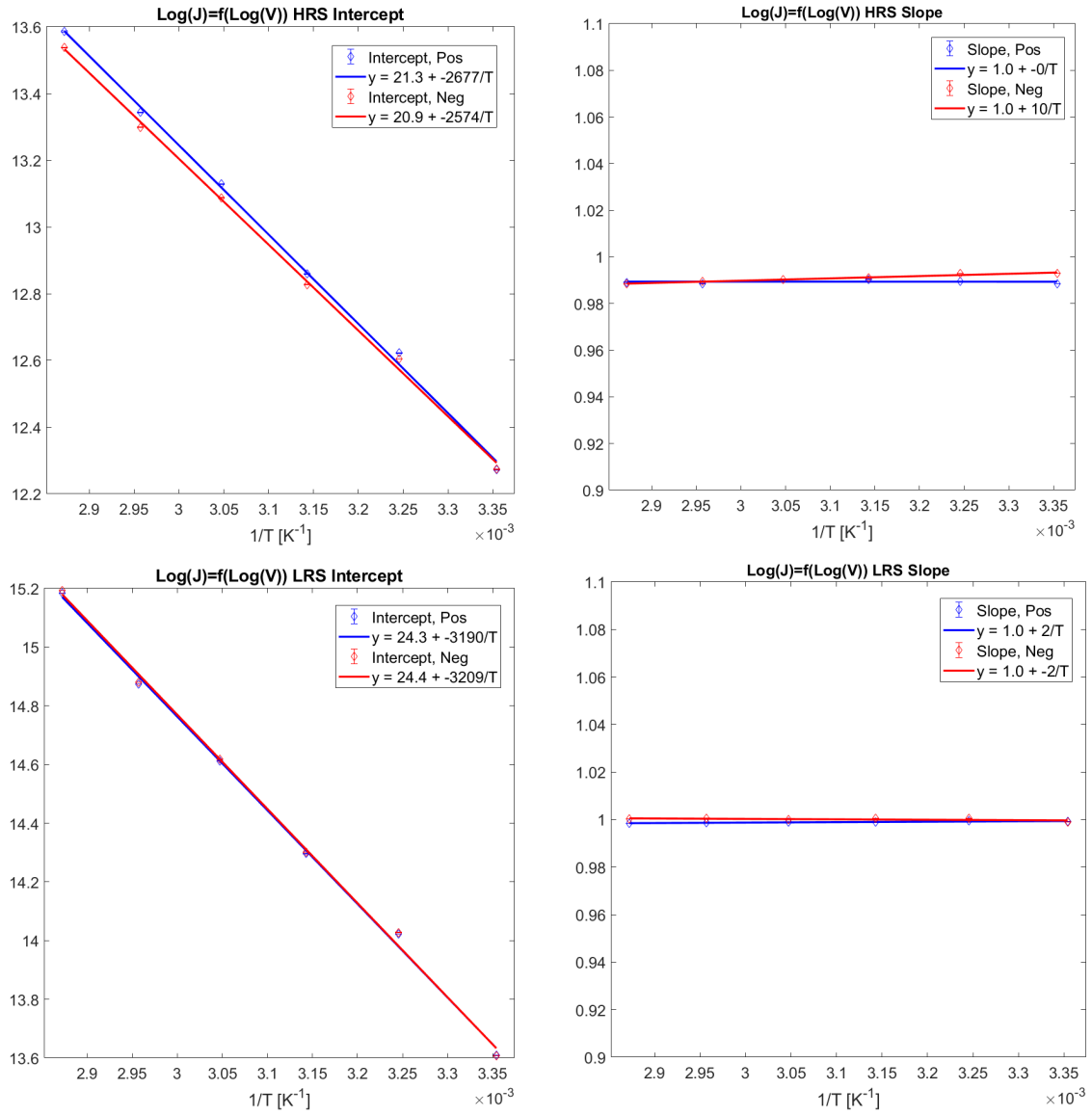


Figure 4.2: Arrhenius plot of intercept and slope, for both HRS and LRS.

The trends show a conduction totally symmetric with respect to the swept voltage, in fact current density J is almost identical for positive and negative voltages. Both in HRS and LRS, the slope of $\log(J)$ as function of $\log(|V|)$ is close to 1 and temperature independent, as expected in ohmic conduction. Intercept Arrhenius plot shows a linear trend with respect to $\frac{1}{T}$, where the absolute value of the slope is proportional to $(E_C - E_F)$, while from the intercept the μN_C product can be estimated. Both the electron mobility μ and the effective density of states in the conduction band N_C depend on temperature, but in such a way that their relation compensate each other and the relative product μN_C can be assumed temperature independent [33]. From these measurements, the parameters reported in table 4.1 can be inferred.

Table 4.1: WO_x parameters inferred from ohmic conduction.

	ρ RT ($\Omega\cdot$ cm)	$E_C - E_F$ (eV)	$\mu \cdot N_c$ $(cmVs)^{-1}$
HRS	$4.8 \cdot 10^1$	0.23	$9.2 \cdot 10^{20}$
LRS	$1.2 \cdot 10^2$	0.27	$2.3 \cdot 10^{22}$

The resistivity of the WO_x channel at room temperature suggests a stoichiometry of $x \simeq 3$ [20]. The μN_C product increases of more than one order of magnitude going from HRS to LRS, which can be reasonable considering that the overall conduction is improving.

The band gap of WO_x is expected to be between 2.4 eV and 3 eV [34], thus the extracted energy distance between the bottom of the conduction band and the Fermi level, is coherent with the idea that WO_x behaves as n-type semiconductor [34]. However, in a standard semiconductor, it is known that increasing the n-type doping concentration, hence the electrical conductivity, the Fermi level moves towards the bottom of the conduction band [33]. In this case the inferred $(E_C - E_F)$ energies in HRS and LRS are not coherent with the previous described model, indicating a possible limitation of the analytical model used to fit the channel conduction. In fact, the convolution of several other mechanisms in addition to the ohmic one, can make the latter slightly inaccurate. Another possible reason for this apparently anomalous behavior of $(E_C - E_F)$ distance, is that actually tungsten oxide is not a standard n-type semiconductor, since the doping is due to oxygen vacancies more than external dopants. A more accurate analysis, is necessary to model and simulate the peculiar band structure of this novel metal oxide material.

Finally, the temperature dependencies of HRS, LRS and ON/OFF ratio are extracted and reported in figure 4.3.

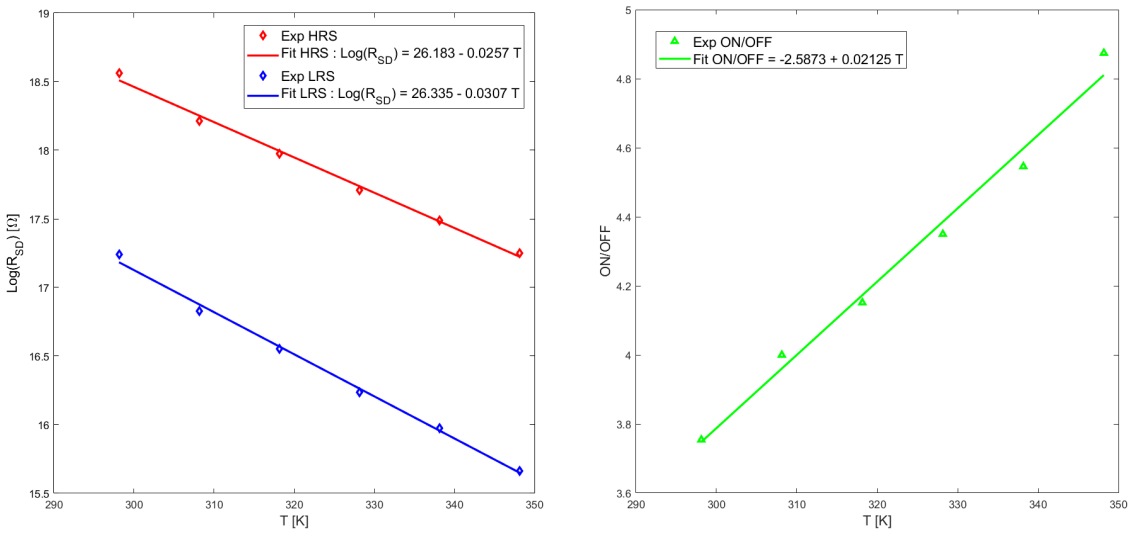


Figure 4.3: Temperature dependence of HRS, LRS and ON/OFF ratio of planar representative FeFET.

Channel resistance, both in HRS and LRS, decreases exponentially with the increase of temperature, as expected for ohmic conduction [23], while the ON/OFF ratio shows an increasing trend with respect to the temperature.

Considering that the total channel resistance can be approximated by two channels in parallel, one in which the sheet carrier density and thus the resistivity is modulated upon polarization switching, and a bulky one with a constant resistivity [19], increasing the temperature, hence the electron concentration in the conduction band of WO_x [33], is expected to reduce the polarization screening length, decreasing the extension of the modulated resistance, hence the ON/OFF ratio. However, ON/OFF data in figure 4.3, do not reflect the predicted behavior, and show a stronger temperature decreasing trend for the LRS with respect to the HRS, from which the increasing temperature trend of the ON/OFF ratio originates. This different temperature dependence may be due to the fact that increasing the temperature means increasing the mobility of oxygen ions in HZO, hence during the reset operation, since a negative voltage is applied to the gate, oxygen ions can be moved from HZO to WO_x, causing a light oxidation of the latter. By contrary, during set operation, since a positive voltage is applied to the gate, WO_x channel can be slightly reduced due to oxygen ion motion from WO_x to HZO. This mechanism may explain the lighter temperature-decreasing trend of the HRS compared to the LRS, which causes an increasing dynamic range with respect to the temperature.

4.1.2 Gate conduction

The gate stack is composed by HZO packed between WO_x and TiN. Since the work functions of these two material are different, a non symmetric behavior is expected during set and reset cycles. In particular, TiN 10 nm thick has a work function of $\Phi \simeq 4.7\,\text{eV}$ [35], while the one of slightly reduced WO_x is expected to be around 5.5 eV [36]. Since by definition the Fermi levels E_F of each material have to be aligned at the thermodynamic equilibrium, for the same absolute value of the applied gate voltage, an higher current is expected during the set operation with respect to reset. In fact, during the reset, a negative voltage is applied on the gate, causing the net motion of electrons from TiN to WO_x, hence part of the applied voltage is used first to make the band flat, and then to bend it in the other direction. On the other hand, during set operation, since a positive voltage is applied on the gate, the band bending already present at thermodynamic equilibrium, due to different work functions, is enhanced and an higher current is expected. This trends are confirmed by experimental measurements, provided in figure 4.4.

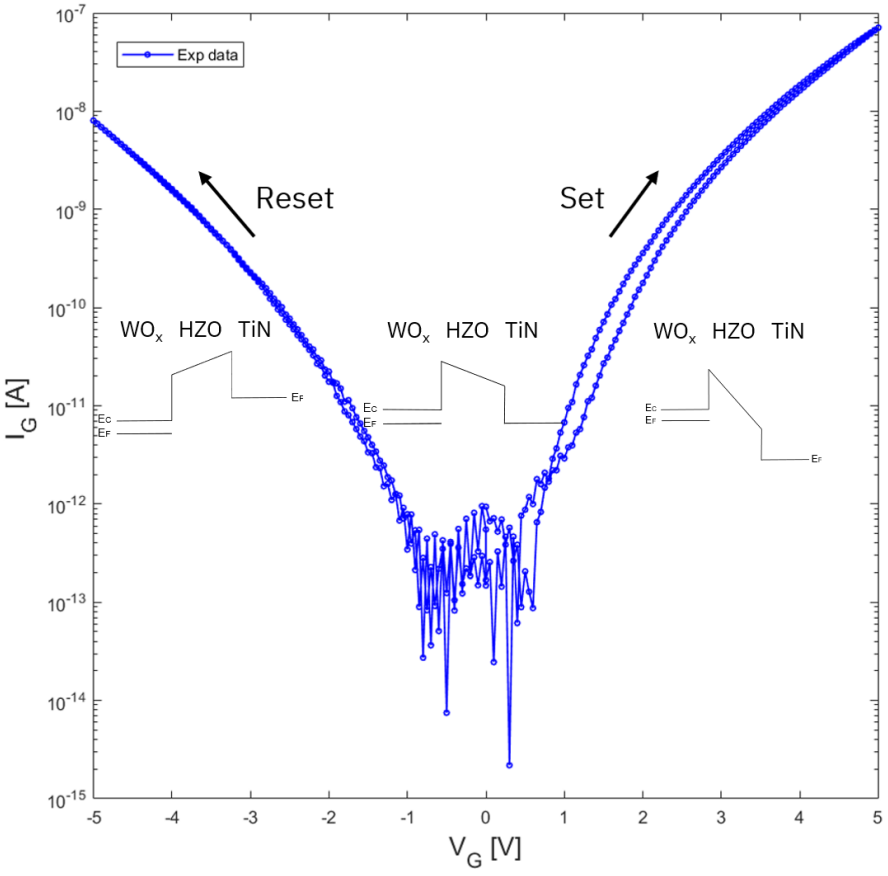


Figure 4.4: Gate current as function of applied voltage. The inserts show a qualitative representation of band bending at thermodynamic equilibrium, set and reset.

To study the conduction through the gate stack, all the physical mechanisms detailed in sub-section 2.1.1 have been used to fit the experimental data. In addition, since HZO is much more resistive than WO_x, the conduction current is basically only due to the former, since the vertical stack can be thought as the series of two resistances, where the WO_x one, 4 nm thick, is totally negligible with respect to that of 10 nm thick HZO.

The comparison of the conduction currents and mechanisms measured through the channel and through the gate, allows to determine if the conduction between the source and the drain actually occurs in WO_x 10 µm wide channel, or if the current flows vertically across HZO, only 10 nm thick, travels horizontally through the TiN and flows back across HZO. If the latter case occurred, hence WO_x was not involved in the conduction between source and drain, since both the channel and the gate transports would be dominated by HZO, the two conduction mechanisms would be the same.

The mechanism that better fits the experimental data is the Modified Schottky Emission. In figure 4.5 and 4.6, the Modified Schottky Emission fits of the representative device are reported after both a set and a reset operation.

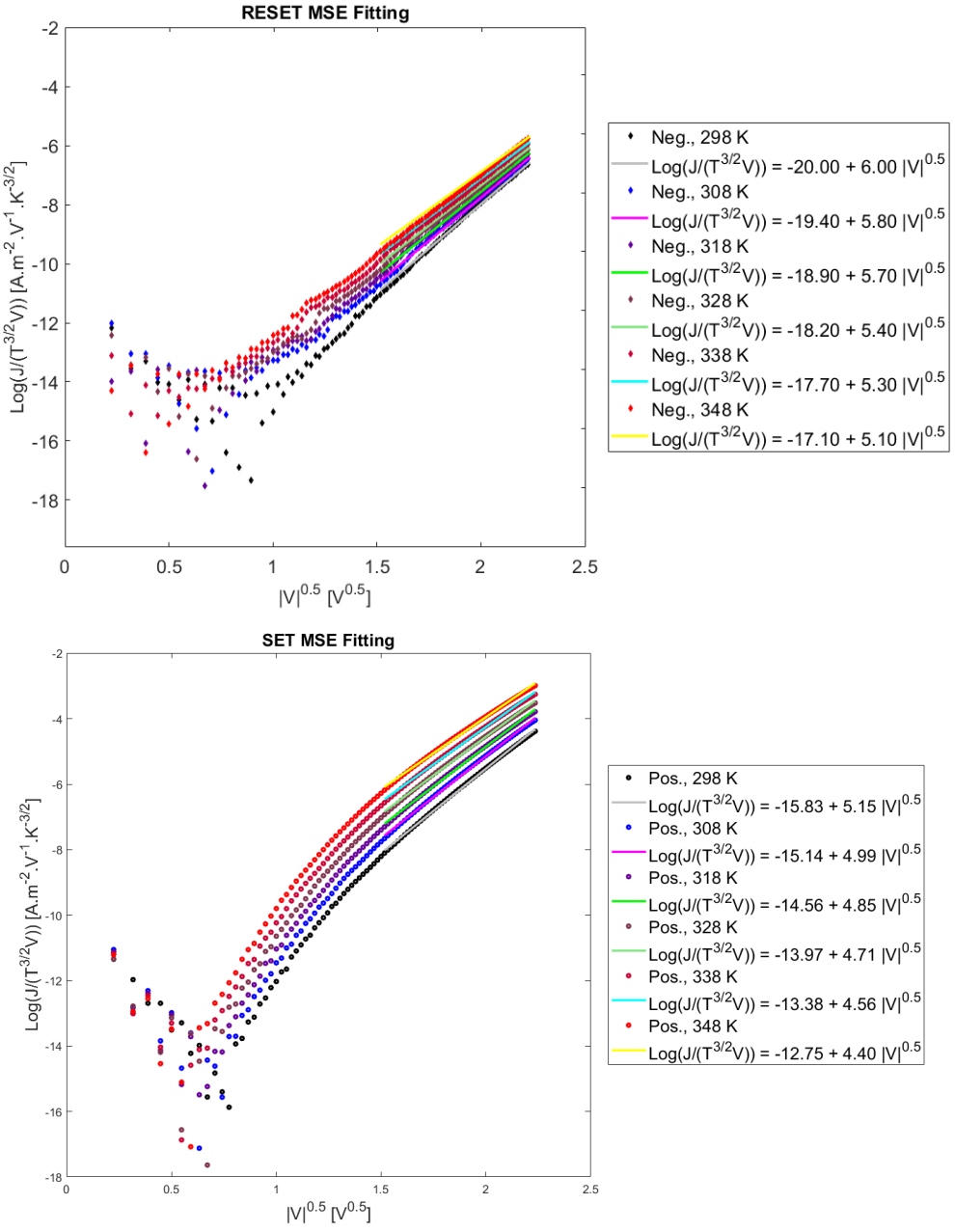


Figure 4.5: Experimental data and MSE conduction fits at different temperatures, during set and reset cycles.

