Modeling and simulation of valence-change based resistive switching

Von der Fakultät für Elektrotechnik und Informationstechnik
der Rheinisch-Westfälischen Technischen Hochschule Aachen

zur Erlangung des akademischen Grades einer

Doktorin der Ingenieurwissenschaften

genehmigte Dissertation

vorgelegt von

Elhameh Abbaspour, M. Sc.

aus Iran

Berichter: Universitätsprofessor Dr.-Ing. Christoph Jungemann
Universitätsprofessor Dr.-Ing. Rainer Waser

Tag der mündlichen Prüfung: 09.09.2019

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.

I hereby declare that I have created this work completely on my own and used no other sources or tools than the ones listed, and that I have marked any citations accordingly.

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Aachen, September 2019
Elhameh Abbaspour

Abstract

Among the current nonvolatile memory (NVM) devices Redox-based resistive switching RAM (ReRAM) possesses unique features such as promising scalability, low energy consumption, compatibility with CMOS technology and fast switching speed. These features make ReRAM a promising alternative for the main memory application of the future computer systems. ReRAM has also the capability of performing logic and arithmetic operations beyond data storage and can accelerate neural network applications.

In this work a Kinetic Monte Carlo (KMC) simulator is presented to investigate the electroforming, set and reset processes in ReRAM based on the valence change mechanism in real time and size scales. This simulator is based on a detailed microscopic physical model that explains the formation and dissolution of an oxygen-deficient/oxygen-vacancy-rich filament in the resistive switching oxide material. The most important chemical and physical processes involved in the KMC model are introduced and formulated. In contrary to all other proposed KMC based models, generation and recombination of oxygen vacancies can only happen due to an oxygen exchange reaction along the active electrode/oxide interface. Further, the role of different interface conditions on the dynamics of the filament growth and switching characteristics is investigated. Special attention is paid to consider proper charge transport mechanism in different stages of the switching processes to calculate the leakage current along the oxide. It is demonstrated that this alternative model could successfully reproduce $I-V$ characteristics observed in experiment. It can also simulate the reliable generation and annihilation of the conductive filament during a large number of switching cycles, where a considerable resistance window between the low and high resistance states is maintained during the whole switching time.

The developed model is then employed to study the variability of the resistive switching by performing statistic and transient simulations. Specifically two intrinsic sources of variability, namely the cycle to cycle variability and random telegraph noise (RTN)-related fluctuations, are investigated. The role of different forming and switching conditions like the current compliance and the features of the applied voltages during the forming, set and reset processes on switching variability is studied.

Kurzfassung

Unter den aktuellen Bauelementen mit nichtflüchtigem Speicher (NVM) verfügt das Redox-basierte resistive Switching RAM (ReRAM) über einzigartige Merkmale wie vielversprechende Skalierbarkeit, geringer Energieverbrauch, Kompatibilität mit CMOS-Technologie und schnelle Schaltgeschwindigkeit. Diese Eigenschaften machen ReRAM zu einer vielversprechenden Alternative für die Hauptspeicheranwendung zukünftiger Computersysteme. ReRAM kann auch über die Datenspeicherung hinaus logische und arithmetische Operationen ausführen und neuronale Netzwerkanwendungen beschleunigen.

In dieser Arbeit wird ein Kinetic Monte Carlo (KMC) -Simulator vorgestellt, der die Electroforming-, Set- und Resetprozesse in ReRAM auf der Grundlage des Valenzänderungsmechanismus in realistischen zeit- und in Größenmaßstäben untersucht. Dieser Simulator basiert auf einem detaillierten mikroskopischen physikalischen Modell, das die Bildung und Auflösung eines oxygen-deficient/oxygen-vacancy-rich Filaments im resistiven Schaltoxidmaterial erklärt. Die wichtigsten chemischen und physikalischen Prozesse des KMC-Modells werden vorgestellt und formuliert. Im Gegensatz zu allen anderen vorgeschlagenen KMC-basierten Modellen kann die Erzeugung und Rekombination von Oxygenvacancies nur aufgrund einer Oxygen Austauschreaktion entlang der aktiven Elektrode / Oxid-Grenzfläche erfolgen. Ferner wird die Rolle verschiedener Grenzflächenbedingungen für die Dynamik des Filamentwachstums und die Schaltcharakteristik untersucht. Besondere Beachtung wird auf den richtigen Ladungstransportmechanismus in verschiedenen Phasen der Schaltvorgänge gelegt, um den Leckstrom durch des Oxids zu berechnen. Es wird gezeigt, dass dieses alternative Modell, die im Experiment beobachteten I-V-Eigenschaften, erfolgreich reproduzieren kann. Es kann auch die zuverlässige Erzeugung und Vernichtung des leitenden Filaments während einer großen Anzahl von Schaltzyklen simulieren, wobei während der gesamten Schaltzeit ein beträchtlicher Widerstandsunterschied zwischen den Zuständen niedrigen und hohen Widerstands eingehalten wird.

Das entwickelte Modell wird dann verwendet, um die Variabilität des resistiven Schaltens zu untersuchen, indem statistische und transiente Simulationen durchgeführt werden. Insbesondere werden zwei intrinsische Variabilitätsquellen untersucht, nämlich die Variabilität von Zyklus zu Zyklus und Fluktu-

ationen im Zusammenhang mit Random telegraph Noise (RTN). Die Rolle der verschiedenen Forming- und Schaltbedingungen wie die Stromconformität und die Eigenschaften der angelegten Spannungen während desForming-, Set- und Resetprozesses für die Schaltvariabilität werden untersucht.

Contents

Abstract **ii**

Kurzfassung **iii**

1 Introduction to memory devices **1**
1.1 Volatile memory devices . . . 1
1.2 Nonvolatile memory devices . . . 2
1.2.1 Flash memory . . . 4
1.2.2 Emerging NVMs . . . 10
1.2.3 Introduction to ReRAM . . . 12
1.2.4 ReRAM memory architecture . . . 19
1.2.5 Performance of ReRAM . . . 20
1.2.6 Multilevel cell operations . . . 26

2 Physical model description of VCM **29**
2.1 Basics and assumptions . . . 29
2.1.1 Interface model . . . 31
2.2 Conduction mechanism . . . 34
2.2.1 Trap assisted tunnelling conducting mechanism . . . 35
2.2.2 Drift mechanism . . . 40
2.3 Electrical potential . . . 43
2.3.1 Numerical method for solving the Poisson equation . . . 44
2.4 Joule heating . . . 45

3 KMC simulation of VCM **47**
3.1 Basics of Kinetic Monte Carlo method . . . 48
3.2 Developing a KMC simulator for VCM ReRAM . . . 50
3.2.1 Processes involved in KMC . . . 50
3.2.2 Simulation procedure . . . 53
3.3 Result and discussion . . . 56
3.3.1 2D system . . . 56
3.3.2 3D system . . . 63

3.3.3 Comparison of 2D and 3D results 74

4 Variability of the switching parameters 75
4.1 Cycle-to-Cycle variability 75
4.2 Random telegraph noise 84

5 Summary and Outlook 89
5.1 Summary 89
5.2 Outlook 90

Bibliography 91

Chapter 1

Introduction to memory devices

Memories are divided into two main branches: 1) the volatile memories, which despite the very high read and write speed lose data when the power is turned off. Static random-access memory (SRAM) and dynamic random-access memory (DRAM) are two well-known examples of volatile memories 2) the nonvolatile memories like ReRAM or ferroelectric RAM or Flash, which can retain the stored information even without power supply with a balanced programming and reading performance [1]. Because of this capability the nonvolatile memories offer a broad application in many functional information processing systems. Fig. 1.1 illustrates the memory tree of the main volatile and nonvolatile classifications. In this graph a branch of NVM called the prototypical includes the memories from the ones with prototype test-chips to the even commercialized ones. The memories in the emerging branch of this tree are mainly made out of novel materials such as chalcogenides, metal oxides, etc., which despite the simple two terminal structure they usually have a conduction mechanism beyond the classical mechanisms, such as ion hopping, phase transition, etc.

In the following some of the memory types shown in the memory tree with a special attention to the NVMs are discussed.

1.1 Volatile memory devices

SRAM is a fast volatile memory device with low power consumption. Despite the quite small market for SRAM it is very important and used in almost all memory chips and is specially suitable in cache applications. A conventional 6 transistor SRAM, which consists of six metal oxide semiconductor field effect transistors (MOSFET) and used in CMOS technology, is shown in Fig. 1.2(a). The two cross coupled inverters (M1, M2, M3 and M4) are connected in series with two access transistors (M5 and M6). The represented 6 transistor SRAM structure can store only one bit. This increases the required area and the cost

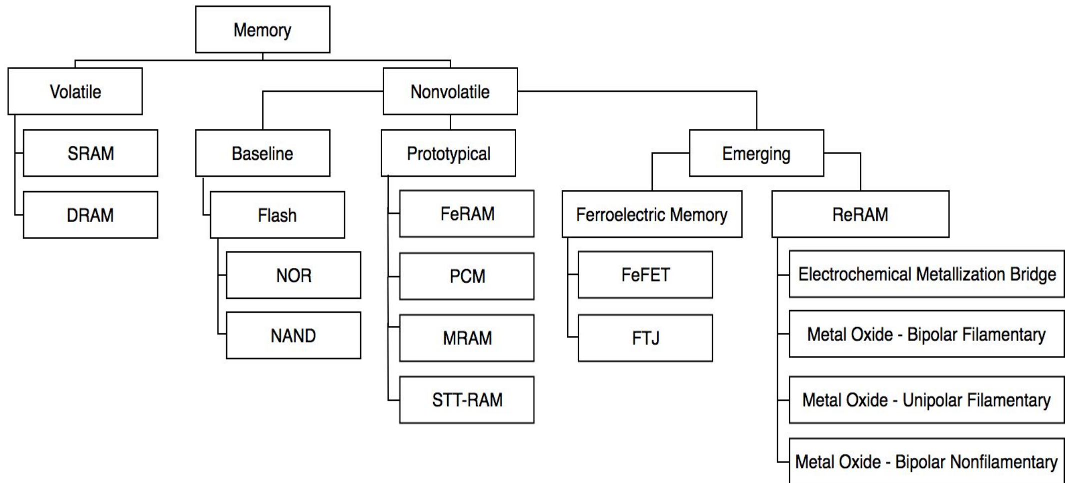


Figure 1.1: Memory tree including the emerging NVMs. [2]

of the processing. Thus, fewer transistors per cell are preferred to decrease the cost and improve the scalability.
DRAM is another type of volatile memories, that only consists of one MOSFET and one capacitor as shown in Fig. 1.2(b). This simple structure can store one bit of data, which allows DRAM to reach very high densities. Because of this DRAM is used much more often than SRAM specifically in the main memory and graphic cards despite its more complicated circuitry. Because of the charge depletion through the capacitor in DRAM structure, it needs to be charged periodically to maintain information. That's why the power consumption of DRAM is relatively high. There are some other types of upcoming volatile memories as future candidates to replaces SRAM and DRAM like Z-RAM (formerly called compcache), Advanced RAM (A-RAM) and twin-transistor random access memory (TTRAM).

1.2 Nonvolatile memory devices

The idea of NVM was first proposed by Kahng and Sze in 1967, where they fabricated a floating gate, in which semi-permanent charge storage was possible [3]. This floating gate was charged up under an application of a high electric field through an outer gate. The charges remained there even after removing the power supply because of a much lower back transport probability. This device worked as a bistable memory with a metal-insulator-semiconductor (MIS) field effect transistor as a non-destructive read-out structure. The holding time of this memory was around an hour. Fig. 1.3 shows the energy band diagram of the floating gate structure to store the charge. In this structure (Fig. 1.3a) the

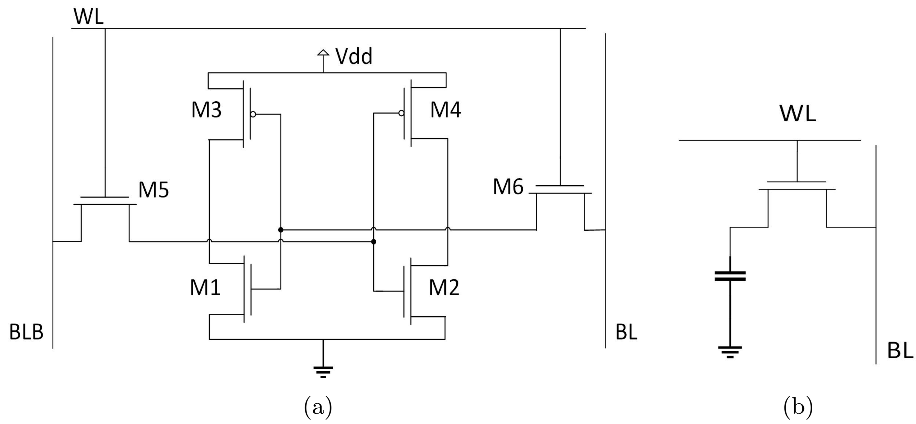


Figure 1.2: Basic circuit structure of a (a) SRAM made up of 6 MOSFETs and a (b) DRAM

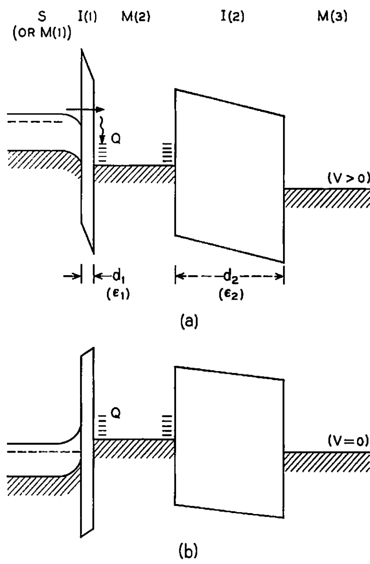


Figure 1.3: Energy band diagram of a floating gate device with a S-M-I-M structure. For the sake of measurement the S layer was replaced by a metal layer (M1). (a) a positive voltage was applied to M(3) causing charge accumulation in the floating gate (M(2)) (b) the applied voltage was removed. © 2019 IEEE. Reprinted with permission from [3].

charges are accumulated in layer M(2) (the floating gate) if a positive voltage is applied to the external gate M(3). The accumulated charges stay in the floating gate even after removing the applied voltage (Fig. 1.3b) assuming that the electron transport across I(2) is small [3]. The schematic of the device fabricated by Kahng and Sze is shown in Fig. 1.4.

A continuous advancement of the memory technologies is required due to

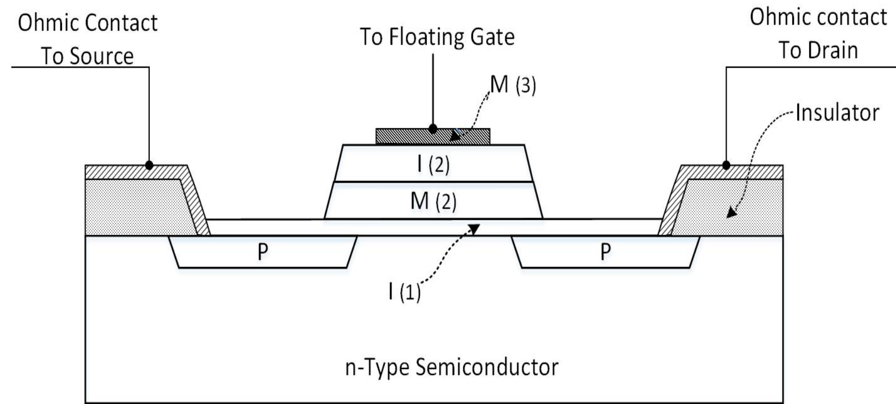


Figure 1.4: A floating gate device fabricated using an insulated gate field effect transistor (IGFET). The numbers on the layers correspond to those shown in Fig. 1.3. © 2019 IEEE. Reproduced with permission from [3].

the fast growing of data generation. Ultra low power, high density and low cost data storage are the current market trends for the hardware requirements. The new emerging NVMs like ReRAM with lower power consumption and cost and improved flexibility may become more important technology enablers for efficient and intelligent hardware systems [2]. Recently, the Yole Développement announced that after 15 years of developing the phase change memories (PCM), one of the emerging NVM technologies, has finally taken off in stand-alone applications. Fig.1.5 represents the NVM market revenues breakdown per application [4]. Fig. 1.6 shows the evolution of the revenue share of different types of the emerging NVMs including ReRAM, PCM and MRAM for stand-alone and embedded applications. However, the market of the emerging NVMs is still limited and smaller than DRAM and FLASH.

Among the known NVMs here we discuss Flash memory as today's the most cost-effective and dominant solid-state nonvolatile memory technology and four prominent emerging NVM candidates in more detail.

1.2.1 Flash memory

Unlike SRAM and DRAM, Flash memories don't require power supply to retain information. The main reasons that make Flash memory one of the most suitable choices for NVM applications are: 1) high density, which lowers the cost 2) the possibility of being written and erased on system more than 100000 times with minimum granularity [1]. Because of the explosive growth of the electronic portable devices like mobile phones, laptops , digital cameras and so on over the past two decades the market for Flash memories has risen up dramatically [5].

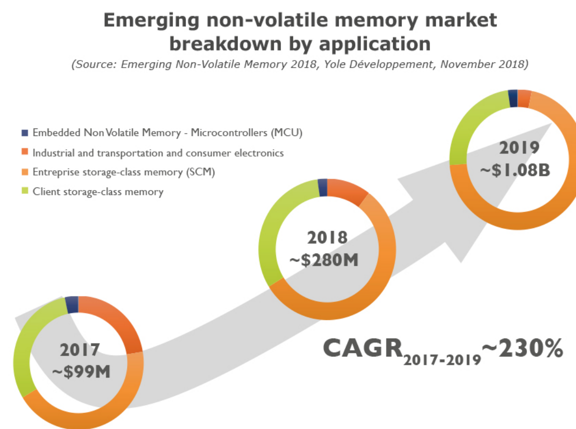


Figure 1.5: The market of the emerging NVMs reported by Yole Développement [4].

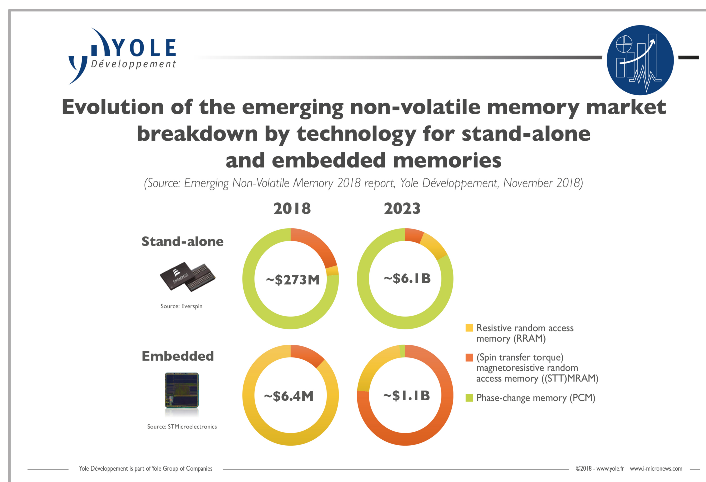


Figure 1.6: Stand-alone vs. embedded applications of NVMs [4].

The Flash memory market revenues worldwide from 2013 to 2021 (in billion U.S. dollars) are shown in Fig. 1.7.The forecasted revenue of Flash market in 2021 is almost tripled in comparison to 2013. In the following section the operation principle of a Flash cell is briefly discussed.

Basic operation mechanism

A Flash cell is basically a special type of MOSFET, which consists of two gates separated by an insulating layer as shown in Fig. 1.8. The floating gate (FG), laying between the control gate (CG) and the silicone substrate, is electrically

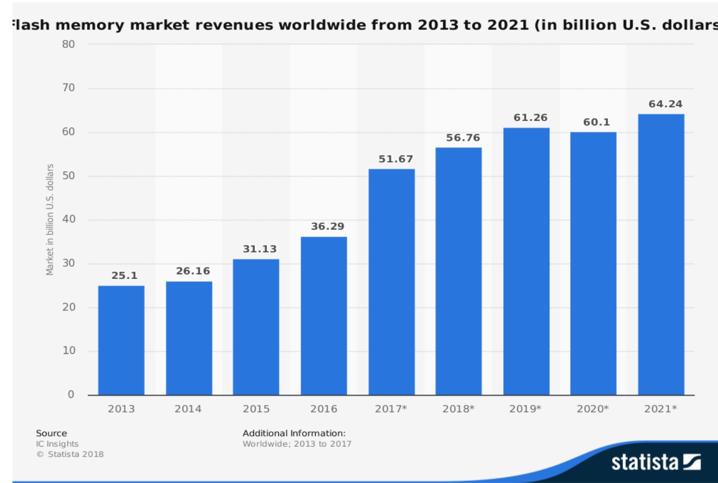


Figure 1.7: Flash memory market revenue [6].

completely isolated. Because of this the injected charge to FG can be trapped and stored there and cause threshold voltage (V_T) modulation. When some charges are stored in FG, V_T is shifted as follow,

$$\triangle V_T = V_T - V_{T0} = -\bar{Q}/C_{FC}, \tag{1.1}$$

where $\bar{Q}$ is the stored charge in FG, V_{T0} is the threshold voltage when $\bar{Q} = 0$ and C_{FC} is the capacitance between FG and CG. If the stored charge is negative, V_T is shifted in positive direction [7]. If the applied read voltage is grater than V_{T0} and less than V_T, the cell current will be zero for non-zero $\bar{Q}$. We can change the memory state and have a non-zero current by having zero $\bar{Q}$. The negative $\bar{Q}$ is associated with the logical value zero and positive or zero $\bar{Q}$ is equivalent to the logical state one. This is schematically shown in Fig. 1.9. The dielectric between the FG and transistor channel (gate dielectric) is called tunnel oxide because of Fowler Nordheim tunnelling of electrons through it. There is another layer of dielectric between FG and CG that is usually thicker than the tunnel oxide. The main functionality of this layer is to improve the quality of the tunnel oxide. Flash memories can be classified in terms of the access type into two classes: 1) parallel 2) serial. Different programming and erasing mechanisms also divide Flash memories in different categories: 1) Fowler-Nordheim tunnelling 2) channel hot electron 3) hot holes 4) source side hot electrons. Among all of these different classes of Flash two main industry standard memories are discussed here: NOR, suitable for both data and code storage[8], NAND, optimized for the data storage [9]. These two types of Flash memories are discussed as next in more detail.

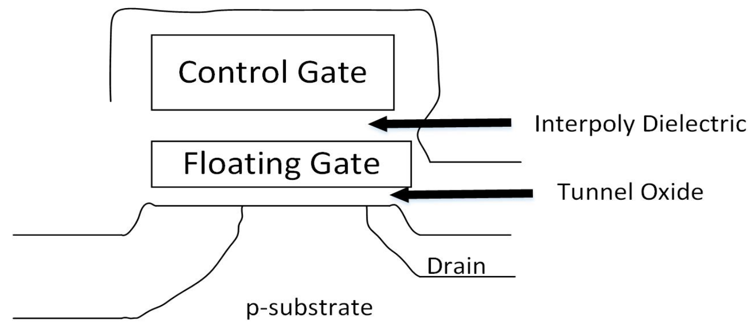


Figure 1.8: Cross section presentation of a Flash cell [1].

NOR and NAND Flash cells

Based on the array structure, Flash memories are divided into NOR and NAND cells. The equivalent circuits of these two types of Flash cell are represented in Fig. 1.10. In this figure the WL, which stands for word line, is a line that connects the gate of the Flash cells sharing the same gate electrode. On the other hand BL, which stands for bit line, connects the drain of the cells sharing the same drain electrode. The NOR cell consists of some FG transistors connecting in parallel, while in the NAND cell the FG transistors are connected serially. During the NOR programming operation charges are moved into FG using hot electron injection (HEI). According to eq. 1.1 this causes a V_T modulation. Two sets of pulses are applied to the control gates and drains simultaneously while the source is grounded. This can be done by applying the pulses to the WL and BL. A NAND flash is programmed by Fowler Nordheim electron tunnelling into FG. This quantum mechanical tunnelling is induced by applying a strong electric field across the thin oxide between CG and FG. In this case a large voltage is applied to WL, while BL is grounded and the source is floating. The injected electrons in FG shift V_T in the positive direction. The amplitude and length of the applied pules and the thickness of the oxide between FG and CG are the factors, which can control the V_T shift.
The erase operation in NOR cells needs applying a high voltage to the source, while CGs are grounded and drains are floating. The Fowler Nordheim tunnelling of charges form FG to the source causes this electrical erase. The source junction needs to be specially designed to undergo the high applied electric field without breaking down. The erasing process for the NAND cells is more straight forward than NOR. All cells can be erased simultaneously by applying a large voltage to the substrate and grounding WL and keeping BL and sources floating. In this case electrons can tunnel back to the channel. After the erasing operation all cells are read to check whether the whole block were erased successfully or not. If not the erasing and reading operations are repeated. This process repeats until

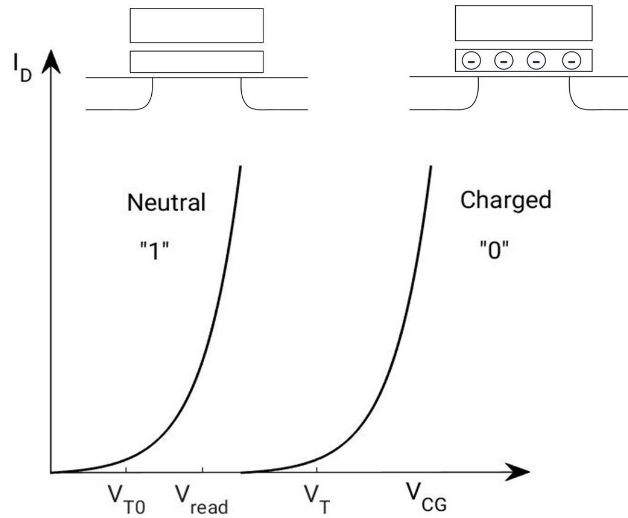


Figure 1.9: $I - V$ characteristics of a Floating gate MOSFET in neutral and charged states [1].

achieving a threshold voltage less than the "erase verify level“ for all the cells within the block.

Performance of Flash memories

Dealing with a large number of memory cells in many arrays requires considering the parameter fluctuations between cells. One usual way to estimate the device to device variability of Flash is to compare the threshold voltage of the cells. The dispersion in the threshold voltage distribution depends on the operation mode (i.e. program or erasure) and its mechanisms. For example the threshold voltage distribution after UV erasure is quite narrow and symmetrical in comparison to the distribution after the Fowler Nordheim erasure. Although one can explain the exponential tail of the erased distributions as defective cases, which can be minimized by process optimization, but the rest of the many tail cells are not trivial to explain. There are different models to explain the tail cells in the distribution. For example one model introduces randomly distributed positive charges in the tunnel oxide, which act as donor-like bulk oxide traps, as the source of the tail cells existing [10].
As a NVM, Flash memory is expected to retain the data for at least ten years. The reason for data loss in Flash is the charge loss in FG, which can be due to several possibilities like the existence of: 1) defects in the oxide between CG and FG or in the tunnelling oxide 2) de-trapping of charges from the insulators around FG 3) mobile ion contamination [1]. In order to have a better data reten-

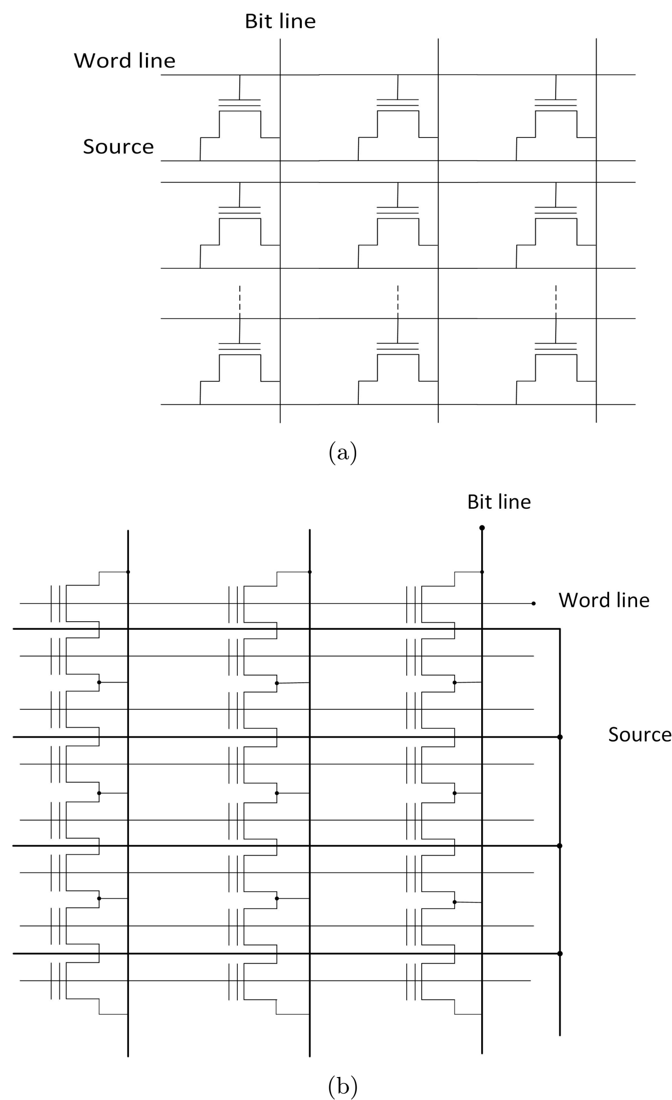


Figure 1.10: The equivalent circuit of a) NOR and b) NAND Flash memories.

tion, the charge loss in FG should be as minimum as possible. Down scaling the cell size and subsequently reducing the capacitance can increase the sensitivity even to small charge changes and cause unwanted read operation and data loss. Therefore data retention can be severely deteriorated.

Programming/erasing endurance determines how many program and erase cycles the device can sustain before confronting the operational failure. The typical erase/program endurance for Flash products is 10^5 cycles. The tunnel oxide degradation during the cycling can wear out the cell performance and subse-

quently the cell endurance [11].
Scaling is a fundamental issue for Flash memories, which will eventually become the bottleneck of the future application of theses devices. The nonvolatility of Flash cells influences the scalability of the thin oxides placed between FG and CG (interpoly dielectric) and between FG and transistor channel (tunnel oxide). The thickness of the tunnel oxide is limited to 6 nm by the direct tunnelling mechanism. In realistic cases this thickness increases by 1 to 2 nm due to the trap assisted tunnelling in the oxide [12]. The scaling limit for interpoly dielectric based on reports is around $12 - 13$ nm [13].

1.2.2 Emerging NVMs

Over the past few years the application of Flash memories was limited by serious trade-off between a couple of issues like: 1) scalability of the lateral device dimension 2) maintaining the coupling ratio between CG and FG to ensure passing enough voltage to FG during the program and erase operations 3) the stress induced leakage current caused by a large voltage application across an ultra-thin oxide during the program operation 4) the parasitic interference between the retained charges in closely packed cells in arrays [14, 15]. The Flash researchers are already struggling with maintaining the current level of the device reliability and performance while increasing the density. However the amount of difficulties in scaling to future technology nodes makes Flash memories unlikely to fit into the opening up new applications. Therefore there is a demand for a new generation of NVM that can overcome the scalability issues of Flash. Among the emerging NVMs here we discuss the advantages and challenges of the four most promising NVMs and their potential applications: 1) Phase change memory (PCM) 2) Spin-transfer-torque random-access-memory (STTRAM) 3) Ferroelectric-FET (FeFET) memory 4) redox-based random-access-memory (ReRAM).

PCM

The PCM benefits from the large resistance difference between crystalline (low resistance) and amorphous (high resistance) states of the phase change materials [16]. The transition from amorphous to crystalline state is the set process and limits the speed of PCM. The reverse transition is the reset process and a power limiting step. The properties of the phase change materials, which exist in an amorphous and one or sometimes several crystalline phases determine the functionality and success of PCM technology to a large extent [17]. The general behaviour of phase change materials in the scaled dimensions is desirable. Phase transition in very small dimensions like several nanometres, improved endurance and longer retention, lower thermal conductivity, which means lower power consumption, etc., are some examples of the advantages of the phase

change materials in the scaled dimensions. The limiting factor for PCM cell size is mainly the access devices, which need relatively large switching current [2]. Several access devices like bipolar junction transistors and vertical transistors have been considered to increase the PCM storage density. One of the main PCM failures happens when the void formation in the electrode prevents the cell from switching to the low resistance state. Another main failure mechanism is the elemental segregation caused by the electromigration. This results in poor data retention when the cell can not switch to the high resistance state or after switching can not remain in this state.
Among the novel NVMs, PCM is one of the most developed ones having promising performance like switching speed less than 100 ns and cycling endurance greater than 10^9. However using PCM as a replacement to DRAM is a difficult challenge. Because for that a combination of requirements (extremely high cycling endurance $> 10^{16}$ and very fast switching time in the order of nanoseconds) are required that PCM has not achieved yet. Although very fast switching has been achieved by some phase change materials, but the extremely high cycle endurance remains as a serious challenge [17]. There are some reports of improving the cycling endurance to more than 10^{15} by extremely down scaling the size of the phase change materials [18]. PCMs have shown potential for new applications such as neuromorphic computing and phase change logic and storage class memory.

STTRAM

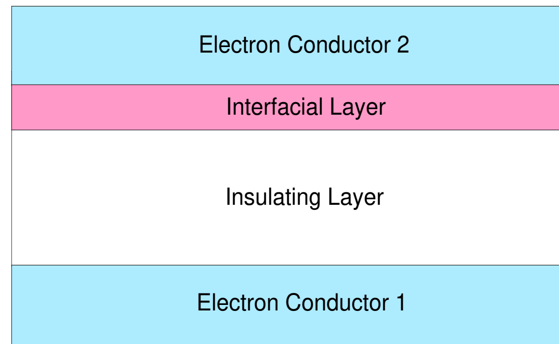
STTRAM is a more efficient and more scalable version of the conventional field-switching magnetoresistive RAM (MRAM). The structure of the memory cell consists of a magnetic tunnel junction (MTJ) with two ferromagnetic layers separated by a thin oxide tunnel barrier. The data is retained in the magnetic state of the MTJ and read back regarding the corresponding resistance of the MTJ. The high and low resistance states correspond to the parallel and anti-parallel orientations of the ferromagnetic layers [19]. The highest performance among NVMs belongs to STTRAM with write speed less than 10 ns and more than 10^{12} endurance cycle. The large size of the access transistors used with STTRAM can increase the bit cell size extremely, however it is still smaller than SRAM.
STTRAMs still face some serious challenges like considerable variability increasing with scaling and small off/on resistance ratio. The device to device variation, which can be caused by the fabrication process like smoothness, etching damages, etc,, is also quite high.
There are continuing advancement in STTRAM design and performance. Two recent examples are the control of interfacial anisotropy using the electric field and spin current generation with the spin Hall effect [2].

