Lund University

Department of Electrical and Information Technology

Nano Electronics Group

HfO_2 and ITO Resistive Random-Access Memory

Author:
Mattias Åstrand

Supervisor:
Prof. Lars-Erik Wernersson
Co-supervisors:
Dr. Karl-Magnus Persson
Dr. Rainer Timm

Thesis submitted for the degree of Master of Science

Autumn 2019 - Spring 2020

Lund
University

Abstract

The purpose of this work is of evaluating the choice of HfO_2 and ITO as the dielectric and the top electrode in high performance resistive random-access memory (RRAM), respectively. The study is twofold, as it quantifies performance according to standard figures of merit for this technology, as well as provides insight into the physics of current conduction for this material choice. Results are achieved thanks to the probing of structures processed by the Nano Electronics group in Lund, followed by direct observations on the produced $I - V$ characteristics and more involved data analysis. Specifically, different models according to different mechanisms of current conduction in a dielectric are fitted to the measured data, and physical parameter extraction is used as a mean to evaluate goodness of fit. It is discovered that several modes that were thought to potentially describe conduction in RRAM can instead not be the case, for confusing (if not impossible) parameters are extrapolated for them. An important result is thus that no matter how great a fit may be numerically, one cannot simply claim that a given mechanism of conduction is a good physical interpretation of a studied system without corroborating its fit with relevant parameter extraction. It is concluded that the only reasonable and available interpretation of measured devices is that of Ohmic type conduction, followed by space-charge (at higher voltages) in the high resistance state of RRAM, and solely Ohmic type in the low resistance state. The latter relates to not observing conduction through a barrier, which was expected as HfO_2 and ITO RRAM has previously been found to be modulated by a self-compliance effect. Based on this interpretation, a computer model is developed in Python, and its output is put to the test by comparison with data from a different sample (i.e. not related to the observations that lead to the creation of the model itself). The universality of the model is found to be satisfying, hinting at the achieved interpretation being adequate. Concerning raw performance of the studied devices, resistive switching at low and steady voltages is observed, which relates to envisioning low power operation for HfO_2 and ITO structures.

Popular Science

The computer memory technology that we currently rely on is either fast, when it comes to writing and reading programmed states, yet volatile (i.e. it loses its programmed state if power is not being supplied to it), or non-volatile yet slow in terms of write operations and facing issues with further scaling (reduction in size). It should then be obvious that finding a replacement for the above is an important task, as volatility is an undesired trait that implies power consumption, and slow and non-scalable devices are inevitably going to become a bottleneck for new, high performance systems. Multiple new technologies have been proposed, yet a common opinion seems to be that resistive random-access memory (RRAM) is the most promising alternative for the future of computer memory.

To comprehend the benefits of RRAM a good approach is to first understand how it works, and only then draw conclusions about its implications for the current state of memory applications. Imagine the task of having to make a stream of water pass through a patch of compact soil. The latter, in its pristine state, will impose a high resistance to water flow, as no paths are available in it for water to cross it. However, if a high enough water inlet pressure is applied, the soil will eventually reshape and a conductive path will arise. Subsequently, one could use a shovel to pat on the soil and cause one end of the patch to become compact again, and after that apply a high pressure to reopen the path for facilitated water flow. That is, one can switch between high and low water conductivity by applying appropriate stress on the soil. If we now think of the soil as an insulator, of the water as electrons, and of soil manipulation as the application of electrical bias, we suddenly begin to understand RRAM basics.

The reshaping of soil resembles quite well the principles of operation of RRAM, and understanding how these principles are put to practice becomes easier once a clear picture of an RRAM cell is had in mind. The latter may be visualised as a sandwich-like structure, consisting of an electrode-insulator-electrode stack. Given an appropriate initial electrical bias, a conductive filament arises in the insulator, and the latter's pristine high electrical resistance is changed to a low resistance; this is known as a FORM operation. The tip of the filament can then be ruptured to reinstate high resistance (RESET), and regrown to yield low resistance (SET). High and low resistance are the "0s" and "1s" of RRAM, and achieving them makes the programming of cells possible.

The switching between resistance states has been shown to correlate with the drift of oxygen ions and of vacancies that are left behind by moving oxygen. The response of these entities to electric fields in RRAM, which arise when different biases are applied, is very fast, and allows for writing speeds down to no more than a billionth of a second! Moreover, programmed states are permanent, and do not require power to be retained, and the simplicity of cell design makes scalability a straightforward task for this technology. Further benefits may be achieved with appropriate material selection. In fact, it has been shown that using hafnium oxide as the insulator and indium-tin-oxide as an active electrode results in the possibility to operate RRAM units at biases down to a fraction

of a volt. This, together with current limitation, opens the door for the realisation of low power-consumption devices! All in all, a study dedicated to characterising RRAM units featuring promising materials seems like a perfect starting point for deepening our knowledge about systems that may make the necessary difference in the electronics of the world of tomorrow.

Acknowledgements

First of all, I would like to thank professor Lars-Erik Wernersson for promptly welcoming me into the Nano Electronics group, and ultimately making this project a reality. I would also like to thank Dr. Karl-Magnus Persson for the countless hours of discussion, and for providing assistance 24/7, and Dr. Rainer Timm for readily accepting to co-supervise this thesis, for being available throughout the academic year and providing great guidance through an outsider's perspective. Further, I am grateful to have gotten the chance to know better and work closely with with a fellow student, Oresthes, who I now consider a dear friend of mine.

The pursuit of a master's degree in physics inevitably came with some stressful moments, and it would not have been as easy to overcome these had it not been for the love and support that I am constantly given by my family and friends. I am lucky to be surrounded by people that care about me so much, and cannot begin to describe my gratefulness. I want to thank my brother and best friend Lucas, my parents Giuliana and Pär, and my girlfriend Maria for talking to me every day, pushing me to do better, and reminding me that no matter what, they are always there for me. I also want to thank my family in Stockholm, Koster, Göteborg, and Rome - we all live so far from each other, yet we manage to keep contact, and our joint calls make me feel like I am never alone. Last but not least I want to thank my dear friends form Arolo Sueg and the D-Crew for distracting me when necessary and keeping me sane - you always manage to add an extra layer of spice and fun to my free time. I am incredibly happy to have met you, for everything that we have been through together, and am thrilled about what is yet to come!

Contents

Abstract i
Popular Science ii
Acknowledgements iv
List of Figures vii
List of Tables ix
Acronyms and Abbreviations x

1 Introduction **1**
1.1 Discovery and Development of RRAM 1
1.2 Current Memory Technology 3
1.3 Benefits (and Drawbacks) of RRAM 3

2 Scientific Background **5**
2.1 Programming Operations 5
2.2 Classification 6
2.2.1 Non-polar RRAM 6
2.2.2 Bipolar RRAM 6
2.2.3 Anion-type RRAM 7
2.2.4 Cation-type RRAM 8
2.2.5 Carbon-based RRAM 8
2.2.6 RRAM Comparison 8
2.3 HfO_2 and ITO RRAM 9
2.4 Unoptimised Operation and Detrimental Effects 10
2.5 Charge Transport Mechanisms 11
2.5.1 Electrode-Limited Mechanisms 11
2.5.2 Bulk-Limited Mechanisms 12

3 Method **16**
3.1 Figures of Merit 16
3.1.1 Basic Quantities and How to Obtain Them 16
3.1.2 Benchmarking RRAM 17
3.2 Device Processing 19
3.3 Measurement Setup and Procedure 20
3.4 Data Analysis 20
3.4.1 Extrapolating Figures of Merit 21
3.4.2 Identifying Characteristic Modes of Conduction 21
3.4.3 DC Characterisation - Coding in Python 22
3.5 RRAM Modelling 23

4 Results and Discussion **25**
4.1 DC Measurements . . . 25
4.1.1 FORM Operation . . . 25
4.1.2 Switching Behaviour . . . 26
4.1.3 Current Conduction - Qualitative Analysis . . . 28
4.1.4 Current Conduction - HRS Quantitative Analysis . . . 30
4.1.5 Current Conduction - LRS Quantitative Analysis . . . 34
4.2 RRAM Modelling . . . 35

5 Conclusion and Outlook **38**

A Python Code for Plotting of and Fitting to DC Measurements **40**
A.1 Imported Libraries . . . 40
A.2 SET- and RESET-point Identification . . . 40
A.3 Making and Plotting Linear Regressions . . . 40
A.4 Plotting and Fitting in Schottky Scale . . . 41
A.5 Plotting and Fitting in Other Scales . . . 41

B Parameter Extraction - Formulas and Code **42**
B.1 Schottky Emission . . . 42
B.2 Thermionic Emission . . . 42
B.3 Pool-Frenkel Effect . . . 43
B.4 Hopping Conduction . . . 43
B.5 Mott-Gurney Law (SCLC) . . . 43
B.6 Ohmic Conduction . . . 44
B.7 Note on Other Physical Quantities . . . 44
B.8 Reliability of Extracted Parameters . . . 45

C Python Model **47**
C.1 Imported Libraries . . . 47
C.2 SET- and RESET-point Generating Functions . . . 47
C.3 Current Defining Functions . . . 47
C.4 SET Curve Generation . . . 47
C.5 RESET Curve Generation . . . 48
Bibliography . . . 49

List of Figures

1.1 A visualisation of the growth in research on RRAM, resistive switching and memristors since Chua's publication (dashed green), enhanced after HP group's paper (dashed red). The vertical axis on the left counts the number of publications released every year on these topics, whist that on the right shows the interest (in terms of accumulated citations) that Chua's paper (green) and HP's group (red) have recently obtained. Publication and citation data were obtained from Web of Science [4]. 2

2.1 A representation of the different programming operations in RRAM: FORM, RESET and SET. 5
2.2 A simple representation of how a CF is formed in an anion-type RRAM, such as the structure considered in this thesis. 9
2.3 Band diagram that qualitatively shows the difference between the three discussed electrode-limited conduction mechanisms. 13

3.1 A representation of typical DC characteristics of RRAM. 17
3.2 Definition of the TiN BE by means of ALD, UV lithography and etching. . 20
3.3 ALD of the HfO_2 switching layer, followed by TE definition with UV lithography and ITO sputtering. 20
3.4 $I-V$ characteristics of a simulated RRAM device. 22
3.5 $I-V$ characteristics of a simulated RRAM device, represented in different scales: linear (Ohmic) scale (a); Mott-Gurney law (b). 23
3.6 Linear fit to the HRS in the SET sweep of a simulated device. 23

4.1 FORM DC measurements for the considered devices (a), alongside the empirical cumulative distribution function for the form events (b). 25
4.2 $I-V$ characteristics of a normal working device (a), a very reproductive device (b) and a non-working device (c). In blue are curves that are thought to be quite centred, thus representing their respective sample well. 26
4.3 Empirical cumulative distribution functions as calculated by considering the average set and reset behaviour of the devices that were measured in this thesis. 28
4.4 Plots of the pre-SET $I-V$ characteristics of the device in figure 4.2a in the scales relevant to: Schottky emission (a), Fowler-Nordheim tunnelling (b), thermionic field emission (c), Pool-Frenkel effect (d), hopping (e), ohmic conduction (f), Mott-Gurney law (g), Frenkel enhanced SCLC (h). Finally, a plot in log-log scale is included (i) to facilitate the visualisation of power laws. 29
4.5 Post-SET $I-V$ characteristics of the device in figure 4.2a. 30

4.6 A visualisation of how LRS resistance R_{LRS} and filament cross section S behave as a function of parameter analyser current compliance. 34
4.7 Reproduction of RRAM $I-V$ characteristics from a given device with a Python model based on structures from a different sample. (a), (b) and (c) show attempts at retracing different SET (blue) and RESET (orange) behaviours with the aid of the model, as compared to real measurements (dashed grey). HRS model parameters are of $R_{HRS} = 1.4\,\mathrm{M\Omega}$ (a), $19\,\mathrm{M\Omega}$ (b), $200\,\mathrm{G\Omega}$ (c); $\mu = 2.5{\cdot}10^{-5}\mathrm{m^2V^{-1}s^{-1}}$ (a), $1.6{\cdot}10^{-5}\mathrm{m^2V^{-1}s^{-1}}$ (b), $2{\cdot}10^{-10}\mathrm{m^2V^{-1}s^{-1}}$ (c); $x =$ ca. 1 nm for (a), (b), (c); $S =$ ca. 1 nm^2 for (a), (b) and (c) alike, given that the LRS is similar for the three ($R_{LRS} =$ ca. $10\,\mathrm{k\Omega}$). 36
4.8 Reproduction of the same characteristics as in figure 4.7 (a) and (b), assuming same μ ($5{\cdot}10^{-6}\mathrm{m^2V^{-1}s^{-1}}$) and varying x (0.6 nm and 1.4 nm for (a) and (b), respectively). 37

List of Tables

4.1 Physical parameter extraction results. 31

B.1 Values of unknown parameters as reasonably guessed and/or researched for the purpose of parameter extraction in this work. 45

Acronyms and Abbreviations

ALD	Atomic layer deposition
BE	Bottom electrode
CBRAM	Conductive-bridge random-access memory
CF	Conductive filament
CMOS	Complementary metal-oxide-semiconductor
CPU	Central processing unit
DC	Direct current
DRAM	Dynamic random-access memory
HD	Hyperdimensional (as in hyperdimensional computing)
HDD	Hard disk drive
HRS	High resistance state
ITO	Indium tin oxide
LRS	Low resistance state
MIM	Metal-insulator-metal
MLP	Multi-level programming
Ox-RRAM	Oxygen-based RRAM
PR	Photoresist
RRAM	Resistive random-access memory
RW	Resistance window
SCLC	Space charge limited current
SRAM	Static random-access memory
SSD	Solid state drive
TE	Top electrode
UV	Ultraviolet

Chapter 1

Introduction

Ever since the invention of the transistor, technology has advanced at a rapid pace. This has been possible especially thanks to the formulation and continuous realisation of Moore's law [1], which states that the number of transistors per chip area doubles approximately every two years. Transistors are the fundamental building blocks of computer processors, and more of them working together in a confined area inevitably results in more computational power. It is thus easy to focus on the benefits of research on existing and new transistor technologies, as success in this area translates to development of technological equipment in terms of computational capabilities. However, more computations go hand in hand with an increasing demand for excellent memory technology, both to support processors during computations, and to store the outputs of the latter in a permanent manner. Therefore, one should think of research dedicated to memory technology as just as important as that on transistors. In fact, if it were not for the realisation of better computer memory, technological advances would come to a halt, bottlenecked by extremely fast processors having to communicate with much slower and perhaps not capable enough storage units.

This thesis sets itself the task of exploring the benefits brought forth by an emerging technology, resistive random-access memory (RRAM), and understanding how this may change the future of computer memory. In particular, stacks of hafnium dioxide (HfO_2) and indium tin oxide (ITO) are considered, and it is investigated whether this material choice results in optimal devices thanks to the evaluation of relevant metrics. Moreover, computational and simulational tools are developed to gain insight on the behaviour of RRAM while under electrical bias, and to provide a basis for circuit simulations in further studies where computer code is necessary to represent RRAM.

The remainder of this chapter is dedicated to discussing the current state of the memory market, some history on RRAM and how this may revolutionise computer memory as we know it; chapter 2 focuses on giving more scientific context and explaining the theory that is necessary to interpret the main results of this thesis; the purpose of chapter 3, instead, is of familiarising the reader with the experimental and computational procedures that were used to collect results; the latter are presented and discussed in chapter 4, whilst an overall conclusion and morals are drawn in chapter 5.

1.1 Discovery and Development of RRAM

The history of RRAM begins as early as 1971, when Leon O. Chua published a paper describing the mathematics of a circuit element that had so far been neglected: the

"memristor" [2]. Chua's first observation was that known two-terminal circuit elements, the resistor, the capacitor and the inductor, could be described mathematically thanks to some relationships that exist between fundamental circuit variables: current I, voltage V, charge q and flux-linkage φ (integral of applied voltage over time). His second observation was that out of all possible combinations of circuit variables and their relationships, one relationship in particular had not been studied yet: that between q and φ. Further in his paper, he came to theorising and proving that this relationship should be described by equation 1.1, where M is the "memristance" of a given memristor.

$$M(q) = \frac{d\varphi(q)}{dq} \tag{1.1}$$

Given that M is in units of Ω, its dependence on q may be interpreted as a memristor having a variable resistance, which is dependent on the amount of charge that has flowed through it over time. This variable resistor may be programmed by "switching" between low and high resistance states, and the memristor used as a memory cell.

In the beginning, Chua's publication did not receive much attention, as no memristor was realised in practice to corroborate his work. The area of research on resistive switching remained dormant for a few decades, until HP labs' group of Dimitri B. Strukov et al. was able to apply Chua's theory in the description of crossbar metal-insulator-metal (MIM) arrays [3]. In fact, they were able to experimentally verify a change in resistance in the titanium oxide of their MIM stacks, fitting Chua's predictions. Ever since this breakthrough, research in resistive switching devices has flourished, and memory cells relying on this technology were donned with the term RRAM (specifically, by SHARP corporation). The growth in interest in RRAM can be visualised thanks to figure 1.1, which quantifies the efforts in research on resistive switching and RRAM after the above publications, and how relevant these publications have become.

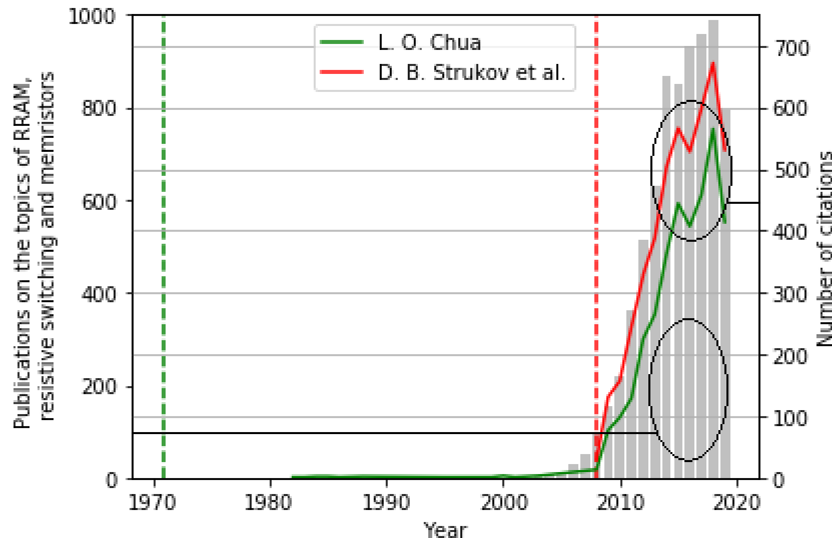


Figure 1.1: A visualisation of the growth in research on RRAM, resistive switching and memristors since Chua's publication (dashed green), enhanced after HP group's paper (dashed red). The vertical axis on the left counts the number of publications released every year on these topics, whist that on the right shows the interest (in terms of accumulated citations) that Chua's paper (green) and HP's group (red) have recently obtained. Publication and citation data were obtained from Web of Science [4].

1.2 Current Memory Technology

The reasons why an advance in memory technology is necessary stem from the non-ideal characteristics of current computer memory. One can start by subdividing the latter into two groups according to the task it fulfils for a computer: there is primary storage, which acts as the memory component directly available to the central processing unit (CPU), and secondary storage, which is used to store processed data in the long run. The former is required to have extremely fast programming (read and write) speeds, as slow operation is a direct bottleneck to computations performed by the CPU. The latter, instead, must focus on its capability to retain data in the long run, and its read and write speeds are but secondary traits. Because of this distinction, different technologies are being implemented for different storage purposes in modern computers.

