Department of Electronic & Electrical Engineering
UNIVERSITY COLLEGE LONDON

Resistive Switching in Silicon-rich Silicon Oxide

Adnan MEHONIĆ

Supervisor : Dr Anthony KENYON

Dissertation submitted
for the degree of Doctor of Philosophy

December, 2013

I, Adnan Mehonic confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Over the recent decade, many different concepts of new emerging memories have been proposed. Examples of such include ferroelectric random access memories (FeRAMs), phase-change RAMs (PRAMs), resistive RAMs (RRAMs), magnetic RAMs (MRAMs), nano-crystal floating-gate flash memories, among others. The ultimate goal for any of these memories is to overcome the limitations of dynamic random access memories (DRAM) and flash memories.

Non-volatile memories exploiting resistive switching – resistive RAM (RRAM) devices – offer the possibility of low programming energy per bit, rapid switching, and very high levels of integration – potentially in 3D. Resistive switching in a silicon-based material offers a compelling alternative to existing metal oxide-based devices, both in terms of ease of fabrication, but also in enhanced device performance.

In this thesis I demonstrate a redox-based resistive switch exploiting the formation of conductive filaments in a bulk silicon-rich silicon oxide. My devices exhibit multi-level switching and analogue modulation of resistance as well as standard two-level switching. I demonstrate different operational modes (bipolar and unipolar switching modes) that make it possible to dynamically adjust device properties, in particular two highly desirable properties: non-linearity and self-rectification. Scanning tunnelling microscopy (STM), atomic force microscopy (AFM), and conductive atomic force microscopy (C-AFM) measurements provide a more detailed insight into both the location and the dimensions of the conductive filaments.

I discuss aspects of conduction and switching mechanisms and we propose a physical model of resistive switching.

I demonstrate room temperature quantisation of conductance in silicon oxide resistive switches, implying ballistic transport of electrons through a quantum constriction, associated with an individual

silicon filament in the SiO_x bulk.

I develop a stochastic method to simulate microscopic formation and rupture of conductive filaments inside an oxide matrix. I use the model to discuss switching properties – endurance and switching uniformity.

Contents

1 Introduction to Digital Memories **7**

Memory Devices . . . 8

Classification of Matrix-based Memories . . . 9

Read Only Memory . . . 10

Random Access Memory . . . 11

Static Random Access Memory and Dynamic Random Access Memory . . . 12

Ferroelectric Random Access Memory . . . 14

Nano-electro-mechanical Random Access Memory . . . 15

Magnetoresistive Random Access Memory . . . 16

Phase Change Random Access Memory . . . 17

Organic Random Access Memory . . . 18

Flash Memory . . . 19

Ferroelectric FET Memory . . . 20

The Memristor . . . 21

Memristor Model . . . 22

The HP Memristor . . . 25

Redox Based Random Access Memories . . . 27

Phenomenological Description of ReRAM . . . 27

Requirements for ReRAM as Non-volatile Memory . . . 31

Electrochemical Metallization Cell Memory . . . 36

Valence Change Memory . . . 37

Thermochemical Memory Cells . . . 39

Introduction to Resistive Switching in Silicon Oxide . . . 40

2 Experimental Techniques and Sample Preparation **44**

Electrical Characterisation . . . 44

Electrical Probe Station . . . 44

Impedance Spectroscopy . . . 45

Structural Characterisation . . . 46

Scanning Tunnelling Microscopy . . . 46
Atomic Force Microscopy . . . 48
Transmission Electron Microscopy . . . 51
Sample Preparation . . . 53
1st, 2nd and 3rd Generation . . . 53
4th Generation . . . 55

3 Resistive Switching in Active Silicon Oxide 56

Electrical Characterisation . . . 56
Current-voltage Characteristics . . . 57
Bipolar Switching Mode . . . 57
Unipolar Switching Mode . . . 63
Multi-level and Gradual Switching . . . 66
Temperature Dependence . . . 68
Current-time Characteristics . . . 70
Impedance Spectroscopy . . . 70
Nonlinearity and Self-rectification . . . 71
Structural Characterisation . . . 72
AFM and STM Characterisation . . . 72
C-AFM Characterisation . . . 74
TEM Characterisation . . . 78

4 Quantum Conductance in Silicon Oxide ReRAM Devices 81

Mesoscopic Electron Transport . . . 81
Introduction to Quantum Conductance . . . 82
Introduction to Quantum Conductance in RRAM Devices . . . 82
Conduction Quantisation in Silicon Oxide RRAM . . . 83
Reproducibility of G-V Curves . . . 88
Nonlinear Components of the Conductance . . . 88
Theoretical Model – Nonlinear Quantum-Point Transport Mode . . . 93
Distinction between the Half-integer and Integer Quantisation RRAM Systems . . . 96
Conclusion . . . 100

5 Phenomenological Model of Resistive Switching in Silicon Oxide 101

Introduction . . . 101
Conduction Mechanism . . . 101

Fowler-Nordheim Tunnelling 106
Poole-Frenkel Conduction Mechanism 107
Trap-assisted Tunnelling 108
Conduction Mechanism Results 109
Switching Model 115
Multi-channel Conduction Using Quantum Point Contact
Theory 120
Model Description and Fitting Results 120

6 Stochastic Simulation of Resistive Switching in Oxide RRAM 130
Introduction to the Stochastic Model 130
Model Description 131
Simulation Results 136
Electroforming 137
Reset Process 138
Set Process 139
Endurance Testing 140
Resistive Switching without an External Oxygen Ion Source 141
Unipolar Switching 141
Bipolar Switching 143
Resistive Switching with Unlimited Oxygen Vacancies Source 144

7 Conclusion and Future Plans 146

Acknowledgements

I would like to use this opportunity to thank everybody who supported me during my doctorate.

Firstly, I am deeply grateful to my supervisor Dr Tony Kenyon. His continuous support, advice and knowledge have made my time at UCL one of the best experiences of my life – both academically and personally. His understanding and patience have made the whole process very enjoyable. For all this time I felt very lucky that I have had not only an ideal professional mentor but also a friend.

I am also grateful to all my collaborators who have made a big contribution to my work and from whom I have learnt a lot. I thank Dr Sébastien Cueff, Dr Christophe Labbé, Prof Richard Rizk and Prof Enrique Miranda who have been great collaborators from the start of the project. I would like to thank Dr Vasillios Albanis and UCLb who greatly helped my work through our strategic and financial partnership.

Special thank you goes to my parents for all their kind support, love and understanding. I thank my sister Aida for sharing the ups and downs during this period. I thank Nejra for being the best support I could ask for.

Finally I thank all my friends in Bosnia and the UK for their love and support.

Chapter 1

Introduction to Digital Memories

This chapter provides an overview of digital memories. Different memory concepts are classified into several groups. The basic physical principles governing the processes of different memory types are briefly discussed. Typical memory properties of each type are compared.

The special focus is on resistive random access memories (RRAM) and, more specifically, on redox-based resistive random access memories (ReRAM).

RRAM is often seen as a special case of a general class of technologies that exhibit memristance[1] – a resistance that depends on the past history of the device (a "memristor"[2]). There is some controversy over the definition of memristor. RRAM is classified as a type of memristive system rather than a memristor. This is discussed later in the chapter. Memristors can be non-volatile memory elements, novel logic gates, or elements in neural networks.

There is therefore considerable interest in memristor/RRAM commercialisation as replacement for flash memory and other non-volatile memory systems.

As successive generations of NAND technology strive to boost performance, processing and physical limitations make it increasingly difficult to shrink device size and increase bit-count per cell without encountering power dissipation and speed problems.

RRAM offers a promising solution, requiring far fewer mask steps than existing flash technologies, with far higher areal densities. The International Technology Roadmap for Semiconductor (ITRS)

International Roadmap Committee report into emerging post-NAND memory technologies (2010) states: "*...incentives for developing RRAM technologies include their low cost-per-bit, low-voltage (and low power dissipation) operation, high endurance, and their plausible suitability for integration in crossbar arrays stacked in multiple levels in 3D.*"[3]

Several RRAM technologies are currently under investigation, including phase-change materials (chalcogenides, perovskites, Ge sulphide and selenide), and, currently most promising, metal oxides (NiO, TiO, HfO_2) in which resistance is switched through the formation and destruction of conductive filaments. Work is moving rapidly from fundamental research to development. However, issues around understanding the switching mechanism, device reliability and processing remain critical. The consensus is that switching is predominantly due to the migration of oxygen vacancies under the application of external fields, although Joule heating plays an important role too. Much of the microscopic detail of the switching process is largely conjecture.

Memory Devices

Information storage represents a crucial step in passing human experience and knowledge in time and space. Number tracking evolved from the system of counting using small clay tokens to impressing a stylus into clay to first pen writings to printing via moveable parts and finally electronic printing. Similar scenarios occurred for sound and picture recording. In 2011 it was estimated that humankind has stored more than 295 billion gigabytes of data since 1986[4].

Generally speaking, electronic storage systems are classified into two groups – random access devices and mass storage devices. This is shown in Table 1.1 [5].

	Mass Storage Devices	Random Access Devices
System	Access unit Storage medium	Matrix of conductor lines Storage elements are at nodes
Addressing of Data	Positioning of the read/write unit and storage medium	Application of address signals (columns and rows)
Data exchange	Mechanical Optical Magnetic Electric	Electronic access via matrix

Table 1.1: Classification of electronic storage systems

Random access devices are controlled in a parallel way, providing a faster access. They have a matrix organisation. Every memory cell is addressed by appropriately addressing two address lines – bit line and word line.

Mass storage devices are more suitable for storing larger amounts of data. In this case, access of data is sequential. This means that the operational speed is slower and it depends on the physical position of accessed data.

There is a clear trade-off between the operational speed and the overall storage capacity. Different types of memories are used for different purposes – from very fast, low capacity cache memories to much slower higher capacity permanent memories like hard disk drives.

Although there is a very clear difference between these two types there are many memory concepts sharing the properties of both. For example, solid-state memories (SSD) are addressed in the same way as hard-disk drives (HDD), but they have an internal matrix organisation like flash memories (an example of random access memory).

Classification of Matrix-based Memories

Different memory states are represented by different physical states of a memory unit. In most cases, these physical states are either different charge states or different resistance states.

Matrix-based memories are further classified into two types – random access memories (RAM) and read only memories (ROM). The first type is used as a main memory in semiconductor devices such as personal computers, mobile phones, portable electronics. The second type is used when the access time needs to be as short as possible. Memory capacity of ROM is lower compared to RAM. ROM is typically used for the instruction storage.

These two types of memories are further divided in several categories as shown in Figure 1.1.

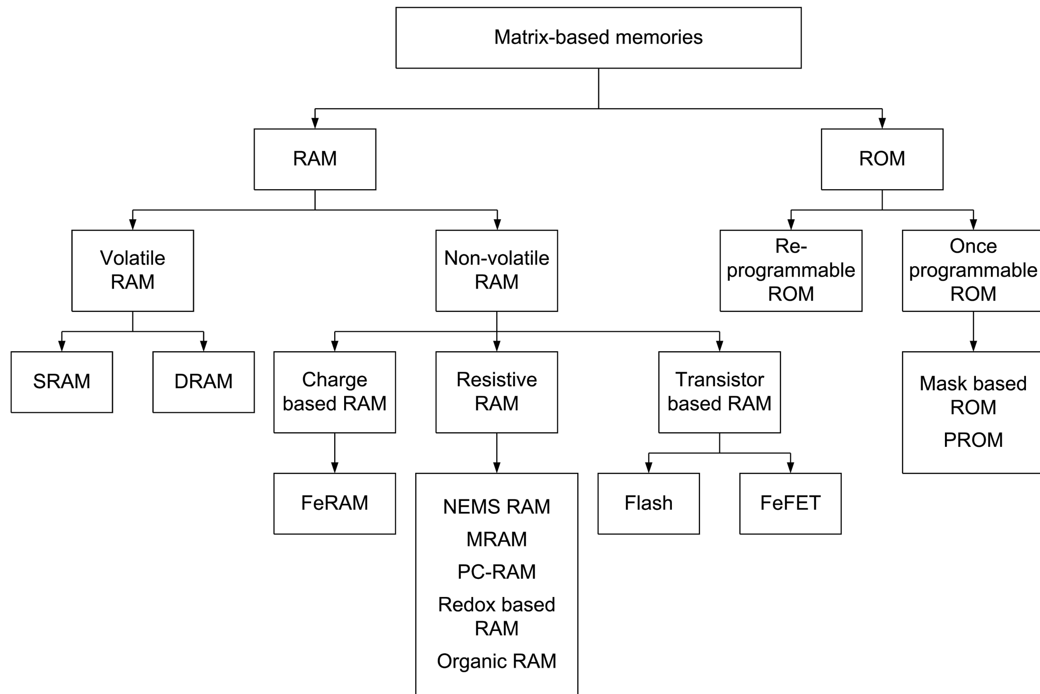

Figure 1.1: Classification of Matrix-based memories. After [5].

Read Only Memory

Read only memory (ROM) is classified into two groups – re-programmable ROM and once programmable ROM.

In the case of once programmable ROM, the state of the memory is defined during the first programming step or the last fabrication step. In either case these states are permanent and they cannot be changed. When defining the memory state during the first programming step, the metal bridges at the nodes of the matrix are fused. Therefore either fused or intact nodes represent "0s" and "1s". The matrix can also be defined during the last fabrication step – different mask layouts represent different states (mask-based ROM).

In case of re-programmable ROM the matrix's configuration can be changed. The process of reprogramming can be done either electrically – electrically erasable PROM (EEPROM), or by application of UV light that discharges all units – erasable PROMs (EPROM).

Random Access Memory

When it comes to random access memory (RAM) there are two types of matrix organisation – passive matrix organisation and active matrix organisation.

Matrix memory organisation is shown in Figure 1.2. The rows of address lines (called word lines – WL) are perpendicular to the columns of address lines (called bit lines – BL). At the node of every intersection there is a memory unit/resistive switch (RS). In case of passive organisation there is no additional active switch. However the nonlinear element (NLE), usually a diode, in series to RS is used.

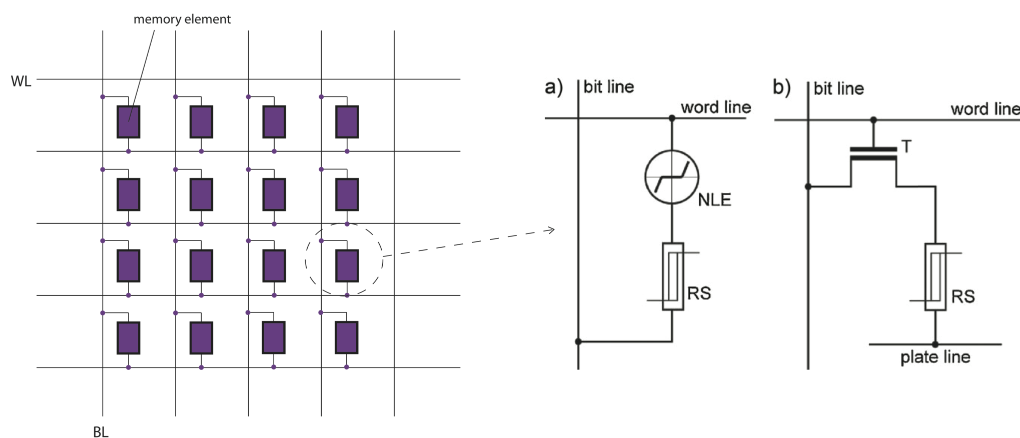

Figure 1.2: Memory crossbar array. (a) Passive memory matrix. (b) Active memory matrix. After [6].

Selecting a different WL and BL line accesses a single memory unit. Every cell that is located in the same WL or BL experiences a fraction of the addressing signal – even non-selected cells. To prevent an unintentional switching event (switching between two logic states – "1"and "0"), a well-defined method (threshold values) is required. Furthermore, during the reading process the occurrence of many parasitic current sneak paths is possible. If system is not well adapted, a faulty reading will occur.

A passive memory matrix allows a high device density (which is only limited by the size of a memory unit), low fabrication costs and the possibility of 3D architectures – stacking of multi layers in a 3D configuration. However, in the case of RRAM with passive matrix ogranisation, sneak currents through non-selected cells limit the maximum size of the matrix.

Active memory matrix, as an addition to a memory cell, includes an active switch. This active

switch is a selector device, usually a transistor (T), that overcomes the problem of sneak paths and fractional stressing of non-selected cells. Although this greatly helps in respect of the maximum size of an individual matrix, it comes at the cost of an additional element.

RAM is further classified into two groups – volatile RAM and non-volatile RAM.

In the case of volatile RAM the memory state is lost once the power supply is removed. For continuous retention of data an external power supply is required. There are two types of volatile RAM – static RAM (SRAM) and dynamic RAM (DRAM).

Static Random Access Memory and Dynamic Random Access Memory

SRAM uses a bistable latch configuration. The two stable states represent logical "1" or "0". In contrast to dynamic RAM static RAM does not require a periodical refreshment of data. Still, data is lost when the power supply is removed – therefore SRAM is still classified as a volatile type of memory.

The typical cell consists of 6 MOSFET transistors, and has two stable states. The states are stored in four transistors – 2 cross-coupled inverters formed by transistors M_1 – M_4 as shown in Figure 1.3(a). Transistors M_5 and M_6 provide an access to the inverters.

In the case of DRAM, a capacitor charge is used to represent a memory state. Every bit state (zero or one) is represented by the charge of the capacitor (charged or discharged capacitor). To prevent loss of data caused by the constant discharging of capacitor (through the leakage current), the states of capacitors are periodically refreshed. This requirement for a refresh process defines DRAM as a dynamic type of memory. DRAM is very simple compared to other types of memories – it requires only one transistor and one capacitor per bit. Consequently the density of DRAM is very high – billions of bits can be fitted on a single chip.

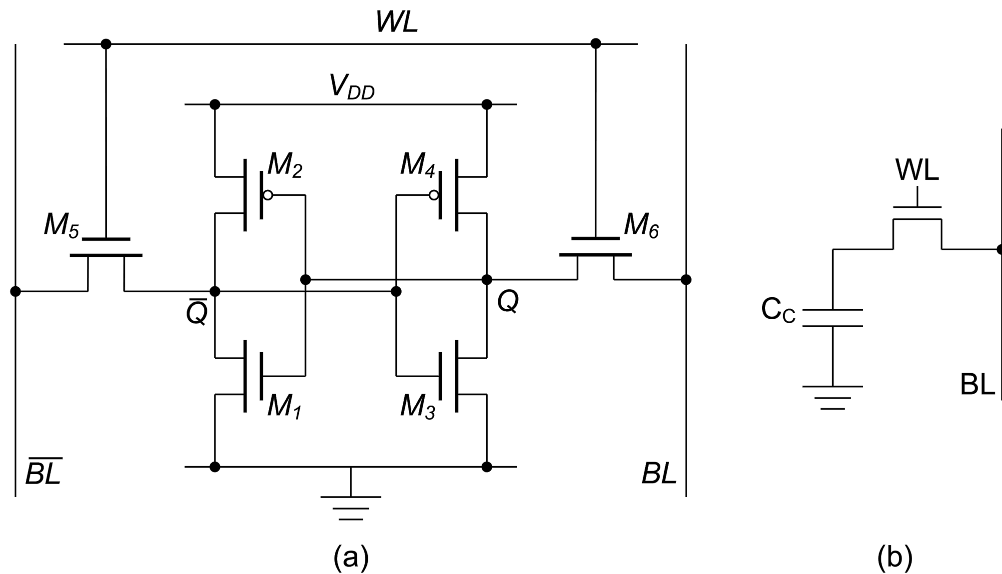

Figure 1.3: (a) SRAM memory cell. (b) DRAM memory cell.

DRAM is much faster than non-volatile types of memory (e.g. flash) – it operates on the nanosecond scale (to write or read a bit state), while flash requires at least tens of microseconds.

Although DRAM uses much smaller space than equivalent SRAM (capacitor and selector in the case of DRAM, compared to six transistors in the case of SRAM), SRAM is still used for fast memory like cache. The access time of SRAM is shorter than the access time of DRAM (approximately 50ns in case of DRAM and 10ns in case of SRAM).

Non-volatile RAM does not require a power supply for the retention of data. The most used type of non-volatile RAM is a flash memory.

There are still a few drawbacks of flash memory despite it being one of the most used memory types. These include slow access times compared to volatile RAM – either DRAM or SRAM, limited number of write-erase cycles (normally around 100,000), and in some cases limited random addressability. Due to these intrinsic drawbacks of a flash, many alternative approaches are being developed. Although the early results are promising (compared to flash or even SRAM), they are still not mainstream.

Non-volatile ROM can be classified into three groups based on main physical principle of data storage: charge-based RAM, resistive RAM and transistor-based RAM.

The operation of charge-based RAM relies on different levels of charge stored in the memory cell. These states are non-volatile, as they do not require a power supply. A typical example is ferroelectric RAM (FeRAM).

Ferroelectric Random Access Memory

FeRAM uses a typical MOSFET structure. The main difference compared to a typical MOSFET is that the insulator between the gate and the transistor body is a ferroelectric. Different polarisation states of the ferroelectric represents different memory states. The amount of charge flowing into the device depends directly on the polarisation state. When an external electric field is applied dipoles inside the ferroelectric align themselves with the field that is produced by small shifts of electronic charge in the crystal structure. The dipoles retain their state even after the charge is removed.

During the reading process the ferroelectric is forced into a particular state. If the forced state is the same state as the ferroelectric was before, nothing will happen. If there is a change of the state a short current pulse will be generated due to re-orientation of atoms. The reading process is destructive and it is required to re-write the state after reading. These states are non-volatile. The switching speeds (around 150 ns) are much faster than the ones of typical flash. Although data retention is significantly improved in the last few years, the main issue with FeRAM is lower storage density and higher cost compared to high volume memories that are already on the market.

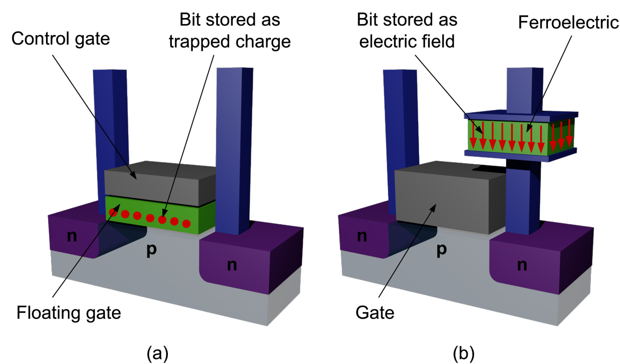

Figure 1.4: Schematic representation of the conventional floating gate transistor – (a) Flash memory cell. (b) FeRAM cell. After[7].

Nano-electro-mechanical Random Access Memory

Resistance-based RAM relies on different resistance states of the memory cell. There are many different types of RRAM. These include nano-electro-mechanical RAM (NEMS RAM or NRAM), magnetoresistive random-access memory (MRAM), phase change RAM (PC-RAM or PRAM), redox-based RAM (ReRAM), and organic RAM.

The memory state in nano-electro-mechanical RAM (NRAM) is represented by the position of carbon nanotubes (CNT). When nanotubes are in contact the resistance state is low, which represents the logic state “1” and when there is no contact the resistance is much higher, representing logic state “0”.

An NRAM memory cell is shown in Figure 1.5(a).

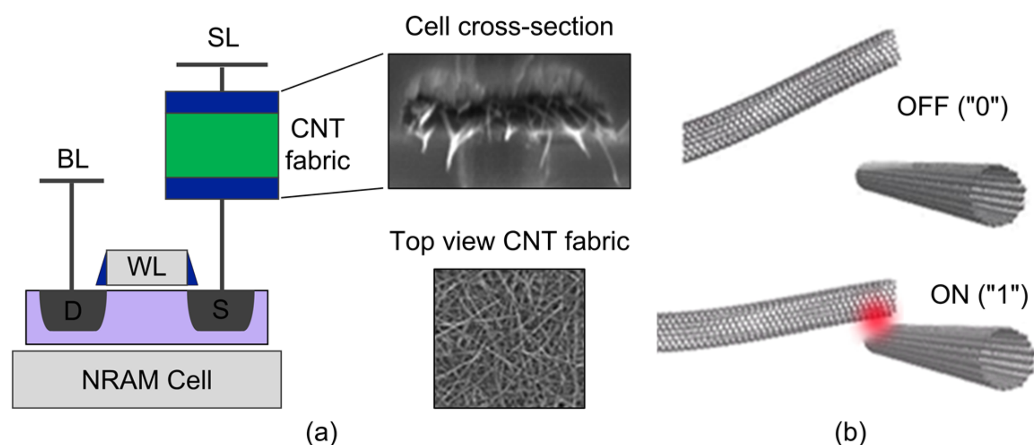

Figure 1.5: (a) Basic memory cell structure. (b) Representation of memory state – OFF state (CNT are not in contact), ON state (CNT in contact). After[8].

The carbon nanotubes are sandwiched between two metal electrodes. A small electric pulse is used to bring the nanotubes in contact or to separate them – as shown in Figure 1.5(b).

When two nanotubes are separated a short pulse greater than the read pulse will cause an electrostatic attraction between the CNTs, and they will be brought into contact. This represents the set operation (the transition from “0” to “1”). The nanotubes stay in contact due to Van der Waals forces. For the reset process (the transition from “1” to “0” state) a short reset pulse is used to generate a phonon energy high enough to separate them.

The density of NRAM is defined by the feature size of the electric contacts. The size of the CNT

is much smaller than this and does not contribute to the overall cell size. Consequently, only the lithography process limits the memory density. Theoretically, this means that NRAM density can surpass the density of DRAM while being non-volatile and performing similarly to SRAM (in terms of speed and power). The technology is still in development by company Nantero.[8]

Magnetoresistive Random Access Memory

A magnetisation state is used to represent the memory states of magnetnoresistive RAM (MRAM). The basic cell structure is shown in Figure 1.6.

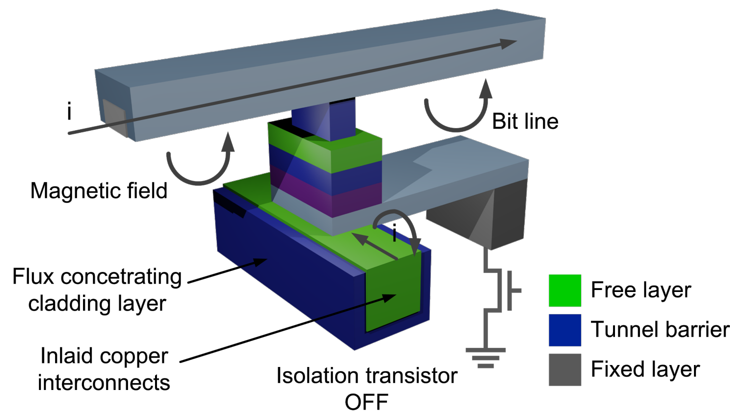

Figure 1.6: The structure of an MRAM cell with bit and word lines. After[9].

The memory cell consists of two ferromagnetic plates separated by an insulating layer. One of the plates is fixed and the magnetic moment of the plate is permanently set. The magnetic moment of the second plate can be changed by an external magnetic field. The magnetic moments of the two plates can be either in parallel or in anti-parallel to each other.

These two cases represent the two memory states – "1" and "0".

Every memory cell is sandwiched between two lines – BL and WL. An induced magnetic field is generated at the junction when a current passes through the lines. This field can change the polarisation state of the cell. The state of the cell is read by means of electrical resistance of the cell. An anti-parallel arrangement of two magnetic moments will provide a higher resistance of the cell.

MRAM is still in the early stage of its development, but initial results indicate that MRAM can be as fast as SRAM. Since the 1990s MRAM has been a continuous subject of both industry and academic research. It is often seen as a "universal memory" to replace not only flash but also SRAM and DRAM.

Phase Change Random Access Memory

Many materials can be found in two distinct phases – amorphous and crystalline. They are commonly fabricated in either one or other phase, but there are some particular materials which exhibit fast transitions from one phase to the other. These materials exhibit a large electrical and optical contrast between these two phases with high stability in either of them, which makes them very interesting for memory-like applications.

The idea of media storage in phase-changing materials dates back to the 1960s with the preposition that chalcogenides might be used as the storage media. But this idea became more technologically viable only with the discovery of fast crystallisation in certain materials. The best results have been reported for $Ge_2Sb_2Te_5$ (GST), which is still the most used phase changing material[10, 11]. Optical data storage based on these materials already had a great success (e.g. Blu-Ray disks and other rewritable optical disks).

A Phase Change Memory (PCM) consists of cells which could either pass current, when they are in the ON state (crystalline state), or block the current when in the OFF state (amorphous state). By heating the cell via a tiny heather/resistor it is possible to change state either to amorphous or crystalline. Different voltage pulses control the material crystallisation – either to allow crystals to grow from the melted mass, or to prevent the growth. These processes happen at the localised area and they are very fast – a few nanoseconds. A schematic of a typical PCM cell is shown in Figure 1.7(a). Intense heating and abrupt cooling afterwards turns the material into the amorphous phase. Longer, weaker heating turns the material into the crystalline phase, as shown in Figure 1.7(b).

Although PCRAM can already be found in volume production[12] several challenges are still present.

High temperature sensitivity of PCM is still the main drawback. Moreover, the existing commercial memories are proving to be more economical. Nevertheless due to other excellent properties (in particular the scaling prospect) it is expected that PCM will play an important role in the general memory market in the near future.

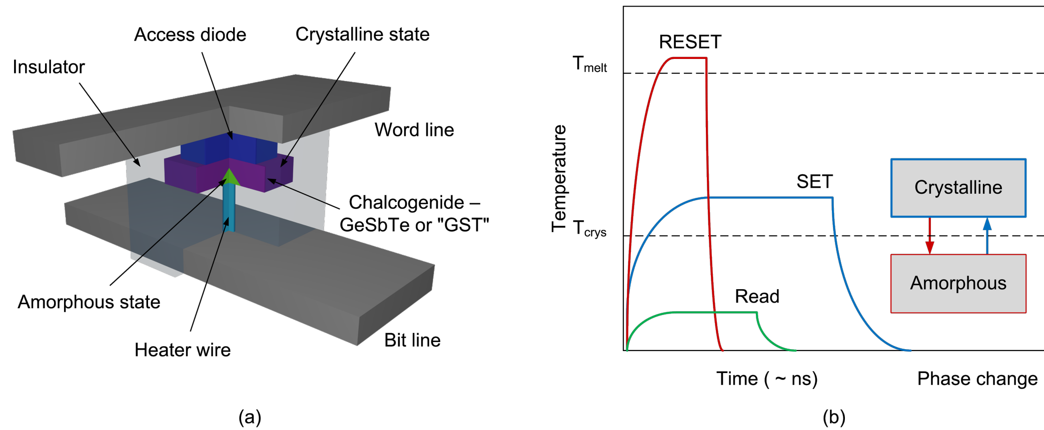

Figure 1.7: (a) Cross-point PCM memory cell. (b) Typical T-t curves of PCM element.

Organic Random Access Memory

Another type of RRAM is organic RAM. The basic unit cell is an organic material sandwiched between two electrodes. Organic RAM has already shown good memory properties. At least two stable memory states, good endurance, good retention, fast switching, good ON/OFF ratio have been reported[13, 14].

The main issue with organic RRAM is the still quite unclear physical picture of the switching process. Various organic materials have shown resistive switching when a voltage bias is applied[15]. These materials include small molecules, polymers and composites containing nanoparticles.

A few structures have been explored:

- Single organic layer sandwiched between the contacts;
- Multi-layer structures containing at least two types of organic material;
- Spin-cast polymer nanoparticles embedded in host material.

The main advantage of organic RRAM is its low fabrication cost, scalability and flexibility, while poor reliability is the main issue.

Once the physics behind the switching is better understood this particular technology might find its place in non-volatile memory market – especially in new flexible memory applications. At the moment organic RRAM is still in its early stages of development.

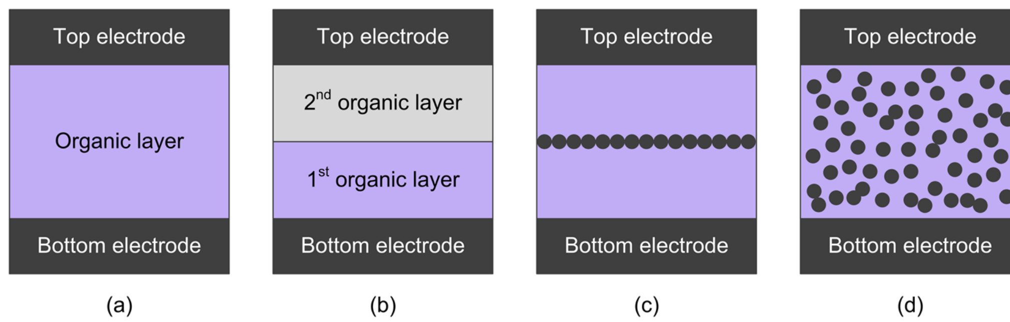

Figure 1.8: Typical structures of organic memory cells: (a) a single-layer structure without nanoparticles (NPs), (b) a bilayer structure containing two kinds of polymers, (c) a structure with nano-traps buried in the middle of an organic layer, and (d) a polymer-NP composite with NP traps randomly distributed throughout the entire host polymer. After[16].

Flash Memory

Data is stored in the gate of a selecting transistor in the case of transistor based RAM. Typical examples are flash memory and ferroelectric field effect transistors (FeFETs).

Flash memory, based on floating gate transistors, still dominates the non-volatile memory market. The basic flash memory unit is the floating-gate MOSFET transistor.

A schematic of a MOSFET transistor is shown in Figure 1.9. The floating gate (FG) is electrically isolated from the external control gate and drain and source. As there is no resistive connection between FG and the control gate, the charge stored in the FG remains the same for long periods of time. To modify the charge stored in the FG a high voltage is applied to the control gate. Charging or discharging of the FG occurs through Fowler-Nordheim tunnelling and hot-carrier injection mechanisms. Different amounts of charge stored in FG modify the threshold voltage and this represents different memory states.

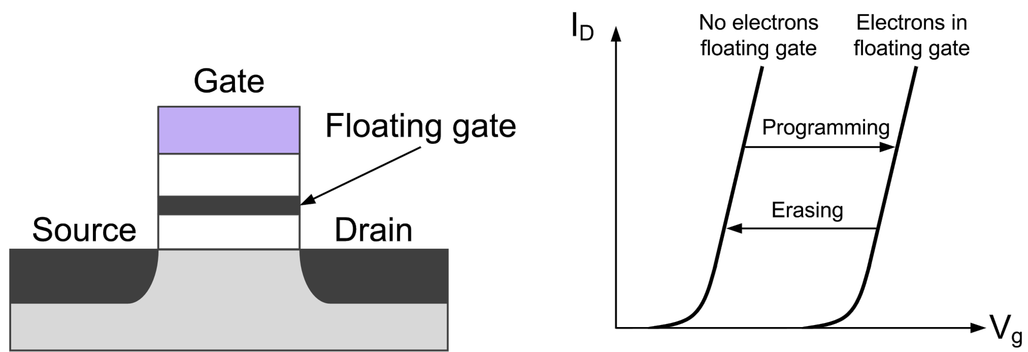

Figure 1.9: Floating gate transistor and the basic working principle of flash memory.

Ferroelectric FET Memory

Ferroelectric FET, similarly to ferroelectric RAM, uses ferroelectric material instead of the gate dielectric of a MOSFET.

In contrast to FeRAM, the reading process in a FeFET is not destructive. By application of an external field the polarisation of the ferroelectric can be changed. Two different polarisations represent two logic states, and different polarisation states yield different threshold voltages of the FET.

The FeFET is much faster, and operates at significantly lower voltages, than flash. A disadvantage of FeFET is low data retention – the 10 years limit, required for non-volatile memories, is still not reached.

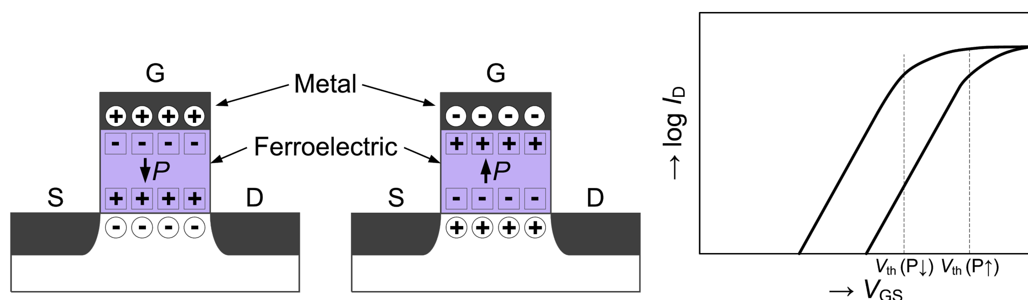

Figure 1.10: Schematic of FeFET cell and corresponding I_D-V_{GS} characteristics.

The Memristor

Although Leon Chua first introduced the name "Memristor" in 1971, the concept of electrically variable resistance is much older. In a recent commentary publication[**?**] Leon Chua (with other authors) states: "*The memristor is not an invention. Rather it is a description of a basic phenomenon of nature that manifests itself in various dissipative devices, made from different materials, internal structures and architectures. We end this historical narrative by noting that even though the memristor has seen its light of joy only recently in 2008, and has been recognized as the fourth circuit element along with the resistor, capacitor and inductor, it actually predates the resistor, which was formally published by Ohm in 1827, and the inductor, which was formally published by Faraday in 1831*".

Here is a brief historical background:

- In 1960, the device called the "Memistor" was introduced by Bernard Widrow from Stanford University. His device had three terminals and the conductance between the two terminals was controlled by the time integral of the current that goes into third terminal.
- In 1968, three years before Chua's paper on "Memristors – The Missing Circuit Element", a paper entitled "Switching phenomena in titanium oxide thin films"[**?**] was published reporting similar results to these of the HP memristor work that came 40 years after, in 2008.
- In 1971, Leon Chua defined a "Fourth Circuit Element" calling it the "Memristor" [2].
- In 1976, Leon Chua published another paper giving the mathematical generalisation of the model of the Memristor[17].
- In 1990, variable resistance was observed in tungsten oxide that can be electrically programmable[18].
- In 2000, research group from IBM's Zurich Research Laboratory reported reproducible resistive switching in oxide thin films[19].
- In 2008, HP reported a direct link between resistive switching in their titanium oxide device and Chua's model [20].

Between the initial work of Chua in 1971 and the HP report on titanium oxide based memristor in 2008 many papers have reported memristor-like behaviour, but only the HP report found a direct link between resistive switching in their device and Chua's model. They found a link between the

resistance switching in two-terminal device and Chua's mathematical model. Since this announcement, the number of reports of memristors has increased dramatically. Globally, the number of publications on memristors has jumped from ≈49 in 2008 to ≈2400 in 2013.

Memristor Model

In circuit theory there are three basic circuit elements and four circuit variables. The fundamental circuit variables are current i, voltage v, flux φ and charge q.

Five well-understood relations relate these four fundamental variables, but there are six possible combinations. Three basic circuit elements are defined as two-terminal devices relating the two fundamental variables. In that respect, there is the resistor R, the capacitor C and the inductor L. They are defined by the following relations:

$$dV = Rdi \tag{1.1}$$

$$dq = Cdv \tag{1.2}$$

$$d\phi = Ldi \tag{1.3}$$

Variables are also related as

$$i = \frac{dq}{dt} \tag{1.4}$$

$$v = \frac{d\phi}{dt} \tag{1.5}$$

Looking at symmetry in Figure 1.11 it is immediately obvious that there is one missing circuit element to complete the symmetry. This element should relate charge q and flux φ. Chua predicted that such element exists and he called it a memristor.

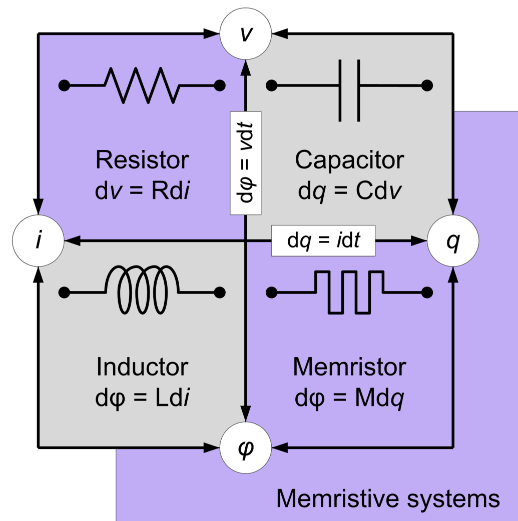

Figure 1.11: Relations between fundamental circuit variables.

The memristor gives a functional relation between the flux and charge. The commonly used definition of a memristor is that it is a two-terminal circuit element in which the flux between the two terminals is a function of the amount of electric charge that has passed through the device [10].

The derivation of memristance is given by:

$$\phi = f(q) \tag{1.6}$$

$$\frac{d\phi}{dt} = \frac{df(q)}{dq}\frac{dq}{dt} \tag{1.7}$$

Voltage is given by:

$$v(t) = \frac{d\phi}{dt} \tag{1.8}$$

so

$$v(t) = M(q)i(t) \tag{1.9}$$

where

$$M(q) = \frac{df(q)}{dq} \tag{1.10}$$

where M stands for memristance and its unit is the same as for resistance. This value linearly relates voltage and current if the charge does not change. Consequently, memristance M becomes resistance in the case that it does not change.

Memristance is the fundamental property of the memristor. Resistance of a memristor decreases if the charge flows in one direction, and it increases if it flows in the opposite direction. When the charge flow stops, the memristor "remembers" the last resistance and it resides in that state until the charge flow starts again.

There are many physical systems that exhibit the same phenomenological behaviour as the memristor. These are not restricted on electrical systems only. The mechanical example of a memristor is a water pipe whose shape depends on direction of water's flow.

Memristive systems are generally described by two equations:

$$v = M(w, i)i \tag{1.11}$$

and

$$\frac{dw}{dt} = f(w, i) \tag{1.12}$$

where w is a set of state variables, M and f are function of time, and v and i are voltage and current [17].

The memristor is a passive circuit element, meaning it can only dissipate energy; it cannot store any energy. As a consequence, memristance must be non-negative. The Φ-q curve needs to be monolithically increasing and the slope of the curve gives the memristance.

The main fingerprint of a memristor is the "pinched hysteresis curve" in the I-V characteristics. This pinched loop shrinks when the frequency of the input voltage increases.The memristor becomes a resistor once the frequency approaches infinity. When the input voltage is zero, the current passing through memristor has to be zero too (Figure 1.12). It is often said that "if a system exhibits such behaviour it is a memristor" [21], although there is some ongoing debate over this claim.

There is some controversy if certain types of devices exhibiting variable resistance are indeed memristors. Chua in his recent paper states: "*Resistance switching memories are memristors*" and "*All 2-terminal non-volatile memory devices based on resistance switching are memristors, regardless of the device material and physical operating mechanisms. They all exhibit a distinctive "fingerprint" characterized by a pinched hysteresis loop confined to the first and the third quadrants of the v-i plane whose contour shape in general changes with both the amplitude and frequency of any periodic "sine-wave-like" input voltage source, or current source. In particular, the pinched hysteresis loop*

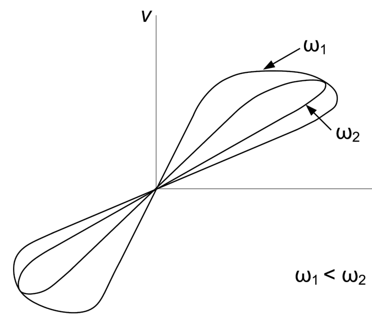

Figure 1.12: The pinched hysteresis loop and the loop shrinking with the increase in frequency.

shrinks and tends to a straight line as frequency increases. Though numerous examples of voltage v_s. current pinched hysteresis loops have been published in many unrelated fields, such as biology, chemistry, physics, etc., and observed from many unrelated phenomena, such as gas discharge arcs, mercury lamps, power conversion devices, earthquake conductance variations, etc., we restrict our examples in this tutorial to solid-state and/or nano devices where copious examples of published pinched hysteresis loops abound. In particular, we sampled arbitrarily, one example from each year between the years 2000 and 2010, to demonstrate that the memristor is a device that does not depend on any particular material, or physical mechanism. For example, we have shown that spin-transfer magnetic tunnel junctions are examples of memristors. We have also demonstrated that both bipolar and unipolar resistance switching devices are memristors."[19].

In a recent publication, Valov et al put a special emphasis on the distinction between memristors, memristive systems and resistive memories: "*Recently, the serendipitous discovery of the link between redox-based nanoionic-resistive memory cells and memristors and memristive devices has further intensified the research in this field. Here we show on both a theoretical and an experimental level that nanoionic-type memristive elements are inherently controlled by non-equilibrium states resulting in a nanobattery. As a result, the memristor theory must be extended to fit the observed non-zero-crossing I-V characteristics.*"[22]. In other words this brings the necessity for the revision of the original memristor theory to include the ReRAM. Demonstration of the internal voltage within the ReRAM clearly violates the memristor postulate as a passive circuit element.

The HP Memristor

In 2008, a research group working in HP Labs published the paper entitled "The missing memristor found". Stanley Williams and his group were working on thin titanium oxide films. It was an MIM

structure consisting of a thin (a few nanometres in thickness) titanium dioxide film rich with oxygen vacancies sandwiched between two platinum electrodes.

Titanium dioxide is intrinsically an insulator, meaning it has a very high resistance. Making the oxide richer with oxygen vacancies increases the conductivity. Engineering the stoichiometry of the film, during the fabrication process, produces different profiles of oxygen vacancies. The oxygen vacancies behave as positively charged species. By applying an external voltage bias it is possible to drift the oxygen vacancies.

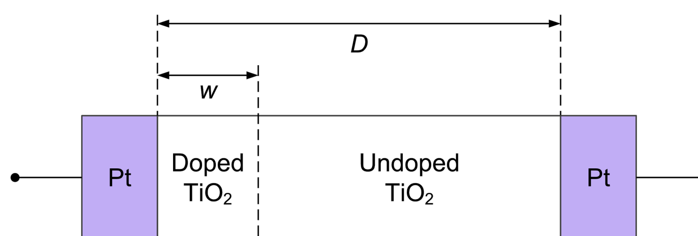

Figure 1.13: Schematic of the HP memristor. After[23].

In a simplified model, the whole oxide structure can be described as if it consists of two parts. One part is pure TiO_2, while the second part is titanium dioxide rich with oxygen vacancies, $TiO_{2\text{-}x}$. Rearrangement of oxygen vacancies defines the position of boundary between these two regions.

The first part TiO_2 is highly resistive while the second part $TiO_{2\text{-}x}$ is more conductive due to the presence of oxygen vacancies. When an external voltage is applied, positively charged oxygen vacancies will drift towards the negatively biased electrode changing the position of the boundary between two regions. The resistance of the whole film will change. If the oxygen vacancies are moving towards the left electrode the overall resistance will increase, as the boundary will be pushed further towards the left electrode. Similarly, if the vacancies are moving in the opposite direction the resistance will decrease.

The direction of vacancy drift is controlled by the external voltage, thus the resistance of the device is controlled by voltage. When the external voltage is turned off the device keeps the same resistance, as the boundary between the two parts does not move.

The HP device can be mathematically described as:

$$M(q) = R_{off}\left(1 - \frac{R_{on}}{\beta}q(t)\right) \tag{1.13}$$

where $\beta = \frac{D^2}{\mu_D}$; μ_D is the average drift velocity, D is the thickness of the oxide film, R_{off} and R_{on} are two states of ON and OFF resistance and $q(t)$ is the charge passing through the device.

Redox Based Random Access Memories

A local redox process governs the resistive switching in redox based RAM (ReRAM).

This redox process is triggered by an external voltage bias. Depending on whether the process is thermally or electrically driven there are several different types of ReRAM.

Both thermal and electrical effects are always acting simultaneously, but typically only one of them dominates. The three main types of redox RRAM are: electrochemical metallization cells (or conductive bridge) ReRAMs (ECM), valence change ReRAMs (VCM) and thermochemical ReRAMs (TCM).

Phenomenological Description of ReRAM

Resistive switching is the main phenomenon seen in these devices. A thin insulating film can reversibly change its electrical properties – from an insulating state to a conducting state under the application of an external electrical stimulus. At least two stable states are achievable – normally called the high resistive state (HRS), or OFF state, and the low resistive state (LRS), or ON state. In many cases more than two stable states are observed[24, 25].

