Experimental and Simulation Study of Resistive Switches for Memory Applications

Feng Pan

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2013-38
http://www.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-38.html

May 1, 2013

Copyright © 2013, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

Experimental and Simulation Study of Resistive Switches for Memory Applications

By

Feng Pan

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Vivek Subramanian, Chair

Professor Tsu-Jae King Liu

Professor Ronald Gronsky

Fall 2012

Abstract

Experimental and Simulation Study of Resistive Switches for Memory Applications

By

Feng Pan

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Vivek Subramanian, Chair

In recent years, resistive random access memory (RRAM) has gained significant attention as one of the promising candidates for next generation memory applications. This is due to its anticipated advantages versus Flash technology with respect to high density, low power and fast read and write speed. The main operation mechanism of these devices is a resistance change induced by filament formation through metal-cations or oxygen vacancies.

In the first part of this work, a Kinetic Monte Carlo (KMC) simulator is built to study the filament formation process in an electrochemical metallization (ECM) RRAM. This simulator takes into account most important physical and chemical processes such as oxidation, reduction, metal crystallization, adsorption, desorption and ionic transportation. The characteristics of the forming voltage, forming time and switching speed are investigated. In addition, studies on filament overgrowth and on-state resistance distribution are presented. Further, filament topography, which strongly influences device properties, is studied under different device operation conditions. The simulator also predicts that depending on the strength of lateral electric field, the conductive filament can break at various locations during the RESET process. The simulation results are verified by experiments conducted on $Ag/Ag_2S/W$ and $Cu/H_2O/Cu$ systems.

In the second part of this work, RRAM memory devices based on amorphous Yttria stabilized Zirconia (YSZ) are systematically studied. The effects of different top electrodes of Au, Cu, Ni, Al and Ti are investigated. And the characteristics and the mechanisms of Ti/YSZ and Cu/YSZ are studied in details. It is found that Ti/YSZ has much better endurance, retention and reliability than Cu/YSZ. The underlining physics driving this behavior is investigated. In addition, it is found that Ti/YSZ has very smooth transition in the RESET stage and the off state resistance exponentially increases with an increase of erase voltage. Based on those properties, a multi-level programming (MLP) cell is realized that shows good endurance. The underlying physics that makes the MLP possible for Ti/YSZ is investigated. Finally, it is shown that an incremental step pulse programming (ISPP) technique can significantly increase the device endurance and reliability. Furthermore, it can optimize the tradeoff between resistance programming window and device lifetime.

Acknowledgements

Pursuing PhD in UC Berkeley is an extraordinary experience for me that I have never had before. It is a journey not only to drink in knowledge from book but also to learn from others and realize the weakness of my own and subsequently find a way to improve. Throughout this process there are many people I feel thankful to. And without their support, this work would not have been possible.

I feel deeply indebted and thankful to my parents. I would like thank my Dad and Mom for raising me up and thank for their enormous love and sacrifice for me throughout my life. Without their boundless love and support, I would not achieve what I have today. I would especially thank my father. I would never forget the hardships he had in his life for the family. Furthermore, I would like thank my wife. She is such a nice person and has almost perfect personality and character. Coming across with her is my greatest fortune in my life. I also want to sincerely thank my Dad in law and Mom in law for their selfless support for taking care my daughter and my family. They sacrificed their time and even health in order to let my wife and I focus on research and study. I also would like thank my daughter. She is my hope and my joy source and moreover, she motivates me to work harder.

I would like heartily to thank my advisor Vivek Subramanian for his enormous support and guidance from the very beginning I joined the group. He is one of the greatest advisors I have ever encountered. He gives me the freedom to do the research topic I like. And he never gives students excessive pressure. Life is always balanced in our group. In addition, from him I am not only learning how to systematically conduct scientific research but also learning how to be an integrate, optimistic, mature person which is utmost important for my future career and personal life.

Next, I would like thank my colleague Dr. Shong Yin. I have continued his previous work on non volatile memory which I enjoy and perhaps will further continue on this topic after I graduate. Shong is such a nice people that he is willing to help me any time I need help. And he helps me go through my toughest time at the beginning of my graduate life.

I am also thankful to Kyle Braam, Jaewon Jang, Teymur Bakhishev, Huangyuan Tseng, Hongki Kang, Steven Volkman, Dan Soltman, Lakshmi Jagannathan and all the rest friends in organic group. Thanks for being friends and thanks for the help and encouragement given to me throughout years. I also feel very grateful to Dr. Nuo Xu for his inspiring ideas and great help.

In addition, I would like thank my qualifying exam and dissertation committee members: Prof Tsu-Jae King Liu, Prof Ronald Gronsky and Prof Sayeef Salahuddin.

Last but not least, I would like to thank all the people throughout my life who once loved me and cared about me. Thank you very much!

Table of Contents

Chapter 1: Introduction to Nonvolatile Memory Devices 1
1.1 Memory background 1
1.2: FLASH 8
1.2.1: Floating Gate Device Operation Mechanism 9
1.2.2: NAND and NOR FLASH 11
1.2.3: The Application and Scaling Limitations for FLASH 15
1.3: Introduction to Resistive Random Access Memory 16
1.3.1 RRAM Background 17
1.3.2 Classification of RRAM Devices 18
1.3.3: Introduction to Electrochemical Metallization RRAM 19
1.3.3: Introduction to Oxygen Vacancies Based RRAM 25
1.3.4: The Equivalence Between ECM and VO Based RRAM 31
1.4: Performance and Scaling Perspective of RRAM 32
1.4.1: RRAM Performance 32
1.4.1: RRAM Scaling 33
1.5: This Work 38
1.6: References 39
Chapter 2: Kinetic Monte Carlo Simulation on Electrochemical Metallization RRAM 43
2.1: Introduction to Kinetic Monte Carlo Method 43
2.2: Establishment of a KMC simulator for ECM RRAM 47
2.2.1: Basic Chemical and Physical Processes and Assumptions 47
2.2.2: Simulation Procedures 50
2.3: Simulation Results and Discussion 52
2.3.1: Filament Formation Study 52
2.3.2: Relationship Between Forming Voltage and Device Size 57
2.3.3: I-V Characteristics and SET Switching Properties 59
2.3.4: RESET Stage Characteristics 62

2.3.5: The Effect of Material Properties 65
2.4: Summary 66
2.5: References 67
Chapter 3: Yttria Stabilized Zirconia Based RRAM 69
3.1: Introduction to Yttria Stabilized Zirconia 69
3.2: Experimental Procedures 72
3.3: Effect of Different Top Electrodes and Switching Mechanisms 75
3.4: Ti/YSZ and Cu/YSZ Devices 77
3.4.1: I-V Characteristics for Ti/YSZ and Cu/YSZ 77
3.4.2: Endurance, Retention and Uniformity of Ti/YSZ and Cu/YSZ 83
3.4.3: Noise Analysis for Ti/YSZ and Cu/YSZ 87
3.5: Summary 92
3.6: References 92
Chapter 4: Multi-level Programming and Impact of Programming Methodology on Device Reliability 95
4.1: Introduction 95
4.2: Multi-level Programming of Ti/YSZ in RESET Stage 96
4.3: Impact of Programming Methodology on Device Reliability 99
4.4: Summary 102
4.5: References 103
Chapter 5: Conclusions and Future Work 105
5.1: Conclusion 105
5.1.1 KMC Simulation for ECM based RRAM 105
5.1.2 YSZ based RRAM 106
5.1.3 Multi-level Programming of Ti/YSZ based RRAM 107
5.1.4 Impact of Programming Methodology on Device Reliability 108
5.2: Future Work 109
5.3: References 109

Chapter 1: Introduction to Nonvolatile Memory Devices

1.1 Memory background

Electronic memory devices have vast applications in almost every area of modern life. Essentially, all modern electronic devices have memories either embedded or externally attached. In addition, the emerging of the personal electronics such as laptop computers, digital cameras, smart phones, tablets and other entertainment devices in the past two decades have resulted in a dramatic increase in the demand for such memory devices. Consumers' consistent desire to have better and cheaper memory has motivated researchers to pursue the continued advancement of memory technology. In the following paragraphs, the properties that a good memory device should have are described.

First, the density of a memory should be as large as possible such that it can store a large amount of data in a given die area. This means the cost per bit of information stored can be minimized. This significantly benefits the consumers in that they can satisfy their storage needs at a cheaper price. In fact, it is hard to imagine that in the year 1966, it would incredibly cost $\$8\times10^{12}$ to have a one gigabyte DRAM, whereas today it is only around $2 [1]. This shows the tremendous importance of high density memory.

Second, a memory device should have good endurance. Endurance determines how many cycles the device can sustain during write and erase operations. A device with large endurance can be repeatedly used without encountering operational failures. Clearly, this is critical because consumers don't need to replace the memory device frequently.

Third, a good memory device should have long retention. Retention defines how long data can be stored inside the memory devices without being lost inadvertently. A good memory should have as long a retention time as possible, so that once the memory is written, the system doesn't need to periodically back up or refresh the data.

Next, it is also good for the memory device to have fast programming and access speed so that the time spent on reading data from and writing data into the memory can be minimized. This is especially important when the speed is the primary concern. For example, in the computational arena, there is a trend towards replacing traditional hard disk drives (HDD) with solid state drives (SSD), which are based on FLASH technology. One of the reasons for doing so is that SSD has a faster access time. Thus, it can significantly increase computation speed and hence improves the computer efficiency.

Finally, the power consumption of the memory should be as low as possible. This is especially important for mobile devices such as smartphones, tablets and digital cameras. A low

power device reduces the need for battery charging and brings a significant convenience to customers.

In reality, it is very difficult for a memory device to have all the merits listed above. Depending on the application, every type of memory always has a particular focus. However, the merits mentioned above serve as guidelines motivating people to do research and try to come up with new memory devices that can have as many of the aforementioned properties as possible. In the following paragraphs, the classification of memory devices is discussed. In addition, special attention is paid to phase change memory, which is one of the emerging memory devices with which people hope to replace the current floating gate memory device technology. Finally, due to the extreme importance of floating gate memory devices in current technology, they are discussed comprehensively in the next section.

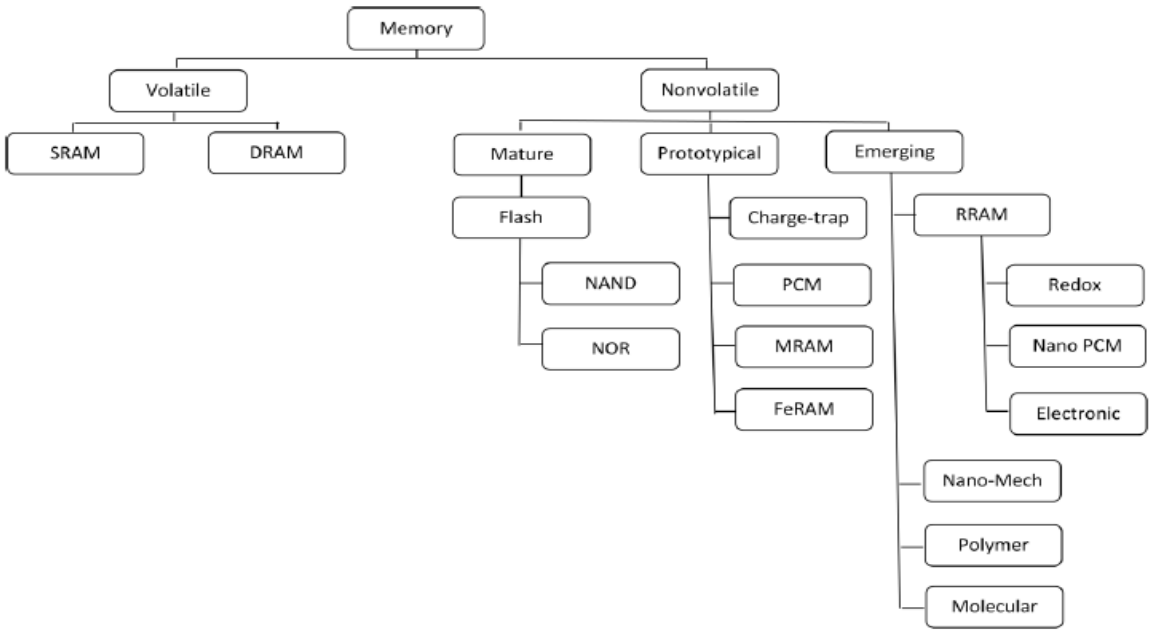


Figure 1.1. Memory classifications (An Chen – GLOBAL FOUNDRIES) [2]

Generally memory devices can be classified into two categories based on their operational behavior as shown in Fig. 1.1 [2]. One class is volatile memory and the other type is non-volatile memory (NWM). Volatile memory, by definition, has to be connected to a power supply during the entire operation period. It will lose all the stored information (bit stored in each memory element) once the power is turned off. One of the most important volatile memories is static random access memory (SRAM), which consists of six transistors for each memory element as shown in Fig. 1.2 (a). Essentially, it is two inverters connecting in series with two access transistors attached to each input of the two inverters. As can be seen, due to the complicated

structure involving six transistors per bit, the density of SRAM is poor. Dynamic random access memory (DRAM) on the other hand has a much simpler structure as shown in Fig. 1.2 (b). It consists of a transistor in series with a capacitor which can naturally give rise to two memory states. The first state is the one when the capacitor is charged and the other one is when the capacitor is empty. Ideally, if there was no leakage current passing through the capacitor when it is in charged state, DRAM can hold the charge forever and information will remain even if the power is off. In reality, however, due to the leakage current present all the time, charges stored on the capacitor can only stay for several milliseconds. Therefore, periodic refreshment is needed to keep the information. The leakage current could come from several factors. First of all, the leakage path could be due to the non-ideal insulating properties of the dielectric between the two capacitor metal plates, where a finite current can always flow through the capacitor. The second leakage source is the subthreshold conduction of the transistor which results a finite a leakage current even if the world line is unselected. As can be seen due to its simple structure, DRAM has a much larger density than SRAM. Because of this advantage, the main memory in personal computer is DRAM.

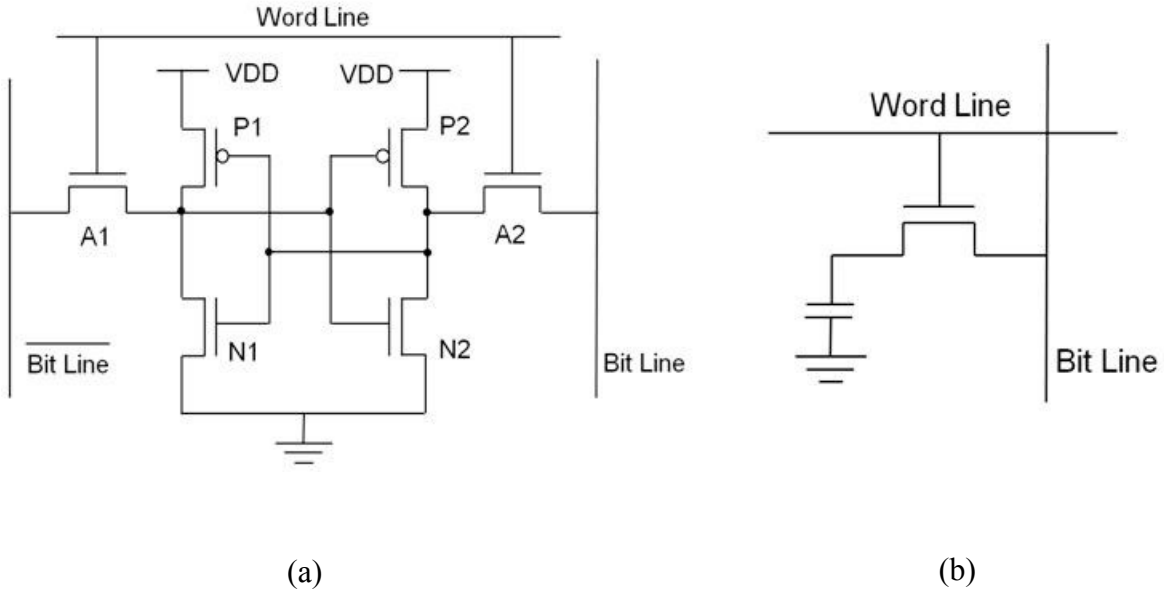


Figure 1.2. (a) Structure of SRAM (b) Structure of DRAM

NWM, on the other hand, does not require power to sustain information. Today, the main commercial NVM technology is Flash, which is short for “flash electrically erasable programmable read-only memory”. The word “read-only” is a bit confusing since nowadays those memories can be erased and programmed many times. The details of Flash will be discussed in the next section. Now, a question may naturally arise. Since NWM can sustain data

even without power supply, why do people still use volatile memory? The answer to this question is related to memory hierarchy as shown in Fig. 1.3 [3]. SRAM, even though it is volatile with very low density, has very fast program, erase and access time. Thus, it is used inside the processor chip and acts as cache. DRAM which has larger density than SRAM but slower speed is used as RAM for a computer. Traditional NVM such as Flash, even though it has much larger density than both SRAM and DRAM, is much slower than both of them. Thus, Flash is typically used for external storage of data instead of storing programming code inside a computer chip.

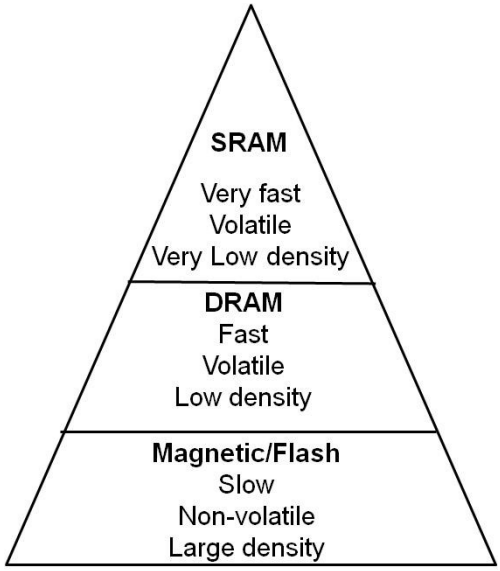


Figure 1.3. Memory Hierarchy

The initial motivation for people to study other types of NVM is to further improve the memory density such that people can store more information on a given size chip. Flash technology, as will be discussed soon in the next section, faces serious downscaling bottlenecks for device feature sizes less than 20nm [4]. Currently, lots of research has been done on phase change memory (PCM) [5, 6]. At a fundamental level, phase change memory utilizes joule heating to induce a phase change inside the material (germanium-antimony-tellurium (GST) as an example [7, 8]), so that the resistance of the material will change accordingly. Typically, it involves physical transitions between amorphous and crystalline states. Fig. 1.4 (a) shows the schematic of a conventional PCM cell. As can be seen the unit cell consists of both top and bottom electrodes with a phase change material sandwiched in between. One unique feature for the archetypal PCM cell is that there is a heater embedded in the insulator layer between the phase change material and the bottom electrode. The purpose of this heater is to increase the programming efficiency and reduce the programming power, as will be discussed soon.

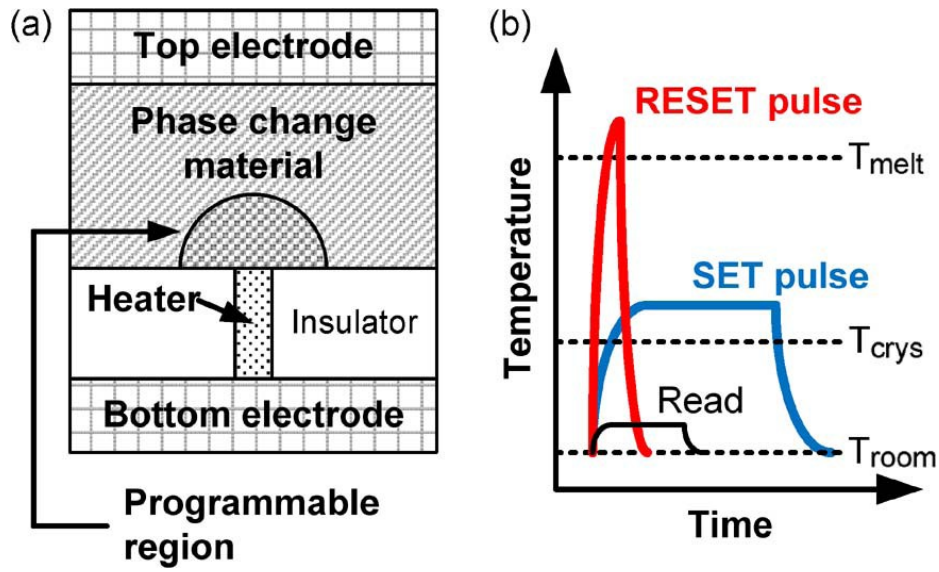


Figure 1.4. (a) A schematic of a conventional PCM cell. (b) The temperature evolution inside the PCM device under different operation regimes [5]

As fabricated, the phase change material is in a crystalline state, because the processing temperature of the metal interconnect layer is high enough to crystallize the phase change material [5]. The crystalline structure generally shows a low resistivity property (memory at a low resistivity state is commonly described as being in the on state). In order to switch the device from the on state to the off state (high resistance state), which is defined as a RESET process, a large and fast voltage/current pulse has to be applied across the top and bottom electrodes such that the temperature inside the PCM material can rapidly rise up to the melting temperature, T_{melt} as shown in Fig. 1.4 (b) due to the joule heating effect. This high temperature will randomize the atomic order inside the PCM material, causing the material to transition from a crystalline structure to an amorphous state. One important note is that the falling edge of the applied pulse has to be very sharp as shown in Fig. 1.4(b). This means the device has to be quenched rapidly from T_{melt} to a lower temperature below the crystallization temperature $T_{crys,}$ in order to prevent crystallization from occurring. In this process, a heater can be added to localize the current and thus enhance the temperature in the local programmable region as shown in Fig. 1.4 (a). In this case only part of the PCM material is converted to the amorphous state. However, since this portion of material is in series with the crystalline material on the top, it effectively determines the overall resistance of the cell [5]. By adoption of the heater, due geometry effects, the current density at the heater and PCM material interface is much higher than the current density at other locations. Thus the temperature at the interface region reaches T_{melt} first, which make the PCM material become amorphous. Subsequently, the boundary of the amorphous region starts to grow and a droplet shape region,

defined as a programmable region, is formed inside the PCM material as shown in Fig. 1.4 (a). The significant advantage of this approach is a reduction in the programming current and power due to the current localization effect.

Next, in order to switch the PCM cell from an amorphous to a crystalline state, defined as a SET process, an annealing process is applied. During this process, a moderate voltage/current pulse with longer duration is applied. The temperature of the PCM cell rises up to T_{crys}, but stays below T_{melt}. T_{crys} refers to the temperature at which the PCM material can crystallize. Generally, the crystallization of the material takes time, thus it is very important that the pulse be long enough that the temperature inside the PCM material is held high for sufficient time such that the crystallization process can be completed. After this process, the device transitions from the high resistive state to a low resistive state.

The reading process is relative straightforward compared with the SET and RESET processes. During the reading process, a small voltage/current detection pulse is applied across the electrodes. The resultant temperature inside the PCM material is well below T_{crys} so that the memory state of the device is not disturbed.

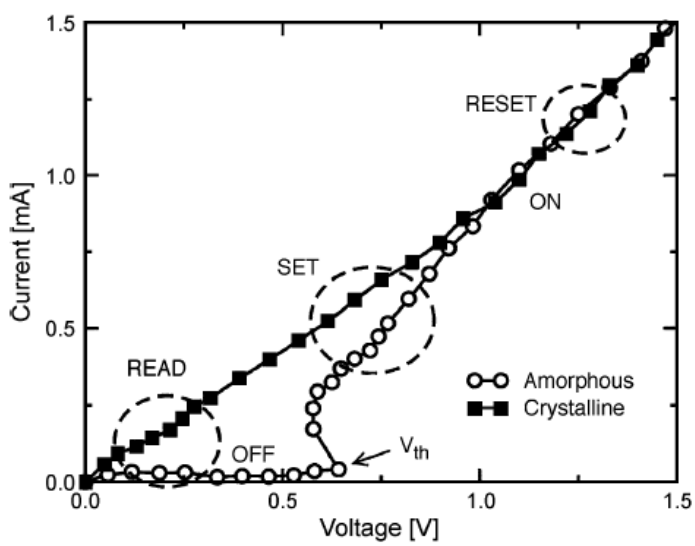


Figure 1.5. The I-V characteristics of a typical PCM cell under different operation regimes [5]

Fig 1.5 shows the typical current-voltage characteristics of a PCM device [5]. Assuming the device is initially at the off state, since the the resistivity of the amorphous material is large, there is little current flowing through the device. The device is turned on when the voltage reaches a threadhold value (V_{th}). This is accompanied by a large increase in current. This is indicated by SET process as shown in the figure, and the device reaches the on state. In the RESET stage, an even higher voltage is applied. If the pulse drops rapidly, the device will reach the off state. However, if the voltage is ramped down gradually, there is enough time for the

material to crystallize and the device will stay at the on state. In addition, it is interesing to see that there is snap-back behavior in the I-V characteritics shown in Fig. 1.5 and resulting in a negative resistance. F. Buscemi and colleagues [9] conducted a Monte Carlo simulation and conclude that when the electron hopping process inside the PCM material is not fast enough, charge is accumulated close to the contact regions, which generates a counter electric field. The net effect is a decrease of voltage as the current increases which results in the snap-back phenomenon. Thus, the snap-back behavior is a non-equilibrium behavior.

One problem associated with PCM is the power consumption. In order to change the crystal structure of the film, a large amount of current has to be applied to the device to make sure that enough heat is generated. Fig. 1.6 shows the relationship between the RESET current and effective device contact area. As can be seen, even though the RESET current decreases as the device contact area decreases, the required current density is more or less constant around 40MA/cm^2. It is estimated that even for a PCM with size reduced to 15nm on a side, 40µA RESET current is still needed [2]. This limits the applications of PCM, particularly with respect to today's low-power electronic devices such as laptops and mobile devices.

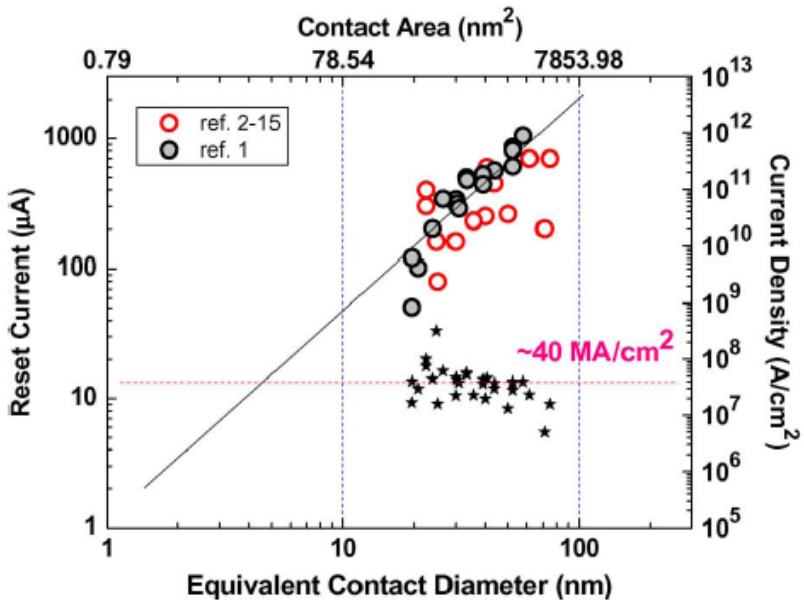


Figure 1.6. The relationship between RESET current and equivalent device contact diameter [5]

It will be ideal if people can find a kind of universal memory device which is non-volatile like FLASH, offers fast programming and access time like SRAM, and at the same time has the combined advantages of high density and low power consumption [10]. If such a universal memory exists, it can not only replace FLASH but also possible replace DRAM such that it could

be directly integrated into computer chips and thus can significantly improve computational performance. Resistive random access memory (RRAM) is such a candidate memory.

1.2: FLASH

Flash memory has had a dramatic market growth in the past two decades. Table 1 shows the rise of the Flash memory market in the past 20 years [11]. As can be seen within only fifteen years, the share of Flash memory in total memory market increased from 0.3% to 34.4% and by year 2006 Flash alone had occupied 8.1% of entire semiconductor market. The rapid growth of Flash memory is attributed to two of its major properties which ideally suit the consumer electronics market [11]. The first one is that Flash is small and reliable. The second feature is its non-volatile property. Due to those features, Flash has become an ideal candidate for electronics requiring mobility and miniaturization. It is typically used in USB flash drives, solid-state drives and memory cards which can be found in various modern electronic products such as smartphones, digital cameras, tablets, laptops and personal computers.