Embedded static random-access memory (SRAM), working together with dynamic random-access memory (DRAM) directly linked to the CPU via memory "buses", are the technologies that dominate the market when it comes to primary storage, ought to their excellent read and write speeds. The main issue with these technologies is that they are volatile, meaning that they demand for power in order to be able to retain programmed states. Secondary storage, instead, consists of drives of non-volatile memory such as hard disk drives (HDDs) and solid state drives (SSDs). The latter are preferred as they present impressive read speeds (which outperform those of HDDs), however, they lag behind primary storage units in terms of write speed, and could thus be improved upon. Furthermore, the technology on which SSDs are mostly based on, the NAND-flash, relies on the possibility to scale transistors for new device generations to be realised, and so do SRAM and DRAM. A recent realisation has thus been that these technologies will eventually face high bit error rates due to leakage that inevitably arises from too aggressive transistor scaling [5].

All in all, demonstrating a non-volatile memory technology capable of outperforming the NAND-flash in terms of speed and scalability would result in an impressive step forward for secondary storage. Moreover, if such novel technology were to present similar programming speeds to SRAM and DRAM, it could also take over the primary storage scene, and result in a revolution in computer memory as we know it - all computations and storage necessities would be taken care of by a single non-volatile technology.

1.3 Benefits (and Drawbacks) of RRAM

It is certainly interesting to highlight that the problems that have been introduced for the NAND-flash (being comparatively slow at writing and hard to scale) and for SRAM and DRAM (being volatile and also difficult to scale) do not necessarily affect RRAM. In fact, the latter consists of MIM stacks that, thanks to their simplicity in design and fabrication as compared to transistor-based technology, may be cost-efficiently scaled without too many difficulties (assuming that no series selector is strictly required for each RRAM unit in a memory array). Furthermore, RRAM is a non-volatile technology, meaning that unlike SRAM and DRAM, it does not lose its programmed state after a power cut. Ought to the nature of the ionic-drift mechanisms that make resistive switching possible (more on this in chapter 2), programming of RRAM cells is extremely fast (in the order of a nanoseconds [6]), matching that of SRAM and DRAM and thus making RRAM a competitive technology. This same nature also allows for cell operation at low voltages given appropriate material selection. Together with low current levels this opens the doors

for low power consumption in future memory applications.

Amongst other benefits, one has the possibility to achieve multi-level programming (MLP) - given a large enough resistance window of operation, ideally, more than two resistance states may be attained. It is also worth mentioning that materials that are typically investigated for the purpose of RRAM realisation, as those in this thesis, are perfectly compatible with the complementary metal-oxide-semiconductor (CMOS) technology, which is nothing less than the fundamental building block of modern circuitry. Finally, thanks to low power consumption, high scalability and non-volatility of RRAM cells, these seem like the perfect basis for machines running alternative computational methods such as hyperdimensional (HD) computing [7] and neuromorphic computing [8]. These are fundamentally different from traditional computing, and rely on the formation of extensive vectors to represent data (HD computing) or on making multiple simultaneous manipulations on a given input, imitating neuro-biological structures (neuromorphic computing). Such computing styles would benefit from the realisation of high performance RRAM, since they cannot be successfully realised with current non-volatile technology (given the slow writing speed of the NAND-flash and its sub-optimal scalability).

At this point, one may ask why RRAM is not yet being implemented in everyday devices, given the ample amount of benefits listed so far. The reason for this often boils down to sub-optimal reproducibility. Therefore, for the sake of demonstrating that RRAM is indeed a valid technology, and to ensure a positive future for computer memory, research on RRAM should focus on studying promising material stacks and their inherent switching behaviour for the realisation of high performance memory units. This is precisely why this thesis is dedicated to the study of HfO_2 and ITO RRAM structures, as they inherit all the major RRAM attributes described above, low voltage operation in particular [9], and may improve upon the current state of result-reproducibility (both within a device and between devices) thanks to a self-compliance phenomenon [10]. Moreover, regardless of how promising measurements may actually be, the availability of a physical based model of RRAM in this thesis allows for it to be directly considered and implemented in further projects. A study that leads to the formulation of a model is of great help to the RRAM community, as much is yet to be discovered about the leading mechanisms in this technology.

Chapter 2

Scientific Background

The principle on which RRAM is based is resistive switching, which is achieved by programming memory units with different electrical biases so that different resistance levels may be attained at separate moments in time. In order to describe RRAM, a reasonable approach would then be of discussing the basics of operation of resistive switching first, so to give the reader an idea of what procedures need to be performed, and then move on to how programming is possible thanks to some insight into the underlying physics of resistive switching. As a disclaimer, one should mention that many different, and sometimes clashing, interpretations have been proposed for the phenomena that contribute towards resistive switching in RRAM. Multiple factors, such as material choice, device engineering and processing in the laboratory, contribute towards overall device behaviour. Therefore, different research groups may very well come to different conclusions, which then apply to their structures specifically and possibly not to others. Keeping this in mind, an overview of relevant physical aspects will be discussed in the following, with particular emphasis put on descriptions that seemingly apply well to HfO_2 and ITO RRAM, and on the methods that need to be developed in order to deepen one's understanding of RRAM in general.

2.1 Programming Operations

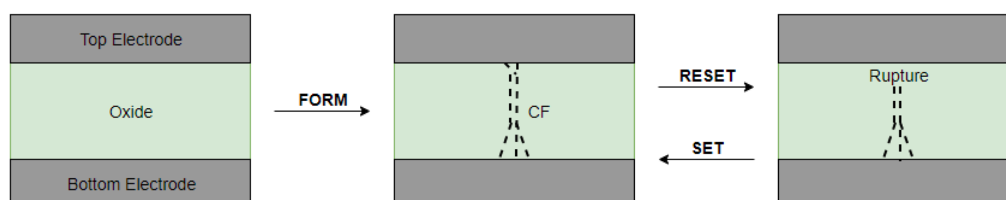


Figure 2.1: A representation of the different programming operations in RRAM: FORM, RESET and SET.

A MIM structure, with the insulator in its pristine state (as seen in the left-hand side of figure 2.1), exhibits a high resistance to current conduction from one metal layer (electrode) to the other - this is simply due to the nature of the insulator (often an oxide) separating them. If a dielectric breakdown were to happen in the insulator, however, resistance to current conduction would surely diminish. Supposing that such breakdown is reversible by means of appropriate device biasing, it becomes clear how one could arbitrarily change the resistance state in a MIM-like structure. Starting with a pristine

structure, one could define "electroforming" (often simply referred to as FORM) as the programming operation that results in a soft dielectric breakdown in the insulator due to the formation of a conductive filament (CF) through it - the FORM operation thus initiates the programmability of an RRAM unit by changing its pristine high resistance state (HRS) to a low resistance state (LRS). On top of this, two additional operations may be defined in order to interchangeably switch between LRS and HRS. These are the RESET, which causes a rupture in the CF and thus resistance to become high, and the SET, which instead repairs the filament and once again lowers the cell's resistance - a visualisation of this is available in figure 2.1. High and low resistance are the fundamental units (the "1's" and "0's") of RRAM programming, and the possibility to switch between these efficiently through SET and REST operations is what makes RRAM so interesting in the eyes of researchers and the industry.

2.2 Classification

Now that the mechanisms which allow for RRAM operation have been introduced, one can proceed to consider the physical properties that make these possible, and thereafter arrive to classifications of cells according to different inherent behaviour. Firstly, it is important to consider what effect an electrical bias has on a given RRAM cell, specifically, if bias polarity matters towards cell operation. From this, what usually arises is a distinction between two types of RRAM: non-polar and bipolar.

2.2.1 Non-polar RRAM

The former type is characterised by irrelevance of voltage polarity when it comes to cell programming - this signifies that both SET and RESET are achieved with the same polarity. The principal contributor towards resistive switching in non-polar RRAM is believed to be thermal dissolution (Joule heating) [11]. In this picture, once a high enough voltage drop occurs over an RRAM cell, a CF is formed, and the biasing can be stopped. Once biasing is initiated again, however, current starts flowing through the CF and resistive heating inevitably impacts the structure. The point of maximum heat is the centre of the filament - it being the farthest from the electrodes (which input and output charge efficiently, given their conductive nature and greater size as compared to the CF itself). Therefore, excessive heat build-up eventually results in a RESET operation, dictated by thermal dissolution of the CF (near its centre) in the insulating layer of the stack. Then again, the CF may be repaired with yet another biasing operation; given that current flow is hindered by the filament being incomplete, resistive heating does not have a detrimental impact on such SET operation.

It may be worth mentioning that non-polar RRAM has a subclass referred to as unipolar RRAM, which also presents SET and thermally driven RESET under the same bias polarity, yet only operates in one polarity (either positive or negative).

2.2.2 Bipolar RRAM

The other type of RRAM, based on bias polarity observations, differs from the above in that polarity is crucial in determining what operation is being performed on a given RRAM cell. In fact, bipolar RRAM requires opposite polarities for SET and RESET. The leading mechanism in this type of RRAM is the migration of ions inside the structure,

as their charge makes them manipulable by different biases and bias polarities [12]. A filament may thus be formed by directing ion redistribution in one direction, rectified by switching voltage polarity, and repaired by switching once again. This thesis concerns itself with bipolar RRAM, as this is easier control than cells that rely on the somewhat unorganised dissolution of CFs by heat.

Other than classification according to bias response, RRAM may also be categorised according to the active components which contribute towards resistive switching by means of formation, rupture and reparation of a conductive path. This leads to three new categories: anion-type, cation-type and carbon-based RRAM [6].

2.2.3 Anion-type RRAM

When the active components in an RRAM cell boil down to oxygen ions and (charged) oxygen vacancies, the cell is said to be of anion-type. It is thus common to also call this Ox-RRAM. Thanks to the application of a given bias, oxygen ions in the oxide layer of a cell drift towards the electrode that is linked to the biasing terminal of opposite sign to the charge of the moving ions themselves (simply due to electrostatics). Oxygen ions that have been relocated populate vacancies in the approached electrode, or simply come to form an oxygen rich layer in the adjacent region. A direct effect of this is that, similarly to the creation of an electron-hole pair in semiconductor theory, the removal of oxygen ions from their original position in the oxide leaves oxygen vacancies behind, which act as if charged in order to screen the missing oxygen ions. Such vacancies then also relocate thanks to the driving force imposed by the applied bias, and may construct a conductive path in the oxide. This is the equivalent of a FORM operation in Ox-RRAM, and can easily be visualised thanks to figure 2.2. Subsequently, by reversing the bias polarity, oxygen ions and vacancies are driven back towards their place of origin, and recombination of ions and vacancies occurs. This results in the rupture of the vacancy CF, and thus equals a RESET operation. Finally, the bias polarity may be switched again to repair the filament, and thus SET the device (this dependence on polarity translates to Ox-RRAM being predominantly bipolar) [12]. It is to be noted that the point of rupture is a discussed topic amongst research groups, although the most common interpretation is that the CF brakes at the interface with the top electrode (i.e. the electrode onto which the resetting bias is applied, the bottom electrode being grounded) [13, 14].

The above may lead to the conclusion of there being a predefined directionality to follow for correct operation of anion-type RRAM devices. However, unless restricted by the capabilities of its electrodes (of which materials may specifically be chosen to be inert and thus not easily contribute towards overall resistive switching properties of a device), it may not matter whether a device is biased in one direction or the opposite to yield CF formation - only after the FORM operation may the "correct" biasing direction arise, following how oxygen ions and vacancies were relocated during such process. Furthermore, one could come to think that anion-type RRAM is mostly of bipolar type. Although ion migration is indeed dictated by the applied bias and its polarity, heating may have an impact on the dissolution of CFs and may very well be the leading force in determining switching in a given cell, making it non-polar. Therefore, one can conclude that the nature and optimal switching conditions of an RRAM device are indeed difficult to interpret, and that such properties result from the whole stack rather than the insulating layer on its own. These remarks have now been made for anion-type RRAM, but should be kept in mind for any further category.

2.2.4 Cation-type RRAM

Similar to the anion-type RRAM is the cation-type, which differs from the former simply in that rather than there being oxygen ions relocating, and oxygen vacancies forming a CF, a "conductive bridge" is now formed by means of redistribution of metallic ions from an active electrode [15]. It should not come as a surprise that this type of RRAM is most often referred to as conductive-bridge random-access memory (CBRAM).

2.2.5 Carbon-based RRAM

Finally, one also has RRAM which owes its resistive switching property to carbon-based CFs. In such RRAM, appropriate biasing may lead to the alignment of conductive sp^2 hybrids into a conductive path, hence yielding a LRS in a given device. A HRS is instead achieved when biasing results in the formation of an insulating sp^3 carbon layer [16].

Other than the above, another carbon based switching mechanism exists, and this relies on graphene oxidation. Specifically, carbon rings in a graphene-doped switching layer may incorporate oxygen, which results in stretched carbon-carbon bonds and increased resistance due to increased oxygen content and hopping distance for charge carriers [6].

2.2.6 RRAM Comparison

As a transition from the above to what is specifically considered in this thesis, it is relevant to compare the mentioned RRAM types and conclude which one may be most interesting to look at. More specifically, one can outline why CBRAM and carbon-based RRAM will not be discussed further, and then highlight what aspects of Ox-RRAM make it more worth one's while.

Starting with CBRAM [17], this has been shown to generally operate at low voltages and have large windows of operation, supposedly allowing for MLP. However, devices often suffer from unreliability in the reset, as the spread in HRS can become large over many cycles of resistive switching [18]. Moreover, CBRAM devices often rely on metals which are incompatible with the silicon-dominated industry, such as copper or silver [19]. Moving on to carbon-based RRAM, the aforementioned technologies promise great functionality in the long run, given the simplicity of the chemical composition of active species (and thus good understanding of the underlying physics), low cost, high stability and low operation voltage [6]. Specifically concerning graphene oxide RRAM, great result reproducibility is predicted, together with excellent scaling capabilities (due to graphene being two-dimensional) and the possibility for MLP [20]. The issue with carbon-based technology in general (not only concerning RRAM) lies in the novelty of the field of research, and recurring difficulties in finding procedures that allow for high yield, good control and high quality of production throughput [21].

When it comes to Ox-RRAM, ultra-low operating voltages are achievable (lower than for the other RRAM types), with good reproducibility and processing that is compatible with industry standards (making this a comparatively more interesting technology than CBRAM) [6]. Furthermore, the practical realisation of Ox-RRAM is not hindered as much by the novelty of the field as for carbon-based RRAM. Perhaps the only question mark in state-of-the-art Ox-RRAM lies in the reliability of mass production - a concern that applies to all RRAM and has halted any realisation of the technology from permanently reaching the market. What needs to be demonstrated is a device that displays the great properties achievable by Ox-RRAM, maintains the same switching behaviour throughout

its lifespan, and behaves equally to devices that have been processed in parallel. This thesis makes an attempt at this by studying the promising qualities of HfO_2 and ITO structures, and most importantly, it tries to explain what is happening in the LRS and HRS, such that further interpretation may lead researchers to understand what needs to be done to improve upon RRAM. Furthermore, the realisation of a physical based model offers the possibility to consider studied devices in computer simulations where one may want to reproduce that which was observed in the laboratory, and see if it scales well when coupled to a realistic scenario [22].

2.3 HfO_2 and ITO RRAM

As demonstrated by groups who have been working on HfO_2 and ITO RRAM [9, 10], this material choice potentially allows for the realisation of high quality anion-type bipolar RRAM. This is thanks to extremely low operation voltages, and the possibility to have self-limitation during switching, which allows for better control on CF growth and thereafter good reproducibility. These outstanding results are not accidental, and may be explained through some underlying physics.

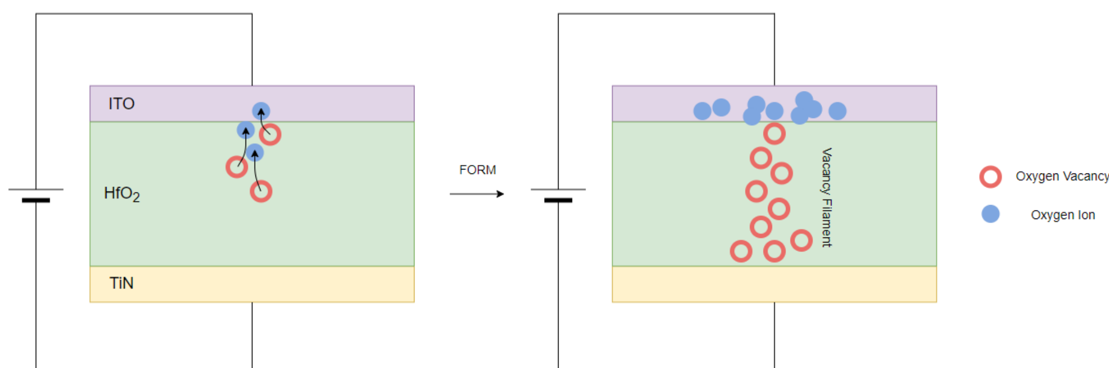


Figure 2.2: A simple representation of how a CF is formed in an anion-type RRAM, such as the structure considered in this thesis.

As previously introduced, Ox-RRAM is dominated by the formation of a CF of oxygen vacancies (which is aided by local electric field enhancement at the tip of a growing filament [23]), as depicted in figure 2.2. Earlier it was simply claimed that recombination of vacancies with oxygen ions causes a rupture in the CF during RESET, yet more details should be included in a satisfying description of Ox-RRAM. In fact, it is not obvious that oxygen ions should drift to the active electrode in a given device, and that charged vacancies arise from this. A careful consideration of the energy of formation of defects, in fact, reveals that in an oxygen-rich scenario (e.g. the case for an isolated oxide slab) the most likely defect resulting from the application of an external bias is the oxygen interstitial. Such a defect is detrimental to an RRAM structure as it would be an unwanted source of oxygen for recombination with vacancies, resulting in unreliable control over switching properties. Fortunately, by adding a "scavenging layer" on top of the oxide, an oxygen-poor state is attained, which increases considerably the energy of formation of interstitials. Oxygen vacancies, instead, become energetically inexpensive (i.e. achieved with low operating voltages), and may contribute towards device operation [23]. In the case of this thesis, the scavenging layer is the ITO, which caps the HfO_2 insulating layer

and acts as the active RRAM electrode. The choice of ITO is a result of it being an excellent oxygen ion absorbent [6, 24], and it being highly compatible with HfO_2, as explained below.

Another concerning aspect is the possibility to form vacancies that are hard to control. Taking into account the band structure of HfO_2 [23], it follows that a number of different regions exist, corresponding to vacancies of different charge. The choice of scavenging layer has an impact on what type of vacancy results upon biasing, following the energy region that is reached by the scavenger's own work function. That of ITO (4.2-5 eV [25]) perfectly aligns with the energy range in which twice positively charged oxygen vacancies are formed in HfO_2. This is a great advantage when it comes to device operation since it translates to vacancies that can be controlled by driving electric fields in RRAM switching.

All in all, HfO_2 and ITO are chosen together to yield high performance devices because of the possibility to create oxygen vacancies at ultra-low voltages, which can be controlled thanks to their charged state. Furthermore, this material choice is well compatible with industry fabrication standards, making it interesting for the simple reason that it could directly be implemented in parallel with existing technology. Common RRAM nomenclature is top electrode (TE) for the active electrode in the MIM-like structure, and bottom electrode (BE) for the opposite electrode. The TE in this thesis is thus made out of ITO, whilst the BE is of titanium nitride (TiN). Note that the latter's purpose is of merely conducting electrical current once a CF is formed, hence the choice of an inert material.