The basic memory unit is a two-terminal device consisting of two conducting electrodes and the switching thin film sandwiched inbetween. This is shown in Figure 1.14.

When the two electrodes are biased, the thin film can change its resistance. This change of resistance can be very large – typically several orders of magnitude.The initial transition from the insulating to the conducting state is called electroforming. After electroforming, the device can be cycled between the two stable states (LRS and HRS). This is achieved by applying appropriate voltage pulses. The pristine device, device before the electroforming, is typically more resistive than the HRS.

In many cases electroforming a requires higher voltage than the switching process. This process can

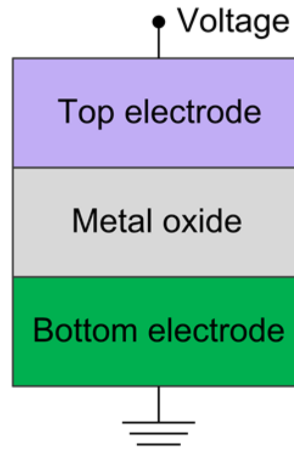

Figure 1.14: Schematic of ReRAM cell.

be seen as the controlled breakdown of an insulator.

During the electroforming, a conductive filament is produced inside the insulating film (in case of filamentary ReRAM). The formation of the filament is demonstrated by in situ transmission electron microscopy measurements[26] and conductive atomic force microscopy measurements[27]. Also it is found that the conductivity of the LRS does not scale with the electrode size. Typically, there is only one dominant filament and the switching affects only that single filament[28]. During the switching process, only a small fraction of the filament is changing.

The voltage level required for the electroforming process depends on the thickness of the film[29]. This is expected, as the electroforming is (in principle) a soft breakdown of the film. Breakdown occurs when the threshold electric field is achieved, thus it is thickness dependent.

The further growth of the formed filament is thermally driven due to the high local Joule heating. External current compliance is generally required during electroforming to prevent an extensive local heating and permanent damage to the film. The following switching process is not thickness dependent, as the switching, after the initial electroforming, occurs at localised point on the filament.

Another, less common type is interface-type switching. In this case the conductivity of LRS depends on electrode size. The switching is driven by the formation of a tunnel barrier or Schottky barrier across the whole interface between the contact and the switching layer[30, 31].

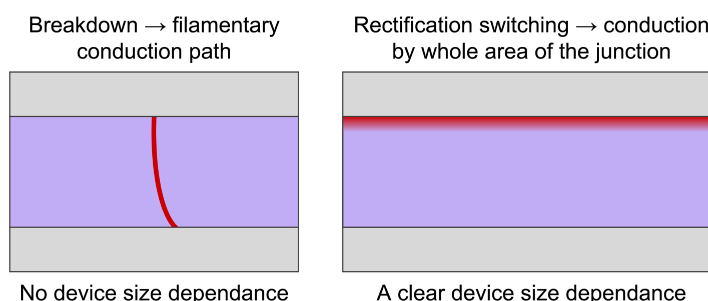

Figure 1.15: LHS: Filamentary resistive switching. RHS: Interface resistive switching.

Transitional metals such as TiO_x[32], VO_x[33], NiO_x[34], CuOx[35], ZnO_x[36], ZrO_x[37], HfO_x[38], TaO_x[39], WO_x[40], and $SrTiO_x$[41] are the most used switching materials, though other materials including organic materials have also been used. Even SiO_x[42, 43], which is best known for its good insulating properties, has shown good resistive switching properties.

We can differentiate two types of switching regarding if the electrodes play an active role in the switching events. If the electrodes play an active role in the switching process it is the extrinsic type of switching and the ECM is the typical example of this type. If the electrodes do not play an active role in switching (but they can still affect the switching properties) is reffered to as intrinsic switching. VCM and TCM are two typical examples of this type.

In oxide ReRAM devices, there are two commonly reported switching modes: unipolar and bipolar mode. This is shown in Figure 1.16.

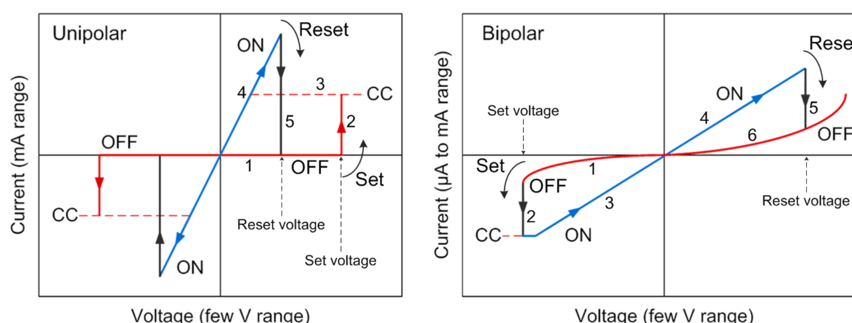

Figure 1.16: LHS: Unipolar switching mode. RHS: Bipolar switching mode. Current compliance (CC) is used to prevent the hard breakdown (permanent damage) during the set process.

In case of the unipolar mode, the resistance change occurs irrespective of bias polarity, whereas in the case of bipolar mode it is necessary to use opposite polarities. Certain materials have shown the coexistence of both types of switching[44].

The transition from one state to the other is commonly called the set process if the transition is from the OFF to the ON state, and the reset process if the transition is from the ON to the OFF state.

In the case of unipolar switching a current compliance is used during the set process to prevent the hard breakdown. During the reset process the current compliance is removed (to allow the high current to pass at the lower voltages) and a transition to the HRS is achieved. In some cases the reset voltage is higher than the set voltage (though the opposite is more typically observed) and therefore the current compliance is not required.

Figure 1.16(LHS) shows the sequence of device operation in the unipolar mode. For this mode only one polarity is used. I will focus on positive polarity in this description. Starting from the higher resistance state (OFF state), there is a very low current passing through the device (part of the I-V curve noted with the part "1"). After the threshold voltage (set voltage) is reached there is a sudden current jump to the level of current compliance (part "2"). Sweeping the device back to the origin does not switch the device back to the higher resistance state, but device stays in the lower resistance state (ON state) (I-V curve would follow the part "3" which is the level of current compliance). To reset device back to the higher resistance state the current compliance is removed and device is swept from the origin. After the certain current level is reached (at the reset voltage) device switches back to the higher resistance state (part "4" and "5"). This sequence can be repeated many times and device can be switched from one to the other resistance state.

In the case of bipolar switching the set and the reset are obtained in the opposite polarities and a certain device asymmetry is required.This asymmetry can occur during the electroforming process, or it can be achieved by using different materials for the two electrodes. The current compliance is still typically required during the set process (to prevent the hard breakdown).

Again starting from the higher resistance state, device is swept in the negative bias (part "1" in Figure 1.16(RHS)). Similarly to unipolar switching, after certain threshold voltage is achieved, device switches to the lower resistance state and current jumps to the current compliance level (part "2"). Sweeping the device back to origin does not switch the device back to the higher resistance state (part "3"). In the case of bipolar switching to switch the device back to the higher resistance state the opposite polarity (compared to the polarity of the set process) is used. In this case that is the positive polarity. Device is swept to the set voltage in the positive polarity without the current compliance (part "4"). After the reset voltage is achieved, device switches back to the higher resistance state (parts "5" and "6"). To switch the device back to the lower resistance state the

sequence is the same (begining with the parts "1").

In practical applications of the resistive switching memories, transitions are induced with the short pulses rather than with the voltage sweeps.

Two different processes dominate these two types of switching. If the electric field is the dominant factor in the switching process, it can be described as an electrochemical. Electric field causes the drift of the charge species (oxygen ions and vacancies) inside the switching film. In this case bipolar switching is observed. If the Joule heating (generated by the passing current) is the dominant factor of the switching process, it can be described as a thermochemical. Oxygen ions and oxygen vacancies move along a temperature gradient. In this case unipolar switching is observed. Three voltage pulses are used for the device operation in either case. A read voltage pulse is used to read the device state; a set voltage pulse is used for the transition from HRS to LRS; a reset voltage pulse is used for the transition form LRS to HRS.

Requirements for ReRAM as Non-volatile Memory

There are a few general requirements for all non-volatile memories, and this includes ReRAM as well. The requirements are resistance ratio between the states, endurance and retention.

- The set voltage should be in the range of a few hundred mV to a few V. It should not be too low (less than 100mV) in order to be compatible with scaled CMOS. The upper limit of set voltage gives ReRAM an advantage over other non-volatile memory types. Flash normally operates at much higher voltages (>10V) compared to typical ReRAM (a few V).
- The set process should be fast enough to outperform flash. The typical set pulse is shorter than 30ns. This is on a par with DRAM and much faster than flash (typically around 10µs). Very fast set processes of around 300ps have been demonstrated[45].
- Ideally, reading voltage should be 10 times lower than the writing voltage to prevent an unintentional change of the memory state and to still be appropriate for the circuit design (detection by sense amplifiers). Reading current should not be less than 100nA for easy and fast detection. Reading time should be at least fast as the set time, and preferably faster.
- The resistance ratio between the states should be at least 10. This allows a cost-effective design of the reading amplifiers.

- Endurance should be on a par with flash – typically at least 10^3, although higher endurance is normally required. Higher endurance on a par with DRAM or SRAM (10^{12}) has already been reported[39].
- Data retention of 10 years under a thermal stress of 85°C is a typical requirement for non-volatile memories.
- High density is needed to be competitive with the aggressive scaling of the present memory technologies.

A crosspoint (crossbar) structure provides the best scaling prospect. Passive crossbar memories consist of two layers of the perpendicular electrodes with the active switching material between the crosspoint. This is shown in Figure 1.17.

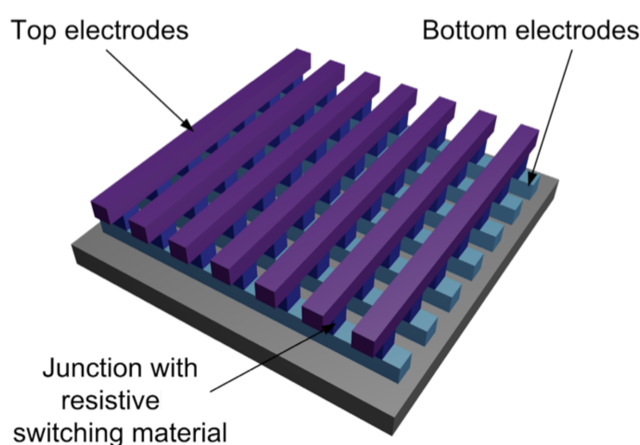

Figure 1.17: Crossbar array structure. Two layers of perpendicular lines with the switching material inbetween.

Any memory unit can be addressed by applying an appropriate voltage across the two perpendicular arrays.

In the ON state the selected unit gives enough current to be read/sensed by the current amplifier. To obtain a non-destructive reading two lines are biased with the total voltage difference smaller than the threshold set voltage. Lines are typically biased with opposite polarities and the same voltage level. During the reading perpendicular lines are biased with $-V_{red}/2$ and $+V_{red}/2$, where V_{red} is the reading voltage ($0>V_{red}>V_{set}$, where V_{set} is the set voltage).

A similar process is used in case of writing and only selected cell is affected while common-line cells remain in the same state.

In conventional active memories a selector device is integrated together with the memory cell. The selector is usually a transistor and the scaling of the whole structure is limited by the transistor scaling. In contrast to transistor scaling, where two critical dimensions are present and an aggressive scaling is not possible, ReRAM systems provide scaling in just one dimension (the thickness of the switching layer). The whole structure can be stacked in the third dimension providing much higher density if a high temperature process is not required during the fabrication.

Nonlinearity of the ReRAM unit is crucial for implementation of passive crossbar memories; it decreases sneak currents passing through the semi-selected devices. Nonlinearity is typically achieved by putting a diode in the series with the memory unit.

Apart from these solutions (integration of an external diode or engineering some nonlinearity of the active material), an interesting approach not requiring nonlinearity in the ON state is proposed. This approach is called complementary resistive switching (CRS) and the main idea is to use two bipolar resistive units connected in anti-series[46].

Four different states are possible: both units are in the HRS; both units are in the LRS; the first unit is in the LRS and the second unit is in the HRS; the first unit is in the LRS and the second unit is in the HRS. To prevent a leakage current one of the memory units is always in the HRS. The drawback of this approach is higher complexity and the destructive manner of the reading process.

Logical state "0" is represented by the state in which the first unit is in the HRS and the second unit is in the LRS. Logical state "1" is represented by the state in which the first unit is in the LRS and the second unit is in the HRS.

In order to write the state "0", positive voltage bias larger than $V_{th,2}$ is applied ($V_{th,2} = |V_{th,set}| + |V_{th,reset}|$, where $V_{th,set}$ is the set volage and $V_{th,reset}$ is the reset voltage of the individual unit). If the CRS is already in the state "0", nothing will happen. The first unit is in the HRS and the negative bias (note that the first unit is inverted with respect to the second unit) does not change its state, while the second unit is in the LRS, and thus the positive bias does not change its state either. If the CRS is in the state "1", positive voltage larger than $V_{th,1}$ will switch the second unit to the LRS and the CRS will temporarily be in an all-ON state (both units are in the LRS). When the voltage increases above $V_{th,2}$, the first unit switches to the HRS so the CRS changes its state

to the state “0”.

Writing the state “1” is equivalent to writing the state “0” by applying the negative voltage bias larger than $V_{th,4}$ (where $|V_{th,4}| = |V_{th,2}|$).

In order to read the state, positive bias bias larger than $V_{th,1}$ (where $V_{th,1} = V_{th,set}$) and smaller than $V_{th,2}$ is applied. If CRS is in the state “0”, a small current at the output is detected as the first unit is in the HRS and the states do not change. If CRS is in the state “1”, positive voltage bias larger than $V_{th,1}$ and smaller than $V_{th,2}$ switches the second unit to the LRS and a large current is detected (as both units are in the LRS). This indicates the logical state “1” for CRS. It is necessary to recover CRS to the state “1” because the state of the second unit is changed. Negative voltage bias larger than $V_{th,4}$ is applied and CRS is recovered to the state “1”. This approach is shown in Figure 1.18.

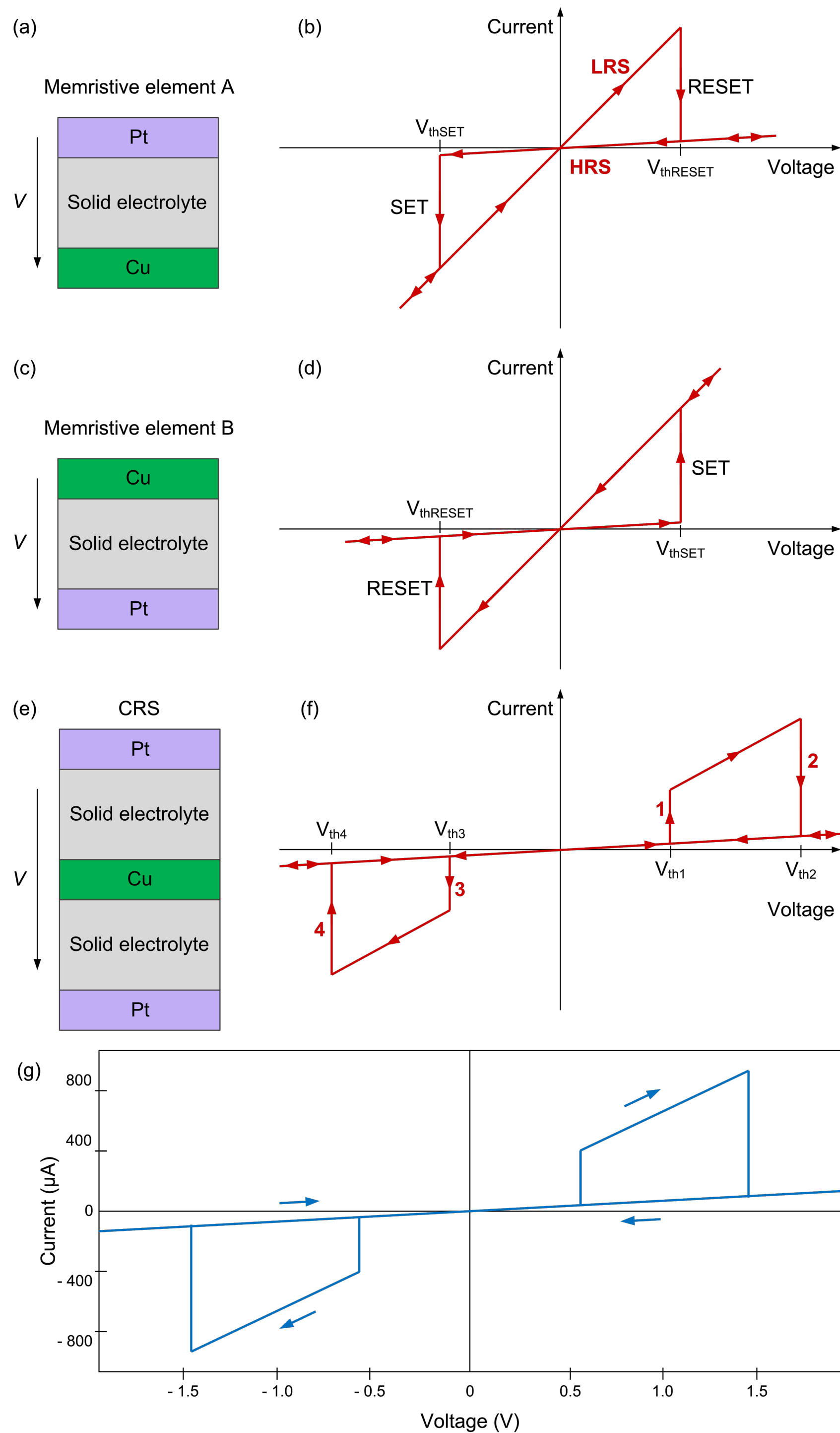

Figure 1.18: (a) Bipolar memristive element A with a Pt/solid electrolyte/Cu stack. (b) I-V characteristic of memristive element A. The resistance can be toggled between the LRS and HRS by exceeding $V_{th,SET}$ and $V_{th,RESET}$. (c) Bipolar memristive element B with a Cu/solid electrolyte/Pt stack. (d) I-V characteristic of memristive element B. (e) CRS resulting from the combination of memristive element A and B. (f) I-V characteristic of a CRS. (g) Measured I-V curve. After[46].

ReRAM has the potential to become not just the alternative to flash or DRAM but to become a universal memory. Universal memory would need to combine all the good characteristics of all types of memories including non-volatility, high endurance (higher than 10^{12} cycles), sub-nanosecond writing speed, long retention time (>10 years) and nanometre dimensions. Although most of these characteristics are already reported, high-density memory requires high yield and small variation in switching behaviour. The required level is still not achieved.

Electrochemical Metallization Cell Memory

Electrochemical metallization cell (ECM) memories are also called Conductive Bridge RRAMs.

The main part of the memory cell is electrochemically active metal electrode, usually Ag or Cu. The other electrode is chemically inert – such as Pt, Ir, TiN, Au, W. Thin insulating layer is sandwiched between two electrodes. This layer is typically either a solid electrolyte such as Ag_2S, Cu_2S or an insulator such as WO_3, GeS, SiO_2. The thickness of the insulating layer is typically a few tens of nm, but can range from 5nm to several hundreds of nm. The basic principle of operation is shown in Figure 1.19.

In the pristine state the cell shows high resistivity. There is no metallic material from the active electrode present in the insulating layer. During the forming and the SET process a positive voltage is applied to the active electrode. After the application of an electrical bias the process of anodic oxidation occurs. Metallic cations are formed at the site of the active electrode and they begin to propagate into the insulating layer under the electric field. When cations reach the passive electrode the chemical reduction takes place and the filament growth starts. The filament grows from the passive electrode towards the active electrode. Once the filament bridges the two electrodes the low resistive state is reached as much higher current can pass through the conductive filament. Further growth of the filament is controlled by the external current compliance.

Typically, only one complete filament is produced as the voltage immediately drops when the first contact is made and current compliance is reached. Nevertheless, the formation of several filament seeds is possible during the formation process. The presence of grain boundaries and accumulated defects within the insulating layer serves to facilitate the formation of a filament. The radial growth of the filament is controlled by the current compliance and different resistivity levels can be achieved. This means that multi-level operation is possible by controlling the filament growth.

Many different processes can occur during the reset. Once the electric conduction is sufficiently high a large current passes through the narrow filament producing local Joule heating. This Joule heating will be most pronounced at the narrowest point – presumably near the active electrode. High Joule heating leads to the rupture of the filament through temperature assisted atom diffusion. Once the rupture occurs and a gap is produced the electrochemical process dominates the rest of the reset process.

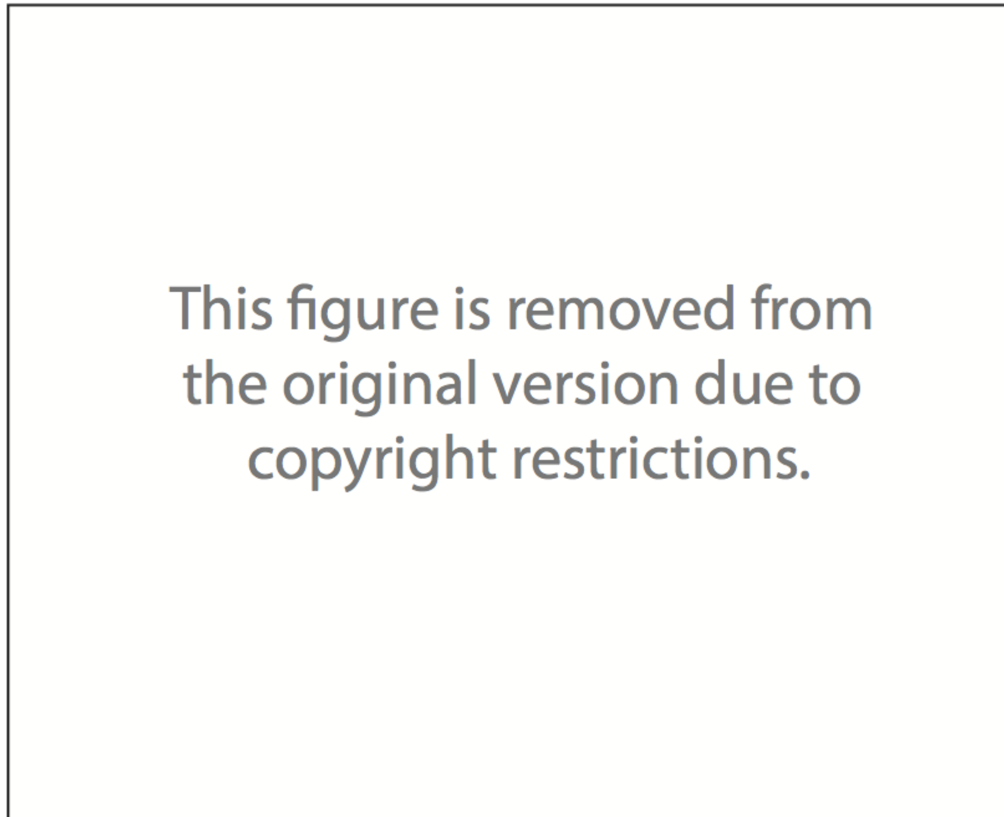

Figure 1.19: Typical I-V curve and basic principle of operation of ECM. (a) OFF-state; (b), (c) Set process; (d) ON-state; (e) Reset process. Taken from [5].

Valence Change Memory

A VCM memory cell consists of an active electrode, a mixed ionic-electronic layer and an Ohmic passive electrode. In the case of VCM the active electrode does not provide a diffusion of metallic ions to form the filament, but rather provides an active site where the switching takes place.

Material with low oxygen affinity is used for the active electrode (e.g. TiN, Ir, Pt). Transition metal oxides like TaO_x, TiO_x, WO_x, HfO_x are the most popular choice for the switching layer. These transition oxides are typically non-stoichiometric and oxygen deficient. Fully oxidised layers can also

be used and an electroforming process is used to produce an oxygen deficient conduction channel. Low work function materials with high oxygen affinity are used for the Ohmic electrode.

The basic principle of operation, together with a characteristic I-V curve, is shown in Figure 1.20.

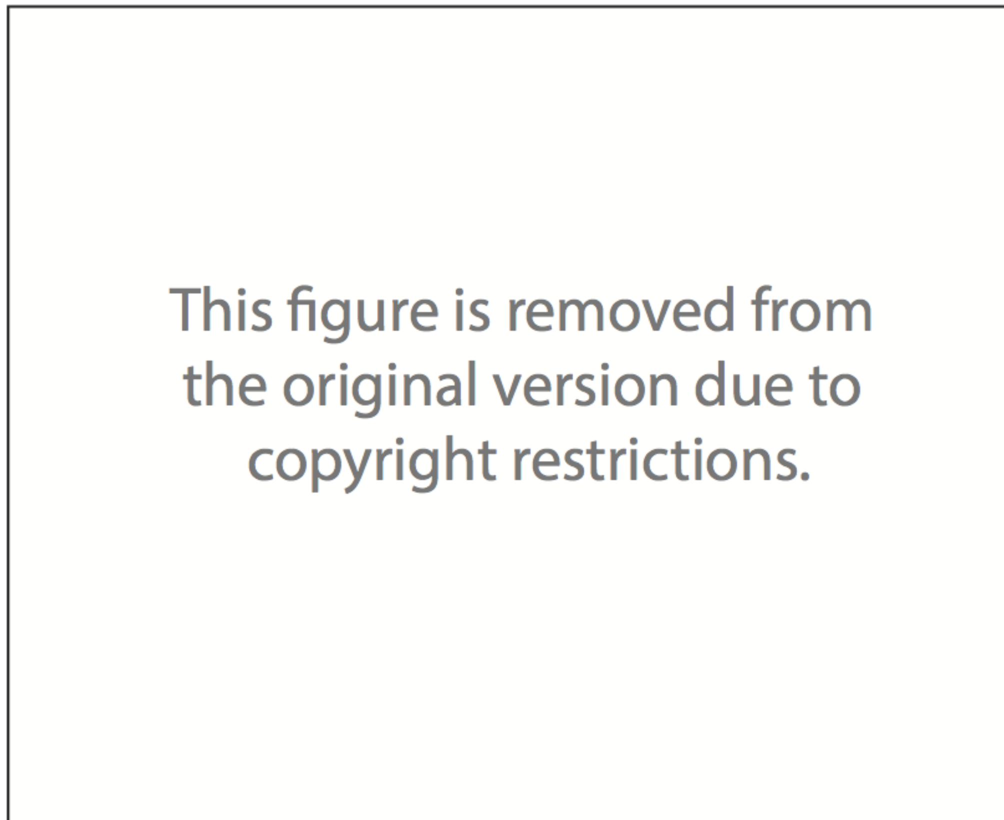

Figure 1.20: Typical I-V curve and basic principle of operation of VCM. (a) OFF-state; (b) Set process; (c) ON-state; (d) Reset process. Taken from [5].

The electroforming step is crucial for most VCM systems. The forming voltage is typically significantly higher than the set voltage. Electroforming is a local reduction of the insulating layer. Both voltage polarities can be used. During the forming a high local temperature is generated due to the significant currents and large voltages. This temperature can cause local morphological changes or a phase transition of the material (e.g. formation of oxygen deficient Magnéli phase in case of TiO_2).

After electroforming, a conductive n-type channel is produced. The following switching process takes place near the active electrode (the electrode with the lower oxygen affinity). A tunnelling barrier is typically present between the conductive filament and the active electrode. By varying the barrier height it is possible to achieve different resistance states. During the set process oxygen ions are removed from this region - consequently tunnelling and thermionic current is increased and

barrier height is reduced. During the reset process oxygen ions are moved back into the region. This increased the barrier height and the overall resistance of the VCM cell.

Thermochemical Memory Cells

The thermochemical memory cell (TCM) is dominated by the thermochemical process in contrast to ECM and VCM where the electrochemical process is the dominant one. Another difference to ECM and VCM is that TCM is operated in the unipolar mode – the polarity of both set and reset is the same. A temperature gradient affects the local stoichiometry through a local redox processes.

Many different materials are used for the switching layer, most often NiO, Al_2O_3, CoO, CuO, TiO_x, Fe_2O_3, ZrO_x. Even SiO_2 is used as the switching material in TCM. The same material is used for both electrodes as there is no need for asymmetry as in case of ECM and VCM. The reason for this is that oxygen ions and oxygen vacancies diffuse along the temperature gradient and the position of the switching point can be at the arbitrary place along the filament – not necessary near the interface. A typical I-V curve with physical principle of switching is shown in Figure 1.21.

A dramatic current increase is observed due to thermionic breakdown. An increase in the local temperature leads to permanent material change and formation of a conductive filament.

Current compliance is required to prevent the hard breakdown and nonreversible transition to LRS. Current compliance is removed in the reset process. This allows a higher current to pass at much lower voltages, leading to a partial rupture of the filament and recovery of the HRS.The conductivity of the HRS is usually larger than the conductivity of the pristine device as the filament is not completely dissolved and current can still flow through the part of the filament.

The set process is similar to the electroforming process, but it typically occurs at lower voltages. Current compliance is needed to prevent the hard breakdown just as in case of the electroforming. During the forming/set process a reduction of oxidation states occurs. Most transition metal oxides show much lower resistivity in the lower oxidation states.

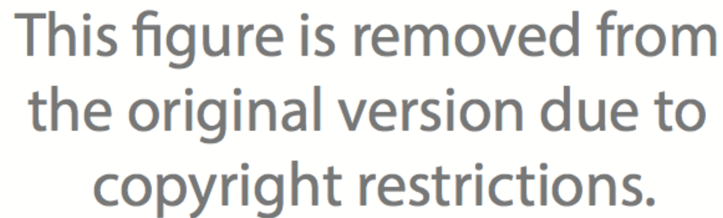

Figure 1.21: Typical I-V curve and basic principle of operation of TCM. (a) Electroforming process; (b) ON-state; (c) Reset process; (d) OFF-state; (e) Set process. Taken from [5].

Introduction to Resistive Switching in Silicon Oxide

Silicon dioxide (silica) is one of the most abundant materials on the planet Earth – silicon and oxygen together account for 75% of earth's crust. In nature it can be found in crystalline or amorphous form – sandy areas are a good source of silicon dioxide (e.g. beaches, deserts). Also it can be found as quartz in the Earth's crust.

Silicon dioxide has a very wide spectrum of applications, especially in electronics and photonics, such as in manufacturing of semiconductor devices. In its quartz form, silicon dioxide has piezoelectric properties and it is used in the TV and radio industry for reception and transmission of signals. Silica is used for production of all kinds of glass. It is used in the food industry as an additive to powdered food as a flow agent.

The native oxide of silicon is an extremely stable compound, stable both in water and at elevated

temperatures. It is an excellent electrical insulator, and for many years it has been used as insulating layer in MOSFETs. Silicon dioxide forms a very good electrical interface with its silicon substrate. Silicon dioxide is one of the most widely used materials in the CMOS industry.

Reports of resistive switching in defective SiO_2 date back to the 1960s[47], and the so-called soft breakdown of silicon oxides has been of interest for some time as a model to understand oxide failure in silicon microelectronics. However, the underlying physics of resistive switching in silicon oxide is relatively poorly understood.

Although one of the earliest reports of resistive switching in thin films is about resistive switching in thermally grown silicon dioxide, silicon dioxide is not the most popular choice for resistive switching material. Far more popular are transition metal oxides such as HfO_x, TiO_x, Ta_2O_5, NiO.

Silicon dioxide (or silicon oxide) was regarded as a passive, insulating element and for applications in RRAMs it is used either in ECM systems or doped with metallic species during the fabrication process. When used in ECM systems it supports the diffusion of metallic ions from the active electrode and does not play an active role in switching. Only recently has the idea of using silicon oxide as an active material for resistive switching been reinvestigated.

Typically, a Cu active electrode is used in silicon oxide based ECM. The switching is attributed by the formation and dissolution of a Cu filament within the silicon dioxide matrix[48]. An Ag active electrode is used in the case of amorphous silicon[49]. Similarly, amorphous silicon supports the diffusion of Ag ions to form the conductive filament. The silicon dioxide matrix may itself be doped with metallic ions, the diffusion of which can modify the resistivity of the matrix[50].

However, in CMOS processing diffusion of metallic ions is undesirable, as this could potentially endanger the operation of surrounding electronics. The intrinsic resistive switching of silicon oxide is therefore a more appealing mechanism. Such a system was first demonstrated by Yao et al.[51] in 2010, who reported a switchable silicon conductive path formed on the vertical surface of a silicon-rich silica pillar. No metals were present in the device, and the authors reported intrinsically unipolar switching. The switching process was attributed to the voltage-driven formation of silicon nano-crystals (NCs) in the surface of the silicon oxide. These nano-crystals will eventually produce a conductive filament that is able to support current flow. Very promising switching properties were reported: high ON/OFF ratios ($>10^5$), fast switching (sub-100 ns), and good endurance (10^4 write-erase cycles). Figure 1.22[51] shows device schematics and characteristic I-V curves together with programming pulsing.

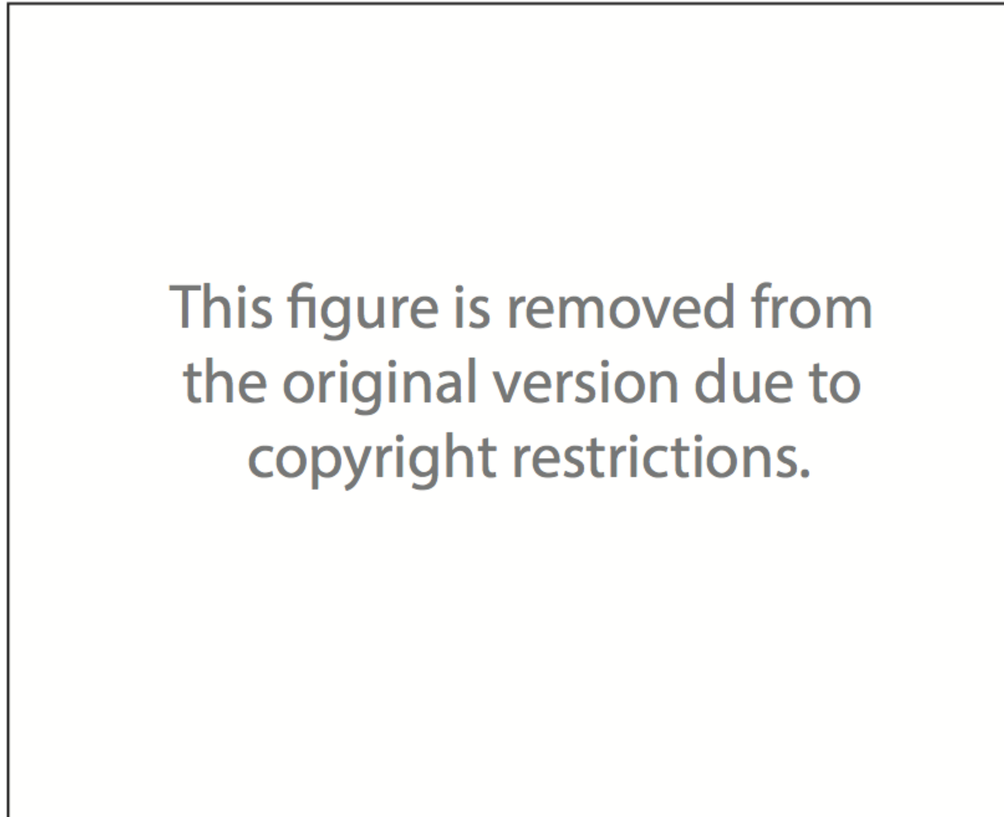

Figure 1.22: Device structures and electrical characterisations. (a) Schematics of the poly-Si/SiO_x/poly-Si devices (50µm diameter) and the electrical characterisation setup. (b) An optical top-view image of the device arrays (top panel) and a scanning electron microscope (SEM) image of the cross section (bottom panel). (c) Characteristic I-Vs in a formed device: (blue curve) starting from an OFF state, the conductance of the device suddenly increases at +3.5V and drops at +8V. During a subsequent reverse voltage sweep (gray curve), the conductance goes back to a higher value at +8V, producing a current hysteresis. Corresponding "read", "set", and "reset" regions are defined based on these conductance changes. The vertical dashed line indicates the ON/OFF ratio at 1V. Inset: device resistance changes after +6 and +13 V programming voltage pulses. (d) (Top panel) A series of voltage pulses of +13, +1 (5 times), and +6 V serve as reset, read (5 times), and set operations, respectively. (Bottom panel) currents corresponding to each voltage pulse in the top panel. Taken from [51].

Interestingly, I-V curves are different to typical I-V curves reported either in the case of bipolar or unipolar switching. Switching is intrinsically unipolar but current compliance is not used and the reset voltage is higher than the set voltage, in contrast to more conventional unipolar switching. These early results are very promising as they fulfil many of the requirements for non-volatile memories.

However, the operation of such a device is possible only under vacuum due to the oxidation of silicon conductive pathways on the device surface under ambient conditions. The authors reported no evidence of high contrast controllable resistive switching in continuous films of silicon oxide: instead, silicon nano-filaments at vertical surfaces of mesa structures were proposed as the switching elements and any exposure to oxygen leads to oxidation and rupture of conductive filament.

In this thesis, resistive switching in the bulk silicon oxide is examined. This switching is not restricted to vacuum operation or to the surface of the active material.

Chapter 2

Experimental Techniques and Sample Preparation

This chapter provides an overview of the experimental techniques and the sample preparation procedure.

Electrical Characterisation

Electrical Probe Station

The equipment that we used for the I-V characterisation is the Keithley 4200-SCS Semiconductor Characterisation System. It consists of a characterisation probe station, an analyser module and the appropriate software (Keithley Interactive Test Environment). The characterisation probe station has four probes and a highly conductive sample platten. Two probes were used to generate I-V and I-t curves. Electrical characterisation was done in ambient conditions.

Impedance Spectroscopy

Impedance spectroscopy is a technique used to characterise electrical properties of materials. The sample is biased with an alternating voltage source, and the sample impedance is measured as a function of source frequency.

Impedance, in general terms, is defined as a complex ratio of the applied alternating (AC) voltage and the passing current. Impedance (Z) as a function of input frequency is plotted in the complex plane, called Cole-Cole plot. These are impedance plane plots (see Figures 2.1(b) and (e)). Admittance (inverse of impedance) plane plots are commonly used together with Cole-Cole plots (see Figures 2.1(c) and (f)). These plots can be related to the microstructure properties of the measured material. The overall resistivity can be modelled as an equivalent circuit consisting of resistors and capacitors connected in series and parallel (see Figures 2.1(a) and (d)). When plotted on a Cole-Cole plot, the curve shape of impedance can consist of several semi-circles indicating different equivalent circuits. By fitting the curve shape with a particular model it is possible to obtain the values for resistors and capacitors in the circuit. This is used to determine if there is only one or several conduction paths. This can be further related to bulk or grain boundary properties of the material.

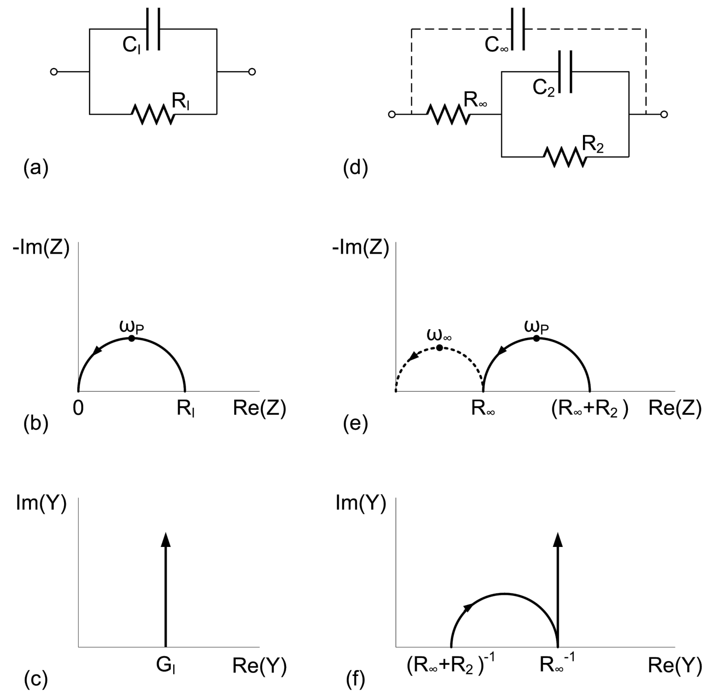

Figure 2.1: Parts (a) and (d) show two common RC circuits. Parts (b) and (e) show their impedance plane plots and (c) and (f) their admittance plane plots. Arrows indicate the direction of increasing frequency. After [52].

Structural Characterisation

Scanning Tunnelling Microscopy

In 1986 the Nobel Prize in physics was awarded to Gerd Binning and Heinrich Roher for their invention of scanning tunnelling microscopy (STM) in 1981. Since its innovation, STM was an important tool for characterisation of surface structures. The breakthrough of this technique is its ability to study surface properties with atomic resolution.

The basic concept of the STM is quantum tunnelling of electrons between a sharp microscope tip and the investigated surface (the tip that we used is a platinum-rhodium or tungsten needle). The scanned surface must be conductive. A voltage is applied between the surface and the tip. When the needle tip is sufficiently near to the surface the electrons are able to tunnel through the vacuum (ambient - air in our case) barrier in between (typical distances are around 1nm). This current is called tunnelling current and it is based on quantum-mechanical tunnel effect. When the tunnelling current starts to flow and a tunnelling contact is established, the tip is moved across the surface by piezoelectric scanning unit. The scanning unit is typically capable to scan the areas in range of a few nm up to several µm. The scan provides a microscopic image of the spatial variation of the tunnel current.

STM can reveal informations about the atomic-scale structures that must correspond to electrical states. These are the states from which electrons can tunnel into the tip or into which the electrons can tunnel from the tip. The electrons can tunnel across the potential barrier (between the tip and the sample) even when the electron's energy is lower than the barrier height. The corresponding tunnelling probability is proportional to e^{-kz} , where z is the tip-sample distance and k is a decay constant. Furthermore $k = [2m(W_b - W)/\hbar]^{0.5}$, where m is electron mass, W_b is the potential barrier and W is the energy of the tunnelling electron.

When applied bias is small (a few V), only electronic states very near the Fermi level (a few eV around Fermi level) are excited and tunnel across the barrier. Tunnelling also requires that there is a corresponding empty level of the same energy in which the electron can tunnel to. This implies that the current can be related to the density of available or filled states in the sample. Tunnelling current depends on the number of electrons between the Fermi level and a few eV in the sample, and the number of corresponding free states at the tip (in the case when the electrons tunnel from the sample into the tip). The tunnelling current increases with the number of available states.

Electrons in the tip tunnel into empty states in the sample when the applied bias is positive, while for the negative bias, electrons tunnel out of occupied states in the sample into the tip. This is shown in Figure 2.2.

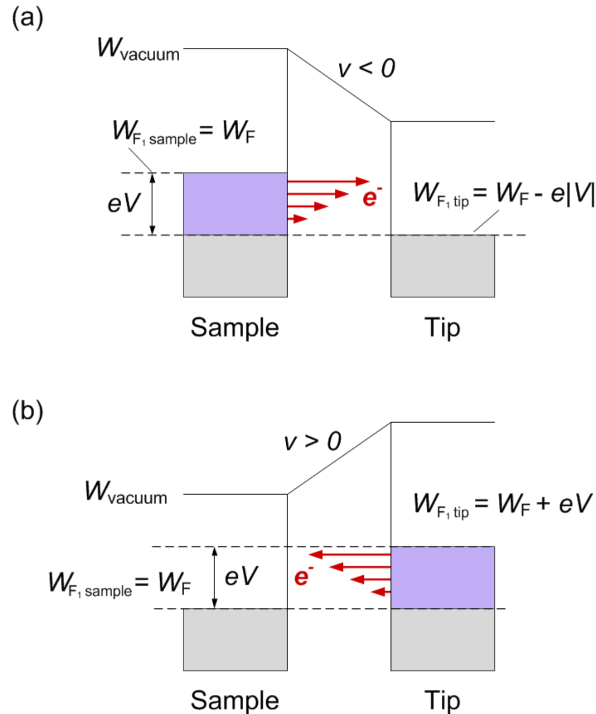

Figure 2.2: Electrons whose energy states range between W_F and $W_F + eV$ contribute to the current. The figure shows the cases of (a) a negative voltage applied to the sample and (b) a positive voltage applied to the sample. After [5].

There are two main modes of operation. In the constant-current mode the current is held constant, and as the tip moves across the surface it must move up and down, following the surface topography. In this way atomic resolution can be achieved. A precise vertical motion of the tip is achieved with piezoelectric cylinders. In the constant-height mode the tip height is held constant and the current is monitored. As the tunnelling current is exponentially dependent on the distance this mode gives us very high sensitivity and thus single atoms can be resolved.

The STM equipment consists of a very sharp tip, piezoelectric motors and the feedback loop. In the first mode of operation, when the tip is driven across the surface, the feedback loop allows the tip to move parallel to the surface. In the second mode the feedback loop is slowed down, thus the tip is held at a constant height. This mode is only applicable if the monitoring surface is flat, otherwise there is the possibility of tip and surface collision.

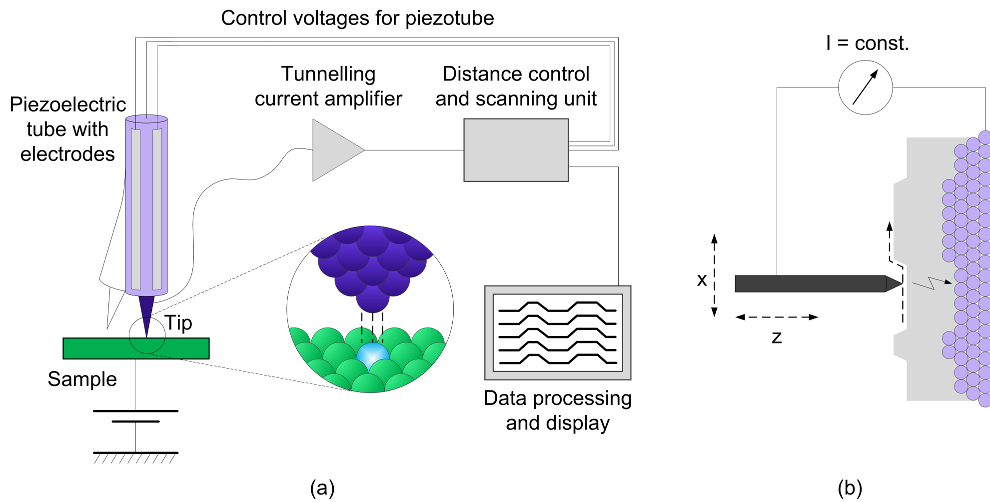

Figure 2.3: (a) Schematic view of an STM. (b) Feedback between the tip and monitored current allows the tip to move in the z-direction according to surface topology.

Atomic Force Microscopy

In 1986, Calvin Quate and Christophe Gerber invented atomic force microscopy (AFM) several years after the STM was invented. Similar to STM, the AFM measures the interactions between the moving tip and the surface. The interaction is measured at very short distances (0.2nm to 10nm).

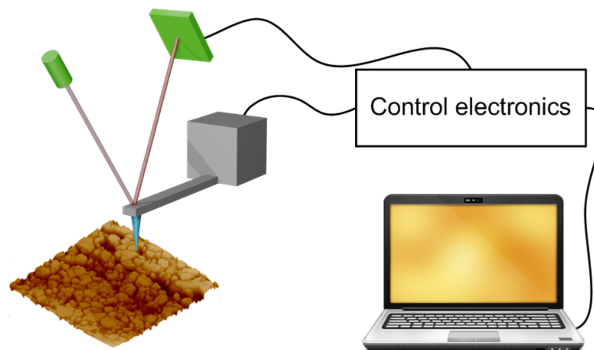

Figure 2.4: AFM feedback loop and control electronics.