Table 1.1. The rise of the Flash memory market [11]

	Flash memory market (USD Million)	Flash memory market annual percentage growth	Flash memory as percentage of total semiconductor market	Flash memory as percentage of total memory market
1990	35		0.1	0.3
1991	135	286	0.3	1.0
1992	270	130	0.5	1.8
1993	640	106	0.8	3.0
1994	865	35	0.9	2.7
1995	1,860	115	1.3	3.5
1996	2,611	40	2.0	7.2
1997	2,702	3	2.0	9.2
1998	2,493	-8	2.0	10.8
1999	4,561	83	3.1	14.1
2000	10,637	133	5.2	21.6
2001	7,595	-29	5.5	30.5
2002	7,767	2	5.5	28.7
2003	11,739	51	7.1	36.1
2004	15,611	33	7.3	33.1
2005	18,569	19	8.2	38.3
2006	20,275	9	8.1	34.4

In this section, the operation principle of Flash and its limitations are discussed. First of all, the basic physics of its building block floating gate device is described. Next, NAND and NOR, two types of Flash structures, are discussed in detail. The discussion includes their programming and erases mechanisms as well as physical properties like endurance, retention

and reliability. Finally, the limitations affecting the future development of Flash are emphasized.

1.2.1: Floating Gate Device Operation Mechanism

The basic building block of FLASH is a special type of metal oxide semiconductor field effect transistor (MOSFET) as shown in Fig. 1.7 (a) [12]. The major difference compared with a normal MOSFET is that a second gate named "floating gate (FG)" is inserted between the control gate (CG) and silicon substrate. The purpose of this floating gate is to modulate the device threshold voltage (V_{TH}). If there are electrons trapped inside the FG, V_{TH} will shift towards the positive direction as shown in Fig. 1.7 (b) by an amount of:

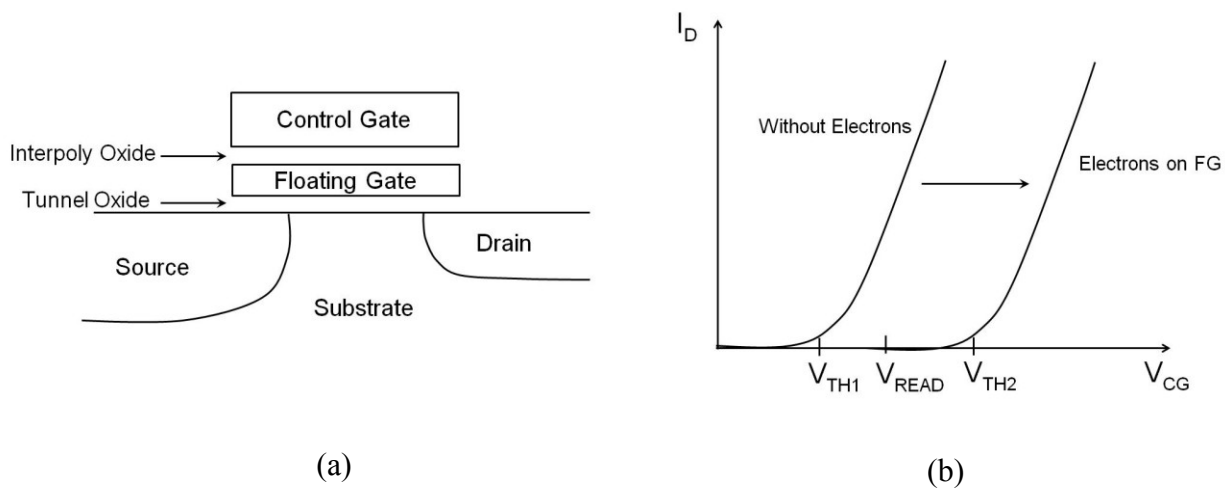


Figure 1.7. (a) The structure of floating gate device. (b) Typical I-V characteristics of a floating gate device with and without electron on the floating gate.

$$\Delta V_{TH} = \frac{-\Delta Q_{FG}}{C_{FC}} \tag{1}$$

Where ΔQ_{FG} is the electron charges trapped on the FG and C_{FC} is the coupling capacitance between CG and FG. Thus, if a read voltage (V_{READ} shown in Fig. 1.7 (b)) is applied, there will be no current flowing for devices which have electrons trapped on their FGs, yet a significant amount of current flowing for devices which don't have electrons on their FGs. Therefore, by doing so, two memory states are created.

There are two ways to inject electrons (program) into the FG as shown in Fig. 1.8 (a). The first one is to use hot carrier injection (HCI) and the second method is by adopting F*owler-Nordheim* (FN) tunneling shown in Fig. 1.8 (b). HCI is done by applying a large positive voltage above 10V to both the CG and the drain electrode, which drives the transistor into saturation

mode. A very large electric field is therefore created near the junction of the drain and substrate corner. Electrons in this region gains significant kinetic energy and subsequently become "hot". Once electrons have enough energy, they can overcome the potential barrier represented by the tunnel oxide, and traverse through the tunneling oxide and get trapped inside the floating gate. One obvious disadvantage of this method is power consumption. In order to create enough hot carriers, a large drain current has to flow though the device which results in very poor power efficiency.

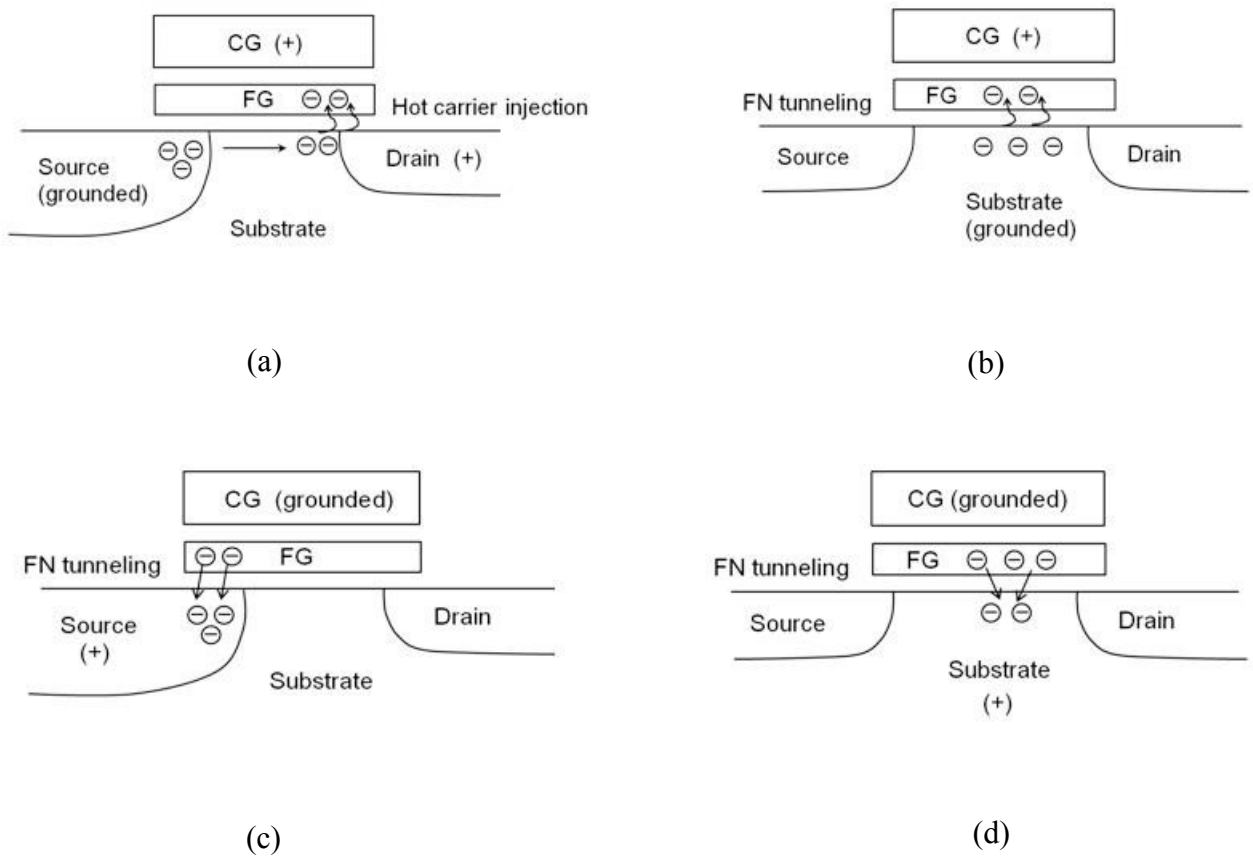


Figure 1.8. To program the Floating gate device by (a) hot carrier injection method and (b) by FN tunneling. and to erase the device by (c) FN tunneling at source side and (d) FN tunneling at channel.

Another type of programming method is based on FN tunneling as shown in Fig. 1.8 (b). In this case, the CG is positively biased and the substrate is grounded while the source and drain are kept floating. The strong transverse electric field between the channel and the CG enable electrons to tunnel from the inversion layer formed in the substrate directly into FG. The major advantage of this method is that FN tunneling consumes much less power than that of HCI. The approximate mathematical formula to describe FN tunneling can be achieved by using the Wentzel-Kramers-Brillouin (WKB) approximation [13]

$$J = \frac{q^3 F^2}{16\pi^2 h^2 \Phi_B} \exp\left[-\frac{4(2m_{ox}^*)^{1/2} \Phi_B^{3/2}}{3\hbar q F} \right] \quad (2)$$

Where Φ_B is the barrier height and m_{ox}^* is the effective mass in the dielectric and F is the electric field through the oxide. As can be seen, the current density is very sensitive to electric field and this can cause a variation problem with respect to the V_{TH} as will be discussed later.

There are two methods to erase the electrons on the FG. Both of them are based on FN tunneling as shown in Fig. 1.8 (c) and (d). The first method is to apply a positive voltage at source side and leave the CG grounded. The electrons can then tunnel through the oxide and reach the source region. Since a large positive voltage is present on the source side, the junction between the source and substrate is highly reversely biased. Therefore, to prevent junction breakdown, the doping concentration of the source cannot be too high. Practically, the source depth is designed deeper than the drain side with less doping concentration to prevent breakdown [12]. A second problem associated with this erase operation is that as device size keeps scaling down, the overlap region between the FG and the source side become smaller and smaller. Thus the erase process gradually becomes less efficient. To overcome this problem, a second erase methodology has been adopted as shown in Fig. 1.8 (d). In this case, the substrate instead of the source is positively biased and the CG is grounded. The electrons can directly tunnel into the channel region. Furthermore, a symmetric source and drain configuration can be utilized.

Finally, even though FN tunneling is power efficient, the exponential dependence of tunnel current causes critical problems of process control. Any small variation in oxide thickness can cause significant variation in V_{TH} during the programming and erase steps. Thus, good process control is needed to minimize the oxide surface roughness [12].

1.2.2: NAND and NOR FLASH

A typical FLASH memory is based on one of the array structures shown in Fig. 1.9. The NAND module contains a series of floating gate devices connecting in series while the NOR module has them connecting in parallel. Fig. 1.10 [12] shows the layouts of those two structures. The layout is drawn based on one micron technology design rules, for convenience, but is generally applicable across all technology nodes. In the example, each memory array consists of eight single cells. As can be seen, the average cell area occupied by NOR Flash is 29.6 μm^2, whereas only 12.9 μm^2 is needed for NAND Flash. It is obvious that the density of NAND FLASH is higher than that of NOR, particularly since every transistor in the string does not require its own source and drain contact. This is one of the major advantages of NAND over a NOR structure. In fact, if F represents the smallest feature size at any given technology node, the individual cell size in NAND structure is $2F \times 2F = 4F^2$, and it is at least $5F \times 2F = 10\ F^2$ for NOR.

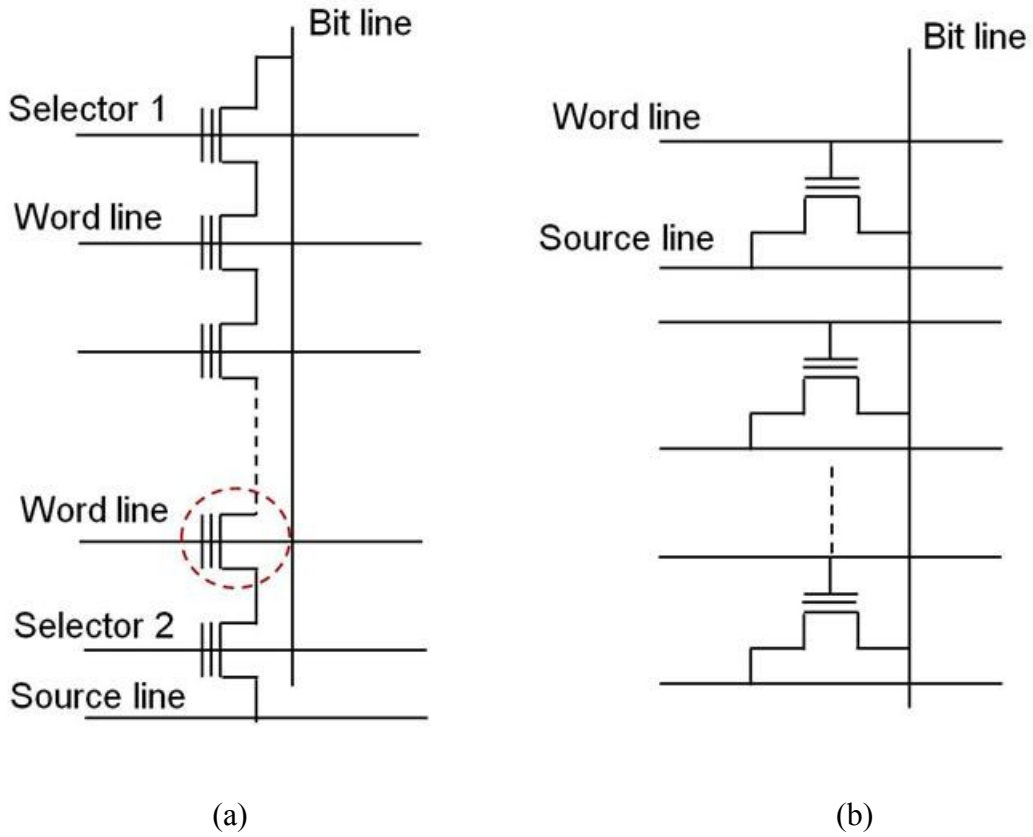


Figure 1.9. (a) The NAND array structure of FLASH (b) the NOR array structure.

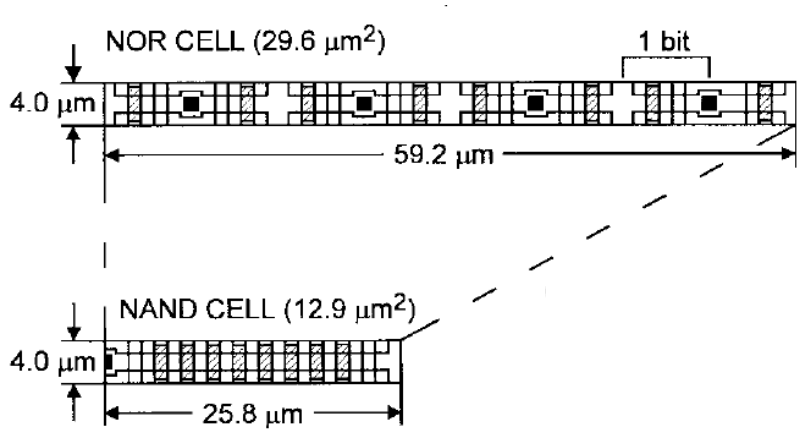


Fig. 1.10. Layout comparisons between NAND and NOR Flash [12].

The read operation of NAND FLASH is done as follows. First, the BL is pre-charged to ~1V and the select gates of the block are enabled. Next, the selected WL is held at 0V and all the other deselected WL are biased at a voltage which is higher than the highest V_{TH} such that the information can go through. If the selected cell has a V_{TH} smaller than 0V, the BL will discharge and a page buffer circuit will read the data as "1". If the selected cell has a V_{TH} larger than 0V, there is no current flowing through and the BL stays at 1V and the page buffer reads the data as 0 [14]. Because of the NAND structure one obvious disadvantage is the slow read speed. For example, in order to read a cell at the bottom of the chain as shown in Fig. 1.9(a), circled by dash line, signal has to pass through all the other cells on top of it, which generates significant amount of parasitic resistance and capacitance and therefore the RC time delay is large. On the other hand, such problem does not exist in NOR FLASH structure which offers the random access to every memory cell inside the array. Because of access time difference, together with the difference in density, NOR FLASH is more attractive for applications such as program-code storage, while NAND is more suitable for video or audio storage which do not need fast random access but need large storage density [15].

To program and erase the memory array is much more complicated than the read operation. Generally before the programming step, all the cells in both NAND and NOR array have to be erased first [16]. It is straightforward to do the erase operation for NAND architecture, in which all cells in the module are erased at same time and this is achieved by applying large positive voltage to the substrate and 0V to the word line while keeps the source line and bit line floated so that the electrons can tunnel back to the channel. For NOR array, 0V is applied to CG and a large positive voltage to the source side so that electrons can tunnel into the source region.

To program a cell inside a NAND array, the selector 1 is enabled such that the bit line is connected to ground while selector 2 is disabled so that the source line is kept floating. Then a high voltage (~20V) is applied to the CG and electrons can subsequently tunnel into the FG by the FN tunneling mechanism. To program a NOR device, a high voltage is applied to both word line and bit line of the selected devices, based on HCI, electrons can overcome the barrier of oxide and get trapped in the FG. Because of the difference in the programming methodology, NAND FLASH is much more power efficient than NOR array.

It has to be mentioned that because of its parallel array structure, leakage path problems have to be avoided in NOR memory array and in practice a program "0" step has to be applied during the erase process of NOR Flash. Fig. 1.11 illustrates the entire erase procedure of a NOR device [16]. Before the erase operation, the V_{TH} distribution throughout the memory array is illustrated in Fig. 1.11 (a). After that, a "program 0" step has to be performed shown in (b) such that the V_{TH} of all the devices is raised, which reduces the over-erase problem in the following steps. Next, an erase step is performed as shown in (c). As can be seen, this erase step causes a large distribution of V_{TH} which can go beyond the acceptable range of V_{T1} and V_{T2} respectively. Therefore additional steps have to be performed to squeeze all the individual V_{TH} values into this allowable range. Thus, an erase verification procedure has to be applied and this is shown in (d). However, at this moment some devices are over-erased as shown in (d) and V_{TH} of some devices are less than V_{T1} and even go negative. This should be absolutely avoided in the NOR array since cells with negative V_{TH} cause significant leakage current when the device is

unaddressed (e.g. the word line voltage is 0). Therefore, as a final step, a “soft programming” has to be done to shift all the V_{TH} within the range.

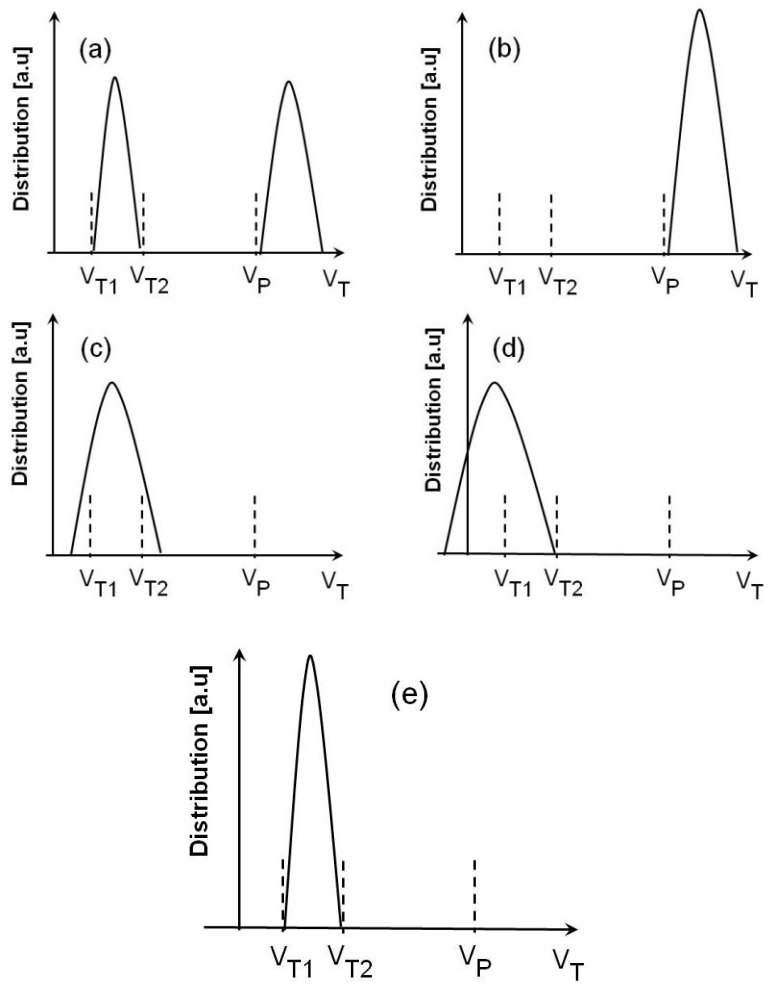


Figure 1.11. Illustration of threshold voltage distribution in the erase process for NOR Flash [16] (a) before erase (b) after a “program 0” step (c) after a single erase pulse (d) after the erase verify procedure (e) after soft-programming.

Table 1.2 Performance comparison between NAND and NOR Flash

	Density	Read	Erase	Program	Access Method
NAND	High	slightly slower	Milliseconds	Fast	sequential
NOR	Low	Fast	Seconds	Slow	Random

Due to the complicated erase procedure, NOR takes a much longer time to erase than does NAND. Table 1.2 compares the performance of NAND and NOR. Due to the differences in their performances, NOR is more suitable for code storage and NAND is more suitable for data storage.

1.2.3: The Application and Scaling Limitations for FLASH

Based on the fundamental operation mechanism of FLASH, there are several limitations from both application and scaling perspectives. First, as discussed above it is impossible to achieve properties like fast programming, access and high density together with random access at same time. NAND Flash offers the advantages of faster programming and high density yet it doesn't have the random access capability and the read speed is slow. NOR Flash, on the other hand, has random access capability; however, it is inferior to NAND in all the other aspects. Thus, it is desirable to have kind of device which has all the merits as a memory device.

There is another fundamental challenge for Flash for future NVM application. In fact, scaling will become the eventual bottleneck for all such devices. Even though NAND Flash has higher density than NOR, it faces lots of difficulties to continue scaling down. People generally believe that the feature size of traditional Flash is impossible to scale down to below 20nm [17]. Those limitations generally are related to its fundamental operation principle. First of all, the interpoly dielectric thickness must scale with tunnel oxide to maintain adequate coupling of the applied voltage to the tunnel dielectric [17]. To ensure sufficient voltage can be passed to the floating gate during program and erase operations, the gate coupling ratio (the capacitance ratio of the CG to FG capacitor to the total FG capacitance) has to be greater than 0.6 [17]. As devices scale down, the CG needs to wrap around the FG to provide enough capacitance. However, it is very difficult to achieve such a structure when the bit line spacing becomes 20nm or less. In addition, there is also a limitation on the thickness of the tunneling oxide. If the tunneling oxide is too thin, electrons can, even without any external electric field, easily tunnel back to the channel. Therefore, the device retention properties could be severely degraded. The same principle is also true for interpoly oxide. Furthermore, because of the difficulties in scaling down the thickness of oxides, the reduction of the operating voltage also becomes difficult.

In addition, there are additional factors that set a fundamental scaling limitation on the device channel length. For devices that are programmed by HCI and erased by electrons tunneling from the FG to the source side, a minimum overlap between drain and FG as well as between source and FG is needed. This sets a limitation on how small the channel length can be. In addition, even for device based on FN tunneling between the FG and the channel, the channel length cannot be too small in order to prevent leakage current caused by various short channel effects. Finally, as the size of device goes smaller, the number of electrons stored in the FG will be very small and cause severe random telegraph noise [17] and furthermore, result in associated reliability problems.

1.3: Introduction to Resistive Random Access Memory

As stated in the previous section, due to the scaling limitations of current Flash technology, it is desirable to discover new technologies such that the scaling can be continued further. More than that, it is very attractive to find one kind of memory which has the combinational advantages of both SRAM and Flash such that it not only has fast program, erase and access time but also has good non-volatility as well as high density characteristics. This type of memory can be called universal memory or storage class memory [18]. If this type of memory is realized, the memory hierarchy will be fundamentally changed as shown in Fig. 1.12.

As shown from bottom to top, currently we have tape, magnetic/Flash, DRAM and Cache as well as SRAM. Along this direction, the speed of the memory goes up while the density decreases. If a universal memory can be realized, such a memory could replace the Flash, the DRAM and part of the Caches. The magnetic disk is only used for archive purpose due to its extremely high density while the SRAM is only adopted when extremely fast speed is required. As a result, the memory hierarchy becomes much simplified.

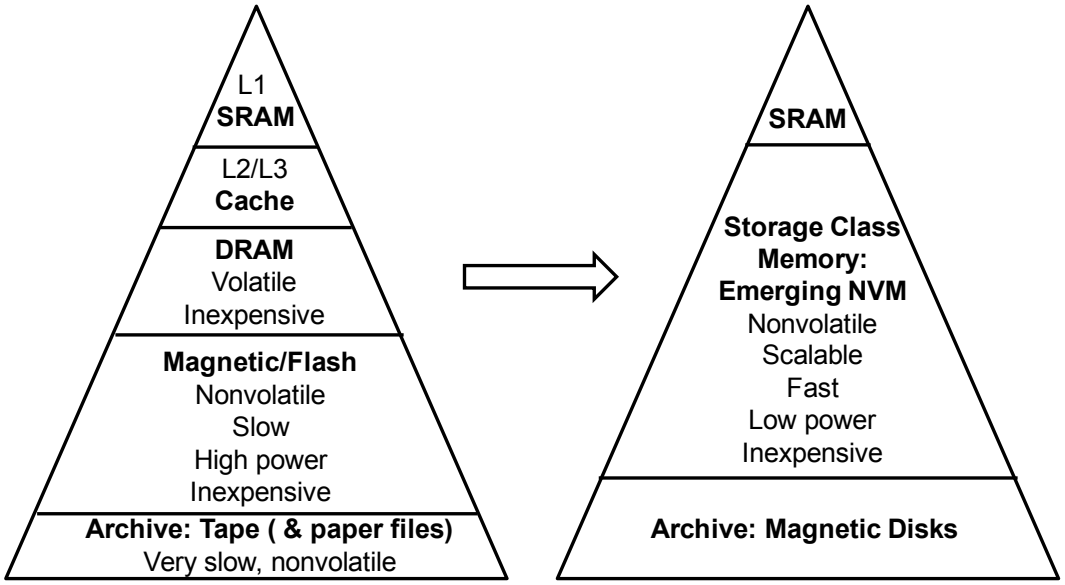


Figure 1.12. The modification of memory hierarchy upon realization of a universal memory (storage class memory) [18]

The requirements of such universal memory will be nonvolatile, scalable, fast, low power and inexpensive. People are trying various technologies and developing new physical concepts to make this type of memory possible. One of most promising candidates the so called Resistive Random Access Memories (RRAM) will be introduced in this chapter.

1.3.1 RRAM Background

The history of the resistive change effect in RRAM can be traced back to the 1960s. In 1964, P.H. Nielsen [19] and colleagues first found the resistive switching effect in an Au/SiO/Au structure. They further suggested that such an element could be used as a memory device which could be accessed nondestructively. Several years later, several researchers also found similar phenomena in an Al/SiOx/Au system [20, 21]. Due to stability problems and the emergence and successful development of Si based memories [22], the resistive switching memories never developed until quite recently. In 2002, W.W. Zhuang and colleagues [23] demonstrated an RRAM device based on PCMO. They first showed that such a device is capable of providing memory characteristics that are potentially more favorable than that of traditional FLASH devices. After that, in 2004, Samsung Electronics [24] demonstrated resistive switching behavior in various transition metal oxides (TMO) such as NiO, TiO_2, HfO_2 and ZrO_2. The devices delivered operating voltage below 3V, programming current around 2mA with 10^6 cycles of endurance. Ever since, RRAM has become one of the hottest research topics in both academia and industry. The main motivation for people to investigate RRAM is due to its anticipated advantages, which almost cover all the merits of an ideal memory such as large endurance, long retention, fast access speed, low power and most importantly high density.

Similar to PCM, the operation of RRAM, through applying voltage/current pulses, induces a resistance change in the device system. In addition, the RRAM device can be repeatedly switched between high and low resistance states. Unlike PCM, however, in which the resistance change is achieved by modification of the crystal structure of the resistive layer, the operation of RRAM generally involves electrochemical processes together with the physical migration of metal cations or oxygen anions.

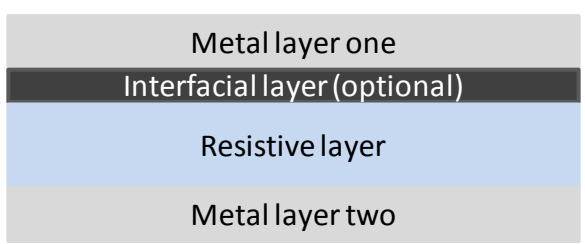


Figure 1.13. basic structure of RRAM

Fig. 1.13 shows the typical structure of a RRAM device. Generally, the structure is simple and consists of several layers. Two of them are top electrode and bottom electrode with a resistive layer sandwiched in between. And there is another optional interfacial layer which sometimes is used to improve the RRAM device performance. It is worthy to note that the basic structure is very similar to that of a PCM cell. One major difference between them is that in RRAM, the additional heater structure is eliminated. This greatly simplifies the fabrication process flow. As will be discussed later, even though there is no heater structure, RRAM is much more power efficient when compared with PCM.