It should be noted that filament rupture and growth are processes of stochastic nature, ought to the fact that they rely on the occurrence of events at the nanoscale and beyond. This translates to potentially different CF manipulation conditions throughout the lifespan of a device, and device engineering is crucial to set up a structure that minimises the impact of CF variability on reproducibility. In this thesis, the choice of ITO as the TE aims at doing this also. Other than the mentioned compatibility benefits, there is the possibility to achieve self-limitation of CF growth [10], which is surely an advantage when it comes to have good control over resistive switching. This is an effect believed to be linked to the formation of a Schottky barrier corresponding to an oxygen rich ITO region at the interface between HfO_2 and ITO, resulting from the absorption of oxygen ions by the latter during FORM and SET operations. This separates the CF and the TE and hinders excessive filament growth.

2.4 Unoptimised Operation and Detrimental Effects

Once a satisfying structure has been realised in the laboratory, what remains to be done is to characterise it, highlight its positive aspects and draw conclusions that are relevant to the development of the field. What must then be avoided is the misinterpretation of the capabilities of a given device as a result of inappropriate data collection. In fact, inappropriate operation may lead to premature device failure, linked to either the HRS or the LRS not being attained any longer.

The major contributors towards premature failure are unbalanced SET and RESET [26], and current overshoot [27], and the reason why is straight forward. In the former case, one can imagine how applying a much higher voltage in one of the two fundamental programming operations (SET or RESET), may have a detrimental effect on the other operation. This is because a too large SET may result in a CF that cannot be easily ruptured, especially if the RESET bias is low. Conversely, if the RESET is exaggerated, the CF of a given structure may be rectified so much that a subsequent SET may

not be able to repair it, especially in the case of low biases. In order to avoid unbalanced RRAM operation, it is important to make sure that the chosen SET and RESET biasing conditions do not require to be adjusted after an arbitrarily small amount of cycles of programming, and can instead be sustained for a cell's whole lifespan.

When it comes to current overshoot, a given device may momentarily attain a current level much higher than that of the desired operation range due to too slow response of the system used to retrieve data - for instance, a parameter analyzer. As already predicted by Ochua [2], resistance in RRAM is a direct consequence of how much charge is allowed to flow through it. In the case of current overshoot, this translates to a too large filament being instantiated, which may not be easily ruptured by further operations. To minimise this detrimental effect one can increase the rise and fall times of voltage pulses applied to a device - given that the parasitic capacitance from which current overshoot originates couples current to the derivative of voltage in time, slow voltage ramping results in small current overshoot [28]. Alternatively, and perhaps more convincingly, a selector may be utilised in series with a studied RRAM unit, allowing current to flow through the RRAM and the selector itself only up to a very specific compliance limit. The most sophisticated option here is a transistor, given the advanced state in research of such technology. A one transistor one RRAM (1T1R) is thus set up, with compliance decided by means of arbitrary gate biasing of the transistor. It should be mentioned that having a transistor in series with an RRAM cell makes the overall unit scalability potential much lower. However, the scalability of such a system is still greater than that of the NAND-flash since the RRAM component can be scaled much more aggressively than the capacitor which attains a given charged state in NAND-flash (this would leak and result in unacceptable error rates if scaled as much).

2.5 Charge Transport Mechanisms

As a last section in this chapter, some explanation as to why current conduction may occur in a dielectric is provided. This will allow for the interpretation of RRAM current-voltage characteristics, an essential step in the formulation of a compact model of RRAM. By compact it is meant that thanks to some valid simplifications and straight forward current equations, an accurate reproduction of experimental data (reality) is possible. It is thanks to review articles such as that of F.-C. Chiu [29] that it was possible to collect all the mechanisms that are presented below.

2.5.1 Electrode-Limited Mechanisms

When current conduction is modulated by phenomena that are related to the contact between an electrode and the dielectric, it is said that conduction is electrode-limited.

Schottky Emission

As studied by Sir O. W. Richardson [30], heating has an impact on charge transport through a material, as electrons may be supplied with sufficient energy to overcome existing extraction energy-costs and thus contribute towards current conduction. The fulcrum of Richardson's work lies in the formulation of the dependence of current on temperature and work-function W, as written in equation 2.1.

$$J = A\,T^2\,\exp\left(\frac{W}{k_BT}\right) \tag{2.1}$$

In the above, A is a pre-exponential factor given by the Richardson constant A_0 times a material specific factor λ_r (between 0 and 1, typically 0.5) [30], T is temperature and k_B is Boltzmann's constant. Now, if an external electric field F were to be included in the picture, equation 2.1 would be changed to equation 2.2 [29], where $\phi_B = W/q$ (q being the fundamental charge) is the field-free barrier height, and the square root term in the exponent takes into account the modulation to the barrier by the applied field.

$$J = A\,T^2\,\exp\left(\frac{-q(\phi_B - \sqrt{qF/4\pi\epsilon_r\epsilon_0})}{k_BT}\right) \tag{2.2}$$

Here, ϵ_r and ϵ_0 are the relative permittivity and permittivity of vacuum, respectively. What is described in equation 2.2 is the principle of Schottky emission. This is expected to have an impact on charge transport in RRAM when a CF has not yet been fully formed, hence a portion of oxide hinders transport to and from a partial CF and the adjacent electrode. Furthermore, it might come into play in the case of formation of a Schottky barrier between the TE and a fully formed CF.

Fowler-Nordheim Tunnelling

Also referred to as (cold) field emission, Fowler-Nordheim tunnelling consists of conduction through a triangular potential barrier by means of tunnelling. Specifically, the barrier, which may be identified with a dielectric separating two conductors, might experience a reduction in width due to the application of an external field. Tunnelling of charge carriers is thus facilitated and can be shown to follow equation 2.3 [29] (assuming that the carriers tunnel without help of any thermal excitation).

$$J = \frac{q^2F^2}{8\pi h\phi_B}\exp\left(\frac{-8\pi\sqrt{2qm^*}}{3hF}\phi_B^{3/2}\right) \tag{2.3}$$

Here h is Planck's constant, and m^* is the effective mass of the tunnelling charge carrier.

Thermionic Field Emission

In between Schottky emission and field emission, one has the phenomenon of thermionic field emission, which consists of nothing more than Fowler-Nordheim tunnelling aided by a thermal energy contribution. Current arising from this follows equation 2.4 [29], where $\hbar = \frac{h}{2\pi}$ is the reduced Planck constant.

$$J = \frac{q^2F\sqrt{m^*k_BT}}{8\pi^{5/2}\hbar^2}\exp\left(-\frac{q\phi_B}{k_BT}\right)\exp\left(\frac{\hbar^2q^2F^2}{24m^*k_B^3T^3}\right) \tag{2.4}$$

A schematic representation of the above three mechanisms is proposed in figure 2.3.

2.5.2 Bulk-Limited Mechanisms

If current mostly depends on the electric properties, as well as the processing quality and resulting energy domains, of the dielectric slab in an analysed structure, one refers to conduction as bulk-limited.

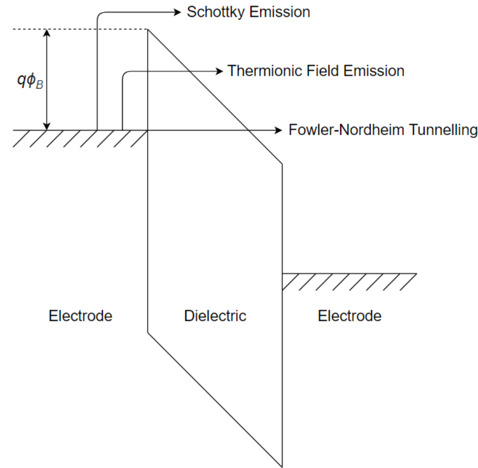


Figure 2.3: Band diagram that qualitatively shows the difference between the three discussed electrode-limited conduction mechanisms.

Pool-Frenkel Effect

When a bias is applied over a dielectric, electrons that are "trapped" in localised states in the material may gain sufficient energy to excite to the conduction band of the material, and contribute towards current conduction. If one defines ϕ_T as the trap energy level, the contribution to current that is had because of excitation of charges from these traps according to the driving force of an externally applied field F is given by equation 2.5 [31, 29].

$$J = q\mu N_C F \, \exp\left(\frac{-q(\phi_T - \sqrt{qF/\pi\epsilon_r\epsilon_0})}{k_B T} \right) \tag{2.5}$$

In the above, μ is charge carrier mobility, and N_C is the effective density of states in the conduction band of the dielectric.

Hopping

Similar to the Pool-Frenkel effect is the phenomenon of hopping. This consists of current arising due to trapped electrons gaining enough energy to leave their trap states (again, ϕ_T) and "hop" to subsequent traps. If one takes into account the concentration n of electrons in the conduction band of the dielectric through which conduction is had, the frequency f of thermal vibrations at trap sites (which has an impact on electrons leaving trap states), and assumes that traps are separated by a distance a, hopping may be quantified with equation 2.6 [29, 32, 33].

$$J = qanf \exp\left(\frac{qaF}{k_B T} - \frac{q\phi_T}{k_B T} \right) \tag{2.6}$$

Ohmic Conduction

Perhaps the most straightforward mechanism to take into account is Ohm's law, which simply states that current I in an electrical conductor is linear with applied voltage V, with proportionality being described by conductance G, or alternatively resistance R (inverse of conductance), as written in equation 2.7 [34].

$$I = GV = V/R \tag{2.7}$$

Given that what allows for conduction in RRAM in its LRS is a conductive path, it is believed that ohmic conduction, and hence equation 2.7, should be a relevant entity in the description of charge transport once a CF has been formed.

Thinking of the CF in an RRAM structure as a metallic rod allows for the estimation of the latter's cross section, an entity that gives an idea of the scale at which phenomena occur in RRAM, as well as being essential to the extrapolation of further physical parameters. Specifically, from classical electrodynamics, one has that the resistance R to current flow in a metallic object is proportional to the length t, and inversely related to the cross section S of the object itself, as given in equation 2.8.

$$R = \rho \frac{t}{S} \tag{2.8}$$

Here, ρ is resistivity, which has been extensively studied for many metals. In the case of CFs in the RRAM structures in this thesis, one expects the filaments to consist of Hf, given that they correspond to strips of HfO_2 depleted of oxygen. By using an extrapolated LRS resistance R_{LRS} as R, the thickness of the HfO_2 layer as t, and ρ_{Hf} as resistivity, an estimate for S may be obtained. It is to be noted that ρ_{Hf} is affected by the scale of the study [35], thus care must be taken in choosing the correct value of resistivity in computations.

Space Charge Limited Current

When charge is conducted in a non-ideal electrical conductor, such as the oxide layer in RRAM, excessive charge accumulation in a confined region in space may take place and have an impact on the relation between I and V. Accumulation may result in charge being more appropriately described as a continuum, rather than as an ensemble of singularities (c.f. the concept of energy band in solids vs. energy levels in atoms). It is then said that space charge is attained, and current is space charge limited (SCLC). The first description of this effect was outlined by C. D. Child [36], who focused on current conduction through a plane parallel vacuum diode, arriving to the conclusion that SCLC results in $I \propto V^{3/2}$. This result is interesting yet not directly applicable to the case of RRAM since conduction here does not occur in vacuum. As proposed by N. F. Mott and R. W. Gurney, in semiconductors and insulators, the relation rather follows the square of V, as written in equation 2.9 (the Mott-Gurney law) [37].

$$J = \frac{9}{8} \epsilon_r \epsilon_0 \mu \frac{V^2}{t^3} \tag{2.9}$$

Here, t is the thickness of the conducting piece of dielectric. The problem with equation 2.9, when it comes to interpreting conduction in RRAM, is that it makes some perhaps unreasonable assumptions, amongst which one has that current shall not be affected by traps and energetic disorder. This is not likely (yet not impossibly) the case in a resistive switching device, since the core functionality of the latter lies in the rearrangement of ions, inevitably followed by the formation of defects in the insulating layer.

Frenkel Enhanced SCLC

As discussed by Murgatroyd in [38], equation 2.9 may be extended so to account for the contribution by charges that are localised at trap states in the studied dielectric. The contribution arises from the fact that trapped charges find themselves in potential wells in the energetic landscape of the dielectric. When an external field is applied, this energy profile is tilted such that the effective well height is reduced - some carriers may escape their trap states, thus resulting in more free carriers and an increase in current throughout the dielectric. This is known as Frenkel enhancement. In the case of room temperature HfO_2, it can be calculated that such a contribution only really matters when a field in the order of magnitude of MV/m is applied. When it comes to RRAM, biases of up to a few volts are applied over structures with insulating layers just a couple of nanometres thick. This corresponds to fields well above the mentioned requirement for Frenkel effect to yield a contribution, and one should expect this to come into play when characterising RRAM.

The modified expression for SCLC, enhanced by Frenkel effect, can be approximately found to be as in equation 2.10 [38], where θ_0 is the ratio of filled traps to all available trap states.

$$J = \frac{9}{8}\mu\epsilon\frac{V^2}{t^3}\theta_0 \exp\left(\frac{0.891}{k_B T}\sqrt{\frac{q^3 V}{\pi\epsilon t}}\right) \tag{2.10}$$

Chapter 3

Method

When it comes to practice, it is important to present results that may directly be compared with the work of other research groups such that a given RRAM structure may be put into perspective, and conclusions may be drawn regarding its impact towards technological development. The beginning of this chapter is thus dedicated to the introduction of typical RRAM figures of merit, which are implemented in this thesis for the characterisation of HfO_2 and ITO RRAM structures. Particular attention is dedicated to those quantities that are essential to the physical interpretation of RRAM, and thus play a role in the creation of a physical based model. Other figures are merely introduced for the sake of completeness. Following this, a description of the processing steps taken in the laboratory to realise the studied devices is laid out. Finally, the self-developed computational tools used to analyse gathered data are presented, together with a description of the carried out computer modelling. The latter are the key to a better physical understanding of RRAM, and thus help to highlight how to handle this technology for further development.

3.1 Figures of Merit

Thanks to publications such as that of Mario Lanza's group [39] and Philip Wong's group [40], a somewhat standardised procedure for RRAM characterisation has been formulated for the community to use as a point of reference. This involves the consideration of figures that are essential in highlighting the outstanding, or insufficient, properties of studied structures, as discussed below.

3.1.1 Basic Quantities and How to Obtain Them

Direct current (DC) $I-V$ characteristics of RRAM are perhaps the type of measurement which helps the most towards the interpretation of conduction through a given structure, hence it is a good habit to include an arbitrarily well behaved curve in the report of an RRAM study. It is best not to present an averaged curve, due to fluctuations that are typical for contemporary RRAM devices, hence the possibility to create artefacts or lose valuable information in the averaging process. Instead, the curve should be one that captures overall device behaviour, and thus displays the most similarities with other curves. Since this is a subjective choice, it is good practice to present multiple curves (corresponding to different cycles of operation, or different devices - depending on the situation and point to be made) and highlight one of them.

Figure 3.1 below gives the reader an idea of how a straightforward RRAM DC measurement looks like. Such a measurement arises once a cell has undergone its FORM process, in which a CF is created for the first time through the application of a large electrical bias. $I - V$ characteristics thus lead to two key resistive switching properties of RRAM, namely the SET and RESET voltages, V_{SET} and V_{RESET}, respectively. These correspond to the voltages at which the transitions from HRS to LRS, and vice versa, occur, and give a clear idea about how to optimally bias an RRAM cell to achieve resistive switching.

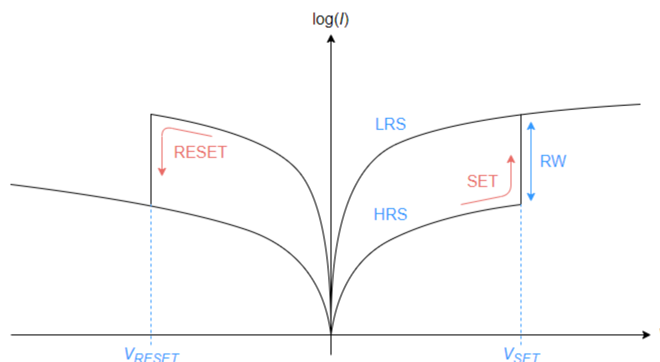


Figure 3.1: A representation of typical DC characteristics of RRAM.

Experimental data may not be as smooth as in figure 3.1. Specifically, the SET and RESET transitions may not be abrupt, and instead occur over some voltage window. One then needs to define a general way to extract V_{SET} and V_{RESET} - this can be done by considering the voltage which corresponds to a noticeable change in current for V_{SET}, and by noting the voltage that corresponds to maximum current during a negative voltage sweep for V_{RESET} [41].

Other than V_{SET} and V_{RESET}, figure 3.1 displays a clear resistance window (RW) between the HRS and the LRS. This can be quantified by looking at the so called ON/OFF ratio of the device in question. As suggested by the name of the quantity, this corresponds to looking into the current levels in a device in its ON (low resistance) state versus its OFF (high resistance) state, and can be extrapolated by taking the ratio of LRS current to HRS current. Alternatively, one can correlate current and voltage as in equation 2.7, extrapolate resistance values for the HRS (R_{HRS}) and LRS (R_{LRS}), and take the ratio of R_{HRS} to R_{LRS}, given the inverse relationship between current and resistance.

3.1.2 Benchmarking RRAM

Endurance and Retention

Endurance is the term used to indicate the amount of programming that a cell can sustain before failing. If one defines an RRAM programming cycle as a SET operation followed by a RESET, endurance would correspond to the number of cycles that may be successfully performed before the operated RRAM cell stops switching. This figure is quantified by comparing consecutive R_{HRS} and R_{LRS} over the span of the studied cycles. Note that a minimal RW needs to be defined to identify the point at which a device is considered to have failed.

The property that makes RRAM a non-volatile technology is its capability to supposedly retain a programmed state for an indefinite amount of time. Therefore, for devices

that are under development, it is essential to verify that programmed states are maintained. Such a retention study is performed on the LRS of cells, given that the HRS is closer to the pristine state of the cells themselves, and is therefore naturally maintained for longer; the used metric is LRS current versus time. Since it would be cumbersome to test devices in the timescales that are required for practical applications (in the order of years!), a common practice is to take measurements at elevated temperatures and extrapolate the behaviour at room temperature, or at expected on-chip temperature (ca. 85 °C) [42].

The above quantities are most definitely related to the underlying physics of resistive switching - e.g. should a poor filament be formed in a device, both endurance and retention would suffer from it. Interested phenomena are, however, not as easily deduced from endurance and retention tests than from simple DC characterisation. Because of this, the reader is invited to consult publications on these figures of merit on HfO_2 and ITO RRAM, should they desire to learn more on the topic [43]. This paper will instead proceed with only considering what is directly necessary for the physical interpretation of RRAM.

Variability

An obvious prerequisite for RRAM to be considered by the industry is the possibility to mass-produce devices that may consistently be programmed with the same conditions and will reliably produce the same results. A study is thus necessary to ascertain that investigated structures are suitable - both under the aspect of cycle-to-cycle variability, and cell-to-cell variability. In both cases, the to-be-considered quantities are those described in section 3.1.1. Specifically, for cycle-to-cycle variability and device-to-device variability alike, one can recur to showing a superposition of $I-V$ curves, as well as histograms or cumulative distribution functions of V_{SET} and V_{RESET}, and/or R_{LRS} and R_{HRS}. In this thesis, variability within a device is discussed thanks to subsequent $I-V$ curves plotted together, and that between devices is highlighted thanks to differences in $I-V$ curve features, and empirical cumulative distribution functions of SET and RESET voltages.

Scalability

One final aspect to consider is the change in behaviour of structures when the only varied parameter is cell size. It has been previously shown that scaling is not only a general requirement for nanotechnology, but is actually beneficial to RRAM as it helps to increase the RW of scaled devices thanks to reduced leakage. This in fact lowers current in the HRS, whilst current in the LRS is close to unchanged due to CF size not necessarily depending on device area [40, 44, 45, 46, 47, 48, 49, 50]. Due to unavailability and/or unreliability of working devices of different sizes to those that are considered in this work, scalability will not be further discussed.

