Nanoionic memristive phenomena in metal oxides: the valence change mechanism

Regina Dittmann, Stephan Menzel & Rainer Waser

To cite this article: Regina Dittmann, Stephan Menzel & Rainer Waser (2022): Nanoionic memristive phenomena in metal oxides: the valence change mechanism, Advances in Physics, DOI: 10.1080/00018732.2022.2084006

To link to this article: https://doi.org/10.1080/00018732.2022.2084006
Nanoionic memristive phenomena in metal oxides: the valence change mechanism

Regina Dittmann a,b,∗, Stephan Menzel a,c and Rainer Waser a,b,d

a Peter Grünberg Institute PGI-7, Electronic Materials, Forschungszentrum Jülich, Jülich, Germany
b Jülich-Aachen Research Alliance, Section Future Information Technology (JARA-FIT), Aachen, Germany
c IBM Research – Zurich, Rüschlikon, Switzerland
d Institute für Werkstoffe der Elektrotechnik IWE-2, RWTH Aachen University, Aachen, Germany

This review addresses resistive switching devices operating according to the bipolar valence change mechanism (VCM), which has become a major trend in electronic materials and devices over the last decade due to its high potential for non-volatile memories and future neuromorphic computing. We will provide detailed insights into the status of understanding of these devices as a fundament for their use in the different fields of application. The review covers the microscopic physics of memristive states and the switching kinetics of VCM devices. It is shown that the switching of all variants of VCM cells relies on the movement of mobile donor ions, which are typically oxygen vacancies or cation interstitials. VCM cells consist of three parts: an electronically active electrode (AE), often a metal with a high work function, in front of which the switching occurs, a mixed ionic-electronic conducting (MIEC) layer consisting of a nanometer-scale metal oxide or a stack of different metal oxides, and an ohmic counter electrode (OE). After an introduction to definitions and classification, the fundamentals of solid-state physics and chemistry associated with VCM cells are described, including redox processes and the role of electrodes. The microscopic changes induced by electroforming, a process often required prior to resistive switching, are described in terms of electronic initialization and subsequent changes in chemistry, structure, and conductivity. The switching process is discussed in terms of switching polarity, geometry of the switching region, and spectroscopic detection of the valence changes. Emphasis is placed on the extreme nonlinearity of switching kinetics described by physics-based multiscale modeling, ranging from ab initio methods to kinetic Monte Carlo and finite element models to compact models that can be used in circuit simulators. The review concludes with a treatment of the highly relevant reliability issues and a description of the failure mechanisms, including mutual trade-offs.

Keywords: redox-based resistive switching; valence change mechanism; nanoionics; mobile donors; electroforming; filamentary switching; area-dependent switching; counter-eighth/eightwise switching; nonlinear kinetics; multiscale simulations; memristive devices; memristors

Contents

1 Introduction 3
2 Definitions and general classification 4
  2.1 Physical storage principles 4
  2.2 Nanoionic redox-based memristive elements: modes of operation 7
  2.3 Classes of nanoionic redox-based resistive switching phenomena 10
    2.3.1 Survey of generic processes in redox-based memristive cells 10
    2.3.2 Types of redox-based memristive cells 12
    2.3.3 General features of valence change memories 12
    2.3.4 Requirements of memory cells and beyond 14

∗Corresponding authors. Email: r.dittmann@fz-juelich.de (Regina Dittmann); st.menzel@fz-juelich.de (Stephan Menzel)

© 2022 Informa UK Limited, trading as Taylor & Francis Group
2.4 Historical aspects 17
2.5 From memristive effects to non-volatile resistive switching 19

3 Relevant properties of metal oxides and resistive switching cells 20
  3.1 Phase diagrams – full-valent and lower-valent phases 21
  3.2 Electronic structure and band diagrams 21
  3.3 Lattice disorder
    3.3.1 Schottky- and Frenkel disorder 23
    3.3.2 Acceptors, donors, and electronic compensation schemes 27
  3.4 Extended defects – dislocations and grain boundaries 28
  3.5 Amorphous and nanocrystalline oxides 30
  3.6 Ionic transport by drift and diffusion 31
  3.7 Redox processes 34
  3.8 Electronic transport – mobilities and correlation effects 35
  3.9 Types of barriers and their modulation – the fundament of resistive switching 37
  3.10 Electrode contacts
    3.10.1 Formation of Schottky barriers and ohmic contacts 38
    3.10.2 Current transport across the barrier 39

4 Electroforming processes 40
  4.1 Typical procedures for electroforming 41
  4.2 Microscopic changes induced by electroforming
    4.2.1 Conductivity changes 43
    4.2.2 Chemical changes 45
    4.2.3 Structural changes 47
  4.3 Microscopic mechanism during forming
    4.3.1 Electronic effects initializing the forming 51
    4.3.2 Redox reactions at the electrode interfaces 52
    4.3.3 Ionic concentration polarization processes 55
  4.4 Filament growth direction 57
  4.5 Forming-free systems 60

5 Valence change switching mechanism 62
  5.1 Survey of the switching systems 62
  5.2 Valence change in the switching region – the essential metal insulator transition 62
  5.3 Memristive states 68
    5.3.1 Bulk switching models 70
    5.3.2 Interface-dominated switching models 72
  5.4 Filamentary switching
    5.4.1 Bulk switching models 81
    5.4.2 Interface-dominated switching models 84
  5.5 Area-dependent switching
    5.5.1 Prerequisites for area-dependent switching 92
    5.5.2 Examples of C8W area-dependent resistive switching systems 94
  5.6 Eightwise switching 102
  5.7 Switching kinetics
    5.7.1 Switching voltages and times 110
    5.7.2 Origin of the ultra-nonlinear kinetics 114
    5.7.3 General considerations 116
    5.7.4 Consequences of the ultra-nonlinear switching kinetics and influence of a series resistance 121
    5.7.5 SET kinetics 125
1. Introduction

Resistive switching or memristive phenomena denotes reversible changes of the electrical resistance of 2-terminal elements which occur upon electrical stimuli in a non-volatile fashion [1]. Reversibility means that repeated applications of suitable stimuli adjust the resistance value repeatedly between two or more levels. Non-volatility reflects the fact that the resistance change remains for a (long) retention time after the stimulus has been released. Phenomenologically, the stimulus affects an internal state variable of the element which controls the resistance. From an engineering point of view, the internal states and the related properties of the 2–terminal elements can be described by a theory of memristive devices [2]. The required switching speed and the retention times depend on the area of application and will be discussed later.

The fundamental physical principles of memristive elements and, hence, the nature of the internal state variable can be manifold such as, for example, nanomechanical, magnetoresistive, or phase-change related. This article will exclusively cover memristive elements based on nanoionic redox phenomena, i.e. ionic motion over nanoscale dimensions in 2-terminal elements which leads to local redox phenomena and, in turn, affect the resistance of the elements. An emphasis will be placed on those redox phenomena which occur as a valence change of ions in metal oxides.

Throughout the article, we will use the term Redox-based Resistive Switching Random Access Memory (usually called ReRAM), when binary devices (i.e. devices which utilize two logic levels) are addressed, despite the fact that these elements may be used for functions beyond pure memory, i.e. in the area of logic functions in various computational concepts. The term memristive will preferentially be used to describe the general characteristics of these elements and for applications of multilevel or analogue features. In a more general sense, we will use the terms: resistive switching or memristive phenomena, elements, cells, and devices. The nanoionic redox phenomena are placed in the perspective of other memristive devices in the framework of matrix-based data memories in Sec 2.1.

Historically, the first papers on resistance switching in oxides were published in the 1960s although similar electrochemical effects involving valve metal oxides date back even further. A large range of materials in a metal–insulator–metal (MIM) configuration have been reported to show non-volatile resistance switching, comprehensively reviewed by Dearnaley et al. [3], Oxley [4], and Pagnia [5]. The research activity on resistive switching decreased in the 1980s
because inherent problems were encountered in the elucidation of the microscopic mechanism and because of the advantages of Si semiconductor-based memories (DRAM, EEPROM, Flash) at that time which quenched the need for alternative memories. The research activity gained momentum again in the late 1990s, triggered by Asamitsu et al. [6], Kozicki et al. [7], and Beck et al. [8], covered by reviews such as [9–15] and in a more comprehensive monography recently published by Ielmini and Waser [16]. Over the years, all the major semiconductor research centers and all semiconductor manufacturers have initiated large research and development projects on redox-based memristive elements. In fact, given the number of papers and citations per year, today this field constitutes one of the megatrends in solid state physics and in nanoelectronics engineering. Some more details on the history will be sketched in Section 2.4.

In this article, we intend to review the background knowledge and the breakthroughs in this stimulating research field, and we have organized the material in the following manner: In Chapter 2, we provide the relevant definitions, a general classification, and an overview of the performance parameters of redox-based memristive elements. Chapter 3 addresses the physics of metal oxide cells which is required to understand the underlying microscopic mechanism. Prior to reversible resistive switching, the elements often require an electroforming process which is closely related to the memristive phenomenon and which is described in Chapter 4. Chapter 5 is devoted to the details of the valence change mechanism during the redox-based memristive switching, covering filamentary and area-dependent switching events as well as the origin of the non-linearity of the switching kinetics. Chapter 6 outlines the most relevant aspects of the reliability. Chapter 7 provides a conclusion and a brief discussion of the remaining open questions. The article can be regarded as a significantly extended and updated text of those parts of the review in Ref. [13] which cover the valence change mechanism in metal oxide cells (sometimes also called Oxide-ReRAM or OxRAM), complemented by aspects from related chapters in Ref. [16]. It has clear focus on the physics of the memristive states and the memristive switching of the valence change mechanism in metal oxide cells.

It should be mentioned that the authors of this article are not in a position to cover every detail of the nanoionics redox-based memristive phenomena in metal oxides, not only because of the huge diversity of findings and explanations but also because of the rapid advances in this field. There has been a vast number of excellent publications by many research groups all over the world in recent years, that it is simply not possible to acknowledge everyone adequately. We apologize for any conspicuous omissions with the related work cited here and for any technical deficiencies in this article.

2. Definitions and general classification

2.1. Physical storage principles

For electronic memories, the simplest and most efficient storage principle is based on physical states that can be written and read electrically by addressing a storage element in a dense matrix. Matrix-based memories (Figure 1) are based today on semiconductor chips and can be grouped into read and write random access memory (RAM) and read only memory (ROM). ROM can be further divided into once-programable ROM and re-programable ROM, typically utilizing metal–oxide–semiconductor field effect transistors (MOSFET) with an additional floating gate as storage element based on an invention by Kahng and Sze in 1967 [17]. The floating gate is differently charged during the programming sequence to shift the threshold voltage of the MOSFET channel in a non-volatile fashion. Before a new programming sequence, the information must be erased by discharging the floating gates of all memory cells. This discharge is performed
Figure 1. Categories of matrix-based solid-state memory principles showing the most relevant types. The abbreviations are explained in the text. Please note that some boundaries between categories are not sharp.

either by UV light in erasable PROM (EPROM) or electrically in electrically erasable PROM (EEPROM).

RAM devices are conventionally classified into volatile RAM and non-volatile RAM. Volatile elements will lose their information when the operating voltage is removed. **Volatile RAM** types traditionally comprise static RAM (SRAM) employing flip-flop-based latches as storage elements and dynamic RAM (DRAM), which uses a tiny dielectric capacitor with two different charge storage levels to represent the binary information. Due to the unavoidable self-discharge of capacitors and the limited barrier properties of the select transistor in the OFF state, the information needs to be refreshed periodically. The refreshing period is a fraction of a second (ms range). **Non-volatile RAM** technologies can be grouped into those which are based on transistors, those which use elements for storing electric charge, and those which rely on a change of the resistance. The transistor-based non-volatile principle in the ubiquitous **Flash memories** is identical to the EEPROM but with a different, much more simple and cost-effective architecture and relatively short re-programming times. As an alternative, polarization charges of ferroelectric gate oxides of a FET are employed in **ferroelectric FET (FeFET)**. Charge-based non-volatile RAMs are also utilizing ferroelectrics (Ferroelectric RAM, FeRAM) as dielectrics in separate capacitors similar to DRAM. The group of **non-volatile RAMs** based on the principle of resistance changes comprise of a large variety of different physical effects (Figure 2).

In the first approach, resistance-based memory (or: memristive) storage principles can be arranged into three categories, based on magnetic effects, electrostatic effects, and atomic configuration effects. **Magnetic effects** emerge from the tunnel magnetoresistance in magnetic tunnel junctions (MTJ) [18,19] in which the tunneling probability depends on the direction of the magnetization giving rise to the Stoner-Wohlfarth MRAM. A combination with the spin-transfer torque (STT) phenomenon [20–22] led to memory elements called STT-MRAM [23].

**Electric / electronic effects** either rely on the ferroelectric polarization affecting the resistance of a 2-terminal device such as a ferroelectric tunneling junction (FTJ) [25,26] first proposed by Esaki in 1971 [27], on a ferroelectric diode first realized by Blom et al. in 1994 [28] or they
Figure 2. Survey of resistance-based non-volatile memories, also called memristive elements or memristive RAM. Abbreviations and operating principles are explained in the text. Please note that such a classification cannot be done without ambiguity. The type of phenomena and their weighing is qualitatively based on the number of publications, the degree of development, and the potential seen by the committee of the International Technology Roadmap for Semiconductors [24]. The VCM-type elements and the area highlighted in red are the topic of this review paper. (Color figure available online).

operate through an electrostatic trapping/detrapping mechanism. In the latter case, a different geometry beyond a simple MIM structure must be employed, i.e. a substructure in the I-layer must be present. The reason is that one needs to have the trap site out of the flow of electrons, because otherwise its charge state cannot be kept in a non-volatile fashion which has been clearly demonstrated by Schroeder et al. [29]. This can be demonstrated by a floating gate transistor, which is turned into a two-terminal device by connecting the source and the gate, called MemFlash by Ziegler et al [30]. Possibly, the so-called nanometallic cell concept proposed by the group of I-Wei Chen [31,32] follows the same principle, although there are still many open questions. It is argued that nanosized dispersions of electronically conductive phases, e.g. metals such as Pt, in insulators such as SiO₂ are responsible for the resistive switching effect. Electric and magnetic functionalities are combined in the idea of multiferroic tunnel junctions (MFTJ) in one device [33–36].

The effects based on atomic configuration can be subdivided in the following manner: Nanoelectromechanical (NEM) memories can be regarded as electrostatically driven nanoscale relays [37]. A variety of effects have been observed in organic molecules [38,39]. Phase-change memories exploit the different resistivity of the amorphous and the crystalline phases of dedicated higher chalcogenides [40]. A local structural phase transition between two crystalline phases in narrow band gap Mott insulators is regarded as the underlying principle in Mott transition memories [41]. Redox-based memristive effects rely on resistance changes caused by ion motion.
over nanometer dimensions and can be subdivided into thermodinanical mechanisms (TCM), valence change mechanisms (VCM, which are emphasized in this review article), and electrochemical metallization mechanisms (ECM). Details of this subdivision are provided in Section 2.3.2.

It should be mentioned that there are numerous concepts beyond the 2-terminal devices, which rely on non-volatile physical effects in 3-terminal configurations, ranging from FeFET [26,42] to several other devices which exploit ion motion and redox process such as the electrochemical RRAM (ECRAM) [43], ionic floating gate-transistor [44], the so-called bulk-RRAM [45], and more.

2.2. Nanoionic redox-based memristive elements: modes of operation

A redox-based memristive element, i.e., a ReRAM cell, is generally built by an MIM structure, composed of an ion or mixed ion-electron conducting (i.e., resistive) material I sandwiched between two (possibly different) electron conductors M as electrodes. The materials I are oxides, higher chalcogenides, or other (at least partially) ionic solids. Figure 3(a) illustrates an example of the current response of a ReRAM cell upon a (triangular or sinusoidal) voltage excitation in a bipolar operation mode (see below), showing the typical hysteretic $I-V$ loops. The corresponding $R-V$ curves are shown in Figure 3(b).

Although memories are operated by pulses in real device applications, Figure 3 can be used to explain the basic memory operation. By applying appropriate programming or write voltage pulses $V_{wr}$, a cell in its High Resistance State (HRS) or OFF state can be SET to a Low Resistance State (LRS) or ON state or RESET back into the HRS. In the literature, the RESET is sometimes called an ‘erase’ operation. We assign the logic ‘0’ state to the HRS and the logic ‘1’ state to the LRS. In numerous cases multilevel switching has been demonstrated, i.e., intermediate resistance states (IRS) have been established in order to realize, for example, multiple bits per cells. The state of the ReRAM cell is detected by applying a (small) read voltage $V_{rd}$, which will not change the resistance state.

Depending on the specific type of ReRAM, the picture has to be broadened and different operation modes have to be used. In this sense, Figure 4 shows schematically characteristic $I-V$ diagrams recorded by periodic voltage sweeps (left) and pulse sequences with voltage pulse excitation and current responses (right). By far most device applications will use the pulse mode. Yet, the $I-V$ sweep mode is helpful for obtaining an overview of the characteristics and for determining the (very) approximate threshold voltages $V_{th}$ for the pulse operation. It has already been briefly mentioned that ReRAM cells often need an electroforming (short: forming) step prior to their first write/read operation. This electroforming step involves different voltage and current levels than the write/read operations.

Most ReRAM systems reported in the contemporary literature such as the valence change and the electrochemical metallization variants (VCM and ECM, see Section 2.3.2) are operated in the bipolar resistive switching (BRS or BS) mode (Figure 4(a)). Starting in the HRS, a SET process can be triggered by a voltage $|V_{SET}| > |V_{th1}|$ and leads to the LRS. Often current compliance (cc) is used for the SET operation in order to avoid damage to the cell and to optimize the operation. A read operation is performed at a much smaller voltage magnitude $V_{rd}$ to detect the current while avoiding a detectable change of the state. A voltage signal $V_{RES}$ of opposite polarity and an amplitude $|V_{RES}| > |V_{th2}|$ is used for the RESET process to switch the cell back into the HRS.

The unipolar resistive switching (URS or US) mode (Figure 4(b)) is characterized by the fact that all write and read operations can be performed with only one voltage polarity. This is known for phase change memories and also for the thermochemical variant of ReRAM (TCM, see Section 2.3.2). Starting in the HRS, the SET process takes place at a voltage $V_{SET} > V_{th1}$. 
Figure 3. Schematic illustration of (a) the current response to a voltage excitation of a ReRAM cell. A bipolar cell is shown in this example. Such curves are typically recorded by (triangular) voltage sweeps of low to medium frequencies. Above threshold voltages for the SET and RESET operation, the cell changes its resistance and retains the change when the excitation is released. The $I$--$V$ curves are transformed into resistance--voltage ($R$--$V$) curves in (b). The light curves indicate that intermediate resistance values can be written, too. Notes: It is often also possible to use a current excitation and record the voltage response. In any case, the characteristics of the hysteretic curves depend strongly on the specific system and may deviate considerably from the example shown here. Unipolar cells are excited by only one voltage (or current) polarity. Reprinted with permission from [13], © 2012 by Wiley-VCH.

with a LRS current limited by a current compliance (cc). It is important that the cc is released in the RESET process with $V_{\text{RES}} > V_{\text{th2}}$, so that the current can exceed the cc value which leads to change back into the HRS. The read operation is performed at a small voltage $V_{\text{id}}$ as in the bipolar operation. If the unipolar switching can be performed with both voltage polarities, e. g. in symmetric stacks, it has been also referred to as apolar switching [46].

The **complementary resistive switching** (CRS or CS) mode (Figure 4(c)) can be obtained by connecting two BRS-type ReRAM cells in an antiserial manner as suggested by Linn et al. [47]. In some cases, such a CRS behavior is also obtained by suitable processing and operation of single ReRAM cells [48]. Typically, the state of a CRS cell cannot be read at small voltages because the cell then always appears to be in a HRS. The state of the cell is only recognized at voltages $V > V_{\text{th1}}$. A read voltage $V_{\text{id}} > V_{\text{th1}}$ will lead to a higher current (upper $I$--$V$ trace in Figure 1(c), left) in the case of a logic ‘1’ state, and to a lower current (lower $I$--$V$ trace in Figure 4(c), left) in the case of a logic ‘0’ state. The write ‘0’ is achieved by a positive voltage $V_{\text{wr}} > V_{\text{th2}}$, and a write ‘1’ is obtained by a negative voltage $V_{\text{wr}} < V_{\text{th1}}$. Because of the relatively high read voltage $V_{\text{id}} > V_{\text{th1}}$ and the corresponding currents, the internal state is affected by
Figure 4. The three most common operation modes (bipolar, unipolar, complementary) of different types of ReRAM elements shown for the $I$–$V$ sweep operation (left) and the pulse operation (right). Details are described in the text. Please note that the elements are non-volatile. At first glance, in a simplified manner, the CS (CRS) mode resembles the so-called threshold switching which shows a hysteresis above a certain voltage bias but which disappears at voltages below this bias. The difference is the fact that the information is lost in the case of a threshold switch while it is maintained in a CRS cell and can be read-out in the indicated manner.

The read operation, i.e. the read voltage destroys the logic state (so-called Destructive Read-Out, DRO). As a consequence, the last logic state needs to be re-written into the cell after every read operation. This is the same situation as in the case of the standard (volatile) DRAM cells. It has
been also demonstrated that more than one bit can be stored in a CRS device. To this end, a current compliance during the SET operation ($V < V_{\text{th},2}$ or $V > V_{\text{th},4}$) [49] can be used or different voltage amplitudes in the RESET ($V > V_{\text{th},2}$ or $V < V_{\text{th},4}$) [50]. In addition, the information can be not only stored in the resistance but also the orientation of the filament, doubling the amount of information stored. This, however, requires a two-phase destructive read-out [50]. For a Non-Destructive Read-Out (NDRO) of CRS cells, a capacitive read-out has been proposed and demonstrated [51].

For bipolar operation, as realized in the valence change memory ReRAM variant covered in this article, the MIM system needs to have some asymmetry, such as different electrode materials, $M'$ and $M''$, or a dedicated voltage polarity during the initial electroforming step, in order to show bipolar switching behavior. In metal oxide cells, in which the reduction/oxidation of the metal oxide controls the memristive mechanism, there is an interplay between electrochemical and thermochemical redox processes determining if the unipolar TCM or the bipolar VCM mechanism dominates. If electrochemical redox processes prevail, VCM switching is observed, and if thermochemical redox processes dominate, TCM switching is encountered. Sometimes ReRAM cells can be toggled between unipolar and bipolar schemes depending on the operation conditions [52–54]. In addition, a transition between the VCM and ECM has been shown [55–57]. Furthermore, several VCM material systems are known which can change between a BS and a CS operation mode [48,58–61] as we shall see in Chapter 6.

2.3. Classes of nanoionic redox-based resistive switching phenomena

2.3.1. Survey of generic processes in redox-based memristive cells

Microscopic processes which may be relevant for the electroforming and the switching process of ReRAM cells are schematically sketched in Figure 5. It shows an MIM cell under voltage bias giving rise to a corresponding current through the cell. The metal electrodes $M'$ and $M''$ just carry electronic current while the resistive material $I$, which is an ionic solid $MX$ here, may carry both, electronic and ionic currents. Given the current direction in Figure 5 the ionic current may consist of anions $X^-$ and of cations $M^+$ moving to the left and to the right, respectively. In addition, ions $M^+$ may stem from the anode metal. The relative current contributions strongly depend on the type of ReRAM cells and the operation conditions. Local temperature increase due to Joule heating will typically occur in the interior of the MX layer and/or close to a contact. The ionic partial current in the MX leads to electrochemical reactions, oxidation at the anode and reduction at the cathode – the specific electrochemical interface reaction is determined by the type of ReRAM cell. Despite these ionic processes which may happen at the electrodes, the metal electrodes are defined in a way that they carry only electronic currents. The ionic current may be, at least partially, blocked at the electrode interfaces. This leads to so-called concentration polarization, also called kinetic demixing, i.e. an accumulation of the mobile ions near one electrode and a depletion near the other. Except in the (typically very narrow) space charge regions, this process is compensated by local redox reactions, i.e. a change in the average valence of the counter ions of the MX phase. As another result of the concentration polarization, phase transformations and space charge regions can occur. Phase transformation may, of course, also result from the temperature increase due to Joule heating, and from electrochemical interface reactions. The main driving forces for the processes just described can be summarized as gradients of the electrochemical potential $\nabla \eta$ and the temperature $\nabla T$. Often the interaction of both gradients drives the switching operation. In some cases, in particular when phase transformations are involved, one also has to consider mechanical stress gradients.

An ionic current in MX during SET and RESET is common to all redox-based memristive cells, and a corresponding redox process in the cell. It is essential that the state of matter, i.e. the
atomic configuration is changed by the nanoionic redox processes during switching (and during the preceding electroforming process), and that these different states of matter lead to different resistance values of the overall cell. Furthermore, it is required that the change of the state of matter is reversible for a large number of SET/RESET cycles. Consider a solid-oxide fuel cell for comparison. Several of the described processes, such as ion conduction in MX as well as oxidation and reduction at the electrode interfaces take place in such a cell, too. However, the processes are stationary and do not lead, in the ideal case, to a change of matter. In ReRAM cells, as said, a change of the state of matter or, in other words, a change in the atomic configuration is essential. It can be described in terms of memristive devices (see Section 2.5).

In the case of non-destructive read-out (NDRO) operation, the read operation of a ReRAM cell should not change the state of matter (i.e. BS and US mode in Figure 4). As a consequence, processes such as the ionic current contribution in MX and temperature activated processes must be negligible during read operations.

For any discussion of the stability of states and the long-term reliability (Chapter 6), one should keep in mind that only one of the states, ON state (LRS) or OFF state (HRS) or any intermediate state, can be thermodynamically stable. Due to the nature of redox-based resistive switching, the other state(s) must be metastable. They are frozen-in after a kinetically fast (i.e. temperature- and/or field-accelerated, see Section 5.7.2) switching event. The reason for only one state being the thermodynamically stable state is due to the fact that there may be only one arrangement of ions and atoms that has the lowest free energy. This is different to ferroelectric and ferromagnetic systems in which (vector!) states with opposite polarization direction may both be thermodynamically stable, as described by the Landau-Ginzburg-Devonshire theory. The situation in ReRAM is similar to PCM, where only the crystalline ON state is thermodynamically stable. It should be mentioned that in ReRAM cells, both states may be metastable, e.g. characterized by a frozen-in enrichment or a frozen-in depletion of ions in front of an electrode interface.
2.3.2. Types of redox-based memristive cells

As mentioned before, three major types of ReRAM cells can be distinguished which involve ion migration and coupled redox processes in the MIM cell [62]. Firstly, the bipolar electrochemical metallization memory effect (ECM, also called Conductive Bridge RAM, i.e. CBRAM) relies on an electrochemically active electrode metal such as Ag, the drift of the highly mobile Ag\textsuperscript{+} cations in the ion conducting I-layer, their discharge at the (inert) counter electrode leading to a growth of Ag dendrites which form a highly conductive filament in the ON state (LRS) of the cell. Upon reversal of polarity of the applied voltage, an electrochemical dissolution of these filaments takes place, resetting the system into the OFF state (HRS). For a detailed review on ECM systems, the reader is referred to Ref. [63]. Secondly, the valence change memory effect (VCM) occurs in a wide range of metal oxides as band insulators and is triggered by a migration of ions constituting the metal oxide. The mobile ions may be anions, such as oxygen anions (which are usually described by the motion of the corresponding vacancies, i.e. oxygen vacancies) or cations, such as cation interstitials. Sometimes the nature of the mobile ion is not known. Since both, oxygen vacancies and cation interstitials, are donor-type defects one may regard the process as being controlled by mobile donors. A subsequent change of the stoichiometry leads to a redox reaction expressed by a valence change of the cation sublattice and a change in the electronic conductivity in front of an electrode interface. This bipolar memory switching is induced by voltage pulses, where the polarity of the pulse determines the direction of the change, i.e. reduction or oxidation. In specific cases, the distinction between ECM and VCM-type switching may disappear [55]. A third type relies on a thermochemical memory effect (TCM, also called fuse-antifuse memory) due to a current-induced increase of the temperature, which leads to a redox-related change of the stoichiometry along a discharge filament, and a subsequent freezing-in of this ON state (LRS). A differently shaped current pulse disrupts the conductive filament again to return the cell into the OFF state (HRS). For a detailed review on TCM systems, the reader is referred to [64].

Apart from the three major redox-based types, also so-called Mott-based switching can occur. In contrast to band insulators, Mott insulators exploit the competition between the Coulombic repulsion energy of transition metal d- or f-orbitals in transition metal oxides and the bandwidth (or electron kinetic energy) of incompletely filled bands. Reversible insulator-to-metal-transition (IMT) can be induced by pressure, temperature, and voltage. Voltage-induced IMT leading to non-volatile states are known as memristive processes. The injection of dopant ions such as oxygen vacancies as mobile donors (i.e. the extraction of oxygen ions) can lead to an electron doping and, hence, to a change in the filling of the corresponding bands. Such the voltage-induced filling-controlled IMT is also a kind of a VCM-type resistive switch, which occurs in a Mott insulator instead of a band insulator. A comprehensive review of Mott-transition based memristive switching is provided by Ref. [65].

As mentioned, this review will focus on VCM systems. The other types will only be mentioned if there are relevant aspects in conjunction with VCM systems.

2.3.3. General features of valence change memories

In the case of valence change memory (VCM) cells, the MIM system consists of an electronically active interface (electronically active electrode, AE), at which the switching takes place, a mixed ionic–electronic conducting (MIEC) I-layer, and an ohmic counter electrode (OE). Sometimes, the conducting (MIEC) I-layer is subdivided into an oxygen exchange layer (OEL) at the OE and a switching layer (SL) at the AE. Many variants are known.
A **microscopic analysis** of ReRAM cells reveals that the electrodes and, in particular, the material ‘I’ between the metal electrodes M in the MIM structure of the cell participate in different manner in the resistive switching process. Typically, MIM structures are sandwich stacks. But also lateral MIM structures on surfaces and metal electrodes as scanning probe tips are used, in particular for research purposes. Without addressing the physics and chemistry of the switching at this point, one can distinguish categories of the location of the switching event in a MIM stack. Along the path from one electrode to the other electrode (which we will call the **parallel direction**, because it is parallel to the direction of the current flow), one can distinguish the following switching locations:

- at/near one of the electrode interfaces which we will call **interface-type switching**; as we shall discuss, this type of switching is the most frequent one,
- in the (approximate) center between the electrode interfaces which we will call **bulk switching** because the switching happens somewhere in the bulk of the I-layer,
- involving the entire path between the electrode interfaces; this rarely reported case will be subsumed under bulk switching as well.

In the perpendicular direction i.e. in the plane of cross section of the cell (which we will call the **perpendicular direction**, because it is perpendicular to the direction of current flow) one can distinguish:

- filamentary switching, i.e. a switching by the formation and dissolution of a single **conducting filament**; multifilamentary switching which is sometimes observed will be subsumed in this category as well
- area-proportional switching or **area-dependent** switching, i.e. a switching event that takes part over the entire cross section of the cell so that the currents become proportional to the cross-sectional area of the cell. Because this switching typically happens close to one electrode, it is also called interface-type switching in the literature [66]. However, this should not be confused with the interface-type switching introduced above. In this article, we will stick to the notation introduced here.
- In addition, there are reports about bulk-type switching in which apparently the entire volume of the material ‘I’ participates in the switching process [67]; we will subsume these cases under area-dependent switching.

Figure 6 illustrates these scenarios schematically. Of course, this brief description represents a very simple, coarse-grained classification only. In reality, many intermediate cases have been reported such as multiple filament switching or switching areas that are relatively extended spots without spreading over the entire cross section of the cell. It should be pointed out further that these categories and combinations thereof do not occur equally. The most frequent resistive switching described in the literature is filamentary switching (with evidence provided, for example, in Refs. [68–89]). If such a cell shows a bipolar operation, there are usually clear indications that the switching is localized in both, in the perpendicular direction (filamentary) and in the parallel direction (near one of the electrode interfaces).

The simple ReRAM concept might resemble a switch, which can be settled in two different configurations to let a current flow or not. Although apparently simple, physical processes responsible for the switching have not been fully explained at the time of the publication of this article, not to mention the great complexity of achieving a working technology based on the ReRAM concept (e.g. a crossbar non-volatile memory or a logic circuit), and outperforming existing technologies. The purpose of this article is to address the fundamental physics and chemistry concepts of the VCM-type memristive phenomena and ReRAM devices.
Here, we will briefly describe a filamentary VCM-type memristive switching based on an n-conducting MIEC oxide. A typical $I-V$ characteristic is shown in Figure 7, together with sketches of the switching mechanism. In the OFF state (Figure 7(A)), the filament consists of the n-conducting MIEC oxide (called plug) and a potential barrier (called disc) in front of the left electrode. Upon application of a negative $V_{SET}$ to the active interface, mobile donor ions (i.e. oxygen vacancies or cation interstitials) from the plug part of the filament migrate towards the active interface (Figure 7(B)), which results in a significant decrease in the barrier height and width due to a local reduction process, which turns the cell into the ON state (Figure 7(C)). For the RESET, a positive voltage $V_{RES}$ is applied to the active interface which repels the mobile donor ions from the active interface (Figure 7(D)), leading to a local re-oxidation, and turns the cell into the OFF state again.

If the voltage of the active electrode (AE) is displayed and the voltage of the ohmic electrode (OE) is regarded as the reference potential (often called ground potential), as done in Figure 7, then the switching polarity will be called counter-eightwise (C8W), because the $I-V$ curve has a drawing direction which is against that of the handwriting of a (tilted) ‘8’. The opposite switching polarity is called eightwise (8W). It should be noted that both switching polarities have been found in the same SrTiO$_3$ cell as first reported in Refs. [70,90]. It was possible to reveal the exact location of both events. This coexistence of both switching polarities has later been confirmed [91] and also found for other oxides such as TiO$_2$ [92,93], or Ta$_2$O$_5$ [94].

2.3.4. Requirements of memory cells and beyond

The target application of memristive devices (ReRAM) is either non-volatile memories or logic circuits in which memory functionalities are integrated, such as FPGA-type and neuromorphic architectures. Based on the circuit requirements of high-density memories such as non-volatile Flash or volatile DRAM, one can collect a number of requirements for ReRAM cells [11]:

- **Write operation**: The write voltage $V_{wr}$ should be in the range of a few hundred mV (not lower, in order to be compatible with scaled CMOS) to a few V. The latter limit gives a great advantage over Flash which suffers from high programming voltages. The write time, i.e. the length of write voltage pulses $t_{wr}$ is typically < 30 ns. This allows for a competition with DRAM specifications and outperforms Flash, which has a programming
Figure 7. Schematic current–voltage characteristic of a Pt / ZrO$_x$ / Zr valence change memory (VCM) cell, recorded with a triangular voltage sweep. Here, the Pt represents the active electrode (AE) and Zr the ohmic electrode (OE). The insets A to D show the different stages of the switching procedure. The sketches are consistent with simulations discussed later in this paper. The green spheres indicate mobile donor ions (here: oxygen vacancies), the purple spheres indicate Zr ions in a lower valence state. Please note that only oxygen vacancies are mobile while Zr ions are immobile and may only change their valance state. (A) OFF state (HRS); (B) SET process; (C) ON state (LRS); (D) RESET process. Adapted and reprinted with permission from [13], © 2012 by Wiley-VCH. (Color figure available online).

speed of some 10 μs. In special cases, $t_{wr} < 100$ ps have been demonstrated for ReRAM cells [95,96], which is similar to high-performance SRAM cells.

- **Read operation:** Read voltages $V_{rd}$ need to be significantly smaller than write voltages $V_{wr}$ in order to prevent an unintentional write during the read operation. Because of constraints by circuit design of the sense amplifier, $V_{rd}$ cannot be less than approximately one tenth of $V_{wr}$. An additional requirement originates from the minimum read current $I_{rd}$. In the ON state, $I_{rd}$ (of the ON state) should be at least 100 nA to allow for a detection after a fast charging of the bit line capacitance according to $I = CV$. The bit line capacitance is in the order of (at least) 30 fF. The read time $t_{rd}$ must be in the range of $t_{wr}$ or preferably shorter.

- **Resistance values and resistance ratio:** The resistance values $R_{LRS}$ and $R_{HRS}$ (or: $R_{ON}$ and $R_{OFF}$) have to match the circuit requirements which may vary strongly on the application of the circuit. Resistance ratios ($R_{HRS}/R_{LRS} > 10$) are preferred to allow for small and highly efficient sense amplifiers and, hence, reasonably cost-effective ReRAM chips. The $R_{HRS}$ and $R_{LRS}$ values must take into account the resistance of the bit line and the select devices (e.g. select transistor channel) so that the $R_{HRS}/R_{LRS}$ ratio of the ReRAM cell may need to be larger [97]. In principle, an $R_{HRS}/R_{LRS}$ ratio of as small as approx. 1.3 can be
utilized by dedicated circuit design as demonstrated for MRAM. However, this comes on the expense of much larger periphery circuits.

- **Endurance**: Contemporary Flash shows a maximum number of write cycles between $10^3$ and $10^7$, depending on the type. ReRAM typically has the potential to show similar or even much higher **endurance** (Chapter 6).

- **Retention**: A data **retention time** $t_{\text{ret}}$ of at least 10 years is required for universal non-volatile memories. This retention time must be kept at a thermal stress up to the specified upper operation temperature of the circuit (e.g. 85°C) and a small electrical stress such as a constant stream of $V_{\text{rd}}$ pulses.

- **Read/write disturb immunity**: The ReRAM devices should not change its state while an adjacent cell is operated (read or written).

- **Operation energy per bit**: The energy required to write a cell, i.e. to change its resistive state, must be competitive to that of the conventional device to be replaced.

- **Scalability**: The geometrical size to which a cell can be miniaturized before it encounters inherent (physical) limits is called the scalability limit. In particular for stand-alone memories, any emerging technology such as ReRAM has to compete with contemporary Flash.

- **Stackability**: The option to stack several layers of cells on top of one another by fabrication technology is called stackability. This is related to the scalability. For instance, contemporary Flash only shows a high device density on the chip because of the fact that a large number (i.e. more than hundred) of layers are stacked. The individual Flash cell has already approached its physical scaling limit because of the minimum thickness of the required tunneling barriers and additional dielectric layer thicknesses [98].

- **Multilevel storage**: This refers to the option to store more than one bit of information in one cell. This is also highly desirable, in particular, for e.g. neuromorphic compute circuits.

Obviously, the performances required of memory cells depend on the specific application. In particular, VCM-type ReRAM are interesting for **Storage Class Memories** as they bridge the latency gap between the DRAM and the mass storage devices [99–103]. Yet, there are some basic guidelines and some interrelations which need to be considered. Often, the optimizations of different **performance parameters** contradict one another and trade-offs are needed. For example, the requirement of long retention times and short write times in combination with write and read voltages in a comparable range (the $V_{\text{wr}}/V_{\text{rd}}$ should not exceed approx. 10 for circuit design reasons) represents a huge demand, sometimes referred to as the **voltage–time dilemma**. In the worst case of a fast non-volatile memory, a binary state should not change under permanent $V_{\text{rd}}$ pulse trains for 10 yrs, while it must change within a very short time ($t_{\text{wr}}$) if a $V_{\text{wr}}$ pulse is applied. This dilemma must be solved by kinetics of the switching process between binary states which must show a **non-linearity of the switching kinetics** of more than 15 orders of magnitude. In many cases, it has been possible to find systems in which this non-linearity is actually realized. There are processes in solid-state physics showing such huge non-linearities, for example, the voltage dependence of tunneling through barriers, the temperature-controlled phase transformation, and threshold voltage controlled electrochemical processes. The physical mechanisms that offer the required huge non-linearity of their kinetics are discussed in Section 5.8. In some cases, the requirement imposed by the voltage-time dilemma is significantly relaxed. This holds for memory applications in which ReRAMs are supposed to replace DRAM due to their better scaling, in applications which do not require maximum retention but a high endurance.
Figure 8. Measured $I–V$ characteristics for a Nb–Nb$_2$O$_5$–Bi MIM stack showing bistable resistive switching (Adapted and reprinted with permission from [108], © 1965 by the American Institute of Physics). The device is initially in a low resistance state (a) due to the previous forming operation. Reset transition to the high resistance is shown for negative applied voltage (b), while set transition to the low resistance appears at positive voltage (c). The inset shows the $I–V$ curves of three stable states, a high resistance state H and a low resistance state L and an intermediate state M (d).

2.4. Historical aspects

Phenomena today known as resistive switching have been studied in various solid-state materials since the early 1960s. Typically, early reports of resistive switching emphasized the negative differential resistance (NDR) observed in the $I–V$ characteristics of MIM structures. The ‘I’ material in the MIM stack investigated in this period stands for one of various binary and multinary oxides, chalcogenides as well as group-IV, III–V and II–VI semiconductors. In 1962, Hickmott reported a large NDR in MIM stacks with oxide insulator, such as SiO$_x$, Al$_2$O$_3$, Ta$_2$O$_5$, ZrO$_2$, and TiO$_2$ [104]. The NDR effect was explained by a space-charge-limited current in the oxide layer. In this model, the trapping of electrons in localized states influences the leakage current, which is then called space-charge-limited current [105,106].

Bistable resistive switching was first reported in 1964 in NiO thin films on Ni substrate, where the switching was believed to be due to the formation and rupture of a nickel metallic filament in the NiO layer sandwiched by two electrodes [107]. Later in 1965, bistable resistive switching between two stable resistance states was shown in Nb$_2$O$_5$ [108]. Figure 8 reproduces the reported $I–V$ curves for a Bi/Nb$_2$O$_5$(125 nm)/Nb cell. A first positive voltage sweep initiated the forming of the initially insulating stack and the device turned into a low resistance state (a). Application of a negative voltage led to a RESET transition to high resistance (b), while application of a positive voltage causes the SET transition to the initial low-resistance state (c). The bipolar switching was called bistable because both states were stable after releasing the voltage stimulus.

Bistable resistive switching has also been reported for thin films of Ta$_2$O$_5$ [109], SiO [110], TiO$_2$ [111], Al$_2$O$_3$ [112], and for ZnSe–Ge heterostructures [113]. The first report on what we today classify as ECM-type switching dates back to 1976 when Hirose and Hirose observed Ag dendrites being formed and dissolved between the Ag and Au electrodes in a bipolar operation mode of lateral Ag/As$_2$S$_3$/Au cells [114]. This early period of research faded in the late 1970s. Obviously, the interest in this area decreased because of the overwhelming progress of
the Si-based integrated circuit technology, in particular, the Flash memories. In 1967 Kahng and Sze invented the floating gate MOSFET as a non-volatile three-terminal device [17], which was developed into the EPROM (Erasable Programable Read-Only Memory) in the 1970s and into the Flash memory in the 1980s. Flash memories have grown into the dominating non-volatile solid-state storage since the turn of the millennium. It may be speculated that another reason for the decrease in research in metal oxides and related compounds was the lack of progress in understanding and controlling the resistive switching phenomena possibly due to insufficient analytical tools at that time. The period has been reviewed comprehensively by Dearnaley et al. [3], Oxley et al. [4], and Pagnia et al. [5].

A new era in research on resistive switching gradually started in the mid-1990s. The Tokura group found electrically triggered resistive switching in Pr$_x$Ca$_{1-x}$MO$_3$ (PCMO) while investigating the magnetoresistive properties of this material [6], and Zhuang et al. used this material to fabricate the first integrated ReRAM memory, a 64-bit array, using a 500 nm CMOS technology [115]. In 2000, the IBM Zurich lab reported the resistive switching of perovskite-type zirconates, including many properties which are essential to NVM applications [8]. In the ECM-type area, Kozicki, Mitkova et al. started to study the Ag–GeSe systems in the late 1990s [7], while the Aono group published their first report on so-called atomic switches in 2001 [116]. These devices make it possible to control the electrochemical formation and dissolution of a, for example, an Ag atomic bridge in a nanogap between a mixed electronic–ionic conducting Ag$^+$ electrolyte and a metal electrode with the precision of Landauer conductance quantization [117]. In 2004, Samsung successfully demonstrated a high-density ReRAM chip using a 180 nm technology. It was based on unipolar switching Pt/NiO/Pt cells with an endurance of $10^6$ SET/RESET cycles (which today would be assigned to the TCM category). These and related papers have been the beginning of an unprecedented rise of R&D activities which led to the mega-trend which we encounter in the 2010s. TCM-type devices suffered from limited endurance. A major breakthrough was the introduction of thin reactive metal layers on one electrode side [118], which led to bipolar switching cells with improved properties. This led to the success of bipolar switching memristive cells.

While there has been a basic understanding of the ECM mechanism from the beginning, a broad spectrum of mechanisms has been suggested as underlying mechanisms for the resistive switching in the various metal oxide systems. In 2006 Waser’s group was able to clarify the effect as a motion of oxygen ions and a coupled valence change in the cation sublattice on the nanometer scale at structural defects in the crystal lattice of the metal oxides near one electrode [119]. For this reason, the expression valence change memory effect (VCM) was suggested for bipolar metal oxide systems [11], extending the more detailed classification of the nanoionically driven, redox process based resistive switching memories [9]. The first spectroscopic evidence for the valence change was published for the dopants of Cr-doped SrTiO$_3$ by Janousch et al. [120] and for the host cation in TaO$_x$ based cells by Z. Wei et al. [121]. The filamentary nature of an oxygen-deficient phase generated during the forming process was shown by Kwon et al. [122] using high-resolution transmission electron microscopy. In 2008, William’s group at the Hewlett-Packard Labs discovered that the electrical characteristics of resistive switching elements can be described in terms of the theory of memristive devices [123]. This theoretical concept was published in 1976 by Kang and Chua [2] as a generalized version of the theory of memristors introduced by Chua in 1971 [124], and this link led to a further increase in the international research activities (see Section 2.5).

Further historical notes will be given in the forthcoming individual sections of this article. The technological realization has advanced to the point that a VCM-type 32 GB ReRAM chip has been presented by Sandisk and Toshiba in 2013 [125], and an ECM-type 16 GB ReRAM chip has been introduced by Micron and Sony in 2014 [126]. Moreover, VCM-type ReRAM devices have
been successfully co-integrated into CMOS circuits for in-memory computing [127] Recently, even 3D HfO2-based memristor circuits have been employed to built complex neural networks [128].

2.5. From memristive effects to non-volatile resistive switching

Throughout this article, we will use the term *memristive* in a qualitative manner denoting phenomena and devices in which the resistance state is memorized until the next stimulus that is strong enough to change the state (SET pulse).

In some journals, it has become common to describe such non-volatile resistively switching devices in the framework of *memristors* defined by Chua in 1971 [124] and, more general, *memristive systems* introduced by him and Kang in 1976 [2]. In order to clarify some confusion encountered in the literature (to which some recent redefinitions, e.g. [129], may have contributed), we will briefly discuss the issue here. The *memristor* has been originally defined in 1971 by the relationship between the charge $q$ and the flux $\varphi$ (as the time integral of the voltage). In today’s terminology one would write:

$$
\dot{q} = I
$$

This definition establishes the memristor as a fourth passive element apart from the resistor, the capacitor, and the inductor in the relation of the voltage $V$, the current $I$, the charge $q$, and the flux $\varphi$. However, until today, the memristor is a purely hypothetic element which is not represented by any simple device (‘simple’ means that electronic circuits which emulate its behavior are excluded). Furthermore, even if it existed it would not be useful as a non-volatile resistively switching device because the time derivative of the charge $q$ as the inner state variable is just proportional to $I$, while non-volatile resistively switching devices demand a very strong non-linearity (and other issues).