1.3.2 Classification of RRAM Devices

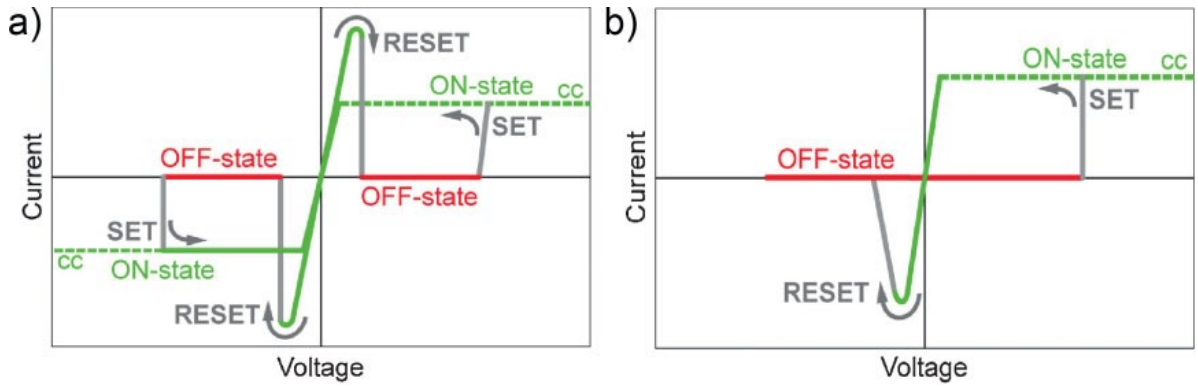


Figure 1.14. Two operation schemes for RRAM (a) unipolar device (b) bipolar device [25]

There are many ways to classify the various kinds of RRAM devices. Two approaches are particularly important and will be discussed here. The first approach is to classify RRAM devices based on the symmetry of I-V characteristics during the SET and RESET operation as shown in Fig. 1.14 [25]. Fig. 1.14 (a) illustrates one type of RRAM device, in which both the SET and RESET operation can be achieved by applying either positive or negative voltage. In other words, there is no polarity dependence of the device I-V characteristics. This type of device is called a unipolar device. It is worthwhile to point out that for such device, a current compliance (CC) has to be applied during the SET operation to avoid excessive heat generation and thus protect the filaments created as part of the SET process. On the other hand, in the RESET process, no CC is applied and a large current is intentionally generated to break the conductive filament and switch the device back to a high resistance state. Thus, for unipolar devices, the RESET current is generally larger than the SET current. Fig. 1.14(b) demonstrates the other type of RRAM devices where a stable SET operation is only possible when a positive voltage is applied across the device. On the other hand, the voltage needed for the RESET operation is always opposite to that used in the SET process. In this particular example, the RESET is only achievable when the voltage is negative. This type of device operation mode is called bipolar. As will be discussed in the following sections, the RESET current in bipolar devices is much less than that of unipolar

devices, and thus it is more power efficient. The current mainstream technology of RRAMs is based on the bipolar type of devices.

Next, based on the operation principle and device physics, RRAMs can be classified into two types of devices. One is metal cation based electrochemical metallization (ECM) RRAM. The other type is oxygen vacancies based (VO) RRAM. For ECM RRAM, metal cation generally is much more diffusive than oxygen anion or it easily precipitates such that in the on state, a conductive metallic bridge is formed. On the other hand, for VO based RRAM, oxygen anion have a significant role which is much more diffusive than metal cation. And the nature of the conductive filament in the on state is attributed to the traps formed by the VOs.

In the following sections, the ECM RRAM will be discussed first and then a description of VO based RRAM is presented. Finally, as will be shown, there are no fundamentally physical differences between those two type devices. Ergo, a unified theory can exist.

1.3.3: Introduction to Electrochemical Metallization RRAM

1.3.3.1 I-V Characteristics and Switching Mechanisms for Three Types of ECM RRAM

By definition, the operation of ECM based RRAMs typically involves the diffusion and redox electrochemical reaction of metal cations. In the on state, a metallic-like conductive filament is formed which brings the device from a high resistance state (HRS) to a low resistance state (LRS). Lots of materials have been explored for RRAM applications and even for ECM based RRAM, and thus, the material choice for each layer is not limited. Depending on origin of the metal cations, ECM RRAM can be further classified into three groups. In the first group, the metal cations are supplied only from the electrode. For examples, for system like Cu/SiO_2/Pt [26] and Cu/Ta_2O_5/Pt [27], Cu cations are from the Cu electrode. For the second group, the cations are supplied purely from the resistive layer. For example, for system like Pt/NiO/Pt [28], the Ni cations are from the NiO layer. Finally, in system like Ag/Ag_2Se/Pt [29], the metal cation Ag^+ can be supplied both from the electrode and the resistive layer. In the following sections, the operation principles for these devices are discussed.

1.3.3.1.1 I-V Characteristics of Cu/SiO2/Pt

Fig. 1.15 shows the I-V characteristics of Cu/SiO2/Pt [26]. The system starts from the initial state (a). After a positive voltage ramp is applied on the Cu electrode, Cu atoms on the Cu electrode start to get oxidized and become Cu^+/Cu^{2+} ions. Those cations, under the applied electric field, subsequently diffuse through the SiO_2 layer and get accumulated on the Pt electrode where the crystallization process occurs. This process continues until a narrow Cu filament eventually is formed. At the moment when the filament shorts the top and bottom electrode as shown (b), a significant current flows through the device and the device goes from HRS to LRS. As can be seen, generally a current compliance (CC) is imposed in this process to avoid too much current flowing through the device. Excessive current will generate significant joule heating which can damage the metallic filament. Opposite to the SET process, in the

RESET process a negative voltage is applied on the Cu electrode. Therefore, the Cu filament gets oxidized and starts to dissolve. The oxidized Cu^{+}/Cu^{2+} from the filament migrates back to the Cu electrode and gets precipitated there. As long as the filament no longer connects the top electrode, there is a sharp decrease in the current as shown in (d) and the device goes from LRS to HRS. In addition, it is noted that in this particular case the I-V shows bipolar characteristics, where the SET voltage (V_{SET}) is positive whereas the RESET voltage (V_{RESET}) is negative. Further, it is shown that the RESET current (I_{RESET}) is more or less same as the SET current (I_{SET}), i.e., around 250nA. As will be discussed later, as a general experimental observation, $I_{RESET} \approx I_{SET}$ is valid for most bipolar devices regardless of their chemical compositions and switching mechanisms.

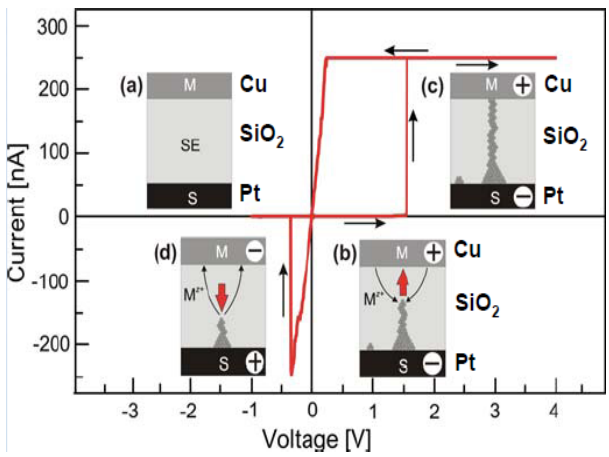


Figure 1.15. I-V characteristics of Cu/SiO2/Pt [26]

There are two major advantages of bipolar devices. First of all, since V_{SET} and V_{RESET} have different polarity, there is a larger margin between them and therefore the device reliability is typically significantly improved. Second, as mentioned above, the RESET process in bipolar devices is a combinational effect of both chemical and thermal processes, and therefore the I_{RESET} is reduced which in turn minimizes the device power consumption.

1.3.3.1.2 I-V Characteristics of Pt/NiO/Pt

Fig. 1.16 shows the I-V characteristics of a Pt/NiO/Pt device [28], which belongs to the second type of ECM devices in that the metal cations all comes from the NiO layer. In this example, inert Pt is used for both top and bottom electrodes. When electrons are injected into the NiO resistive layer by applying a voltage across the two electrodes, Ni cations can capture the electrons and thereby reduce to a metallic state. The resulting small metallic precipitate will enhance local electric field distribution and more and more precipitates will be formed; eventually a metallic filament is created which shorts both top and bottom electrodes. Therefore, the device switches from HRS to a low resistance state LRS. This process is called a forming process as indicated inside Fig. 1.16 (a). The forming process is the first cycle of the

device operation in which the device is driven to LRS from its pristine HRS state. The voltage at which the device goes to LRS is called the forming voltage (V_{Form}). For Pt/NiO/Pt, the RESET process is done mainly by thermal joule heating effect (with the help of chemical oxidation). In this operation, there is no CC limit applied on the device, and therefore a large current can flow through the metallic filament. A large current in turn can generate large heat flow which can melt the filament and make the device go to HRS again.

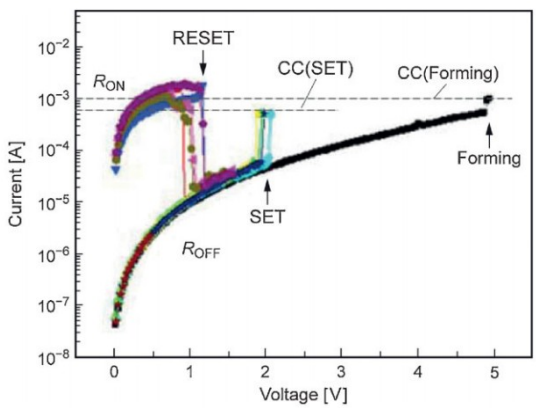


Figure 1.16. I-V characteristic of a Pt/NiO/Pt stack [28]

Russo [30] gives a detailed description of the RESET process as shown in Fig. 1.17. First of all, due to the poor thermal conductivity of the dielectric layer, the middle part of the Ni metallic filament is more likely to have higher temperature when a current is flowing. Therefore, it has a larger dissolution rate which in turn makes the filament thinner and resistance higher. The change of resistance modifies the distribution of voltage and current and a positive feedback loop is formed. The filament will eventually break and cause a sharp transition in the RESET stage. In addition, it is interesting to note that the voltage needed to turn on the device again in the following cycle is less the than V_{Form}. The reason is that unlike the first forming cycle, in the following cycles only part of filament is broken and therefore a smaller voltage is enough to turn on the device. Thus, as a general rule the V_{SET} typically is less than V_{Form} for all RRAM cells.

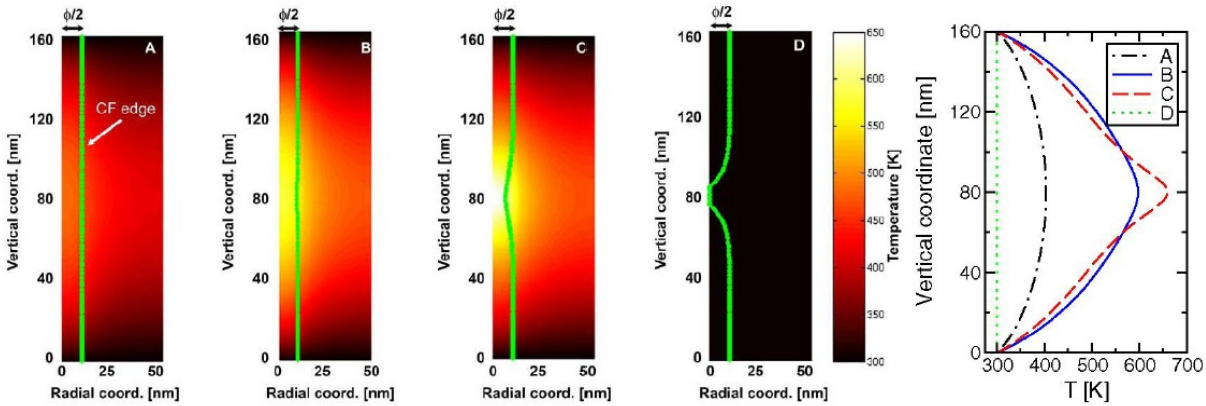


Figure 1.17. The metallic Ni filament dissolution process by joule heating in Pt/NiO/Pt stack [30].

1.3.3.1.3 I-V Characteristics of $Ag/Ag_2Se/Pt$

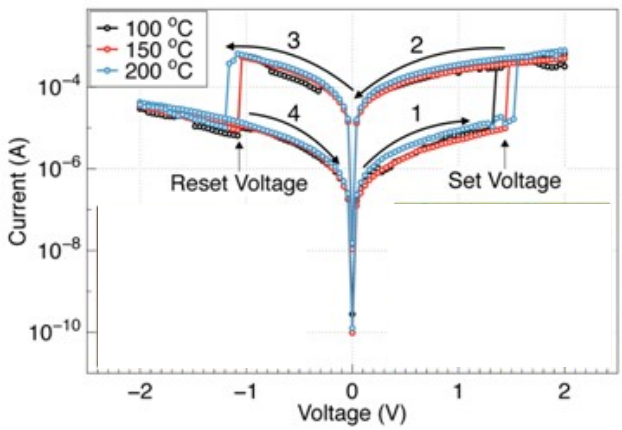


Figure 1.18. I-V characteristic $Ag/Ag_2Se/Pt$ stack [29]

Finally, Fig. 1.18 shows I-V characteristics of an $Ag/Ag_2Se/Pt$ stack, which belongs to the third group of ECM RRAMs, where the metal cation can be supplied both from the Ag electrode and from the Ag_2Se resistive layer. Such systems also include $Ag/Ag_2S/Pt$ [31] and $Ag/Ag_{33}Ge_{20}Se_{47}/Ni$ [32]. As can be seen, the I-V characteristic of $Ag/Ag_2Se/Pt$ is almost identical to that of the aforementioned $Cu/SiO_2/Pt$ stack. Both show bipolar switching characteristics. This is expected, since all the physical and chemical processes involved in the $Ag/Ag_2Se/Pt$ stack are almost identical to the processes involved in the $Cu/SiO_2/Pt$ stack. The only difference is that Ag_2Se can supply the Ag^+ ions that participate in the metallization process, whereas SiO_2 only serves as an ionic transportation medium for metal cations. Because of this Ag_2Se has a very large ionic conductivity and this causes a voltage distribution which is different from SiO_2 devices. This will be discussed more in details in section 1.3.3.3.

1.3.3.2 Comparisons Between Unipolar and Bipolar Switching Modes

As already been shown in the previous section, ECM devices can have both unipolar and bipolar operational modes. Generally, bipolar devices have two advantages over unipolar devices. The first one is that bipolar devices are more power efficient. During the RESET process, because of the combination of the thermal and electrochemical effects in bipolar devices, less I_{RESET} is needed whereas only a thermal effect is involved for unipolar devices, necessitating a larger I_{RESET}. Second, because the V_{SET} and V_{RESET} in bipolar devices have different polarity, bipolar devices have larger voltage margin, which makes their operation more reliable.

In addition, for unipolar devices, the SET-RESET competition [33] has to be avoided as shown in Fig. 1.19 [33], whereas no such problems present for bipolar devices. The SET-RESET competition is a phenomenon where either the SET or RESET processes are not stable. For example, if the SET process is not stable, once the device is driven to the SET stage it can automatically go to the RESET stage. The SET-RESET competition can be avoided by connecting a lager series resistor in the SET stage, whereas a smaller resistor is connected in the RESET operation. The details are shown in Fig. 1. 19 and are discussed below.

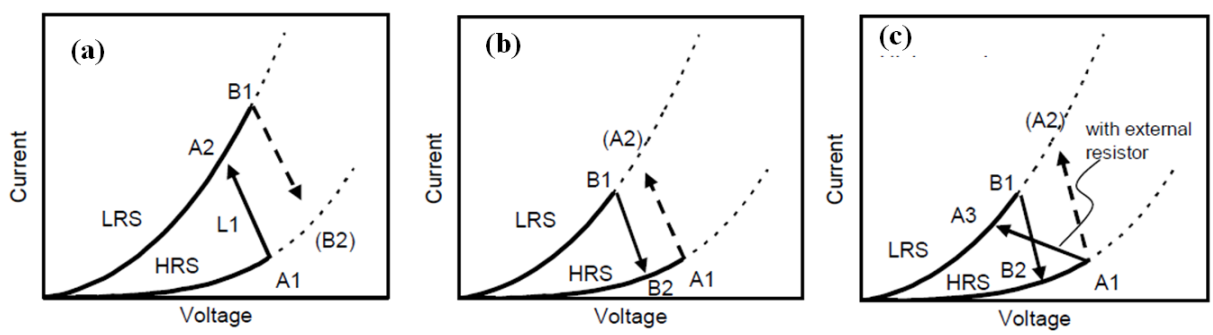


Figure 1.19. Several switching modes in unipolar RRAM (a) stable SET (b) stable RESET (c) both stable SET and RESET [33]

Fig 1.19(a) shows a situation where an identical series resistor (or no external resistor) is connected to a unipolar device in both SET and RESET stage. Its resistance is small and can be viewed as system parasitic resistance. Starting from the HRS, the device switches to LRS at point A; after that, due to the voltage drop on the parasitic resistance, the voltage on the resistive layer reduce to A2. In order to RESET the device, a larger voltage and current has to be applied and at point B1 the device switches back to HRS and stays at position B2. However, voltage at B2 is larger than SET voltage A1, and thus, the SET-RESET competition occurs. The device most likely will not stay at HRS but will switch back to LRS. This corresponds to the unstable RESET. A similar situation can occur in the SET stage as well, as shown in Fig. 1.19(b) where the RESET

process is stable and the SET stage is not stable. For a reliable operation, both SET and RESET have to be stable. The only way to achieve this goal is to connect a large series resistor (equivalent to a lower CC) during the SET stage operation and a small resistor (equivalent to a larger CC) in the RESET stage operation as shown in Fig 1.19(c). In this mode, after RESET, the voltage across the device is less than V_{SET} (as indicated by point B2 and A1) such that the RESET is stable. On the other hand, after the SET operation, the voltage across the device is less than V_{RESET} (as shown by point B1) and therefore accidental RESET is also prevented so that the SET operation is also stable. Even though adopting a smaller series resistor for RESET can stabilize the unipolar device operation, it makes the device operation more complicated.

In summary, because of the power, voltage margin and the SET-RESET competition issue, the current main technology of RRAM is based on bipolar devices.

1.3.3.3 Voltage distribution inside ECM devices

In the above discussion, it is shown that in the Cu/SiO_2/Pt stack, the SiO_2 just serves as a transportation layer for oxidized anode ions passing through from the anode to the cathode [34]. However, for the Ag/Ag_2Se/Pt stack, Ag_2Se itself consists of a large number of Ag cations that can participate in the crystallization process. They can then get replenished from the Ag electrode. In this case, the Ag_2Se layer serves as an electrolyte that has large ionic conductivity. Due to the ionic conductivity difference, the voltage distribution in Ag/Ag_2Se/Pt can be significantly different from the voltage distribution in a Cu/SiO_2/Pt stack. For Cu/SiO_2/Pt, it is possible that most of the applied voltage is dropped across the SiO_2 layer, whereas only a very small voltage is dropped across the Ag_2Se layer in an Ag/Ag_2Se/Pt stack.

In fact, depending on the conductivity of the dielectric layer and the ability of the metal cations to precipitate, the voltage distribution across the resistive layer can be classified into bulk transport limited and crystallization limited operating regimes as shown in Fig. 1.20. In the bulk transport limited regime, most of the voltage drops across the dielectric (system like Cu/SiO_2/Pt). Thus, for this type of system, it is almost impossible to get the resistive switching effect when the device thickness significantly increases (for example in the micron range). On the other hand, for systems like Ag/Ag_2S/Pt, most of the voltage drops across the electrode and dielectric interface where the electron charge transfer process occurs, and only a very small portion of voltage drops across the resistive layer. For such systems, metallic formation can be directly observed even when the dielectric becomes extremely thick (for example beyond the millimeter range).

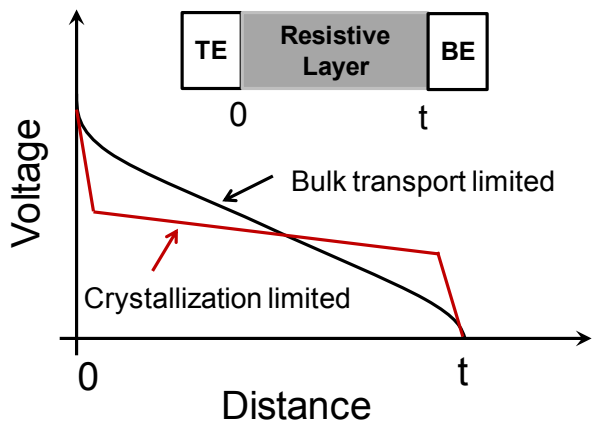


Figure 1.20. The voltage distribution across the resistive layer in bulk transport limited and crystallization limited cases

1.3.3: Introduction to Oxygen Vacancies Based RRAM

For ECM based RRAMs, metal cation has an important role in the formation of conductive bridge. On the other hand, when metal cations are less diffusive or very difficult to precipitate, O^{2-} anions may have the dominant effects on device performance. Unlike ECM based RRAM, the physics of VO based devices are quite complicated and there are still lots of unknowns. Despite the vagueness, several basic mechanisms and theories still exist which can at least explain part of the device behaviors.

1.3.3.1: The Switching Mechanisms and Models for Oxygen Vacancies Based RRAM

For VO based RRAMs, VO has an important impact on the conductive bridge formation process inside the memory cell. In these systems, VO is typically more diffusive than metal cations. In addition, the SET and RESET operations typically involve VO generation and annihilation. When the VO concentration is large, the cell has low resistance. On the other hand, when the VO concentration is low, the cell resistance increases. Thus, the key to understand VO based RRAMs is to understand how the VO affects the conductivity of the memory cell and how the VO concentration can be modulated by applied voltage.

For low and moderate bandgap materials like TiO_2 and Ta_2O_5, VO typically acts as an n-type dopant [35]. Thus, when the VO concentration increases, the electron concentration also increases. On the other hand, for large bandgap materials like HfO_2 and ZrO_2, VO typically acts as traps. In this case, electron hopping conduction becomes important. In the following sections both conduction mechanisms are discussed.

1.3.3.1.1: Oxygen Vacancies Act as Dopants

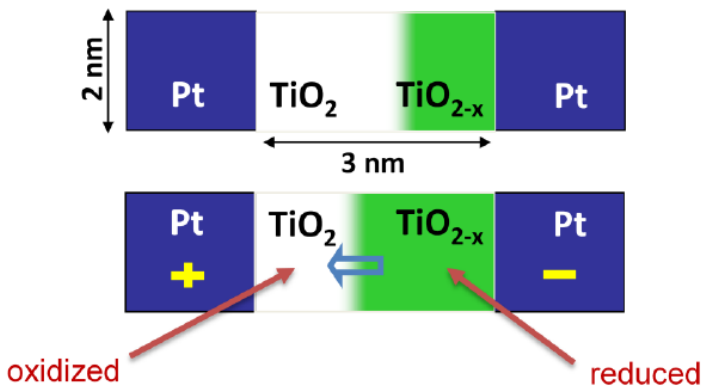


Figure 1.21. switch mechanism of $Pt/TiO_2/Pt$ cell [2]

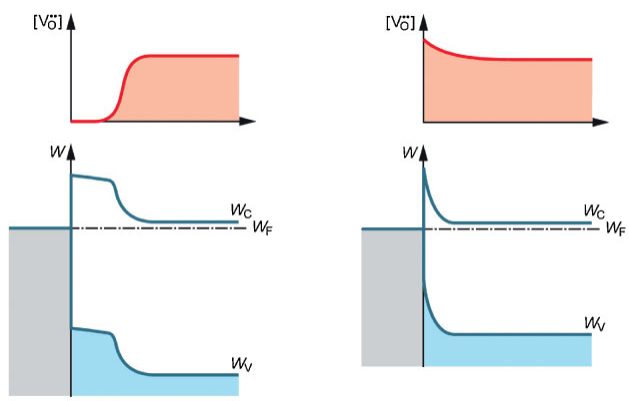


Figure 1.22. The VO distribution and the band diagram for $Pt/TiO_2/Pt$ system [25]

TiO_2 is one of the most commonly studied materials for VO based RRAMs. Fig. 1.21 shows the evolution of the chemical composition inside the TiO_2 layer when a voltage is applied on a $Pt/TiO_2/Pt$ stack [2]. Fig. 1.22 shows the distribution of VO and the band diagram for such a system. As can be seen, when a positive voltage is applied on the left side at the Pt electrode, O^{2-} ions migrate towards the left under the electric field, as indicated by the arrow and leave the VOs behind. Hence TiO_2 becomes TiO_{2-x} (green region in Fig. 1.21). Since VOs are n-type dopants for materials having moderate bandgap such as TiO_2 [35], the increase of VO

concentration can make the dielectric more electronically conductive. This is demonstrated in Fig.1.22. As shown when the VO concentration increases, the energy difference between the Fermi level and the conduction band decreases, which indicates an increase in the electron concentration. For this reason, the region containing TiO_{2-x} is also called a virtual electrode. As this process continues, the width of the tunneling barrier formed between the left Pt electrode and the virtual electrode decreases. When the tunnel barrier becomes thin enough; there will be a significant current flowing through the device. As a result, the device switches from HRS to LRS.

During the RESET process, a negative voltage is applied on the left electrode which pushes O^{2-} ions back into the TiO_x film and subsequently recombines them with VOs. Thus, the VO concentration decreases at the interface between the left Pt electrode and the TiO_x layer. Therefore, the width of the tunneling barrier increases, which leads to a significant increase in the cell resistance.

1.3.3.1.2: Oxygen Vacancies act as traps

The above theory works well for a dielectric with moderate bandgap. However, for large bandgap dielectric material such like HfO_2 and ZrO_2, VO doesn't act as a dopant but rather behaviors like trap. In this case other theories are developed [36, 37]. One of the frequently used models is so called Ion-Transport-Recombination model [37] shown in Fig. 1.23.

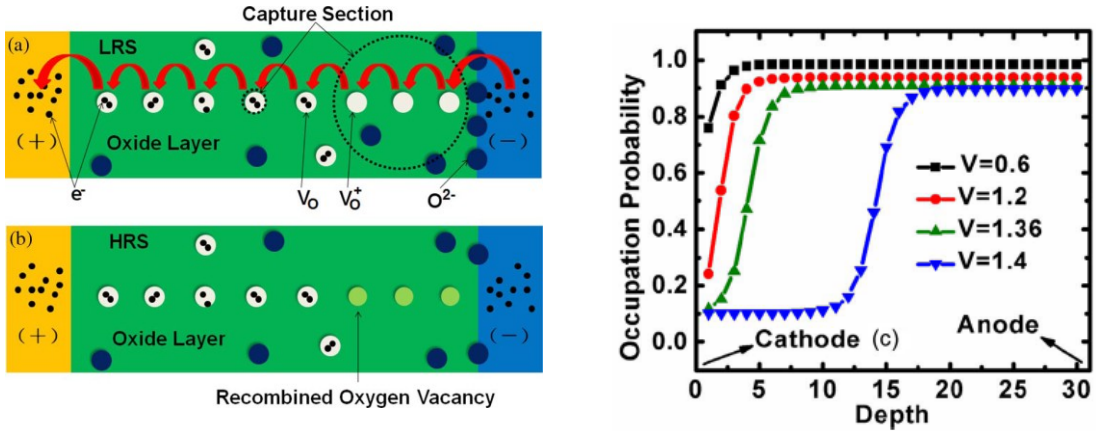


Figure.1.23. (Left) Illustration of electron transport in LRS and HRS (right) illustration of electron occupation probability at different VO sites in HRS [37].

In this model, the LRS is represented by the percolation through a series of VOs denoted by the white circles in the left figure. The conduction current is due to the electron hopping through those VO sites. In addition, during the forming and SET process a positive voltage is applied on the electrode (shown in blue) which causes a large amount of O^{2-} anions to flow

towards the electrode and subsequently get adsorbed there. During the RESET process, the electron occupation probability will significantly decrease at the anode side assuming that the electron hopping rate between VO sites are much faster than the hopping rate between VO and electrode. As a result, the VO carries net positive charge which can then combine with the O^{2-} ions released from the cathode (shown in blue). Therefore, the original dielectric structure is restored and a gap is created between the filament and electrode which leads a significant current reduction. As a result, the device switches from LRS to HRS.