Verification of HfO_2 and ITO Core Characteristics

As previously mentioned, it has been demonstrated that HfO_2 and ITO RRAM devices are in principle capable of operating at ultra-low voltages, and have an inherent quality of self-limitation that should contribute to great device control. The former of these attributes is easily verified by direct inspection of the $I-V$ characteristics of the studied RRAM devices; given that a satisfactory amount of switching cycles present SET and

RESET voltages of a fraction of a volt, ultra-low voltage operation is confirmed. The self-limitation characteristic, instead, needs to be considered more closely.

In a straight forward $I - V$ measurement, performed with a parameter analyser with in-built current compliance, self-limitation should be highlighted by current modulation in the LRS of the device. Specifically, it should be possible to confirm that a barrier is formed between CF and TE, effectively stopping CF growth and having an impact on current conduction. If this is not the case, it could be that the compliance level set by the measuring apparatus is so low that the self-limitation regime is not reached, and CF growth is halted by the imposed compliance rather than by the formation of an oxygen barrier in the structure.

In this project, the above is investigated by means of a LRS vs. parameter analyser compliance study, and the impacts of self-compliance and artificial compliance are ascertained. Specifically, the LRS resistance R_{LRS}, linked to filament cross-section S, is considered as a function of compliance, and trends in the following curve are discussed, taking into account direct observations from the $I - V$ characteristics (e.g. if it is evident that there is no self-compliance).

3.2 Device Processing

The structures that are considered in this thesis were realised according to a procedure developed by the Nano Electronics group in Lund [51]. A final device consists of a TiN-HfO_2-ITO stack on a silicon substrate. It is to be noted that the focus of this thesis does not lie in the processing of devices, but rather their physical characterisation. Therefore, the implemented fabrication techniques are mentioned albeit not discussed, and the reader is invited to consult with the provided references for further information.

The fabrication process begins with the selection of an adequate substrate, namely a silicon (Si) slab. This provides a well known, robust, and inexpensive base for processing, with improved surface roughness thanks to a layer of thermally grown native SiO_2 of the thickness of 200 nm. On top of this, the BE is defined following the steps illustrated in figure 3.2. A 60 nm layer of TiN is deposited on the substrate by means of atomic layer deposition (ALD) [52]. The precursors are TDMA-Ti and a 300 W N_2 plasma, supplied at a rate of 300 s per cycle. Subsequently, spin coating [53] is used to cover the TiN with photoresist (PR) for subsequent ultraviolet (UV) lithography [54]. The PR is a photo-sensitive material, and via exposure to UV light through a mask (which imposes a desired BE pattern), it is activated such that, thanks to a development step (with TMAH), it may be removed wherever not necessary. Once the PR has been patterned, a dry etch [53] (based on SF_3) is performed to impose the mask pattern to the underlying TiN. Finally, the BE material is freed from the PR coating thanks to the removal of the latter with acetone, and any remaining trace or unintentional compounds that might have been created due to PR and etchant coming into contact (e.g. teflon) are surely removed by oxygen ashing [55].

Following the above, a 3 nm HfO_2 layer is deposited by 300 K ALD using TEMA-Hf and a 300 W O_2 plasma, at 80 s per cycle and with 30 sccm oxygen flow. This is shown in figure 3.3, which then proceeds to highlight how the processing is completed. A spacer layer is deposited and patterned by means of UV lithography, analogously to the above description, and the top electrode material is deposited with sputtering [56], resulting in a 20 nm ITO layer. The latter is conveniently coated in gold for better contacting when probing for device characterisation. All in all, a cylindrical TiN-HfO_2-ITO structure is

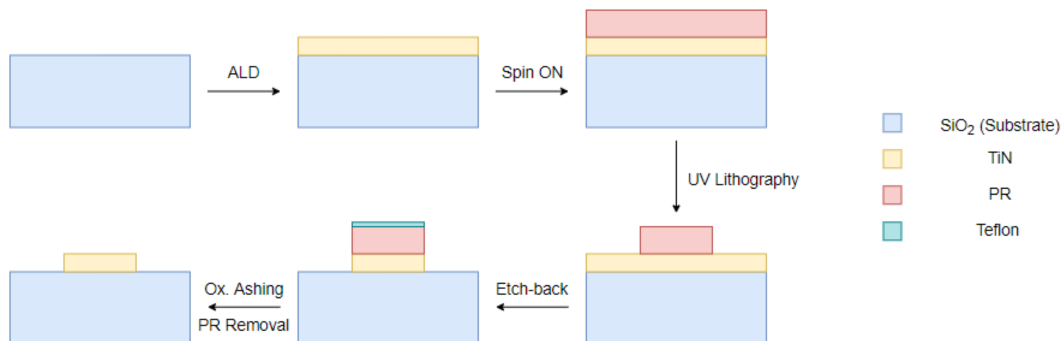


Figure 3.2: Definition of the TiN BE by means of ALD, UV lithography and etching.

realised, with a diameter of 10 µm and thus a footprint of almost 80 µm^2.

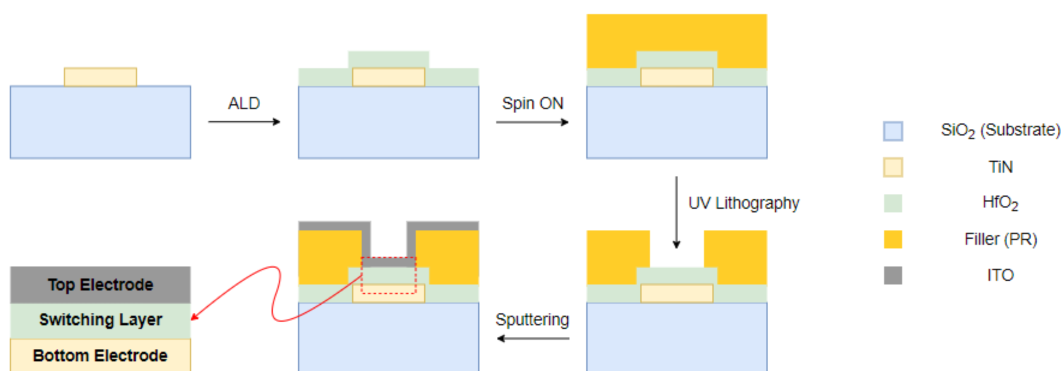


Figure 3.3: ALD of the HfO_2 switching layer, followed by TE definition with UV lithography and ITO sputtering.

3.3 Measurement Setup and Procedure

In order to electrically characterise RRAM structures in this project, a cascade probe station [57] was used. Specifically, two probes were implemented to make contact with the electrodes of a device, and DC sweeps were performed to measure the $I - V$ curves corresponding to its resistive switching behaviour.

A Keithley parameter analyser [58] was used for biasing via the station's probes, and for the collection of resulting currents. Before performing definitive measurements, the used probes were tested either by shorting them through contact with a copper strip, or by performing measurements over a reference resistor. In both cases, one needed to verify that resistances measured according to voltage-current linearity would not differ from the expected values (0 Ω for the short, 50 Ω for the used resistor) by more than a few Ω. In case of erroneous results, different probes were used until the test resistances could be verified.

3.4 Data Analysis

Results obtained from the probing of RRAM structures consisted of tables (xls) of voltage values with corresponding current levels. These could then be implemented in some data

analysis (performed with Python) to extract fundamental characteristics and figures of merit of RRAM, and to gain a deeper understanding of the underlying physics of resistive switching. A characterisation study as such paves the way for the realisation of a computer model which is capable of reproducing realistic RRAM behaviour under electrical bias, as discussed in section 3.5 below, and later in section 4.2.

The first step taken in data analysis in this project was that of writing an algorithm capable of understanding the features of an RRAM $I-V$ curve. This was done by setting up a program to identify the SET and RESET points in a sweep, and then divide the latter into four main sub-regions of interest: HRS and LRS in the SET sweep (before and after the SET event, respectively), and LRS and HRS in the RESET sweep (before and after the RESET event, respectively).

3.4.1 Extrapolating Figures of Merit

Taking inspiration from the explanation in section 3.1.1, it was possible to automate the localisation of SET and RESET events. SET points were identified thanks to an updating variable corresponding to the change in current with every voltage step, normalised by the smaller current of the two in this difference. Specifically, V_{SET} was set to the largest instance of this updated variable throughout a given SET sweep - this corresponds to a large change in resistance, and the normalisation makes sure that the switch is not confused with a step taken late in the sweep, where current values are large due to large bias. RESET points, instead, were identified as voltages corresponding to current maxima in RESET sweeps.

SET and RESET points corresponding to subsequent cycles within a device were gathered, and averages were computed. Further, given that an adequate amount of devices was characterised, enough data was recorded to carry out a meaningful device-to-device variability evaluation. Specifically, empirical cumulative distribution functions were generated, using V_{SET} and V_{RESET} averages from each device.

3.4.2 Identifying Characteristic Modes of Conduction

Thanks to the identification of SET and RESET points, the $I-V$ curves at hand were split into LRS and HRS segments for each programming operation. Thereafter, each segment could be analysed in more detail to understand what type of conduction has a contribution in the different resistance states of RRAM. Specifically, each segment was plotted in different scales - for example, following the description in section 2.5, it was chosen to plot $\ln I$ vs. $\sqrt{V}$ to potentially highlight a contribution due to Schottky emission. Any clear linear relations in such plots indicated that the corresponding mechanisms of conduction potentially provided an adequate description of conduction through the structure that was being characterised.

Following the above qualitative study, quantification was performed by means of linear regression fitting and extrapolation of physical parameters that are meaningful towards specific interpretations of conduction. Fits were applied in those regions where linearity was evident to the naked eye, and different behaviours in different biasing regions were identified.

3.4.3 DC Characterisation - Coding in Python

The remainder of this section is dedicated to showcasing how the developed python code (see appendix A) may be implemented for the characterisation of RRAM devices. Specifically, an artificial dataset is generated and put to the test, such that code features may be discussed without yet exposing the results of this project.

The hereby implemented dataset consists of a simulated voltage sweep that results in resistive switching between two resistance states of 1 kΩ and 100 kΩ, respectively. Current values are generated as if linearly dependent with voltage, and inversely related to resistance (i.e. Ohm's law applies). Furthermore, to add a touch of credibility, some random normally distributed noise is added. The generated data is saved in an xls spreadsheet, with format matching that of data collected in the laboratory.

Having established the to-be-analysed file, what remains to be done is simply to feed this into the developed python algorithms to first gain a qualitative understanding of the studied device, and then extrapolate physical parameters through fitting. Two types of function are thus necessary, one for qualitative and the other for quantitative analysis (see A.4). These are run separately for each region of interest in a given $I-V$ curve - the former type is run first, and based on conclusions that can be drawn from plotting results, one or more functions are handpicked from the latter type for pertinent data extrapolation.

Before being able to use the above functions, it is obvious that the aforementioned sub-regions of interest must be isolated. Therefore, the first generated output is a plot of the input curve, in which SET and RESET points have been identified and used to split the curve itself into subsections, as seen in figure 3.4.

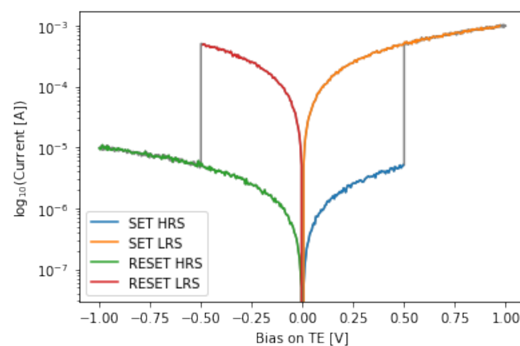
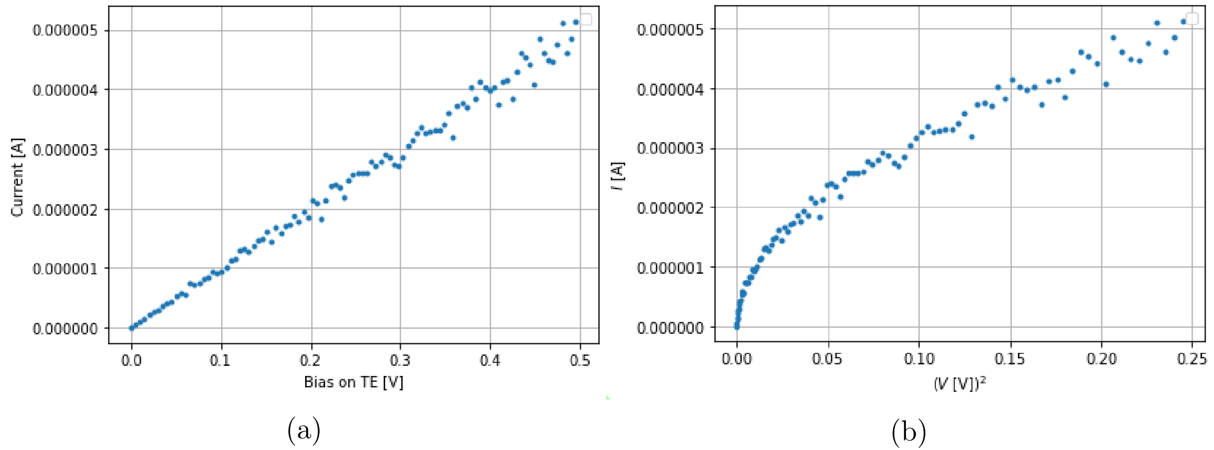


Figure 3.4: $I-V$ characteristics of a simulated RRAM device.

Following is the plot of each subsection in different scales - this should facilitate the user's task of identifying whether there is linearity in any given region. Figure 3.5 displays this kind of output for the subsection of HRS in the SET sweep (marked in blue in figure 3.4). Specifically, an example each for interesting (3.5a) and an uninteresting (3.5b) plot is given.

Given the results obtained from mere plotting, one may decide to personalise the adopted fitting program, such that only fits that are pertinent to the analysed RRAM are performed. For instance, in the case of the hereby discussed artificial data, it is obvious that there is linearity in current and voltage (figure 3.5a), hence the relevant fit is as displayed in figure 3.6. The secondary output of the fitting algorithm is a collection of fitting parameters - more specifically, the slope and the intercept corresponding to the performed fit. In the case of Ohmic conduction, the slope of I vs. V corresponds



(a) (b)

Figure 3.5: $I - V$ characteristics of a simulated RRAM device, represented in different scales: linear (Ohmic) scale (a); Mott-Gurney law (b).

to the inverse of resistance. The computer-extracted slope of the fit in figure 3.6 is of $1.01 \cdot 10^{-5}$A/V, thus the estimate of $R_{HRS} = 99\,\mathrm{k}\Omega$. Taking into account that there was added noise, this is a good interpretation of the actual input for artificial data generation ($R_{HRS} = 100\,\mathrm{k}\Omega$).

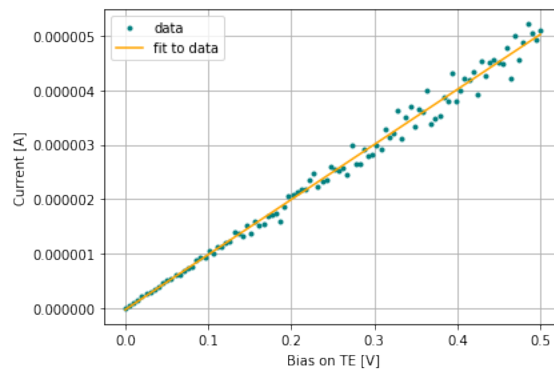


Figure 3.6: Linear fit to the HRS in the SET sweep of a simulated device.

In reality, relations that are more complicated than Ohm's law have to be considered when characterising an RRAM device. For instance, through the slope of an I vs. V^2 linear relation, one may deduce the mobility of charge carriers in the switching layer of an RRAM device, as follows directly from equation 2.9. When it comes to linearity in $\ln I$ vs. $\sqrt{V}$, instead, (local) temperature may be computed from the intercept of a linear regression, and effective emission barrier height may be inferred from its slope, as explained by equation 2.2.

3.5 RRAM Modelling

The last piece to the puzzle is to use gained physical understanding to create a tool that corroborates that which is concluded from data analysis, and is of help to further RRAM studies, e.g. the implementation of this technology in arrays or other circuits. In fact, if a computer model is set up following identified charge transport mechanisms, and satisfying reproducibility is obtained with model results as compared to real life measurements, it

is likely that a reasonable interpretation of RRAM has been met. Moreover, the model itself becomes useful to members of the research group in Lund for computer simulations where their own devices need to be represented.

Now that the experimental and computational methods of this thesis have been presented, the text will move on to the discussion of obtained results. Chapter 4 starts off with showcasing core $I-V$ characteristics of a number of devices from the sample that was studied in this project, immediately giving the reader the possibility to appreciate the variability of the devices themselves. Further, the results of qualitative analysis of representative $I-V$ curves are presented, followed by fitting with the current equations that are introduced in section 2.5. Physical parameter extraction is implemented to ascertain goodness of fit and determine the best possible interpretation for current conduction in RRAM. The phenomenon of self-compliance in HfO_2 and ITO RRAM is considered in a separate study, and it is evaluated whether it is noticeable in the measured devices. Finally, the model that was developed based on the above is discussed, and its universality is put to the test by comparison with data from a different sample.

Chapter 4

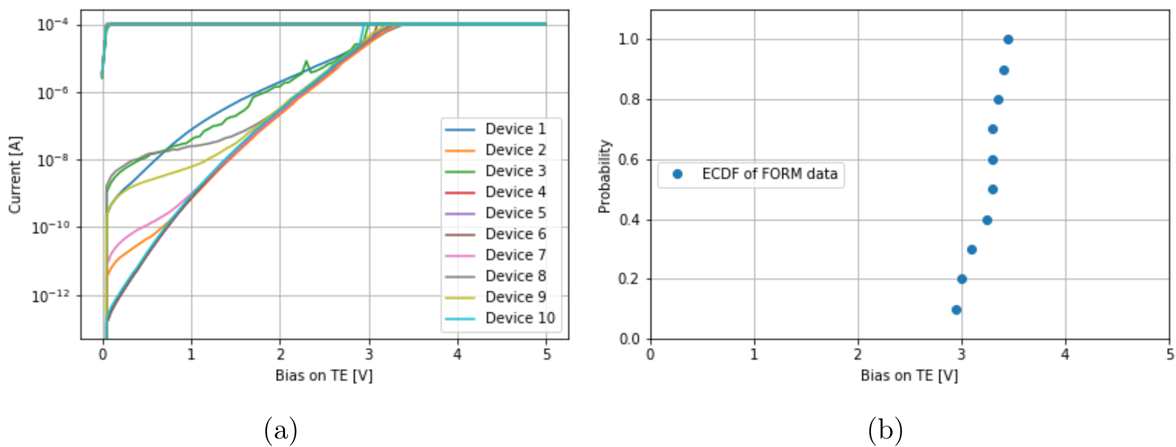
Results and Discussion

From the sample that was selected for measurement taking in this project, ten devices were fully characterised and are the subjects of the below discussions.

4.1 DC Measurements

4.1.1 FORM Operation

Had it not been for an initial electroforming step, none of the considered devices would have shown resistive switching properties. It is thus interesting to look at how the different devices performed during the FORM operation, specifically, to understand whether similar conditions can be expected from different devices. In fact, the ability to initialise a device consistently at known conditions is likely of crucial importance to any further application of RRAM. Figure 4.1 is thus proposed, showing how forming occurred in the characterised devices.



(a) (b)

Figure 4.1: FORM DC measurements for the considered devices (a), alongside the empirical cumulative distribution function for the form events (b).