The tip is attached to a mechanical cantilever-type spring. Piezoelectric elements enable very precise and controlled scanning. An image is gathered from the map of the detected forces (ranging from 10^{-13} to 10^{-6} N) at each point of the surface. The cantilever is commonly made of silicon or silicon nitride and when the tip is sufficiently close to the surface there is a deflection of the cantilever according to Hook's law. These deflections can be used to control surface-tip distance on an atomic level. A laser beam monitors the cantilever deflection; the reflected laser beam from a cantilever is

detected by a photodiode array, which directly gives the deflection.

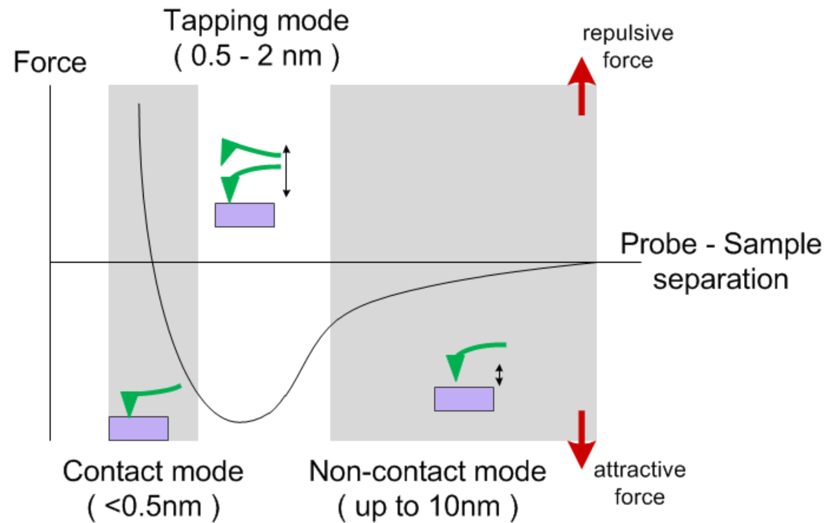

Figure 2.5: Forces between the tip and the surface in function to tip-sample distance.

At the very short distances (typically a few angstroms) there is a very strong repulsive force between the tip and the surface. Origin of this force is exchange interaction that occurs when the electronic orbitals are overlapped at atomic distances. In this case it is said that the tip and surface are in contact.

The dominant interaction at longer distances is Van der Waals interaction/force. Other interactions, more evident at much longer distances, include electrostatic, capillary and magnetic forces.

Van der Waals forces lead to an attractive interaction between the tip and the sample surface. This interaction can be described as the interaction between the time-dependent dipole moments of the atoms in the surface and atoms at the tip. Although the centres of gravity of electronic charge density and charge of nucleus overlap when averaged over time, the separation of centres of gravity spatially fluctuates in every moment. This means that the atoms' dipole moments fluctuate. The dipole moment of the first atom can induce a dipole moment in the neighbouring atom and this dipole moment acts back at the first atom. Fluctuating dipole moments produce the dipole-dipole interaction. Generally this interaction is proportional to z^{-6} for the small distances and z^{-7} for the larger distances.

Similarly to STM, the distance dependence of the Van der Waals interaction is used to achieve a high resolution of AFM. However this dependence is much weaker compared to the exponential dependence of the tunnelling current in the case of STM. This means that the sensitivity of AFM

is lower than in the case of STM.

There are three main AFM modes of imaging. The main difference between these modes is the tip-sample distance during the scanning process.

The first mode is the contact mode. In this mode the cantilever is moving across, feeling the forces from the surface. These forces are repulsive forces dominantly. By keeping the force constant (deflection of the cantilever) it is possible to obtain the image of the sample surface topography. Advantage of this mode is fast scanning. However soft samples are not suitable for this mode as they can be damaged during the scanning.

The second mode is called the tapping mode. In this mode piezoelectric motors are used to cause the cantilever to oscillate at its resonant frequency above the sample. During these oscillations, tip is touching ("tapping") surface of the sample. Oscillating amplitude is kept constant by keeping the constant tip-surface interaction. In this way the surface topography is obtained. This mode is more suitable for softer samples that are easily damaged (i.e. biological samples). Disadvantage of this mode is slower operational mode.

The third mode is called the non-contact mode. In this mode cantilever is kept completely above the surface. Similarly as for the tapping mode, the cantilever oscillates at the resonant frequency. The forces between the tip and the surface change the resonant frequency. A feedback loop can hold the surface-tip distance constant. Advantage of this mode is extremely low forces "felt" by the tip (typically 10^{-12} N), which significantly increases the tip lifetime. Disadvantage of this mode is lower resolution compared to two other modes.

In case of AFM, a sample does not need to be conductive in order to be scanned. AFM is used to study many different types of samples including metals, plastics, glasses, semiconductors and biological samples. However the resolution of AFM is typically lower compared to STM. One of the reasons for this is the fact that the AFM tips are not atomically sharp. There is always an artifact of tip geometry when the scanned features are approaching the dimensions of the used tip (typically less than 10nm).

Conductive atomic force microscopy (C-AFM) is a special type of AFM which uses a conductive tip coating (e.g. Pt or Ir) on a conventional silicon tip. There are 2 output channels; the first gives a deflection of the tip (same as for the AFM) and the second gives a tip current. The current is flowing through the metal-coated tip of the microscope and the conducting sample (typically from a

few pA to a few hundreds of nA). Negative voltage is applied to the sample and electrons flow from the sample to the tip. Electrons tunnel similarly as for the case of STM. As the tip moves across the surface, AFM topography is obtained on the first output channel by vibrating the tip (it is possible to scan in contact mode too), while on the second output channel the current map is acquired. In this way it is possible to study the correlation between spatial features and their conductivity. C-AFM also enables spectroscopic measurements of the sample. It is possible to obtain I-V curves at any studied position on the sample. The main advantage of C-AFM over standard electrical measurement techniques is the high spatial resolution (i.e. C-AFM measurements on polycrystalline thin films can identify differences in the conductivity between grain boundaries and the interior of the grains). For the less conductive samples C-AFM is more suitable compared to STM. Scanning can be initiated even when the current does not flow between the tip and sample.

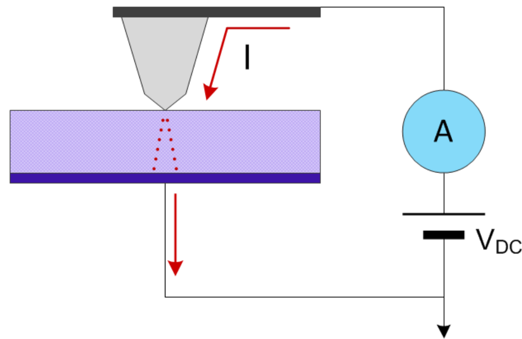

Figure 2.6: Schematic of conductive atomic force microscope (C-AFM)

Transmission Electron Microscopy

Transmission Electron Microscopy (TEM) is another microscopy technique widely used in nanotechnology. A beam of electrons interacts with an exposed sample. The sample needs to be thin enough for electrons to pass through it (typically 50nm or less).

The transmitted beam of interacting electrons is collected and this yields the image of the specimen. Due to the small de Broglie wavelength of an electron TEM can provide much greater resolution compared to light microscopes. State of the art TEMs are able to resolve atomic scale. A TEM operates similarly to a light microscope, but instead of visible light it uses electrons. Schematic of TEM microscope is shown in Figure 2.7.

TEM can be divided into three components: the illumination system, the lens system and the

imagining system. The illumination system consists of the electron gun (which produces the beam of accelerated electrons) and the condenser lenses. Electrons are accelerated to energies in the range of 20keV to 1MeV. Condenser lenses focus the beam of electrons with a selected diameter. The lens system consists of several electron lenses. Electron lenses focus the parallel rays of electrons at the constant focal length. The typical spot size is smaller than $3\mu m$ This is the part where the beam interacts with the sample and where the creation of the images and the diffraction patters take place. The imaging system uses additional lenses to magnify the images. These are focused on the viewing screen (CCD or TV cameras).

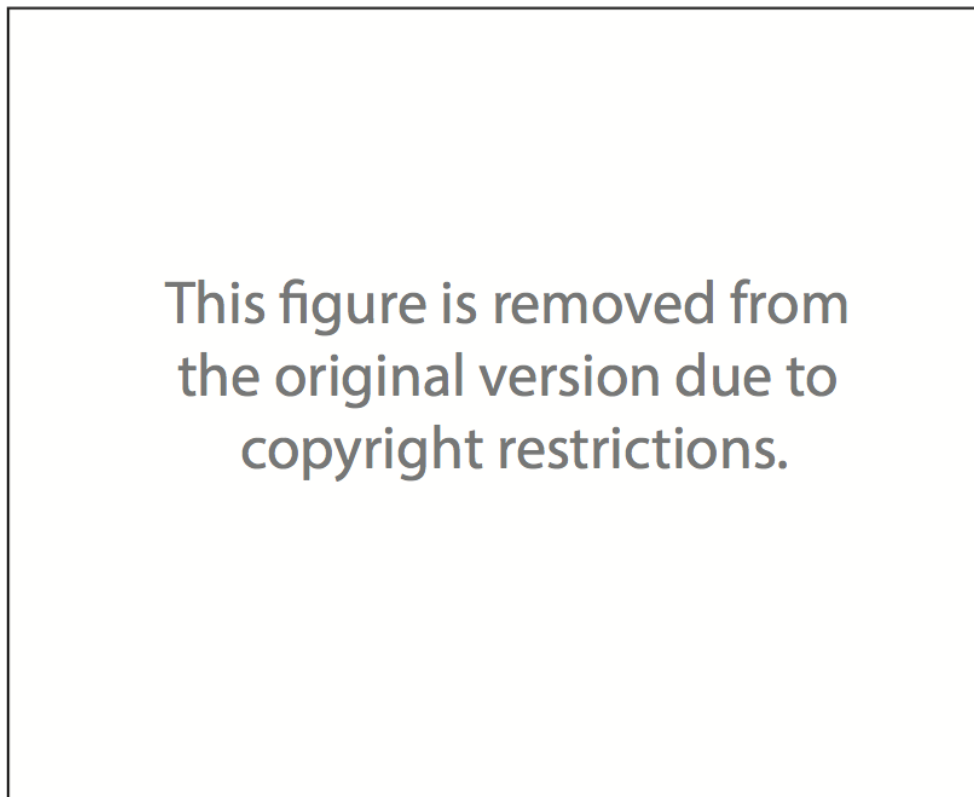

Figure 2.7: Schematic of TEM system. Taken from [53].

There are several different TEM operation modes. The most common mode is the bright field imaging mode. An image is generated in such a way that the regions of higher atomic density appear dark, while less dense regions appear brighter. In this mode the information is gathered directly by absorption of electrons in the sample. In the case of the dark field imaging mode the unscattered beam is excluded from the image. This produces the images in which the specimen is usually bright while the background is dark (as the beam is not scattered in this region).

Another operation mode is the diffraction mode. In this mode the objective aperture is placed in the back focal plane. The specific reflection can be selected, and only the electrons with the selected

reflection are projected. In this way only the parts of the sample that cause the specific reflection will be shown in the image. This also produces the diffraction patters, which could yield important information about the structure of the specimen. If the specimen is the single crystal particle a regular area of spots is produced, while if the specimen consists of randomly oriented crystalline particles a ring pattern is formed.

Sample Preparation

There have been four generations of devices fabricated up to date.

The first generation was made as a simple MOS capacitor, where the bottom Si substrate serves as a bottom electrode and Ohmic contact. Indium Tin Oxide (ITO) was used for deposition of top electrodes. The second generation of devices was also made in MOS design with the difference that the top electrode was n-type poly Si. The purpose of this is to avoid any possible (regardless how unlikely) presence of metallic species within the switching silica layer. The third generation was made to the identical process as the first generation, only the content of excess silicon in the active silicon oxide layer is varied, as well as the thickness of the active layer. Finally, the fourth generation was made in the more popular MIM design in which TiN is used for both top and bottom electrode.

1st, 2nd and 3rd Generation

Devices contain thin SiO_x layers sandwiched between a p-type silicon wafer and n-type top electrodes (a range of electrode sizes was used – from 1.5 mm $\times$ 4 mm to 125 μm $\times$ 125 μm). We investigated two configurations.

The first and third generation were made with indium tin oxide (ITO) top electrodes, deposited at room temperature. The second generation was made with n-type polycrystalline silicon top electrodes requiring a 950°C anneal to optimize conductivity. The device structure is schematically shown in Figure 2.8.

The active layer is identical in all three generations.

Layers were grown on p-type B-doped Si substrates by magnetron co-sputtering. Two confocal

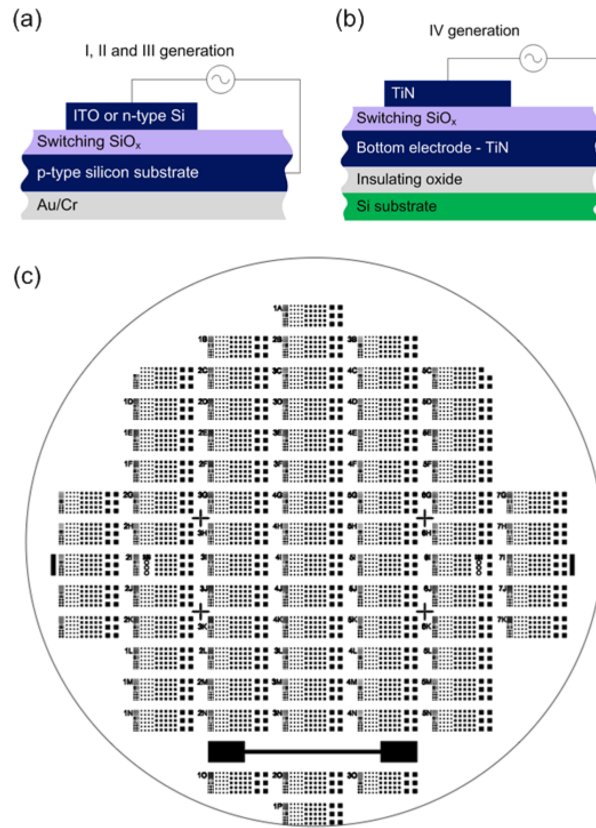

Figure 2.8: (a) Device schematic for the first 3 generations of devices. (b) Device schematic for 4th generation of devices. (c) Photolitographic mask used to pattern top contacts.

cathodes were used: SiO_2 and Si. The whole process was done under a pure Ar plasma. By applying the RF power it is possible to precisely control the thicknesses of layers and also to incorporate Si in wanted concentrations in these thin layers.

The layers were annealed for 1 hour in nitrogen at 900°C or 500°C. Thicknesses were measured using spectroscopic ellipsometry. X-ray photoelectron spectroscopy (XPS) is used to obtain the composition of the samples. We used a Perkin-Elmer PHI-5500 instrument (using Al Ka radiation). Samples were sputtered with an Ar^+ ion beam at 4keV.

For the second generation, 185nm of n-type silicon (phosphorous doped, resistivity 10mWcm) was deposited on top of the SiO_x layer by Low Pressure Chemical Vapour Deposition (LPCVD). After growth, samples were annealed at 950°C for 30 minutes in nitrogen to activate the dopants and achieve a final resistivity of 1mΩcm. After a buffered HF dip to remove surface oxide, 100nm of chrome was sputtered onto the front side, photoresist spun on, and the sample exposed with a chrome-on-glass mask (see Figure 2.8(c)). The mask contained a number of different electrode

sizes – from 1.5mm × 4mm to 125μm × 125μm. The photoresist was developed, and then the exposed chrome etched. Removal of the remaining photoresist was followed by a buffered HF etch and the samples were etched in a TMAH, 25% solution at 60°C, for 40sec. Finally, the chrome was etched, and the samples rinsed and blow-dried in N_2.

For the first and third generation 70nm-thick ITO layers were deposited by sputtering, followed by photolithographic contact definition using the same mask. The ITO layer was etched using hydrochloric acid. Chrome-gold Ohmic contacts were provided on the back sides of all wafers by evaporation (10nm Cr, followed by 100nm Au).

So for the first three generations, an asymmetric structure (n-type top electrodes – n-Si or ITO; p-type substrate) allows us to define two regimes in our MOS structure. Applying a negative potential to the top electrode allows higher currents to flow (negative bias), while a positive potential results in lower currents (positive bias).

4th Generation

For the fourth generation of devices the more common MIM design was used. Boron doped p-type silicon wafers with pre-processed thermal SiO_2 were used. The thickness of the thermal oxide was 4μm. For the bottom electrode the uniform layer of TiN was deposited by sputtering. The thickness was around 100nm. Deposition was done at room temperature. The active layer of silicon rich silicon oxide was deposited similarly to that in the first three generations with magnetron co-sputtering. The thickness was kept constant at around 30nm. Deposition temperature was 250°C. The excess silicon was around 11%. The top layer of TiN was identical to the one used for the bottom electrode. Standard lift-off was used to define the top contacts. One set of samples in this generation was annealed at 600°C and the second one was not annealed.

Chapter 3

Resistive Switching in Active Silicon Oxide

In this chapter the focus is on resistive switching in silicon oxide. The measurement results are divided into two groups. The first group presents the structural properties of the silicon oxide and the second group is focused on the electrical and switching characterisation of the devices.

Electrical Characterisation

The first generation of samples is described in the sample preparation section. These samples are made in MOS configuration. The thickness of the oxide layer for the first generation of devices is 35nm. These samples have Indium-Tin Oxide (ITO) top contacts.

The thickness of the active silicon oxide layer in the second generation of devices is 37nm (for the results shown in this chapter). These samples have n-type poly silicon top contacts.

The thickness of the oxide layer in the third generation is 22nm, and the top contacts are made of ITO. The first three generations all used p-type silicon substrate as the bottom electrode.

The fourth generation of samples is made in MIM design where both contacts are made of 100nm thick TiN. The thickness of the active layer is 30nm.

Current-voltage Characteristics

Bipolar Switching Mode

Before any switching is possible the device needs to be electroformed. This electroforming is done only once with the pristine device. The device is swept to high voltage (up to 30V). During the sweep several current peaks are observed and the whole I-V curve is unstable. Once a high voltage is reached, (typically >20V for devices thicker than 25nm), the device switches to a low resistive state (LRS), after which the curves become much smoother. This is shown in Figure 3.1.

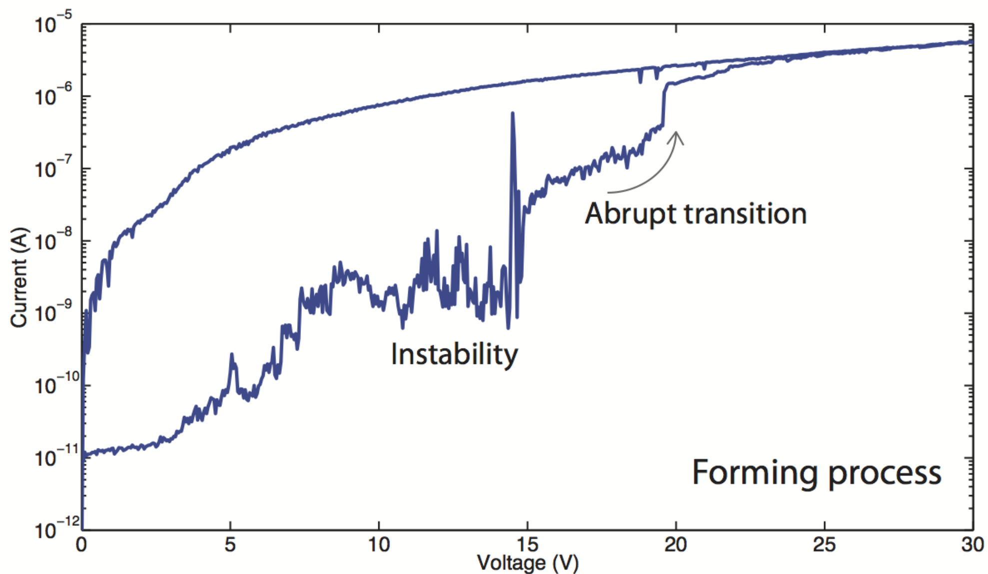

Figure 3.1: Electroforming process prior to resistive switching. Blue line is a double sweep curve (from 0V to 30V and from 30V to 0V) Obtained from a device from the 1st generation.

A structural change occurs during the electroforming process, most likely a local reduction process (removal of oxygen from silicon oxide). The low resistive state after the electroforming is stable, and the device remains in this state after the voltage bias is removed. In the case of MOS design (generations I,II and III) the electroforming process is done in the positive bias and no current compliance is needed. The reason why there is no need for current compliance comes from the fact that p-type Si substrate behaves as a self-compliance. The maximum current is limited by minority carriers (electrons) in the substrate. For MIM design, however, the current compliance is needed.

After the electroforming process, the device can be switched between two states – high resistive state (HRS) and low resistive state (LRS).

Figure 3.2 shows a typical I-V curve from a 1st generation device. Hysteresis is evident in both positive and negative bias.

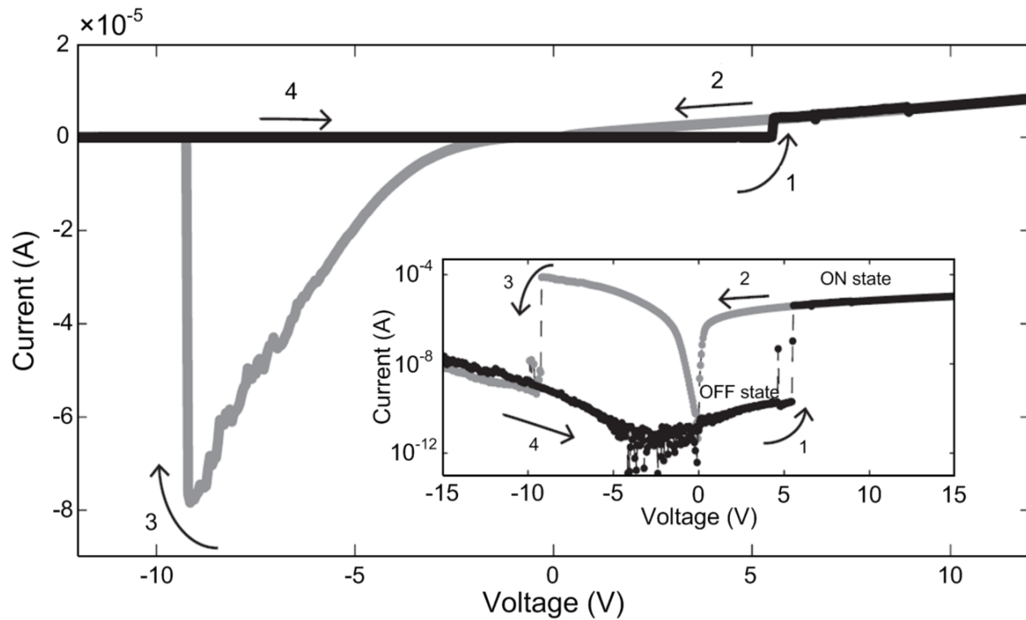

Figure 3.2: Typical hysteresis I-V curve in bipolar switching mode. (Inset: Logarithmic representation). Obtained from a device from the 1st generation.

In the positive bias (top electrode biased positively with respect to the silicon substrate), the high resistive state (HRS/OFF) switches to the low resistive state (LRS/ON) at a threshold voltage (black line in the positive bias). This is the set process. Reducing the voltage below threshold does not switch the device to its initial state, and much larger currents flow (grey line).

Resistance contrast between the two states is up to six orders of magnitude in devices studied to date; four to five orders are typical. Transitions between states are abrupt (faster than the sampling time of our measurement) and do not depend on voltage sweep speed. The results shown are obtained with an acquisition time of around 15ms per point (the maximum resolution of our equipment), which is equivalent to a sweep rate of 3.33 Vs^{-1} with 300 points sampled.

After switching, the device stays in the LRS until a negative bias (silicon substrate biased positively in respect to the top electrode) is applied and certain critical voltage is reached at which a sudden current drop is observed. This is the reset process. This sequence can be repeated many times.

The ON current does not show a clear tendency to scale with contact size (we tested this with 3 contact sizes of 125μm × 125μm, 250μm × 250μm and 500μm × 500μm), suggesting carrier transport via individual conductive pathways. We note that both ON read current and reset current

could be increased if much larger contacts are used (2mm $\times$ 2mm or larger). In this case it is likely that a number of parallel conduction paths or conductive network is formed.

Set and reset voltages can vary from the values shown in Figure 3.2, depending on the history of the devices (duration and magnitude of the last applied voltage bias). However, short voltage pulses of $\pm$15V or greater almost always trigger switching in healthy devices. These switching processes are repeatable and the states are stable for at least 120 hours at room temperature.

This is the bipolar mode of programming. Opposite polarities are used for the set and the reset processes. Physically this type of memory falls within the class of Valence Change Memories.

In such memories, during the electroforming process a local reduction process occurs forming the conductive channel. This conductive channel/filament is likely to consist of oxygen vacancies that can drift when electric field is applied.The switching site is likely to be near one of the interfaces between the filament and the electrode. At the interface a tunnelling barrier can form and the effective height of the barrier will define the resistance state. After the filament is formed, the negative bias will push the oxygen vacancies from the interface between the substrate and the filament. After the oxygen vacancies are repelled from this side the local redox process can occur. This results in increase of the barrier height and the HRS is achieved.

Positive bias will push the oxygen vacancies back towards the substrate and the tunnelling barrier will be reduced again. This results in reduction of barrier height and recovery of the LRS.

This is the convectional description of switching process of Valence Change Memories, but many variations are possible. Specific particularities about our system will be discussed in the next chapter.

In some cases it is possible to return the device to a partially OFF state (HRS) with an intermediate resistance after the reset process, as shown in Figure 3.3. This behaviour is even more pronounced with second and fourth generation of devices.

We note that the set process is also achievable under negative bias, but its occurrence is less likely (and normally occurs only for larger contacts) than in the case of positive bias. This is consistent with the model of Valence Change Memories in which the active side is more likely to be formed near the interface with the electrode with lower oxygen affinity. In our case the oxygen affinity of the p-type silicon substrate is lower than the oxygen affinity of the ITO electrode.

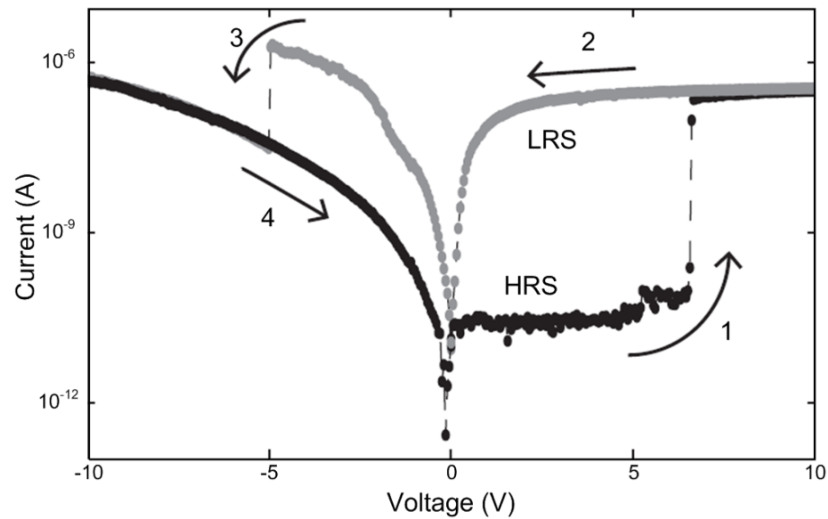

Figure 3.3: I-V characteristics with the reset process to intermediate state (line 4). Obtained from a device from the 1st generation.

The second generation devices with the top n-type poly Si show similar switching behaviour. A typical I-V curve is shown in Figure 3.4.

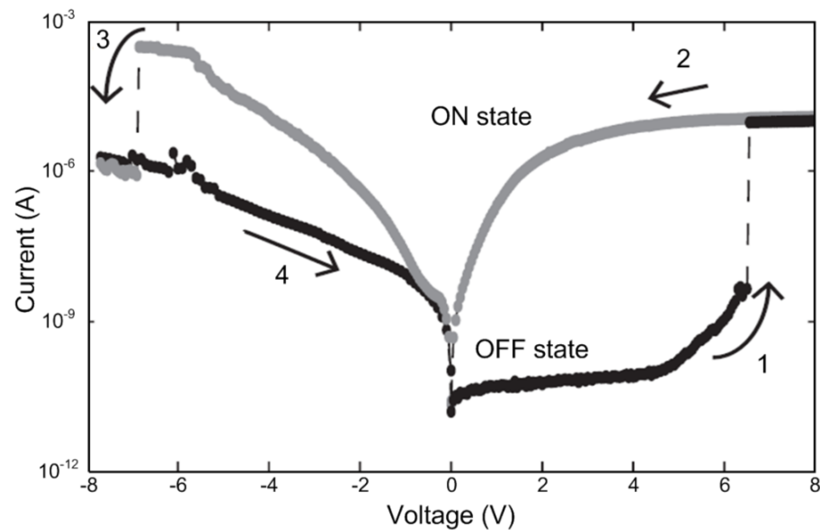

Figure 3.4: I-V characteristics for a device with a poly-Si top contact. Obtained from a device from the 2nd generation.

In the case of devices with top poly-Si contacts the reset process occurred less frequently than for devices with top ITO contacts. Also the endurance of these devices was much lower. This can be attributed to an additional annealing step at 950°C required to maximize the conductivity of the n-type poly-Si layer. This is likely to produce high density of silicon nanoinclusions within the film as a result of phase separation, allowing strong conductive filaments to form. Consequently higher currents are induced during the switching process and additional structural changes could occur

leading to permanent damage of the sample.

Figure 3.5 illustrates sequential cycling between high and low resistance states (device programming) for a 1st generation device. For the results shown in Figure 3.5 a programmed sweep mode was used with a sweep speed of around 50ms per point (the maximum resolution of our equipment in this mode). The applied voltages were: +20V (set), -20V (reset), and +2V (read). In the best devices programming can be achieved using 10V pulses as short as 90ns (the limit of our equipment resolution – for these measurements a stand-alone pulse generator was used, rather than the Keithley 4200).

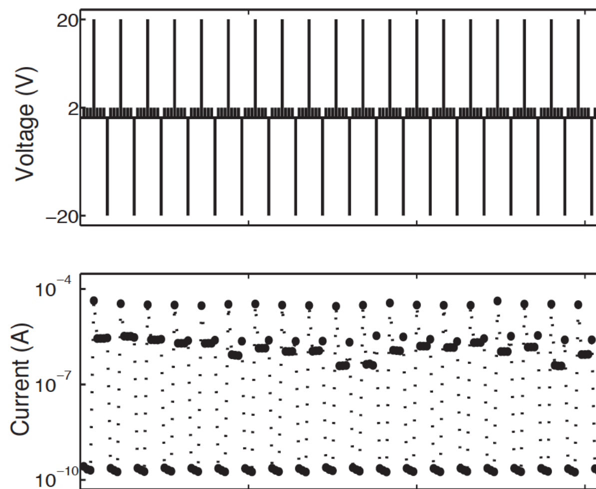

Figure 3.5: Switching cycles using voltage pulses of +20, -20, and 2 V for setting, resetting and reading, respectively. Obtained from a device from the 1st generation.

Reset current in these devices can be up to 100mA (typical is around 10mA, and in the best devices this can be lower than µA), which is still sufficiently low for memory applications. The read operation uses approximately an order of magnitude less current in the ON state. The variations in reset current are likely an effect of non-homogeneous film deposition.

The devices of generation 3 exhibit even lower programming voltages and currents. This is shown in Figure 3.6.

Figure 3.6 shows a typical I-V curve for bipolar switching between two states (using a thinner sample – a device from the 3rd generation). Programming voltages around 3.5V initiate resistive switching. In this switching regime, devices are not swept to voltages higher than ±7V, in order to retain low voltage and current operation. The set process occurs from as low as 3V, depending on the

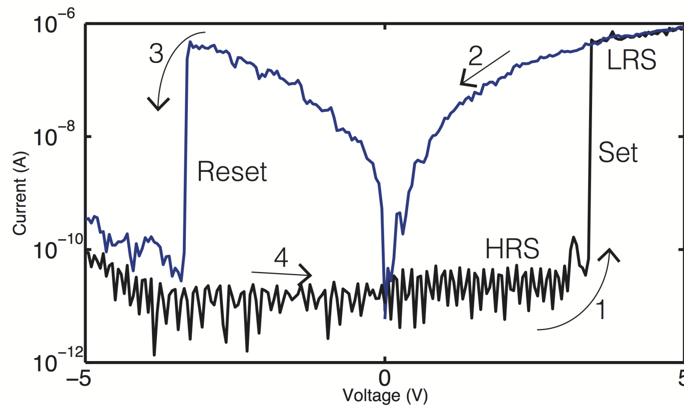

Figure 3.6: Typical I-V curve in bipolar mode. Obtained from a device from the 3rd generation.

previous negative sweep; we note that the set voltage is highly dependent on the device history (specifically the negative sweep). However, keeping sweep voltages in the range of ±7V provides low switching voltages with a narrow and stable distribution. Lower programming voltages and currents are obtained compared to sample of the 1st generation.

Figure 3.7(a) shows sequential cycling (device programming). A programmed sweep mode was used with a sweep speed of around 50ms per point (the equipment resolution in this mode). Applied voltages were +20V (set), -10V (reset), and -1V (read).

Another important requirement for ReRAM is narrow distribution of resistance within a single level. Figure 3.7(b) shows resistance distribution in both states. Note that the wider distribution in the HRS is probably affected by the extremely low current level read in HRS (current level close to the resolution of the equipment).

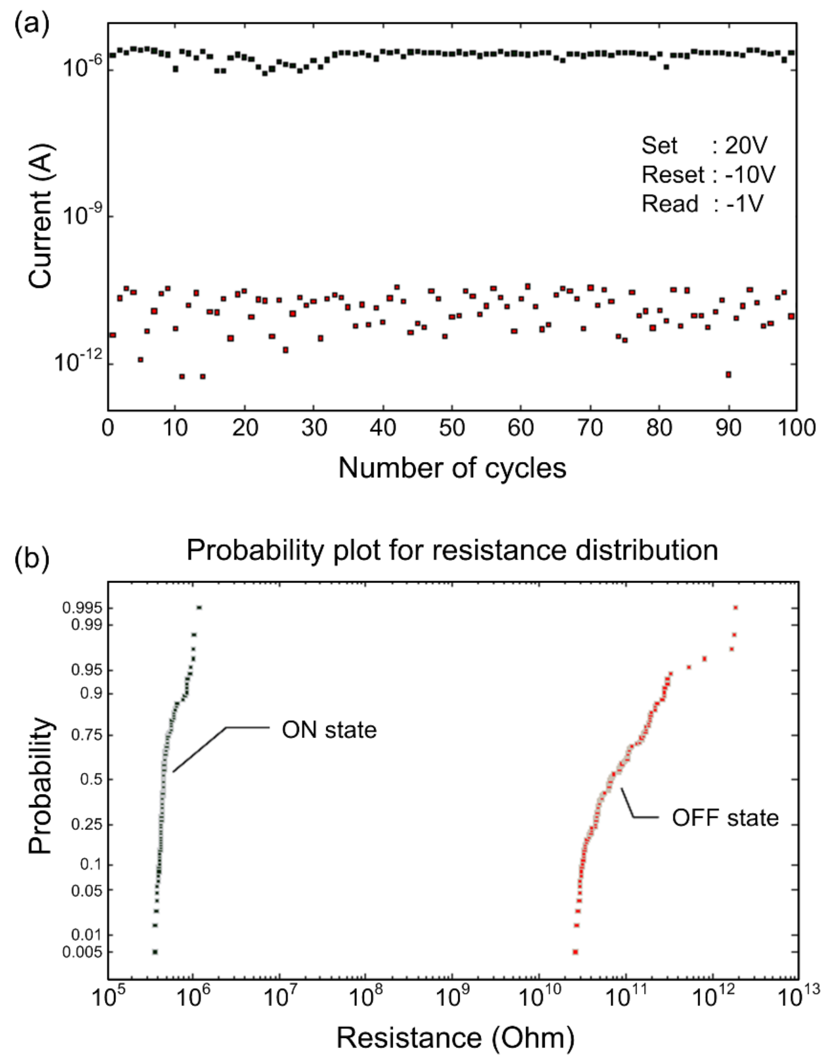

Figure 3.7: (a) Device cycling between HRS and LRS state. The resistance state is shown after every cycle. (b) Resistance distribution in LRS (ON) and HRS (OFF) calculated from 100 cycles. Results obtained from a device from the 3rd generation.

Unipolar Switching Mode

Most metal oxides exhibit either bipolar or unipolar switching. The main driving force for bipolar switching is electric field that causes the drift of ions inside the matrix. In case of unipolar switching the more dominant factor is thermal energy. Joule heating generated by high current causes high local temperatures.

We have found that silicon oxide exhibits the co-existence of both types of switching.

We can further differentiate two types of unipolar switching. In the case of conventional unipolar switching current compliance is needed; the set voltage is larger than the reset voltage. Figure 3.8 shows this type of unipolar switching.

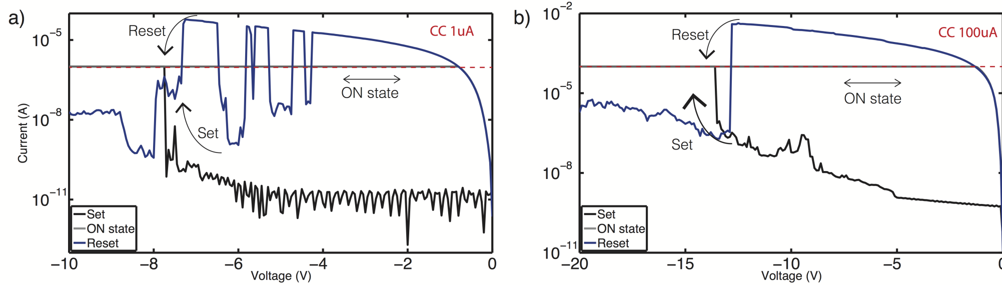

Figure 3.8: Results obtained from a device from the 3rd generation. (a) Unipolar device switching with low current compliance (1µA). (b) Unipolar device switching with higher current compliance (100µA).

Figures 3.8(a) and 3.8(b) show unipolar switching in the same devices. The black curves, labelled 'set', show the transition from HRS to LRS. For stable switching in this mode, current compliance (CC) is needed, in contrast to bipolar operation. Without the current compliance hard breakdown can occur, or the device can exhibit multiple transitions between the two states. These multiple transitions suggest competition between set and reset processes reported previously. High Joule heating tends to reset the device to the HRS, while the high field tends to set it in the LRS. This effect can be seen clearly in Figure 3.8(a) (blue curve) once the current compliance is removed: several transitions between different resistive states occur before the device finally fully resets at around -8V.

In Figures 3.8(a) and 3.8(b) the current compliance is set to 1µA and 100µA, respectively. The reset current is 20-60 µA (reset starts at 20µA and the device fully resets at 60µA) for a current compliance of 1µA, and 4.2mA for CC of 100µA. This proportionality between the current compliance and the reset current has been observed previously for a number of ReRAM systems, and is described in the work of Ielmini[54]. We note here that unipolar switching shows a wider distribution of set and reset voltages than is the case for bipolar switching. This is expected, due to the higher currents (and consequently temperatures) involved in Joule heating-driven transition processes in this mode: the shapes of the formed conductive filaments are likely to differ significantly from one switching cycle to another.

The second type of unipolar switching is if the set voltage is lower than the reset voltage. For this type of switching there is no requirement for current compliance.

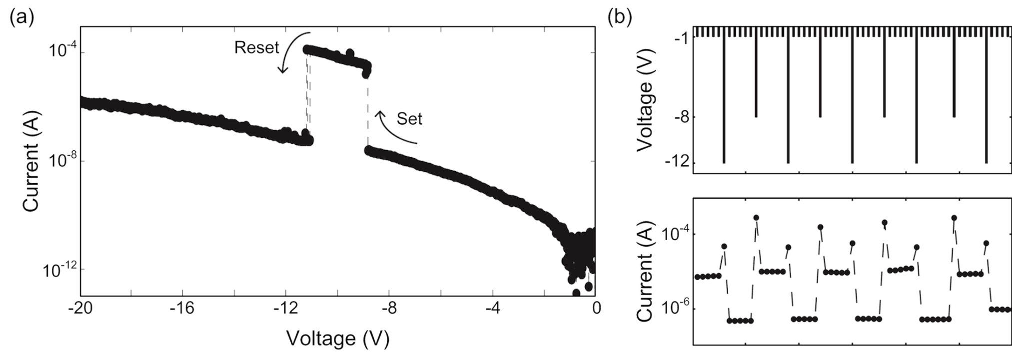

Figure 3.9: Untypical unipolar switching (250mm × 250mm contact size). Results obtained from a device from the 1st generation.

Figure 3.9(a) demonstrates unipolar switching of the device under negative bias. Both set and reset processes can be seen, with reset occurring once the current has reached a critical value. The switching process is thus inherently unipolar.

Figure 3.9(b) shows unipolar programming of a device, showing that it is possible to repeatedly set and reset the resistive states using one polarity of voltage pulse. However, it should be noted that only a small number of devices exhibited stable switching in this mode. This is likely to be related to competition between the two processes, and the window between the set and the reset voltage is relatively small in this case.

For the devices of the fourth generation with MIM design only the unipolar switching mode is observed. This is a consequence of the inert type of the both electrodes and the symmetric design.The switching (both set and reset processes) is thermo-chemically driven. This is the TCM type of ReRAM. The typical I-V curves are shown in Figure 3.10.

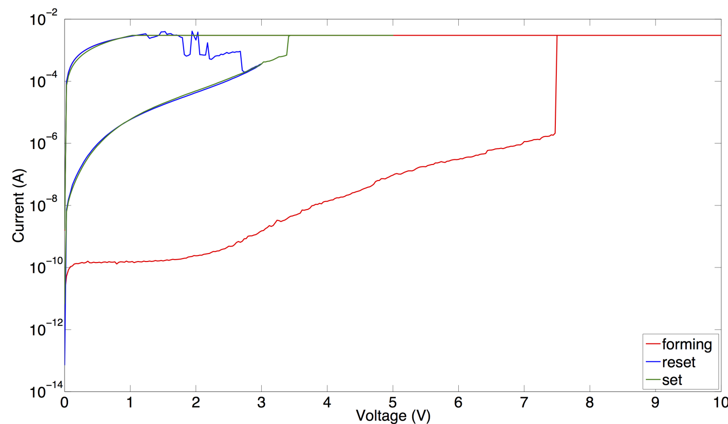

Figure 3.10: I-V curves obtained from a device from the 4th generation (MIM design).

Multi-level and Gradual Switching

Multi-level bipolar switching is demonstrated in Figure 3.11(a) for a 3rd generation device. The initial set process from HRS to LRS can be followed by a second transition to an even lower resistive state, providing at least three stable states. In negative bias, two reset processes occur, reversibly switching the device in two steps to the HRS. As the window for set transitions is usually wider than that for the complementary reset transitions, we use two different levels of positive voltage pulses to set the device either to the first LRS or the second LRS.

Figure 3.11(b) shows three-level programming with three different voltage pulses. We used a +12V pulse to set the device to the first LRS, +20V to set it to the second LRS (either from the first LRS or the HRS), and -10V to reset the device to the HRS either from the first or the second LRS. 4V was used for stable non-destructive reading.

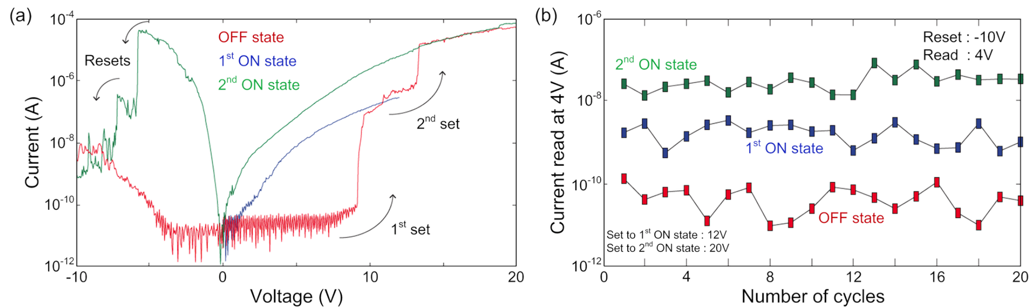

Figure 3.11: (a) I-V characteristics showing multi-level switching. (b) Three-level pulse programming. Results obtained from a device from the 3rd generation.

The LRS states are stable, as in the case of two-level programming, with a contrast between states of around one order of magnitude. Reading at negative voltages (e.g. -1V) can provide a higher contrast, although the correspondingly higher reading current can slightly modify the first LRS.

In the case of bipolar switching, in addition to abrupt changes in device resistance, we also observe a more gradual change shown in Figure 3.12 by sequentially sweeping the device to positive voltages with a current limit set. Starting in the LRS (after an initial sweep to switch the device from HRS to LRS), we performed three sequential sweeps in positive bias. For the first one we set the current compliance to 100µA; subsequent sweeps were performed without a current limit. Three I-V curves were obtained.

The gradual modification of device resistance can be observed during the reset process. Figure 3.13 shows the gradual reset process during unipolar switching. Starting from the LRS and sweeping the

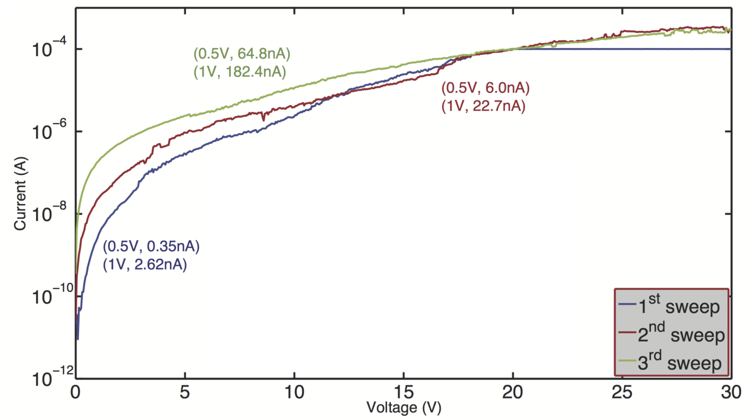

Figure 3.12: Gradual increase in LRS conductivity with subsequent positive sweeps.

device to progressively higher voltages without setting the current compliance many different stable resistance states can be achieved. The final reset process is usually more abrupt, after which no more transitions can be seen.

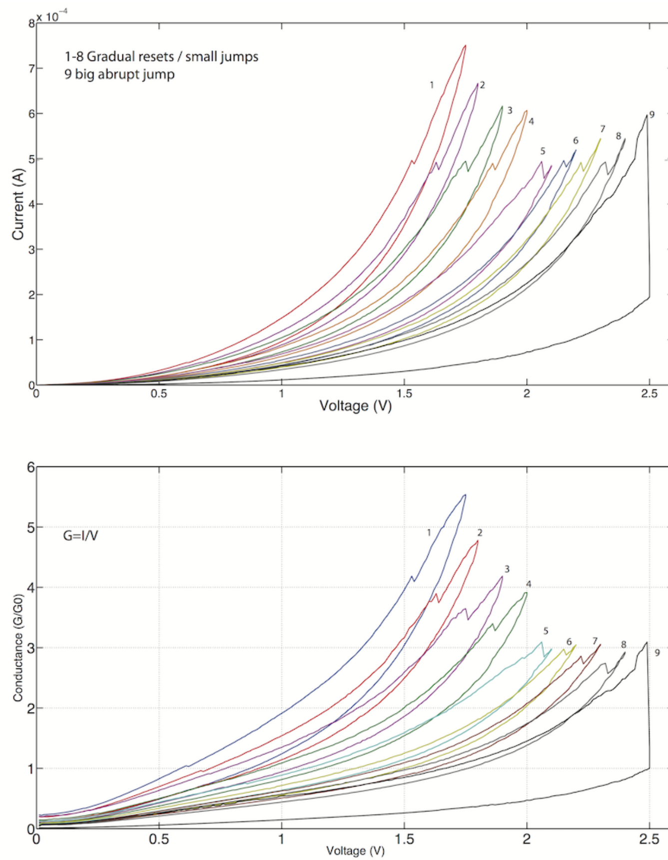

Figure 3.13: Gradual modification of device resistance during the reset process. Results obtained from a device from the 4th generation.