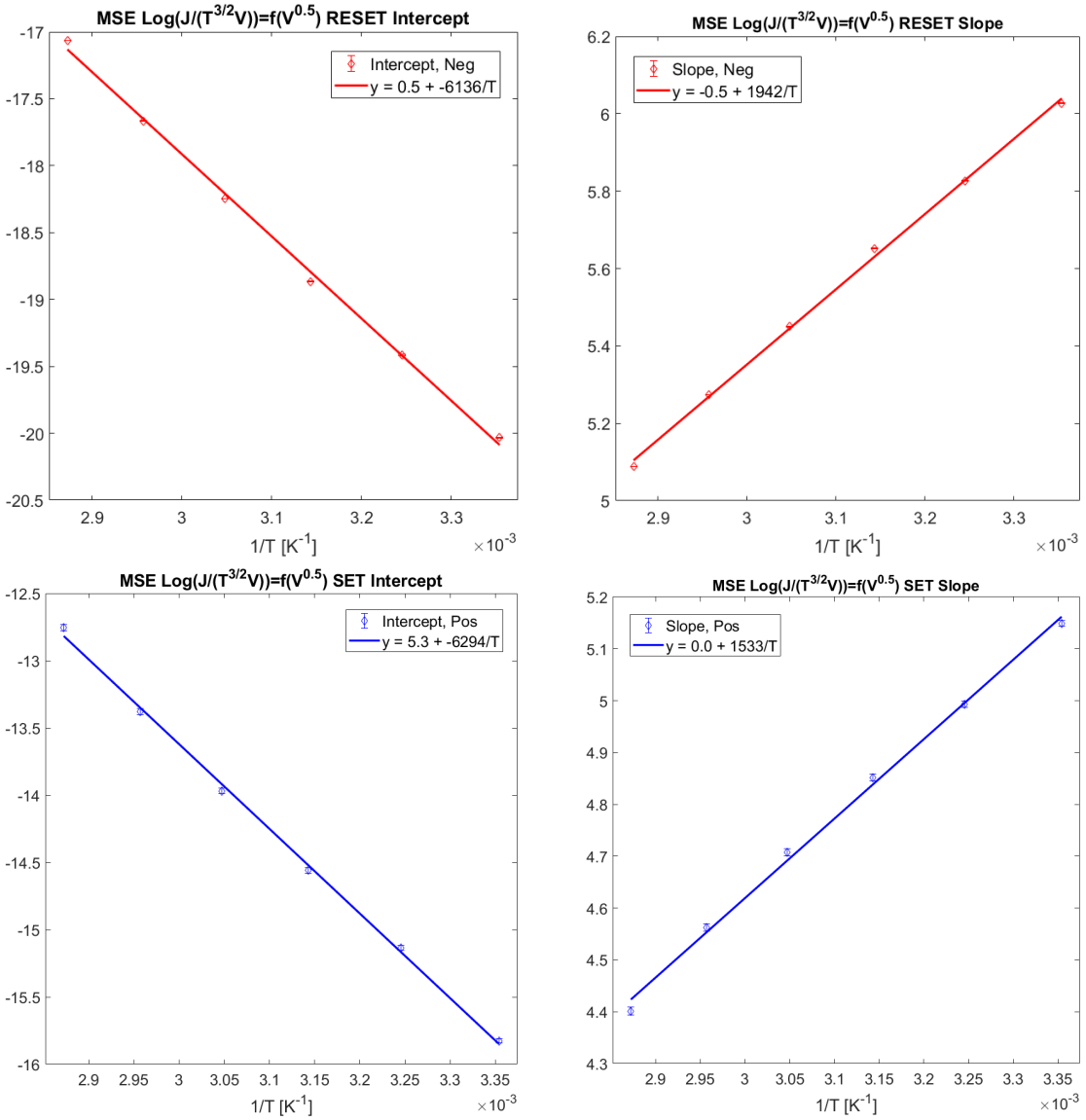


Figure 4.6: Arrhenius plot of intercept and slope, during both gate set and reset.

From equation 2.7, the product $\mu \cdot (\frac{m^*}{m_0})^{3/2}$ and the dynamic dielectric constant ϵ_r can be measured from the intercept Arrhenius plot, while the Schottky barrier height $q\phi_B$, can be measured from the slope Arrhenius plot. From these measurements, the parameters reported in table 4.2 can be inferred.

Table 4.2: HZO parameters inferred from MSE conduction.

	$q\phi_B$ (eV)	ϵ_r	$\mu \cdot (\frac{m^*}{m_0})^{3/2}$ $(cm^2 s^{-1} V^{-1})$
SET	0.54	8.2	$6.42 \cdot 10^{-5}$
RESET	0.53	5.1	$5.45 \cdot 10^{-7}$

The extracted dynamic dielectric constant ϵ_r, defined as the square of the optical refractive index $\epsilon_r = n^2$ [37], is found to be different for set and reset operations. This can be linked to the change in electrostatic screening of the polarization charges within the HZO itself. However, the predicted values of ϵ_r are reasonable if compared to the optical refractive index of hafnium oxide $n_{HfO_2} = 2.1$ [38]. The slightly higher intensity of Schottky barrier height during the set operation with respect to reset is coherent with the higher work function of WO_x with respect to that of TiN. In fact, since the electron affinity of HZO is unique, it is possible to assert that the following relation has to be always fulfilled:

$$q\phi_{WO_x} - (E_C - E_F)_{WO_x} - q\phi_B^{SET} = q\phi_{TiN} - q\phi_B^{RESET} \tag{4.1}$$

where $q\phi_{WO_x}$ and $q\phi_{TiN}$ are the work functions of WO_x and TiN respectively and $(E_C - E_F)_{WO_x}$ is the distance between the bottom of conduction band and the Fermi level in tungsten oxide. However, according to these values, the electron affinity, defined as the energy obtained by moving an electron from the vacuum to the bottom of the conduction band, of HZO is expected to be around 4 eV, which could be the case for a substoichiometric HZO [39].

In conclusion, through temperature dependent DC electrical characterization, it is possible to provide insights about the nature of the conduction in such complex materials. However for a detailed analysis and parameter extraction, a more accurate analysis, using numerical simulation rather than analytical models, is suggested.

4.2 Fin-FeFET

Fin based Ferroelectric-Field Effect Transistor (FinFeFET) performances are mostly determined by the aspect ratio of the fins. In fact, enhancing the aspect ratio, which means increasing the height and/or decreasing the width of the fins, all the advantages of 3D structures, become more and more important, hence the overall performances are expected to improve.

However, the achievable channel aspect ratio is limited both by the processing capability of modern tools and the mechanical properties of the channel material, which in this case is tungsten oxide. On the other side, no processing limitation affects the achievable number of fins in parallel. However, since this number is proportional to the current flowing in the transistor's channel, hence to the electrical conductance of the device, the overall switching performances, such as the resistance window, may shift away from the targeted range, if it is too large.

Knowing this, several geometries of FinFeFETs are processed and characterized, to find out the best trade off in terms of fin's geometry and number. In table 4.3 all the investigated sizes of FinFeFETs are reported.

Table 4.3: FinFeFET Fabricated Geometries.

Length (nm)	Number Fins	Width (nm)
100	1	2
200	5	4
500	10	8
	20	10
	40	30

In particular, for all the possible combinations of length, number and width of fins shown in table 4.3, five identical FinFeFETs are processed, to have enough devices to characterize that specific size combination. The height of WO_x fins is kept constant to 30 nm. Each block contains 375 devices, and 4 identical blocks are placed in a chip of 2 cm × 2 cm area.

The design of all those devices is performed with a python script exploiting the Library *Ipkiss* [40]. In figure 4.7 the design of a single block and of a single FinFeFET is shown using the KLayout software.

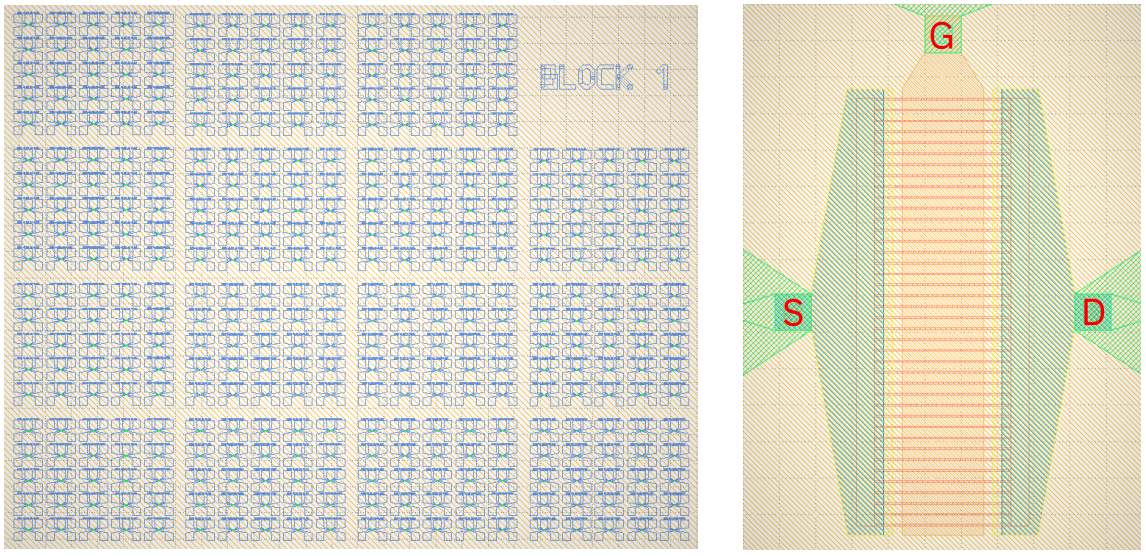
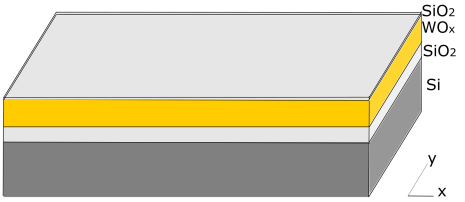
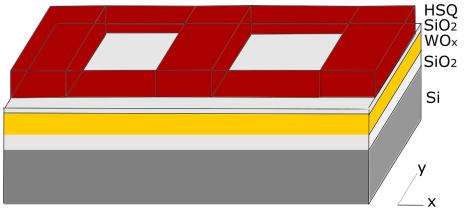
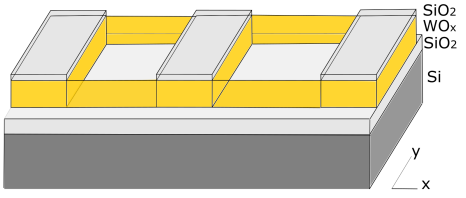
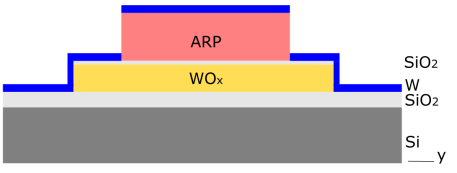
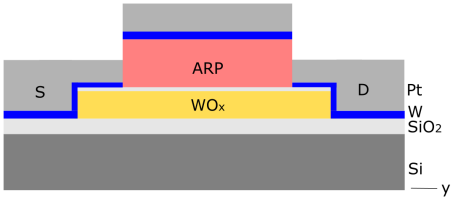


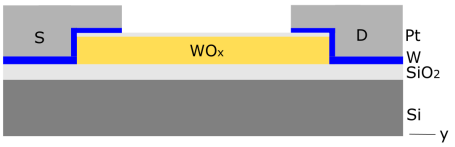
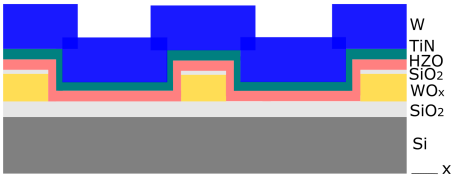
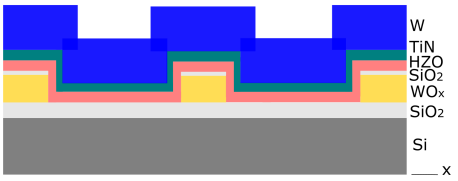
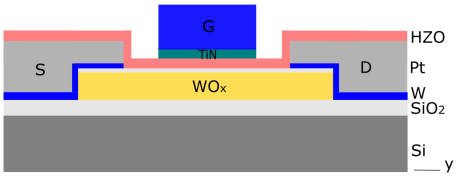
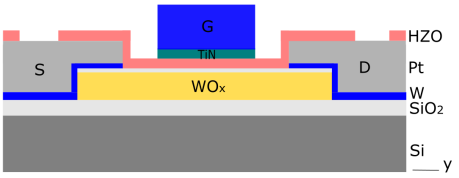
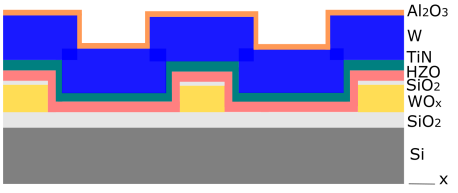
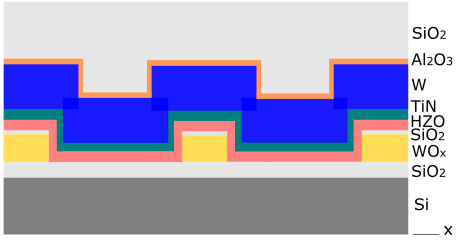
Figure 4.7: On the left, the GDSII format of a single block containing 375 devices is shown, while on the right a zoomed view of a single FinFeFET with L = 500 nm, N = 40 and W = 8 nm is reported as an example.

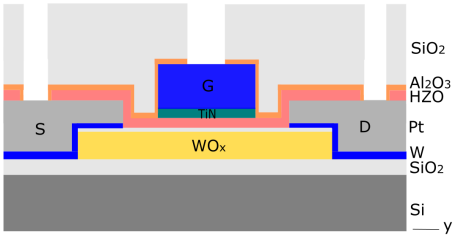
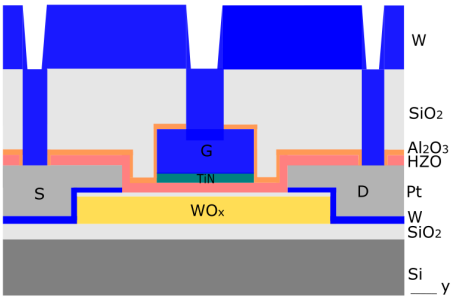
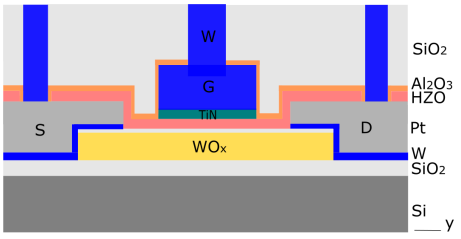
The process flow of FinFeFETs consists of about 60 steps, 7 of which are electron beam lithographic exposures. In addition, since the required resolution of fins is below 10 nm, the bottom gate approach used for planar FeFETs can not be pursued, since it would require the most critical step, the patterning of fins, at the end of the whole process, when all the non-idealities of the previously performed processing steps are summed up.

In table 4.4 the main steps of FinFeFET fabrication are depicted.