As it can be seen, compliance (set by the parameter analyser) is reached approximately in the same point in voltage in all devices. This is taken to be the FORM point, as it can be seen that in the back-sweeps, current remains in compliance until very low voltages, which indicates that a filament has indeed been formed. Variability is satisfying, as pointed out

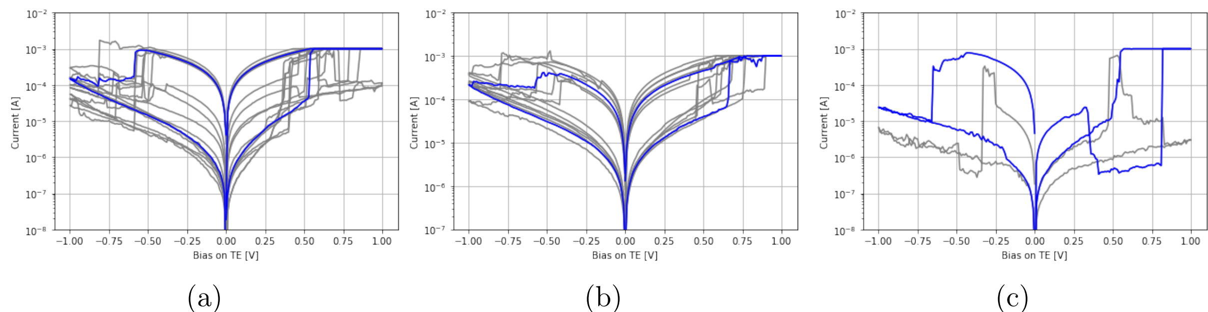
by the empirical cumulative distribution function in figure 4.1b - the FORM operation can be consistently achieved below 3.5 V, in a narrow (ca. 0.5 V) window.

A few remarks should be made about the curves in figure 4.1a. It can be noted how the beginning of the different sweeps does not follow the exact same trajectory, i.e. there is a discrepancy (of up to ca. four orders of magnitude between certain curves) in current versus voltage that is linked to some pristine devices already conducting much better than others. However, it is undeniable that a FORM occurs in all devices regardless of how conductive they are before the soft dielectric breakdown. Furthermore, it should be noticed how current increase, prior to FORM, is quite drastic. It is often the case that current remains very low prior to FORM, and a noticeable "jump" (of a few orders of magnitude) suddenly occurs once the FORM point has been reached. What drives the devices in this thesis to behave as they do is likely an effect of gradual CF formation (which ensures a steady increase in current, and the fact that once compliance is reached, a filament has indeed been achieved) and high leakage linked to the footprint of the devices themselves. Should the switching layer cross section (here in the order of tens of μm^2) be reduced, current before FORM would inevitably be reduced, without compromising current levels whilst a CF is sustained. This is because filament size is in principle independent of device size, but is rather determined by biasing conditions.

Satisfying FORM reproducibility hints at the devices being suitable for the implementation in applications which demand simultaneous initialisation of cells in an array. If this were not the case, devices would have to undergo separate forming processes prior to their implementation, which is cumbersome to say the least. A last important remark to make is that, as it can be seen in figure 4.1, voltages of over 3 V need to be sustained for a FORM operation to go through. This must be kept in mind when implementing pristine RRAM units in parallel to other circuit components (e.g. fragile transistors), as the initialisation of the former could compromise the latter.

4.1.2 Switching Behaviour

Moving on to how devices performed after their respective FORM operation, the $I - V$ characteristics of a few subsequent programming cycles, for three devices in particular, is proposed as a first result - further discussion stems from the analysis of such curves. The reason for showcasing three measurements in particular is to distinguish between general device behaviour, reproductive behaviour and undesirable behaviour.



(a) (b) (c)

Figure 4.2: $I - V$ characteristics of a normal working device (a), a very reproductive device (b) and a non-working device (c). In blue are curves that are thought to be quite centred, thus representing their respective sample well.

What is seen in figure 4.2a is the case for eight devices out of ten (the figure corre-

sponds to the measurement of a single device; however, its behaviour is similar to that of others and is thus used as representative), whilst figures 4.2b and 4.2c correspond to a particularly well behaved device, and one that does not switch successfully, respectively. That nine out of ten characterised devices are working is per se a quite satisfying result. It would be inconceivable to implement RRAM in circuitry without high yields, and speaking of yield, it is not only important to make sure that devices are switching - one should consider reproducibility of measurements, both within a device (cycle-to-cycle) and between different devices. Variability may be considered in terms of the basic RRAM properties that are introduced in subsection 3.1.1, as well as in terms of the conduction mechanisms that take place in each device. The former is discussed in the remainder of this subsection, whilst the latter is the topic of subsections 4.1.3 and 4.1.4, where qualitative and quantitative analysis of meaningful curves (one for each device, selected just as for the blue curves in figure 4.2) is performed.

A clear difference between figure 4.2a and 4.2b lies in the magnitude of current throughout different sweeps within a device. Specifically, the device in 4.2b maintains the same current level throughout subsequent programming operations, both in low and in high resistance state, whilst that in 4.2a attains different levels at different stages of the cycling. This is most surely related to variations in the CF between cycles, and results in the RW to vary as well - not only in size, but in location. In fact, if measured at 0.1 V, the device in figure 4.2b consistently displays an RW of over an order of magnitude, stretching from well below 10 µA (HRS) to almost 100 µA (LRS), whilst the device in figure 4.2a sometimes attains a larger RW, yet its location is somewhat undefined between different cycles. This is an undesirable property, as too large of an uncertainty in RW may lead to the inability to correctly determine the state attained by a given RRAM unit. Imagine setting up a very simple algorithm for state determination, involving current level measurement at a certain voltage (small enough in magnitude to not cause a change of state - say 0.1 V) and sorting states according to whether they fall short or exceed a certain threshold. The interpretation offered by such an algorithm may then be erroneous as the set threshold may work in some cases, and fail in others where the RW has changed a lot from its expected value. It goes without saying that this is problematic towards RRAM implementation, as state misinterpretation inevitably leads to computational errors.

Generally speaking, RRAM devices tend to perform better (in terms of reproducibility of $I-V$ characteristics) the more cycling they undergo (provided that a current compliance is imposed, regulating filament formation - whether this is naturally occurring in HfO_2-ITO structures, or if it is provided with a selector device or a parameter analyser). This reproducibility is inevitably related to device life-expectancy, and is what has allowed for endurance measurements of over a million cycles here in Lund [22]. Therefore, the situation is likely not as severe as portrayed in the above, even more so when another aspect of variability is considered: that in switching points. In fact, the spread in V_{SET} and V_{RESET} between devices and cycles is quite reduced, i.e. resistive switching is seen to consistently occur in well established regions in voltage. This argument is supported by figure 4.3, which shows the regions of voltage in which resistive switching is expected to occur, together with corresponding switching likelihood.

A direct observation inspired by figure 4.3, and perhaps foretold by figure 4.2, is that the spread in the SET point is greater than that in the RESET point. This in turn translates to filament regrowth being more unpredictable than rupture. The explanation as to why this is happening likely boils down to the intimate relation between SET and RESET themselves. Specifically, a SET operation heavily depends on the RESET

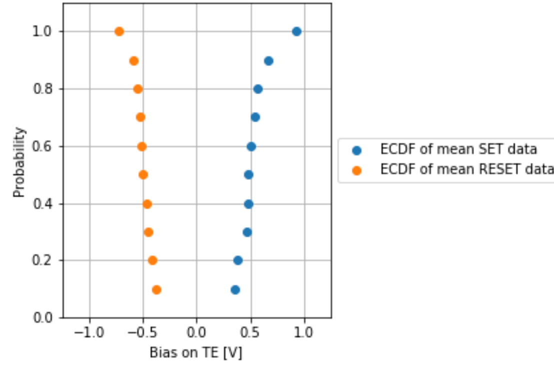


Figure 4.3: Empirical cumulative distribution functions as calculated by considering the average set and reset behaviour of the devices that were measured in this thesis.

operation that preceded it, given that the state in which an RRAM cell is left after rupture requires suitable biasing for regrowth. The starting point for a RESET operation is a well defined filament, hence it is understandable that the spread in V_{RESET} is small. On the contrary, the starting point for a SET operation is a partially dissolved filament, a state that may vary from cycle to cycle depending on the continued voltage drive after passing the reset point simply due to the nature of the species (ions and vacancies) that make resistive switching possible.

Regardless of the above, the discrepancy in V_{SET} and V_{RESET} width is quite small. Furthermore, one should appreciate the mean of the latter values. In fact, this is of around half a volt in magnitude for SET and RESET alike, hinting at the applicability of the low voltage arguments in favour of research on HfO_2 and ITO RRAM raised in the previous chapters.

4.1.3 Current Conduction - Qualitative Analysis

Plots of current and voltage for the blue curve in figure 4.2a are hereby presented in different scales, leading towards a discussion on what fits are more relevant for the devices in this thesis. The plots are for before (figure 4.4) and after (figure 4.5) the SET operation (positive voltage in figure 4.2a); SET and RESET properties were observed to be quite symmetric and it is deemed redundant to include both. Likewise, plots are not included for other devices; all measured devices displayed qualitatively similar behaviour, and showing this for one of them is deemed sufficient for the sole sake of supporting the choice of what fits to make. Some extra words are spent, where necessary, to point out features displayed by plots of other devices as compared to those that are hereby presented.

As it can be seen, a number of different scales are included for the HRS (pre-SET, figure 4.4), whilst only one plot is displayed for the LRS (post-SET, figure 4.5). This is because no single mode could safely be singled out as the most likely case for conduction in the HRS, due to apparent linearity in numerous regions in different scales. On the contrary, in the LRS, linearity of current in voltage was the only real contender. This is per se quite reassuring, given that ohmic-type conduction is to be expected from a metallic filament once this is formed in a device. However, this also translates to the foreseen self-compliance effect in HfO_2-ITO devices to not be noticeable. Self-compliance is thus further investigated, and discussed in subsection 4.1.5. All in all, RRAM LRS simply demands for linear regressions, whilst interpretation of the HRS is more complicated and

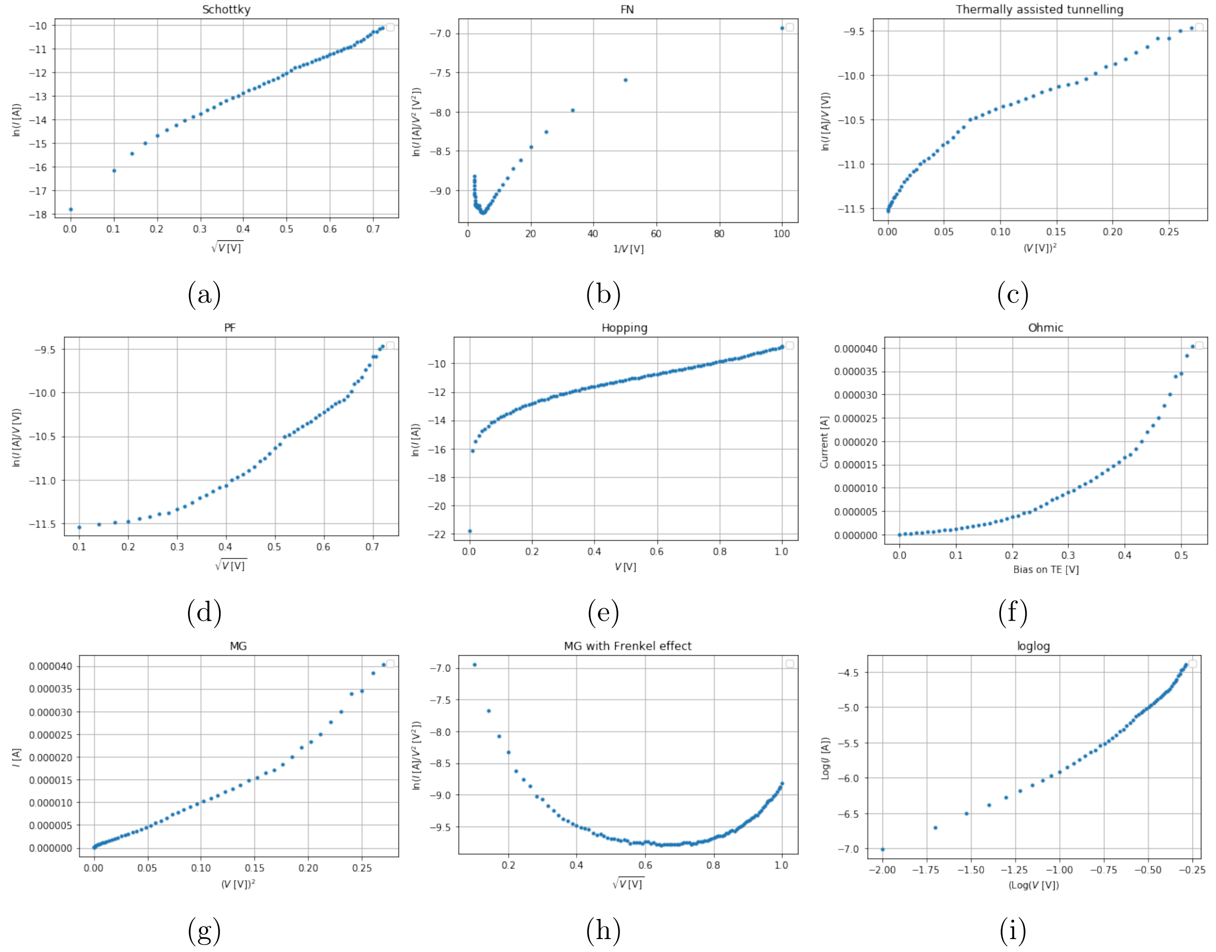


Figure 4.4: Plots of the pre-SET $I-V$ characteristics of the device in figure 4.2a in the scales relevant to: Schottky emission (a), Fowler-Nordheim tunnelling (b), thermionic field emission (c), Pool-Frenkel effect (d), hopping (e), ohmic conduction (f), Mott-Gurney law (g), Frenkel enhanced SCLC (h). Finally, a plot in log-log scale is included (i) to facilitate the visualisation of power laws.

needs further discussion.

As seen in figure 4.4h, saying that Frenkel enhanced SCLC contributes to current conduction is a stretch. Therefore, this conduction type is disregarded in fitting. Likewise, it is decided to not consider Fowler-Nordheim tunnelling further. Whilst it could be argued that that linearity is the case for $\ln(I/V^2)$ vs. $1/V$ in a somewhat reduced region in voltage, this is even less so the case in other devices, making its consideration somewhat irrational. When it comes to the other two electrode-limited mechanisms, i.e. Schottky and thermionic field emission (figures 4.4a and 4.4c, respectively), linearity is quite noticeable with the increase of voltage. Similarly, it can be argued that the Pool-Frenkel effect (figure 4.4d) has an impact for voltages above 0.25 V, or perhaps hopping (figure 4.4e) for voltages exceeding a certain threshold (0.3 V). For the remaining bulk-limited mechanisms, it would seem that the Mott-Gurney law (figure 4.4g) follows linearity in its respective scale for quite a broad voltage spectrum, whilst ohmic type conduction (figure 4.4f) has no good linear fit, if not for very low voltages. Thanks to figure 4.4i, further insight can be gained: up to a tenth of a volt, one has linearity with a small slope, and after this, the slope rises. It can be extrapolated that in the range

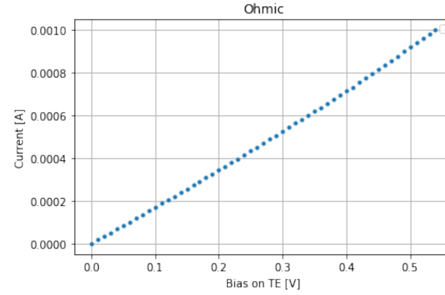


Figure 4.5: Post-SET $I-V$ characteristics of the device in figure 4.2a.

0-0.1 V the slope is almost exactly one, which seems to hint at ohmic type conduction being a good description for these low voltages. This is the case for all devices, just as the Mott-Gurney law applying well. However, for the latter, it is often the case that a good linear fit in the respective scale arises only starting from from ca. 0.1 V or more (at lower voltages, the plot follows a square root, again indicating that ohmic conduction is a good interpretation).

In conclusion, it is reassuring that at least one mode has been found to represent devices well when starting to apply a bias to the HRS (i.e. ohmic type conduction), whilst for increasing voltages, fitting is required. Again, note that fitting is only performed where linearity in a given scale is met, which according to experience gained from a thorough consideration of all samples boils down to the following approximate intervals: above 0.1 V for Schottky emission; above 0.2 V for thermionic emission; between 0.2 and 0.6 V for the Pool-Frenkel effect; above 0.3 V for hopping; between 0.1 and 0.6 V for the Mott-Gurney law (in pre-SET fitting, fits are carried out only up to the SET-point). Small tweaks are made to the intervals where necessary, to always ensure good linearity. Based on the credibility of extracted parameters, it can be concluded whether one mode of conduction is more likely to describe RRAM in this thesis than another.

4.1.4 Current Conduction - HRS Quantitative Analysis

To determine whether a given mechanism describes conduction in RRAM HRS well, linear fits are performed for all the modes that seem to display linearity in their respective scales. Thanks to the formulas and the Python code listed in Appendix B (which follow directly from the theory in section 2.5), different physical parameters are extracted, and their credibility is hereby evaluated.

Given that the considered devices are quite large (footprint-wise), it is deemed important to extrapolate parameters as if there were two possibilities for conduction. The first being from the tip of a ruptured filament to the TE; the second directly from BE to TE. The reason for this is that the role of the ruptured CF in the HRS is somewhat unknown - it is reasonable to think that its presence alone allows for higher current conduction than in a pristine HfO_2 state, hence motivating the study of conduction from CF to TE. However, it could also be that, given the width of the RRAM cells, conduction throughout the whole (FORMed) HfO_2 layer is dominant over contribution from a single (ruptured) CF. Parameter extraction results assuming conduction from CF tip to TE are gathered in table 4.1, alongside results assuming BE-to-TE conduction. Distinction is made between device values for pre-SET and post-RESET states, as the difference in biasing direction might have an impact on some conduction mechanisms. Furthermore, for each entry in

the table, a range going from lowest to highest extrapolated value is reported to give the reader an idea of device-to-device variability.

Table 4.1: Physical parameter extraction results.

Mechanism	Parameter	SET avg	RESET avg	Device Range
CF tip-to-TE:				
Schottky emission	T [K]	402	504	[251,784]
	$q\phi_B$ [eV]	0.0864	0.105	[0.0223,0.242]
Thermionic emission	T [K]	2360	3050	[1590,3580]
	$q\phi_B$ [eV]	0.0168	0.0526	[-0.417,0.872]
Pool-Frenkel effect	T [K]	2200	2350	[970,4130]
	$q\phi_T$ [eV]	-0.564	-0.550	[-1.85,0.0557]
Hopping	T [K]	2090	3130	[978,4270]
	$q\phi_T$ [eV]	-4.01	-6.10	[-8.33,-1.68]
Ohmic conduction	R [kΩ]	244	107	[10.2,1440]
Mott-Gurney law	μ [m^{2}V^{-1}s^{-1}]	$2.46\cdot10^{-5}$	$1.52\cdot10^{-5}$	[$3.18\cdot10^{-7}$,$5.18\cdot10^{-5}$]
BE-to-TE:				
Schottky emission	T [K]	232	291	[145,453]
	$q\phi_B$ [eV]	0.331	0.414	[0.227,0.643]
Thermionic emission	T [K]	1140	1470	[765,1720]
	$q\phi_B$ [eV]	1.35	1.76	[0.963,2.14]
Pool-Frenkel effect	T [K]	1270	1350	[560,2386]
	$q\phi_T$ [eV]	1.21	1.32	[0.640,3.40]
Hopping	T [K]	698	1040	[326,1420]
	$q\phi_T$ [eV]	-0.423	-0.670	[-1.01,-0.137]
Ohmic conduction	R [kΩ]	244	107	[10.2,1440]
Mott-Gurney law	μ [m^{2}V^{-1}s^{-1}]	$1.56\cdot10^{-10}$	$9.50\cdot10^{-11}$	[$3.89\cdot10^{-12}$,$3.55\cdot10^{-10}$]

Starting with Schottky emission, it is reasonable to conclude that this mechanism is unlikely, yet not impossible, to yield conduction in the studied RRAM cells on its own. In the CF-to-TE consideration, it suggests that a barrier in the order of 0.1 eV separates CF and ITO, which seems like an underestimation. A simple consideration of work functions of Hf [59] and ITO [25], electron affinity [60] and band gap [23] of HfO_2 (the latter giving an idea of likely Fermi level location) yields that the barrier should rather be in the order of eV. This is an educated guess, albeit the real case band structure is somewhat unknown, and the extrapolated barrier height should be given the benefit of the doubt.