Temperature Dependence

The nature of the filaments formed after the unipolar or bipolar switching modes are different, so in the analysis of the temperature dependence we differentiate these two modes.

Figure 3.14 shows conduction dependence on temperature in both LRS and HRS (from a 1st generation device) for the case of the bipolar mode. The result shows typical semiconductor behaviour (resistance decreases with increased temperature), suggesting no metallic conduction in the LRS. Temperature increase does not affect the HRS enough to be sensed by our equipment. Contrary to this, systems exhibiting metallic filament conduction show a decrease in the LRS conduction with increasing temperature.

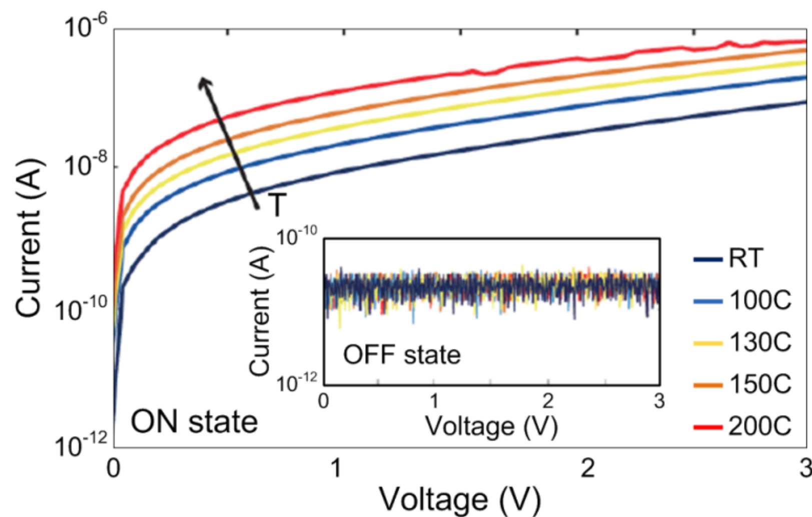

Figure 3.14: LRS and HRS dependence on increasing temperature in bipolar switching. Results obtained from a devices from the 1st generation.

When devices are heated up to 200°C during test, the conductance in the LRS shows a reversible increase, confirming the non-metallic nature of the conduction path. As the current is determined by the conductive pathway rather than the availability of carriers from the substrate, we conclude that the pathways are either silicon or oxygen vacancies precolation paths, rather than diffused metal ions from the ITO top contact, or from contamination of the active layer.

The result can also suggest that the filament created in bipolar mode consists of a channel that is not fully formed and that significant conduction increase can occur through a thermally activated hopping process.

In the case of the unipolar mode a much weaker dependence on ambient temperature is observed (see Figure 3.15). This suggests that the nature of filament formed after the unipolar switch is different to the one formed after the bipolar switch.

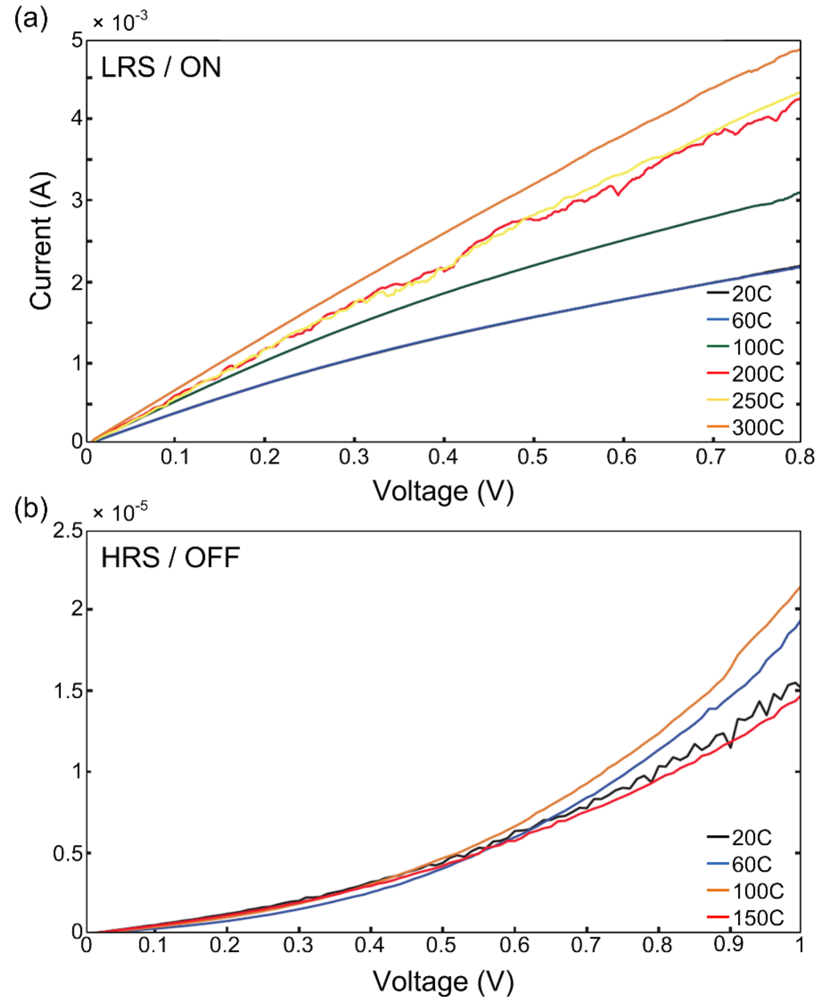

Figure 3.15: LRS and HRS dependence on increasing temperature in unipolar switching. Results obtained from a device from the 4th generation.

In the LRS a small overall increase in the conductance is observed. This increase is much lower than in the case of bipolar switching. Also, it is apparent that in case of unipolar switching the I-V curve is much more linear compared to bipolar switching. In the HRS there is no strong temperature dependence up to 150°C. This suggests that the conduction mechanism is not thermally activated in case of unipolar switching. The filament created in unipolar mode is much more homogeneous (compared to the filament produced in bipolar mode). In this more homogeneous filament, the potential barrier seen by the tunnelling electrons reduces as the band offset between the conduction bands of the silicon nanoinclusions and the substoichiometric oxide shrinks, and hence the resulting I/V curve becomes more linear.

Current-time Characteristics

Figures 3.16(a) and 3.16(b) show current-time graphs under constant voltage biases of -8V and -10V for the 2mm × 2mm contact: it is evident that formation (set) and destruction (reset) processes act in opposition.

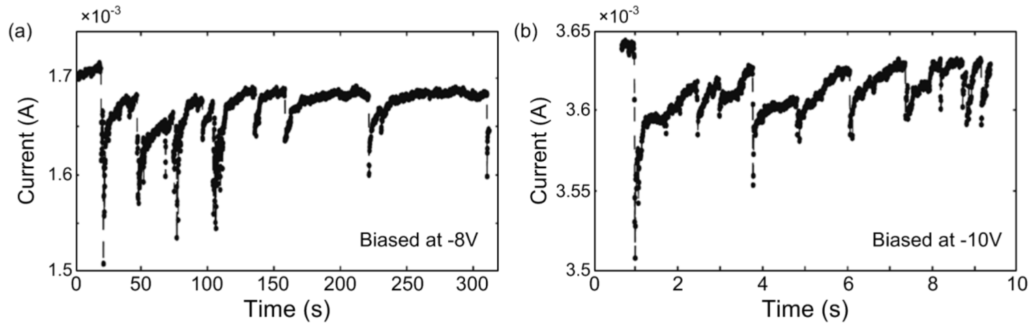

Figure 3.16: Current-time graphs under a constant voltage bias of (a) -8V and (b) -10V. Results obtained from a device from the 1st generation.

In the case of larger contacts, it is likely that multiple conduction paths are formed, resulting in increased reset negative voltage and current. Clear transitions between two states can be seen with rapid current drops (up to 10%) followed by slower, exponential increases. Recovery speed is proportional to the applied voltage: the average time constant decreases from 10sec to around 0.5sec on changing the bias from -8V to -10V. Higher negative voltages reset the device without repetitive transitions to the LRS. However, the LRS is stable if the bias is lower than 6V.

We distinguish two conditions: high field and low current (positive bias) in which pathway formation is favoured, and high field and high current (negative bias) in which destruction dominates. We therefore propose a switching mechanism based on competing field-driven formation and current-driven destruction of conductive pathways.

Impedance Spectroscopy

Impedance Spectroscopy reveals different conduction mechanisms in LRS and HRS, and is used to examine the nature of a conducting path.

In the HRS, a single arc suggests an equivalent circuit model with a single parallel capacitor and resistor ($R = 2.81 \times 10^{8}\Omega$, $C = 2.61 \times 10^{-12}F$).

This high resistance is governed by the bulk properties of the SiO_x layer. However, in the LRS, two clear arcs are seen, suggesting additional parallel resistances and capacitances appearing in series. This is consistent both with formation and destruction of conductive pathways, and with previous studies of carrier transport in such films. The latter can be modelled as shown in the inset of Figure 3.17(b) ($R_1 = 1.68 \times 10^6 \Omega$, $C_1 = 1.72 \times 10^{-10} F$, $R_{eq} = 2.28 \times 10^8 \Omega$, $C_{eq} = 3.03 \times 10^{-10} F$ (equivalent capacitance of series capacitors)).

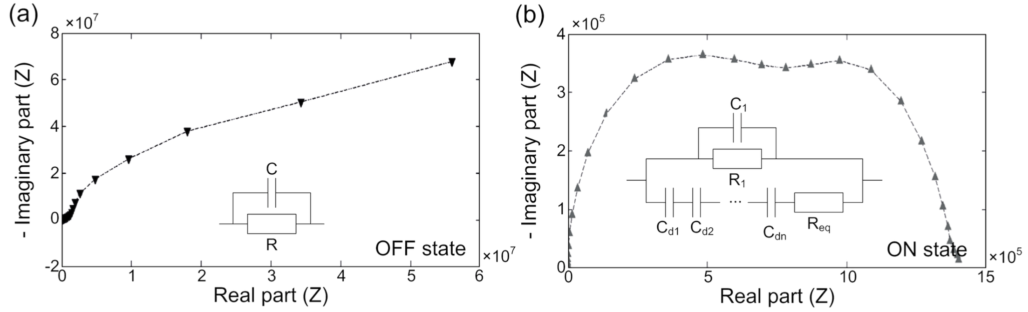

Figure 3.17: Cole-Cole plots with equivalent circuits under 1V in (a) HRS and (b) LRS. Obtained from a device from the 1st generation.

The imaginary part of the impedance remains negative at frequencies between 1Hz and 10^7Hz, suggesting no inductive behaviour. Inductance would be typical for metallic filaments, which is not the case for our devices from the 1st generation.

Nonlinearity and Self-rectification

As discussed in the introduction, two highly desirable properties of resistive switching devices are self-rectification and non-linearity. These allow prevention of faulty device reading due to leaky current paths in passive crossbar arrays without the requirement to add a diode in series with each ReRAM element, as is the case for most metal oxide-based systems.

Non-linearity is usually defined as $\frac{I(V_{red})}{I(V_{red}/2)} \gg 2$. In the case of bipolar switching in our devices we can obtain a ratio of up to 10. This is shown in Figure 3.18(a). This ratio decreases as resistance in the LRS decreases, which is consistent with the trap-assisted tunnelling model (this model will be described in the Chapter 5) in which the current-voltage relation is highly non-linear. With growth of a conductive filament, non-linearity decreases as the silicon sub-oxide gaps are filled with more oxygen vacancies (silicon nanoinclusions). Continual growth of the conductive filament into the more continuous form causes a reduction in non-linearity. While conductive filaments formed after the unipolar switch are more continuous, and hence show poor non-linearity, excellent self-rectifying

behaviour is observed.

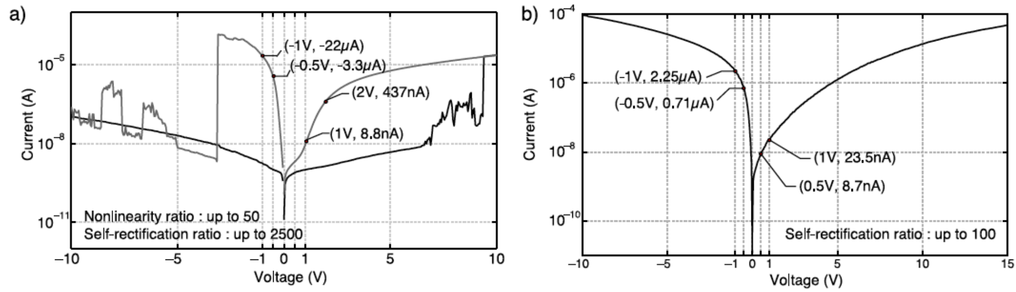

Figure 3.18: (a) Self-rectifying and nonlinearity behaviour in a sample with top poly-Si electrode. (b) Self-rectifying behaviour in a sample with top ITO electrode after unipolar switch. Obtained from a device from the 3rd generation.

In Figure 3.18(b) we can see that the difference between currents flowing in positive and negative bias is around two orders of magnitude for a read voltage between 0.5V and 1V. Similar behaviour has been previously reported for unipolar resistive switching in hafnium oxide[55], and it is believed that the formation of a Schottky barrier at one of the interfaces (ITO/SiO_x or SiO_x/Si) is the cause.

Structural Characterisation

AFM and STM Characterisation

It is well established that films grown by sputtering often exhibit columnar or granular growth. Boundaries between adjacent columns can extend through the whole thickness of the film, effectively connecting the top and bottom of the active layer.

Atomic force microscopy (Figure 3.19(b)) of sample surfaces indicated periodic 5-10 nm high circular dome shaped surface features that varied in diameter from 10 to 50 nm. STM of these structures showed high conductivity at the edges of the features, and low conductivity at the centre (Figure 3.19(c)). Such pathways are around 5-30 nm in diameter, suggesting that devices may be scaled down to nanometer dimensions to achieve very high levels of integration. Figure 3.19(a) shows I-V measurements of the two regions, highlighting the difference in conductivity. Figure 3.19(d) is a schematic of the film structure, showing columnar structures with silicon nanoinclusions nucleating at column boundaries. Such structure is consistent with both AFM and STM results. These results have been reproduced in around 15 different scan areas.

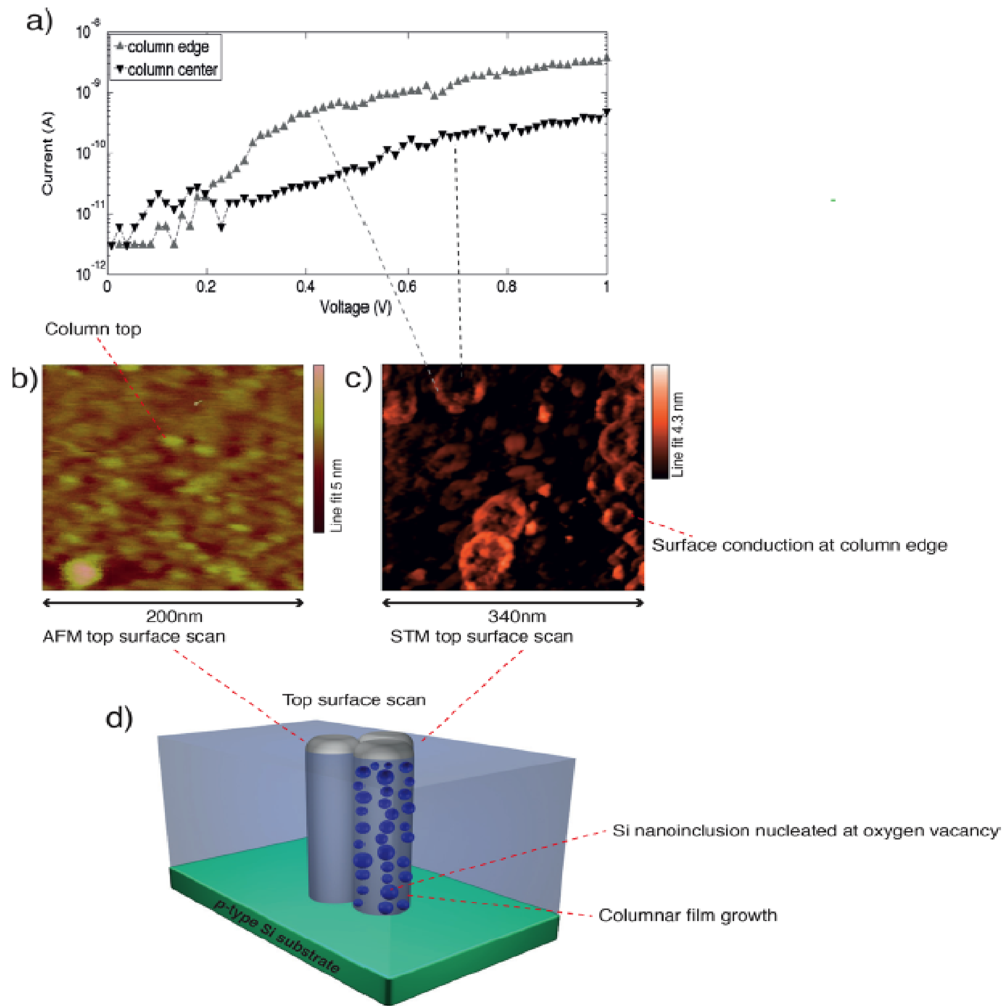

Figure 3.19: (a) STM I-V curves for the edge point. (b) Atomic force microscopy scan of the surface top side showing surface features attributed to the tops of growth columns. (c) Scanning tunnelling microscopy scan of a sample surface (different area to that in (b)) showing enhanced conductivity at column edges. (d) Schematic of columnar structure of switching film and switchable site. Obtained from a device from the 1st generation.

Placing the atomically sharp STM tip above a column edge and applying the appropriate voltage allows us to switch the material between HRS and LRS. Figure 3.20 shows results obtained using the STM tip on a portion of the sample without a top electrode, demonstrating resistive switching between two distinct resistive levels. Note that in this case a voltage of only -1.8V triggers the set transition.

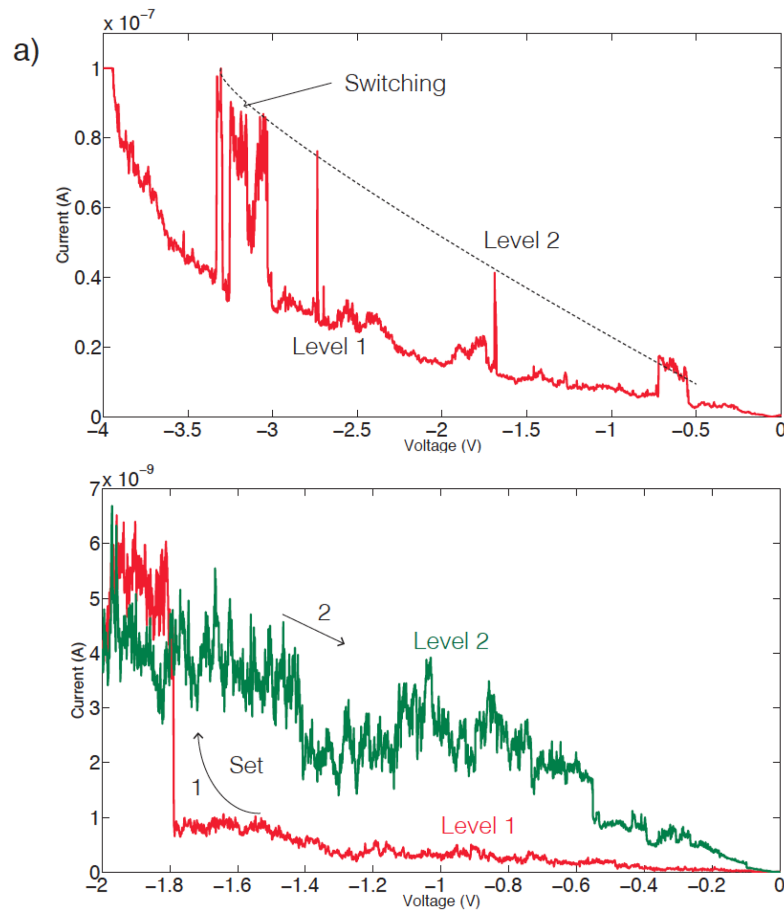

Figure 3.20: I-V curve obtained with STM spectroscopy showing resistive switching between two distinct states. Obtained from a device from the 1st generation.

C-AFM Characterisation

Conductive Atomic Force Microscopy (C-AFM) is used in order to examine the shape and size of the filaments produced after the electroforming process. Measurements were taken using a Vecco Dimension 3100 AFM.

The filament produced after electroforming is typically very small and tracking an individual filament is a challenging task. The size of the smallest top contact is 10µm × 10µm. Tracking a nanometre-scale filament (after the removal of the top contact) on that scale is practically very hard. Instead we used a contactless area of the wafer to electroform the filament. We probed the wafer with a 1µm × 1µm probe tip. To be able to mark the area where the filament is formed we intentionally induced damage to the wafer during electroforming. Extensive heating (due to high current) generates enough heat to leave a visible mark on the wafer. Note that this procedure results in hard breakdown

and the device cannot be switched back to the HRS. Nevertheless, this gives useful information about the filaments that are formed at these extreme conditions (without current compliance).

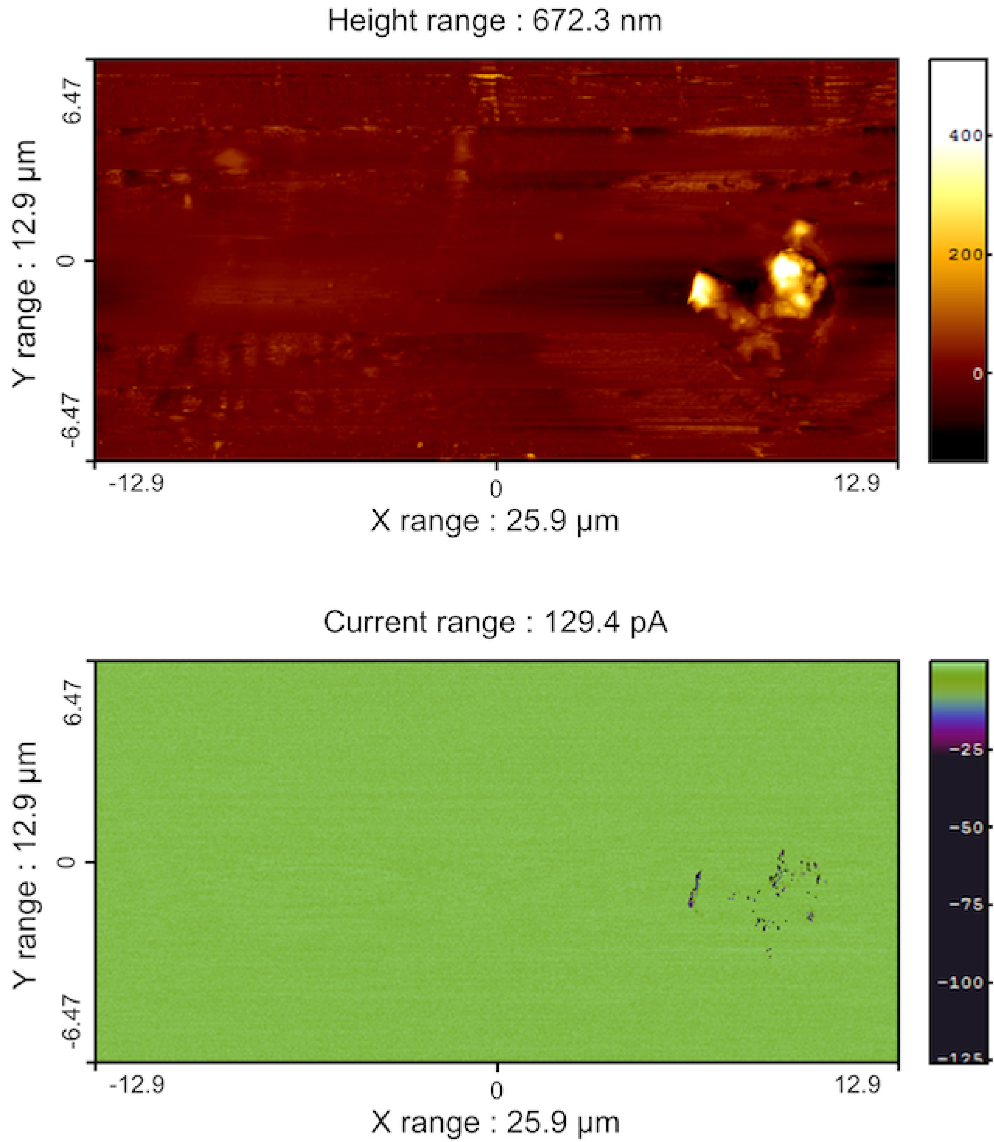

Figure 3.21: Top image: AFM scan of burned area. Bottom image: C-AFM of the same area. Obtained from a device from the 4th generation.

Figure 3.21 shows the burned area of the wafer where the filament is formed. The damaged wafer area is spotted in the AFM scan as a bump of about 670nm in height. The corresponding C-AFM scan shows the area with increased conductivity. Many individual areas are observed and many different filaments might be formed. This is likely the effect of tip's irregular shape and irregular contact with the sample.

To better examine the filament shapes and sizes we performed additional AFM and C-AFM scans over a smaller area.

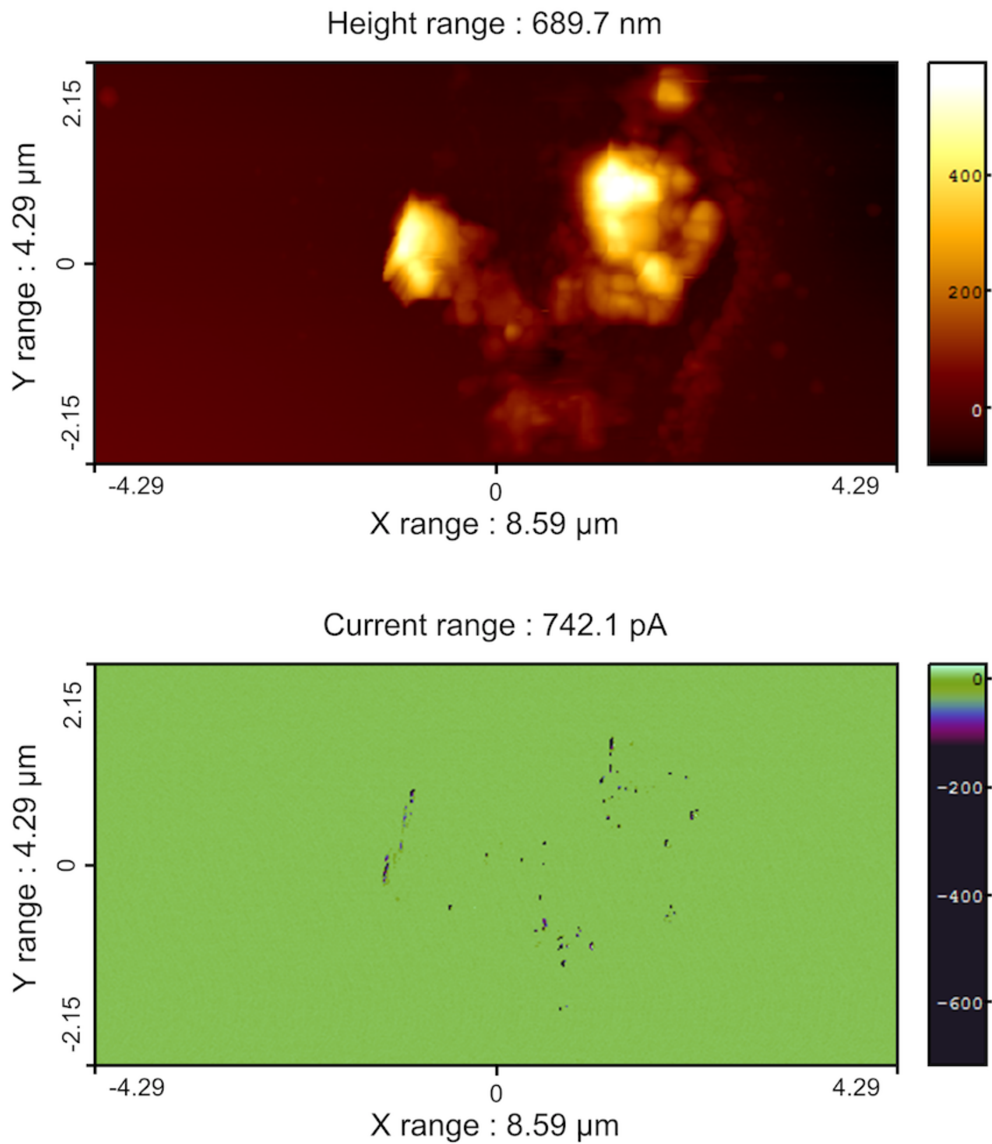

Figure 3.22: Top image: Zoomed AFM scan (8.9µm × 4.3µm). Bottom image: C-AFM of the same area.

In figure 3.22 it is quite obvious that the areas of increased conductivity are much more localized than the areas where the wafer is damaged.

After we further decrease the scanned area the AFM scan starts to lose its form and it is hard to distinguish the edges of damaged areas. However, C-AFM starts to reveal more details about the filaments.

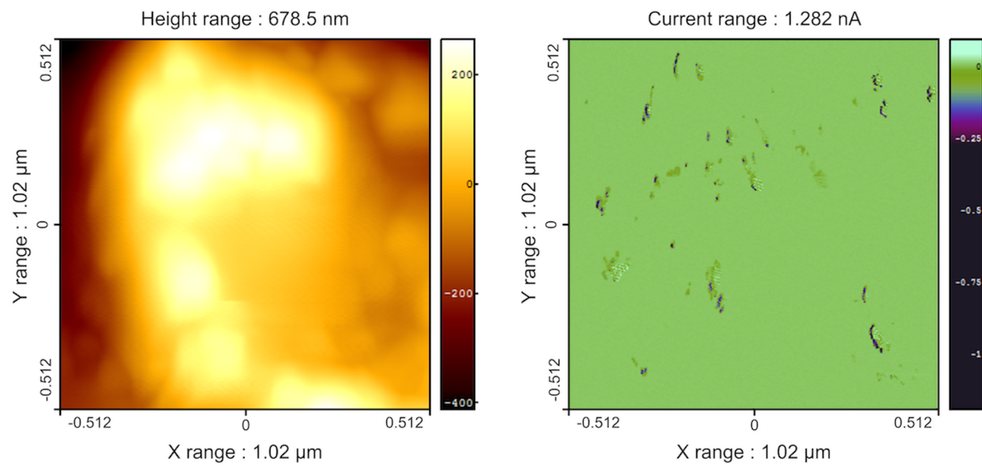

Figure 3.23: Left image: Further zoomed AFM scan (1µm × 1µm). Right image: C-AFM of the same area.

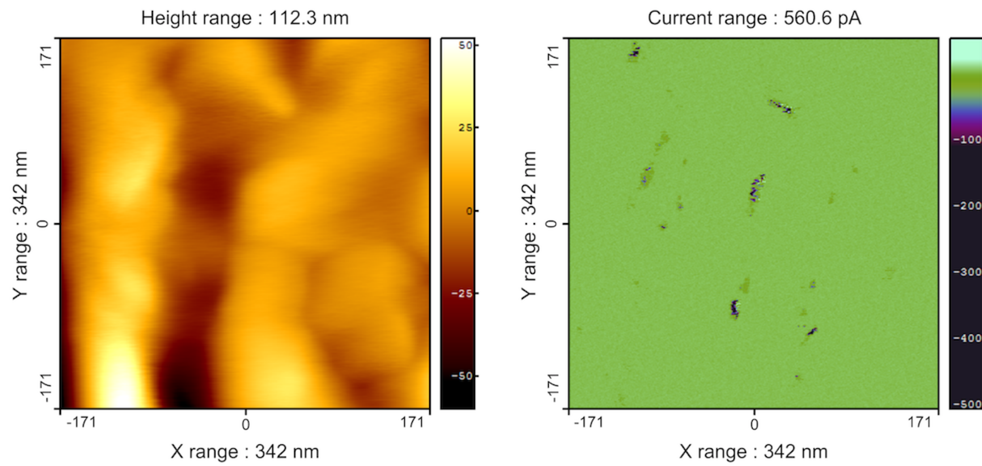

Figure 3.24: Left image: Further zoomed AFM scan (340nm × 340nm). Right image: C-AFM of the same area.

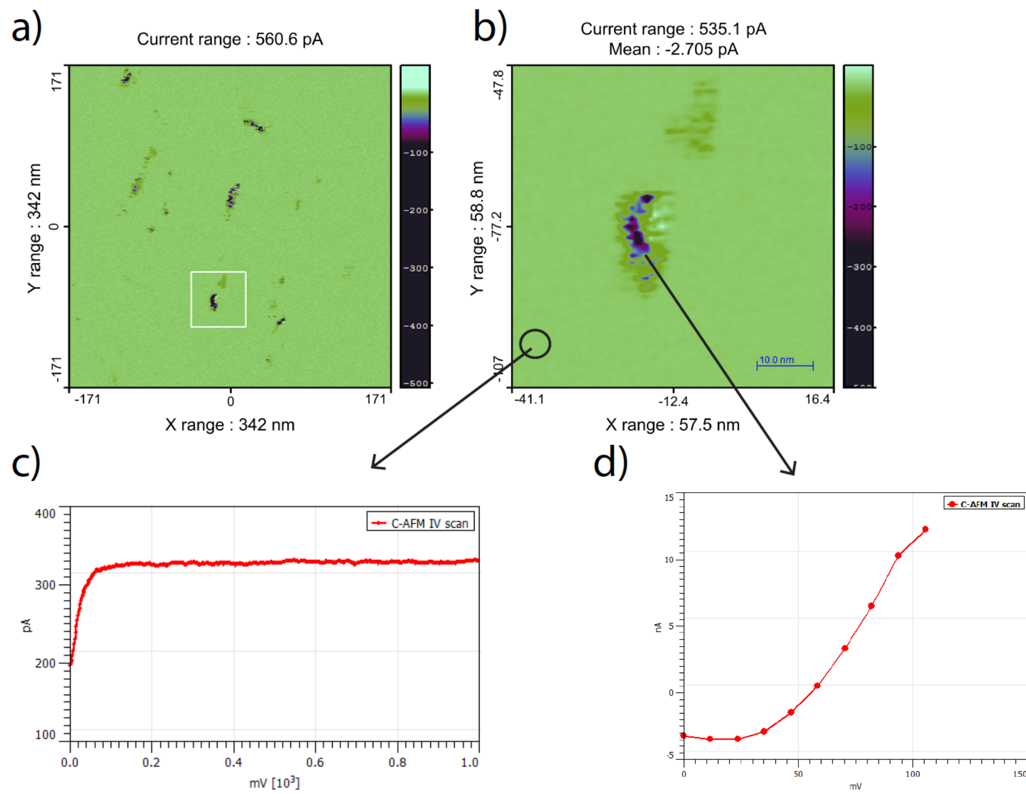

Figure 3.25: (a) C-AFM scan (340nm × 340nm). (b) C-AFM scan (60nm × 60nm). (c) I-V curve on filament-free area. (d) I-V curve on filament area.

Eventually it is possible to measure the dimension of filaments. Most filaments are smaller than 10nm. To further confirm this we took I-V curves from the areas of high and low conductivity. The I-V curve in Figure 3.25(c) corresponds to an area of low conductivity – an area without filaments - and the I-V curve in Figure 3.25(d) corresponds to an area with the formed filament. In the latter case current achieves the compliance level above 100mV, while in the first case there is no increase in current above the noise level.

TEM Characterisation

An important insight in the microstructure of the prepared samples can be obtained with the Transmission Electron Microscopy (TEM) technique. The main problem that might arise doing precise measurements with these techniques is a poor contrast between Si and SiO_2. Consequently, it is very challenging to observe Si nanoclusters in an SiO_2 matrix. Normally, only Si-nanoclusters that are crystallised are observable. For the tested devices the highest used temperature is 900°C.

This temperature is still below the threshold temperature for crystallisation of Si-ncs in thick films. The thinner films normally require higher temperatures for crystallization. This effect is shown in Figure 3.26 where nanoclusters are observable only in thicker sample annealed at 1100°C. TEM measurements are done in collaboration with Dr Sebastien Cueff.

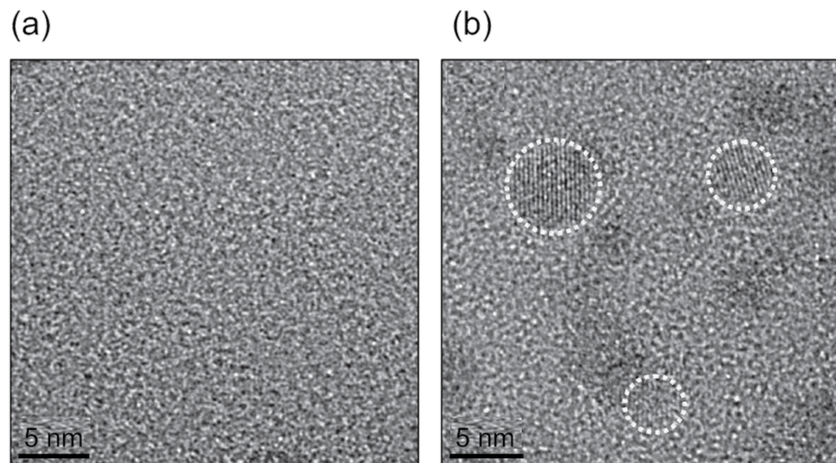

Figure 3.26: Transmission electron microscope images of samples deposited at 500°C for two different thicknesses. (a) 50 nm and (b) 1,400 nm. In "thin" film (a) no Si-nc was detected throughout the whole area of the sample, while in "thick" film (b) numerous well-crystallised Si-ncs are seen with diameter as high as 5nm. Results are taken from PhD thesis of Dr Sebastien Cueff [56].

It is also found that there is an inhibition of the agglomeration and formation of Si-nanoclusters near the substrate. The possible reason for this phenomenon is the proximity of substrate and resulting mechanical stress of the matrix. This is visible from the results obtained with Dark Field TEM (Figure 3.27) where amorphous Si nanoclusters could be detected.

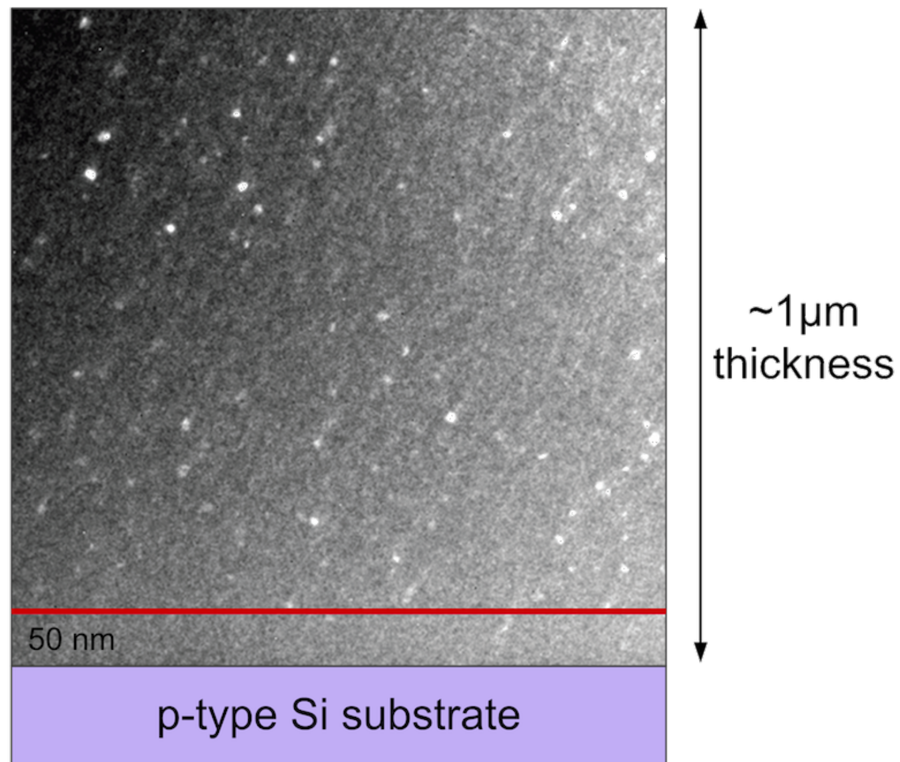

Figure 3.27: Dark field TEM shows bright spots – amorphous Si-nc that diffract incoming flux. Results are adopted from PhD thesis of Dr Sebastien Cueff [56].

The directional arrangement of Si nanoclusters (Figure 2.31) suggests that the film grown by sputtering exhibits a columnar structure. Note it is observed that there is the lack of silicon inclusions near the silicon substrate. In region to around 50nm from the substrate there is a very few silicon inclusions. This is likely the consequence of the mechanical stress produced between the silicon oxide film and the silicon substrate. However, this does not mean that the smaller inclusions (not observable by TEM) cannot be present in this area.

Chapter 4

Quantum Conductance in Silicon Oxide ReRAM Devices

Mesoscopic Electron Transport

The conventional treatment of electron transport in conductor relies on the Boltzmann model. The origin of resistivity is the scattering between the electrons and ions. Electrons thermally move inside the conductor. Moreover, every conductor has a certain number of impurities (lattice imperfections) that further increase the resistivity.

Nevertheless, in the nanometre-scale conductors the Boltzmann model is not valid. Systems in which the Boltzmann model is not appropriate are called mesoscopic systems. Many different conditions can lead to unsuitability of the Boltzmann model. Generally, if the length of the sample is comparable to the electron elastic mean free path the standard treatment of electron transport is not appropriate. In metals the Fermi wavelength is defined as the de Broglie wavelength of the electron at the Fermi energy ($\lambda = 2\pi/k_F$, where k_F is Fermi wave vector). In metals if the Fermi wavelength of electron is comparable to the dimension of the metal conductor than the Boltzmann model is also not appropriate.

Introduction to Quantum Conductance

Quantum wires (QW) are conductors with widths that are comparable with the Fermi wavelengths of the electron. In the case when the length of the QW is much larger than the elastic mean free path many scattering events will occur as electrons are passing through the wire. On the other hand, if the length of the wire is smaller than the elastic mean free path, the wires are called quantum point contacts (QPC). In this case there are no scattering events apart from the boundary scattering at the edges of the wire.

Somehow counterintuitively, in the case of QPC a finite, non-zero resistivity is still observed. Experiments show that when the ambient temperature is down to a few Kelvins the overall resistance of the ballistic QWR is quantised in steps of G_0, where $G_0 = 2e^2/h$.

Introduction to Quantum Conductance in RRAM Devices

Of particular interest, especially in the field of quantum information processing, are those systems that exhibit quantised conductance at room temperature.

One common example is that of a MEMS switch opening a metal-to-metal micro-contact until a sub-micrometre metallic bridge is formed between the electrodes (Poulain et al.[57]). Crossbar structured nanodevices have also demonstrated quantised conductance through the formation of atomic point contact switches. For example, Terabe et al.[58] report a quantised conductance atomic switch (QCAS) consisting of fixed Ag_2S-coated Ag wires and Pt wires. Switching arises from the formation and annihilation of a point-contact atomic bridge, via the reversible growth of a silver nanoprotrusion formed at the surface of Ag_2S at the crossing point between the two wires. Similarly, Wagenaar et al.[59] demonstrated quantisation of conductance in Ag_2S thin films using a Scanning Tunnelling Microscope (STM) Pt tip in contact with the thin film. In the context of resistive switching systems, microcrossbar structured AgI electrochemical metallisation (ECM) cells studied by Tappertzhofen et al.[60] exhibited multilevel resistive switching where, in the low resistive state, discrete resistance changes suggested a single atomic point contact. More generally, metal oxide-based RRAM devices also show quantisation of conductance in nanoscale conductive filaments.

Conduction Quantisation in Silicon Oxide RRAM

Reports in the literature of quantum conductance in RRAM devices variously mention quantisation in integer or half-integer multiples of G_0.

There is a need to reconcile these results, to understand which devices exhibit half-integer quantisation, and to relate this to the physics of the resistance switching process. We observe quantum conductance in silicon oxide-based ReRAM devices in which multilevel conductance steps are separated by half-integer multiples of G_0.

Conductance steps are likely to be associated with individual conductive filaments, as is the case in metal oxide systems. However, the more complex carrier transport in these devices, along with the inherent nonlinearity of semiconductor system (silicon and silicon oxide), suggests the origin of the half-integer quantisation in these and other ReRAM devices.

Prior to switching, devices must undergo an electroforming step. This serves as an initial nucleation of filaments within virgin material. By applying an appropriate voltage across a fresh device, filament growth is initiated and the device switches from an essentially insulating state (IS) to a low resistance state (LRS). A subsequent reset of the device changes it to a high resistance state (HRS) with a resistance lower than that of the initial virgin material. Further set/reset processes cycle the material between the HRS and LRS. Note that the conductivity of the filament formed after switching is strongly dependent on the compliance limit set during the electroforming step[54]. Figure 4.1 shows an initial electroforming event for one of our devices, which occurs at around -5.5V. The lower threshold voltage compared to the devices from the 1st generation is the effect of the thiner active layer (16nm compared to 35nm).

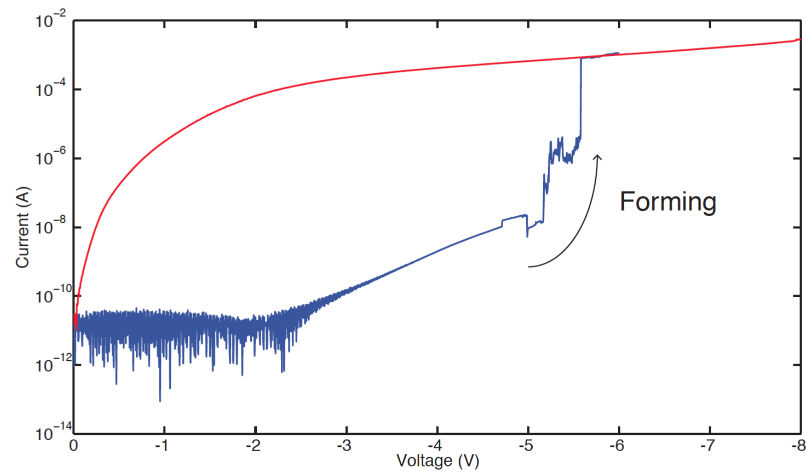

Figure 4.1: Electroforming of device conduction, followed by resistive switching. The device is initially in a high resistive state (HRS – blue line). Increasing the voltage produces a transition to a low resistance state (LRS) at around 5.5V, at which point a filament has formed. Subsequent reset operations do not return the device to its initial state, suggesting that the set/reset transition does not involve the complete destruction of the conductive filament.

As discussed in the previous chapter, devices shown schematically in the inset of Figure 4.2, consist of a thin (16nm) silicon-rich silica (SiO_x) layer sandwiched between a p-type silicon substrate and an n-type polycrystalline silicon top contact. These are the devices from the 2nd generation.