FeFET

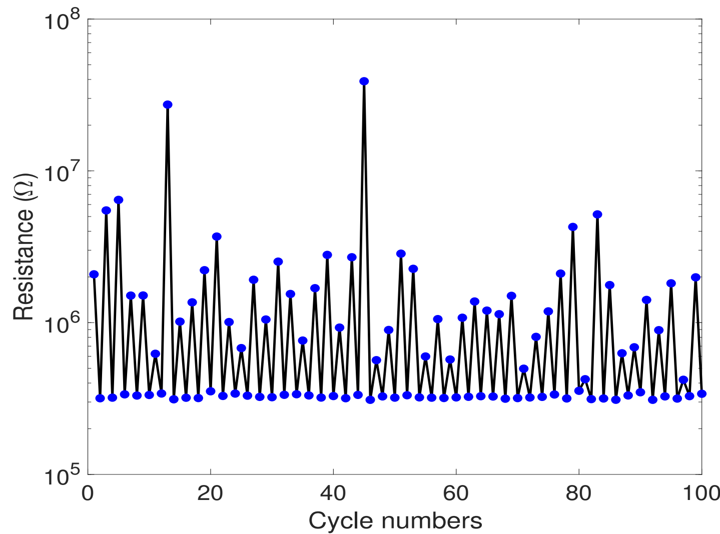
In a FeFET device the FET channel conductance is modulated by ferroelectric polarization of a ferroelectric layer, which is added to the gate dielectric. The relatively short retention time of FeFET, which has hindered its development as a viable NVM for years, is attributed to two main reasons: 1) depolarization field 2) the gate leakage current and trapping of carriers in the gate dielectric stack [20]. Recent discovery of HfO_x-based FeFET has brought back the interest in FeFET. Engineering ferroelectricity in HfO_x thin films allows manufacturing a highly scaled CMOS compatible FeFET with the performance close to that of DRAM [21].

1.2.3 Introduction to ReRAM

The observation of the resistive switching phenomena in the insulators dates back to the 1960s [22, 23]. Despite the intensive studies until the early 1980s, resistive memories lost the battle to the capacitive memories due to the poor stability, high operation voltage and current, and low film handling technique. The recent progress in the material science brought the attentions to the resistive memories again. In 2004 Samsung reported the fabrication of a simple binary transition metal oxide (TMO) based resistive RAM, having promising scalability, high CMOS compatibility and reasonable performance even at high temperatures like 300 ℃ without any considerable degradation. The device operated successfully under 3V bias voltage and 2mA switching current for different TMO materials and electrodes [24]. Up till now ReRAM has hold its place among the hot topics in the area of the memory applications owing to low power, high density, fast access speed, long retention and multi-bit operation.
ReRAMs are expected to be the memory technology that can be simply integrated with the conventional CMOS technology with no further dedicated process or facility. This can be done by selecting a set of TMO and electrode materials compatible with the fabrication environment for the conventional CMOS technology and process temperature, so that the memory fabrication can be performed on top of the metal layers or inside the contact vias of a MOSFET source and drain in a CMOS chip. There are some density multiplying technologies like multi-level or multi-layer cell (MLC) structure that are used as an alternative to increase the memory density despite the scaling limits. ReRAM is also a good candidate for such density multiplying technologies because of its simple structure and low process temperature and CMOS compatibility [25].
ReRAM has a simple capacitor-like metal-insulator-metal (MIM) structure as shown in Fig. 1.11(a). In the simplest case it consists of an insulating layer sandwiched between two electron conductors. Sometimes there is another optional interfacial layer between one electrode and the insulating to improve the



(a)



(b)

Figure 1.11: a) The schematic of MIM structure for ReRAM b) Resistance change of ReRAM between two high and low levels for the successive switching cycles.

performance of ReRAM. During the operation of a ReRAM the cell resistance changes between two high and low levels as shown in Fig. 1.11(b). This resistance modulation is known to be due to the local modification of the materials caused by the electrochemical processes along with the physical diffusion of metal cations or oxygen vacancies. The choice of the material of each layer can have a significant effect on the switching behaviour of ReRAM.

ReRAM classifications

There are different ways to classify ReRAM, depending on the type of the conduction path, the polarity of the switching or the operation principle. Here three important classifications of ReRAM devices are introduced.

Filamentary and area dependent switching One classification is based on the conduction path. Depending on the type of the localization of the chemical modification responsible for the change of the conductance, ReRAM can be categorized into two classes as shown in Fig. 1.12: (a) filamentary switching, (b) area dependent switching [26]. In the filamentary switching usually a conductive

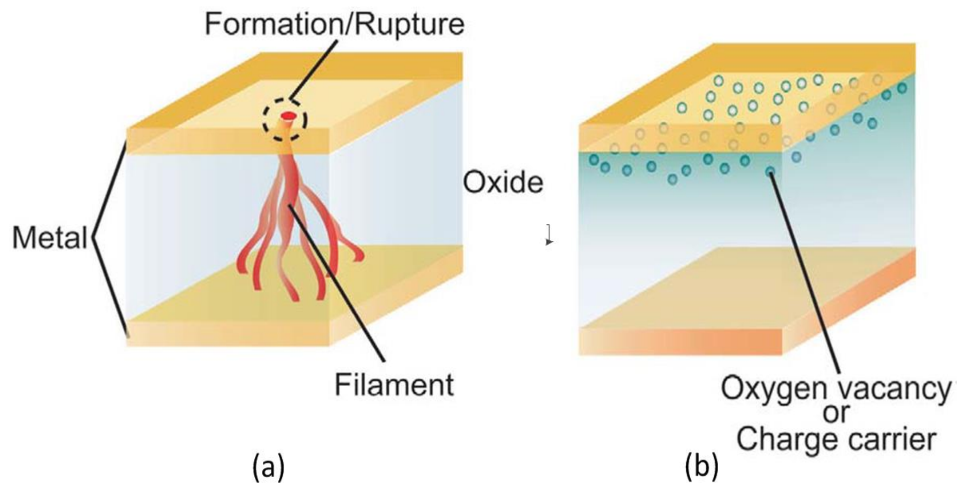


Figure 1.12: a) Filamentary b) area dependent switching mechanism. Copyright (2019) Elsevier. Reprinted with permission from [26].

filament (CF) is formed during the electroforming process. The CF formation process is a soft breakdown phenomena that creates a locally degraded region with high defect concentration [27]. The electroforming step is an initial process that happens only once in a cell's lifetime, however it has a significant effect on the memory cell performance. Fig. 1.13 shows a virgin cell, where it contains only a few initial defects, therefore it is highly insulating and the cell is switched off. The state of this cell at the end of the forming process is also shown in this figure, where a filamentary conducting path has been generated due to the soft dielectric breakdown that connects the top and bottom electrodes. In this case the switching happens locally and is independent of the electrode area. The structure of the oxide based ReRAM (OXRAM) consists of a TMO as the insulating layer like HfO_x or TiO_x sandwiched between two metallic electrodes. In this type of ReRAM the CF is made out of metallic impurities or oxygen vacancies, which are responsible for charge transport [26]. Conduction bridge RAM

(CBRAM) is another type of ReRAM, which has the filamentary switching [26]. In CBRAM usually Ag or Cu based interface layers are used to inject Ag or Cu cations into a chalcogenide or oxide electrolyte layer like GeSe or SiO_2 to create CF.
The experimental observations by Fujiwara et al. in Fig. 1.14 clearly show the formation of a filament like structure within the metal/CuO/metal stack during the forming process [28]. They showed that the cell resistance decreases after the CF formation and returns almost to the initial value when a part of this CF is ruptured with a focused ion beam. This observation indicates the role of this filament like structure as the conducting path in the dielectric.

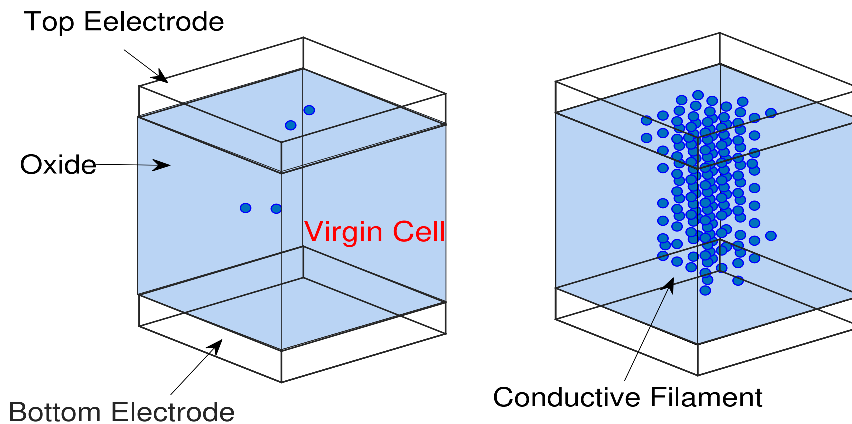
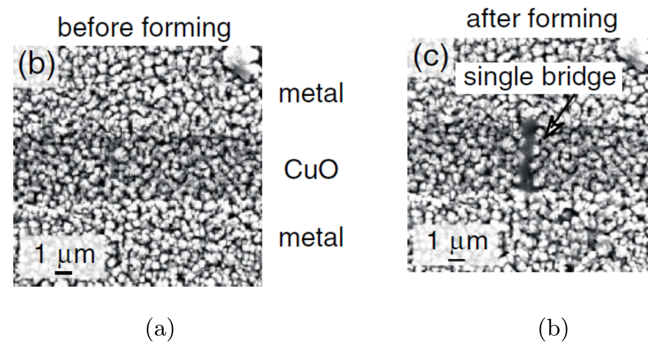


Figure 1.13: Transformation of a virgin cell to the conducting phase during the filamentary electroforming process

In the area dependent case the switching takes place more or less homogeneously along the whole area of the electrode-oxide interface as shown in Fig. 1.12(b). This type of uniform switching was observed in some perovskite-type oxides such as PCMO and TaO_x /TiO_2 bilayers [29, 30], where a local chemical reaction happens along the interface between two separate materials, for example the oxygen exchange reaction driven by electric field that happens between a reactive top electrode and the oxide layer. The general principle of the area dependent switching is as follow: When a positive voltage is applied to an electrode, the oxidation of the electrode/oxide interface takes place by the oxygen ion drift in the reverse direction of the applied electric field from the oxide toward the electrode. Therefore, the material of this electrode should be relatively reactive like Al or Sm and inert materials like Pt should be avoided. Because of the uniform switching across the whole interface area the switching characteristics like the on/off resistance values and the set and reset currents



(a) (b)

Figure 1.14: a) Scanning electron microscope (SEM) image of part of a Ni/CuO/Ni device before forming b) after forming, where a bridge like structure has been created between two electrodes. Copyright (2008) The Japan Society of Applied Physics. redrawn with permission from [28]

depend on the device area [27].
The general shape of the $I - V$ characteristics is different for the filamentary and area dependent switching. The set operation in the filamentary switching is usually an abrupt process [31], while in the area dependent switching it is more smooth and gradual [29].

Unipolar and bipolar switching ReRAMs with filamentary switching can themselves be classified based on the switching mode into: (a) unipolar and (b) bipolar switching. Fig. 1.15 shows the $I - V$ characteristics of the bipolar and unipolar switching modes. In the unipolar mode the switching direction depends on the amplitude of the applied voltage and not on the voltage polarity. The set and reset processes in this case happen at the same voltage polarity. In this case the thermally activated redox reactions are responsible for the creation and rupture of the CF. During the reset process the tip of the filament is oxidized and a gap with maximum temperature is created. The reduction of the metal oxide in the gap region in the set process caused by Joule heating reconstruct the ruptured CF.
In the bipolar switching the set and reset processes require different polarity of the voltage. The CF formation and rupture in bipolar ReRAMs are explained by ion migration driven by the local electric filed under the effect of the temperature. The migration of defects toward the electrode with the negative applied voltage in the reset process results in the CF rupture, while the reverse process in the set operation causes the reconstruction of the CF. The coexistence of the unipolar and bipolar switching modes in the same ReRAM device has been also frequently observed [32].
The combination of the temperature and electric field driving forces results in a less reset current for bipolar devices than unipolar ones, which are only under the

thermal effect. Therefore, bipolar ReRAMs have in general less power dissipation than unipolar ReRAMs. The reliability of the bipolar switching is also higher than unipolar one. The reason is the larger voltage margin between the set and reset voltages because of their different polarity. These two advantages of the bipolar ReRAM over unipolar one make it more attractive in practice.
In this work we will focus on the filamentary type of the resistive switching in bipolar mode.

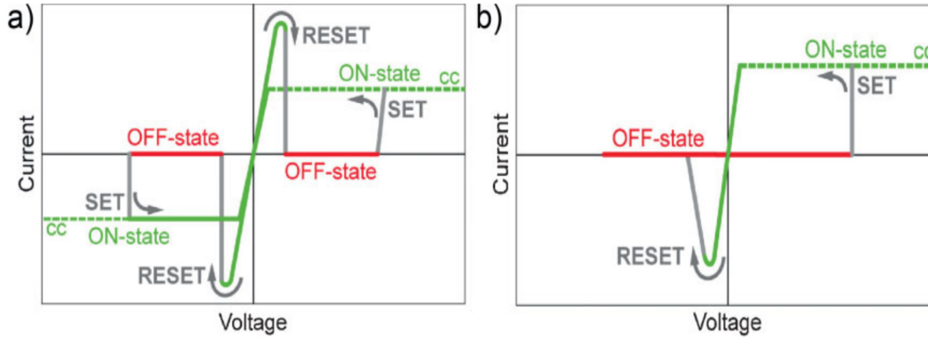


Figure 1.15: $I-V$ characteristics of the (a) unipolar switching and the (b) bipolar switching. Copyright (2019) Wiley. Reprinted with permission from [33]

ECM and TCM and VCM cells The ReRAM devices can also be divided based on the operation principle and device physics into three classes: (a) electrochemical metallization mechanism (ECM), (b) thermochemical mechanism (TCM) and (c) valence change mechanism (VCM).
The operation of ECM cell is basically based on the migration of the highly mobile metal cations across the ion conducting layer and their electrochemical redox reaction. The growth of a metal-like conductive filament during the electroforming/set process transfers the cell from high resistance state (HRS) to low resistance state (LRS). The electrochemical dissolution of this CF resets the device into the off state. The origin of the metal cations can be different depending on the material choice for each layer: 1) dissolution of the active electrode releases the metal cations [34] 2) the dielectric layer supplies the metal cations [35] 3) metal cations are generated both by the oxidation of the electrode and the dielectric layer [36]. ECM devices can show both unipolar and bipolar behaviour. The basic principle of an ECM cell operation is depicted in Fig. 1.16. The memory cell shown in this figure consists of an electrochemically active electrode (Ag), an electrochemically inert electrode (Pt) and an ion conducting insulating layer in between of these two electrodes [33]. In this case the application of high enough positive voltage to the anode Ag dissolves it according to the following

equation and generates Ag^+ cations,

$$Ag \rightarrow Ag^+ + e^- \quad (1.2)$$

The generated Ag^+ cations after desorption from the anode surface move in the direction of the applied electric field across the ion conducting layer. These cations after reaching to the surface of the cathode Pt are reduced as follow and adsorbed.

$$Ag^+ + e^- \rightarrow Ag \quad (1.3)$$

As a result of the described processes a metallic CF starts to be created. It continues to grow until it connects the surrounding electrodes and brings the cell to the on state. If a sufficiently large negative voltage is applied to the active electrode, a reverse sequence of the above described processes happens, which results in the CF rupture and returns the cell to off state. The resistive switching

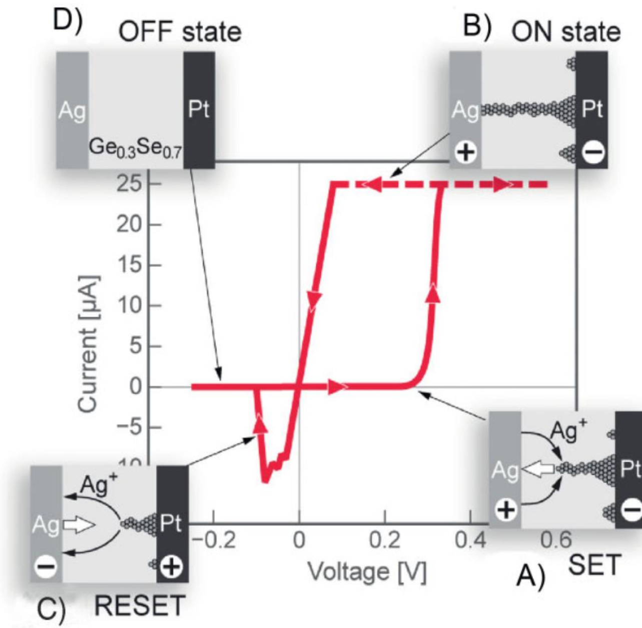


Figure 1.16: Typical $I - V$ characteristics of an ECM cell with Ag/Ag-Ge-Se/Pt structure. The insets show the different stages of the switching process (A to D). Copyright (2019) Wiley. Reprinted with permission from [33]. © 2019 IEEE. Reprinted with permission from [37].

in TCM cells is based on the change of the stoichiometry due to the thermal effects, thus showing unipolar behaviour. This thermochemical switching usually happens in some transition metal oxides like NiO [24]. The typical $I - V$ characteristics for a Pt/NiO/Pt TCM cell is shown in Fig. 1.17 including the forming process limited by a current compliance at 1 mA. The resistive switching in TCM cells is known to be of filamentary type switching [33].

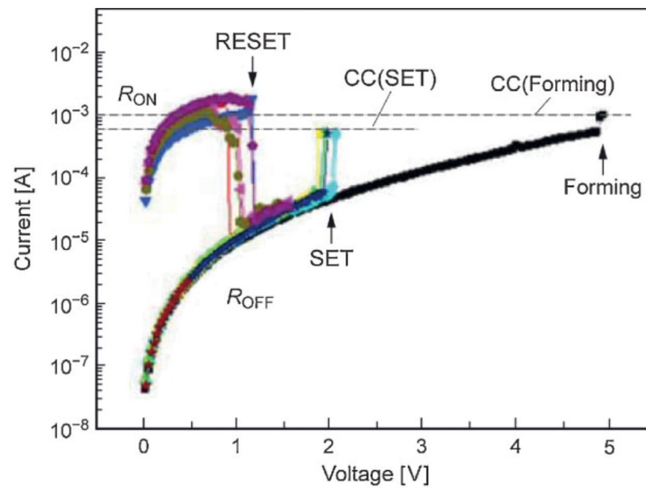


Figure 1.17: Typical $I - V$ characteristics of a Pt/NiO/Pt TCM cell including the forming, set and reset processes. Copyright (2019) Wiley. Reprinted with permission from [33] and [35].

The switching process in VCM cells is based on the oxygen vacancy ($V_O^{\bullet\bullet}$) generation and migration. The variation of $V_O^{\bullet\bullet}$ concentration can cause different valence states of the metal cations and consequently change the electronic conductivity. The naming of this class of memories refers to the valence change effect [33]. This type of ReRAM will be discussed in detail in the next chapter.

1.2.4 ReRAM memory architecture

The ReRAM cells are organized in a memory cell array structure, which provides the short access time as shown in Fig. 1.18. The rows and columns of this cell array are usually called word lines and bit lines, respectively. To perform the read and write operations the word lines and read lines are connected to some electronic amplifiers located at the periphery of the storage matrix. Depending on the characteristics of the ReRAM cells and constraints like energy, cost or delay a ReRAM array can be: a) a passive cross point architecture with no access devices (Fig. 1.18(a)). This is the simplest case, where a ReRAM cell directly connects the word line and a bit line at each node. In this design every single cell can be accessed by applying the proper voltage to the word line and bit line that this cell connects to each other. The ideal situation is that all the current flows through the selected cells by applying the voltage across the word lines and bit lines. However, in the passive cross-bar matrix there will be a leakage current flow also in all the other cells in the selected word line and bit line because of the lack of any isolation of the selected cell [38]. b) an active matrix, which is used to avoid the non-addressed cells experiencing part of the

signal in a selected row or column by adding an active switch at each node. The typical choice for this switching element is a select transistor. The conventional design for a ReRAM array has one dedicated MOSFET transistor for each cell called 1T1R structure (Fig. 1.18(b)). When a row is activated, the exclusive access to each cell can be controlled by the access transistors without disturbing other cells in that row. The benefit of the active matrix over passive matrix comes at the cost of the more circuit complexity [39]. The size of the access transistors used in the ReRAM array structure is usually larger compared to DRAM because of the higher operating current of ReRAMs. This can increase the area and consequently the cost. However the 1T1R structure is in general an energy efficient choice by generating excellent isolation due to the very high resistance in the saturation region and is preferred to other alternatives of the selector devices [40, 41]. In the 1T1R structure the parasitic capacitance between the transistor and ReRAM cell, which can be quite large, should be minimized. This parasitic capacitance can cause current overshoot during the forming and set process which can in turn increase the maximum reset current and consequently the power dissipation. The maximum reset is usually in the order of the current compliance in the forming and set processes. One can estimate the amount of the current overshoot by measuring the difference between the maximum reset current and the nominal value of the compliance current. The greater difference between these two currents means the higher current overshoot [41]. Anything that can modulate the parasitic capacitance can affect the current overshoot. The role of the device structure in determining the current overshoot was reported by Gu et al. [42]. A ReRAM cell with a pillar structure shows less current overshoot than a concave structure, which has a large parasitic capacitance due to having extra supporting oxides. Another structure that minimizes the parasitic capacitances of the interconnects between the transistors and ReRAM cells is the direct fabrication of a ReRAM cell on top of the MOSFET's source or drain contacts.

1.2.5 Performance of ReRAM

In this section the main performance metrics of ReRAM devices are thoroughly discussed.

Variability

The variability of the switching characteristics of ReRAM is one of the main challenges facing this memory device and a strong obstacle in front of its mass production. In general the variability issue can be divided to relevant types: 1) cycle to cycle variability that shows the device stability during the cycling. Significant parameter fluctuations are observed in both the set and reset processes

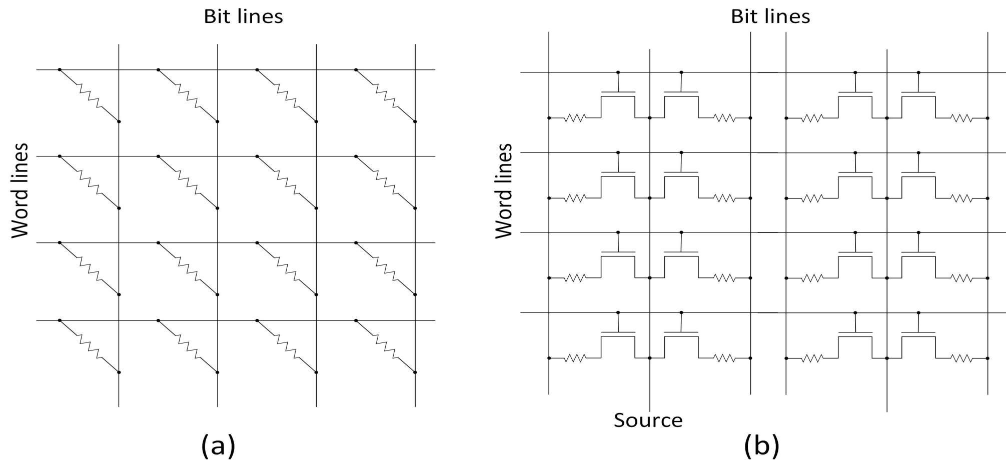


Figure 1.18: ReRAM array architecture: (a) without access device crossbar structure (b) 1T1R structure

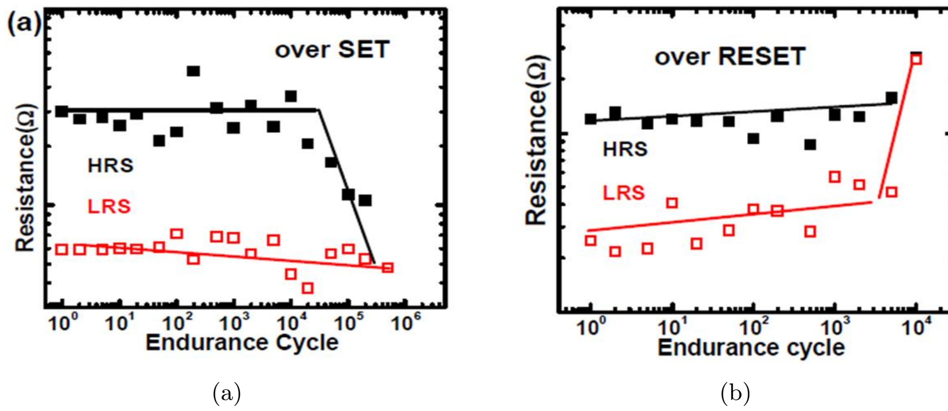
in terms of the switching voltages and resistances. Specifically in a memory array the existence of a noticeable tail bit in the resistance distribution can reduce the off/on resistance ratio and consequently cause a problem for the multilevel programming of ReRAM devices. The origin of the cycle to cycle variations is not clearly understood yet but there are some reports that relate it to the stochastic nature of the oxygen vacancies [43]. The fluctuations during the HRS are more crucial than the LRS. The reason is that the LRS variation comes from the CF size or the number of the CFs in the case of multi-filamentary switching. But the HRS variation is due to the fluctuation of the gap size between the tip of the filament and the electrode, which can change the current by an exponential dependency on the tunnelling distance. 2) device to device variability that characterizes the uniformity within a memory array. The device to device variability mainly depends on the fabrication process and uniformity control of the processes. The spatial or chemical variations such as the fluctuations in the oxide thickness, electrode/oxide interface roughness, grain boundaries, bond polarization factor, that can affect the energy barrier of oxygen vacancy generation, and so on from one cell to another cell can affect the performance of ReRAM devices significantly [44]. For example smoothing the electrode contacts by a chemical-mechanical polish can improve the uniformity of the ReRAM devices in a memory array [45]. A lot of methods have been proposed to improve the uniformity of ReRAM by material engineering. One type of these methods is the electrode/oxide interface engineering by introducing proper buffer layers [46]. Another method is confining the filamentary switching paths. This can be done by introducing some seeds to enhance the local electric field, such as inserting Pt

nanocrystals into the TiO_x memory cell [47]. The CF path confinement can be also done directly by reducing the cell size. Besides to the material engineering and process improvement, another possible way to improve the switching uniformity is optimizing the programming methods in terms of the applied voltages during the forming, set and reset processes. The amplitude and length of the applied pulse can be effective parameters on controlling the uniformity of the switching characteristics[48].
The focus of this work will be on the cycle to cycle variability, which is more of an intrinsic type of ReRAM features related to the stochastic motion of atoms.

Reliability

One of the strongest concern about ReRAM devices is the reliability issue. The operation of ReRAM during cycling requires repeated creation, migration and annihilation of atoms inside the device at different electric field and temperature regimes. This can result in degradation of the electrodes and active materials and consequently the device performance [49].
Reliability of a memory device in general can be characterized by the endurance and retention metrics. The endurance of a memory device refers to the maximum number of the program/erase cycles before the off/on resistance ratio falls down to a certain value causing a hard error. Usually this unreliability means happening of a failure state. There are different types of failure events: 1) the HRS resistance tends to decrease during the cycling and in the extreme case it might reach the order of the LRS resistance value. In this case the device is stuck in the LRS and unable to switch back to the HRS and a reset failure takes place [50] 2) on the other hand it is also possible to have a set failure, causing a high resistance tail in the LRS distribution and a stuck in HRS [51]. Fig. 1.19 shows the typical endurance degradations observed in experiment for both stuck in the LRS and HRS cases, which in this figure are denoted as "over set" and "over reset", respectively 3) the third failure mode in a memory device is the falling resistance window between LRS and HRS because of increasing the LRS resistance and decreasing the HRS resistance. An intermediate resistance level is observed in this case as shown in Fig. 1.20. The measured resistance values in the LRS and HRS in this figure show almost a constant window between them for about 1.7×10^5 cycles, then the resistance window collapses causing a failure state [52]. The device endurance is improved in the 1T1R array structure because of eliminating the half selected disturbance. The best reported ReRAM endurance is up to 10^{12} cycles [31]. This is much larger than the typical endurance of the Flash memories, which is between 10^3 to 10^7 cycles.

The retention time indicates how long a memory cell can retain the stored data. The universal acceptable retention time for an NVM is more than 10 years even at the operating temperature up to 85 °C. Most of the current ReRAM



(a) (b)

Figure 1.19: Measured resistance values in the LRS and HRS as a function of the number of cycles showing the endurance degradation for the stuck in the a) LRS and b) HRS cases. © 2019 IEEE. Reprinted with permission from [50].

devices can reach to this retention time standard. Fig. 1.21 compare the LRS resistance (R_{on}) for a HfO_2 ReRAM at different temperatures. The value of R_{on} increases over time until it reaches a value as large as the resistance order in the HRS, causing a failure state. The role of the temperature on R_{on} retention can also be investigated in this figure, where by increasing the temperature R_{on} falls faster toward R_{off} [53].

Scalability

Device scaling is a great concern for any memory technology. It should offer a continued growth of density for several years in the industrial roadmap. ReRAM is considered as an optimal candidate for the future NVM applications with the potential scalability to the nanometer regime. This potential comes from the filamentary switching concept, where the active device area is defined by the area of the CF, which can be as small as few atomic units.
There have been several projects aiming to scale down the size of resistive switching based memories. Park et al. aggressively reduced the size of the ReRAM device to sub-5nm. The effective electrode radius was confirmed by conductive-atomic force microscopy (C-AFM) and focused ion beam-transmission electron microscopy (FIB-TEM) analysis to be in the range of 2-3 nm [54]. The final sub-5 nm device structure is shown in Fig. 1.22(a). The top part of the Cu CF is the active area of the device. The current mapping image of C-AFM in Fig. 1.22(b) indicates the 2 $\sim$ 3 nm Cu bottom electrode. The DC switching characteristics and the more than 3 orders of on/off current ratio are shown in Fig. 1.22(c), 1.22(d), respectively. In another work the size of the CF in the localized switching in a NiO-based ReRAM was reported by Lee et al. to be

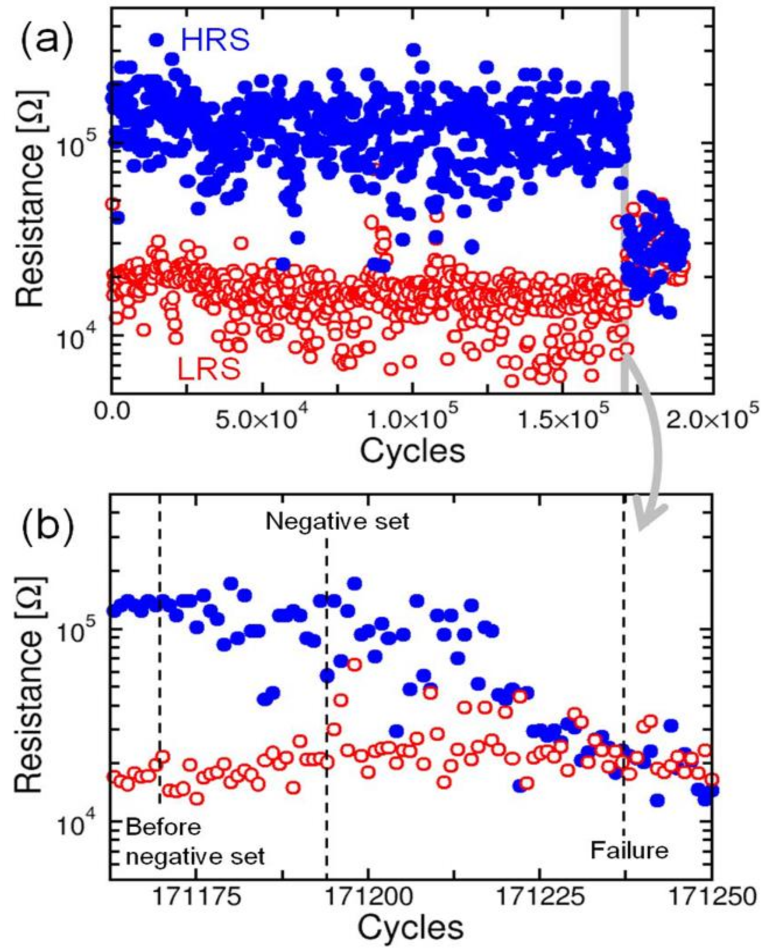


Figure 1.20: (a) Endurance degradation observed in the experiment, where the resistance falls into an intermediate level between LRS and HRS (b) a close up of (a) where the failure occurs. © 2019 IEEE. Reprinted with permission from [52].

lower than 10 nm, providing the scaling of ReRAM to sub-10 nm dimension [55]. Govoreanu et al. announced integrating a Hf/HfO_x resistive element stack with an area of less than 10×10 nm^2 in a 65 nm CMOS process with a good performance [56]. Chen et al. fabricated successfully a 1 Kb 1T1R array with a 30×30 nm^2 ReRAM showing a robust switching characteristics [57].

Scaling can also affect the switching parameters. The HRS resistance increases inversely with down-scaling the cell size, while the LRS resistance depends only slightly on the cell area because of the filamentary conduction path in the LRS that is independent of the area. So the device down-scaling has the benefit of increasing the off/on resistance ratio [24, 40, 58]. The maximum reset current

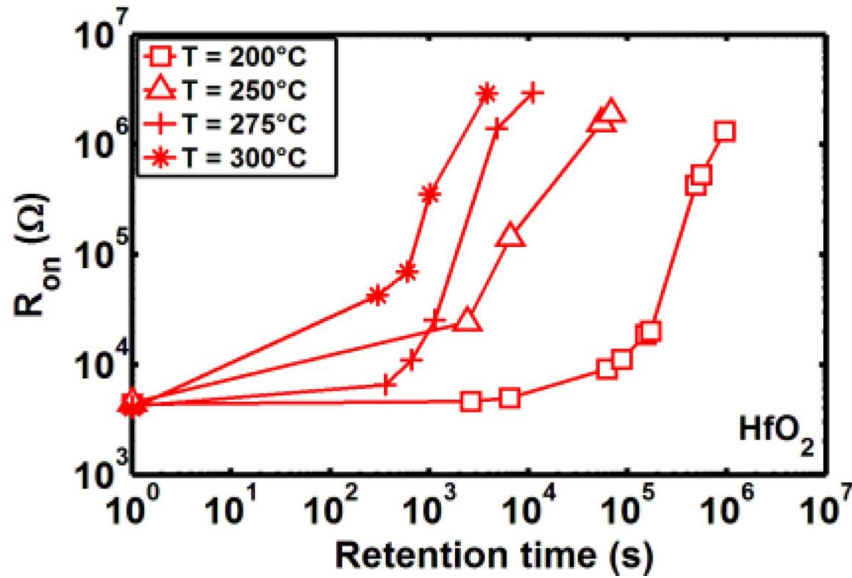


Figure 1.21: Mean retention of the LRS resistance of a HfO_2-based ReRAM at four different temperatures T. © 2019 IEEE. Reprinted with permission from [53].

is another switching parameter that can be influenced by the device scaling. The maximum reset current is an important parameter because it usually determines the peak power consumption. The general reported scaling trend of the reset current shows only a slight reduction with the device down-scaling [24, 40]. This means a significant increase of the reset current density and consequently the power consumption is expected when the devices is scaled down. To overcome this problem it is usually tried to reduce the reset current by decreasing the set compliance current, since there is an almost linear relationship between them [59]. The parasitic capacitances should be minimized to avoid the overshooting effect, which causes the maximum reset current to be higher than the nominal applied compliance current [41]. The lower set current means higher LRS resistance and lower off/on resistance ratio. This means that there is a limit on the set current compliance reduction not to pass the tolerated resistance window. Fortunately the HRS resistance increases with the down-scaling can compensate the increase of the LRS resistance partially.
The high compatibility of resistive switching based memories with the conventional CMOS technology also allows achieving a high density ReRAM architecture.