This model is simple yet very effective in explaining certain phenomena. The key process in this model involves the flow O^{2-} ions in and out from the blue colored electrode such that the electrode can temporarily store the O^{2-} during SET process yet release the O^{2-} in the RESET process. In the other words, the electrode serves as an oxygen reservoir. In reality, both Ti and TiN are good O^{2-} reservoirs [38, 39].

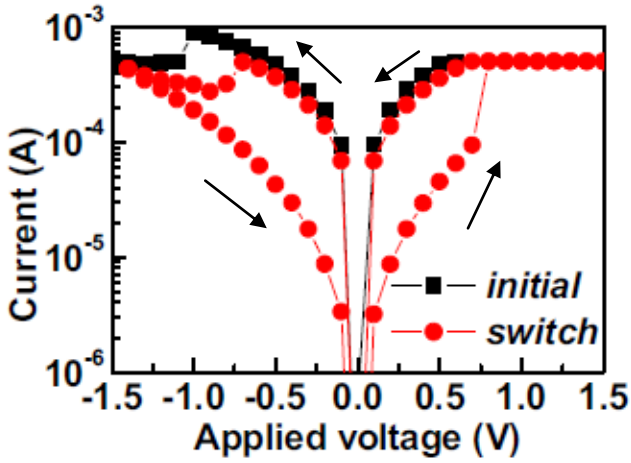


Figure 1.24. The I-V characteristics of a Ti/HfO_2/TiN stack [40]

As an example, Fig. 1.24 [40] shows the I-V characteristics of a Ti/HfO_2/TiN stack. During the measurement, the TiN electrode is always grounded. When a positive voltage is applied on the Ti electrode, O^{2-} ions migrate towards the Ti electrode and get adsorbed there, leaving the VO behind. The VO concentration inside the HfO_2 continues to increase as the voltage increases until a percolation threshold reaches. At around 0.7V there is a sharp increase of current and at this moment, the device is turned on. In the RESET process, a negative voltage is applied on the Ti electrode. Due to the combined effects of joule heating and the presence of the electric field, the adsorbed O^{2-} ions on the Ti electrode start to diffuse back into HfO_2 layer, where they recombine with VOs. Therefore, the trap density inside the HfO_2 decreases. The current gradually decreases and the device switches from LRS to HRS again. It is worthwhile to note that in this case, the device shows bipolar switching characteristics and in addition, the RESET process is a gradual process, which means the conductance of the cell decreases slowly as the

V_{RESET} increases. This is quite different from the ECM cells, which RESET rapidly. This gradual RESET process has important impacts on the multilevel-programming capability of the device as will be discussed in the following chapters.

1.3.3.2: The Atomic Filament Structure of VO Based RRAMs

In the above discussion, the essence of resistive switching is attributed to the modulation of the O^{2-} flow inside the oxide. This picture gives an effective way to describe the overall behavior of the device. However, in reality the migration of O^{2-} typically triggers a phase transition inside the metal oxide such that a transition between an insulating phase and a more conductive phase usually occurs. The model described above is oversimplified and cannot describe the real physical processes happening inside the metal oxide in a mesoscopic scale.

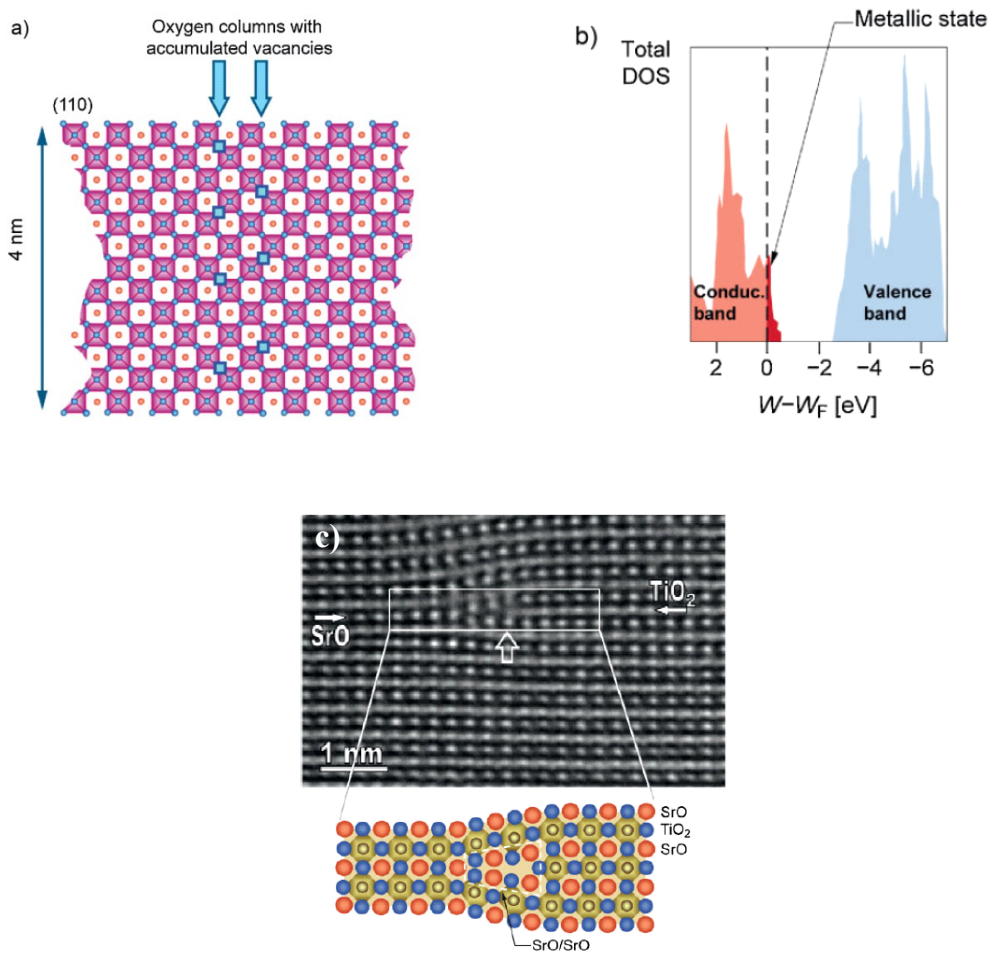


Figure 1.25. (a) The VO accumulation along lines in SrTiO3 [25] [41] (b) Calculated density of states close to the defect region [18] (c) Illustration of lattice structure close to the core of edge dislocations in SrTiO3[25] [42].

As stated above, VOs typically act as n-type dopants in a moderate bandgap metal oxide. As the VO concentration exceeds the equilibrium value, point defects start to interact with each other and form extended defects such as vacancy chains [25]. They can also get accumulated at the cores of edge or screw dislocations [25]. Regardless of the detailed structures of those extended defects, one of the common features is that at the region close to the extended defects a metal rich region is formed and metal cations get reduced. At the same time the band gap shrinks and the Fermi level increases, which lead to an increase of the local electronic conductivity.

Fig. 1.25 illustrates several possible crystal extended defects. In Fig. 1.25 (a) [25, 41] VOs are accumulated along lines in a confined region. Through first principles simulation, it is found that along this defect region, the electrons around Ti ions get delocalized, resulting in a local insulator to metal transition [25]. Fig. 1.25 (c) shows another possibility that VOs are trapped in the core of the edge dislocation in $SrTiO_3$ [25, 42]. In addition, researchers [43] found that compared with the bulk region, VOs can have a dramatic increase of diffusivity along those dislocations. As a result, those dislocations can act as fast transport paths for O^{2-} anions [25] and can form a percolation path which shorts two electrodes to create a LRS inside RRAM cell.

Besides the extended defects, it is reasonable to believe that as VO concentration further increases, with the help of the joule heating during the forming or SET processes, metal oxides can transfer to other more conductive and metal rich phases or suboxides. Researchers [44] have found that Ti_4O_7 is formed in LRS inside the TiO_2 layer for a Pt/TiO2/Pt RRAM cell as shown in Fig. 1.26, by using high resolution TEM. The transition between LRS and HRS is attributed to the creation and rupture of this conductive phase. The observed diameter of those filaments is around 5nm to 10nm. Similarly, in the SET stage, a more conductive TaO_2 phase is observed for RRAMs based onTa_2O_5 [45].

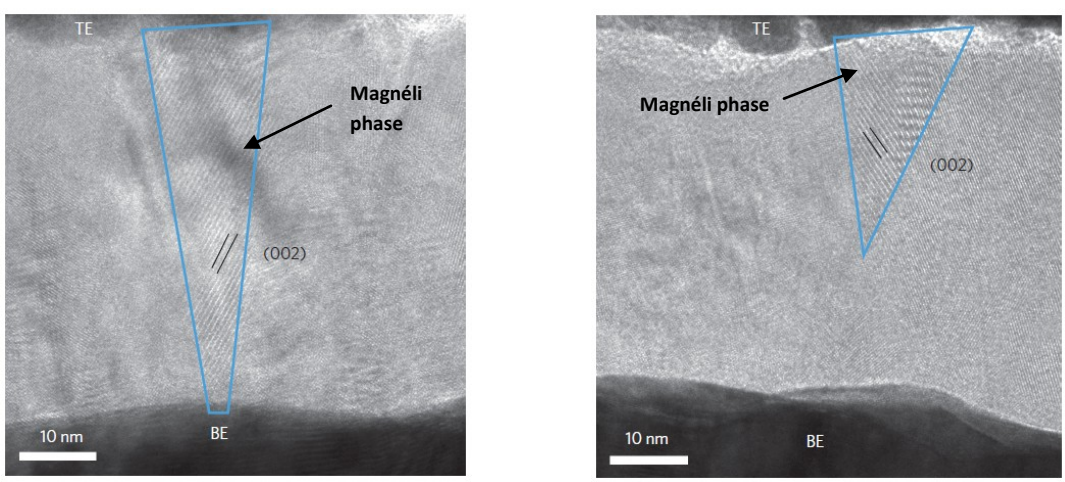


Figure 1.26. TEM image of Magnéli phase after the forming process inside of Pt/TiO_2/Pt cell (Left) and a discontinued Magnéli phase (Right) [44]

1.3.4: The Equivalence Between ECM and VO Based RRAM

Even though in the above sections, RRAM devices are intentionally classified into metal cations based ECM and VO based devices, there are really no essential differences between them. Rather, they are two different ways to describe the same physical and chemical processes inside the metal oxide cells:

- In the SET stage, this is a locally increasing metal element concentration and at same time reducing metal cations process such that metal rich region (or suboxide) can be formed.
- In the RESET stage, this is a locally decreasing metal element concentration and oxidizing the metallic species or cations such that more insulating oxide can be formed locally.

One slight difference between those two systems may be the difference in the diffusivity between metal cations and O^{2-} anions. For system like $Cu/SiO_2/Pt$, Cu cations have larger mobility than O^{2-} anions. On the other hand for $Ti/HfO_2/Pt$, O^{2-} has the larger diffusivity than Hf or Ti cations. Regardless of the difference in diffusivity, for both cases in LRS metal rich region is formed and metal cations get reduced. For example, for $Cu/SiO_2/Pt$, Cu^{1+} or Cu^0 rich region is formed inside the SiO_2 because of the migration, accumulation and subsequently reduction of Cu^{2+} or Cu^{1+} element coming from copper electrode and for $Ti/HfO_2/Pt$ stack, Hf^0, Hf^{1+}, Hf^{2+} and Hf^{3+} rich region is formed inside HfO_2 due to the depletion of local O^{2-} and the reduction of Hf^{4+}. Therefore, besides the difference in diffusivity there is no essential difference between those systems.

Depending on the detailed chemical composition, the metal rich region formed in the SET stage can have either metallic, semiconducting or electron hopping based conduction features. And their conductivities vary. For a more metallic type of filament, it has a positive temperature resistance coefficient due to the increased phonon scattering from lattice; therefore, the resistance of the device increases as the temperature increases. For filaments having semiconducting or electron hopping based conduction features, a negative temperature resistance coefficient is observed; therefore, the conductivity increases as the temperature increases. And a typical system would be like $Ti/HfO_2/Pt$.

1.4: Performance and Scaling Perspective of RRAM

In the previous section, the basic physics of RRAM are described. In this section, the performance and scaling perspective of RRAM are comprehensively discussed. First, the performances of the latest reported RRAM devices are shown and the discussion focuses more on endurance, retention, switching speed. Next, special attention is paid to the scaling issues, with a focus on the memory array structure, selector device and power consumption. Finally, the necessity of multi-level programming is discussed.

1.4.1: RRAM Performance

Table 1.3 summarizes the major achievements in recent RRAM technology developments [46]. The first row indicates the switching type. As can be seen most of the devices show the bipolar switching feature. As discussed earlier, bipolar device generally has less RESET current than unipolar device and therefore is more power efficient.

Table 1.3. RRAM performance [46]

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

The next two entries are device structure and cell area which both are related to scaling properties of RRAM and will be discussed in next section. Here, several important device performance characteristics such as speed, endurance, retention and HRS/LRS ratio are introduced. Speed is the time needed to program and erase the memory device. In this case, it is the minimum time required to establish and annihilate the conductive filament inside RRAM.

Nowadays the demonstrated speed of RRAM is less than 10ns which is much faster than that of Flash (in tens of µs range). In addition, the operating voltage for a typical Flash is at least larger than 5 volts, whereas only less than 2V is required for RRAM.

One of the problems for RRAM is the RESET current. Typically the peak RESET current for RRAM is around 10 to 100µA. On the other hand, for Flash technology, even though the program and erase voltage are larger than that of RRAM, since the programming and erasing are based on FN tunneling current the magnitude of the current is extremely small. Thus, the overall power consumption for Flash is very small. However, for RRAM this problem can be alleviated if the switching time can go down to the nanosecond range so the overall energy consumption is small for one switching cycle. In addition as will be discussed further, a reduction of SET current will also lower the RESET current making RRAM power efficient and suitable to build memory array.

One of the most significant advantages of RRAM is of its superior endurance. Researchers have demonstrated endurance of more than 10^{12} cycles based on TaOx devices [47], which is much larger than typical Flash endurance of 10^4 to 10^6 cycles. Such large endurance makes the universal type of memory possible. Finally, another important metric for memory device is retention and it indicates how long the memory cell can hold the data. Typically a requirement of 10 years retention time is required for non-volatile device at room temperature. Fortunately, most RRAM devices demonstrated so far have such long retention characteristics. In a retention measurement, the test is done at elevated temperature such as 80ºC, 150ºC and 200ºC. Then the result is extrapolated to the room temperature to get the estimation. Such measurement shortens the test period significantly.

From above discussion, it can be seen that RRAM has most of the advantages as a memory device. Yet, there is still a critical factor that has to be considered for it to be feasible for future memory applications. This is the capacity of scaling. As discussed earlier, planar Flash is very difficult to scale down to 20nm. Even though in recent years, people have suggested 3-D flash structures [48], it is extremely difficult to scale to 11nm [48]. Thus, the chance of RRAM being successful in the future strongly depends on its scaling ability.

1.4.1: RRAM Scaling

In this section, the scaling perspectives of RRAM are discussed. For scaling, there are several issues that have to be considered. First, how small can a single RRAM device be made, since this determines the ultimate scaling limit for RRAM device? Second, is it possible to integrate RRAM devices into a memory array while still maintaining the high density features of the same? Additionally, is the power density of the RRAM memory array acceptable for today's low power electronics applications? Finally, is it possible for RRAM to achieve multi-level programming, which can effectively increase the storage density? In this section, these three issues will be discussed.

Currently, the smallest RRAM device has been demonstrated by Imec [49]; the researchers fabricated 10nm×10nm Hf/HfO_x Crossbar structure RRAM. The resulting device satisfies most of

the memory requirements. For example, the switching speed is less than 10ns, endurance is more than 10^7 cycles, on/off ratio is larger than 50, and the retention time can reach 30 hours at 200°C.

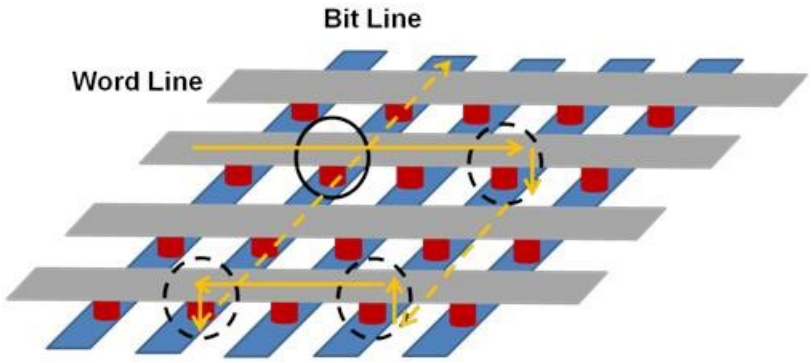


Figure 1.27. Schematic of 2D array structure and leakage path problem.

Even though the size of single RRAM can be made smaller than 10nm by 10nm, the array structure which can efficiently integrate all the single cells together must be considered. In order to obtain a minimum area size of $4F^2$, where F is the smallest feature size at a given technology node, the crossbar structure must be adopted as shown in Fig. 1.27 Unfortunately, this structure alone has the sneak path problem and cannot be directly adopted for application.

The leakage path problem is illustrated in Fig.1.27. Suppose the memory cell we want to access (circled by solid black line) is in HRS, therefore it is expected that a low output current will flow through the corresponding WL and BL. However, if other adjacent cells (circled by the dashed black lines) are in LRS, a leakage path as indicated by the yellow arrow will form, which will cause a significant amount of current to flow through the addressed WL and BL and create an error.

In order to suppress the leakage path problem, a selector has to be attached to each device such that only the selected device is able to get accessed. As shown in Table 1.3, 1T-1R structure is commonly used for this purpose. Here, 1T stands for one transistor and 1R represents one RRAM cell. Therefore, 1T-1R structure utilizes a transistor as the selector device.

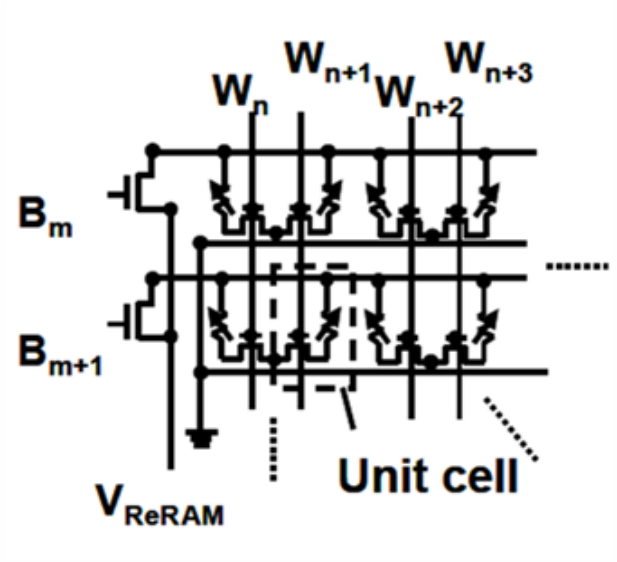


Figure1.28. Schematic of 1T-1R structure [50].

Fig.1.28 [50] shows a typical 1T-1R structure. In this configuration, RRAM cells (indicated by arrow and the resistor symbols) are connected in series with transistors. Even though this structure can effectively eliminate the sneak path problem, it significantly decreases the memory array density since the transistor is very chip area consuming. To approach the targeted $4F^2$ size for a memory unit cell, the selector has to be stacked on top of RRAM structure and the size of selector has to be comparable to that of RRAM.

Based on above discussion, a diode could be a good choice. However, the traditional diode can only pass voltage and current signals in one direction. Thus, only unipolar devices can adopt diode based selectors. However, as discussed in previous sections, unipolar RRAM devices typically consume much more power than bipolar device. Thus, it is necessary to have a bidirectional diode-like device which can serve as the selector for bipolar devices. Recently, several groups have demonstrated such selectors [51, 52]. Fig. 1.29 [51] demonstrated a bidirectional diode-like selector based on Ni/TiO_2/Ni. Due to the symmetric structure, the Schottky barrier is identical from either direction. As a result, a symmetric Schottky emission current can be obtained as shown in the left figure. In addition this stack is readily combined with Ni/HfO2/Pt cell and the final I-V characteristic is shown in the right figure. During the reading process, the selected WL and BL are grounded and connected to VDD correspondingly with all other lines floated. As has been demonstrated, such a structure is able to achieve maximum memory array size [51].

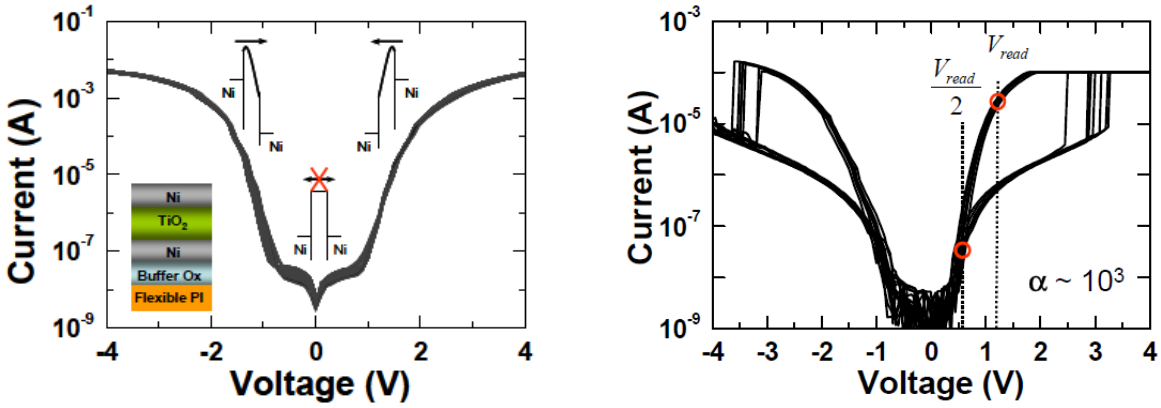


Figure 1.29. A bidirectional selector based on Ni/TiO_2/Ni (left) and the overall performance of a bidirectional selector integrated with Ni/HfO2/Pt RRAM (right) [51]

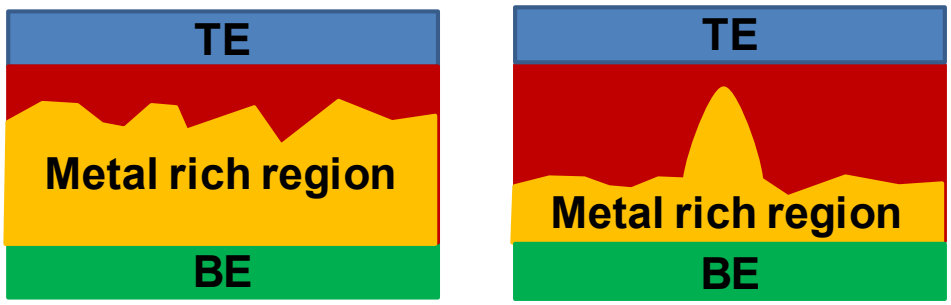


Figure 1.30. Two scenarios in the HRS. (left) a homogenous conductive region (right) a filamentary conductive region.

The power consumption is another issue for an RRAM memory array. To be feasible for use in a large-scale array structure, I_{RESET} has to be less than 100µA. Typically, for various types of RRAMs, if the parasitic capacitance is eliminated I_{RESET} linearly scales down with I_{SET} and they satisfy the relationship of $I_{RESET} \approx I_{SET}$ [53]. Thus, for low power applications a low I_{SET} is required. However, a lower I_{SET} generally results in a larger R_{on} which may reduce the on/off resistance window. Some research results [54] show that the R_{off} can effectively increase as the device size scales down, which can help maintain the programming window as R_{on} increases. Yet, there are still controversies regarding this, since other results [55] indicate an independence of R_{off} on device area. The dependence of R_{off} on device area really depends on the material properties and detailed switching process of a device. Even though R_{on} of RRAM is widely accepted as filamentary type conduction and thus independent of device area, once the filament is ruptured in the RESET process the off state can have two scenarios. As shown in Fig. 1.30, one

possibility is that the remaining filament (metal rich region or virtual electrode) is homogenously distributed throughout the oxide layer, and in this case R_{off} increases as device area scales down. Other the other hand, if the remaining filament has a topography like the one shown in the right figure, where a dominant filament remains in the oxide, R_{off} can be independent of device size.

Finally, another important property of RRAM to consider is its multilevel programming (MLP) capability. This directly determines if RRAM is a viable replacement for the current Flash technology, since Flash has well established MLP ability. Given the same device area, a device with MLP capability effectively has larger bit storage density. Generally, there are two ways to achieve MLP capability for RRAM cell. One is to adjust the current compliance (CC) during the SET process. One problem with this approach, however, is the capacitive discharge effect [53]. The discharging current can cause large variations in R_{on}. Even though the 1T1R structure can effectively eliminate this effect [53], it significantly reduces the RRAM array density. On the other hand, for a diode based selector structure, the capacitive discharging effect still exists.

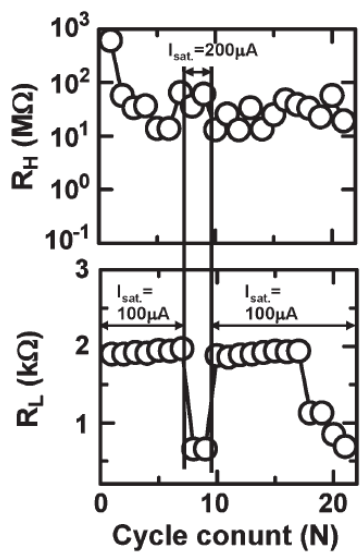


Figure 1.31. The history dependent instability of R_{on} [55]

Another problem to this approach is the history dependent R_{on} instability problem [55] as shown in Fig 1.31. At the beginning, the CC is set to 100μA, and a R_{on} of 2KΩ is obtained. Next, when the CC increases to 200μA, R_{on} decreases to 500Ω as expected. After that, if the CC reduces back to 100μA again, the R_{on} also increases to 2kΩ as before. However, as cycling goes on, the R_{on} tends to decrease to 500Ω even if the CC is still maintained at 100μA. Obviously this history dependent R_{on} instability can cause severe programming errors.

The second way to achieve MP is through the RESET process. Through this approach, there is no capacitive discharging effect involved and it is also a voltage-controlled operation and easy to implement. However, for this technique to be feasible, it would be ideal that the I-V transition in the RESET stage is smooth enough that the off-stage resistance (R_{off}) can gradually increase with applied voltage. In the later part of this dissertation, a multilevel programming cell based on amorphous yttria stabilized zirconia (YSZ) will be studied.

1.5: This Work

This dissertation mainly contains two parts. The first part is a Kinetic Monte Carlo simulation study on the ECM type of RRAM. In the second part, experimental work is carried out for both ECM and VO based RRAM. The material used in this study is yttria stabilized zirconia. In addition, a programming methodology used to improve device uniformity, reliability and endurance is also studied.

Chapter 2 describes the physical and chemical processes inside an ECM based RRAM. A Kinetic Monte Carlo based simulator is built. Through this simulator, several physical phenomena are studied. This simulator is mainly used to simulate the electroforming and SET process inside such a device. In addition, experiments based on Ag/Ag_2S and Cu/H_2O are conducted to verify the simulation results.

Chapter 3 studies RRAM memory properties based on Ti/YSZ and Cu/YSZ devices. A comprehensive study on the mechanism and device resistive switching properties is carried out. The underlining physics is discussed. Finally, noise measurements are conducted to compare the reliability between Ti/YSZ and Cu/YSZ device.

Chapter 4 investigates the multi-level programming capability of Ti/YSZ in the RESET stage. Also the impact of the programming methodology on the device uniformity, reliability and endurance are investigated. It is found that by adopting incremental step pulse programming technique, a significant improvement in device endurance and reliability can be obtained.

Chapter 5 is the summary and describes possible future work.

1.6: References

[1] http://www.archivebuilders.com/whitepapers/22045p.pdf "Cost of Semiconductor RAM over 50 years: 1970 to 2020".

[2] Assessment of the Potential & Maturity of Selected Emerging Research Memory Technologies Workshop & ERD/ERM Working Group Meeting (April 6-7, 2010)

[3] http://en.wikipedia.org/wiki/File:ComputerMemoryHierarchy.svg

[4] "Process integration, devices and structure" *International technology roadmap for semiconductors* 2011 Edition

[5] H.-S. Philip Wong, Simone Raoux, SangBum Kim, Jiale Liang, John P. Reifenberg, Bipin Rajendran, Mehdi Asheghi, and Kenneth E. Goodson "Phase change memory" *Proceedings of IEEE* Vol.98, No. 12, December 2010

[6] F. Pellizzer, A. Benvenuti, B. Gleixner, Y. Kim, B. Johnson, M. Magistretti, T. Marangon, A. Pirovano, R. Bez, G. Atwood, *Symp. VLSI Technology*, 2006, 150.

[7] M. Boniardi, A. Redaelli, A. Pirovano, I. Tortorelli, D. Ielmini, and F. Pellizzer, "A physics-based model of electrical conduction decrease with time in amorphous Ge2Sb2Te5", *Journal of Applied Physics*, vol. 105, no. 8, 084506, 2009.