The general *memristive system* introduced in 1976 is a two-terminal device defined by a more complex state-dependent Ohm’s law and a state equation:

$$
V(t) = R(x, I, t) \cdot I(t)
$$

$$
\dot{x} = f(x, I, t)
$$

$x$ is an internal state variable or, more general, a vector of $n$ internal state variables $x = (x_1, x_2, \ldots, x_n)$. Internal state variables may be the temperature of the device, a magnetization, a chemical composition, etc. According to Equation (2), the resistance $R$ will be non-linear $R(I)$ and its value at any time $t$ will depend on the entire past history of the device, because of

$$
x(t) = \int_{-\infty}^{t} f(x, I, \tau) d\tau.
$$

In an $I$–$V$ diagram, memristive systems show a hysteresis when $V$ or $I$ are used as a periodical stimulus. The hysteresis loop is pinched, i.e. it goes through the origin. The shape of the hysteresis loop will depend on the frequency $\omega$ of the periodic stimulus. For $\omega \to \infty$ it will approach a linear resistance, and for $\omega \to 0$ it will approach a non-linear resistance, as shown in Figure 9. In this original definition of Leon Chua presented here, the memristive system represents, for instance, any type of thermistor (temperature-dependent resistor), or any other conceivable two terminal device. In order to use the concept of memristive systems to describe specifically non-volatile resistively switching devices, additional requirements must be introduced:
Figure 9. Schematic $I$-$V$ curves of a pinched hysteresis loop of a memristive system subject to a periodic stimulus [130]. The size of the hysteresis depends on the frequency of the voltage: at low frequencies, $\omega_3$, memristive systems typically behave as non-linear resistors with different resistances for the two polarities, at intermediate frequencies, $\omega_2$, they exhibit pinched hysteresis loops, and at high frequencies, $\omega_1$, they typically operate as linear resistors. On the plot, $\omega_1 \gg \omega_2 \gg \omega_3$. Adapted and reprinted with permission from [130], © 2011 by Taylor & Francis.

(i) $x$ must include a material dependent state variable such as the magnetization, the crystallographic phase, or the length of a conducting filament formed by an internal redox process;

(ii) $x$ must have a lower and upper limit, $x_{\text{min}} < x < x_{\text{max}}$;

(iii) since $\dot{x}$ describes the kinetics of the switching process, the function $f$ in $\dot{x} = f(x, I, t)$ must be highly non-linear in order to reflect a solution for the voltage-time dilemma. Note that in case of ReRAMs, the current $I$ corresponds to the ionic current rather than the electronic (or total) current.

The concept of memristive systems has been applied to resistance-based RAMs for the first time by Strukov et al. [123], where the authors implicitly included the additional requirements (i) and (ii), while requirement (iii) has been indicated in Ref. [131]. It should be noted that in the general definition according to Equation (2) memristive systems (which is also called generalized memristor [129]) do not represent the fourth passive element. A controversy developed over the years in which it has been discussed if the memristor has been found or not [132], and a simple test has been proposed to clarify about whether an ideal memristor does actually exist or it is a purely mathematical concept [133,134].

3. Relevant properties of metal oxides and resistive switching cells

The resistive I-phase in a MIM cell is involved in the redox processes considered in this chapter. Typically, the I-phase is an ionic material, i.e. a solid compound consisting of atoms of different electronegativity and, as a consequence, the chemical bonds show a significant ionic bond character. In many cases, our ionic materials will be metal oxides. For this reason, we will use them to provide a brief introduction to relevant properties of ionic materials. It should be noted that, more precisely, the I-phase is a mixed ionic-electronic conducting material (MIEC) as already mentioned in Chapter 2. We will mainly focus on crystalline solids and briefly mention amorphous materials in Section 3.2.2. Of course, only a very brief sketch of this topic can
be provided in which those aspects are mentioned which are relevant for VCM-type resistive switching. Comprehensive treatments are found e.g. in refs [135,136].

3.1. Phase diagrams – full-valent and lower-valent phases

In stoichiometric compounds such as MgO, the elemental composition can be represented by a ratio of well-defined (small) integers which correspond to the integer valences (or: oxidation states) of the atoms involved. For example, the Mg:O ratio is 1:1 and the valences are +2 for the Mg atom and −2 for the O atom. Please note that these valence numbers represent a simplification which is justified only for highly ionic compounds. For compounds with a significant covalent bond contribution, such as SiO₂, the Si atom can only formally be attributed a valence of +4. The precise description requires a spatial electron density distribution around the atom under consideration.

The Gibbs free energy of formation \( \Delta G_f \) of the reaction to form a stable oxide from metal and molecular oxygen

\[
x \text{M} + \frac{1}{2} \text{O}_2 \rightleftharpoons \text{MO}
\]

will always be negative (because this is the definition of ‘stable’). The entropy term \( \Delta S_f \) within the expression

\[
\Delta G_f = \Delta H_f - T \Delta S_f
\]

however, will be negative for solid oxides because of the high entropy of molecular oxygen as a gas. As a consequence, the magnitude of \( \Delta G_f \) decreases with increasing temperature (Figure 10). The slope is given by \( \Delta S_f \) (approx. \( -0.92 \pm 0.16 \text{ meV/K} \)).

Because of the negative entropy of formation, obviously oxides with lower valence of the cations and finally the metals will become stable for decreasing \( p_{\text{O}_2} \) at a given \( T \) (see reaction (4)) and for increasing \( T \) at a given \( p_{\text{O}_2} \).

A more detailed description of solid oxides includes the thermodynamics of mixtures between different oxide phases and the discussion of phase diagrams (Figure 11). These diagrams reveal on the one hand, for example, temperature – composition regions in which solid solutions occur, i.e. homogeneous phases with mixed cation valences in which the average valence varies with the O content. On the other hand, they show regions in which phases are separated, i.e. heterogeneous phase mixtures of stoichiometric compounds are present. In a fictitious oxide MO₂ (Figure 11), M₂O, MO, M₂O₃, and MO₂ have been considered, for which the corresponding valences of the cations M\(^{z+}\) are \( z = 1, 2, 3, \) and 4.

In many cases such as TaO\(_x\) and HfO\(_x\) there are no stable intermediate phases between the metal and the fully oxidized metal oxide [138,139].

3.2. Electronic structure and band diagrams

In solid state physics, the electronic structure and the formation of energy band in solids (insulators, semiconductors and metals) are described by two theoretical approaches: the nearly-free electron model and the tight-binding model. The formation of the energy bands in the insulator MgO is shown in Figure 12. It shows how the energy levels of (fictitious) free ions (Figure 12(a)) change when the ions form a crystal lattice and every ion is surrounded by ions of the opposite charge. This Madelung effect leads to a significant stabilization of the solid (Figure 12(b)). The interaction of the atomic orbitals results in the formation of energy bands (Figure 12(c)), which are shown in density of states diagram in Figure 12(d). The (filled) energy levels of the 2p-orbitals of \( O^{2-} \) ions typically form the valence band while the (empty) 3s-orbitals of the \( Mg^{2+} \) ions form the conduction band. The same holds true for TiO\(_2\) and for SrTiO\(_3\).
Figure 10. Temperature dependence of the free energy of formation of various transition metal oxides according to Equation (4). The energy axis is also shown in equivalent equilibrium oxygen partial pressures \( p_{O_2} \) at \( T = 1273 \) K. The color code shows phases of identical elements. Each line shows the coexistence of an oxidized and a reduced compound. For simplicity only the oxidized form is indicated. The diagram can be used to determine the thermodynamically stable phase for a given \( p_{O_2} \) and \( T \) (redrawn from Ref. [137]). (Color figure available online).

Figure 13 shows a more realistic density of states diagram for \( \text{SrTiO}_3 \) obtained through a density functional theory (DFT) calculation. The conduction band is mostly formed by the Ti 3d states with a small contribution from the oxygen 2p resulting from the finite covalence of the system. The valence band is formed by the oxygen 2p band with small contributions from Sr and Ti.

It should be noted that in transition metal oxides in most cases the d-orbitals of the metal cations contribute to the conduction band. However, due to correlation effects the d-band is often split, resulting in a more complex band structure such as that the oxygen 2p band is energetically in between the two Hubbard bands of the metal 3d states.

The electronic structure of other classes of oxides such as PCMO will be introduced in the corresponding context of the following chapters.
Figure 11. Phase diagram of a (fictitious) metal oxide MOx which shows several stable compositions. Dark blue: homogeneous solid solution phase, bright blue: heterogeneous mixture of the two neighboring solid phases, green: mixture of solid and liquid phases, yellow: liquid. Please take into account that in a translation into Figure 10 the heterogeneous regions (bright blue) collapse into the lines shown in Figure 10. Furthermore, in order to maintain a given composition on increasing temperatures in Figure 11, one would need to adjust the pO2 significantly, as shown in Figure 10. Arrows: also those compositions, such as MO2, which are stoichiometrically well defined (sometimes called ‘fine compounds’), i.e., which are not involved in solid solutions with neighboring phases, show a certain stoichiometric width δ for MO2±δ because of the formation of point defects (Section 3.3). Although δ increases significantly with the temperature, for most oxides it reaches only small values in the order of 10^{-3} even close to the melting temperature. Adapted and reprinted with permission from [13], © 2012 by Wiley-VCH. (Color figure available online).

3.3. Lattice disorder

For entropic reasons, also the stoichiometric compounds have got a certain stoichiometric width at high temperatures, i.e. a (typically very narrow) compositional region in which they are non-stoichiometric and represent solid solutions. In many cases, this stoichiometric width δ is only a very small deviation from the ideal stoichiometric composition as e.g. in TiO2−δ where δ is approx. 0.001 at T = 900 K.

In these cases, it is more convenient to discuss just the deviations from the stoichiometry instead of dealing with the overall composition. Microscopically, these deviations originate from point defects in the crystal lattice. In general, vacancies, interstitials, and substitutional defects are known. Because of the relatively large size of O^{2−} ions compared to most cations, oxygen vacancies occur frequently while oxygen interstitials are energetically less favorable, and, hence, occur in rare cases only. Cation defects may occur as vacancies or interstitials. We will give a very brief introduction to the thermodynamics of point defects here. A more detailed introduction can be found, e.g. in the comprehensive text book of J. Maier [136].

3.3.1. Schottky- and Frenkel disorder

For a (fictitious) model oxide MO, we will consider the so-called Schottky disorder which involves oxygen vacancies V_o^− and cation vacancies V_M^+. Here * (') denotes a positive (negative)
Figure 12. Sketch of the formation of energy bands of MgO (Adapted and reprinted with permission from [140], © 1987 by Oxford University Press). (a) Energy levels of free ions. (b) Energy levels of ions in the crystal field. (c) Band structure by overlap of atomic orbitals. (d) Density of states (strongly simplified).

Figure 13. Sketch of the formation of energy bands of SrTiO$_3$ as calculated by DFT. The density of states (DOS) and the local DOS with Sr, Ti, and O atoms shown in green, blue and red, respectively. Adapted and reprinted with permission from [141], © 2016 by Forschungszentrum Jülich. (Color figure available online).
charge relative to the ideal lattice according to the Kröger-Vink notation [142] which will be used throughout this paper. The terms density and concentration will be used as synonymous expressions. For the species i, they will be abbreviated by \( N_i \) or \([i]\). In accordance with the literature, the concentrations of electrons and holes will get special symbols: \( n \equiv N_e \equiv [e] \) and \( p \equiv N_h \equiv [h] \). In a simplified picture, the following defect equilibria (i.e. the reaction equations and the corresponding law of mass action equations) must be considered in order to describe the lattice disorder of the oxide:

- The **Schottky equilibrium** describes the formation of anion and cation vacancies (typically at very high temperatures at which the lattice is formed):

  \[
  M_M + O_O \rightleftharpoons V_O^{\ast} + V_M^{\ast} + ^{'}\text{MO}(s)^{'} \quad [V_O^{\ast}] \cdot [V_M^{\ast}] = K_S = K_S^0 \exp(-W_S/k_B T) \quad (6)
  \]

  Here, ‘MO(s)’ denotes the segregation of MO at external or internal surfaces of the crystalline phase. The oxygen vacancies \( V_O^{\ast} \) and cation vacancies \( V_M^{\ast} \) can be regarded as intrinsic (or: native) dopants, where \( V_O^{\ast} \) are doubly ionized donors, and \( V_M^{\ast} \) act as doubly ionized acceptors.

- The **oxygen exchange reaction** with the ambient atmosphere

  \[
  O_O \rightleftharpoons \frac{1}{2} O_2(g) + V_O^{\ast} + 2e' \quad (pO_2)^{1/2} \cdot [V_O^{\ast}] \cdot n^2 = K_O = K_O^0 \exp(-W_O/k_B T) \quad (7)
  \]

  shows that the extraction of molecular oxygen leaves electrons behind. It should be noted that a similar equation can be formulated in which cation interstitials occur instead of oxygen vacancies.

- The **ionization reactions** of the defects, for example,

  \[
  V_O^{\ast} + e' \rightleftharpoons V_O' \quad [V_O'] \cdot n^{-1} \cdot [V_O^{\ast}]^{-1} = K_{V_O} = K_{V_O}^0 \exp(-W_{V_O}/k_B T) \quad (8)
  \]

  \[
  V_M^{\ast} + h' \rightleftharpoons V_M' \quad [V_M'] \cdot p^{-1} \cdot [V_M^{\ast}]^{-1} = K_{V_M} = K_{V_M}^0 \exp(-W_{V_M}/k_B T) \quad (9)
  \]

- Furthermore, there is the **electronic equilibrium** which controls the generation and recombination of electron–hole pairs and is mainly determined by the band gap \( W_G \)

  \[
  n \cdot p = K_e = K_e^0 \exp(-W_G/k_B T) \quad (10)
  \]

Please note that we do not include foreign dopants here, in order to keep this introduction simple. Foreign dopants will be mentioned wherever required in the following text.

Figure 14(a) illustrates the defect concentrations as a function of the oxygen partial pressure \( pO_2 \) at a high temperature at which all equilibria are active. As mentioned, such a defect diagram can be regarded as a magnification of a (near-stoichiometric) homogenous phase within a phase diagram (Figure 11). There are upper and lower boundaries on the \( pO_2 \) scale (not shown in Figure 14) at which the defect diagram is terminated because the homogenous, single-phase situation is left (boundaries shown in Figure 11) and additional phases need to be taken into account. For some oxides, only part of the defect diagram can be experimentally observed, for example, the regime in which the \( n \)-conduction dominates. Again, as a general trend, the \( n \)-conduction increases (\( p \)-conduction decrease) if the oxygen partial pressure is decreased. Transition metal oxides are of particular interest for this chapter. The light transition metal oxides show band gaps between 1.4 eV (CuO [143]) and approx. 5 eV (ZrO2). Oxides with a very high band gap of 8–9 eV, such as the main group oxides Al2O3 and SiO2, typically show extremely low electronic carrier concentrations and the oxides can be described as real insulators.
Any of these reactions (6) to (10) can only contribute to the composition of the phase actively, if the mobility (diffusivity) of the defects involved over the length scale of the specimen is sufficiently high. While this is the case for the equilibria (8) to (10) even for cryogenic temperatures because they can act by the motion of electronic defects whereas the ionic defects may stay immobile, the two other reactions become active only above a critical material dependent temperature because ionic defects must be mobile. So, for the equilibria (6) and (7) to get established one needs two processes. Firstly, since particle exchange across interfaces is involved, the thermally activated exchange rate must be high enough. Secondly, ionic defect must move within the crystal lattice and the motion of ions is always a thermally activated hopping conduction (see Section
3.2.3). Figure 14(a) reveals that (undoped) oxides have to be regarded as **mixed ionic electronic conductors (MIEC)** at high temperatures.

Figure 14(b) shows the defect concentrations for the same oxide as in Figure 14(a), however, after *quenching* the system to room temperature (at every point on the $pO_2$ axis). It is important to note that for most oxides, the native donors (here: $V_{O}^{••}$) have shallow energy levels, i.e. $W_{VO}$ is very close to the conduction band, while native acceptors (here: $V_{M}^{′′}$) have energy levels $W_{VM}$ which are deep in the band gap. As a consequence, for the $n > p$ regime, the oxide remains highly $n$-conducting after quenching. For the for $p > n$ regime, the $p$-conductivity drops by many orders of magnitude due to Equation (9), after quenching and an insulating oxide is obtained. This fact is essential for the resistive switching discussed later.

Instead of the Schottky disorder, the so-called **Frenkel disorder** is established at high temperatures in some oxides. This involves again cation vacancies as native acceptors, but cation interstitials as native donors – instead of the oxygen vacancies in the Schottky disorder.

$$M_{M} \rightleftharpoons V_{M}^{•} + M_{i}^{••} \quad [M_{i}^{••}] \cdot [V_{M}^{•}] = K_{F} = K_{F}^{0} \exp(-W_{F}/k_{B}T) \tag{11}$$

All fundamentals and the consequences for resistive switching can be described for the Frenkel disorder in the same manner as for the Schottky disorder. For instance, the oxygen exchange reaction will be:

$$O_{O} + M_{M} = \frac{1}{2}O_{2}(g) + M_{i}^{••} + 2e^{′} \quad (pO_{2})^{1/2} : [M_{i}^{••}] \cdot n^{2} = K_{O} = K_{O}^{0} \exp(-W_{O}/k_{B}T) \tag{12}$$

i.e. the law of mass action shows the same oxygen partial pressure dependencies except within the oxide cation interstitials are involved instead of oxygen vacancies.

3.3.2. **Acceptors, donors, and electronic compensation schemes**

As first mentioned in Chapter 2, the oxygen vacancies act as **mobile donor dopants** in $d^0$-transition metal oxide. Alternatively, also cation interstitials can serve as mobile donors. In the example of SrTiO$_3$ and TiO$_2$, oxygen vacancies are **shallow donors** (Figure 15) due to the high permittivity of the oxide. As a consequence, the electrons compensating the positive relative charge of the oxygen vacancy will be released to the $d^0$-band of the transition metal at $T = 300$ K. Consequently, the valence of the transition metal changes. Within this picture, the electron concentration $n$ in the conduction band is given by the oxygen vacancy concentration $c$ and the corresponding charge state $z$. Then, the oxide conductivity can be calculated by

$$\sigma = n\mu_{n} = zec_{V_{O}^{••}}\mu_{n} \tag{13}$$

where $\mu_{n}$ is the electron mobility. Considering a filamentary region, it is straightforward that the LRS would be represented by a relatively high homogeneous oxygen vacancy concentration along the filament axis in the first approximation. In contrast, the HRS results in an inhomogeneous distribution. When the oxygen vacancy concentration and in turn the electron concentration increases, the atomic orbitals of the reduced transition metal cations will overlap. According to the Mott-criteria, the former semiconducting oxide will turn into a metal. It should be noted that oxygen vacancies as mobile donors have a very different role than substitutional donors in SrTiO$_3$. Donor doped SrTiO$_3$ (n-STO) is obtained by a substitutional accommodation of higher valent cations with suitable ionic radii, such as Nb$^{5+}$ on Ti$^{4+}$ sites or La$^{3+}$ on Sr$^{2+}$ sites. The diffusivities of these donors are $> 10$ orders of magnitude smaller than that of oxygen vacancies even at 1000 K. Thus, substitutional donors can be considered immobile up to relatively high
Figure 15. Band diagram of a donor doped semiconductor showing the density of states DOS for the conduction band (CB), the donor states (D) and the valence band (VB), as well as the distribution function $f$ for a relatively high temperature. $n$ denotes the concentration of free electrons in the CB, while $p$ is the hole concentration in the VB.

Independent of the nature of the donors, the electron mobility in donor-doped SrTiO$_3$ is of polaronic nature showing moderate values (approx. 6 cm$^2$/Vs at $T = 300$ K) limited by phonon scattering [146–151]. There are other oxides such as HfO$_x$ and TaO$_x$ in which oxygen vacancy energy levels are deeper in the band gap and/or spread over a significant energy interval within the band gap. Electronically, this may lead to a sub-band conduction [152] (see Section 3.8 for more details).

With respect to the oxygen non-stoichiometry, it should be noted that there is a continuous transition between a thermodynamically controlled small concentration of oxygen vacancies and metastable situations with large oxygen deficiencies which are kinetically frozen-in and well beyond the solubility limits (see Section 3.1).

The relationship between the stoichiometry (i.e. a O-deficiency) and the electronic transport is discussed i.e. in [153–155].

### 3.4. Extended defects – dislocations and grain boundaries

In addition to point defects which can be regarded as 0D defects in the lattice, there is a hierarchy of higher dimensional defects. Dislocations are 1D defects, which can be classified into edge and screw dislocations. A regular sequence of edge dislocations forms a low angle grain boundary. Grain boundaries are 2D lattice defects and can be classified into coherent twin boundaries in which the ions in the grain boundary plane belong to the regular lattice on both sides, low angle grain boundaries, and large angle grain boundaries without symmetry relationships between the two adjacent lattices (Figure 16(a,b)). In highly pure materials, grain boundaries may be atomically sharp, e.g. without any secondary phase. However, in some materials a very thin disordered
Figure 16. Schematic sketch of internal boundaries in a perovskite lattice $\text{ABO}_3$. (a) large-angle grain boundary, (b) low-angle grain boundary, (c) anti-phase boundary.

(amorphous) phase is detected along large angle grain boundaries [136]. Point defects typically segregate at extended defects and change the stoichiometry in the defect region. That affects also impurities if they are present in the material. Depending on the extent of the segregation, this process may lead to a phase separation e.g. to the formation of secondary phases at grain boundaries. Typical planar defects for complex $\text{ABO}_3$ perovskite thin films are anti-phase boundaries (Figure 16(c)). They are formed either when the substrate is not singly terminated or when either the A or the B site atoms segregate to the surface during the thin film growth process. For $\text{SrTiO}_3$, the segregation of Sr during the growth results in the formation of both, vertical and horizontal antiphase boundaries [156].

Redox processes may also lead to the generation of extended defects of different dimensionality. The chemical reduction of $\text{TiO}_2$ first leads to the introduction of additional oxygen vacancies and/or Ti interstitials. Beyond a concentration of approx. 0.01 at% these point defects start to accumulate in random shear planes, known as Wadsley defects. Upon further reduction, the Wadsley defects order themselves and form Magnéli phases $\text{Ti}_n\text{O}_{2n-1}$ [157].

In the case of ionic solids, extended defects such as grain boundaries and dislocations are often charged. This leads to a space charge region adjacent to the extended defect. An example is shown in Figure 17. The space charge density $r$ in a cross-section of a large angle grain boundary in a slightly acceptor-doped (insulating) $\text{SrTiO}_3$ is provided by a finite-element simulation [158] and molecular dynamics simulation [159]. Low angle grain boundaries can be regarded as a chain of dislocations which show a very similar space charge effect in $\text{SrTiO}_3$ as large angle grain boundaries [160].

Due to the defect configuration shown in Figure 17, the positive GB charge also can attract electrons and lead to an enhanced n-conduction along dislocations and grain boundaries in $\text{SrTiO}_3$ and related band insulators [161].

Until now, we have only discussed the crystalline state. It should be mentioned that amorphous phases are always metastable, i.e. show a higher free energy than their crystalline counterpart (which may be a homogenous phase or a phase mixture). Therefore, they cannot be considered in equilibrium phase diagrams. It should be noted that amorphous compounds play a particularly important role in phase change memories.
Figure 17. Simulated spatial profiles along an axis perpendicular to the grain boundary of the space charge density \( r(x) \) for 0.2 at.% acceptor-doped SrTiO\(_3\) at a temperature \( T = 500 \) K with an interface density of positively charged donor-type grain boundary states of \( 2.5 \times 10^{14} \) cm\(^2\) [158]. These positive interface states are trapped, immobile oxygen vacancies. The negative space charge is formed by ionized acceptors.

3.5. Amorphous and nanocrystalline oxides

For many decades amorphous insulators and semiconductors are of high importance in gate dielectrics of MOS transistors and in Silicon-Oxide-Nitride-Oxide-Silicon (SONOS) flash memories. Amorphous silicon in which the dangling bonds are passivated by hydrogen (a-Si:H) is used for solar cells and for thin film transistors in LCDs. More recently, amorphous oxide semiconductors based on ternary or quaternary oxides of post-transition metals such as In–Sn–O, Zn–Sn–O, or In–Ga–Zn–O found technological use. Typically, amorphous oxide semiconductors are produced by physical vapor deposition or chemical vapor deposition (such as atomic layer deposition, ALD) at relatively low temperatures in order to avoid the crystallization of the material. The electronic structure of amorphous semiconductors has been described in e.g. Ref. [162–164]. The broad distribution of coordinations, bond angles and the lack of a long-range order results in tails at the conduction and valence band edges into the band gap (Figure 18). These tail states are localized states which occur in very high concentration. Because of this fact, heterovalent doping of amorphous semiconductors shows almost no effect on the electronic conductivity. It is more appropriate to describe the electronic structure by the overall stoichiometry which determines the Fermi energy in the material. In addition to band conduction, there is hopping conduction between the localized tail states, known as the impurity band conduction in highly doped crystalline semiconductors. In recent years, the means of investigation of amorphous semiconductors have greatly improved, on the experimental side by advanced spectroscopic methods and on the theoretical side by molecular dynamics and stochastic quenching [165,166].

The crystallographic structure and the electronic structure of nanocrystalline oxides lays in-between crystalline and amorphous oxides. In all cases, the grains are so small that any space charge layers at the grain boundaries of nanocrystalline oxides strongly overlap.
Figure 18. Schematic band diagram of an amorphous semiconductor according to Anderson [163] showing the conduction band and the valence band as well as the localized tail states. The boundaries between band and tail states are called the mobility edges. Adapted and reprinted with permission from [163], © 1975 by the American Physical Society.

3.6. Ionic transport by drift and diffusion

The random walk theory relates the self-diffusion coefficient \( D_i \) of an ion \( i \) to the drift velocity \( v_D \) by

\[
D_i = \frac{1}{2} a v_D = \frac{1}{2} a^2 \omega_0 \cdot \exp \left( -\frac{\Delta W_i}{k_B T} \right). \tag{14}
\]

Here, the so-called attempt frequency \( \omega_0 \) is given by the vibration of the ion in its potential well, and \( a \) is the hopping distance, i.e. the distance of the neighboring lattice site. Often, the pre-exponential term is condensed into an empirical factor \( D_{0,i} = \frac{1}{2} a^2 \omega_0 \). Please note that, unlike electrons and holes, the hopping distance of ions and atoms can never exceed this nearest-neighbor hopping distance, i.e. they cannot become hot in solids because of their mass and size. \( \Delta W_i \) denotes the activation energy of the hopping, the minimum energy barrier on the path of the hop. In oxides, activation energies \( \Delta W_i \) between approx. 0.3 eV (for highly ion conducting oxides) and 3 eV are observed.

In the following, we will discuss few more aspects using the diffusion of oxygen ions as the most important example. These ions move by a vacancy mechanism (Figure 19). We will first discuss the self-diffusion coefficient \( D_O \) of oxygen ions. Within very good approximation, these can be determined by isotope tracer experiments, which reveal the tracer diffusion coefficients. In order for an oxygen ion (i.e. a tracer ion) to jump, there must be a vacancy at a nearest neighbor site. The probability of an oxygen vacancy on a nearest neighbor site of an oxygen ion under consideration is

\[
p_V = \frac{N_V}{N_O} \tag{15}
\]

where \( N_V \) and \( N_O \) are the concentrations of oxygen vacancies and of the total number of oxygen sites, respectively, in the lattice. For not too high vacancy concentrations (i.e. up to few percent), \( N_O \) is equal to the number of occupied oxygen sites in good approximation. Under this assumption, a vacancy will always find a partner to jump. As a consequence, the relation of the
self-diffusion coefficient oxygen ions $D_O$ and that of the oxygen vacancies $D_V$ is given by:

$$D_O = \left(\frac{N_V}{N_O}\right)D_V$$

(16)

where $D_V \gg D_O$ and $D_O$ is determined by the concentration of oxygen vacancies, while $D_V$ is independent of this concentration. This relation explains, for example, the orders of magnitude difference in oxidation and reduction rates of acceptor- and donor-doped titanates [167].

In the next step, we have to take into account that charged species (ions and vacancies) cannot diffuse independently of other charges in a concentration gradient (i.e. in a so-called chemical diffusion experiment), because otherwise the charge neutrality would be violated. In other words, a chemical diffusion of ions is always an ambipolar diffusion in which the ambipolar diffusion coefficient $D*$ is determined by the self-diffusion coefficients and the concentrations (for details see [135,136]). However, in a chemical reduction experiment of oxides, the diffusion of oxygen vacancies will be accompanied by the diffusion of approx. the same number of electrons ($n \sim N_V$). In this case, $D*$ will be determined by the diffusion coefficient of the slower species, e.g. $D* \sim D_V$.

In the case of the vacancy mechanism, the oxygen ion conductivity is given by

$$\sigma_O = 2e\mu V N_V$$

(17)
where $\mu_V$ denotes the electrical mobility of the oxygen vacancies and the pre-factor 2 stems from the double charge of $V_o^\bullet$. The mobility and the self-diffusion coefficient are related by the Nernst-Einstein equation:

$$\mu_V = \frac{ze}{k_B T} D_V. \quad (18)$$

The ion mobility becomes field-dependent at high fields by considering a field-lowering of the activation energy for ion hopping $D_{W_i}$. This leads to the so-called Mott-Gurney law [168] (cf. Equation (45) in section 5.7.2) which has been refined by Genreith-Schriever and De Souza [169].

Please note that no ambipolar effects have to be taken into account in the ion transport by a field-induced drift in a region of constant concentrations in contrast to the transport by chemical diffusion in a concentration gradient.

Lastly, temperature gradients could also lead to a movement on the ionic defects, also called \textit{thermo-diffusion}, \textit{Soret effect} or \textit{thermophoresis}. Temperature gradients can occur in VCM cells due to local Joule heating. As the metal electrodes serve as a heat sink, temperature gradients are expected in the vertical direction. Moreover, in filamentary systems temperature gradients are expected in the lateral direction. The thermo-diffusion coefficient $D_T$ is coupled to the self-diffusion coefficient according to

$$D_T = S_T D_V \quad (19)$$

where $S_T$ is the Soret coefficient. It is often assumed that the Soret coefficient is positive. Thus, positive ions would move towards the hot spot. From theory, however, the situation is more complicated and the sign of the Soret coefficient is not strictly defined [170]. Up to now, we have discussed the motion of ions in a regular (so-called \textit{bulk}) lattice.

If extended defects are present, these might significantly impact the over-all transport of ions. In a first approximation, it is useful to consider the regions of the extended defects empirically as a different phase, with different diffusion coefficients and activation energies of the ions. Sometimes this phase must be further subdivided into the core of the extended defect (e.g. core of the dislocation, core of the grain boundary) and a possible (depletion or enrichment) space charge region surrounding it. Depending on the connectivity of these phases and on the transport properties of the bulk lattice ‘phase’ as well as the ‘phase’ represented by the extended defects, the overall result on the average ion transport is extremely different (as sketched in Figure 20). For details, the reader is referred to e.g. [158,167,171].

The same first approximation can be applied to a material composed of crystalline and amorphous regions. In the glass phase itself, the ion transport can be described in a similar way as for crystals, i.e. ions jump to available nearest neighbor sites. Still, the picture must be modified because the bond angle and lattice distances of the nearest neighbor positions show a certain scatter due to the amorphous nature. As a consequence, there will also be activation energies and jump distances which show a certain scatter around mean values.

In order to discuss the relevance of extended defects for the resistive switching process, one needs to consider the specific material and its microstructure because the situation may vary between the extreme cases shown in Figure 20. A comprehensive treatment of the ion transport in conjunction with dislocations is given by De Souza [173]. In addition, the electronic conduction must be considered since this may lead to temperature increase which, in turn, affects the ion diffusion. An impressive example of this effect is discussed in Ref. [174].

With respect to antiphase boundaries in SrTiO$_3$ thin films, it has been shown that both the formation energy for Sr vacancies and that the energy barrier for Sr diffusion is strongly reduced [175]. Therefore, Sr diffusion is strongly enhanced along antiphase boundaries.
Figure 20. Sketch of the O diffusion in a simplified oxide ceramics. A moderate O diffusion coefficient in the grain boundaries is assumed. (a) If the bulk lattice shows very high O diffusion, the grain boundaries act as barriers impeding the O transport. This is found, for instance, in highly acceptor-doped ZrO$_2$ used in solid-oxide fuel cells and in acceptor-doped SrTiO$_3$ model systems [158,172]. (b) If the bulk lattice shows very low O diffusion, the grain boundaries may act, relatively seen, as fast tracks for O transport [167,171]. On the absolute scale the average O transport is, however, much slower than in (a). Please note that this picture is strongly simplified, e.g. space charge effects, the realistic microstructure, etc. are not taken into account.

3.7. Redox processes

A redox process denotes a coupled reduction and oxidation reaction, i.e. an electron transfer reaction where the reduction is the uptake of electrons and the oxidation is the release of electrons by atoms. Often, the atoms involved completely lose or win one or more electrons in their outermost electron shell, i.e. there is an integer change in their valence. This is clearly the case for localized electrons, for example, in metal ions in aqueous solutions or in the ions of an ionic solid. The picture of integer valence changes still holds when covalent bond contributions appear. However, as soon as electron delocalization by metallic bond contributions has to be taken into account, such as in non-stoichiometric compounds with delocalized electrons, redox reactions may act on a large ensemble of electrons and the result of the redox process may be a minor change in the Fermi energy level and/or the electronic band structure. Formal valence changes (per atom) would be fractional numbers in this case.

The mere transfer of electrons such as the charge injection of electrons into the floating gate of a Flash transistor or the trapping/detrapping of electrons by dopant centers in a semiconductor can, in principle, also be considered as redox reactions. However, in all practical cases redox processes are accompanied by transport of atoms or ions. For example, a simple redox process is the rusting of iron in air where solid Fe and O$_2$ gas turns into solid Fe$_2$O$_3$. Another example, which involves only solid phases and which is closer to what we need to discuss in this chapter, is the reaction of a mixture of Al metal and Fe$_2$O$_3$ particles:

$$2\text{Al}^0 + \text{Fe}^{3+} + 3\text{O}_2^{2-} \rightarrow \text{Al}^{3+} + 3\text{O}_2^{2-} + 2\text{Fe}^0 \quad (20)$$

where the formal valences are noted as superscript at the element symbols. Since only stoichiometric compounds are involved in this example, these valences are (small) integers. Please note that the oxygen ions do not take part in the electron exchange, i.e. they keep their valence $-2$, although their transport is essential for the redox reaction.
The oxidation of a non-noble electrode such as Ti by the in-diffusion of oxygen ions is a very common redox reaction which occurs during the electroforming with a positive bias at that electrode:

$$\text{TiO}_2(\text{MIEC}) + \text{Ti(electrode)} \rightarrow \text{TiO}_{2-x}(\text{MIEC}) + \text{TiO}_x(\text{electrode})$$  \hspace{1cm} (21)

$\text{TiO}_x$ indicates that oxygen has partially oxidized the Ti electrode. TiO$_x$ is electronically conducting.

Electrochemical reactions are redox processes in which the electron transfer of the oxidation and the reduction are separated by an external electron conductor. In electrochemical metallization memory (ECM) cells for example, the electrochemical oxidation of Ag, the transport of Ag$^+$ through an electrolyte, and the electrochemical reduction of Ag$^+$ ions back to Ag metal at the counter electrode is an example of a simple electrochemical redox process.

In all types of resistive switching memories described in this chapter, the redox reactions must be coupled to the transport of ions (typically over lengths of few nanometers only) – this is why the mechanisms are sometimes summarized as **nanoionic redox processes**.

Another aspect is the charge neutrality, which has to be considered for all redox reactions. Despite the fact that redox processes are electron transfer reactions, the system involved may not accumulate charges, i.e. it must maintain its (overall) charge neutrality. In electrochemical cells, this typically involves the transport of counter ions which themselves do not take part in the electron transfer reaction. On the nanoscale, within the spatial extension of space charge regions, this rule may be violated. The consequences for resistive switching cells will be discussed in the subsequent sections.

### 3.8. Electronic transport – mobilities and correlation effects

The electrical conduction through a ReRAM MIM cell is, in general, determined by the conductance of the MI interfaces and by the conductance of the bulk of the resistive I-layer. The ionic contribution to the bulk conductance has been dealt with in Section 3.6. Here we will discuss the electronic contribution. If an electron is injected from the cathode into the conduction band of the oxide, it will drift within the oxide before it is ejected into the anode. As these transport mechanisms occur in series, the current flow will be dominated by the most limiting (resistive) one. As the transport across the interfaces is polarity-dependent an asymmetric $I$–$V$ characteristic with respect to voltage polarity is expected. This asymmetry can vary from very asymmetric cases if only one interface is limiting and thermionic emission dominates to only slightly asymmetric cases if the electron transport through the barriers is dominated by electron tunneling.

The combined influence of the interfaces and the bulk on the total conductance of a MIM cell has to be considered. In the literature, ReRAM cells have sometimes been discussed in the same way as electron transport through the gate dielectrics of Flash memories. The major difference, however, is the orders of magnitude higher conductance level in ReRAM cells, in particular in the ON state (LRS). A brief estimation shows that a read access time of $< 100 \text{ ns}$ requires a read current $I_{rd} > 100 \text{ nA}$ in the ON state (LRS), if the bit line capacitance $C_{BL}$ is in the order of 30 fF and the read voltage is 0.3 V (see: Introduction to Part V, in [176]). For a cell cross section of $30 \times 30 \text{ nm}^2$, this translates into high current densities of $> 10^4 \text{ A/cm}^2$ which the ReRAM cell must be able to support.

In order to discuss how these current densities can be achieved in MIM cells, we will briefly describe the mechanisms of electronic conduction in the bulk of solids, before we include the electrode interfaces.

More details to this topic are found in textbooks of solid state physics and, for example, in Ref. [177].
Figure 21. Illustration of the different conduction mechanisms in the bulk of the oxide. The red bars in the band gap show defect-induced trap states. The electrons can be transported through the oxide by (i) phonon-assisted trap-assisted tunneling, (ii) nearest-neighbor hopping, (iii) variable range hopping, or a (iv) sub-band transport. (Color figure available online).

In metals and doped (inorganic) semiconductors including some oxides such as ZnO, we encountered band conduction characterized by a strong orbital overlap, a broad, partially filled conduction band and, as a consequence, highly delocalized electrons with mobilities of $10^2 \ldots 10^4$ cm$^2$/Vs. Their mean free path lengths are between several nanometers and macroscopic distances depending on the density of scattering centers (including phonons). If the orbital overlap is smaller, there is a significant interaction between the electron and the lattice, i.e. the electronic carrier polarizes the lattice in its vicinity. The entity of the electronic charge and the distorted lattice is called a large polaron, which shows mobilities of $0.1 \ldots 10$ cm$^2$/Vs with no or only very weak temperature dependencies. If electrons are localized, they are called small polarons, show mobilities $< 10^{-2}$ cm$^2$/Vs, and move by thermally activated hopping. As discussed in Section 3.3 defect states due extrinsic impurities or intrinsic nonstoichiometry are present in oxide materials. These defects may form localized defect states within the band gap enabling an electron hopping transport via these defect levels as illustrated in Figure 21. This defect state need to be deep-level in order to ignore band transport. In contrast to polaron transport, the trapping potential landscape is static. It is important to note that electron hopping has to occur from a filled site to an empty site. Thus, the Fermi level must lie within the range of the defect energy levels. The average hopping distance depends on temperature and it can be larger than the average distance between a pair of neighboring defect sites. Therefore, this mechanism is also called variable range hopping, (ii) and (iii). In order to consider the defects as localized neighboring defects, they should be far enough apart. If they are close to together (high defect concentration), sub-bands within the band gap might form. The electron transport could then occur via half-filled sub-bands (iv).

In addition to this very coarse classification, anisotropy effects as well as various types of correlation effects have to be taken into account, depending on the type of resistive materials. For ReRAM, we are interested in resistive materials that will show in the reduced state relatively high electron mobilities and high electron concentrations in order to fulfill the requirements mentioned above. Because of the small dimensions of ReRAM cells, mesoscopic effects such as ballistic
Figure 22. Three prototypical barriers for the electronic transport shown by a material composition and an energy $W(x)$ profile. (a) A neutral barrier built by the intrinsic region of $n$-doped semiconductor. (b) A space charge depletion barrier at the contact of a $n$-type semiconductor and a high-workfunction metal, (c) a tunneling barrier by a high-bandgap tunnel oxide (TO) between a $n$-type semiconductor and a metal.

transport, scattering at the interfaces, and Landauer conductance levels have to be considered in a more detailed analysis (see Chapter 3 by T. Heinzel in [176]).

3.9. Types of barriers and their modulation – the fundament of resistive switching

In order to understand resistive switching in general, and VCM-type switching in particular, it is important to note that there must be a barrier for the electronic transport in the ‘I’-layer of the MIM stack. In a simplified view, we can distinguish three types of barriers which may play a role in resistive switching:

(a) Neutral barriers
(b) Space-charge depletion barriers, and
(c) Tunneling barriers.

A neutral barrier is formed, for example, in a semiconductor with a $n$-$i$-$n$ doping profile, i.e. a donor-doped region on the left, an undoped intrinsic region in the middle and another donor-doped region on the right (Figure 22(a)). Obviously, the intrinsic region (which may also be just a more lightly doped region) represents a neutral barrier for the electronic transport. If you assume that the donor dopants on one side are mobile (e.g. oxygen vacancies or cation interstitials) and on the other side they are fixed (e.g. substitutional donors) then you can decrease and increase the width of the intrinsic region by the application of an electric field and, hence, control the resistance. It should be mentioned that the formation of a $pn$-junction in a dielectric oxide due to a concentration polarization of oxygen vacancies was also explained by a neutral barrier which was diminished in the course of the resistance degradation process [178]. The modulation of neutral barriers leads to the bulk switching mentioned in Chapter 2 and described in more detail in Chapter 6.

A space charge depletion barrier is typically formed at electrode contacts if the work function of the metal is larger than the electron affinity of the oxide (Figure 22(b)). Both types of barriers, neutral barriers and space charge depletion barriers, are overcome by thermally excited electrons which is called a thermionic emission.

A tunneling barrier (Figure 22(c)) occurs if the barrier thickness is small enough so that electron tunneling (tunneling emission) will occur. Modulation takes place by thickness or by charging the tunnel junction, i.e. changing the energetic height of the barrier. The tunneling barrier
itself may be either neutral or it may be charged [66]. Typically it consists of a higher bandgap material, e.g. a tunnel oxide (TO).

Of course, these three types of barriers are prototypical, limiting cases. In reality, they hardly appear in a pure form. For example, a space charge depletion barrier may get thin enough so that tunneling occurs too. In general, in the case of a barrier formed at the interface between an oxide and a high-workfunction metal we will speak of Schottky barrier, comprising both a space charge depletion barrier effect and a tunneling barrier effect, as will be explained in more detail in Section 3.10. And in the case of a neutral barrier, any change of a dopant concentration will lead to space charge formation and, hence, possibly to a contribution by a space charge depletion barrier.

3.10. Electrode contacts

3.10.1. Formation of Schottky barriers and ohmic contacts

The conductance of the interface between a metal electrode and the resistive layer is determined by the work function of the metal, charged interface states, and the electron affinity of the resistive layer. It may range from pronounced barriers due to space charge depletion effects, very thin barriers which can be overcome by direct tunneling, ohmic contacts, and galvanic contacts if metal filaments are formed in the resistive layer.

The differences in the work functions of the metal electrode and the oxide layer gives rise to the formation of an electrostatic barrier. The nominal barrier height $e\phi_B$ is given by

$$e\phi_B = W_{F,m} - e\chi,$$

where $W_{F,m}$ is the work function of the metal electrode and $\chi$ is the electron affinity. If a metal–oxide–metal structure is considered, the difference in the work function of the two metal electrodes will lead to an internal electric field within the oxide material. Without any additional space charge within the oxide this electric field will be constant. According to Section 3.3 charged ionic defects will be present, e.g. doubly charged oxygen vacancies compensated by two free electrons in the conduction band. In this case, a band bending will occur according to Poisson’s equation and an electron depletion zone will form close to the metal electrode. The thickness of this depletion zone depends on the concentration of charged ionic defects $c$ according to

$$d_{\text{depl}} = (2\varepsilon\phi_s/(ec))^{1/2}$$

where $z$ is the charge number and $\varepsilon$ is the oxide permittivity and the surface potential $\phi_s$. For a very high concentration $c$ of ionic defects, the depletion zone is very small and a very narrow electrostatic barrier results enabling tunnel injection from the metal electrode to the oxide conduction band.

An example of an electrostatic potential profile is shown for Pt/Fe: SrTiO$_3$/ Nb: SrTiO$_3$ MIM structures using electron holography in combination with numerical drift-diffusion simulations [179] (Figure 23). The measured potential profile does not only confirm the Pt electrode as the electrically active electrode but also show a penetration of the electric field into the top few nanometers of the Nb: SrTiO$_3$ bottom electrode, presumably because of acceptor-type Sr vacancies at the Nb: SrTiO$_3$ substrate interface.

It should be noted that in some cases, the character of an electrode may change depending on its stoichiometry. This is typically encountered for TiN. In its fully stoichiometric form, this is a metal with a relatively high work function and a relatively low oxygen affinity so that it can be used as an electronically active electrode (AE). On the other hand, if TiN is deficient in N, i.e. TiN$_{x<1}$ (which may happen due to the synthesis conditions) it turns into an ohmic electrode (OE) because of the surplus of Ti in the TiN$_{x<1}$ material.
Advances in Physics

3.10.2. Current transport across the barrier

In order to inject electrons into the oxide material the interfacial barrier needs to be overcome as illustrated in Figure 24. It shows a situation in-between Figure 22(b) and Figure 22(c). At temperatures \( T > 0 \) K thermally excited electrons could surpass the barrier via thermionic emission (TE) over the barrier. If the barrier is very narrow electrons can surpass the barrier by tunneling into the conduction band. Here, it is distinguished between field emission (FE), also called Fowler–Nordheim tunneling, and thermionic field emission (TFE). In the former case, mainly electrons close to the Fermi level in the metal electrode tunnel, whereas for TFE the main current contribution is from thermally excited electrons tunneling at an energy higher than the Fermi level in the metal electrode. It depends on the applied voltage magnitude and the voltage polarity which mechanism dominates. In forward direction of the metal/oxide junction, thermionic emission gets more probable. In contrast, the dominant mechanism might change from TE over TFE to FE for increasing voltage magnitude under reverse bias. The discrimination between TE, TFE and FE is very coarse. In reality, the current transport is a function of the density of states in the metal and the oxide material in form of the supply function \( N(W) \) and the transmission probability \( T(W) \) at a certain energy according to

\[
J = \frac{A^*}{k_B} \int_{-\infty}^{\infty} N(W)T(W) \, dW
\]  

(22)

with \( A^* \) being the effective Richardson constant. In case of thermionic emission, the transmission probability would be one. For the tunneling regime, the transmission probability is often simplified using the Wentzel-Kramers-Brioullin WKB approximation. The effective barrier height for electrons is not identical to the nominal barrier height. In the presence of an electric field at the metal/oxide interface, the image force of the approaching electron leads to a barrier lowering (also known as Schottky barrier lowering). The electric field at the interface is particularly high in the presence of a space charge. Considering oxygen vacancies as prevalent ionic defect, this means that a high concentration of oxygen vacancies close to the metal/oxide interface not only leads to thinning of the barrier but also to barrier lowering. The highest current is thus expected
Figure 24. Illustration of the barrier-related transport across a metal semiconductor interface. This is a typically Schottky-type metal-semiconductor interface biased in forward direction (a) and backward direction (b). Here a degenerate donor-doped semiconductor is considered. $W_F$ denotes the Fermi level, $W_C$ is the conduction band edge in the semiconductor, $V_F$ ($V_R$) is the absolute voltage applied in forward (reverse) direction of the interface, and $W_m$ describes the energy level at which the main electron transport takes place.

if the oxygen vacancy concentration is high close to the interface. In an extreme case, the barrier is so small and thin that the electron injection does not limit the current anymore. A so-called Ohmic contact is formed. Besides injection into the conduction band also injection into/from empty/filled electron trap sites is conceivable. Another possibility is trap-assisted tunneling into the conduction band via a defect state within the band gap. In both cases, the Fermi level needs to lie in the range of the defect energy levels.

For a resistive switching system, in general, we need to consider the contribution of both, the potential barriers and the bulk electron transport. The limiting (most resistive one) will determine the overall device resistance as these transport mechanisms will occur in series. During switching the contribution may also change. In the LRS state, for example, the barrier might be so thin that the overall current is dominated by the bulk contribution. In the HRS state, however, the barrier is re-established and limits the current transport. This picture is consistent with the VCM switching model shown in Figure 7.