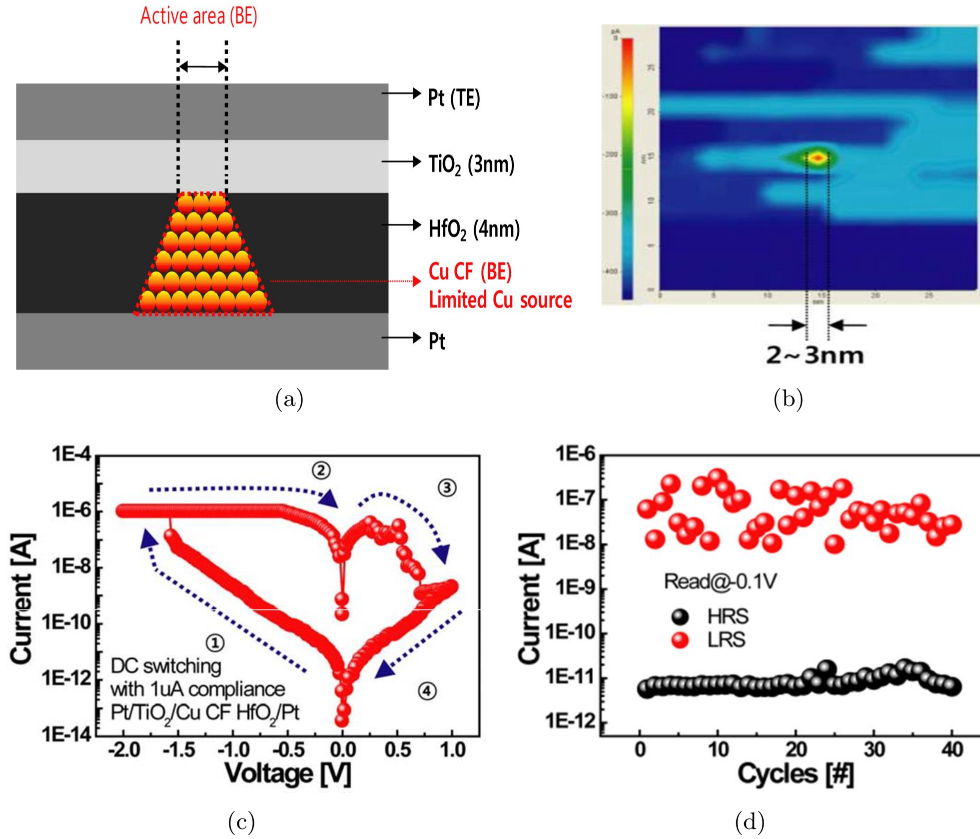


Figure 1.22: (a) Extremely down-scaled ReRAM device structure (b) current mapping image of C-AFM analysis confirming the less than 5 nm effective electrode (c) DC $I-V$ characteristic (d) HRS and LRS current distribution during cycling. © 2019 IEEE. Reprinted with permission from [54].

1.2.6 Multilevel cell operations

For a high density memory application the optimal usage of the layout area of the memory device is critical. ReRAM must show the multilevel cell (MLC) operation capabilities and 3D vertical architecture to compete with the Flash technology. The MLC operation of ReRAM is achieved by controlling the LRS and HRS resistance states. The LRS resistance can be modulated by the set current compliance, while the HRS resistance can be changed by the reset stop voltage [60]. In the LRS the overshoot effect should be minimized to have better control on the resistance states. Most of the materials used in ReRAM devices like TaO_2 and HfO_2 show MLC operation capability [61, 40]. Prakash et al. achieved 7 LRS levels with the same HRS ensuring 3-bit per cell storage in a TaO_x-based ReRAM. This 7 LRS resistance levels were obtained by changing the compliance current between 30 μA and 300 μA. For a successful MLC oper-

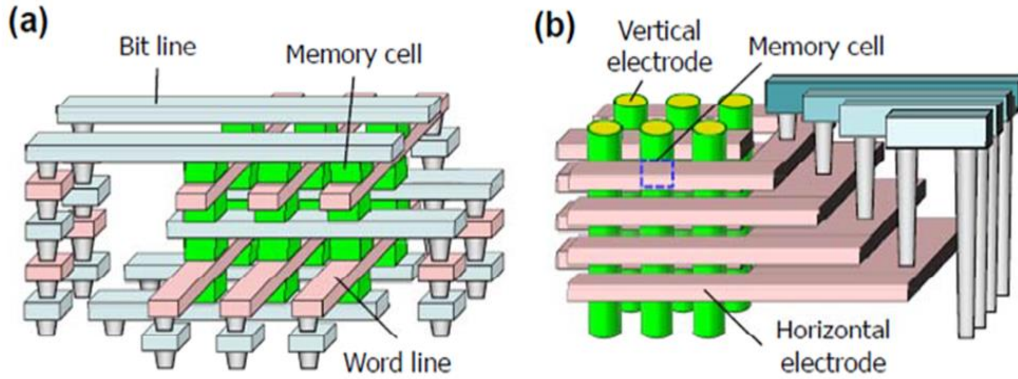


Figure 1.23: Schematic picture of (a) 3D horizontal crosspoint ReRAM (b) Vertical ReRAM. © 2019 IEEE. Reprinted with permission from [62].

ation some conditions should be met: 1) the on/off current ratio should be high enough for every resistance level without degrading the switching uniformity 2) the cycling endurance should be acceptable 3) thermal stability of the stored data at each state.
It is crucial to implement ReRAM technology in 3D scheme in order to achieve high densities comparable to NAND Flash. Fig. 1.23 shows the schematic picture of the horizontal 3D crossbar array and the vertical ReRAM (VRAM) [62]. The horizontal 3D structure consists of stacking several layers of 2D memory arrays. The simple and compact structure of the 3D crossbar array enables high density memory design, which can form $4F^2$ array footprints. In the vertical 3D structure each memory cell is located between a vertical and a horizontal electrode. Regarding the fabrication cost, VRAM is more efficient due to the limited number of critical masks and not needing all the costly fabrication processes required for 3D crossbar array like double patterning technology (DPT) or EUV litho [63]. On the other hand VRAM has also a better scaling feature because of using vertical NAND like structure, where multi-layers are formed at the same time [64]. The similar problem of both 3D crossbar array and VRAM is the voltage loss along electrode lines and the large number of current sneak paths, which makes the selection of a specific cell difficult [62].

Some of the recent achievements in ReRAM technology has been summarized in Table. 1.1. This table introduces some published data of the device structure, cell area and performance characteristics in binary metal-oxide ReRAM devices.

This work introduces a microscopic physical model of the resistive switching process in a valence change memory cell. A kinetic Monte Carlo code is developed to simulate the whole switching process in experimental time and length

	NiO IEDM 2004	Cu_xO IEDM 2005	Ti:NiO IEDM 2007	TaO_x IEDM 2008	Ti/HfO_x IEDM 2008	Ti/HfO_x IEDM 2009 &2010	WO_x IEDM 2010	ZrO_x/HfO_x IEDM 2010	N:AlO_x VLSI 2011	TaO_x/Ta_2O_5 VLSI 2011	Hf/HfO_x IEDM 2011
switching type	unipolar	bipolar	unipolar	bipolar	bipolar	bipolar	bipolar	bipolar	bipolar	bipolar	bipolar
structure	1T-1R	1T-1R	1T-1R	1T-1R	1T-1R	1T-1R	1T-1R	1R	1T-1R	1R	1T-1R
cell area (μm^2)	~0.2	~0.03	~0.49	~0.25	~0.1	0.0009 (30nm)	0.0036 (60nm)	0.0025 (50nm)	~1	~9000	0.0001 (10nm)
speed	~5μs	~50ns	~5ns	~10ns	~5ns	~0.3ns	~50ns	~40ns	N/A	~10ns	~10ns
peak voltage	<3V	<3V	<3V	<2V	<1.5V	<2.5V	<3V	<2V	<2V	<2.5V	<1.5V
peak current	~2mA	~45μA	~100μA	~170μA	~25μA	~200μA	~1mA	~50μA	~50nA	~30μA	~50 μA
HRS/LRS ratio	>10	>10	>90	>10	>100	>100	>10	>10	>100	>100	>10
endurance	10^6	600	100	10^9	10^6	10^{10}	10^6	10^6	10^5	10^{12}	5x10^7
retention	300h@ 150℃	30h@ 90℃	1000h@ 150℃	3000h@ 150℃	10h@ 200℃	28h@ 150℃	2000h@ 150℃	28h@ 125℃	28h@ 125℃	3h@ 200℃	30h@ 250℃

Table 1.1: Performance of the binary metal-oxide ReRAM. © 2019 IEEE. Reprinted with permission from [65].

scales with considering a great amount of microscopic details.
Chapter 2 describes the conduction and switching characteristics in a valence change memory device. Detailed mechanisms of the electroforming and switching processes are covered.
Chapter 3 introduces a kinetic Monte Carlo simulator to model the whole switching process in valence change memories. All the physical and chemical processes involved in the switching process are presented and formulated. The simulation results of the electroforming and switching processes are given and compared to the experimental results.
Chapter 4 is devoted to the analysis of the variability of the resistive switching by performing statistic and transient simulations.
Chapter 5 is the summary and discusses about possible future work.

Chapter 2

Physical model description of VCM

In the VCM cells usually the metal cations are less diffusive or the oxidation of the electrode metal or reduction of the oxidized metal is difficult. Different experiments have shown that oxygen vacancies are the mobile species in these type of ReRAMs [66]. Unlike ECM cells the physics of VCM ReRAMs are quite complicated and there are still many open questions. In this chapter the physical mechanism of VCM switching is discussed, focusing on the most widely accepted models.

2.1 Basics and assumptions

The formation and dissolution of an oxygen-deficient filamentary region is widely accepted to be the switching mechanism in VCM cells. It means that there is a high vacancy concentration in the conducting filamentary region. There is a strong belief that the change in the concentration of vacancies in the CF is the reason behind the resistance change of the oxide in VCM cells, however there are different models to describe the mechanism of this change in VO concentration: (a) by generation of Frenkel defects in the oxide bulk (b) by oxygen exchange reactions at the metal/oxide interface. In model (a) it is proposed that positively charged oxygen vacancies ($V_O^{\bullet\bullet}$) are generated within the bulk due to the release of an oxygen ion O^{2-} from a regular lattice site O_O^x to an interstitial position O_i'', thus forming an $V_O^{\bullet\bullet}-O_i$ Frenkel Pair [67, 68]. In this model the only mobile particle is the interstitial oxygen, which tends to move away from $V_O^{\bullet\bullet}$ in order not to recombine with it. The recombination of the interstitial oxygen with a charged oxygen vacancy during the reset operation results in the dissolution of the filament. Several kinetic Monte Carlo (KMC) models have been developed based on this mechanism. These KMC models were mainly developed to simulate

the electroforming process or simple $I-V$ characteristics [69, 70, 71, 72, 73, 74]. A major problem of model (a) is that the generation of $V_O^{\bullet\bullet}-O_i$ Frenkel pairs is either not energetically favourable within an oxide film or even if the pairs are generated they are not stable [75, 76]. The defect formation energies, migration barriers and electrical energy levels of the oxygen vacancies and interstitials were calculated in HfO_2 and a couple of other oxides in [75]. Fig. 2.1 demonstrates the formation energy of different defects inside a HfO_2 film as a function of the Fermi energy (E_f) in two situations: (I) oxygen rich (O-rich) limit, where as shown in Fig. 2.1(a) the oxygen interstitial (noted by I_O in the figure) has the lowest formation energy and is the dominant defect. (II) oxygen poor (O-poor) limit: the presence of a metal electrode or a metallic scavenging layer creates an O-poor limit by shifting the chemical potential of oxygen (μ_O) from O-rich limit ($\mu_O = 0$ eV) close to the μ_O of the metal electrode or the scavenging layer in equilibrium. In the case of Hf as the scavenging metal it was found that μ_O was lowered by 5.9 eV. The vacancy formation energy in stoichiometric oxides depends on μ_O [77], therefore this reduces the formation energy of a neutral oxygen vacancy from 6.1 eV to only 0.2 eV [75], while the formation energy of the oxygen interstitials inside the bulk remains large, making the probability of this process near zero. The shaded areas in Fig. 2.1 show the relevant range of E_f in the device controlled by the work function of the relevant electrodes.
The second major problem is the recombination of the Frenkel pairs. It has been shown by simulation that the Frenkel pair should recombine spontaneously in a ps time frame without the need to overcome an energy barrier [78]. It was then argued that a Frenkel pair formation can be stabilized close to the existing oxygen vacancies, which are charged with 4 electrons [68, 79]. As soon as the Frenkel pair forms, 2 of the 4 electrons within the adjacent V_O'' will jump into the newly generated positively charged $V_O^{\bullet\bullet}$, which will then become neutral. Thus, there is no Coulombic attraction between the negatively charged interstitial oxygen and the generated V_O^x preventing a recombination. In this case, however, the question arises how the reset operation can be described in this model.

The model described above does not consider any redox-reactions at the electrode/oxide interface. In the literature, however, there are several indications that interface reactions play an important role. Sowinska et al. showed that the Ti anode of a Ti/HfO_2/TiN cell oxidizes during the electroforming process using hard x-ray photoelectron spectroscopy [80]. Kim et al. demonstrated that the reset dynamics of Pt/Ta_2O_5/Me cells is influenced by the choice of the Me electrode material [81]. It was concluded that the observed effects originate from an oxygen exchange reaction at the Ta_2O_5/Me interface. For a $SrTiO_3$-based VCM cell it was even shown that the switching process itself has to be explained by an interfacial oxygen exchange reaction at the Pt/$SrTiO_3$ interface [82]. Moreover, it was shown that the creation of $V_O^{\bullet\bullet}$ defects at an oxidizable electrode is more favourable than within the bulk [76, 83]. In the oxygen exchange

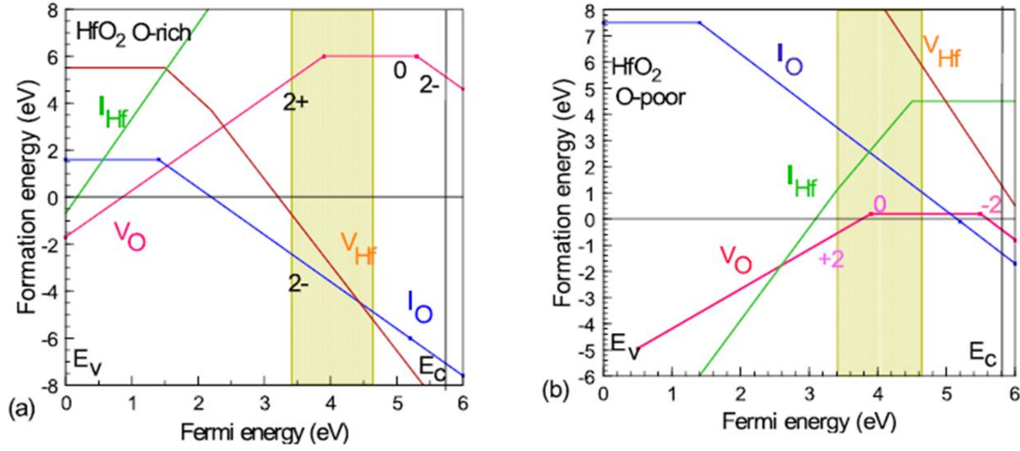


Figure 2.1: Formation energy of defects inside HfO_2 in the (a) O-rich and (b) O-poor limit. Reprinted from [75], with the permission of AIP Publishing.

model, $V_O^{\bullet\bullet}$ are created due to an oxygen exchange reaction at the anode [84, 33]. Oxygen is extracted and released as O_2, which is consistent with the observation of bubble formation at the anode [84, 85, 86]. Alternatively, the oxygen exchange reaction leads to an oxidation of the anode material [80]. The positively charged $V_O^{\bullet\bullet}$s can be treated as mobile donors and move within the applied electric field. It has been shown by physical continuum simulation that a redistribution of defects can explain the observed resistive switching phenomena [87, 88, 89, 90]. Including an oxygen exchange reaction into the physical continuum model, the electroforming process could also be successfully simulated [91].

2.1.1 Interface model

In contrary to all other KMC models published so far on this topic [69, 92, 70, 93], a new model is introduced in this work that is based on the interface reactions. In this model the formation/annihilation of $V_O^{\bullet\bullet}$s is considered to occur due to an oxygen exchange reaction at one of the electrodes (anode/cathode) solely. The $V_O^{\bullet\bullet}$s are treated as the only oxygen defects within the oxide material and are considered to be mobile. It will be shown that this model generates the forming and switching characteristics that are similar to the ones of a Pt/HfO_2/Hf/Pt cell emphasizing that the oxygen exchange model is well suited to describe the forming as well as the set and reset operations in filamentary switching VCM cells.

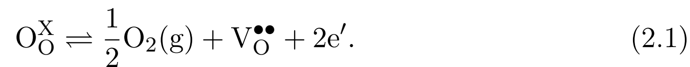
Material and device properties

The structure of a VCM cell is rather simple and typically consists of an insulating layer sandwiched between two metallic electrodes. In practice there is also a thin metal layer next to one of the electrodes that scavenges oxygen from the oxide to form $V_O^{\bullet\bullet}$. Bipolar filamentary resistive switching is the typical switching description of the VCM and has been demonstrated for many transition metal oxides such as HfO_2 [56], Ta_2O_5 [94], $SrTiO_3$ [95, 85] or TiO_2 [96] as the insulating layer. Among a lot of transition metal oxides, that show the valence change based switching, Hafnium oxide (HfO_x) as a high-k dielectric has drawn a lot of attentions and the focus throughout this work will be laid on this material [65]. HfO_2 based ReRAMs have good CMOS compatibility. Furthermore, these cells do not need the injection of extrinsic conducting defects because their operation is based on intrinsic defects (oxygen vacancies). Excellent performance of HfO_2 based ReRAMs such as high speed operation, large on/off current ratio and reliable switching endurance have been reported [40, 56].
Depending on the polarity of the applied voltages the electrodes play the role of the cathode or anode. The choice of the electrode material is as important as the insulating layer and its effect on the resistive switching has been reported in various studies. In order to evaluate the time dependant distribution of $V_O^{\bullet\bullet}$ the nature of the interface between electrodes and the oxide should be identified. The interface between the electrode and the oxide can vary from highly blocking to highly transparent for oxygen ions. Our assumption here is that one of the electrodes is easily oxidizable and creates a contact that has a reasonable degree of transparency to the oxygen ion. Along the interface of this electrode with the oxide $V_O^{\bullet\bullet}$ can be generated and leave behind oxygen ions that can diffuse into the electrode and create a reservoir there. This electrode is usually referred to as oxygen exchange layer (OEL) or defect reservoir or active electrode. The reasonable degree of transparency of anode here means that it shouldn't be highly blocking for oxygen ions so that the formation of oxygen gas out of the accumulated oxygen ions can cause the appearance of serious deformation or defect in the anode. On the other hand it shouldn't also be highly non-blocking, so that the easy drift and diffusion of oxygen ions in and out of the switching layer doesn't allow the formation of an oxygen ion reservoir in the anode. The other electrode is assumed to be inert so that we can neglect the $V_O^{\bullet\bullet}$ generation there [97].
The choice of the electrode material can also stablish an ohmic or a Schottky contact depending on the work function of the material. The active electrode should build a Schottky contact with the oxide. Otherwise in the case of an ohmic contact the current can flow fluently and reach the upper limit before that any breakdown happens [33, 97].
Another point that should be considered in the material selection for the elec-

trodes is the reaction with oxygen. If oxygen reacts easily with the electrode material, a metal-oxide layer is created along the contact area. The formation of this metal-oxide layer can hinder the further extraction of oxygen from the switching oxide layer during the CF formation and prevent fluent resistive switching. Lee et al. investigated the effect of various metal electrodes on the resistive switching of NiO thin films and reported the crucial effect of the generated interfacial metal-oxide layer on the switching properties of the memory cell [98]. In this work any reaction of the electrode materials with oxygen is neglected.

Switching mechanism

The physical model considered in this work to simulate the electroforming, set and reset processes in a typical VCM memory is based on the switching mechanism shown in Fig. 2.2. In this figure the bottom electrode (BE) plays the role of anode or cathode when a positive or negative voltage is applied to it, respectively. This model explains the switching process by an interfacial oxygen exchange reaction at the BE/oxide interface. This means that the formation/annihilation of $V_O^{\bullet\bullet}$ can only take place along the BE/oxide interface according to the oxygen exchange reaction, which is described below in Kröger-Vink nomenclature [33],



$$O_O^X \rightleftharpoons \frac{1}{2}O_2(g) + V_O^{\bullet\bullet} + 2e'. \tag{2.1}$$

Where, O_O^X denotes an oxygen ion in the regular lattice site. This reaction involves formation of oxygen gas.
During the forming and set processes, the newly generated $V_O^{\bullet\bullet}$s are injected to the oxide and migrate in the direction of the applied electric field from the anode toward the cathode. Depending on the relation between the diffusion rate of $V_O^{\bullet\bullet}$s inside the oxide and the generation rate of them at the interface, $V_O^{\bullet\bullet}$ either accumulate close to the cathode or anode. If the $V_O^{\bullet\bullet}$ generation energy barrier is high enough, they tend to pile up close to the cathode. Since $V_O^{\bullet\bullet}$s are donors, the place, where they accumulate, becomes n-conducting. This will reduce the local electric field in this region and hence the diffusion rate of $V_O^{\bullet\bullet}$s. In this case the direction of the CF growth is from the cathode toward the anode during the forming and set processes. On the other hand an extra oxygen scavenging layer can result in a very low energy barrier of $V_O^{\bullet\bullet}$ generation and cause $V_O^{\bullet\bullet}$ accumulation close to the anode. The growth of the CF between the cathode and the anode inside the oxide eventually transfers the cell from high resistance state (HRS) to low resistance state (LRS). The rupture of the CF, which transfers the cell from LRS to HRS, happens during the reset process, where a reverse bias is applied to the active electrode. In this case $V_O^{\bullet\bullet}$s migrate back toward the active electrode/oxide interface and recombine with the excess oxygen ions stored in the active electrode. The driving forces involved in the ReRAM operation in this

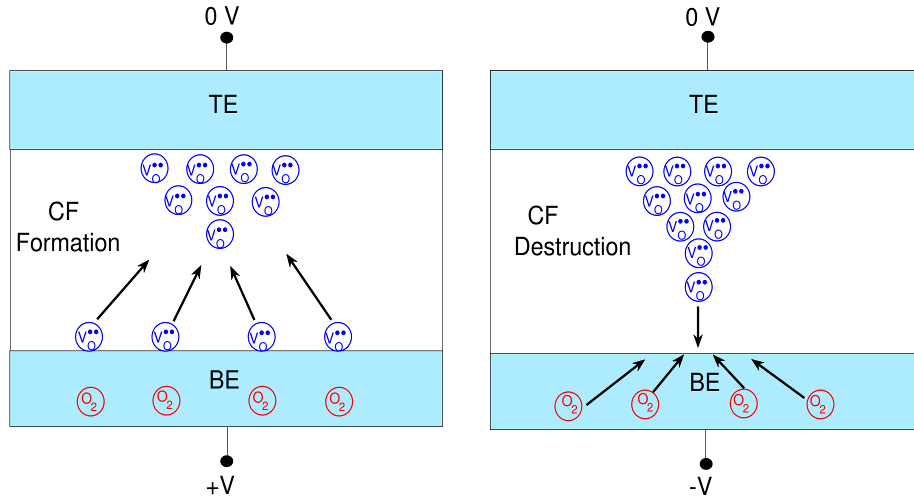


Figure 2.2: Schematic of the CF formation mechanism during the forming/set processes and the CF destruction during the reset process. During the CF formation $V_O^{\bullet\bullet}$s are generated along BE/oxide interface, leaving behind an oxygen reservoir in the BE and then migrate toward TE under an applied positive voltage. During the CF rupture the accumulated oxygen ions in the BE recombine with the $V_O^{\bullet\bullet}$ migrating towards the BE under an applied negative voltage.

model are the electric field, temperature and temperature gradient. Electrons are injected from the cathode toward the anode in the opposite direction of $V_O^{\bullet\bullet}$ migration as shown in Fig. 2.2. In this model the localized $V_O^{\bullet\bullet}$s induce defect states within the band gap of the reduced HfO_{2-x}, which act as trapping centres and assist the electron hopping transport among them.

2.2 Conduction mechanism

The understanding of the electronic transport in the high-k layer is an important prerequisite to simulate the behaviour of VCM cells and a major goal of this work. In this section we discuss the possible charge transport mechanisms along the oxide. In general the charge transport mechanisms can be divided into two main categories: (a) intrinsic mechanisms, which are related to the existence of the dielectric with no defects or impurities like direct tunnelling (b) extrinsic mechanisms, related to the role of the defects inside the oxide that can play a role in charge transport such as defect-defect tunnelling. The intrinsic mechanism in this case can be described as the direct tunnelling of the electron. The direct tunnelling here refers to the coherent tunnelling through the bandgap of the dielectric material. The barrier height and width are the conduction band offset (CBO) and the thickness of the dielectric, respectively. The role of the direct

tunnelling was assumed to be negligible in this model based on the results of a KMC transport simulation [99]. Therefore, the conduction model proposed in this work is based on the extrinsic mechanisms.
Different regimes of VCM cell operation i.e. HRS and LRS correspond to different levels of the $V_O^{\bullet\bullet}$ concentration inside the oxide. The charge transport depends on the concentration of $V_O^{\bullet\bullet}$ and hence the operation mode of the cell. In the HRS due to the low $V_O^{\bullet\bullet}$ concentration (large distance between vacancies) there is no coupling between the $V_O^{\bullet\bullet}$-mediated mid-gap defect states [100]. In this case we can use the trap assisted tunnelling (TAT) method to calculate the leakage current through the oxide. In the LRS that $V_O^{\bullet\bullet}$s are very close to each other the metallic conduction should be taken into account. These two mechanisms will be discussed in more detail in next sections and are the main conduction mechanisms driving the charge transport in the introduced model in this work.

2.2.1 Trap assisted tunnelling conducting mechanism

The set of the electron transport mechanisms related to the presence of localized defect states considered in this model includes: (i) electron tunnelling from an electrode into a defect state (ii) electron tunnelling from a defect state into an electrode (iii) electron hopping between defect states.

Tunnelling between defects and electrodes

The first step in the charge transport through the oxide is step (i) and it is finished by step (ii). The tunnelling process of electrons between electrodes and defects can be either elastic or inelastic. The inelastic case involves the emission or absorption of some lattice phonons with certain energy. In this work the elastic model will be used for more simplicity, reducing the number of free parameters and being computationally effective. Fig. 2.3 demonstrates the elastic injection and emission of electrons into and out of the defects, respectively. The electron injection rate from an electrode to a trap and vice versa can be stated as a function of the tunnelling probability, T_p, through the oxide [101]. T_p is defined as the probability that a particle, described by a wave package, can tunnel through a potential barrier. T_p of an electron with energy E_e through an arbitrary shaped potential V(x) (Fig. 2.4) can be calculated using the Wentzel-Kramers-Brillouin (WKB) approximation [102],

$$T_p = e^{-2\int_{x_0}^{x_1} \frac{1}{\hbar}\sqrt{2m^*(V(x)-E_e)}dx}, \tag{2.2}$$

where x_0 and x_1 are the classical turning points at which $V\left(x = x_0, x_1\right) = E_e$, m^* is the tunnelling effective mass and depends on the dielectric material. The hopping rate of an electron from an electrode to the trap i, P_{Ei}, and vice versa, P_{iE}, can be formulated as follows [103],

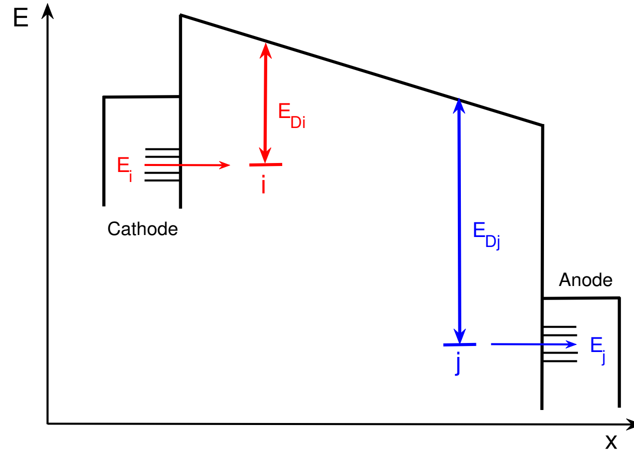


Figure 2.3: The elastic electron injection from an electrode to a defect and vice versa.

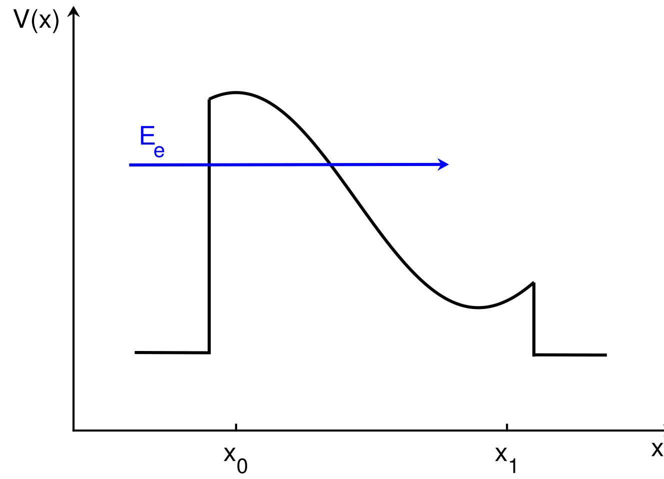


Figure 2.4: Tunnelling of an electron with energy E_e through a potential barrier $V(x)$.

$$P_{Ei} = P_0 T_p^{Ei} \int_{E_i^e}^{+\infty} F(E) g(E) dE, \tag{2.3}$$

$$P_{iE} = P_0 T_p^{iE} \int_{-\infty}^{E_i^f} (1 - F(E)) g(E) dE, \tag{2.4}$$

where P_0 is the the coupling factor between the electrode and trap i. E_i^e and E_i^f are the energies of an empty and a full trap i, respectively determined by [69],

$$E_i^e = E_{i0}^e - qV_i^H, \tag{2.5}$$

$$E_i^f = E_{i0}^f - qV_i^H, \tag{2.6}$$

where E_{i0}^e and E_{i0}^f are the energies of the trap i when no external voltage is applied. V_i^H is the homogenous potential at site of trap i. $F(E)$ denotes the Fermi-Dirac distribution,

$$F(E) = \frac{1}{1 + \exp[E - (E_f - qV_E)]}, \tag{2.7}$$

where E_f is the Fermi level of the electrodes and V_E is the applied voltage to the electrodes. $g(E)$ is the energy density of states in the electrodes. Assuming parabolic bands $g(E)$ is given by [104],

$$g(E) = \frac{m^*(2m^*E)^{1/2}}{\pi^2\hbar^3}. \tag{2.8}$$

In 2D eq (2.3) and (2.4) are simplified as

$$P_{Ei} = P_0 g_0 T_p^{Ei} \int_{E_i^e}^{+\infty} F(E)dE, \tag{2.9}$$

$$P_{iE} = P_0 g_0 T_p^{iE} \int_{-\infty}^{E_i^f} (1 - F(E))dE, \tag{2.10}$$

where g_0 represents the number of states at the given energy in the electrode. The product of $P_0 g_0$ is fitted to the experiment [69]. The integral term in eq (2.3), $\int_{E_i^e}^{+\infty} F(E)g(E)dE$, represents the number of the filled states in the electrode above E_i^e, that can inject electrons into an empty trap with the energy of E_i^e below the conduction band of HfO_2. Similarly the integral term in eq (2.4), $\int_{-\infty}^{E_i^f}(1 - F(E))g(E)dE$, represents the number of the empty states below E_i^f, which can accept electrons from a filled trap with the energy of E_i^f below the conduction band of HfO_2.

T_p^{Ei} and T_p^{iE} are the electron tunnelling probabilities from an electrode to a trap and vice versa, respectively, and determined by WKB approximation in eq (2.2) [69],

$$T_p^{Ei} = e^{-2\int_{x_m}^{x_i} \frac{1}{\hbar}\sqrt{2m^*(E_c - qV^H(x) - E_i^e)}dx}, \quad E_c - qV^H(x) - E_i^e > 0 \tag{2.11}$$

$$T_p^{iE} = e^{-2\int_{x_i}^{x_m} \frac{1}{\hbar}\sqrt{2m^*(E_c - qV^H(x) - E_i^f)}dx}, \quad E_c - qV^H(x) - E_i^f > 0 \tag{2.12}$$

where x_m and x_i are the electron positions in the electrode and trap i, respectively. E_c is the conduction band energy of HfO_2 and $E_c - qV^H(x)$ is the energy barrier V(x) in the eq (2.2).