Table 4.4: FinFeFET Process Flow

Step	Process Description	Tool	Ideal Schematic View
1	Deposition of WO_x	ALD	
2	Crystallization of WO_x	RTA	
3	Deposition of SiO_2	ALD	
4	Fin Definition in HSQ	E-beam	
5	Etching of SiO_2 and WO_x	ICP	
6	Deposition of W	Sputter	
7	Deposition of Pt	Evaporator	

8	Source-Drain Lift-Off	DMSO	
9	Conformal Deposition HZO - TiN	ALD	
10	Deposition of W Gate	Sputter	
11	Crystallization HZO	ms-FLA	
12	Gate Definition	RIE	
13	Open VIAs through HZO	ICP	
14	Etch stop Al_2O_3 Deposition	ALD	
15	Passivation SiO_2	PECVD	

16	Open VIAs through Passivation	RIE	
17	Deposition of W	Sputter	
18	Etching of W	RIE	

In the following sub-sections, the developed process flow with the main fabrication challenges and solutions, as well as the electrical characterization and further possible optimizations, are detailed.

4.2.1 Fabrication of Fin-FeFETs

The starting point of the FinFeFET fabrication is a 2 cm × 2 cm Si substrate, with 500 nm SiO_2 and 2 nm Al_2O_3 on top. The very first step consists in the deposition of 30 nm thick tungsten oxide layer by plasma enhanced atomic layer deposition. This deposition, which exploits $(BuN)_2W(NMe_2)_2$ and O_2 plasma as precursors, occurs at T = 300 °C and has a deposition rate of 0.47 Å/cycle. Due to low processing temperature, which allows the BEOL compatibility, the deposited WO_x film is sub-stoichiometric and not crystalline.

In order to avoid misalignments when several lithographic steps are involved, e-beam markers have to be defined on top of the sample. The markers are made of gold, with titanium as adhesion layer, and a lift-off technique is used for their definition. In particular, a double e-beam resist layer made of 320 nm thick PMMA AR-P 617.06 at the bottom, and 114 nm thick PMMA AR-P 672.03 on top, is spinned on the WO_x layer, whose surface was previously cleaned and dehydrated for 5 min at T = 180 °C. For both the resists, a soft bake step of 5 min at T = 180 °C is performed after coating, to remove the remaining solvent and stabilize the resists.

After coating, the double resist layer is exposed with electron beam lithography. Since the polarity of the double layer is positive, the exposed regions are more soluble in the developer. The PMMA-developer used contains Methyl Isobutyl Ketone, known as MBIK, and Isopropyl Alcohol, named IPA, in a ratio 1:2 respectively.

At this stage, both titanium and gold are deposited with electron beam evaporator. In particular, 5 nm of Ti and 100 nm of Au are evaporated. The actual lift-off process is performed using Dimethyl sulfoxide (DMSO) as solvent, heated at T = 130 °C to accelerate the process.

After the definition of e-beam markers, the next step consists in crystallizing and making stoichiometric the ALD WO_x layer. To do so, rapid thermal annealing in O_2 atmosphere is performed. In particular, the sample is heated up to T = 350 °C for 30 min, with 50 sccm O_2 flow. This step is crucial for the following reasons:

- **From WO_x to WO_3**:
  It allows to oxidize the tungsten oxide from sub-stoichiometric state to fully oxygen enriched one. This step, together with the following crystallization of HZO with ms-FLA, mainly affects the final resistivity of the channel material, hence the resistive window for HRS and LRS, which is a figure of merit of memristors. In fact, increasing the oxygen concentration, causes a transition from metal to insulator in tungsten oxide [41]. Since during the following crystallization of HZO in its ferroelectric phase, tungsten oxide is heavily reduced, without this RTA oxidation step, the final channel material would be too conductive, resulting in a low performance memristor.

- **Crystallization of WO_3**:
  Crystalline tungsten oxide is found out to be processed in a more controlled manner with respect to its sub-stoichiometric phase. For instance, the undesired etching rate in salty developers, is lower for the crystalline phase with respect to the sub-stoichiometric one.

In order to prove that tungsten oxide is stoichiometric and crystalline, grazing incidence X-Ray diffraction analysis is performed. In particular, knowing that 2θ and ω are the detection angle and the incident angle respectively, a GIXRD scan is performed doing a 2θ scan from 20° to 45°, with a step of 0.02°, an integration time of 1 s per step and an incident angle $\omega = 0.5°$. In figure 4.8, the experimental tungsten oxide X-Ray diffraction pattern is shown.

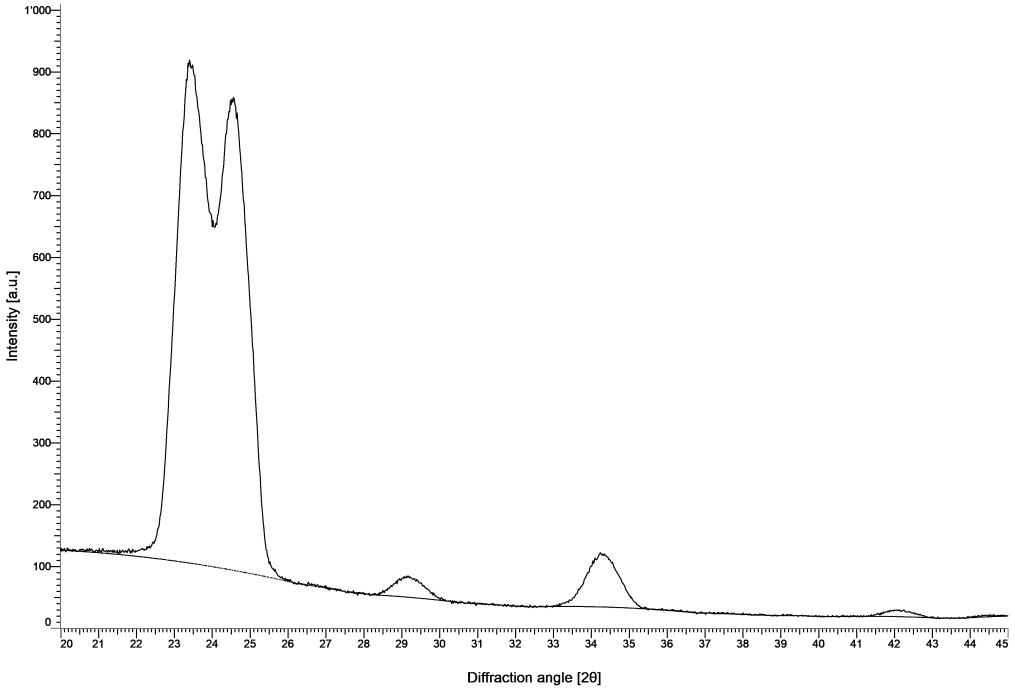


Figure 4.8: GIXRD scan of tungsten oxide after RTA crystallization and oxidation.

Materials Project open-access database is used to label the diffraction pattern peaks. In particular, this tool, based on density functional theory, computes the predicted X-Ray diffraction pattern of the desired material. According to this, the diffraction peaks at $\sim 23.5°$, $\sim 24.5°$, $\sim 29.1°$, $\sim 34.3°$ and $\sim 42.1°$ can be attributed to (001), (200), (111), (220) and (221) Miller indices of tetragonal $P42_1m$ phase of WO_3, respectively. In fact, the calculated X-Ray Diffraction pattern of this WO_3 phase is the one that best matches the experimental data.

In the following paragraphs, the description of FinFeFET process flow continues highlighting the main challenges faced, as well as the implemented solutions.

Fins Definition

The first crucial processing challenge is the definition of the fins into the resist, with a desirable resolution below 10 nm. A first test has been carried out using 365 nm thick ARN 7520.17 resist. In figure 4.9, both the e-beam exposed GDS layer and an SEM image after resist development with the developer AR 300-47, are shown.

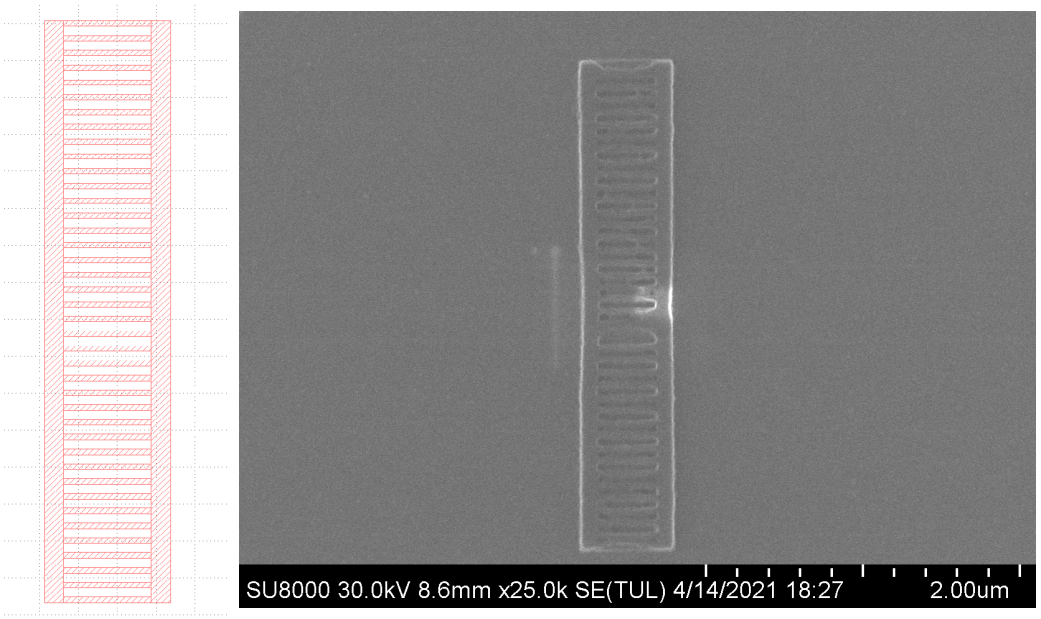


Figure 4.9: On the left, GDS layer used for e-beam exposure, while on the right an SEM image after resist development. The bright areas in SEM image are an artifact due to charge-up effect. The FinFeFET shown size is L = 200 nm, N = 40 and W = 30 nm.

From figure 4.9, it is evident that the used resist is not suitable for the desired resolution, since the fins, even those 30 nm wide, hence with less critical requirements in terms of resolution, are irregular.

Instead of ARN 7520.17, Hydrogen Silsesquioxane 2% resist, known as HSQ, is adopted. This negative resist, characterized by an achievable resolution of ~ 10 nm, after e-beam exposure cross-links so much that resist stripping becomes almost impossible. In fact, since its chemical composition is $[HSiO_{3/2}]_n$ [42], the exposed regions behaves similarly to SiO_2, being inert to stripping solutions as Acetone or DMSO.

Before spinning HSQ 2%, to both improve the resist adhesion and protect WO_x from the HSQ developer, a 3 nm thin layer of SiO_2 is deposited at T = 300 °C by plasma enhanced ALD. If the SiO_2 layer was not deposited, the HSQ salty developer based on NaOH 1% and NaCl 4% would slightly etch the underneath crystalline WO_3 causing HSQ delamination and undesired channel etching.

After a surface dehydration for 5 min at T = 180 °C, 40 nm thick HSQ 2% layer is spinned and exposed by e-beam lithography. After the development, an SEM inspection is performed, whose result is shown in figure 4.10.

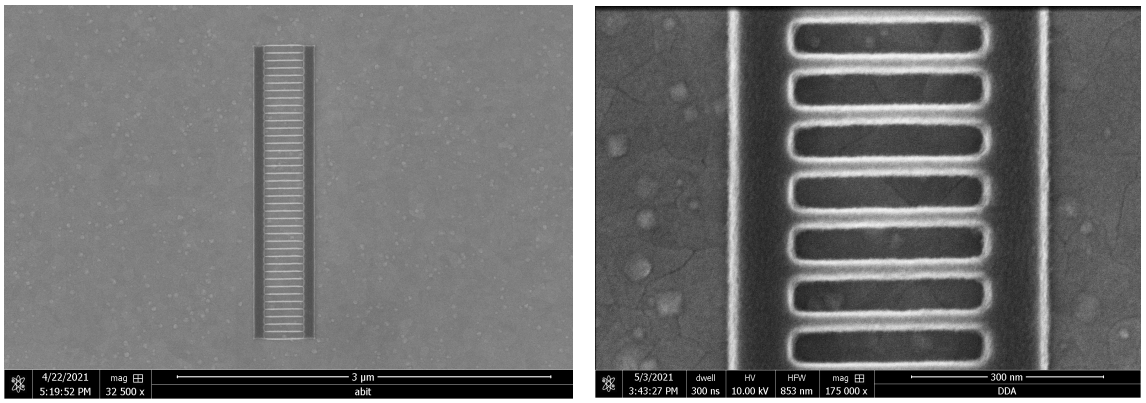


Figure 4.10: SEM images after HSQ development using the HSQ salty developer based on NaOH 1% and NaCl 4%. The shown FinFeFET's size is L = 100 nm, N = 40 and W = 30 nm.

The pattern in HSQ is well defined and an evident improvement is visible if compared to the ARN 7520.17 resist. However, the resolution of HSQ is found to be between 8 nm and 10 nm. In fact, HSQ fins having 2 nm nominal width collapsed, and the ones nominally 4 nm, 8 nm and 10 nm wide actually are, on average, 8 nm, 10 nm and 11 nm respectively.

Etching WO_3

The following challenge consists in transferring the fins-pattern from HSQ to WO_3 through a dry etching process. In order to etch both the thin layer of SiO_2 and WO_3, inductively coupled plasma reactive ion etching is used. The etching recipe consists of 15 sccm flow of both CHF_3 and SF_6, and an RF power of 1200 W. In figure 4.11 an SEM image of the sample after 60 s WO_3 ICP-RIE etching, is shown.

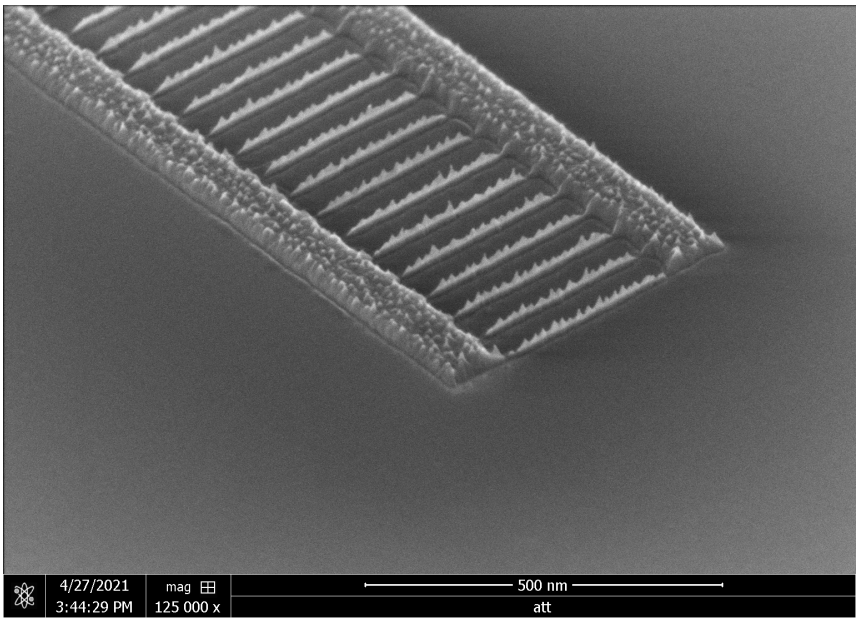


Figure 4.11: SEM image after WO_3 etching process with ICP-RIE. The shown FinFeFET's size is L = 100 nm, N = 40 and W = 10 nm.

After the ICP-RIE etching process, the surface of fins appears rough and irregular. A possible explanation is that, since both the HSQ thickness and the ICP-RIE etching rate are not perfectly uniform, in some spots all HSQ is removed and there is an amplification of surface morphology due to etching first of thin SiO_2 layer, and then of WO_3 fin. In fact, the used ICP-RIE etching recipe etches WO_3 faster than HSQ.

A solution could be to decrease the etching time, to keep a thin layer of HSQ on top of fins avoiding the surface roughness shown in figure 4.11. However, since it is necessary to etch all WO_3 between the fins, a decrease of etching time is not possible. For this reason, the implemented solution consists in increase the HSQ thickness, from 40 nm to 60 nm, in order to still have a thin layer of it on top of the fins when the etching process is finished. The implication of this HSQ layer on top, which is unavoidable to have high resolution and flat fin's surface, is that the advantage of 3D channel geometry is partially lost, since the accumulation/depletion of electrons in tungsten oxide channel occurs only on the two vertical sides if HSQ is on top. However, since the height of fins is larger than their width, the loss in channel control is expected to be negligible. In figure 4.12, an SEM image of the WO_3 fins and a FIB cross-section are reported.