4. Electroforming processes

The majority of metal oxides used in redox-based resistive switching cells are highly insulating materials in their pristine state and a so-called electroforming process is needed to transform them into a reversible switchable material with increased electronic conductivity. There are a few exceptions such as p- or n-doped conducting oxides which often show area-dependent switching to be described in Section 5.5 and forming-free systems which will be discussed in Section 4.5. Electroforming of insulating oxides is somehow similar to (controlled) dielectric break-down processes induced by the random generation of defects during strong electrical biasing. In particular, electroforming can be regarded as a soft-breakdown process, in the sense that the breakdown process is limited and can, at least, be partly reversed by an electrical stimulus. Dielectric breakdown in SiO$_2$ or other gate oxides is an undesired event which causes failure in CMOS circuits and has been studied in detail in the framework of CMOS reliability [180–183]. These references
describe a process in which the electrons get hot in the electric field and cause a bond-breaking at the anode leading to a tree-formation which grows towards the cathode. Electron injection and transport by the different mechanisms are described in Chapter 3. In other words, defects are created either at the anode, as mentioned, or (less likely) according to other references at arbitrary positions within the dielectric, resulting in a breakdown when the defects form a percolation path that short-cuts the two electrodes [184–188]. Based on the similarity to dielectric breakdown phenomena, electroforming has often been described by models restricted to the field-induced creation of lattice defects at arbitrary positions [189–194]. However, for transition metal oxides besides electron injection, ionic processes (described in detail in Chapter 3) such as ionic motion, redox-reactions and anodic oxidation processes, often mediated by Joule heating and thermal runaways, have to be considered to describe the changes of the conductivity during electrical biasing. The differences between the two forming models (creation of defects only at the anode vs. creation at arbitrary positions) have been discussed in detail in [195].

In this chapter, we start with describing typical electroforming procedures used in the community (Section 4.1) as well as the experimental findings about the changes in the electrical conductivity, chemical changes and structural changes taking place during electroforming (Section 4.2). Afterwards, we will describe the status of understanding of the microscopic mechanisms of the forming process, starting with the electronic effects initializing the electroforming (Section 4.3.1), followed by redox reactions at the electrode interfaces (Section 4.3.2) and the ionic concentration polarization (Section 4.3.3). We will then discuss the two different growth directions of the filament during electroforming (Section 4.4) and outline ways towards forming-free systems (Section 4.5). Most of the physical modeling and simulation is postponed until Chapter 6 because the resistive switching is more in the focus of this review paper and the switching processes are in parts similar to the electroforming processes.

4.1. Typical procedures for electroforming

In order to transform as-prepared insulating MIM cells into a reversible switchable state, certain electrical treatments are necessary which might vary strongly from system to system. For single crystals and thick thin films, the application of a dc-voltage over a considerable time is required before a steep increase of the leakage current takes place as sketched in Figure 25(a) [119,196,197]. For many VCM-type oxide thin film devices, electroforming takes place during the first cycle of the current–voltage ($I$–$V$) curve. As depicted in Figure 25(b), the first branch of the $I$–$V$ curve differs significantly from the reversible subsequent branches hinting on an irreversible forming process during the first biasing step. In order to study the dynamics of the process, some research groups have conducted the electroforming in a pulse mode, by applying rectangular voltage pulses with different pulse height or length [198–201]. As an alternative to the voltage-controlled mode, forming can also be performed in a current-controlled mode, as explained in [202,203].

It is important to note that electroforming might also go along with a deterioration of the oxide and/or the metal electrodes [69,199,205,206]. In order to prevent the deterioration or even an irreversible breakdown of the cells during voltage controlled forming, the current flow has to be carefully limited by a current compliance. Since the current compliance of conventional voltage sources is often too slow to limit the current during fast thermal runaways, which might take place during electroforming, uncontrolled and severe current overshoots may occur [207]. This problem may be solved by an ultrafast current compliance [208,209]. Alternatively, the current can be limited by a series resistor implemented on the chip or at the measurement tips. In particular, in some stack configurations such as Ta$_2$O$_5$/Ta cells, internal series resistances act as internal current limiters [210,211].
Figure 25. (a) Current–time dependence of a 750 nm thick Fe-doped STO thin film cell which shows a forming step after around 40 s during the application of a $+10 \text{ V}$ dc-bias [196]. (b) $I$–$V$ characteristics of TiN/HfO$_x$/TiO$_x$/TiN cell showing the initial forming cycle and a subsequent RESET/SET cycle. Here, the forming has been conducting into the LRS, limited by a current compliance (cc). Adapted and reprinted with permission from [204], © 2011 by IOP Publish. Ld.

Figure 26. Impact of film thickness and crystallinity on the area scaling of the forming voltage of HfO$_x$ cells [212].

The most controllable way is to use a series connection with a MOS transistor very close to the memristive cell, to avoid large parasitic capacitances, so called 1T-1R structures, where the maximum current during forming and SET is maintained by the saturated drain current adjusted by the applied gate voltage [118].

The forming voltage usually decreases with the oxide film thickness [69,213] and increases with cell size (see Figure 26) [212,214]. In addition, the forming voltage for polycrystalline films was found to increase stronger with decreasing cell size than in amorphous films, which is likely due to the more homogeneously distributed ‘defects’ in the amorphous films on the nanoscale. When employing a current-controlled forming procedure, it has been observed that the forming current is proportional to the linewidth of the cross-bar cells implying an increased current flow along the perimeter of the cell [203].

Furthermore, a large variety of investigations prove that both, point defects as well as extended defects have a key influence on the forming ability of thin film devices. In particular,
systematic variations of the metal/oxygen content in HfO$_x$ [215,216] and Ta$_2$O$_{5-x}$ [217–219] thin films indicate a decrease of the forming voltage with increasing O deficiency. Furthermore, grain boundaries have been identified as preferential forming sites [78,220] and polycrystalline thin films exhibit similar forming voltages as amorphous thin films with the same thickness (see Figure 26). This is consistent with the observation that grain boundaries provide leakage paths and promote dielectric breakdown in HfO$_2$ thin films [221,222], see Chapter 3. Single crystalline thin films are generally much more difficult to form than polycrystalline thin films and extended defects such as vertical stacking faults [90] and misoriented thin film regions formed at the existence of substrate screw dislocations [223,224] act as preferential forming sites. A variety of studies furthermore show a decrease of the forming voltage with increasing electrode roughness which is attributed to a local enhancement of the electric field and an increased defect density [225,226]. Moreover, different approaches have been used to pinpoint the filament position by modifications of the material stack that locally enhance the electric field such as embedding nanoparticles into the electrode [227] or the dielectric film [228,229], or by employing pyramidal-shaped electrodes [230]. Moreover, methods to confine the ionic motion by providing holes in an oxygen-blocking graphene interlayer [231] and by fabricating nanocavities along misfit dislocations [232] have been employed to guide the ionic filament formation process.

As mentioned in Chapter 2, MIM cells usually contain one high work function metal electrode such as Pt, Ir, Au (and TiN if in the fully stoichiometric phase) providing a Schottky-contact (called the \textbf{electronically active electrode} AE in this paper) and one oxidizable metal electrode with low work function providing an ohmic-like contact (called the \textbf{ohmic electrode} OE) assuming an n-type oxide. At the oxidizable metal electrode, such as Ti, Ta, W, Hf, and others, a redox-process takes place, resulting in the formation of a metal oxide interface layer and an increase of the oxygen vacancy concentration in the oxide layer, as described in Section 3.7. As mentioned above TiN is often produced in research labs as a substoichiometric nitride. In this case, it typically acts as an ohmic electrode because of the reaction of the excess Ti. In industry, the processing is well controlled and TiN is typically used in its stoichiometric composition as the active electrode.

Because of this redox-process, the forming voltage depends on the thickness relation between oxide and metal layer thickness [212,233,234] and will be discussed in detail later in this chapter in the context of forming-free devices.

The forming process has a key impact on the subsequent switching properties such as SET and RESET characteristics [235–238], and device variability [239]. A variety of forming protocols have been introduced in the literature in order to optimize certain device properties [199,240–242]. Furthermore, forming in different gas atmospheres [202] and at elevated temperatures [243] has been investigated. In particular, forming of HfO$_x$ cells at elevated temperatures reduces the forming time and voltage and improves the memory window [243].

For \textit{symmetric} Pt/TiO$_2$/Pt cells, it has been shown that the forming procedure induces a pronounced asymmetry of the two oxide-electrode interfaces resulting in strongly asymmetric $I$–$V$ curves [203,244]. Furthermore, the polarity of the forming voltage determines the switching polarity of the subsequent switching process [202].

It is very important to note, that it is essential if the process is driven by a field effect or a temperature effect or a combination of both. This aspect will be discussed later in detail.

\section{Microscopic changes induced by electroforming}
\subsection{Conductivity changes}
Since breakdown phenomena have in general a filamentary character, it is likely that electroforming might also be locally confined. The simplest approach to distinguish between filamentary and
homogeneous forming is to investigate the scaling of the cell current with the area after the forming step. The independence of HRS and LRS on the device area can be regarded as a proof for filamentary switching [10]. However, one has to keep in mind that the HRS resistance is often determined rather by the leakage current of the cells than by the filament resistance and might therefore scale with the cell area even in the presence of filaments. Furthermore, internal series resistances that limit the cell current might even result in an area-scaling of the LRS resistance. As alternative methods to prove filamentary forming, cells have been divided into two pieces after electroforming [245], or parallel connected cells have been separated after electroforming [196] in order to demonstrate that forming is locally confined. In this section, we describe different methods which have been employed to investigate the conductivity in VCM cells after forming in a spatially resolved way.

Imaging of conducting filaments in VCM devices by scanning probe techniques such as conducting atomic force microscopy (LC-AFM) is hindered by the short-cut of filaments by the top electrode. This approach has been employed for the detection of metallic filaments in ECM devices [246] but it is unsuitable for VCM cells due to the much lower conductivity of the filaments with respect to the top electrode. Possible approaches to circumvent the short-cut by the top electrodes is to use the moveable conductive tip as nanoscale top electrode [78] or to peel-off the top electrode prior to the LC-AFM investigation [70,247]. One powerful alternative technique to identify local changes of the resistivity underneath the electrode that does not require a delamination step is the use of pressure-modulated conductive tip AFM (PMCM) [69], which is illustrated in Figure 27(a) [248]. Figure 27(b) shows the resistance map of a TaOx-based cell indicating the formation of one single conducting filament underneath the electrode.

Alternatively, resistive contrast imaging in a scanning electron microscope, namely, electron beam induced current (EBIC) imaging or electron beam absorbed current (EBAC) imaging, which are prominent techniques for the failure analysis in the semiconductor industry, have been employed for the detection of filaments in memristive oxide devices [89,249–251]. By employing EBAC imaging, filaments in the order of 10 nm size have been detected in 0.5 × 0.5 μm² Ir/Ta2O5/TaOx/TaN cells (Figure 28(b)) [250]. It is important to note, that a nearly linear decrease of the filament size with the forming voltage could be detected by using this technique as depicted in Figure 28.
Figure 28. EBAC images of 0.5 × 0.5 μm² Ir/Ta₂O₅/TaOₓ/TaN cells for HRS with large filament (a) and LRS with small filament (b) produced by high and low forming voltage, respectively [250] (c) Dependence of the filament size on the current compliance of TaOₓ cells determined by electron beam electron-beam absorbed current imaging [250]. Adapted and reprinted with permission from [250], © 2011 by IEEE.

Depending on the specific material stack system, the device dimensions and most importantly the exact biasing procedure during forming, 2D current imaging techniques have visualized filament sizes ranging from several μm [249,252,253], over several hundreds of nm [254] down to the 10 nanometer scale [250]. In many cases, the conductivity distribution within the filament is not homogeneous, hinting on a complex substructure of the filament or a conglomeration of several smaller filaments.

The above described 2D imaging techniques provide only information about a mean cross-section of the conductivity of formed devices. However, the exact filament shape will play a crucial role in the subsequent switching properties. A tomography technique based on LC-AFM provides 3D filament visualization by scratching through the oxide layer while simultaneously mapping the conductivity [255–257]. Based on this technique, 3D filaments have been visualized in SiOₓ thin films addressed by a μm size tungsten tip [257]. The detected filament shape ranges from tubular to conical shape depending on the distance to existing grain boundaries promoting filament growth. Furthermore, these investigations suggest that many partial filaments are produced in the early stage that finally converge to a larger single filament [257]. A precise control of the forming current by a selector transistor in nanosized Hf/HfO₂/TiN crossbar memory cells succeeded in the formation of filaments with conical shape, visualized by the above-mentioned tomography technique (Figure 29(a)), shrinking from 39 nm² at the top Hf layer to 8 nm² at the bottom TiN electrode (Figure 29(b)) [256]. As can be seen in Figure 29(c), the filament size can be reduced by the reduction of the programming current.

4.2.2. Chemical changes

In order to investigate if the local changes in the conductivity induced during electroforming can be attributed to chemical changes such as the formation of oxygen vacancies and the corresponding change of the metal valence state (see Chapter 3), a variety of spectroscopic investigations have been performed over the last decade [258]. However, the direct detection of changes in the oxygen concentration within conducting filaments is strongly impeded by their small spatial extension, which requires spatially resolved spectroscopic techniques. Furthermore, X-ray spectroscopic techniques based on the detection of electrons such as photoelectrons or secondary electrons are generally surface sensitive [258] and can therefore not be applied to detect changes within the metal oxide underneath metal electrodes of usual thickness. Besides this, tiny changes
Figure 29. Tomographic reconstruction of filaments in 800 nm × 200 nm sized Hf/(5 nm)HfO2/TiN cross-bar memory cells using the scapel AFM method. (a) 2D current map using LC-AFM showing a zoom into the cell region containing the filament after scratching away part of the stack. The lower figure shows the current tomography of the device stack that is compiled by 2D LC-AFM scans after removing layer by layer using the scalpel AFM. The surfaces are iso-current surfaces. The blue color illustrates the constriction of the filament (scale bar 2 nm); (b) 2D current maps at two different locations z along the stack (left panels) and C-AFM linescans (right panels) to determine the effective CF size. (c) Dependence of the narrowest cross-section of the filament (so-called quantum point contact (QPC) constriction area) as a function of the programming current. Adapted and reprinted with permission from [256], © 2015 by the American Chemical Society.

of the oxygen concentration can induce considerable large changes of the cell resistivity (see Chapter 6 for more details). Therefore, the change of the oxygen content might be below the detection limit, in particular for small filaments induced by low power electroforming procedures. Irrespective of these constraints, a variety of examples for successful spectroscopic proofs of redox-processes taking place during electroforming exist in the literature and will be outlined in the following.

Bulk sensitive, X-ray absorption near-edge spectroscopy (XANES) of the Cr K-edge succeeded to identify oxygen vacancies in the first coordination shell of the Cr metal ions in Cr-doped SrTiO3 [259,260]. By scanning the sample under a micro-focused X-ray beam, spatially resolved analysis of the oxygen vacancy concentration was performed, revealing the accumulation of oxygen vacancies between two lateral electrodes on Cr-doped SrTiO3 single crystal [261,262]. A similar approach was used to investigate MIM structures of Fe-doped thin film SrTiO3 cells by micro-focused X-ray absorption spectroscopy [263]. Figure 30 shows the Fe K-edge spectra recorded on different positions of an electroformed device, on a reference film as well as on a device where a complete breakdown was induced during the electrical treatment. Clear changes of the pre-edge features are visible that can be attributed to a different amount of oxygen vacancies in the first coordination shell of the Fe dopants [263]. In particular, the oxygen vacancy concentration has the lowest value for the reference Fe-doped SrTiO3 thin film and the highest concentration of the breakdown spectrum. Based on the pre-edge intensity at 7122 eV, an oxygen vacancy map of the device is determined and depicted in Figure 30. It shows clear evidence for the creation of an oxygen vacancy rich filament during electroforming. Furthermore, it is important to note that the oxygen vacancy concentration beneath the electrode (see blue spectrum in Figure 30(a)) is significantly higher than in the thin films reference (see black spectrum in Figure 30(a)). One can therefore conclude that electroforming results in both, a net increase of the oxygen vacancy concentration underneath the whole electrode as well as the formation of a vacancy-rich filament, which might be formed at a later stage of the electroforming procedure. This might
be explained by multiple filament configuration [224] or a broad front of oxygen vacancies that moves in the early stage of electrical biasing. However, at certain defective positions of the film [223, 224], the oxygen vacancy formation or their movement might be enhanced, resulting in the formation of spikes in the moving oxygen vacancy front. The details of the different processes leading to the formation of substoichiometric filaments will be discussed below.

The loss of oxygen within filamentary regions ranging from several 100 nm to several μm has also been identified by other X-ray spectromicroscopic techniques for TiO₂ cells [264–268], Ta₂O₅ cells [248, 252, 269], SrTiO₃ cells [205, 224, 253, 254, 270, 271] and SiOₓ cells [272]. For filaments in the sub-50 nm regime, transmission electron microscopy (TEM) based techniques have to be employed to detect chemical changes. However, for TEM-based techniques, it is challenging to identify the filament position within the generally much larger memristive cells. Calka et al have identified a 20 nm extended oxygen poor region (up to 50% loss of O) in the vicinity of the top interface in TiN/HfO₂ cells by OK-edge EELS mapping [78]. For these experiments, the filament has been produced at a predefined region by the tip of a conductive AFM. Guided by EBAC filament detection in 500 nm × 500 nm sized Ir/Ta₂O₅/TaOₓ/TaN cells (Figure 31(a)), TEM analysis (Figure 31(b)) and OK-edge EELS mapping identified oxygen deficient filaments in the order of 15 nm (Figure 31(c)) [250].

By employing in-situ TEM analysis during forming and switching of Pt/SiO₂/Ta₂O₅₋ₓ/TaO₂₋ₓ cells, the formation of multiple substoichiometric Ta oxide filaments could be identified [77]. In particular, 1–2 nm size filaments consisting of TaO₂₋ₓ are detected within the SiO₂ layer after electroforming (see Figure 32), which are subsequently oxidized during the RESET process.

4.2.3. Structural changes

Besides the pure release of oxygen from the lattice and the accommodation of the non-stoichiometry by point defects such as oxygen vacancies or cation interstitials, electroforming might go along with the formation of extended defects or significant structural changes. The structural changes could on the one hand be inherently connected with the reduced oxygen content and the resulting structural change according to the phase diagram (see Chapter 3). In particular, some systems exhibit oxygen vacancy ordered phases such as Magnéli phases (see detailed discussion later) or Brownmillerite phases [273] and electroforming results in a local transition.
Figure 31. Filament detection in Ir/Ta$_2$O$_5$/TaO$_x$/TaN cells. (a) EBAC image of the cell exhibiting a conductive filament. (b) TEM image of the cross sectional structure. (c) EELS OK edge mapping [250]. Reprinted with permission from [250], © 2011 by IEEE.

Figure 32. In-situ TEM observation of electroforming in Ta$_2$O$_{5-x}$/SiO$_2$/Pt cells. (a) HAADF-STEM image of the Pt/SiO$_2$/Ta$_2$O$_{5-x}$/TaO$_{2-x}$/Ta$_2$O$_5$ interface in the virgin state. (b) HAADF-STEM image of the Pt/SiO$_2$/Ta$_2$O$_{5-x}$/TaO$_{2-x}$/Ta$_2$O$_5$ interface after forming. The substoichiometric TaO$_{2-x}$ filaments are marked in yellow. Adapted and reprinted with permission from [77], © 2013 by Springer Nature Limited. (Color figure available online).

to the respective phases. On the other hand, Joule heating might induce significant structural changes such as crystallization or phase separation during electroforming.

Since heating is generally not strictly confined to the conducting filament region, structural changes could take place within much larger regions. This might provoke an overestimation of the filament size as demonstrated by LPCM measurements [69] or a spurious assignment of structural changes to conducting filaments [274]. One example for such a side effect is the melting of the top electrode [205] and a related change of the grain size of the top electrode [201] observed during electroforming. The position of the structural change in the electrode might be a local hot spot in the electrode, which position is not necessarily identical with the filament position. It is interesting to note that Strachan et al. [264] have shown that the grain size of the Pt has changed within the whole Pt bottom electrode due to Joule heating during electroforming.

By employing soft-X-ray spectroscopic techniques, it has been shown that crystallization of the fully oxidized amorphous TiO$_2$ to anatase in Pt/TiO$_2$/Pt cells takes place in the whole cell area and not only in the filament region [264–266]. However, the spatial extension of the crystallization process depends strongly on the exact biasing procedure, in particular the power required for electroforming [266]. Carta et al. [268] have demonstrated a complex distribution of different TiO$_2$ phases within their electroformed cell and have identified crystalline anatase TiO$_2$ as well as orthorhombic TiO$_2$ adjacent to the main reduced area suggesting that the temperature has locally increased up to 1000 K. On the other hand, if the virgin thin films are already strongly substoichiometric and do not require a strong forming step, these TiO$_{2-x}$ cells can be operated over many cycles without significant crystallization effects [275].
One exceptional property of TiO$_2$ compared to other memristive oxides is, that a variety of substoichiometric phases are thermodynamically stable (see Chapter 3) and might be formed as a result of oxygen release if sufficient Joule heating occurs. In TiO$_2$, oxygen vacancies tend to order and form Wadsley-type defects, which are nucleation points for the formation of Magnéli phases Ti$_{n}$O$_{2n-1}$ during electroreduction [276]. In-situ TEM analysis on TiO$_2$ single crystals could prove the reversible formation and dissociation of Wadsley defects with electrical biasing (see Figure 33) [277].

TXM analysis of cross section lamellae of electroformed TiO$_2$ cells indicates the presence of distinct disordered substoichiometric TiO$_x$ phases localized to a 100 nm-sized region [268]. Furthermore, the crystalline Magnéli phase Ti$_4$O$_7$ has been identified in electroformed TiO$_2$ devices by several groups [122,265]. In particular, in-situ forming in the TEM was employed to perform structural investigations of the formed filaments and to subsequently characterize their electrical properties. The corresponding set-up is shown in Figure 34(a). By performing nanodiffraction analysis on the crystallized conical shape filament region, the observed spots could be assigned to the Ti$_4$O$_7$ Magnéli phase (b). This was confirmed by the observation of a metal-to-insulator transition near 130 K detected by temperature dependent resistivity measurements at the filament position [122]. The observed reversible formation/dissociation of Wadsley defects does not appear to correlate to resistive switching phenomena at these length scales as shown by Kamaladasa et al. [82]. It was found that the defect zones reversibly reconfigure in a manner consistent with charged oxygen vacancy migration responding to the applied bias polarity.

As explained below, the electronic initialization of the forming process is followed by local heating which may end up in a thermal runaway (and needs to be limited by a fast current compliance to avoid the destruction of the sample). The electron conductive path may be located in extended defects as described in Sec 3.4, and may act as a local heating element. This might explain earlier findings of apparent switching at the position of dislocations in SrTiO$_3$ [119]. The related temperature acceleration of the switching event has been explained in detail in Ref. [278].

For most other binary oxides such as HfO$_2$ and Ta$_2$O$_5$ no phases with ordered oxygen vacancies are known, with all other suboxides being metastable (see Chapter 3). Because of the high melting point of Ta$_2$O$_5$, virgin thin films are generally amorphous. Interestingly, nanocrystals observed in the vicinity of the conductive regions induced by electroforming consist of fully oxidized Ta$_2$O$_5$. Suboxides such as TaO$_{2-x}$ induced during electroforming remain amorphous [77,248,252] what has been attributed to the low solubility of oxygen vacancies in the crystalline phase [248].

For ZnO thin films, in-situ TEM studies proved the formation of conical-shaped filaments containing Zn dominated ZnO$_{1-x}$ with hexagonal close packed structure within a matrix of ZnO with wurzite structure [75].
For crystalline SrTiO$_3$ it has been shown that Joule heating during electroforming results in a plastic deformation by the formation of shear planes. As a result, a strong increase of the densities of dislocations has been observed in electroformed SrTiO$_3$ single crystals [279] after electroforming. Moreover, at very high local temperatures Sr segregates to the surface and forms SrO protrusions in the vicinity of the conducting filament [205,253,280]. Although this local phase formation is not prerequisite for the switching ability of the filament, the SrO coverage strongly improves the stability of the oxygen-deficient filament against reoxidation [254,270] as will be discussed in more detail in Chapter 7.

Atomically resolved analysis of similar SrTiO$_3$ cells by STEM/EELS revealed that the forming process induces an agglomeration of several 10 nm size Ti$^{3+}$-rich, oxygen deficient filaments instead of a single large filament [281]. As can be seen in Figure 35(a), these conical shaped filaments reside in different evolutionary stages. According to their considerable amount of Ti$^{3+}$ (see Figure 35(b)), all filaments might contribute to the current in the formed state. Figure 35(c) shows an atomically resolved image of the filament marked in Figure 35(a) displaying the atomic details of the filament as well as the surrounding film together with a further enlarged image of the boundary region marked in red. The reduction of the intensity within the filament region compared to the bulk can be reproduced by image simulations using structure models with both Sr- and O vacancies, confirming the loss of Sr in the filament during the SrO formation on the surface [280].

In summary, intensive studies on the microscopic changes during electroforming have been performed over the last decade. Conducting, oxygen deficient filaments have been detected for the most common VCM materials such as HfO$_2$, Ta$_2$O$_5$, TiO$_2$ and SrTiO$_3$ with diameters ranging from μm size down to a few nanometers, depending on the biasing procedure and the exact material stack configuration. There exists experimental evidence that the shape of the filaments can have a rather complex shape and could consist of rather networks of nanoscale filaments than on a single larger one. The observed melting or recrystallization of the metal electrodes, hint on strong Joule heating during the forming process. As a result of local heating, both recrystallization and phase transformation have been observed in VCM cells during electroforming.
Figure 35. (a) Annular bright field STEM of electroformed Fe-doped SrTiO\textsubscript{3} thin film cells depicting several filament in different evolutionary stages; (b) Ti\textsuperscript{3+} false color maps determined from EELS analysis of the Ti L-deg; (c) Two different magnifications of the STEM image of a, together with the projected unit cell of SrTiO\textsubscript{3} (Sr: green, Ti: red; O:blue) \cite{281}, \textcopyright 2017 by the American Chemical Society. (Color figure available online).

4.3. Microscopic mechanism during forming

4.3.1. Electronic effects initializing the forming

Although strong experimental evidence exists that electroforming is connected with ionic movement and concurring redox reactions, the processes are obviously triggered electronically in initially highly insulating devices at room temperature. As described in detail in Chapter 3, either high electric fields, high temperatures or both are needed to provide sufficient ion mobility to activate ionic processes. A detailed study of the early stage of electroforming in Ta\textsubscript{2}O\textsubscript{5-x}-based and TiO\textsubscript{2-x}-based devices have indeed shown that electronic processes play a dominant role in initiating the forming procedure \cite{201} and will be presented in the following.

Figure 36 shows the voltage across a Pt/Ta\textsubscript{2}O\textsubscript{5-x}/Pt (a) and a Pt/TiO\textsubscript{2-x}/Pt device (b) as a function of time. The individual curves (1)–(4) correspond to the trace during a single voltage pulse applied to the same device via a series resistance \cite{201}. Each pulse had the same amplitude but slightly different duration, allowing for the interruption of the process at different stages, a few nanoseconds prior to the completion of forming. The gradual decrease of voltage in the beginning is associated with the decrease of the resistance with increasing Joule-heating. The rapid drop between 45 and 55 ns and 7 and 9 \textmu s corresponds to the beginning of the electroforming process in the Ta\textsubscript{2}O\textsubscript{5-x}-based and the TiO\textsubscript{2-x}-based device, respectively. No permanent change of the device takes place after the pulses (1)–(3). Therefore, the initial part of the sharp resistance drop is volatile and is suggested to be electronic in nature \cite{201}. In pulse (4), the resistance drop is permanent and the electroforming process can be regarded as completed. For the TiO\textsubscript{2-x}-based devices, the voltage pulses (2)–(4) lead to morphological changes of the Pt top electrode as shown in Figure 36(c), which seems to originate from a kind of recrystallization in a circular area of different diameter. This process indicates locally confined Joule heating. In the initial stages of the volatile resistance change the current flow is quite homogeneous, whereas the sharp resistance decrease corresponds to a localized electronic instability \cite{201}. The incubation time, i.e. the time until the sharp resistance drop occurs, shows a highly nonlinear temperature and voltage dependence as illustrated in Figure 37 for a 60 nm-thick Ta\textsubscript{2}O\textsubscript{5-x} device and a 15 nm-thick TiO\textsubscript{2-x} device. The dashed line marks the thermal time constant of the device. Above this line, Joule heating occurs instantaneously and the slope in this regime becomes very steep indicating the existence of a threshold voltage. Below this line, the incubation time is influenced by the thermal time constant of the device setup and a 1/E relation is apparent \cite{200}. At very high
voltages, the effective activation barrier of the volatile effect is zero and the incubation time saturates. Field-induced nucleation of a conducting filament was proposed as origin of the incubation time [200]. This hypothesis, however, requires the appearance of a volatile second phase during electroforming, which has not been proven yet. One can conclude from these experiments that electroforming may be initiated through purely electronic and reversible events, to be followed by a heating effect which triggers oxygen vacancy (or cation interstitials) movement and/or structural changes. It should be noted that the purely electronic process does not need to be the initial step in all systems. As the electronic current is very localized, a strong temperature gradient is produced. It was speculated in the literature that this causes thermo-diffusion effects which leads to perpendicular oxygen vacancy movement or even metal diffusion, which finally forms the filament [88,282,283]. As temperature gradients also evolve along the parallel direction, the parallel movement of ionic defects was discussed in terms of thermo-diffusion alone [284]. It should be noted that all simulations including the charge state of ions show a dominant contribution of ionic drift in the parallel direction [285,286]. Thus, thermo-diffusion as dominant mechanism might be more reasonable for non-charged particles.

4.3.2. Redox reactions at the electrode interfaces

Based on the various experimental proofs for the formation of oxygen vacancies (or cation interstitials) during electroforming reported in 5.2, it is evident that oxygen is released from the oxide lattice during electroforming by anodic oxidation as explained in detail in Chapter 3. This could take place either by the oxidation of an oxidizable metal electrode or by the release of oxygen
Figure 37. Incubation time as a function of applied bias and stage temperature for (a) TaO$_x$ and (b) TiO$_x$ based crossbar devices. The horizontal dashed line denotes the value of the thermal time constant of devices with uniform current flow. Adapted and reprinted with permission from [200], © 2015 by American Institute of Physics.

gas according to the reactions in Equations (23) and (24) at an inert electrode such as Pt.):

\[
O_x^+ + 2M_M \rightleftharpoons \frac{1}{2}O_2(g) + V_0^+ + 2M_M' \tag{23}
\]

\[
O_x^+ \rightleftharpoons \frac{1}{2}O_2(g) + V_0^+ + 2e' \tag{24}
\]

Both equations describe the same process. Oxygen is released to the gas phase, indicated by (g), and a doubly ionized oxygen vacancy acting as mobile donors as well as two electrons are left behind. In the first Equation (23), these electrons are trapped by two cations M reducing their valence. In the second equation, the two electrons are considered as free (e.g. in the conduction band). In some metal oxides, an alternative oxygen release reaction is found:

\[
O_x^+ + 2M_M' \rightleftharpoons \frac{1}{2}O_2(g) + 2M_I^+ + 2e' \tag{25}
\]

Here, cations are transferred into interstitial positions and now act as mobile donors instead of the oxygen vacancies. This is observed, e.g. for TaO$_x$. Obviously, reaction (25) leads to a shrinkage of the oxide volume which can in fact be observed [219].

One first hint on the release of oxygen gas during electroforming was the observation of local deformations of the top electrode [119]. The degree of deformation of the electrode during oxygen gas evolution is correlated with the elastic properties of the electrode material on the one hand and on the amount of oxygen gas on the other hand. Due to latter reason, the deformation is more pronounced for µm scale cells than for nanoscale cells and depends on the amplitude and duration of the forming step [69]. As can be seen in Figure 38 for 60 µm size Pt/TiO$_2$/Pt cells, the occurrence of bubble formation is strongly dependent on the bias polarity and consistent with the anodic oxidation of lattice oxygen at the positively biased top electrode. For negative bias applied to the top electrode, only a few bubbles are formed (Figure 38(b)) since it is more difficult to form gas at the bottom interface underneath the TiO$_2$ layer. However, the bubbles remain after removing the bias (Figure 38(c)) since it is difficult for the gas to escape. An opposite bias leads to a shrinking of the previous bubbles resulting from the reincorporation of the gas in the TiO$_2$ lattice. Furthermore, new small bubbles are formed underneath the top Pt electrode (Figure 38(d)), which grow in number and size when the positive voltage is maintained.
Figure 38. Optical image of Pt/TiO$_2$/Pt cells. Both the Pt bottom electrode and the TiO$_2$ film are blanket films and the Pt top electrode defines the junction area. Gas bubble formation under negative bias (b)–(c) and under positive bias (d)–(h), AFM image (g) of the electrode deformation remaining when the bias is removed. Reprinted from [69] with permission, © 2009 by IOP Publishing Ltd.

(Figure 38(e)). A larger or longer applied voltage causes a growth and agglomeration of bubbles (Figure 38(h)). After the removal of the bias, the bubbles disappear within seconds (Figure 38(f)) leaving behind remnant eruption features on the top electrode (Figure 38(g)). We can conclude that the oxygen gas can escape at the eruption of the Pt electrode, at the device edges or by grain boundary diffusion within the Pt [287–289].

For dense crystalline cells with oxygen blocking bottom electrodes such as epitaxial SrTiO$_3$ on Nb-doped single crystals [196], forming with negative voltages on the top electrode is not possible since oxygen cannot be released at the bottom electrode. For amorphous Pt/TiO$_2$/Pt cells, it has been observed that negative forming induces a local delamination of the TiO$_2$ layer from the substrate [247, 264] as a result of the oxygen gas evolution at the bottom interface.

However, one should keep in mind that electrode deformations are not solely induced by oxygen gas evolution but also by structural changes in the oxide material as will be discussed in detail later in this chapter resulting in protrusions on the surface as well as to cavities within the metal oxide thin film [268]. Therefore, it is important to note that the evolution of oxygen gas during forming has recently been explicitly confirmed by mass spectrometry for SiO$_x$ cells with TiN top electrode [272].

For non-noble metal electrodes, the metal might be oxidized during the anodic oxidation according to Equations (3)–(21) instead of by the release of oxygen (Section 3.7). This has been explicitly proved for Ti/HfO$_x$/TiN devices by *in operando* hard X-ray photoelectron spectroscopy analysis by the observation of a shift of the chemical weight from metallic Ti to Ti$^{2+}$ and Ti$^{4+}$ [290]. In general, oxidizable metals form an oxide layer at the interface already during the metal deposition (see Chapter 2). For devices with one oxidizable electrode and one noble metal electrode it is often observed that forming by positive biasing of the oxidizable electrode results in more reliable switching. In this case, the oxygen can be easily accommodated in the oxidizable electrode or within the so-called oxygen exchange layer at the electrode interface (Chapter 6). During forming, an oxygen exchange reservoir is formed near the filament. Scalpel scanning probe microscopy indicated that the lateral dimensions of the oxygen reservoir is 2–5 times larger.
than the conducting filament depending on the material combination [87]. This oxygen exchange reservoir furthermore provides a kinetic barrier for the re-oxidation of the filament during subsequent operation. This effect will be discussed in more detail with respect to reliability issues in Chapter 7.

4.3.3. Ionic concentration polarization processes

In a first approach to an understanding of the electroforming process one may ask what happens to a mixed electronic–ionic conducting oxide, such as SrTiO$_3$ in the low temperature regime, when a dc voltage is applied to an MIM system. While the electrode interfaces M/I are considered to be sufficiently transparent for electronic carriers, the answer to this question mainly depends on the ionic interface reaction. Theoretically, one may distinguish between two cases. If both electrodes are ideally non-blocking, i.e. the reaction rates at the interfaces are much faster than the ionic transport within the I-layer, no big change in the concentration profiles will take place, since oxygen will enter at the cathode and exit at the anode, as in solid-oxide fuel-cells [136]. If, however, the electrode interfaces partially or completely block the ionic currents a typical concentration polarization occurs in the oxide. For low voltages, a linear concentration gradient of oxygen vacancies builds up [291]. For high voltages, however, the situation changes dramatically due to nonlinearities in the transport processes. Recent studies indicate a complex interplay of ionic and electronic processes during the electrodegradation in Fe-doped SrTiO$_3$, which is very sensitive to impurities [292]. The process has been visualized by recording the optical transmission image of a slightly Fe acceptor doped SrTiO$_3$ single crystal under dc voltage and temperature stress [293]. Please note that Fe$^{4+}$ ions in SrTiO$_3$ show a strong brownish-red color in transmission, while Fe$^{3+}$ ions are almost colorless. The drift of O vacancies towards the cathode regions leads to a very pronounced concentration polarization which shows up as an electrocoloration process (Figure 39). Local changes in the valence states of the Fe ions allow for a detailed observation of this process. Starting with a dark yellow homogenous color a dark brownish-red region and a light yellow, almost colorless region evolve under DC voltage bias at the anode and cathode, respectively (Figure 39(a–d)) [279]. Concurrently the current through the sample increases (Figure 40). The dark brownish-red color in the anodic region is due to an oxidation of Fe ions into the valence state $+4$ indicating the depletion of O vacancies. Because of defect equilibria, there is also an enhanced $p$-conductivity in the brownish-red region [178].

In the cathodic region, an almost colorless region evolves showing Fe ions in the lower valence state (+3) due to the enrichment of O vacancies. The cathodic region shows a pronounced $n$-conductivity. When the color fronts have met, a $pn$-junction has formed that is biased in the forward direction [178]. Therefore, the overall resistance of the MIM structure degrades. This is a typical example of the reduction (shrinkage) of a neutral barrier which in the initial stage is represented by the material ‘I’. The position of the $pn$-junction area depends on the annealing oxygen partial pressure, the dopant concentration, and the annealing temperature [294].

To explain the electroforming process in resistive switching oxides a net reduction of the oxide thin film has to occur. In the context of the concentration polarization, this will happen when there is an anodic reaction and this anodic reaction is faster than the cathodic one [178]. Typically the anode does not completely block the ionic partial current, as exemplified by an oxygen release at the electrode and eventually the formation of entrapped O$_2$ gas bubbles [69,119,294–296]. In contrast to the earlier studies [297], this process is only observed at times beyond the full concentration polarization (Figure 39(d)). In this case, the $n$-conducting cathodic (almost colorless) region, sometimes called ‘virtual cathode’, propagates further towards the anode. It protrudes into the dark brownish-red region in a filamentary manner until the brownish-red region is vanished almost completely (Figure 39(e–j)). This process is accompanied by a further
current increase until the current compliance is reached at point h (Figure 20). Under current control the voltage across the sample decreases indicating a further resistance degradation. As the bright yellow, almost colorless state indicates a reduction of Fe$^{4+}$ to Fe$^{3+}$, the whole sample is reduced. This process can be regarded as the completed electroforming process. A numerical simulation model, including the drift-diffusion equation for oxygen vacancies, the Poisson equation and the set of mass action equations, was used to interpret the formation and progression of the color fronts in Figure 39 [178,279], fully consistent with the optical images (Figure 21) and later complementary studies [298] as shown in Figure 41.

Similar studies on the concentration polarization as a function of voltage and time have been performed for TiO$_2$ single crystals [299]. When subjected to an applied voltage of 15 V, a time-dependent increase and saturation in the leakage current is observed, which is associated with...
an accumulation of point defects and an attendant decrease in stoichiometry at the cathode electrode. Cathodoluminescence spectroscopy shows that Ti interstitials dominate the point defect redistribution process. Under larger applied voltages, of around 30 V, the resistivity increases as a function of time. This behavior is associated with the condensation of point defects into a region of extended defects and Magnéli phases near the cathode, sufficient to increase the bulk stoichiometry and resistivity.

4.4. Filament growth direction

Based on the discussions in the Sections 4.3.2 and 4.3.3 the electroforming forming process involves an anodic oxidation to release oxygen from the oxide matrix followed by the migration of the left behind oxygen vacancies (or cation interstitials) within the MIEC, which leads to a local reduction of the oxide material. Alternatively, it involves the reaction of the oxide with the ohmic electrode which also reduces the oxide. This means that electroforming can, indeed, be achieved with both voltage polarities as explained e.g. in Refs. [69,203]. As mentioned above, the local reduction of the oxide might be accompanied by the formation of stable or meta-stable oxygen deficient phases. To drive these processes within the time scale of typical forming experiments sufficient thermal energy is required. This energy has to be delivered by local Joule heating. A prerequisite for Joule heating is a current confinement induced by the electrical processes explained in Section 4.3.1.

Yalon et al. investigated the filament growth direction during the electroforming process of HfO₂-based devices using a special metal–insulator–semiconductor bipolar transistor structure, which allowed to detect the filament growth direction from electrical measurements [300]. The authors showed that the filament grows from the cathode to the anode if the anodic oxidation happens at the chemically inert electrode, which is in this experiment either a Pt electrode or the semiconductor base electrode. In contrast, the filament grew from the anode to the cathode when a chemically active electrode material (here Ti) is used.

These experimental findings have been interpreted in terms of the rate of the anodic oxidation reaction compared to the speed of ion migration within the oxide layer based on drift diffusion modeling [301,285]. Marchewka et al. investigated the electroforming process using a 2D axisymmetric simulation model, which includes the drift-diffusion of oxygen defects within
Figure 42. (a) $I-V$ characteristics during electroforming for (a) a high anodic oxidation rate and (b) a low anodic oxidation rate. (c) and (d) show the evolution of the maps of the oxygen-vacancy concentration in the switching layer at points A to J as marked in the $I-V$ characteristics in (a) and (c), respectively. The anode is at the bottom interface and the cathode is at the top boundary of the maps. At the anode, anodic oxidation occurs consuming O from the oxide layer leaving behind positively charged oxygen vacancies that can migrate towards the top cathode during electroforming. The black line in the concentration maps marks the isoline of $N_{\text{VO}} = 1 \times 10^{20} \text{cm}^{-3}$ and illustrates the filament evolution. Adapted and reprinted with permission from [285], © 2017 by IEEE.

The material, anodic oxidation reactions, Joule heating, a band model for electron conduction and electron tunneling processes (details of the model will be given in the Chapter 5). Figure 42(a,c) show the simulated $I-V$ ($T-V$) characteristics for one high and one low oxidation rate, respectively. The corresponding evolution of the defect concentration within the switching layer is illustrated in Figure 42(b,d). It is observed that the filament growth direction changes while changing the oxidation rate. For a high oxidation rate (cf. Figure 42(b)) in comparison to the ion motion, the filament grows from the anode to the cathode. In contrast, the filament evolves from the cathode to the anode for a low oxidation rate. This finding is consistent with the experimental findings of Yalon et al. [300] If the rate of the anodic oxidation reaction is slow compared to the migration of oxygen vacancies within the switching layer, the filament will grow from the cathode to the anode as discussed for the concentration polarization in Sec 4.3.3. If the rate of the oxygen exchange reaction is fast in comparison to the $V_{\text{O}}^{\bullet\bullet}$ migration, the filament growth direction is reversed. With respect to the experiment of Yalon et al. described before, the oxidizable Ti anode facilitates a fast oxidation reaction and the filament grows from the Ti anode to the semiconductor cathode. In contrast, the oxygen evolution reaction rate at the inert Pt anode is slow and the filament grows from the cathode to the anode. When the semiconductor base electrode serves as anode, there is no difference in the oxidation rate of both samples. The rate is slow
compared to the ion migration and the filament grows from the Ti (or Pt) cathode to the semiconductor base anode. The same conclusion was shown for a 3D KMC model of the electroforming process considering anodic oxidation and ion migration [302]. This explanation of the filament growth direction is very similar to the one for ECM cells [303].

The choice of the electrodes (inert or oxidizable) does not only affect the filament growth direction, but also the resulting resistance state of the electroformed device. In the following, we will briefly discuss the microscopic mechanism behind the forming of the HRS and the LRS. We assume a cell with a chemically inert but electronically active electrode and an ohmic counter electrode with a high oxygen affinity (e.g. a cell Pt/TiO₂/Ti). Electronically, this MIM cell corresponds to a Schottky diode.

Starting from an initial resistive state (pristine state) of very high resistance (Figure 43(a)), a positive forming voltage will lead to an electroforming into the HRS (Figure 43(d–f)). The Schottky diode is biased in the forward direction. As already mentioned above, the anodic oxidation reaction will release O₂ gas (and may lead to entrapped gas bubbles underneath the Pt electrode, Figure 43(e)):

\[
O_2 \rightarrow 2e^- + \frac{1}{2}O_2(g) + V_{0}^{\star}
\]  

(26)

Oxygen vacancies \( V_{0}^{\star} \) are injected into the oxide and drift rapidly towards the cathode due to the high electric field. Since there is no redox reaction at the cathode interface, the \( V_{0}^{\star} \) start to accumulate near the cathode, compensated by electrons. Along with the further accumulation of \( V_{0}^{\star} \), the n-conducting cathodic region, i.e. the virtual cathode, propagates towards the anode. When only a relatively small potential barrier (disc) remains, the resistance of the cell decreases significantly (usually limited by a compliance current) and the electroforming into the HRS is completed. This process can also be used if one starts from a symmetric cell of high work function electrodes, e.g. Pt/TiO₂/Pt, because a virtual cathode is formed during the reduction process.

A negative forming voltage will lead to an electroforming into the LRS (Figure 43(b, c)). The anodic oxidation reaction will lead to an oxidation of the Ti electrode in this case. Oxygen vacancies \( V_{0}^{\star} \) are formed and drift towards the Pt cathode. They start to accumulate near the cathode, compensated by electrons. This accumulation proceeds until the n-conducting region bridges the electrodes and the electroforming (limited by a compliance current) is terminated in the LRS of the cell. The difference in the process for forming into the HRS is, that the anode is a low work-function metal so that no effective barrier remains at the end of the process. In addition, Ti consumes oxygen ions also purely chemically, so that no fully oxidized TiO₂ layer is left. A partial oxidation of the electrode material was observed for TiN electrodes in a TiN/TaO₂ system, where an oxynitride is probably formed [284]. It was also shown that the degree of electrode oxidation depends on the thickness and oxygen affinity of the electrode material for ZrO₂ and TaO₂ as oxide material and Hf and Ta as oxidizable electrodes [304]. Celano et al. revealed a higher resistive electrode region on top of the formed filament for Hf/HfO₂ and Ta/Ta₂O₅ interfaces using scalpel AFM [87]. According to this study, the size of the resistive region depends on the electrode thickness, which is consistent with the results of Kindsmueller et al. as mentioned before.

In both cases, the overall process is a reduction of the oxide. Typically, this process is non-uniform and leads to the formation of n-conducting filaments. Very high local temperatures occur during the forming process, due to the localized currents and the relatively high voltages. This accelerates the forming process and may lead to morphological changes [305] as well as the formation of new, oxygen deficient phases such as Magnéli phases in TiO₂ (Figure 34) [122], as discussed above. After the release of the forming voltage (e.g. by the current compliance), the concentration profiles get frozen in.
4.5. Forming-free systems

In the device application, electroforming procedures described in the previous sections are undesirable because of CMOS incompatible, energy-inefficient high voltages and the time-consuming necessity of addressing every individual cell. Furthermore, as mentioned in the previous sections, forming often results in deformations of the oxide or the electrode materials. Therefore, it is highly desirable to reduce the forming voltage or to get rid of it completely by the fabrication of forming-free devices.