[8] B.-J. Bae, S. Kim, Y. Zhang, Y. Kim, I.-G. Baek, S. Park, I.-S. Yeo, S. Choi, J.-T. Moon, H.-S. P. Wong, and K. Kim, "1D thickness scaling study of phase change material (Ge2Sb2Te5) using a pseudo 3-terminal device," in *Proc. IEEE International Electron Devices Meeting*, pp. 93–96, 2009

[9] Fabrizio Buscemi, Enrico Piccinini,Rossella Brunetti,Massimo Rudan, and Carlo Jacoboni "Monte Carlo simulation of charge transport in amorphous chalcogenides" *Journal of Applied Physics* 106, 103706, 2009.

[10] S.Ahn et.al, "Write current reduction in transition metal oxide based Resistance change Memory" *Advanced Materials* 2008, 20, 924–928

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

[12] Paolo Pavan, Roberto Bez, Piero Olivo, and Enrico Zanoni "Flash Memory Cells—An Overview" *Proceedings of the IEEE*, vol. 85, No. 8, August 1997.

[13] M. Lezlinger and E. H. Snow, "Fowler-Nordheim tunneling into thermally grown SiO2," *Journal of Applied Physics*, vol. 40, no. 1, pp.278–283, 1969.

[14] "Memory Technology Overview" EE231 Lecture notes, Section 8, Vivek Subramanian

[15] A. Tal: Two Technologies Compared: NOR vs. NAND White Paper http://139.138.48.19/pdf/NAND/MSystems/MSystems_NOR_vs_NAND.pdf

[16] P. Cappelletti, C. Golla, P. Olivo and E. Zanoni "Flash Memories" Kluwer academic publishers, second printing 2000.

[17] International Technology Roadmap for Semiconductors (ITRS), ITRS 2005 edition, http://www.itrs.net/reports.html

[18] Matt Marinella, "How Advances in Nanotechnology Will Enable the Next Generation of Exascal Computers", 8th Annual Full Day Symposium Emerging Non-Volatile Memory Technologies, April, 6, 2012.

[19] P.H. Nielsen and N.M. Bashara "The Reversible Volage-Induced Initial Resistance in the Negative Resistance Sandwich Structure" *IEEE Transactions on Electron Devices*, Vol 11, May 1964, P.243.
[20] J. G. Simmons and R. R. Verderber, "Newconduction and reversible memory phenomena in thin film insulating films" Proc. R. Soc. A., vol. 301, pp. 77–102, 1967.
[21] C. J. Varker and E. M. Juleff "Electron beam recording in SiO2 with direct read-out using the electron beam induced current at a p-n junction" *Proceedings of IEEE*, vol. 55, no. 5,pp. 728–729, 1967.
[22] D. J. Wouters "Oxide Resistive RAM (OxRRAM) for scaled NVM application" IMST 2009 Tutorial 9 Sept, 2009.
[23] W. W. Zhuang, W. Pan, B. D. Ulrich, J. J. Lee, L. Stecker, A. Burmaster, D. R. Evans, S. T. Hsu, M. Tajiri, A. Shimaoka, K. Inoue, T. Naka, N. Awaya, K. Sakiyama, Y. Wang, S. Q. Liu, N. J. Wu, and A. Iganatiev "Novell colossal magnetoresistive thin film nonvolatile resistance random access memory (RRAM)" *Tech. Dig. International Electron Devices Meeting*, San Francisco, CA, 2002, pp. 193–196.
[24] I. G. Baek, M. S. Lee, S. Seo, M. J. Lee, D. H. Seo, D.-S. Suh, J. C. Park, S. O. Park, H. S. Kim, I. K. Yoo, U.-I. Chung, and J. T. Moon, "Highly scalable non-volatile resistive memory using simple binary oxide driven by asymmetric unipolar voltage pulses" in *Tech. Dig. Int. Electron Devices Meeting*., San Francisco, CA, 2004, pp. 587–590.
[25] R. Waser, R. Dittmann, G. Staikov and K. Szot , " Redox-Based Resistive Switching Memories – Nanoionic Mechanisms, Prospects and Challenges", *Advanced Materials* ,2009, 21, page 2632.
[26] C. Schindler, G. Staikov and R. Waser "Electrode kinetics of Cu-SiO_2-based resistive switching cells: overcoming the voltage-time dilemma of electrochemical metallization memories," *Applied Physicas Letters*, 94. 072109. 2009.
[27] T. Sakamoto, et al., "A Ta_2O_5 solid-electrolyte switch with improved reliability," *Symposium on VLSI Technology Digest,* 2007, pp. 38–39.
[28] J.-B. Yun, S. Kim, S. Seo, M.-J. Lee, D.-C. Kim, S.-E. Ahn, Y. Park, J. Kim, H. Shin, "Random and localized resistive switching observation in Pt/NiO/Pt" *Phys. Status Solidi* (RRL) 2007, 1, 280
[29] Jaewon Jang, Feng Pan, Kyle Braam, and Vivek Subramanian "Resistance Switching Characteristics of Solid Electrolyte Chalcogenide Ag_2Se Nanoparticles for Flexible Nonvolatile Memory Applications" *Advanced Materials,* Volume 24, Issue 26, pages 3573–3576, July 10, 2012
[30] U. Russo, D. Ielmini, C. Cagli, A. Lacaita, S. Spic:ga, C. Wiemer, M. Perego, M. Fanciulli, IEEE Technical Digest International Electron Devices Meeting, 2007, 775.
[31] K. Terabe, T. Hasegawa, T. Nakayama and M. Aono, "Quantized conductance atomic switch," *Nature*, vol.433, 2005, pp. 47–50.
[32] M. N. Kozicki, M. Park and M. Mitkova, "Nanoscale memory elements based on solid-state electrolytes," *IEEE Transactions on Nanotechnology*. Vol.4 No.3. 2005, pp. 331–338.
[33] Y. Hosoi, Y. Tamai, T. Ohnishi, K. Ishihara, T. Shibuya, Y. Inoue, S. Yamazaki, T. Nakano, S. Ohnishi, N. Awaya, I. H. Inoue, H. Shima, H. Akinaga, H. Takagi, H. Akoh, and Y. Tokura, "High speed unipolar switching resistance RAM (RRAM) technology," *Tech. Dig. Int. Electron Devices Meeting*, 2006, p. 30.7.

[34] F. Pan and V. Subramanian, “A kinetic Monte Carlo study on the dynamic switching properties of electrochemical metallization RRAMs during the SET process,” in *Proc. Int. Conf. Simulation of Semiconductor Processes and Devices (SISPAD),* 2010, pp. 19–22.
[35] Knauth, P. & Tuller, H. L. “Electrical and defect thermodynamic properties of nanocrystalline titanium dioxide” *Journal of Applied Physics*, 85, 897–902 , 1999.
[36] Vandelli, L.; Padovani, A.; Larcher, L.; Broglia, G.; Ori, G.; Montorsi, M.; Bersuker, G.; Pavan, P.; , "Comprehensive physical modeling of forming and switching operations in HfO2 RRAM devices," Electron Devices Meeting (IEDM), 2011 IEEE International , vol., no., pp.17.5.1-17.5.4, 5-7 Dec. 2011
[37] Bin Gao; Bing Sun; Haowei Zhang; Lifeng Liu; Xiaoyan Liu; Ruqi Han; Jinfeng Kang; Bin Yu; , "Unified Physical Model of Bipolar Oxide-Based Resistive Switching Memory," Electron Device Letters, IEEE , vol.30, no.12, pp.1326-1328, Dec. 2009
[38] M. Fujimoto, H. Koyama, M. Konagai, Y. Hosoi, K. Ishihara, S. Ohnishi, and N. Awaya, “TiO2 anatase nanolayer on TiN thin film exhibiting high-speed bipolar resistive switching” *Applied Physicas Letters,* 89, 223509 (2006).
[39]Chih-Yang Lin; Chen-Yu Wu; Chung-Yi Wu; Tzyh-Cheang Lee; Fu-Liang Yang; Chenming Hu; Tseung-Yuen Tseng; , "Effect of Top Electrode Material on Resistive Switching Properties of Film Memory Devices," *Electron Device Letters, IEEE* , vol.28, no.5, pp.366-368, May 2007
[40] Yu-Sheng Chen, Tai-Yuan Wu1,Pei-Jer Tzeng, Pang-Shiu Chen, Heng-Yuan Lee
Cha-Hsin Lin, Frederick Chen, and Ming-Jinn Tsai “Forming-free HfO2 Bipolar RRAM Device with Improved Endurance and High Speed Operation2 *Tech. Dig. Int. Electron Devices Meeting* p37, 2009.
[41] C.L. Jia, M. Lentzen, K. Urban “Atomic-Resolution imaging of oxygen in perovskite ceramics” 7 FEBRUARY 2003 VOL 299 *SCIENCE*
[42] C. L. Jia, L. Houben and K. Urban “Atom vacancies at a screw dislocation core in SrTiO”3, *Philosophical Magazine Letters*, Vol 86, page 683, 2006
[43] K. Szot, W. Speier, R. Carius, U. Zastrow, and W. Beyer “Localized Metallic Conductivity and Self-Healing during Thermal Reduction of SrTiO3” *Physical review Letters*, 18 February 2002.
[44]D. H. Kwon, K. M. Kim, J. H. Jang, J. M. Jeon, M. H. Lee, G. H. Kim, X. S. Li, G. S. Park, B. Lee, S. Han, M. Kim, and C. S. Hwang, “Atomic structure of conducting nanofilaments in TiO2 resistive switching memory” *Nature Nanotechnology*, vol. 5, pp. 148–153, Feb. 2010.
[45] Z. Wei, Y. Kanzawa, K. Arita, Y. Katoh, K. Kawai, S. Muraoka, S. Mitani, S. Fujii, K. Katayama, M. Iijima, T. Mikawa, T. Ninomiya, R. Miyanaga, Y. Kawashima, K. Tsuji, A. Himeno, T. Okada, R. Azuma, K. Shimakawa, H. Sugaya, T. Takagi, R. Yasuhara, K. Horiba, H. Kumigashira, and M. Oshima, “Highly reliable TaOx ReRAM and direct evidence of redox reaction mechanism” *IEEE Int. Electron Devices Meeting*, 2008, pp. 293–296.
[46] H.-S. Philip Wong, Heng-Yuan Lee, Shimeng Yu, Yu-Sheng Chen, Yi Wu, Pang-Shiu Chen, Byoungil Lee, Frederick T. Chen, and Ming-Jinn Tsai “Metal–Oxide RRAM” *Proceedings of the IEEE* Vol.100, No.6, June 2012
[47] M. J. Lee, C. B. Lee, D. Lee, S. R. Lee, M. Chang, J. H. Hur, Y. B. Kim, C. J. Kim, D. H. Seo, S. Seo, U. I. Chung, I. K. Yoo, and K. Kim, “A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta_2O_{5-x}/TaO_{2-x} bilayer structures” *Nature Materials*, vol. 10, pp. 625–630, Aug. 2011.

[48]Peter Clarke "RRAM set to follow 3-D flash, says IMEC" EE Times news & Analysis,10/12/2011. http://www.eetimes.com/electronics-news/4229617/RRAM-set-to-follow-stacked-flash?pageNumber=0
[49] B. Govoreanu, G. S. Kar, Y. -Y. Chen, V. Paraschiv, S. Kubicek, A. Fantini, I. P. Radu, L. Goux, S. Clima, R. Degraeve, N. Jossart, O. Richard, T. Vandeweyer, K. Seo, P. Hendrickx, G. Pourtois, H. Bender, L. Altimime, D. J. Wouters, J. A. Kittl, and M. Jurczak, "10 × 10 nm2 Hf/HfOx crossbar resistive RAM with excellent performance, reliability and low-energy operation," *Electron Devices Meeting (IEDM), 2011 IEEE International* 2011, pp. 31.6.1–31.6.4
[50] Y. Sakotsubo, M. Terai, S. Kotsuji, Y. Saito, M. Tada, Y. Yabe and H. Hada "A New Approach for Improving Operating Margin of Unipolar ReRAM using Local Minimum of Reset Voltage" *IEEE Sym. on VLSI Tech*., Dig., pp. 87-88, 2010
[51] J.-J. Huang, Y.-M. Tseng, W.-C. Luo, C.-W. Hsu, and T.-H. Hou, "one selector-one resistor (1S1R) crossbar array for high-density flexible memory applications", *in Tech. Dig. Int. Electron Devices Meeting*, Washington, DC, 2011, pp. 733–736.
[52] A. Kawahara, R. Azuma, Y. Ikeda, K. Kawai, Y. Katoh, K. Tanabe, T. Nakamura, Y. Sumimoto, N. Yamada, N. Nakai, S. Sakamoto, Y. Hayakawa, K. Tsuji, S. Yoneda, A. Himeno, K. Origasa, K. Shimakawa, T. Takagi, T. Mikawa, K. Aono "An 8Mb Multi-Layered Cross-Point ReRAM Macro with 443MB/s Write Throughput" in *Proc. Int. Solid-State Circuits Conf*., San Francisco, CA, 2012, pp. 432–434.
[53] K. Kinoshita, K. Tsunoda, Y. Sato, H. Noshiro, S. Yagaki, M. Aoki, and Y. Sugiyama, "Reduction in the reset current in a resistive random access memory consisting of NiOx brought about by reducing a parasitic capacitance" *APPLIED PHYSICS LETTERS* 93, 033506 , 2008
[54] H. Y. Lee, P. S. Chen, T. Y. Wu, Y. S. Chen, C. C. Wang, P. J. Tzeng, C. H. Lin, F. Chen, C. H. Lien, and M.-J. Tsai, "Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO_2 based RRAM," *Electron Devices Meeting (IEDM), 2011 IEEE International*, 2008, pp. 297–300
[55] M. Terai, Y. Sakotsubo, S. Kotsuji, and H. Hada "Resistance Controllability of Ta_2O_5/TiO_2 Stack ReRAM for Low-Voltage and Multilevel Operation" *IEEE ELECTRON DEVICE LETTERS*, VOL. 31, NO. 3, MARCH 2010

Chapter 2: Kinetic Monte Carlo Simulation on Electrochemical Metallization RRAM

The nature of RRAM is characterized by randomness. For anion-based electrochemical metallization (ECM) RRAMs, the device operation involves chemical reactions and physical diffusion. Both of these are random at the atomic scale. For oxygen vacancy (VO) based RRAM, random trap generation is included as well. All of those result in random creation and annihilation of conductive filaments, which directly affect the RRAM characteristics. Among various simulation tools, Kinetic Monte Carlo (KMC) can be an effective approach. It can effectively capture the random nature of RRAM devices and at the same time not cost too much computation power and time.

The basic physical and chemical processes of ECM RRAM are investigated in this chapter. First, a brief introduction to the KMC simulation technique is provided. Next, a KMC simulator is established to simulate the metallic filament formation process in a typical ECM RRAM. All the chemical and physical processes involved in the KMC are discussed. The characteristics of the forming voltage, forming time and the so called "Voltage–time dilemma" are studied. In addition filament topographies, which strongly influence device properties, are studied under different device operation conditions. Further, studies on filament overgrowth and on-state resistance are presented. Finally, the simulation results are verified by experiments.

2.1: Introduction to Kinetic Monte Carlo Method

There are generally two major techniques used to understand the microscopic kinetics of an electronic material system. One is molecular dynamics (MD) and another one is Kinetic Monte Carlo (KMC). MD simulation consists of a numerical, step-by-step, solution of the classical equations of motion [1]. Thus, calculation of the forces acting on every particle at each time step is needed. Those forces are generally obtained from the system potential energy and an interatomic potential for atoms has to be chosen in advance. Once the potential and the initial condition of a system are chosen, the whole system will naturally evolve in time.

One problem with MD, however, is the time limitation problem. In order to capture the atomic movement details, the time step should be short enough (10^{-15}s) to describe the atomic vibration of atoms. Thus, the particular simulation time is limited to 1µs [1]. On the other hand if the time evolution of a system which we are interested is much longer, MD becomes impractical. KMC, however, can successfully overcome this time scale limitation. One of the major differences between KMC and MD is that in KMC instead of focusing on the atomic vibration at a small time step, KMC only focuses on the key processes which can actually transit the system to a new state. For example, consider an atom diffusion problem: the atom can stay at position A for quite a long time before it hops to a new position B. During this time interval,

the atom just stays at position A and thermally vibrates around it and there is nothing interesting thing happening during this time interval. Thus, if we use MD to simulate this process, it will waste lots of time just to describe the atomic vibration around point A. By adopting KMC, however, we can ignore this boring part and only focus on the real transition which is the hopping process for the atom from position A to position B. Thus, KMC can allow us to simulate a physical process that has a much longer time scale.

In addition, most of the systems being considered are memoryless. That means once a particle enters a state i, it does not remember which state it came from. This rule is true for a physical system in which the time a particle takes to transit from state *m* to state *n* is much shorter than the time it takes to stay at state *n*. Thus, during each short increment of time, it has the same probability of finding an escape path as that in the previous increment of time. This gives rise to a first order process with exponential decay statistics [1]. The probability that the particle remains at same state at time t is [1]:

$$p_{i,stay}(t) = \exp(-\Gamma_i t) \tag{1}$$

Where Γ_i is the transition rate for the ith particle. Equation (1) is the cumulative distribution function (CDF). From above equation, the probability density distribution function (PDF), the time derivative of CDF, can be obtained easily:

$$p_i(t) = \Gamma_i \exp(-\Gamma_i t) \tag{2}$$

Very often a real physical system is an ensemble of a large number of particles. In most cases, all of them can be assumed independent. According to the probability theory if particles have the transition rates of Γ_1, Γ_2, Γ_3 …. Γ_n and if they are independent, the probability of the occurrence of next event also follows the exponential distribution with the transition rate equal to the sum of all the individual transition rate of the particles [1].

$$\Gamma_{tot} = \sum_i \Gamma_i \tag{3}$$

Similarly the CDF of the time when the next event occurs follows the exponential distribution and can be described by equation:

$$p(t) = \Gamma_{tot} \exp(-\Gamma_{tot} t) \tag{4}$$

The average transition time is [1]:

$$\tau = \int_0^\infty t p(t) dt = \frac{1}{\Gamma_{tot}} \tag{5}$$

A common algorithm to implement KMC is so called Variable Step Size Methods (VSSM) [2]. It is simple and efficient. The general procedures are listed below [2].

Variable Step Size Method:

1. Initialize

- Generate an initial configuration
- Set the time t to the initial value.
- Choose conditions when to step the simulation.

2. Update the system time

- Generate a time interval Δt when no transition takes place

$$\Delta t = -\frac{1}{\Gamma_{tot}} \ln r \tag{6}$$

Where r is a random number uniformly distributed on the unit interval.

- Update system time to t = t + Δt.

3. Update system configuration

- Select a transition with a probability of $\frac{\Gamma_i}{\Gamma_{tot}}$ and update the system configuration.

4. Continuation

- If the stop conditions are fulfilled, then stop. If not, repeat step 2.

As can be seen from above description, a KMC simulation requires quantitative information about the transition rates of all the chemical and physical processes inside a system. Those rates are critical for the successful implementation of a KMC simulation.

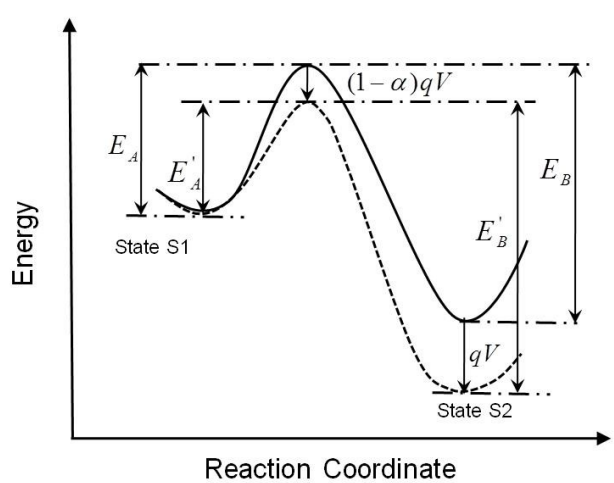


Figure 2.1. The energy diagram for a transition between states S_1 and S_2

Quite often, the transition rates are obtained from the transition state theory (TST), where a reaction is considered to be a transition between adjacent local energy minima [3]. As shown in Fig. 2.1 [4], the solid line represents the state diagram when there is no external electric field and the dash line on the other hand represents the state diagram when there is an external electric field present across the system. Let us first consider the situation when there is no electric field. The particle at state S_1 can be thermally excited and overcome the activation energy barrier E_A to reach the state S_2. And this is considered as the forward transition. Similarly, if a particle resides at state S_2, it also has the opportunity to reach state S_1. In this case, however, the reverse activation energy barrier E_B is obviously larger than E_A. The difference between E_A and E_B is the energy difference between states S_1 and S_2. In this case, the backward transition is not favored compared with forward transition.

If Boltzmann statistics is assumed to govern the chemical and physical processes, the forward transition rate can be written as [5, 6]:

$$\Gamma_f = v_o \exp\left(-\frac{E_A}{k_B T}\right) \tag{7}$$

Where Γ_f is the forward transition rate and v_o is a frequency factor, which can be assumed to be the atomic vibration frequency roughly equal to 10^{13} s^{-1}. k_B is Boltzmann factor. T is the absolute temperature. Similarly the reverse transition rate can be expressed as:

$$\Gamma_r = v_o \exp\left(-\frac{E_B}{k_B T}\right) \tag{8}$$

All the activation energies values have to be modified in the presence of electric field as shown in the dashed line. The electric field can come from both the ion space charge distribution and the external voltage supply. The E_a for the forward and reverse transitions have to be modified by $-\alpha q\Delta$ and $(1-\alpha)q\Delta$ respectively, where $q\Delta$ is the potential change across two states. Note that α is the symmetry factor, which describes the location of barrier plateau with respect to two final states. It is typically 0.5. Thus, whenever, there is electric field present, the activation energies and the transition rates have to be modified.

The description above briefly goes through the fundamental principles of KMC. Even though KMC is straightforward to implement, it have several limitations as pointed out by Battaile [3]. The author points out that first it is very difficult, even using the most sophisticated quantum mechanical methods to accurately obtain the chemical and physical rate parameters. Thus any mechanism is, by necessity, approximate at best. Second, it is also virtually impossible to know all the reactions (even just the important ones) in advance. Therefore, any reaction mechanism represents a limited and simplified version of the real chemistry. In the other words, KMC cannot generate new physics and all the physics has to be defined before the simulation. Thus, the author pointed out that the challenge in simulating system in atomic scale is often not in developing the algorithm but rather in constructing a meaningful chemical reaction mechanism.

2.2: Establishment of a KMC simulator for ECM RRAM

2.2.1: Basic Chemical and Physical Processes and Assumptions

As discussed in the previous chapter, a typical ECM RRAM consists of a layer of electrochemically active metal as an anode, an insulator layer and finally a layer of inert metal as a cathode. The insulator layer can either serve as an electrolyte, which has a large concentration of anode type metal ions and can thus contribute to filament formation processes (for example, systems such as $Ag/Ag_2S/Pt$ [7] and $Ag/Ag_2Se/Au$ [8]) or just serve a transportation layer for oxidized anode ions passing from the anode to the cathode (for example, systems such as $Cu/SiO_2/Pt$ [9] and $Cu/Ta_2O_5/Pt$ [10]). The simulator discussed in this chapter focuses on the second class of ECM RRAM described above, where the insulator layer only serves as transportation medium for the oxidized anode ions. In this way, the complicated electrolyte structure can be ignored. However, as will be shown, because both types of devices have almost identical physical processes the simulation result works well for both types of devices.

The various physical and chemical processes included in the KMC simulation are shown in Fig. 2.2. Both anode and cathode, oxidation/reduction reactions ($M \leftrightarrows M^+ + e$) are included and additionally ion surface diffusion, desorption and adsorption processes are also included. Finally, an isotropic electrolyte with zero electron conductivity is assumed, which implies that the bulk ionic diffusion E_a is the same in every direction and there is no electron current through the electrolyte.

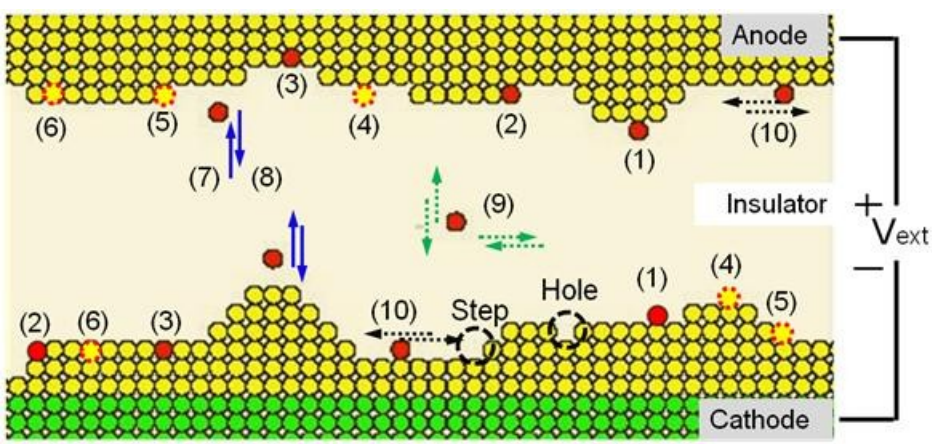


Figure 2.2. Processes included in the KMC simulation. Oxidation at (1) adatom site (0.65 ev), (2) step site (0.7 ev), (3) hole site (0.75ev). Reduction at (4) adatom site (0.35 ev), (5) step site (0.3 ev), (6) hole site (0.25 ev). (7) adsorption (0.15 ev). (8) desorption (0.3 ev). (9) bulk diffusion (0.15 ev). (10) surface diffusion (0.2 ev).

As shown in the Fig. 2.2, after a cation gets adsorbed on the cathode surface, it can reduce to form an adatom, and it can also reach a step or hole site through surface diffusion. The activation energies (E_a) of various processes are structure dependent [5]. As shown in Fig. 2.3, for example, E_a for ions that are reduced at adatom sites is larger than its value when they are reduced at step or hole sites. Therefore, ions are more likely to be reduced at step or hole sites.

Before proceeding, it is important to state the assumptions this simulator makes. In fact, most of the chemical and physical processes in the KMC are phenomenological. That means those processes are approximations of a real system. There are several reasons for this. First, most of the chemical processes are extremely complicated and may involve multiple reactions and charge transfer processes. It would be impractical to spend computation resources on such detailed processes. Second, as stated above even by using the most advanced first principles simulator, there are numerous fundamental processes people do not understand yet and for some other processes, only a very rough qualitatively description can be given. Despite those limitations it is still possible to capture the essence of the problem and make a satisfactory simulation for the entire filament formation process. In the next several paragraphs some basic assumptions and approximations are described.

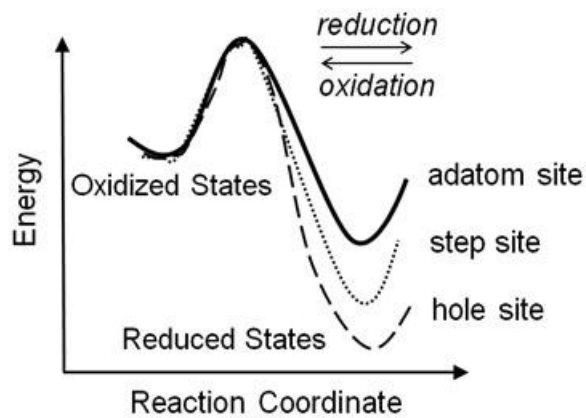


Figure 2.3. Structural dependency of activation energy

The first thing to consider is the metal (such as Cu) oxidation processes at the metal-dielectric interface (such as at Cu/SiO_2 interface). This, relatively simple-looking problem, is still not well understood yet. B.G. Willis [11] did an excellent experiment trying to clarify the transport mechanism of copper inside SiO_2. He summarizes four possible mechanisms as shown in Fig. 2.4. (a) thermal diffusion from the electrode into the dielectric, (b) copper oxidation via reduction of SiO_2 and subsequent transport of the ion under the electric field, (c) copper atoms under an electric field lose the electrons and becomes ions, then those ions are injected into the dielectric with the help of electric field (d) ''chemical oxidation'' of copper at the interface via ambient gases such as H_2O and O_2 or via dielectric out-gassing, and transport of the ions under electric bias.