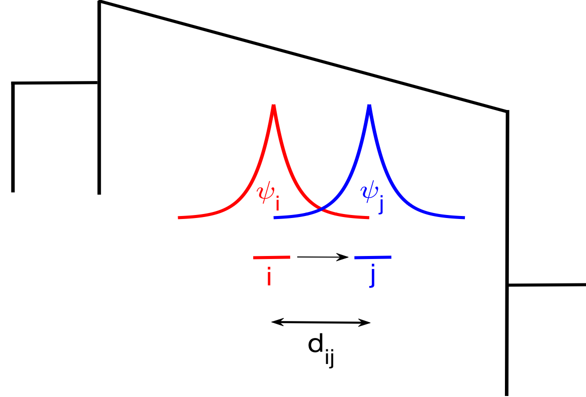


Figure 2.5: Schematic of the electron tunnelling between two adjacent states i and j with distance d_{ij}. The wave functions of the trap states, that are so far from each other that there is no coupling between them, are demonstrated by ψ_i and ψ_j

Defect-Defect tunnelling

In this section the electron hopping process from one defect to another one is investigated. In order to have the the hopping process between defects, the distances between them should be large enough so that we can consider them as localized individual states. In this case the electron involved in the charge transfer is localized at the defect site and there is no coupling between the initial and final states. With this condition the TAT theory can describe the charge transport inside the oxide [100]. Fig. 2.5 demonstrates the defect-defect tunnelling process.

The hopping rate of an electron, h_{ij}, from one localized defect state at site i to another defect state at site j with the distance of d_{ij} is calculated by the Abraham-Miller Formula,

$$h_{ij} = \begin{cases} \nu_{0n} \exp\left[-\frac{d_{ij}}{a_0} + \frac{q(V_j^H - V_i^H)}{k_B T}\right] & V_j^H \leq V_i^H \\ \nu_{0n} \exp\left[-\frac{d_{ij}}{a_0}\right] & V_j^H > V_i^H, \end{cases} \tag{2.13}$$

where V_i^H and V_j^H are the homogenous potentials at sites of traps i and j, respectively. q is the elementary charge, ν_{0n} is the electron characteristic vibration frequency, T is the local temperature and k_B is the Boltzmann constant. The electron hopping barrier between two traps is considered in this work to depend on the homogenous solution of the boundary value problem for the potential as

explained in [105]. This is a good approximation when the trap concentration is low enough and the Coulomb interactions between traps can be neglected. The potential induced by a charge itself also does not modifiy the energy barrier [106]. This means with no voltage applied, electrons trapped in different vacancy sites will face the same barrier for transferring to an unoccupied trap state.
a_0 in eq (2.13) is the attenuation length of the localized wave function, i.e. $\exp(-r/a_0)$. The normalized wave function of a trapped electron in a defect with a delta-like defect potential is taken to be [101],

$$\psi(r) = (K/2\pi)^{1/2} \exp(-Kr)/r, \tag{2.14}$$

where $K^2 = 2m^* E_D/\hbar^2$ and E_D is the defect depth, meaning how far this defect is located below the conduction band of HfO_2. Depending on the occupancy of the defect i, E_D is

$$E_D = E_i^e \qquad \text{for an empty defect,} \tag{2.15}$$

$$E_D = E_i^f \qquad \text{for a filled defect.} \tag{2.16}$$

From eq (2.14) it is obvious that $a_0 = 1/K$, therefore,

$$a_0 = \frac{\hbar}{\sqrt{2m^* E_D}}. \tag{2.17}$$

TAT current solver

The leakage current through the oxide in the HRS using the TAT method, I_{TAT}, can be calculated as [107]

$$I_{TAT} = e \sum_{i=1}^{N} [p_i P_{iA} - (1 - p_i) P_{Ai}], \tag{2.18}$$

where e is the electron charge, N is the total number of traps, p_i is the electron occupational probability of trap i and P_{iA} and P_{Ai} are the electron hopping rate from trap to the anode and vice versa, respectively. This formula gives the current flow through the anode, but one can also calculate the current through the cathode, which should give the same result. To obtain p_i we need to solve the master equation in the quasi-steady state [107],

$$(1 - p_i) \sum_{j=1, j \neq i}^{N} p_j h_{ji} - p_i \sum_{j=1, j \neq i}^{N} (1 - p_j) h_{ij} \tag{2.19}$$

$$+ (P_{Ci} + P_{Ai})(1 - p_i) - (P_{iC} + P_{iA}) p_i = 0.$$

As boundary conditions, it is assumed that the electron occupational probabilities in the cathode and anode are 1 and 0, respectively. The master equation is solved using the Newton-Raphson iteration scheme, which is the generalization of the Newton method to many variables. If eq (2.19) for the trap i is formulated as $f_i(\vec{p}) = 0$, a system of N non-linear equations should be solved as $\vec{f}(\vec{p}) = 0$, where $\vec{p}$ is an N dimensional vector of the electron occupational probabilities. The following iteration is then used to obtain $\vec{p}$,

$$\vec{p}_k = \vec{p}_{k-1} - [\hat{\mathcal{J}}|_{\vec{p}_{k-1}}]^{-1}\vec{f}(\vec{p}_{k-1}), \tag{2.20}$$

where $\hat{\mathcal{J}}|_{\vec{p}_{k-1}}$ is the $N \times N$ Jacobian matrix calculated at $\vec{p}_{k-1}$,

$$[\hat{\mathcal{J}}|_{\vec{p}_{k-1}}]_{ij} = \frac{\partial f_i}{\partial p_j}|_{\vec{p}_{k-1}}, \tag{2.21}$$

where k is the number of the iteration. The iteration is stopped when a predefined error limit is reached.
Fig. 2.6(a) shows the simulation result of the distribution of the electron occupational probability along the CF during the set process, where the CF has not completely formed yet. During the set process a positive voltage is applied to the bottom electrode(anode). p has a higher value close to the cathode(top electrode), from where the electrons are injected into the traps. This is due to the larger hopping rate between the electrodes and traps than the rate between two traps. These rates depend on the material properties of the oxide and electrode and the coupling factor between them. This explains well the dependency of ReRAM performance on the material choice for the electrodes and the dielectric. The probability values decrease with increasing the absolute values of the applied voltage as shown in Fig. 2.6(b).

2.2.2 Drift mechanism

During the final stage of the forming and set processes usually the $V_O^{\bullet\bullet}$ concentration is so high that we cannot neglect the strong interaction between adjacent $V_O^{\bullet\bullet}$s. In this situation it is not possible any more to use the TAT mechanism, where $V_O^{\bullet\bullet}$ play the role of independent defects assisting the electron transport. In this case we switch to drift mechanism (Fig. 2.7) and solve the drift-diffusion current continuity equation to obtain the local electron density n. Here we assume the quasi-steady state, where $\partial n/\partial t = 0$. The generation/recombination rate of electrons (G−R) was also neglected. With this assumptions the continuity equation for electrons is given by,

$$\nabla \cdot \vec{J}_n = 0. \tag{2.22}$$

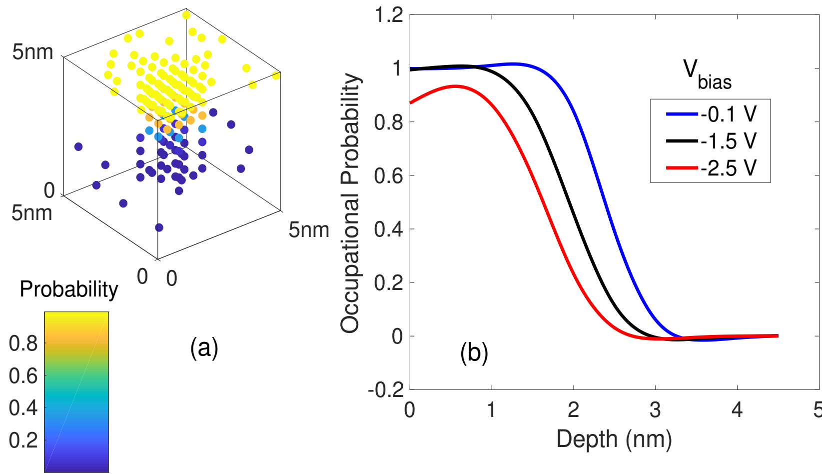


Figure 2.6: (a) The electron occupancy profile along CF during the set process. The value of this probability is higher for the traps close to the cathode. (b) The averaged value of the electron occupancy probability along the oxide thickness under different applied negative voltages in the HRS [108].

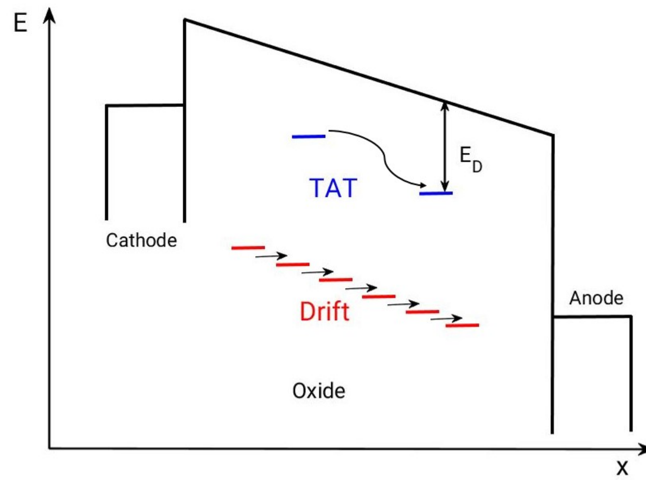


Figure 2.7: Multiple electron conduction mechanism along the oxide in ReRAM depending on the distance between defects inside the oxide.

$\vec{J_n}$ is the drift-diffusion current density,

$$\vec{J_n} = -q\mu_n(n\nabla V - V_T\nabla n), \tag{2.23}$$

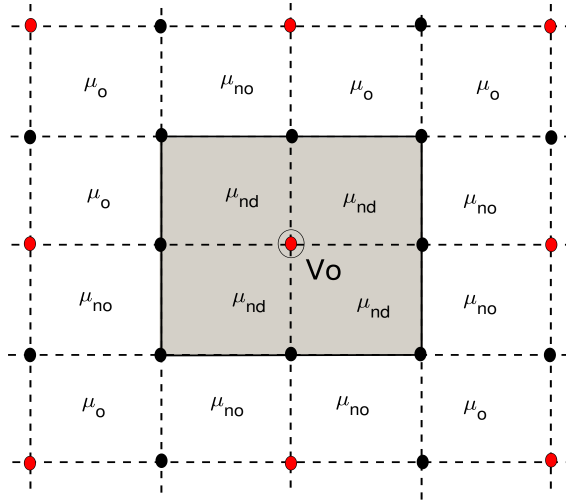


Figure 2.8: 2D mesh to solve the drift-diffusion equation. Red points are the grid points, where $V_O^{\bullet\bullet}$ can reside.

where V is the local potential, μ_n is the electron mobility and V_T is the thermal voltage. The electron mobility can differ for the oxide and the CF. The mobility in oxygen deficient regions, which contain $V_O^{\bullet\bullet}$ is set to $\mu_n = \mu_{nd}$ and is higher than the mobility in the rest of the oxide, $\mu_n = \mu_{no}$. Fig. 2.8 shows the rectangular 2D grid to discretize drift-diffusion equation containing the grid nodes, where the electron density is defined. The lattice constant of $Hf0_2$, which determines where $V_O^{\bullet\bullet}$ can sit is twice the lattice constant of the grid to solve the drift-diffusion equation. The red coloured points in Fig. 2.8 are the allowed grid points on which $V_O^{\bullet\bullet}$ can sit. The electron mobility is defined in every grid primitive and if there is an $V_O^{\bullet\bullet}$ in a grid point the assumption is to set the electron mobility in the primitives surrounding this grid point to μ_{nd}. In Fig. 2.8 these cells are highlighted in gray.

To solve the drift-diffusion equation the Scharfetter-Gummel discretization method was employed to avoid the instability that finite volume method can cause [109]. To have a unique solution we need to apply the proper boundary conditions (BC). Dirichlet boundary condition is considered here along the oxide-electrode interfaces, due to the metallic contact, and is given below,

$$n_{cont} = n_0 T_p, \tag{2.24}$$

where n_{cont} is the electron density at the electrode-oxide contact. n_0 is the electron density in the electrode that can be injected into the oxide via the $V_O^{\bullet\bullet}$s and T_p is the tunnelling probability from the electrode into $V_O^{\bullet\bullet}$ and is calculated using eq (2.2).

2.3 Electrical potential

The local electrical potential, V, inside the oxide is calculated by solving the Poisson equation

$$\nabla \cdot (\varepsilon \nabla V) = -\rho, \tag{2.25}$$

where $\varepsilon = \varepsilon_r \varepsilon_0$. ε_r is the dielectric constant of HfO_2, ε_0 is the permittivity of vacuum and ρ is the space charge density. If $V_O^{\bullet\bullet}$s are considered as delta-like defect points, then ρ can be expressed as

$$\rho = q_i \delta(r - r_i), \tag{2.26}$$

where q_i and r_i are the charge and location of defect i. q_i is defined as the vacancy charge state, which is in this case doubly positive charged plus the negative charge due to the occupation of $V_O^{\bullet\bullet}$ by electron. Two sets of BCs were applied to have a unique solution: 1) Dirichlet BC along a metallic contact, where the potential BC at the interface, V_{ci}, is set to a fixed value,

$$V_{ci} = V_a + \psi_{Bi}, \tag{2.27}$$

where V_a is the applied voltage to the electrode and ψ_{Bi} is the built-in potential and defined as the difference between the Fermi level of the electrode and oxide materials and in this work is considered as a constant value. 2) homogenous Neumann BC along the rest of the boundaries of the solution domain, which means that there is no normal variation of the potential at these boundaries. In the finite volume approach, which will be discussed later in 2.3.1, this is ensured by setting the permittivity outside of the solution domain to zero.
In order to obtain the electron hopping rates in the TAT model, the homogenous potential is needed. To calculate the homogenous potential the space charge should be set to zero, therefore the Poisson equation reduces to the Laplace equation,

$$\nabla \cdot (\varepsilon \nabla V_H) = 0. \tag{2.28}$$

The BC to solve this equation is the same as the BC of eq (2.25).
In the case of the drift-diffusion model the following system of equations should be solved self-consistently to give the potential and electron density,

$$\nabla \cdot (\varepsilon \nabla V_k) = -(\rho_{vac} - q n_{k-1}) \tag{2.29}$$

$$\nabla \cdot \vec{J}_{nk} = 0, \quad \vec{J}_{nk} = -q\mu_n (n_k \nabla V_k - V_T \nabla n_k),$$

where ρ_{vac} is the space charge density of $V_O^{\bullet\bullet}$. This system of equations is solved by iterating over the equations. k is the number of iterations. First the Poisson equation is solved to obtain V_k, while the electron density n_{k-1} is given by the previous iteration. This process stops when a certain error limit is reached.

2.3.1 Numerical method for solving the Poisson equation

The finite volume method is used to solve the Poisson equation numerically [110]. Using this method instead of the finite element method ensures the principle of the flux conservation. In this method the solution domain, D, is partitioned into finite volumes, where each volume contains a grid point. The solution of differential equations, in this case the electrical potential, is assigned to each grid point, while the material properties like the permittivity is defined for the finite volume. Fig. 2.9 shows a two dimensional Delaunay mesh in Cartesian coordinates to discretize the Poisson equation. The finite volume of grid point ij, which is highlighted in gray, by definition contains all points of the solution domain, which are closest to this point. Every grid primitive is a rectangle labeled by its upper left corner node. The volume integration of the Poisson equation can be transformed into a surface integral using the divergence Gauss' law,

$$\int_{D_{i,j}} \nabla \cdot (\varepsilon \nabla V) \mathrm{d}V = \oint_{\partial D_{i,j}} (\varepsilon \nabla V) \cdot \mathrm{d}\vec{A}, \tag{2.30}$$

where $D_{i,j}$ is the domain of the finite volume around grid point (i,j), $\partial D_{i,j}$ is the boundary of $D_{i,j}$ and $\mathrm{d}\vec{A}$ is the normal vector oriented out of the volume. The surface integral is discretized with splitting $\partial D_{i,j}$ into eight pieces, labeled in Fig. 2.9 by $1, 2, ..., 8$, according to the different boundaries and the material properties of $D_{i,j}$. Assuming a constant flux density over each piece of $\partial D_{i,j}$ yields

$$\oint_{\partial D_{i,j}} (\varepsilon \nabla V) \cdot \mathrm{d}\vec{A} = \sum_{k=1}^{8} (\varepsilon \nabla V)_k \cdot \vec{A}_k. \tag{2.31}$$

For example for $k = 1, 2$ the flux through surfaces 1 and 2 is formulated as,

$$(\varepsilon \nabla V)_1 \cdot \vec{A}_1 + (\varepsilon \nabla V)_2 \cdot \vec{A}_2 = \tag{2.32}$$

$$(\varepsilon_{i-1,j-1} \frac{V_{i,j} - V_{i,j-1}}{y_j - y_{j-1}} \vec{e}_y) \cdot (-\frac{x_i - x_{i-1}}{2} \vec{e}_y) + (\varepsilon_{i,j-1} \frac{V_{i,j} - V_{i,j-1}}{y_j - y_{j-1}} \vec{e}_y) \cdot (-\frac{x_{i+1} - x_i}{2} \vec{e}_y).$$

The right hand side of eq (2.32) can be represented as $C_{i,j}^{i,j-1}(V_{i,j} - V_{i,j-1})$, where $C_{i,j}^{i,j-1}$ is defined as the capacitance per unit length in 2D,

$$C_{i,j}^{i,j-1} = \frac{1}{2} (\varepsilon_{i-1,j-1} \frac{x_i - x_{i-1}}{y_j - y_{j-1}} + \varepsilon_{i,j-1} \frac{x_{i+1} - x_i}{y_j - y_{j-1}}). \tag{2.33}$$

In this manner the flux through all other surfaces can be calculated and the final discrete form of the surface integral in eq (2.31) is given by

$$C_{i,j}^{i,j-1} V_{i,j-1} + C_{i,j}^{i-1,j} V_{i-1,j} + C_{i,j}^{i,j} V_{i,j} + C_{i,j}^{i+1,j} V_{i+1,j} + C_{i,j}^{i,j+1} V_{i,j+1} = D_{i,j} \rho_{i,j}, \tag{2.34}$$

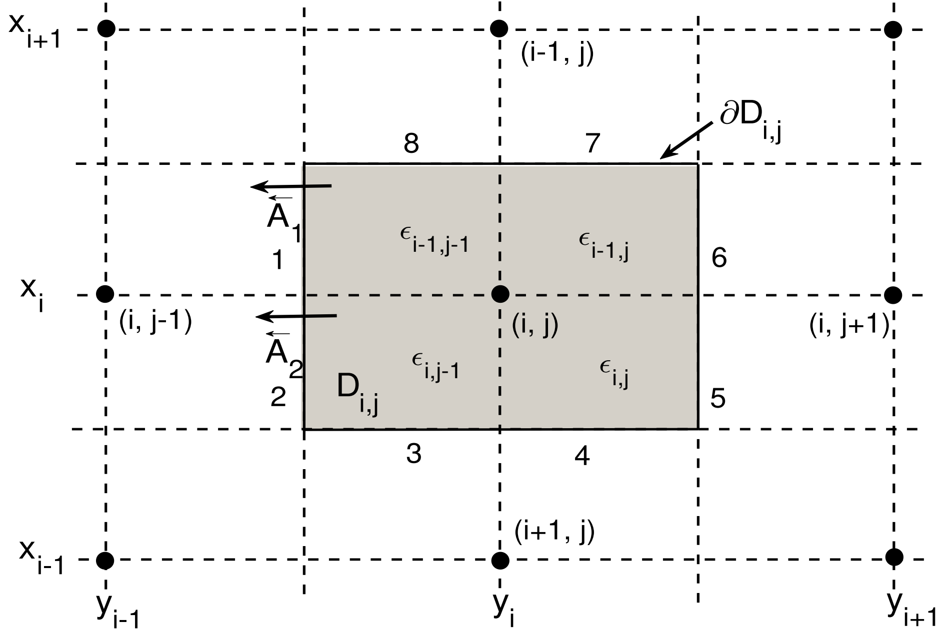


Figure 2.9: Two dimensional finite volume of grid node i, j.

where $D_{i,j} = \frac{x_{i+1}-x_{i-1}}{2}\frac{y_{i+1}-y_{i-1}}{2}$ and $\rho_{i,j}$ is the space charge density on the grid node i, j. $D_{i,j}\rho_{i,j}$ gives the charge per unit length, $q_{i,j}$.
The following system of linear equations should be solved then to obtain the potential in grid points,

$$\hat{C}\vec{V} = \vec{Q}, \tag{2.35}$$

where $\hat{C}$ is the capacitance coefficient matrix, $\vec{V}$ and $\vec{Q}$ are vectors containing the potentials and the charges on the grid nodes, respectively.

2.4 Joule heating

Joule heating has been suggested to play an important role in the forming and switching processes. Here we aim to calculate the local temperature profile by solving the Fourier's heat transfer equation,

$$\nabla \cdot (k_{th}\nabla T) = -g, \tag{2.36}$$

where k_{th} is the oxide thermal conductivity and g is the rate of the heat generation. The current and potential information of the cell can be used to calculate g, which is the product of the electric field and the current density. To solve

eq (2.36) we considered homogenous Neumann boundary condition along where there is no metal contact so that there is no heat flow through these boundaries,

$$\frac{\partial T}{\partial n} = 0, \tag{2.37}$$

where n is the direction normal to the surface. The interface between the metal electrode and oxide was modelled by an external thermal resistance [111]. Similar to the Poisson equation, eq (2.36) is solved using the finite volume method.
All the values of the parameters used in this chapter are listed in table. 2.1. These are the default values and will be used in the next chapters unless it is explicitly mentioned that another value is chosen for them.

Table 2.1: Values of the main physical parameters related to HfO_2 used in chapter 2

Parameter	Definition	Value
m^*	Tunnelling effective mass	$0.1m_0$
$E_C - E_{i0}^e$	Trap energy of an empty $V_O^{\bullet\bullet}$s below the conducting band of HfO_2	1.8 eV
$E_C - E_{i0}^f$	Trap energy of a filled $V_O^{\bullet\bullet}$s below the conducting band of HfO_2	1.95 eV
E_F	Fermi level of the top and bottom electrodes below the conduction band of HfO_2	-1.9 eV
ν_{0n}	Characteristic vibration frequency of an electron	10^{12} Hz
μ_{nd}	Electron mobility in oxygen-deficient HfO_2	8×10^-5 m^2/Vs
μ_{no}	Electron mobility in HfO_2	8×10^-7 m^2/Vs
V_T	Thermal voltage at room temperature	26 mV
ε_r	Dielectric constant	25
k_{th}	Oxide thermal conductivity	0.5 W/Km

Chapter 3

KMC simulation of VCM

Various types of computational models have been developed to help to understand and predict the switching characteristics of ReRAM. Finite element method (FEM) and kinetic Monte Carlo (KMC) are two examples of the models that have received strong attention, since they enable the simulation of device conduction and switching characteristics. At the same time they provide a microscopic view of the switching mechanism. In the FEM, transport equations are solved numerically for electrons and ions, while all variables are treated as continuous. As opposed to FEM, KMC deals with discrete quantities like the number of defects. FEM and KMC lay between detailed ab initio Molecular Dynamics (MD) and compact modelling regarding the efficiency and the accuracy. MD approach focuses on the atomic movement and provides quantitative details of the ionic migration and generation and the dependency of the thermal and electrical conductivities on the defect concentration. However, this approach is only practical to simulate very short time scales in comparison to the experimental scales. The continuum model is suitable for large-scale simulations of circuits by providing analytical models with an adjustable degree of simplifications depending on the circuit size. Fig. 3.1 schematically compares different modelling approaches of ReRAM devices.

Among the above mentioned computational models we found KMC as the most suitable one to simulate the ReRAM operation, most importantly because of the random nature of ReRAM devices, which can be captured by KMC. The ReRAM operation involves random processes such as $V_O^{\bullet\bullet}$ generation and annihilation and physical diffusion of $V_O^{\bullet\bullet}$ leading to the random CF formation and dissolution. KMC is the best approach to include the resultant variability to allow for a realistic description of the statistical effects. Complexities at the level of individual charge carriers can be handled in reasonable computation power and time using the KMC method [112]. This allows to simulate the operation of a ReRAM device at experimental time and length scales, which considers a

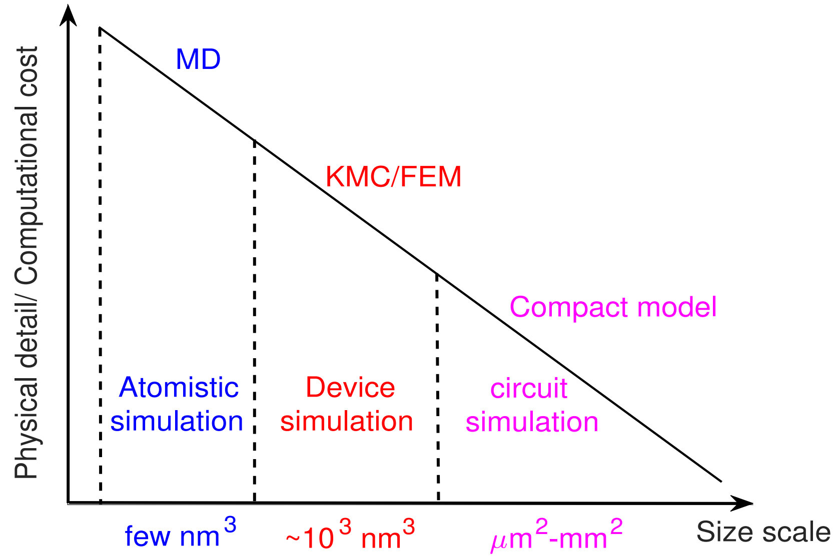


Figure 3.1: Comparison of the main physical modelling approaches to investigate ReRAM devices with respect to the computational cost and the size scale.

highly accurate description of the switching process. A brief introduction to the KMC method seems helpful here before moving to present our KMC simulator of the forming and the switching processes.

3.1 Basics of Kinetic Monte Carlo method

In order to model the dynamics of the physical system, it is considered as a macroscopic state at any instant of time. A state to state transition determines the time evolution of the physical processes. All the transitions that are important to describe the dynamics of the system should be identified and associated with a transition rate R.
There are several known algorithms of KMC that all give the same result. Among them a simple yet very efficient one is the variable step size method [113]. The procedure of this algorithm is given below,

1. **Defining the initial and end conditions:**
   These conditions include determining the initial distribution of particles, initial value of time and the condition to stop the simulation.

2. **Calculating the transition rates depending on the physics of the system under investigation:**

These rates are generally derived from the transition state theory (TST), where a chemical or physical reaction is considered as a transition of the system between adjacent local energy minima [114]. In a specific case the hopping process of a particle from state i to state j is considered as shown in Fig.3.2. If it is assumed that there is no electric field applied, the system attempts to escape the energy barrier between S_i and S_j by thermal excitation. Assuming that all the chemical and physical processes are governed by Boltzmann statistics [106], the rate of this transition can be formulated as below,

$$R_{ij} = \nu_0 \exp\left(\frac{-E_A}{k_B T}\right) \tag{3.1}$$

where ν_0 is the average attempt frequency, E_A is the energy barrier to reach from S_i to S_j and k_B is the Boltzmann factor. This step must be repeated at every simulation flow to update the previous values of rates.

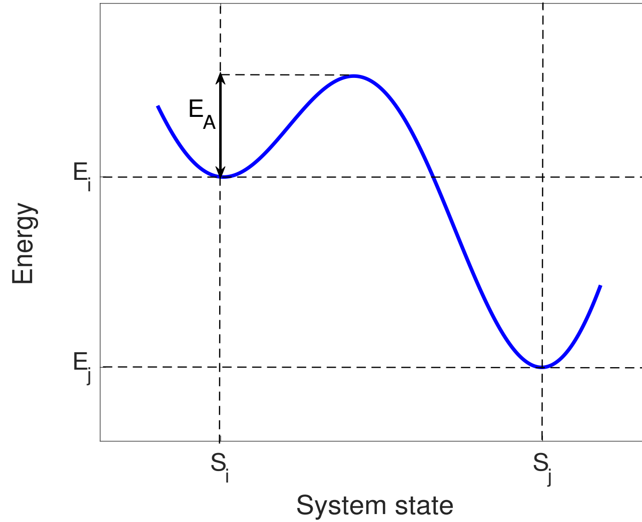


Figure 3.2: The energy diagram of transition between two energy states S_i and S_j.

3. **Determining the next process and calculating the time increment:** To select one event among the total number of possible N events a uniform random number r_1 is generated between 0 and 1. The rate of the chosen event must satisfy the following relation [115],

$$\sum_{i=1}^{n-1} R_i < r_1 \sum_{i=1}^{N} R_i < \sum_{i=1}^{n} R_i, \tag{3.2}$$

where R_i is the rate of event i and n is selected as the next event.
At each time interval Δt for a given simulation step only one possible event

can be carried out. Δt is variable and depends on the system's state at each simulation step and is calculated using the following formula [115],

$$\Delta t = -\frac{\ln(r_2)}{\sum_{i=1}^{N} R_i}, \tag{3.3}$$

where r_2 is another uniform random number between 0 and 1. The simulation time can then be updated accordingly as $t = t + \Delta t$.

4. **Carrying out the next transition:**
   The selected process is executed and the system is changed accordingly.

5. **Checking the stop condition:**
   One or more conditions must be defined to end the simulation like a limit on time or some properties of the system.

3.2 Developing a KMC simulator for VCM ReRAM

3.2.1 Processes involved in KMC

In order to implement a KMC simulation successfully the essential steps are to define all the involved chemical and physical processes and formulating the corresponding transition rates. Knowing all the involved reactions and accurately obtaining the rate parameters are practically impossible. Therefore the goal here is to construct a meaningful set of processes with the best possible approximation of the corresponding rates.
The various processes included in the developed KMC simulator of a VCM cell in this work are represented in Fig. 3.3 and the description of all processes are listed below,

(1) $V_O^{\bullet\bullet}$ generation at the electrode-oxide interface

The ReRAM device is an open system, where the number of defects, in this case $V_O^{\bullet\bullet}$, may not be conserved due to the oxygen exchange. As explained in chapter 2, in our model the oxygen exchange reaction can only happen at the active electrode-oxide interface. The other electrode is assumed to be inert and to block the oxygen. For such a blocking contact $V_O^{\bullet\bullet}$s are simply accumulated at the interface. According to eq (3.1) the transition rate associated with the $V_O^{\bullet\bullet}$ generation process is formulated as

$$R_G = \nu_0 \exp\left(-\frac{E_G - \alpha a q F}{k_B T}\right), \tag{3.4}$$

with $E_A = E_G - \alpha a q F$, where q is the elementary charge and E_G is the energy barrier of $V_O^{\bullet\bullet}$ generation at room temperature and no external applied electric

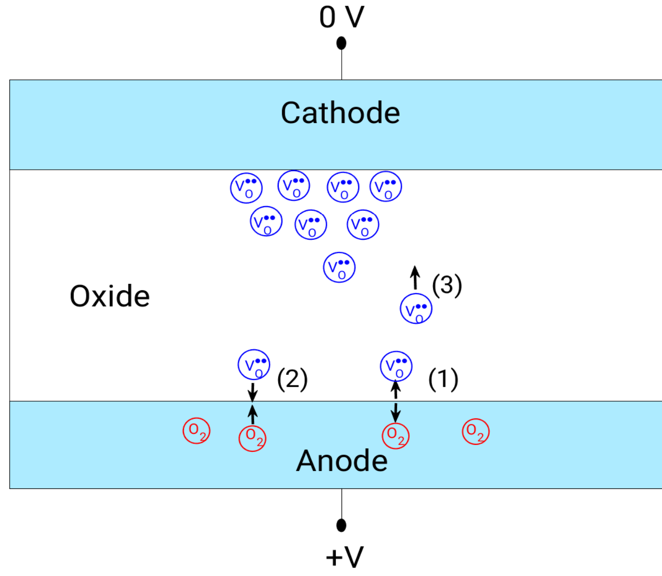


Figure 3.3: Processes involved in the KMC simulation: (1) generation of a new vacancy at the interface and release of an oxygen ion, which then diffuses into the electrode and is stored there in the form of oxygen gas (2) vacancy diffusion inside the oxide (3) recombination of a vacancy at the interface with an oxygen ion coming from the oxygen gas stored in the electrode.

field. The value of E_G is usually relatively high and reduced by the locally induced electric field, F, ($\vec{F}$ is approximated with only one component normal to the electrode/oxide interface) and the local temperature, T [67]. If the $V_O^{\bullet\bullet}$ generation is counted as the forward process, then the $V_O^{\bullet\bullet}$ recombination is the reverse process. In this case the generation and recombination energy barriers are modified by $-\alpha aqF$ and $-(1-\alpha)aqF$, respectively. α is the symmetry factor, which is usually close to 0.5 and a is the lattice constant.

(2) $V_O^{\bullet\bullet}$ recombination at the electrode-oxide interface

Similar to the generation process, the recombination of $V_O^{\bullet\bullet}$ with the oxygen ions coming from the oxygen reservoir inside the electrode is expected to happen at the electrode-oxide interface. The corresponding transition rate according to eq (3.1) is

$$R_R = \nu_0 \exp\left(-\frac{E_R - (1-\alpha)aqF}{k_B T}\right), \tag{3.5}$$

with $E_A = E_R - (1-\alpha)aqF$, where E_R is the energy barrier of $V_O^{\bullet\bullet}$ recombination at room temperature and no applied voltage bias. F and T are two factors assisting lowering the effective recombination barrier.