It can be seen that for some device (lower limit of the given range) a temperature below that of the room is expected. This is not physically sound, and is the same reason why Schottky emission is believed not to be a good description should BE-to-TE conduction occur instead. Compared to the temperatures that are extrapolated for other mechanisms (discussed below), however, the estimates which arise from Schottky emission fitting do not seem incredible after all - they are at least in the order of the room. It is thus difficult to draw a decisive conclusion about this mechanism.

Moving on to Thermionic emission, incredibly high temperatures and extremely low barrier heights are predicted in the case of CF-to-TE conduction, making this an improper description. In the bottom-to-top electrode case, instead, barrier height estimation is much more reasonable. However, related temperatures are again too high to be true.

Similarly, the Pool-Frenkel effect is just as incapable of describing the HRS of RRAM. In fact, assuming filament-to-ITO conduction only, negative trap "depths" would arise, which is yet another senseless parameter extraction. Assuming that BE-to-TE conduction is instead the case leads to trap sites of depth $q\phi_T$ in the same order of the expected barrier height $q\phi_B$, which is an overestimation, together with temperatures that are surely much higher than those attained in reality.

A remark to make about the Pool-Frenkel effect, and its respective parameters as described in table 4.1, is that two somewhat unknown parameters had to be assumed for computations, namely the effective density of states N_C in the conduction band of HfO_2 (which determines the strength of the effect as it has a direct impact on the possibility for electrons to be emitted from trap states) and electron mobility. For the former, it was assumed that parabolicity in the band structure near the conduction band minimum applied isotropically, such that a single value of effective mass could be used [60] and density of states estimated by assuming that the Fermi level is distant from the conduction band minimum, as done in course books such as [31] and mentioned in the footnotes of appendix B. For the mobility, it was assumed that anything up to 1 $cm^2V^{-1}s^{-1}$ would be within reason (this being the approximate electron mobility in amorphous silicon [61]). Given that it was observed that the lower the mobility input into the parameter extraction, the worse the results, the largest possible mobility was assumed for the sake of discussion.

Parameter extraction for hopping current also requires educated guessing of unknown quantities entering computations, amongst which the only real threat is electron density n in the conduction band. Just as for mobility in the case of Pool-Frenkel interpretation, it was observed how a low input n resulted in unreasonable parameter extraction, hence a somewhat large n was fixed in computations, as written in table B.1. Given that negative trap barriers arise for both CF-to-TE and BE-to-TE, it is unthinkable that hopping could successfully describe the HRS of RRAM on its own.

When it comes to Ohmic-type conduction, there is no difference in parameter extraction for CF-to-TE and BE-to-TE. This is because the extracted parameter is resistance, which follows directly from the slope of the $I-V$ characteristics, without the need of taking length and cross-section of the conducting region into account. Shape would instead come into play in the extrapolation of a more universal quantity, such as resistivity (or conductivity). The reported estimates are credible, as in 10 to 1000 kΩ is an acceptable range for current conduction in a poor conductor. The span of the range itself would lead one to think that there is some noticeable device-to-device variability. However, one should know that the upper limit of the range relates to a single device, observed to have a somewhat weak CF and thus less reliable switching when compared to other devices. Excluding this device from the picture reduces the variance, and leads one to the conclusion that R_{HRS} in the order of 100 kΩ is then actually a good description for most devices.

Finally, the applicability of the Mott-Gureny law in describing the HRS of RRAM in this thesis remains to be discussed. As it can be seen from table 4.1, the obtained estimates of mobility are in the order of 0.1 $cm^2V^{-1}s^{-1}$ assuming CF-to-TE conduction, whilst several orders of magnitude lower should BE-to-TE be the case. The former seems like a reasonable interpretation; the latter demands further discussion.

As proposed in literature [29, 31], a modification to equation 2.9 could come into play if trap states in the structure were to prevent all electrons to contribute to current flow during conduction. Specifically, one could take into account "missing charge" by multiplying by a term that is nothing more than the ratio of free charge carrier density

to total (free and trapped) carrier density. One could then see μ in the bottom half of table 4.1 as an effective mobility (the product of actual mobility and the above ratio), suddenly making the low estimate more credible.

Unfortunately, there is a flaw in the above argument, which links back to some of the theory in section 2.5, and explains why this modification to the Mott-Gurney law was not discussed earlier in the manuscript. In fact, should trap states play a role in determining current conduction through the HfO_2 layer, at the attained bias levels over the small (nm thick) characterised structures, it is expected that Frenkel-enhancement should eventually take place, and current be described by equation 2.10. Given what is observed in subsection 4.1.3, this is likely not the case, hence SCLC through the bulk of the HfO_2 layer of the structures in this thesis seems less and less likely to be an accurate description.

On the contrary, should HRS conduction take place mainly from CF tip to TE, the reduced size of the volume in which current flow is attained (ca. a cylinder of cross section equal to that of the CF, and height equal to the gap separating ruptured CF and TE) likely makes contribution by traps less important, thus the original expression for the Mott-Gurney law could apply, as given by equation 2.9. This further supports the credibility of the estimates of mobility in the upper half of table 4.1.

All in all, it would seem like, out of all the considered the possibilities, SCLC from ruptured CF to TE is the best achievable interpretation to current conduction in the HRS of the devices in this thesis, after an initial ohmic conduction region. However, one should not neglect the possibility for there being some flaw in the above arguments; mainly, it could be that multiple modes of conduction are contributing simultaneously, and the adopted approach would not be capable of singling these out. For instance, some yet to be understood background effect could exist, and subtracting it from raw data could lead to new fits with credible parameter extraction results.

Alternative Approach

The above discussion is based on the parameter extraction that arises from the assumption of some thickness for the region of conduction. For BE-to-TE considerations, this is merely the thickness of the HfO_2 layer, whilst in the study of CF tip-to-TE conduction a fraction of the latter is inferred. What follows for multiple conduction mechanisms is the extrapolation of temperature, which can then be interpreted as reasonable or not.

It stands to reason that this workflow could be reversed, with temperature used as an input for computations, followed by the evaluation of estimates for the thickness of the conduction region. Given the mediocre success of most modes in describing the HRS of RRAM, it is foreseen that the assumption of a reasonable temperature would result in the extrapolation of unreasonable CF rupture lengths and HfO_2 layer thicknesses. This is verified by carrying out the reversed parameter extraction on one of the available devices: assuming room temperature, one obtains estimates that exceed the known thickness of the HfO_2 layer, both in BE-to-TE and CF tip-to-TE computations. Specifically, Schottky emission predicts a conduction region thickness in the same order of the HfO_2 layer thickness, yet above this known value; thermionic field emission and hopping predict an order of magnitude more, and the Pool Frenkel effect overestimates by two orders of magnitude. As previously discussed, amongst these temperature dependent mechanisms, Schottky emission is perhaps the most credible, although it is difficult to conclude whether it does indeed provide a contribution.

4.1.5 Current Conduction - LRS Quantitative Analysis

As mentioned in subsection 4.1.3, the LRS of the devices in this thesis is accurately described by Ohm's law, given that current is linear in voltage. It therefore holds that it could not be verified whether a barrier had formed in the devices after CF creation, such that further formation would be halted, i.e. the phenomenon of self-compliance that has previously been shown to exist in HfO_2 and ITO structures could not be observed.

It was speculated that the above could have been the case due to the compliance that is set by the parameter analyser, which might determine CF dimensions before self-compliance could have any impact. Thereafter, a study of the LRS as a function of artificial compliance was conducted, of which results are displayed in figure 4.6.

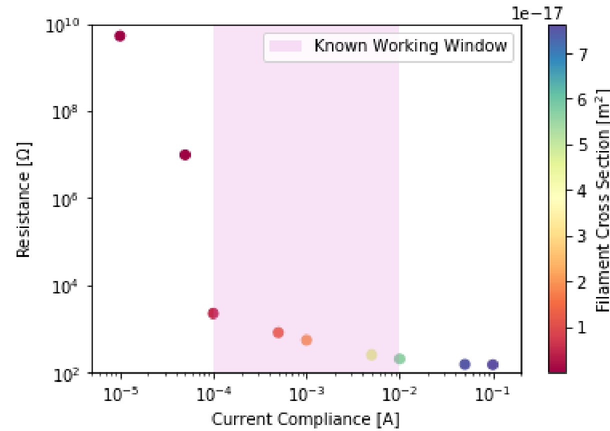


Figure 4.6: A visualisation of how LRS resistance R_{LRS} and filament cross section S behave as a function of parameter analyser current compliance.

LRS values outside the shaded area in figure 4.6 should be taken with a grain of salt. In fact, devices were observed to not switch, or behave poorly, outside of this compliance region: using a too low compliance results in a weak FORM, hence an unreliable CF from the start; a large compliance, instead, results in irreversible breakdown, i.e. a CF so strong that it cannot be reliably ruptured and regrown by subsequent programming operations. The data points within the shaded area are averages over subsequent cycles of RRAM operation, whilst those outside simply follow from the LRS attained directly after a FORM operation (unreliability that might have followed from unpredictable switching is thus disregarded).

It can be seen how a trend of decrease in resistance (and thus filament cross section increase) with current compliance is attained. This points at there being truth in the speculation that led to this study - if self-compliance was the case, then parameter analyser compliance would not make a difference. At compliance levels greater than 10^{-2} A, it can be seen how the LRS seems to stabilise, hinting at self-compliance potentially having an impact at these current levels. The fact that devices do not switch in this regime, however, indicates that self-compliance is somewhat irrelevant in this thesis - alone, it does not provide enough support to the filament growth process to limit CF size efficiently and thus yield the possibility for controlled switching.

The fact that RRAM programming relies on artificial compliance rather than self-compliance in the known working window is likely a direct effect of device footprint. If self-compliance were to have an impact, a significant barrier would need to exist between the CF and the TE in the LRS, linked to reallocation of oxygen ions that leave the HfO_2

layer, where the CF is thus formed. If the device in question is wide, oxygen might disperse more at the interface between HfO_2 and ITO, resulting in either a thinner or a more porous barrier (or both). It is thus foreseen that device scaling would likely have an impact on deciding the role of self-compliance in RRAM devices. Moreover, it may be argued that the ohmic-type conduction that is uniquely observed in this RRAM LRS study is a consequence of the somewhat high level of Sn doping in the ITO used in Lund. This is commercial grade, with 10% Sn (weight percent), as compared to doping below 5% used by other groups [10, 62].

4.2 RRAM Modelling

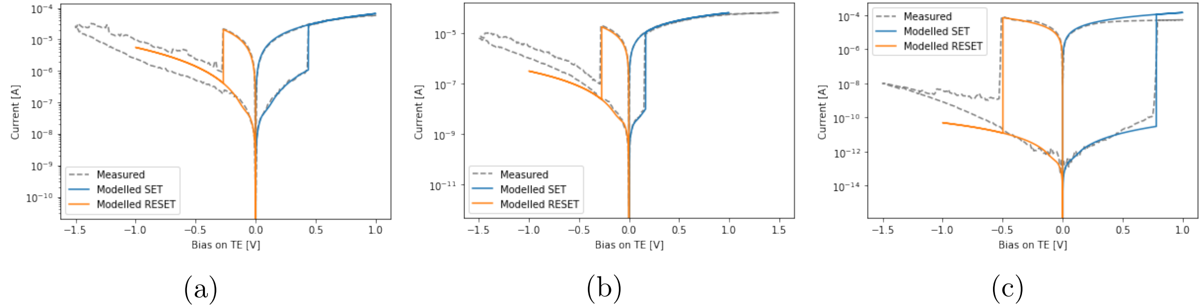
Following that which was concluded in the previous sections, a simple model is set up in Python with an Ohmic conduction-driven LRS, and Ohmic conduction, followed by the Mott-Gurney law in the HRS, with a smooth change from one conduction mechanism to the next occurring between 0.05 V and 0.15 V. The code for this is reported in appendix C. Note that this interpretation of conduction has been observed in previous studies [63], with the LRS explained by conduction through a metallic filament, and the HRS owing its behaviour to a change from conduction due to few thermally generated carriers in the HfO_2, to a space charge domain from plentiful injected carriers [29].

Before proceeding towards an evaluation of the model, two remarks are made. First and foremost, the model is in principle incapable of predicting RRAM behaviour at high voltages in the post-RESET HRS. This is because here, as hinted at in subsection 4.1.3, the Mott-Gurney law does not necessarily provide a good fit, and no other mechanism could be identified as relevant towards conduction. Secondly, the intercepts of Ohmic and SCLC fits have not yet been discussed. These can be seen as a (small) voltage independent background, which finds its origin in noise from the parameter analyser, as an artefact from data fitting, and/or other unknown processes. If necessary, this shift can be accounted for in the model by addition of a constant.

The model is thus put to the test by comparison with data from a new device (different sample). The difference between this and the previously studied ones is in the area of the device itself (ca. $7\,\mu m^2$, i.e. an order of magnitude smaller than the others). Further, the device chosen for corroboration shows the possibility for MLP, and it is interesting to see whether the compiled model can describe different resistance levels with the same interpretation for conduction. Some modelling results are presented in figure 4.7.

It is interesting to see that, assuming that the parameters that are mentioned in figure 4.7 are reasonable, the fidelity of the model is quite satisfactory, with the exception of high voltages in the post-RESET HRS, and the pre-SET HRS in figure 4.7c. The latter is probably for the same reason as for the former, namely the Mott-Gurney law not being a good description of current conduction at increasingly high voltages in the HRS of RRAM. Given that in the cases of figures 4.7a and 4.7b, a SET is achieved at a low voltages (below 0.5 V), one does not incur the problem of the Mott-Gurney law eventually not being a good fit anymore.

When it comes to interpreting the discrepancy between the model and the new data, one can follow a study analogous to that in section 4.1 to see whether a reasonable interpretation arises for large biases on a given HRS. Considerations made for figure 4.7a are somewhat inconclusive, albeit those on figures 4.7b and 4.7c seem to point at Frenkel enhanced SCLC to be the case. This is because local temperatures that are not incredibly hot are extrapolated (600-700 K), and $\theta_0\mu$ values of one to two orders of magnitude less



(a) (b) (c)

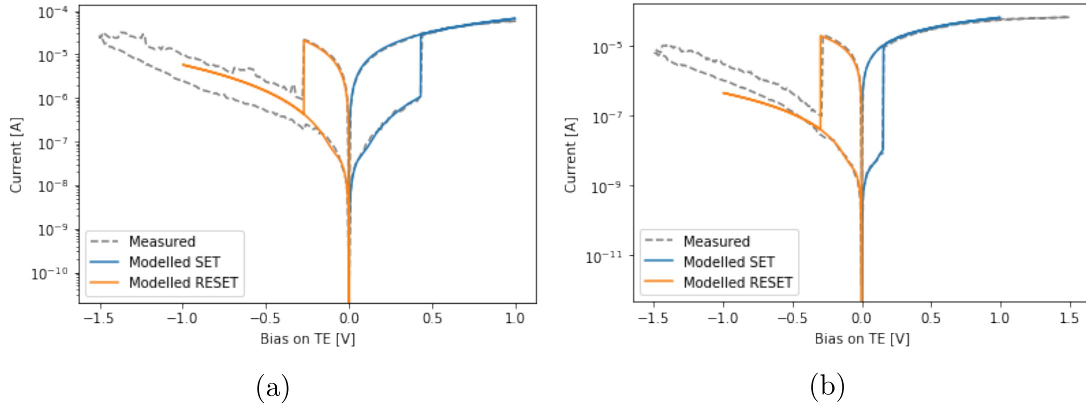
Figure 4.7: Reproduction of RRAM $I-V$ characteristics from a given device with a Python model based on structures from a different sample. (a), (b) and (c) show attempts at retracing different SET (blue) and RESET (orange) behaviours with the aid of the model, as compared to real measurements (dashed grey). HRS model parameters are of $R_{HRS} = 1.4\,\mathrm{M\Omega}$ (a), $19\,\mathrm{M\Omega}$ (b), $200\,\mathrm{G\Omega}$ (c); $\mu = 2.5{\cdot}10^{-5}\mathrm{m^2V^{-1}s^{-1}}$ (a), $1.6{\cdot}10^{-5}\mathrm{m^2V^{-1}s^{-1}}$ (b), $2{\cdot}10^{-10}\mathrm{m^2V^{-1}s^{-1}}$ (c); $x =$ ca. 1 nm for (a), (b), (c); $S =$ ca. 1 nm^2 for (a), (b) and (c) alike, given that the LRS is similar for the three ($R_{LRS} =$ ca. $10\,\mathrm{k\Omega}$).

than the estimates of mobility at lower biases are found; the decrease may be identified as a direct consequence of the ratio θ_0 coming into play.

On the topic of model credibility, specifically related to input parameter choice (see figure 4.7), it is reasonable to qualitatively believe in those parameters that yield figures 4.7a and 4.7b, as the HRSs in the latter are not far from each other, and resistance and mobility inputs that are not too low are adopted. However, since conduction occurs in the same medium in all measurements (i.e. a stretch of HfO_2) it is perhaps odd to come to the conclusion that different mobilities are the case for the different curves. Of course, deviations in mobility could be had, as this physical quantity is proportional to the average scattering time, which in turn is dependent on defects in the conduction layer (these may change between cycles, and with varying x). Nonetheless, it is interesting to see if the model can handle data reproduction assuming same μ, and varying other parameters. Specifically, in all three reproductions in figure 4.7, it was assumed that a CF-to-TE gap of $x = 1$ nm was the case. For a qualitative study this may pass, however, assuming that a MLP cell owes its multi-level in the HRS to the possibility for a more (or less) retracted filament, it would be appropriate to demand the model to reproduce data assuming approximately same μ, and varying x. This approach is shown in figure 4.8 for the first two curves in figure 4.7.

The third curve could not be reproduced assuming similar mobility (this could be an indication that such low current levels require a different interpretation). Also, for the realisation of figure 4.7c, an extremely low resistance (in the order of $100\,\mathrm{G\Omega}$) and a low mobility ($10^{-10}\mathrm{m^2V^{-1}s^{-1}}$) had to be assumed - it is somewhat difficult to say whether this is reasonable. Therefore, it is rather appropriate to conclude that two out of three curves could be reproduced (again, with the exception of high voltages in the post-RESET regime), with uncertainty in whether it is a change in mobility or a change in CF-to-TE gap (or both) to govern the possibility for MLP.