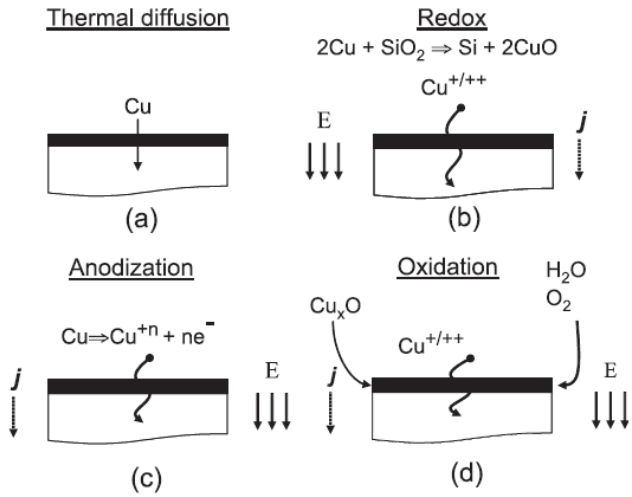


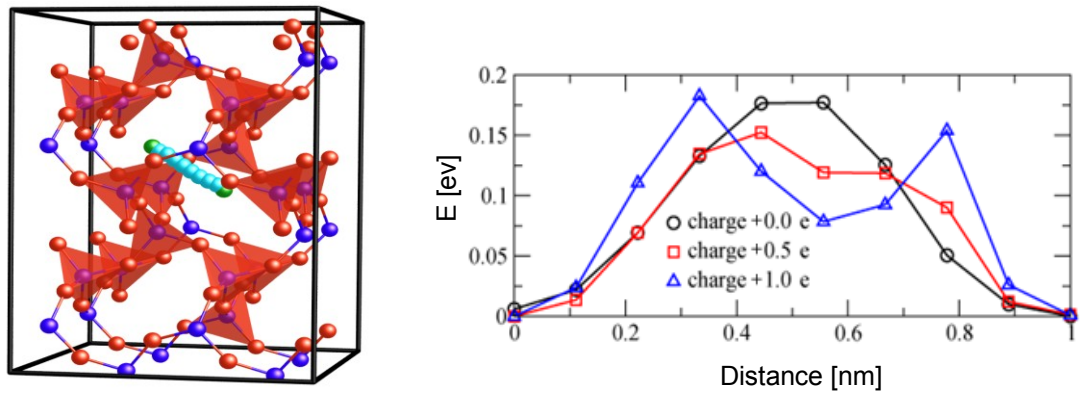
Figure 2.4. Four possible ways of copper injection into SiO_2 [11]

His conclusion is that in contrast to earlier results, he found no evidence of copper diffusion directly from electrodes, but he found a clear signal of copper diffusion from an oxidized copper electrode. That means only processes (c) and (d) are possible. Through further study, Willis claims that only mechanism (d) is active for copper transport at temperatures less than 450-500°C and electric field < 1MV/cm. In the other words, the chemical reaction has to be assisted by either O_2 or H_2O, which comes from the surrounding environment or from the out gassing of the dielectric. In this KMC simulation, process (c) is adopted to reduce the complexity of the system. However, it may be possible to take the O_2 and H_2O into account by assigning a smaller E_a for oxidation to make the oxidation process easier.

In addition, approximations are made for the crystallization process. In fact, the real crystallization processes are also much more complicated than the model used in this KMC simulator. Typically, after an ion gets adsorbed on the surface it can have several choices. First, it can complete the charge transfer process and become an adatom. This adatom can further diffuse to kink or step sites. In the second case, after the ion gets adsorbed, it could just complete a partial charge transfer process to become an adion. This adion then diffuses to a kink or step site and gain the rest of the charge and eventually becomes a neutral atom which finally incorporates into the crystal. In the current KMC simulator, it simplifies the problem by assuming all the ions once adsorbed on the electrode still carry charge. Only adsorbed ions are allowed to diffuse through the surface. This approach on one hand simplifies the problem, on the other hand, it doesn't lose the generality of the crystallization process since it still guarantees that the kink and step sites are stable sites compared with the adatom sites.

Finally, it has to be pointed that the values of various E_a related to chemical reaction are difficult to obtain. As Bronshtein [12] points out the theoretical difficulties in determining those values lies in several factors. First, typically the reactants are strongly coupled to metal electrode. Second, the reaction pathway depends crucially on the interfacial structure. And finally the reaction rates depending on the dynamics of the solvent (solid electrolyte in this

case). All those abovementioned effects make the charge transfer process a challenge for theoretical modeling. Typically, those activation energies are modeled by experimental fitting.



(a) (b)

Figure 2.5. (a) Simulated cell structure for Cu diffusion through *α*-cristobalite phase SiO_2 (b) Diffusion energy barriers for Cu element carrying different charges [14].

Unlike the individual chemical reactions, people can generally describe the bulk diffusion process well because there is no charge transfer process involved. Nevertheless, because of the inherent limitations in density functional theory (DFT) [13], the simulated barrier height is typically lower than the real value. M. Zelený [14] did an Ab initio simulation to determine the copper diffusion barrier through *α*-cristobalite phase SiO_2. The simulated structure and result are shown in Fig. 2.5. The simulated diffusion barrier is between 0.15ev to 0.2ev depending on the charge carried by the Cu element.

In the KMC simulator, the diffusion of neutral atoms is ignored, since there have much lower diffusion rate than the rate of ions when an external electric field is present. In addition, in simulation the diffusion barrier for ions is set to 0.15ev, which can mimic an insulator layer with a small diffusion barrier or can represent the fast ion paths created after the device undergoes several cycles' operation and so that the barrier gets lowered [7]. In addition, in the KMC simulation a larger value of oxidation E_a than that of reduction is used. This represents metals like copper and silver, which have large positive standard electrode potential [15] and therefore find it easier to capture electrons and precipitate but more difficult to get oxidized.

2.2.2: Simulation Procedures

Ideally the simulation should be carried out in 3D. Unfortunately, this would consume too many computational resources and too much time. Thus, only 2D is considered here. In the simulation, as shown in Fig. 2.2, each layer of electrode consists of 150 atoms and the insulator

thickness is 10nm. A periodic boundary condition is applied while solving the Poisson equation; thus, the edge effects are ignored. This is a good approximation when the device width over thickness aspect ratio is large. In addition, since the simulation is carried out in 2D, the out-of-plane electric field is ignored. The Poisson equation is solved by a finite difference method.

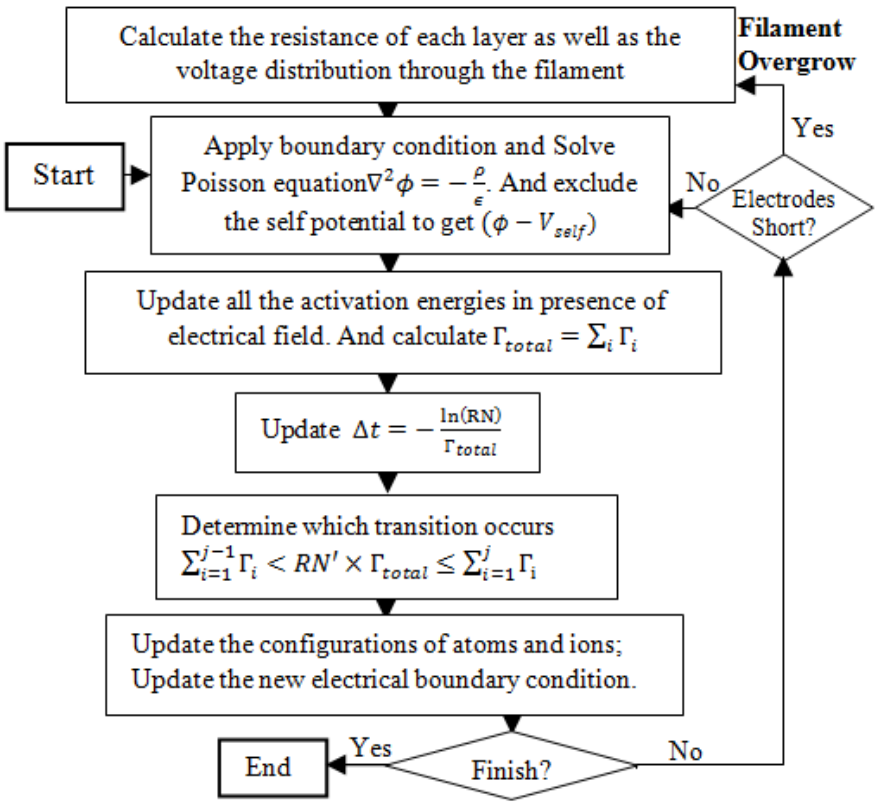


Figure 2.6. The flowchart of the KMC simulation process. RN and RN' are random numbers uniformly distributed between [0 1]

The simulation flow-chart is provided in Fig. 2.6. Starting from the initial configuration, the Poisson equation is solved. After that the self potential is excluded [16]. The self potential is the potential generated by the ion itself. This potential doesn't help lowering its own E_a. After potential information is obtained, the activation energies in the presence of electric field can be updated and both individual and total transition rates can be calculated based on the following equation:

$$\sum_i \Gamma_i = \sum_i \nu \exp(\frac{-qE_{a,i}}{k_B T}) \exp\left[\frac{q(V_{initial} - V_{final,i})}{2k_B T}\right] \exp(\frac{-qV_{self,i}}{2k_B T}) \quad (9)$$

Next, according to equation 6, system time t can be updated. Next, by generating a second random number, a specific transition can then be chosen. Once a process is chosen, the configurations of all the atoms and ions can be updated. Then Poisson equation has to be solved again and the whole simulation process continues.

2.3: Simulation Results and Discussion

2.3.1: Filament Formation Study

Even though the forming process is only carried out once in the device operation, the simulation of the filament formation process is critically important in understanding the basic physics of the memory cell. Actually, the subsequent SET processes can be viewed almost identical to the forming process. The only difference between SET and forming processes are the effective resistive layer thickness as described in the previous chapter. Thus all the conclusions drawn from the simulation of the forming process can be directly applied to the SET process as well.

In addition, certain simulated results such as the voltage-time relationships and the filament topographies can be directly compared with experimental results to verify the correctness of a simulator.

In this section, the KMC simulator is first used to predict the relationship between forming time and applied voltage as well as its associated filament topography. The definition of forming time is the time needed to form a metallic filament from the pristine state. In the other word, it is the time needed at the first cycle of the device operation to bring the memory cell from HRS to LRS.

Fig. 2.7 shows the filament topographies obtained at different voltage levels. Fig. 2.8 shows the time required to form those filaments. And the inset shows literature-reported experimental result from Cu/SiO_2 devices [17].

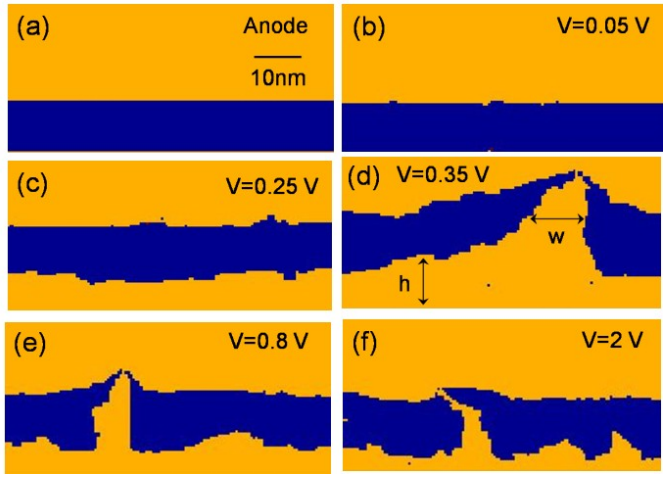


Figure 2.7. Filament topographies obtained at different voltage levels. (a) initial state. (b) V=0.05 V. (c) V=0.25 V. (d) V=0.35 V. (e) V=0.8 V. (f) V=2 V. The arrows in (d) with labels 'h' and 'w' indicate the thickness of isotropic deposition and the average width of the filament respectively [4, 7].

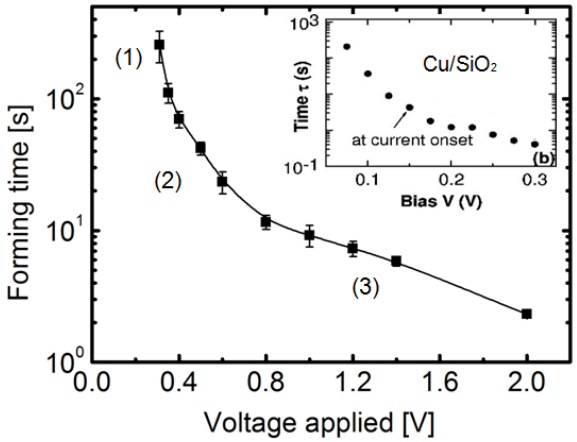
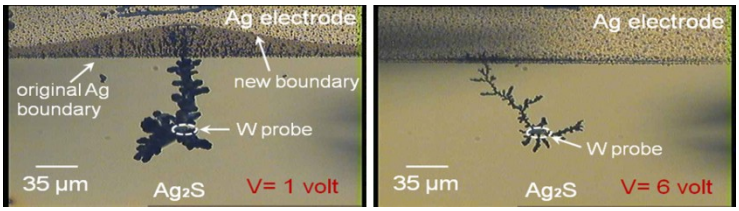


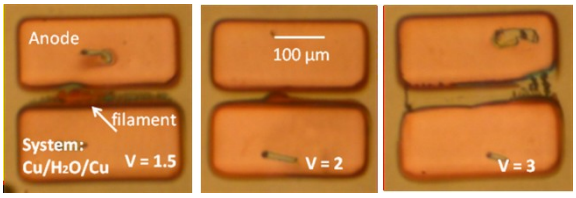
Figure 2.8. The relationship between forming time and applied voltage [4, 7]. The inset shows the experimental results on Cu/SiO2 [17].

As apparent from the simulation results shown in Fig. 2.8, there is a threshold (V_t) in the I-V characteristics. For an input voltage well below V_t, it is almost impossible to get any deposition on the cathode as shown in Fig. 2.7 (b). This is because when the voltage is extremely small, it is very difficult for anode atoms to first be oxidized, then desorbed from (without being reduced back on) the anode and finally be reduced (without being re-oxidized) on the cathode. Next, when the voltage is around V_t, isotropic deposition is observed (Fig. 2.7 (c)) and the distance between the anode and cathode remains essentially constant. Thus, at this stage, it is still impossible to get filament formation. When the voltage is slightly above V_t, a filament is eventually formed. In this bias condition, the width of the filament 'w' is large as shown in Fig.

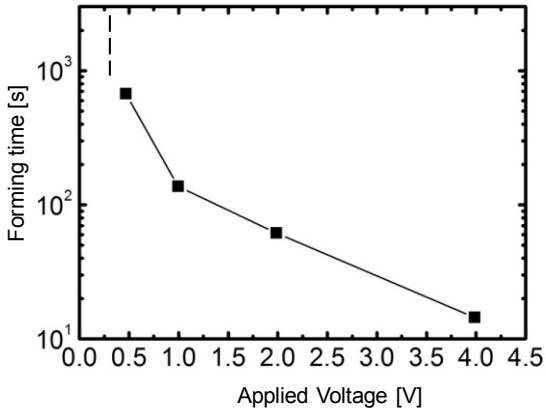
2.7 (d). When the voltage is further increased, the width of the filament decreases significantly as demonstrated in Fig. 2.7 (e) and (f). These results can be explained by the fact that when the voltage is small, the adsorbed cations at the cathode side tends to diffuse to and reduce at more stable step and hole sites rather than at adatom sites, and hence the deposition is isotropic; thus, there is no effective gap shrinkage between the anode and cathode. A larger voltage, however, make the adatom formation easier; thus the filament grows faster (Fig. 2.8) and its width decreases. To confirm these observations, experiments were performed on lateral cells built suing both Cu/H_2O and Ag/Ag_2S [4, 7]. In fact, Cu/H_2O closely resembles our simulated system, since H_2O doesn't provide any cations but only acts as a transportation medium for them. Fig. 2.9 shows that the experiments match the predictions from simulation. When the voltage is small, very wide filaments are obtained. One the other hand, when the voltage is large a much narrower pattern appears.



(a)



(b)



(c)

Figure 2.9. Filament topographies at different Voltages levels for (a) $Ag/Ag_2S/W$ and (b) $Cu/H_2O/Cu$. (c) The experimental results of forming time vs applied voltage obtained in $Ag/Ag_2S/W$ system [4].

Fig. 2.8 indicates that the forming time, t_{Form}, generally decreases as the voltage increases. In addition we observe three regimes in the graph. The shape of this curve is a result of competition between the various aforementioned chemical and physical processes and depends on the individual activation energies, which themselves are material-dependent properties. The general trends, however, can be roughly explained by considering the Butler-Volmer equation (BV) [15].

$$i = i_0 \left\{ \exp\left[\frac{\alpha nq}{k_B T} \eta_s \right] - \exp\left(-\frac{(1-\alpha)nq}{k_B T} \eta_s \right) \right\} \tag{10}$$

$$\eta_s = V - U \tag{11}$$

In the two equations above, η_s is the surface overpotential, which is the potential difference between the electrode (relative to the solution) potential minus the value at equilibrium (U) [15]. It is also the potential of the electrode minus that of a reference electrode of the same kind and located adjacent to the surface [15]. V is the applied voltage. α is the symmetry factor which is same as defined in previous section.

The applied voltage is the sum of the surface overpotentials at the anode and cathode interfaces as well as the potential drop across the dielectric. At region 1 and 2 shown in Fig. 2.8, the applied voltage is small and the surface overpotentials at the electrode interfaces are also correspondingly small. Thus from the BV equation, the current, and hence the cation flux through the dielectric, is linearly dependent on surface overpotentials. The forming time is roughly equal to the number of cations needed to form the filament divided by the average cation flux. As discussed above, the surface overpotential has a dramatic impact on the filament topography, which eventually determines the number of cations needed for the filament. When the surface overvoltage is small, the deposition tends to be isotropic and both the deposited film and the anode surface tend to be smooth. Consequently, there is no filament-induced effective shrinkage of the distance between the cathode and the anode. Therefore, we expect that a large number of cations have to be supplied from the anode in order to form a filament and hence the forming time increases strongly as well. In region 2, the cation flux increases linearly with surface overpotential and the number of cations needed also decreases due to the filament topographic effect. Thus t_{Form} decreases as voltage increases. Finally when the voltage is further increased as shown in region 3, the surface overpotential is large. The BV equation then reduces to the Tafel approximation [15];

$$i = i_0 \exp\left[\frac{\alpha nq}{k_B T} \eta_s \right] \tag{12}$$

the cation flux is then exponentially dependent on surface overpotential. Furthermore, under a large surface overpotential the adsorbed cations can easily be crystallized at adatom sites thus the number of cations needed for the filament plays a minor role. Therefore, we expect to see an exponential relationship between t_{Form} and applied voltage as shown in region 3.

Through the above discussion it seems the BV equation alone can describe the entire process. This would then suggest that there is no need to implement KMC simulation. In fact, the BV equation itself cannot fully describe the whole system for the following reasons [7]: first, the BV equation only deals with the charge transfer between the electrode and the dielectric interface and doesn't include the ion transport processes on the deposited metal surface and inside the dielectric layer. Thus, other transport equations such as drift-diffusion equation describing how cations transport through the dielectric layer and metal surface have to be coupled to the BV equation. The entire problem is a function of time and position. Second, in such a nano-device the charge distribution is discrete. Thus physical quantities like continuously varying charge concentrations functions become ambiguous. By using KMC, however, this ambiguousness can be easily solved.

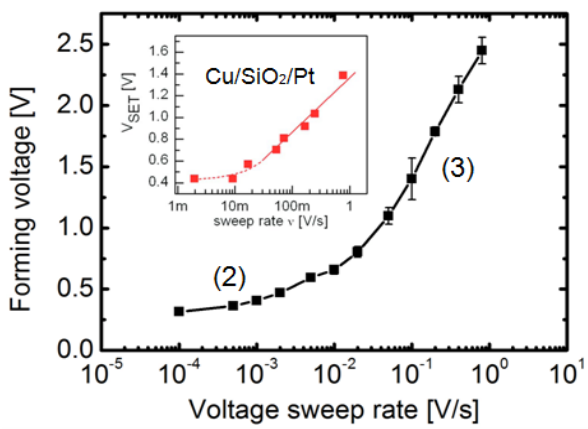


Figure 2.10. Simulated V_{Form} vs voltage sweep rate [4 ,7]. The inset shows the experimental results on $Cu/SiO_2/Pt$ system [9]

Next, the relationship between forming voltage and voltage sweep rate is explored. The simulated results are shown in Fig. 2.10 [7]. There are several features on this graph. First, there should be a threshold voltage V_t as discussed above no matter how small the voltage sweep rate is. This can be seen in the figure that as the voltage sweep rate keeps decreasing, the V_{Form} approach a constant. Next, when the voltage sweep rate increases slightly, we can see that the forming voltage also increases slightly. Finally, when the voltage sweep rate is large, the V_{Form} logarithmically depends on the voltage sweep rate.

It is worth to note that C. Schindler [9] gives a mathematical model to describe the logarithmical relationship between V_{Form} and voltage sweep rate. Schindler assumes the applied voltage is approximately equal to the surface overpotential by assuming both oxidation at anode and the ionic transport inside the dielectric layer is not the rate limiting process. The total charge needed in the forming stage Q_{Form} can be further expressed as

$$Q_{FORM} = \int_0^{V_{Form}} \frac{I_f}{v} dV = \int_0^{V_{Form}} \frac{i\pi r_f^2}{v} dV \tag{13}$$

where I_f is total current going through the device, i is the current density and v is the voltage sweep rate and is equal to

$$v = \frac{dV}{dt} \tag{14}$$

In addition, when the applied voltage is large, Tafel approximation is assumed. Putting equations 12 to 13 together, the following relationship can be obtained:

$$V_{Form} = \frac{kT}{\alpha nq} \ln v + \frac{kT}{\alpha qn} \ln \frac{Q_{Form}\alpha nq}{i_0 \pi r_f^2 kT} \tag{15}$$

This equation gives a logarithmic relationship between the forming voltage and voltage sweep rate. From experimental results based on $Cu/SiO_2/Pt$, extracted α is around 0.1 [9], which is significantly less than $\alpha \approx 0.5$ for an electrochemical system with liquid electrolyte. In addition, from the KMC simulated results, even though α is intentionally set to be 0.5, the extracted α from the simulated result is only around 0.05. Those significant discrepancies indicate Butler-Volmer equation alone can't accurately describe the entire system. In addition, in Schindler's approach all the voltage is assumed to be dropped at the interface between the cathode and SiO_2 where the crystallization process occurs. This assumption is also questionable since both the ionic transport process inside the dielectric and the oxidation process at the anode side are also likely to be rate limiting factors. Further detailed studies are needed to find the impacts of various physical and chemical processes on α.

2.3.2: Relationship Between Forming Voltage and Device Size

The relationship between V_{Form} and device size is studied in this section. The main reason to study the impact of device size on V_{Form} is to study the scaling properties of RRAM devices. Scaling includes both the reduction of the resistive layer thickness and also the reduction of the device area. From the simulation, it can be shown that when the device becomes thin enough, the forming process can be effectively eliminated, which is a very desirable property. In addition, it also can be used to verify the simulator by comparing with experimental results.

Devices with different dielectric thicknesses are first simulated as shown in Fig. 2.11. For this simulation, the voltage ramping rate is 1V/s. As can be seen, V_{Form} linearly depends on the

insulator thickness. The inset shows result of experiments done in a Cu/SiO_2/Pt system [9]. The experimental result matches well with simulation. It indicates that the system is purely electric field dependent and as long as the electric field is constant, systems with different dielectric thickness will have identical behavior. The reason for this linearity can be explained as follows- the effective voltage drop between the initial and final states as shown in Fig. 2.1 is the product of the average electric field across the two states and the effective distance between those two states. Typically, the effective distance between the initial and final states is the same for a given system at various bias conditions. Taking the ionic bulk diffusion as an example, the metal ion roughly only hops across a distance of one lattice constant each time unless the electric field is so strong that the ion passes through multiple sites at one time. Thus, for most situations, as long as the electric field is maintained constant, the total voltage drop across the initial and final states and thus the change of transition rate will be identical. Since the electric field is inversely dependent on the film thickness, the forming voltage is therefore linearly dependent on the film thickness.

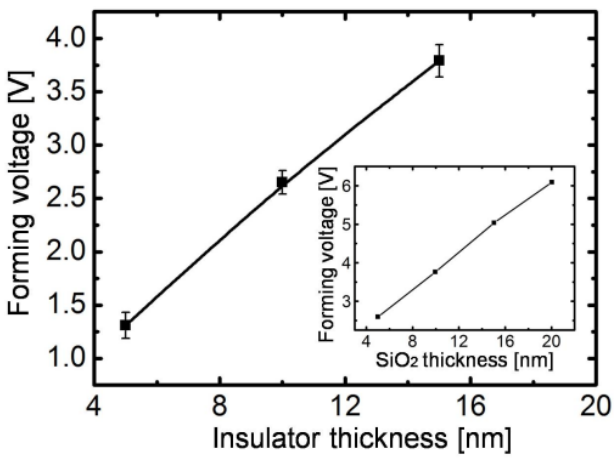


Figure 2.11. Simulated V_{Form} vs. insulator layer thickness. The sweep rate is 1V/s. Inset shows experimental result for Cu/SiO_2/Pt system [9].

Next, the relationship between V_{Form} and device area was investigated. In addition, the cathode surface roughness effect is included in the simulation and this is done based on an autocorrelation function method [18]. As show in Fig 2.12 (a), in the simulation Gaussian autocorrelation is used to generate a surface roughness profile. The correlation length is 10nm, and the root mean square (rms) value is set to1nm. The simulated surface roughness pattern is shown in the green colored region in Fig 2.12 (b). As expected, it is more favorable to form a filament at places where the effective distance between two electrodes is shortest, since the effective electric field is strongest at those positions. Both simulated and experimental results

(Fig. 2.12 (c)) [10] shows that V_{form} approaches a constant when the size of device is small, and gradually reduces as the size increases.

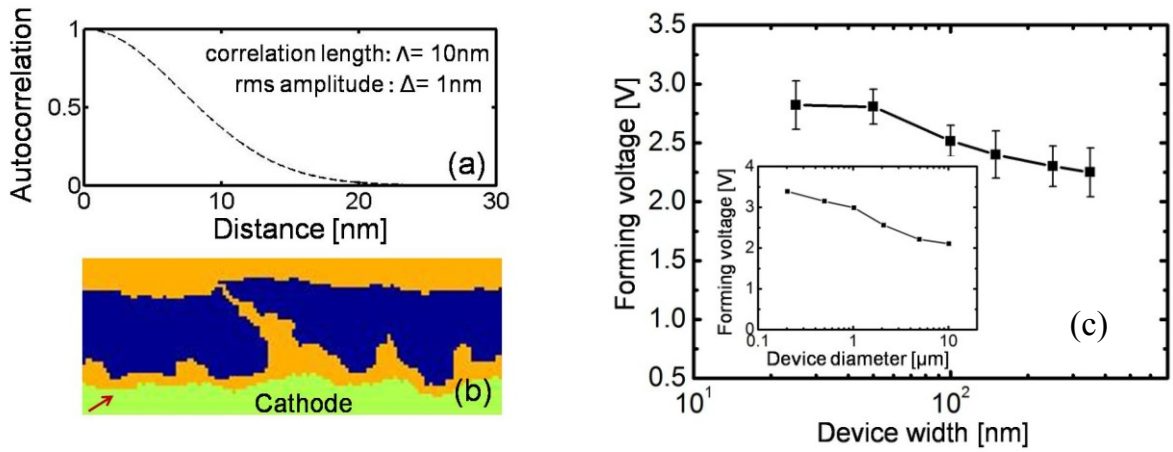


Figure 2.12. (a) The Gaussian autocorrelation used for surface roughness simulation. Due to simulation time limitation, the correlation length is 10nm, which is smaller than the real value. (b) Simulated pattern. The arrow shows the cathode surface roughness. (c) Simulated forming voltage vs. device width and experimental result [10] for $Cu/Ta_2O_5/Pt$ system (inset).

2.3.3: I-V Characteristics and SET Switching Properties

After the filament topography, voltage-time relationship and device size studies, this section mainly focuses on the I-V characteristics and the SET switching statistics. As discussed in the beginning of the chapter, due to the random nature of RRAM. KMC can be an effective approach which captures those stochastic features of RRAM devices without too much consumption of computation power during the simulation process. The merit of KMC is demonstrated in this section. As will be shown, it naturally produces statistical distributions that describe the device dynamic behavior. Also, by counting the charges arriving at one side of the electrode, the I-V relationships can be extracted. The I-V behavior, which is one of most important characteristics of RRAM devices, can thus be obtained by KMC as well.

This section first studies the I-V characteristics of the ECM RRAM together with a focus on the filament overgrowth phenomenon. Next, the relationship between applied voltage and SET time is investigated. Finally, the impact of the SET current compliance (CC) on the distribution of the filament resistance is obtained.