(3) $V_O^{\bullet\bullet}$ diffusion inside the oxide layer

The positively charged $V_O^{\bullet\bullet}$s are treated as the only oxygen defects within the oxide material and are considered to be mobile under the applied electric field. The transition rate of $V_O^{\bullet\bullet}$ diffusion inside the oxide based on eq (3.1) is given by the following equation,

$$R_D = \nu_0 \exp\left(-\frac{E_D - q\Delta V/2 - k_B \Delta T/2}{k_B T}\right), \tag{3.6}$$

with $E_A = E_D - q\Delta V/2 - k_B\Delta T/2$, where ν_0 is the characteristic vibration frequency and E_D is the diffusion energy barrier of $V_O^{\bullet\bullet}$ at room temperature and no external applied voltage bias. E_D in this model is independent of the charge state of $V_O^{\bullet\bullet}$. The effective value of E_D is modified by the potential difference, ΔV [116], the local temperature, T, and the temperature gradient across two neighbouring lattice sites [117]. In this model $V_O^{\bullet\bullet}$ jump is only possible between two nearest neighbouring sites. ΔV is defined as following,

$$\Delta V = V(r) - V_{self}^r(r) + V_{im}^r(r) - (V(r') - V_{self}^r(r') + V_{im}^r(r')), \tag{3.7}$$

where $V(r)$ is the local potential at the initial vacancy site r and $V(r')$ is the local potential at the neighbouring site r', that vacancy intends to jump to it. An extra term should be added to the local potential difference to remove the effect of the potential generated by the ion itself, which is called the self potential, $V_{self}^r(r)$ [106]. In order to obtain $V_{self}^r(r)$ the Poisson equation should be solved, where there is only a single charge at site r and the rest of the space charge is ignored and the applied voltages on both electrodes are set to zero. The self potential of an ion doesn't help lowering its own diffusion barrier. $V_{self}^r(r)$ is usually a considerable value and if it is not subtracted from the local potential, it can cause a large error in calculating the diffusion barrier lowering. $V_{im}^r(r)$ is the image potential due to the Coulomb interactions of the ion at r with its image charges in the electrodes. By removing $V_{self}^r(r)$ the image potential is also removed. We should add back this term to consider the corret potential at site r. The potential energy of repetitive image charges up to order n_{im} due to the charge q at site r with a distance x from one electrode and $L - x$ from other electrode, where L is the distance between two electrodes, is given by,

$$qV_{im}^r(r) = \frac{q}{16\pi\varepsilon}\left(\sum_{i=-n_{im}}^{n_{im}} \frac{-q}{|2x - 2iL|} + \sum_{i=-n_{im}, i\neq 0}^{n_{im}} \frac{q}{|2iL|}\right). \tag{3.8}$$

In this work $n_{im} = 50$ is considered, which is more than sufficient.
$V_{self}^r(r')$ is the potential at r' due to the ion charge at r and no other space charges and zero applied voltages on both electrodes. This term should be subtracted

from $V(r')$ to take into account the effect of neglecting the ion at r. On the other hand we should again add back the potential generated at r' due to the image charges of the ion at r. Depending on the position of the neighbouring sites of the ion, which always have one lattice constant a distance from the ion, three cases are considered: (1) when r' has the same distance x from the electrode,

$$qV_{im}^{r}(r') = \frac{q}{16\pi\varepsilon}\left(\sum_{i=-n_{im}}^{n_{im}} \frac{-q}{\sqrt{a^2+(2x-2iL)^2}} + \sum_{i=-n_{im},i\neq 0}^{n_{im}} \frac{q}{\sqrt{a^2+(2iL)^2}}\right), \tag{3.9}$$

(2) when r' is a lattice constant a closer than r to the electrode,

$$qV_{im}^{r}(r') = \frac{q}{16\pi\varepsilon}\left(\sum_{i=-n_{im}}^{n_{im}} \frac{-q}{|2x-a-2iL|} + \sum_{i=-n_{im},i\neq 0}^{n_{im}} \frac{q}{|a-2iL|}\right), \tag{3.10}$$

(3) when r' is a lattice constant a further than r to the electrode,

$$qV_{im}^{r}(r') = \frac{q}{16\pi\varepsilon}\left(\sum_{i=-n_{im}}^{n_{im}} \frac{-q}{|2x+a-2iL|} + \sum_{i=-n_{im},i\neq 0}^{n_{im}} \frac{q}{|a-2iL|}\right), \tag{3.11}$$

ΔT in eq (3.6) is simply given as

$$\Delta T = T(r') - T(r), \tag{3.12}$$

where $T(r)$ and $T(r')$ are local temperatures at the initial and the final sites of $V_O^{\bullet\bullet}$. This term indicates that vacancies tend to move toward the hotter regions [117].

3.2.2 Simulation procedure

The simulation procedure of the forming, set and reset processes is shown in Fig. 3.4. The choice of the dimension of the simulated structure is aligned with the purpose of our simulation, which can vary between 2D and 3D. A 3D simulation is run when we need to have more precision and compare our results with the experimental results. In order to investigate the variation of device parameters, which requires a larger number of repeated simulations the 3D model is still efficient enough to use. To simulate larger devices and random telegraph noise effect we choose to implement the model in 2D to keep an acceptable balance between computation load and accuracy.
The simulation starts with an initial random distribution of $V_O^{\bullet\bullet}$ inside the simulated region representing the weak spots of an as-fabricated HfO_2 ReRAM cell, where the breakdown can occur during the forming. These weak spots align with

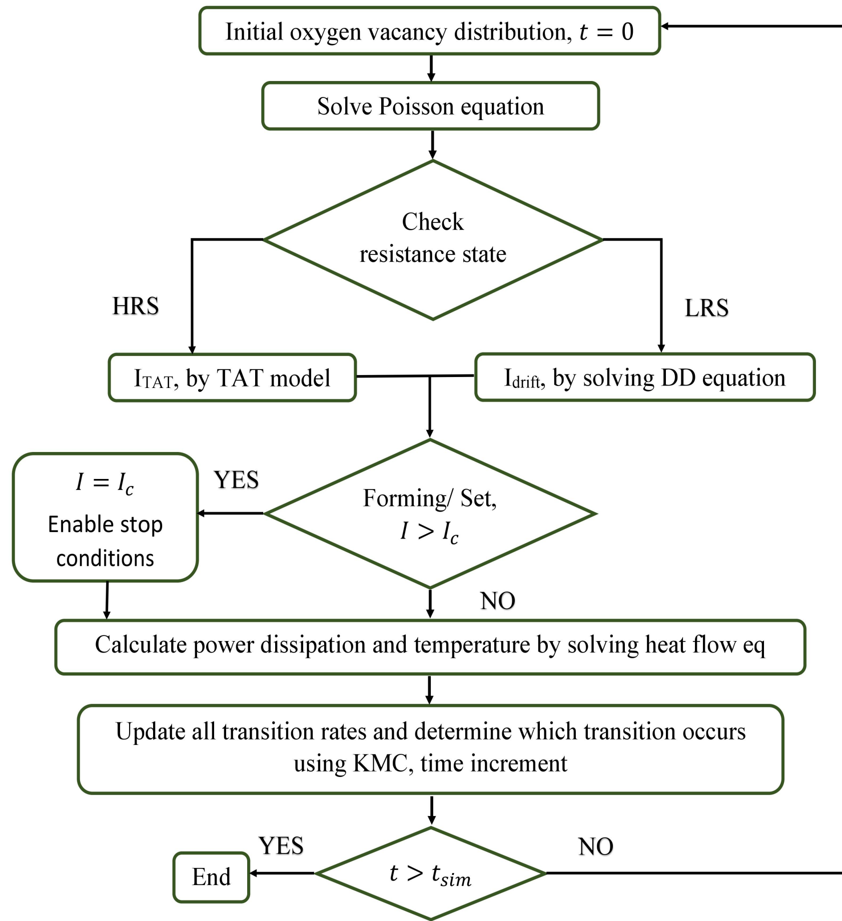


Figure 3.4: Flowchart of the simulation flow of the forming and switching processes.

the grain boundaries(GB) in the dielectric, which represent a natural conduction path [67]. The initial local temperature in the solution domain is set to the external temperature, $T_0 = 300$ K. The bias voltage can be applied either linearly increasing or decreasing with time with a specific rate (DC sweep analysis) or in a pulse mode (transient analysis). Next the Poisson equation is solved considering the applied voltage and the space charge to obtain the local electrical potential. The next step is calculating the leakage current which depends on the $V_O^{\bullet\bullet}$ concentration. For a low $V_O^{\bullet\bullet}$ concentration corresponding to HRS the current continuity equation is solved self consistently with the Poisson equation and the current is obtained with the TAT solver. For higher $V_O^{\bullet\bullet}$ concentration corresponding to LRS the drift-diffusion equation is solved self consistently with the Poisson equation. The criteria to switch from TAT to drift conduction mech-

anisms are: (1) the number of $V_O^{\bullet\bullet}$ inside the CF is bigger than a predefined value, N_0 (2) the tunnelling gap between the tip of the CF and the electrode is less than a predefined length, d_0. Usually during the forming and set processes, the leakage current through the oxide is limited by a current compliance, I_C, using a control circuit to avoid any physical damage to the device due to excessive heat generation. At each simulation step if the running process is forming or set, the calculated current is compared to a predefined I_C. If the current is higher than I_C the next step depends on if the current overshoot is taken into account or not. If the overshoot is neglected, the current is limited to the nominal I_C value and the forming or set process ends and the applied voltage is ramped down quickly to zero. The case with the overshoot is not considered in this work, but briefly discussed in 3.3.2. For the reset process there is no current limitation, but we define a reset stop voltage, which is the same for all the switching cycles during a simulation. After having the current and potential information in the next step the heat flow equation is solved to obtain the local temperature. Now we have all the required information to calculate or update the transition rates involved in the KMC method and can determine the next transition process and the time increment. The whole simulation is completed when a predefined simulation time is reached.
All the values of the parameters used in this chapter that always have a constant value are listed in table 3.1.

Table 3.1: Values of the main physical parameters related to HfO_2 used in chapter 3

Parameter	Definition	Value
ν_0	characteristic vibration frequency of $V_O^{\bullet\bullet}$s	10^{13} Hz
α	symmetry factor	0.55
a	lattice constant	0.25 nm
E_R	energy barrier of $V_O^{\bullet\bullet}$ recombination	1.0 eV
N_0	boundary value for the number of $V_O^{\bullet\bullet}$ to switch the conduction mechanism from TAT to drift	115
d_0	boundary value for the tunnelling gap to switch the conduction mechanism from TAT to drift	0.5 nm

3.3 Result and discussion

The developed KMC model is used to simulate the forming and switching processes both in 2D and 3D. First the results of the 2D model are presented, then we upgrade it to the 3D.

3.3.1 2D system

Forming process

Electroforming is an initialization process and is required to trigger on the switching property for the subsequent cycles. Despite happening only once during the lifetime of a cell, the forming process and the conditions applied to it can affect the switching properties of the device significantly. Therefore it is crucial to avoid any pre-assumptions about this process and simulate it from scratch. The DC sweep analysis of the forming process was carried out in a 2D structure using the KMC model. For the entire 2D simulation it is assumed that $E_D = 0.7$ eV and $E_G = 3.0$ eV. During the HRS, where still the $V_O^{\bullet\bullet}$ concentration is low and TAT is the major conduction mechanism, the electron occupational probability should be calculated in order to obtain the leakage current. Fig. 3.5 demonstrates the evolution of the electron occupational probability during the forming step at different voltage regimes, starting from the low voltage at stage (I) to the high voltage at stage (IV). With all the values for the parameters mentioned in table 2.1 and table 3.1 the calculated electron occupational probability close to the cathode (inert electrode) is less than the anode (active electrode) and the probability close to the cathode decreases by increasing the applied voltage.

Fig. 3.6 shows the simulation result of the forming characteristics in a metal-HfO_2-metal ReRAM device. The simulated device size is 10×10 nm^2. A positive voltage is applied to the active electrode, while the inert electrode is grounded. This positive applied voltage is linearly ramped up with time. The unit of the calculated current using eq (2.18) in 2D should be A/m, because the unit of the point charges in 2D is C/m. The assumption here is a uniform distribution of the space charge in the third direction. In the demonstration of current values in 2D, the current in the thickness of 1.7 nm in the third dimension is considered, which is the value used in our programming to normalize variables with the unit of meter. The corresponding evolution of the CF inside the oxide and the temperature profile in 4 stages named A-D are represented in Fig. 3.7. These stages belong to different voltage regimes, which are marked on the forming $I-V$ curve in Fig. 3.6. At initial stage A, where the applied voltage is still quite low, the memory cell is in the HRS with no established CF. Only there are a few $V_O^{\bullet\bullet}$s, which were initially distributed at some random GBs inside the oxide. With increasing the applied voltage corresponding to the stage B the electric field becomes stronger and the energy barrier of the generation of

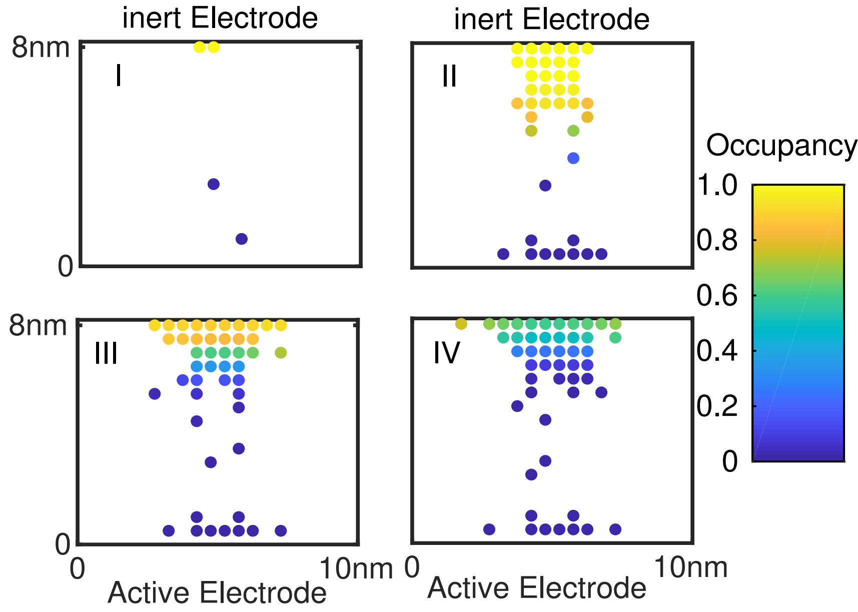


Figure 3.5: Electron occupational probability during the forming process at different stages of the applied voltages [118].

a new $V_O^{\bullet\bullet}$ decreases. This results in creation of new $V_O^{\bullet\bullet}$s inside the oxide and migration in the direction of the applied electric field toward the inert electrode. At the same time the current also increases when there are more traps inside the oxide to assist the electron transport. The higher current and electric field results in higher power consumption and consequently the temperature rises by Joule heating effect. The power dissipation, which is the product of the electric field and the current density, increases for higher applied voltages. The increasing temperature and electric field at stage C cause faster generation of $V_O^{\bullet\bullet}$, which again boost the current and temperature. This positive feed back between electric field and temperature and vacancy generation rate continues until stage D, when the current reaches a predefined I_C. The forming process is completed at this stage and the corresponding time and voltage are known as forming time and forming voltage, respectively. The forming time and voltage are two important parameters, which can affect the switching properties. In this simulation the I_C value is set to 100μA. Varying I_C can change the final shape of the filament after the forming process, which can affect the entire switching characteristics.

To investigate the effect of I_C on CF properties, the forming process was simulated for 3 cases: $I_C = 100$ μA, $I_C = 500$ μA and $I_C = 1000$ μA. Fig. 3.8 shows the simulated CF shape after the leakage current is equal to I_C and the

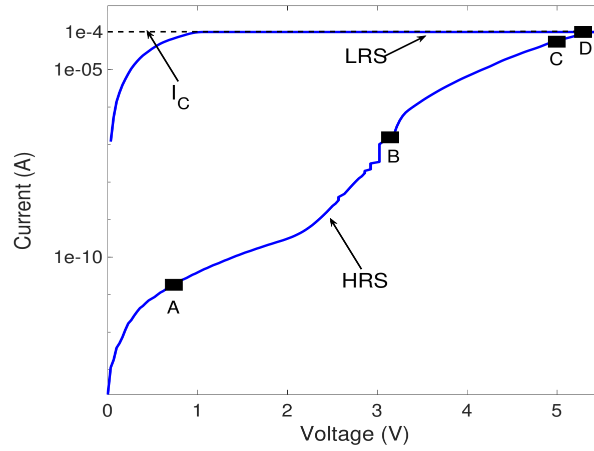


Figure 3.6: I-V plot of the forming process in a 2D structure with $I_C = 100\mu$A [119].

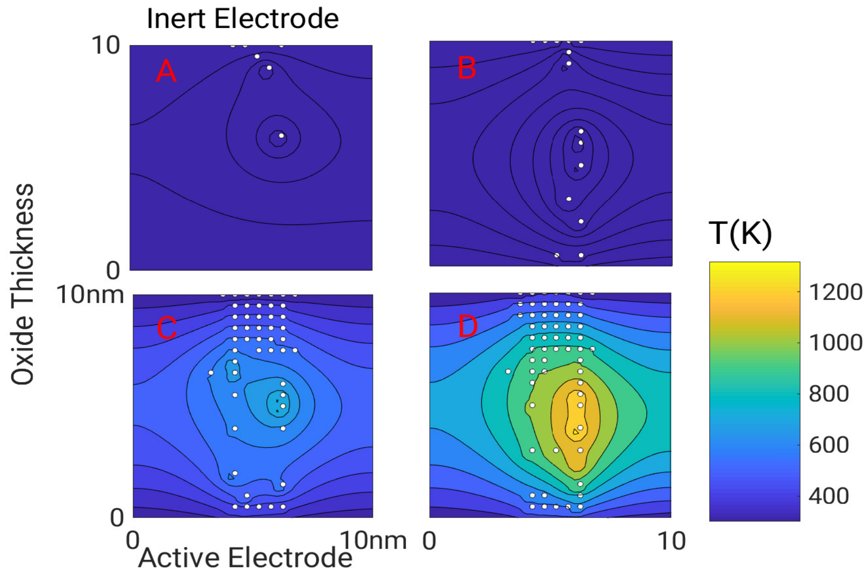


Figure 3.7: The evolution of the $V_O^{\bullet\bullet}$ distribution and temperature in the oxide during the forming process in a 2D device at different stages A($V = 0.5$ V), B($V = 3.0$ V), C($V = 5.0$ V), D($V = Vf = 5.3$ V) marked on the forming $I - V$ in Fig. 3.6 [119].

forming process is completed. For higher I_C the CF has a higher number of $V_O^{\bullet\bullet}$, which makes it more stable and later in the switching process causes less variability. One of the switching properties effected by the value of I_C is the resistance of the cell in the LRS as shown in Fig. 3.9. In this figure a different set of values for the current compliance than the values in Fig. 3.8 is chosen as:

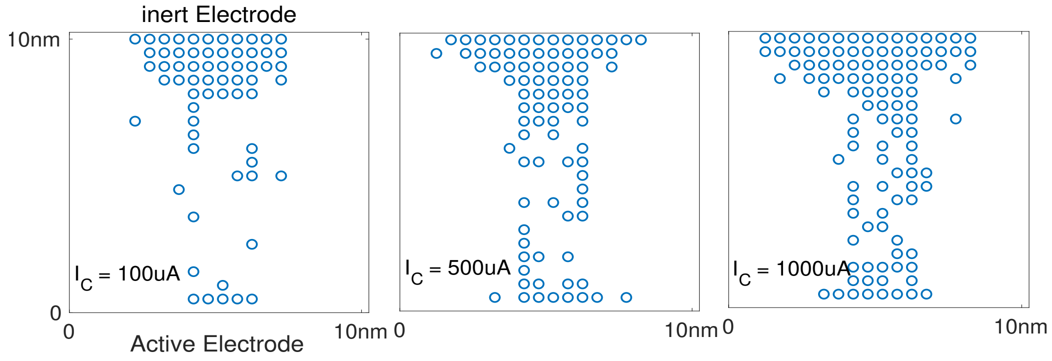


Figure 3.8: CF generation at the end of forming process for different current compliances, $I_C = 100\ \mu$A, $I_C = 500\ \mu$A and $I_C = 1000\ \mu$A [119].

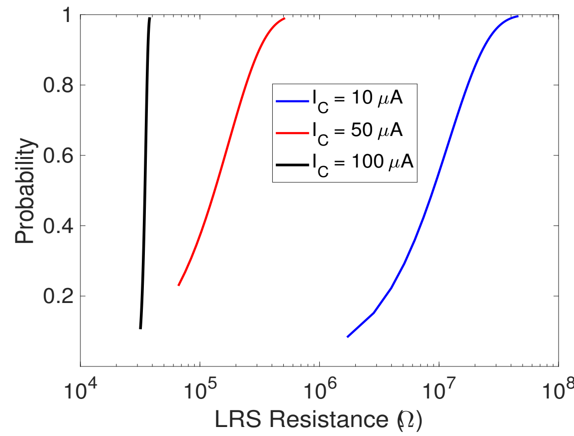
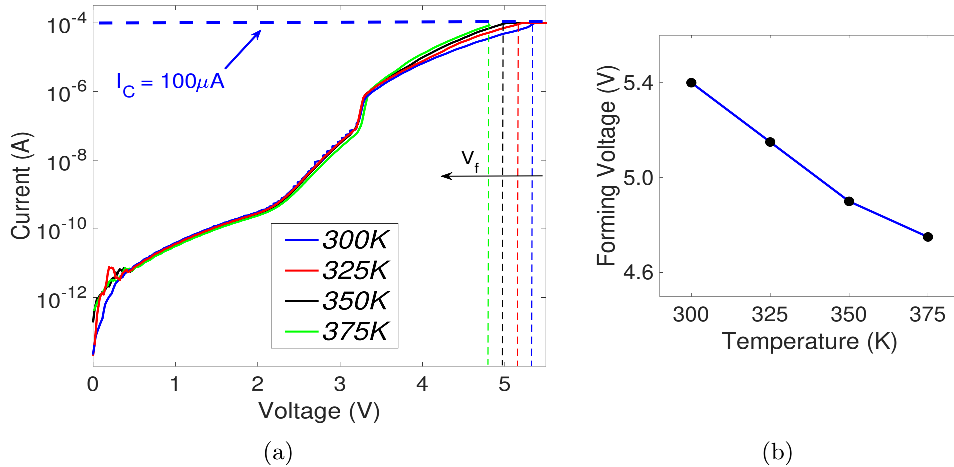


Figure 3.9: LRS resistance for different current compliances $I_C = 10\ \mu$A, $I_C = 50\ \mu$A and $I_C = 100\ \mu$A [119].

$I_C = 10\ \mu$A, $I_C = 50\ \mu$A and $I_C = 100\ \mu$A, since for each current compliance the LRS distribution of 100 switching cycles is simulated. For very large values of I_C this could be computationally very expensive and choosing a smaller value for I_C works well for our purpose. For higher I_C the LRS resistance has less variation. This is consistent with the observation made in Fig. 3.8.

The effect of the external temperature on the forming process was studied as shown in Fig. 3.10. The DC sweep analysis of the forming process under different external temperatures are demonstrated in Fig. 3.10(a). Fig. 3.10(b) depicts the relation between the forming time and the external temperature. The higher external temperature leads to lower forming voltage, hence a lower forming time. This can be explained by the strong dependency of the $V_O^{\bullet\bullet}$ generation rate at



(a) (b)

Figure 3.10: (a) Forming $I-V$ characteristics in a 2D cell at different external temperatures, where the vertical dashed lines align with the forming voltages at each curve (b) forming voltage as a function of the external temperature [119].

the anode/oxide interface and the diffusion rate in the oxide as presented in eq 3.4 and 3.6 on temperature. For higher temperature the Hf-O bond breakage happens at lower voltages, and the $V_O^{\bullet\bullet}$ diffusion is accelerated by lowering the diffusion barrier, therefore the forming voltage decreases.

The characteristics of the forming voltage and forming time i.e., the so called "Voltage-time dilemma" is a particular issue in resistive switching devices [33]. The time scale of the electroforming process can be significantly different depending on the voltage sweep rate. The dependency of the forming time and voltage on the voltage sweep rate was investigated as shown in Fig. 3.11. The results show a significant dependence of the forming time on the voltage sweep rate. For a change in the sweep rate from 0.01 to 0.1 V/s the forming time changes by one orders of magnitude. In contrast to the forming time, the forming voltage does not significantly change with the sweep rate. This might be due to the high activation energy for $V_O^{\bullet\bullet}$ generation, which leads to a low probability for $V_O^{\bullet\bullet}$ generation unless a threshold voltage is exceeded.

The transient simulation of the device switching under the applied pulse voltage is performed as another possible voltage mode in order to study the switching kinetics. For this purpose a transient simulation was performed for extremely short durations of the voltage pulse to investigate the influence of the voltage pulse amplitude on the forming characteristics. Fig. 3.12 shows the Weibull distribution of the forming time for different applied voltage pulse amplitudes. The forming time decreases significantly for a small increase of the applied voltage amplitude. This highly non-linear relation between the forming time and the amplitude of the applied voltage is due to the dramatic increase of

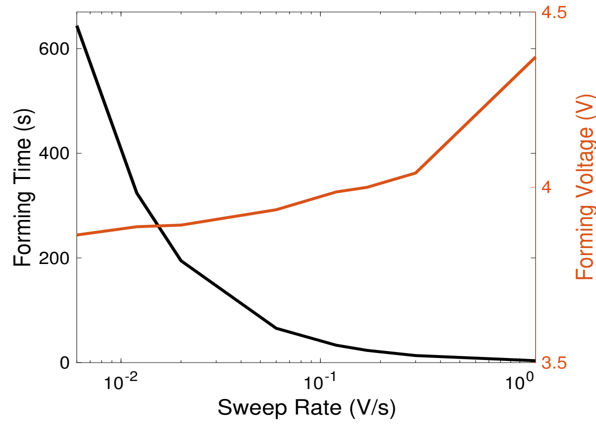


Figure 3.11: Variation of the forming time and voltage as a function of the applied voltage sweep rate [118].

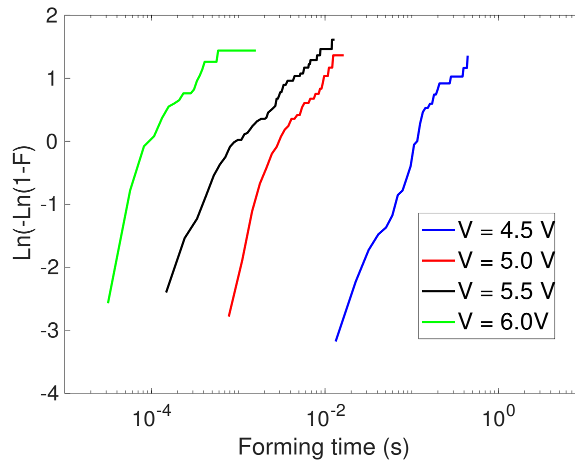


Figure 3.12: Weibull plot for different applied pulse amplitudes [118].

the generation and diffusion rate of $V_O^{\bullet\bullet}$. Fig. 3.12 also shows a trade-off between voltage down scaling and fast programming due to a significant increase of the forming time when the voltage amplitude is reduced. For example, for a voltage decrease of 25% the forming time increases by about one orders of magnitude.

E_G can differ significantly depending on the electrode/oxide interface conditions. Fig. 3.13(a) shows the simulation results of the forming $I-V$ characteristics for different values of E_G and the constant $E_D = 0.7$ eV. The applied voltage is ramped up with the sweep rate of 1 V/s. The corresponding evolution of the number of $V_O^{\bullet\bullet}$ as a function of the applied voltage is depicted in Fig. 3.13(b). Decreasing the formation energy barrier of $V_O^{\bullet\bullet}$ results in the mass generation of

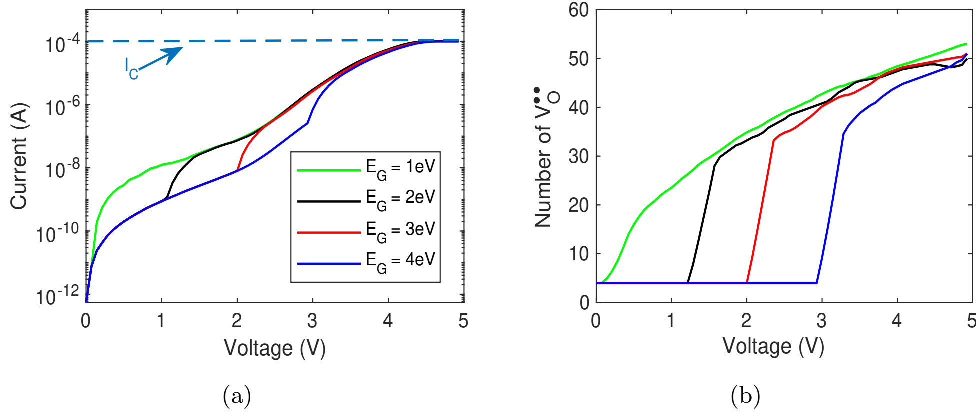


Figure 3.13: (a) Forming $I-V$ curves and (b) number of $V_O^{\bullet\bullet}$ as a function of the applied voltage during the forming process for different $V_O^{\bullet\bullet}$ formation energy [118].

new $V_O^{\bullet\bullet}$s at lower voltages and increasing the current in the pre-forming stage. After a specific increase of the voltage all curves follow the same trace, because of the same diffusion barrier for $V_O^{\bullet\bullet}$.

Set and reset processes

After the forming process is completed the operation of the ReRAM cell continues with the simulation of the set and reset processes. The CF that already exists inside the oxide after the forming process is ruptured partially during the reset process under an application of a negative ramped voltage and the cell is switched off. During the set process a positive voltage is applied and the CF is reconstructed and the cell is switched on. The negative and positive voltages during the reset and set processes are applied to the active electrode, while the inert electrode is always grounded. The simulated $I-V$ curve of the full switching process including the forming step and a couple of successive set and reset cycles is demonstrated in Fig. 3.14. In the set process an abrupt current increase in a specific voltage is observed. This corresponds to the reconstruction of the CF and transition of the cell from HRS to LRS. This voltage, which is referred to the set voltage and indicated on the $I-V$ curve with V_s, is an important switching parameter. The CF shape at the end of the set process is depicted in Fig. 3.15. During the reset process the CF is dissolved due to the application of a negative voltage. The current evolution during the reset is gradual as shown in the simulated $I-V$ curve. The shape of the CF at the end of the reset process is represented in Fig. 3.15.

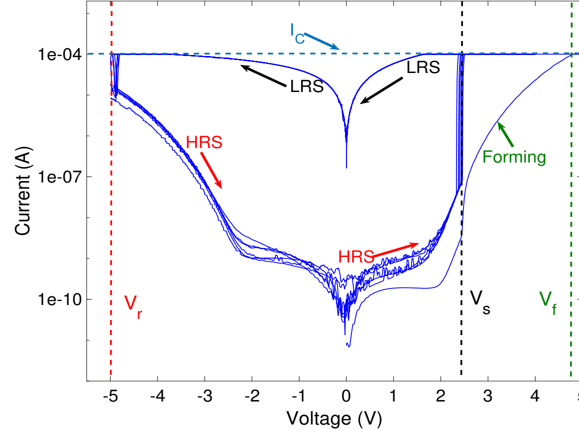


Figure 3.14: I-V plot including the initial forming and following set and reset processes in a 2D structure with $I_C = 100\mu$A [119].

3.3.2 3D system

Forming process

The simulation result of the current for a 3D structure during the forming process is shown in Fig. 3.16(a), where $I_C = 100\,\mu$A, $E_G = 3.0$ eV and $E_D = 0.7$ eV. The size of the simulated device is decreased in this case in comparison to 2D due to higher computational cost and is $5 \times 5 \times 5$ nm^3. Fig. 3.16(b) shows the $V_O^{\bullet\bullet}$ distribution profile during the forming in different voltage stages, also marked on the I-V curve. The $V_O^{\bullet\bullet}$ distribution at stage A contains only a few number of $V_O^{\bullet\bullet}$s corresponding to the high resistance state and low leakage current. At stage B by increasing the applied voltage some new $V_O^{\bullet\bullet}$s are generated at the bottom electrode (Anode) and move toward the top electrode (cathode) in the direction of the existing electric field. During this process the hopping rates of the electrons between cathode and $V_O^{\bullet\bullet}$s decrease dramatically, while at the same time the occupational probability of $V_O^{\bullet\bullet}$ increases significantly and this results in a gradual increase in current according to Eq. 2.18. In the next stages C and D a high number of $V_O^{\bullet\bullet}$ are generated due to the significant increase of the electric field and temperature. This causes an abrupt increase of current until it stops by the condition applied on I_C. The forming process is completed here leaving behind a CF connecting the TE and BE. Note that the sweep rate of the ramped voltage applied during the forming process is 0.33 V/s.

Fig. 3.17 shows the simulated temperature profile in a 2D vertical cross section of the cell at the initial and final stages of the forming process. At the initial stage the small electric field and current give low power dissipation and conse-

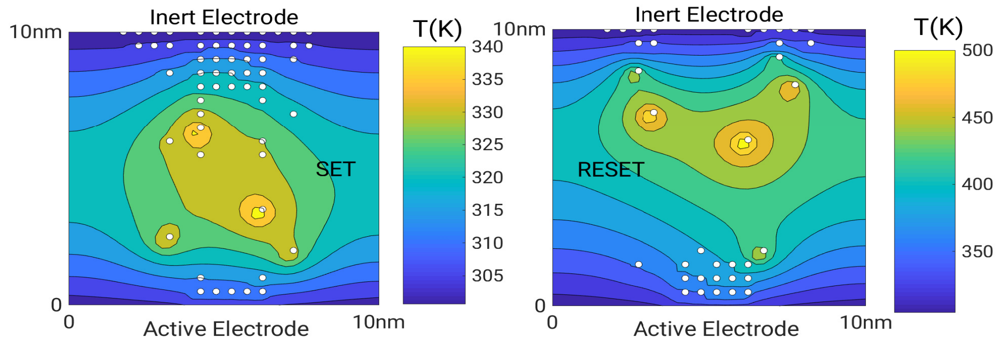


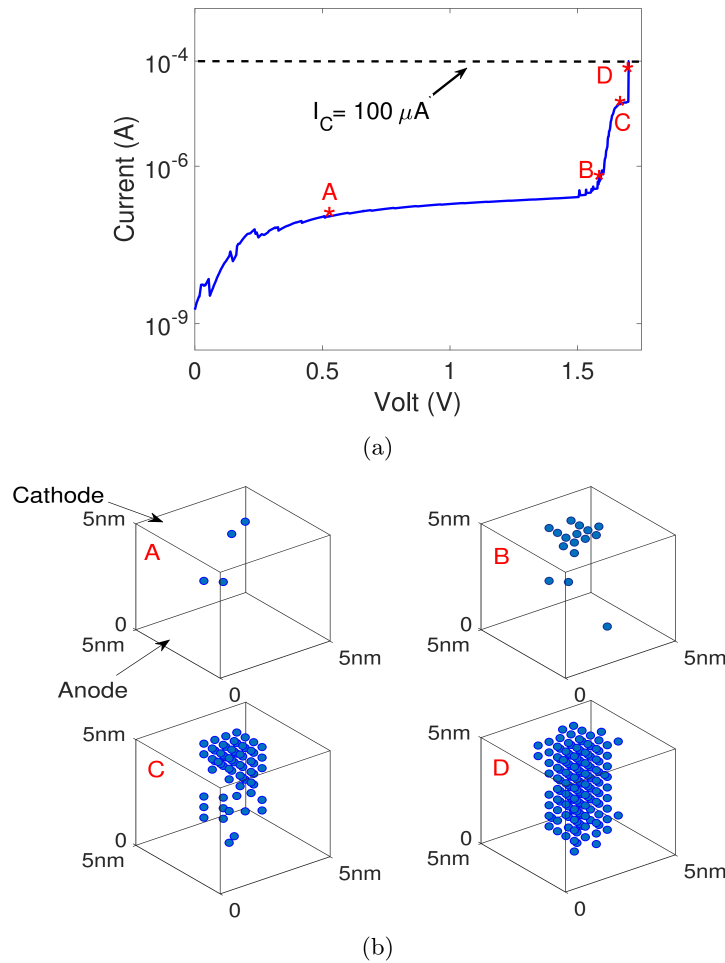
Figure 3.15: CF shape and temperature distributions at the end of set ($V_{set} = 2.3$ V) and reset ($V_{reset} = -5$ V) processes [119].

quently low temperature. The temperature increases considerably by creation of the CF at the final stage which corresponds to high electric field and current and power dissipation.