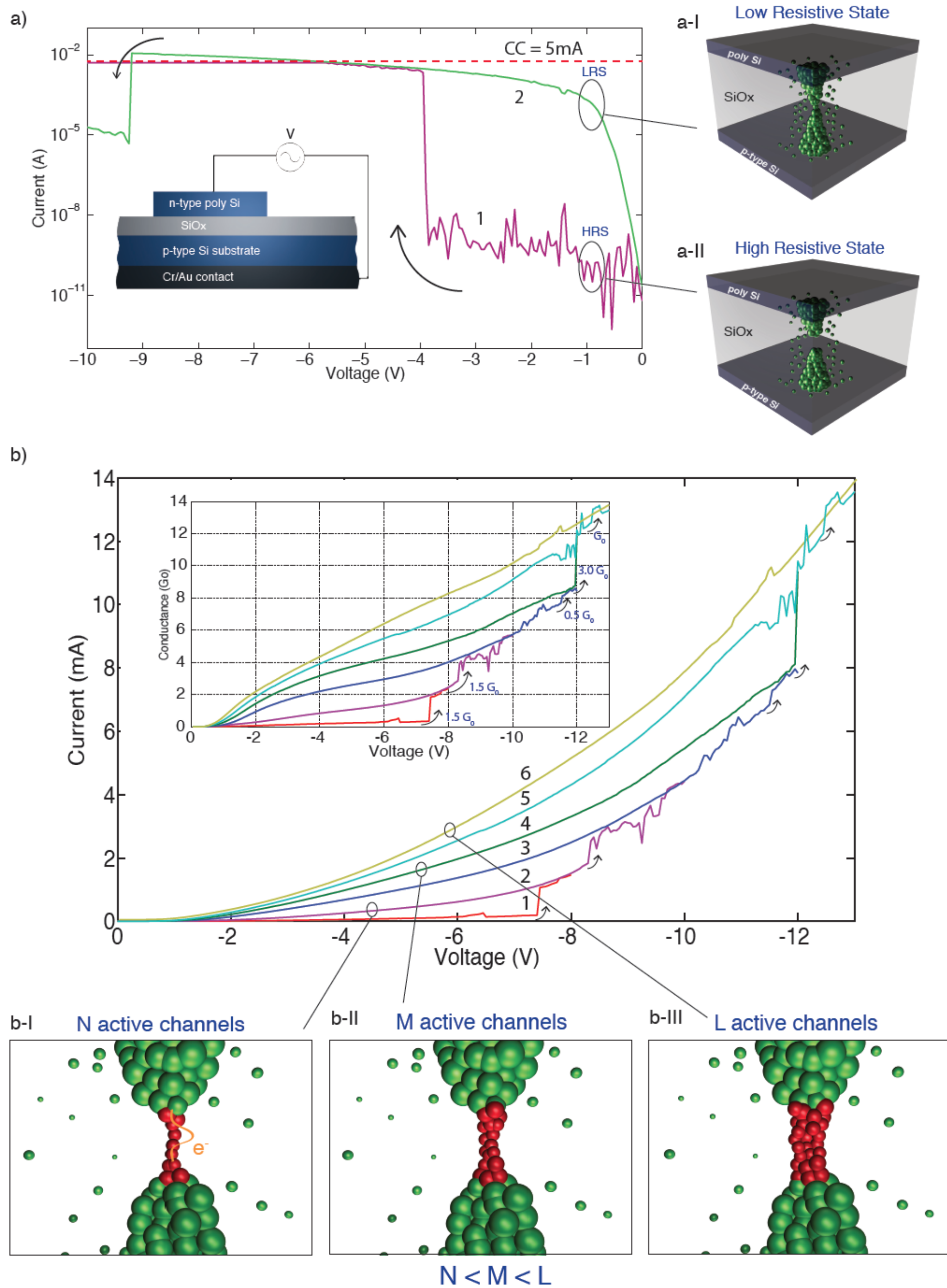

Figure 4.2: Unipolar switching of a device. (a) Magenta line: initial voltage sweep, with a current compliance (CC) limit of 5mA, showing a transition from OFF to ON state at -4V. Green line: subsequent voltage sweep of the ON state, with current compliance removed. Arrows indicate the direction of voltage sweep. Inset: Device schematic cross-section. (a-I) Schematic showing the formation of a continuous conductive filament within the oxide matrix, corresponding to the low resistive state of the device. (a-II) Schematic showing a break in the filament, corresponding to the high resistive state of the device. (b) Current-voltage characteristics of multiple conduction states during sequential voltage sweeps of a single device. These conduction states are seperated by half-integer multiples of G_0as noted in the G-V plot in inset. Inset: corresponding conductance-voltage graph. (b-I, b-II, b-III) show schematics of the growth in filament thickness during successive voltage sweeps to progressively higher voltages.

Devices exhibited both bipolar and unipolar resistive switching. Here the focus is on unipolar operation in which transitions between resistive states all occur for the same polarity of applied voltage. After an electroforming step (similar to the process discussed in the previous chapter), an initial High Resistive State (HRS) switches to a Low Resistive State (LRS) at a threshold voltage typically around -4V (Figure 4.2(a)) by the reversible formation of a conductive filament within the insulating film.

STM measurements of the diameter of filaments at the top of the oxide layer show them to be in the range of a few nanometres to tens of nanometres (as discuessed in the previous chapter) – well within the range for which quantum effects should be observable.

The filaments are formed by the diffusion of oxygen vacancies under the application of an external electric field between top and bottom electrodes, resulting in a single continuous or semi-continuous conducting filament bridging the two electrodes. Filamentary resistive switching occurs at a weak (narrow) point, which is likely to be close to one or other of the electrodes. Such point presents a quantum constriction, leading to one-dimensional confinement of carriers during transport, and consequent quantisation of conductance.

After switching the device to a LRS a sequence of voltage sweeps is applied to progressively higher voltages, during each of which multiple current jumps were seen, as shown in Figure 4.2(b). These jumps are much smaller than the transition between the HRS and LRS, which can be around five orders of magnitude. The overall conductance slightly increased with each consecutive sweep.

This may be due to a progressive thickening of the conductive filament, as illustrated schematically in Figure 4.2. As the voltage is swept to progressively higher and higher values, more and more oxygen ions are removed from the region surrounding the filament, uncovering a larger volume of conductive tissue. This is illustrated in Figure 4.2(b-I) by thickening of the filament (represented in red).

We gradually increased the maximum voltage bias until no further abrupt changes were detected, corresponding to a fully formed conductive path with no further filament thickening.

During these scans around one thousand abrupt conduction steps were recorded, and the distribution of their conductances plotted in Figure 4.3(b).

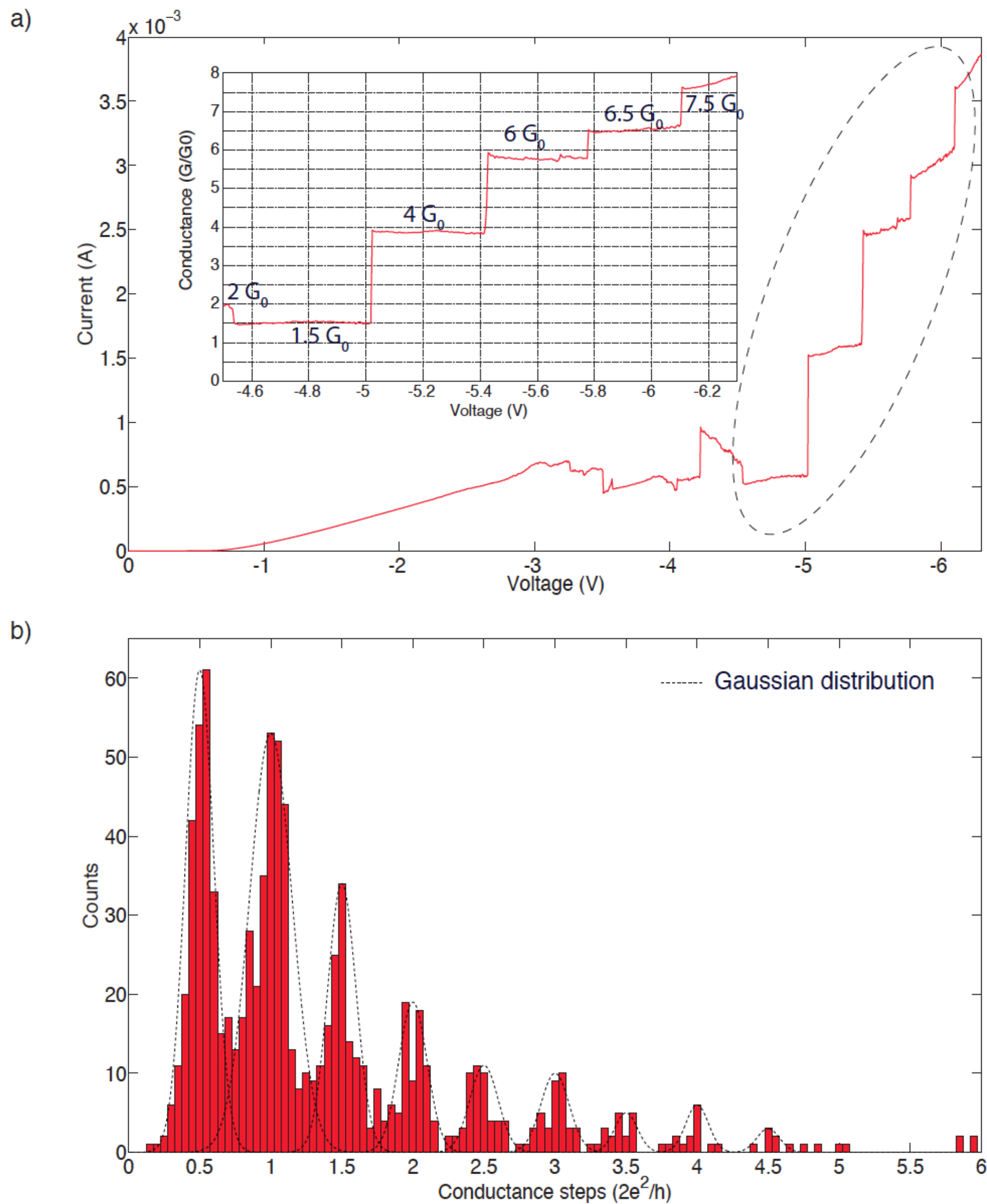

Figure 4.3: Quantised conductance steps. (a) Current-voltage and (inset) conductance-voltage curves for a device showing linear behaviour in the voltage range between -4.5V and -6.1V (expanded in the inset). Several level conductance plateaux can be seen at half-integer multiples of G_0. Note the deviation from linearity in the G-V curve above 6.1V and $G=7.5G_0$. (b) Histogram of conductance changes during ~1,000 conductance steps. Clear peaks are evident at half-integer multiples of G_0, which have been fitted with a series of Gaussian distributions as a guide to the eye (dotted lines).

The histogram can be fitted with a set of normal distributions with peaks positioned at half-integer multiples of the quantum conductance unit G_0, confirming the signature of ballistic carrier transport in a one-dimensional constriction.

Reproducibility of G-V Curves

Seventy individual devices were measured, for each of which a number of I-V curves was recorded. Some curves yielded more steps than others, but data from all curves were aggregated in order to generate the histogram shown in Figure 4.3(b). Example curves from nine different devices are shown in Figure 4.4 to give an indication of the device-to-device variation in conductance jumps. In Figure 4.4 we reproduce G-V curves for nine representative devices. Despite the variation between individual curves, statistical analysis of conductance data from these, and 61 further devices, yields the histogram shown in Figure 4.3(b).

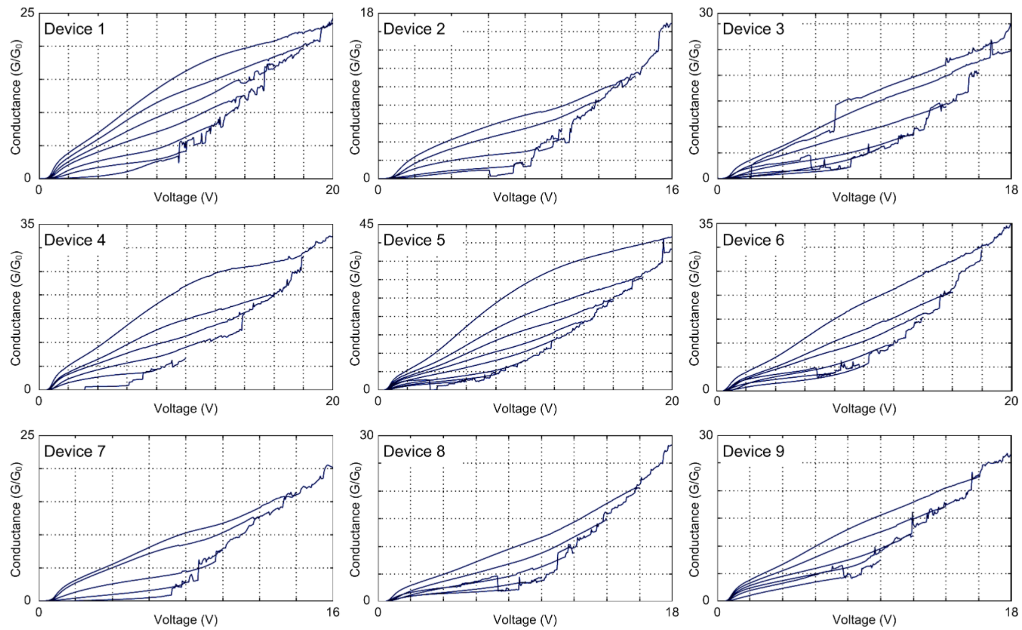

Figure 4.4: Representative conductance-voltage curves from nine different devices. Data for the conductance histogram came from such measurements performed on 70 samples. Despite the variation in individual curves, clear conductance steps are apparent, which, when statistically analysed, exhibit quantised steps.

Nonlinear Components of the Conductance

The nonlinear conductance curves in Figure 4.2(b) suggest contributions from both linear and nonlinear conduction channels, the former of which exhibits quantisation. The degree of linearity of device response depends strongly on the strength of the conductive filament – those supporting higher currents tend to be more linear, in keeping with our previous observations.

In the case of unipolar switching of SiO_x, when sufficient current flows through the filaments their behaviour becomes more Ohmic, suggesting the formation of a highly conductive phase of silicon, or of a mini-band in the silicon oxide band gap associated with dangling bonds or oxygen vacancies; devices become progressively more linear as the voltage is swept to even higher voltages. The observation of a conductive phase of silicon is supported by TEM data from Yao et al.[61], who suggest on the basis of high-resolution Transmission Electron Microscopy studies that local Joule heating can transform the silicon filament into a highly conductive Si-II/Si-XII phase that has a much lower resistivity than that of bulk silicon. The formation of a dangling bond related mini band is supported by results from Wang et al.[62]. Whichever is the case, it is clear that the Ohmic contribution to conductance comes from a conductive material phase whose linearity depends on the past history of applied voltages.

Figure 4.3(a) shows current-voltage (I-V) and conductance-voltage (G-V) results from a highly linear device. In this case the steps in the G-V curve occur clearly at half-integer multiples of G_0 for applied voltages between -4.5V and -6.1V. Above this the curve becomes nonlinear.

These results imply that device conductance may be expressed as a sum of linear and nonlinear terms:

$$I(V) = NG_0V + f(V) \tag{4.1}$$

where f(V) represents a nonlinear background current through semiconducting tissue. This background conduction does not significantly change during successive voltage sweeps. Instead, a set of discrete, quantised jumps in the linear component is observed. We may suppose that the Ohmic contribution to conductance comes from a highly conductive filament core, while the nonlinear component may arise from an interface region around the filament within which phase separation of the conductive channel has not fully progressed.

Figure 4.5 shows two conductance-voltage curves (G_1 and G_2) measured sequentially on the same device. Curve G_1 exhibits an abrupt change in conductance of $5/2G_0$ at a threshold voltage of around 6.6V. The subsequent voltage sweep (G_2) shows the device has moved to a lower resistance state. Note that the both curves are highly nonlinear; we may therefore model the device conductivity as parallel linear and nonlinear components, with the linear component due to a highly conductive filament core, and the nonlinear component due to conduction through semiconducting tissue surrounding the core. A subtraction of the two curves (G_2-G_1) yields a change only in the linear component, as switching will affect only the conductive filament core, and should have little effect on the surrounding semiconducting tissue. In case of Valence Change Memory (VCM) or

Thermochemical Memory (TCM) systems such as that described here, switching occurs at a highly localised region of the conductive filament – it is not a case of the whole filament being disrupted[5]. Instead, only a fraction of the filament at a weak point is altered (red section in Figure 4.2). Such a weak point is likely to be at one of the electrodes. An increasing nonlinear component is also observed for metallic nanocontact break junction devices in which the conductance quantisation is analysed at room temperature[63]. In this case it was reported that the nonlinear component is almost independent of the nanocontact conductance and more pronounced at the higher voltages, much like in our system.

For an ideal system, the result of the G_2-G_1 subtraction (for the results in Figure 4.5(b)) would be a step of magnitude $5/2G_0$. However, at low voltages there is considerable nonlinearity in the G-V curve due to silicon band bending, as has been previously described in the work of J. Suñé at al.[64]. Nevertheless, for voltages between 3.5V and 6.5V there is good agreement with the expected result (Figure 4.5(a) red line), confirming that switching affects primarily the largely Ohmic highly conductive filament core. An equivalent circuit of the overall cell resistance is shown schematically in Figure 4.5(c). It is a parallel of the nonlinear component (R_{nonlin}) that arises from background conduction of semiconducting tissue surrounding the conductive filament core and switching resistance R_{switch}.

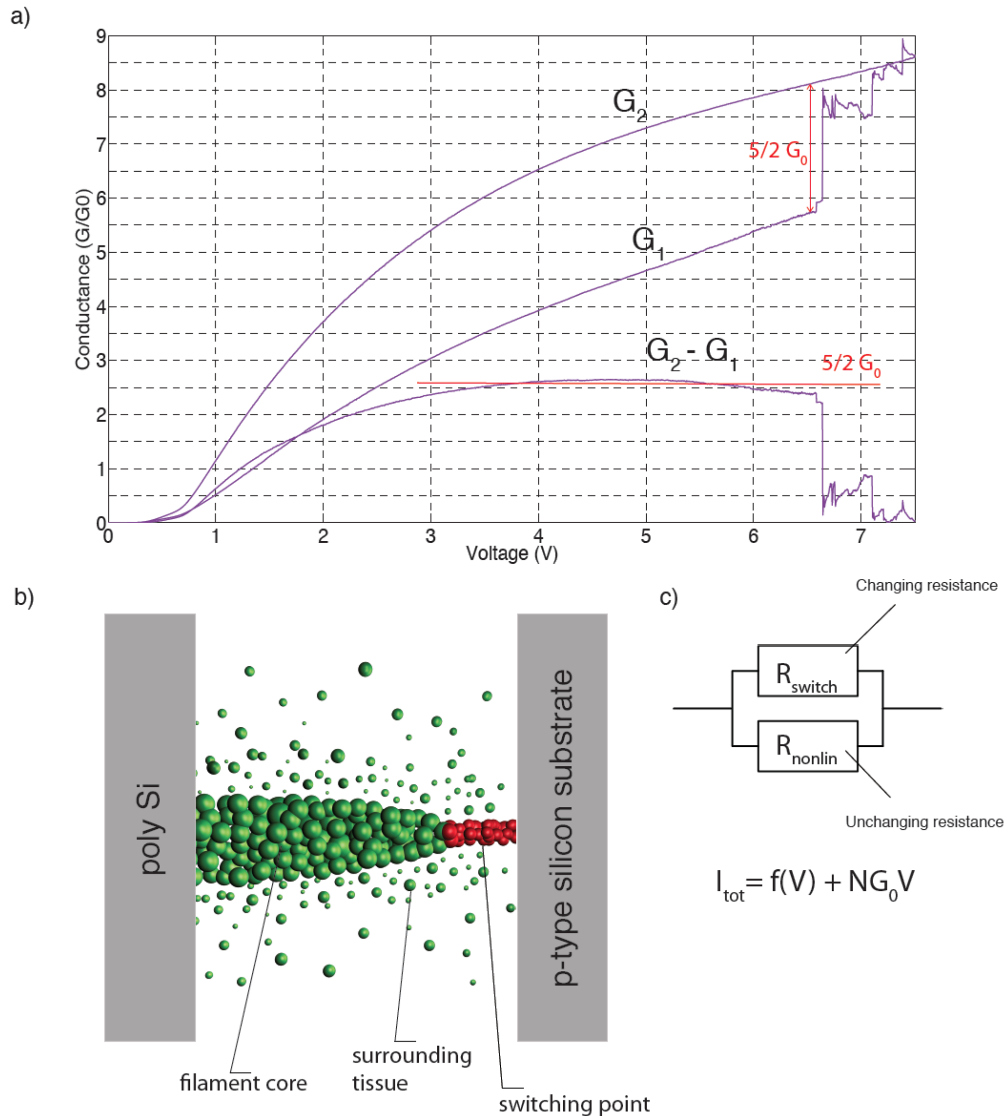

Figure 4.5: (a) Conductance-voltage curves (G_1 and G_2) taken sequentially for a single device, showing an abrupt transition of $5/2G_0$ at around 6.6V. Subtracting these curves produces a step of $5/2G_0$, demonstrating the deconvolution of the linear switching component from the overall nonlinear conductance. (b) Schematic of filament, showing the quantum constriction close to one of the electrodes. R_{switch} is the resistance of the quantum constriction, and R_{nonlin} represents the background nonlinear resistance of semiconducting tissue surrounding the highly conductive filament core. (c) Equivalent electrical circuit of the schematic shown in 4.5(b).

Low Temperature Measurements

To further examine conduction in the ON state low temperature measurements down to 77K were performed. If the overall conduction consists of two components – linear (Ohmic) and non-linear (semiconducting) – the semiconducting component will be significantly decreased at 77K as free carriers are frozen out. Figure 4.6 shows current-voltage curves taken at 77K and at room temper-

ature. In contrast to conduction at the room temperature, that at 77K is largely linear for voltages greater than -2.25V, and the overall conduction given by the gradient corresponds closely to one quantum of conductance (G_0 = 7.75×10^{-5}S); the slightly higher measured value (8.1×10^{-5}S) can be explained by residual free carriers at 77K.

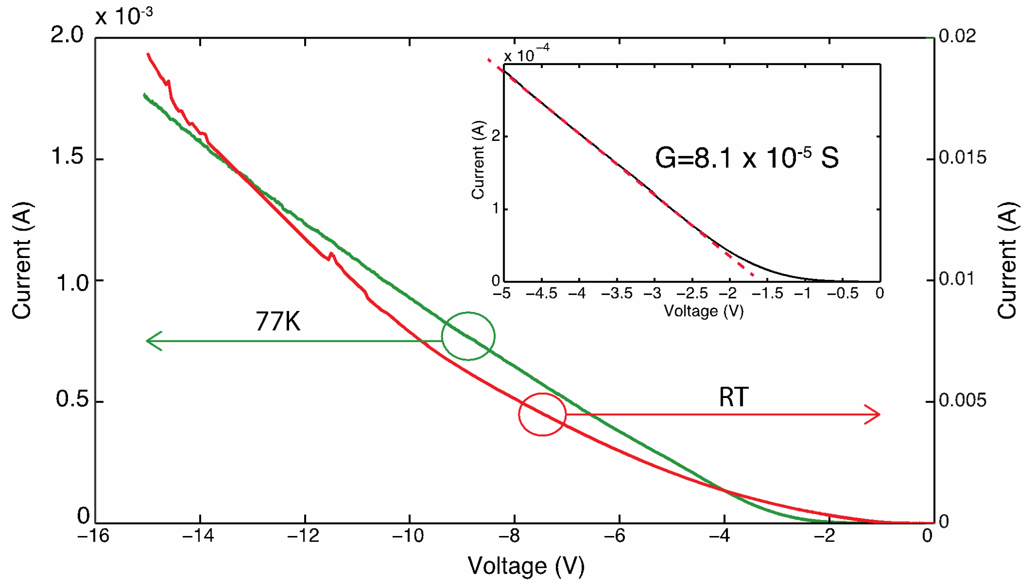

Figure 4.6: Temperature dependence of current-voltage characteristics. Current-voltage curves for a single device measured at room temperature (red line) and at 77K (green line) show a reduction in the nonlinear conduction component at low temperature, consistent with the reduction in free carriers in semiconducting tissue. Inset: the gradient of the 77K curve yields a conductance of 8.1×10^{-5}S, close to G_0 (7.75×10^{-5}S). The small difference can be due to residual conduction through semiconducting tissue.

Note also that when the sample was cooled from room temperature to 77K the overall current measured at a bias of -5V, decreased from 1.87mA to 0.23mA. The overall decrease in the conductivity (by a factor of approximately 8) is too small to be related to strongly temperature-dependent conduction mechanisms such as Poole-Frenkel emission or Thermionic emission where current density J depends on temperature as $J \propto \exp\left(-c/T\right)$, or to be related to conduction through an intrinsic semiconductor in which the concentration of carriers depends on temperature as $n \propto \exp\left(\frac{-E_g}{2kT}\right)$, where E_g is the silicon band gap.

However, if the filament is a highly conductive Si-II or Si-XII phase, as suggested by Yao et al.[61], or a dangling bond related mini band, as suggested by Wang et al.[62], such a result is to be expected as the filament core would behave metallically and have significant conductivity even at low temperatures.

The unipolar switching mechanism in these devices is thermochemical in that Joule heating is required to reset devices, and local temperature plays a role in increasing the mobility of oxygen ions and/or oxygen vacancies during the set process. Consequently, we have not observed set/reset processes or current jumps at low temperatures (e.g. 77K).

Theoretical Model – Nonlinear Quantum-Point Transport Mode

The switching behaviour depicted in the I-V characteristics of the low resistance state (LRS), can be explained within the framework of Landauer theory for mesoscopic systems.

The very high degree of lateral confinement of carriers within the quantum constriction (Figure 4.7(a)) produces a set of discrete one-dimensional sub-bands in the conduction band of the constriction, through which electrons flow.

This is illustrated schematically in Figure 4.7(b). When the spacing between the sub-bands is greater than k_BT and the applied bias is greater than the sub-band energy spacing (and therefore also greater than k_BT), steps are seen in the resulting conductance/voltage characteristics. As the width of the constriction increases, more conduction modes are allowed; for each additional mode the conductance of the device jumps by one unit of G_0.

To account for the observed half-integer G_0 plateaux we modify the preceding argument by considering the nonlinear quantum-point conduction model. We first assume that voltage drops equally at both sides of the quantum constriction shown in Figure 4.7(a). Importantly, a prerequisite for the observation of quantum conductance is for electron transport through the constriction to be ballistic, in which case no voltage is dropped within the constriction. Instead, an applied field generates a difference in chemical potentials between the left and right electron reservoirs. Following the finite-bias Landauer approach, the influx of electrons from each electrode is expressed in terms of the transmission probability of the constriction, the number of available conduction modes (sub-bands) and the Fermi distribution.

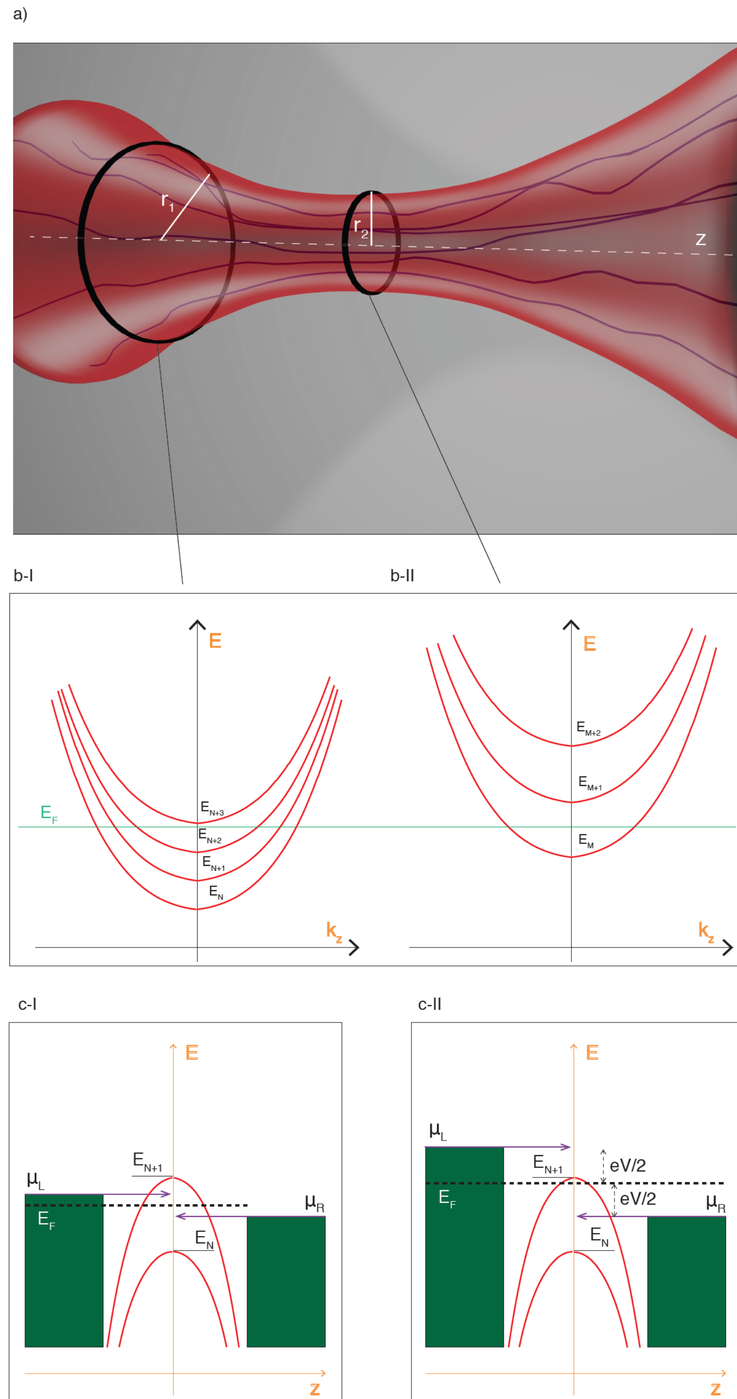

Figure 4.7: The effect of a quantum constriction on conductance. (a) Schematic diagram of the quantum constriction at the heart of the filament. (b-I) Illustration of the dispersion curve of the first four electronic sub-bands at the edge of the constriction. (b-II) The first three sub-bands at the centre of the constriction where the confinement is stronger, causing a spacing-out of the subbands. (c-I) Situation in which the difference in chemical potential between the L and R reservoirs is small, and hence the L- and R-going electron modes both fall within the same sub-band. (c-II) Larger difference in chemical potentials (e.g. a higher bias), causing the L- and R- going modes to fall into different sub-bands.

The current coming from the left electrode (reservoir) is:

$$I_L = \frac{2e}{h} \int_{-\infty}^{eV/2} T(E)N(E)f(E - \frac{1}{2}eV)dE \tag{4.2}$$

where T is the transmission probability, $N(E)$ is the number of sub-bands in the constriction, $f(E)$ is the Fermi distribution and V is the bias voltage. Similarly, the current coming from the right electrode is:

$$I_R = \frac{2e}{h} \int_{-\infty}^{-eV/2} T(E)N(E)f(E + \frac{1}{2}eV)dE \tag{4.3}$$

We approximate the transmission probability for electrons above the energy minimum of the sub-band to one, and that of the electrons with energy below this to zero. Adopting the zero temperature limit, currents from both sides are:

$$I_R = \frac{2e}{h} \int_{-\infty}^{ev/2} \sum_i H(E - E_i)dE = \frac{e^2}{h}N_R V = \frac{1}{2}G_0 N_R V \tag{4.4}$$

$$I_L = \frac{2e}{h} \int_{-\infty}^{-ev/2} \sum_i H(E - E_i)dE = -\frac{e^2}{h}N_L V = -\frac{1}{2}G_0 N_L V \tag{4.5}$$

where N_R and N_L are the numbers of occupied sub-bands accessed from the right and left sides, respectively, and H(E) is the Heaviside unit step function. The total current is:

$$I = I_R - I_L = \frac{1}{2}(N_R + N_L)G_0 V \tag{4.6}$$

If $N_R + N_L$ is an odd number quantised plateaux at half integer multiples of G_0 will be observed rather than the usual integer values. In general, therefore, we can say that systems in which the difference in chemical potentials between the two electron reservoirs is small only integer quantisation of conductance will be seen.

In case of these device, the voltage bias across the device prior to the observation of conductance steps is much higher than that reported in metal oxide based ReRAM devices. In this case the chemical potential of the left reservoir (μ_L) can rise above the energy E_{N+1} while that of the right reservoir (μ_R) will still be between E_N and E_{N+1}, where E_N and E_{N+1} are minimum energies for the Nth and $(N+1)th$ sub-bands.

This is shown in Figure 4.7(c-II). When there is an odd number of conduction modes transporting current in only one direction, half-integer G_0 plateaux are observed. Miranda et al.[65] have provided a similar theoretical treatment for the electron transport in CeO_x-based resistive switching devices. Early reports of the nonlinear quantum point transport mode can be found in work of Glazman and Khaetskii[66] , Xu et al.[67], and Patel et al.[68]

Distinction between the Half-integer and Integer Quantisation RRAM Systems

The model (eq. 4.2 – 4.5) allows us to reconcile the various reports of integer- and half-integer quantisation of conductance observed in RRAM devices in the literature.

Both are reported, with half-integer quantisation being variously ascribed to non-unity transmission probabilities, absorbed and adsorbed defects on metallic filaments, and scattering. There is thus some uncertainty surrounding the origin of half-integer quantisation – an uncertainty that needs to be resolved.

It is clear from the preceding discussion that the key quantity governing the type of quantisation seen is the difference in chemical potentials between the two reservoirs ($\Delta\mu$).

At this point we can consider in more detail the mechanism by which conductive filaments are formed during resistive switching. Broadly, there are two classes of resistive switching devices in which quantisation of conductance is reported: electrochemical metallization (ECM) and valence change memory (VCM) as discussed in Chapter 1. In the former, conductive filaments form by the diffusion into the switching matrix of metal ions from one or other of the metallic electrodes. In the latter, application of an external field to the switching layer induces a redox reaction, resulting in the migration of oxygen vacancies within the active matrix, revealing a filament of conductive tissue connecting the two electrodes. In a sense, we may thus describe ECM switching as extrinsic to the switching medium, while VCM is intrinsic.

In these devices, unipolar switching can be assigned to a third class of mechanism: thermochemical (TCM). In this case, although the initial HRS-LRS transition is driven by a redox reaction similar to that of VCM, the reset process is one of Joule heating – hence thermochemical. Nevertheless, we may still classify the switching as intrinsic.

We now return to the issue of integer and half-integer quantisation of conductance – specifically, how this relates to the filament switching mechanism. Here we note the metallic nature of the filament in ECM devices. In the majority of reports of conductance quantisation in RRAM devices the filament is silver, and in all of these cases integer quantisation is seen. This is consistent with the impossibility of maintaining a large difference in chemical potential between the electron reservoirs in the case of a highly conductive filament. On the other hand, most VCM systems exhibit half-integer quantisation, implying that filaments formed by valence change mechanisms can sustain a large $\Delta\mu$ between the reservoirs. This is to be expected if we consider the formation of VCM filaments to be more stochastic than that of those resulting from ECM.

The latter are driven by the sequential aggregation of metallic cations around an initial nucleation site on one electrode. Field driven diffusion of metal ions results in a continuous metal filament bridging the electrodes. However, in VCM devices the growth of filaments is more uneven, and the resulting conductive pathway is an admixture of metal cations and oxygen vacancies.

This is reflected in the higher resistivity reported for VCM devices – they do not generally require the setting of current limits during switching, in contrast to ECM devices that sustain much higher currents and hence are prone to breakdown if the current is not restricted. Figure 4.8 is a schematic representation of conductive filaments in ECM and VCM switching, emphasising the intermixing of cations and oxygen vacancies in the latter.

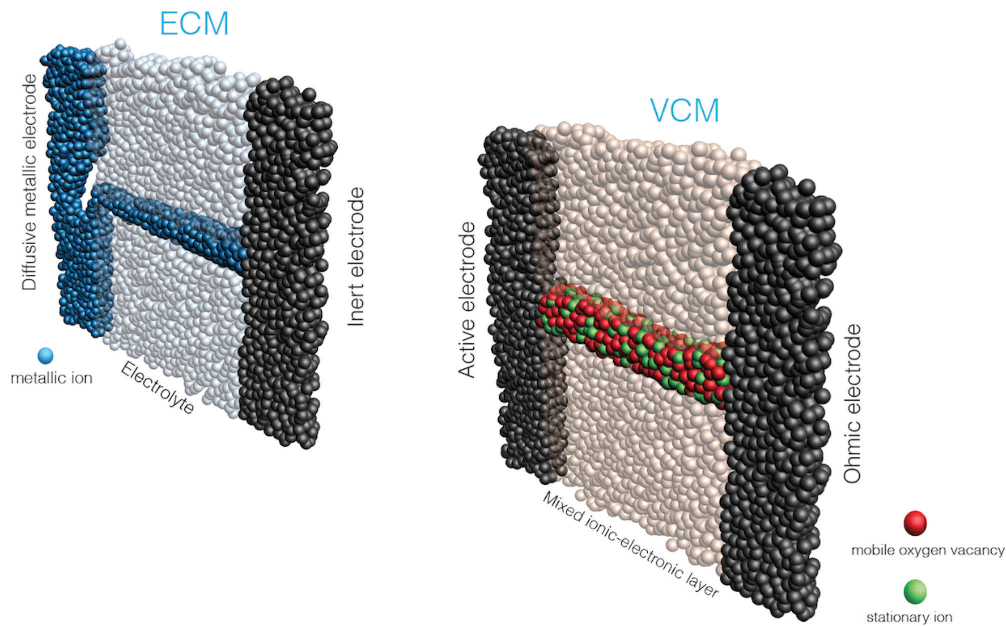

Figure 4.8: Schematic representation of conductive filaments formed in ECM (electrochemical metallisation) and VCM (valence change memory) systems. LHS: in electrochemical memory systems the conductive filament is a continuous metallic path way formed by the diffusion of metal ions from an active electrode (for example, Ag) into the dielectric layer. RHS: in the case of valence change memories, diffusion of oxygen vacancies and associated redox reactions within the dielectric layer result in a more mixed filament, typically with a higher resistivity than that of ECM systems.

Table 4.1 summarises the literature reports of conductance quantisation in resistive switching devices. We note that the early reports of switching in amorphous silicon and in vanadium pentoxide did not specify switching mechanism, and the determination of filament type was unclear in both cases.

Nevertheless, we can surmise that V_2O_5 is a VCM material, in line with other transition metal oxides, which is consistent with the observed half-integer quantisation of conductance. In the case of a-Si:H, although there is undoubtedly some contribution from metallic diffusion from the electrodes, it is likely that this is not the full story, given the complexity of the a-Si:H system. We can reasonably suppose that a more mixed filament is responsible for half-integer quantisation in this case as well, as indeed reported by the authors. In the case of hafnium oxide, the authors show linear conduction with high currents, which they ascribe to the formation of thick metallic filaments. This, again, is consistent with the observed integer quantisation of conductance.

System	Switching	Filament	Quantisation level
metal/a-Si:H/metal [69]	-	?	$\frac{1}{2}G_0$
$V/V_2O_5/V$ [70]	-	?	$\frac{1}{2}G_0$
$Nb/ZnO_x/Pt$ [71]	ECM	Nb or V_0	G_0 or $\frac{1}{2}G_0$
$ITO/ZnO_x/ITO$ [71]	VCM	V_0	$\frac{1}{2}G_0$
$W/CeO_x/SiO_2/NiSi_2$ [65]	VCM	V_0	$\frac{1}{2}G_0$
$Ag/Ta_2O_5/Pt$ [72]	ECM	Ag	G_0
Ag_2S or Cu_2S (vacuum gap) [58]	ECM	Ag	G_0
$Ag/Ag_2S/Pt$ (STM tip) [59]	ECM	Ag	G_0
Pt/AgI/Ag [60]	ECM	Ag	G_0
$Ag/GeS_2/W$ [73]	ECM	Ag	G_0
$Pt/HfO_2/Pt$ [74]	VCM	V_0 or metallic	G_0

Table 4.1: Literature reports of quantised conductance in resistive switching systems, showing the division into those systems that show integer and half-integer G_0 quantisation. V_0 refers to oxygen vacancies.

Conclusion

Studying the quantisation of conductance in filamentary resistive switches offers an insight into the ultimate scaling limits of RRAM, as quantisation can only be seen when the diameter of the narrowest point of the filament (the quantum constriction) is smaller than the Fermi wavelength of the electron in the medium. Assuming the conductive filament in our devices to be silicon, the upper limit on the filament diameter is therefore between 35nm and 112nm. This is consistent with our previous STM results that demonstrated the nanometre-scale size of conductive regions in a switched SiO_x film.

In conclusion, these results demonstrate room temperature quantisation of conductance in silicon oxide resistive switches, implying ballistic transport of electrons through a quantum constriction, associated with an individual silicon filament in the SiO_x bulk.

Because filaments are formed by a valence change process they are less conductive than if they were metallic, and therefore the resulting large $\Delta\mu$ between the electron reservoirs ensures that conductance steps are half-integer multiples of G_0.

This model enables to resolve a puzzle in the literature over the origin of half-integer quantisation. It predicts which resistive switching systems will exhibit integer or half-integer quantisation, the former being principally metallic ECM devices, the latter predominantly VCM with correspondingly higher resistivity.

Furthermore, unlike similar reports of quantisation in metal oxide materials, which are purely linear, in the case of tested devices conductance results can be deconvolved into parallel linear and non-linear terms. We associate these with conduction via conductive silicon nanofilaments in parallel with current flow through semiconducting tissue, perhaps surrounding the conductive nanofilament. These measurements can thus provide useful information on the microscopic structure of the switching layer.

Chapter 5

Phenomenological Model of Resistive Switching in Silicon Oxide

Introduction

In this chapter we propose the phenomenological model that governs the resistive switching. The model is based on the basic physical picture of the main ReRAM switching types – more specifically VCM and TCM systems. Differentiation is made between the bipolar and unipolar switching modes.

Conduction mechanism analysis provides an useful insight of the nature of LRS and HRS. In light of results demonstrated in Chapter 4, a multi-channel conduction model based on quantum point contact theory is suggested.

Conduction Mechanism

The ideal insulator should have zero conductance, but thin layers (order of 10-100 nm) are not ideal and have small conductance. This conductance highly depends on both the temperature and the

applied electric field. An estimate of electric field strength through the insulator is expressed by the basic relation with respect to applied voltage and layer thickness:

$$E_i = \frac{V}{d} \tag{5.1}$$

where V is the applied voltage across the insulator and d is the insulator thickness.

This equation also assumes that the oxide charges are negligible and that the flat-band voltages and the semiconductor band bending are small in compare to the applied voltage. The next table summarizes the basic conduction mechanisms in insulators together with their dependence on the temperature and voltage. This dependence is very useful for the analysis of the experimental results with the aim of specific mechanism identification.

Process	Expression	Voltage and temperature dependance
Tunnelling	$J \propto E_i^2 \exp\left[-\frac{4\sqrt{2m^*}\,(q\phi_B)^{3/2}}{3q\hbar E_i}\right]$	$\propto V^2 \exp\left(-\frac{b}{V}\right)$
Thermionic Emission	$J = A^{**}T^2 \exp\left[-\frac{q\left(\phi_B - \sqrt{qE_i/4\pi\varepsilon_i}\right)}{kT}\right]$	$\propto T^2 \exp\left[\frac{q}{kT}\left(a\sqrt{V} - \phi_B\right)\right]$
Poole-Frenkel Emission	$J \propto E_i \exp\left[-\frac{q\left(\phi_B - \sqrt{qE_i/\pi\varepsilon_i}\right)}{kT}\right]$	$\propto V \exp\left[\frac{q}{kT}\left(2a\sqrt{V} - \phi_B\right)\right]$
Ohmic	$J \propto E_i \exp\left(-\frac{\triangle E_{ac}}{kT}\right)$	$\propto V \exp\left(-\frac{c}{T}\right)$
Ionic Conduction	$J \propto \frac{E_i}{T} \exp\left(-\frac{\triangle E_{ac}}{kT}\right)$	$\propto \frac{V}{T} \exp\left(-\frac{d'}{T}\right)$
Space-charge-limited	$J = \frac{9\varepsilon_i \mu V^2}{8d^3}$	$\propto V^2$

Table 5.1: Basic conduction mechanism in insulators.[75]

- A^{**} – effective Richardson constant (related to thermionic emission)
- Φ_B – barrier height
- E_i – electric field in insulator
- ε_i – insulator permittivity
- m^* – effective mass
- d – insulator thickness
- $\triangle E_{ac}$ – activation energy of electrons

- $\triangle E_{ai}$ – activation energy of ions
- a, b, c, d' – constants

Under high field, tunnelling is the most common mechanism for conduction through insulators. This tunnelling is driven by the quantum mechanical effect that the electron wave function can penetrate through a finite potential barrier. As we can see in the table 5.1 it has a strong dependence on the voltage, but it is also independent of the temperature.

Tunnelling could be direct tunnelling where electron is tunnelling through the complete width of barrier, or Fowler-Nordheim tunnelling where electron tunnels only through a part of the barrier width (i.e. a trapezoid potential).

Thermionic emission (Schottky emission) is the process where flow of charge carriers is induced by heat. This emission occurs when the thermal energy given to carriers is larger than the potential energy of barrier resisting the carrier flow. The Poole-Frenkel mechanism is driven by emission of trapped electrons into the conduction band. Thermal excitation is responsible for the release of electrons from the traps.

Ohmic behaviour typically occurs at high temperatures and low fields. Thermally excited carriers hop from one isolated state to the other. The dependence on temperature is exponential.

For the case of DC ionic conduction conductivity is decreased when an electric field is applied. The reason for this is that the ions cannot readily be injected into or extracted from the insulator. Over time carriers will build up near the interfaces (metal-insulator and conductor-insulator), so the potential distribution will be distorted. When the external field is removed a large internal field still remains, causing the ions to flow back to their equilibrium positions. Evidence for this is the hysteresis behaviour in I-V curves.

Current flow also occurs when the carriers are injected into an insulator that does not contain compensating charges. This current is called space-charge-limited and it is proportional to the square of the applied voltage.

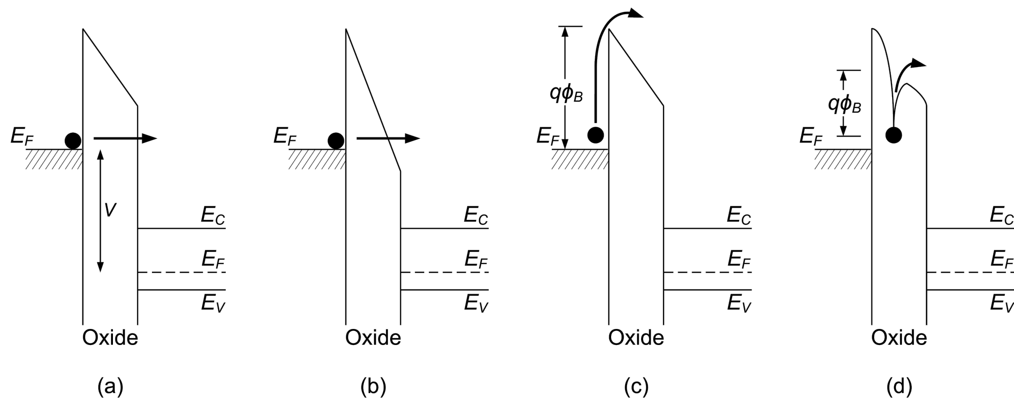

Figure 5.1: Energy-band diagrams showing different conduction mechanism: (a) direct tunnelling, (b) Fowler-Nordheim tunnelling, (c) thermionic emission, (d) Poole-Frenkel emission.[75]

It is important to note that the different conduction mechanisms are not totally independent from each other. It is possible that different conduction mechanisms are present under different conditions (temperature and voltage range) within a single insulator. The most commonly reported conduction processes for ReRAM devices are Poole-Frenkel, Fowler-Nordheim and Trap assisted tunneling (TAT). These three are described in more details in the following section.

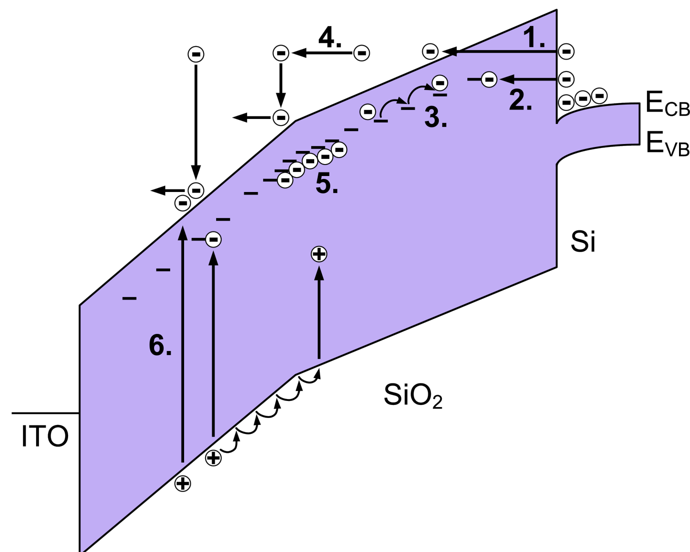

Figure 5.2: Band model of an ITO/SiO2/Si structure demonstrating injection and conduction mechanisms in electroluminescent SiO2. (1) Fowler-Nordheim tunnelling, (2) Trap-assisted tunnelling (TAT), (3) Hopping or Poole-Frenkel conduction, (4) free movement in the SiO2 conduction band including scattering events, (5) charge trapping, (6) impact ionization or trap assisted impact ionization.[76]

Fowler-Nordheim Tunnelling

Tunnelling is the most common conduction mechanism if a high electric field is applied. It is the result of a quantum process in which the electron wave function can penetrate through a potential barrier.