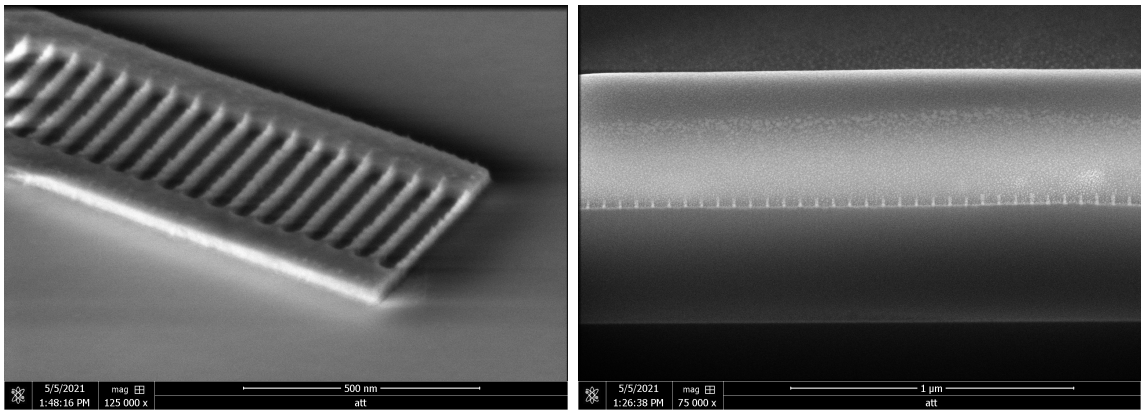


Figure 4.12: On the left, an SEM image of tungsten oxide fins, while on the right a FIB cross-section. The shown FinFeFET's size is L = 100 nm, N = 40 and W = 10 nm.

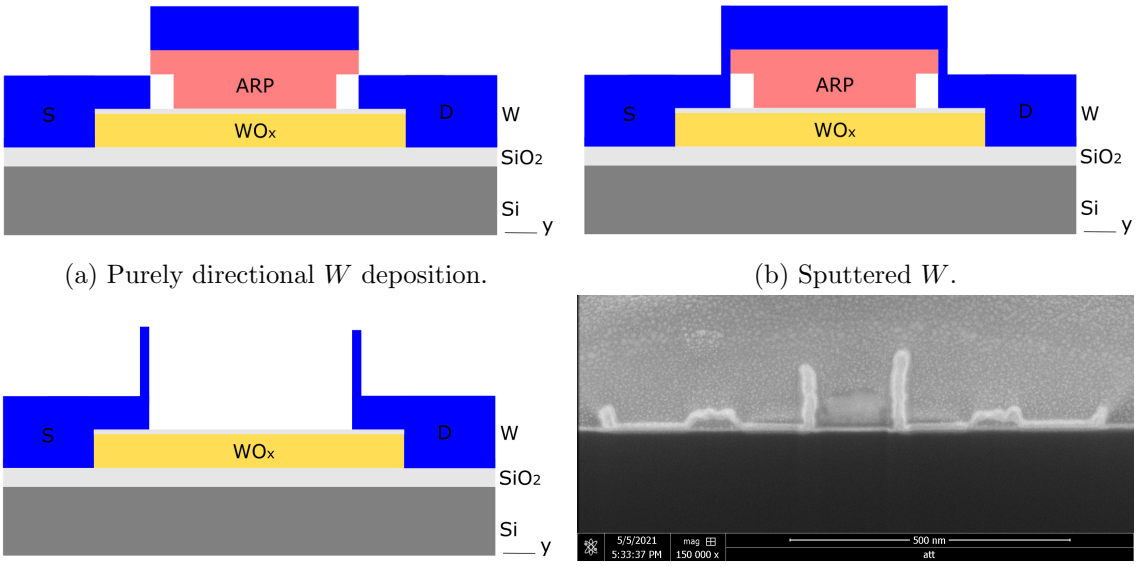
The surface of the fins is flatter than before, even if a light roughness is still present and may be attributed to the not ideal ICP-RIE process. The FIB cross-section view shows that the fins have considerable aspect ratio.

Source and Drain Definition

The definition of source and drain electrodes is achieved through a lift-off process. To do so, a double e-beam positive resist layer made of 325 nm thick PMMA AR-P 669.04 at the bottom, and 265 nm thick PMMA AR-P 672.05 on top, is spinned onto the sample, whose surface was previously dehydrated for 5 min at T = 180 °C. For both the resists, a soft bake step of 5 min at T = 180 °C is performed. After the e-beam exposure, resist development is performed using a solution containing IPA and H_2O in 7:3 ratio.

The first choice as source and drain material is tungsten, since it forms a fine interface with tungsten oxide. For this reason, 40 nm thick W layer is sputtered using 120 W as DC-power. The lift-off process is done using DMSO solvent, heated at T = 130 °C.

However, since sputtering is a not directional deposition technique, even with the double layer approach, after the deposition, tungsten covers also the vertical sidewalls of the resist. Therefore, lift-off causes the formation of vertical metallic shapes along the sidewalls, known as *lift-off ears*, which stands upwards from the surface. If the ears remain on the surface, they can cause unwanted connections or can even fall over the surface. In figure 4.13 a schematic explanation of this phenomenon, as well as the experimental FIB cross-section view, is provided.



(a) Purely directional W deposition.

(b) Sputtered W.

(c) Lift-off of sputtered W.

(d) FIB cross-section.

Figure 4.13: Schematic representation of W-ears problem after lift-off.

To address this problem, e-beam thermal evaporation, which is a deposition technique more directional than sputtering, is used. However, due to its high melting point, tungsten is notoriously difficult to evaporate. For this reason, platinum (Pt) is chosen as the source and drain electrode material. To keep the fine interface between tungsten and tungsten oxide, just 5 nm thick W layer is sputtered, and then 50 nm thick Pt layer is thermally evaporated.

In figure 4.14 both the GDS layout used to define source and drain contacts, and the result after the lift-off of W-Pt layers, are shown.

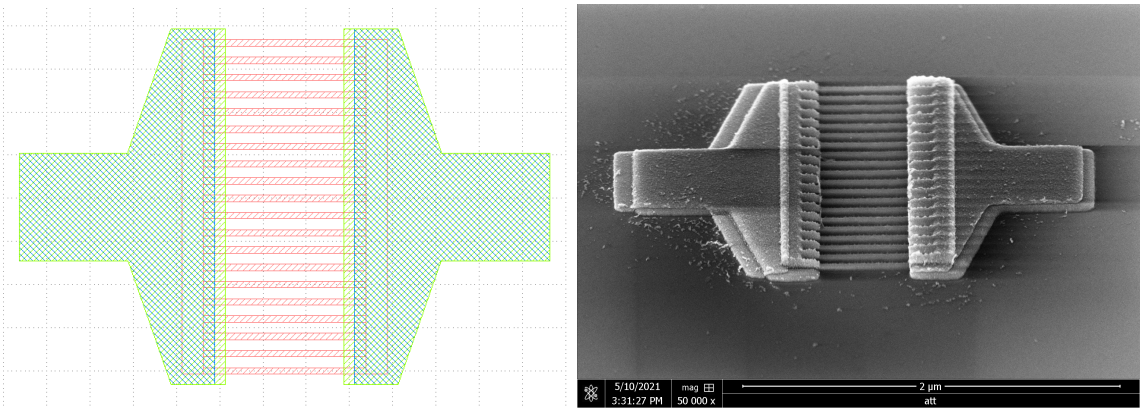


Figure 4.14: On the left, GDS layout used for e-beam exposure, while on the right an SEM image after W-Pt lift-off. The shown FinFeFET's size is L = 500 nm, N = 20 and W = 30 nm.

The step and the light misalignment visible in figure 4.14 are due to a double lift-off process. In fact, first 30 nm thick Pt layer was deposited, but then since the source and drain electrode covering of fins was not complete, another 20 nm thick Pt layer

was evaporated.

After source and drain definition, a 10 nm thick layer of Hafnium Zirconium Oxide (HZO), which is the material responsible for the memristor switching mechanism, is deposited by plasma enhanced ALD using alternating cycles of tetrakis-(ethylmethy lamino)hafnium (TEMAH) and ZrCMMM ((MeCp)-2Zr(OMe)(Me)) precursors at T = 300 °C. Afterwards, without venting the ALD chamber, further 10 nm of titanium nitride (TiN) is immediately deposited on top using tetrakis-(dimethylamino)titanium (TDMAT) as precursor. In fact, for ferroelectric HZO crystallization, using low thermal budget millisecond flash lamp annealing technique or conventional rapid thermal annealing, a capping layer such as TiN or W is necessary [43].

Finally, to avoid TiN oxidation, 40 nm thick W layer is sputtered as gate electrode material. Flash lamp annealing (FLA), with a preheat of 120 s at T = 375 °C and a flash energy density of 70 J/cm^2, is performed to crystallize Hafnium Zirconium Oxide in its orthorhombic phase, the only one showing ferroelectric properties [44].

During flash lamp annealing, in addition to HZO ferroelectric crystallization, the reduction of WO_3 occurs. This allows the transition from WO_3 insulating material to WO_x, which exhibits n-type semiconductor properties. Probably, both SiO_2 underneath and HZO on top, are responsible of the oxygen reduction of WO_3. This process finally determines the resistivity of the channel material, hence the resistive window of the memristor.

In order to prove the correct crystallization of Hafnium Zirconium Oxide after FLA, grazing incidence X-Ray diffraction analysis, shown in figure 4.15, is carried out.

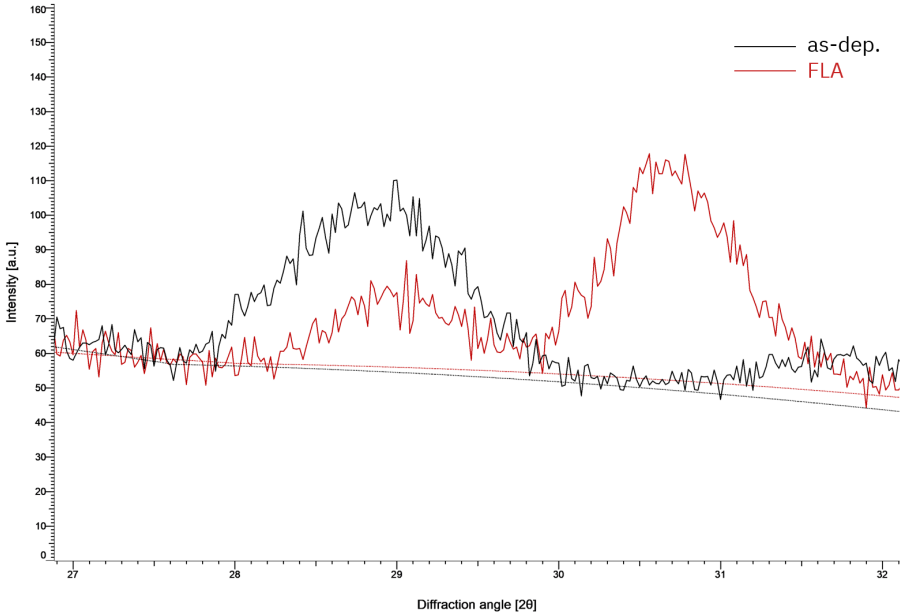


Figure 4.15: The black and red curves show the GIXRD pattern before and after HZO crystallization respectively.

The peak at $\sim$ 30.8°, which appears after FLA, originates from the overlapped

orthorhombic (111) and tetragonal (011) phases of HZO [44]. The peak at $\sim 29°$, already present before FLA and only shaped during it, may be due to (111) plane in WO_3 tetragonal $P42_1m$ phase, as suggested by *Material Project* DFT based XRD pattern simulation.

To pattern the gate electrode, negative resist 160 nm thick AR-N 7520.073 is used. However, before resist spinning, the sample is dipped inside Surpass 4000 water based adhesion promoter, to mitigate the risk of resist delamination. Soft bake of 90 s at T = 85 °C is performed. After the exposure by e-beam lithography, the development is carried out with a solution made of AR 300-47 and water, in a ratio of 4:1. However, especially for configuration having L = 100 nm, several ARN resist delamination events and resist thinning below nominal length, are registered, anticipating possible low performances in terms of switching capability, hence ON/OFF ratio, for this configuration.

Afterwards, the gate electrode is defined etching both W and TiN layers, using Argon sputtering in between, from undesired regions. The etching is performed using a reactive ion etching recipe, which consists of 30 sccm flow of both SF_6 and N_2.

In figure 4.16 both the GDS layout used to pattern gate contact, and the result after W-TiN RIE process and successive resist stripping in acetone, are shown.

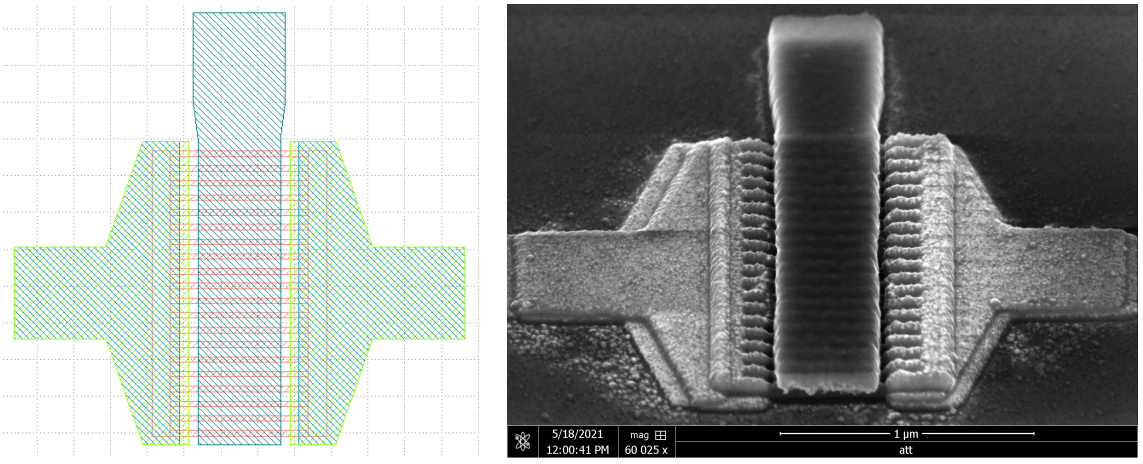


Figure 4.16: On the left, GDS layout used for e-beam exposure, while on the right an SEM image after W-TiN etching process using RIE. The FinFeFET shown size is L = 500 nm, N = 20 and W = 30 nm.

At this step, source and drain VIAs, defined using positive resist 265 nm thick AR-P 672.05, are opened through HZO by ICP-RIE etching based on a CF_4 chemistry. Finally, to protect the active area of the memristor from the surrounding environment, a passivation layer is necessary. In particular, first 5 nm thick Al_2O_3 layer by plasma enhanced ALD and then 100 nm thick SiO_2 by PECVD, are deposited.

Openings in the passivation are defined in correspondence of source, drain and gate VIAs, using positive resist 265 nm thick AR-P 672.05, and then etched by reactive ion etching based on a CHF_3 chemistry. Aluminium oxide, which behaves as a

stopping layer during SiO_2 etching, is then removed by a wet etching process in MIF726 developer.

Finally, a 150 nm thick W layer is first sputtered, and then patterned with RIE, to act as ultimate interconnection layer and to fabricate contact pads.

4.2.2 Electrical Characterization

The fabricated FinFeFETs are electrically characterized, to estimate performances such as the resistive range, the ON/OFF ratio, the number of intermediate states and the linearity of weight update. To do so, measurements of pristine resistance of all the processed devices are carried out, then on those showing the targeted resistance, first $R_{DS} - V_{write}$ DC measurements are performed, then potentiation and depression by 5 µs pulses with progressive amplitude is proved. Since the amount of devices to be measured is considerable, a wafer map containing the relative positions of all the FinFeFETs is generated to allow automatic measurements.

For pristine resistance reading, a voltage sweep back and forth between -0.21 V and 0.21 V, with a step of 20 mV, is applied between source and drain, keeping the gate floating. Since the channel is ohmic in that voltage range, the resistance of each device is extracted. To have statistically relevant measurements on how the pristine resistance depends on the three swept parameters (length, number and width of fins), 5 devices nominally identical are measured for each configuration. The configuration characterized by nominal width of 2 nm is not reported in the following analysis being the yield close to zero. In fact, at the limit of HSQ's resolution, such thin structures would require further optimisation for the adhesion, exposure and development conditions. Neglecting this configuration, the number of devices characterized in terms of pristine resistance is 300. The overall processing yield is of 77.7%, with 67/300 devices showing an open circuit behavior.