The forming voltage is an important parameter obtained from the forming $I - V$ characteristics. The oxide thickness is one of the factors affecting the forming voltage. Fig. 3.18 shows the simulation result of the forming $I - V$ curves for different oxide thicknesses indicated by d_{ox}. A clear increase of the forming voltage for thicker oxides is observed in this figure as expected. The diffusion path for $V_O^{\bullet\bullet}$ in the oxide increases for thicker oxides, which requires a longer time in the case of a DC sweep analysis, hence a higher voltage to form a CF connecting top and bottom electrodes. The linear relation between the forming voltage and oxide thickness in the $V_{form} - d_{ox}$ curve in Fig. 3.18 is consistent with the experimental results in literature (e.g. [120]).

The total time of the voltage ramp, as one of the forming conditions, was studied to have a better understanding of the characteristics of the forming voltage and forming time. The time scale of the electroforming process can be significantly different depending on the voltage sweep rate. Fig. 3.19 shows the variation of the forming time and voltage with the voltage sweep rate for $\alpha = 0.55$ and $\alpha = 0.7$. α is the symmetry factor introduced in eq 3.4. A high non-linearity is observed in the variation of the forming time so that for one decade increase of the sweep rate from 10^{-3}V/s to 10^{-2}V/s (for $\alpha = 0.55$) the forming time decreases by a factor of 10, while the forming voltage increases only 4% under the same condition. A higher value of α, which results in a higher impact of the applied electric field on the lowering of the $V_O^{\bullet\bullet}$ generation barrier, decreases the value of the forming time and voltage, as well as their dependency on the voltage sweep rate.

As mentioned in section 3.3.1, the current compliance is another forming condition that affects the CF shape and consequently the forming and switching



(a)

(b)

Figure 3.16: (a) Simulated $I-V$ characteristics of the forming process in a 5 nm thick cell with $I_C = 100\,\mu$A under an applied ramped voltage with the rate of 0.33 V/s (b) evolution of the $V_O^{\bullet\bullet}$ (blue balls) distribution and creation of the CF during the forming process at different stages A($V = 0.5$ V), B($V = 1.6$ V), C($V = 1.66$ V), D($V = V_f = 1.7$ V) marked on the $I-V$ curve [108].

properties. Fig. 3.20 shows the shape of the 3D CF at the end of the forming process for $I_C = 10\ \mu$A and $I_C = 400\ \mu$A, respectively. Results show that the higher I_C causes a CF with higher $V_O^{\bullet\bullet}$ density, which results in more stability in the following switching cycles.

Set and reset processes

The simulated $I-V$ characteristics including the forming process and a couple of switching cycles are demonstrated in Fig. 3.21(a), where the reset process is

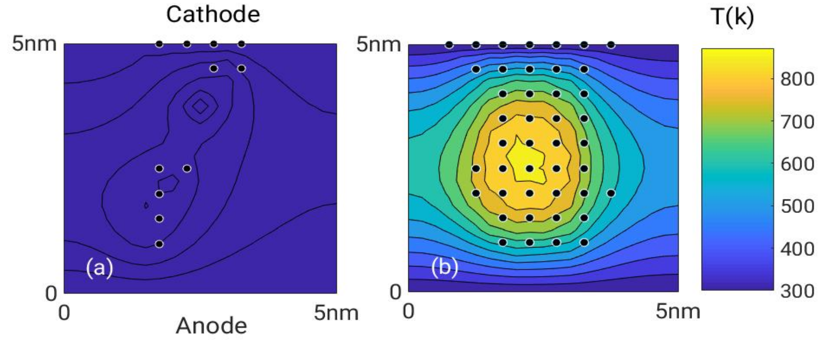


Figure 3.17: The temperature profile inside the oxide (a 2D cross section) in the (a) initial and (b) final stages of the forming process [108].

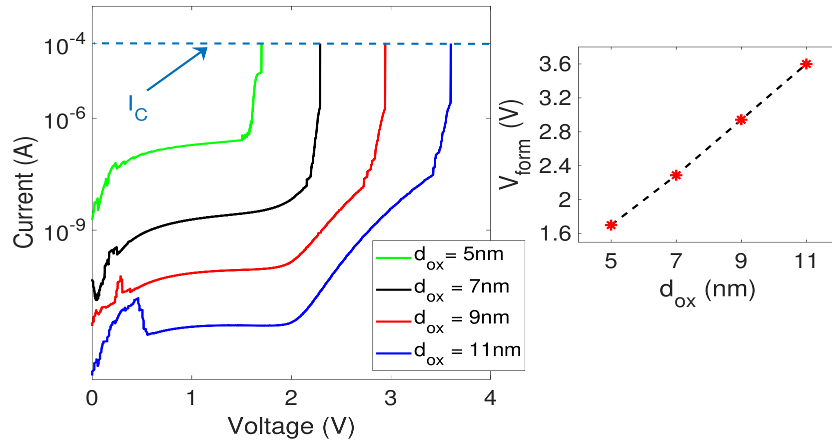


Figure 3.18: (a) simulated forming $I-V$ characteristic for devices with different oxide thicknesses, d_{ox} (b) dependency of the forming voltage on the oxide thickness [108].

limited with the reset stop voltage $V_{reset} = -2.7$ V and the forming and set processes with $I_C = 100\ \mu$A. The rates of the applied ramped voltages for the forming, set and reset processes are 0.33V/s, 2V/s and −3V/s, respectively. The choice of these sweep rates can affect the forming and switching properties significantly (cmp. Fig. 3.19). At stage A shown on the $I-V$ curve in Fig. 3.21(a), where the applied voltage is equal to V_{reset}, the switching from LRS to HRS happens, however the gradual drop of the current continues till stage B due to the further dissolution of CF and decreasing of the magnitude of the applied voltage. From B to C the creation or recombination of $V_O^{\bullet\bullet}$ is negligible due to the low value of the applied voltage. The current variation between these stages is because of the voltage change and shifting of the remaining part of CF from

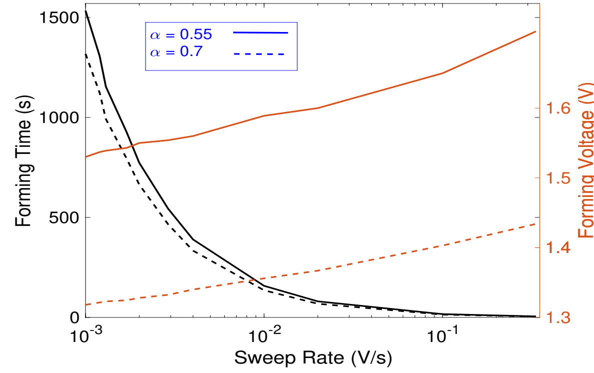


Figure 3.19: Dependence of the forming time and forming voltage on the voltage sweep rate for two cases of $\alpha = 0.55$ and $\alpha = 0.7$ [108].

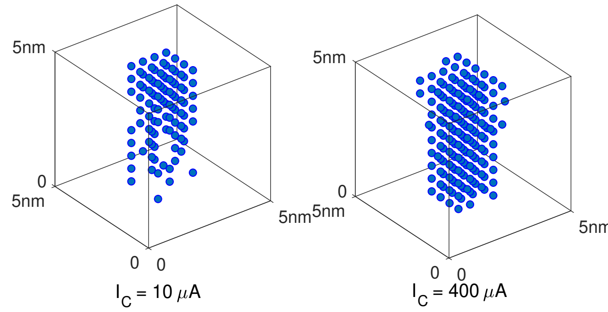
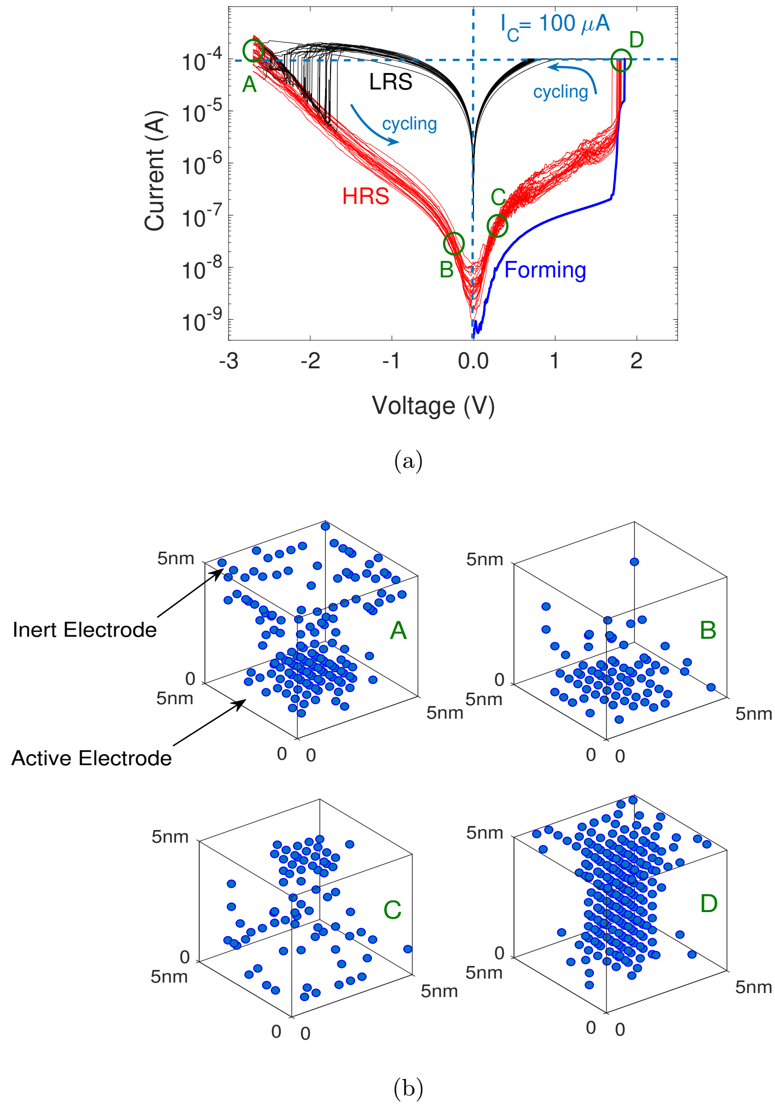


Figure 3.20: Comparison of the created CF at the end of the forming process for $I_C = 10\ \mu$A and $I_C = 400\ \mu$A [108].

bottom electrode toward top electrode due to the reversed electric field. The generation of a large number of $V_O^{\bullet\bullet}$s at stage D due to the high electric field and temperature causes an abrupt current increase and sets the cell to LRS. Fig. 3.21(b) shows the distribution of $V_O^{\bullet\bullet}$ at the initial and final stages of the reset and set processes. During the reset process a part of the CF close to the top electrode is partially annihilated and during the set process this part is recreated again.

The simulated switching process is pretty stable, so that after 100 cycles there is a reliable filament regeneration and annihilation resulting in a considerable window between LRS and HRS resistances. Fig. 3.22(a) shows the CF shape at the end of the last set and reset processes of the total 100 cycles. The filament shape after the set process is as expected for a normal successful transition to LRS and no degradation to the HRS is observed. After the last reset process



(a)

(b)

Figure 3.21: (a) Simulated $I-V$ characteristics including the forming and multiple set and reset processes (b) Partial rupture and reconstruction of CF during the reset and set processes in different stages A, B, C and D shown also on the simulated $I-V$ curve [108].

the CF has been dissolved sufficiently to switch off the cell without any failure. The evolution of the number of $V_O^{\bullet\bullet}$ during the simulated 100 switching cycles is depicted in Fig. 3.22(b), which shows consistent generation and recombination of $V_O^{\bullet\bullet}$s.

The relation of E_G/E_D is an important factor, which depends on the interface conditions. For example using of an oxygen exchange layer between the

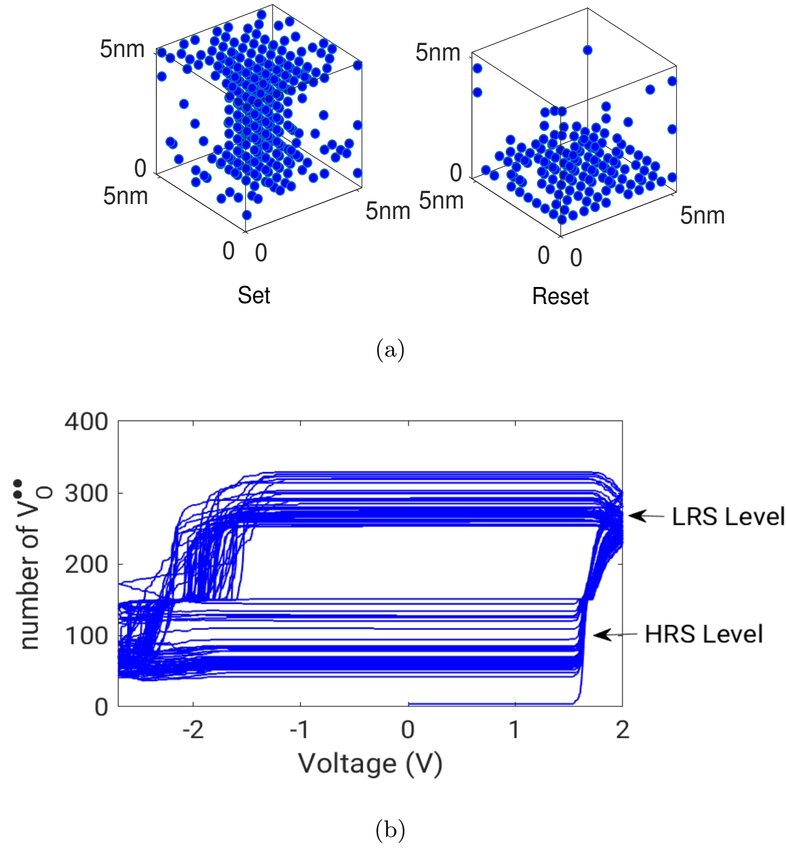
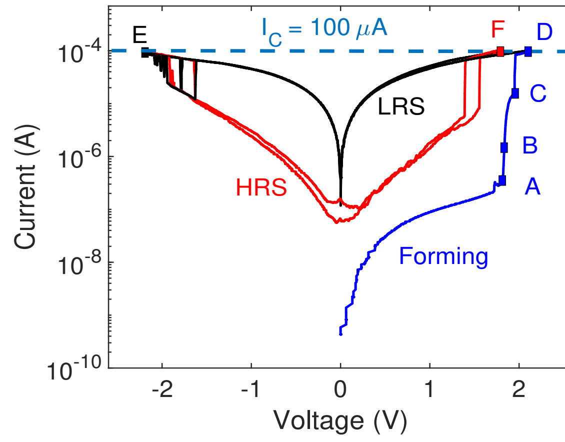
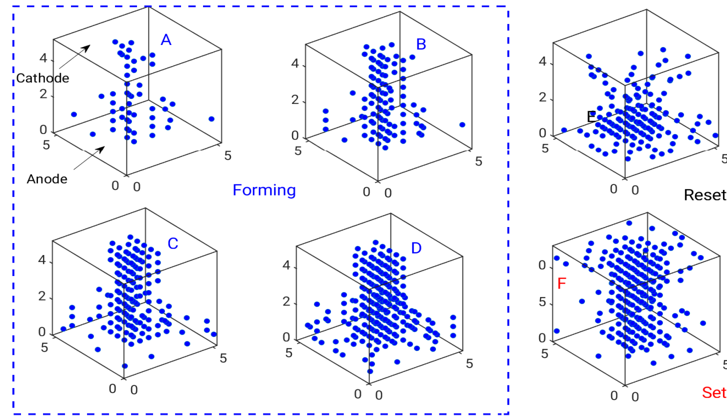


Figure 3.22: (a) Variation of the number of $V_O^{\bullet\bullet}$ during 100 switching cycles (b) shape of the CF at the last set and reset processes.

active electrode and the dielectric (i.e. Hf/HfO_2 stack) can result in a rather smaller generation energy for $V_O^{\bullet\bullet}$. The values considered for E_G and E_D in the simulation results shown in Fig. 3.16 and Fig. 3.21 are 3.0 eV and 0.7 eV, respectively. For this set of values the direction of the CF growth during the forming and set processes is from top electrode (cathode) toward the bottom electrode (anode). For a different interface condition these values can be different. We tried a different set of values for generation/diffusion barriers, i.e. $E_G = 2.2$ eV and $E_D = 1.1$ eV and the result is shown in Fig. 3.23. In this case the generation at the interface is faster compared to the $V_O^{\bullet\bullet}$ migration inside the oxide, which leads to a higher density of $V_O^{\bullet\bullet}$ close to the anode during the forming process at different stages. This is the opposite of the the CF evolution shown in Fig. 3.16. This result is consistent with the experimental observations of the different growth directions [121]. The corresponding $I-V$ curve including the forming step and a couple of the switching cycles for the new set of values for



(a)



(b)

Figure 3.23: (a) Simulated $I-V$ curve including the forming and a couple of the switching cycles for the values of $E_G = 2.2$ eV and $E_D = 1.1$ eV (b) Corresponding evolution of the CF shape during the forming and set and reset processes [108].

E_G and E_D is demonstrated in Fig. 3.23(a). The evolution of CF during the forming step and the shape of CF at the end of the reset and set processes are represented in Fig. 3.23(b).

As discussed in 2.2.1 the relation of the electron hopping rate between two $V_O^{\bullet\bullet}$s (h_{ij}) and between a $V_O^{\bullet\bullet}$ and an electrode (P_{Ei}) can determine the electron distribution inside the filament. Two scenarios are possible here: (I) if in average P_{Ei} is larger than h_{ij}, then electrons tend to accumulate close to the electrode, where they are injected to the oxide, otherwise (II) the electrons travel faster through the filament and cause a higher electron occupational probability/electron density close to the electrode, where they want to jump to it.

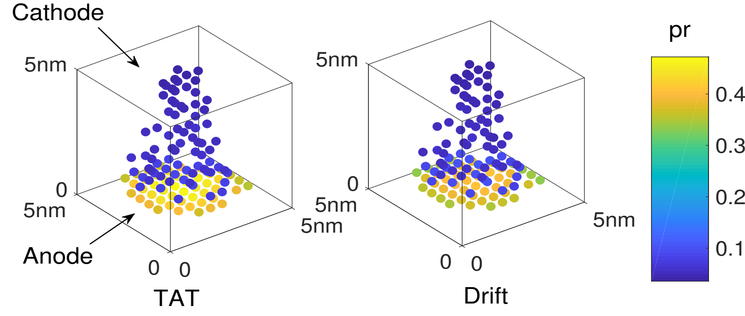


Figure 3.24: Electron occupational probability right before (TAT) and after (Drift) HRS to LRS transition during the forming process.

Which scenario happens, depends on the defect depth and the tunnelling barrier between the electrodes and the oxide. In the results introduced till now based on the values considered for these parameters the scenario (I) was true. In a different study we kept the tunnelling barrier between the electrode and the oxide as before but increased h_{ij} in average by lowering the defect depth and increasing the characteristic vibration frequency of electrons. The distribution of the electron occupational probability using the developed TAT model just before the HRS to LRS transition during the forming process is represented in Fig. 3.24. A low electron occupation region is formed close to cathode, because of fast electron hopping between $V_O^{\bullet\bullet}$s. The electron distribution right after the HRS to LRS transition, where the charge transport mechanism changes to drift-diffusion, is depicted in the same figure. A consistent transition of the electron distribution between two conduction mechanisms can be observed. In both cases electrons tend to be more present close to the anode, to which the positive voltage is applied, than close to the cathode, which is grounded.

The simulation result of the $I-V$ curve including the forming step and a couple of switching cycles are demonstrated in Fig. 3.25(a). The applied voltages in the forming, set and reset processes are ramped voltages with the sweep rates of 0.25 V/s, 1.5 V/s and -1.5 V/s, respectively. This is a typical form of the $I-V$ curve expected from the experimental results as shown in Fig. 3.25(b), which shows the measured $I-V$ curve for a HfO_2-based device. The difference between calculated and experimental $I-V$ characteristics results from the different pre-forming 'virgin' states. In the HfO_2 sample the virgin state is highly insulating whereas an initial leakage path via predefined oxygen-vacancy mediated defects was assumed for the simulation. A higher initial resistance causes a higher forming voltage. The inset of Fig. 3.25(b) shows the device structure used in the experiment, where the bias is applied to the Hf electrode that is oxidizable and the Pt electrode is grounded. This is consistent with the switching/forming

polarity in the simulation.
The values considered for the generation and diffusion barriers of $V_O^{\bullet\bullet}$ are $E_G =$ 1.6 eV and $E_D = 0.9$ eV, respectively. With this set of values the evolution of the CF during the forming process is demonstrated in Fig. 3.25(c). The direction of the CF growth is from cathode toward anode.

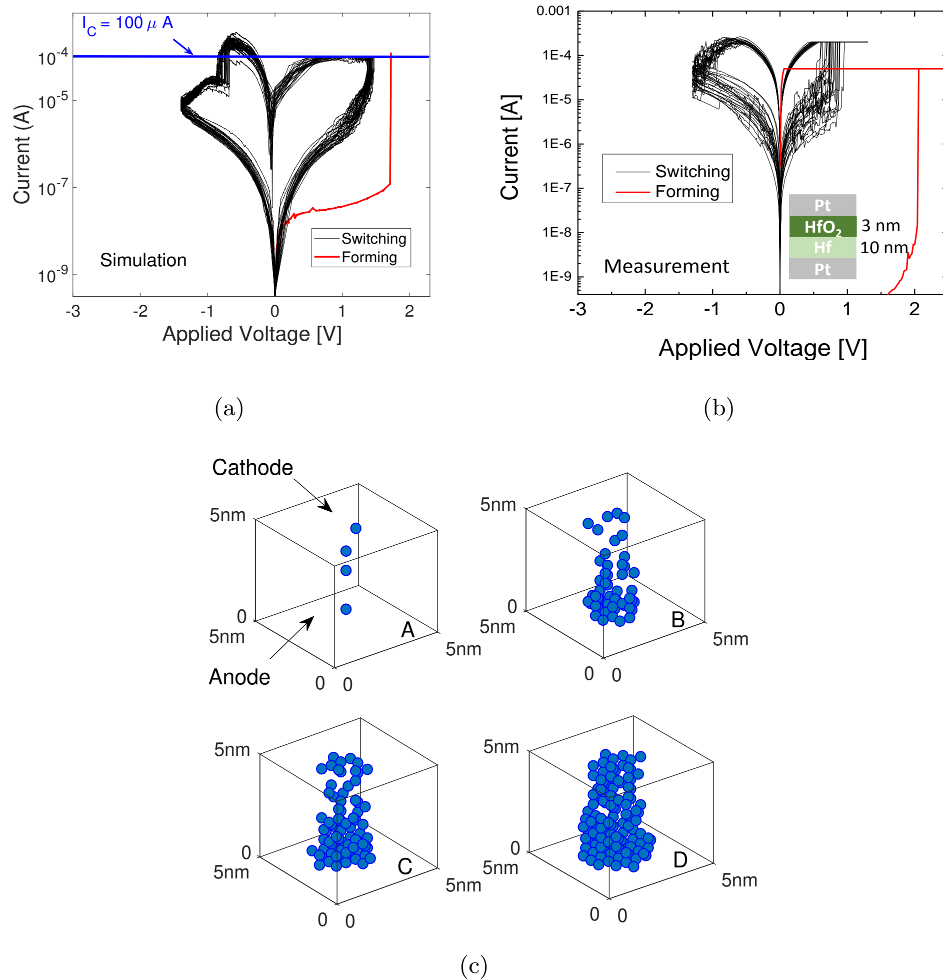


Figure 3.25: (a) Simulated and (b) measured $I-V$ characteristics including the forming step and a couple of the set and reset cycles (c) evolution of the CF during the forming process at different stages of the applied voltage, $V_A = 0.5$ V, $V_B = 1.85$ V, $V_C = 1.95$ V and $V_D = 2.0$ V.

Current overshoot

As mentioned before I_C is the maximum allowed current during the filament formation to avoid irreversible physical damage to the device. In practice, a current limiter is used in the cell to enforce the cell current to the I_C value. Depending on the response speed of the current limiter to the rapid increase of the leakage current at the end of the filament formation and transition from HRS to LRS, a current overshoot might happen [41, 122]. Fig. 3.26 shows the $I - V$ characteristics of the first reset after the forming with the same initial conditions for three cases: no overshoot, overshoot with two RC time constants, $t_{os} = 100$ ns and $t_{os} = 1000$ ns. When there is an overshoot effect the current at the end of the forming step exceeds the nominal value for I_C. Since there is a linear relation between the maximum forming current and the maximum reset current [41], a higher maximum reset current is observed when there is an overshoot. The longer t_{os} also results in higher maximum reset current. The comparison of the $\mathrm{V_O^{\bullet\bullet}}$ distribution at the end of the forming process for no overshoot and overshot with $t_{os} = 100$ ns in Fig. 3.26 shows that during the overshoot time due to the high current and temperature a considerable number of $\mathrm{V_O^{\bullet\bullet}}$ are generated and the filament continues to grow, which causes a higher forming current and consequently a higher maximum current during the reset process.

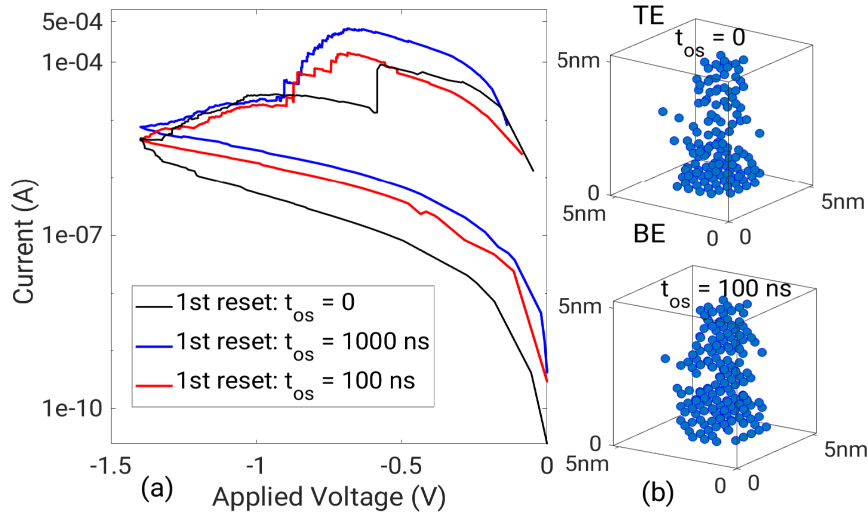


Figure 3.26: (a) Simulation result of the $I - V$ characteristics including the first reset process after the forming process with the same initial conditions for $\tau = 0$, $\tau = 100$ ns and $\tau = 1000$ ns (b) shape of the CF after the forming process for $t_{os} = 0$ and$t_{os} = 100$ns.

3.3.3 Comparison of 2D and 3D results

The comparison of the forming characteristics in 2D and 3D models in Fig. 3.6, 3.16(a) and 3.25(a) shows the abrupt current increase in 3D, which is also observed in the experiment (Fig. 3.25(b)), while the current increase in 2D is more gradual as also reported in the 2D KMC study of ReRAM in [69]. The general shape of the switching characteristics in 2D and 3D shown in Fig. 3.14 and 3.25(a) and comparing them with experiment in Fig. 3.25(b) justifies the considerable higher accuracy of the 3D model. The reason for this is the approximation made in 2D and lack of freedom in the third dimension. Ignoring the variations in the third dimension can also reduce the stochastic property of an ReRAM cell. The current fluctuations specifically in the reset process, which are more important, in 3D results are higher than 2D.

Chapter 4

Variability of the switching parameters

Achieving reliable ReRAM devices faces several challenges that are needed to be addressed and solved including switching variability. The Controllability of the set and reset characteristics through different forming and switching conditions is one of the key issues that needs to be studied. The stochastic nature of the resistive switching process is a source of the variability of the switching parameters. Here we focus on two types of the intrinsic variability sources: 1) the cycle to cycle variability that characterizes the device stability during the switching time 2) random telegraph noise (RTN) affecting the ReRAM behavior during the read process.

4.1 Cycle-to-Cycle variability

The set and reset characteristics in general show some variations over different switching cycles. To study the effect of these fluctuations on the switching parameters, the probability distribution of the resistance in the HRS and LRS for 100 switching cycles is demonstrated in Fig. 4.1(a). In general the fluctuation in the HRS is higher than compared to the LRS, which is consistent with the experimental findings of the HfO_2-based cell shown in Fig. 4.1(b). The origin of the variations in the LRS and HRS causes this difference. The variation of the CF characteristics like the cross section, shape and concentration of $V_O^{\bullet\bullet}$ inside the CF can cause parameter fluctuations in the LRS. The HRS fluctuations come from the variation of the barrier thickness between the partially dissolved CF tip and the TE, which results in an exponential change of the electron tunnelling probability (Eq. 2.2) and consequently the leakage current. The comparison of the experimental data and the simulation results shows a difference in the R_{HRS}/R_{LRS} ratio and the shape of the HRS I-V branch. This difference might

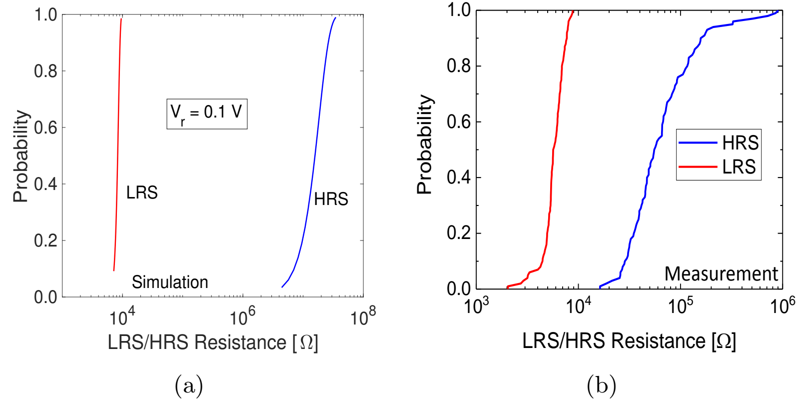


Figure 4.1: (a) Simulated and (b) measured probability distribution of the resistance in the HRS and LRS with $I_C = 100$ μA, $V_{reset} = -2.7$ V and $V_{\mathrm{read}} = 0.1$V [108].

originate in the assumption of TAT as the dominating conduction mechanism in the HRS. The conduction mechanism within the physical device might be more complicated and several conduction mechanisms coexist depending on the local $V_O^{\bullet\bullet}$ concentration and the applied voltage.

In the following the effects of different forming and switching conditions on the switching parameter variation is discussed. All the results in the cycle-to-cycle variation studies are simulated using the 3D KMC model.

Time evolution of set

The influence of the applied set voltage in ramped and pulse mode on the switching characteristics is studied in this section. Fig. 4.2 shows the simulation results of the dependency of the set time on the applied ramped voltage during the set process. The set time here is defined as the time, at which the cell current is equal to the current compliance. 50 switching cycles are simulated under different values of the applied voltage sweep rate. Fig. 4.2(a) demonstrates the probability distribution of the set time for one decade increase of the sweep rate from 2 V/s to 20 V/s. The mean value of each distribution is plotted as a function of the voltage sweep rate in Fig. 4.2(b). The set time in average decreases with increasing the sweep rate, while the variance in the probability distribution of the set time increases.

The set time variation under the pulse voltage application is studied and the simulation results are represented in Fig. 4.3. Fig. 4.3(a) shows the probability distribution of the set time over 50 switching cycles for different values of the pulse amplitude of the applied set voltage. The corresponding average value of each distribution as a function of the applied pulse amplitude is depicted in Fig. 4.3(b). Based on these results, doubling the set voltage amplitude causes

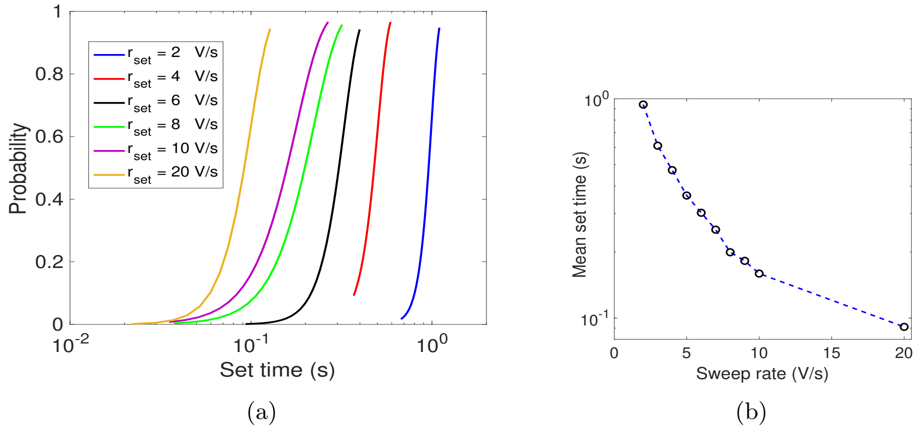


Figure 4.2: (a) Probability distribution of the set time of 50 switching cycles for different values of the applied voltage sweep rate during the set process (b) mean set time relation with the applied voltage sweep rate.

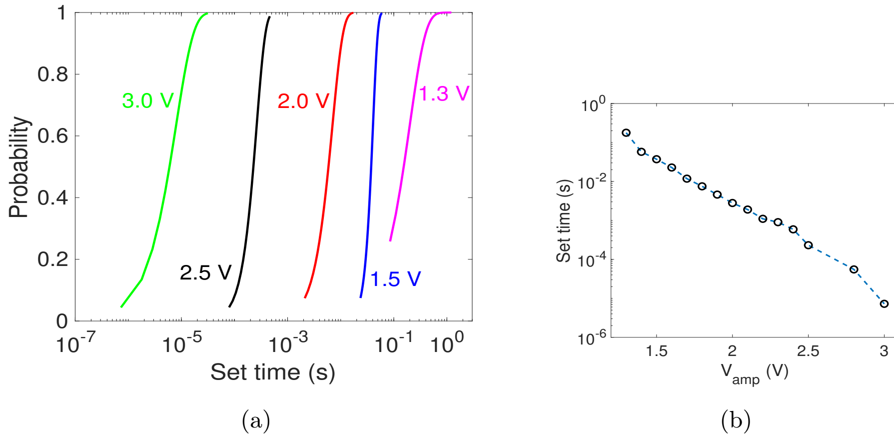


Figure 4.3: (a) Probability distribution of the set time of 50 switching cycles for different values of the voltage pulse amplitude (V_{amp}) during the set process (b) mean set time relation with the amplitude of the applied pulse voltage.

a reduction of the set time by around five orders of magnitude. This high non-linearity comes from the $V_O^{\bullet\bullet}$ diffusion barrier modulation by the electric field and temperature. Increasing the applied voltage and consequently increasing the electric field and local temperature can drastically reduce the diffusion barrier of $V_O^{\bullet\bullet}$ along the oxide and result in significant set time reduction.