Tunnelling has strong dependence on the applied field, but it is independent of temperature. Generally there are two ways of tunnelling - direct tunnelling and Fowler-Nordheim tunnelling. The difference is that for direct tunnelling, the carriers will tunnel through the complete width of barrier, in contrast to Fowler-Nordheim where the carriers only travel through the partial width of the barrier. Fowler-Nordheim tunnelling relies on a trapezoidal potential barrier, the width of which depends exponentially on the applied filed. It is the most common and dominant conduction mechanism in MOS structures, especially for those with thicker oxide layers. When electrons (or more generally – carriers) get to the insulator they are free to move either in conduction or valence band. The expression connecting the current density and the electric field in the layer is:

$$J_{FN} = C_{FN} E^2 \exp\left(-\frac{4}{3}\frac{\sqrt{2m^*}}{q\hbar}\frac{(q\phi_B)^{3/2}}{E}\right) \tag{5.2}$$

The most convenient way to check this mechanism is to plot experimental I-V characteristic in the form of $ln(\frac{J_{FN}}{E^2})$ versus $\frac{1}{E}$. This representation is called a Fowler-Nordheim plot. Values of the effective mass of the electron are usually known (for SiO_2 it is $m* = 0.42m$) and by fitting the data to a straight line we are able to extract values for the barrier height. The simple condition for the tunnelling is that the product of electric field and layer thickness is larger than the barrier height.

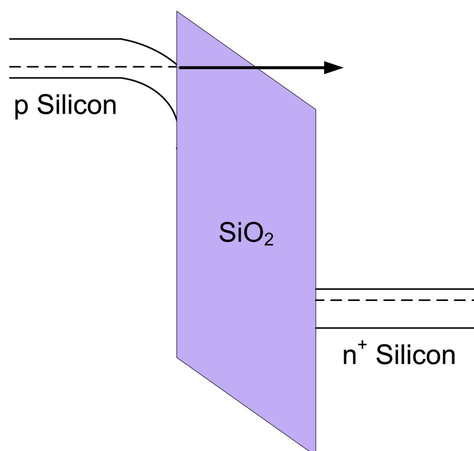

Figure 5.3: Fowler-Nordheim tunnelling across the MOS structure.

Poole-Frenkel Conduction Mechanism

It is possible to conduct carriers through insulators without tunnelling. In semiconductors with a high density of structural defects (often the case in deposited insulators) carriers are not free to move as is the case with the Fowler-Nordheim conduction. These structural defects cause additional energy states near the band edges (traps). Traps are responsible for the capture and emission processes, which restrict current flow. The Poole-Frenkel conduction mechanism treats electrons that are trapped in localized states. The electrons can eventually gather enough energy through thermal fluctuations to be able to leave the localized state and move to the conduction band. If this happens, the electrons will travel through the material as long as they do not get stuck in another localized state. Normally the rate of this excitation of electrons to the conduction band is not high at room temperature, but if a high electric field is applied, electrons would not need as much thermal energy to be excited. The total energy is obtained from both the thermal energy and the applied electric field. The conduction mechanism is driven by the standard expression for the drift current:

$$J = qn\mu E \tag{5.3}$$

where q is a charge of an electron, n is the number of electrons per cubic centimeter, μ is mobility of electrons and E is an applied electric field.

The difference is that the carrier density is dependent on the depth of the trap and it is related to the applied electric field as:

$$n = n_0 \exp\left[-\frac{q}{kT}\left(\phi_B - \sqrt{\frac{qE}{\pi\varepsilon}}\right)\right] \tag{5.4}$$

This will give the final expression relating the current density and the electric field.

$$J_{PF} = qn_0\mu E \exp\left[-\frac{q}{kT}\left(\phi_B - \sqrt{\frac{qE}{\pi\varepsilon}}\right)\right] \tag{5.5}$$

where ε is a dielectric constant of the material.

A modification of this formula could be made taking into account the non-ideality factor. This modification involves the constant m that can take values between 1 and 2. This factor is similar to the ideality factor that appears in the case of the diode current. It depends on the fabrication

process and semiconductor material.

$$J_{PF} = qn_0\mu E \exp\left[-\frac{q}{kT}\left(\phi_B - \sqrt{\frac{qE}{m^2\pi\varepsilon}}\right)\right] \tag{5.6}$$

ds

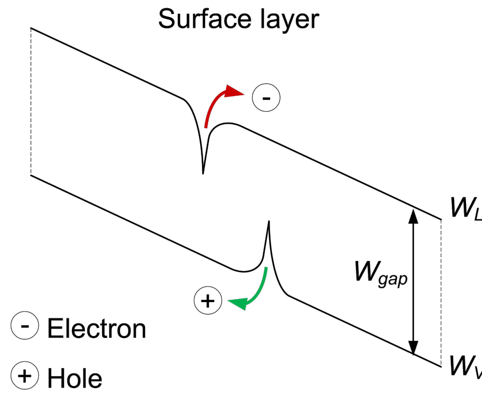

Figure 5.4: Poole-Frenkel conduction mechanism – lowering of the potential barrier due to the applied electric field.

Trap-assisted Tunnelling

Fowler-Nordheim tunnelling is based on a one-step tunnelling process. However, it is possible that defects inside the insulator layer allow the tunnelling processes to consist of two or even more steps. This process usually happens when the traps are introduced by stress. The stress can be caused by high electric fields applied across the layer (as in case with the erasing and writing cycles in EEPROMs). Quantum tunnelling is exponentially dependent on the barrier width. This means that the probability for carriers to go over the energy barrier is lower for the thicker samples. Nevertheless, in the presence of traps this energy barrier could be split in two paths. Carriers would sequentially tunnel across the thinner barriers, and thus the probability of conduction would be greater. Generally the TAT process could go in two ways – elastic and inelastic TAT process (meaning with and without energy loss). In highly stressed materials it is common to have more than one oxide trap which will lead to more than a two step process.

Several theoretical models have been proposed for the trap assisted tunnelling. A simplified TAT expression relating tunnelling current density with trap barrier height has been proposed and shown

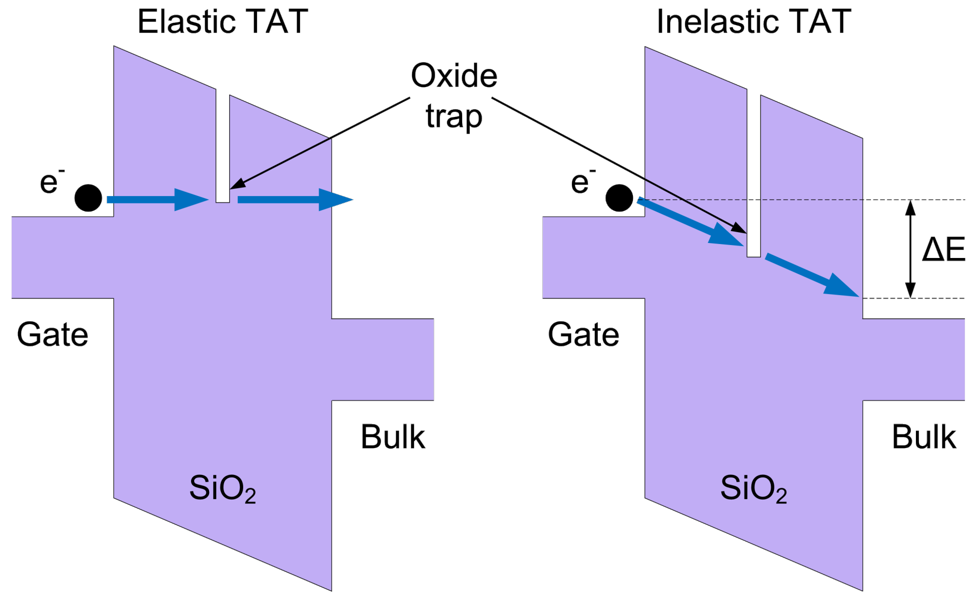

Figure 5.5: Energy band diagrams showing the two cases of Trap Assisted Tunnelling – elastic/without energy loss and inelastic/with energy loss.

to be valid at the high electric fields[77]:

$$J_S \propto \exp\left[-\frac{8\pi\sqrt{2qm^*}}{3hE}\Phi_t^{\frac{3}{2}}\right] \tag{5.7}$$

where $m^* = 0.42$ is the electron effective mass, h Planck's constant and Φ_t the trap barrier height. The value for the trap barrier height can be extracted from the above equation.

Conduction Mechanism Results

To better understand the switching mechanisms and differences between the two switching modes (bipolar and unipolar) in our devices we performed an analysis of the conduction mechanism in both modes.

The conduction mechanism in bipolar switching mode is analysed first. The biploar switching is shown in Figure 5.6. Results are obtained from the devices of the 1st generation.

Most probable conduction mechanisms (most commonly reported) are: Ohmic conduction (only in the case of LRS), Poole-Frenkel emission, Schottky emission and Trap Assisted Tunnelling. Other possible conduction mechanisms such as direct tunnelling do not show good fit with our results.

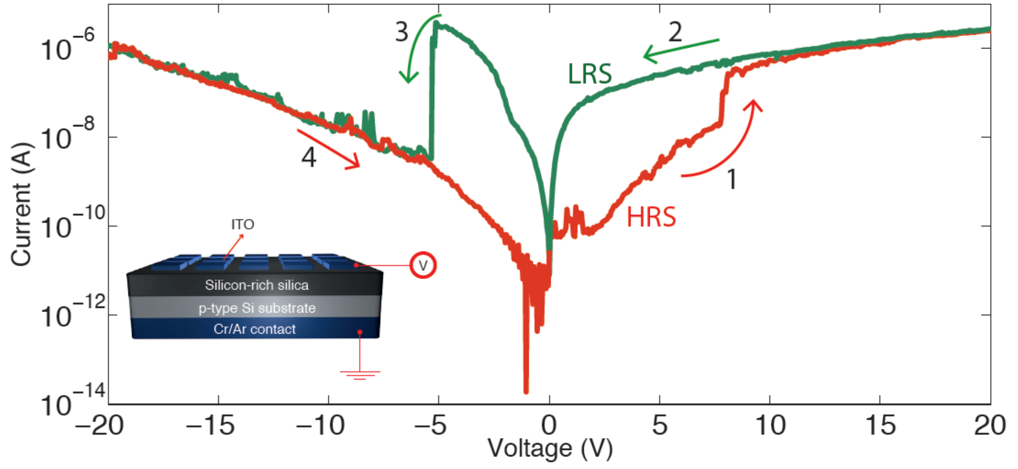

Figure 5.6: I-V characteristics. Arrows indicate the direction of voltage cycling. Inset: Device schematic. Obtained from a device from the 1st generation.

Fowler-Nordheim tunnelling does provide some reasonable fits but only in the much higher electric field regimes (near the hard breakdown threshold voltage – not shown here). It is also aparent that the curves are highly non-linear, which discounts Ohmic conduction.

Poole-Frenkel emission is field enhanced thermal emission from trap to trap in the oxide material.

$$J_{PF} \propto E \exp\left[-\frac{q}{kT}\left(\phi_B - \sqrt{\frac{qE}{\pi\varepsilon^2}}\right)\right] \tag{5.8}$$

where J_{PF} is current density, E is electric field across oxide material, k is Boltzmann constant, T is temperature in Kelvins, Φ_B is barrier height, q is electron charge and ε is the dielectric constant.

Schottky emission represents thermal excitation of electrons over the potential barrier to the conduction band.

$$J_S \propto \exp\left[-\frac{q}{kT}\left(\phi_B - \sqrt{\frac{qE}{4\pi\varepsilon^2}}\right)\right] \tag{5.9}$$

The third conduction mechanism we explore is Trap Assisted Tunnelling. It is possible that inside the insulator layer the tunnelling process comprises two or even more steps (through two or more defects). Silicon nanoinclusions could take the role of defects in this case. Inclusions may be formed in the silicon oxide matrix after the phase separation induced by the post annealing treatment.

$$J_S \propto \exp\left[-\frac{8\pi\sqrt{2qm^*}}{3hE}\Phi_t^{\frac{3}{2}}\right] \tag{5.10}$$

where m^* is electron effective mass, h is Planck constant and Φ_t is trap barrier height.

Using the above relations it is possible to extract the values for relative dielectric constant in the case of Poole-Frenkel/Schottky fit and the values for trap barrier height in the case of Trap Assisted Tunnelling. Figure 5.7 provides fits for all three conduction mechanisms in both LRS and HRS.

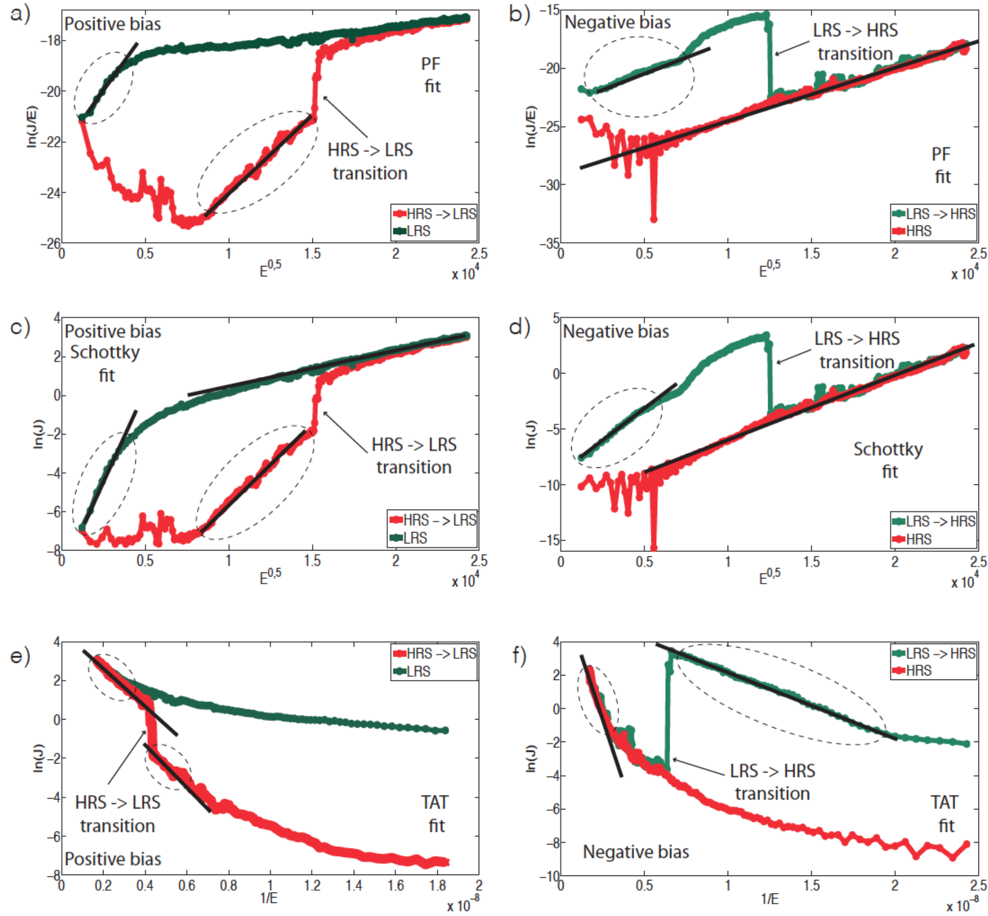

Figure 5.7: Data (red and green) and fitting lines (black). (a) Poole-Frenkel fit for positive bias, (b) Poole-Frenkel fit for negative bias, (c) Schottky fit for positive bias, (d) Schottky fit for negative bias, (e) Trap Assisted Tunnelling fit for positive bias, (f) Trap Assisted Tunnelling fit for negative bias.

Table 5.2 shows the fitting results obtained for each fit in two different states (LRS and HRS) and two polarities. In the case of Poole-Frenkel conduction mechanism and Schottky emission the dielectric constant is extracted, while in the case of trap assisted tunnelling the trap barrier height is extracted.

Notwithstanding that a reasonable curve fit can be found for the Poole-Frenkel conduction mechanism, the obtained values for relative dielectric constant are much higher than those theoretically expected for both silicon dioxide (dielectric constant 4) and pure silicon (dielectric constant 12). This suggests that Poole-Frenkel hopping can possibly happen but with some difficulty or in combination

	Bias	HRS	LRS
Poole-Frenkel hopping	Positive	$\varepsilon_R = 23$	$\varepsilon_R = 6.3$
	Negative	$\varepsilon_R = 42$	$\varepsilon_R = 21.5$
Schottky Emission	Positive	$\varepsilon_R = 3.7$	$\varepsilon_R = 53.3(0.46)$ high field (low field)
	Negative	$\varepsilon_R = 7.3$	$\varepsilon_R = 0.59$
Trap Assisted Tunnelling	Positive	$\theta_t = 0.40eV$	$\theta_t = 0.42eV$
	Negative	$\theta_t = 0.85eV$	$\theta_t = 0.22eV$

Table 5.2: Fitting results of conduction mechanism analysis.

with some other conduction mechanism.

The Schottky emission fit provides much closer dielectric constant values to those theoretically estimated. For the HRS the value of 3.7 is close to the known value of 4 for silicon dioxide, which suggests that in HRS the conduction is mainly governed by direct emission from electrodes to the conduction band of silicon oxide.

Trap assisted tunnelling shows very good fits for the LRS state. The obtained values of 0.22eV and 0.42eV for the trap barrier heights are similar to those reported in DiMaria's work[78] (0.4eV). For the HRS, a good fit is obtained under the high electric field. High field allows tunnelling across a much thicker oxide layer, which is equivalent to a higher trap barrier height. The obtained value is 0.85eV. According to these results we postulate the following conduction mechanism. In the LRS, the silicon nanoinclusions are connected by conduction paths formed after the removal of oxygen. In the HRS after the rupture of the conduction path, the main conduction mechanism is thermal excitation of electrons over the interface barrier from contacts to silicon oxide conduction band – Schottky emission. This can occur in parallel with Poole-Frenkel emission. Trap assisted tunnelling can also play an important role in the HRS under stronger fields. A schematic of this mechanism, together with energy band diagrams is shown in Figure 5.8.

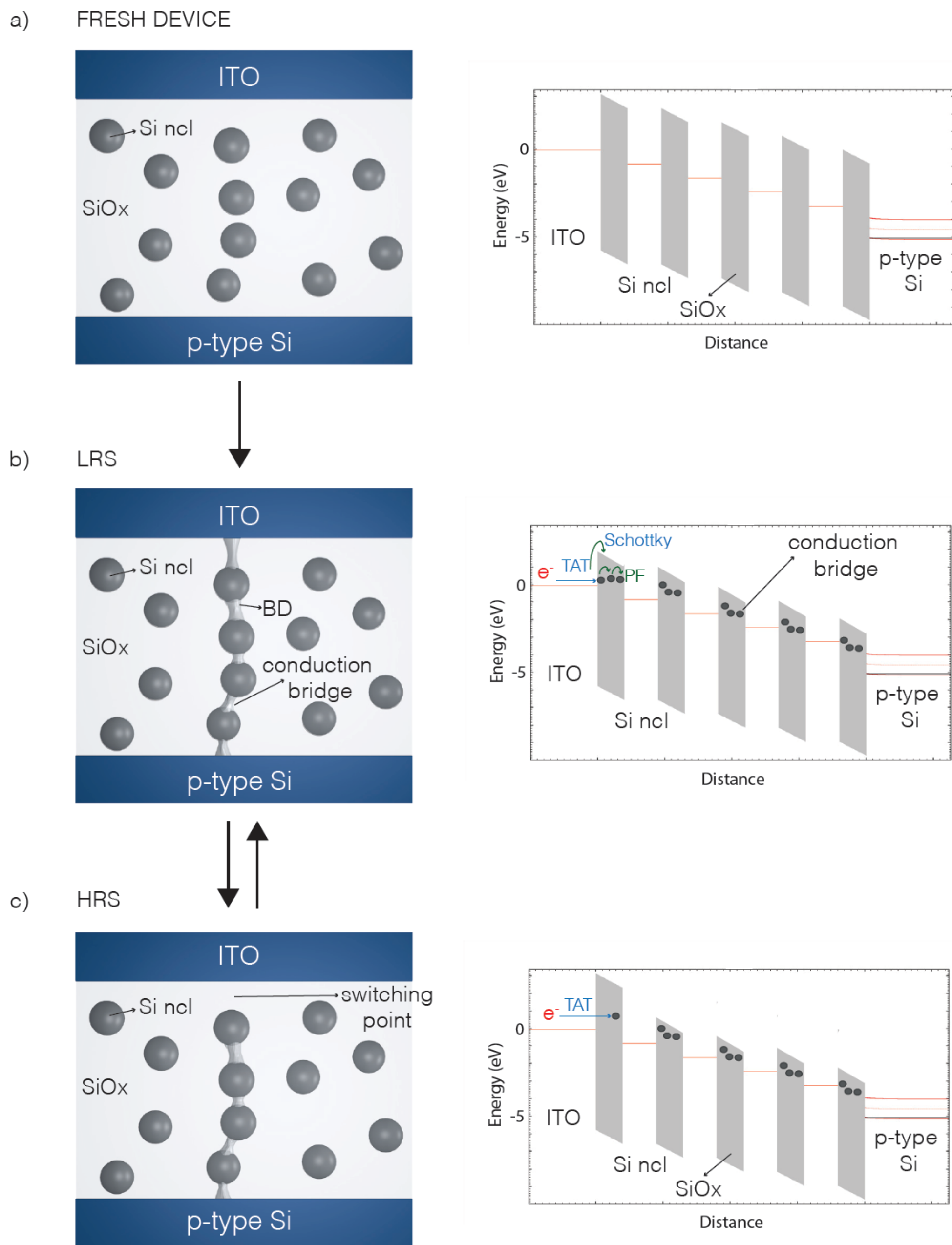

Figure 5.8: Schematic (left) and energy band diagram (right). (a) Fresh device, (b) LRS and (c) HRS.

As noted in Chapter 3, in addition to the abrupt changes in device resistance during bipolar switching, by sequentially sweeping the device to positive voltages with a current compliance limit set, we can also observe the more gradual change shown in Figure 5.9. In this case, three distinct I-V curves were obtained and analysed. The inset of Figure 5.9(a) clearly shows that the three curves are non-linear, which discounts Ohmic conduction. Other conduction mechanisms, such as Poole-Frenkel

and hopping do not provide a good fit. The obtained values from Poole-Frankel fit and equation, for the dielectric constants, are 59, 111, and 254 for the first, the second and the third sweep respectively. Again these values are drastically larger than the known values for dielectric constants of both silicon dioxide and pure silicon (4 and 12).

The trap assisted tunnelling model best describes carrier transport in the case of bipolar switching. Similar behaviour is observed in hafnium-oxide ReRAM devices[79]. Our physical interpretation of this result is that carrier transport occurs via tunnelling between silicon nanoinclusions within the SiO_xlayer, the silicon nanoinclusions taking the role of traps. It is possible to extract the value for trap barrier height from equation.

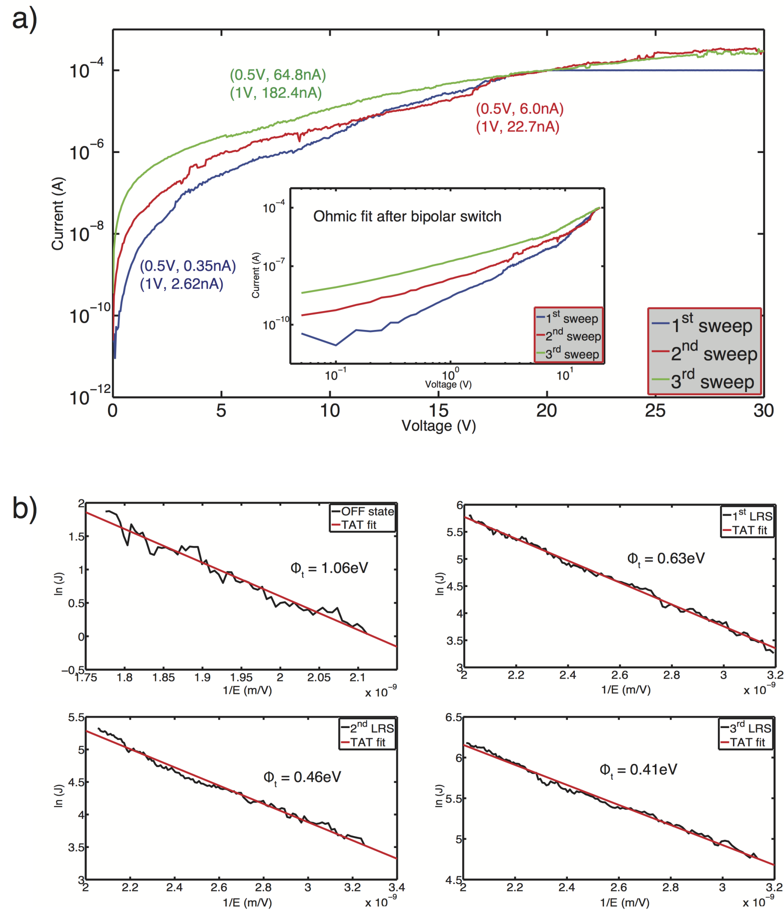

Figure 5.9: (a) Three different I-V curves obtained after three sequential sweeps. (b) TAT fits in the high resistive state and different low resistive states after gradual bipolar conduction increase.

The fitting curves for the initial HRS and the three consecutive sweeps in the LRS are shown in Figure 5.9. The extracted values for trap barrier heights are 1.06eV, 0.63eV, 0.46eV, and 0.41eV for the HRS and the first, second, and third LRS, respectively.

Figure 5.10(a) shows a gradual increase of barrier height as a function of device resistance. In the case of the unipolar switching mode, the LRS conduction can be fitted with either an Ohmic model (Figure 5.10(b)) or trap assisted tunnelling with a much lower trap depth of 0.11eV; I-V results are not sufficient to distinguish between the two mechanisms. HRS conduction after the unipolar switch again shows the best fit with the trap assisted tunnelling model.

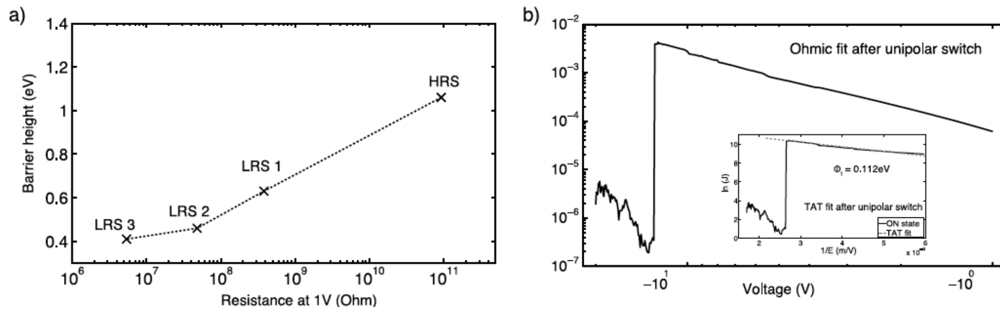

Figure 5.10: (a) Gradual increase of barrier height as a function of device resistance, (b) Ohmic and TAT (inset) fits in LRS after unipolar switch.

These results suggest that, in the case of unipolar switching, the conductive filament is more continuous (the inter-inclusion regions are far more silicon-rich), and might, in the case of Ohmic conduction, be formed out of a continuous metallic phase of silicon, as recently reported by in situ TEM results of conductive Si filaments at the surface of SiO_x[61].

Switching Model

We note that silicon-rich silica is a metastable material that readily segregates into silicon and silicon dioxide[80, 81]. It contains a high concentration of oxygen vacancies[82], which can be driven by high temperature annealing and consequent diffusion of vacancies and silicon to form silicon nanoclusters. In the initial stages of annealing, sub-nanometre clusters nucleate at oxygen vacancies[83], increasing in number and growing by Ostwald ripening at longer times and/or higher temperatures as silicon diffuses.

We propose that voltages applied across our devices drive oxygen vacancy migration, producing a field-driven phase separation of the active layer[84]. This separation is enhanced by structural defects such as nano-scale cracks or inclusions[81]. Significantly (as described in Chapter 3), films grown by sputtering typically exhibit columnar or granular growth[85]. Boundaries between adjacent growth columns constitute structural defects that can nucleate phase separation. The rate of nucleation, as

well as the number density and size of silicon nanoinclusions produced by the migration of oxygen vacancies will increase with applied field until a critical point, at which percolation pathways can be formed along the column boundaries (corresponding to the set process). Current transport in case of bipolar switching mode is dominated by trap assisted tunnelling, suggesting that conductive pathways through our material are not continuous filaments, but instead a sequence of separate but neighbouring silicon nanoinclusions (analogous to aggregated oxygen vacancies).

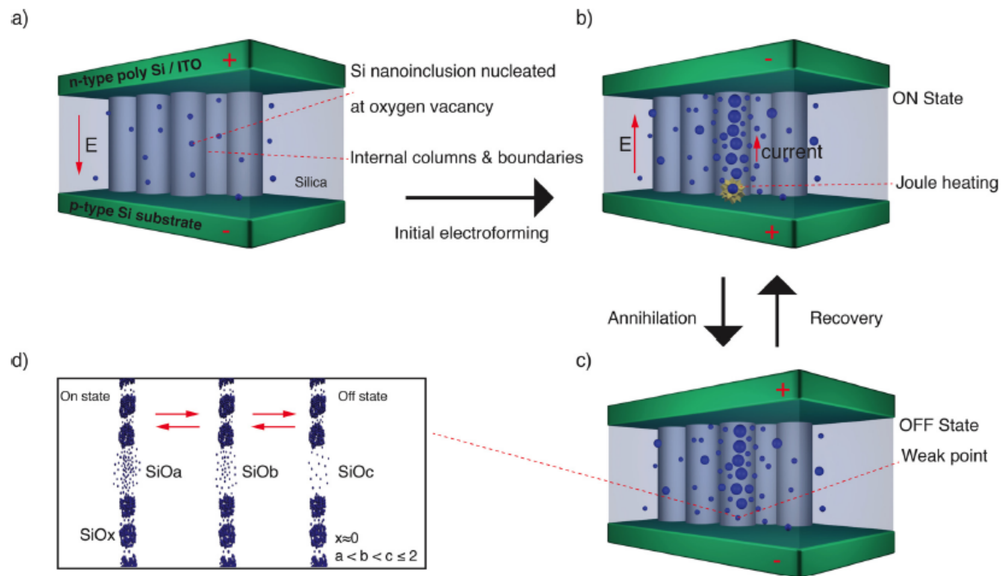

Figure 5.11: Schematic of one cycle process. (a) Pristine state before applying the electric field, showing as-grown silicon nanoinclusions nucleated at oxygen vacancy sites. (b) LRS after the chain/filament formation, showing extra silicon nanoinclusions produced by field-driven migration of oxygen vacancies. (c) Annihilation process due to electric field (in case of bipolar mode) or Joule heating (in case of unipolar mode). (d) Silicon and oxygen distribution at the weak point.

We emphasise that the material is a sub-oxide of silicon with a smaller band gap than stoichiometric SiO_2, resulting in a lower barrier height – intermediate between the conduction band offset of the Si/SiO_2 interface (3.3eV) and zero. Bearing in mind that we consider the conductive pathway as being formed of a chain of oxygen vacancies and silicon nanoinclusions, the larger trap depth in HRS can be ascribed to a reduction in the stoichiometry of the sub-oxide gaps in the conductive pathway.

The low barrier height of traps (found from the TAT fit) is typical for conduction through oxygen vacancy defects. Such a low value is understandable if we assume that transport is through a semi-continuous array of adjacent silicon nanoinclusions in a sub-oxide matrix. The stoichiometry of the interinclusion matrix varies with applied field and Joule heating. The proposed process is shown schematically in Figure 5.11.

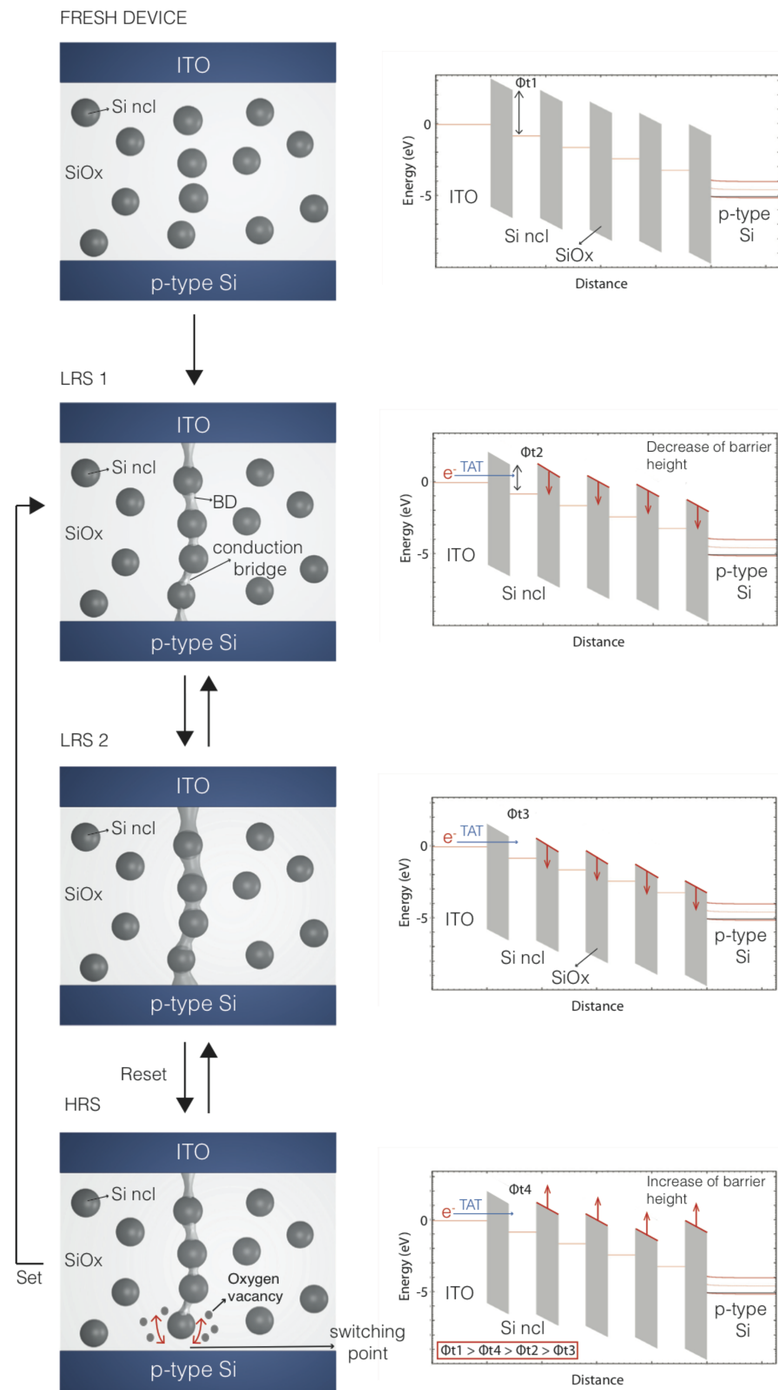

Figure 5.12: Schematic of filament growth and gradual lowering and modification of barrier height.

Although two switching modes (bipolar and unipolar) are likely have a common root, we still make a few distinctions. This is also suggested by the different conduction mechanism for the two modes.

The embedding of more and more oxygen vacancies in the conductive filament causes a gradual

decrease in the trap depth with consecutive voltage sweeps. The main driving force in bipolar switching is the drift of oxygen vacancies under the influence of the external electric field to form or dissolve the conductive filament. This process is schematically shown in Figure 5.12(a).

On the other hand, the LRS conduction taking place after the unipolar switch can be fitted with an Ohmic model (Figure 5.10(b)) or trap-assisted tunnelling with a much lower trap depth of 0.11eV. These results suggest that, in the case of unipolar switching, the conductive filament is more continuous, and might, in the case of Ohmic conduction, be formed out of the metallic phase of silicon reported by in-situ TEM results of conductive Si filaments at the surface of SiO_x[61]. This is a thermal effect as the consequence of the higher current passing during the unipolar switching mode. The difference between the two switching modes is shown in Figure 5.13.

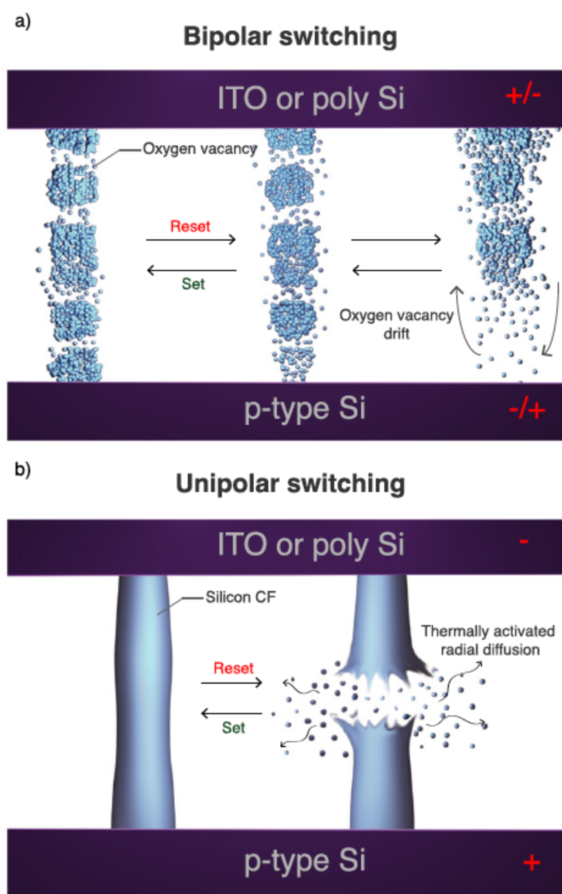

Figure 5.13: Schematic representation of (a) bipolar switching by the vertical drift of oxygen vacancies under the applied field, and (b) unipolar switching, including thermally activated conductive filament dissolution due to radial diffusion.

We note that full unipolar switching has not been seen for all devices studied. We relate this with the position of the weak (switching) point.

Rather than the conductive filament being semi-continuous through the film, we suppose that there is a region located close to the top or bottom of the film that is depleted of silicon nanoinclusions. Such structure has been reported in metal oxide films containing conductive nanofilaments in which the initial electroforming step generates an excess of vacancies close to the anode[86]. As a result, a highly resistive region forms close to the interface. As shown in Figure 5.14, two configurations of the high resistance state are possible in this case, dependent on the polarity of the electroforming step. Configuration 1 shows a switching point near the ITO/n-type poly silicon – SiO_x interface; configuration 2 shows it near the SiO_x– substrate interface. Arrows show the movement direction of oxygen ions under the applied field. In the first configuration the set process occurs more easily in the negative bias as oxygen ions move downward and oxygen vacancies (silicon nanoinclusions) are left in the insulating (switching) gap, making it more conductive. Similarly, for the second configuration the positive bias will more effectively initiate the set process. It is known that the silicon/silica interface is more favorable for the formation of oxygen vacancies[87] and thus more porous – consequently, in our system the second configuration is more plausible.

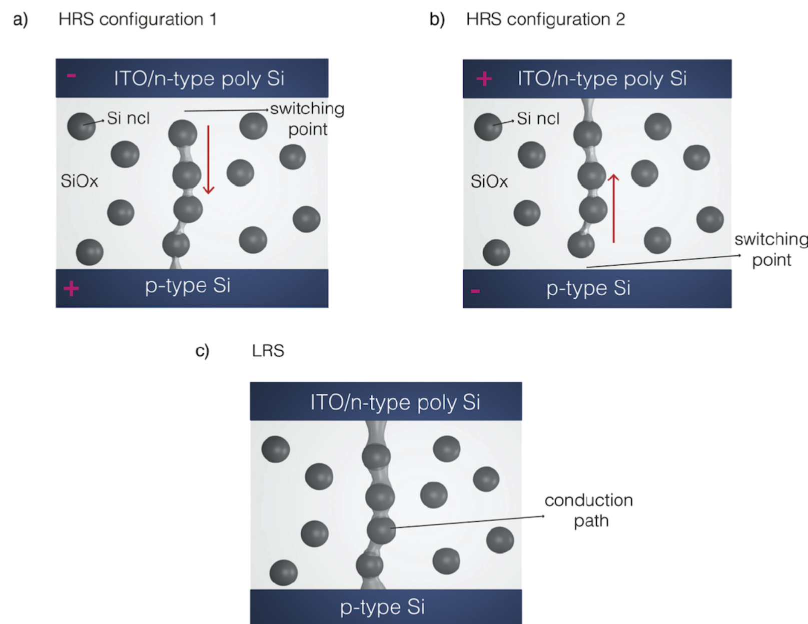

Figure 5.14: Different positions of switching point due to depletion of oxygen vacancies close to one interface (a) high resistive state: switching point near ITO/poly-Si–SiO_x interface; (b) high resistive state: switching point near SiO_x substrate interface; (c) low resistive state.

We note that previous reports of resistive switching in SiO_x have proposed a redox process[87], though at the time little direct supporting evidence was available. However, it is well known that oxygen vacancies can be highly mobile in oxide matrices, and recent detailed work on resistive switching in metal oxides proposes a key role for oxygen vacancies in the formation of conductive

filaments. Although Joule heating is the main driving force for the reset process in the case of unipolar switching, the electric field undoubtedly plays an important role, especially in the case of bipolar switching. As the oxygen vacancies behave as positively charged species, vacancies will be pushed toward the substrate during the set process (positive bias), recovering the broken path. In the reset process (negative bias) energy from the high local Joule heating overcomes the binding energy of vacancies, and the electric field consequently pushes them toward the top electrode, rupturing the conduction path.

Multi-channel Conduction Using Quantum Point Contact Theory

As discussed in Chapter 4, the overall conduction can be divided into two parts – a linear part that comes from the filament's core, and a nonlinear part that comes from background semiconductor tissue. The next model was developed by Prof Enrique Miranda, from the Universitat Autònoma de Barcelona, Spain. In collaboration with our group and me, Prof Miranda has expanded the model to account the main ideas presented in the Chapter 4. The main change with respect to the model developed in [88] is that it accounts for the two components of the conductive filament.

I provided the experimental data and Prof Miranda has tested his hypothesis by fitting the data. First, Prof Miranda's model will be described and then the fitting results will be presented.

Model Description and Fitting Results

A simple analytic model for electron transport through filamentary-type structures is proposed. The model is based on a mesoscopic description and is able to account for the linear and nonlinear components of conductance that arise from both fully and partially formed conductive channels spanning the dielectric film. Channels are represented by arrays of identical scatterers whose number and quantum transmission properties determine the current magnitude in the low and high resistance states. Proposed model reproduces both the experimental current-voltage (I-V) characteristics and also the normalized differential conductance $\frac{dln(I)}{dln(V)} - V$ curves of devices under test.

Several reports [89, 71, 90, 91] have pointed out that the filaments exhibit conductance values close to the quantum conductance unit $G_0 = \frac{2e^2}{h}$ (or $\frac{G_0}{2}$), where e is the electron charge and h is Planck's constant. This is a well-known feature of atomic-sized constrictions[92]. This phenomenon is also demonstrated in Chapter 4.

As described in Chapter 4, the conduction can be written as $I(V) = NG_0V + f(V)$, where N is an integer or half-integer number and $f(V)$ a nonlinear function of the applied voltage. While the constant conductance term is attributed to a highly conductive filament core (see Figure 5.15(a)), $f(V)$ is assigned to electron transport through a region around the core in which the conductive channel has not fully formed (see Figure 5.15(b)). In Prof Miranda's model, $f(V)$ is consistent with these ideas, with the support of experimental results from the devices from the 4th generation.

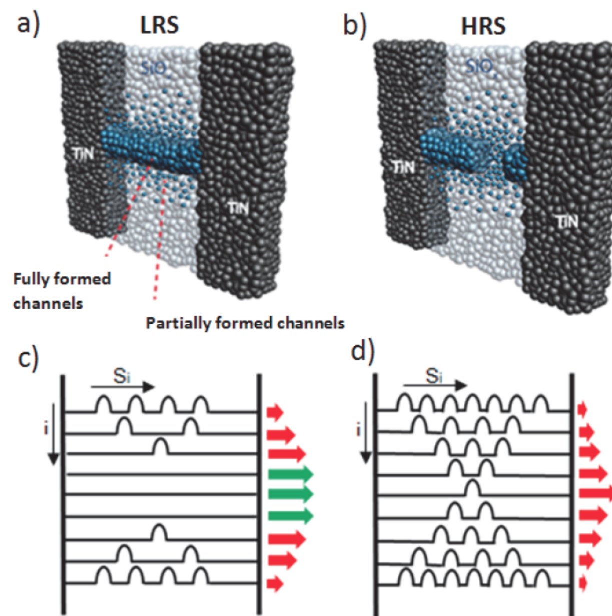

Figure 5.15: a) Schematic showing a filamentary leakage current path in the low resistance state (LRS) of the device composed of fully (core region) and partially (cloud of defects or vacancies) formed channels. b) Schematic showing a break in the filament within the oxide matrix, corresponding to the high resistive state (HRS) of the device. c) and d) Arrangement of scatterers along the electron transport channels associated with cases a) and b), respectively. The green and red arrows represent the transmission probability of fully- and partially-formed channels, respectively. Si is the number of scatterers along the channel i.

The devices were initially electroformed with a maximum voltage ramp to around 7-8 V. The subsequent unipolar set process was typically a single abrupt jump in current at voltages around 3-4 V. This is shown in Chapter 3, in Figure 3.10.

Using the finite-bias Landauer approach for mesoscopic systems[92] and the additive property of the transmission probability for a series combination of identical scatterers[93], the current that flows through a multichannel conductive structure (see Figures 5.15(c) and 5.15(d)) can be expressed as:

$$I(V) = \frac{2e}{h}\sum_i \int_{-\infty}^{\infty} \frac{[f(E - eV_i/2) - f(E + eV_i/2)]}{1 + S_i[(1 - t(E)/t(E))]} \tag{5.11}$$

where the sum runs over the total number of channels that constitute the filament; E is the energy; f the Fermi function; S_i the number of scatterers along the channel i; V_i the voltage drop across the channel i, and $t(E)$ the transmission probability for an individual scatterer.

In this model, which simulates the filament as a bunch of monomode nanowires, two types of channels coexist: fully and partially formed. By a fully formed channel we mean a perfectly transparent ($t(E) = 1$) conduction channel with $S_i = 0$, whereas by a partially formed channel we mean a nanosized bridge with $S_i > 0$, and therefore $t(E) < 1$.

This approach is also compatible with the presence of isolated channels located far from the main constriction since no specific requirement about the spatial arrangement of the current pathways is made. Note that in eq.(5.11) the current is calculated assuming that the applied bias is symmetrically distributed at both sides of the constriction, but asymmetrical conduction structures can be modeled as well[94]. For simplicity, it is also assumed that the phase-relaxation length is much shorter than the distance between the scatterers[93]. Similar one-dimensional filamentary structures with scatterers have been the subject of numerous publications in the past[95, 96, 97] and it has been shown that a complete understanding of electron transport through these structures requires a quantum approach[98, 99, 100]. The scatterers are assumed to be described by inverted parabolic barriers of height Φ_0. These parabolic barriers could represent the potential bottlenecks between two neighbor atoms.