The current information can be done by counting the number of metal ions reaching the cathode surface during a given time interval. The current is calculated by the total charge

divided by the time duration. Here, the filament overgrowth effect is also simulated. In order to simply the filament overgrowth problem, a flat anode surface is adopted. This is a not bad approximation as can be seen from Fig. 2.7, which shows that the filament topography is mainly determined by the cathode [19]. Fig. 2.13 shows the simulated I-V result. As can be seen, there are several distinctive regions on the curve. The additional simulation results show that the current in region 1 is proportional to the electrode area while that in regions 2, 3 and 4 is area-independent. This is expected, since the number of immature filaments formed in region 1 is proportional to the electrode area and thus the ionic current is also proportional to the same, while regions 2-4 are dominated by local filament conduction and are thus area-independent. The current has a sharp turn on in region 2, when the filament shorts two electrodes. Once a short is formed, the ohmic electron current flows. Since the conductivity of the filament is not ideal, there is still a large voltage drop across it and thus the chemical process continues – the filament grows wider and its conductivity keeps increasing until a steady state is achieved with the external test system. In the KMC simulation, the voltage distribution across each layer of the filament has to be updated in every step.

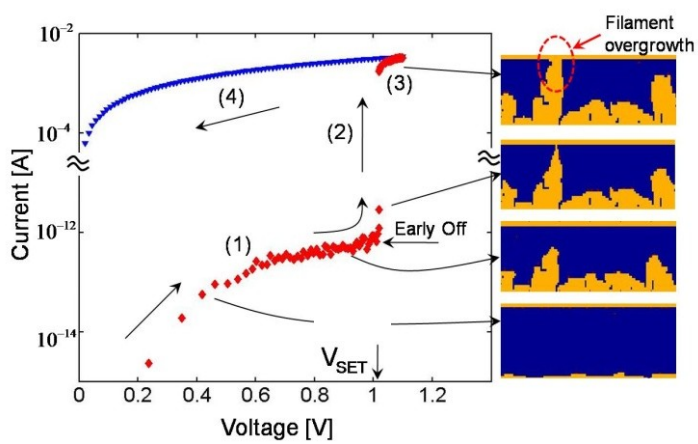


Figure 2.13. The KMC simulated I-V and related filament shape. The voltage sweep rate is 3V/s. Region (1) is ionic current while (3) and (4) are ohmic current. The resistivity of metal filament is set to $5\times10^{-8}\Omega$m. The external system resistance is taken as $R_{ext}=100\Omega$.

Besides the forming stage, the SET process can also be simulated. The difference between forming and SET processes is shown in Fig 2.14. At the forming stage, the entire filament is formed from the cathode side. However, during the actual SET and RESET switching operation, it only takes a few layers of atoms on top of the filament to turn on/off the device. Thus, the on-state switching time (t_{SET}) of the RRAM is far less than t_{FORM}, which is the time required to form an entire filament, rather than that required to form the last few layers required to complete the residual filament from a previous SET/RESET cycle.

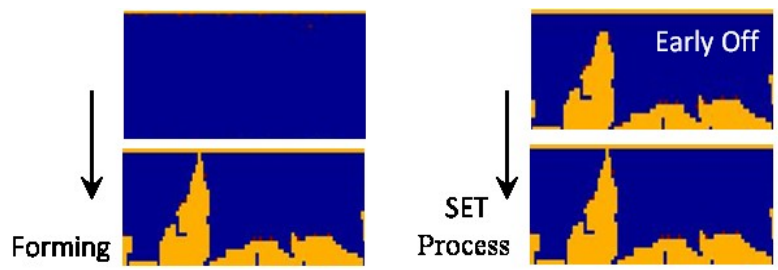


Figure 2.14. Comparison between Forming and SET processes.

In order to simulate the SET process, the initial configuration of the system is chosen as shown in right figure of Fig. 2.14, where six layers of atoms are removed. Fig. 2.15 (a) shows the t_{SET} distribution at different input voltages steps. In order to clearly see the distribution tail, a Weibull plot is adopted. As shown, at 2V at least 60µs is needed in order to turn on 99% of the devices, whereas only 3 µs is needed at 3V. The experimental results [10] done on Cu/Ta_2O_5 matches with the simulation result well. In addition, the mean switching time shows an inverse exponential dependence on voltage. This is expected since the SET process is similar to the forming process described in the previous section.

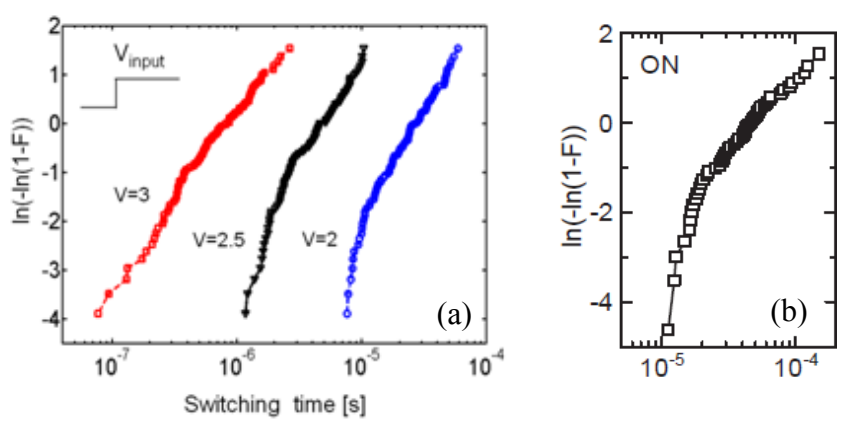


Figure 2.15. (a) Simulated switching time t_{ON} distribution under different input conditions (b) experimental results in Cu/Ta_2O_5 system [10].

In a real circuit application, both voltage and current operation modes are available. The KMC simulator is also capable of simulating the device operation behavior under the current operation mode. In this simulation, the same configuration (Fig. 2.14) as what was used in the voltage mode is adopted. At each time step, the resistance of the entire system is calculated,

and the voltage which is equal to the product of current and resistance is updated at every step at every layer of the filament. Fig. 2.16 (a) shows the simulation result. As can be seen, the filament on-state resistance (R_{on}) varies with control current due to the filament overgrowth effect. Generally, large programming current results in a smaller R_{on} with less variation. It should be noted that thermal effects are ignored for the simulation of the filament overgrowth effect. This is a good approximation when the actual filament resistance is large or a series protecting resistance is connected to the device so that there is only a limited amount of current passing through the device without a large amount of heat dissipation. Finally, figure (b) shows the experimental results obtained in Cu/Ta_2O_5 system.

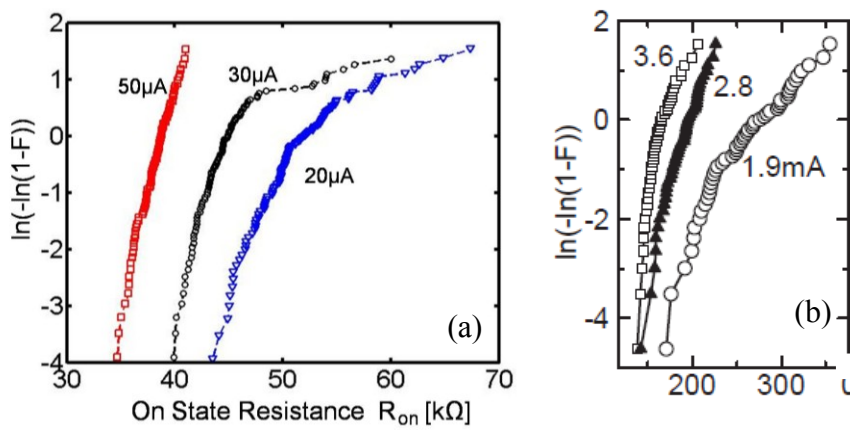


Figure 2.16. (a) Simulated on-state resistance distribution under different current input conditions (b) the experimental results for the Cu/Ta_2O_5 system [10]

2.3.4: RESET Stage Characteristics

During the RESET stage, a negative voltage is applied to the reactive metal electrode. Thus, the anode in the SET process becomes the cathode, and the cathode becomes the anode. Because of the reversal of the voltage polarity, the established metallic filament during the SET process starts to dissolve and drift back to the cathode (the reactive metal electrode). This process continues until the filament breaks. Once the filament breaks, there is a sharp resistance transition from low resistance state (LRS) to high resistance state (HRS).

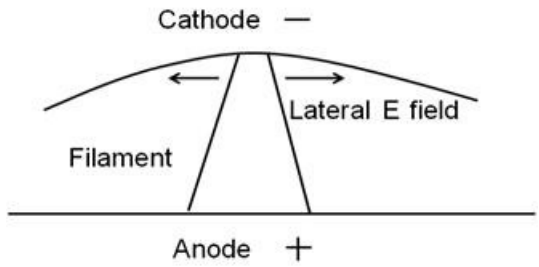


Figure 2.17. Illustration of side electric field near the cathode region in the RESET stage.

There are several important factors to be considered during the RESET process. The first one is the joule-heating effect [20], since the temperature can accelerate the chemical and physical processes. The second factor is the cathode-side topography, which determines the lateral electric field distribution at the filament tip close to the cathode side as shown in Fig. 2.17. As will be seen, this lateral electric field determines where the filament eventually breaks. The current KMC simulator doesn't have a temperature dependence implemented. As stated earlier, it works for the situation that the actual filament resistance is large or a large series protecting resistance is connected to the device so that there is only a limited amount of current passing through the device without a large amount of heat dissipation.

Secondly, it is very time and computation consuming to fully take the cathode topography into account during the filament overgrowth and RESET stages. In order to mimic the lateral electric effect, the following approximations are made. When the lateral electric field is small, the dissolved metal ions have difficulty diffusing out from the filament tip region; instead they prefer to drift along the vertical direction and accumulate at the cathode directly in front of the filament tip. Thus, in this simulation, it is assumed that metal ions can be accumulated at the filament tips. When the lateral field is large, however, beside diffusing in the forward direction, the metal ions can also diffuse laterally. Thus in the simulation, it is assumed that metal ions have difficulty to be accumulated at the front of the filament tip.

Fig. 2.18 shows the simulated I-V characteristics and related filament topographies. The voltage sweep rate is fixed at 1V/s. The external resistance is set to 30 $k\Omega$ with filament resistivity of $5\times10^{-6}\Omega$m. Simulation starts from condition (1) and ends at condition (8). As stated earlier, to simplify the problem a flat anode is adopted. The current in region (1) is ionic. It has a sharp turn on at $V=V_{Form}$, after that ohmic current flows. Since the filament has finite conductivity, there is still some voltage drop on it and hence filament overgrowth starts--the filament grows wider ((2) to (3)) and the resistance keeps decreasing until the voltage sweeps back to zero. On the other hand, when voltage sweeps to negative the filament starts to dissolve. After the filament breaks at V_{RESET} around 0.4V the over-dissolution process starts.

For the remaining cycles, since there is undissolved filament available, the SET voltage (V_{SET}) in the subsequent cycles is much smaller than V_{Form}. It is worthwhile to note that Fig. 2.18 (a) shows the filament breaks at the top where there is no metal ions accumulation on the cathode side. This is the case when the lateral electric field is large. On the other hand, Fig. 2.18 (c)

shows another scenario where metal ions are allowed to be accumulated and precipitate at the cathode side. Under this circumstance the filament may possibility break in the middle as indicated by the white circle in Fig. 2.18 (c). During overgrowth and dissolution processes, the voltage distribution in each layer of the filament has to be updated every step and serve as new boundary conditions in KMC.

Next, Fig. 2.19 shows the relationship between V_{RESET} and I_{SET}. In this simulation, a temperature effect is also included. As can be seen, generally filament formed at a larger I_{SET} has a larger V_{RESET} as well since more atoms have to be dissolved in the RESET process. By varying the temperature (T), the V_{RESET} varies significantly. Higher T accelerates the dissolution process and results in a much smaller V_{RESET} and less varied distributions.

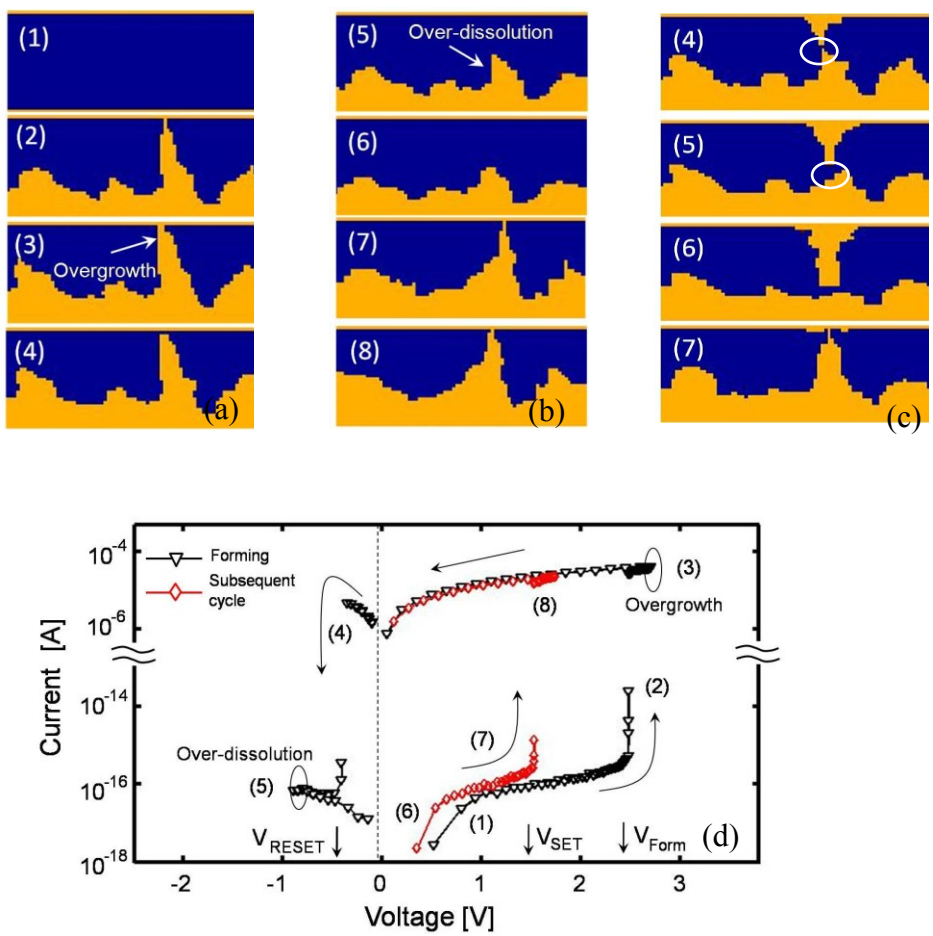


Figure 2.18. KMC Simulated I-V characteristics and related filament topographies (for the first scenario). The voltage sweep rate is 1V/s. Simulation starts from (1) and ends at (8). All the critical physical parameters are indicated in the figures. Overgrowth and over-dissolution are clearly seen And V_{SET} in the subsequent cycle is much less than $V_{Form.}$ The external resistance is set to 30kΩ. Filament resistivity is $5\times10^{-6}\Omega m$

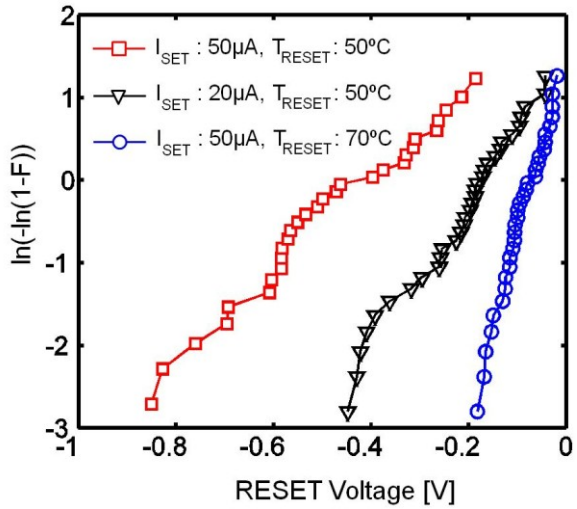


Figure 2.19. V_{RESET} distribution at different SET control current and RESET temperature levels.

2.3.5: The Effect of Material Properties

Besides the I-V characteristics, the simulation is also able to investigate the impact of material properties on device behavior [19, 20]. Fig. 2.20 shows the impact of variations in surface diffusion and reduction activation energies on filament shape. When the activation energy of surface diffusion is large it is very difficult for the adsorbed ions to diffuse to a hole or step site. Thus, the filament more looks like a tree shape with lots of branches as shown in Fig. 2.20 (a) and this is very similar to electric breakdown [21]. On the other hand, figure (b) shows another situation where the surface diffusion energy barrier is small and hence adsorbed ions have enough time and are very easily diffuse to either step or hole sites. Thus, the filament topography is more like electro-chemical plating. Therefore, we see that a wide range of switching modes can be simulated using this methodology by appropriate choice of the system physical parameters.

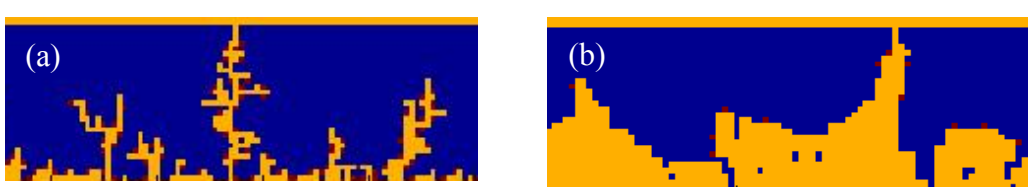


Figure 2.20. Simulated Filament shape formed under different conditions. (a) when the activation energy E_{surf} is large while E_{adatom} is small (b) when E_{surf} is small and E_{adatom} is large.

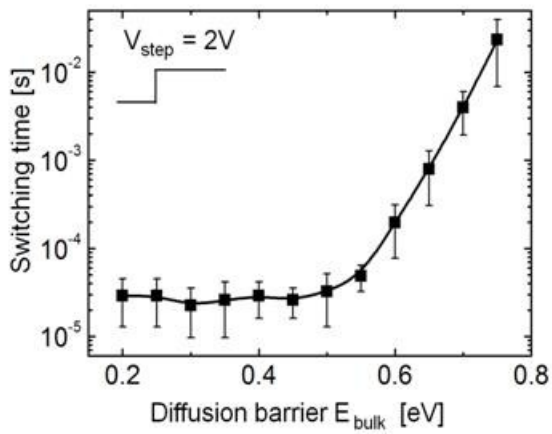


Figure 2.21. Simulated switching time versus diffusion barrier of the solid electrolyte.

Finally, the effect of the bulk diffusion energy barrier is studied. Fig. 2.21 shows the effect of E_{bulk} on the device switching properties. When the bulk diffusion energy barrier is large, the switching time exponentially depends on E_{Bulk}. Under this condition, the bulk diffusion is the rate limiting process for the entire system, therefore, the switching time t_{ON} decreases as the migration barrier of the insulator decreases (smaller E_{bulk}). However, when the E_{Bulk} further decreases, the oxidation and metal crystallization become rate-limiting processes and the switching time saturates as shown in the figure.

2.4: Summary

In this chapter, a Kinetic Monte Carlo simulator is established to study electrochemical metallization RRAM. The simulator is more focused on the forming and SET processes. Various relationships among the forming voltage, forming time, voltage sweep rate and filament topographies are studied. It is found that without filament overgrowth effect, the average width of the filament formed at large applied voltage is much narrower than that of the filament formed at small voltage condition. Experiments on Ag/Ag_2S and Cu/H_2O systems validate the simulation results. In addition, the RESET process is also simulated. It is found that it depends on the presence of the lateral electric field close to the cathode, and the filament can break at different locations. Finally, the impact of variations in surface diffusion and chemical reduction energy barriers on filament shape are also studied.

2.5: References

[1] A. F. Voter, “Introduction to the Kinetic Monte Carlo Method,” to be published in Radiation Effects in Solids, K. E. Sickafus and E.A. Kotomin, eds., Springer, 2005.

[2]A. P. J. Jansen “An introduction to Monte Carlo simulations of surface Reactions” arXiv:cond-mat/0303028v1

[3] C.C. Battaile “Kinetic Monte Carlo simulation of chemical vapor deposition” Annual Review of Materials Research**,** 2002. 32: 297-319

[4] F. Pan, S. Yin, V. Subramanian “A Comprehensive Simulation Study on Metal Conducting Filament Formation in Resistive Switching Memories” 3rd *IEEE International Memory Workshop (IMW),* 2011

[5] A. Saedi “A study on mutual interaction between atomistic and macroscopic phenomena during electrochemical processes using coupled finite difference-kinetic Monte Carlo model: Application to potential step test in simple copper sulfate bath” Journal of *Electroanalytical Chemistry* 588 (2006) 267-284

[6] R. Pornprasertsuk, T. Holme, Friedrich B. Prinz “Kinetic Monte Carlo simulations of solid oxide fuel cell” *Journal of the electrochemical Society*, 156 (12) B1406-B1416 (2009).

[7] F. Pan; S. Yin; V. Subramanian, "A Detailed Study of the Forming Stage of an Electrochemical Resistive Switching Memory by KMC Simulation," *Electron Device Letters, IEEE*, vol.32, no.7, pp.949-951, July 2011

[8] J. Jang, F. Pan, K. Braam, and V. Subramanian “Resistance Switching Characteristics of Solid Electrolyte Chalcogenide Ag_2Se Nanoparticles for Flexible Nonvolatile Memory Applications” *Advanced Materials*, Volume 24, Issue 26, pages 3573–3576, July 10, 2012

[9] C. Schindler, G. Staikov and R. Waser “Electrode kinetics of Cu-SiO_2-based resistive switching cells: overcoming the voltage-time dilemma of electrochemical metallization memories,” *Applied Physicas Letters*, 94. 072109. 2009.

[10] T. Sakamoto, N. Banno, N. Iguchi, H. Kawaura, H. Sunamura, S. Fujieda, K. Terabe, T. Hasegawa, and M. Aono, “A Ta2O5 solidelectrolyte switch with improved reliability,” *Symp. on VLSI Tech., Dig. Tech*. Papers, pp. 38-39, 2007.

[11] B.G. Willis, D.V. Lang, “Oxidation mechanism of ionic transport of copper in SiO2 dielectrics” *Thin Solid Films*, 467 pages: 284-293, 2004

[12] Michael D. Bronshtein, Renat R. Nazmutdinov and Wolfgang Schmickler “An approach to optimised calculations of the potential energy surfaces for the case of electron transfer reactions at a metal/solution interface” *Chemical Physics Letters* 399 (2004) 307–314

[13] Aron J. Cohen, Paula Mori-Sánchez, Weitao Yang “Insights into Current Limitations of Density Functional Theory” *Science* 321, 792, 2008

[14] Martin Zelený J. Hegedus, A. Foster, and R. M. Nieminen , D. A. Drabold and S. R. Elliott “A model for a non-volatile memory material: Ab initio study of Cu diffusion in SiO_2”, *Material Research Society Spring Meeting* 2011.

[15] John Newman and Karen E. Thomas-Alyea “Electrochemical Systems” third edition, Wiley-Interscience, 2004

[16] Abhijit U. Modak, Mark T. Lusk “Kinetic Monte Carlo simulation of a solid-oxide fuel cell: I. Open-circuit voltage and double layer structure” *Solid State Ionics* 176 (2005) 2181-2191
[17] Y. Bernard, P. Gonon, and V. Jousseaume, “Resistance switching of Cu/SiO2 memory cells studied under voltage and current-driven modes,” *Applied Physics Letters*, vol. 96, no. 19, pp. 193 502-1–193 502-3, May 2010.
[18] S. Kaya, A.R. Brown, A. Asenov, D. Magot, T. Linton "Analysis of Statistical Fluctuations due to Line Edge Roughness in Sub-0.1µm MOSFETs” *Simulation of Semiconductor Processes and Devices (SISPAD), International Conference on,* 2001.
[19] F. Pan and V. Subramanian, “A kinetic Monte Carlo study on the dynamic switching properties of electrochemical metallization RRAMs during the SET process,” *in Proc. Int. Conf. Simulation of Semiconductor Processes and Devices (SISPAD),* 2010, pp. 19–22.
[20] F. Pan, V. Subramanian “ Kinetic Monte Carlo simulation of resistive switching and filament growth in electrochemical RRAMS” *Device Research Conference*, 2010, page 255.
[21] L.Niemeyer, L Pietronero and H.J.Wiesmann, “Fractal dimension of dielectric breakdown,” *Physical Review Letters* 52, 1984, pp.1033–1037.

Chapter 3: Yttria Stabilized Zirconia Based RRAM

This Chapter studies the characteristics of RRAM devices based on amorphous Yttria Stabilized Zirconia (YSZ) metal oxide. In the first section, a brief introduction on YSZ is presented. Next, the effects of different top electrodes on the device performance are compared, and the underlining physics is investigated. Further, two kinds of devices, Cu/YSZ and Ti/YSZ, are studied in detail. This includes a comprehensive comparison of device performance parameters such as endurance, retention, reliability and so on. It is found that Ti/YSZ is much better than Cu/YSZ in most of the device characteristics. Physical explanations are given. In addition, this chapter also gives a brief introduction on 1/f noise and the noise level inside Cu/YSZ and Ti/YSZ are systemically investigated. It is found that generally Ti/YSZ has much lower noise level than Cu/YSZ and is more reliable for RRAM applications.

3.1: Introduction to Yttria Stabilized Zirconia

Yttria stabilized zirconia (YSZ) is a type of ceramic in which yttria (Y_2O_3) is added into zirconia (ZrO_2) to stabilize the crystal structure. Above 2370ºC, undoped ZrO_2 has cubic structure. It transforms to tetragonal phase between 1170 to 2370 ºC [1, 2]. When temperature drops below 1170 ºC it becomes a monoclinic phase. Due to this phase transition, undoped ZrO_2 usually cracks during the cooling process. By addition of material such as Y_2O_3, it allows the crystal to be stabilized in a cubic phase at room temperature.

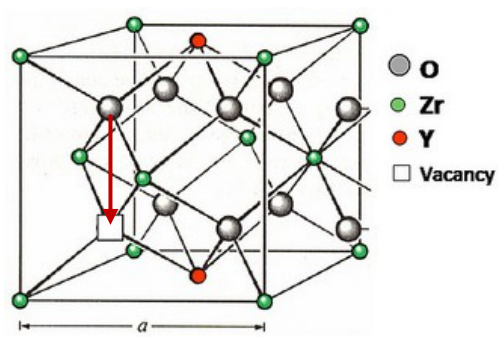


Figure 3.1. The crystal structure of YSZ [3]

The conventional cell structure of YSZ is show in Fig. 3.1. As shown, two Y^{3+} ions replace two Zr^{4+} in the zirconia lattice. Because of the Y^{3+} carries different number of charges from Zr^{4+}, one

O^{2-} ion has to be removed in order to maintain the charge neutrality. In Kroger-Vink notation, this process is written as [3]:

$$Y_2O_3 \rightarrow 2Y_{Zr}^{t} + 3O_O^{x} + \ddot{V}_O \tag{1}$$

Where Y_{Zr}^{t} indicates *Y* is incorporated into *Zr* site. $\ddot{V}_O$ is the vacancy in the oxygen site which carries two positive charges. O_O^{x} is the lattice oxygen which does not carry any net charge. Because of the oxygen vacancies created, YSZ typically has very large oxygen ionic conductivity at high temperature. The O^{2-} ions can hop through those vacancy sites as shown by the red arrow in Fig. 3.1. In addition, its ionic conductivity depends on the concentration of yttria. R. Pornprasertsuk et.al. [4] did an excellent KMC simulation study on the ionic conductivity of YSZ. Part of the results is shown in Fig. 3.2. The YSZ has peak conductivity when Y_2O_3 has a concentration around 8mol% and it gradually decreases after that as shown in Fig. 3.2 (a). This reduction of conductivity can be explained by the change of O^{2-} migration barrier. The experiments and simulation show that the migration barrier for oxygen ions is roughly a constant when the concentration of Y_2O_3 is less than 6 to 8mol%, yet it increases steadily after that as shown in Fig. 3.2 (b). The reason for the activation energy change is that when the Y_2O_3 concentration goes higher, O^{2-} have to more likely pass through Y-Y and Y-Zr edges instead of through space between Zr-Zr. Y^{3+} has a large ionic size than Zr^{4+}. Thus it makes the diffusion of O^{2-} more difficult [4].

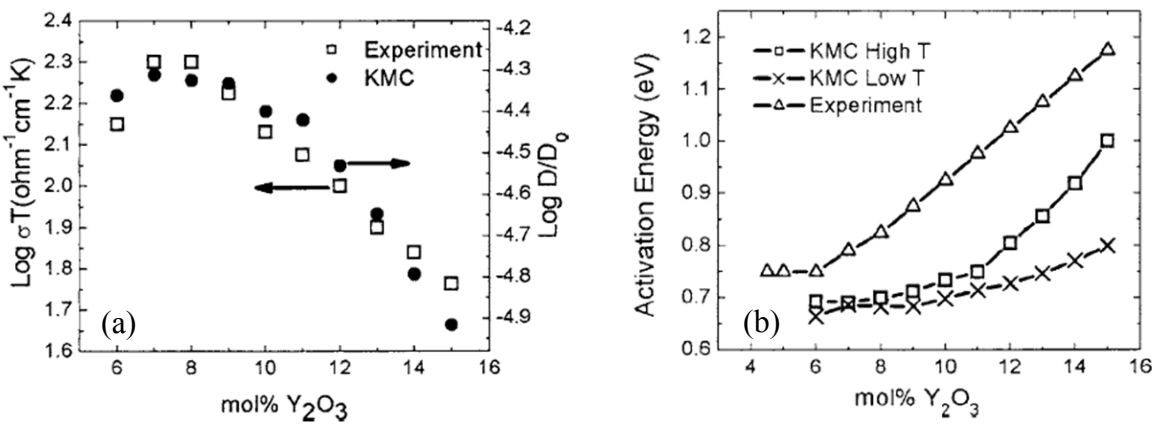


Figure 3.2. (a) The conductivity of YSZ vs the concentration of Y_2O_3 inside ZrO_2 (b) The relationship between the migration energy barrier of O^{2-} and the concentration of Y_2O_3 [4].