Current compliance effect

Practical usage of ReRAM devices requires a lot of conditions including low power consumption. Therefore the high maximum current during the reset pro-

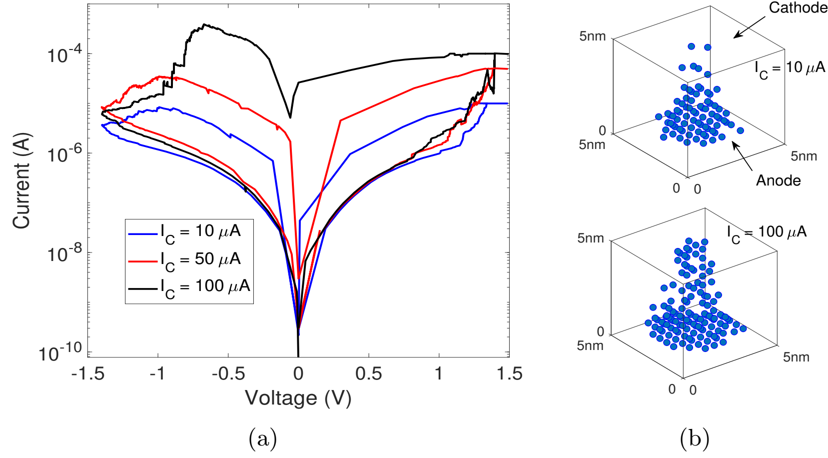
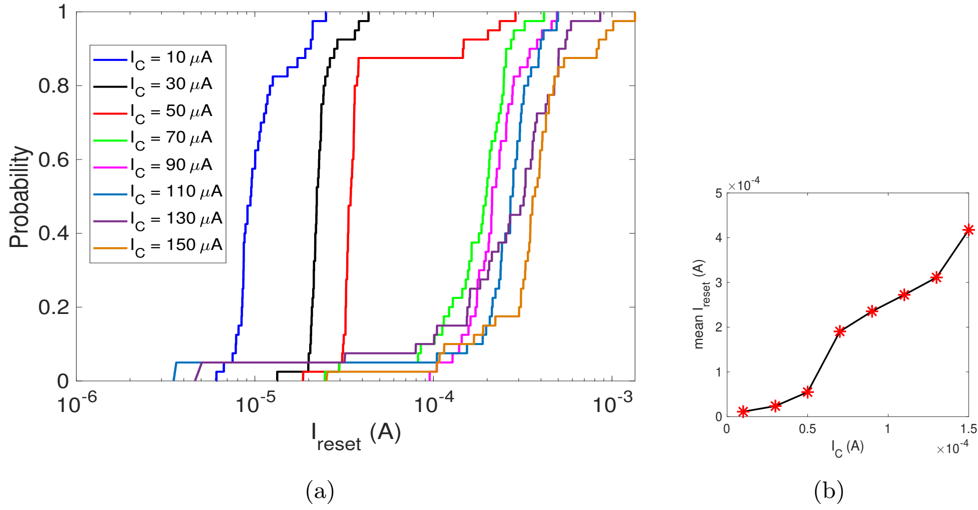


Figure 4.4: (a) $I-V$ characteristics including one successive reset and set process for different I_C values (b) CF shape at the end of the forming process for $I_C = 10\mu$A and $I_C = 100\mu$A.

cess, I_{reset}, can be an important limiting factor. Paying special attention to the factors affecting I_{reset} is of great importance. A well-known parameter to control I_{reset} is the maximum allowed current during the forming and set processes, I_C, [32, 59]. Fig. 4.4(a) shows the simulated switching $I-V$ characteristics for three different values of I_C. The dependency of I_{reset} on I_C can be seen in this figure, where by increasing I_C form 10 μA to 100 μA, I_{reset} goes up. The reason of this relation between I_{reset} and I_C can be explained by the simulation results in Fig. 4.4(b), which shows the CF shape at the end of the forming process for $I_C = 10\mu$A and $I_C = 100\mu$A. The CF for the case $I_C = 100\mu$A has clearly a higher density of $V_O^{\bullet\bullet}$. A denser CF during the reset process before switching to HRS causes a better electron conduction and a higher I_{reset}.

To have a better understanding of the I_C effect on I_{reset}, 50 switching cycles were simulated for a set of different I_C values and the result is demonstrated in Fig. 4.5. The probability distributions of I_{reset} are shown in Fig. 4.5(a) for different I_C changing from 10 μA to 150 μA. The larger I_C results in higher I_{reset} variation. I_{reset} strongly depends on the shape of the CF after the forming and set processes, which is controlled by I_C. Higher I_C causes a CF with a higher number of $V_O^{\bullet\bullet}$, which can fluctuate in and out of the CF and consequently increase the variation of the reset current. To have a better comparison the corresponding mean value of each I_{reset} distribution is also depicted as a function of I_C in Fig. 4.5(b). The results confirm the increase of the average value of I_{reset} by increasing I_C.

The choice of I_C can also affect the LRS parameters like the LRS resistance. Fig. 4.6 shows the simulation result of the LRS resistance distribution for different I_C. Lower I_C results in a higher resistance value and more variance in the



(a) (b)

Figure 4.5: (a) Probability distribution of the maximum reset current (I_{reset}) for different I_C values (b) mean I_{reset} as a function of I_C.

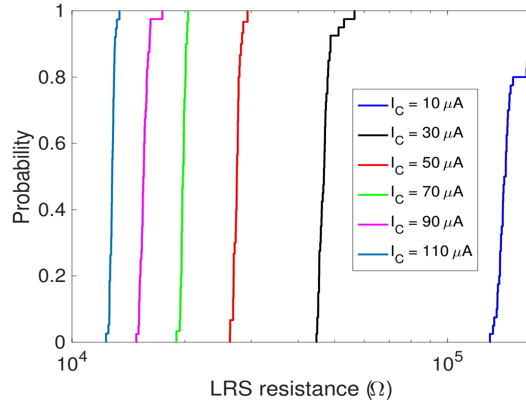


Figure 4.6: Probability distribution of the LRS resistance for different I_C values.

probability distribution. The generation of new $V_O^{\bullet\bullet}$s when a positive set voltage is applied can happen at much smaller voltages in comparison to the forming process. The reason is the residual of the CF remaining after the reset process, which does not exist in the virgin cell. The shape of the dissolved CF and the density of $V_O^{\bullet\bullet}$ inside the CF after the reset process plays an important role in the dynamics of the CF regrowth inside the oxide during the set process. With increasing applied voltage, more and more $V_O^{\bullet\bullet}$s are generated and the CF continues to grow and becomes more stable. The smaller I_C corresponds to lower $V_O^{\bullet\bullet}$ concentration, where the generation or rearrangement of few $V_O^{\bullet\bullet}$s can result in a considerable current change, hence higher variability from one cycle to

another is expected. On the other hand for higher I_C the CF is more stable and less variation in the LRS resistance is observed. The measured LRS resistance and I_{reset} as a function of I_C for various ReRAM material in [123] confirm the trends shown in Fig. 4.5 and Fig. 4.6.

Reset stop voltage effect

Unlike the set process, the end of the reset process is controlled by a stop voltage, V_{reset}. The choice of V_{reset} can be critical regarding its effect on the switching properties and variability of ReRAM devices. In this section the role of V_{reset} is studied in ramped and pulse voltage modes. Fig. 4.7 shows the simulated $I-V$ characteristics for different V_{reset}, where a ramped voltage is applied. Every

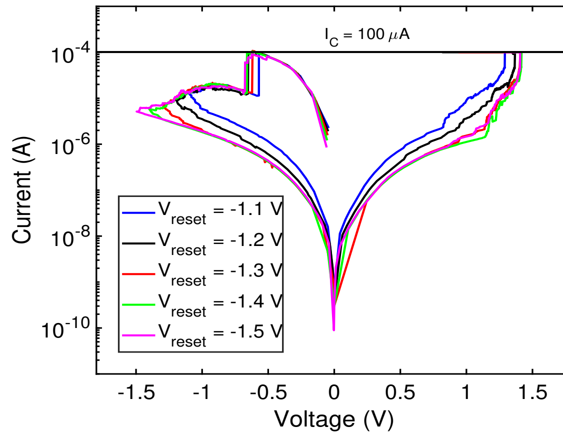


Figure 4.7: Switching $I-V$ characteristics including one set and one reset cycle for different values of the applied stop reset voltages (V_{reset}).

reset process in this figure started immediately after the forming process with the same initial conditions resulting in the same shape of the CF at the end of the forming. Increasing V_{reset} from -1.1 V to -1.5 V results in a lower current at V_{reset}. This can be explained by the simulation results represented in Fig. 4.8, which show 2D cross sections of the memory cell to compare the CF shape in three cases: (a) end of the forming process, where $V = V_{form} = 2$ V and the CF connects the top and bottom electrodes (b) end of the reset process, where $V = V_{reset} = -1.1$ V. In this case the part of the CF close to the inert electrode has been partially dissolved (c) end of the reset process, where $V = V_{reset} = -1.5$ V. The created gap between the tip of the filament and the inert electrode in this case is larger than case (b). The higher absolute value of V_{reset} increases the recombination rate and diffusion rate of $V_O^{\bullet\bullet}$s, resulting in a larger tunnelling gap and consequently lower tunnelling current. Fig. 4.7 also shows part of the

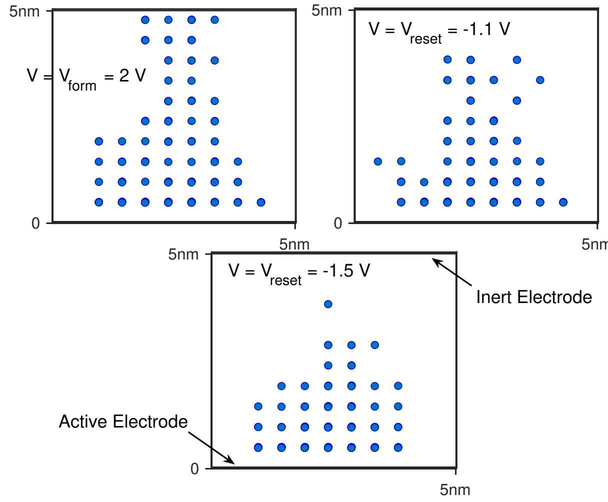


Figure 4.8: Shape of the CF at different applied voltages: $V = V_{\text{form}} = 2$ V, where the creation of the CF is completed and the device has switched to LRS, $V = V_{reset} = -1.1$ V, where the CF has been partially dissolved at the end of the reset process, $V_{reset} = -1.5$ V, where the CF has been dissolved even more due to the higher applied reset voltage. V_{form} and V_{reset} were all applied to the active electrode, while the inert electrode is grounded.

$I - V$ curve after the reset and during the set process. Increasing V_{reset} from -1.1 V to -1.5 V also increases the set voltage. The reason is the residual CF after the reset, which should be reconstructed during the set process. The more the CF is dissolved during the reset, a higher set voltage is required to rebuild it.

Next the effect of V_{reset} on the switching properties and variability was investigated in the ramped and pulse voltage modes. The distributions of the reset current for 50 switching cycles in the ramped voltage mode are shown in Fig. 4.9(a). Each distribution belongs to a different V_{reset}, while the applied voltage during the set process is the same for all cases. The mean and relative standard deviation (STD) of the reset current for each distribution as a function of V_{reset} are depicted in Fig. 4.9(b) for a better analysis. Increasing V_{reset} results in a lower average value for the reset current and also a lower STD, which means less variability. This trend was also observed in the experiment [52]. The average value of the reset current reduces with increasing V_{reset} as expected because of the growing tunnelling gap. By increasing V_{reset}, the reset current reaches a stable plateau as depicted in Fig. 4.7 and only small fluctuations are observed. In lower V_{reset} the gap between the tip of the CF and the electrode, which can be quite different from cycle to cycle, is smaller. Therefore any $V_O^{\bullet\bullet}$ fluctuation in this area can cause a considerable current fluctuation. The effect of V_{reset} on the LRS and HRS resistance values is shown in Fig. 4.9(c). The probability

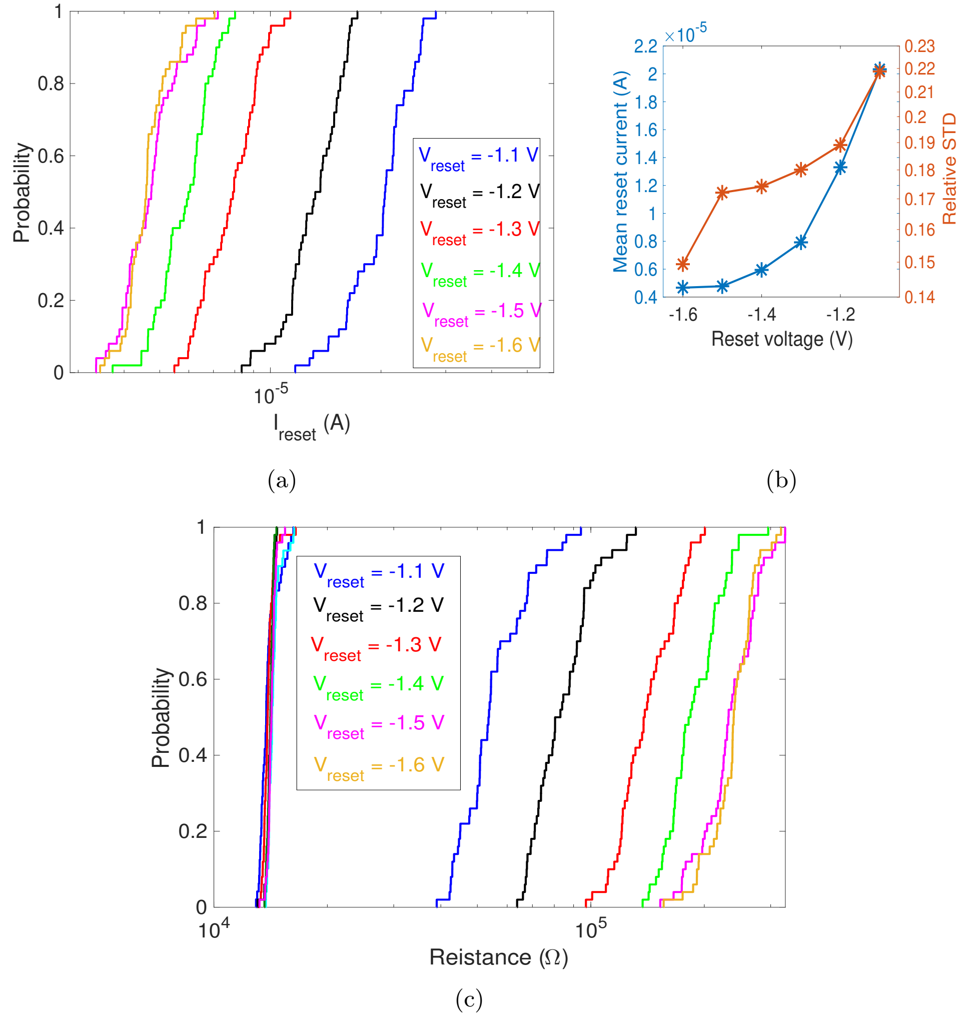


Figure 4.9: (a) Probability distribution of the leakage current at the end of the reset process, where $V_{applied} = V_{reset}$ for different values of V_{reset} for 50 switching cycles (b) mean value and relative standard deviation (STD) of the reset current (c) cumulative distribution of the LRS and HRS resistances for different values of V_{reset}.

distributions of the resistance in the HRS and LRS for different V_{reset} show that in contrast to the HRS, the LRS is almost independent of the V_{reset} value. Application of a larger V_{reset} causes a larger gap with higher resistivity within the CF, resulting in a larger HRS resistance. The variation of the HRS resistance is slightly higher for smaller V_{reset}. Generally as shown in Fig. 4.9(c) the resistance variation in the HRS is higher than in the LRS and the resistance window between the LRS and HRS increases with increasing the V_{reset} value.

The switching process under an applied pulsed voltage for different values

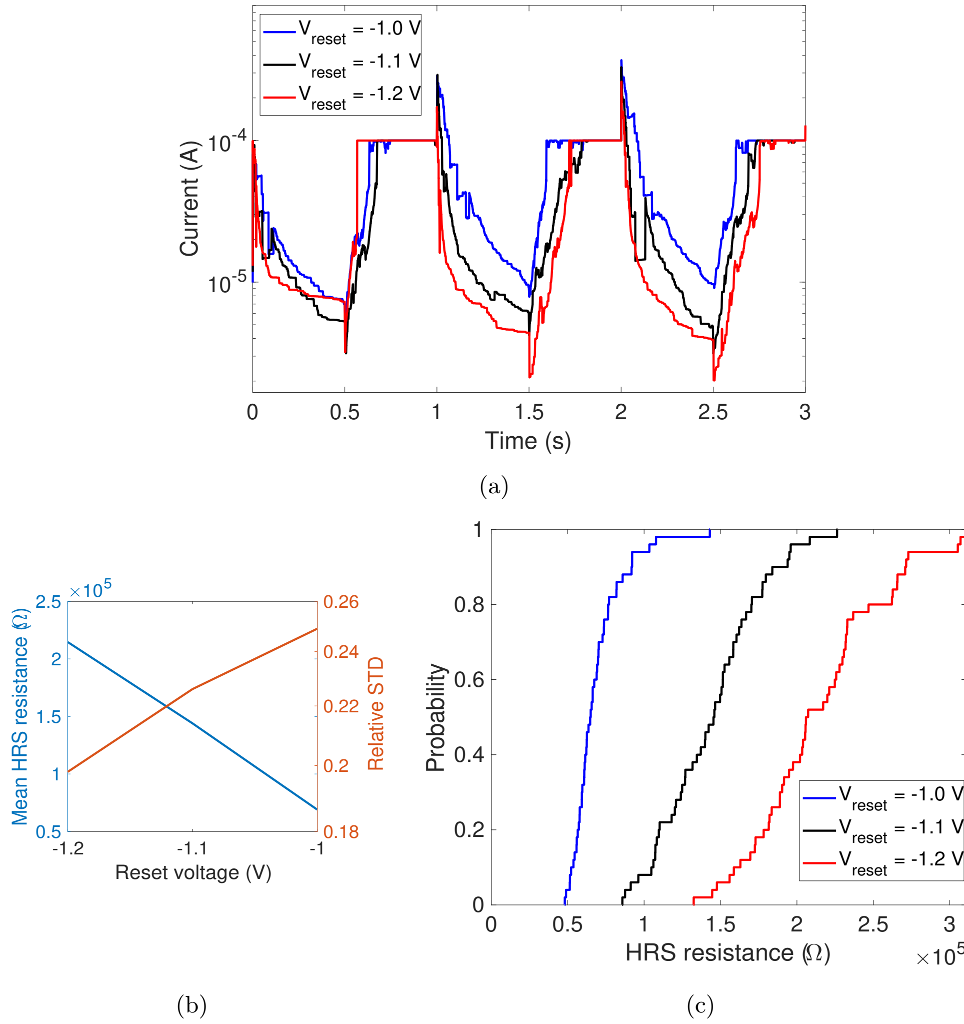


Figure 4.10: (a) $I-t$ characteristics under an applied pulse voltage for three different reset voltage pulse amplitudes and a fixed set voltage, $V_{\text{set}} = 1.3$ V (b) probability distribution of HRS resistance for different reset voltage pulse amplitudes (c) corresponding mean and relative standard deviation (STD) for each HRS resistance distribution.

of the reset pulse amplitudes was simulated. The applied voltage during the set process is a pulsed voltage with the amplitude of $V_{\text{set}} = 1.3$ V, while the amplitude of the reset voltage takes three values, $V_{reset} = -1.0$ V, $V_{reset} = -1.1$ V and $V_{reset} = -1.2$ V. The pulse duration for both set and reset voltages is 0.5 s. The simulation result of the transient current is shwon in Fig. 4.10(a). As we saw in the case of the applied ramped voltage, the current at the end of the reset process decreases by increasing the absolute value of V_{reset} due to the larger tunnelling gap at the end of the reset process. The current profile during

the reset process looks more noisy than the set process. The distribution of the HRS resistance for three different V_{reset} and the mean and relative STD of each distribution are shown in Fig. 4.10(b) and 4.9(c). These results are consistent with our observation in Fig. 4.9(c). Here also increasing V_{reset} causes higher mean value of the HRS resistance, while the variations decrease.
MLC operation of ReRAM, which makes it suitable for high density memory application, is achieved by controlling the LRS and HRS resistance states. Modulating the LRS resistance by the set current compliance and the HRS resistance by the reset stop voltage is a common property of ReRAM devices ([60]) and is reported for many materials [124]. This multilevel resistance state distributions are captured by the simulations presented in Fig. 4.6, Fig. 4.9(c) and Fig. 4.10(c).

4.2 Random telegraph noise

RTN, which is known as a major reliability concern in electronic devices, is a kind of electronic noise, that consists of rapid step-like transitions between two or more discrete levels of a measured quantity like voltage, current or resistance. The current measured during the reading operation in a ReRAM cell shows RTN related fluctuations, which are more problematic during the HRS than LRS. This is because of the origin of the RTN fluctuations, which are generally related to the capture and emission of electrons by vacancies. This is consistent with the presented physical model that considers the trap assisted tunnelling (TAT) process as the main conduction mechanism in the HRS, where the individual $V_O^{\bullet\bullet}$s assist the electronic conduction by trapping and de-trapping of electrons.
Studying the RTN effect requires a large number of repeated simulations, for this reason the simulation results presented in this section are in the 2D structure. The KMC simulation result of the $I-V$ characteristics in a 2D ReRAM cell was already represented in Fig. 3.14 including the forming process and a couple of the switching loops.
Fig. 4.11 shows the current fluctuation during the HRS under an applied ramped voltage with very low sweep rate (-6 mV/sec). Two types of fluctuations in this figure can be observed: 1) larger jumps, which are due to the $V_O^{\bullet\bullet}$ fluctuations causing structural changes of the CF 2) smaller jumps placed between larger jumps, which are generated by electron trapping/de-trapping at $V_O^{\bullet\bullet}$ sites. The latter case is responsible for the $V_O^{\bullet\bullet}$ induced RTN variability. The structural change of the CF here means the addition of a new $V_O^{\bullet\bullet}$ into CF or the removal of a $V_O^{\bullet\bullet}$ from the CF. In order to focus only on the RTN effect on the ReRAM operation the structural changes which usually happen at larger applied voltages should be excluded from the current fluctuations [125].

The random change of the read current in the LRS and the HRS under a constant applied voltage is demonstrated in Fig. 4.12. The RTN pattern, known

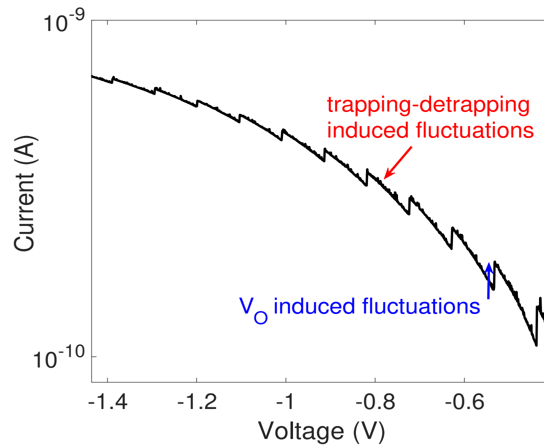


Figure 4.11: Current fluctuation under an applied negative very slowly ramped voltage during the HRS. The smaller jumps correspond to the repeatable trapping/de-trapping of an electron at a vacancy site causing the RTN. The larger jumps correspond to the vacancy diffusion in and out of the conductive filament leading to the structural changes [126].

as the abrupt current fluctuation between discrete levels, is obviously observed in the HRS. In this plot $\Delta I/I$ is defined as $(I_{max} - I_{min})/I_{min}$. As shown in this figure the $\Delta I/I$ value for the HRS is 7.5%, while this value is only 0.66% for the LRS. This higher RTN-related instability in the HRS comes from the activation and de-activation of the individual vacancies assisting the TAT process.

The power spectral density (PSD) profile, which one can define as the strength of the variations energy as a function of frequency (In other words, it shows at which frequencies variations are strong and at which frequencies variations are weak), is plotted versus frequency in the HRS and LRS in Fig. 4.13(a), was calculated by averaging over 100 simulations. It shows that in both resistance cases we have a behaviour close to $1/f$ noise, whose PSD is proportional to $1/f$, and the absolute value of the PSD for the HRS is more than 2 orders of magnitude greater than the LRS. Fig. 4.13(b) shows a couple of the realizations of the RTN process during the read time in the HRS and the average signal. It shows that the average value of the current does not change with time, ensuring that there is no violation of stationarity.

The RTN fluctuations can be affected by the forming and switching conditions. An important factor to influence the fluctuation of the read signal is the read time. If it is assumed that the PSD is proportional to $1/f^{(1+\alpha)}$, where α is the factor showing the deviation from $1/f$ noise, then the variance of the read current, $\sigma^2_{I_{read}}$, can be calculated using MacDonald's formula [127],

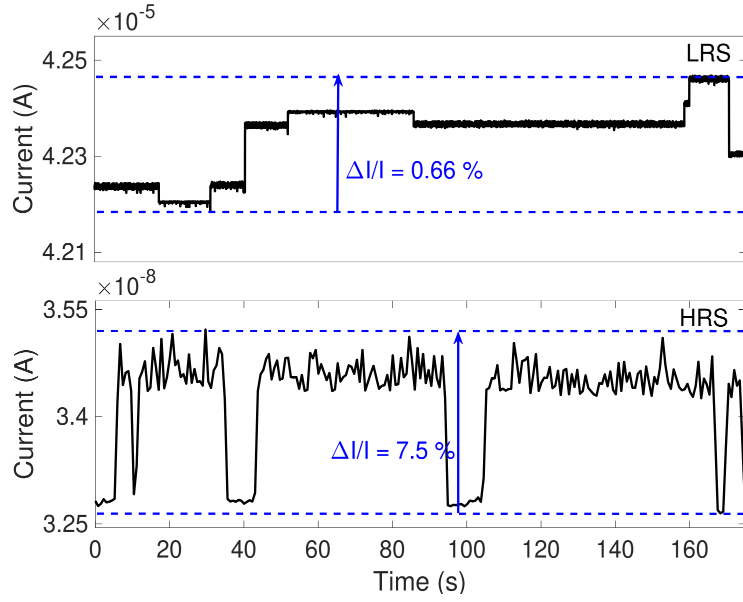
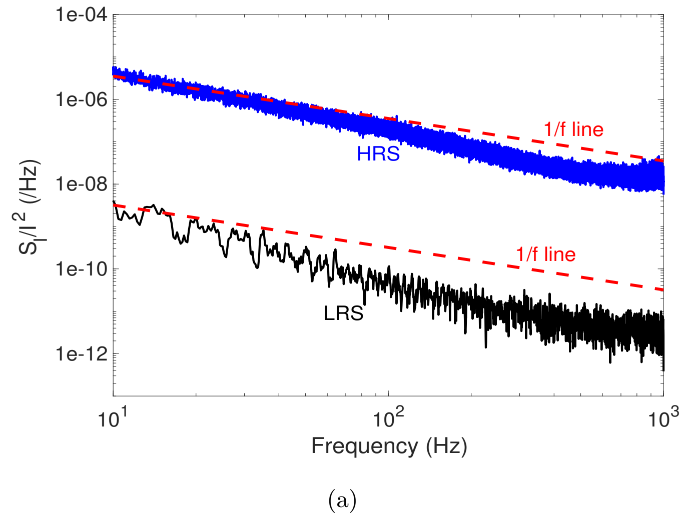


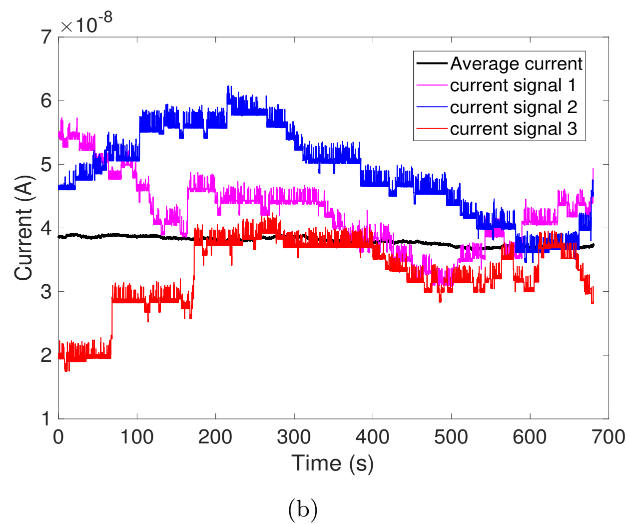
Figure 4.12: ΔI versus time under a constant applied read voltage voltage in the LRS and HRS [126].

$$\sigma^2_{I_{read}} = \frac{1}{\pi t_{read}} \int_{f_{min} t_{read}}^{\infty} \frac{k}{\left(\frac{\nu}{t_{read}}\right)^{1+\alpha}} \frac{1-\cos(\nu)}{\nu^2} d\nu, \tag{4.1}$$

where k is a proportionality constant, $\nu = f t_{read}$ is the normalized frequency and the lower cut-off frequency $f_{min} = t_{read}^{-1}$. Using the above formula, one can conclude that $\sigma_{I_{read}} \propto t_{read}^{\alpha/2}$. This shows that the mean square of the fluctuations increases with the duration of each realization of the reading process [128]. This effect is plotted in figure 4.14. It shows the distribution of the read current fluctuations for different read times in the HRS, where the ΔI increases with the read time.



(a)



(b)

Figure 4.13: (b) Average PSD profile of RTN versus frequency in the HRS and LRS (average of 100 simulations) (c) a couple of the current fluctuation signals during read time in the HRS and the average of 100 processes [126].

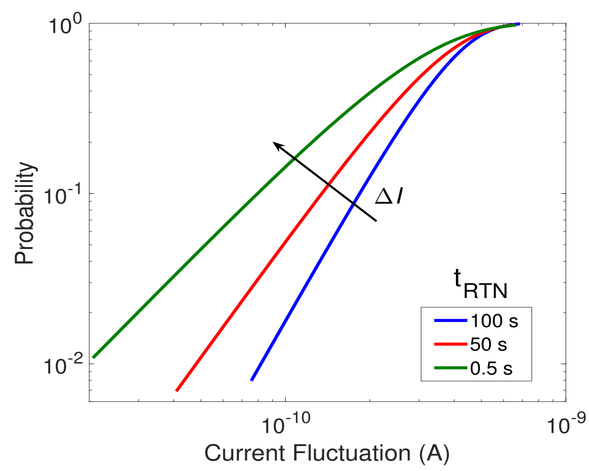


Figure 4.14: RTN-related current fluctuations, ΔI, increase with read time growth [126].

Chapter 5

Summary and Outlook

5.1 Summary

A comprehensive KMC simulation of valence change based ReRAM in 2D and 3D structures was presented in this work. A set of relevant physical and chemical processes involved in the KMC model such as $V_O^{\bullet\bullet}$ generation, recombination and diffusion were introduced and the corresponding rates were formulated. The driving forces modulating these rates were electric field, temperature and temperature gradient. This KMC simulator is based on an alternative physical model, which in contrast to all other previous KMC models limits the $V_O^{\bullet\bullet}$ generation and recombination to the active electrode/oxide interface. Using the proposed model the creation of a CF during the forming and set processes and its rupture during the reset process were succefully simulated. The charge transport mechanism during the switching process was investigated and a proper mechanism depending on the $V_O^{\bullet\bullet}$ concentration inside the CF was assigned in every step of the simulation. For a low $V_O^{\bullet\bullet}$ concentration a deterministic TAT solver was developed and for a high $V_O^{\bullet\bullet}$ concentration the drift-diffusion equation was solved to calculate the current. This KMC model made it possible to simulate the forming and a large number of the set and reset cycles with a reliable filament regeneration and annihilation. The reproduced $I-V$ characteristics of the forming step and successive switching cycles using this model show a similar trend as the experimental data of a HfO2-based device.
The variability of the switching process was investigated using the proposed KMC model. The cycle to cycle fluctuations of key parameter like the resistance in LRS and HRS and maximum reset current under different forming and switching conditions such as the current compliance or the applied set and reset voltages were studied thoroughly. The RTN-related fluctuations were also investigated as an intrinsic source of variability affecting the ReRAM behavior during the reading process.

5.2 Outlook

The developed KMC model to investigate the electroforming, SET and RESET processes in VCM based ReRAM can be used to study the performance of these devices and focus on the main challenges. Understanding the origin of these operational challenges and its mechanisms can lead to development of methods and solutions to improve the performance metrics of ReRAM such as: reliability, switching endurance and on/off resistance ratio. KMC as a fast yet accurate method can handle the simulation process of this study, which requires a large number of repeated simulations, in reasonable computational power and time.
In this work the intrinsic sources of variability in ReRAM devices were introduced and comprehensively studied but still the device-to-device variability, where the focus is on the variation of the local chemical structures, as an extrinsic type of ReRAM feature is missing. Some of these parameters, whose impact of their fluctuation on the memory cell performance can be studied, are oxide thickness, electrode/oxide interface roughness, grain boundaries and bond polarization factor that can affect the energy barrier of oxygen vacancy generation.
The potential scalability to the nanometer regime identifies ReRAM as an optimal device from the viewpoint of scaling. Studying the scaling trend of the switching parameters like the HRS and LRS resistances and RESET current can be identified as a hot topic for future study.
Finally, the presented physical model in this work can be extended and improved in the following areas: 1) Understanding the details of the oxygen exchange reaction along the oxide/electrode interface and the driving forces affecting that 2) existing electron conduction model to consider multiple conduction mechanisms that can coexist inside the insulating layer.

Bibliography

[1] R. Bez, E. Camerlenghi, A. Modelli, and A. Visconti, "Introduction to flash memory," *Proceedings of the IEEE*, vol. 91, no. 4, pp. 489–502, April 2003.

[2] A. Chen, "Emerging nonvolatile memory (NVM) technologies," in *2015 45th European Solid State Device Research Conference (ESSDERC)*, Sept 2015, pp. 109–113.

[3] D. Kahng and S. M. Sze, "A floating gate and its application to memory devices," *The Bell System Technical Journal*, vol. 46, no. 6, pp. 1288–1295, July 1967.

[4] Y. Développement, "Emerging non-volatile memory," *Lyon: Yole Développement*, Nov 2018.