The results of this analysis is shown in figure 4.17, where the missing configurations represent not working devices. FinFeFETs having the same number of fins but different widths are graphically scattered around the nominal value N_{fins} to allow a direct comparison, while data with same L are encoded with the same color code.

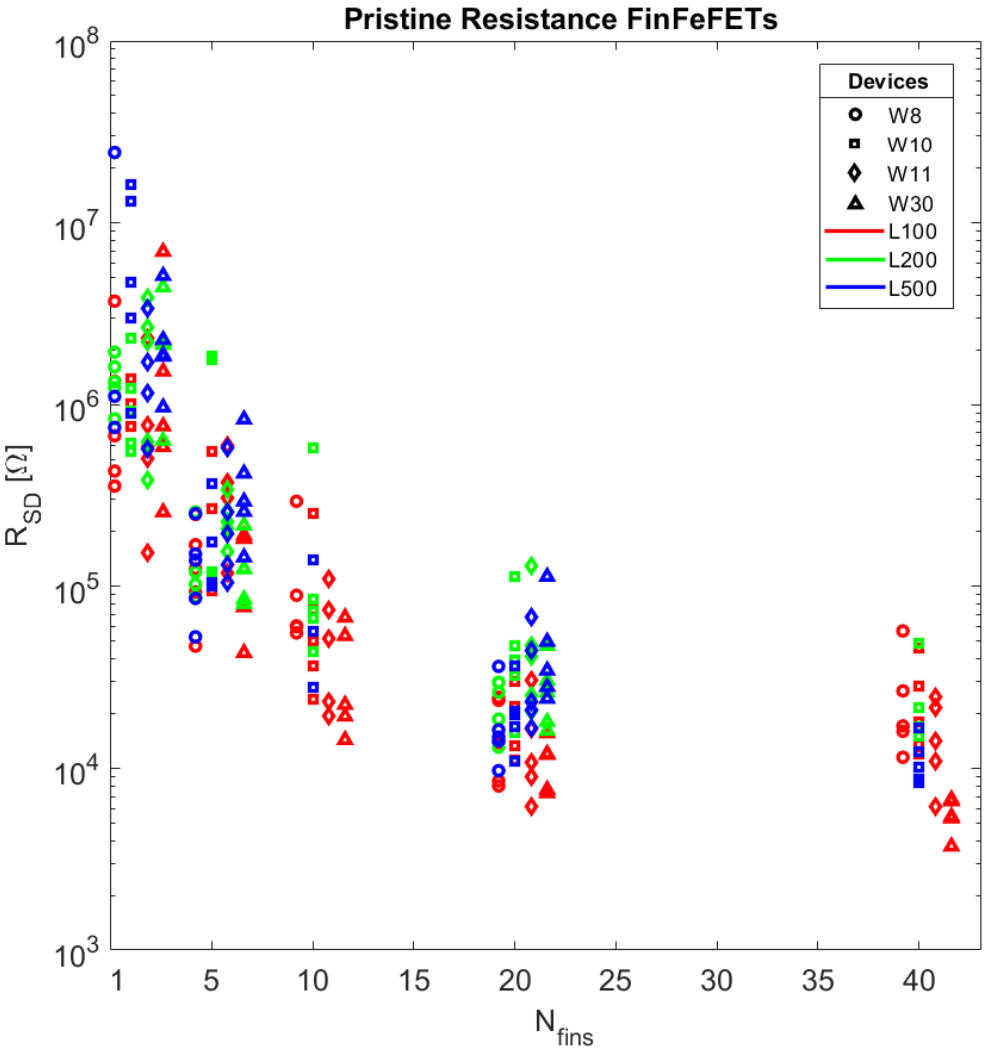


Figure 4.17: Pristine resistance of FinFeFETs as a function of length (L), number (N) and width (W) of fins.

A decreasing trend of the pristine resistance as a function of number of fins is evident in figure 4.17, which is coherent since an higher number of fins in parallel entails a lower overall resistance. Regarding the length of fin, it is expected that longer it is, more resistive the channel will be. This trend is not always fulfilled, suggesting processing variability and similar performances especially between configuration with L = 200 nm and L = 500 nm.

Finally, the resistance is supposed to decrease by increasing the nominal width of fins. However also this trend is not always reported in the experimental data. This can be explained considering that width is the most critical dimension in terms of resolution. In fact, due to resolution limitations of both e-beam lithography and HSQ resist, it is known from FIB cross sections that the devices having nominal widths of 10 nm, 8 nm and 4 nm are actually 11 nm, 10 nm and 8 nm wide respectively, hence comparable resistances are expected. In addition, it has to be considered that since

the device dimensions are heavily scaled, physical phenomena may scale differently with respect to the macroscopic counterpart.

After mapping the pristine resistance of all the processed devices, further electrical characterization in DC and with pulses is performed just on the 60 devices having one fin, different lengths and different widths. In fact, it is expected that multiple fin configurations are just a convolution of what happens in the single-fin one. In addition, from figure 4.17 it is evident that the single-fin FinFeFETs can guarantee the highest resistive range, which is a figure of merit of memristors.

In particular, the routine used applies a DC voltage V_{write} on the gate of the memristor, with source and drain as common reference, and then read the channel with $V_{read} = 200\,\text{mV}$, keeping the gate floating. This couple of operations are repeated sweeping V_{write} from $-4\,\text{V}$ to $4\,\text{V}$ back and forth, with a step of $250\,\text{mV}$. This kind of measurements allow to extract information such as the switching window of the memristor, so called ON/OFF ratio, as well as the coercive voltage V_c and the values of High Resistive State and Low Resistive State.

To avoid any wakeup effects, before $R_{DS} - V_{write}$ DC measurement, polarization in HZO is cycled up and down 100 times by applying a sequence of $4\,\text{V}$ and $-4\,\text{V}$ on the gate, keeping source and drain as common terminals.
In figure 4.18, the channel resistance as a function of the writing voltage, is reported for two representative samples.

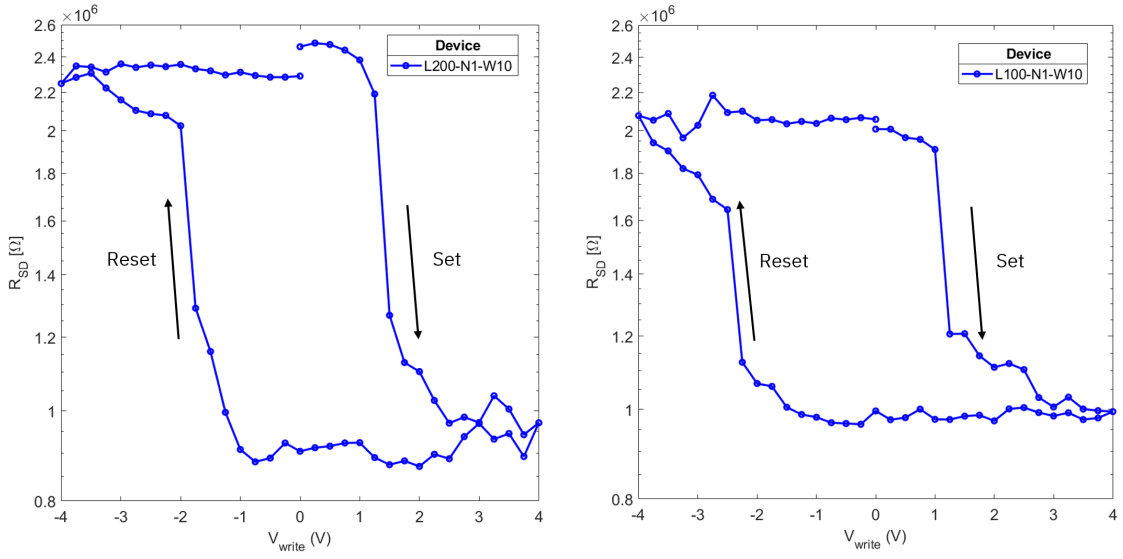


Figure 4.18: Channel resistance R_{SD} after the application of write voltage V_{write} of varying amplitudes.

Set and reset operations occur with a positive and negative programming voltage on the gate respectively. Both the high and low resistive states are stable and a coercive voltage $V_c \simeq 2\,\text{V}$ is necessary to change the state of the memristor. The light asymmetry in the hysteresis $R_{DS} - V_{write}$ loop is due to the imprint in the ferroelectric layer [45]. For the devices shown in figure 4.18, the ON/OFF ratios are

slightly larger than 2, and the resistive window is, as expected for $N_{fins} = 1$, around $M\Omega$s.

All the 60 devices with $N_{fins} = 1$, are characterized, as in figure 4.18, to measure the ON/OFF ratio. In figure 4.19 the distribution of this figure of merit among all the single fin FeFET, is provided. FinFeFETs having the same fin's length but different widths are graphically scattered around the nominal value L_{fins} and encoded with different color codes, to allow a better representation.

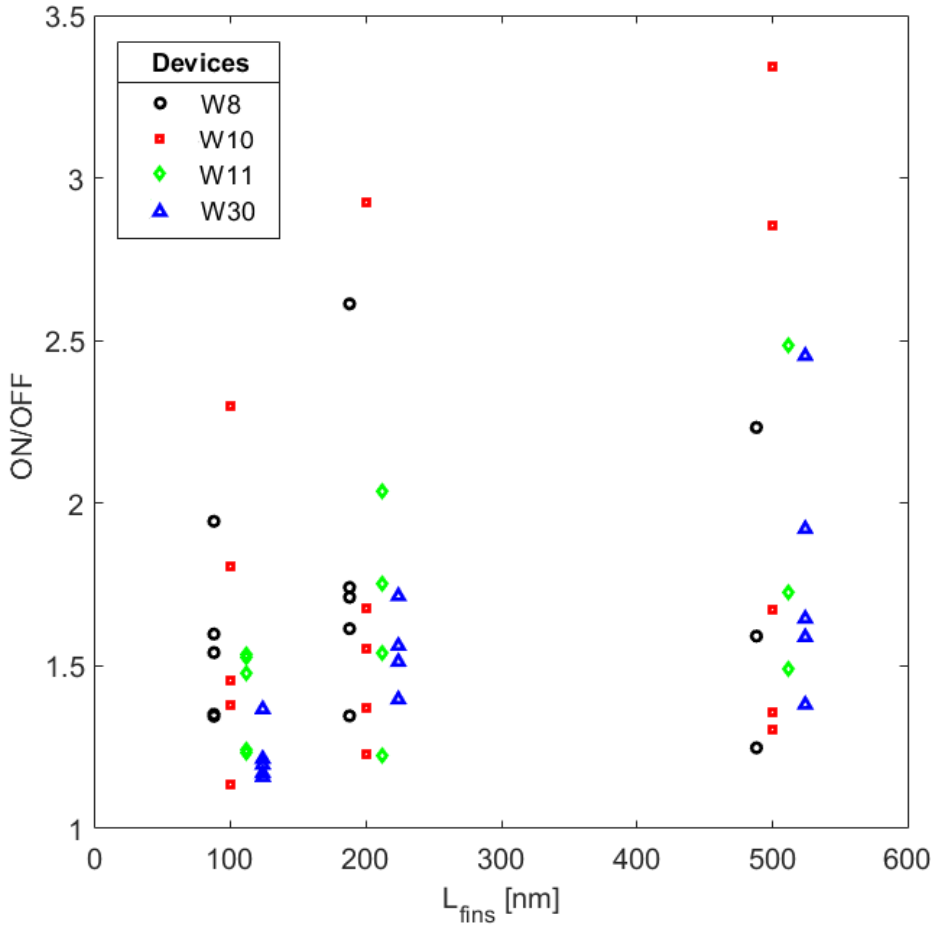


Figure 4.19: Distribution of ON/OFF ratio among one fin FinFeFET.

Considering just these 60 devices, the processing yield is 95%, with 57/60 devices that switch between an high and low resistive state. The variability in the ON/OFF distribution, may be due to several factors, such as the not yet mature and reproducible process flow, as well as inhomogeneous WO_x material, and a critical fin's dimension comparable to the size of the ferroelectric domains in HZO, which is found to be of the order of the thickness of the ferroelectric film [19]. The light increasing trend of the ON/OFF ratio with the length of fins, can be explained considering that, on average, devices with L = 500 µm are processed better than those with L = 100 µm, as explained in sub-section 4.2.1.

After DC characterization, in order to estimate the dynamics of voltage controlled partial polarization switching in FinFeFETs, hence the multistate nature of the memristor, pulsed characterization is performed by applying voltage pulses of varying amplitudes V_{write} and keeping a fixed pulse duration of 5 µs. V_{write} pulses, generated by a Waveform Generator Fast Measurement Unit (WGFMU) and RSU module of a Agilent B1500, are applied directly to the gate through a triax cable, while grounding both the source and the drain. For potentiation, V_{write} is increased from 1 V to 5 V, and for depression, decreased from -1 V to -5.5 V, with 50 mV steps. A slightly higher voltage is used for the depression to compensate the imprint effect in HZO. After each pulse, the channel resistance R_{SD} is measured, keeping the gate floating and applying an IV sweep from -200 mV to 200 mV to the source, while having the drain connected to the ground. R_{SD} is then determined by averaging the resistance at $\pm$ 200 mV.

In figure 4.20, 10 potentiation and depression cycles of a FinFeFET with L = 500 nm, N = 1 and W = 30 nm, are shown.

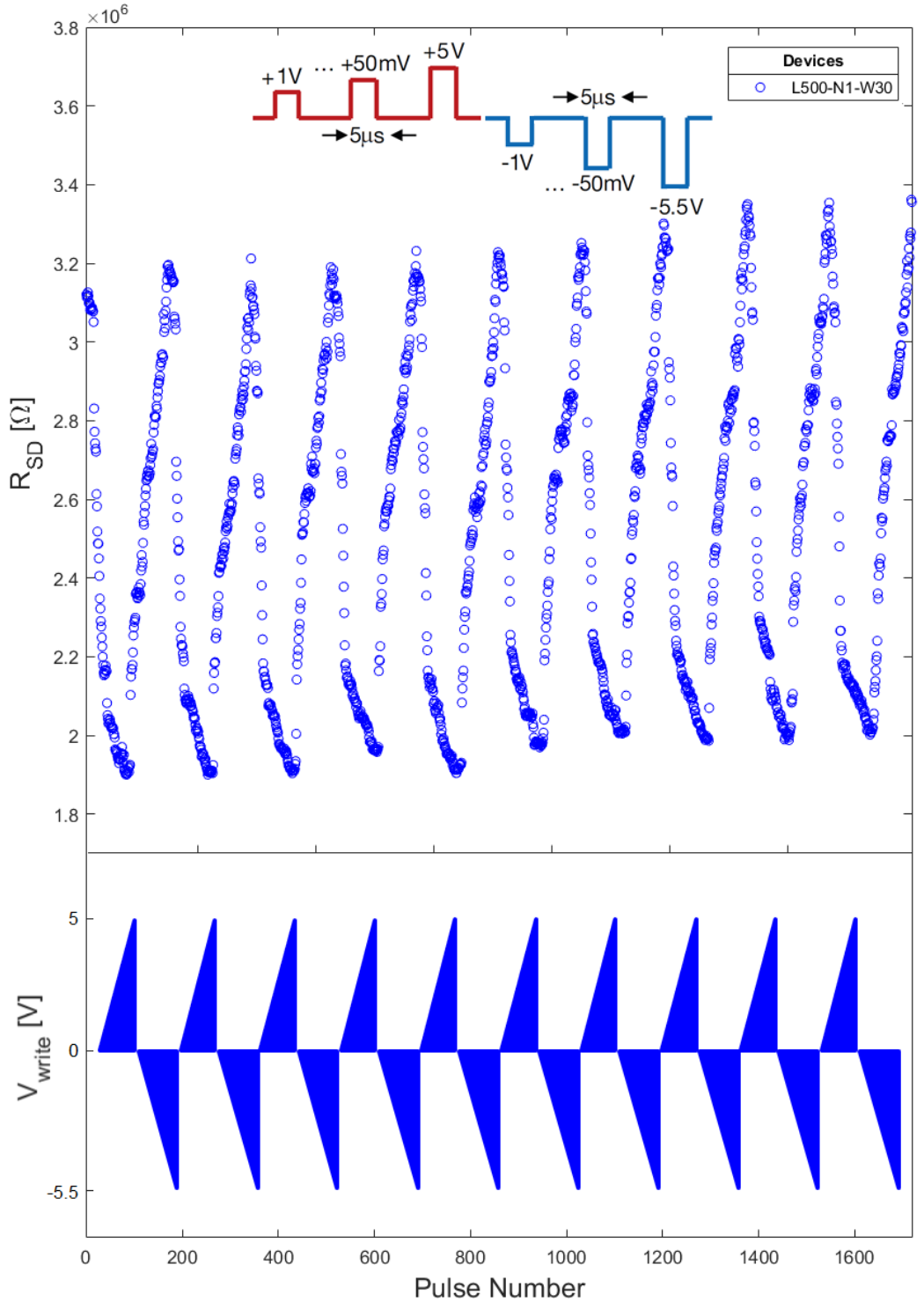


Figure 4.20: Multiple potentiation and depression cycles of the FinFeFET channel resistance R_{SD} with varying pulse amplitudes V_{write} and constant pulse widths of 5 µs. The bottom panel shows the corresponding write pulse sequence. After each pulse, R_{SD} is measured.