Tuning of the oxygen vacancy concentration in HfO$_2$$_{−x}$ and Ta$_2$O$_5$$_{−x}$ thin films by the deposition conditions showed that the forming voltage systematically decreases with increasing oxygen vacancy concentration [215,218,219]. Since an increased oxygen vacancy concentration goes along with a higher electronic conductivity, Joule heating and temperature-accelerated ionic motion becomes significant at lower voltages. If the forming voltage approaches the switching voltage, the devices can be regarded as forming-free. However, one should keep in mind, that for ex-situ electrode deposition processes, the forming voltage of the device is governed rather by the oxidized surface layer than by the sub-stoichiometric oxide layer underneath [215]. Similar
Figure 44. (a) Dependence of the forming voltage of HfOx/Hf cells on the oxide thickness for a 10 nm thick Hf cap electrode. Inset: Weibull-plot for the SET process for the 1st and the next set events. The cell can be regarded as forming-free; (b) Dependence of the forming voltage on the Hf cap layer thickness for different HfO2 device areas; [212], © 2011 by IEEE.

bilayers of substoichiometric TiOx and thin stoichiometric TiO2 have been grown intentionally to realize forming-free Pt/TiO2/TiOx/Pt cells [69,275].

As mentioned previously, in cell stacks with oxidizable electrodes, a redox process with the metal electrodes takes place and results in an increase of the oxygen vacancy concentration in the oxide layer resulting also in bilayer stacks of stoichiometric and substoichiometric oxides. It has been shown for Nb-doped SrTiO3/SrTiO3/Ti devices that the degree of oxidation of the Ti electrode depends on the relative thickness of SrTiO3 to the Ti cap layer [233]. In turn, the oxidation state of Ti ions in the SrTiO3 decreases with increasing Ti cap layer thickness. If the Ti thickness exceeds a certain value, the devices become forming-free. This effect has been confirmed for a large variety of other layer stack combinations such as HfO2/Hf [212], Ta/Ta2O5 [306], Ti/TiO2 and W/TiO2 [307]. Figure 44(a) shows a linear dependence of the forming voltage on the HfO2 thickness of HfO2/Hf cells for a given Hf layer thickness of 10 nm. Moreover, Figure 44(b) shows a decrease of the forming voltage with increasing Hf electrode thickness. Stacks with 2 nm-thick HfO2 can be regarded as forming-free [212,308,309] as can be seen from the set statistics in the inset of Figure 44(a). Besides the increased oxygen vacancy concentration in the 2 nm-thick layer induced by the interface reaction, the reduction of the forming voltage has also been attributed to the increased electric field in the thinner oxide layer stacks [212]. This is consistent with the observation of forming-free Pt/TiO2/Pt cells with 4 nm-thick stoichiometric TiO2 layers [69].

These methods to reduce the forming voltage, however, also lead to an undesirable increase of the leakage currents and, in turn, a decrease in the resistance window [215,233] as well as a degradation of the reliability and retention properties [310]. In an alternative approach, that prevents these drawbacks, nanoscale forming-free Ta2O5 and HfO2 devices have been produced by ion implantation (e.g. with O ions) into the switching oxide layer. With optimized implantation dose, which differs between different materials, as-fabricated devices are initially in the LRS with a narrow RON distribution. The forming-free devices show similar resistive switching properties
as reference devices without any degradation of electrical performance such as $R_{OFF}/R_{ON}$ values in the order of 20 with a low current compliance of 50 μA [311–313].

5. Valence change switching mechanism

5.1. Survey of the switching systems

There are several ways to arrive at a VCM-type MIM system. Three typical approaches are sketched in Figure 45. In any case, an electrode material with a low oxygen affinity and a relatively high work function (e.g. Pt, Ir, TiN) is used on the electronically active interface (AE) side. For the oxide thin film, there are different strategies: (a) the oxide film is a homogeneous monolayer and composed of e.g. TiO$_x$, TaO$_x$, WO$_x$, HfO$_x$, SrTiO$_3$. In the case of fully oxidized, highly insulating oxides, an electroforming step is essential, which generates an oxygen-deficient, n-conducting oxide at the side of the ohmic electrode. (b) In order to support such a structure already during processing, the bi-layer concept has been introduced. In the case of homogeneous bi-layer, an oxygen deficient, n-conducting layer is deposited on the ohmic electrode (sometimes called the oxygen exchange layer, OEL), and a second, fully oxidized layer of few nanometer thickness of the same oxide (called the switching layer, SL) is processed on the side of the active electrode. Examples are TiO$_2$/TiO$_{(x<2)}$ [245], and Ta$_2$O$_5$/TaO$_{(x<2.5)}$ [314]. (c) Alternatively, in a heterogeneous bi-layer concept, the second layer is made from another oxide with a larger formation energy and/or a larger band gap. Examples are Al$_2$O$_3$/TiO$_{(x<2)}$ [92,315], HfO$_2$/TiO$_{(x<2)}$ [204,316] or other transition metal oxides [317].

For the ohmic electrode (OE), an electrode metal with a low work function and a high oxygen affinity is preferred [318]. In most MIM systems, an electroforming process is required before resistive switching can be conducted. If a high work function/low oxygen affinity metal is used as the counter electrode, the electroforming is essential in order to create the ohmic counter electrode. The electroforming process as well as the possibility to create forming-free devices has been discussed comprehensively in Chapter 4.

The classification of VCM-types according to the switching geometry, i.e. filamentary vs. area-dependent switching, and according to the switching polarity, i.e. counter-eightwise (C8W) switching vs eightwise (8W) switching has already been briefly introduced in Chapter 2. In addition, we have mentioned the complementary switching (CS or CRS). Figure 46 shows the two main switching geometries and the two switching polarities as independent categories and, in fact, examples for all four cases have been found, although with very different occurrences.

Chapter 6 is organized as follows: Examples for the valence change during the switching are presented in Section 5.2. The two major memristive states, i.e. the LRS and the HRS, are described in Section 5.3, subdivided into sections on bulk switching models (emphasizing neutral barriers) and interface-related conduction models (comprising space charge and tunneling barriers). After these general sections, we will discuss the filamentary C8W switching and the corresponding switching models in Section 5.4. Because of its importance we call the C8W switching the standard resistive switching mode. Section 5.5 will deal with the area-dependent switching starting with the prerequisites of this switching geometry followed by typical examples. Subsequently, in Section 5.6, the 8W switching will be discussed taking into account both geometries. After these VCM classes, the highly non-linear kinetics of the VCM-type switching will be described in Section 5.7. We will conclude this chapter by a section on the ultimate physical limits to scaling (Section 5.8).

5.2. Valence change in the switching region – the essential metal insulator transition

Within this chapter, we will give an overview over the experimentally observed changes of the atomic and electronic structure taking place during resistive switching in different types of
memristive cells. Whereas electroforming (see Chapter 5) can be a rather slow process which goes along with significant, at least partly, irreversible changes of the atomic and electronic structure of the material, the switching process has to be fast and in the best case completely reversible. Therefore, less severe changes of the material can be expected and the ionic movement will likely be much more localized than during the forming procedure. Since memristive cells, both filamentary as well as area-dependent, often contain several oxide sublayers (layers with different oxygen stoichiometry or a different metal oxide, see Figure 45) as well as two different electrodes, it is also important to clarify which layers and interfaces undergo chemical changes and how does this influence the overall cell resistance in the different resistance states. In order to elucidate the filamentary switching mechanisms, it has to be clarified which regions of the filament created during electroforming are modified subsequently during the switching process and which exact changes of the electronic and atomic structure take place.

Figure 45. Typical MIM stacks with different concepts for the I-layer. The shade of green indicates the degree of O deficiency. (a) Homogeneous monolayer, including a gradient of the degree of reduction; (b) homogeneous bi-layer; (c) heterogeneous bi-layer.
Figure 46. Classification of VCM-type cells according to the switching polarity and the switching geometry, modified after [319]. Note: As will be explained in the text, the switching polarity in a given material such as SrTiO$_3$ results from a competition between the speed of the ion transport in the oxide and the rate of the ion exchange reaction at the interface. Therefore, the microstructure of the oxide has a significant influence.

<table>
<thead>
<tr>
<th>Switching Geometry</th>
<th>Filamentary switching</th>
<th>Area-dependent switching</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counter-eighthwise (C8W)</td>
<td>Standard TiO$_2$, Ta$_2$O$_5$, HfO$_2$, etc.</td>
<td>e.g. TO/PCMO (TO = Al$_2$O$_3$, YSZ)</td>
</tr>
<tr>
<td>Eighthwise (8W)</td>
<td>e.g. SrTiO$_3$, SrTiO$_3$:Nb at high current</td>
<td>e.g. SrTiO$_3$:Nb at low current</td>
</tr>
</tbody>
</table>

Observation of **filamentary switching** requires a pronounced spatial resolution in order to reveal the microscopic switching mechanism. Whereas XPS analysis on large area cells has been successful in detecting valence changes in area-dependent switching systems, it failed to detect valence changes in filamentary systems such as HfO$_2$ [320] and TiO$_2$ [321] as a result of the insufficient amount of modified material. Other spectroscopic fingerprints of resistive switching detected by XPS analysis is a shift in the Fermi energy in HfO$_2$ [320] and a change of the Ti/O content in Pt/TiO$_2$/TiO$_2$ cells [267].

By employing spatial resolved techniques, single large filaments with a size of 1–2 μm have been identified in SrTiO$_3$ thin film cells by PEEM analysis after Au electrode removal [270]. By comparing SrTiO$_3$ cells in different resistive states, it could be clearly proved that a significant amount of Ti$^{3+}$ is present in the LRS whereas the HRS contains mainly Ti$^{4+}$. This switching-induced valence change from Ti$^{4+}$ to Ti$^{3+}$ has been confirmed for Nb:STO single crystal/Au cells [254]. By employing the same experimental techniques, it has been shown that a valence change from Ta$^{5+}$ to Ta$^{4+}$ takes place within a μm sized multifilamentary region at the Ta$_2$O$_{5-x}$/Pt interface of Ta/Ta$_2$O$_{5-x}$/Pt cells [252] as depicted in Figure 47. All these experiments could explicitly prove valence changes on the metal cation sites in the oxide at the active high work-function metal electrode side as underlying switching mechanisms.

Due to the considerable variation of filament diameter, shape, position and oxygen vacancy concentration from sample to sample, it is difficult to assign tiny differences to the switching process if the samples investigated in different resistive states are not identical. Therefore, some groups have developed *in operando* techniques to investigate one and the same sample during repeated electrical biasing. By using photoelectron transparent graphene electrodes, *in operando* PEEM analysis of SrTiO$_3$ thin film devices has been successfully performed [271]. Figure 48 (panels a,b) show a PEEM image of a filament in the LRS (a) and the HRS (b) identified by a contrast in the O K edge which can be assigned to a change in the charge carrier density induced by a local change in the carrier concentration. Repeated switching of the cells resulted in a reversible change of the contrast in the PEEM image. Figure 48(d) shows the oxygen vacancy concentration, determined by a comparison of the O K edge spectra in the filament region with reference spectra from the literature, after repeated biasing. The comparison with the cell resistance in
Figure 47. (a)–(d) PEEM analysis of a TaTa$_2$O$_{5-x}$/Pt device after Pt delamination. False color maps showing the distribution of Ta$_4^+$ (red) and Ta$_5^+$ (blue) for the LRS (a) and HRS (b). Ta 4f core level spectra from the dendrite-like structure (red curve) and the reference region (blue curve) for an LRS (c) and HRS (d) cell. The spectra were fitted with two components, one for Ta$_2$O$_{5-x}$ (blue) and one for the low binding energy component TaO$_2$ (red), the total fit is presented by the pink curve. Adapted and reprinted with permission from [252], © 2015 by Wiley-VCH. (Color figure available online).

Figure 48(c) proves a systematic change of the carrier concentration with the cell resistance. By inserting the spectroscopically determined oxygen vacancy concentrations into a Schottky barrier model [322] of the SrTiO$_3$/graphene interface, the measured current–voltage curves could be nicely reproduced. As a result of the strong change of the shape of the Schottky-barrier with the oxygen vacancy concentration, a change of factor 2 in the concentration results in a change in the cell resistance of 2 orders of magnitude (cf. Section 5.3.2, Figure 54).

As the majority of oxide systems, Ta$_2$O$_5$ is generally regarded as a filamentary or a multifilamentary switching system since no scaling of the LRS current with the cell area is observed and a variety of investigations show that electrical and chemical changes are locally confined [248,252,269] (see for example Figure 47). In operando STEM analysis of Pt/SiO$_x$/Ta$_2$O$_5$/TaO$_{2-x}$ cells has provided atomic scale evidence of multifilamentary redox-processes taking place during switching [77]. As previously mentioned in Chapter 5, these investigations gave evidence for the formation of multiple sub-stoichiometric TaO$_{2-x}$ filaments. During subsequent switching, these filaments are oxidized in the HRS and reduced in the LRS as can be seen in the contrast change in the regions marked with yellow arrows in Figure
Figure 48. In operando PEEM measurements on SrTiO$_3$ cells (a, b) PEEM images of the filament region in LRS (a) and HRS (b) at the O K edge at a photon energy of 531.6 eV. (c) Resistance change between the different switching sequences; (d) Related change of the oxygen vacancy concentration during repeated switching; Adapted and reprinted with permission from [271], © 2016 by Springer Nature Limited.

Although this confirms the valence change at the Pt/Ta$_2$O$_5$ interface to be the dominant redox-process with respect to the observed resistance change between LRS and HRS, detailed HAADF-STEM analysis of the whole specimen region could reveal that chemical changes also take place locally within the bottom TaO$_{2-x}$ film irrespective of the reduced voltage drop in this region. In particular, as depicted in Figure 49(c, d), an increase of conducting Ta rich clusters is observed within the TaO$_{2-x}$ bottom layer in the LRS. The complete microscopic picture of the switching process deduced from these investigations is summarized in the sketch in Figure 49(e).

**Electrode reactions** (see Section 3.7) during fabrication as well as during electrical biasing imply that they might play a role during the switching process as well. Indeed, *in operando* X-ray photoelectron spectroscopy on Ti/HfO$_2$/TiN devices gave experimental hints that the Ti electrode is directly involved in the switching process. Figure 50 shows the Ti core levels of Ti/HfO$_x$/TiN cells in different resistive states [290]. During electroforming, a significant decrease of the Ti metal contribution as well as an increase in the Ti$^{4+}$ contribution has been observed. The difference between LRS and HRS is less pronounced, however, a small shift of the spectral weight from high Ti valence states (Ti$^{3+}$) to low valence Ti states (Ti$^{1+}$) has been observed during SET operation (see Figure 50(d)). The observed switching polarity is consistent with a reversible redox-reaction between a TiO$_x$ layer at the bottom electrode and the HfO$_x$ switching layer, resulting in a change of the band-bending at the Ti/HfO$_x$ interface [290].

For **area-dependent switching** which will be discussed in detail in Section 5.5, spectroscopic investigations employing XPS techniques without any pronounced spatial resolution have been successful in detecting changes of the electronic structure during resistive switching. For amorphous GaO$_x$ thin film cells (Pt/GaO$_x$/ITO) Aoki et al. [67] succeeded to show that different resistive states, which have been adjusted by the programming voltage polarity and amplitude,
Figure 49. In-situ TEM switching experiments on Pt/SiO$_2$/Ta$_2$O$_{5-x}$/TaO$_{2-x}$ cells. (a, b) HAADF-STEM images of the Pt/SiO$_2$/Ta$_2$O$_{5-x}$ interface in HRS and LRS. The filaments where a contrast change in the SiO$_2$ is observed are marked in yellow. (c,d) Pseudocolor maps of the TaO$_{2-x}$ region in HRS and LRS. Yellow: Ta rich phase (conducting paths). Blue: non-conducting clusters (e) Sketch of the switching process in Pt/SiO$_2$/Ta$_2$O$_{5-x}$/TaO$_{2-x}$ cells. Adapted and reprinted with permission from [77], © 2013 by Springer Nature Limited. (Color figure available online).

Nicely scale with the amount of Ga suboxide underneath the whole electrode area. Figure 51 shows the XPS Ga 3d core level spectra for differently programed cells recorded after electrode delamination. In the first sweep, a contribution from metallic Ga is visible which might be oxidized during biasing and, therefore, disappears for the subsequent sweeps. Afterwards, a reversible change of the relation between Ga$^{3+}$ and Ga$^+$ can be detected which is consistent with a concentration polarization (see Section 4.3.3) within the GaO$_x$ thin film. In particular, a negative bias results in the attraction of an oxygen deficiency (by the motion of ions) to the Pt electrode, increases the amount of Ga$^+$ and switches the cell to the LRS. In turn, positive biasing retracts the oxygen vacancies from the Pt electrode, reduces the amount of Ga$^+$ and switches the cell to the HRS.
In summary, these examples show clearly that a change in the valence of the cations in VCM cells do not only occur during electroforming but also during bipolar switching.

5.3. Memristive states

As shown in the previous subsection the different resistance states are characterized by a change in the valence state in the cation sublattice. This change in the valence state is induced by the movement of mobile ions that act as mobile dopants. While this observation has been proven experimentally as discussed in the previous section, it is not clear in which way the electronic...
Figure 51. PEEM analysis of Pt/GaOx/ITO thin film cells after electrode delamination [67]. Ga 3d corelevel spectra are depicted for cells programmed with different voltages. The black curves show the raw data and the red curves the fits taking into account contributions from different Ga valence states (P1:Ga$^{3+}$, P2:Ga$^{+}$, P3&P4:Ga$^0$). Adapted and reprinted with permission from [67], © 2014 by Springer Nature Limited. (Color figure available online).

Conduction processes are modulated by this atomic reconfiguration. Different models are discussed in the literature (cf. the comprehensive reviews of Lim et al. [323] and Funck et al. [324]), which can be categorized roughly into bulk switching (modulating a neutral barrier) and interface-dominated switching (modulating space charge and tunneling barriers) models, see Section 3.9. This will be addressed in this section.

In addition, the electronic conduction mechanism may change during switching. In the LRS, a bulk conduction mechanism (cf. Section 3.8) might dominate whereas the overall resistance is dominated by an interface-related conduction mechanism in the HRS. Furthermore, variations in the conduction mechanism for similar layer stacks might arise from the use of different fabrication methods, crystallinity and the actual atomic configuration during switching. In addition, the dominant conduction mechanism probably differ in filamentary systems and area-dependent ones. Keeping this in mind any experimental conduction mechanism study on ReRAMs is probably only valid for the presented specific device in the resistance regime under investigation. The analyses of the prevailing conduction mechanism might be obscured by occurring local Joule heating or atomic reconfigurations during measurement. In the following, different conduction models proposed in the literature will be discussed.
It must be clearly stated that the description of a separation into bulk or interface conduction is strongly simplified. In reality, there is always a blend of both types in which one of the contributions dominates. Our notation tries to highlight the dominating contribution. Furthermore, independent of this fact, the area-dependent switching is sometimes called interface-type switching [66]. However, as we explain in our paper, also filamentary switching may be interface dominated.

5.3.1. **Bulk switching models**

Bulk switching models discuss how the atomic configuration changes the overall resistance without taking into account the different metal/oxide interfaces. In these models, it is hard to explain a dependence of the conduction state on the choice of electrode materials. As mentioned above, it is important to have an asymmetry in the electrode contacts in order to obtain a bipolar switching either by metals with different work functions / oxygen affinities or by the introduction of a virtual cathode during electroforming (see Chapter 5).

Kamiya et al. reported [325] that from the perspective of Mott criteria (see Section 3.3.3), it follows directly that LRS can be described at the atomic level by a chain of oxygen vacancies (oxygen vacancy filament), as shown in Figure 52 for TiO₂, HfO₂ and Al₂O₃-based resistive switches. The calculated partial density of states seems to indicate an electronically conducting channel along these oxygen vacancy chains. However, it must be clearly stated that the electrons which carry the current always are located in the atomic orbitals of the adjacent cations, typically d-states in the case of transition metal oxides [326]. This fact is often hidden or covered in the corresponding papers. By disrupting this chain locally, a non-continuous partial charge density results and the HRS is obtained. As the calculations were performed at 0 K, the proposed HRS is probably still well-conducting at operating temperatures. Moreover, a chain of oxygen vacancies may overestimate the local oxygen vacancy concentration dramatically. In fact, a lower local concentration may already lead to the experimentally observed current densities, in particular when considering that the switching takes place on a larger area (a few nanometer-sized filaments to area-dependent switching).

In filamentary switching systems, the conductance depends highly on the local atomic configuration. Thus, it is quite reasonable that discrete conduction steps appear when the filament dimensions are as small as a few nanometers. Degraeve et al. observed the occurrence of discrete conduction steps in a TiN/HfO₂/Ti cell [327,328]. To describe these resistance states, the authors of this study introduced a quantum point conduction (QPC) model. In this model, a narrow constriction made of a few oxygen defects is placed between to bigger reservoirs of oxygen defects.
and thus called it Hour glass model. The constriction is characterized by a parabolic energy valley in radial direction to the electron path and a parabolic barrier height along the electron path. The electrons can only pass this constriction at certain energy levels, energetic positions of which depend on the number of atoms in the construction. A QPC model has been also applied by other groups to SiOx [329], HfOx [330], and TaOx [331]. The QPC model shows a very good fit to the experimental data, but is a purely mathematical model. Whereas the fit is very good, the QPC model does not give any information about the position of the constriction within the oxide layer. The model is based on the Landauer-Büttiker formalism, which can be applied to describe the transport across the metal/oxide interfaces as well [324]. Thus, the barrier does not need to be within the oxide (away from the interface) it can be also interpreted as the barrier at the metal/oxide interfaces. The discrete levels predicted by the QPC model are sometimes viewed as parallel physical conduction paths [330]. In fact, it could be also parallel paths occurring at different energy levels. Moreover, the occurrence of ballistic transport at the operating temperatures > 300 K is doubtful from a theoretic point of view [332]. Nevertheless, the occurrence of discrete resistance steps can be attributed the finite size of the filament and the limited number of possible atomic configurations.

In order to identify the dominant electron conduction mechanism temperature-dependent measurements of the conductivity for different resistance states are inevitable. Fang et al. [333] reported on the measured temperature-dependence of a Pt/HfO2/TiN device for differently programmed resistance states (read at 0.1 V) from 300 K down to 4 K. For all programmed states LRS, IRS, HRS a semiconducting behavior is observed. Two different regimes were found: a saturation regime with a small temperature dependence up to 77 K and a stronger temperature-dependence at higher temperatures. The authors of this study propose that this transition can be explained by a transition between variable range hopping and nearest neighbor hopping, whereas the oxygen defects induce the defect states. Graves et al. investigated the temperature and field dependence of Pt/Ta2O5/Ta cells in different resistance states at low voltages [334]. The measurement data showed a very weak temperature dependence of the states below room temperature and a strong current increase with temperature above room temperature. The data was interpreted by two parallel conduction modes, Schottky emission with a high temperature dependence and Mott hopping showing a low temperature dependence. While the Schottky emission is suggested to be only weakly state-dependent, the Mott hopping is linked by the authors to the resistance state, i.e. the number of defects in the oxide layer. Trap-assisted tunneling TAT is quite often reported as prevailing conduction mechanism in the HRS. In this case, oxygen vacancies are considered as trap states and the electrons tunnel from one trap to the other. In the multi-phonon TAT model proposed by Bersuker et al. [194,335] the capture and emission of electrons leads to a heat dissipation at the position of the defect by energy relaxation. By assuming multiple current paths along trap states the authors could fit the temperature dependence of a HfOx based stack. This conduction model required the traps to lie deep in the band gap in order to prevent the excitation of electrons into the conduction band. As discussed above and in Chapter 3, the electrons provided by the oxygen vacancies appear in typically the d-states of the transition metal ion. Thus, the electron transport should occur by hopping between different reduced valent transition metal sites (subband conduction) or as transport in the conduction band. Moreover, the assumption of pure TAT transport leads to a very high amount of oxygen vacancies in the LRS to explain the high currents [336–338]. For high concentration of vacancies, and thus electrons, the Fermi-level will be closer to the conduction band. In this case a transition from TAT to band conduction should be expected and a smaller VO concentration is possibly required to explain the LRS conduction. For very reduced TaOx, with x ≈ 1, a metallic conduction can appear [339,340]. Such reduced oxide may appear under switching currents of mA which are too high for a memory operation. Nevertheless, the study shows that
an oxidized Ta electrode will be still highly conductive and can be used as oxygen exchange layer.

5.3.2. **Interface-dominated switching models**

The conduction mechanisms in conjunction with the bulk switching discussed in the previous section neglect the electron injection from the metal to the oxide material. The choice of the electrode material, however, plays already an important role during forming (Chapter 5). In addition, to enable a bipolar switching mechanism, the cell has to be asymmetric (see Section 2.2) [11]. In fact, quite often asymmetric $I–V$ characteristics with respect to the voltage polarity are reported and the modulation of the electron transport across one of the metal/oxide interfaces has been proposed as the dominating resistance switching mechanism [11,119,244,271,318,324,341–350].

Yang and co-workers fabricated a four-pad geometry with two Ti and two Pt electrodes on a TiO$_2$ crystalline sample as shown in Figure 53(a) [245]. In contact with the Ti electrode the TiO$_2$ is supposed to be locally reduced and an ohmic contact should result as illustrated in (Figure 53(b)). The Pt/TiO$_2$ contact, however, is supposed to form a Schottky contact. By measuring the current flowing between different pads, the authors could support this hypothesis and identify the blocking interface. Whereas the $I–V$ curve measured between the two Ti contacts exhibits an ohmic resistance, the $I–V$ curves measured between one Pt and one Ti contact showed a rectifying characteristic with the rectifying contact located at the Pt/TiO$_2$ interface (Figure 53(c)). The $I–V$ characteristics measured between the two Pt contacts exhibit a back-to-back diode characteristic.

In further experiments, the authors performed switching between one Pt and one Ti electrode and identified that the switching took place at the Pt/TiO$_2$ interface by modifying its current blocking properties [245]. In a further study, Yang et al. demonstrated that the chemical properties of the metal/oxide interface determine the barrier height rather than the work function difference [318]. For a non-noble electrode, the electrode is oxidized and the oxide underneath is reduced. The resulting high concentration of positively charged oxygen vacancies lowers the Schottky barrier and eventually an ohmic contact results. For noble metals, no redox reaction occurs and the Schottky barrier remains.

As the conduction mechanism depends on the local sub-stoichiometry (e.g. the oxygen vacancy concentration) it is detrimental to know the exact distribution in the LRS and HRS. Bäumer et al. used in-operando PEEM measurements to analyze the change in the donor concentration during resistive switching in a Graphene/SrTiO$_3$(20 nm)/Nb-doped SrTiO$_3$ VCM cell [271] (cf. Figure 48). By acquiring spatially resolved O K-edge images, a switching filament of about 500 nm in diameter can be identified. Moreover, the charge carrier concentration in the upper 2–3 nm of the filament can be quantified to be $1.5\cdot10^{21}$ electrons per cm$^{-3}$ in the LRS and $6.7\cdot10^{20}$ electrons per cm$^{-3}$ in the HRS. Trap-assisted transport processes can be excluded since the experimentally determined position of the defect levels is situated below the Fermi-energy [351]. The extracted filament dimensions and carrier densities illustrated in Figure 54 (panels a and b) were used as input into a device simulation model that includes drift-diffusion transport of the electronic carriers in the thin SrTiO$_3$ thin film and the Nb:SrTiO$_3$ bottom electrode and a Schottky barrier at the top electrode interface. The simulated $I–V$ characteristics show a very good agreement with the experimental data (Figure 54(c)). The simulation results reveal that the resistance change between LRS and HRS of more than 2 orders of magnitude by changing the electronic carrier concentration by only a factor of 2–3 can be explained by the dopant induced barrier modulation at the graphene/SrTiO$_3$ interface (cf. Figure 54(d)). Due to the high dopant concentration, the electrostatic barrier is very thin and electron tunneling dominates over thermionic emission. This leads to very symmetric $I–V$ characteristics with respect to the voltage
Figure 53. Junctions on single-crystal TiO$_2$ show the role of the interfaces in determining the electrical behavior. (a) Schematic of the electrodes and the single crystal. Four adjacent pads were deposited as pairs of Pt (blue) and Ti/Pt (yellow) contacts. (b) Energy diagram showing the low oxygen vacancy concentration under the Pt pads maintaining the Schottky-like barrier (denoted by a rectifier) between Pt and TiO$_2$ to produce rectifying junctions, whereas the high vacancy concentration at the interfaces under the Ti/Pt pads collapses the Schottky-like barrier and produces ohmic contacts (denoted by a resistor). $F_b$ and $w$ are the electronic barrier height and width. (c), The four-probe $I-V$ curves between the combinations of the four pads in (a). The insets to these $I-V$ diagrams are the corresponding equivalent circuit diagrams consisting of two electronic elements (rectifier or resistor) in series. Adapted and reprinted with permission from [245], © 2008 by Springer Nature Limited.

polarity for both resistance states. It is important to note that this system shows a 8W switching polarity with the graphene top electrode being the AE (cf. Section 5.7).

Funck et al. investigated the temperature and field-dependence of the interface-dominated conduction of a Pt/SrTiO$_3$/Nb-doped SrTiO$_3$ system. Figure 55(a) shows the temperature and field-dependence for negative voltages applied to the Pt AE at different programmed resistance states. Two major features can be extracted from the experimental data. At lower voltage amplitudes the current increases strongly with temperature and below a certain, so-called intersection voltage, the trend is inverted. This intersection voltage shifts to lower voltages (i.e. to higher voltage amplitudes) for more resistive states. In the forward direction of the Pt/SrTiO$_3$ interface, the current increases exponentially, and then saturates, accompanied by a transition in the temperature dependence. This complex behavior was analyzed using DFT-NEGF simulations for a negative polarity [349] and using a continuum model for all observed states [350]. Based on the simulation results, a consistent model for the experimental observations could be deduced. In reverse direction and high voltage amplitudes (cf. Figure 55(c)), the electrons can tunnel directly at the Fermi-edge into the conduction band of the SrTiO$_3$. In addition, the transport of the electrons in the conduction band becomes dominating and the current decreases with increasing temperature due to phonon scattering. At lower voltage amplitudes, however, the conduction band edge lies above the metal Fermi level. Thus, the electrons need a thermal activation to be injected into the conduction band (cf. Figure 55(d)). This thermal activation accounts for the increase in current with increasing temperature in this regime. The transmission through the barrier is determined by the width of the triangular barrier, which is determined by the width of the
space charge zone. For a decreasing oxygen vacancy concentration, the space charge zone widens and the effective barrier increases based on the Schottky-lowering effect. Thus, lower concentrations will lead to higher resistive states. The change of the temperature dependence is attributed to the change of the dominating conduction mechanism (interface to bulk-limited). For higher resistive states, the interface becomes more dominating and thus the intersection regime is shifted to higher voltage magnitudes. In the forward direction of the Pt/SrTiO$_3$ interface, the increasing voltage shifts the conduction band upwards and the tunneling barrier thickness decreases, leading to an exponential increase in current (c.f. Figure 55(e)). The electrons can tunnel directly at the kink of the conduction band edge to the metal. The temperature dependence in this configurations stems from the thermal activation of the electrons at the Nb:STO/SrTiO$_3$ interfaces. As for the negative polarity, the dominating conduction mechanism changes at high voltages to band limitation, which leads to the change of the temperature-dependence. It has to be noted that the electrons tunnel through the Pt/SrTiO$_3$ interface in all cases and there is no contribution of thermal emission over the barrier.

In a further study, the conduction states of thin Pt/HfO$_{2-x}$/Hf cells with different oxygen defect concentrations were investigated using DFT-NEGF calculations [324]. It is observed that
Figure 55. Temperature and field-dependent study of multilevel switching of Pt/SrTiO$_3$ based devices; Adapted and reprinted with permission from [350], © 2020 by the American Physical Society. (a) I–V–T measurements of several resistance states programmed. (b) Experimental LRS current voltage measurement for different temperatures for the Schottky forward direction. (c)–(f) Schottky barrier sketch at different applied voltages. The curved red line describes the thermal activation process. The white dashed arrow indicates the tunneling path. The black line describes the conduction band. The $\Delta \Phi_{B0}$ indicates the barrier lowering effect.

the oxygen vacancies induce defect states deep within the band gap around the Fermi levels of the metals in the stack. The electron transport occurs in this case by tunneling via the defect states, i.e. an interface-limited trap-assisted tunneling mechanism, which was also proposed by S. Yu et al. [348]. S. Yu and co-authors proposed that the electrons tunnel to the trap nearest to the AE. It was, however, shown by Funck et al. that the electron tunnel to the defect state at the metal Fermi level that is nearest to the Pt electrode [324]. The position of this defect is determined by the extension of the space charge zone at the Pt/HfO$_2$ interface. Thus, the nearest defect state may be at a too high energetic level.

Finally, a comprehensive model for the electron transport in VCM cells was proposed in Ref. [324]. Two different types were identified (Figure 56). The so-called type 1 conduction comprises materials with shallow oxygen vacancy defect states as for example SrTiO$_3$ or TiO$_2$ [92]. In type 1 devices, thermally excited electrons tunnel from the metal into the conduction band as discussed
Figure 56. The mechanisms of the two suggested conduction mechanisms are sketched within the figure.

Type 1 describes the band transport, where the electron transfer at the active electrode (AE) is given by tunneling into the conduction band. This transport is expected for oxygen vacancies with a shallow defect state and results in a strong exponential $I-V$ dependence. In Type 2, the electron transport across the active electrode is obtained by a tunneling from the oxygen vacancy defect states at the end of the Schottky depletion zone to the metal electrode. This conduction mechanism is expected for deep oxygen vacancies and results in a less strong $I-V$ dependence compared to Type 1. Adapted and reprinted with permission from [324, 352], © 2020 by the American Physical Society.

before. This leads to a high temperature dependence. The resulting barrier has a more triangular shape, which leads to highly non-linear (exponential) $I-V$ characteristics. In the so-called type 2 devices (e.g. comprising HfO$_2$, ZrO$_2$ or Ta$_2$O$_5$), the oxygen vacancies induce deep defect states within the band gap. In this case, the electron transport can be described by an interface limited TAT mechanism at the Fermi level. This leads to a very weak temperature dependence of the electron transport. In addition, the high barrier has a more trapezoidal shape and thus, the $I-V$ characteristic becomes more linear. The switching is attributed in both cases to the change of the defect concentration close to the active interface. An increase in the defect concentration leads to a shortening of the depletion zone and thus a reduction of the effective tunneling length. Hence, the resistance decreases. It should be noted that this ‘static’ picture of switching is independent of the switching polarity, i.e. 8W and CW8 switching.

In summary, it is fair to say that determining the correct conduction mechanism in the different states of a ReRAM devices is challenging, due to the different positions of the defect states on the energy scale, in space and at the interface. In this respect, the Pt/SrTiO$_3$ example was particularly lucky because it was possible to determine all relevant properties experimentally and use them as input for simulations.
Based on the previous discussions, the authors conclude that the electron conduction in VCM cells is dominated by an injection of electrons through a barrier at the AE into the conduction band or a sub-band/single defect states within the band gap.

5.4. Filamentary switching

The filamentary switching, in particular with the C8W polarity, can be regarded as the most frequently observed resistive switching mode. For this reason, we call this the standard resistive switching mode as mentioned above. It has been mentioned already in Chapters 2 and 3 and it was treated in many details in Chapter 5 on electroforming. This type of switching is also the most mature one. In 2011, Samsung reported on a Pt/Ta2O5/TaOx/Pt system with record endurance data [314]. Panasonic released the first commercial Ta2O5 filamentary, C8W switching VCM cell product, i.e., a microcontroller with embedded ReRAM, in 2013. [353]. Later, Panasonic ReRAM devices at the 28 nm CMOS node [251]. Small ReRAM modules are also available from Fujitsu. WeebitNano is offering ReRAM based on filamentary switching SiO2 [354], TSMC integrated filamentary ReRAM at the 40 nm node [355,356], the 28 nm node [357], and the 22 nm node [358]. In addition, CMP foundry offers HfO2 based filamentary VCM cells integrated with 130 nm CMOS technology through LETI [359,360].

As a typical example, the resistive switching of TaOx VCM cells is shown in Figure 57. In the corresponding studies, the impact of the OE metal such as W, Ta, Ti, and Hf [361] as well as the opportunity of a multilevel resistive switching (i.e., a 3-bit cell) by controlling the RESET process in a Pt/TaOx/W/Pt cell [362] were investigated. It was found that W and Ta with a modest free energy for metal oxide and, associated, a modest formation energy of oxygen vacancies at the TaOx/OE interface leads to highly stable resistive switching behavior. This has shown that obviously also the oxygen exchange reaction at the OE metal electrode plays an important role in determining the resistance states. In line with this observation, Ti and Hf as OE metals lead to a too strong oxygen exchange reaction and, as a consequence, too strongly reduced TaOx layer and a stuck in the LRS during the first switching cycle [361].
As described in detail in Chapter 5, there have been a manifold of proofs of the formation of conducting filaments during the electroforming process using spectroscopical, electron microscopical, as well as scanning probe techniques. In particular, a spectroscopical evidence of a filamentary switching has been given already in Section 5.2 in Figure 48 by in-operando PEEM measurements on SrTiO₃ cells in which an electron-transparent graphene top electrode has been used.

Moreover, we reported on the use of a pressure-modulated conductive tip AFM (PMCM) technique in order to visualize the location of the conducting filament through a top electrode (Figure 27 in Section 4.2). Figure 58 complements this description by showing the reversible bipolar switching utilizing the same technique for a Pt/TiO₂/TiO₂₋ₓ/Pt where the Pt/TiO₂ constitutes the active electrode. The filament location clearly visible in the LRS (ON state) vanishes after switching to the HRS (OFF state). Details are reported in Ref. [69].

As a background for the description which follows, we will introduce the resistance components of a (simplified) filamentary system. When the cell resistance $R_{\text{cell}} = R_{\text{HRS}} \cdot R_{\text{LRS}}$ is discussed, the composition of the filament resistance $R_{\text{fil}}$ by the serial connection of the resistance of the disc (barrier) region at the active interface $R_{\text{disc}}$ and the resistance of the plug $R_{\text{plug}}$
Figure 59. Sketch of the resistance components which make up the cell resistance $R_{\text{cell}}$. In parallel to the filament resistance $R_{\text{fil}}$ there is the resistance of the unformed, pristine film, $R_{\text{pris}}$. The filament resistance $R_{\text{fil}}$ is composed of the plug resistance $R_{\text{plug}}$ and the $R_{\text{disc}}$ resistance. Only $R_{\text{disc}}$ contributes to the resistive switching. Reprinted with permission from [13], © 2012 by Wiley-VCH.

(including the ohmic counter electrode) has to be taken into account, i.e. $R_{\text{fil}} = R_{\text{disc}} + R_{\text{plug}}$. In addition, in parallel there is the resistance of the pristine area, $R_{\text{pris}}$, around the filament (Figure 59):

$$R_{\text{cell}} = \frac{(R_{\text{disc}} + R_{\text{plug}})R_{\text{pris}}}{R_{\text{disc}} + R_{\text{plug}} + R_{\text{pris}}}$$

(27)

It can be assumed that only the $R_{\text{disc}}$ part of $R_{\text{cell}}$ will change during the resistive switching. Typically, $R_{\text{plug}}$ is small compared to disc resistance in the HRS, i.e. $R_{\text{disc,HRS}} \gg R_{\text{plug}}$, so that the HRS/LRS resistance ratio of the filament is determined by the usual large HRS/LRS swing of $R_{\text{disc}}$. However, for large area cells the HRS/LRS resistance ratio deteriorates because of $R_{\text{pris}}$. For this reason, $R_{\text{HRS}}/R_{\text{LRS}}$ ratio of a cell usually improves with lateral down scaling. A word of caution should be added to note that the expression $R_{\text{fil}} = R_{\text{disc}} + R_{\text{plug}}$ is only an approximation because the resistance contributions influence one another due to the nanometer dimensions. It should be noted that the disc is sometimes called the gap region throughout this text.

In a brief anticipation to the detailed presentation later in this section, the band diagram of the active interface in the LRS and HRS is shown schematically in Figure 60. Because of the small dimensions, the exact potential landscape will be determined by the local density of states (LDOS) which deviates from this macroscopically defined band picture. For a qualitative discussion of the states, however, the band picture approximation is sufficiently precise. In the HRS, a significant energy barrier exists which originates from the fact that the disc region is fully oxidized. For homogeneous layer systems (Figure 45(a)), this corresponds to a full depletion of oxygen vacancies. In the case of heterogeneous bi-layer systems, there are either no charge carriers in the disc layer or a certain density of negative ions ($O^{2-}$) which further increase the barrier height. The temperature dependence of $R_{\text{HRS}}$ typically shows a thermally activated characteristic, as has been described in detail above. In our picture, it will be mainly determined by $R_{\text{disc,HRS}}(T)$ if the cells – and, hence, the conduction through $R_{\text{pris}}$ – are not too large. Presumably $R_{\text{disc,HRS}}(T)$ is caused by a thermionic transmission (TE) over the barrier. An additional current contribution may be due to direct tunneling or Fowler–Nordheim tunneling i.e. field emission (FE) as sketched in Figure 24. In any case, as expected for any potential barrier, $R_{\text{disc,HRS}}$ shows a strong voltage dependence, i.e. it is a highly nonlinear resistor. This gives rise to pronouncedly nonlinear $I$–$V$
characteristics (as shown in many examples in this paper and contributes to the Joule heating effect required for fast switching kinetics (Chapter 7).

During SET, oxygen ions are removed from the disc. For homogeneous layer systems, this is identical with an injection of oxygen vacancies which will be (at least partially) compensated by electrons. In the case of the heterogeneous bi-layer systems, the overall process is the same while the vacancy picture might be more complicated, depending on the type of oxide. In any case, the extraction of oxygen ions will lead to a (chemical) reduction of the disc region which in turn will result in a significant decrease in the barrier width and/or height (Figure 60). In the LRS, as a consequence, there is a pronounced increase of the thermionic conduction and/or of the tunneling conduction. $R_{\text{LRS}}$ is determined by the series combination $R_{\text{disc,LRS}}$ and $R_{\text{plug}}$, while $R_{\text{pris}}$ can be neglected. Now, due to the reduced barrier, the situation will be given by $R_{\text{disc,HRS}} \ll R_{\text{plug}}$, and therefore, the temperature dependence of $R_{\text{LRS}}$ will be dominated by the latter. $R_{\text{plug}}(T)$ is supposed to be weak because the n–conducting oxide represents a semiconductor in the saturation regime or (more often) a degenerate semiconductor which can be described as a metal (Chapter 3). During RESET, oxygen ions are attracted into the disc, leading to the HRS again.

In order to explain the switching mechanism, different models have been proposed. A model that can explain the $I–V$ characteristics of a VCM cell needs to account for the ionic processes and the electron transport. The ionic processes comprise the movement of ionic defects i.e. mobile donors such as oxygen vacancies or cation interstitials within the switching layer and local redox reactions. An important property is the nonlinearity of the ionic processes that drive the resistance change. In order to identify the dominating ionic processes, the switching kinetics of the investigated device needs to be analyzed over a wide voltage range. This aspect will be discussed in Section 5.7. Here, the focus is on the simulation of the $I–V$ characteristics obtained by applying triangular voltage sweep. To model the electronic transport the ionic dopant distribution needs to be linked to the electronic transport mechanism. Based on the respective underlying conduction model, we will discuss bulk- and interface-dominated switching models in the following.
5.4.1. Bulk switching models

Larentis et al. developed the first fully dynamic axisymmetric 2D model that is capable of simulating the $I$–$V$ characteristics of a VCM cell [363]. In this model a TiN/HfO$_x$/TiN VCM cell was considered. Here, the local conductivity $\sigma$ is a function of the donor dopant concentration $N_D$, which can be oxygen vacancies, whereas space-charge layer and tunneling conduction mechanisms are neglected. Instead, the modulation of a neutral barrier is considered. The dopant movement is modeled using the drift-diffusion equation

$$\frac{\partial N_D}{\partial t} = \nabla (D \nabla N_D - \mu E N_D).$$

(28)

Here, $D$ is the diffusion constant of the mobile ions, and $\mu$ their mobility. The electric field $E$ and the local temperature $T$ are calculated solving the continuity equation

$$\nabla \sigma \nabla \varphi = 0$$

(29)

for the potential $\varphi$ coupled with Fourier’s heating law

$$-\nabla k_{th} \nabla T = \sigma |\nabla \varphi|^2$$

(30)

where $k_{th}$ denotes the thermal conductivity. The initial dopant concentration was assumed to be homogeneous in a filamentary region, thus describing the LRS. By applying a positive voltage, the positive dopants at the hottest spot within the filament start to migrate within the electric field. This leads to an ‘opening’ of the filament close to the anode as illustrated in Figure 61(a). The resulting depleted gap region describes a region with low donor concentration, which is still as in the area surrounding the filamentary region. By reversing the polarity the dopant concentration in the depleted gap region increases again and the LRS is achieved. The corresponding simulated $I$–$V$ characteristic reproduces the experimental behavior of a TiN/HfO$_x$/TiN device quite nicely as shown in Figure 61(b), which demonstrates that the redistribution of dopants, such as oxygen vacancies, can describe the switching in general. The simulation model features an abrupt SET switching and the multilevel programming during the gradual RESET by changing $V_{stop}$. For a certain voltage $V_{stop}$, the local electric field and temperature drop so much that no further lengthening of the depleted gap region occurs. In order to lengthen the depleted gap region even more a higher voltage needs to be applied. The stop voltage also influences the following SET voltage. When the depleted gap region is larger, most of the voltage drops in this region. Thus, the temperature and electric field outside the depleted region is smaller and it is harder to move dopants from outside the depleted gap into it. Thus, a higher voltage is required to drive the SET process. As the simulated cell is completely symmetric, the homogenous filament assumed as initial configuration in the previous simulation can be RESET with both voltage polarities. Thus, it is straightforward to simulate CS behavior using the same model [48]. Figure 62 shows the simulated $I$–$V$ characteristic and the corresponding concentration contours for a CS device. The simulation starts with the filament in the HRS. By applying a positive voltage, the mobile dopants move downwards, the filament becomes more homogenous resembling the LRS. By further increasing the voltage, more dopants move downwards and a filament ‘gap’ opens close to the top electrode and the device is again in the HRS. By reversing the applied voltage polarity, the cell will undergo the same transition HRS $\rightarrow$ LRS $\rightarrow$ HRS. During complementary switching, the electrically active interface changes. By using a current compliance that prevents the RESET transition after the SET process, bipolar switching can be obtained again [49,364].

In this model, the temperature at the metal/oxide interfaces was set to 300 K, which implies an infinitely fast heat transport via the electrodes. Thus, dopants always remain at the electrodes.
Figure 61. Resistive switching of a TiN/HfOₓ/TiN device after electroforming. (a) 3D contour plot of doping density, corresponding to points having \( N_D = 6 \times 10^{20} \text{ cm}^{-3} \), illustrating the evolution of the conducting filament (CF) shape during the gradual RESET transition, providing an insight into the length of the depleted gap \( \Delta \). (b) Measured and (c) calculated \( I-V \) curves, showing RESET \( (V > 0) \) and SET \( (V < 0) \) transitions, obtained by applying triangular voltage sweeps. First, the RESET sweep, applied to an initial ON state with \( R = 400 \Omega \), is interrupted at \( V_{\text{stop}} \) for preparing a HRS state with variable \( R \). Then, the SET sweep is applied, showing that \( V_{\text{SET}} \) increases with \( V_{\text{stop}} \), hence with \( R \). Adapted and reprinted with permission from [363], © 2012, IEEE.

and the filament breaks closer to the middle of the oxide layer. When the heat dissipation via the electrodes is considered for solving the heat transport equation, the temperature at the metal/oxide interfaces is high enough to enable ion movement away from the interface [286,365–367]. A further issue with this bulk switching model is the complete symmetry of the equation system with respect to the metal electrodes. As a result, the model will predict that CS behavior will occur in every ReRAM stack. There are, however, many stacks that always show bipolar switching, in particular in the case of asymmetric electrode materials. The described model has been adapted by Kim et al. to simulate bipolar, multilevel and complementary switching in Pt/TaOₓ/Ta₂O₅/Pt bilayer structures [286,365]. When the TaOₓ base layer is very oxygen-deficient, the device exhibits typical bipolar switching. In this case, the TaOₓ base layer contains a high concentration of oxygen vacancies and can act as an infinite reservoir of vacancies, i.e. as an oxygen exchange layer. For a higher oxygen content in the TaOₓ layer, complementary resistive switching could be simulated [365]. The transition from pure bipolar to complementary switching is thus explained as a result of the reduction in the oxygen vacancy inhomogeneity across the switching layer. While the previously described modeling approach was developed for filamentary switching,
Aoki et al., however, applied a very similar model for area-dependent switching in GaO<sub>x</sub>-based cells [67].