[5] F. Yinug, "The rise of the flash memory market: Its impact on firm behavior and global semiconductor trade patterns," 2007.

[6] "Statista," https://www.statista.com/statistics/553556/worldwide-flash-memory-market-size/.

[7] P. Pavan, R. Bez, P. Olivo, E. Zanoni, and S. Member, "Flash memory cells-an overview," in *Proc. IEEE*, 1997, pp. 1248–1271.

[8] S. Lai, "Flash memories: where we were and where we are going," in *International Electron Devices Meeting 1998. Technical Digest (Cat. No.98CH36217)*, Dec 1998, pp. 971–973.

[9] F. Masuoka, M. Momodomi, Y. Iwata, and R. Shirota, "New ultra high density EPROM and flash EEPROM with NAND structure cell," in *1987 International Electron Devices Meeting*, Dec 1987, pp. 552–555.

[10] C. Dunn, C. Kaya, T. Lewis, T. Strauss, J. Schreck, P. Hefley, M. Middendorf, and T. San, "Flash EPROM disturb mechanisms," in *Proceedings of 1994 IEEE International Reliability Physics Symposium*, April 1994, pp. 299–308.

[11] P. Cappelletti, R. Bez, D. Cantarelli, and L. Fratin, "Failure mechanisms of flash cell in program/erase cycling," in *Proceedings of 1994 IEEE International Electron Devices Meeting*, Dec 1994, pp. 291–294.

[12] R. Moazzami and C. Hu, "Stress-induced current in thin silicon dioxide films," in *1992 International Technical Digest on Electron Devices Meeting*, Dec 1992, pp. 139–142.

[13] S. Mori, Y. Y. Araki, M. Sato, H. Meguro, H. Tsunoda, E. Kamiya, K. Yoshikawa, N. Arai, and E. Sakagami, "Thickness scaling limitation factors of ONO interpoly dielectric for nonvolatile memory devices," *IEEE Transactions on Electron Devices*, vol. 43, no. 1, pp. 47–53, Jan 1996.

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

[15] S. K. Lai, "Flash memories: Successes and challenges," *IBM Journal of Research and Development*, vol. 52, no. 4.5, pp. 529–535, July 2008.

[16] G. Burr *et al.*, "Phase change memory technology," *J. Vac. Sci. Technol. B*, vol. 28, pp. 223–262, Mar./Apr 2010.

[17] S. Raoux, F. Xiong, M. Wuttig, and E. Pop, "Phase change materials and phase change memory," *MRS Bulletin*, vol. 39, no. 8, p. 703–710, 2014.

[18] I. S. Kim, S. L. Cho, D. H. Im, E. H. Cho, D. H. Kim, G. H. Oh, D. H. Ahn, S. O. Park, S. W. Nam, J. T. Moon, and C. H. Chung, "High performance PRAM cell scalable to sub-20nm technology with below 4F2 cell size, extendable to DRAM applications," in *2010 Symposium on VLSI Technology*, June 2010, pp. 203–204.

[19] J. M. Slaughter, N. D. Rizzo, J. Janesky, R. Whig, F. B. Mancoff, D. Houssameddine, J. J. Sun, S. Aggarwal, K. Nagel, S. Deshpande, S. M. Alam, T. Andre, and P. LoPresti, "High density ST-MRAM technology (Invited)," in *2012 International Electron Devices Meeting*, Dec 2012, pp. 29.3.1–29.3.4.

[20] T. P. Ma and J.-P. Han, "Why is nonvolatile ferroelectric memory field-effect transistor still elusive?" *IEEE Electron Device Letters*, vol. 23, no. 7, pp. 386–388, July 2002.

[21] J. Muller, T. S. Boscke, S. Muller, E. Yurchuk, P. Polakowski, J. Paul, D. Martin, T. Schenk, K. Khullar, A. Kersch, W. Weinreich, S. Riedel, K. Seidel, A. Kumar, T. M. Arruda, S. V. Kalinin, T. Schlosser, R. Boschke,

R. van Bentum, U. Schroder, and T. Mikolajick, "Ferroelectric hafnium oxide: A CMOS-compatible and highly scalable approach to future ferroelectric memories," in *2013 IEEE International Electron Devices Meeting*, Dec 2013, pp. 10.8.1–10.8.4.

[22] T. W. Hickmott, "Low[U+2010]frequency negative resistance in thin anodic oxide films," *Journal of Applied Physics*, vol. 33, pp. 2669–2682, Sep 1962.

[23] J. Gibbons and W. Beadle, "Switching properties of thin NiO films," *Solid-State Electronics*, vol. 7, no. 11, pp. 785–790, 1964.

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

[25] I. G. Baek, D. C. Kim, M. J. Lee, H. . Kim, E. K. Yim, M. S. Lee, J. E. Lee, S. E. Ahn, S. Seo, J. H. Lee, J. C. Park, Y. K. Cha, S. O. Park, H. S. Kim, I. K. Yoo, U. Chung, J. T. Moon, and B. I. Ryu, "Multi-layer cross-point binary oxide resistive memory (OxRRAM) for post-NAND storage application," in *IEEE International Electron Devices Meeting, 2005. IEDM Technical Digest.*, Dec 2005, pp. 750–753.

[26] A. Sawa, "Resistive switching in transition metal oxides," *Materials Today*, vol. 11, no. 6, pp. 28–36, 2008.

[27] D. Ielmini and V. Milo, "Physics-based modeling approaches of resistive switching devices for memory and in-memory computing applications," *Journal of Computational Electronics*, vol. 16, no. 4, pp. 1121–1143, Dec 2017.

[28] K. Fujiwara, T. Nemoto, M. J. Rozenberg, Y. Nakamura, and H. Takagi, "Resistance switching and formation of a conductive bridge in metal/binary oxide/metal structure for memory devices," *Japanese Journal of Applied Physics*, vol. 47, no. 8R, p. 6266, 2008.

[29] S. Li, D. S. Shang, J. Li, J. L. Gang, and D. N. Zheng, "Resistive switching properties in oxygen-deficient $Pr_{0.7}Ca_{0.3}MnO_3$ junctions with active Al top electrodes," *Journal of Applied Physics*, vol. 105, pp. 033 710 – 033 710, 03 2009.

[30] Y.-F. Wang, Y.-C. Lin, I.-T. Wang, T.-P. Lin, and T.-H. Hou, "Characterization and modeling of nonfilamentary Ta/TaO_x/TiO_2/Ti analog synaptic device," *Scientific Reports*, vol. 5, p. 10150, 2015.

[31] M.-J. Lee, C. B. Lee, D.-S. Lee, S. R. Lee, M. Chang, J. H. Hur, Y.-B. Kim, C.-J. Kim, D. H. Seo, S. Seo, U.-I. Chung, I.-K. Yoo, and K. Kim, "A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta_2O_{5-x}/TaO_{2-x} bilayer structures," *Nature materials*, vol. 10 8, pp. 625–30, 2011.

[32] F. Nardi, S. Balatti, S. Larentis, and D. Ielmini, "Complementary switching in metal oxides: Toward diode-less crossbar RRAMs," in *2011 International Electron Devices Meeting*, Dec 2011, pp. 31.1.1–31.1.4.

[33] R. Waser, R. Dittmann, G. Staikov, and K. Szot, "Redox-based resistive switching memories-nanoionic mechanisms, prospects, and challenges," *Advanced Materials*, vol. 21, no. 25-26, pp. 2632–2663, 2009.

[34] 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*, vol. 94, no. 7, p. 072109, 2009.

[35] J.-B. Yun, S. Kim, S. Seo, M.-J. Lee, D.-C. Kim, S.-E. Ahn, Y. Park, J. Kim, and H. Shin, "Random and localized resistive switching observation in Pt/NiO/Pt," *physica status solidi (RRL) – Rapid Research Letters*, vol. 1, no. 6, pp. 280–282, 2007.

[36] J. Jang, F. Pan, K. Braam, and V. Subramanian, "Resistance switching characteristics of solid electrolyte chalcogenide Ag_2Se nanoparticles for flexible nonvolatile memory applications," *Advanced Materials*, vol. 24, no. 26, pp. 3573–3576, 2012.

[37] C. Schindler, M. Meier, R. Waser, and M. N. Kozicki, "Resistive switching in Ag-Ge-Se with extremely low write currents," in *2007 Non-Volatile Memory Technology Symposium*, Nov 2007, pp. 82–85.

[38] C. Xu, D. Niu, N. Muralimanohar, R. Balasubramonian, T. Zhang, S. Yu, and Y. Xie, "Overcoming the challenges of crossbar resistive memory architectures," in *2015 IEEE 21st International Symposium on High Performance Computer Architecture (HPCA)*, Feb 2015, pp. 476–488.

[39] R. Waser, *Nanoelectronics and Information Technology*, 3rd ed. Wiley-VCH, 2012.

[40] H. Y. Lee, P. S. Chen, T. Y. Wu, Y. S. Chen, C. C. Wang, P. J. Tzeng, C. H. Lin, F. Chen, C. H. Lien, and M. . Tsai, "Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO_2 based RRAM," in *2008 IEEE International Electron Devices Meeting*, Dec 2008, pp. 1–4.

[41] K. Kinoshita, K. Tsunoda, Y. Sato, H. Noshiro, S. Yagaki, M. Aoki, and Y. Sugiyama, "Reduction in the reset current in a resistive random access memory consisting of NiO_x brought about by reducing a parasitic capacitance," *Applied Physics Letters*, vol. 93, no. 3, p. 033506, 2008.

[42] P. Gu, Y. Chen, H. Lee, P. Chen, W. Liu, W. Chen, Y. Hsu, F. Chen, and M. Tsai, "Scalability with silicon nitride encapsulation layer for Ti/HfO_x pillar RRAM," in *Proceedings of 2010 International Symposium on VLSI Technology, System and Application*, April 2010, pp. 146–147.

[43] S. Yu, X. Guan, and H. . P. Wong, "On the switching parameter variation of metal oxide RRAM-part II: model corroboration and device design strategy," *IEEE Transactions on Electron Devices*, vol. 59, no. 4, pp. 1183–1188, April 2012.

[44] M. Bocquet, D. Deleruyelle, H. Aziza, C. Muller, J. Portal, T. Cabout, and E. Jalaguier, "Robust compact model for bipolar oxide-based resistive switching memories," *IEEE Transactions on Electron Devices*, vol. 61, no. 3, pp. 674–681, March 2014.

[45] H. Y. Lee, Y. S. Chen, P. S. Chen, P. Y. Gu, Y. Y. Hsu, S. M. Wang, W. H. Liu, C. H. Tsai, S. S. Sheu, P. C. Chiang, W. P. Lin, C. H. Lin, W. S. Chen, F. T. Chen, C. H. Lien, and M. . Tsai, "Evidence and solution of over-RESET problem for HfO_X based resistive memory with sub-ns switching speed and high endurance," in *2010 International Electron Devices Meeting*, Dec 2010, pp. 19.7.1–19.7.4.

[46] S. Yu, B. Gao, H. Dai, B. Sun, L. Liu, X. Liu, R. Han, J. Kang, and B. Yu, "Improved uniformity of resistive switching behaviors in HfO_2 thin films with embedded Al layers," *Electrochemical and Solid-State Letters*, vol. 13, no. 2, pp. H36–H38, 2010.

[47] W.-Y. Chang, K.-J. Cheng, J.-M. Tsai, H.-J. Chen, F. Chen, M.-J. Tsai, and T.-B. Wu, "Improvement of resistive switching characteristics in TiO_2 thin films with embedded Pt nanocrystals," *Applied Physics Letters*, vol. 95, no. 4, p. 042104, 2009.

[48] Y. S. Chen, H. Y. Lee, P. S. Chen, P. Y. Gu, C. W. Chen, W. P. Lin, W. H. Liu, Y. Y. Hsu, S. S. Sheu, P. C. Chiang, W. S. Chen, F. T. Chen,

C. H. Lien, and M. . Tsai, "Highly scalable hafnium oxide memory with improvements of resistive distribution and read disturb immunity," in *2009 IEEE International Electron Devices Meeting (IEDM)*, Dec 2009, pp. 1–4.

[49] D. Ielmini, "Resistive switching memories based on metal oxides: mechanisms, reliability and scaling," *Semiconductor Science and Technology*, vol. 31, no. 6, p. 063002, may 2016.

[50] P. Huang, B. Chen, Y. J. Wang, F. F. Zhang, L. Shen, R. Liu, L. Zeng, G. Du, X. Zhang, B. Gao, J. F. Kang, X. Y. Liu, X. P. Wang, B. B. Weng, Y. Z. Tang, G. . Lo, and D. . Kwong, "Analytic model of endurance degradation and its practical applications for operation scheme optimization in metal oxide based RRAM," in *2013 IEEE International Electron Devices Meeting*, Dec 2013, pp. 22.5.1–22.5.4.

[51] Y. Y. Chen, B. Govoreanu, L. Goux, R. Degraeve, A. Fantini, G. S. Kar, D. J. Wouters, G. Groeseneken, J. A. Kittl, M. Jurczak, and L. Altimime, "Balancing set/reset pulse for $> 10^{10}$ endurance in HfO_2/Hf 1T1R bipolar RRAM," *IEEE Transactions on Electron Devices*, vol. 59, no. 12, pp. 3243–3249, Dec 2012.

[52] S. Balatti, S. Ambrogio, Z. Wang, S. Sills, A. Calderoni, N. Ramaswamy, and D. Ielmini, "Voltage-controlled cycling endurance of HfO_x-based resistive switching memory," *IEEE Transactions on Electron Devices*, vol. 62, no. 10, pp. 3365–3372, Oct 2015.

[53] B. Traore, P. Blaise, E. Vianello, H. Grampeix, S. Jeannot, L. Perniola, B. D. Salvo, and Y. Nishi, "On the origin of low-resistance state retention failure in HfO_2-based RRAM and impact of doping/alloying," *IEEE Transactions on Electron Devices*, vol. 62, no. 12, pp. 4029–4036, Dec 2015.

[54] J. Park, W. Lee, M. Choe, S. Jung, M. Son, S. Kim, S. Park, J. Shin, D. Lee, M. Siddik, J. Woo, G. Choi, E. Cha, T. Lee, and H. Hwang, "Quantized conductive filament formed by limited Cu source in sub-5nm era," in *2011 International Electron Devices Meeting*, Dec 2011, pp. 3.7.1–3.7.4.

[55] M.-J. Lee, S. Han, S. H. Jeon, B. H. Park, B. S. Kang, S.-E. Ahn, K. H. Kim, C. B. Lee, C. J. Kim, I.-K. Yoo, D. H. Seo, X.-S. Li, J.-B. Park, J.-H. Lee, and Y. Park, "Electrical manipulation of nanofilaments in transition-metal oxides for resistance-based memory," *Nano Letters*, vol. 9, no. 4, pp. 1476–1481, 2009.

[56] B. Govoreanu, G. S. Kar, Y. Chen, V. Paraschiv, S. Kubicek, A. Fantini, I. P. Radu, L. Goux, S. Clima, R. Degraeve, N. Jossart, O. Richard, T. Vandeweyer, K. Seo, P. Hendrickx, G. Pourtois, H. Bender, L. Altimime,

D. J. Wouters, J. A. Kittl, and M. Jurczak, "10×10 nm^2 Hf/HfO_x crossbar resistive RAM with excellent performance, reliability and low-energy operation," in *2011 International Electron Devices Meeting*, Dec 2011, pp. 31.6.1–31.6.4.

[57] Y. S. Chen, H. Y. Lee, P. S. Chen, P. Y. Gu, C. W. Chen, W. P. Lin, W. H. Liu, Y. Y. Hsu, S. S. Sheu, P. C. Chiang, W. S. Chen, F. T. Chen, C. H. Lien, and M. . Tsai, "Highly scalable hafnium oxide memory with improvements of resistive distribution and read disturb immunity," in *2009 IEEE International Electron Devices Meeting (IEDM)*, Dec 2009, pp. 1–4.

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

[59] S. Seo, M. J. Lee, D. H. Seo, E. J. Jeoung, D.-S. Suh, Y. S. Joung, I. K. Yoo, I. R. Hwang, S. H. Kim, I. S. Byun, J.-S. Kim, J. S. Choi, and B. H. Park, "Reproducible resistance switching in polycrystalline NiO films," *Applied Physics Letters*, vol. 85, no. 23, pp. 5655–5657, 2004.

[60] S. Yu, Y. Wu, and H.-S. P. Wong, "Investigating the switching dynamics and multilevel capability of bipolar metal oxide resistive switching memory," *Applied Physics Letters*, vol. 98, no. 10, p. 103514, 2011.

[61] A. Prakash, J. Park, J. Song, J. Woo, E. Cha, and H. Hwang, "Demonstration of low power 3-bit multilevel cell characteristics in a TaO_x-based rram by stack engineering," *IEEE Electron Device Letters*, vol. 36, no. 1, pp. 32–34, Jan 2015.

[62] S. Park, M. K. Yang, H. Ju, D. Seong, J. M. Lee, E. Kim, S. Jung, L. Zhang, Y. C. Shin, I. Baek, J. Choi, H. Kang, and C. Chung, "A non-linear ReRAM cell with sub-1µA ultralow operating current for high density vertical resistive memory (VRRAM)," in *2012 International Electron Devices Meeting*, Dec 2012, pp. 20.8.1–20.8.4.

[63] I. G. Baek, C. J. Park, H. Ju, D. J. Seong, H. S. Ahn, J. H. Kim, M. K. Yang, S. H. Song, E. M. Kim, S. O. Park, C. H. Park, C. W. Song, G. T. Jeong, S. Choi, H. K. Kang, and C. Chung, "Realization of vertical resistive memory (VRRAM) using cost effective 3D process," in *2011 International Electron Devices Meeting*, Dec 2011, pp. 31.8.1–31.8.4.

[64] H. S. Yoon, I.-G. Baek, J. Zhao, H. Sim, M. Y. Park, H. Lee, G.-H. Oh, J. C. Shin, I.-S. Yeo, and U.-I. Chung, "Vertical cross-point resistance change memory for ultra-high density non-volatile memory applications," in *2009 Symposium on VLSI Technology*, June 2006, pp. 26–27.

[65] H. S. P. Wong, H. Y. Lee, S. Yu, Y. S. Chen, Y. Wu, P. S. Chen, B. Lee, F. T. Chen, and M. J. Tsai, "Metal Oxide RRAM," *Proceedings of the IEEE*, vol. 100, no. 6, pp. 1951–1970, 2012.

[66] L. Goux, P. Czarnecki, Y. Y. Chen, L. Pantisano, X. P. Wang, R. Degraeve, B. Govoreanu, M. Jurczak, D. J. Wouters, and L. Altimime, "Evidences of oxygen-mediated resistive-switching mechanism in TiN/HfO_2/Pt cells," *Applied Physics Letters*, vol. 97, no. 24, p. 243509, 2010.

[67] G. Bersuker, D. C. Gilmer, D. Veksler, P. Kirsch, L. Vandelli, A. Padovani, L. Larcher, K. McKenna, A. Shluger, V. Iglesias, M. Porti, and M. Nafría, "Metal oxide resistive memory switching mechanism based on conductive filament properties," *Journal of Applied Physics*, vol. 110, no. 12, p. 124518, 2011.

[68] D. Z. Gao, A.-M. El-Sayed, and A. L. Shluger, "A mechanism for Frenkel defect creation in amorphous SiO_2 facilitated by electron injection," *Nanotechnology*, vol. 27, no. 50, p. 505207, 2016.

[69] X. Guan, S. Yu, and H.-S. Wong, "On the switching parameter variation of metal-oxide RRAM;part I: Physical modeling and simulation methodology," *Electron Devices, IEEE Transactions on*, vol. 59, no. 4, pp. 1172–1182, 2012.

[70] A. Padovani, L. Larcher, O. Pirrotta, L. Vandelli, and G. Bersuker, "Microscopic modeling of HfO_x RRAM operations: from forming to switching," *IEEE Transactions on Electron Devices*, vol. 62, no. 6, pp. 1998–2006, June 2015.

[71] A. Padovani, D. Z. Gao, A. L. Shluger, and L. Larcher, "A microscopic mechanism of dielectric breakdown in SiO_2 films: An insight from multiscale modeling," *Journal of Applied Physics*, vol. 121, no. 15, p. 155101, 2017.

[72] B. Butcher, G. Bersuker, D. C. Gilmer, L. Larcher, A. Padovani, L. Vandelli, R. Geer, and P. D. Kirsch, "Connecting the physical and electrical properties of hafnia-based RRAM," in *2013 IEEE International Electron Devices Meeting*, 2013, pp. 22.2.1–22.2.4.

[73] Y. Zhao, P. Huang, Z. Chen, C. Liu, H. Li, B. Chen, W. Ma, F. Zhang, B. Gao, X. Liu, and J. Kang, "Modeling and optimization of bilayered TaO_x RRAM based on defect evolution and phase transition effects," *IEEE Transactions on Electron Devices*, vol. 63, no. 4, pp. 1524–1532, April 2016.

[74] F. M. Puglisi, L. Larcher, A. Padovani, and P. Pavan, "Bipolar resistive RAM based on HfO_2: Physics, compact modeling, and variability control," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 6, pp. 171–184, June 2016.

[75] Y. Guo and J. Robertson, "Materials selection for oxide-based resistive random access memories," *Appl. Phys. Lett.*, vol. 105, no. 22, pp. 223 516/1–5, 2014.

[76] A. O'Hara, G. Bersuker, and A. A. Demkov, "Assessing hafnium on hafnia as an oxygen getter," *Journal of Applied Physics*, vol. 115, no. 18, p. 183703, 2014.

[77] J. Robertson and S. J. Clark, "Limits to doping in oxides," *Phys. Rev. B*, vol. 83, p. 075205, Feb 2011.

[78] S. Clima, Y. Y. Chen, C. Y. Chen, L. Goux, B. Govoreanu, R. Degraeve, A. Fantini, M. Jurczak, and G. Pourtois, "First-principles thermodynamics and defect kinetics guidelines for engineering a tailored RRAM device," *Journal of Applied Physics*, vol. 119, p. 225107, 06 2016.

[79] S. R. Bradley, A. Shluger, and G. Bersuker, "Electron-injection-assisted generation of oxygen vacancies in monoclinic HfO_2," *Phys. Rev. Applied*, vol. 4, p. 064008, 12 2015.

[80] M. Sowinska, T. Bertaud, D. Walczyk, S. Thiess, M. A. Schubert, M. Lukosius, W. Drube, C. Walczyk, and T. Schroeder, "Hard x-ray photoelectron spectroscopy study of the electroforming in Ti/HfO_2-based resistive switching structures," *Appl. Phys. Lett.*, vol. 100, no. 23, p. 233509, 2012.

[81] W. Kim, S. Menzel, D. J. Wouters, Y. Guo, J. Robertson, B. Roesgen, R. Waser, and V. Rana, "Impact of oxygen exchange reaction at the ohmic interface in Ta_2O_5-based ReRAM devices," *Nanoscale*, vol. 8, pp. 17 774–17 781, 2016.

[82] D. Cooper, C. Baeumer, N. Bernier, A. Marchewka, C. L. Torre, R. E. Dunin-Borkowski, S. Menzel, R. Waser, and R. Dittmann, "Anomalous resistance hysteresis in oxide ReRAM: Oxygen evolution and reincorporation revealed by in situ TEM," *Adv. Mater.*, vol. 29, p. 1700212, 2017.

[83] S. Clima, B. Govoreanu, M. Jurczak, and G. Pourtois, "HfO_x as RRAM material-first principles insights on the working principles," *Microelectronic Engineering*, vol. 120, pp. 13–18, 2014.

[84] J. J. Yang, F. Miao, M. D. Pickett, D. A. A. Ohlberg, D. R. Stewart, C. N. Lau, and R. S. Williams, "The mechanism of electroforming of metal oxide memristive switches," *Nanotechnology*, vol. 20, p. 215201, 2009.

[85] K. Szot, W. Speier, G. Bihlmayer, and R. Waser, "Switching the electrical resistance of individual dislocations in single-crystalline $SrTiO_3$," vol. 5, pp. 312–20, 05 2006.

[86] A. Mehonic, M. Buckwell, L. Montesi, M. S. Munde, D. Gao, S. Hudziak, R. J. Chater, S. Fearn, D. McPhail, M. Bosman, A. L. Shluger, and A. J. Kenyon, "Nanoscale transformations in metastable, amorphous, silicon-rich silica," *Advanced Materials*, vol. 28, no. 34, pp. 7549–7549, 2016.

[87] A. Marchewka, B. Roesgen, K. Skaja, H. Du, C.-L. Jia, J. Mayer, V. Rana, R. Waser, and S. Menzel, "Nanoionic resistive switching memories: On the physical nature of the dynamic reset process," *Advanced Electronic Materials*, vol. 2, no. 1, pp. 1 500 233/1–13, 2016.

[88] A. Marchewka, R. Waser, and S. Menzel, "Physical simulation of dynamic resistive switching in metal oxides using a Schottky contact barrier model," in *2015 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2015, pp. 297–300.

[89] ——, "A 2D axisymmetric dynamic drift-diffusion model for numerical simulation of resistive switching phenomena in metal oxides," in *2016 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2016, pp. 145–148.

[90] S. Larentis, F. Nardi, S. Balatti, D. C. Gilmer, and D. Ielmini, "Resistive switching by voltage-driven ion migration in bipolar RRAM-part II: modeling," *IEEE Transactions on Electron Devices*, vol. 59, no. 9, pp. 2468–2475, Sept 2012.

[91] A. Marchewka, R. Waser, and S. Menzel, "Physical modeling of the electroforming process in resistive-switching devices (talk)," in *2017 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2017.

[92] L. Larcher, A. Padovani, O. Pirrotta, L. Vandelli, and G. Bersuker, "Microscopic understanding and modeling of HfO_2 RRAM device physics," in *2012 International Electron Devices Meeting*, 2012, pp. 20.1.1–20.1.4.

[93] A. Makarov, V. Sverdlov, and S. Selberherr, "Stochastic modeling of bipolar resistive switching in metal-oxide based memory by Monte Carlo technique," *Journal of Computational Electronics*, vol. 9, no. 3, pp. 146–152, 2010.

[94] T. Sakamoto, K. Lister, N. Banno, T. Hasegawa, K. Terabe, and M. Aono, "Electronic transport in Ta_2O_5 resistive switch," *Applied Physics Letters*, vol. 91, no. 9, p. 092110, 2007.

[95] R. Oligschlaeger, R. Waser, R. Meyer, S. Karthäuser, and R. Dittmann, "Resistive switching and data reliability of epitaxial (Ba,Sr)TiO_3 thin films," *Applied Physics Letters*, vol. 88, no. 4, p. 042901, 2006.

[96] K. M. Kim, B. J. Choi, Y. C. Shin, S. Choi, and C. S. Hwang, "Anode-interface localized filamentary mechanism in resistive switching of TiO_2 thin films," *Applied Physics Letters*, vol. 91, no. 1, p. 012907, 2007.

[97] K. M. Kim, D. S. Jeong, and C. S. Hwang, "Nanofilamentary resistive switching in binary oxide system; a review on the present status and outlook," *Nanotechnology*, vol. 22, no. 25, p. 254002, may 2011.

[98] C. B. Lee, B. S. Kang, A. Benayad, M. J. Lee, S.-E. Ahn, K. H. Kim, G. Stefanovich, Y. Park, and I. K. Yoo, "Effects of metal electrodes on the resistive memory switching property of NiO thin films," *Applied Physics Letters*, vol. 93, no. 4, p. 042115, 2008.

[99] W. Stehling, E. Abbaspour, C. Jungemann, and S. Menzel, "Kinetic Monte Carlo modeling of the charge transport in a HfO_2-based ReRAM with a rough anode," in *2017 17th Non-Volatile Memory Technology Symposium (NVMTS)*, Aug 2017, pp. 1–4.

[100] O. Pirrotta, A. Padovani, L. Larcher, L. Zhao, B. Magyari-Koepe, and Y. Nishi, "Multi-scale modeling of oxygen vacancies assisted charge transport in sub-stoichiometric TiO_x for RRAM application," in *2014 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Sep. 2014, pp. 37–40.

[101] I. Lundström and C. Svensson, "Tunneling to traps in insulators," *Journal of Applied Physics*, vol. 43, no. 12, pp. 5045–5047, 1972.

[102] D. Bohm, *Quantum theory.* Englewood Cliffs, N.J. : Prentice-Hall, 1951, includes index.

[103] L. Lundkvist, I. Lundström, and C. Svensson, "Discharge of MNOS structures," *Solid-State Electronics*, vol. 16, no. 7, pp. 811–823, 1973.

[104] L. Larcher, "Statistical simulation of leakage currents in MOS and flash memory devices with a new multiphonon trap-assisted tunneling model," *IEEE Transactions on Electron Devices*, vol. 50, no. 5, pp. 1246–1253, May 2003.

[105] D. Ielmini and Y. Zhang, "Analytical model for subthreshold conduction and threshold switching in chalcogenide-based memory devices," *Journal of Applied Physics*, vol. 102, no. 5, p. 054517, 2007.

[106] A. U. Modak and M. T. Lusk, "Kinetic Monte Carlo simulation of a solid-oxide fuel cell: I. Open-circuit voltage and double layer structure," *Solid State Ionics*, vol. 176, no. 29, pp. 2181 – 2191, 2005.

[107] B. Gao, B. Sun, H. Zhang, L. Liu, X. Liu, R. Han, J. Kang, and B. Yu, "Unified physical model of bipolar oxide-based resistive switching memory," *IEEE Electron Device Letters*, vol. 30, no. 12, pp. 1326–1328, Dec 2009.

[108] E. Abbaspour, S. Menzel, A. Hardtdegen, S. Hoffmann-Eifert, and C. Jungemann, "KMC simulation of the electroforming, set and reset processes in redox-based resistive switching devices," *IEEE Transactions on Nanotechnology*, vol. 17, no. 6, pp. 1181–1188, Nov 2018.

[109] D. L. Scharfetter and H. K. Gummel, "Large-signal analysis of a silicon read diode oscillator," *IEEE Transactions on Electron Devices*, vol. 16, no. 1, pp. 64–77, Jan 1969.

[110] A. F. Franz, G. A. Franz, S. Selberherr, C. Ringhofer, and P. Markowich, "Finite boxes–A generalization of the finite-difference method suitable for semiconductor device simulation," *IEEE Transactions on Electron Devices*, vol. 30, no. 9, pp. 1070–1082, Sep. 1983.

[111] S. Lee and K. P. Moran, "Constriction/spreading resistance model for electronics packaging," in *ASME/JSME Thermal Engineering Conference*, vol. 4, 1996.

[112] A. F. Voter, "Introduction to the kinetic Monte Marlo method," *Radiation Effects in Solids*, pp. 1–23, 2007.

[113] R. Nieminen and A. Jansen, "Monte Carlo simulations of surface reactions," *Applied Catalysis A: General*, vol. 160, no. 1, pp. 99 – 123, 1997, kinetic Methods in Heterogeneous Catalysis.

[114] C. C. Battaile and D. J. Srolovitz, "Kinetic Monte Carlo simulation of chemical vapor deposition," *Annual Review of Materials Research*, vol. 32, no. 1, pp. 297–319, 2002.

[115] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of Computational Physics*, vol. 22, no. 4, pp. 403 – 434, 1976.

[116] B. Butcher, G. Bersuker, L. Vandelli, A. Padovani, L. Larcher, A. Kalantarian, R. Geer, and D. C. Gilmer, "Modeling the effects of different forming conditions on RRAM conductive filament stability," in *2013 5th IEEE International Memory Workshop*, May 2013, pp. 52–55.

[117] D. B. Strukov, F. Alibart, and R. Stanley Williams, "Thermophoresis/diffusion as a plausible mechanism for unipolar resistive switching in metal–oxide–metal memristors," *Applied Physics A*, vol. 107, no. 3, pp. 509–518, Jun 2012.

[118] E. Abbaspour, S. Menzel, and C. Jungemann, "The role of the interface reactions in the electroforming of redox-based resistive switching devices using KMC simulations," in *2015 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Sep. 2015, pp. 293–296.

[119] ——, "KMC simulation of the electroforming, set and reset processes in redox-based resistive switching devices," in *2016 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Sep. 2016, pp. 141–144.

[120] D. Ito, Y. Hamada, S. Otsuka, T. Shimizu, and S. Shingubara, "Oxide thickness dependence of resistive switching characteristics for Ni/HfO_x/Pt resistive random access memory device," *Japanese Journal of Applied Physics*, vol. 54, no. 6S1, p. 06FH11, 2015.

[121] E. Yalon, I. Karpov, V. Karpov, I. Riess, D. Kalaev, and D. Ritter, "Detection of the insulating gap and conductive filament growth direction in resistive memories," *Nanoscale*, vol. 7, pp. 15 434–15 441, 2015.

[122] J. Song, D. Lee, J. Woo, Y. Koo, E. Cha, S. Lee, J. Park, K. Moon, S. H. Misha, A. Prakash, and H. Hwang, "Effects of reset current overshoot and resistance state on reliability of RRAM," *IEEE Electron Device Letters*, vol. 35, no. 6, pp. 636–638, June 2014.

[123] D. Ielmini, "Modeling the universal set/reset characteristics of bipolar RRAM by field- and temperature-driven filament growth," *IEEE Transactions on Electron Devices*, vol. 58, no. 12, pp. 4309–4317, Dec 2011.

[124] L. Goux, Y.-Y. Chen, L. Pantisano, X.-P. Wang, G. Groeseneken, M. Jurczak, and D. J. Wouters, "On the gradual unipolar and bipolar resistive switching of TiN/HfO_2/Pt memory systems," *Electrochemical and Solid-State Letters*, vol. 13, no. 6, pp. G54–G56, 2010.

[125] D. Veksler, G. Bersuker, B. Chakrabarti, E. Vogel, S. Deora, K. Matthews, D. C. Gilmer, H. F. Li, S. Gausepohl, and P. D. Kirsch, "Methodology for the statistical evaluation of the effect of random telegraph noise (RTN) on RRAM characteristics," in *2012 International Electron Devices Meeting*, Dec 2012, pp. 9.6.1–9.6.4.

[126] E. Abbaspour, S. Menzel, and C. Jungemann, "Random telegraph noise analysis in redox-based resistive switching devices using KMC simulations," in *2017 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Sep. 2017, pp. 313–316.

[127] D. K. C. MacDonald, *Noise and fluctuations: an introduction.* New York, Wiley, 1962.

[128] D. Veksler, G. Bersuker, B. Chakrabarti, E. Vogel, S. Deora, K. Matthews, D. C. Gilmer, H. . Li, S. Gausepohl, and P. D. Kirsch, "Methodology for the statistical evaluation of the effect of random telegraph noise (RTN) on RRAM characteristics," in *2012 International Electron Devices Meeting*, Dec 2012, pp. 9.6.1–9.6.4.