Both the high and low resistive states of the device shown in figure 4.20, tends to slightly drift towards higher resistance values, but the relative ratio remains constant. No wake-up effect is visible. The response of the memristor to pulses is stable, showing a HRS of $\sim$ 3.2MΩ and a LRS of $\sim$ 1.9MΩ, hence an ON/OFF ratio $\simeq$ 1.7. With respect to DC characterization, almost all the devices show a decreased resistive window, hence ON/OFF, which may be due to the short programming pulses and less movement (oxidation and reduction) of the WO_x channel.

To take into account the cycle to cycle variability, all the potentiation and depression cycles are averaged. In fact, averaging over several cycles, allows to show multiple intermediate states with their corresponding standard deviation. In figure 4.21, R_{SD} averaged on all cycles as a function of relative pulse number and corresponding V_{write} is shown.

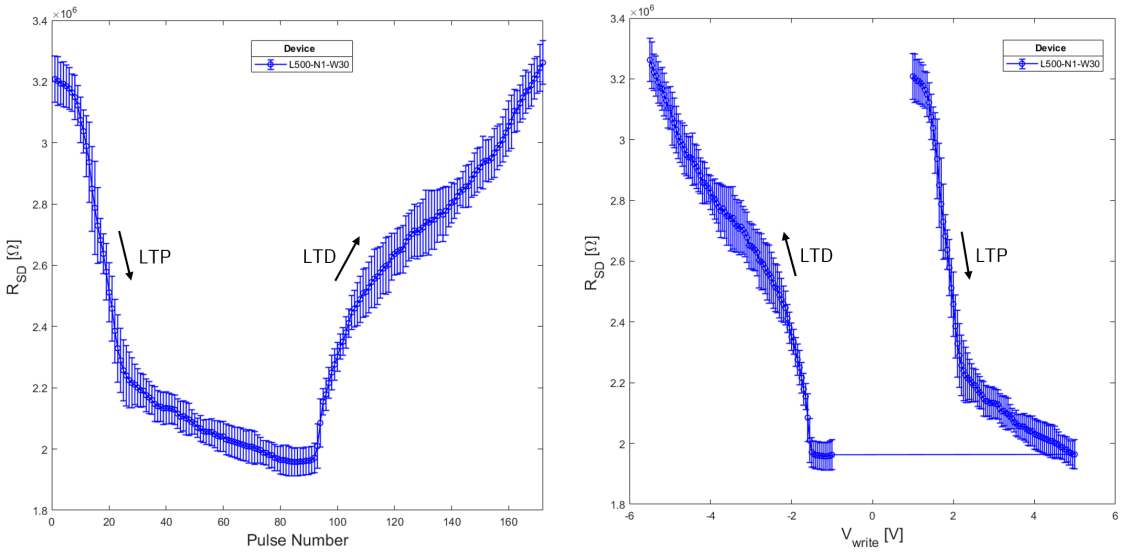


Figure 4.21: Channel resistance R_{SD} averaged over 10 potentiation and depression cycles, considering V_{write} varying from 1 V to 5 V for potentiation and from −1 V to −5.5 V for depression, keeping 5 µs as pulse width.

As shown in figure 4.21, by reducing the range of V_{write}, the resistive switching window is considerably reduced. On the other hand, a light decreasing trend of the resistive switching window is noticed for shorter pulse width, down to 100 ns. However, since it is expected that even shorter pulses could successfully program this device [46], this effect can be due to set-up limitation. For this reason, all the analysis has been performed using 5 µs width pulses.

In addition, in figure 4.21, the number of intermediate states is defined by the potentiation and depression step size, which can be further reduced to increase the resolution. However, even if not all the states are differentiable since the overlapped standard deviation ranges, they are monotonic increasing and decreasing, which is desirable for online learning.

Fitting the depression range from −1 V to −5.5 V and the potentiation range from 1 V to 5 V, allows to extract the weight update characteristic, thus conductance as

a function of pulse number, which can be used as input for MLP simulator and NeuroSim, to emulate the online learning/offline classification scenario with MNIST handwritten dataset in a 2-layer multilayer perceptron (MLP) neural network based on memristors [47], [48]. To do so, experimental weight update data are fitted using MATLAB script *nonlinear-fit.m* developed by Pai-Yu Chen et all [48].

In particular, to model the nonlinear weight update, the conductance change with number of pulses (P) is described by the following device behavioral model [47], [48]:

$$G_{LTD/LTP} = B_{LTD/LTP} \cdot (1 - \exp\left\{-\frac{P}{A_{LTD/LTP}}\right\}) \tag{4.2}$$

$$B_{LTD/LTP} = \frac{G_{max} - G_{min}}{1 - \exp\left\{\frac{-P_{max}}{A_{LTD/LTP}}\right\}} \tag{4.3}$$

where G_{max}, G_{min} and P_{max} can be directly extracted from the experimental data, which represents the maximum conductance, minimum conductance and the maximum pulse number required to switch the device between the minimum and maximum conductance states [47], [48]. The parameter A controls the behavior of weight update and it is proportional to the linearity of the curve, while B is a function of A that fits the function within the range of G_{max}, G_{min} and P_{max}. The parameters A and B can be different for depression and potentiation. To simplify the representation, the conductances G_{max} and G_{min} as well as the total number of pulses P_{max}, equal to 91 and 81 respectively for potentiation and depression, are normalized.
Finally, after the fitting is done, a look-up table provided by Pai-Yu Chen et all [48] can be used to extract the corresponding nonlinearity from the normalized A value.

In figure 4.22, the normalized conductance as a function of the normalized pulse, is shown for both Long Term Potentiation (LTP) and Long Term Depression (LTD).

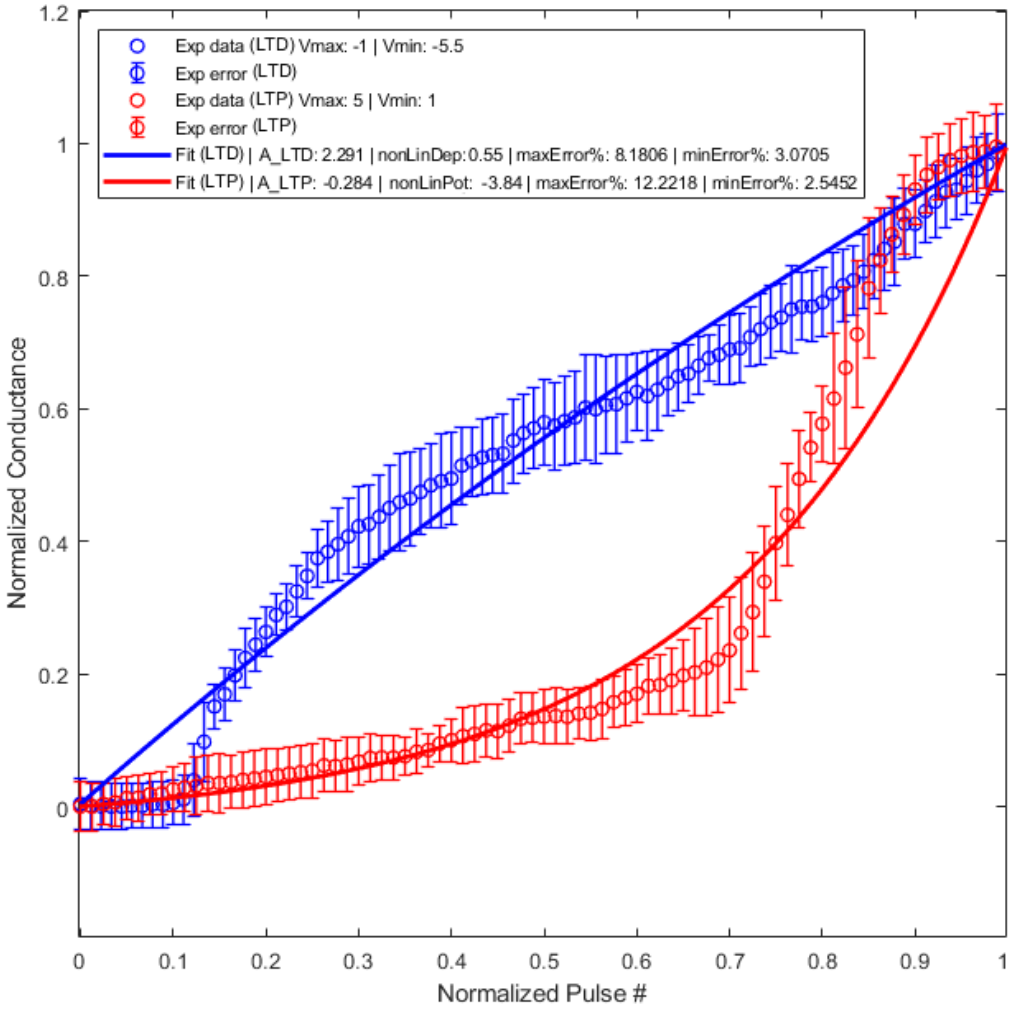


Figure 4.22: Experimental data and relative fit using the device behavioral model of the nonlinear weight update described in equation 4.2, for both potentiation and depression.

A good linearity in the weight update is proved for both potentiation and depression cycles. Since the impact of non-linearity on online learning accuracy is very critical and high accuracy can be achieved only with small non-linearity [47], FinFeFET memristors, as the one characterized in figure 4.22, show promising features for implementation in pseudo-crossbar array.

4.2.3 Possible Optimizations

The main limitation of fabricated FinFeFETs is the heavy reduction of WO_3 semiconductor channel during the crystallization of HZO in its ferroelectric phase with flash lamp annealing.

However, since the switching mechanism, due to polarization screening at the $WO_x - HZO$ interface, becomes more and more negligible by decreasing the channel's resistivity, a more resistive WO_x is desired to further optimize the performances of FinFeFETs.

To do so, the following two main strategies have been employed:

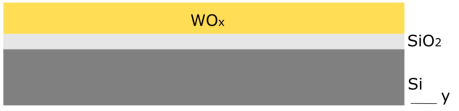
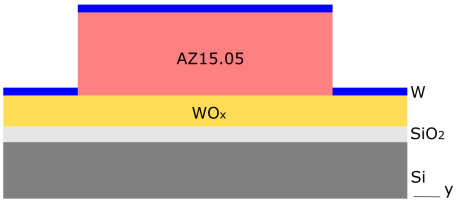
- **RTA oxygen saturation of HZO**:
  After the deposition of HZO by ALD and before FLA crystallization, a rapid thermal annealing step in O_2 atmosphere, is performed in order to have an oxygen saturation of HZO. This could result in less oxygen reduction of tungsten oxide by HZO during its ferroelectric crystallization.
- **Increased oxygen plasma time during ALD of HZO**:
  Increasing the time of oxygen plasma step in each ALD cycle of HZO, may allow the deposition of a fully oxygen saturated HZO, which could result in less oxygen reduction of tungsten oxide by HZO during its ferroelectric crystallization.

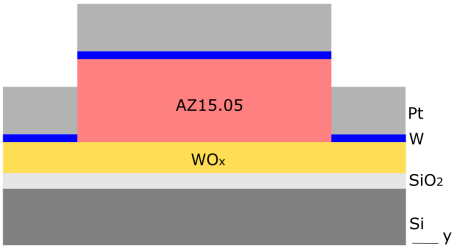
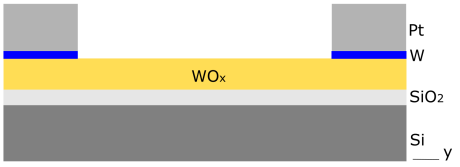
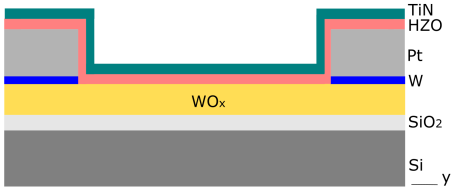
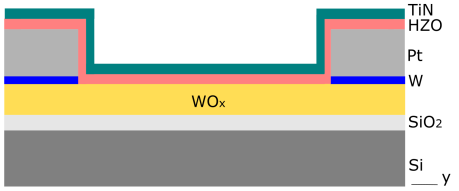
However, since the entire processing of FinFeFET is time expensive and based on e-beam lithography, planar test structures, based on laser lithography, with WO_x as semiconducting channel material, and HZO-TiN gate stack on top, are processed.

Afterwards, CTLM measurements and GIXRD analysis are performed to extract the resistivity of WO_x after HZO crystallization with the first, and prove the correct crystallization in orthorhombic phase of HZO with the second.

To have a common reference to compare and evaluate the different strategies, a sample with standard process flow, summarized in table 4.5, is produced and characterized. Then, the two possible approaches are implemented in two other samples, with slightly modified process flow. It is important to point out that channel properties, such as the electrical resistivity, may behave differently in optical micrometric planar or e-beam nanometric fin structures. This is the reason for which a common reference is needed and the resistivity of already processed FinFeFET channel can not be used to fairly evaluate the two proposed strategies.

Table 4.5: CTLM structures process flow

Step	Process Description	Tool	Ideal Schematic View
1	Deposition of WO_x	ALD	
2	Crystallization of WO_x	RTA	
3	Deposition of W	Sputter	

4	Deposition of Pt	Evaporator	
5	Source-Drain Lift-Off	DMSO	
6	Conformal Deposition HZO - TiN	ALD	
7	Crystallization HZO	ms-FLA	
8	Etching TiN	RIE	
9	Open VIAs through HZO	ICP	

Using the process flow depicted in table 4.5, CTLM test structures are fabricated changing both the gap spacing and the inner contact diameter. In particular, 5 different spacings are available in the layout, yielding 5 different resistance values. In addition, for each gap, circular structures with 3 different diameters are processed. In figure 4.23 a CTLM block is shown.

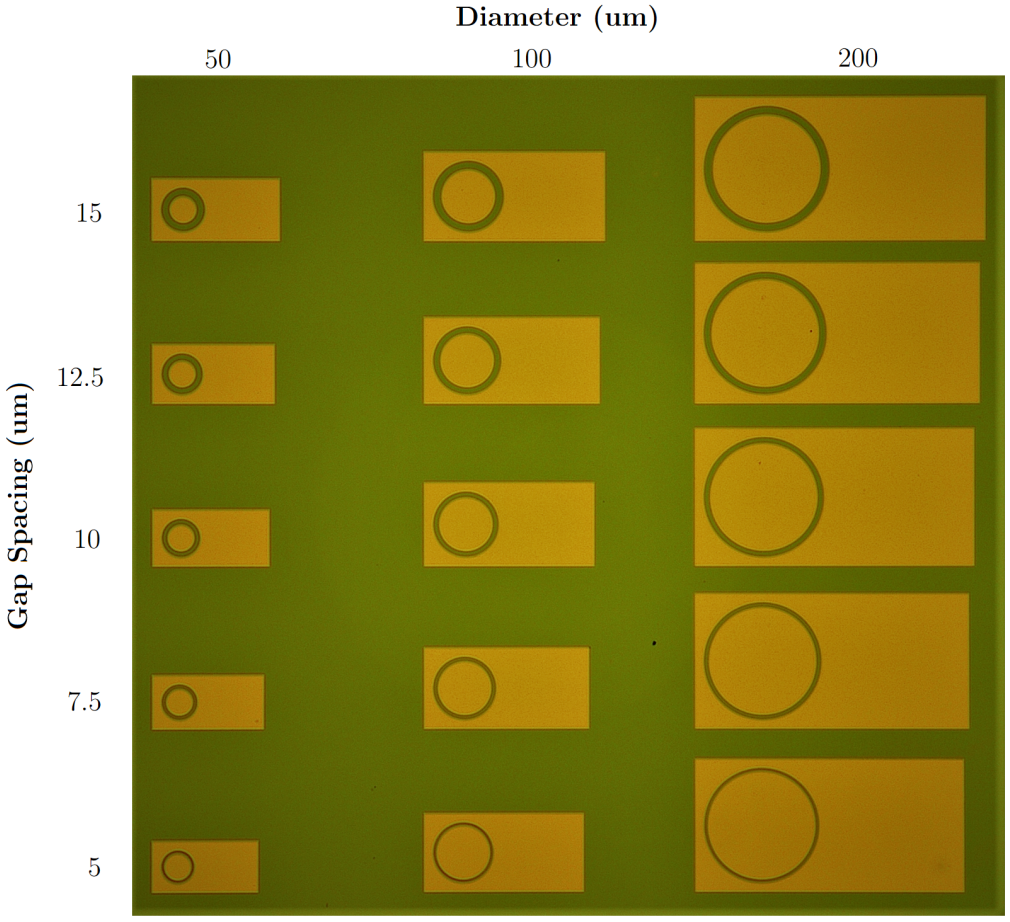


Figure 4.23: Planar view of fabricated CTLM layout.