While the previous model is based on the ion migration in the parallel direction, sometimes a resistive switching mechanism and electroforming process (cf. Chapter 5) based on a perpendicular movement of the defects due to temperature gradients have been proposed [88,282,284,368,369]. So far, no dynamic simulation model could be presented that consistently describes the SET and RESET switching based on such thermo-diffusion processes. In contrast, the simulation model that included thermo-diffusion did not show a dominating contribution of thermo-diffusion to the switching process [286,366,367]. Typically, the field dependent migration of the charged particles dominates the process. This is also consistent with the bipolar operation principle. The perpendicular temperature gradient, however, will have the same sign independent of the applied voltage polarity, which makes it hard to explain consistently a bipolar switching operation. Nevertheless, thermo-diffusion might play a role during the formation of the filament as discussed in Chapter 5 or can explain the diffusion of non-charged species in the systems [284].

Several kinetic Monte Carlo (kMC) models for the simulation of VCM cells have been developed [191,192,194,338,370,371]. All these models speculate on the generation and recombination of O<sub>i</sub>–V<sub>O</sub> anti-Frenkel pairs within the bulk of the oxide layer. While the first models only cover the formation of the conductive filament, the model was further improved to cover the first RESET process after forming and later the whole switching cycle. In the most advanced version of this model, the electronic conduction mechanism is described as a trap-assisted tunneling (TAT) mechanism, in which the oxygen vacancies act as electron traps. When the oxygen vacancy
concentration exceeds a certain limiting value, the electron conduction mechanism changes from TAT to electron drift in the conduction band. The kMC model we just described has been used to simulate the forming and subsequent switching process of a TiN/TiO$_x$/HfO$_{2-x}$/TiN device stack [191]. The general switching behavior, such as switching/forming voltages and the switching polarity, can be reproduced using this model. The current levels, however, differ significantly. During the forming process, for example, the current jump occurs at currents higher than 1 μA, whereas it appears at approximately 100 pA in the simulation (as shown in [191]). In addition, the slope of the $I$–$V$ curves in the experimental data is much steeper, which might indicate a different conduction mechanism as implemented in the simulation model. The RESET current is much lower and the switching more abrupt as in the experimental data (cf. blue curves). This might be related to a too small resulting filament in the simulations or an incorrect conduction mechanism. It should be noted that stable switching was only achieved in this model if an oxygen exchange was considered at the TiO$_x$/HfO$_{2-x}$ interface. One major problem of these kMC models that rely on the VO generation/recombination mechanism is that the resulting anti-Frenkel pairs are in fact very unstable in oxides. It has been demonstrated that using MD simulations that VO–Oi anti-Frenkel pairs should recombine without activation energy within the timeframe of ps [372] and, hence, they are completely irrelevant for the process. Moreover, ab-initio simulations of a Hf/HfO$_2$ system revealed that the formation of oxygen vacancies is much more favorable to occur at the Hf/HfO$_2$ interface by incorporation of O into the Hf metal than the formation of V$_{O–O_i}$ in the HfO$_2$ bulk [373]. Furthermore, Schie et al. [374]employed atomistic simulations based on molecular static and molecular dynamic approaches which show that the critical field strength to create an anti-Frenkel pair exceeds the dielectric breakdown voltage of the most insulating oxides. Furthermore, the ps recombination times were confirmed. We therefore conclude that there is no physically reasonable route for generating anti-Frenkel pairs in oxides such as HfO$_2$.

Abbaspour et al. developed an alternative kMC model that allows the generation of oxygen vacancies only as a result of an oxygen exchange or oxygen evolution reaction at one of the electrodes [302,375–377]. Moreover, the conduction model included the electrostatic barriers at the interface. Thus, we will discuss this model in the following section.

The dynamic Hourglass model could be seen as a very simplified version of a multidimensional KMC model [327]. The model assumes two reservoirs which are connected by a narrow constriction as mentioned in section 5.3.1. The amount of defects within the reservoir defines the conduction via the QPC model. By applying a voltage defects are jumping from the reservoirs to the constriction and back. In this way, the model could reproduce the dynamic switching behavior, and in particular, the gradual RESET behavior. As the jumps of the defects are considered to occur with a certain probability also resistance fluctuations could be explained [378]. The existence of a narrow constriction was motivated by the shape of the filaments observed with 3D AFM tomography (cf. Figure 29). This model is quite abstract and the physical origin of the reservoirs is not clear. In principle, one of the reservoirs (the smaller one) could be associated with the AE, the constriction describes the transport through AE/oxide barrier and the second reservoir could be interpreted as the filament ‘plug’ region. As already discussed in section 5.3.1 in conjecture with the QPC model, the constriction does not need to be considered as geometric constriction, but could be also considered as a small number of energy channels for the electron transport, e.g. through a metal/oxide barrier. With this interpretation of the reservoirs the Hourglass model in fact becomes an interface-dominated switching model.

5.4.2. Interface-dominated switching models

In the previously presented bulk switching models the influence of contact potentials at the metal/oxide interface on the conduction mechanism is neglected and only a concentration
polarization of the ionic defects is considered. The importance of this electron transport mechanism is evident in the asymmetric current characteristics observed in the HRS of different ReRAM cells. [271,318,379,380] A first 1D dynamic model including electron transport across a Schottky-type barrier was developed by Noman et al. to investigate the retention properties of ReRAM devices [381]. However, since this model does not account for Joule heating, it cannot be employed to study the switching dynamics. Lee et al. presented a similar model including Joule heating to investigate the switching polarity of the hysteresis loop, i.e. 8W vs. C8W-switching, as a function of donor distribution. In their approach, however, the conduction in an insulating bulk zone (depletion region) behind the Schottky contact is neglected, which makes their conclusions questionable [343]. In 2015 Marchewka et al. presented fully dynamic 1D and 2D models that couples electronic and ionic transport, Joule heating and Schottky-barrier related transport [322,366]. To calculate the potential within the oxide layer $\psi$ the Poisson equation

$$\nabla (\epsilon \nabla \psi) = -e(n - N_{\text{VO}}^+ - 2N_{\text{VO}}^{2+})$$  \hspace{1cm} (31)$$

is solved. Here, minority carries are neglected and it is assumed that the oxygen vacancies are twofold ionizable and $N_{\text{VO}}^+$ ($N_{\text{VO}}^{2+}$) denotes the singly (doubly) ionized oxygen vacancy concentration. This equation is solved along with the steady-state drift-diffusion equation for electrons

$$\nabla (\mu_n n \nabla \psi - D_n \nabla n) = \pm \partial j_{\text{n,tunnel}} / \partial x$$  \hspace{1cm} (32)$$

the time-dependent drift diffusion equation for the doubly ionized oxygen vacancies

$$\partial N_{\text{VO}}^{2+} / \partial t - \nabla (\mu_{\text{VO}} N_{\text{VO}}^{2+} \nabla \psi - D_{\text{VO}} \nabla N_{\text{VO}}^{2+}) = -R_{\text{VO},2}$$  \hspace{1cm} (33)$$

the rate equation for the immobile singly ionized oxygen vacancies

$$\partial N_{\text{VO}}^+ / \partial t = -R_{\text{VO},1}$$  \hspace{1cm} (34)$$

and the rate equation for the neutral oxygen vacancies

$$\partial N_{\text{VO}}^0 / \partial t = -R_{\text{VO},0}$$  \hspace{1cm} (35)$$

In Equations (32)–(35), $\epsilon$ is the permittivity of the oxide, $n$ the electron concentration, $\mu_n$ ($\mu_{\text{VO}}$) the electron (oxygen vacancy) mobility, $D_n$ ($D_{\text{VO}}$) the electron (oxygen vacancy) diffusion coefficient and $N_{\text{VO}}^0$ the neutral oxygen vacancy concentration. $R_{\text{VO},2}$, $R_{\text{VO},1}$ and $R_{\text{VO},0}$ represent the reaction rates that are derived from the donor ionization reactions along with oxygen vacancy ionization statistics [366]. The local temperature in the device structure is estimated according to Equation (50) in the 1D model [322] and the static heat equation is solved in the 2D model [366]. The temperature enters in the reaction rates, mobilities and diffusion coefficients.

The electron transport across the metal–insulator interfaces has two different contributions, as mentioned in Section 3.10: Thermionic emission occurs for electrons with energy $E_x$ higher than the conduction band edge maximum $E_{C,\text{max}}$ and tunneling occurs for energies between the conduction band minimum and maximum. The thermionic emission contribution is calculated by
integration over the respective energies according to

\[ j_{\text{TE}}(x_i) = \frac{A^* T}{k_B} \int_{E_{C}(x_i)}^{\infty} N_{\text{supply}}(E_x) dE_x \] (36)

while the tunneling contribution is calculated as

\[ j_{\text{tunnel}}(x) = \frac{A^* T}{k_B} \int_{E_{C,min}(x)}^{E_{C,max}(x)} T(E_x) N_{\text{supply}}(E_x) dE_x \] (37)

In (36) and (37), \( A^* \) denotes the effective Richardson constant, \( N_{\text{supply}}(E_x) \) the supply function that describes the supply with carriers and is derived by integration of the occupancy function on both sides of the barrier, and \( T(E_x) \) is the transmission coefficient that is calculated using the WKB approximation. The tunneling current enters as a local generation/recombination rate into the right-hand side of the drift-diffusion equation for electrons (32). This set of equations is complemented by appropriate boundary conditions as outlined in [322,366]. In addition, image-force-induced barrier lowering in the presence of the oxygen vacancies is included within the 1D model.

The 1D model was used to simulate the general switching behavior of a MIM stack [322]. In a first simulation asymmetric barrier heights of 0.7 and 0.1 eV were chosen, which resemble the electrically active (AE) and the ohmic interface (OE), respectively. Starting from an initial homogeneous donor distribution a first RESET sweep was simulated as illustrated as black curve in Figure 63 leaving the cell in the HRS. The subsequent bipolar voltage sweeps shown in blue features an abrupt SET switching at negative voltages and a gradual RESET transition at positive voltages. This switching polarity is consistent with the C8W-switching mode. After the initial RESET sweep the oxygen vacancy concentration is low at the active interface, which results in a high barrier at the active interface as indicated by the color curve in Figure 63(c,d). During the SET sweep the oxygen vacancies are accumulated at the active interface and depleted at the ohmic interface (green curve). Thus, the barrier is reduced at the active interface and increases a little bit at the ohmic interface. Overall, the barriers are lower than in the HRS and the cell is in the LRS. When the voltage is swept to zero the concentration profile relaxes a little bit as long as the temperature is high enough to allow for ion movement and then it becomes frozen-in (orange curve). During the RESET sweep the oxygen vacancies are driven towards the ohmic electrode. Due to the depletion of oxygen vacancies at the active electrode, the electron barrier at the active electrode interface is restored and the cell is again in the HRS (red curve). The model has been further applied for simulating multilevel switching by changing the RESET stop voltage. Different multilevel states could be obtained as the concentration profiles freeze in when the voltage is swept to zero. By increasing the RESET stop voltage the oxygen vacancy concentration becomes less at the active electrode and the electrostatic barrier increases [322]. In this model, the bipolar C8W-switching mode is a result of the asymmetry in the work function of the metal electrode. A change in the oxygen vacancy concentration close to the high work function metal has a huge impact on the overall current transport, whereas a change in concentration at the ohmic (low work function metal) interface has only a small one. Thus, the interface between the high work function metal and the oxide is dominating the overall current conduction of the MIM cell and therefore it is called the electrically active interface.

As the bipolar switching behavior results from an asymmetry in the work function of the two metal electrodes, it is straightforward that using two identical metals leads to an electrically symmetric stack and complementary switching should result. In a further simulation study, Marchewka et al. varied the asymmetry of the barrier heights [322]. Figure 64 shows the simulated \( I-V \) characteristics for a symmetric barrier configuration with 0.7 eV high barriers on
both sides. After the initial RESET shown in black the oxygen vacancies are depleted at the left electrode and accumulate at the right electrode. This leads to a high electron barrier at the left electrode and a low one at the right electrode. Overall, the MIM cell is in the non-conducting HRS/LRS. By applying a negative bias to left electrode, the oxygen vacancies are redistributed and a more homogenous distribution is reached slightly above \(-1\) V (yellow). In this case, the barriers at both electrodes are low and the overall sample is in the conducting LRS/LRS. When the negative voltage is further increased oxygen vacancies accumulate at the left electrode and deplete at the right electrode. After sweeping back to zero the obtained concentration profile is the mirror-image of the distribution before the negative voltage sweep. Now, there is a low/high barrier at the left/right interface and a high barrier at the right interface and the cell is non-conducting LRS/HRS state. By applying a positive voltage sweep, the cell undergoes the transition LRS/HRS \(-\) LRS/LRS \(-\) HRS/LRS. This cell behavior resembles the complementary switching mode \([48,58,344,382]\). It was demonstrated that an MIM cell with slightly asymmetric barriers lead to complementary switching with an \(I-V\) crossing \([322]\). These findings fit very well to the experimentally observed transition from complementary switching to pure bipolar switching in a Pt/Ta2O5/Ta stack obtained by increasing the Ta electrode thickness \([58,383]\). In that study, the features of the slightly asymmetric barrier case were observed. Moreover, Yang et al. demonstrated that the rectification polarity of a Pt/TiO2/Pt system can be reversed by biasing the sample with positive or negative polarity for some time. This coincides with the discussed simulations of an electrically symmetric stack.

The presented simulation model was applied to filamentary switching, but in general it could also be applied to area-dependent systems (cf. Section 5.5). The area-dependent switching BiFeO3 system presented by Schmidt et al. showed very similar characteristics \([344]\). In case of area-dependent switching the impact of Joule heating would be smaller and the systems should switch at higher voltages.
Figure 64. Numerical drift-diffusion model of a coupled electronic-ionic transport through MIM structure with symmetric electrodes which leads to a CS-type switching. Adapted and reprinted with permission from [322], © 2015 by IEEE. (a) Sketch of the MIM structure, (b) simulated I–V characteristic, (c) Conduction band edge for four different states and (d) corresponding oxygen vacancy concentration.

Abbaspour et al. developed a kMC model that allows the generation of oxygen vacancies only as a result of an oxygen exchange or oxygen evolution reaction at the ohmic electrodes [302, 375–377] in contrast to the models mentioned in the previous section. The AE in this model is assumed to be chemically inert. The current transport is described as interface-limited TAT mechanisms at low defect concentration, but changes two a band conduction model at high defect concentration. The current transport leads to a local power dissipation \( g(x,y,z) \), which gives rise to local Joule heating. The local temperature \( T \) is calculated using Fourier’s law according to

\[
k_{th} \nabla^2 T(x,y,z) = -g(x,y,z)
\]

where \( k_{th} \) denotes the thermal conductivity. The dynamics of the electroforming and resistive switching are simulated using a kinetic Monte Carlo approach. The generation of oxygen vacancies at the ohmic electrode is described by the generation rate

\[
R_G = \nu_0 \exp\left(-\frac{\Delta W_G - \alpha aeE}{k_B T}\right)
\]

whereas the recombination rate reads

\[
R_R = \nu_0 \exp\left(-\frac{\Delta W_R - (1 - \alpha)aeE}{k_B T}\right),
\]

and the diffusion/migration rate of the oxygen vacancies within the oxide layer is calculated by

\[
R_D = \nu_0 \exp\left(-\frac{\Delta W_D - ae\nabla V_{ij} - k_B a \nabla T}{k_B T}\right).
\]

In Equations (33)–(35) \( \nu_0 \) denotes the vibration frequency, \( a \) is the lattice constant, \( \alpha \) is the symmetry factor, which is usually close to 0.5, \( E \) is the locally induced electric field, and \( \Delta W_G \)
Advances in Physics

Figure 65. Simulated $I$–$V$ characteristics of a Pt/HfO$_2$/Hf/Pt cell including the forming and multiple SET and RESET processes. (c) The 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. Adapted and reprinted with permission from [302], © 2018 by IEEE.

and $\Delta W_R$ are the energy barriers for the generation and recombination of $V_{\text{O}}^{\bullet\bullet}$ at the OE/oxide interface. The migration barrier $\Delta W_D$ of the $V_{\text{O}}^{\bullet\bullet}$ is modulated by the potential difference $\nabla V_{ij}$ and the temperature gradient $\nabla T$ across two defect sites. The electric field $E$ is calculated by solving Poisson’s equation.

With this model, the forming SET and RESET switching have been simulated. Figure 65 shows the simulated forming behavior and subsequent 100 switching cycles. The forming and SET show a very abrupt current increase, while the RESET shows a more gradual behavior. In the model, the voltage is applied to the ohmic electrode and not the active electrode. Thus, forming and SET appear at positive voltage, and the $I$–$V$ curve shows a C8W switching behavior. The distribution of the defects in the oxide during switching shown as (A)–(D) supports the C8W behavior. The OE is at the bottom of the shown oxide region, whereas the AE is located at the top. Subfigure (C) illustrates the HRS configuration of the cell. The defect concentration close to the AE is depleted and the defects are located close to the OE. When a positive voltage is applied to the OE, the defects migrate to the top and further defects are generated at the bottom OE interface (D). In the LRS, a strong filament has formed ‘connecting’ the two electrodes (A). When a negative voltage is applied, the oxygen vacancies are annihilated (recombination) at the OE and the defects migrate to the bottom electrode (B) leading again to the HRS (C). As mentioned in Chapter 4, this model can also reproduce the opposite growth directions of the two filaments during electroforming by varying the ratio of generation/recombination rate to migration rate. For both cases, the switching takes place at the AE consistent with C8W switching [302]. Later, it was shown that the model also reproduces the nonlinear switching kinetics, and the multilevel programming using the gradual RESET or a current compliance during the SET operation [377].

5.5. Area-dependent switching

Up to this point, we have dealt with VCM-type memristive switching where a filament is first formed (Chapter 5) and then switched (Section 5.4) typically between an electrode and the tip of the filament. In the present chapter, we will describe memristive systems in which no filament
Figure 66. Two examples of area-dependent switching cells. The area dependence of the resistance values in high and low resistance states (HRS, LRS) is shown. Red: Pt/YSZ/PCMO/SrRuO$_3$ (red) memory cells. Adapted and reprinted with permission from [384], © 2017 by IEEE. YSZ represents a 2.8 nm thick layer of Y-doped ZrO$_2$ acting as a tunnel barrier, and PCMO is a 20 nm thick (Pr,Ca)MnO$_3$ layer epitaxially grown on SrRuO$_3$. Blue: TiN/a-Si/TiO$_x$/TiN cells, 8 nm a-Si layers 8 nm TiO$_x$ layers (Adapted and reprinted with permission from [385], © 2015 by IEEE.) In both examples, the resistance values for the HRS and the LRS are linearly dependent on the area, suggesting that the resistive switching takes place over the entire area of the interface. (Color figure available online).

formation takes place and, as a consequence, the resistance values scale with the electrode area. As mentioned above, this type of switching is called area-dependent switching in this review paper, while the names area-scaling switching or homogeneous interface-type switching are also used in the literature. The latter notation is because the switching is supposed to take place at the entire interface between the metal electrode and the oxide.

A specific characteristic of the area-dependent switching is the scaling of the device resistance with the area of the cell [384–388]. Figure 66 shows the resistance values in the low and high resistance states (LRS and HRS) for two different types of cells, which are described in more detail below. In the two examples, both resistance states have resistance values that are inversely proportional to the cell area, indicating that the memristive switching indeed takes places over the entire cross section of the electrode. Accordingly, in Figure 6, in the view perpendicular to the current flow, they represent those cells in which the switching occurs over the entire area. Area-dependent memristive switching is typically found for forming-free cells and for cells that have to be formed into the HRS, starting from a rather well conducting initial state.

A clear advantage of area-dependent RS systems compared to filamentary RS systems is their typically much better reproducibility, i.e. they exhibit a much more narrow distribution of the resistance values [385,389,390] provided the fabrication technology is optimized. Figure 67 provides a comparison of the device-to-device variation of the cells on a wafer. In addition, the cycle-to-cycle variation is similarly tight. Because these two types of variabilities are among the biggest challenges of memristive switching systems for application in storage or neuromorphic computing, the characteristics of area-dependent systems offer a significant advantage. In addition to the tighter distributions, the position of cells within the distribution is quite stable in a
Figure 67. Comparison of low resistive state (LRS) and high resistive state (HRS) cumulative resistance distributions (CDF) for area-dependent switching cells (AD) and conducting-filament switching cells (CF). As variability is depending on current level, and is increasing for low current levels, both set of distributions are for the same low switching current of 5 μA (corresponding to the maximum RESET current for the AD cells and SET 1T1R current compliance for CF cells). Data for AD cells are for a $40 \times 40 \text{ nm}^2$ size TiN/a-Si/TiO$_x$/TiN (a-VMCO) cell [385] and for CF cells are for a $20 \times 20 \text{ nm}^2$ TiN/HfO$_2$/Hf/TiN cell [393]. Please note that the resistive curves of the filamentary cells have been rescaled to the median values of the LRS and HRS curves for better comparison. Median (50%) values are for the AD cells: $R_{\text{LRS}} = 10^{6.8}$ Ohm, $R_{\text{HRS}} = 10^{7.8}$ Ohm; and for the CF cells: $R_{\text{LRS}} = 10^5$ Ohm, $R_{\text{HRS}} = 10^{6.4}$ Ohm. Adapted and reprinted with permission from [393], © 2013 by IEEE.

cycle-to-cycle comparison, while cells of filamentary ReRAM tend to change their position in the distribution drastically which renders error correction very difficult [66,391]. Moreover, area dependent systems show in general a gradual transition between the resistance states for both SET and RESET which may enable multi-bit storage within a cell and is highly interesting for neuromorphic and analog computing concepts. On the other hand, area-dependent systems show some drawbacks with respect to the RS kinetics as will be discussed in Section 5.7. In addition, scaling area-dependent switching systems leads to very low amount of defects in the switching interface volume and thus increased variability [392].

In the literature, a number of reports have proposed an understanding of the mechanisms involved in the area-dependent RS, such as electromigration (drift/diffusion) of ionic point defects like oxygen vacancies [70,394–400], trapping of charge carriers (holes or electrons) [401,402], and interfacial Mott transitions [403–407]. As briefly explained in Chapter 2, only those mechanisms which involve a configuration change of atoms or ions will be able to support both, reasonably low resistances ($\approx \text{k}\Omega$) and long retention times (years) at the same time in 2-terminal devices, while purely electronic trapping models suffer from what we can call the voltage–time dilemma. There have been some reviews of area-dependent RS based on collections of empirical results from the literature (e.g. [10,66]) without, however, elaborating in more detail on the guidelines which lead to this type of switching.

In this chapter, we will first discuss such physical guidelines for situations under which area-dependent RS seems to be favored over filamentary RS. Then, we will describe some examples of VCM-type materials and cells for which area-dependent switching has been reported, and we will try to interpret the findings in the light of the general guidelines.

It is noteworthy that also donor-doped SrTiO$_3$ (n-STO) cells at low currents exhibit an area-dependent RS characteristics. This system will be described in Section 5.6 in the context of the 8W switching polarity.
5.5.1. Prerequisites for area-dependent switching

Area-dependent RS requires an internal geometry of the memristive cell that is characterized by a uniform layer thickness of the memristive layer in which the switching takes place. Any significant deviation from such uniform layer thickness would lead to an inhomogeneous distribution of the resistances and, hence, of the electric fields across the plane of the cell.

Let us continue to assume that in VCM-type cells there is no repeated and reversible phase transition during the RS process – unlike in electrochemical metallization cells (ECM) or phase change material (PCM) cells. Then the internal geometry is determined by two factors: Firstly, it is determined by device fabrication. In the case of forming-free cells, this is the only factor to set up the internal geometry. The fabrication has to ensure that the memristive layer(s) of the cell are produced with a uniform layer thickness. Secondly, if an electroforming procedure is required for the specific cell, then this process also affects the internal geometry because it typically involves a phase transition or, at least, the formation of regions of different conductivities separated by relatively distinct boundaries through stoichiometry polarization (see Chapter 5). Empirically it has been found that when forming is required for area-dependent RS cells, it is typically a forming into a HRS. We will try to understand this aspect by discussing the motion of boundaries in a generic manner.

If you consider the motion of any phase boundary in matter, there is a net flux $J$ of a physical quantity at the boundary that is related to the phase transition. This flux is the driving force for the motion of the boundary and it is proportional to the gradient of an intensive thermodynamic variable $Y$. Few examples from other areas of science: (1) the crystallization of a melt is driven by the flux of the heat of crystallization away from the boundary caused by a temperature gradient, $\text{grad } T$. (2) the electro-crystallization of a metal is driven by the flux of cations towards the boundary caused by an electrical potential gradient, $\text{grad } \phi$, in the electrochemical double layer in front of the boundary. (3) the crystallization of a material from a supersaturated solution is driven by a flux of particles $i$ caused by a gradient of the chemical potential, $\text{grad } \eta_i$. In a very simplified picture, one now can derive under what circumstances curvatures in the phase boundary will get enhanced during their motion (roughening), i.e. when they will tend towards dendrite formation, and under what circumstances curvatures will tend to smooth out during the motion of the boundary (smoothening). Figure 68 illustrates the two possible scenarios. Figure 68(a) shows the initial situation as a cross section between two phases, blue and red, with a slightly curved boundary. In Figure 68(a), the time evolution of an increase in the curvature is sketched. The reason is that the gradient in $Y$ in the red phase controls the net flux for moving the phase boundary. This gradient in $Y$ gets enhanced during the growth of the blue phase. Because the gradient in $Y$ is higher in front of convex areas of the blue phase, the flux is higher at this location, and these areas will grow faster than other areas. As a consequence, the curvature increases with time, which is a precondition for dendritic growth. An example is the electro-crystallization of Ag metal from an Ag$^+$ containing electrolyte. In front of a convex part of the metal surface, the gradient of $\phi$, (i.e. the electric field) will be higher than in concave part, and this higher gradient will give rise to a higher flux. As a consequence, the curvature increases during growth. In Figure 68(c), the opposite evolution is sketched. Here, the gradient in $Y$ in the blue phase is controlling the net flux for moving the boundary. Because this gradient decreases with time, the motion of the phase boundary will lead to a reduction of the curvature. An example is the growth of a dense oxide layer on valve metals in air, for example in the case of Al, Ti, Zr, Ta, etc. The driving force for the oxide layer growth is the gradient of the electrochemical potential of the mobile ions in the layer. The gradient and, hence, the flux will be lower where the oxide layer is thicker, and it will be higher where the oxide is thinner. As a consequence, the process will always lead to a smoothening of the metal/oxide phase boundary, i.e. to a uniform oxide layer thickness.
Figure 68. Simplified illustration of the change in the curvature of a phase boundary during its motion. (a) Initial situation – sketch of the curved cross-section of a boundary between a blue and a red phase. In (b) and (c) the motion of the phase boundary in x-direction from left to right is shown, illustrated for three moments in time, $t_1 < t_2 < t_3$. The driving force for the motion of the boundary is proportional to the gradient of an intensive thermodynamic variable $Y$, as described in the text. (b) Roughening process – the driving force is a net flux caused by a gradient in $Y$ in the red phase. This gradient in front of the boundary increases during the boundary motion. (c) Smoothening process – the driving force is a net flux caused by a gradient in $Y$ in the blue phase. This gradient at the boundary decreases during the boundary motion. (Color figure available online).

In summary, a roughening of phase boundaries is expected when the gradient driving the motion of the boundary increases during the motion, a smoothening is expected when the gradient decreases during the motion of the boundary. Of course, this is an extremely simplified picture. In the real world, the surface energy of the boundary, the anisotropy of physical quantities of the phases, and many more factors play a decisive role. For the example of the crystallization
from a normally and supercooled melt, the smoothening and the roughening boundary motion are comprehensively discussed in Ref. [408].

In this simplified picture it becomes clear that an electroforming into LRS typically belongs to the cases in which the boundary between the more and less conductive phase roughens (Figure 68(b)), often leading to a dendritic filament formation.

This is because the gap of the higher resistance is reduced, and as a consequence the electric field increases during the process. In many cases, this is accompanied by a local temperature increase which amplifies the growth of the filament (see Chapter 5, Electro-formation). In contrast, an electro-formation of a memristive cell with a relatively low initial resistance into a HRS belongs to the cases in which a smoothening of the boundary may take place (Figure 68(c)). We will encounter this below when we discuss examples of area-dependent memristive cells.

As mentioned above the division into filamentary and area-dependent VCM-type cells is often encountered, but it is not strict. For instance, multifilamentary scenarios in which the RS takes place in a regular manner with a periodicity on the nanometer scale are also covered here, because they show a clear area-scaling on the μm to mm scale [77,205]. In other cases, the area-dependence does not extend over the entire area rather than parts of it. This could be described as a multifilamentary switching geometry [409,410]. Furthermore, in many cases in the literature the area dependence has not been studied or no clear area-scaling of one or both RS states is found. In fact, there are indications that sometimes more than one RS mechanism is active. These mechanisms superimpose each other and, possibly, interact in an unforeseeable manner.

It should be mentioned that the resistance of the filament in the LRS might be so low that the LRS is determined by the series resistance of the metal electrodes. In this case, the LRS might scale with the device area for certain geometries and might mimic an area-dependent change of the resistance during switching.

Another essential ingredient of area-dependent RS is the fact that local positive feedback, e.g. due inhomogeneous Joule heating, is avoided. The non-linearity of the kinetics (cf. Section 5.8) cannot be supported by temperature enhancement in the case of area-dependent RS because any heating will typically lead to a self-accelerated and localized thermal runaway which is the signature for filamentary switching. For this reason, only the high-field enhancement of the ion mobility remains as the origin of the non-linearity. As a consequence, the switching may only take place in thin layers such as barrier oxide layers and depletion space charge layers because in such thin layers sufficiently high fields can be build up at voltages of a few volts applied externally, while thicker resistive layers will not change their resistance state. As a result of the high fields at the barrier region, it might happen that the break-down field of the barrier is reached when high operation voltages are applied. As a result, the area-dependent switching mode is transformed into a filamentary mode in an irreversible manner [254,411,412]. This effect is a typical endurance failure for area-dependent switching devices. However, a sharp SET process during normal operation is also reported for nominally area-dependent type of materials [413–415] hinting on the formation of a filament or even arrays of filaments in the interface barrier.

A drawback of the fact that only field enhancement supports the non-linearity of the RS kinetics is a limitation in the non-linear regime, leading to limited retention times and/or slow switching. This will be discussed in more detail in Section 5.7. On the other hand, the lack of thermal acceleration results in a non-abrupt, gradual transition between the resistance states.

5.5.2. Examples of C8W area-dependent resistive switching systems

As a first example of area-dependent RS, we will discuss cells with mixed-valent manganites. These oxides represent a class of ABO₃ perovskite-type oxides, which show a large diversity in their electronic and magnetic properties [416–420]. Their crystal structure is given by a distorted
Figure 69. Cross-sectional TEM images of a Ti/PCMO/SRO junction. After forming (into a HRS), a thin amorphous TiO\textsubscript{y} layer (∼10 nm) evolved between the Ti electrode and PCMO layer. Adapted and reprinted with permission from [400], © 2009 by the American Physical Society.

ABO\textsubscript{3}, where Mn is placed on B–site and the A-site is shared by trivalent rare earth cations (R) such as La, Pr or Nd and divalent alkaline earth cations (A) such as Ca, Sr or Ba. The composition (R\textsubscript{1−x}A\textsubscript{x})MnO\textsubscript{3} can be considered as a solid solution of RMnO\textsubscript{3} and AMnO\textsubscript{3}, in which the ratio of R to A determines the ratio of Mn\textsuperscript{3+} and Mn\textsuperscript{4+}. In manganites, the 3d-orbitals of Mn are split due to both Jahn–Teller effect and electron correlations. That results in a complex band structure which varies strongly with doping concentration and oxygen content (see e.g. [416,417,419,421–423]). The end-members RMnO\textsubscript{3} (i.e. x = 0) and AMnO\textsubscript{3} (i.e. x = 1) are reasonably good insulators, while solid solutions show p-conductivity for x < 0.8. The electronic carriers can be described by small polarons, because the holes are somewhat localized at the dopant position, represented as Mn\textsuperscript{4+}, and move by the so-called double exchange mechanism (see, e.g. [424,425], and references cited therein).

The lattice disorder in (R\textsubscript{1−x}A\textsubscript{x})MnO\textsubscript{3} is determined by the densely packed perovskite crystal lattice which means that interstitial oxygen ions can be neglected and that the Schottky disor-der is favored over the Frenkel disorder. If one starts to substitute R\textsuperscript{3+} by A\textsuperscript{2+} in RMnO\textsubscript{3}, the A ions act as acceptors which are compensated by holes as mentioned above. Thermal or electrochemical reduction leads to the introduction of O vacancies near the Mn sites which changes predominantly the hole mobility due to a change in the double-exchange mechanism and, hence, in the p-conductivity [66,418] and only slightly the hole concentration [426].

Resistive switching has been extensively investigated for (Pr\textsubscript{1−x}Ca\textsubscript{x})MnO\textsubscript{3} (short: PCMO) with x = 0.3–0.5. Typically, memristive PCMO cells have been fabricated with an electrode such as Pt or SrRuO\textsubscript{3} on one side, which forms an ohmic contact with the p-type oxide. At the other electrode, a barrier oxide is inserted that facilitates an area-dependent RS. This barrier oxide layer is either deliberately deposited during fabrication or it appears as the result of a redox reaction between the PCMO and a valve metal electrode. Numerous combinations of PCMO with valve metals are reported in literature such as Ti [427], Al [428], Ta [429], W [430] [431], Mo [432] and others [433–436]. Another possibility is the use of the pn-junction between PCMO and the n–conducting Nb-doped STO [437]. In the following, we will discuss in detail the Ti/PCMO system.
For Ti/(Pr$_{0.5}$Ca$_{0.5}$)MnO$_3$/SrRuO$_3$ cells, an ultra-thin amorphous TiO$_y$ layer of $< 1$ nm thickness was observed at the Ti/PCMO interface in the pristine state [400,438]. Obviously, this layer is caused by a reduction of the adjacent PCMO by the Ti metal, i.e. O vacancies are introduced into the PCMO. Electroforming is conducted by applying a positive voltage of $> +5$ V at the Ti side which transforms the cells into the HRS. During this process, there is further phase transformation by an electrochemical oxidation of the Ti metal to TiO$_y$ which grows up to a uniform thickness of approx. 10 nm (Figure 69). Because the gradient of the electrochemical gradient in the a-TiO$_y$ and the motion of the phase boundary point in opposite directions, this represents the case of a smoothening boundary motion (Figure 68(c)). As explained, this is a prerequisite of area-dependent RS. After forming, the subsequent RS operations do not lead to any significant change in the thickness of the a-TiO$_y$ barrier oxide layer, but rather in the average Ti valence and, hence, the barrier transmissivity as will be discussed in detail in the following.

When a negative voltage bias is applied to the Ti top electrode, a SET operation is conducted, i.e. the resistance state is converted from HRS to LRS, corresponding to a counter-eightwise (C8W) polarity of these cells. In this case, negatively charged O ions are expected to drift from the a-TiO$_y$ layer into the PCMO layer (and annihilate O vacancies in the latter). During the RESET, O vacancies are expected to drift back to the PCMO layer from the a-TiO$_y$ layer. In the ionic picture, the ion exchange (of a fraction $z$ of O ions) should lead to redox reaction according to

$$\text{SET} \rightleftharpoons \text{RESET}$$

$$\begin{align*}
(\text{Pr}_{0.5}^{3+}\text{Ca}_{0.5}^{2+}) (\text{Mn}_{0.5}^{3+}\text{Mn}_{0.5}^{4+})\text{O}_2^- + \text{Ti}^{4+}(y/2)\text{O}_y \\
(\text{Pr}_{0.5}^{3+}\text{Ca}_{0.5}^{2+}) (\text{Mn}_{0.5+2z}^{3+}\text{Mn}_{0.5-2z}^{4+})\text{O}_{2-2z}^- + \text{Ti}^{4+}((y+z)/2)\text{O}_{y+z}
\end{align*}$$

with $y + z \leq 2$, where the left side of the equation denotes the LRS and the right side the HRS. Because the ion motion is only facilitated by a high electric field and not supported by any temperature enhancement in the area-dependent RS, only the outermost part, i.e. the screening layer of the PCMO layer participates in this ion exchange.

This process has been studied with hard X-ray photoelectron spectroscopy (HAXPES) combined with an electrical characterization (Figure 70) [414,438]. This analysis succeeded to prove that during electroforming, a redox-reaction takes place between Pr$_{1-x}$Ca$_x$MnO$_3$ and the TiO$_2$ interface layer formed during the growth of the Ti top electrode [414,438]. The Ti 2p core levels of the cells contain a variety of oxidation states ranging from fully oxidized Ti$^{4+}$ to metallic Ti$^0$ (Figure 70(a,c)) confirming the formation of an oxide layer at the Pr$_{1-x}$Ca$_x$MnO$_3$ interface observed by TEM analysis [438,400]. In the Mn 2p spectra, a corresponding change of the spectral weight indicates the reduction of Mn during the deposition of the Ti layer (Figure 70(b)). The PCMO reference sample (Figure 70(c)) shows the change of the spectral weight of the different Ti valence states for different resistance states. It can be clearly seen that the Ti$^{4+}$ contribution is dominant but decreases with decreasing cell resistance ($R_{\text{formed}} > R_{\text{HRS}} > R_{\text{LRS}} > R_{\text{pristine}}$) while the spectral weight of lower Ti valence states increases. In contrast, in the Mn 2p spectra, a corresponding change of the spectral weight can only be observed between an untreated PCMO reference sample and various resistance states. However, the convolution of the O 1s spectra (see Figure 70(d)) shows a clear trend with respect to the contribution from the TiO$_{2-x}$ and the Pr$_{1-x}$Ca$_x$MnO$_3$. This analysis confirms a shift of oxygen from the Pr$_{1-x}$Ca$_x$MnO$_3$ to the TiO$_{2-x}$ during electroforming. During switching to the LRS mainly the oxygen content of the TiO$_{2-x}$ is decreased which is consistent with an increase of the metallic Ti contribution in the LRS. At the same time, the oxygen from the Pr$_{1-x}$Ca$_x$MnO$_3$ shows a slight increase during SET. These findings by HAXPES and TEM combined with band structure considerations indicate that the TiO$_y$
Advances in Physics

Figure 70. Photoelectron spectra of Pt/Ti/PCMO/SRO, sorted from low to high resistances, i.e., for the pristine state (PS), the low resistance state (LRS), the high resistance state (HRS), and the formed state (FS). (a) Ti 2p spectra. The weight of the lines of the different Ti valence correspond to the stoichiometry of the TiO_y layer formed at the Ti/PCMO interface. (b) Mn 2p spectra. An untreated PCMO sample is included for reference. Area percentages of the constituent components of the Ti 2p (c) and (d) the O 1s spectra in the different resistive states (pristine, formed, LRS, HRS) determined by a fitting analysis of the core level spectra. Adapted and reprinted with permission from [414], © 2014 by Wiley-VCH.

Layer acts as a tunneling barrier which is in series with (polaronic) resistances of the depletion layer of the PCMO, the bulk of the PCMO, and the ohmic PCMO/SRO contact. A quantitative analysis shows that the tunneling barrier dominates the \( I-V \) characteristics (for voltages well below the RS) for the HRS and the formed state, whereas it is of minor importance for the LRS and the pristine state [414]. Based on these results, a band diagram can be sketched which helps to understand the switching mechanism (Figure 71). According to Equation (4) a more oxidized TiO\(_{y+z}\) corresponds to a higher barrier related to the HRS, while a reduced TiO_y with a strongly suppressed band gap, represents the LRS, which is dominated by the contribution of the polaronic transport at the in the oxygen depleted PCMO interface layer.
In operando TEM studies on PCMO devices with TiN electrode have shown the formation of a TiO$_x$N$_y$ at the interface [434]. The thickness of this interface layer is modulated in a gradual and reversible manner by both, the applied voltage amplitude and the number of voltage pulses.

A related example of this class of area dependent PCMO devices is based on the controlled introduction of a high tunneling barrier by depositing an oxide with a sufficiently high band offset between the mixed-conducting manganite and to the electrode [439,440], instead of growing the barrier oxide by oxidizing a valve metal electrode during electroforming as discussed above. In accordance with Ref. [66], we will call this the tunnel ReRAM concept. This concept was developed by Unity Semiconductor under the name of conductive metal oxide (brand name: CMOx). Later, this work was continued by Rambus Labs. A possible advantage over other types of ReRAM devices seems to be that current levels for the SET, RESET, and READ operations can be pre-adjusted by the thickness of the deposited tunnel oxide material Figure 74.

Figure 72 shows the device structure and a TEM cross section displaying the different layers of the tunnel ReRAM device. A thin tunnel oxide with a thickness of 2–3 nm is sandwiched between a noble metal electrode such as Pt and a conductive metal oxide as, for example, (Pr,Ca)MnO$_3$. The tunnel barrier is typically a high–k oxide such as ZrO$_2$ or HfO$_2$. We will call this side of the cell the active side. The opposite side of the conductive metal oxide is contacted by a metal that provides an ohmic contact (such a Pt in the case of PCMO).

In this device structure, the choice of a metal with no oxygen affinity on the active side is essential. In contrast to valve metals such as Ti (Section 5.5.2) this prevents any additional formation of an oxide layer and the tunnel oxide thickness remains at its value adjusted during cell fabrication. Because of this concept, no electroforming step is required, in contrast to cells described in Section 5.5.2. However, it has reported that a certain cycling procedure (conditioning) of the cells is needed before stable switching is established [384]. This procedure does not result in any abrupt change of the resistance as in the case of a classical forming process.

Figure 73 displays the quasi-static hysteretic $J$–$V$ characteristics of the tunnel ReRAM based on PCMO. The $J$–$V$ characteristic is highly non-linear for both positive and negative polarities. When a positive voltage is applied to the active electrode, the device current decreases by a factor
of 10–100. Applying negative voltages reverses the resistance change, i.e. this device shows a typical C8W switching polarity with respect to the contact at the tunnel oxide. In contrast to filamentary memory devices, which often show an abrupt SET process, the transition between HRS and LRS is continuous. Furthermore, an external current limiter, commonly a transistor, is not required.

Electrode feature sizes have been studied from 20 μm down to 30 on 300 mm wafers. The result demonstrated a linear relationship between current and area indicating a clear area-dependent RS. The control of the device currents with tunnel oxide thickness is shown in Figure 74. An exponential relationship between tunnel oxide thickness, here represented by the sputter time deposition time, and the current densities is observed. The characteristic behavior as well as shape of the \(I–V\) curve and temperature dependence (not shown here) can be best described by a trap-assisted tunneling mechanism.

The typical cycling endurance of the tunnel ReRAM devices is in the order of \(10^3–10^6\) cycles, depending on the material combination and the SET/RESET conditions [439]. Data retention ranges from several days to months at elevated temperatures and is a function of the material system and the SET/RESET conditions.
In the tunnel ReRAM presented here, the change in resistance can be explained by a field-driven exchange of oxygen between the PCMO and the tunnel oxide and the resulting resistance change [439]. This is similar to the process described in Section 5.5.2., except the fact that the tunnel currents always dominate the overall cell resistance. If a positive voltage is applied to the electrode at the active side, oxygen ions are moved from the PCMO into the tunnel oxide. After reversing the polarity, oxygen ions are pushed back by the top electrode into the PCMO. It has been proposed in the literature [66] that due to the negative charge, the height of the tunnel barrier increases if oxygen is shifted into the tunnel oxide and the electron tunneling current is reduced, i.e. the device is turned in the HRS. The mechanism and the corresponding band diagrams including a charge induced change in the tunnel barrier height are sketched in Figure 75(a). However, it is important to note that TEM-EELS studies have revealed a much more complex mechanism, including also a structural change in the YSZ barrier that could also result in a change of the barrier height upon switching [384]. Moreover, the change of the oxygen vacancy concentration within the YSZ barrier would also result in a modulation of the transmission through the tunnel barrier by the modulation of the trap density.

Capacitance measurements indicate a 3–5% decrease in the device capacitance in the HRS compared to the LRS [66]. This may be interpreted as the formation of a very thin high resistive layer in the PCMO. The layer forms during the RESET operation, when oxygen is moved from the PCMO layer into the tunnel oxide and the PCMO is depleted in oxygen, in the same way as we discussed it in Section 5.5.2.

A comparable study using YSZ as a thin tunnel oxide in combination with PCMO is reported for an integrated W plug as the electrode at the YSZ side [430,441]. An Al₂O₃ tunneling barrier has been employed in the same arrangement [442].
In addition to the systems described up to now, there are a number of reports in which no barrier oxide has been included in memristive PCMO cells. Here it is expected that either the depletion space charge region of pn junctions, e.g. between p-type PCMO and n-type donor-doped SrTiO$_3$, acts as the memristive barrier or a stoichiometry change within the manganite layer leads to a region of higher resistivity and sufficient ionic conductivity [390,425,443–449].

RS studies on other manganite systems than PCMO include (La,Sr)MnO$_3$ The electrical contact properties have been reported, e.g. in Refs. [415,450–453], (La,Ca)MnO$_3$ [387], and TbMnO$_3$ [454]. The electronic properties, often including RS properties, of manganite surfaces without top electrode have been studied by scanning probe techniques [395,399,455–458] as well as by spectroscopic techniques [452,459].

Besides heterostructures of p-conducting oxides, also n-conducting oxides have been employed to fabricate area-dependent switching layer stacks. The most prominent example is Nb-doped SrTiO$_3$ that forms a Schottky-contact with high work function metals such as epitaxial SrRuO$_3$ (SRO) [402] and Pt [386]. These systems will due to its 8W switching polarity be discussed later in Section 5.6. Moreover, bilayers of substoichiometric and therefore n-conducting oxides such as TaO$_x$ [412] and TiO$_x$ [460] with a stoichiometric HfO$_2$ tunnel barrier exhibit area-dependent switching.

Moreover, Hansen et al. [388] reported on an area-dependent RS device based on n-conducting materials with a layer sequence Al/Al$_2$O$_3$/NbO$_x$/Au, in which the Al$_2$O$_3$ layer is a tunnel barrier and the NbO$_x$ the memristive layer. Neither an initial forming process nor a current compliance were required for recording the resistive switching of these cells. Using cells of very different electrode areas it has been confirmed that they, indeed, show an area-dependent switching. Using an appropriate device model, Ref. [388,461] demonstrates that the experimental $I$–$V$ curves can be adequately reproduced.

A comparable concept has been reported by Govoreanu et al. using a TiN/Al$_2$O$_3$/TiO$_2$/TiN cell [462] and a TiN/a–Si/TiO$_2$/TiN cell [385], in which Al$_2$O$_3$ and amorphous Si (a–Si), respectively, act as tunneling barriers. The TiN electrode constitutes a Schottky barrier. The cells exhibit forming-free characteristic, no need of current compliance, low device-to-device variability, multilevel capability, and a clear area-dependent switching behavior.
A similar situation is found for BiFeO$_3$ (BFO) thin films that are donor doped by Ti on one side of the film and stacked between two high work-function metals. Crystalline BFO thin films typically show an n-conductivity after been grown by PLD at elevated temperatures and annealed in oxidizing atmospheres, so that high work-function metals such as Au and Pt form Schottky contacts [463]. The n-conductivity of BFO may be caused by a residual O vacancy background concentration. In a range of studies [344,464–469], BFO films of few hundred nanometers thickness have been grown on Pt bottom electrodes on sapphire or SiO$_2$/Si wafers. After annealing, Au top electrodes were deposited by sputtering. In the absence of Ti doping, the MIM stacks showed typical anti-serial double Schottky diode $I–V$ characteristics without resistive switching [465]. However, if Ti doping is introduced into the BFO film near the Pt electrode, a significant increase in the current levels and a pronounced resistive switching hysteresis occurs for positive voltages applied at the Au contact. In all cases, no electroforming has been necessary to evoke the resistive switching. Figure 76 shows a sequence of voltage sweeps and the typical current response. Positive voltages lead to a SET, while negative voltages lead to a RESET operation. As we shall see, this corresponds to a C8W switching polarity because the switching event occurs at the (here: grounded) bottom (Pt) electrode. The first sweep with an opposite voltage polarity executes the switching operation (traces 1 and 9 to $−8$ V for RESET, traces 5 and 13 to $+8$ V for SET), while subsequent sweeps into the same direction show current traces which fall perfectly onto of each other (such as traces 3 and 4 which follow trace 2). Obviously, the $I–V$ characteristic does not depend on the bias sweeping direction for a given state and no current compliance is required for the operation. In addition, there are no abrupt current changes, in contrast to typical SET processes for filamentary VCM cells. It is suggested that the $I–V$ characteristics and the resistive switching behavior can be explained by the doubly donor doped BFO film and the Schottky contacts at the Au and the Pt electrodes [470]. The doping of the BFO film by Ti donors is highly asymmetric, because Ti is accumulated near the Pt electrode in three different cases of experimental preparation of the cells.