It is also important to point out that the LRS in figure 4.7c is somewhat misrepresented. However, this simply relates to the setup used when measuring this device as compared to others in this thesis. In fact, current compliance was imposed through a transistor, and it would seem like the compliance in figure 4.7c resulted as more limiting than that in figures 4.7a and 4.7b. This is further supported by the pre-RESET regime



(a) (b)

Figure 4.8: Reproduction of the same characteristics as in figure 4.7 (a) and (b), assuming same μ ($5{\cdot}10^{-6}\mathrm{m}^2\mathrm{V}^{-1}\mathrm{s}^{-1}$) and varying x (0.6 nm and 1.4 nm for (a) and (b), respectively).

in figure 4.7c, as transistor compliance is essentially removed in the RESET sweep, and here the measurement agrees well with the model.

Finally, it is worth mentioning that the model is capable of generating realistic V_{SET} and V_{RESET} values, following expected distributions for the latter. For instance, if a satisfyingly large sample of devices were to be characterised, and cumulative distribution functions were established for their SET and RESET points, these could be used as model inputs. This feature is not implemented in the above - the scope of the corroboration study was to recreate some fixed curves, hence respective V_{SET} and V_{RESET} values were used in the model instead of distributions.

Chapter 5

Conclusion and Outlook

In this report, a number of HfO_2-ITO RRAM structures are discussed, and conclusions are drawn for their performance, both in terms of benchmarking, and the underlying physics.

Direct observation of measured $I - V$ characteristics shows the possibility to FORM devices in the same voltage range, followed by appreciable variability in SET- and RESET-point, and low values for V_{SET} and V_{RESET}, hence confirming that this material choice is suitable for the pursuit of low-power operation. LRS inspection, however, leads to the conclusion that the foreseen effect of self-compliance does not have an impact here. It is speculated that further scaling (even more than in the device used for model corroboration - this also only showed Ohmic type conduction in the LRS) is necessary for this trait to come in play. No self-compliance translates to the need of an external device to control current levels in RRAM and ensure that a permanent state of dielectric breakdown is never reached - this has a toll on overall unit scalability. It is important to dedicate more time and effort to study self-compliance in HfO_2 and ITO RRAM, as this is one of the most important features for this material combination to be chosen over others.

The fit-to-data approach that is adopted in this work ultimately brings to the realisation that observing linearity in a given scale is not enough to conclude that the related conduction mechanism is a good description for conduction in RRAM. This seems like a somewhat common practice in the community, but should really be supported by appropriate parameter extraction computations to show that the proposed interpretation is indeed correct. Further, as foreseen by Yu Shimeng [42], fit to data may not be enough for the interpretation of current conduction in RRAM. Parameter extraction is an arduous task in that potentially unknown quantities come into play in computations, hence a more complete study would include results from complementary laboratory and computational techniques. For example, transmission electron microscopy could be used to corroborate assumptions that are hereby made for filament shape and size, whilst density functional theory could perhaps be used to investigate electronic structure, and lead to new ideas on how to tackle the interpretation of current conduction.

The above also links to the realisation of an accurate model. In this thesis, a reasonable interpretation has been reached, although the loose ends that were discussed in section 4.2 could be tightened by reducing uncertainty in model parameters. Regardless, the proposed interpretation implies another important result: it is impossible to interpret RRAM by assuming that a single conduction mechanism is capable of describing the HRS in its entirety; rather, one needs to combine different modes in different voltage regions to achieve a full interpretation.

As briefly discussed in section B.7, an addition to this thesis could be the formulation

of a reasonable background for conduction through RRAM in its HRS. Subtraction of this from raw data, followed by the same fitting study, could lead to an interpretation different from that which is hereby proposed. The reason why it was decided not to pursue this path lies in the fact that it would have added a number of new uncertainties (all the parameters that are necessary to the background) to the already faced uncertainties in the fitting followed by parameter extraction; it sounds interesting, yet it would likely lead to more speculation and confusion.

The fitting of a model to experimental data is a necessary step in the interpretation of unknown phenomena. On the way towards perfect understanding, heuristic conclusions might be necessary, exemplified by Max Plank's work in interpreting black-body radiation: this thesis does not claim to have an interpretation that works on all RRAM, albeit it establishes some tools for interpretation, and arrives to a plausible conclusion. Further corroboration of proposed results, with data from more HfO_2 and ITO samples, is necessary, and would be the most obvious extension to this thesis.

This work focuses on a few aspects of RRAM, although much is yet to be discussed about other benchmarking results (endurance, retention, scalability) and how these relate to the underlying physics. A nice addition would be the study of the correlation between HRS parameters and RESET-point. It is in fact the RESET which determines the HRS, whilst the SET leads to a supposedly well defined CF state regardless of V_{SET}, although a study of the dependence of the LRS on the SET-point might also be interesting. Furthermore, little to no investigation has been done on the impact of the BE material on device performance. Substituting TiN with some other electrode material could lead to a change in performance, both in terms of benchmarking and conduction mechanisms.

All in all, this study of HfO_2 and ITO RRAM scratches the surface of what can be understood about this material choice, not to mention of RRAM in general. However, it attempts to make the most rigorous explanation that is possible given what is known for sure, and provides the community with a point of reference for future studies. Also, the development of a model that seems to be qualitatively correct for the representation of multiple $I-V$ domains is of help to the research group in Lund; fine tuning of the model (ensuring that high voltages in the HRS are not misrepresented) inevitably yields a tool capable of quantitatively representing the RRAM at hand. This is likely the safest way to computationally reproduce the behaviour of locally grown structures - relying on models developed by other groups means relying on the interpretation of their own structures, which may be vastly different due to the variability that is inherent to RRAM.

Appendix A

Python Code for Plotting of and Fitting to DC Measurements

A.1 Imported Libraries

```
import numpy as np
import pandas as pd
import scipy.stats as sps
import matplotlib.pyplot as plt
```

A.2 SET- and RESET-point Identification

```
def find_SET(V_list, I_list):
    '''Finding SET values'''
    I_change = []
    V_switch = 0
    for k in range(5,len(V_list)):
        I_change.append(I_list[k] / I_list[k-1])
        shift_max_ind = I_change.index(max(I_change))
        max_ind = shift_max_ind + 4
        if (V_list[max_ind]) > 0:
            V_switch = (V_list)[max_ind]
    return(V_switch)


def find_RESET(V_list, I_list):
    '''Finding RESET values'''
    V_switch = 0
    for j in range(len(V_list)):
        if I_list[j] == max(I_list[0:100]):
            V_switch = V_list[j]
    return(V_switch)
```

A.3 Making and Plotting Linear Regressions

```
def linear_fit(v_data,i_data):
    m = sps.linregress(v_data,i_data)[0]
    q = sps.linregress(v_data,i_data)[1]
    return (m,q)
```

```
def plot_line(data_points,fit_function):
        x_values = data_points
        y_values = [fit_function(x) for x in x_values]
        plt.plot(x_values, y_values, 'orange', label='fit to data')
```

A.4 Plotting and Fitting in Schottky Scale

```
def plot_schottky(v_data,i_data):
    '''Plotting in lnI ~ sqrtV scale'''
    sqrtV = []
    for v in v_data:
        sqrtV.append(np.sqrt(abs(v)))
    lnI = []
    for i in i_data:
        lnI.append(np.log(i))

    plt.plot(sqrtV,lnI,'.')
    plt.title('Schottky')
    plt.xlabel(r'$\sqrt{V \: [\mathrm{V}]}$')
    plt.ylabel(r'$\ln{(I \: [\mathrm{A}])}$')
    plt.legend()
    plt.grid()
    plt.tight_layout()
    plt.show()


def fit_schottky(v_data,i_data):

    sqrtV = []
    for v in v_data:
        sqrtV.append(np.sqrt(abs(v)))
    lnI = []
    for i in i_data:
        lnI.append(np.log(i))

    plt.plot(sqrtV,lnI,'teal',marker='.',ls='',label='measured data')

    slope, intercept = linear_fit(sqrtV,lnI)[0], linear_fit(sqrtV,lnI)[1]
    fit_func = lambda v: slope*v + intercept

    plot_line(sqrtV,fit_func)

    plt.xlabel(r'$\sqrt{V \: [\mathrm{V}]}$')
    plt.ylabel(r'$\ln{(I \: [\mathrm{A}])}$')
    plt.legend()
    plt.grid()
    plt.tight_layout()

    return (slope, intercept)
```

A.5 Plotting and Fitting in Other Scales

The code which was implemented for plotting and fitting in other scales is analogous to that in section A.4, with the main difference being that what is displayed on the axes of the generated plots changes (e.g. I vs. V^2 for the Mott-Gurney law, as compared to $\ln \frac{I}{V}$ vs. V^2 for thermionic emission).

Appendix B

Parameter Extraction - Formulas and Code

The equations in section 2.5 are hereby rearranged for direct parameter extraction. Furthermore, each mentioned reiteration is presented alongside its Python equivalent, implemented in this project using fitting results as input.

B.1 Schottky Emission

Relevant scale: $\ln I$ vs. $\sqrt{V}$. From the slope, temperature T is deduced (equation B.1), whilst the intercept provides an estimate of emission barrier height $q\phi_B$ (equation B.2).

$$T = \frac{1}{k_B \,\text{slope}} \sqrt{\frac{q^3}{4\pi\epsilon_r\epsilon_0 x}} \tag{B.1}$$

$$q\phi_B = k_B T \left[\ln\left(SA^*T^2\right) - \text{intercept}\right] \tag{B.2}$$

In Python:

```
def extract_schottky(expslope, expintercept):
    T = 1/(k_B * expslope) * \
          np.sqrt(q**3 / (4*np.pi*epsilon_r*epsilon_0*x))
    q_phi_B = (np.log(S *A_0 * lambda_R * T**2) - expintercept) * \
               k_B * T
    return T,q_phi_B
```

B.2 Thermionic Emission

Relevant scale: $\ln \frac{I}{V}$ vs. V^2. From the slope, temperature T is deduced (equation B.3), whilst from the intercept, one can extrapolate the emission barrier height $q\phi_B$ (equation B.4).

$$T = \left(\frac{\hbar^2 q^2}{24 m x^2 k_B^3 \,\text{slope}}\right)^{1/3} \tag{B.3}$$

$$q\phi_B = k_B T \left[\ln\left(\frac{Sq^2\sqrt{mk_BT}}{8\hbar^2\pi^{5/2}x}\right) - \text{intercept}\right] \tag{B.4}$$

In Python:

```python
def extract_TAT(expslope, expintercept):
    T3x2 =  (hbar**2 * q**2) / (24 * m_0 * k_B**3 * expslope)
    T = (T3x2 / x**2)**(1/3)
    q_phi_B = (np.log((S * q**2 * np.sqrt(m_0 * k_B * T)) / \
                (8 * hbar**2 * np.pi**(5/2) * x)) - expintercept) * \
                k_B * T
    return T,q_phi_B
```

B.3 Pool-Frenkel Effect

Relevant scale: $\ln\frac{I}{V}$ vs. $\sqrt{V}$. From the slope, one may deduce temperature T (equation B.5), whilst from the intercept, the depth of trap sites $q\phi_T$ follows (equation B.6).

$$T = \frac{1}{k_B \,\text{slope}}\sqrt{\frac{q^3}{\pi\epsilon_r\epsilon_0 x}} \tag{B.5}$$

$$q\phi_T = k_B T\left[\ln\left(\frac{Sq\mu N_C}{x}\right) - \text{intercept}\right] \tag{B.6}$$

In Python:

```python
def extract_PF(expslope, expintercept):
    Trootx = 1 / (k_B * expslope) * np.sqrt(q**3 / \
            (np.pi * epsilon_r * epsilon_0))
    T = Trootx / np.sqrt(x)
    q_phi_T = (np.log((S * q * mu * N_C)/(x)) - expintercept) * \
                k_B * T
    return T,q_phi_T
```

B.4 Hopping Conduction

Relevant scale: $\ln I$ vs. V. Temperature T follows directly from the slope (equation B.7), whilst trap barrier height $q\phi_T$ is implied by the intercept (equation B.8).

$$T = \frac{aq}{xk_B \,\text{slope}} \tag{B.7}$$

$$q\phi_T = k_B T\left[\ln\left(Sqanf\right) - \text{intercept}\right] \tag{B.8}$$

In Python:

```python
def extract_hopping(expslope, expintercept):
    Tx_div_a = q / (k_B * expslope)
    T = Tx_div_a * a / x
    q_phi_T = (np.log(S * q * a * n * f) - expintercept) * \
                k_B * T
    return T,q_phi_T
```

B.5 Mott-Gurney Law (SCLC)

Relevant scale: I vs. V^2. Charge carrier (electron) mobility μ is determined from the slope, as seen in equation B.9.

$$\mu = \frac{8}{9}\frac{x^3}{S\epsilon_r\epsilon_0}\text{slope} \tag{B.9}$$

In Python:

```
def extract_MG(mg_slope):
    mu = (8 * x**3 * mg_slope) / (9 * S * epsilon_r * epsilon_0)
    return mu
```

B.6 Ohmic Conduction

Relevant scale: I vs. V. Resistance R to current conduction in the sample follows from the slope, as proposed by equation B.10. If measured in the LRS of RRAM, it is identified with filament resistance and can be used to estimate filament cross section as explained by equation 2.8. Otherwise it simply corresponds to resistance in HfO_2.

$$R = \frac{1}{\text{slope}} \tag{B.10}$$

In Python:

```
def extract_ohmic(ohm_slope, hrs=0):
    R = 1/ohm_slope
    if hrs==0:
        S = thickness * rho / R
    else:
        S = 0
    return R,S
```

Note that above, all computations are written for conduction from the tip of a broken filament to the TE in a given RRAM device. Should one want to consider bulk conduction (through the whole dielectric slab) instead, the only thing to do is replace x with the thickness of the HfO_2 layer, and S with the area of the device.

B.7 Note on Other Physical Quantities

To perform the estimates discussed in the above sections, a number of different physical parameters and constants are required. Many of these are unknown, but can be reasonably guessed to understand the validity of a given fit to data. In table B.1, values that were assumed for the sake of order of magnitude estimation in this thesis are listed. Comments and references are added where necessary.

Table B.1: Values of unknown parameters as reasonably guessed and/or researched for the purpose of parameter extraction in this work.

Quantity	Value
$m^* = m_{e,\mathrm{HfO_2}}$	0.11 m_e [60]
t	3 nm [1]
Device area	$7.85 \cdot 10^{-11}$ m^2 [1]
a	ca. 1 nm [2 3]
μ	10^{-10}-10^{-4} m^2 V^{-1} s^{-1} [4]
N_C	ca. 10^{24} m^{-3} [5]
n	10^{-16}-10^{16} m^{-3} [6]
f	ca. 10^{13} Hz [7]
x	ca. 1 nm [2]
CF cross section = S	ca. $2 \cdot 10^{-17}$ m^2 [8]
ρ	$3800 \cdot 10^{-9}$ Ω m [35]

B.8 Reliability of Extracted Parameters

As explained in section 4.1.4, only two extremes are considered when working towards an interpretation of conduction in RRAM: CF tip-to-TE vs. a dominating background that arises due to direct BE-to-TE conduction. Here, some further motivation as to why this is done is proposed. Specifically, these are the only two cases that can transparently be studied with the proposed method of fitting followed by parameter extraction, and together, they give an idea of whether something in between is instead the case. Should a "mixed" state of conduction be the case, where CF-to-TE is just as relevant as BE-to-TE background, one would need to come up with a reasonable interpretation of such background to subtract from raw data to then be able to perform meaningful fits. This is a guessing game even more dangerous than what is done in section B.7, as it involves not only guessing an appropriate conduction mechanism for the background, but also those parameters that are then necessary for the chosen mechanism (this seems like a study with many loose strings, however, it could be an interesting followup to this work). On the contrary, what is proposed in this thesis aims at directly realising the role of the background, by assuming that it is either the main contributor, or somewhat irrelevant. If any of the results in the bottom half of table 4.1 had been found to be credible, it

[1] Parameter known from processing.

[2] As guessed from device geometry.

[3] consulting publications such as [32].

[4] The two extremes are the lowest value extrapolated from SCLC fits to data, and electron mobility in amorphous Silicon, respectively. Pool-Frenkel interpretation relies on this. Specifically, if high mobility fails to yield physically meaningful parameters, this conduction mechanism may be discarded (as low mobility was correlated with worse parameter extraction).

[5] As suggested by $N_C = 2\left(\frac{2\pi m^* k_B T}{h^2}\right)^{3/2}$ [31].

[6] Wide range to consider. Highly (exponentially) dependent on the unknown position of the Fermi level in HfO_2. If hopping current parameter extraction fails to meet expectations, assuming the higher limit of concentration, hopping is to be discarded as an option for current conduction.

[7] Standard room temperature value, as suggested by $k_B T/h$.

[8] As suggested by Ohmic type conduction interpretation of RRAM LRS. Note that different S's may be computed for all evaluated devices simply by considering respective LRSs - this is done and the result is used as input into the different parameter extraction formulas.

would have meant that the background likely had an important impact towards overall HRS conduction, and the chase for a mixed-state interpretation would have inevitably started. Given that this was not the case, it was (somewhat safely) assumed that the BE-to-TE contribution could be neglected.

It is also worth it to spend some words on how large of an error one expects on the physical parameters that are extrapolated in this thesis. Specifically, this is a relevant consideration given that some quantities in table B.1 had to be guessed.

Even though educated guesses have been made, supported by arguments as in section 4.1.4 and in the footnotes of section B.7, small deviations are still possible, and these inevitably yield different results from those in section 4.1.4. Mainly, the guessing of the gap x, separating the tip of a ruptured filament from the TE, is the most obvious error contribution to the estimation of reasonable quantities. Given the geometry of the studied RRAM, 1 nm seems like a good estimate for x; however, one could argue that anything in the range between 0.5 and 1.5 nm is reasonable (less and more seem unlikely as they would yield high leakage and poor conduction, respectively). Parameter extraction following the formulas in this appendix then yield that percentage errors of around 50 % are possible, i.e. the results in section 4.1.4 should be taken as order of magnitude estimates (they are definitely not exact), but all the drawn conclusions still apply.

μ and n are other, highly unknown, parameters that need to be assumed. However, a maximum is supposed for them, as it was observed that lowering them yielded arguably worse estimates. Given that even these maxima are not enough to yield credible estimates (even when accounting for the error allowed by x), it is irrelevant to speak further of the errors that might arise from changing μ and n, as these are unidirectional, and cannot bring to a different interpretation from that which is concluded in section 4.1.4. Other necessary parameters (f, N_C, S, ρ, a) are even less unfounded (supported by references), and errors arising from their choice are not discussed.

All in all, it can be concluded that parameter extraction in this thesis is far from perfect, and would be somewhat disappointing if its scope were to accurately determine given quantities. However, for the purpose of understanding whether a conduction mode is relevant or not, it is more than adequate and fulfils its duty.