In this case, $t(E)$ reads:

$$t(E) = \{1 + exp[-\alpha(E - \phi_0)]\}^{-1} \tag{5.12}$$

where α is a constant related to the longitudinal shape of the barrier, and more specifically to the second derivative of the potential profile[101]. Eq.(5.11) can be easily integrated in the zero-temperature limit so as to give the approximate expression:

$$I \approx G_0\left[NV + \frac{2}{e\alpha}exp(-\alpha\phi_{eff})sinh\left(\frac{e\alpha V}{2}\right)\right] \tag{5.13}$$

where N is the number of formed channels. The effective barrier height Φ_{eff} in eq.(5.13) is given by the expression:

$$\phi_{eff} = \phi_0 - ln\left(\sum_i S_i^{-1}\right)^{1/\alpha} \tag{5.14}$$

$\Gamma = \sum_i S_i^{-1}$ is called the configuration factor of the constriction, which reflects the particular arrangement of scatterers that constitute the filament. It depends on number of individual scatters within the single channel (S_i is the number of scatters along the channel i).

As a very simplistic view of the problem, the first term of eq.(5.13) is associated with the core of the constriction (LRS), whereas the second term, which we identify with $f(V)$, corresponds to the current flowing through the cloud of partially formed filaments surrounding the central region (HRS).

According to eq.(5.13), in the absence of a core region (see Figure 6.2(d)), the gap in the constriction can be represented by a single scatterer with an effective barrier Φ_{eff}, which simplifies the conduction problem enormously. As expected, eq.(5.14) reveals that, while the presence of scatterers diminishes the current flow by increasing the potential barrier height, the generation of parallel current paths acts in the opposite manner.

Figure 5.16 illustrates some fitting results using eq.(5.13). Notice that log-linear and log-log axes are used to highlight the agreement and deviation of the experimental data from the theoretical model. Remarkably, Γ is necessary to justify the low effective barrier heights used in eq.(5.13), which otherwise would be incompatible with a tunnelling process.

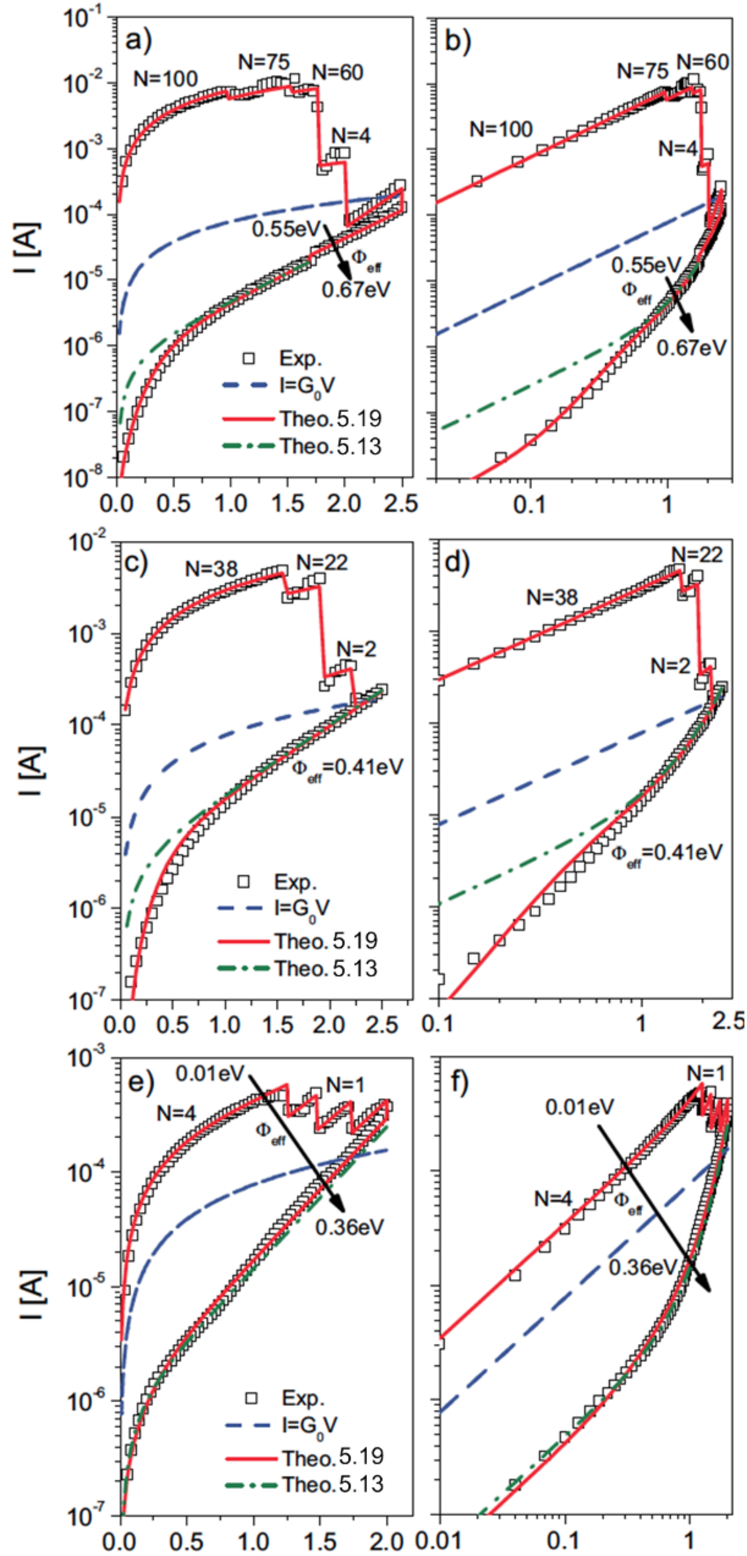

Figure 5.16: Typical current-voltage characteristics with multiple reset events. N is the number of formed channels and Φ_{eff} the effective barrier height in eqs.(5.13) and (5.19). The symbols are experimental data, the red solid line corresponds to the model using eq.(5.19), while the green dot-dashed line was calculated using eq.(5.13). The blue dashed line is the I-V characteristic for a monomode ballistic conductor.

For the sake of simplicity a constant value $\Phi_0 = 1eV$ is considered and its statistical spread is neglected[91]. In Figure 5.16, the conduction characteristic for a monomode ballistic conductor $I = G_0 V$ has been included for comparison.

Figure 5.17 shows several hysteretic I-V loops (0V→V_{max}→0V with increased maximum voltage V_{max} from 1 V to 1.5 V). Contrary to what is observed in Figure 5.16, no jumps are detected in the I-V curve, which indicates the progressive generation of new scatterers or, equivalently, the modification of Φ_{eff}. As proposed in [88], the dynamic behaviour of Φeff as a function of the applied voltage determines the transition from HRS to LRS and vice-versa.

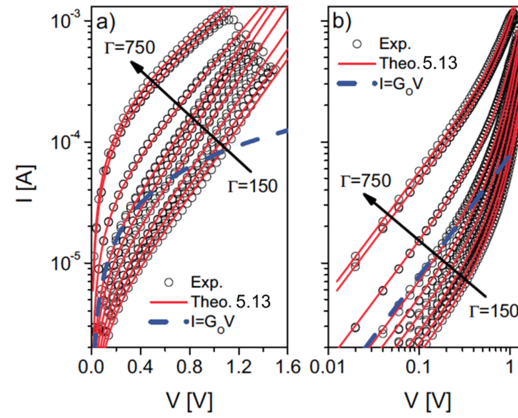

Figure 5.17: Typical hysteretic current-voltage characteristics without abrupt jumps. The symbols are experimental data, the red solid line corresponds to the model using eq.(5.13). The blue dashed line is the I-V characteristic for a monomode ballistic conductor. Γ is the constriction configuration factor.

The model described here makes no reference to this issue but $\Phi_{eff}(V)$ can be incorporated into the model as a coupled equation as is done for memristive systems[1].

In order to achieve further insight into the electron transport mechanism proposed above, the nonlinear term in eq.(5.13) is further discussed. According to this framework, the conductance of an individual scatterer, g_S, can be calculated as[93]:

$$g_s(E) = G_0 \frac{t(E)}{1 - t(E)} = g_s^0 exp\left(\alpha E\right) \tag{5.15}$$

where $g_s^o = g_s(E = 0) = G_0 exp(-\alpha\phi_0)$ is the conductance at the equilibrium energy E=0.

Hence, the conductance of the constriction G_C which arises from the bunch of partially formed

channels, can be written as:

$$G_C(E) = \sum_i \left(\sum g_s^{-1}(E)\right)^{-1} = g_s^0 \Gamma exp(\alpha E) \tag{5.16}$$

Expression (5.16) stems from the scatterers in series and the additive property of the transmission probability already used in eq.(5.11). Consequently, the average conductance of the constriction $< G_C >$reads:

$$\langle G_C(V)\rangle = \frac{1}{eV}\int_{-eV/2}^{+eV/2} G_C(E)dE = G_0 exp(-\alpha\phi_{eff}) sinhc(\frac{e\alpha}{2}V) \tag{5.17}$$

where the function $sinhc(x)$ is defined as $sinh(x)/x$. Substituting eq.(5.17) into eq.(5.13) and taking into account the N formed channels, we obtain:

$$I(V) = (NG_0 + \langle G_C\rangle)\, V \tag{5.18}$$

which expresses the total current as a combination of the core current contribution and that associated with the nearby cloud of scatterers. Notice that as the scatterers become transparent (α $\ll 1$ in eq.(5.17)), $< GC >\approx$ Γ$G0$ is a constant and the linear behaviour of the I-V curve is recovered.

Remarkably, even though eq.(5.13) can be formally rearranged as eq.(5.18) without invoking <G_C>, the latest formulation did not include the contact conductances, as eq.(5.13) implicitly does, and which arise from the mismatch of the energy states occurring at the connections between the metal reservoirs and the channels[93]. This coincidence is a consequence of the approximation $t(E) \approx S_i^{-1} exp\left[\alpha(E - \phi_0)\right]$ used to obtain eq.(5.13). In order to improve the fitting results in the HRS low-bias region (see Figures 5.16(a), 5.16(b), 5.16(c) and 5.16(d)) the correction factor $V_0(V) = Atanh(BV)$ can be introduced into eq.(5.13) as:

$$I \approx G_0 \left\{ NV + \frac{2}{e\alpha} exp(-\alpha\phi_{eff}) sinh\left[\frac{e\alpha}{2}(V - V_0)\right]\right\} \tag{5.19}$$

where A and B are fitting constants. Note that this additional potential drop V_0 is not required for the formed channels.

A similar expression for the effective potential drop has been used before in other RS devices to deal with the semiconducting behaviour of the bottom electrode[65], but no clear explanation in the present context has been found yet.

A dependence of the effective barrier height on the applied voltage cannot be ruled out. Expression (5.19) is chosen this way because it satisfies $V_0(V = 0) = 0$; $V_0(V \ll 1)$ is a linear function of$V$, and $V_0(V \gg 1)$ is a constant which does not affect the exponential dependence of the HRS current in the high-bias range. The fact that the voltage correction factor is unnecessary to fit some HRS I-Vs (see Figures 5.16(d), 5.16(e), 5.17(a) and 5.17(b)) indicates that the faster roll-off of the experimental current for $V < 1V$ is related to particular features of the filaments and not to an essential limitation of the model. Notice that if eq.(5.19) is used, a lower Φ_{eff} is required to compensate for the lower effective applied bias.

Finally, with the aim of analyzing in more detail the experimental I-V curves, the normalized differential conductance $g(V)$ is used. For the two limiting cases we are dealing with, using eq.(5.13), $g(V)$ reads:

$$g(V) = \frac{dln(I)}{dln(V)} = \frac{V}{I}\frac{dI}{dV} = \begin{cases} 1 & N \gg 1 \\ \frac{\alpha e}{2} V coth\left(\frac{\alpha e}{2} V\right) & N = 0 \end{cases} \tag{5.20}$$

which is independent of Φ_{eff}, and therefore of the configuration factor Γ.

This is illustrated in Figure 5.18 for three cases of particular interest: Figure 5.18(a), Figure. 5.18(b), and Figure 5.18(c) correspond to Figure 5.16(c), Figure 5.16(e), and Figure 5.17(a), respectively. In Figure 5.18(a), the results using eqs.(5.13) and (5.19) are compared. In Figure 5.18(b), departures from $g(V) \approx 1$ in the LRS case would indicate the presence of a nonlinear conductance term. In Figure 5.18(c), the set of I-V hysteretic loops is represented.

The only free parameter is α. Notice also that in all the cases $g(V \ll 1) \approx v1$, regardless of the conduction mode. This linear behaviour of the I-V characteristic is a well-known feature of RS devices[102] that our model captures nicely.

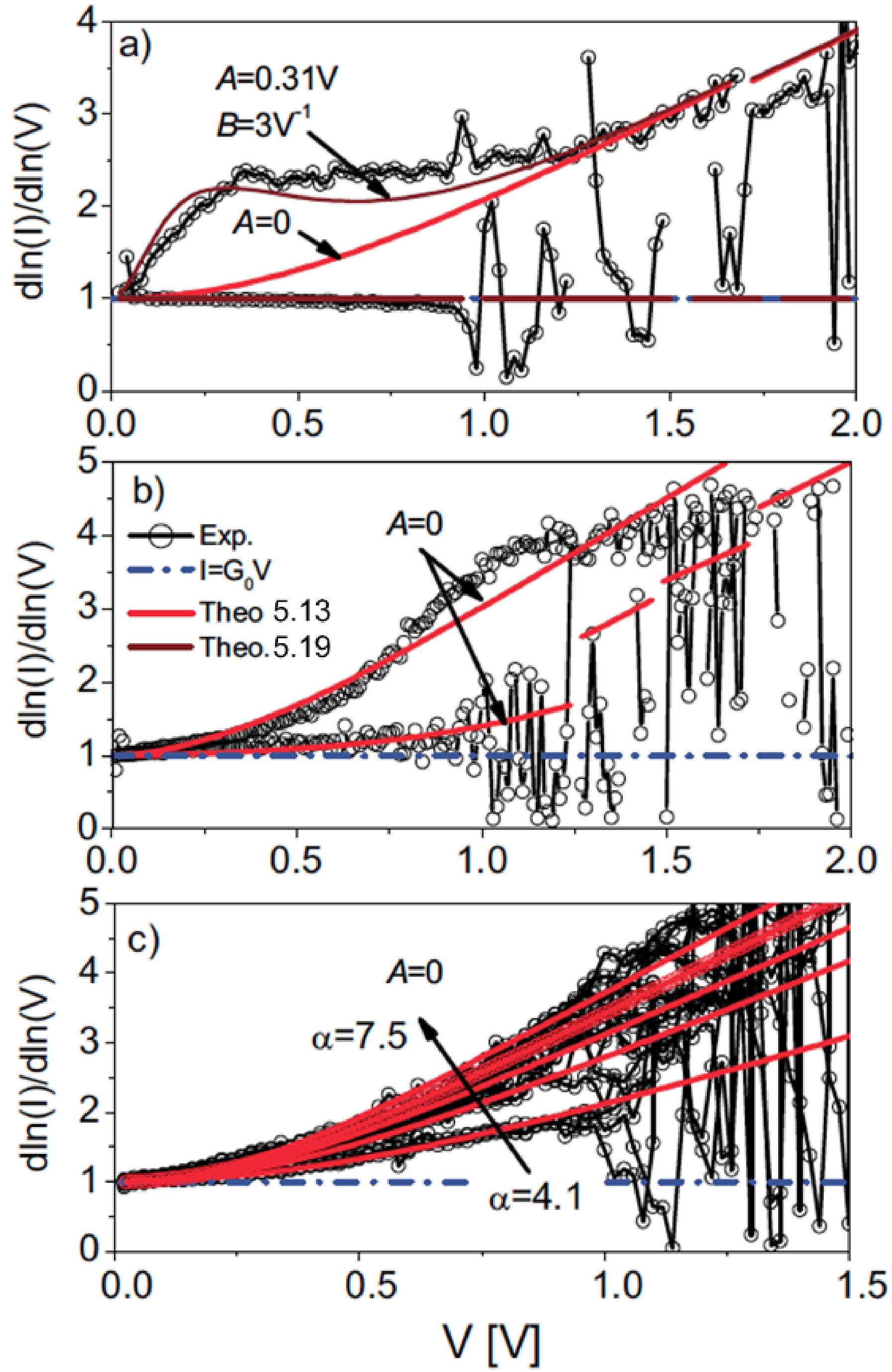

Figure 5.18: Normalized differential conductance-voltage characteristics for three cases of particular interest: a) corresponds to the I-V curve shown in Figure 5.16(c), b) corresponds to the I-V curve shown in Figure 5.16(e), and c) corresponds to the I-V curves shown in Figure 5.17(a). Notice that the correction factor V_0 (eq.(5.19)) is only used in the first case. The blue dashed line is the g-V characteristic for a monomode ballistic conductor.

In summary, a simple analytic model for the I-V characteristics is proposed. The model relies on the theory of mesoscopic conductors and considers the filamentary electron transport pathways within the dielectric film as chains of scatterers whose transmission properties collectively determine not only the magnitude of the current (low and high resistance states) but also its functional dependence on the applied voltage (linear or exponential). The stochastic nature of the atomic arrangement that forms the constriction (configuration factor) is overridden by assuming an effective tunnelling barrier.

Chapter 6

Stochastic Simulation of Resistive Switching in Oxide RRAM

Introduction to the Stochastic Model

The intrinsic nature of the switching process is stochastic. Each switching cycle produces a different configuration of the conductive filament. In the case of redox-based RAM the formation and the rupture of the conductive filament occurs through the rearrangement and movement of oxygen ions and oxygen vacancies. The shape, size and geometry of the filament play a crucial role in determining the switching properties of the device. The properties of the pristine device also affect the nature of the formed filaments. The stochastic nature of the filament formation/rupture presents an additional challenge in analyzing and optimizing the device properties. In this chapter, a stochastic self-consistent model is developed to simulate the formation, set and reset processes on microscopic level. The model is based on resistor breaker network[103] and percolation theory[104]. It is possible to simulate the processes occurring during the switching events.

Model Description

Before I proceed with the model description, a brief reminder of the basic physical picture behind the switching process is provided.

The electroforming step is crucial for most ReRAM systems. The forming voltage is typically significantly higher than the set voltage. Electroforming is a local reduction of the insulating layer triggered by drift or diffusion of oxygen ions. Both voltage polarities can be used. During the forming a high local field is generated due to the large applied voltages. This electric field can cause local morphological changes or a phase transition of the material (e.g. the formation of an oxygen-deficient Magnéli phase in the case of TiO2).

After electroforming, a conductive n-type channel (filament) is produced. Subsequent switching processes take place at the localised part of the filament. A tunnelling barrier is typically present between the two parts of the conductive filament. By varying the length of the gap (barrier height) it is possible to achieve different resistance states.

During the set process oxygen ions are removed from this region - consequently tunnelling and thermionic current is increased and barrier height is reduced. During the reset process oxygen ions are moved back into the region. This increases the barrier height and hence the overall resistance of the ReRAM cell.

In some cases, once a conductive filament has been established, further increasing the applied voltage (without the current compliance) beyond a certain threshold results in a high current flowing through the filament that is sufficient to raise the local temperature through Joule heating to the point at which the filament is disrupted. Temperature-driven migration of oxygen vacancies out of the filament break the current path and return the oxide to a high resistive state. In such a case unipolar switching is seen, in which field-driven filament formation and current-driven filament disruption occur in the same polarity. The reason for this is that oxygen ions and oxygen vacancies diffuse along the temperature gradient and the position of the switching point can be at the arbitrary place along the filament.

A typical I-V curve with physical principle of switching is shown in Figure 6.1. In VCM systems (bipolar switching), oxygen ion drift is dominated by the influence of electric field, while in the case of TCM systems (unipolar switching) the diffusion of oxygen ions is primarily caused by local Joule

heating. Current compliance is required to prevent the hard breakdown and irreversible transition to LRS. Current compliance is removed in the reset process. This allows a higher current to pass at much lower voltages, leading to a partial rupture of the filament and recovery of the HRS. The conductivity of the HRS is usually larger than the conductivity of the pristine device as the filament is not completely dissolved and current can still flow through the part of the filament. The set process is similar to the electroforming process, but it typically occurs at lower voltages. Current compliance is needed to prevent the hard breakdown just as in case of the electroforming. During the forming/set process a reduction of oxidation states occurs. Most transition metal oxides show much lower resistivity in the lower oxidation states.

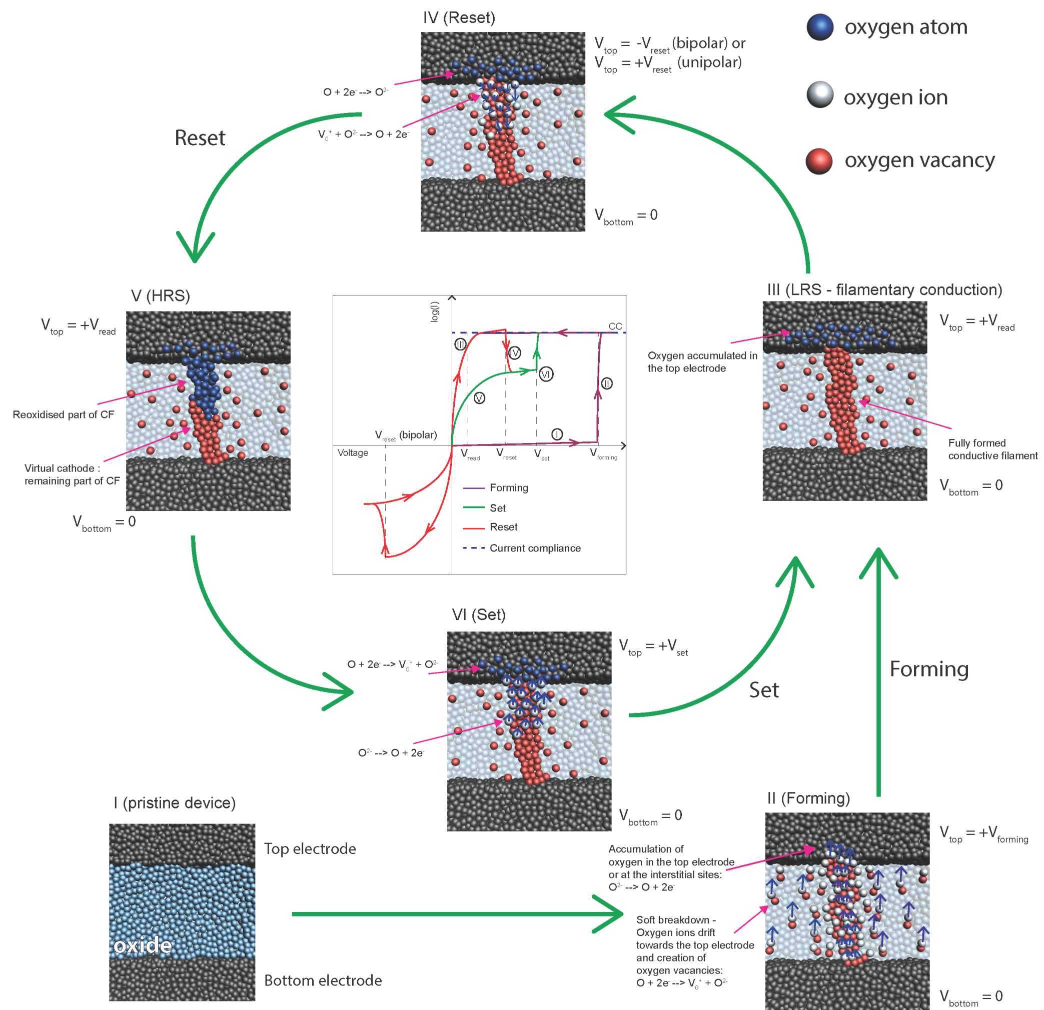

Figure 6.1: Schematic of the switching process in intrinsic filamentary resistive switching. Application of an appropriate voltage with a current compliance limit to the pristine device (I) forms the oxide (II) into a low resistive state (LRS) (III) by moving oxygen ions to one of the electrodes, generating a conductive filament bridging the two electrodes as oxygen vacancies are left behind. Removing the current compliance and applying a suitable voltage then disrupts the filament (IV) through thermally-assisted migration of oxygen, retuning the device to a high resistive state (HRS) (V) – note this is generally less resistive than the initial pristine state because only a section of the filament is disrupted. The device may then be returned to the low resistive state on the application of a suitable voltage with a current compliance limit set to avoid overheating (VI).

In this model, the formation and the rupture of the conductive filament are modeled through the movement of oxygen ions and generation of oxygen vacancies. The conductive filament consists of oxygen vacancies. The overall conductivity of the memory cell is dependent on size and shape of the filament.

Movement of oxygen ions is described by the general probabilistic equation[105]:

$$P_a = f \exp\left(-\frac{E_a - \triangle\varphi_1}{k_B T_{loc}}\right) \tag{6.1}$$

P_a is the probability of ion hopping between the two sites, f is vibration frequency of the oxygen, E_a is the activation energy, $\Delta\varphi_1$ represents the modification of the activation energy due to the potential difference between two sites, k_B is Boltzmann's constant and T_{loc} is the local temperature.

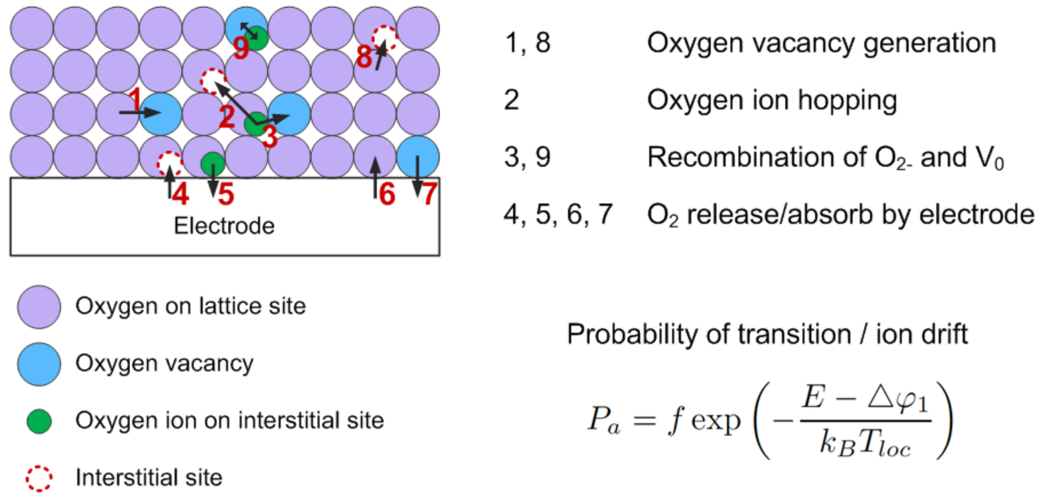

Figure 6.2: Schematic representation of the physical processes.

Figure 6.2 is a schematic representation of the different hopping processes. At the lattice site two states are possible: oxygen state (pink in figure) and oxygen vacancy state (blue in figure).

Apart from the lattice sites, interstitial sites can exist. Every interstitial site can accommodate an oxygen ion. In Figure 6.2 a pink circle represents an oxygen at the lattice site, while a blue circle represents an oxygen vacancy, a blank orange circle represents an unoccupied state at interstitial site, and a green circle represents an occupied state (oxygen ion) at an interstitial site.

Hopping of the oxygen ions can occur through different processes. Hopping of an oxygen ion from an oxygen state on the lattice site is followed by the generation of an oxygen vacancy. This transition is only possible if there is an available state to accommodate the incoming ion.

Generally, the transition probabilities are modified by the decrease/increase of the transition barrier due to the applied voltage bias and by the change of the local temperature.

Process 1 represents the hopping of an oxygen ion from the oxygen state to the next oxygen vacancy state. A new oxygen vacancy will be generated while the old oxygen vacancy will be annihilated through the recombination of oxygen ion and oxygen vacancy. This process is modeled with equation 6.1 where the activation energy is E_a. If there is an adjacent interstitial site that is unoccupied the oxygen ion can move to that site and that represented by process 8, modeled with an activation energy of E_b which is usually smaller than E_a (interstitial site being closer to the hopping oxygen ion).

The equivalent ion hopping processes from the interstitial sites are represented by process 2 (ion hopping through the interstitial sites) or process 3 (the recombination process of ion and oxygen vacancy).

As the electrodes are usually considered as the source/drain of oxygen ions the process of the exchange of oxygen between the electrodes are represented by processes 4,5,6 and 7. Finally, if there is an occupied interstitial site near the oxygen vacancy the recombination process occurs (process 9).

To generate the I-V characteristics we adopted the random circuit breaker model[103] shown in Figure 6.3. This model assumes that between any two sites there is a resistor with one of the two specific values. These two resistor values represent the microscopic equivalent of the HRS and LRS resistance values, so $R_1 \ll R_2$. R_1 stands for the lower resistance value; this value models the resistance between the two oxygen vacancies, while R_2 stands for the much higher resistance value and models the other configurations including resistance between two oxygen states or oxygen state and oxygen vacancy.

When the percolation path of oxygen vacancies between the two electrodes is formed, the LRS is achieved. If the path is broken the HRS is recovered. These simulate the set/formation and reset processes.

Figures 6.4 shows the chart of the main steps in the simulation. The input in the simulation is a preconfigured distribution of oxygen vacancies. This is the pristine state of the device.

In the second step the current map and the potential map are calculated. The local currents are

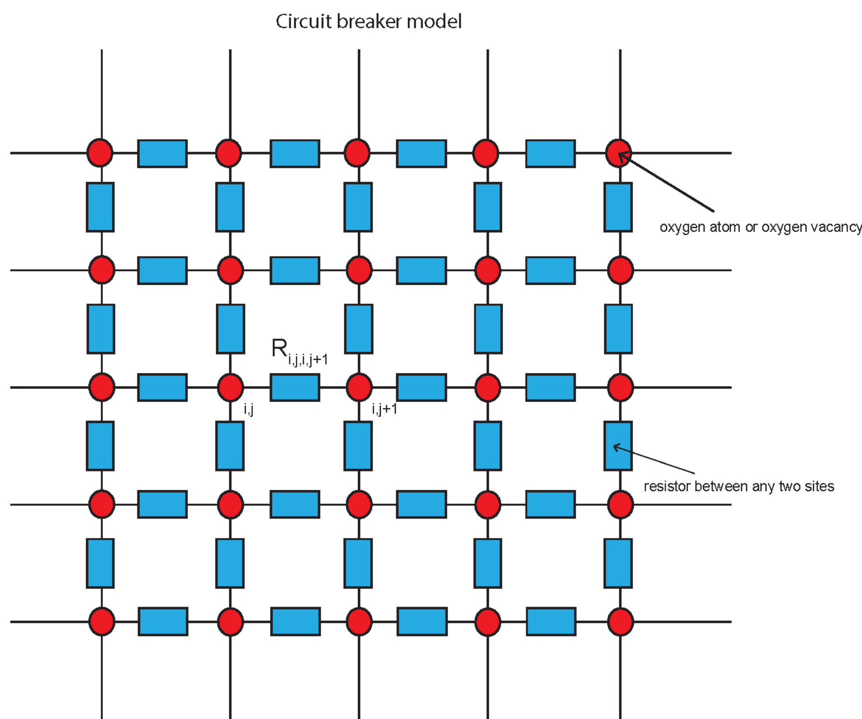

Figure 6.3: Circuit breaker model. The model assumes that between any two sites there is a resistor with one of the two specific values.

potentials are calculated using mesh current method. There are two components that contribute to the overall current in the case of HRS (broken filament). The first component is the current calculated by Ohm's law – dividing the bias voltage and equivalent resistance of the whole network. The second component takes into account a tunneling current in the case when the filament is ruptured and there is a tunneling gap.

For this component the following formula is used:

$$I = G_0 \left\{ \frac{2}{e\alpha} \exp\left(-\alpha\varphi_{eff}\right) \sinh\left[\frac{e\alpha}{2}\left(V - V_0\right)\right] \right\} \tag{6.2}$$

The formula is described in more details in Chapter 5.

The effective tunneling barrier is exponentially dependent on the width of the filament gap. For calculation of the shortest path between the two electrodes and the tunneling gap the A* search algorithm[106] is used. After this step the local temperatures are calculated in step 3. The local temperature is calculated as the Joule heat generated within the resistor. Once the local potentials and local temperatures are calculated it is possible to calculate the probabilities of the transitions (1-9). For every site multiple transitions are possible and a stack of the transitions is generated

with the transitions with the highest probabilities on the top. This is done in step 4. After the individual probabilities are calculated the transitions occurs according to a stochastic process. A random number (positive, not greater than 1) is generated and it is compared with the stacked probabilities. After every comparison the new random number is generated. The random number can take any value between 0 and 1 (including 0 and 1) with the same probability. The probabilities are compared from the highest to the lowest. If the probability of the certain transition is greater than the randomly generated number, transition takes place and the next site is examined. After every transition the new configuration matrix is generated. For effectiveness of the code the local currents, local potentials and local temperatures are calculated once the whole matrix is reconfigured.

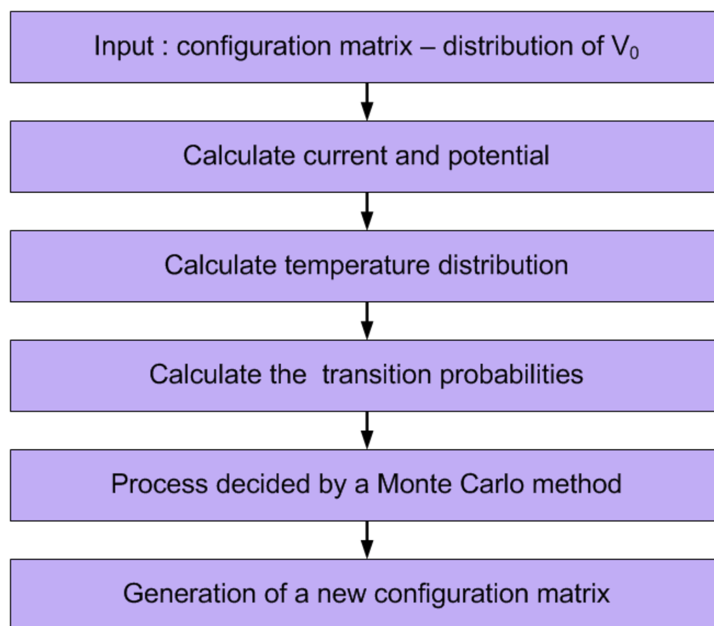

Figure 6.4: Algorithm flow chart.

The model allows a direct generation of I-V curves. Modeled I-V curves are compared to experimental results and validity of the model and its assumptions are tested.

Simulation Results

In the simulation we assume that the pristine device does not contain any predefined oxygen vacancies and that the generation of oxygen vacancies occurs with the application of electrical bias.

Electroforming

Figure 6.5 summarizes the simulation of the formation process. The device is swept until the current level increases above a critical value (in this case 10mA). This is equivalent to setting a current compliance.

Very few (if any) oxygen vacancies are generated at the lower voltages (lower than 15V). This is shown in Figure 6.5(a). Once a certain voltage level is achieved the accelerated generation of oxygen vacancies occurs. At this point oxygen vacancies start to form a filament. The filament grows with the further voltage increase. This is shown in Figures 6.5(b)-(d).

It is possible that during the growth of the initial filament a secondary filament seed is produced. Once the filament bridges the two electrodes a sudden current jump is observed – which represents the transition to the LRS. At this point voltage bias is removed. This is represented in Figure 6.5(e).

If the current is not limited, meaning voltage bias is present after the current jump, the filament that bridges the electrodes starts to grow radially. This is observed in Figure 6.5(f). This causes the hard breakdown.

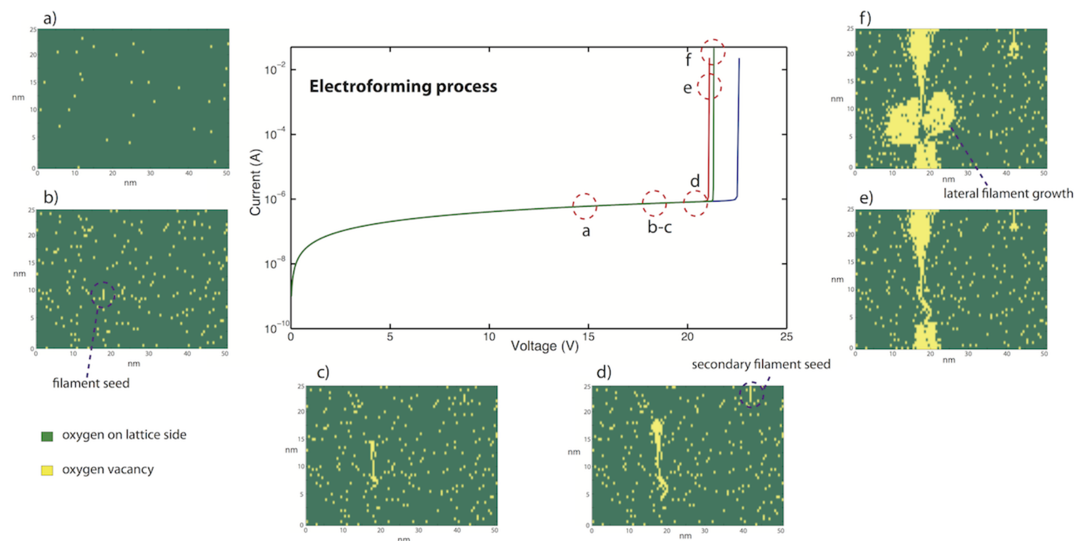

Figure 6.5: Simulation results of the forming process. (a) Only a few oxygen vacancies are generated at the lower voltages. (b) Filament seed is formed. (c) Growth of the filament. (d) Secondary filament seeds can be formed. (e) Filament is completely formed at the threshold voltage. (f) If the voltage bias is not removed, the filament starts to grow laterally.

Formation of the filament (from the pristine device) is governed by stochastic processes and consequently the shape of the filament varies from device to device. Nevertheless some general properties (e.g. radial growth, LRS conductivity) are determined by current compliance.

Reset Process

After the filament is formed and device switches to LRS, we simulate the reset process.

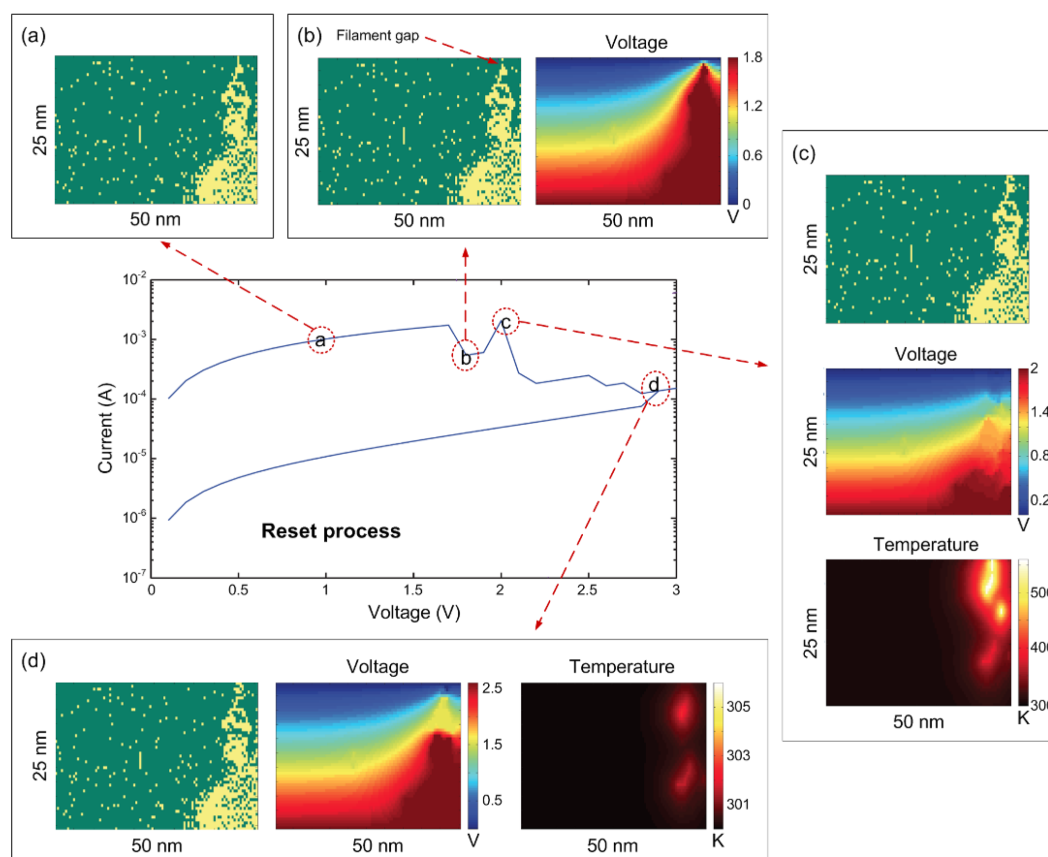

Figure 6.6: Simulation results of the reset process. (a) The filament is fully formed and the temperature is generated by Joule heating. (b) Once the current reaches the critical level filament gap is produced by re-oxidation. (c) If the filament gap is small, the filament can be temporally recreated. Higher Joule heating is generated, rupturing the filament. (d) If the filament gap is wide enough, the stable state is achieved (the HRS).

Figure 6.6 shows the simulation results of the reset process. Starting from the LRS the device is swept to 3V. No processes occur until some critical current level is achieved. This current is always larger than the current compliance level of the previous formation/set process, but usually not by more than a factor of 2.

After the critical current level is achieved the device's conductivity starts to fluctuate. In Chapter 3 we described this behaviour as the competition between set and reset processes. In Figure 6.6(b) a smaller current drop is observed and the filament gap is produced. This is easily observed in the potential map in which the potential difference is largest in the gap area. If the gap is not too large it is possible that the filament recovers and the LRS is reestablished. This is observed in Figure 6.6(c). The filament is reconstructed and there are no filament gaps. The potential map shows that

the potential is now much more uniformly distributed across the filament.

With recovery of the filament a high current is once again achieved which produce a high local Joule heating. This is shown in Figure 6.6(c) in the temperature map. The narrowest parts of the filament experience the largest heating (around 600K). This is the direct consequence of the Joule heating which is dependent on resistivity of the path. The resistivity of the narrowest part is the largest while the current remains the same. The probability of oxygen ions hopping in these parts is increased and a larger gap will be consequently produced. This is observed in Figure 6.6(d).

When the large filament gap is produced the current drops significantly as well as the local heating – consequently the HRS is achieved. The more uniform distribution of potential in the gap area is also the direct consequence of the gap size. If the voltage is not increased above the critical value (set voltage) the device remains in this state, as the probability of ions hopping is low.

Set Process

After the device is switched to the HRS it remains in this state until it is swept again to higher voltages. The set process is shown in Figure 6.7.

Starting from HRS the device is swept to higher voltages until a certain current level is achieved when the voltage bias is removed. This is the same procedure as for the initial electroforming process. No processes occur until the larger voltage is achieved, at which point current fluctuations are again observed. This voltage is always larger than the reset voltage. This is observed in Figure 6.7(c) where a small current drop is followed by sudden current increase until the current compliance level is achieved.

Figures 6.7(a) and 6.7(b), respectively, show the HRS and the LRS (after the set process). Figure 6.7(d) shows the temperature map during the set process. The maximum temperature during the set does not increase significantly over room temperature (300K) and we conclude that the potential distribution (electric field) plays much more important role during the set process.

Figure 6.7(e) shows the filament recovery process. Starting from the HRS (Figure 6.7(e1)) during the set process the filament recovers and the gap decreases until no gap is present. The electric field across the gap increases as the gap decreases. Consequently, the gap decrease works as a sort of positive feedback during the set.

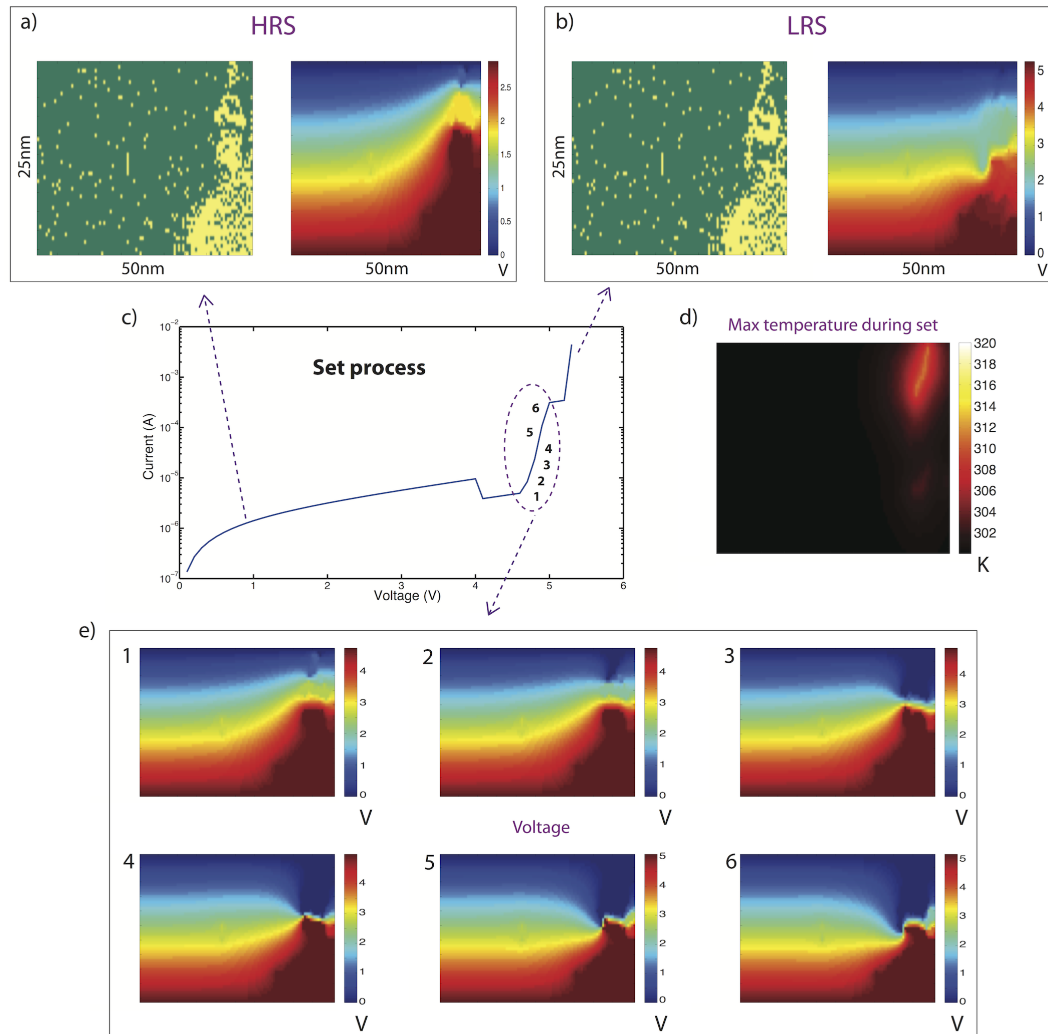

Figure 6.7: Simulation results of the set process. (a) At the lower voltages the HRS is stable as the electric filed is not high enough to trigger the oxygen ions movement. (b) At the final stage of the set process the filament is recreated. (c) I-V curve showing the set process. (d) Maximum temperature during the set process does not go above 320K confirming that the set process is mainly driven by the electric field. (e) Once the threshold voltage is achieved the set process (recreation of the filament) is observed by closing the filament gap.

Endurance Testing

Endurance is one of the most important properties of any memory type and ReRAMs are not an exception. For most cases the failure is observed as the failure of the HRS where the device remains in LRS and reset process cannot be achieved.

With the model we can simulate the device's endurance by testing different assumptions, more specifically the availability of the oxygen ions.

Resistive Switching without an External Oxygen Ion Source

Unipolar Switching

The I-V curves shown in Figure 6.8 are obtained for the case in which there is no other source of oxygen ions apart from the oxygen ions located at the interstitial sites during the preceding set process. We examine the unipolar switching mode first.