Endurance tests show excellent reproducibility of the states to $3 \times 10^4$ cycles, with extremely narrow distributions (Figure 77). Which is a typical property of area-dependent VCM cells. A retention test shows a decay of the $R_{HRS}/R_{LRS}$ ratio which stabilizes at a lower level after some time [467].

As for all area-dependent RS systems, the switching kinetics is solely controlled by the field-enhancement of the O vacancy mobility because even the highest current densities are not sufficient to lead to any relevant temperature increase. The switching kinetics become nonlinear for voltage amplitudes $> |5|$ V. The electric field is mainly focused over a part of the depletion layer close to the metal electrodes. This effective layer can be estimated as few nm. The switching times reduce from approx. 100 ms to 500 ns if the voltages are increased to $|20|$ V, following the Mott-Gurney law in reasonable approximation [470].

### 5.6. Eightwise switching

We have defined the RS polarity in Section 2.3.3 in the following way: First, one suggests what side of the MIM is mainly responsible for the redox-based resistive switching process. The electrode at that side is called the electrically active electrode (AE). Second, one defines the applied voltage $V$ to this active electrode. Third, the circulation of the $I–V$ curves on a linear scale is compared to the typical drawing of the character ‘8’ by hand. The normal VCM-type switching originates from an attraction (repulsion) of mobile donors by a negative (positive) voltage at the AE and, as a consequence, a more (less) electron transparent Schottky barrier. The corresponding $I–V$ curve has a counter-eightwise (C8W) RS polarity as explained in Section 2.3.3.
Figure 76. (a) Sequence of applied ramping voltages; (b) resulting $I$–$V$ curves of the Au/BFO(600 nm)/Pt/Ti/Sapphire MIM structure. The numbers 1–16 label successive ramping voltages and corresponding current traces on a logarithmic scale, and the arrows indicate the scanning direction of the applied ramping voltages. The inset in (b) exhibits the schematic setup for the electric measurements. Adapted and reprinted with permission from [467], © 2014 by the American Chemical Society.
However, there are many reports about an opposite, e.g. eight-wise (8W) RS polarity observed at typical metal–oxide cells [55, 70, 90, 91, 253, 254, 343, 346, 386, 396, 398, 402, 409, 471–494]. Most of the reports on 8W-RS have been for Schottky contacts at n–doped SrTiO$_3$ and for thin epitaxial SrTiO$_3$ films grown on n–doped SrTiO$_3$. The first comprehensive reports were published by Fujii et al. [402] and Sim et al. [386] in 2005. They studied the RS of a Schottky contact between a Nb-doped SrTiO$_3$ single crystal as a n-type semiconductor and high workfunction metals such as epitaxial SrRuO$_3$ (SRO) [402] and Pt [386], respectively.

Most interestingly, Shibuya et al. [90] reported that it is possible to invoke a C8W switching and an 8W switching at the same memristive cell just by changing the voltage amplitude. This phenomenon has been confirmed by several subsequent reports [70, 91, 93, 480, 495]. Münstermann et al. used a LC-AFM technique to reveal the location of the two switching processes after peeling-off the top electrode [70]. They used a 500 nm 1at% Fe-doped SrTiO$_3$ thin film on an electronically conducting 1at% Nb-doped SrTiO$_3$ substrate and a Pt top electrode which provided the Schottky junction of this MIM system. To activate RS, a forming process was applied to the devices as described in Chapter 4. The subsequent $I$–$V$ characteristics are shown in Figure 78. A stable C8W switching is observed for the sequence of branches 1–2–3–4–1. If the negative voltage scan is extended, however, a higher resistance state is reached (branch ‘5’) which remains upon the sweep to positive voltage (Branch ‘6’). This sequence 1–5–6–4–1 is a clear signature of an 8W switching. It was possible to invoke both types of RS repeatedly by adjusting the negative voltage amplitude.

A very similar behavior was found for Pt/TiO$_2$/Ti/Pt cells, where the left-most Pt electrode is the AE (Figure 78(b)) [92]. Similar to the case described in Figure 78(a), the resistance states of both modes are non-volatile and share one common state; i.e. the HRS of the C8W mode which equals the LRS of the 8W-mode. A transition between the opposite hysteresis loops is possible by voltage control in a polarity defining region. Specifically, 8W BRS in the TiO$_2$ cells is a self-limited low-energy non-volatile RS process. Additionally, the 8W RESET process enables the programming of multilevel HRS. A similar transition was reported for Pt/Ta$_2$O$_5$/TiN cells [94] and for HfO$_2$ cells [496]. It should be noted that an 8W polarity and a C8W polarity have also
Figure 78. (a) $I$–$V$ characteristics of Pt/SrTiO$_3$: Fe/SrTiO$_3$:Nb cells. The voltage $V$ is applied to the Pt electrodes. Both resistive switching polarities are observed: the C8W polarity (green) and the 8W polarity (orange). Adapted and reprinted with permission from [70], © 2010 by Wiley-VCH. (b) Coexistence of two bipolar switching modes with opposite polarities in a Pt (BE)/6 nm TiO$_2$/Ti/Pt (TE) nanocrossbar cell of $(100 \text{ nm})^2$ size. The voltage is applied to the Pt electrode (AE). The switching orientation is marked by arrows, and the capital letters define states or switching events. The inset shows the zoomed 8W BRS loop (red curve) on a linear scale. Adapted and reprinted with permission from [92], © 2018 by the American Chemical Society. (Color figure available online).

been found in STM-tip induced resistive switching experiments in ultrathin TaO$_x$ cells, linked to the transport of anions and/or cations [493].

The above-mentioned papers are consistent with a filamentary geometry of the 8W switching based on a pad area independence of the LRS current, the independence of the Schottky depletion capacitance on the resistance state, and/or a pronounced inhomogeneity of the Schottky barrier height [478,479,486,491,497]. Münstermann et al. showed that although confined to restricted areas, 8W switching is generally more laterally extended than C8W switching in the same
cells [70]. There are also reports about a multifilamentary conductance on the surface of Nb-doped SrTiO$_3$ single crystals observed by LC-AFM where patches of approx. 20–30 nm diameter are found to show resistive switching [485]. While inhomogeneities in the dopant distribution can be excluded to explain this effect [498], the origin of this multifilamentary conductance characteristic is not yet clear.

Besides this, well-defined area-dependence of 8W switching has been reported for Schottky junctions on n-doped SrTiO$_3$ crystals using SrRuO$_3$ electrodes [346], Pt electrodes [343,475,488,489] or other noble (high work function) metal electrodes [472].

This controversy has been resolved by reports describing that both, area-dependent and filamentary switching may occur for the same cell, depending on the electrical conditions. Yang et al. report about an area-dependent RS of virgin Pt/Nb: SrTiO$_3$ junctions at low currents, while after electroforming at higher forward voltages clear indications of a filamentary RS is observed based on a detailed analysis of the $I$–$V$ and $C$–$V$ characteristics [491]. This is confirmed by Bäumer et al. who detected microspectroscopically a localized significant reduction of Ti$^{4+}$ to Ti$^{3+}$ after electroforming and a locally high-temperature phase separation of a SrO phase at the same location. A unique correlation with LC-AFM has proved higher conductance at this location, clearly identifying the conducting filament as a result of the local heating during the electroforming process [254].

We will now turn to possible explanations of the microscopic mechanism of the observed 8W switching phenomenon. In the literature, frequently there are suggestions of pure electronic mechanisms [10,346,402,479,480,483,484,488]. These models suggest that traps at the interface of the active electrode and the oxide or in the space charge layer within the oxide are charged/discharged by electrons during the switching. The mechanism of trapping/detrapping is important to arrive at a specific RS polarity. A positive voltage at the electrode would lead to a RESET process if the electrons were electrostatically attracted from the interior of the (conducting) oxide into the traps, resulting in a C8W polarity. A RESET process for an 8W polarity is achieved, if the electrons are injected from the electrode into the traps by, e.g. Fowler–Nordheim tunneling when a negative voltage is applied to the electrode, as e.g. explained by the so-called MemFlash (Section 2.1). A full MemFlash operation is, of course, not possible for the MIM cells discussed here, because this requires a dedicated lateral structure. In many cases of the MIM cells in the references mentioned, a relatively short retention time of few seconds to few thousand seconds is reported. In combination with low current densities, the systems might overcome the RC time dilemma described in Chapter 3, so indeed, a purely electronic mechanism might hold in such cases. However, for longer retention times and higher current densities, purely electronic mechanisms are not conceivable as mentioned in Section 2.1 [29]. And there are clear indications for a nanoionic process as we will discuss now.

Within the possible nanoionic processes, we can identify four different reasons for an anomalous (8W) switching polarity.

The first reason can simply be a mistake in the assignment of the active side of the cell. As explained in Section 2.2, the asymmetric character of the bipolar RS cannot be matched by a fully symmetric cell, so that one of the two electrode interfaces is more active in the RS process. If the electronically active electrode (AE) is assigned to the wrong side, a C8W switching will appear as an 8W polarity by mistake. The change of the AE has been explained in great detail in references [244,307]. It should be mentioned that it is often not an easy task to identify the AE, because of the tiny dimension of many cells and because chemical interface reactions may modify the characteristics of an electrode interface. Only in few cases, the electrostatic potential profile between the electrodes is experimentally determined, allowing for an unequivocal assignment of the electrode in front of which the highest electrical field changes occur.
The second option is an O vacancy model (or, more general, a mobile donor model) which is similar to the common C8W model, however, applied to p-conducting oxides instead of n-conducting oxides. If the background or base conducting relied on holes instead of electrons, the effect of a concentration polarization of mobile donors would reverse. As a consequence, an 8W switching polarity would be observed instead of a C8W polarity. Mixed-valent manganites \((R_{1-x}A_x)\text{MnO}_3\) are usually p-conducting oxides. In \(Pr_{1-x}\text{Ca}_x\text{MnO}_3\), the conductivity decreases with increasing oxygen vacancies as a result of the disturbed superexchange and the resulting hole delocalization [426]. Indeed, 8W switching has been reported in Ag/\(Pr_{1-x}\text{Ca}_x\text{MnO}_3/YBa_2\text{Cu}_3\text{O}_7\) cells where Ag acts as active electrode [445]. Even nominally symmetric Pt/\(Pr_{1-x}\text{Ca}_x\text{MnO}_3/Pt\) showed bipolar 8W switching with respect to the top electrode [397]. This has been attributed to the high ohmic oxygen deficient \(Pr_{1-x}\text{Ca}_x\text{MnO}_3\) formed at the top electrode during sputtering of the Pt top electrode which induces an inherent asymmetry in the cell and determines the Pt top electrode as active electrode. Rozenberg and co-workers proposed a model for resistive switching in p-type transition metal oxides, which is based on the movement of oxygen vacancies and a resulting local conductance change [499,500]. In contrast to the n-type oxides discussed in Section 3.3, the local resistance increases with increasing oxygen vacancy content. With this model, 8W-switching bipolar \(I-V\) characteristics are obtained [499,500].

The third possibility was suggested as a derivation of the C8W switching mechanism and a very close attraction of the O vacancies to the active electrode interface as first suggested by Muenstermann et al. [70]. Starting from a LRS in which the O vacancies as mobile donors are spread towards this electrode, a negative voltage would attract the O vacancies into a layer right at the interface (Figure 79). They would be mobile because of the high field in this region and the Mott-Gurney field enhanced mobility, which leads to a separation from the O vacancy reservoir in the bulk of the oxide layer. These are assumed to be much less mobile so that a (high-resistive) region of low O vacancy concentration between the bulk and the O vacancy-rich layer immediately behind the electrode builds up, giving rise to the HRS.

Lee et al. offered an electrostatic semiconductor model including O vacancies as mobile donors which seemed to support this model in a quantitative manner [343,487,502]. However, the electron drift and diffusion through the depleted region and the oxide was not included in the model. Thus, it is questionable if LRS and HRS are correctly attributed to a certain O vacancy.

Figure 79. Suggested origin of the 8W switching observed in Pt/Fe:SrTiO\textsubscript{3}/Nb:SrTiO\textsubscript{3} cells. Top: In the LRS, the O vacancy concentration is assumed to be quite homogeneous as reported for the LRS in C8W switching cells. A negative voltage at the Pt electrode is suggested to attract the O vacancies from a certain depth separating them from the bulk of the oxide layer at the right side (dark blue layer at the Pt electrode). Adapted and reprinted with permission from [70], © 2010 by Wiley-VCH. Bottom: Oxygen vacancy concentration profile across the device cross-section for the LRS and the HRS within the suggested model [501].
distribution. Furthermore, the model is challenged by the question how the O vacancies close to the Pt electrode interface in the HRS can be repelled again in a subsequent SET operation, because most of the field would be located in the high-resistive region of low O vacancy concentration. This question gets even more severe for a suggested deposition of O vacancies at the very interface or inside a SrRuO$_3$ electrode, based on a first principles calculation [398]. A modification of the model is the assumed clustering of O vacancies near the electrode interface giving rise to lower lying energy states which do not contribute to the donor concentration anymore [489].

A fourth model considers the electrochemical exchange of (molecular) oxygen. This oxygen exchange at the electrode is composed of an evolution and a reduction reaction, and is sometimes referred to as an electrocatalytical model. It corresponds to the electroformation into an HRS at an electrically active noble electrode, as explained in Section 4.4. This process has been suggested as a reversible RS mechanism for various oxides such as TiO$_2$ [503], Si rich SiO$_x$ [272,504] Nb:SrTiO$_3$ [254], CeO$_2$ [477], SrRuO$_3$ [493], La$_{0.8}$Sr$_{0.2}$MnO$_3$ [415], Pr$_{1-x}$Ca$_x$MnO$_3$ [425,447,505], Sr$_{1-x}$MnO$_3$ [506,507] and TaO$_x$ [55,493]. The idea of this model is an oxygen evolution reaction upon a positive voltage applied to the electrode:

$$\text{O}^2- \xrightarrow{\text{oxidation}} \frac{1}{2}\text{O}_2(g) + 2e^- \quad (43)$$

leading to a SET process and a LRS of the cell, and the reverse reaction, i.e. a reduction of molecular oxygen from the ambience upon a negative voltage applied:

$$\frac{1}{2}\text{O}_2(g) + 2e^- \xrightarrow{\text{reduction}} \text{O}^2- \quad (44)$$

leading to a RESET process and a HRS of the cell. In both cases, a certain threshold voltage is required at the electrode interface in order to start a significant reaction. This threshold voltage is connected to the specific exchange current density at the interface of the electrode and the oxide material, and the current density is described by the corresponding Butler–Volmer equation (see Section 5.7.2) of this process. Let us consider the major difference of this 8W model compared to the normal C8W model for a positive voltage at the electrically active electrode (AE). The oxidized species, here the molecular oxygen, leaves the cell upon the oxidation, so that the reduced (and, hence, electron conducting) oxide can spread fully to the active electrode and diminish the Schottky barrier to achieve a LRS (see Figure 80(b)). Such a situation has been described in the context of the electroforming in Section 4.4, in which we mentioned the formation of bubbles underneath the electrode during the formation into the HRS. In the case of the C8W switching and a positive voltage to the AE, the oxidized region of the oxide was blocking the electron conduction path between the electrodes because of the concentration polarization (Section 5.5), so that a HRS of the cell is obtained. The opposite happens upon a negative voltage applied (see Figure 80(b)). In principle, both types of processes, oxygen drift (leading to the concentration polarization) and the O exchange can occur concurrently. The RS polarity is then determined by the faster process, while the other process is rate determining.

This oxygen exchange model has been experimentally confirmed for 8W switching STO devices by operando TEM studies [501]. The oxygen vacancy profile determined experimentally by STEM-EELS analysis in the LRS (see Figure 81), indicating an accumulation of oxygen vacancies at the Pt-STO interface, is in contradiction to the third model sketched in Figure 79, but consistent with the LRS resulting from an oxygen release process as sketched in see Figure 80(a). Moreover, the flat oxygen vacancy profile determined by STEM-EELS analysis in the HRS (see Figure 81) is in contradiction to the pronounced vacancy concentration dependence of the third model sketched in Figure 79, but consistent with the HRS resulting from an oxygen release
process as sketched in see Figure 80(b). Moreover, the total oxygen content in the LRS shown in Figure 81 is significantly reduced, indicating a complete release of oxygen from the lattice instead of a redistribution as for the third model sketched in Figure 79. This model of an oxygen exchange process for 8W switching devices sketched for the whole operation cycle in Figure 80 has later been explicitly confirmed by $^{18}$O tracer diffusion experiments combined with SIMS analysis [509] that proved the incorporation of oxygen into the lattice from the surroundings during the RESET process.

A major challenge for the application of devices based on this oxygen exchange model is, of course, the storage of the molecular oxygen in the LRS. Obviously, memory cells in information technology cannot rely on a constant open contact to the ambient air. There have been ideas to store the oxygen gas in small pockets underneath the electrode in nano-bubbles or in the grain boundaries of the metal electrode. For example, it is known that Pt may transport and storage oxygen in the grain boundaries of the metal [287,288]. Still, it is hardly conceivable that such a concept is ideally suited for fast and reliable non-volatile memory operation in integrated circuits.

5.7. Switching kinetics

To achieve long-term stability on the one hand and fast switching within nanoseconds on the other hand, ReRAMs require highly nonlinear switching kinetics (Section 2.3.4). For example, the state of the memory cell should not change while a constant low read-voltage $V_{rd}$ is applied.
Figure 81. Operando STEM-EELS analysis of an 8W switching STO thin film device: Upper panel: intensity of the L₃ e₉ edge as a function of depth into the specimen for the LRS and HRS to a depth of 55 nm into the device. Lower panel: oxygen concentration profiles acquired by integrating the oxygen EELS edge for the same region in the HRS and LRS. Adapted and reprinted with permission from [501], © 2017 by Wiley-VCH.

for up to ten years, but the state should change within nanoseconds when the write voltage $V_{wr}$ is applied. As mentioned in Section 2.3.4, $V_{rd}$ cannot be less than approximately one tenth of $V_{wr}$. This requirement has also been called the voltage-time dilemma [11,278,510]. Thus, the nonlinearity of the switching process should cover more than 15 orders of magnitude in time. In order to understand how this high nonlinearity can be achieved in VCM cells, the switching kinetics has to be studied over many orders of magnitude in time. Such experiments did not only prove that the voltage-time dilemma can be solved, but also helped to identify the dominating physical processes that drive the resistance change. In this section, the switching kinetics, the transient behavior during SET and RESET and the ultimate switching speed of VCM cells is addressed. A special emphasis is given on the underlying physical processes that determine the switching kinetics. Here, only the switching kinetics of VCM cells will be discussed. A review on the switching kinetics of VCM cells in comparison to ECM and PCM cells was given by Menzel et al. in 2015 [510].

5.7.1. Switching voltages and times

The switching dynamics of VCM cells can be most suitably studied using pulse experiments. In these experiments, the voltage pulse amplitude is varied and either the change of the resistance or ideally, the transient current response is monitored. Figure 82 (panels a and b) shows a typical SET transient and RESET transients of a filamentary switching Pt/SrTiO₃/TiN [511] and a Pt/Ta₂O₅/Ta cell, [366], respectively. During the application of a negative voltage applied to the electrically active Pt electrode, the current transient (red solid line) exhibits first a current plateau at a low current level of about $-20\ \mu A$, which resembles the HRS. After 6 μs the current jumps abruptly to the LRS level of about $-380\ \mu A$. The time from the beginning of the SET pulse to the abrupt current jump is defined as SET switching time $t_{SET}$. Such an abrupt SET transition is typical for filamentary switching systems. The abrupt switching event implies that a slight change in resistance leads to positive feedback, which self-accelerates the switching process. After the abrupt switching event the change in conductance is more or less zero, which means that the additional energy put into the system does not lead to further switching as discussed by
Strachan et al. [512]. In contrast to the abrupt SET transition, the RESET transition is typically gradual as shown exemplary by the RESET transients of a Pt/Ta2O5/Ta cell for different pulse voltage amplitudes. During the voltage pulse, the current decays gradually to a lower value. This shape of the current decay implies that negative feedback exists that slows down the process with increasing time. As the resistance change is gradual, it is not easy to define a RESET time, but the experimental results clearly show that the transition becomes faster when the voltage amplitude is increased.
The situation is somewhat different for area-dependent RS systems. Figure 82 (panels c and d) shows the current change in the area-dependent switching TiN/a-Si/TiO2/TiN cell upon application of 1 μs long pulse with varying SET/RESET voltage amplitudes [385]. This data set only gives the resistance after pulse application, but still a gradual behavior for both, SET and RESET transitions, is apparent. A similar gradual switching has been observed for an area-dependent switching Pt/BiFeO3/Au cell [470]. In this case, the gradual switching was observed on a timescale from 500 ns to 100 ms (cf. Figure 82(d)). Thus, the dynamic behavior of the SET transition is different for area-dependent and filamentary switching systems. Whereas positive feedback leads to the abrupt SET transition in filamentary switching systems, such positive feedback does not exist for area-dependent switching systems. A potential explanation for this difference will be discussed in Section 5.7.5.

The SET kinetics for different filamentary systems can be easily compared as the switching time can be easily defined. For gradual switching processes, the switching time is often defined as the time required to achieve a certain resistance change. When the defined resistance change varies, the switching time changes, too. Thus, a fair comparison between different systems is not possible for gradual switching processes such as the RESET in filamentary systems and SET/RESET in area-dependent systems. A suitable figure-of-merit for comparing the switching kinetics data of different devices is the voltage increment required to switch one decade faster $\Delta V_{10x}$, i.e. the inverse slope in a log$(t_{SET})$–$V_{app}$ diagram. The smaller $\Delta V_{10x}$ is the steeper is the slope in this plot.

Figure 83 shows SET and RESET switching times for filamentary and area-dependent switching systems as a function of applied voltage. In Figure 83(a), the published switching kinetics data of TiN/HfOx/TiN [513,514], TiN/Ti/HfOx/TiN [515], TiN/HfOx/AlOx/Pt [516], Ti/HfOx/Pt [517], Pt/TiOx/Pt [518], Pt/TaOx/Ta [519], and Pt/SrTiO3/Ti [520] filamentary switching devices is illustrated, as compiled in [510]. The observed inverse slopes are all quite similar in a range of $\Delta V_{10x} = 40–240$ mV/dec. Moreover, most of the devices show only a single slope except for the TiN/HfOx/TiN cells of Ielmini et al. (red open squares) [514], the HfOx data published by Cao et al. (red filled squares) [517], and the SrTiO3 data of Fleck et al. (black open squares) [520]. However, most of the studies only cover less than five orders of magnitude in switching time and another slope might be hidden in the voltage regimes not studied. Figure 83(b) displays the RESET kinetics data extracted from the RESET transients of the filamentary switching Pt/Ta2O5/Ta/Pt device shown in Figure 82(b). Here, the RESET time was defined as the time after which the current decrease is 50% of the total current decay during the 1 μs pulse. The inverse slope is $\Delta V_{10x} = 257$ mV/dec and thus a little bit larger than for the typical SET values. The switching kinetics data for the area-dependent TiN/a-Si/TiO2/TiN system was extracted from a series of pulse measurements as shown in Figure 82, panels c and d. For a certain pulse width, the voltage is increased/decreased until a target LRS/HRS current was achieved. The corresponding voltage is defined as SET/RESET voltage and the pulse time as switching time. The resulting switching kinetics data for SET and RESET are plotted in Figure 83, panels c and d, respectively. In comparison to the filamentary systems the inverse slopes are approximately $\Delta V_{10x} = 400$ mV/dec for the SET and $\Delta V_{10x} = 150$ mV/dec for the RESET. The switching voltages of the TiN/a-Si/TiO2/TiN system are 2–5 times higher than for the filamentary systems. The switching kinetics at different ambient temperatures of the area-dependent switching Pt/Ti-doped BiFeO3/Au cell is shown in Figure 83(d). The voltage needs to be changed from about 7 V to about 21 V to change the switching time by about six orders of magnitude in time, which corresponds to $\Delta V_{10x} \approx 2.2$ V/dec. These high switching voltages and the rather flat slope in the switching kinetics are quite typical for area-dependent systems (cf. Section 5.5).

The fastest switching times shown for filamentary systems in Figure 83 are about 10 ns. This value is related to the limitations of the measurement setup and does not reflect the physical limit
Figure 83. Switching kinetics for filamentary switching systems (a,b) and area-dependent systems (c,d). (a) SET switching kinetics data for VCM cells of HfO$_x$ (red) [513–517], TiO$_x$ (green) [518], TaO$_x$ (blue) [519], and SrTiO$_x$ (black) [520]. A specific slope $\Delta V/\Delta t$ of each oxide material can be identified in a narrow range. Shifts along the voltage axis for same species are related to an increase of the oxide thickness. For HfO$_x$ and SrTiO$_x$-based cells regimes with different slopes are observed. Adapted and reprinted with permission from [510], © 2015 by Wiley-VCH. (b) RESET switching kinetics of Pt/Ta$_2$O$_5$/Ta/Pt cells showing the decay time $\tau_{50}$ for different RESET voltage amplitudes $V_{pulse}$ extracted from the transient currents. Adapted and reprinted with permission from [366], © 2016 by Wiley-VCH. (c) SET and RESET kinetics of a TiN/a-SiO$_2$/TiO$_2$/TiN cell. Adapted and reprinted with permission from [385], © 2015 by IEEE. (d) SET and RESET kinetics of a Au/BiFeO$_3$/Pt cell. Adapted and reprinted with permission from [470], © 2018 by the American Physical Society. Please note that the voltage polarities are taken from the original references. They are not corrected for our notation according to which the voltage is applied to the electronically active electrode (AE).

of the intrinsic device switching speed. For measuring the switching speed below 1 ns a dedicated measurement setup and sample layout are required. Resistive switching in the sub-nanosecond regime has been qualitatively demonstrated for HfO$_2$- [521], Ta$_2$O$_5$- [95], and AlN-based resistive switches [522]. Figure 84, panels a and b, show the $R_{LRS}/R_{HRS}$ ratio of tapered coplanar waveguide (CPW) a Pt/TaO$_x$/Ta/Pt device and a Pt/ZrO$_x$/Ta resulting from a $-2.24$ V SET voltage pulse with different pulse length ranging from 50 ps to 250 ps [96]. Below 100 ps there
is no successful switching event observed for the TaO$_x$-based device as indicated by a ratio of $R_{\text{LRS}}/R_{\text{HRS}}$ about 1 in Figure 84(a). Around 100 ps, the device start to switch and for pulse lengths larger than 120 ps the device switches successfully in every cycle. In contrast, the ZrO$_x$-based devices show stable SET switching at pulse lengths above 60 ps as shown in Figure 84(b). These results show that the nonlinear switching kinetics reported in [523] continue in the sub-ps regime and cover in total over 15 orders of magnitude in time. Figure 84(c) shows the SET switching kinetics data collected on a Pt/TaO$_x$/Ta device with an area of $15 \times 20 \, \mu$m$^2$ (Ta1 device), of $5 \times 5 \, \mu$m$^2$ (Ta2 device), and of $2 \times 2 \, \mu$m$^2$ (Ta3 device) which clearly shows that the switching time could be further reduced by decreasing the RC times of the cells [524]. In all published sub-nanosecond pulse experiments, the ultimate physical switching speed has not been found and the measurement setup limited the switching speed [96,523–525]. A theoretical analysis based on the different limiting ionic processes proposed that the ultimate switching speed limit of redox-based resistive switching devices is determined by the phonon frequency of the switching material. This would be typically of the order of a few ps [526].

5.7.2. Origin of the ultra-nonlinear kinetics

In general, non-volatile memories require a highly nonlinear switching kinetics in order to be considered as non-volatile and enable a non-destructive read-out. In a Flash memory, for example, a very high voltage is applied ($>10$ V) in order to charge the floating gate by Fowler–Nordheim
tunneling through a triangular barrier in a reasonable short time. During read, a low gate voltage is applied and the tunneling barrier is rectangular. This leads to an orders of magnitude lower leakage current preventing the floating gate discharge. To accomplish this high nonlinearity in tunneling currents, a very high tunneling barrier is required that is sufficiently thick in order to prevent electron tunneling current. The latter condition sets a limit to Flash memory scaling. In order to scale to smaller dimensions, tunneling needs to be ruled out as memory loss mechanism. As the tunneling probability of ions and atoms is practically zero, memories that are based on atomic/ionic configurations can be scaled down to smaller dimensions [527]. In phase-change memories (PCM), the atomic configuration is changed from amorphous to crystalline order. This phase transition is induced by local Joule heating. As the activation energies for amorphization/crystallization are very high, a sufficient nonlinearity is achieved [528]. In ferroelectric memories, the required non-linearity is attributed to the field-induced switching of the remanent polarization of the ferroelectric material [529].

The nonlinear switching kinetics of ReRAMs are determined by the electrochemical and physical processes that are involved in the resistive switching effect. These are: (i) ion migration, (ii) electron-transfer (redox) reactions, and (iii) electrocrystallisation/nucleation [510]. For VCM cells, the latter process might be relevant for the electroforming process. Although reversible topotactic phase transitions have been proved in VCM-type Brownmillerite thin film devices [273], the question whether electrocrystallisation/nucleation processes are rate-limiting for the switching kinetics has not been addressed so far. This is in contrast to ECM cells in which the nucleation process of a metallic filament has been found to limit the switching speed, for example in Ag/AgI/Pt cells at low voltages [530]. Oxygen exchange as a redox reaction occurs, for example, in 8W-switching SrTiO3 systems or C8W-switching systems with oxidizable electrodes [361,531,532]. In all VCM systems, ions need to move in order to induce resistive switching effects. The slowest of these processes will finally limit the switching speed in a certain voltage regime. In fact, the limiting process can be different depending on the voltage and temperature regime.

The ion migration processes can be mathematically described by the Mott-Gurney law for ion hopping as mentioned in Chapter 3:

\[
  j_{\text{hop}} = 2z e a f \exp \left( - \frac{\Delta W_{\text{hop},0}}{k_B T} \right) \sinh \left( \frac{a e}{2 k_B T} E \right)
\]  

(45)

Here \( j_{\text{hop}} \) denotes the ionic current density, \( z \) the charge number of the hopping ion, \( e \) the elementary charge, \( f \) the jump attempt frequency, \( \Delta W_{\text{hop},0} \) the hopping barrier, \( k_B \) the Boltzmann constant, \( T \) the local temperature, and \( E \) the electric field. The hopping distance \( a \) is of the order of interatomic distances, that is, 0.2–0.5 nm. Sometimes, \( a \) is used as a simple fitting parameter in simulation models and unphysically high values of several 10 nm are obtained. For high electric fields \( E > 2 k_B T/(a e) \) an exponential relation between current density and electric field results. In contrast, the ionic current depends linearly on the applied electric field for \( E < 2 k_B T/(a e) \). In the literature, some groups used the local Lorentz field to calculate the ionic hopping current, in order to fit the strong nonlinearity of the switching kinetics [381,533,534]. This assumption, however, is not valid as the electric field experienced by the moving ion spatially fluctuates and the Lorentz field gives only the maximum value as pointed out by Meuffels et al [535]. In fact, the average electric field has to be applied, which is defined as the voltage \( V \) that drops over a distance \( w \), i.e. \( E = V/w \). The distance \( w \) can be the thickness of the switching layer or only a small part of it where the switching takes place (thickness of the disc region).
The current \( j_{et} \) that results from electron-transfer reactions at the metal/ion-conducting layer interface is described by the Butler–Volmer equation

\[
j_{et} = j_{0,et} \exp \left( -\frac{\Delta W_{et}}{k_B T} \right) \left[ \exp \left( \frac{(1 - \alpha)ze}{k_B T} \Delta \varphi_{et} \right) - \exp \left( -\frac{\alpha ze}{k_B T} \Delta \varphi_{et} \right) \right]
\]

(46)

and depends on the activation energy \( \Delta W_{et} \), the current prefactor \( j_{0,et} \), and the electron-transfer overpotential \( \Delta \varphi_{et} \). The charge transfer coefficient \( \alpha \) describes the asymmetry of the redox reaction and takes values between 0 and 1, but is typically close to 0.5. The first exponential term in Equation (6.21) describes the oxidation reaction, while the second exponential term describes the reduction reaction. If the overpotential is zero, both processes occur at the same rate and the redox reaction is in a dynamic equilibrium. For \( \Delta \varphi_{et} > 0 \) the oxidation reaction overweighs, whereas the reduction dominates for \( \Delta \varphi_{et} < 0 \).

Electrocrystallisation describes nucleation and crystal growth in electrochemical systems under the influence of electric fields. In this process, a charge transfer is also involved in the formation of a new phase. The nucleation time \( t_{nuc} \), which is the time required to form a stable nucleus, depends exponentially on the nucleation overpotential \( \Delta \varphi_{nuc} \) according to

\[
t_{nuc} \propto \exp \left( \frac{\Delta W_{nuc}}{k_B T} \right) \exp \left( -\frac{(N_c + \alpha)ze}{k_B T} \Delta \varphi_{nuc} \right)
\]

(47)

Here, \( \Delta W_{nuc} \) is the activation energy of the nucleation and \( N_c \) gives the number of atoms that is required to form a stable nucleus.

To summarize, despite their different physical and electrochemical nature, all of these processes have a similar mathematical form. They obey an Arrhenius-type relation and thus can be exponentially enhanced when the temperature increases. In filamentary switching systems, the current density is potentially high enough to enable Joule heating. Moreover, the heated volume is so small that Joule heating occurs quasi-instantaneous [278,526]. In contrast, area-dependent switching systems exhibit lower switching currents and consequently lower current density. Thus, Joule heating is rather unlikely in those systems. In addition, the heated volume is considerably larger than in filamentary systems and the thermal time constant of the system could affect the switching speed. In addition to thermal effects, the activation energy of these processes can be lowered by an applied electric field/voltage, which results in an exponential dependence on the voltage, at least when the electric field is high enough [510]. It depends on the investigated system to which parts the switching kinetics are field- and temperature-accelerated.

5.7.3. General considerations

As different physical processes can in principle lead to a nonlinear switching kinetics, the question arises how to identify the rate-limiting one. Due to the different physical parameters and their limited reasonable range, different degrees of nonlinearity should arise for the different processes. In the following, two different cases will be discussed: First, it is assumed that Joule heating does not occur and the nonlinearity is solely given by the field acceleration. In a second step, the influence of Joule heating will be considered.

When Joule heating is neglected, the barrier lowering due to the applied electric field/voltage leads to the nonlinear relation between switching speed/time and voltage. In a gedankenexperiment, we assume a pulse experiment, in which only one of the processes limits the switching speed. Then, the slopes \( m \) in a logarithm of the switching time \( \ln(t_{sw}) \) versus switching voltage \( V_{sw} \), should depend on the parameters of the underlying process, as mathematically described by
Equations (47)–(49). According to the derivation in \([510,536]\) the slopes

\[
m_{\text{hop}} = -\frac{aZe}{2k_BT w}, \quad m_{\text{red}} = -\frac{aZe}{k_BT}, \quad \text{and} \quad m_{\text{nuc}} = -\frac{(N_c + \alpha)Ze}{k_BT}
\]  

(48)
can be extracted for the ion hopping, reduction, oxidation, and nucleation process, respectively. Instead of using the slope \(m\) for comparison of different processes, one can also use the voltage increment

\[
\Delta V_{10x} = -m^{-1} \ln(10)
\]  

(49)
that is required to accelerate the switching speed by a factor of 10. The voltage increment and the corresponding slope \(m_{\text{exp}}\) can be extracted from the experiment. The product of \(|m_{\text{exp}}k_BT/(aze)|\) is equivalent to the physical constants, i.e. \(a/w, \alpha, 1 - \alpha, N_c + \alpha\) for hopping, reduction, oxidation and nucleation, respectively. Since the hopping distance should always be smaller than \(w\) and the charge transfer coefficient is between 0 and 1, a product \(|m_{\text{exp}}k_BT/(aze)| > 1\) can only be explained with the nucleation process as limiting factor, when physically meaningful parameters are assumed. Moreover, a physically reasonable value for the ratio \(a/w\) should be smaller than 1/5. In that case, the voltage drops over a distance of not less than \(w = 1.5–2.5 \text{ nm}\) for a reasonable hopping distance of 0.3–0.5 nm. Thus, if \(|m_{\text{exp}}k_BT/(aze)| > 1/5\), ion hopping is not very likely to be the only limiting process under constant temperature. In Figure 85, the calculated slopes of the three different processes for the limiting parameters given in the figure caption are plotted in a log\((t)–V_p\) diagram \([510]\). Similar to the discussion above, the nucleation process shows the highest nonlinearity followed by the electron-transfer reactions (oxidation or reduction) and the ion hopping process. For this example, the obtained voltage increments are 397 mV/dec < \(\Delta V_{10x,\text{hop}}\) < 4.97 V/dec for ion hopping, 33 mV/dec < \(\Delta V_{10x,\text{et}}\) < 298 mV/dec for electron-transfer, and 8 mV/dec < \(\Delta V_{10x,\text{nuc}}\) < 27 mV/dec. In general, more than one process can be involved in the resistive switching. In this case, the slowest process will determine the switching speed and thus the slope in the \(t–V\) diagram. As the processes have different nonlinearity, the limiting process can also depend on the investigated voltage regime. The process with the steepest slope will most likely determine the slope at low voltages, whereas a process with a flatter slope will limit the switching speed at higher voltages. Overall, one would expect that the slope in the \(t–V\) diagram is flattening out when the nonlinearity is only determined by the electric field. For experimental investigations of the switching kinetics, it is necessary to cover the whole operating voltage range and as many orders of magnitude in time as possible in order to identify all underlying limiting processes.

The previous analysis is only valid for systems in which Joule heating does not occur. When the temperature increases due to Joule heating, the switching sweep accelerates even more. Thus, steeper slopes in the log\((t)–V\) diagram are expected in such systems in comparison to those with sole field-acceleration. The effect of local Joule heating will be discussed in the following.

The static local temperature depends on the dissipated electrical energy \(P_{\text{el}}\) and can be estimated using

\[
T = T_0 + R_{\text{th}}P_{\text{el}} = T_0 + R_{\text{th}}I(V)V
\]  

(50)
where \(T_0\) is the ambient temperature and \(R_{\text{th}}\) the equivalent thermal resistance. The latter depends on the thermal conductivities of the conduction area of the resistive switching material, the surroundings and the electrodes, and the geometry of the filament. If the conduction area is small, the temperature increase due to Joule heating is quasi-instantaneous and Equation (50) can be applied \([278,537]\). If this area is large, e.g. several tens of \(\text{μm}\), the heating can no longer be assumed to be instantaneous and it takes some time to achieve a stable temperature distribution.
Figure 85. Illustration of the nonlinearity in the switching kinetics obtained for electric-field enhanced nucleation (red), electron-transfer reaction (green), and ion migration (blue) in the limiting scenarios explained below. The different processes cover a different range of slopes in the $t$–$V$ diagram. For the ion migration curves $a = 0.3$ nm, and $w = 2$ nm and 25 nm are chosen as lower and upper limit. The charge transfer coefficient is chosen in a range between $0.1 \leq \alpha \leq 0.9$. For the nucleation $N_c = 1$, $\alpha = 0.1$ and $N_c = 3$, $\alpha = 0.9$ are used. For all lines $z = 2$ is assumed. Reprinted with permission from [510]. © 2015 by Wiley-VCH. (Color figure available online).

The corresponding thermal time constant depends on the heat capacity, the mass density, the thermal conductivity and the heated volume. While heating can be assumed to be quasi-instantaneous in filamentary switching systems, a thermal time constant should be considered for systems displaying area-dependent switching. In any case, as the current densities in the latter systems are typically low, significant Joule heating is not likely to occur. In the following, the temperature-induced acceleration of the ion hopping process is discussed. The conclusions, however, apply to the other processes, too.

Assuming a linear $I$–$V$ relation and the parameters given in the figure caption of Figure 86, the normalized switching time can be calculated by inserting Equation (50) into Equation (45) according to [536]

$$t_{\text{SET}} = \text{const} \cdot \exp\left(\frac{\Delta W_{\text{hop}}}{k_B(T_0 + R_{\text{th}}/R_{\text{HRS}} V^2)}\right) \left[\sinh\left(\frac{aze}{2k_B(T_0 + R_{\text{th}}/R_{\text{HRS}} V^2) w}\right)\right]^{-1} \tag{51}$$

Switching times calculated for two different activation energies (0.5 eV and 1.4 eV) and, for comparison, in the absence of any Joule heating and plotted in Figure 86. For better comparison, all scenarios are normalized to a switching time of 10 years at a readout voltage of 0.2 V. The effect of Joule heating is clearly visible. Upon increasing the applied voltage to 2 V, the switching time for the case without Joule heating (dashed blue line) decreases only by two orders of magnitude. In contrast, the switching time decreased by about five or fifteen orders of magnitude for $\Delta W_{\text{hop}} = 0.5$ eV or 1.4 eV, respectively. Thus, the nonlinearity of the temperature-accelerated process depends on its activation energy. The higher the activation energy, the higher the expected nonlinearity of the switching process. This is important to note since often strategies to increase the switching speed are proposed that are based on the reduction of the activation energy, not keeping in mind that this will cause a decrease of the non-linearity and the retention time.

In a second calculation, the influence of the $I$–$V$ relation on the switching kinetics is investigated. Figure 87 shows the interdependence of non-linearity and dissipated electrical power for two different current–voltage scenarios: (i) an ohmic behavior with $I = V/R$, and (ii) a diode-like behavior, i.e. $I = I_0(\exp(V/V_0) - 1)$. The parameters are given in the figure caption. The local
Joule heating effect clearly increases the switching speed. For low voltages, the ohmic (blue) and the diode-like (red) scenarios equal the constant-temperature case (black). As soon as Joule heating sets in, the slope becomes steeper than in the case of sole voltage/field acceleration with constant temperature. For the ohmic behavior, Joule heating sets in at lower voltages than for the diode-like behavior as the current is higher at low voltages. The crossing point in \( t–V \) diagram marks the point where the \( I–V \) pair of values are identical for both scenarios. Comparing both scenarios reveals that the nonlinearity of the kinetics is highly dependent on the nonlinearity of the \( I–V \) relation. It is remarkable that the switching kinetics differ strongly in the \( t–V \) diagram but are almost the same in the \( t–P \) diagram. This illustrates the strong power-dependence in the chosen scenario. The small difference in the \( t–P \) diagram can be related to the voltage/field acceleration of the ion hopping process. For \( P < 10 \mu W \) the temperature increase is below 15 K and the influence of Joule heating is small. To achieve the same power, however, a higher voltage has to be applied in the diode-like case than in the ohmic case. As a result, the switching speed is slightly higher. In order to prove that the switching kinetics are power-dependent in a real ReRAM device, it is useful to compare the \( t–P \) curves for different programed initial resistances. If only one process is limiting the switching speed, coinciding \( t–P \) curves should result for different initial resistances. This behavior might change if several processes limit the switching speed.

In the discussion above we focused on the question of how to analyze the slope in the \( \log(t)–V \) diagram. Not only the slope but also the maximum achievable nonlinearity is of interest for obtaining long-term stability and ultrafast switching in a given device. The figure of merit for this requirement is elaborated in the following. As all processes obey a similar mathematical law (cf. Equations (49)–(51)) one can estimate the switching time more generally according to

\[
t_{sw} \simeq \text{const} \cdot \exp \left( \frac{\Delta W - \beta V_{\text{app}}}{k_B T} \right)
\]  
(52)

where \( \Delta W \) is the corresponding activation energy, and \( \beta = k_B T m \), with the slope \( m \) according to Equation (50). The retention limit can be approximated by setting \( V_{\text{app}} = 0 \), which yields

\[
t_{sw,\text{max}} \simeq \text{const} \cdot \exp \left( \frac{\Delta W}{k_B T_{\text{amb}}} \right)
\]  
(53)
Thus, the retention time depends mainly on the activation energy of the process determining stability around 0 V. To obtain the minimum switching time, two limits can be considered. For field-accelerated switching, the effective barrier, that is the nominal barrier lowered by the electric field/voltage, cannot be lower than zero. It is rather approaching zero, which has been sometimes overlooked in literature. In this zero-effective-barrier limit

$$
\lim_{\Delta W \rightarrow 0} t_{sw,\text{min}} \simeq \text{const} \cdot \exp \left( \frac{0}{k_B T_{\text{amb}}} \right) = \text{const}
$$

results. The switching time is in this case mainly determined by the constant pre-factors. Still there will be a relation to the applied voltage, but this is non-exponential. In contrast, for temperature-accelerated systems based on Joule heating, the switching limit is obtained when the temperature tends to infinity:

$$
\lim_{T \rightarrow \infty} t_{sw,\text{min}} \simeq \text{const} \cdot \exp \left( \frac{\Delta W}{k_B T} \right) = \text{const}
$$

In this case, the exponential function will approach unity and the limit is again determined by the constant prefactors. Hence, in both cases, the same limit applies. Assuming only one process to be limiting, one can define the maximum nonlinearity to

$$
NL_{\text{max}} = \frac{t_{sw,\text{max}}}{t_{sw,\text{min}}} = \exp \left( \frac{\Delta W}{k_B T_{\text{amb}}} \right)
$$

independent from the acceleration mechanism, i.e. electric field or temperature. The maximum nonlinearity is only a function of the activation energy and the ambient temperature. Thus, the minimum activation energy that is required to span all orders of magnitude in time between 10 y retention and 10 ns programming can be calculated to

$$
\Delta W_{\text{min}} \geq k_B T_{\text{amb}} \cdot \log \left( \frac{10 \text{ y}}{10 \text{ ns}} \right) = \begin{cases} 
T_{\text{amb}} = 300 \text{ K} & \Rightarrow \Delta W \geq 0.98 \text{ eV} \\
T_{\text{amb}} = 360 \text{ K} & \Rightarrow \Delta W \geq 1.18 \text{ eV} \\
T_{\text{amb}} = 525 \text{ K} & \Rightarrow \Delta W \geq 1.72 \text{ eV}
\end{cases}
$$
In the specifications for the application of a resistive device as a computer memory, 85°C (≈ 360 K) is typically given as operating temperature. In this case the effective barrier of the limiting process needs to be larger than 1.18 eV. This activation energy, however, only gives the required amount of the maximum nonlinearity. To fulfill the 10-y retention requirement the activation energy has to be slightly higher [381].

Having the above general considerations in mind, we review briefly the experimental switching kinetics data shown in Figure 83. For VCM systems in general, the migration of e.g. double-positively charged oxygen vacancies is supposed to be a major factor limiting the switching speed, as oxides are slow ion conductors at room temperature. In addition, oxygen exchange processes at the interfaces may limit the switching speed [532]. In order to explain the slopes for the filamentary switching systems in Figure 83(a) by pure field-acceleration the voltage needs to drop over no more than 0.6–2.1 nm according to Equations (50) and (51). This value seems to be quite low. In fact, the transient currents prior to the abrupt SET transition are typically higher than a few μA and hence Joule heating should occur. Thus, a combination of electric field and temperature acceleration is the most likely scenario. For the area-dependent switching system shown in Figure 83(a) the currents are quite low and thus temperature-acceleration due to Joule heating is not likely to occur. The inverse slope extracted for the SET mechanism of the area-dependent switching TiN/a-SiO2/TiO2/TiN cell is \( \Delta V_{10x} = 150 \text{ mV/dec} \). Assuming only field-accelerated ion migration as limiting factor, the additional voltage has to drop over a region of approximately 1 nm according to Equations (50) and (51). This is again a quite low value and maybe a different process is limiting in this case.