Appendix C

Python Model

C.1 Imported Libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

C.2 SET- and RESET-point Generating Functions

```
def generate_V_SET(expected_V_SET, deviation):
    SET_value = np.random.normal(expected_V_SET, deviation)
    return SET_value

def generate_V_RESET(expected_V_RESET, deviation):
    RESET_value = np.random.normal(expected_V_RESET, deviation)
    return RESET_value
```

C.3 Current Defining Functions

```
def ohmic(V, R):
    I_ohmic = (1/R)*V
    return I_ohmic

def mg(V, epsilon_r, mu, t, A):
    epsilon_0 = 8.8541878128e-12 # units: F / m
    I_mg = (9/8) * epsilon_r * epsilon_0 * \
           mu * (1/t**3) * A * (abs(V)**(2))
    return I_mg
```

C.4 SET Curve Generation

```
def SET(state, V_stop, expected_V_SET, deviation):

    current_state = state
    V_sweep = (np.linspace(0, V_stop, 1000, endpoint=True)).tolist()
    V_back = (np.linspace(V_stop, 0, 1000, endpoint=True)).tolist()
    V_SET = generate_V_SET(expected_V_SET, deviation)
    print('V_SET = {} V'.format(V_SET))

    I_device = []
    I_temp = 0
```

```
    temp = 0

    '''Generating a realistic I-V curve:'''

    if (state != 'LRS'):
        if (state != 'HRS'):
            print('Invalid input state.')
            return True

    for i in range(len(V_sweep)):

        temp = i

        if V_sweep[i] >= V_SET:
            current_state = 'LRS'

        if current_state == 'HRS':
            if V_sweep[i] <= 0.05:
                I_temp = ohmic(V_sweep[i], HRS_R)
            if V_sweep[i] > 0.05 and V_sweep[i] < 0.15:
                I_temp = ohmic(V_sweep[i], HRS_R) * \
                         ((0.15-V_sweep[i])/(0.1)) + \
                         mg(V_sweep[i], epsilon_r, mu, t, A) * \
                         ((V_sweep[i] - 0.05)/(0.1))
            if V_sweep[i] >= 0.15:
                I_temp = mg(V_sweep[i], epsilon_r, mu, t, A)

        if current_state == 'LRS':
            if temp == 0:
                print('Device is already on!')
                return True
            else:
                I_temp = ohmic(V_sweep[i],LRS_R)

        I_device.append(I_temp)

    for i in range(len(V_back)):
        I_device.append(ohmic(V_back[i],LRS_R))

    plt.plot(V_sweep+V_back, I_device,label='Modelled SET')

    return True
```

C.5 RESET Curve Generation

```
def RESET(state, V_stop, expected_V_RESET, deviation):
    current_state = state
    V_sweep = (np.linspace(0, V_stop, 10000, endpoint=True)).tolist()
    V_back = (np.linspace(V_stop, 0, 10000, endpoint=True)).tolist()
    V_RESET = generate_V_RESET(expected_V_RESET, deviation)
    print('V_RESET = {} V'.format(V_RESET))
    I_device = []
    I_temp = 0
    temp = 0

    '''Generating a realistic I-V curve:'''

    if (state != 'LRS'):
```

```
        if (state != 'HRS'):
            print('Invalid input state.')
            return True

    for i in range(len(V_sweep)):

        temp = i

        if V_sweep[i] <= V_RESET:
            current_state = 'HRS'

        if current_state == 'LRS':
            I_temp = ohmic(V_sweep[i],LRS_R)

        if current_state == 'HRS':
            if temp == 0:
                print('Device is already off!')
                return True
            else:
                if V_sweep[i] <= -0.15:
                    I_temp = mg(V_sweep[i], epsilon_r, mu, t, A)

                if V_sweep[i] > -0.15 and V_sweep[i] < -0.05:
                    I_temp = mg(V_sweep[i], epsilon_r, mu, t, A) * \
                             (abs(V_sweep[i] + 0.05)/(0.1)) + \
                             ohmic(abs(V_back[i]), HRS_R) * \
                             (abs(V_sweep[i] + 0.15)/(0.1))

                if V_sweep[i] >= -0.05:
                    I_temp = ohmic(abs(V_back[i]), HRS_R)

        if I_temp >= 1e-3:
            I_device.append(1e-3)
        else:
            I_device.append(abs(I_temp))

    for i in range(len(V_back)):
        if V_back[i] <= -0.15:
            I_device.append(mg(V_back[i], epsilon_r, mu, t, A))

        if V_back[i] > -0.15 and V_back[i] < -0.05:
            I_device.append(mg(V_back[i], epsilon_r, mu, t, A) * \
                            (abs(V_back[i] + 0.05)/(0.1)) + \
                            ohmic(abs(V_back[i]), HRS_R) * \
                            (abs(V_back[i] + 0.15)/(0.1)))

        if V_back[i] >= -0.05:
            I_device.append(ohmic(abs(V_back[i]), HRS_R))

    plt.plot(V_sweep+V_back, I_device,label='Modelled RESET')

    return True
```

Bibliography

[1] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, April 1965.

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

[3] D. Strukov, G. Snider, D. Stewart, and S. Williams, "The missing memristor found," *Nature*, vol. 453, pp. 80–83, June 2008.

[4] Clarivate Analytics, "Web of science." `https://clarivate.com/products/web-of-science/`. Accessed: 2019-12-04.

[5] K. Kim and G. H. Koh, "Future memory technology including emerging new memories," in *2004 24th International Conference on Microelectronics (IEEE Cat. No.04TH8716)*, vol. 1, pp. 377–384, May 2004.

[6] T.-C. Chang, K.-C. Chang, T.-M. Tsai, T.-J. Chu, and S. M. Sze, "Resistance random access memory," *Materials Today*, vol. 19, no. 5, pp. 254 – 264, 2016.

[7] H. Li, T. F. Wu, A. Rahimi, K. Li, M. Rusch, C. Lin, J. Hsu, M. M. Sabry, S. B. Eryilmaz, J. Sohn, W. Chiu, M. Chen, T. Wu, J. Shieh, W. Yeh, J. M. Rabaey, S. Mitra, and H. . P. Wong, "Hyperdimensional computing with 3d vrram in-memory kernels: Device-architecture co-design for energy-efficient, error-resilient language recognition," in *2016 IEEE International Electron Devices Meeting (IEDM)*, pp. 16.1.1–16.1.4, December 2016.

[8] D. Ielmini, "Brain-inspired computing with resistive switching memory (rram): Devices, synapses and neural networks," *Microelectronic Engineering*, vol. 190, pp. 44 – 53, 2018.

[9] C. Pan, T. Chang, T. Tsai, K. Chang, T. Chu, C. Shih, C. Lin, P. Chen, H. Wu, N. Deng, H. Qian, and S. M. Sze, "Ultralow power resistance random access memory device and oxygen accumulation mechanism in an indium–tin-oxide electrode," *IEEE Transactions on Electron Devices*, vol. 63, pp. 4737–4743, December 2016.

[10] C. Ye, C. Zhan, T. M. Tsai, K. C. Chang, M.-C. Chen, K.-M. Chang, T. Deng, and H. Wang, "Low-power bipolar resistive switching tin/hfo2/ito memory with self-compliance current phenomenon," *Applied Physics Express*, vol. 7, p. 034101, February 2014.

[11] U. Russo, D. Ielmini, C. Cagli, A. Lacaita, S. Spiga, C. Wiemer, M. Perego, and M. Fanciulli, "Conductive-filament switching analysis and self-accelerated thermal

dissolution model for reset in nio-based rram," *2007 IEEE International Electron Devices Meeting*, pp. 775–778, 2007.

[12] B. Gao, S. Yu, N. Xu, L. F. Liu, B. Sun, X. Y. Liu, R. Q. Han, J. F. Kang, B. Yu, and Y. Y. Wang, "Oxide-based rram switching mechanism: A new ion-transport-recombination model," in *2008 IEEE International Electron Devices Meeting*, pp. 1–4, December 2008.

[13] S. Yu, B. Lee, and H.-S. P. Wong, *Metal Oxide Resistive Switching Memory*, pp. 303–335. New York, NY: Springer New York, 2012.

[14] P. Huang, X. Y. Liu, B. Chen, H. T. Li, Y. J. Wang, Y. X. Deng, K. L. Wei, L. Zeng, B. Gao, G. Du, X. Zhang, and J. F. Kang, "A physics-based compact model of metal-oxide-based rram dc and ac operations," *IEEE Transactions on Electron Devices*, vol. 60, pp. 4090–4097, December 2013.

[15] M. Kozicki and H. Barnaby, "Conductive bridging random access memory - materials, devices and applications," *Semiconductor Science and Technology*, vol. 31, October 2016.

[16] E. Ahn, H.-S. P. Wong, and E. Pop, "Carbon nanomaterials for non-volatile memories," *Nature Reviews Materials*, vol. 3, p. 18009, March 2018.

[17] K. C. Chang, T. M. Tsai, K.-M. Chang, K.-H. Chen, R. Zhang, Z.-Y. Wang, J.-H. Chen, T.-F. Young, M.-C. Chen, T.-J. Chu, S.-Y. Huang, Y.-E. Syu, D.-H. Bao, and S. Sze, "Dual ion effect of the lithium silicate resistance random access memory," *Electron Device Letters, IEEE*, vol. 35, pp. 530–532, May 2014.

[18] D. Jana, S. Roy, R. Panja, M. Dutta, D. S. Z. Rahaman, R. Mahapatra, and S. Maikap, "Conductive-bridging random access memory: challenges and opportunity for 3d architecture," *Nanoscale Research Letters*, vol. 10, December 2015.

[19] S. Sonde, B. Chakrabarti, Y. Liu, K. Sasikumar, J. Lin, L. Stan, R. Divan, L. Ocola, D. Rosenmann, P. Choudhury, K. Ni, S. Sankaranarayanan, S. Datta, and S. Guha, "Silicon compatible sn-based resistive switching memory," *Nanoscale*, vol. 10, April 2018.

[20] G. Khurana, N. Kumar, J. Scott, and R. Katiyar, "Graphene oxide-based memristor," *IntechOpen*, April 2018.

[21] D. Cazorla-Amorós, "Grand challenges in carbon-based materials research," *Frontiers in Materials*, vol. 1, p. 6, 2014.

[22] O. Theodoridis, "Rram arrays - master's thesis." Lund University, 2020.

[23] Y. Guo and J. Robertson, "Materials selection for oxide-based resistive random access memories," *Applied Physics Letters*, vol. 105, no. 22, p. 223516, 2014.

[24] R. Zhang, K. C. Chang, K.-M. Chang, T. M. Tsai, S.-Y. Huang, W.-J. Chen, K.-H. Chen, J.-C. Lou, J.-H. Chen, T.-F. Young, M.-C. Chen, H.-L. Chen, S.-P. Liang, Y.-E. Syu, and S. Sze, "Characterization of oxygen accumulation in indium-tin-oxide for resistance random access memory," *Electron Device Letters, IEEE*, vol. 35, pp. 630–632, June 2014.

[25] S. Nehate, A. Prakash, P. D. Mani, and K. Sundaram, "Work function extraction of indium tin oxide films from mosfet devices," *ECS Journal of Solid State Science and Technology*, vol. 7, pp. P87–P90, January 2018.

[26] Y. Y. Chen, B. Govoreanu, L. Goux, R. Degraeve, A. Fantini, G. Kar, D. Wouters, G. Groeseneken, J. Kittl, M. Jurczak, *et al.*, "Balancing set/reset pulse for endurance in 1t1r bipolar rram," *IEEE Transactions on Electron Devices*, vol. 59, p. 3243, December 2012.

[27] D. C. Gilmer, G. Bersuker, H. . Park, C. Park, B. Butcher, W. Wang, P. D. Kirsch, and R. Jammy, "Effects of rram stack configuration on forming voltage and current overshoot," in *2011 3rd IEEE International Memory Workshop (IMW)*, pp. 1–4, May 2011.

[28] J. Song, D. Lee, J. Woo, Y. Koo, E. Cha, S. Lee, J. Park, K. Moon, S. H. Misha, A. Prakash, and H. Hwang, "Effects of reset current overshoot and resistance state on reliability of rram," *IEEE Electron Device Letters*, vol. 35, pp. 636–638, June 2014.

[29] F.-C. Chiu, "A review on conduction mechanisms in dielectric films," *Advances in Materials Science and Engineering*, vol. 2014, pp. 1–18, February 2014.

[30] C. Crowell, "The richardson constant for thermionic emission in schottky barrier diodes," *Solid-State Electronics*, vol. 8, no. 4, pp. 395 – 399, 1965.

[31] S. Sze and K. Ng, *Physics of Semiconductor Devices.* Wiley, 2006.

[32] F.-C. Chiu, C.-Y. Lee, and T.-M. Pan, "Current conduction mechanisms in pr2o3/oxynitride laminated gate dielectrics," *Journal of Applied Physics*, vol. 105, no. 7, p. 074103, 2009.

[33] S. Zaima, "Conduction mechanism of leakage current in ta[sub 2]o[sub 5] films on si prepared by LPCVD," *Journal of The Electrochemical Society*, vol. 137, no. 9, p. 2876, 1990.

[34] R. Millikan, E. Bishop, and A. T. Society, *Elements of Electricity: A Practical Discussion of the Fundamental Laws and Phenomena of Electricity and Their Practical Applications in the Business and Industrial World.* American Technical Society, 1917.

[35] D. Ielmini, "Modeling the universal set/reset characteristics of bipolar rram by field- and temperature-driven filament growth," *IEEE Transactions on Electron Devices*, vol. 58, pp. 4309–4317, December 2011.

[36] C. D. Child, "Discharge from hot cao," *Phys. Rev. (Series I)*, vol. 32, pp. 492–511, May 1911.

[37] J. V. D., "Electronic processes in ionic crystals (mott, n. f.; gurney, r. w.)," *Journal of Chemical Education*, vol. 42, no. 9, p. A692, 1965.

[38] P. N. Murgatroyd, "Theory of space-charge-limited current enhanced by frenkel effect," *Journal of Physics D: Applied Physics*, vol. 3, pp. 151–156, February 1970.

[39] M. Lanza, H.-S. P. Wong, E. Pop, D. Ielmini, D. Strukov, B. C. Regan, L. Larcher, M. A. Villena, J. J. Yang, L. Goux, A. Belmonte, Y. Yang, F. M. Puglisi, J. Kang, B. Magyari-Köpe, E. Yalon, A. Kenyon, M. Buckwell, A. Mehonic, A. Shluger, H. Li, T.-H. Hou, B. Hudec, D. Akinwande, R. Ge, S. Ambrogio, J. B. Roldan, E. Miranda, J. Suñe, K. L. Pey, X. Wu, N. Raghavan, E. Wu, W. D. Lu, G. Navarro, W. Zhang, H. Wu, R. Li, A. Holleitner, U. Wurstbauer, M. C. Lemme, M. Liu, S. Long, Q. Liu, H. Lv, A. Padovani, P. Pavan, I. Valov, X. Jing, T. Han, K. Zhu, S. Chen, F. Hui, and Y. Shi, "Recommended methods to study resistive switching devices," *Advanced Electronic Materials*, vol. 5, no. 1, p. 1800143, 2019.

[40] H. Wong, H. Lee, S. Yu, Y. Chen, Y. Wu, P. Chen, B. Lee, F. Chen, and M. Tsai, "Metal-oxide rram," *Proceedings of the IEEE*, vol. 100, pp. 1951–1970, June 2012.

[41] S. Balatti, S. Ambrogio, Z. . Wang, S. Sills, A. Calderoni, N. Ramaswamy, and D. Ielmini, "Pulsed cycling operation and endurance failure of metal-oxide resistive (rram)," in *2014 IEEE International Electron Devices Meeting*, pp. 14.3.1–14.3.4, December 2014.

[42] S. Yu, "Resistive random access memory (rram)," *Synthesis Lectures on Emerging Engineering Technologies*, vol. 2, pp. 1–79, March 2016.

[43] C. Ye, C. Zhan, T.-M. Tsai, K.-C. Chang, M.-C. Chen, T.-C. Chang, T. Deng, and H. Wang, "Low-power bipolar resistive switching TiN/HfO2/ITO memory with self-compliance current phenomenon," *Applied Physics Express*, vol. 7, p. 034101, February 2014.

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

[45] H. Y. Lee, P. S. Chen, T. Y. Wu, Y. S. Chen, C. C. Wang, P. J. Tzeng, C. H. Lin, F. Chen, C. H. Lien, and M. . Tsai, "Low power and high speed bipolar switching with a thin reactive ti buffer layer in robust hfo2 based rram," in *2008 IEEE International Electron Devices Meeting*, pp. 1–4, December 2008.

[46] N. Xu, B. Gao, L. F. Liu, B. Sun, X. Y. Liu, R. Q. Han, J. F. Kang, and B. Yu, "A unified physical model of switching behavior in oxide-based rram," in *2008 Symposium on VLSI Technology*, pp. 100–101, June 2008.

[47] S. Kim, H. Moon, S. Yoo, and Y.-K. Choi, "Resistive switching characteristics of sol–gel zinc oxide films for flexible memory applications," *Electron Devices, IEEE Transactions on*, vol. 56, pp. 696 – 699, May 2009.

[48] H. Lv, M. Yin, X. Fu, Y. Song, L. Tang, P. Zhou, C. Zhao, T. Tang, B. Chen, and Y. Lin, "Resistive memory switching of Cu_xO films for a nonvolatile memory application," *Electron Device Letters, IEEE*, vol. 29, pp. 309 – 311, May 2008.

[49] X. Sun, B. Sun, L. Liu, N. Xu, X. Liu, R. Han, J. Kang, G. Xiong, and T. Ma, "Resistive switching in CeO_x films for nonvolatile memory application," *Electron Device Letters, IEEE*, vol. 30, pp. 334 – 336, May 2009.

[50] M. Yang, J. W. Park, T. Ko, and J.-K. Lee, "Bipolar resistive switching behavior in ti/mno2/pt structure for nonvolatile memory devices," *Applied Physics Letters*, vol. 95, pp. 042105–042105, July 2009.

[51] K.-M. Persson, M. S. Ram, M. Borg, and L.-E. Wernersson, "Investigation of reverse filament formation in ito/hfo2-based rram," 2019.

[52] *Fundamentals of Atomic Layer Deposition*, ch. 1, pp. 1–31. John Wiley & Sons, Ltd, 2013.

[53] N.-T. Nguyen, "Chapter 4 - fabrication technologies," in *Micromixers (Second Edition)* (N.-T. Nguyen, ed.), Micro and Nano Technologies, pp. 113 – 161, Oxford: William Andrew Publishing, second edition ed., 2012.

[54] R. Cirelli, G. Watson, and O. Nalamasu, "Optical lithography," in *Encyclopedia of Materials: Science and Technology* (K. J. Buschow, R. W. Cahn, M. C. Flemings, B. Ilschner, E. J. Kramer, S. Mahajan, and P. Veyssière, eds.), pp. 6441 – 6448, Oxford: Elsevier, 2001.

[55] A. Sarangan, "5 - nanofabrication," in *Fundamentals and Applications of Nanophotonics* (J. W. Haus, ed.), pp. 149 – 184, Woodhead Publishing, 2016.

[56] D. Depla, S. Mahieu, and J. Greene, "Chapter 5 - sputter deposition processes," in *Handbook of Deposition Technologies for Films and Coatings (Third Edition)* (P. M. Martin, ed.), pp. 253 – 296, Boston: William Andrew Publishing, third edition ed., 2010.

[57] Cascade Microtech, Inc., *Summit 11K/12K Probe Stations User's Guide*, August 2001.

[58] Tektronix, *Model 4200A-SCS Parameter Analyzer Reference Manual*, September 2019.

[59] J. Hölzl and F. K. Schulte, *Work function of metals*, pp. 1–150. Berlin, Heidelberg: Springer Berlin Heidelberg, 1979.

[60] S. Monaghan, P. Hurley, K. Cherkaoui, M. A. Negara, and A. Schenk, "Determination of electron effective mass and electron affinity in hfo2 using mos and mosfet structures," *Solid-State Electronics*, vol. 53, pp. 438–444, April 2009.

[61] P. Fauchet, D. Hulin, R. Vanderhaghen, A. Mourchid, and W. Nighan, "The properties of free carriers in amorphous silicon," *Journal of Non-Crystalline Solids*, vol. 141, pp. 76 – 87, 1992.

[62] Q. Li, L. Qiu, X. Wei, B. Dai, and H. Zeng, "Point contact resistive switching memory based on self-formed interface of al/ito," *Scientific reports*, vol. 6, p. 29347, July 2016.

[63] C.-Y. Lin, S.-Y. Wang, D.-Y. Lee, and T.-Y. Tseng, "Electrical properties and fatigue behaviors of ZrO2 resistive switching thin films," *Journal of The Electrochemical Society*, vol. 155, no. 8, p. H615, 2008.