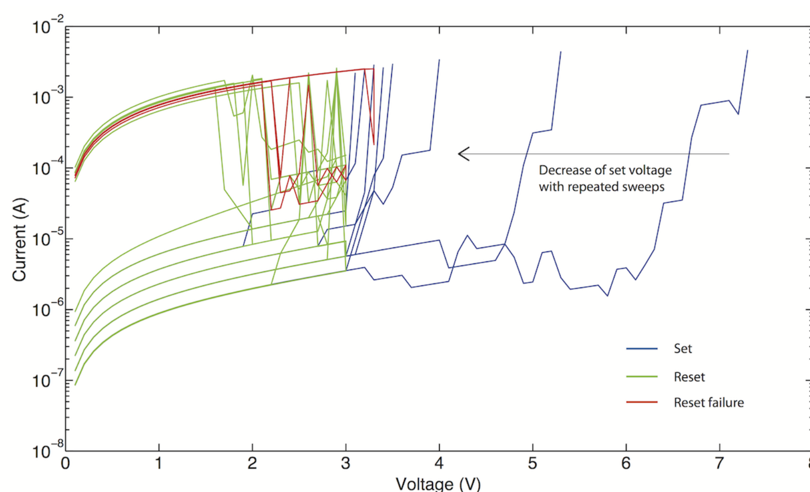

Figure 6.8: Simulation results in unipolar switching mode.

We assume that the pristine device does not contain any oxygen vacancies and the whole matrix consists of oxygen states at lattice sites. During the formation process oxygen ions hop through the matrix, leaving oxygen vacancies behind. We assume that in the proximity of every lattice site there is an empty interstitial site that can accommodate a single oxygen ion.

Oxygen Ions, accommodated at the interstitial sites are used to re-oxidize the filament during the reset process. In this case there is no other oxygen ion source. The filament is re-oxidized through the process of oxygen ion – oxygen vacancy recombination. If there are no available oxygen ions in the proximity of the formed oxygen-vacancy filament, re-oxidation is not possible. This results in the failure of the reset process and the failure of the device operation. Experimentally this always happens after a finite number of switching processes.

Figure 6.8 shows that, for this simulation after 9 switching cycles a failure of the reset cycle occurs. The device cannot be switched back to HRS by applying the standard voltage sweep up to 3V. Figure 6.9 shows the filament structure after every reset process until the reset failure is observed in Figure 6.9(i). It can be seen that the gap position and size changes from cycle to cycle (represented

by red circile). Eventually the device remains in the LRS and the filament can not be re-oxidized as there are no available oxygen ions.

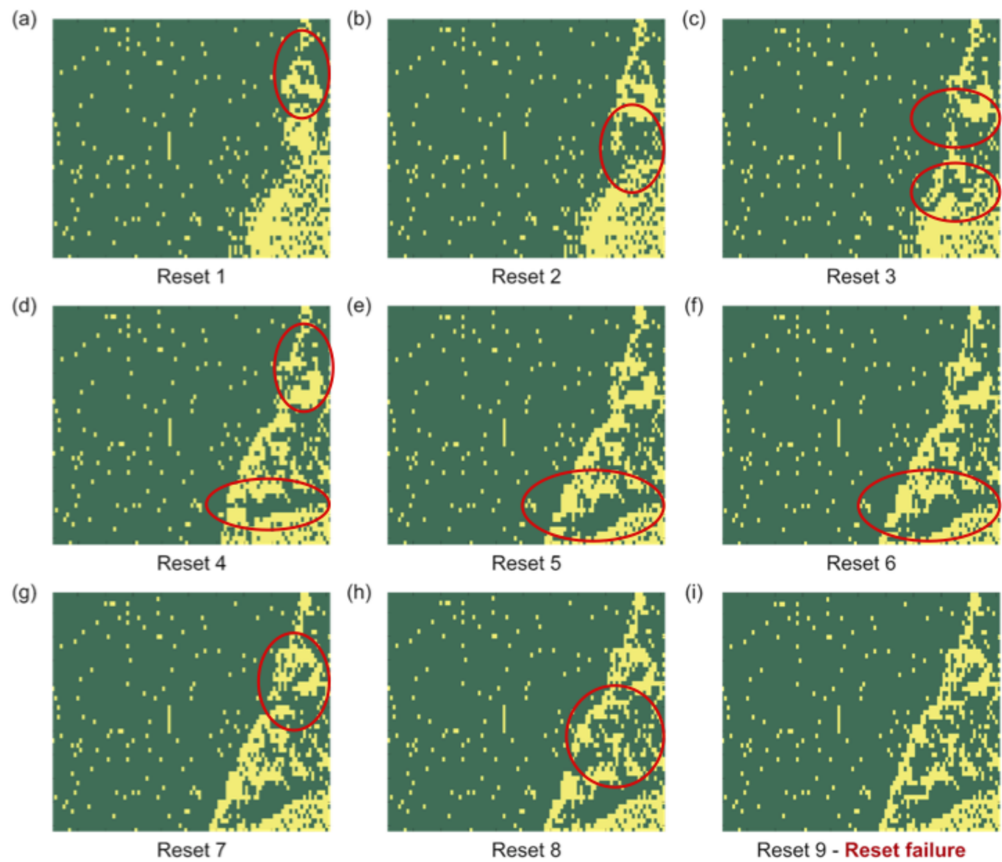

Figure 6.9: Filament structure after every reset process. Red circles represent the position of the gap in the filamnet.

Oxygen ions that are placed at the interstitial sites during the electroforming process (by oxygen ion – oxygen vacancy generation) are represented in Figure 6.10(a) and (b) as green dots. Blue dots represent the empty interstitial sites. The concentration of oxygen ions (at interstitial sites) around the filament is much lower after 9 switching cycles and it is not sufficient for the reset to occur. Consequently, the HRS cannot be restored.

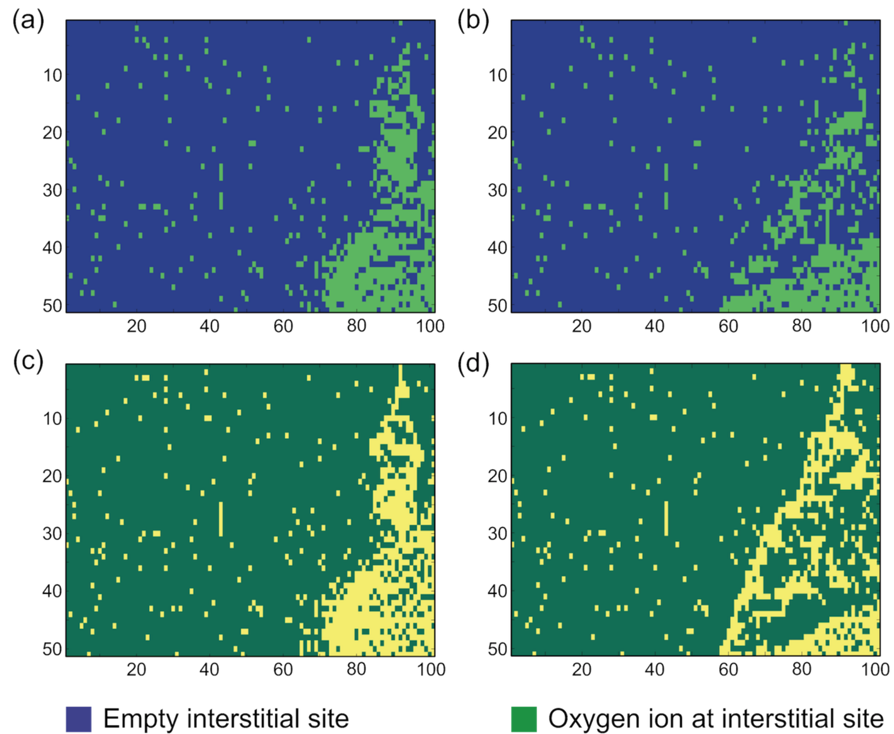

Figure 6.10: Distribution of oxygen ions at the interstitial sites before: (a) first reset process, (b) before the last reset process (failure). Filament structure before: (c) first reset process, (d) before the last reset process (failure).

Bipolar Switching

Figure 6.11 shows I-V curves obtained for the bipolar switching mode. The assumptions are the same as for the case of unipolar switching (no external oxygen source). Switching behaviour in the case of the bipolar mode is very similar to the case of the unipolar mode. Failure of the reset process is the same. The endurance is slightly higher – 14 cycles before the failure compared to 8 cycles before the failure in the case of the unipolar mode.

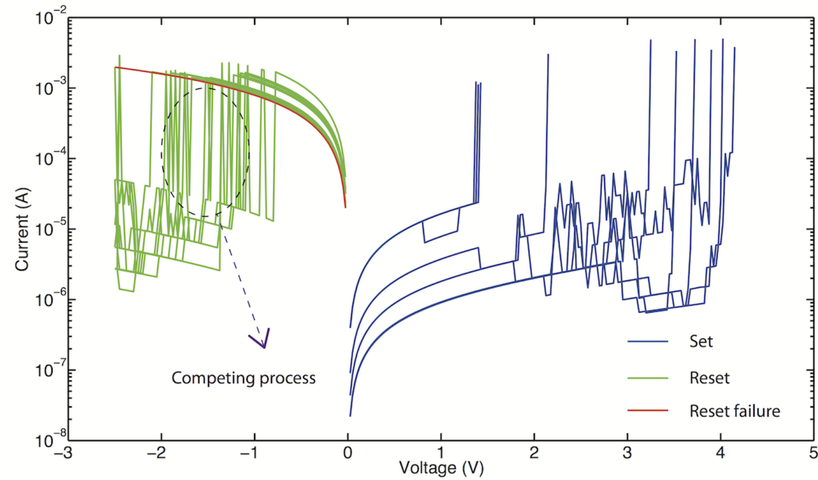

Figure 6.11: Simulation results in bipolar switching mode.

Resistive Switching with Unlimited Oxygen Vacancies Source

In many cases, instead of a single switching layer, a bi-layer configuration is used. The additional layer provides an additional source of oxygen ions and oxygen vacancies. This way device endurance is improved[107].

For the next set of simulations we make the assumption of an unlimited source of oxygen ions. This means that during the set process there is no restriction on site availability for the accommodation of oxygen ion or the availability of oxygen ion during the reset process.

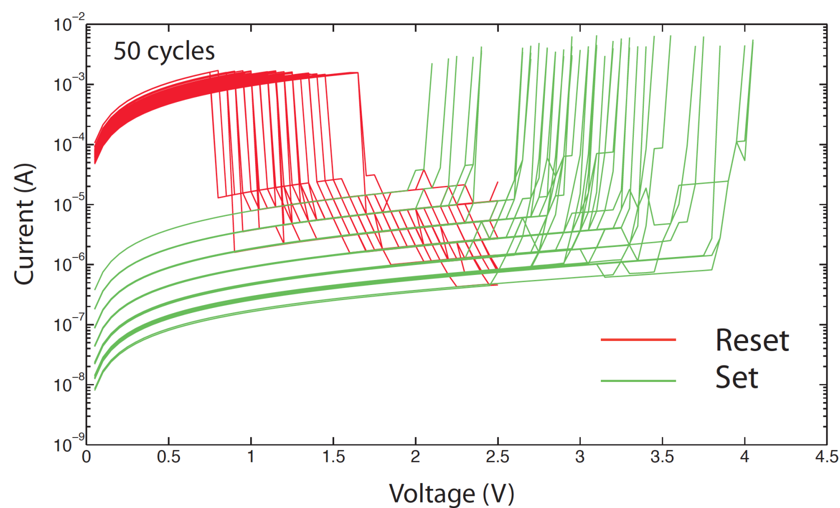

Figure 6.12: Simulation results with unlimited oxygen ions source.

Figure 6.12 shows 50 programming cycles. More stable resistive switching is obtained. The simulation was stopped after these 50 cycles without a sign of endurance failure.

This indicates that the depletion of oxygen ions plays a crucial role in endurance failure. This has been experimentally confirmed by testing the switching in atmospheric, vacuum, and oxygen gas environment[62]. Figure 6.13 shows that the complete reset cannot be achieved in the vacuum environment, probably due to the lack of oxygen necessary for the re-oxidation.

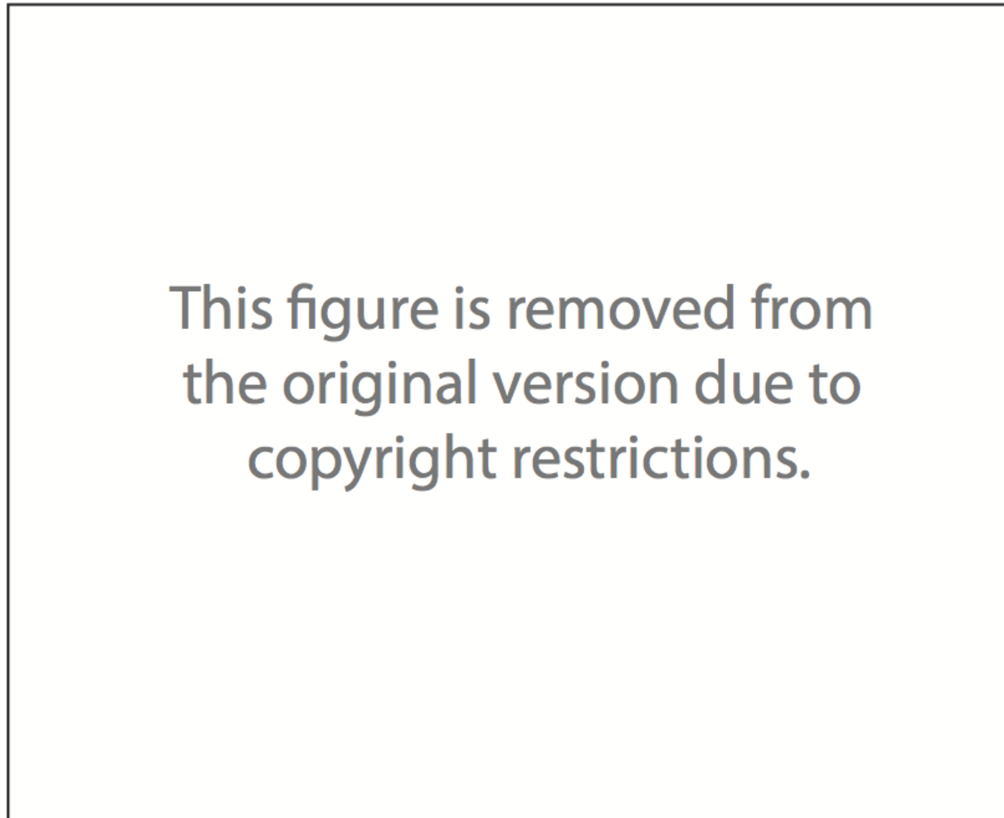

Figure 6.13: Resistive switching behaviours of Pt/SiO0.73/Pt device operating in atmospheric, vacuum, and oxygen gas environment, respectively. Taken from [62].

Chapter 7

Conclusion and Future Plans

I fabricated redox-based silicon oxide memory switching devices. In this thesis, I studied the basic properties of the devices using a set of different experimental techniques and I developed the basic physical model governing the switching processes.

I studied both structural and electrical properties of materials used for fabrication of the devices.

In Chapter 1 I provided an overview of digital memories and discussed the basic physical principles governing different memory types. I put a special focus on resistive random access memories (RRAM) and more specifically, on redox based resistive random access memories (ReRAM). I provided a classification of the main types of ReRAM system and I discussed the main physical processes.

In Chapter 2 I provided an overview of experimental techniques and the samples preparation procedures.

In Chapter 3 I showed the feasibility of the SiO_x memory device by testing devices in both the sweeping and the pulsing mode. I showed the different modes of operation and also demonstrated the electrical adjustment of switching properties. I discussed the advantages of the silicon oxide ReRAM over other metal oxide based ReRAMs; specifically, intrinsic self-rectification and non-linearity of I-V curves. I studied the physical processes behind the switching events by performing a set of scanning tunneling microcopy, atomic force microscopy and conductive atomic force microscopy measurements.

In Chapter 4, I demonstrated the conductance of the devices is quantised in half-integer multiples of G_0. In contrast to other resistive switching systems this quantisation is intrinsic to SiO_x, and is not due to drift of metallic ions. I explained half-integer quantisation in terms of the filament structure and formation mechanism, which allows me to distinguish between systems that exhibit integer and half-integer quantisation.

In Chapter 5 I suggested a physical model that governs the two switching modes. I analysed the conduction mechanism of both the high resistive state and the low resistive state. I tested the experimental results with a model based on quantum point contact theory.

In Chapter 6 I developed a computational stochastic self-consistent model to simulate the formation, set and reset processes on a microscopic level. I discussed the switching uniformity. I also discussed the dependence of the switching properties on the availability of oxygen.

The most prominent model of resistive switching is based on formation and annihilation of a nanometer scale conductive filament. Tracking the filament is a challenging task. In Chapter 3 I showed that this is possible via means of C-AFM and STM measurements. Although very valuable information can be extracted from these measurements they are restricted to the plane projection. The next step should include transmission electron microscopy (TEM) measurements. These should provide a better picture about the shape and size of the whole filament. In-situ TEM can reveal the dynamics of filament evolution. The feasibility of in-situ TEM has already been shown in both ECM[108] and VCM[26] ReRAM. However, poor contrast between silicon dioxide and silicon might provide an additional challenge in case of this particular system.

Wafer yield and switching uniformity should be further improved through fabrication process optimisation. Aggressively scaled devices in the 10s of nm are yet to be tested. Device properties should be tested in crossbar arrays – especially in the passive matrix configuration in which the two discussed properties of SiO_x system are particularly important (self-rectification and nonlinearity in I-V curves). Simple crossbars of 8×8 or 16×16 would provide enough information about the feasibility of diode-less structure where ReRAM cell would provide both selecting and switching.

The quantised conductance discussed in Chapter 4 provides additional information on the filament nano-scale structure. This also brings the possibility of multibit storage or analog-like applications (such as neuromorphic architectures). Precise control over the quantum steps would be crucial step towards such new applications.

Although switching in metal oxides is becoming better understood, comparatively little is known about silicon oxides, other than there are significant differences between metallic and semiconducting systems. In Chapter 5, of particular interest are the differences between unipolar and bipolar switching. I have established already that different mechanisms operate in the two cases, and also that the intrinsic current limiting behaviour of our Si-based devices has an important role to play in allowing us to configure the switching mode purely electrically. With further structural information on the switching material (e.g. HTEM, atom probe microscopy), theoretical models on the atomistic level with focus on defects generation can be developed. Ab initio simulation techniques can help in understanding the role of the specific defects in the electroforming and switching processes.

Finally the simple stochastic model developed in Chapter 6 can be further used to analyse the different effects on resistive switching. Switching properties like switching uniformity and endurance can be directly related to different factors – like availability of oxygen or predefined configuration of oxygen vacancies. These would help in further device optimisation.

Published work

Journal papers:

1. Adnan Mehonic, Sébastien Cueff, Maciej Wojdak, Stephen Hudziak, Olivier Jambois, Christophe Labbé, Blas Garrido, Richard Rizk, and Anthony J. Kenyon. Resistive switching in silicon suboxide films. *Journal of Applied Physics* 111, no. 7 : 074507-074507, 2012.

2. Adnan Mehonic, Sébastien Cueff, Maciej Wojdak, Stephen Hudziak, Christophe Labbé, Richard Rizk, and Anthony J. Kenyon. Electrically tailored resistance switching in silicon oxide. *Nanotechnology* 23, no. 45 : 455201, 2012.

3. Adnan Mehonic, Andrei Vrajitoarea, Sébastien Cueff, Stephan Hudziak, Hume Howe, Christophe Labbé, Ricahrd Rizk, Michael Pepper, Anthony J. Kenyon. Quantum Conductance in Silicon Oxide Resistive Memory Devices. *Scientific Reports* 3, 2013.

4. Enrique Miranda , Adnan Mehonic, Jordi Suñé 1, Anthony J. Kenyon. Multi-channel Conducti on in Redox-based Resistive RAM Modelled Using Quantum Point Contact Theory. *Applied Physics Letters* 103, 222904, 2013.

Conference papers:

1. Adnan Mehonic, Sébastien Cueff, Maciej Wojdak, Stephen Hudziak, Olivier Jambois, Christophe Labbé, Blas Garrido, Richard Rizk, and Anthony J. Kenyon. "Intrinsic Resistive Switching in Bulk SiOx Films." In MRS Proceedings, vol. 1430, pp. mrss12-1430. Cambridge University Press, (2012).

Conference presentations and posters:

1. Adnan Mehonic, Sébastien Cueff, Maciej Wojdak, Stephen Hudziak, Christophe Labbé, Richard Rizk, and Anthony J. Kenyon. Bimodal Resistive Switching in Pure SiOx with Multi-level Capabilities. *EMRS Fall Meeting*, 2012.

2. Adnan Mehonic, Sébastien Cueff, Maciej Wojdak, Stephen Hudziak, Christophe Labbé, Richard Rizk, and Anthony J. Kenyon. Aspects of Semiconductor RRAM. *MRS Spring Meeting*, 2013.

3. Adnan Mehonic, Andrei Vrajitoarea, Sébastien Cueff, Stephan Hudziak, Hume Howe, Christophe Labbé, Ricahrd Rizk, Michael Pepper, Anthony J. Kenyon. Quantized Conductance in Resistive Switching Silicon Oxide. *MRS Spring Meeting*, 2013.

4. Adnan Mehonic, Sébastien Cueff, Maciej Wojdak, Stephen Hudziak, Christophe Labbé, Richard Rizk, and Anthony J. Kenyon. Resistive switching in silicon-rich silica. *IMRC*, 2013.

5. Anthony J Kenyon, Adnan Mehonic. Resistive switching in silicon oxide. *Collaborative Conference in 3D and Materials Research*, 2013.

6. Adnan Mehonic, Sébastien Cueff, Maciej Wojdak, Stephen Hudziak, Christophe Labbé, Richard Rizk, and Anthony J. Kenyon. Resistive RAM from Active Silicon Oxide. *EMRS Spring meeting*, 2013.

7. Anthony J Kenyon, Adnan Mehonic. Resistive switching in silicon oxide containing silicon nanoinclusions. *COST CM1104 meeting on reducible oxides,* 2013.

Patents:

1. Kenyon, Anthony Joseph, and Adnan Mehonic. Oxide Memory Resistor Including Semiconductor Nanoparticles. *WIPO Patent 2013005040*, issued January 11, 2013.

Bibliography

[1] Leon Chua. Resistance switching memories are memristors. *Applied Physics A: Materials Science & Processing*, 102(4):765–783, 2011.

[2] Leon Chua. Memristor-the missing circuit element. *Circuit Theory, IEEE Transactions on*, 18(5):507–519, 1971.

[3] Jim Hutchby and Mike Garner. Assessment of the Potential & Maturity of Selected Emerging Research Memory Technologies, www.itrs.net/Links/2010ITRS/2010Update/ToPost/ERD_ERM_2010FINALReportMemoryAssessment_ITRS.pdf. Accessed 7th May 2013.

[4] Martin Hilbert and Priscila López. The world's technological capacity to store, communicate, and compute information. *Science*, 332(6025):60–65, 2011.

[5] Rainer Waser. *Nanoelectronics and information technology, 3rd edition*. Wiley, 2012.

[6] Rainer Waser, Regina Dittmann, Georgi Staikov, and Kristof Szot. Redox-based resistive switching memories–nanoionic mechanisms, prospects, and challenges. *Advanced Materials*, 21(25-26):2632–2663, 2009.

[7] John Boyd. Mixing memory to speed solid–state drives, http://spectrum.ieee.org/semiconductors/memory/mixing-memory-to-speed-solidstate-drives. Accessed 4th June 2013.

[8] NRAM - Nantero Mission, http://www.nantero.com/mission.html. Accessed 3rd June 2013.

[9] MRAM - Fact Sheet, http://www.freescale.com/files/memory/doc/fact_sheet/MRAMTECHFS.pdf. Accessed 3rd June 2013.

[10] Wei Wang, A. Gibby, Zheng Wang, Tze Wee Chen, Shinobu Fujita, P. Griffin, Yoshio Nishi, and Simon Wong. Nonvolatile SRAM cell. In *Electron Devices Meeting, 2006. IEDM '06. International*, pages 1–4, 2006.

[11] C Vaju, Laurent Cario, Benoit Corraze, Etienne Janod, Vincent Dubost, Tristan Cren, Dimitri Roditchev, D Braithwaite, and Olivier Chauvet. Electric-pulse-induced resistive switching and possible superconductivity in the Mott insulator $GaTa_4Se_8$. *Microelectronic Engineering*, 85(12):2430–2433, 2008.

[12] Micron announces availability of phase change memory for mobile devices, http://investors.micron.com/releasedetail.cfm?releaseid=692563.

[13] Chun Ning Lau, Duncan R. Stewart, R. Stanley Williams, and Marc Bockrath. Direct observation of nanoscale switching centers in metal/molecule/metal structures. *Nano Letters*, 4(4):569–572, 2004.

[14] Peng Guo, Yuan-Wei Dong, Xin Ji, Yin-Xiang Lu, and Wei Xu. Nonvolatile multilevel conductance and memory effect in molecule-based devices. *Electron Device Letters, IEEE*, 28(7):572–574, 2007.

[15] Takhee Lee and Yong Chen. Organic resistive nonvolatile memory materials. *MRS Bulletin*, 37:144–149, 2 2012.

[16] Qi-Dan Ling, Der-Jang Liaw, Chunxiang Zhu, Daniel Siu-Hung Chan, En-Tang Kang, and Koon-Gee Neoh. Polymer electronic memories: Materials, devices and mechanisms. *Progress in Polymer Science*, 33(10):917–978, 2008.

[17] S. Thakoor, A. Moopenn, T. Daud, and A. P. Thakoor. Solid-state thin-film memistor for electronic neural networks. *Journal of Applied Physics*, 67(6):3132–3135, 1990.

[18] A. Beck, J. G. Bednorz, Ch. Gerber, C. Rossel, and D. Widmer. Reproducible switching effect in thin oxide films for memory applications. *Applied Physics Letters*, 77(1):139–141, 2000.

[19] Leon Chua. Resistance switching memories are memristors. *Applied Physics A*, 102(4):765–783, 2011.

[20] Dmitri B Strukov, Gregory S Snider, Duncan R Stewart, and R Stanley Williams. The missing memristor found. *Nature*, 453(7191):80–83, 2008.

[21] K. Tsunoda, Y. Fukuzumi, J. R. Jameson, Z. Wang, P. B. Griffin, and Y. Nishi. Bipolar resistive switching in polycrystalline TiO_2 films. *Applied Physics Letters*, 90(11):113501, 2007.

[22] S. Tappertzhofen S. Schmelzer J. van den Hurk F. Lentz I. Valov, E. Linn and R. Waser. Nanobatteries in redox-based resistive switches require extension of memristor theory. *Nature Communications*, 4(1771), 2013.

[23] Omid Kavehei, Azhar Iqbal, YS Kim, Kamran Eshraghian, SF Al-Sarawi, and Derek Abbott. The fourth element: characteristics, modelling and electromagnetic theory of the memristor, 2010.

[24] Wei-Chih Chien, Ming-Hsiu Lee, Feng-Ming Lee, Yu-Yu Lin, Hsiang-Lan Lung, Kuang-Yeu Hsieh, and Chih-Yuan Lu. Multi-level 40nm WO_X resistive memory with excellent reliability. In *Electron Devices Meeting (IEDM), 2011 IEEE International*, pages 31.5.1–31.5.4, 2011.

[25] Sang-Jun Choi, Gyeong-Su Park, Ki-Hong Kim, Soohaeng Cho, Woo-Young Yang, Xiang-Shu Li, Jung-Hwan Moon, Kyung-Jin Lee, and Kinam Kim. In situ observation of voltage-induced multilevel resistive switching in solid electrolyte memory. *Advanced Materials*, 23(29):3272–3277, 2011.

[26] Jui-Yuan Chen, Cheng-Lun Hsin, Chun-Wei Huang, Chung-Hua Chiu, Yu-Ting Huang, Su-Jien Lin, Wen-Wei Wu, and Lih-Juann Chen. Dynamic evolution of conducting nanofilament in resistive switching memories. *Nano letters*, 13(8):3671–3677, 2013.

[27] Rainer Waser and Masakazu Aono. Nanoionics-based resistive switching memories. *Nature materials*, 6(11):833–840, 2007.

[28] Xuema Li Douglas A. A. Ohlberg Duncan R. Stewart R. Stanley Williams J. Joshua Yang, Matthew D. Pickett. Memristive switching mechanism for metal/oxide/metal nanodevices. *Nature Nanotechnology*, (7):429–433, 2008.

[29] C. Schindler, G. Staikov, and R. Waser. Electrode kinetics of Cu–SiO_2-based resistive switching cells: Overcoming the voltage-time dilemma of electrochemical metallization memories. *Applied Physics Letters*, 94(7):072109, 2009.

[30] Akihito Sawa. Resistive switching in transition metal oxides. *Materials Today*, 11(6):28–36, 2008.

[31] Zhang Ting, Ding Ling-Hong, and Zhang Wei-Feng. Homogeneous interface-type resistance switching in Au/$La_{0.67}Ca_{0.33}MnO_3$/$SrTiO_3$/F:SnO_2 heterojunction memories. *Chinese Physics B*, 21(4):047301, 2012.

[32] Jae Hyuck Jang Jong Myeong Jeon Min Hwan Lee Gun Hwan Kim Xiang-Shu Li Gyeong-Su Park Bora Lee Seungwu Han Deok-Hwang Kwon, Kyung Min Kim. Atomic structure of conducting nanofilaments in TiO_2 resistive switching memory. *Nature Nanotechnology*, (2):148–153, 2010.

[33] Tom Driscoll, H-T Kim, B-G Chae, Massimiliano Di Ventra, and DN Basov. Phase-transition driven memristive system. *Applied Physics Letters*, 95(4):043503, 2009.

[34] D. C. Kim, S. Seo, S. E. Ahn, D.-S. Suh, M. J. Lee, B.-H. Park, I. K. Yoo, I. G. Baek, H.-J. Kim, E. K. Yim, J. E. Lee, S. O. Park, H. S. Kim, U-In Chung, J. T. Moon, and B. I. Ryu. Electrical observations of filamentary conductions for the resistive memory switching in NiO films. *Applied Physics Letters*, 88(20):202102, 2006.

[35] A. Chen, S. Haddad, Y. C. Wu, Z. Lan, T. N. Fang, and S. Kaza. Switching characteristics of Cu_2O metal–insulator–metal resistive memory. *Applied Physics Letters*, 91(12):123517, 2007.

[36] Wen-Yuan Chang, Yen-Chao Lai, Tai-Bor Wu, Sea-Fue Wang, Frederick Chen, and Ming-Jinn Tsai. Unipolar resistive switching characteristics of ZnO thin films for nonvolatile memory applications. *Applied Physics Letters*, 92(2):022110, 2008.

[37] Chih-Yang Lin, Chen-Yu Wu, Chung-Yi Wu, Tzyh-Cheang Lee, Fu-Liang Yang, Chenming Hu, and Tseung-Yuen Tseng. Effect of top electrode material on resistive switching properties of ZrO_2 film memory devices. *Electron Device Letters, IEEE*, 28(5):366–368, 2007.

[38] Heng-Yuan Lee, Pang-Shiu Chen, Ching-Chiun Wang, Siddheswar Maikap, Pei-Jer Tzeng, Cha-Hsin Lin, Lurng-Shehng Lee, and Ming-Jinn Tsai. Low-power switching of nonvolatile resistive memory using hafnium oxide. *Japanese Journal of Applied Physics*, 46(4B):2175–2179, 2007.

[39] Myoung-Jae Lee, Chang Bum Lee, Dongsoo Lee, Seung Ryul Lee, Man Chang, Ji Hyun Hur, Young-Bae Kim, Chang-Jung Kim, David H Seo, Sunae Seo, et al. A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta_2O_5- x/TaO_2- x bilayer structures. *Nature materials*, 10(8):625–630, 2011.

[40] Wei-Chih Chien, Yi-Chou Chen, Feng-Ming Lee, Yu-Yu Lin, Erh-Kun Lai, Yeong-Der Yao, Jeng Gong, Sheng-Fu Horng, Chiao-Wen Yeh, Shih-Chang Tsai, Ching-Hsiung Lee, Yu-Kai Huang, Chun-Fu Chen, Hsiao-Feng Kao, Yen-Hao Shih, Kuang-Yeu Hsieh, and Chih-Yuan Lu. A novel Ni/WO_X/W resistive random access memory with excellent retention and low switching current. *Japanese Journal of Applied Physics*, 50(4):04DD11, 2011.

[41] Dooho Choi, Dongsoo Lee, Hyunjun Sim, Man Chang, and Hyunsang Hwang. Reversible resistive switching of $SrTiO_X$ thin films for nonvolatile memory applications. *Applied Physics Letters*, 88(8):082904, 2006.

[42] Jun Yao, Zhengzong Sun, Lin Zhong, Douglas Natelson, and James M. Tour. Resistive switches and memories from silicon oxide. *Nano Letters*, 10(10):4105–4110, 2010.

[43] Adnan Mehonic, Sébastien Cueff, Maciej Wojdak, Stephen Hudziak, Olivier Jambois, Christophe Labbé, Blas Garrido, Richard Rizk, and Anthony J. Kenyon. Resistive switching in silicon suboxide films. *Journal of Applied Physics*, 111(7):074507, 2012.

[44] Doo Seok Jeong, Herbert Schroeder, and Rainer Waser. Coexistence of bipolar and unipolar resistive switching behaviors in a Pt/ TiO_2/ Pt stack. *Electrochemical and solid-state letters*, 10(8):G51–G53, 2007.

[45] Antonio C Torrezan, John Paul Strachan, Gilberto Medeiros-Ribeiro, and R Stanley Williams. Sub-nanosecond switching of a tantalum oxide memristor. *Nanotechnology*, 22(48):485203, 2011.

[46] Carsten Kugeler Rainer Waser Eike Linn, Roland Rosezin. Complementary resistive switches for passive nanocrossbar memories. *Nature Materials*, (5):403–406, 2010.

[47] D R Lamb and P C Rundle. A non-filamentary switching action in thermally grown silicon dioxide films. *British Journal of Applied Physics*, 18(1):29, 1967.

[48] C. Schindler, M. Weides, M. N. Kozicki, and R. Waser. Low current resistive switching in Cu–SiO_2 cells. *Applied Physics Letters*, 92(12):122910, 2008.

[49] Sung Hyun Jo, Kuk-Hwan Kim, and Wei Lu. Programmable resistance switching in nanoscale two-terminal devices. *Nano Letters*, 9(1):496–500, 2009.

[50] Byung Joon Choi, Albert B. K. Chen, Xiang Yang, and I-Wei Chen. Purely electronic switching with high uniformity, resistance tunability, and good retention in Pt-dispersed SiO_2 thin films for ReRAM. *Advanced Materials*, 23(33):3847–3852, 2011.

[51] Jun Yao, Zhengzong Sun, Lin Zhong, Douglas Natelson, and James M Tour. Resistive switches and memories from silicon oxide. *Nano letters*, 10(10):4105–4110, 2010.

[52] Evgenij Barsoukov and J Ross Macdonald. *Impedance spectroscopy: theory, experiment, and applications*. Wiley. com, 2005.

[53] http://barrett-group.mcgill.ca/tutorials/nanotechnology/nano02.htm. Accessed 20th October 2013.

[54] Daniele Ielmini. Modeling the universal set/reset characteristics of bipolar RRAM by field-and temperature-driven filament growth. *Electron Devices, IEEE Transactions on*, 58(12):4309–4317, 2011.

[55] XA Tran, WG Zhu, B Gao, JF Kang, WJ Liu, Z Fang, ZR Wang, YC Yeo, BY Nguyen, MF Li, et al. A self-rectifying HfO_X-based unipolar rram with nisi electrode. *IEEE electron device letters*, 33(4):585–587, 2012.

[56] Sebastien Cueff. Transfert d'énergie entre nanoclusters de silicum et erbium dans des matrices oxydes et nitrures de si: applications à des diodes électroluminescentes. *PhD Thesis*, 2011.

[57] Christophe Poulain, Guillaume Jourdan, Alexis Peschot, and Vincent Mandrillon. Contact conductance quantization in a MEMS switch. In *Electrical Contacts (HOLM), 2010 Proceedings of the 56th IEEE Holm Conference on*, pages 1–7. IEEE, 2010.

[58] K Terabe, T Hasegawa, T Nakayama, and M Aono. Quantized conductance atomic switch. *Nature*, 433(7021):47–50, 2005.

[59] Jelmer JT Wagenaar, Monica Morales-Masis, and Jan M van Ruitenbeek. Observing "quantized" conductance steps in silver sulfide: Two parallel resistive switching mechanisms. *Journal of Applied Physics*, 111(1):014302–014302, 2012.

[60] S Tappertzhofen, I Valov, and R Waser. Quantum conductance and switching kinetics of AgI-based microcrossbar cells. *Nanotechnology*, 23(14):145703, 2012.

[61] Jun Yao, Lin Zhong, Douglas Natelson, and James M Tour. In situ imaging of the conducting filament in a silicon oxide resistive switch. *Scientific reports*, 2, 2012.

[62] Yuefei Wang, Xinye Qian, Kunji Chen, Zhonghui Fang, Wei Li, and Jun Xu. Resistive switching mechanism in silicon highly rich SiO_X (x 0.75) films based on silicon dangling bonds percolation model. *Applied Physics Letters*, 102(4):042103–042103, 2013.

[63] JL Costa-Krämer, N Garcia, P Garcia-Mochales, PA Serena, MI Marqués, and A Correia. Conductance quantization in nanowires formed between micro and macroscopic metallic electrodes. *Physical Review B*, 55(8):5416, 1997.

[64] J Suñé, E Miranda, M Nafría, and X Aymerich. Modeling the breakdown spots in silicon dioxide films as point contacts. *Applied physics letters*, 75(7):959–961, 1999.

[65] E Miranda, S Kano, C Dou, K Kakushima, J Suñé, and H Iwai. Nonlinear conductance quantization effects in CeO_X/SiO_2-based resistive switching devices. *Applied Physics Letters*, 101(1):012910–012910, 2012.

[66] LI Glazman and AV Khaetskii. Nonlinear quantum conductance of a lateral microconstraint in a heterostructure. *EPL (Europhysics Letters)*, 9(3):263, 1989.

[67] Hongqi Xu. Theory of nonlinear ballistic transport in quasi-one-dimensional constrictions. *Physical Review B*, 47(23):15630, 1993.

[68] NK Patel, L Martin-Moreno, M Pepper, R Newbury, JEF Frost, DA Ritchie, GAC Jones, JTMB Janssen, J Singleton, and JAAJ Perenboom. Ballistic transport in one dimension:

additional quantisation produced by an electric field. *Journal of Physics: Condensed Matter*, 2(34):7247, 1990.

[69] J Hajto, B McAuley, AJ Snell, and AE Owen. Theory of room temperature quantized resistance effects in metal-a-Si: H-metal thin film structures. *Journal of non-crystalline solids*, 198:825–828, 1996.

[70] Eui-Jung Yun, MF Becker, and RM Walser. Room temperature conductance quantization in V/amorphous-V_2O_5/V thin film structures. *Applied physics letters*, 63(18):2493–2495, 1993.

[71] Xiaojian Zhu, Wenjing Su, Yiwei Liu, Benlin Hu, Liang Pan, Wei Lu, Jiandi Zhang, and Run-Wei Li. Observation of conductance quantization in oxide-based resistive switching memory. *Advanced Materials*, 24(29):3941–3946, 2012.

[72] Tohru Tsuruoka, Tsuyoshi Hasegawa, Kazuya Terabe, and Masakazu Aono. Conductance quantization and synaptic behavior in a Ta_2O_5-based atomic switch. *Nanotechnology*, 23(43):435705, 2012.

[73] John R Jameson, Nad Gilbert, Foroozan Koushan, Juan Saenz, Janet Wang, Shane Hollmer, Michael Kozicki, and Narbeh Derhacobian. Quantized conductance in Ag/GeS_2/W conductive-bridge memory cells. *IEEE electron device letters*, 33(2):257–259, 2012.

[74] Xiaojuan Lian, Shibing Long, Carlo Cagli, Julien Buckley, Enrique Miranda, Ming Liu, and Jordi Suñé. Quantum point contact model of filamentary conduction in resistive switching memories. In *Ultimate Integration on Silicon (ULIS), 2012 13th International Conference on*, pages 101–104. IEEE, 2012.

[75] Simon M Sze and Kwok K Ng. *Physics of semiconductor devices*. Wiley. com, 2006.

[76] Lorenzo Pavesi, Sergey Gaponenko, and Luca Dal Negro. *Towards the first silicon laser*, volume 93. Springer, 2003.

[77] Mau Phon Houng, Yeong Her Wang, and Wai Jyh Chang. Current transport mechanism in trapped oxides: A generalized trap-assisted tunneling model. *Journal of applied physics*, 86(3):1488–1491, 1999.

[78] Donelli J DiMaria, DW Dong, C Falcony, TN Theis, JR Kirtley, JC Tsang, DR Young, FL Pesavento, and SD Brorson. Charge transport and trapping phenomena in off-stoichiometric silicon dioxide films. *Journal of applied physics*, 54(10):5801–5827, 1983.

[79] Shimeng Yu, Ximeng Guan, and H-S Philip Wong. Conduction mechanism of TiN/HfO_X/Pt resistive switching memory: A trap-assisted-tunneling model. *Applied Physics Letters*, 99:063507, 2011.

[80] LA Nesbit. Annealing characteristics of Si-rich SiO_2 films. *Applied Physics Letters*, 46(1):38–40, 1985.

[81] GA Kachurin, KS Zhuravlev, NA Pazdnikov, AF Leier, IE Tyschenko, VA Volodin, W Skorupa, and RA Yankov. Annealing effects in light-emitting Si nanostructures formed in SiO_2 by ion implantation and transient preheating. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 127:583–586, 1997.

[82] Sébastien Cueff, Christophe Labbé, Benjamin Dierre, Filippo Fabbri, Takashi Sekiguchi, Xavier Portier, and Richard Rizk. Investigation of emitting centers in SiO_2 codoped with silicon nanoclusters and Er^{3+} ions by cathodoluminescence technique. *Journal of Applied Physics*, 108(11):113504, 2010.

[83] CR Mokry, PJ Simpson, and AP Knights. Role of vacancy-type defects in the formation of silicon nanocrystals. *Journal of Applied Physics*, 105(11):114301–114301, 2009.

[84] A Sarikov, V Litovchenko, I Lisovskyy, I Maidanchuk, and S Zlobin. Role of oxygen migration in the kinetics of the phase separation of nonstoichiometric silicon oxide films during high-temperature annealing. *Applied Physics Letters*, 91(13):133109–133109, 2007.

[85] Michael TP McCann, Damian A Mooney, Mahfujur Rahman, Denis P Dowling, and JM Don MacElroy. Novel, nanoporous silica and titania layers fabricated by magnetron sputtering. *ACS applied materials & interfaces*, 3(2):252–260, 2011.

[86] G Bersuker, DC Gilmer, D Veksler, P Kirsch, L Vandelli, A Padovani, L Larcher, K McKenna, A Shluger, V Iglesias, et al. Metal oxide resistive memory switching mechanism based on conductive filament properties. *Journal of Applied Physics*, 110(12):124518–124518, 2011.

[87] Jun Yao, Lin Zhong, Douglas Natelson, and James M Tour. Intrinsic resistive switching and memory effects in silicon oxide. *Applied Physics A*, 102(4):835–839, 2011.

[88] E Miranda, D Jiménez, and J Suñé. The quantum point-contact memristor. *Electron Device Letters, IEEE*, 33(10):1474–1476., 2012.

[89] Enrique A Miranda, Christian Walczyk, Christian Wenger, and Thomas Schroeder. Model for the resistive switching effect in HfO_2 MIM structure based on the transmission properties of narrow constrictions. *Electron Device Letters, IEEE*, 31(6):609–611, 2010.

[90] Robin Degraeve, Philippe Roussel, Ludovic Goux, Dirk Wouters, Jorge Kittl, Laith Altimime, Malgorzata Jurczak, and Guido Groeseneken. Generic learning of TDDB applied to RRAM for improved understanding of conduction and switching mechanism through multiple filaments. In *Electron Devices Meeting (IEDM), 2010 IEEE International*, pages 28–4. IEEE, 2010.

[91] LM Prócel, L Trojman, J Moreno, F Crupi, V Maccaronio, R Degraeve, L Goux, and E Simoen. Experimental evidence of the quantum point contact theory in the conduction mechanism of bipolar HfO_2-based resistive random access memories. *Journal of Applied Physics*, 114(7):074509–074509, 2013.

[92] Nicolás Agraıt, Alfredo Levy Yeyati, and Jan M Van Ruitenbeek. Quantum properties of atomic-sized conductors. *Physics Reports*, 377(2):81–279, 2003.

[93] Supriyo Datta. *Electronic transport in mesoscopic systems*. Cambridge university press, 1997.

[94] JI Pascual, JA Torres, and JJ Sáenz. Nonlinear ballistic conductance in atomic-scale metallic contacts. *Physical Review B*, 55(24):R16029, 1997.

[95] Xing Wu, Kun Li, Nagarajan Raghavan, Michel Bosman, Qing-Xiao Wang, Dongkyu Cha, Xi-Xiang Zhang, and Kin-Leong Pey. Uncorrelated multiple conductive filament nucleation and rupture in ultra-thin high-K dielectric based resistive random access memory. *Applied Physics Letters*, 99(9):093502–093502, 2011.

[96] Xinjun Liu, Kuyyadi P Biju, Joonmyoung Lee, Jubong Park, Seonghyun Kim, Sangsu Park, Jungho Shin, Sharif Md Sadaf, and Hyunsang Hwang. Parallel memristive filaments model applicable to bipolar and filamentary resistive switching. *Applied Physics Letters*, 99(11):113518–113518, 2011.

[97] Masayoshi Sasaki. An electron conduction model of resistive memory for resistance dispersion, fluctuation, filament structures, and set/reset mechanism. *Journal of Applied Physics*, 112(1):014501–014501, 2012.

[98] RE Borland. One-dimensional chains with random spacing between atoms. *Proceedings of the Physical Society*, 77(3):705, 1961.

[99] Enrique Miranda and J Suñé. Analytic modeling of leakage current through multiple breakdown paths in SiO_2 films. In *Reliability Physics Symposium, 2001. Proceedings. 39th Annual. 2001 IEEE International*, pages 367–379. IEEE, 2001.

[100] Mingzhen Xu and Changhua Tan. Defect band structure investigation of postbreakdown SiO_2. *Applied Physics Letters*, 92(8):082905–082905, 2008.

[101] Kim Hansen and Mads Brandbyge. Current-voltage relation for thin tunnel barriers: Parabolic barrier model. *Journal of applied physics*, 95(7):3582–3586, 2004.

[102] Seunghyup Lee, Heejin Kim, Dong-Jin Yun, Shi-Woo Rhee, and Kijung Yong. Resistive switching characteristics of ZnO thin film grown on stainless steel for flexible nonvolatile memory devices. *Applied Physics Letters*, 95(26):262113–262113, 2009.

[103] Seung Chul Chae, Jae Sung Lee, Sejin Kim, Shin Buhm Lee, Seo Hyoung Chang, Chunli Liu, Byungnam Kahng, Hyunjung Shin, Dong-Wook Kim, Chang Uk Jung, et al. Random circuit breaker network model for unipolar resistance switching. *Advanced Materials*, 20(6):1154–1159, 2008.

[104] Nevill Francis Mott and Ronald Wilfrid Gurney. *Electronic processes in ionic crystals*. Dover New York, 1964.

[105] Shimeng Yu, Ximeng Guan, and H-SP Wong. On the stochastic nature of resistive switching in metal oxide rram: Physical modeling, monte carlo simulation, and experimental characterization. In *Electron Devices Meeting (IEDM), 2011 IEEE International*, pages 17–3. IEEE, 2011.

[106] Daniel Delling, Peter Sanders, Dominik Schultes, and Dorothea Wagner. Engineering route planning algorithms. In *Algorithmics of large and complex networks*, pages 117–139. Springer, 2009.

[107] Myoung-Jae Lee, Chang Bum Lee, Dongsoo Lee, Seung Ryul Lee, Man Chang, Ji Hyun Hur, Young-Bae Kim, Chang-Jung Kim, David H Seo, Sunae Seo, et al. A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta_2O_{5-X}/TaO_{2-X} bilayer structures. *Nature materials*, 10(8):625–630, 2011.

[108] Yuchao Yang, Peng Gao, Siddharth Gaba, Ting Chang, Xiaoqing Pan, and Wei Lu. Observation of conducting filament growth in nanoscale resistive memories. *Nature Communications*, 3:732, 2012.