In order to have statistics and average the extracted electrical parameters on multiple devices, several blocks as the one reported in figure 4.23, are produced. The CTLM measurements are always performed using 4 SMU probes, two to force the current and two to sense the voltage, to have more accurate results. In the following paragraphs, the results for the reference test structures, and for the two proposed alternatives are detailed.

WO_x reference with standard process flow

Reference CTLM test structures with a standard process flow presented in table 4.5, are fabricated. In particular, a standard process flow means that the main steps regarding the crystallization of both WO_x and HZO are the same of previously processed FinFeFETs. The structures consists of a 30 nm thick WO_x layer, crystallized and oxidized at T = 350 °C in O_2 atmosphere using rapid thermal annealing. Then AZ1505 optical resist is used for the W-Pt contact lift-off, which as for FinFeFETs, is done with DMSO heated at T = 130 °C. Afterwards, HZO and TiN are deposited by ALD, 10 nm of both, and then the crystallization occurs by flash lamp annealing. Finally, TiN, used just as capping layer for the HZO crystallization, is removed by reactive ion etching, and VIAs are opened in HZO in correspondence of the two

contacts to be used to probe the WO_x material with CTLM measurements.

In figure 4.24, the measured resistance as a function of the gap spacing, for a specific CTLM structure with diameter of 100 µm, is reported. In particular, the original data, the ones after the application of the correction factor mentioned in sub-section 2.1.2 and the linear fit to corrected data, are shown.

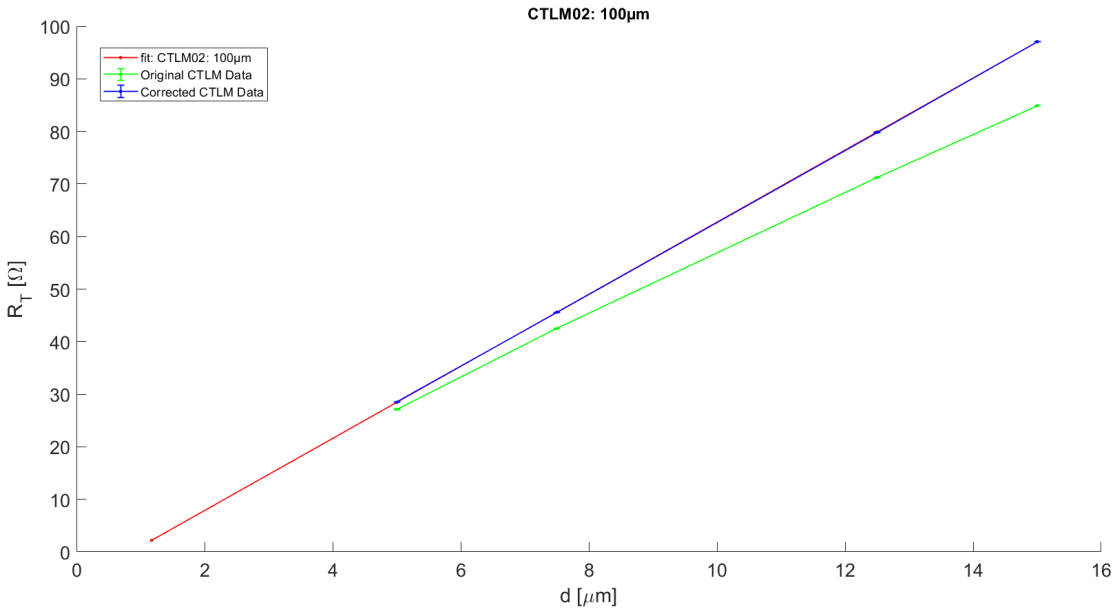


Figure 4.24: Total resistance as a function of gap spacing before and after applying the correction factor. The linear fit well describes the corrected data.

The expected relationship between the resistance and the gap is linear, however since there may be defective circular structures, the data-point with the worst linear fit is always excluded. For this reason in figure 4.24 and in all the following analysis, just 4 of the 5 gaps are considered.

The slightly negative contact resistance R_c predicted by the fit in figure 4.24 is an artifact due to non idealities of the measurements, however being really close to zero, suggests that a good contact between WO_x and $W - Pt$ electrodes is established.

In figure 4.25, all the performed measurements on several CTLM test structures are reported.

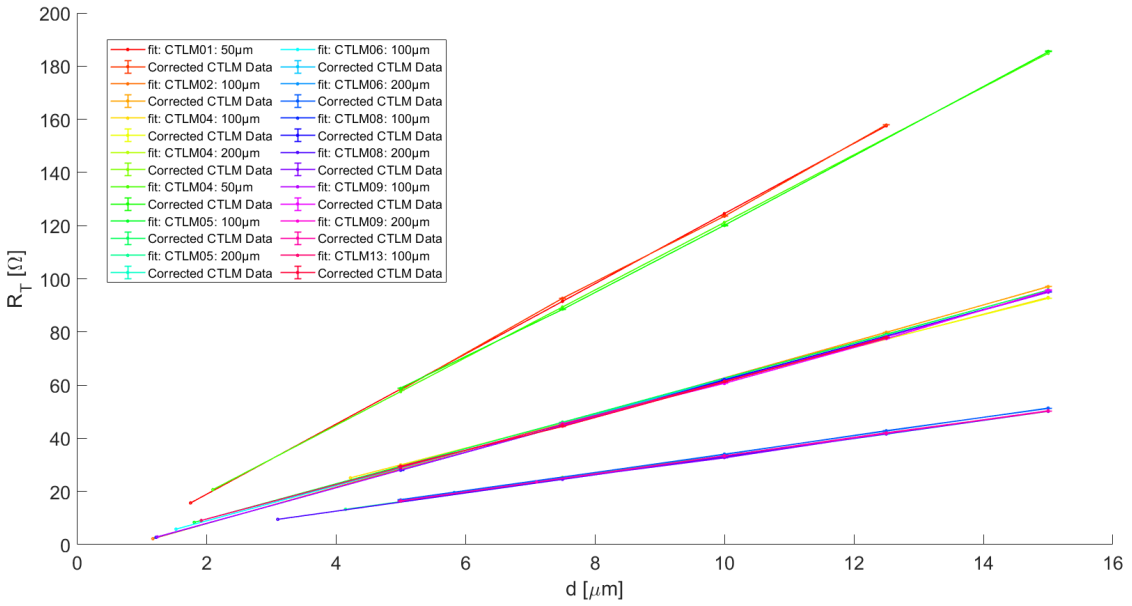


Figure 4.25: CTLM corrected data and their linear fits for all the fabricated standard structures. Three main slopes are evident, related to diameters of 50 µm, 100 µm and 200 µm.

Several identical CTLM blocks, as the one depicted in figure 4.23, are measured, and three main slopes are evident in figure 4.25. The slopes are related to the different diameters of the circular structures. In fact, increasing the diameter causes a decrease of the overall resistance, hence the slope in figure 4.25 is maximum for structures with 50 µm diameter and minimum for those with 200 µm one.

Averaging the CTLM measurements in figure 4.25 allows the extraction of the following electrical parameters of WO_x with the standard process flow:

Table 4.6: WO_x electrical properties with standard process flow in planar structures.

	Rsh (Ω/square)	**Resistivity (Ω· cm)**
Mean	2094	$6.3 \cdot 10^{-3}$
Std. Deviation	56	$1.7 \cdot 10^{-4}$

For FinFeFET optimization, the most interesting electrical parameter in table 4.6 is the electrical resistivity of WO_x. The target of the following analysis is to replicate the same CTLM structures and measurements, just changing some processing steps, and see if WO_x resistivity can be increased, keeping the HZO ferroelectric.

WO_x after the RTA oxygen saturation of HZO

The reduction of WO_3 mainly occurs during crystallization of HZO with flash lamp annealing, hence this strategy consists in performing an additional HZO oxygen saturation step with RTA, just before its crystallization. The process flow of fabri-

cated CTLM test structures is similar to that presented in table 4.5. In this case HZO and TiN, 10 nm of both, are deposited by ALD but not immediately one after the other. In fact after the deposition of HZO, a RTA step of 30 min at T = 300 °C and O_2 flow of 10 sccm, is carried out. Afterwards, TiN layer is deposited and flash lamp annealing, with a pre-heat temperature of 375 °C and flash energy density of 70 J/cm^2, is performed. Finally, the process flow continues as depicted in table 4.5.

CTLM measurements are performed on several structures to investigate how WO_x properties changes with this additional step, keeping all the others identical to the previously characterized reference. In figure 4.26 and in table 4.7, the measurements and the extracted parameters are reported.

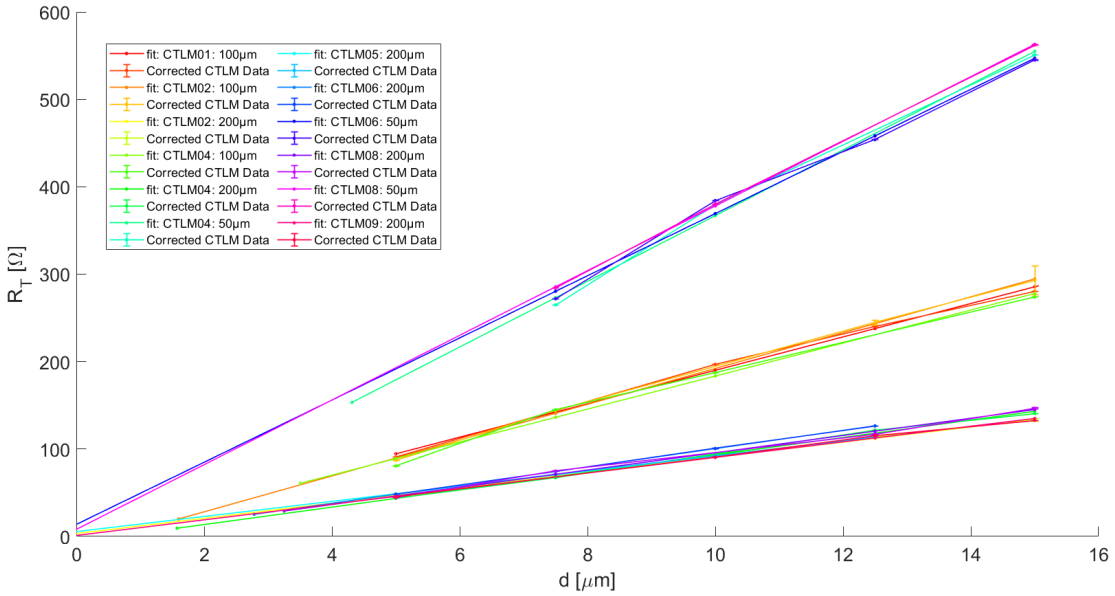


Figure 4.26: CTLM corrected data and their linear fits for all CTLM structures with the RTA additional step. Three main slopes are evident, related to diameters of 50 µm, 100 µm and 200 µm.

Table 4.7: WO_x electrical properties after the 30 min RTA O_2 saturation of HZO.

	Rsh (Ω/square)	**Resistivity (Ω· cm)**
Mean	5933	$1.8 \cdot 10^{-2}$
Std. Deviation	372	$1.1 \cdot 10^{-3}$

Also in this case, the contact resistance R_c is really close to zero, hence a good contact is established. Both sheet resistance R_{sh} and resistivity of WO_x increase of a factor ~ 2.8 with the additional RTA step.

However, this additional step causes a dramatic decrease of the HZO orthorhombic phase content. Grazing incidence X-ray diffraction pattern is shown in figure 4.27, comparing the HZO crystallization in reference sample, and with the additional RTA step.

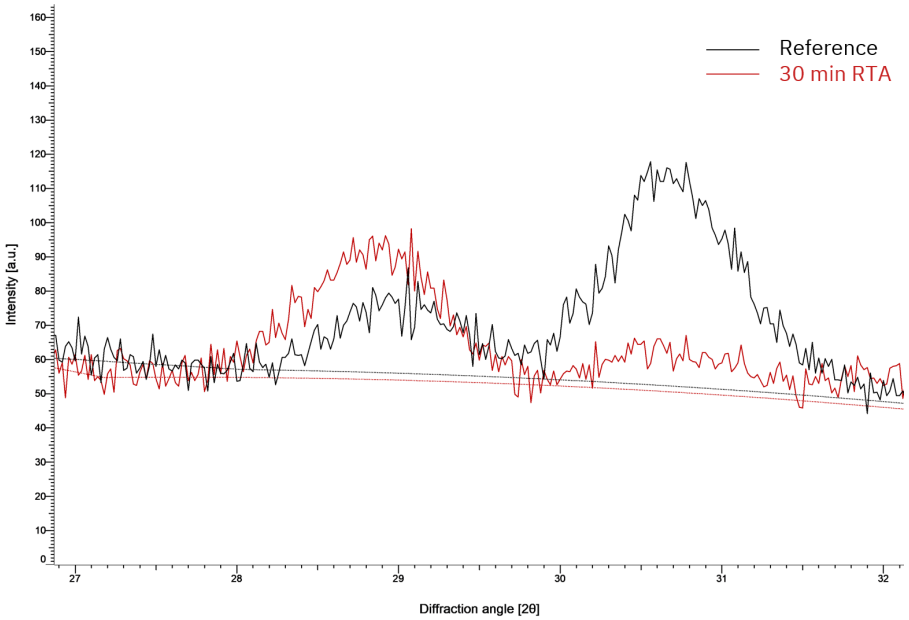


Figure 4.27: The black curve shows the GIXRD pattern of the reference structure with standard process, while the red one that with O_2 based RTA additional step before flash lamp annealing.

The amplitude of the peak at $\sim 30.8°$, due to the overlapped orthorhombic (111), responsible for the ferroelectricity, and tetragonal (011) phases of HZO [44], is considerably decreased. On the other hand, an increase of the peak at $\sim 29°$, probably due to (111) plane of WO_x tetragonal $P42_1m$ phase as suggested by Material Project DFT XRD simulation, is reported.

An additional sample for CTLM measurements is processed, to increase from one side the RTA HZO oxygen saturation time up to 2 h, which could lead to a further increase of the WO_x resistivity, and from the other an increase of FLA preheat temperature and flash energy density to 450 °C and 90 J/cm^2, respectively. However, GIXRD pattern reveals that in this sample the peak at $\sim 30.8°$, hence the ferroelectric properties of HZO, totally disappeared.

This means that even if there is an improvement in terms of WO_x resistivity, this strategy can not be pursued since the ferroelectricity of HZO, responsible of the switching mechanism of FinFeFETs, is removed with this additional RTA oxidation.

WO_x after doubled O_2 plasma time during the ALD of HZO

Additional CTLM test structures are processed basically with the same process flow of the reference sample, depicted in table 4.5, just with a modified ALD step. In fact, during HZO deposition, after the injection of tetrakis-(ethylmethy lamino)hafnium (TEMAH) and ZrCMMM ((MeCp)-2Zr(OMe)(Me)) gaseous precursors, an O_2 RF-plasma is used to act as oxidizer during the deposition. The duration of this plasma based step is doubled in each ALD cycle with respect to the standard reference. As consequence, deposited HZO is expected to be fully O_2 saturated, which could

result in less O_2 reduction of the WO_x during the FLA crystallization.

In figure 4.28 and table 4.8, the CTLM measurements and the extracted parameters are provided.