5.7.4. Consequences of the ultra-nonlinear switching kinetics and influence of a series resistance

A consequence of the ultra-nonlinear switching kinetics of ReRAMs in general are universal SET and RESET characteristics [537–540]. As illustrated in Figure 88(a) different intermediate resistive states (IRS) can be programmed during the SET operation by setting the current compliance value \( I_{CC} \). Independent of the ReRAM stack, the empirical relation

\[
R_{IRS} = V_{ON}/I_{CC}
\] (58)

holds. When the current compliance level is reached during the SET operation, any further decrease in resistance is accompanied with a decrease in voltage. Hence, due to the highly nonlinear switching kinetics, the driving force for the resistance change decreases exponentially. Thus, any further resistance change is strongly suppressed, although it still occurs at a reduced rate. At the voltage \( V_{ON} \), i.e. the voltage at which the current leaves the current compliance during SET, the driving force for the resistance change is virtually zero for the timescale of the experiment. If the timescale changes, e.g. by varying the voltage sweep rate, \( V_{ON} \) changes slightly [539]. This general SET characteristic shows that multiple bits can be stored in a single device. In an integrated ReRAM cell, the current compliance can be realized using a select transistor in series to the ReRAM cell. By changing the gate voltage of the select transistor its drain current is controlled and thus the IRS is determined.

The RESET current scales with the programmed IRS value and thus depends linearly on the SET current compliance value \( I_{RES} = A \times I_{CC} \), where \( A \) is a system-dependent constant prefactor (cf. Figure 88(b)). This empirical relation can be attributed to the almost linear \( I-V \) characteristic of the ON state in most ReRAMs (in most cases) and the nonlinear switching kinetics. The driving force of the RESET process is strongly voltage-dependent and almost independent on the programmed IRS. Thus, the voltage marking the onset of the RESET transition \( V_{RES} \) is constant.
Figure 88. (a) ON resistance vs. SET current and (b) corresponding RESET current vs SET current for VCM, TCM and ECM systems. Experimental data are displayed for a Ag:GeSe ECM-system [542], a Cu:SiO₂ ECM-system [542], a NiO TCM-system [543], a HfOₓ VCM-system (green dots) [118], a bilayer HfOₓ/ZrOₓ VCM-system [544], a TiOₓ VCM-system [545] and a HfOₓ VCM-system (red dots) [546]. TCM and ECM are included for comparison and to show that they display the same behavior. Reprinted with permission from [547], © 2012 by Wiley-VCH.

for one distinct ReRAM. Combining the linear $I$–$V$ characteristics and the almost constant $V_{RES}$ yields

$$I_{RES} = V_{RES}/R_{ON} = V_{RES}/V_{ON} \cdot I_{cc} = A I_{cc}. \tag{59}$$

The constant $A$ is thus identified as the RESET to ON voltage ratio. These two empirical laws have been derived theoretically for two different models. In the first, the resistive switching is explained by the temperature-assisted lateral growth of a conducting filament [541]. The second model describes the switching process as an electric field-driven modulation of a tunneling gap [538,539]. Thus, the universal SET and RESET characteristics are independent of the detailed switching mechanism. The preconditions for the occurrence of this behavior can be summarized as follows [539]:

1) The empirical law $R_{IRS} \propto I_{CC}^{-1}$ has its origin in the highly nonlinear switching kinetics that leads to a huge decrease of the driving force of the resistance change under current control. In this case, the physical origin of the switching kinetics is irrelevant. It is also independent of the physical representation of the LRS state, e.g. variable gap or variable diameter mode. This behavior should therefore be expected for filamentary as well as area-dependent switching systems.

2) The relationship $I_{RES} \propto I_{cc}$ results from the strongly nonlinear switching kinetics and a linear $I$–$V$ relationship of the IRS state. All ReRAMs that fulfill these two conditions also obey the empirical RESET law.

3) The empirical RESET law might also apply to ReRAMs with a nonlinear $I$–$V$ relationship if the current asymmetry in the positive and negative voltage branch is independent of the programed resistance state. If the current asymmetry depends on the resistance state, the ReRAM cell will not obey the empirical RESET law.

Experimentally, the current compliance can also be implemented using variable series resistances. In fact, a transistor in series could be regarded as a realization of such a tunable resistor $R_S$. When the resistance of the switching device approaches the series resistance, the voltage divider between the actual switching device and the resistance changes. The voltage over the
ReRAM cell will decrease and the driving force for the resistance change will drop by orders of magnitude due to the nonlinear switching kinetics. For a voltage sweep experiment, this voltage divider will lead to the SET transition at a certain voltage \([548]\). The current compliance in this case is given when the applied voltage drops completely over the series resistor \(R_S\) according to \(I_{CC} \leq V_{app}/R_S\). The voltage \(V_{app}\) is the applied pulse voltage or the maximum voltage during a voltage sweep. The compliance current is not only given by the series resistance but by the voltage amplitude, too. Thus, storing multiple bits in one ReRAM cell could also be achieved by changing the voltage amplitude for a constant series resistance. This has been demonstrated on sub-ns timescales for a Pt/Ta\(_2\)O\(_5\)/Ta/Pt device \([523,525]\). The series resistance does not need to be an extrinsic one. Many devices show a self-limiting behavior during SET that can be attributed to an internal series resistance. Figure 89(a) shows the \(I-V\) switching characteristics of a nanocrossbar filamentary switching Pt/HfO\(_2\)/TiO\(_2\)/Ti/Pt device for different current compliances \([316,549]\). At low current compliance values, the current jumps abruptly into the current compliance during the SET operation and a gradual RESET transition is observed. When the current compliance level is increased the current starts changing gradually after an abrupt current jump. In this regime, the resistance of the actual action switching region (interface) approaches the resistance of an internal series resistance. Figure 89, an additional effect can be observed. For lower IRS, the RESET voltage shifts to higher absolute values and the switching becomes abrupt. This effect can be explained by the voltage divider effect due to the series resistance, too. Because part of the voltage drops over the series resistance, the RESET transition starts at a higher total voltage. Still the voltage over the active region is constant. For lower IRS values, less voltage drops over the active region and the totally applied voltage needs to be increased in order to initiate the RESET transition. When the RESET transition starts, a self-accelerating effect sets in. When the active device region becomes more resistive, the voltage divider changes and the voltage over this active device region increases. This enhances the driving force and the RESET transition becomes faster, which leads to the abrupt RESET transition. When the internal series resistance is much smaller than the active device region resistance this positive feedback vanishes and the devices shows the typical gradual RESET transition again. Under the assumption that the RESET transition always starts at the same voltage, the internal resistance can be extracted \([316,549]\). By subtracting the voltage drop over this series resistance as suggested by Wouters et al. \([548]\) the ‘true device’ behavior in Figure 89(b) is obtained. For the SET operation, the resistance change starts with a snapback (abrupt SET) followed by a transition at constant voltage, which resembles the gradual SET transition due to the voltage divider effect. The RESET operation starts always at the same voltage. For the low IRS a sharp snapforward occurs which equals the abrupt RESET event. In summary, a series resistance can influence the SET and RESET characteristics significantly. The abrupt SET changes to a gradual one and the gradual RESET becomes abrupt. This behavior has been verified by simulation using the \textit{Jülich Aachen Resistive Switching Tools} JART VCM v1 compact model \([316,550]\), which is based on ion migration, modulation of the interface resistance at the AE, and Joule heating.

As shown in the Figure 89(a), the SET switching process becomes more gradual in the high current regime due to a voltage divider effect. This behavior can be utilized to program different resistance states by using different voltage amplitude. In fact, as the switching is progressing during the whole switching pulse, the programmed state will also depend on the length of the applied pulse. This multilevel programming scheme has been investigated using different pulse lengths and voltages over many orders of magnitude in time. Figure 90 shows the programmed resistance state as a function of the applied voltage \(|V_P|\) and pulse length \(t_p\) \([523]\). It could be
Figure 89. (a) Measured $I(V_{\text{app}})$-characteristic of a Pt/Ti/TiO$_x$/HfO$_2$/Pt ‘HOTO’ device shown in the inset. The current compliance is raised in steps of 100 $\mu$A, starting at 100 up to 800 $\mu$A. For increasing $I_{cc}$ values, the SET event becomes self-limited, the RESET voltages shift to higher negative values, and the RESET behavior changes from gradual to abrupt. (b) Intrinsic switching characteristic $I(V_D)$ of the ‘HOTO’ cell calculated for the current compliance series $I(V_{\text{app}})$ shown in (a). From this analysis, an internal series resistance of $R_S = 1160 \Omega$ and intrinsic SET and RESET voltages of $\pm 0.4$ V, respectively. Adapted and reprinted with permission from [549], © 2016 by IEEE. (Color figure available online).

Figure 90. 3D point plot of the measured data $(R,V_p)$ of a TaO$_x$ CPW cell for a pulse width $t_p$ and fitted behavior of the programed resistance state $R$ using the fit parameters $|V|_{\text{min}}$ and series resistance $R_S$ pursuant to Equation (60). Reprinted with permission from [523], © 2020 by Springer Nature Limited [549].

shown that the programed resistance state can be estimated using

$$R = \frac{R_S}{1 - |V(t_P)|_{\text{min}}/|V_P|}. \quad (60)$$

Assuming a voltage divider, $R_S$ is the series resistance due to the contact resistances and $|V|_{\text{min}}$ is the minimum voltage at which the resistance changes for a given pulse length. The fitted values based on this equation are shown in Figure 90 as well. The minimum voltage follows exactly the trend of the switching kinetics shown for this cell in Figure 84(c). The extracted series resistance is around 150 $\Omega$ for all data points. This study shows that the multilevel switching capabilities are closely linked to the switching kinetics of the device.
Figure 91. Calculated SET time according to Equation (61) depending on the applied voltage for the four different cases in red and blue. The black line corresponds to the mean disc temperature used for the calculations. The experimental data are shown as open circles. Adapted and reprinted with permission from [278], © 2011 by Wiley-VCH. (b) SET times of a filamentary switching nanocrossbar Pt/SrTiO₃/Ti cell as a function of voltage. Experimental data points are shown in blue and in black the model fit based on Equation (61) from [520]. (Color figure available online).

5.7.5. SET kinetics

After the general considerations discussed before, we will now focus the discussion on the origin of the nonlinear SET kinetics of specific VCM systems. As mentioned before, for filamentary systems Joule heating is very likely to occur due to the high current densities. As only a small filamentary region (mainly the disc) is heated, the Joule heating can be considered to be quasi-instantaneous. Evidence for the importance of Joule heating is given by several groups using a combined experimental and simulation approach. Menzel et al. used an electro-thermal model to investigate the SET switching kinetics of SrTiO₃-based filamentary-switching VCM cells [278]. In this study, the switching speed using the drift velocity of the doubly-positively charged oxygen vacancies defined as $v_D = j_{\text{hop}}/(z_i e c)$, where ionic hopping current density $j_{\text{hop}}$ is given by Equation (46). It is assumed that the abrupt SET transition sets in after the oxygen vacancies migrated a particular distance, i.e. the disc thickness as illustrated in Figure 59, which resembles the region of the filament close to the electronically active electrode. The SET switching time in a pulse experiment can then be calculated by $t_{\text{SET}} = l_{\text{disc}}/v_{D}$ and depends on the local disc temperature and the driving electric field within the disc $E_{\text{disc}} = V_{\text{disc}}/l_{\text{disc}}$. In combination with Equation (46)

$$
\Delta t_{\text{SET}} = \frac{l_{\text{disc}}}{2af \exp \left( - \frac{\Delta W_{\text{hop},0}}{k_B T_{\text{disc}}} \right) \sinh \left( \frac{a e c}{2k_B T_{\text{disc}}} E_{\text{disc}} \right)}
$$

results as an estimation for the switching time which is similar to Equation (52). In order to calculate the unknown quantities $T_{\text{disc}}$ and $E_{\text{disc}}$ the heat and the current continuity equation were solved in a 2D axisymmetric domain. Using this model, the experimental SET switching kinetics data could be nicely fitted as illustrated in Figure 91(a). Moreover, the field and temperature contributions to the switching time could be deduced. When only field-acceleration was considered (blue dashed line) the switching speed is only accelerated by less than two orders of magnitude. In contrast, the temperature increase shown as black solid line leads to an acceleration of up to 10 orders of magnitude (red solid line) and is thus the dominating contribution for these systems.
Fleck et al. derived an analytical model to determine the quantities $T_{\text{disc}}$ and $E_{\text{disc}}$ for determining the SET time using Equation (61) [520]. With this model the experimental data of a SrTiO$_3$-based filamentary-switching cell could be accurately described as shown in Figure 91(b). This data set is identical to the one depicted with black open squares in Figure 83(a). Note that Equation (61) is only valid when current and voltage are constant over time. This means the drift velocity is constant with time. Of course, this precondition needs to be verified. This analysis was later applied to switching experiments in the ps regime where the pulse-shape arriving at the cell is strongly affected by charging speed [524] as shown before (Figure 84(c)).

If temperature acceleration dominates, a clear power-dependence of the switching time on the dissipated power should be expected as illustrated in Figure 87. This has been experimentally verified for filamentary switching TaO$_x$-based VCM cells [510,519].

The role of temperature was further verified by analyzing the temperature-dependent switching kinetics of a Pt/Ta$_2$O$_5$/Ta in a special four-point nanocrossbar geometry, in which one of the lines could be used as a nanoh heater enabling heating times in the nanosecond regime [551]. It is demonstrated that the switching speed could be accelerated by orders of magnitudes at low voltages and only little at higher voltages. This is could be consistently explained with the estimation of the SET time discussed above.

Further studies investigating the temperature evolution and confirming the role of Joule heating during switching were given by different groups. Kwon et al used a conductance calibration method to derive the internal temperature during switching [274]. Yalon et al used a special device structure integrating the resistive switching cell into the emitter of bipolar transistor to detect the local temperature [552,553].

The switching model developed by Marchewka et al, which has been presented in Section 5.4.2 was validated by simulating the SET transition of a MIM stack with asymmetric barrier height of 0.7 and 0.1 eV, respectively [322]. Figure 92 shows the simulated SET behavior upon application of voltage pulses with different amplitude. The transient current response in Figure 92(a) resembles the typical behavior observed in experimental data (cf. Figure 82). First, the device current is almost constant until a sharp SET transition appears followed by a second current plateau, which resembles the LRS. The current spike is related to the evolution of the oxygen vacancy concentration. As illustrated in Figure 63 the oxygen vacancies will be depleted at the ohmic electrode. This will lead to a ‘small’ RESET at this interface, which manifests as the current decrease after the spike in the current transients. Due to the rather high asymmetry between the two metal/oxide interfaces, the resulting state is the LRS. The simulation results exhibit a very high nonlinearity in the SET switching kinetics (cf. Figure 92(b)). The origin of this strong nonlinearity is the temperature acceleration of the ionic drift due to Joule heating. As illustrated in Figure 92(c) the temperature during the current plateau before the sharp current jump increases approximately by 400 K when increasing the voltage by 0.4 V. Thus, the sharp SET transition is faster by several orders of magnitude.

Fleck et al. investigated the SET transition of different filamentary switching cells (Pt/SrTiO$_3$/TiN, Pt/TaO$_x$/Ta, Pt/ZrO$_2$/Ta) in more detail [511]. The experimental results revealed that the resistance already decreases gradually before the sharp SET transition as illustrated exemplarily for the Pt/SrTiO$_3$/TiN cell in Figure 93(a). This pre-SET slope, termed preSSs, in fact depends on the applied voltage and is thus related to the SET time, which marks the onset of the sharp SET transition. Figure 93(b) demonstrates that the pre-SET slope changes by orders of magnitudes and is inversely proportional to the SET time. This means that the biggest pre-SET slope should be expected for the fastest SET switching. Indeed, the simulation current transient for 1.34 V shown in Figure 92(a) exhibits such a pre-SET slope. The experimental results could be explained using a compact model that is based on the modulation of a Schottky barrier due to the change in dopant concentration at the Schottky barrier interface [511]. The simulation
Figure 92. Simulation results based on the continuum model of Marchewka et al. (a) Transient currents during the SET operation upon voltage pulse excitation with different amplitudes $V_{\text{pulse}}$ for a bipolar switching structure, (b) SET time as a function of applied voltage, (c) Relationship between set time and temperature $T_{\text{preset}}$ in the oxide layer before the SET operation. Reprinted with permission from [322], © 2015 by IEEE.

results in Figure 93(b) reproduce the experimental data very well. The gradual and the abrupt switching regimes can be explained with this model as follows. In the gradual switching regime, the dopants move slowly and the electron barrier is only modulated slightly. This gradual current increase determines the pre-SET slope. For higher voltages the dopants move faster due to the increased local temperature and the pre-SET slope becomes steeper. The steady increase in current leads to more Joule heating, which in turn enhances the dopant velocity. This process will finally result in self-accelerated thermal runaway process that manifest as abrupt current jump in the SET transient. Later, it was shown that the pre-SET slope increases with temperature in the same way the switching time is reduced by the temperature [551]. Thus, the relation between pre-SET slope and SET time stayed constant. All data points extracted at the different temperatures fall on the same line. Note that the compact model employed in this study [511] gives qualitatively the same result as the continuum model of Marchewka et al. [322]. This compact model, called JART VCM v1(v1b) has been also successfully applied to filamentary switching Pt/HfO$_2$/TiO$_x$/Ti cells [554,555].

As can be observed in Figure 93(a), the transition time of the abrupt switching event is very short but finite. Typically, the transition time is much smaller than the time it takes to reach the thermal runaway regime (also called delay time). For these cases, there is only a small error
by identifying the SET time with the delay time rather than the sum of delay time and transition time. By carefully adjusting the initial state and the pulse length, it was shown that the switching could directly start at the onset of the transition and thus a gradual switching can be achieved also in the regime of the thermal runaway [554]. To this end the pulse length must be shorter than the transition time and the devices resistance needs to be in a limited resistance window.

To emphasize the role of Joule heating in explaining the self-accelerated runaway process, Fleck et al. performed a simulation at constant temperature, i.e. neglecting the role of Joule heating. In that case, the SET transition does not show any abrupt jump but a linear increase in current on a logarithmic time scale. This gradual behavior is in fact consistent with the observed current transients in the area-dependent switching system shown in Figure 82. The question is why there is no self-accelerated process without Joule heating. For a system without Joule heating, such a behavior could only be expected when the electric field would increase during the SET process. The applied voltage, however, is kept constant during the pulsed SET operation. Moreover, in the HRS the voltage drop is more confined within the filament, e.g. in the disc close to the electrically active interface, as in the LRS. Thus, the electric field during the SET transition should even decrease and a gradual switching is expected. In summary, the observation of an abrupt SET switching event indicates a significant influence of Joule heating during the SET switching. Vice versa, the appearance of a gradual SET switching event as in the area-dependent switching systems indicates that Joule heating is not involved.

So far, ion migration is considered as the limiting mechanism for C8W switching systems. In this case, the retention of the programed devices states is described by ion redistribution due to diffusion, at least, as no other process stabilizes a ‘frozen-in’ configuration. The values reported in Ref. [381] are similar to the activation energies required to cover 15 orders of magnitude in switching time, which were calculated using Equation (58) For migration enthalpies smaller than 1 eV, one might wonder whether the goal of 10 year retention can be achieved. However, the retention loss might be determined by another limiting process, which only limits the switching speed at very low voltages. For a filamentary switching Ta2O5 cell, for example, it was demonstrated that the SET switching speed depends on the choice of the...
ohmic electrode [361]. While a faster SET switching was obtained when Ta electrodes were used in comparison to W electrode, the RESET transition became slower for the Ta electrodes than for the W electrodes. This behavior suggests that a redox reaction taking place at the ohmic electrode/oxide interface is involved in the switching event as well. This redox reaction might limit the switching speed at low voltages and it would thus define the retention properties. The physical processes that determine the retention will be discussed in more detail in Chapter 7.

The switching kinetics of SrTiO₃-based 8W switching cells was investigated by Siegel et al [532]. To this end, three different stacks were investigated: a Pt/SrTiO₃/Nb:SrTiO₃ stack, a Pt/SrO/SrTiO₃/Nb:SrTiO₃ and a Pt/Al₂O₃/SrTiO₃/Nb:SrTiO₃ stack. The interlayers were found to improve the retention behavior as in the SrTiO₃-based cell [270] in a previous study (cf. Chapter 7). As the 8W switching mechanism involves the oxygen exchange at the AE, three different processes can be identified as potentially rate-limiting as illustrated in Figure 94(b). The first one is the migration of the oxygen defects in the SrTiO₃. The second is the oxygen exchange at the interface between SrTiO₃ and Pt (or the interlayer). If an interlayer is used, the migration of the oxygen defects in the interlayer could also be rate-limiting. Figure 94(c) shows the SET times as a function of the voltage for these three stacks. Using a similar approach for the estimation of the switching time as in Equation (61) for all three processes, revealed the limiting process. For the Pt/SrTiO₃ cell, the limiting process is the interface exchange, whereas the oxygen defect migration within the interlayers limits the switching speed for the two other systems. The fastest switching is observed for the system without interlayer, but this system also shows the worst retention properties. Thus, it could be shown that there is a tradeoff between data retention and switching speed (cf. Chapter 7).
5.7.6. **RESET kinetics**

In contrast to the SET transition, the RESET transition in VCM devices is typically gradual. This phenomenon can be used to program different intermediate resistance states (IRS) by changing either the RESET ‘stop’ voltage in sweep measurements [235,556–558] or the RESET voltage amplitude in pulse experiments [347,516,559]. Only a few studies on the RESET dynamics have been published so far [512,516,554,559,560].

As discussed before, the RESET transition in VCM devices is typically gradual. This section focuses on the explanation of the RESET switching dynamics and the origin of the gradual RESET phenomenon. In [366] the authors used a combined experimental and numerical modeling approach to address this question. The employed numerical model has been introduced in Section 5.4.2. Here, the axisymmetric model geometry shown in Figure 95 is considered [366]. It comprises a 5 nm thick TaO\(_x\) film sandwiched between a Ta/Pt top electrode and Pt bottom electrode, which is deposited on a SiO\(_2\) substrate. The switching occurs in a filamentary region within the TaO\(_x\) film with high oxygen vacancy concentration. It is assumed that the TaO\(_x\) forms an ohmic-like contact with the Ta electrode (OE) and a Schottky-like contact with the bottom Pt electrode (AE). In order to switch the device from the LRS to the HRS the oxygen vacancy concentration needs to be depleted at the Schottky-like contact as discussed in Section 5.4.2.

This model was applied to analyse the gradual RESET transition in TaO\(_x\)-based VCM cells [366]. The transient currents upon voltage pulses with a rise time of 2 ns, a duration of 1 μs and different voltage amplitudes were simulated. Figure 96(a) shows the simulated current transients (in color) for pulses with amplitude \(-1.3, -1.4, -1.5\) and \(-1.6\) V compared to experimental
data. The model reproduces the experimentally observed transient behavior very well. The point C in each transient marks the decay time $\tau_{50\%}$ when the current drop is half of the total current drop occurring during the pulse, i.e. the difference in currents in point A and E. As shown in Figure 96(b), the decay time depends exponentially on the voltage pulse amplitude, which illustrates the nonlinearity of the RESET switching kinetics [366]. To accelerate the switching speed by one order of magnitude, a voltage increment of 257 mV is required, which is larger than for the SET operation. The measured decay time is reproduced well by the simulation model. By analysing the simulated transient current contributions shown in Figure 96(c), temperature and concentration profiles, the origin of the gradual RESET transition could be identified. Due to the high current density, local Joule heating occurs and the ions drift within the applied electric field toward the ohmic electrode. As the ions redistribute, the potential barrier at the Schottky-like contact is increased and thus the current decreases. This leads to a decrease in temperature. Thus, the total ionic current is reduced (thin solid line) and the driving force for the RESET transition is lowered. In addition, a concentration gradient builds up during the RESET transition and ion diffusion sets in that counteracts the ionic drift built in Figure 96(c). At the beginning of the pulse, the oxygen vacancy concentration is approximately homogeneous and the major contribution to the total ionic current density is the ionic drift (dotted line). Then, a concentration gradient builds up, which gives rise to an ionic diffusion current that counteracts the ionic drift (dashed gray line). Over time, these two current contributions approach an equilibrium state. A third contribution is the thermo-diffusion current (thick solid line), which is driven by the local temperature gradient. However, this current is quite small compared to the other two components. To conclude, the gradual nature of the RESET transition can be explained by the temperature-accelerated oxygen-vacancy motion with the drift and diffusion processes approaching an equilibrium situation, combined with a moderate sensitivity of the current response to the induced contact barrier changes [366].

The bulk-switching model presented in Section 5.4.1 has been also successfully used to simulate the RESET kinetics of a filamentary switching TiN/HfO$_2$/TiN cell [561]. In contrast to the model of Marchewka, the temperature in the rupturing filament does not decrease here, but the electric field drops and slows down the resistance decrease. The temperature does not decrease in this model as the authors assumed that the thermal conductivity is linked to the electrical conductivity via the Wiedemann-Franz law. Thus, the thermal conductivity decreases in the ruptured region and a hot spot occurs. The Wiedemann Franz law can actually only be applied when the heat is carried by electrons. But, in the oxides the main component that determines the thermal conductivity are phonons and thus a constant thermal conductivity is a more realistic assumption [562].

In the previous RESET study, a fixed pulse widths of 1 μs was used. The decay time, which has been used as measure for the RESET transition, is not a fixed quantity but rather depend on the pulse length unless an equilibrium state has been achieved. To reach an equilibrium state the simulation time has been prolonged while keeping all other parameters. Figure 97(a) shows the resulting current transients for four different voltage amplitudes [367]. The time to reach the equilibrium current $I_{\infty}$ depends on the applied voltage amplitude. For higher voltage amplitudes a higher local temperature is generated as the RESET transition becomes faster. In addition, the temperature in equilibrium is higher for higher voltage amplitudes than for lower voltage amplitudes. Such a behavior was observed in filamentary switching Pt/SrTiO$_3$/TiN cells, which were prepared according to [511]. To achieve a time resolution ranging from 1 ns to 0.1 s, pulse trains with increasing pulse lengths were applied and the measured current transients put together. In addition, for each voltage amplitude several combined pulse train transients were measured. The combined transients for different voltage amplitudes are compiled in Figure 97(b). Main features of the simulation results appear in the experimental transients. First, the RESET transition is
Figure 96. Comparison between simulation and measurement of (a) RESET transient currents for pulse voltages of $-1.3$, $-1.4$, $-1.5$ and $-1.6$ V, (b) 50% decay times as a function of pulse voltage, (c) current $I_0$ at the beginning of the pulse and current $I_{1\mu s}$ at the end of the pulse as functions of pulse voltage. Adapted and reprinted with permission from [366], © 2016 by Wiley-VCH. Note that the voltage polarity is opposite to the standard definition in this paper.

faster for higher voltage amplitude. Second, kind of a current plateau appears before the current degrades stronger. This plateau region spans over a longer time for lower voltages. According to the simulation results, this behavior can be explained by the initial doping profile. To model the LRS a slight accumulation of mobile dopants at the Schottky-type interface was assumed. During the RESET transition, the slight accumulation will degrade and a more homogenous distribution will result. In this regime, the current degrades only slightly or it might even increase as the homogenous distribution gives the lowest resistance state [366]. Then, the mobile dopants are depleted at the Schottky-type interface, a concentration gradient builds up and the current decreases.

The pulsed-train programming experiments on HfO$_x$-based VCM cells from Zhao et al. [559] supports the hypothesis of a voltage-dependent equilibrium state. In this study, the authors used RESET pulses with varying amplitude to program IRS. When using a single pulse, a quite high resistance variability was observed. This variability could be significantly reduced by applying a
Figure 97. (a) Simulation of reset equilibria [367] and (b) experimental RESET characteristics on a Pt/SrTiO$_3$/TiN cell for different applied voltages. Adapted and reprinted with permission from [367], © 2016 by IEEE.

pulse-train of identical pulses, thus, increasing the total programming time. If the programming time is prolonged close to the equilibrium, less variability can be expected as the resistance change becomes smaller. In addition, the effect of different initial states becomes smaller [367].

For voltage sweep experiments, Degraeve et al. proposed in 2012 a ‘reset balance line’ that defines the RESET transition [327]. In this case, a drift-diffusion equilibrium is achieved for every voltage in the rising edge of the voltage sweep. When reducing the voltage, the last equilibrium state gets frozen in. In consequence, at a smaller voltage amplitude with RESET polarity the resistance state will degrade towards a better conducting IRS on the ‘reset balance line’. The latter effect has been demonstrated by long-term voltage disturb experiments.

Figure 98 shows a RESET kinetics study of a nanocrossbar Pt/HfO$_2$/TiO$_x$/Ti device [554]. In this study, the RESET time as a function of the initial LRS and the applied voltage were studied. Figure 98(a) shows the programmed initial LRS state. These resistance states are very low ohmic and lie in the regime where the effect of an internal series resistance becomes visible (cf. Figure 89). This is also observed in the RESET current transients upon the application of a 1 V rectangular RESET pulse (Figure 98(b)). First, the current level stays relatively constant until an abrupt RESET event occurs, which is followed by a more gradual RESET transition at lower currents. The delay until the abrupt current drop occurs depends strongly on the initial LRS. The lower the initial resistance is, the longer is the delay time. A change of the delay time of up to 6 orders of magnitude is observed while the initial resistance changes only slightly. These dynamics could be explained using the JART VCM v1b model that is based on the migration of oxygen defects, Joule heating and the change of the interface resistance. It is demonstrated that the change of the delay time is related to the voltage divider effect. For the lowest LRS, only a small part of the voltage drops over the actual device. This results in a very slow switching. For higher LRS the initial voltage drop is higher and device switches faster. As discussed in Section 5.7.4 the voltage over the actual device increases during the progression of the RESET. This leads finally to the abrupt RESET switching event. When the device is a lot more resistive than the series resistance, the positive feedback effect due to the change of the voltage divider stops. Now negative feedback sets in as the decreasing current leads to a decrease in the local temperature. This leads to the gradual RESET behavior towards the end of the pulse. The simulation results show also that the RESET transition time is constant for a specific voltage (see Figure 98(b)). Only the delay time is influenced by the initial state. Figure 98(c) shows the RESET time as
Figure 98. RESET kinetics study of a Pt/HfO2/TiOx/Ti cell. (a) Intermediate LRS of the devices programed by application of a SET pulse of 10 μs and the given voltage amplitude. The black circles show the initial states of the corresponding simulations in (b). (b) RESET current transients at a constant RESET voltage of 1.0 V for the different initial LRSs of (a). Low LRS (high conductance) values lead to pronounced delays during the RESET operation. The simulated transients (black solid lines) are able to fit the variation of the delay by assuming different initial states. The zoom of the simulated transients illustrates that the transition time is state-independent. (c) RESET switching kinetics for the various LRSs programed with different SET pulse amplitudes shown in (a). A delay of up to six orders of magnitude in switching time is observed. The simulations (solid lines) predict the voltage-time dependence well. Adapted and reprinted with permission from [554], © 2019 by the American Institute of Physics. (Color figure available online).

Figure 99 shows the redistribution of oxygen vacancies during the RESET process. Due to the applied electric field, the oxygen vacancies migrate from the electrically active electrode interface (here Pt) to the ohmic electrode interface. Thus, a concentration gradient builds up with the depletion of Vo at the AE and a Vo accumulation at the OE interface (cf. Figure 99(a)). The concentration gradient leads to an oxygen-vacancy diffusion current (cf.
Figure 99. (a) Redistribution of oxygen vacancies under the electric field (E) during the RESET process. The sketches are based on simulations of O vacancy concentration profiles during ReRAM switching, adapted from [366]. (b) Diffusion of oxygen vacancies counteracting the concentration gradient by the drift process. (c) Oxygen exchange with the ohmic electrode resulting in lower concentration of oxygen vacancy at the Pt interface. From [361].

Figure 99(b)) counteracting the drift current. By that, the Vo accumulation at the ohmic electrode will be reduced. As discussed before, an equilibrium state could be achieved that is determined by the applied voltage amplitude. In this way, different IRS can be programed. If an oxygen exchange reaction happens at the ohmic electrode interface as illustrated in Figure 99(c), the high concentration of Vo will be reduced and the back diffusion force hindering a deeper RESET will be lowered. Consequently, more resistive IRS could be programed with oxygen exchange reaction as without oxygen exchange. In the cited work, the RESET process in Pt/Ta₂O₅/Me cells with different ohmic Me electrode was investigated. A higher IRS was found for W than for Ta when used as ohmic Me electrode. W has a higher oxygen chemical potential than Ta and hence the oxygen exchange reaction is expected to be more thermodynamically favored. In turn, the SET transition is faster for Ta electrodes as it can be reduced more easily compared to W electrodes [361].
In this section analyses of the RESET kinetics were only discussed for filamentary switching systems. However, when Joule heating is neglected, a drift-diffusion equilibrium should also be established during RESET. Therefore a gradual RESET process should also occur in area-dependent switching systems. In fact, the measured $I–V$ curves of area-dependent systems exhibit a gradual RESET switching behavior (cf. Section 5.5). It can be assumed that the switching polarity depends on a competition between ionic transport rates in the oxide and the ion exchange rate at the electrode interface.

In summary, it should be stated that the gradual SET as well as the gradual RESET can be physically explained by the simulations described in this section.

5.8. Ultimate physical limits to scaling

One very important question for commercialization is the scalability of VCM cells. Is there a physical limit to how small a VCM cell can be built? And how this limit influenced by the various performance parameters of VCM cells? In this section, we will first show the current state of the experimental findings and then discuss the theoretical limits to the scaling of VCM devices.

The aspects of the integration technology with respect to the scaling prospect follow the conventional lines, because the oxide materials considered for the I-layer in VCM cells are typically very temperature stable and easy to process. As an example, Figure 100 shows a TiN/HfO$_x$/Hf VCM cell with a cross section of 10 nm $\times$ 10 nm [212]. Thermally oxidized WO$_x$ has been used in a 2k VCM-type cell array with 9 nm halfpitch cells in which W acts as the OE and electronically conducting TiON as the AE [563]. A sidewall technique has been employed to fabricate sub-5 nm electrodes and to achieve memristive 1 nm $\times$ 3 nm HfO$_2$ cells [564]. Finally, a 3 $\times$ 3 array of Pt/HfO$_2$/TiO$_x$/Pt memristive elements with a 2 nm feature size and a 6 nm half-pitch has been fabricated using Pt nanofin electrodes to reduce the line resistances [565].

Of course, an increase in the ReRAM device density in a projection onto the plane can be achieved by 3D integration, either by stacking layers of planar arrays of memristive elements on top of each other or by a true vertical arrangement of the ReRAM devices (VRRAM) adapting the 3D vertical NAND technology as suggested in Refs. [566,567]. An interesting fabrication variant has been reported in [568], in which an (unipolar switching) p-Si/SiO$_2$/n-Si structure was
Advances in Physics

Figure 101. (a), A conductivity map of the surface of a SrTiO$_3$ single crystal as recorded by the C-AFM technique. Filamentary paths with enhanced conductance are present on the surface after thermal reduction and re-oxidation under ambient conditions. Inset: spot with a dimension of 1–2 nm, where the main current is concentrated in a region corresponding to the size of the core of a typical edge-type dislocation. (b), Line scan across the selected spot ($D$ denoting distance along AB) showing the dynamic range of the resistance change as a result of the application of a negative tip voltage bias, that is, selective electroformation. Right: Conductivity maps of the selected spot before (upper trace) and after electroformation (lower trace). Adapted and reprinted with permission from [119], © 2006 by Springer Nature Limited.

exploited and a 3D crossbar array with up to five layers and 100 nm memristive elements was built using a fluid-supported Si membranes technique. An 8-layer VRRAM architecture based on VCM-type TiN/TaO$_x$/Ti cells and a self-select scheme to prevent sneak currents was reported in [569]. HfO$_2$-based VCM devices were used in an 8-layer 3D convolutional neural network [128]. In this work, the unique 3D layout was inspired by the 3D CMOL architecture which possesses an excellent stackability [570,571].

In order to approach the physical limits to the scaling of VCM devices experimentally, tip-based scanning probe techniques were employed [572]. An approx. $3 \times 3$ nm$^2$ conducting spot has been detected as the exit of a conductive filament using a conductive-tip AFM (C-AFM) technique [573] after the top electrode has been removed with the scalpel AFM mode [574] on a Ru/Hf/HfO$_2$/TiN cell. An even smaller conducting spot at the exit of a dislocation on the surface of a SrTiO$_3$ single crystal has been studied by C-AFM and a significant conductivity enhancement confined to areas of 1–2 nm was found [119] (Figure 101). The undoped SrTiO$_3$ single crystal has been thermal reduced under a low oxygen partial pressure prior to the experiment. As we have discussed in Section 3.4, extended defects such as dislocations and grain boundaries in band insulators such as SrTiO$_3$ show an enhanced electron conductivity. The same is true for other band insulators [575]. In Ref. [119] it was shown that such exits of dislocations can be resistively switched by the conductive tip. A similar result has been obtained for epitaxially grown SrTiO$_3$ thin films with a SrRuO$_3$ bottom electrode [576]. In contrast to an earlier assumption, the oxygen ions do not move in the core of the dislocation but in the adjacent bulk crystal lattice [159,174,577]. Due to their electronic conductivity, the dislocation can be regarded as a heating rod which facilitates the oxygen vacancy motion in the immediate vicinity [578]. A corresponding change in the stoichiometry in the core of the dislocations has been detected by HTEM and EELS/EDX imaging [579,580].
Figure 102. First principles calculation of an extended linear defect in SrTiO$_3$ [119] introduced by the removal of OSrO units along a chain along the [1] direction. The figure shows a cut along the defect as a real-space image of the charge density of the Ti split-off state (along the defect cutting the two Ti atoms adjacent to the O vacancy defect). The colors indicate electron densities of up to 0.02 electrons (a.u.)$^{-3}$; the isolines are drawn with a logarithmic mesh. In the upper part, the symbols are shown to illustrate the site occupancies and the cut through the crystal lattice. Adapted and reprinted with permission from [119], © 2006 by Springer Nature Limited.

In order to simulate a non-stoichiometric defect by a density functional method, a row of OSrO was removed along the (001) direction in SrTiO$_3$. This leaves an additional electron per two Ti atoms. Although the electrons are mostly localized at the Ti atoms, there is some overlap of the electron density and the formation of an extended state (Figure 102). For such a situation, it is only necessary to remove or re-insert few oxygen ions in order to obtain a sufficient $R_{\text{HRS}}/R_{\text{LRS}}$ ratio.

However, because the electronic conductivity even in the best conducting oxides is significantly lower than for metals such as Ag, we need a more realistic approach if we aim at a more practical scaling limit of VCM cells in a memory array. Specifically, we need to find a conductive filament which supports the required read current of 1 mA in the LRS. If we take a homogeneous monolayer as an example, the oxygen deficiency (donor concentration) in the disc of the filament leads to a residual energy barrier in the LRS (Figure 60(b)). The dominant conduction mechanism will be Fowler–Nordheim tunneling through the barrier. We may start from a simplified cell sketched in Figure 59 and calculate the required donor concentration as a function of the disc diameter while keeping $I_{\text{rd}} = 1$ mA. This has been performed using TiO$_x$ as an example [527]. The results show that a high (but still reasonable) donor concentration of $3 \times 10^{21}$ cm$^{-3}$ leads to a minimum filament diameter of 4 nm. In order to obtain a $R_{\text{HRS}}/R_{\text{LRS}} = 10$, the donor concentration in the disc volume must be lowered by approx. 30 ions to arrive at the HRS.

Up to now, we have only considered the resistive states and the read process. The question is how the switching kinetics is affected by the down scaling of the cell. Figure 103 shows the results of a simulation using the model described in Section 5.7. The filament has been scaled with respect to its diameter and its length. Obviously, the nonlinearity of the switching kinetics does not suffer from the reduction of the minimum feature size.

In order to read the resistance state fast enough, the read current for the LRS should be at least 1 $\mu$A. In addition, the resistance ratio should be high enough to compensate for programming
variability, typically $R_{\text{HRS}}/R_{\text{LRS}} > 10$. Moreover, the resistance state (for non-volatile memory applications) should be stable for up to 10 years (Section 2.3.4). When a low number of defects defines for example the LRS it has to be investigated whether a suitable retention can be achieved. The loss of a few defects might lead to a retention loss. Zhirnov et al. investigated the ultimate scaling limits with respect to the aforementioned constraints theoretically [97,527]. The most relevant case for VCM cells is a scenario, which relies on the interface-switching model described above, and thus will be discussed in more detail in the following. For very small devices with perpendicular dimensions smaller than $L < 20 \text{ nm}$, the conclusions for filamentary switching and area-dependent switching devices should be similar. For these small perpendicular dimensions also the interface to the surrounding material becomes important. Depending on the work function mismatch of the switching material (e.g. TiOx) and the surrounding material (e.g. SiO2), an electron accumulation or an electron depletion layer will evolve. Most often, an electron depletion layer will occur which will be highly insulating and would reduce the actual switching region as illustrated in Figure 104(a). The latter scenario was assumed in the cited studies. Based on an analytical current model for Fowler–Nordheim tunneling including Schottky barrier lowering, the authors calculated the critical defect concentration that is required to achieve an LRS current of 2, and 1 $\mu$A as well as and HRS current of 0.1 $\mu$A in a TiO2 model device. As shown in Figure 104(b), the critical doping concentration (i.e. oxygen vacancy concentration) increases while down-scaling the perpendicular device dimensions. The oxygen vacancy concentration cannot be increased infinitely. The ultimate limit is reached when a phase transition occurs, which might be irreversible. In the present study, the authors assumed a concentration of $3 \cdot 10^{21} \text{ cm}^3$ as the limit, which would correspond to the oxygen content of a Magnéli phase with composition Ti10O19 (cf. Section 3.4). In this case, the scaling limit would be reached for a perpendicular dimension of $L = 4 \text{ nm}$. In the scaling limit the minimum cell size can be estimated as the product of the resulting depletion layer width in parallel direction for the critical defect concentration and the square of perpendicular dimension, which is approx.17 nm$^3$. It was furthermore estimated, how the retention scales with perpendicular scaling. Here, the important quantity is the loss of the LRS over time due to the redistribution of defects within the internal electric field. When the LRS read current drops below 1 $\mu$A, the state was considered as
lost. Figure 104(c) shows the calculated retention times as a function of cell size for 2 different LRS state defect concentrations. The retention time decreases for smaller cells in this scenario, because the initial LRS has a higher resistance and is thus closer to the 1 μA read current limit. For higher concentrations, the calculated retention times are longer but are still lower than 10 y. However, as mentioned above, the retention is linked to the migration enthalpy of the mobile defects (cf. Equation (33)). For an activation energy of approximately 1.3 eV a 10-year retention can be achieved according to the discussed study.

6. Reliability and failure mechanisms

For any application, reliability aspects of VCM devices are of major concern. For memory applications, for example, a long data retention and high endurance are required. Reliability issues appear in the programming operation (switching variability and endurance), the read operations (read-disturb and read noise), and the data storage (retention). To study these effects high statistics are required. As most of the studies on VCM cells are fabricated in research labs rather than in a foundry, studies with high amount of statistics are relatively rare. Nevertheless, failures observed on research devices can shed light on the potential failure mechanisms in the application. The device failures will depend strongly on the operation conditions. In the following sections, we discuss the different reliability categories, starting with the programming related aspects variability (6.1) and the endurance (6), followed by the reading based aspects read noise (0) and read-disturb (6.4), and concluding with data retention. The trade-offs between these aspects are discussed in Section 6.5.

6.1. Variability

The variability of the switching manifests in form of the variability of the programmed resistance states, the switching voltages and SET/RESET times. These quantities can vary from device to
device (D2D variability) and also in one device from cycle to cycle (C2C variability). The D2D variability can be related, in parts, to process variations during fabrication, e.g., different layer thicknesses slightly differing oxygen compositions, defect concentrations etc. Another aspect is the electroforming process which can induce additional D2D variations. Apart from these device specific aspects, also the organization in a memory array could lead to D2D variations during operation. In a 1T1R configuration, there might be variations in the transistors that lead to different programmed levels. In addition, the line resistances in an area might differ depending on the location of the device. The C2C variability originates in parts in the nature of the resistive switching, which is due to the random walk of ions in an irregular atomistic structure (i.e. structural crystal lattice defects) of the switching regions. Besides, this stochastic aspect there is also a deterministic aspect of the switching variability, which will be discussed in this section.

An example of the D2D variability of the LRS and HRS resistance of filamentary VCM Ru/Ta2O5/TiO2/Ru cells is shown in Figure 105 in which the variability in terms of the resistance state is typically characterized by using Weibull plots as illustrated [581]. The memory window can be now defined in a statistical way based on the highest HRS and the lowest LRS. This shows the significance of a statistical analysis of the results. Even if the mean value can be clearly separated the tail bits of the HRS and LRS distribution may overlap, which leads to a read failure as will be discussed in Section 6.4. The example further shows a rather tight distribution of the LRS values, but a higher variability in the HRS. Often it is reported that the LRS follows a normal distribution whereas the HRS shows a log-normal distribution [582]. The different types of distribution may be related to the transition of the conduction mechanism. Alternatively, the smaller number of defects in the HRS may lead to a stronger dependence of the device resistance on the change of the number of single defects and thus a higher variability. Still, additional research is required to clarify this aspect.

Figure 107(a) shows the $I-V$ characteristics of several consecutive cycles of a Pt/HfO2/TiOx/Pt device [555]. The scatter in the SET and RESET voltages as well as the LRS and HRS can be clearly observed. This C2C variability appears also in the SET switching times during a DC voltage pulse as illustrated in Figure 107(c). For a given voltage, the switching time differs up to 4 orders of magnitude in time. The switching variability leads to a pronounced variability in the switching times as shown for Pt/HfO2/TiOx/Pt devices [555]. This means that the switching can be considered probabilistic with respect to time and voltage. Thus for a certain voltage and constant pulse width, a switching probability can be defined. For TiO2-based devices, the switching probability followed a log-normal distribution [583]. Similar distributions have been reported for Pt/ZrO2/Ta devices [584].

As discussed before, the origin of the switching variability can be divided into a deterministic component and a stochastic one. The stochastic component is related to the random walk of the oxygen defects during switching. This means that the ionic defects will move on different paths

---

Figure 105. Example of a Weibull plot of the distribution of $R_{\text{LRS}}$ and $R_{\text{HRS}}$ values of an 1 kbit 1T1R array of Ru/Ta2O5/TiO2/Ru cells. Adapted and reprinted with permission from [581], © 2010 by IEEE.
within the electric field and thus, the HRS and LRS states will be based on different ionic defect configurations from cycle to cycle. Kinetic Monte Carlo (KMC) models (cf. Section 5.4.2), which are in fact based on such stochastic processes, are the ideal models to study the stochasticity of the switching. The KMC models, however, predict a smaller spread in the SET and RESET voltages or the switching times as in the experiment [377,585]. Thus, the stochasticity of the process cannot explain the observed variability alone.

The deterministic component is based on the state-dependence of the switching times/voltages in particular for filamentary VCM devices. On the one hand, the SET time/SET voltage depends on the initial HRS (pristine state). The higher the HRS, the slower is the switching speed or the higher is the required SET voltage [519,554,555]. On the other hand, the RESET time/RESET voltage depends strongly on the programmed LRS. If the device is programmed to a very low LRS, the RESET switching may be delayed by orders of magnitude in time [554]. This effect is mainly due to an internal voltage divider [316] as discussed in Section 6.7. It should be noted that this strong variation of the switching delay due to different initial states is strongly linked to the highly nonlinear switching kinetics and the Joule heating effect in filamentary VCM cells in particular. Such a strong effect is not expected for area-dependent switching cells. The state before SET and RESET is of course a result of the history of the switching and thus depends also on the stochasticity of the process (cf. [586]). In fact, small changes in the defect configuration during the previous SET/RESET, which could be based on the stochasticity of the random jumps can lead to a large variability in the following cycle.