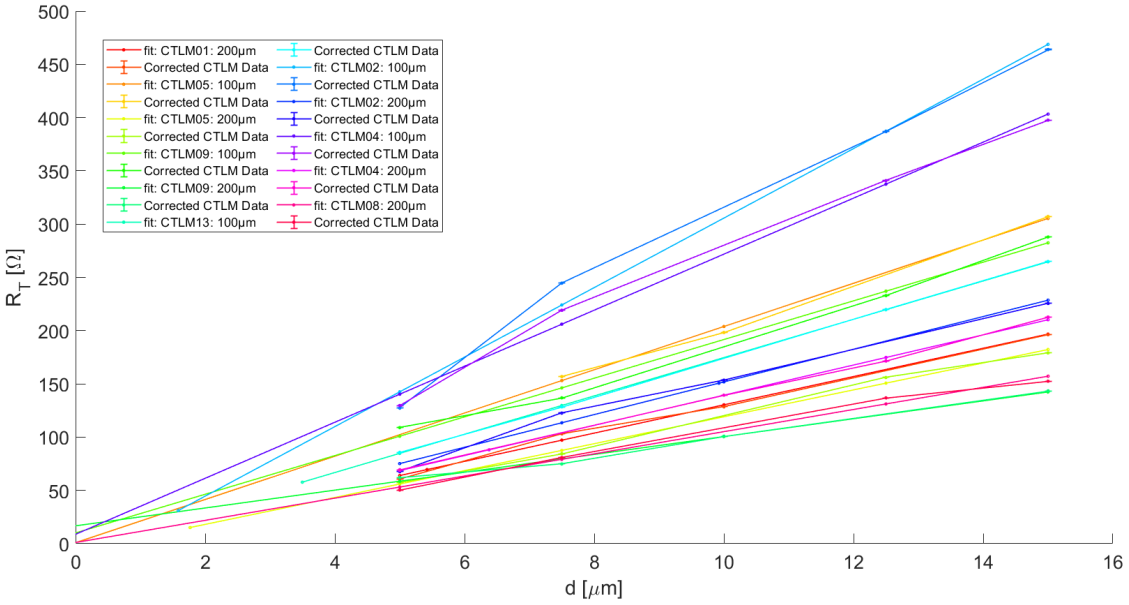


Figure 4.28: The CTLM corrected data and their linear fits for several CTLM structures with modified ALD of HZO.

Table 4.8: WO_x electrical properties with modified ALD of HZO.

	Rsh (Ω/square)	**Resistivity (Ω· cm)**
Mean	7532	$2.3 \cdot 10^{-2}$
Std. Deviation	1714	$5.1 \cdot 10^{-3}$

Also in this case, the contact resistance R_c is really close to zero, hence a good contact is established. Both sheet resistance R_{sh} and resistivity of WO_x increase by a factor of ~ 3.6, with respect to the standard reference, using this modified HZO deposition.

A grazing incidence X-ray diffraction pattern, shown in figure 4.29, is acquired to check the correct crystallization of HZO during standard ms-FLA.

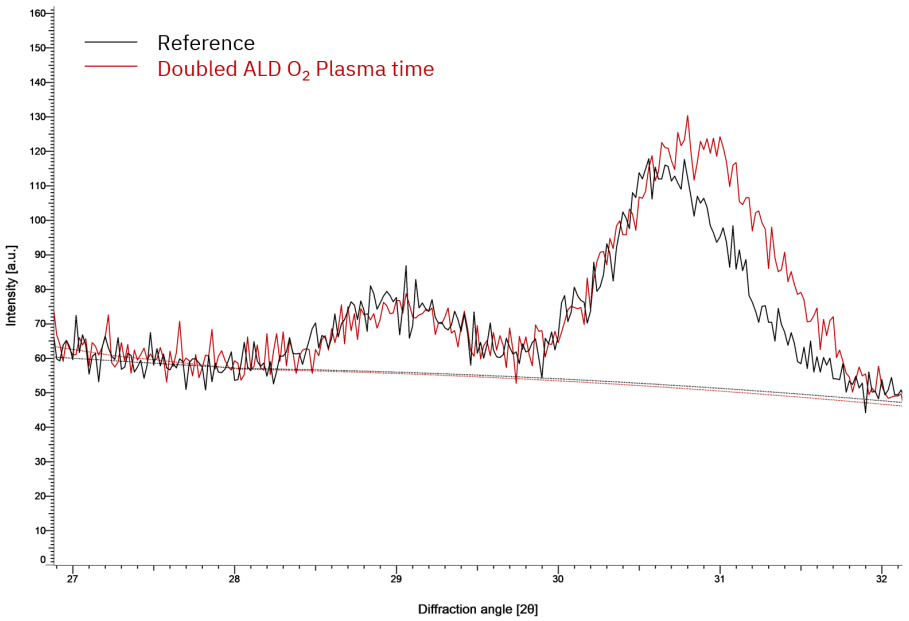


Figure 4.29: The black curve shows the GIXRD pattern of the reference structure with standard process, while the red one that with modified ALD of HZO.

The amplitude of the peak at $\sim 30.8°$, due to the overlapped orthorhombic (111) and tetragonal (011) phases of HZO [44], is slightly increased with respect to the reference sample. This suggests that HZO deposited with this modified ALD step, is still ferroelectric. The peak at $\sim 29°$, probably due to the (111) plane of WO_x tetragonal $P42_1m$ phase as suggested by Material Project DFT XRD simulation, is unperturbed.

A remarkable improvement in terms of WO_x resistivity is proved, as well as the correct crystallization of HZO during standard ms-FLA. This suggests that this strategy can be pursued and FinFeFETs with optimized performances, both in terms of resistive range and switching capability, are expected.

Chapter 5

Conclusion

In this master thesis, planar ferroelectric field effect transistors were electrically characterized, allowing further physical understanding of the transport mechanisms and paving the way for further and more detailed electrical transport simulation. In addition, planar FeFET technology was successfully transferred to a tri-gate sub-10 nm-fin based architecture. $HZO - WO_x$ FinFeFET synaptic devices having an overall footprint $\simeq 10^5$ times smaller than previously characterized planar devices, and showing similar performances in terms of analog switching, were proved. Finally, a possible processing optimization to improve overall performances and face the reduced resistive range, which is the main limitation of processed FinFeFET devices, is provided. Nevertheless, further refinement of the fabrication process, as well as a new metal-oxide solution as channel material, such as tantalum pentoxide instead of tungsten oxide, may be explored in the near future.

Moreover, as next step, the integration of these devices into a functional memristive pseudo-crossbar array, may allow the demonstration of AI algorithms using analog neuromorphic architectures, overcoming the von Neumann bottleneck.

Bibliography

[1] 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.

[2] Qiangfei Xia and J. Joshua Yang. *Memristive crossbar arrays for brain-inspired computing*. 2019. DOI: 10.1038/s41563-019-0291-x.

[3] Tayfun Gokmen and Yurii Vlasov. "Acceleration of Deep Neural Network Training with Resistive Cross-Point Devices: Design Considerations". In: *Frontiers in Neuroscience* 10 (2016), p. 333. ISSN: 1662-453X. DOI: 10.3389/fnins.2016.00333. URL: https://www.frontiersin.org/article/10.3389/fnins.2016.00333.

[4] Daniele Ielmini and H.-S. Philip Wong. "In-memory computing with resistive switching devices". In: *Nature Electronics* 1.6 (2018), pp. 333–343. DOI: 10.1038/s41928-018-0092-2. URL: https://app.dimensions.ai/details/publication/pub.1104412702%20and%20https://re.public.polimi.it/bitstream/11311/1056513/1/nature_18_preprint.pdf.

[5] W.H. Kautz. "Cellular Logic-in-Memory Arrays". In: *IEEE Transactions on Computers* C-18.8 (1969), pp. 719–727. DOI: 10.1109/T-C.1969.222754.

[6] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[7] Rong Gu et al. "Improving Execution Concurrency of Large-Scale Matrix Multiplication on Distributed Data-Parallel Platforms". In: *IEEE Transactions on Parallel and Distributed Systems* 28.9 (2017), pp. 2539–2552. DOI: 10.1109/TPDS.2017.2686384.

[8] Rainer Waser et al. "Introduction to Nanoionic Elements for Information Technology". In: *Resistive Switching*. John Wiley Sons, Ltd, 2016. Chap. 1, pp. 1–30. ISBN: 9783527680870. DOI: https://doi.org/10.1002/9783527680870.ch1. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9783527680870.ch1. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527680870.ch1.

[9] Shimeng Yu. "Neuro-inspired computing with emerging nonvolatile memorys". In: *Proceedings of the IEEE* 106.2 (2018), pp. 260–285. DOI: 10.1109/JPROC.2018.2790840.

[10] Simone Raoux, Wojciech Wełnic, and Daniele Lelmini. "Phase change materials and their application to nonvolatile memories". In: *Chemical Reviews* 110 (1 2010). ISSN: 00092665. DOI: 10.1021/cr900040x.

[11] Wei Wu et al. "Improving Analog Switching in HfO2 Based Resistive Memory With a Thermal Enhanced Layer". In: *IEEE Electron Device Letters* 38.8 (2017), pp. 1019–1022. DOI: 10.1109/LED.2017.2719161.

[12] Jiyong Woo et al. "Improved synaptic behavior under identical pulses using AlOx/HfO2 bilayer RRAM array for neuromorphic systems". In: *IEEE Electron Device Letters* 37 (8 2016). ISSN: 07413106. DOI: 10.1109/LED.2016.2582859.

[13] M. Ye Zhuravlev et al. "Tunneling electroresistance in ferroelectric tunnel junctions with a composite barrier". In: *Applied Physics Letters* 95 (5 2009). ISSN: 00036951. DOI: 10.1063/1.3195075.

[14] Jiantao Zhou, Kuk Hwan Kim, and Wei Lu. "Crossbar RRAM arrays: Selector device requirements during read operation". In: *IEEE Transactions on Electron Devices* 61 (5 2014). ISSN: 00189383. DOI: 10.1109/TED.2014.2310200.

[15] J. Robertson. *High dielectric constant oxides*. 2004. DOI: 10.1051/epjap:2004206.

[16] Tim Böscke et al. "Ferroelectricity in hafnium oxide: CMOS compatible ferroelectric field effect transistors". In: *Electron Devices Meeting, 1988. IEDM '88. Technical Digest., International* 99 (Dec. 2011), pp. 24.5.1–24.5.4. DOI: 10.1109/IEDM.2011.6131606.

[17] T. S. Böscke et al. "Ferroelectricity in hafnium oxide thin films". In: *Applied Physics Letters* 99 (10 2011). ISSN: 00036951. DOI: 10.1063/1.3634052.

[18] Jiong Yan, Yue Kuo, and Jiang Lu. "Zirconium-doped hafnium oxide high-k dielectrics with subnanometer equivalent oxide thickness by reactive sputtering". In: *Electrochemical and Solid-State Letters* 10 (7 2007). ISSN: 10990062. DOI: 10.1149/1.2730720.

[19] Mattia Halter et al. "Back-End, CMOS-Compatible Ferroelectric Field-Effect Transistor for Synaptic Weights". In: *ACS Applied Materials and Interfaces* 12 (15 2020). ISSN: 19448252. DOI: 10.1021/acsami.0c00877.

[20] Toshikazu Hirose, Iwazo Kawano, and Minoru Niino. "Electrical Conductivity of Tungsten Trioxide (WO3)". In: *Journal of the Physical Society of Japan* 33 (1 1972). ISSN: 13474073. DOI: 10.1143/JPSJ.33.272.

[21] Mark Bohr and Kaizad Mistry. "Intel's Revolutionary 22 nm Transistor Technology". In: *intel.com* (2011).

[22] Alan Wadsworth. "The Parametric Measurement Handbook 4th Edition". In: *Keysight* (2017).

[23] Fu Chien Chiu. *A review on conduction mechanisms in dielectric films*. 2014. DOI: 10.1155/2014/578168.

[24] J. G. Simmons. "Richardson-schottky effect in solids". In: *Physical Review Letters* 15 (25 1965). ISSN: 00319007. DOI: 10.1103/PhysRevLett.15.967.

[25] J. H. Klootwijk and C. E. Timmering. "Merits and limitations of circular TLM structures for contact resistance determination for novel III-V HBTs". In: 2004. DOI: 10.1109/icmts.2004.1309489.

[26] Dieter K. Schroder. 2005. DOI: 10.1002/0471749095.

[27] A. Lamberti. "Lecture: Technologies for Nanoscience : Electron Microscopy". In: 2019.

[28] Natasha Erdman and David C. Bell. *SEM Instrumentation Developments for Low kV Imaging and Microanalysis*. 2012. DOI: 10.1002/9781118498514.ch2.

[29] C.A. Volkert and A.M. Minor. *Focused ion beam microscopy and micromachining*. 2007. DOI: 10.1557/mrs2007.62.

[30] F. Giorgis. "Lecture: Materials and characterizations for Micro and Nanotechnologies". In: 2019.

[31] R. Schmidt, M. Parlak, and A. W. Brinkman. "Control of the thickness distribution of evaporated functional electroceramic NTC thermistor thin films". In: *Journal of Materials Processing Technology* 199 (1 2008). ISSN: 09240136. DOI: 10.1016/j.jmatprotec.2007.08.014.

[32] David G. Lishan. "Plasma Etching: Comparing PE, RIE and ICP-RIE". In: *Plasma-Therm LLC* (2020).

[33] S.M. Sze and Kwok K. Ng. *Physics of Semiconductor Devices*. 2006. DOI: 10.1002/0470068329.

[34] Chang Mou Wu et al. *Recent advances in tungsten-oxide-based materials and their applications*. 2019. DOI: 10.3389/fmats.2019.00049.

[35] Steven A. Vitale et al. "Work-function-tuned TiN metal gate FDSOI transistors for subthreshold operation". In: *IEEE Transactions on Electron Devices* 58 (2 2011). ISSN: 00189383. DOI: 10.1109/TED.2010.2092779.

[36] Mathias Mews, Lars Korte, and Bernd Rech. "Oxygen vacancies in tungsten oxide and their influence on tungsten oxide/silicon heterojunction solar cells". In: *Solar Energy Materials and Solar Cells* 158 (2016). ISSN: 09270248. DOI: 10.1016/j.solmat.2016.05.042.

[37] Fu-Chien Chiu and Chih-Ming Lai. "Optical and electrical characterizations of cerium oxide thin films". In: *Journal of Physics D: Applied Physics* 43 (7 2010). ISSN: 0022-3727. DOI: 10.1088/0022-3727/43/7/075104.

[38] A. N. Saxena and K. L. Mittal. "Optical and dielectric constants of hafnium and its anodic oxide films". In: *Journal of Applied Physics* 46 (6 1975). ISSN: 00218979. DOI: 10.1063/1.321959.

[39] Andrey Sergeevich Sokolov et al. "Influence of oxygen vacancies in ALD HfO 2-x thin films on non-volatile resistive switching phenomena with a Ti/HfO 2-x /Pt structure". In: *Applied Surface Science* 434 (2018). ISSN: 01694332. DOI: 10.1016/j.apsusc.2017.11.016.

[40] INTEC Photonics Research Group. "Ipkiss 2.4". In: *Gent University/imec* (2002-2013).

[41] K. Miyake, H. Kaneko, and Y. Teramoto. "Electrical and optical properties of reactively sputtered tungsten oxide films". In: *Journal of Applied Physics* 53 (3 1982). ISSN: 00218979. DOI: 10.1063/1.330649.

[42] David B. Cordes, Paul D. Lickiss, and Franck Rataboul. "Recent developments in the chemistry of cubic polyhedral oligosilsesquioxanes". In: *Chemical Reviews* 110 (4 2010). ISSN: 00092665. DOI: 10.1021/cr900201r.

[43] Éamon O'Connor et al. "Stabilization of ferroelectric Hf x Zr 1-x O 2 films using a millisecond flash lamp annealing technique". In: *APL Materials* 6 (12 2018). ISSN: 2166532X. DOI: 10.1063/1.5060676.

[44] Min Hyuk Park et al. *Evolution of phases and ferroelectric properties of thin Hf 0.5Zr0.5O2 films according to the thickness and annealing temperature*. 2013. DOI: 10.1063/1.4811483.

[45] Y. Zhou et al. "Mechanisms of imprint effect on ferroelectric thin films". In: *Journal of Applied Physics* 98 (2 2005). ISSN: 00218979. DOI: 10.1063/1.1984075.

[46] André Chanthbouala et al. "A ferroelectric memristor". In: *Nature Materials* 11 (10 2012). ISSN: 14764660. DOI: 10.1038/nmat3415.

[47] Pai Yu Chen, Xiaochen Peng, and Shimeng Yu. "NeuroSim: A circuit-level macro model for benchmarking neuro-inspired architectures in online learning". In: *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 37 (12 2018). ISSN: 02780070. DOI: 10.1109/TCAD.2018.2789723.

[48] Pai Yu Chen, Xiaochen Peng, and Shimeng Yu. "NeuroSim+: An integrated device-to-algorithm framework for benchmarking synaptic devices and array architectures". In: 2018. DOI: 10.1109/IEDM.2017.8268337.