In realistic devices, the situation is often much more complicated than a simple random walk in the vicinity of a regular, cylindrical or conical-shaped filament. As described in detail in Chapter 5, the filament structure can be very complex and might vary its shape and size during repeated operation [77,587]. For SrTiO3 devices, it has been observed by operando spectro-microscopy that some devices exhibit variations in the shape of the conductive filament or in the oxygen vacancy distribution at and around the filament (see Figure 106). In other cases, even the location of the active filament changes from one cycle to the next. We propose that both effects originate from the coexistence of multiple (sub-)filaments shown in Figure 49 in Chapter 5 and that the active, current-carrying filament may change from cycle to cycle.

The change of the filament geometry mentioned above was used as motivation to extend the JART VCM v1b model to include switching variability [555]. To this end, the parameters describing the filament geometry and properties, such as radius, length or minimum and maximum defect concentration, are considered to be variable. To model the D2D variability, these parameters are drawn from a truncated Gaussian distribution in physically reasonable range and are kept constant for each device during simulations. In contrast, the parameters are updated from C2C using a random walk algorithm to model the C2C variability. Within one SET/RESET process the parameters are constant. Figure 107 shows the comparison of the variability-aware JART VCM v1b model with experimental data of Pt/HfO2/TiOx/Ti nano-crossbar devices. The model shows a good agreement for different experiments, i.e. I–V sweeps, endurance, SET kinetics and RESET kinetics. In another study, it was shown that the distribution of the C2C variability is very similar for different devices, but the mean values are shifted from device to device [588]. This D2D shift may be reduced by better control of the fabrication process and the forming/SET current compliance. The C2C variability, however, will still be existent.

The previous analysis shows that an important aspect to explain the variability of filamentary VCM is the change of the filamentary structure and/or the competition between different filaments in the oxide layer. For area-dependent switching VCM cells, this aspect is not present. Besides the absence of Joule heating as discussed before, this aspect may explain the reduced variability of area-dependent VCM cells. In total, this leads to a lower variability of area-dependent VCM cells, as outlined in Chapter 6.
Figure 106. PEEM image of a SrTiO$_3$ device in the LRS (a) and of the same device after RESET (b). (c) and (d) PEEM images after one additional SET and RESET operation, respectively. Insets: magnified map of the area around the suspected conductive filaments. Reprinted with permission from [86], ©2017 by the American Chemical Society.

6.2. Endurance

The **endurance** denotes how many write cycles with alternating SET and RESET process can be performed until the LRS or HRS falls out of the predefined acceptance window. Obviously, the endurance is determined by the degree of reversibility of the microscopic switching process. Redox processes never show a complete reversibility due to the statistical variation of the ion motion. The electroforming processes executed prior to the switching include significant irreversible contributions, by changing the morphology of the cell due to the formation of the filament, the evolution of molecular oxygen during the forming of VCM into the HRS, or the consumption of electrode metal by oxidation during forming of the VCM into the LRS. Although the write voltage magnitudes are distinctly smaller than the electroforming voltages, there is a certain probability that these irreversible processes continue to a certain (typically very small) extent during the regular SET or RESET processes. Over many alternating write cycles, these irreversible contributions may accumulate and lead to cells which may get stuck in the LRS or HRS. In addition, the SET and RESET processes themselves may contain irreversible contributions, e. g. due to ions which are placed back on slightly different positions, reactions at the interface of the ion conductor and the electrodes or the sidewalls, gradual growth or dissolution of phases [410], etc. It has for example been shown by TEM analysis that TiN/TaO$_x$/TiN devices that failed by stuck-in-HRS mode exhibited broadening of the filament, Ta depletion and oxygen interdiffusion at interfaces with both electrodes. [83] In devices that failed by stuck-in-LRS mode, a strong Ta-enriched sub-filament bridging the gap at the electrode has been observed. All devices at the end of endurance exhibited interdiffusion of O into TiN and Ti and N into TaO$_x$ [83].

In order to optimize the endurance of ReRAM, it is the aim to select materials and to design and fabricate cells which show a reversibility, e. g. a self-reproduction of the arrangement of atoms in the cell to the highest possible extent. An example of a VCM system is given in Figure
Figure 107. Evaluation of the variability: Measured device characteristics (left column) of a Pt/HfO$_2$/TiO$_x$/Ti cell alongside results obtained using the extended JART VCM v1b compact model (right column). Measured (a) and simulated (b) voltage sourced $I$–$V$ sweeps. Measured (c) and simulated (d) SET kinetics. Measured (e) and simulated (f) RESET kinetics. The blue circles represent the experimental RESET kinetics starting from LRS between 1.85 and 2.22 kΩ while the green points represent the RESET kinetics starting from LRS between 1.52 and 1.67 kΩ. The solid lines represent the RESET kinetics of the deterministic model and the blue box plots show the RESET kinetics for the variability model for an LRS range between 1.92 and 2.04 kΩ while the green box plots show the RESET kinetics of the variability model for an LRS range between 1.58 and 1.67 kΩ. Measured (g) and simulated (h) endurance behavior over 1000 cycles. Adapted and reprinted with permission from [555], © 2020 by IEEE.
Figure 108. Endurance behavior of a Pt/TaOₓ/Ta₂O₅/Pt system. Adapted and reprinted with permission from [590], © 2011 by IEEE, Japan Society of Applied Physics.

108, which shows an endurance of $10^{12}$ cycles. A detailed overview about the endurance obtained for different material combinations so far is given in reference [589].

The SET and RESET transitions are based on the movement of ionic defects (either within the oxide and/or across the interfaces). From this point of view, stable LRS and HRS states can only be achieved if the same number of defects is moved back and forth and similar configurations are obtained. As the switching time/voltage depends on the initial state a small misbalance can lead to endurance degradation over many cycles. To achieve a high endurance, it is thus important to tune the SET and RESET voltages and pulse lengths. Figure 109 shows the endurance behavior of Hf/HfO₂ 1T1R cell for different SET conditions and constant RESET conditions [591]. For the lowest SET voltage, the resistance drifts to higher values over cycling and eventually the device got stuck in the HRS. For a SET voltage of 1 V, SET and RESET are in balance for about 1M cycles. If the voltage is increased even more, the states drift toward the LRS until they the device gets stuck in the LRS. It should be noted the device is in this case typically not destroyed using longer pulse length or an increased SET/RESET voltage can ‘unstuck’ the device.

In an integrated memory array, however, the maximum voltage/pulse length that can be applied is limited. Thus, a so-called over-SET/over-RESET could lead to a permanent stuck-at-LRS/stuck-at-HRS failure. To prevent an over-SET/over-RESET and to program different states reliably, write-verify algorithms have been developed [592–595]. This includes simply applying additional pulses, increasing the pulse length, or increasing the SET/RESET voltage when the programming was not successful. Instead of changing the programming voltage, also schemes adapting the gate voltage in a 1T1R configuration have been employed. The use of additional write pulses will increase the energy consumption of the write process. This poses a trade-off between programming accuracy and write energy. Figure 110(a) shows the distribution of 8 programmed resistance states on a 16 k 1T1R VCM chip using the so-called RADAR algorithm, which reduces the programming pulse count by $2.4 \times$ [595]. The programmed states show some relaxation after baking the chip for 30 min at 130°C, which increased the bit error rate (cf. Figure 110(b)). The relaxation effect will be discussed in more detail in Section 6.4. To optimize the energy efficiency of the programming self-termination circuits have been developed to end the programming directly when the targeted resistance is reached [596,597].

Apart from the attempts to improve the programming endurance by using optimized write voltage schemes, different material engineering approaches have been demonstrated. For example, a $50 \times$ improvement of the switching endurance in TaOₓ-based VCM cells has been shown using Zr ion implantation [313]. It is suggested that due to a reduced formation of oxygen vacancies close to the Zr dopants and an increased diffusion barrier of oxygen close to the Zr dopant,
the stability of the filament is enhanced, i.e. there is less loss of oxygen defects due to the surrounding over time. Theoretical DFT studies showed that p-type dopants in Ta2O5, HfO2, ZrO2 and TiO2 can lower the formation energy of charged oxygen vacancies [598]. In the latter study, it is also proposed that this lowered formation energy leads to reduced formation/SET voltages and longer retention times. In another study on HfO2-based VCM cells, it is suggested that the dopants influence the oxygen scavenging effect at the HfO2/Hf interface and influence, thus, the endurance and switching properties. [599]

Apart from doping, the choice of the ohmic electrode material can change the endurance properties. For example, the endurance of (Ba, Sr)TiO3-based VCM cells could be improved using W electrodes [600]. Using DFT calculations, the oxygen vacancy formation energy at the oxide/metal interface for different OE metals was determined [601]. For HfO2 the defect formation energy is close to zero for Hf electrodes but is increasing for metals with a higher O chemical potential. For Ta2O5, the defect formation energy is negative for Hf, Zr or Ti and then becomes positive for materials with higher O chemical potential. It has been shown that using Hf or Ti OEs with a Ta2O5 VCM cell having a negative formation energy lead to an early RESET failure resulting in a low endurance [361]. For W and Ta electrodes, however, a stable switching behavior was observed. Wiefels et al. investigated the impact of different OE (Hf, Zr, Ti, and Ta) on the endurance properties of ZrO2-based VCM cells [602]. To this end, a writing scheme was developed to reach the endurance limit for each cell. Figure 111(a) shows the statistics of the maximum
Figure 110. Conductance distribution of three bits-per-cell based on 2048 VCM cells per conductance level co-integrated with 130 nm CMOS technology (a) before baking and (b) after a 30-min bake at 130°C. The bit error rate after baking is 0.6% (before ECC). Dashed lines indicate read range boundaries. Adapted and reprinted with permission from [595], © 2021 by IEEE.

endurance for the different stacks. The lowest endurance was obtained for Hf, and Zr electrodes, whereas the endurance was clearly improved when using Ti or Ta electrodes. The mean value of the endurance follows a clear trend with the formation energies of oxygen vacancy defects at the HfO2/M interface (cf. Figure 111(b)). This result shows that an oxygen exchange will likely occur also at the OE interface during switching. This exchange has an important influence of the reliability of the VCM cells. It should be noted, however, that the biggest resistance contrast is still obtained due to a depletion/accumulation of oxygen vacancies close to the AE/oxide interface.

6.3. Read disturb

Read disturb describes the change of the resistance during read operation over many read cycles. For a binary memory, the read disturb in the HRS/LRS occurs mainly when reading with a SET/RESET polarity. Nevertheless, a drift to higher/lower resistive values of the HRS/LRS may lead to permanently stuck resistance states and an endurance failure. In principle, the read-disturb can be considered as an extrapolation of the SET/RESET kinetics to lower voltages as illustrated in Figure 112(a). In order to be able to read the device for a long time without disturbance at low read voltages $V_{rd}$ and switch the device at moderate write voltages $V_{wr}$ in nanoseconds a very high nonlinearity of the switching kinetics is required. This is the so-called voltage time dilemma [11,510] discussed in detail in the switching kinetics section in Chapter 5. As a figure of merit
one can define the nonlinearity $NL$ according to

$$NL = \frac{\log(t_{RD}(V_{rd}))}{\log(t_{SW}(V_{wr}))} = \exp\left(\frac{\Delta W - azE_{rd}/2}{k_BT_{rd}} \cdot \frac{k_BT_{wr}}{\Delta W - azE_{wr}/2}\right).$$  \hspace{1cm} (62)

In Equation (62), $t_{RD}$ is the read-disturb time, $t_{SW}$ is the switching time, $E_{rd}$ ($E_{wr}$) is the working electric field during read (write) and $T_{rd}$ ($T_{wr}$) is the device temperature during read (write). The right-hand side assumes a limitation of the switching speed by ion migration (cf. Section 5.7.2 and ref. [526]). According to Equation (62) a high $NL$ is obtained for high temperature difference between read and write or a high difference in the applied electric field. As discussed in Section 5.7.2, a higher $NL$ is typically obtained for filamentary switching VCM cells as Joule heating plays a crucial role. For area-dependent systems, however, a high $NL$ can only be achieved for a big difference in the electric fields. This suggest that higher switching voltages are required for area-dependent cells to obtain a similar $NL$ as for filamentary VCM cells. An additional factor influencing the $NL$ is the variability in the switching times. As shown in Figure 112(a), the fastest switching events at a specific voltage can be 4–5 orders of magnitude faster than the slowest switching events. For a successful write operation, the slowest event needs to be considered. For the read-disturb, in contrast, the fastest event will already create a failure. The number of tolerable failures is defined by the maximum tolerable bit error rate in an array. This means that the additional variability lowers the effective $NL$ calculated using Equation (62) based on the mean device behavior. As the variability is lower for area-dependent cells, this effect reduces $NL$ less for area-dependent cells than for filamentary switching VCM cells. The above approximation assumes that the switching is quite abrupt, meaning up to a certain time the resistance state only changes little, but then the resistance flips. This is consistent with the observed abrupt SET transition in filamentary VCM cells, but it is not fulfilled for gradual switching systems. In gradual switching systems, the resistance may change gradually upon successive reads. This could lead to higher bit error rates due to read-disturbs in particular when storing more than 1 bit in a cell.

As the switching times at typical read voltages of 0.1–0.3 V are very long it is not easy to access this regime experimentally to study these effects. Thus, typically higher voltages are used and then the read-disturb time is extrapolated to lower voltages. Figure 112(b) shows the change of the LRS current for WO$_x$-based VCM cell for a train of read pulses with varying amplitude in RESET [603]. The current degradation can be separated into two regimes. In the first regime, there is no apparent current change. In the second regime, in contrast, the sense current follows an
Figure 112. Illustration of the read-disturb: (a) SET kinetics according to Figure 107. The upper and lower bound (dashed line) as well as the mean value (solid line) of the switching times are extrapolated to the read regime (gray shaded area). The data was extracted from Figure 107d. Adapted and reprinted with permission from [555], © 2020 by IEEE. (b) Sense current evolutions with cumulative read-disturb time at different \( V_{\text{read}} \) values for TiN/WO\(_x\)/W cells. The sense voltage is \(+0.1\) V. Pulse trains used in characterization are depicted in the inset; Adapted and reprinted with permission from [603], © 2018 by IEEE.

For a bipolar switching cell, one would intuitively only expect a possible disturb-caused switching from the cell HRS (RESET) state to the cell LRS (SET) state for voltage stress having the same polarity as the polarity required for SET programming, and, similarly, an LRS to HRS disturb switching only for voltages with the RESET polarity. For HfO\(_2\)-based VCM cells, however, an abnormal read-disturb has been reported [327]. In this case, an increase in the read current has been observed when applying low voltage pulses with RESET polarity. This effect has been related to the RESET balance due to counteracting ion migration and ion diffusion (cf. [366]). Due to the strong RESET, the device is at the specific read voltage out of ‘balance’ and moves towards this equilibrium at higher current levels over time. As so, this effect should rather be interpreted as a resistance retention issue with the cell resistance moving to an equilibrium state, and the voltage pulses giving the system the required energy to evolve to that state.

6.4. Read noise and programming instability

Read noise or random telegraph noise (RTN) describes the instability of the current level over time during read. This is illustrated in Figure 113 for a filamentary ZrO\(_x\)-based VCM cell. The presence of RTN can cause a large spread in the distribution of the high and low resistance state (HRS, LRS) and induce ‘soft errors’ in reading the wrong memory state (if the memory window is relatively small). Although RTN, which is relevant mainly at read voltage (\( V_{\text{READ}} \sim 0.1\) V) conditions, does not cause irreversible damage unlike endurance test conditions, it affects the variability of the resistance distribution [378,582,595,604–606]. There are three major physical causes of RTN. One cause are steady-state fluctuations involving stochastic electron capture and emission events due to a trap-assisted electron transport with their corresponding time constants (which depend on applied voltage, trap position and trap energy) [607–610].
addition, negatively charged defects close to the conducting filament could lead to a depletion of charge carriers in the filament \cite{604,611}. The third mechanism is based on configurational changes of the filament due to the random jumps of oxygen vacancies due to the finite temperature \cite{582,609,612,613}.

The strength of the RTN, i.e. $\Delta I/I$, based on the latter mechanism, was shown to increase with higher HRS values as the conduction is governed by a lower number of defects \cite{611,614,615}. For LRS values in the regime of typical high compliance levels of 100 $\mu$A – mA, in contrast, a very high number of traps/defects is present. Thus, the RTN is less pronounced in the LRS than in the HRS for typical LRS/HRS current levels (cf. Figure 113). The RTN in the LRS state is generally not a critical issue, unless consider ultra-low power switching devices with forming / SET compliance levels as low as 1 $\mu$A (which currently show very low endurance and retention). On area-dependent VCM cells, studies on RTN are rare. The previous discussion on the number of defects in the switching of area-dependent systems (at least if the cells are large) suggests that RTN is not likely to occur for AD cells. Even if there are statistical fluctuations, the high number of defects will lead to multiple parallel events averaging out the effect of single defects.

As discussed before the read instability can lead to read failures due to a relaxation of the programmed states over time. The programmed distributions appear to be quite stable as shown in Figure 114(a) for the programming of a HfO$_x$-based VCM cell \cite{616}. After an adaptive programming scheme, the tails of the LRS (HRS) are successfully sharpened. After 1 s of relaxation time, however, the states relax and overlap. The final distribution is very stable at least at a timescale where retention effects do not appear (e.g. $>10^4$ s). This effect was also observed on larger arrays \cite{617,618}.

To explain the physical origin of the stable distribution and the microscopic origin of the RTN, Wiefels et al. employed static simulations using a TAT model \cite{582}. As shown in Figure 115(a) a filamentary plug region with a high amount of oxygen vacancies and depleted gap region close to the active electrode is considered to resemble a HRS. By displacing every defect by one
Figure 114. Example of programming instabilities: (a) Distribution of programmed resistances produced by SET and (b) RESET adaptive algorithms. (a) After programming the LRS is restricted to values below 20 kΩ, but after 1 s, another read shows that the states relax and the LRS distribution overlaps with the HRS one (red solid line). The same effect is observed for the HRS in (b). After programming the HRS resistance is higher than 200 kΩ, but after 1 s the states relax to lower values (dashed red line) overlapping with the LRS distribution. Adapted and reprinted with permission from [616], © 2015 by IEEE. (Color figure available online).

lattice site, the influence of this single displacement on the measured current is investigated. The displacement can be split into four different types, perpendicular jumps within the gap region or the filament plug (1), parallel jumps in the filament (2), parallel jumps in the gap region (3) and parallel jumps from the filament plug into the gap region (4). As shown in Figure 115(b), type (4) shows the largest current jump. This can be easily understood in so far as the jump into the gap region reduces the tunneling distance from the electrode to the trap. This type of jump was implemented into the JART VCM v1 model using a state machine that changes the concentration close to the active electrode by ±2, ±1, and 0 (divided by the volume of the gap/disc region) using a state machine [582]. Figure 115(c) shows the current traces obtained with this model. The resulting current jumps agree well with the jumps observed in the experiment. Moreover, the model can reproduce the relaxation after adaptive programming and the observed stable HRS distribution (cf. Figure 115(d)). Based on this model it is concluded that the observed programming instability can be well explained by stochastic jumps of the ionic defects at finite temperatures. The timescale of this effect is in the regime of 1 ms to 1 s, which corresponds with the jump frequency of the ions, and thus, the migration enthalpy. In fact, the required migration enthalpy is too low to explain the observed good retention in these cells. One possible explanation could be that only a few defects are mobile and cause the read instability. The majority of the defects, however, are less mobile and do not jump on the timescale of the read operation. In this case, however, the mobile defects should move away from the filament over time and the RTN should reduce over time, which is not the case. An approach resolving this issue is discussed below in Section 6.5.

6.5. Retention

Retention is the ability of a memory cell to keep a stored information if the cell is not addressed. As ReRAM targets non-volatile memory applications, the retention property of RRAM is a critical reliability aspect. Typical non-volatile memory requires 5–10 years of data retention up to 85–125°C. For any discussion of the stability of states and their long-term reliability, one should
Figure 115. Microscopic origin of the RTN. (a) Exemplary distribution of $V^*_{O}$ (blue) in filamentary plug volume (blue box) and gap/disc volume above. The four possible jump types are illustrated by the colored arrows. (b) Simulated absolute current changes caused by 1000 jumps for different initial configurations of each jump type at a mean read current of approximately 10.5 μA. The black dashed lines represent the jump heights calculated by an FHMM method where the simulation has been fit to. (c) Simulated current traces measured at 0.35 V. The signals all show the characteristic current jumps between distinct levels around a constant median current. Compared with (b), only large jumps of roughly 1 μA are present. (d) Comparison of the experimental and simulated read current distributions. The gray shades show the experimental read current distributions at specific times, while the red, blue, and green curves show the simulated read current distributions. (d, left) and (d, right) show three different simulations with 540 cells. (d, left) Experimental read current distributions at 1, 1.25 and 1.5 s. As previously shown, the distribution stays constant on the considered time scale. While the initial resistance distribution is slightly smaller than the experimental ones, after 1.25 s, the simulated and experimental distributions only show a negligible difference. In (d, right), the simulated cells were shaped after 500 ms and only those between 6 and 12 μA were further considered. The lognormal shape is recovered within 1 s. Adapted and reprinted with permission from [582], © 2020 by IEEE. (Color figure available online).

Keep in mind that at most one of the states, LRS or HRS or any intermediate state, can be thermodynamically stable. The reason for only one state being the thermodynamically stable state is due to the fact that there may be only one arrangement of ions and atoms which has the lowest free energy. This is different to ferroelectric and ferromagnetic systems in which states with opposite polarization direction may both be thermodynamically stable. It could even be that both states may be metastable, e.g. characterized by a frozen-in enrichment or a frozen-in depletion of oxygen vacancies in front of an electrode interface (see switching mechanism described in Chapter 6). They are frozen-in after a kinetically fast (i.e. temperature- and/or field-accelerated) switching event. At any finite temperature (here: 85°C as a typical max. operation temperature), there
is a finite diffusion coefficient of the ions involved and there is a finite reaction rate of possible interface reactions.

For this reason, one strives for materials in which the active components show high activation energies of the hopping processes (Section 5.7.2) so that these processes are frozen-in at the operation temperature of the ReRAM device. In order to predict the retention time, one exploits the fact that diffusion (and reaction) processes are thermally activated. In an accelerated life test, the retention time is measured for (at least two) elevated temperatures and the retention time at operation temperatures is extrapolated in an Arrhenius plot \[619\] (e.g. Figure 116 \[250\]). More sophisticated retention measurement methods in order to consider to the convolution of variability and retention have been suggested in the literature \[620\].

A related category is the data retention during read or read disturb discussed in Section 6.4, defined by a retention time of a cell exposed to a constant stream of \(V_{rd}\) pulses. Microscopically, one has to consider an additional drift under a (relatively) small electric field in addition to the random hops of the ions which represent the diffusion process. Therefore, the question of retention is strongly interlinked to the switching kinetics and the so-called voltage-time dilemma discussed in great detail in Section 6.8. Since the switching kinetics of area-dependent switching cells have a much less pronounced non-linearity than filamentary cells, one can assume that they have comparable poor data retention. This has been confirmed for a few material systems such as Nb doped SrTiO\(_3\) \[254,488\] and PCMO/YSZ cells \[66,439\]. However, systematic studies of the retention failure of area dependent cells are generally missing so far.

In the following, we will therefore concentrate on the progress which has been made with respect to improving the retention time of filamentary VCM cells and with understanding the different types of underlying retention failure mechanisms.

Figure 117 shows the oxygen vacancy map of a filament within a \(\text{Ta}_2\text{O}_5/\text{TaN}\) cell determined by EELS analysis \[250\] presented in Chapter 4. As mentioned above, this oxygen vacancy distribution has been frozen in during electrical biasing and oxygen vacancies might redistribute due to a certain extent afterwards driven by chemical gradients, space charges or by the small electrical field applied during read-out. The possible directions of oxygen movement are sketched by blues arrows in Figure 117. Whereas the drift of oxygen during read-out takes place only in the direction of the field (parallel direction according to the sketch in Figure 117, chemical gradients might be present in all directions. Chemical gradients might induce the diffusion of oxygen in parallel direction, namely, either along the filament or into/out-of the electrode interface or into/out-of the electrode itself. Furthermore, the thin film matrix acts as oxygen reservoir and might induce
oxygen diffusion into the filament in the perpendicular direction. The LRS resistance is often reported as retention failure [250,619,621–623]. This assumption has been confirmed by simulations of the $I$–$V$ characteristics and retention times based on the oxygen diffusion from the side [81,624] or along the parallel direction [625]. Monte Carlo calculations have been employed to predict LRS retention by the oxygen movement in both parallel and perpendicular directions [622].

In [619] it was observed that the LRS resistance HfO$_2$ based cells increases at elevated temperatures. Since the back-diffusion of oxygen into the filament is a thermally activated process, the LRS retention degradation on a large statistical basis follows the classical Arrhenius law dependence with temperature as shown in Figure 118 [619]. It is important to note that the LRS retention becomes increasingly severe with decreasing filament size [250], since a slight reoxidation from the side has increasing impact on the filament conductivity and might even completely resolve the filament over time. Since lower current compliances result in smaller filaments, LRS retention failure becomes also more pronounced as demonstrated in Figure 118 for HfO$_2$ based cells. It is important to note that for filaments in the 100 nm regime and above, the lateral reoxidation plays a minor role [270].

For Ta$_2$O$_5$/TaO$_x$ cells, it has been shown that oxygen diffusion from the cell side walls into the filament is even more severe than the diffusion from the thin film matrix [251]. As shown in Figure 119(a), this process results in a significant number of tail bits which can be assigned to cells where the filament has formed in the vicinity of the side walls, Figure 119(b,c). By employing a sidewall passivation process, the authors could prevent the formation of off-center filaments and thereby solve the LRS retention failure problem as shown in Figure 119(d).

Indeed, HRS retention failure with both, HRS current increase and decrease, have been reported in the literature [250,619,621,623].

In the case of oxygen-transparent electrodes and in particular for electrodes with a considerable solubility for oxygen such as Ta, Hf, Ti the situation becomes more complex since oxygen
Figure 118. Arrhenius plot of LRS retention failure in TiN/HfO₂/Hf/TiN cells for different current compliances. A statistical average over 60 cells is shown for each data point; Adapted and reprinted with permission from [619], © 2012 by IEEE.

Figure 119. (a) Data retention at 85°C of Ir/Ta₂O₅/TaOₓ/TaN cells: Tail bit degradation was observed; (b) EBAC image: A filament is located near the edge; (c) Retention degradation model: Oxygen diffusion into the filament from the edge; (d) Data retention at 85°C of oxidized Ir/Ta₂O₅/TaOₓ/TaN cells: Tail bit degradation was strongly improved by filament control; (e) EBAC image of an oxidized cell: A filament is located in the middle; (f) TEM analysis of cell with and without cell side oxidation. Adapted and reprinted with permission from [251], © 2015 by IEEE.
diffusion into and out-of the electrode as well as the interaction with the oxygen ambient has to be taken into account. For example, for crystalline SrTiO$_3$ – noble metal thin film cells, the LRS retention failure is caused by the reoxidation from the oxygen ambient. [254,270]

As discussed in detail in Chapter 4, an oxide layer is formed at the interface to all oxidizable metal electrodes, which acts as sink or source for oxygen during electroforming and switching. Depending on the employed metal/oxide combination and their thickness relation, the chemical gradients induced during biasing might differ significantly and have a strong impact on the driving forces for retention failure of both HRS and LRS similar to the effects reported to affect the switching kinetics (see Chapter 5) and the endurance (see Section 6.3). Studies on HfO$_{2-x}$ [623] with different metal electrodes show that for Hf and Ti, short retention times are observed owning to its strong thermodynamic ability of extracting oxygen from HfO$_2$, whereas better retention can be achieved for Ta owning to the lower thermodynamic ability of extracting oxygen from HfO$_2$.

Moreover, the data retention in TiN/Ta$_2$O$_5$/Ta cells crucially depends on the Ta thickness according to the different oxygen vacancy distribution at the Ta/Ta$_2$O$_5$ interface. While for the 10 nm thick Ta electrode, the LRS shows a significant resistance drift towards the HRS, the LRS is stable for the 30 nm thick Ta electrode and the HRS drifts towards the LRS over time [626]. This results in a thermodynamically stable LRS for the 30 nm thick Ta electrode and a stable HRS for the 10 nm thick Ta electrode.

Apart from the consideration about the thermodynamic stability of different states, an alternative approach is to increase the constraints for reoxidation to the system by kinetically hindering the oxygen back-diffusion. Noman et al. showed that in the absence of internal electric fields, SrTiO$_3$ cannot exhibit fast switching and long retention times simultaneously [381]. In fact, retention failure after short time was reported for the LRS in homogeneous single crystalline SrTiO$_3$ [627]. Polycrystalline and single crystalline SrTiO$_3$ films with considerable amounts of extended defects, on the other hand, exhibit much better retention behavior [627,628] possibly induced by local variations of the diffusion constants. Furthermore, it has been observed that a phase separation of SrTiO$_3$ into Sr-poor SrTiO$_3$ and SrO (Figure 120(b)) results in a stabilization of the LRS (Figure 120(b)) since oxygen diffusion is strongly hindered in the SrO layer [270]. As a result, a reoxidation of the oxygen deficient SrTiO$_3$ filament is prevented.

A comparable mechanism might take place in HfO$_2$/Hf cells, where the interface reaction results in the formation of an oxygen deficient HfO$_{2-x}$ layer and a Hf electrode with a certain amount of dissolved oxygen. Since ab-initio calculation has shown that the oxygen vacancy mobility is strongly enhanced in oxygen-deficient HfO$_{2-x}$, this layer might act as sublayer where fast diffusion takes place [629]. On the other hand, the Hf layer may act as oxygen reservoir and as oxygen diffusion blocking layer [629] which prevents the reoxidation of the LRS. Besides these intrinsically formed bilayer systems, many groups intentionally grow bilayer systems containing one oxygen blocking layer such as Al$_2$O$_3$ in order to improve the device stability against retention failure [270,630–632]. Systematic studies of the use of interface layers with different oxygen diffusivity confirmed that interface layers with high oxygen diffusivity such as Y-stabilized ZrO$_2$ result in bad retention whereas interface layers with low oxygen diffusivity such as SrO and Al$_2$O$_3$ result in a stable retention[270,633]. This knowledge about the relationship between oxygen diffusion and retention time at the interface can be used to develop rational design rules for stable retention.

In order to evaluate the retention failures, it is important to study large statistics. For 2.5 Mbit HfO$_2$-based VCM cells, it was shown that the mean HRS value moves to higher resistances [634]. Based on the mean value, a retention failure should not be of concern as the read window increases. For these large statistics, however, it was shown that the distribution tilts over time and very few bits will show a decrease of the resistance leading to a retention failure. The question is how these two effects, increase of the mean value and tilting of the distribution can be explained.
Advances in Physics

Figure 120. (a) Time dependence of the LRS and HRS for different SrTiO$_3$ thin film devices showing stable retention or LRS retention failure. (b) Spectroscopic investigation of devices with stable retention and retention failure: the right figures show the comparison of the OK-edge which hints on the formation of SrO in the filament region of the device with stable retention. The PEEM false color on the left depicts the spatial distribution of SrO on the device area. Adapted and reprinted with permission from [270], © 2015 by Springer Nature Limited.

in one model. A KMC model has been employed to study this effect [635]. Depending on the predominant direction of the defect diffusion two different scenarios evolve. If the ions move predominantly in perpendicular direction a drift of the distribution to higher resistance values results. In contrast, if also a significant ion diffusion in parallel direction occurs, the distribution tilts. Hence, the observed tilting and the shift to higher resistance values can be both explained by ion diffusion.

In the above model [635], a multi-domain model with different migration barriers was employed as shown in Figure 121. The filament is divided into a number of small boxes [635]. Within the box, the migration barriers are low, whereas the migration barrier from one box to the other (or to the surrounding) is increased. The retention phenomenon described above is dominated by the jumps out of the boxes and occurs at longer timescales. In contrast, the lower migration barriers for the jumps within the boxes occur on the timescale relevant for read-disturb. The simulated read transients for different initial states show the distinct current jumps as observed in the experiment (cf. Figure 121). Thus, introducing boxes with high jump frequencies and lower jump rates out of the boxes allows us to explain the read-instability and retention using the same mechanism, i.e. jumps of ionic defects. In principle, the read instability can be considered as undirected random jumps of the defects, whereas the retention is driven by the concentration gradient, i.e. ion diffusion. The multi-domain model was motivated by a theoretical work on diffusion in amorphous HfO$_x$ by Schie et al [636]. Using MD simulations, a sub-diffusive behavior was observed for amorphous HfO$_x$. The simulation results were explained by different migration barriers due to the different local configurations. Using KMC simulations it was shown that these different barriers cannot be merely randomly distributed, but grain-like structures are required. Within the ‘grains’ the migration barrier is low, but a higher barrier needs to be overcome if the ion jumps from ‘grain’ to ‘grain’. This is a similar scenario as discussed for the brick-wall model discussed in Chapter 3 on a somewhat larger scale (cf. Figure 20).

It should be noted that the issue of retention failures can only be answered on a large statistical basis. For the application as a memory device, the failures will appear in the very low percentiles
Figure 121. Multi-domain model to explain the read-instability and the retention failure. Left panel: Sketch of the cubic boxes limiting the $V_{\text{O}}$ (red balls) diffusion in the oxide layer of the VCM cell. Exemplary shown is a typical defect distribution in the HRS, where the boxes of the filament plug and gap regions are highlighted in blue and divided by the dark blue interface. The ohmic bottom electrode lays at $z = 0$ in the xy plane. For $(1\, \text{nm})^3$ boxes, the active electrode can be found on top of the blue highlighted boxes. (a) Current evolution over time of exemplary experimental (gray) and simulated (colored) VCM cells in the HRS under applied read voltage. In the zoom-in of the red curve (S1) in panel (b), the highest current jumps (blue arrows) can be identified as parallel (vertical) jumps of $V_{\text{O}}$ close to the filament plug-gap interface. Adapted and reprinted with permission from [635], © 2021 by the American Chemical Society. (Color figure available online).

of the distribution. In research labs, typically, only the mean device behavior can be investigated and the relevant aspects may not be covered.

6.6. Performance trade-offs

Highly promising performances have been demonstrated for memristive devices with respect to switching speed, low power operation and scalability (chapter 6) and with respect to reliability issues such as endurance, variability and retention in this chapter. However, it is important to note that best performance is usually not available within the same material stack and with the same operation conditions. For all filamentary VCM systems there exist several performance trade-offs and the material choice as well as the operation conditions have to be chosen according to the requirements given by the targeted application.

The most important constraint on the device performances is imposed by the aim of low power operation because this strongly reduces the number oxygen vacancies involved in the switching process. Devices with filaments formed and switched with low power result in comparably low $R_{\text{HRS}}/R_{\text{LRS}}$ ratios and are more sensitive to slight variations in the number of oxygen vacancies, resulting in a higher C2C and D2D variability. [637,638] Moreover, devices formed at low power...
usually show a worse retention than devices formed with higher power [633,639] and the low filament temperature also accounts for slow forming or set speed. [637]

On the other hand, devices with softly formed filaments exhibit a much better endurance [640] than devices with strongly formed filament and large $R_{\text{HRS}}/R_{\text{LRS}}$ ratios since they suffer much more from repeated cycling than small filaments. This results in a trade-off between endurance and retention in the most common filamentary VCM devices [623,637,641].

Since the filament shape and strengths is also determined by the material choice, an endurance-retention trade-off becomes also obvious during material optimization. In a direct comparison of HfO$_2$ and Ta$_2$O$_5$ cells, HfO$_2$ showed a better endurance and Ta$_2$O$_5$ a better retention. [641] This was attributed to the higher oxygen vacancy formation energy in HfO$_2$ that thermodynamically promotes the HRS, thus explaining the lower LRS thermal stability in HfO$_2$/Ti cells. In turn, the lower endurance of the Ta$_2$O$_5$ cells was attributed to the lower oxygen vacancy formation energy in Ta$_2$O$_5$ that will promote oxygen vacancy formation during cycling.

A more general model has been suggested by Nail et al [642] to clarify the trade-off between endurance, memory window and retention. They considered material parameters such as the barrier height of diffusion $E_d$, the hopping distance $d_h$ and the formation enthalpy for oxygen vacancies $\Delta H$. In particular, changing $E_d$, $d_h$ and $\Delta H$ controls the switching energy for each SET/RESET cycle, and thus the maximum number of cycles. Consequently, an elevated barrier height $E_d$ increases the switching energy for one cycle (degrading endurance), but also increases the retention activation energy. It should be noted that there is also a related trade-off between switching speed and retention connected with the migration barrier height. A low barrier enables fast switching, but due to the limited achievable nonlinearity of the switching kinetics (Cf. section 6.7.3) the retention time decreases as well. On the other hand, the lower barrier will lead to a less strong thermal feedback upon Joule heating, which may explain the higher cyclability.

With respect to the choice of electrode in HfO$_2$ cells long endurance ( $> 10^{10}$ cycles) could be achieved with Hf and Ti, owning a strong thermodynamic ability to extract oxygen from HfO$_2$. For Ta electrodes, owning a lower thermodynamic ability of extracting oxygen from HfO$_2$, better retention can be achieved. [623]

Very often, bilayer stacks such as Al$_2$O$_3$ and HfO$_2$ [637] or solid solutions of it [643] are used as a good compromise between switching control, memory window, speed and retention. However, one has to be aware that impeding oxygen diffusion in layers stacks in order to improve the retention and the disturb-immunity might at the same time cause a considerable reduction of the switching speed when the diffusion within this layer is the rate-limiting step [532], resulting in a retention-switching speed trade-off.

### 7. Conclusions and outlook

In this review, we have described the current understanding of the underlying physics of VCM-type resistive switching devices. Based on the explanations presented we draw the following main conclusions:

1. Virtually all oxides exhibit VCM-type resistive switching when placed between two (different) metal electrodes. The reason is the occurrence of donor-type ionic defects, typically oxygen vacancies or cation interstitials, and their mobility at (locally) high temperatures and/or fields.

2. Resistive switching is due to the reconfiguration of mobile donor ions throughout the cell, resulting in a modulation of an electrostatic barrier.

3. Electronic conduction is dominated by electron injection at the electronically active electrode (AE) and a transport either via the conduction band or via traps/subbands.
(4) We have explained the two opposite switching polarities, the counter-eightwise (C8W) and the eightwise (8W) switching.

(5) The switching with a C8W polarity is controlled by a concentration polarization of the mobile donors and a change in their concentration in front of the AE (as the main electrostatic barrier) or at an inner interface between oxide layers. An additional oxygen exchange at the ohmic electrode (OE) affects the dynamics and the endurance of the resistive switching.

(6) The 8W switching is based on the dominant contribution of an oxygen exchange at the AE.

Both, C8W and 8W switching, can occur in the same device if the dominating contribution changes as a function of the applied voltage.

(7) A valence change was clearly proven in both cases by spectroscopic methods. In addition, oxygen exchange was also demonstrated.

(8) A further distinction of VCM-type variants was made with respect to filamentary vs. area-dependent switching. Filamentary devices typically require electroforming process. Due to a self-acceleration small inhomogeneities lead to a filamentary (or multi-filamentary) structure in these systems. This is enhanced by local Joule heating. In the absence of a positive feedback (e.g., less Joule heating), area-dependent switching can evolve. Such systems typically do not require a forming process.

(9) Local Joule heating in filamentary devices leads to an ultra-high nonlinearity of the switching kinetics solving the voltage-time-dilemma, reflected in a long retention time and a fast switching at moderate voltages. This, however, comes with strong switching variability (state-dependence). The devices can easily switch in a binary fashion. A gradual switching of the device states is possible in a limited regime too, but not as easily controlled.

(10) Since there is no strong thermal feedback, there is less nonlinearity in the switching kinetics of area-dependent systems. Thus, the retention times are typically reduced (or fast switching cannot be achieved with reasonable voltages). On the other hand, the resistance states can be quite gradually programmed and the systems exhibit less variability.

(11) The gradual SET as well as the gradual RESET can be physically explained as described in this paper.

(12) Physics-based simulation models based on the movement of mobile donors and oxygen exchange at the boundaries can explain the observed bipolar switching properties and the observed failure mechanisms consistently.

Despite significant progress in the microscopic understanding of VCM-type resistive switching, there are still a number of open questions and considerable fundamental research work is still required in the coming years. Although the physics-based modeling has advanced substantially in the last decade, we still lack a true predictive model in which the materials and geometries of a VCM-type cell can be inserted at the atomic level and in which the properties of the corresponding cell could evolve. In other words, a true digital twin of arbitrary VCM-type cells cannot yet be constructed. This is especially true for area-dependent and 8W switching systems, and for the transitions between the different variants. A full understanding of the microscopic physics of VCM cells, in particular the failure mechanisms and the fundamental limits, is considered as a crucial prerequisite, for example, for the exploitation of these systems in future neuromorphic computing.
Acknowledgements

Discussion with our colleagues and students has inspired and shaped this work. In particular, we would like to thank Dirk Wouters (RWTH), Vikas Rana (FZJ), and Victor Zhirmov (SRC Corp., USA) for providing information about the industrial state of art and about scaling aspects, Robert Spatschek (FZJ) and Martin Z Bazant (MIT, USA) for discussions about phase stability, Eike Linn (Vistec Electron Beam GmbH) for discussions about the theory of memristive devices, as well as Stanley Williams (Texas A&M Univ.), John Paul Strachan (FZJ), Thomas Mikolajick (TUD), Daniele Ielmini (Politecnico di Milano, Italy), Susanne Hoffmann-Eifert (FZJ), Ilia Valov (FZJ), Ulrich Böttger (RWTH), and Roger de Souza (RWTH) for numerous discussions on various aspects of this research area. Furthermore, we acknowledge the comprehensive collection of data by our former and present PhD students Benedikt Arndt, Christoph Bäumer, Karsten Fleck, Carsten Funck, Alexander Gutsche, Viktor Havel, Christoph Hermes, Anja Herpers, Andreas Kindsmüller, Camilla La Torre, Florian Lentz, Astrid Marchewka, Lutz Nielen, Christian Rodenbücher, Alexander Schönhals, Katharina Skaja, Moritz von Witzleben, and Stefan Wiefels. We are indebted to Thomas Pössinger for generating and postprocessing all illustrations and for crosslinking all equations, figures, and sections and to Dagmar Leisten for assistance in the illustrations. And we want to heartfully thank Maria Garcia for supporting the project in many technical ways as well as Caroline Zurhelle-Waser for the language check.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the SFB Nanoswitches [grant number SFB 917] funded by the German science foundation (DFG), the MARIE Skłodowska-CURIE ITN ‘Materials for Neuromorphic computing’ (MANIC) and the BMBF project NEUROTEC [grant numbers 16ES1133K and 16ES1134]. One of us (R.W.) acknowledges support by the DFG through the Gottfried Wilhelm Leibniz Prize 2014 [grant number WA 908/30-1].

ORCID

Regina Dittmann http://orcid.org/0000-0003-1886-1864
Stephan Menzel http://orcid.org/0000-0002-4258-2673
Rainer Waser http://orcid.org/0000-0002-9080-8980

References


[87] U. Celano, J. Op de Beeck, S. Clima, M. Luebben, P.M. Koenraad, L. Goux, I. Valov, and W. Vander- 
vorst, Direct probing of the dielectric scavenging-layer interface in oxide filamentary-based valence 

Y. Nishi, and R.S. Williams, Oxygen migration during resistance switching and failure of hafnium 

[89] B.D. Hoskins, G.C. Adam, E. Strelcov, N. Zhitenev, A. Kolmakov, D.B. Strukov, and J.J. McClel-
land, Stateful characterization of resistive switching TiO2 with electron beam induced currents. Nat. 

[90] K. Shibuya, R. Dittmann, S. Mi, and R. Waser, Impact of defect distribution on resistive switching 

mechanisms for epitaxial and ultrathin strontium titanate-based resistive switching bits. ACS Nano 

Hoffmann-Eifert, Understanding the coexistence of two bipolar resistive switching modes with 
OPPOSITE POLARITY IN Pt/TiO2/Ti/Pt NANOIZED ReRAM DEVICES. ACS Appl. Mater. Interfaces 10(35) 
(2018), pp. 29766.

[93] F. Miao, J.J.H. Yang, J. Borghetti, G. Medeiros-Ribeiro, and R.S. Williams, Observation of two resis-
tance switching modes in TiO2 memristive devices electroformed at low current, Nanotechnology 


[95] A.C. Torrezan, J.P. Strachan, G. Medeiros-Ribeiro, and R.S. Williams, Sub-nanosecond switching of 

switching event of VCM-based memories on a picosecond timescale. J. Appl. Phys. 127(20) (2020), 
pp. 204501.


[99] A. Chen, A review of emerging non-volatile memory (NVM) technologies and applications. Solid 

ReRAM cell for storage class memory applications, 2014 IEEE Symposium on VLSI Technology, 
Honolulu, Hi, USA, 2014. doi:10.1109/VLSIT.2014.6894368


[103] G. Burr, B. Kurdi, J. Scott, C. Lam, K. Gopalakrishnan, and R. Shenoy, Overview of candidate device 

(1962), pp. 2669.

Rev. 103(6) (1956), pp. 1648.


(1964), pp. 785.

pp. 106.

184.


[195] S. Menzel and R. Waser, 


[196] T. Menke, P. Meuffels, R. Dittmann, K. Szot, and R. Waser, 


[197] T. Menke, R. Dittmann, P. Meuffels, K. Szot, and R. Waser, 


[200] A.A. Sharma, I.V. Karpov, R. Kotlyar, J. Kwon, M. Skowronsni, and J.A. Bain, 


[201] A. Sharma, M. Noman, M. Abdelmoula, M. Skowronsni, and J. Bain, 


[203] C. Nauenheim, C. Kuegeler, A. Ruediger, and R. Waser, 


[205] C. Lenser, M. Patt, S. Menzel, A. Köhl, C. Wiemann, C.M. Schneider, R. Waser, and R. Dittmann, 


[206] H. Kim and D.W. Kim, 


[207] S. Ambrogio, V. Milo, Z.Q. Wang, S. Balatti, and D. Ielmini, 


[208] H. Hoven, 

Transient processes in resistive switching memory devices at ultimate time scale down to sub-nanosecond range, PhD Thesis, RWTH Aachen University, 2016. doi:10.18154/RWTH-2016-02755

[209] T. Hennen, E. Wichmann, A. Elias, O. Mosendz, R. Waser, D.J. Wouters, and D. Bedau, 


[310] W. Kim, D.J. Wouters, S. Menzel, C. Rodenbücher, R. Waser, and V. Rana, Lowering forming voltage and forming-free behavior of Ta2O5 ReRAM devices, 2016 46th European Solid-State Device Research Conference (ESSDERC), Lausanne, Switzerland, 2016. doi:10.1109/ESSDERC.2016.7599612


doi:10.1109/SISPAD.2015.7292318


[364] A. Schoenhals, S. Menzel, V. Rana, and R. Waser, 3-bit read scheme for single layer Ta2O5 ReRAM, 2014 14th Annual Non-Volatile Memory Technology Symposium (NVMTS), Jeju, Korea (South), 2014. doi:10.1109/NVMTS.2014.7060845


[437] A. Sawa, T. Fujii, M. Kawasaki, and Y. Tokura, Highly rectifying Pr$_{0.7}$Ca$_{0.3}$MnO$_3$/SrTi$_{0.9998}$Nb$_{0.0002}$O$_3$ p-n junction. Appl. Phys. Lett. 86(11) (2005), pp. 112508.


Advances in Physics


Advances in Physics


[630] L. Goux, A. Fantini, G. Kar, Y. Chen, N. Jossart, R. Degraeve, S. Clima, B. Govoreanu, G. Lorenzo, G. Pourtois, D. Wouters, J. Kittl, L. Altimime, and M. Jurczak, Ultralow sub-500nA operating current high-performance TiN\Al2O3\HfO2\Hf\TiN bipolar RRAM achieved through...