Because of its good ionic conductivity, YSZ is typically used for solid oxide fuel cell (SOFC) [4] and oxygen sensor applications [5]. Fig. 3.3 [5] shows a basic structure of an oxygen sensor. Region A is measured gas and region F contains the reference gas which typically is air. In

addition, region B is an optional protective layer that is used to protect electrode C. Electrode C is usually made of platinum and is catalytically active. Region D is the YSZ ceramic. Finally, region E is the reference electrode. During the operation the open circuit voltage V_{Open} follows Nernst equation [5]

$$V_{Open} = \frac{k_B T}{4q} \ln \frac{p^*_{O_2,REF}}{p^*_{O_2}} \tag{2}$$

where $p^*_{O_2,REF}$ is the oxygen partial pressure of reference gas and $p^*_{O_2}$ is the oxygen partial pressure of measured gas. Thus, by measuring the V_{Open}, the oxygen content of measured gas can in principle be obtained.

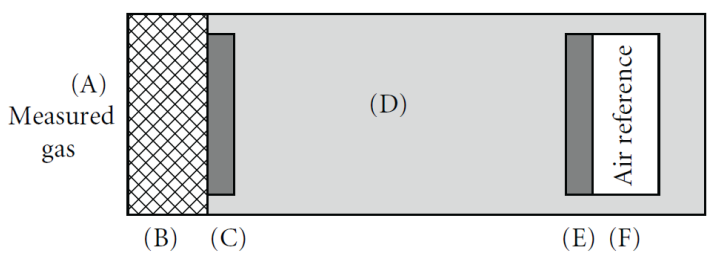


Figure 3.3. Basic structure of an oxygen sensor

Besides its electrochemical applications, YSZ is also a good candidate for high-k dielectric used for semiconductor gate oxide. In fact, it has wideband gap of 5.1-7.8 ev [6] and also has good chemical stability and high resistivity and can be heteroepitaxially grown on Si [7]. For gate dielectric material applications, the amorphous film may be preferred to reduce the leakage current. In fact, research [8] shows that the leakage current of amorphous YSZ films is smaller than that of ZrO_2 due to the addition of yttria.

Even though there are wide applications for YSZ in above-mentioned areas, it has been unstudied for RRAM application. In fact, YSZ has several good properties which in principle are suitable for such types of memory devices. First, YSZ is chemically stable and has very low leakage current, which is good for off state since it may give a larger resistance window between R_{on} and R_{off}. In addition, the high conductivity of O^{2-} in principle make the RESET process very effective since a sufficient amount of O^{2-} from surrounding areas can drift to the VO site in a timely manner to complete the recombination process. Moreover, as will be discussed in a later section, there is another advantage of YSZ. Since both Y and Zr are very difficult to precipitate, a device based on YSZ may have a better chance to have smooth I-V characteristics in the RESET process, which makes it suitable for MLP application at the RESET stage.

In the following sections, the characteristics and mechanisms of RRAM devices based on amorphous YSZ are systematically studied. The detailed experimental fabrication procedures of YSZ devices are described. Next, the effects of different top electrodes of Au, Cu, Ni, Al and Ti are investigated and their device performance parameters are compared. In addition, the characteristics and the mechanisms of Ti/YSZ and Cu/YSZ are studied in great details, which include endurance and retention comparison; the physics that give rises to the observed differences are studied. Next the noise present inside Cu/YSZ and Ti/YSZ is measured. It is found that Ti/YSZ has much lower noise level than that of Cu/YSZ. In particular, Ti/YSZ has a noise recovery capability that makes it more suitable for memory applications.

3.2: Experimental Procedures

Two types of device structures as shown in Fig. 3.4 are fabricated. Device structure I is relatively simple and easy to fabricate. Thus, this structure is used for device screening purpose. For structure I, 30nm 8mol% YSZ was deposited on Au/Cr/SiO_2/Si substrate by physical vapor deposition via an electron-beam evaporation method with a deposition rate of around 0.5nm/s at a pressure of 9e-6 torr. The Au layer serves as a bottom electrode. After that, the YSZ film is treated with UVO for 10 minutes. Subsequently, different top electrodes such as Ti, Al, Au, Cu, Ag and Ni were either thermally evaporated or e-beam deposited at a pressure less than 1e-5 torr and patterned with a shadow mask. The size of all the top electrodes is 100×200um^2.

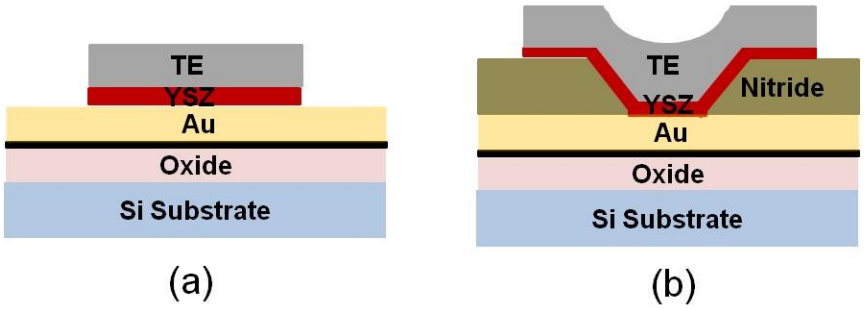


Figure 3.4. Two types device structures

Structure II on the other hand is more complicated and its fabrication process flow is described below. It starts with a 4 inch N type silicon wafer covered by 100nm thermally grown oxide layer and this serves as the device substrate. The first step is the cleaning process. A standard wafer cleaning process is conducted to remove organic contamination and heavy metal ions on the wafer surface. The cleaning process was done by immersing the wafer into

piranha solution, which is made of the mixture of hydrogen peroxide and hot sulfuric acid at 120°C, for 10 minutes. After that, the wafer is rinsed in DI water and followed by a 10 minute spin rinse dry (SRD) process. The wafer status after this cleaning step is shown in Fig. 3.5 (a).

After that the wafer is loaded into a thermal evaporator chamber. Chrome and gold sources are added into separated boats. Then the evaporator is pumped down to a pressure of 100mtor through a mechanical pump. After that a cryopump is turned on to reach a base pressure around 3×10^{-6} torr. During the evaporation process, a 3nm Cr layer was deposited first to serve as an adhesion layer, 80 nm Au layer was then deposited which serves as the bottom electrode. After this step, the wafer status is shown as in Fig. 3.5 (b).

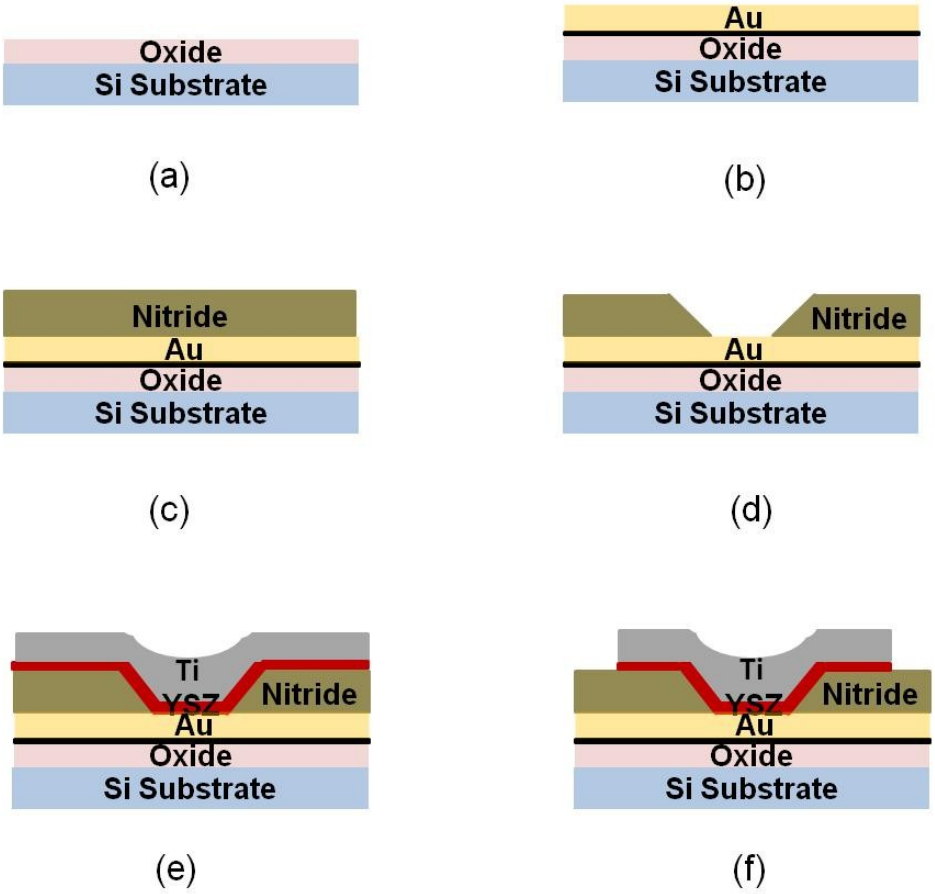


Figure 3.5. Experimental procedures for device structure II

The wafer is then ready for the Si_3N_4 isolation layer deposition. The nitride layer was deposited by plasma enhanced chemical vapor deposition (PECVD). After the wafer is loaded

into the chamber, the system goes through a pre-pump and pre-purge step. Then the chamber pressure is pumped down and stabilized at 0.2 tor. At the same time, the chamber temperature is raised up to 350°C which is the temperature required by the following chemical reactions. The next step is the deposition step. During the deposition, 200 sccm NH_3, 40 sccm SiH_4 and 200 sccm Ar gases circulate inside the chamber and the chamber pressure is maintained at 0.9 torr. At 25W power supply, the deposition rate is roughly 17nm per minute. After 100nm nitride layer was deposited, the chamber goes through another three purge cycles. After this step, the wafer status is shown as Fig. 3.5 (c).

In the following steps, the nitride layer is patterned. The wafer first goes through a dehydration process inside the primeoven for 30minutes, then HMDS vapor is injected inside the oven such that the wafer can be coated with a monolayer of HMDS which improves the photoresist (PR) adhesion to the wafer surface. Immediately after the HMDS coating, the wafer is spin-coated with 1.1µm I-line PR at a spin speed of 4100 round per minute (RPM) and followed by 1 minute soft bake at a temperature of 90°C. After soft bake, the PR is exposed in soft contact mode using a Karl Suss MA6 Mask Aligner. The exposed PR was then developed using OPD 4262 positive PR developer. Immediately after that a postbake process is performed in an oven at 120°C for 30mins. Next, the nitride layer is patterned by plasma dry etching using a 90 sccm SF_6 and O_2 gas mixture at a power of 200W. This gives an approximate etching rate of 125nm/min.

After the etching step, the residual PR is removed by placing the wafer inside a resist strip bath containing PRS-3000 at 80°C for 20 minutes. Then the wafer is rinsed and dried. After this step, the wafer status is as shown in Fig. 3.5 (d). Holes on the Si_3N_4 layer with different areas are therefore created by the patterning process above. The effective area size varies from 64×64 um^2 to 4×4 um^2 and it effectively defines the final device electrode contact size.

In the next step, wafer is loaded into the E-beam evaporator chamber. Ti and YSZ sources are loaded respectively. After the rough and fine pumping processes, the pressure of the chamber reduces to 1e-5 torr. YSZ layers with different thickness ranging from 3nm to 30nm were then deposited. 140nm Ti or Cu top electrodes are deposited subsequently. After this step, the wafer status is as shown in Fig. 3.5 (e).

Finally, another photolithography step is conducted to define the top electrode. The top electrode size is not critical in this case and is set to be 200µm by 200µm. The wafer is then etched in 10% buffer oxide etchant (BOE) to remove the residual Ti and YSZ layer (ion milling process is used to pattern the copper electrode). After that another plasma etching step is performed to finally remove the residual nitride so as to enable probing of the bottom electrode. After all those processes, the final device structure is shown in Fig. 3.5 (f).

3.3: Effect of Different Top Electrodes and Switching Mechanisms

In this section, effects of different electrodes on the device performance are analyzed and the filament conduction mechanisms associated with the various types of electrodes are discussed. The I-V characteristics in Forming, SET and RESET stages are discussed in detail. Next, a detailed comparison of device performance between Cu/YSZ and Ti/YSZ is conducted.

First the effect of various electrodes on the device performance was studied. For this purpose, device structure I is adopted. Devices with Au, Ni top electrodes don't have any switching behavior, and always get stuck at LRS after the forming process. Due to the simple characteristics, they are not further discussed.

Fig. 3.6 shows the V_{Form} distribution for devices with different top electrodes. Device with Au top electrodes have the largest forming voltage and shows a very uniform distribution. The breakdown (BD) electric field is around 5MV/cm, which is comparable to the BD field of 4MV/cm for crystalline YSZ film epitaxially grown on Si substrates [10] and is larger than the BD field of polycrystalline YSZ [11].

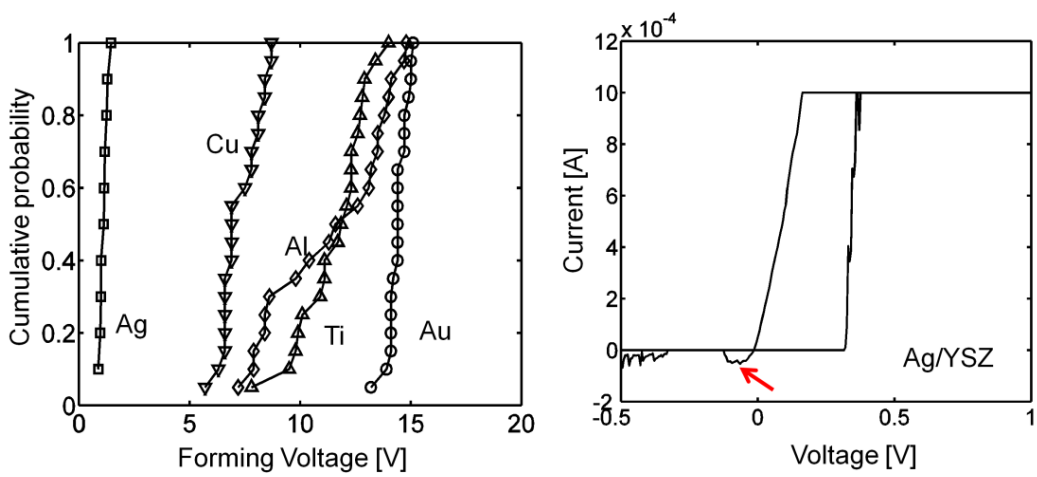


Figure 3.6. (a) The forming voltage distribution for various top electrodes. The device structure I is used. (b) The typical I-V characteristics for Ag/YSZ device. The red arrow shows the very small V_{RESET}

Devices with either Ti and Al top electrodes show a reduced forming voltage with larger distribution. Even though those two kinds of devices have different types of conduction mechanisms (as will be discussed soon), the reduction in forming voltage can be attributed to the chemical reactions at the interface between Ti, Al and YSZ dielectric layer [12]. A very thin layer of Ti and Al at the interface gets oxidized and therefore oxygen vacancies are created

inside the YSZ film which weakens its dielectric strength. The variation can be attributed to the uncontrolled spontaneous chemical reaction.

On the other hand, the devices with Cu and Ag top electrodes show much smaller forming voltage. This indicates that Cu and Ag can easily diffuse into the YSZ films. In addition, Ag has a much larger diffusivity than that of Copper and thus has the smallest forming voltage. Generally, because Ag is too diffusive it lacks good voltage control. In addition, devices with Ag electrodes shows a very small V_{RESET} as shown in Fig. 3.6 (b), which is very vulnerable to noise and read disturbance. Besides, Ag/YSZ can get easily stuck in LRS. Therefore, Ag/YSZ is not further discussed either.

Table I

Device Performance Comparison

Structure I	Yield	Endurance
Ti/YSZ/Au	≈ 60%	$>10^4$
Cu/YSZ/Au	≈ 80%	≈100
Al/YSZ/Au	< 10%	< 10

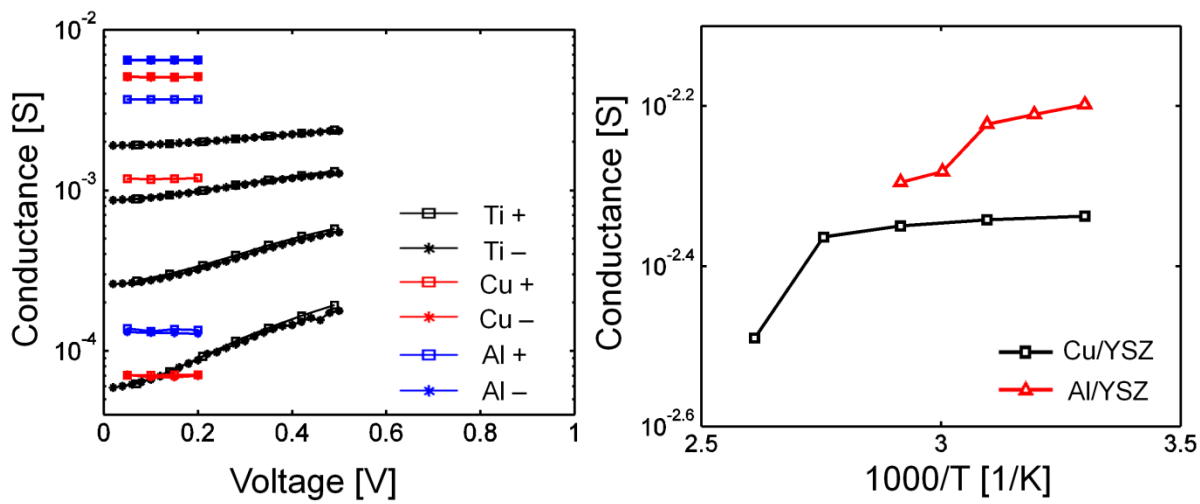


Figure 3.7. (a) Voltage (and its polarity) dependence of the conductance of Ti/YSZ, Cu/YSZ and Al/YSZ at different states. (b) The temperature dependence of the conductance for Cu/YSZ and Al/YSZ

Table 1 shows the yield and endurance comparisons for devices with Ti, Al and Cu top electrodes. Even though Cu/YSZ has the highest yield, Ti/YSZ has much better endurance. Al/YSZ on the other hand, has both very low yield and endurance. A previous report [13] noticed that there is a significant difference in the formation energy of Al_2O_3 and TiO_2, yet the difference of average formation energy per mole of oxygen is not significant. Thus, this alone cannot explain the large difference in both yield and endurance between Ti/YSZ and Al/YSZ.

It was noticed that all the Al/YSZ devices in LRS show perfect ohmic conduction as shown in Fig. 3.7 (a) which is an indication of a metallic type of filament formation. On the other hand, R_{on} of Ti/YSZ shows a voltage dependent behavior, which is an indication of VO based electron hopping conduction. To further confirm our speculation, temperature dependent measurements were carried out. It is found that in the LRS, the conductance of device Cu/YSZ, Al/YSZ decreases as the temperature increases (Fig. 3.7 (b)), which is opposite to Ti/YSZ. Thus, a metallic type of conduction in Al/YSZ is confirmed. Another piece of evidence is that researcher [14] found that for SiO_2 film prepared either by spin-on deposition or PECVD, the extracted diffusion barrier for Al ion is much smaller than that of Cu ions. Thus Al should be much more diffusive. And the overall stable Al/SiO_2 interface (compared with Cu/SiO_2) can only be attributed to the aluminum oxide formed at the interface which needs much more energy to get dissociated and free the Al ions than that required to free the Cu ions from Cu_2O.

Based on the above discussions, an explanation of low endurance of Al/YSZ can be obtained—even though a larger V_{FORM} (compared with Cu/YSZ) is required to free the Al ion from the Al_2O_3, once Al ion is free, it can easily diffuse through the YSZ and form a metallic type filament. However, since Al is much easier to get oxidized and more difficult to precipitate, the formation of a stable filament is difficult, thus both the yield and endurance of Al/YSZ are poor. In addition, as large amount of Al ions get trapped inside the YSZ film, the device is easily stuck at LRS. Due to the poor endurance of Al/YSZ, in the following sections only Ti/YSZ and Cu/YSZ are discussed.

3.4: Ti/YSZ and Cu/YSZ Devices

3.4.1: I-V Characteristics for Ti/YSZ and Cu/YSZ

The forming I-Vs of Ti/YSZ based RRAM devices are measured at different voltage polarity and temperature conditions as shown in Fig. 3.8 (a). The measurement is done at 30ºC, 50°C and 70°C respectively. It can be seen that in both cases the I-V follows the Poole-Frenkle emission [15]:

$$J = C_1 E \exp\left\{ -q[\phi - (qE/\pi\varepsilon_0\varepsilon_d)^{1/2}]/k_B T \right\} \quad (3)$$

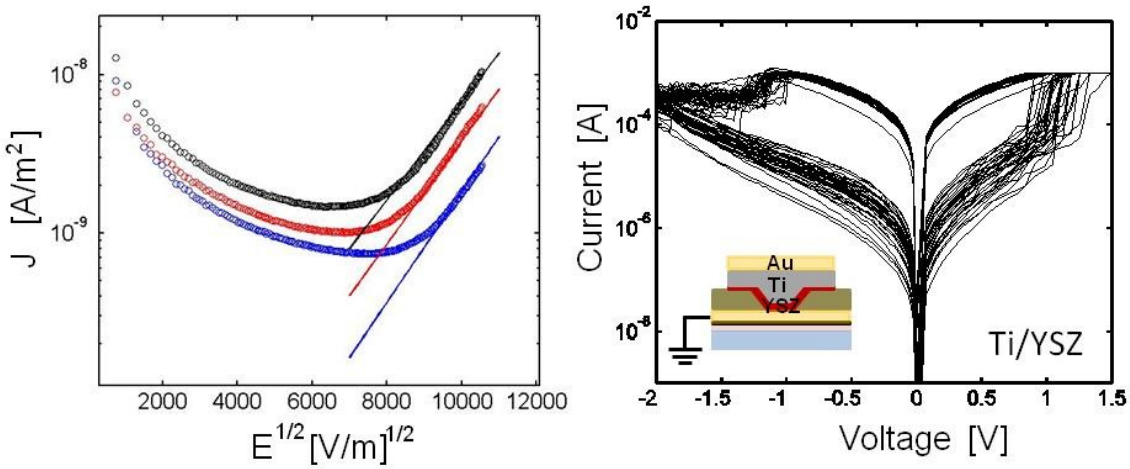


Figure 3.8. (a) Experimental (symbols) and simulated (lines) results of the forming stage I-V for Ti/YSZ. The measurements are taken at 30ºC, 50°C and 70°C respectively. (b) 300 repetitive I-V sweeps for a typical Ti/YSZ device (The sweep count was limited by practical measurement time). YSZ film thickness is 10nm. Device area is 16um by 16um.

where E is the electric field, ϕ is the potential barrier height, and ε_d is the dielectric constant for the insulator and C_1 is a function of density of the trapping centers [15].

When Ti is positively biased, the extracted depth of the effective trap potential barrier is 0.51 ev and the extracted dielectric constant ε_d is 13.02. As mentioned by Sze [15], the extracted dielectric constant value is critical. The space-charge limited conduction effect could also fit the plot of ln*I* vs $V^{0.5}$ type, however, it gives an estimation of ε_d an order of magnitude less than the one obtained from Poole-Frenkel relationship. Thus, C-V measurement was further conducted and the experimentally measured ε_d is 15.2 and is very close to the extracted value. This supports our assumptions and it confirms the current is bulk limited in the forming stage.

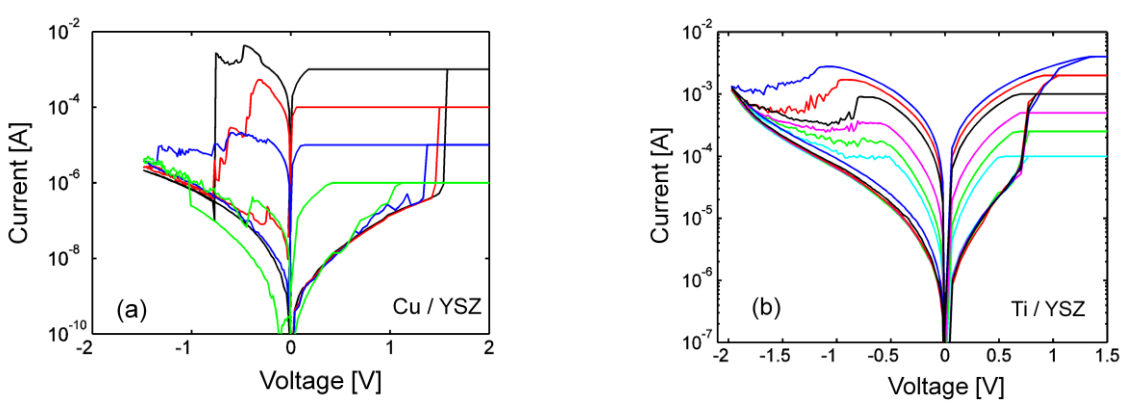


Figure 3.9. The I-V characteristics of (a) Cu/YSZ and (b) Ti/YSZ at different current compliance levels. YSZ film thickness is 10nm. Device area is 16um by 16um.

Fig. 3.8 (b) shows 300 repetitive DC sweeping I-Vs of Ti/YSZ. It shows a very uniform distribution of V_{SET} and V_{RESET}. In addition, Fig. 3.9 shows the I-Vs for Cu/YSZ and Ti/YSZ. For both types of devices, the on-state resistance can be related to SET CC as shown in Fig. 3.10. In addition, the resistance decreases roughly exponentially with the CC. For Ti/YSZ,

$$R_{on} \times I_{CC} \approx 2 \tag{4}$$

For Cu/YSZ,

$$R_{on} \times I_{CC} \approx 0.3 \tag{5}$$

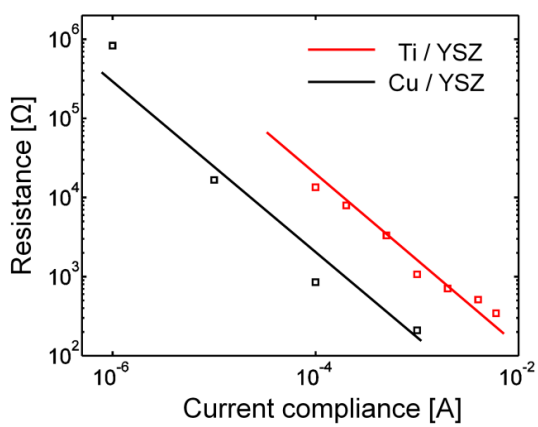


Figure 3.10. The relationship between R_{on} and current compliance for both Cu/YSZ and Ti/YSZ

From equation (4) and (5), it can be seen that Cu/YSZ is more power efficient. This is due to the high diffusivity of Cu and also to the high conductivity of Cu metallic filament.

For memory array applications, smaller I_{RESET} and I_{SET} are desired to reduce the power consumption. In Fig. 3.8 (b), $I_{RESET} \approx I_{SET}$ = 1mA for Ti/YSZ which is too large for a realistic memory array structure. Fortunately, I_{RESET} can be linearly scaled with I_{SET} at least down to 100µA as shown in Fig. 3.11. In addition, since a larger voltage-sweep stop value is adopted for V_{RESET} (4V in this case), the resistance window is much larger than that shown in Fig. 3.9(b). This makes multi-level programming possible as will be discussed later.

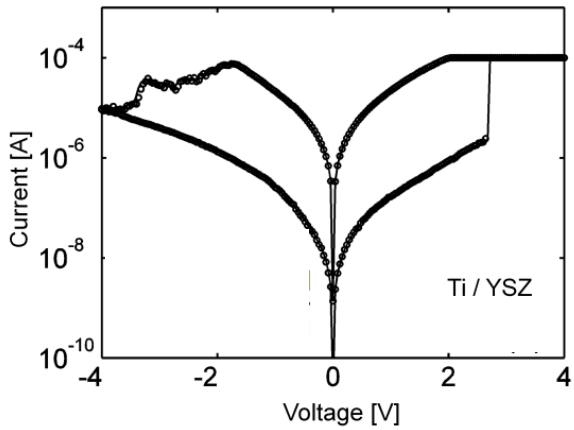


Figure 3.11. I-V characteristics for Ti/YSZ and CC is 100µA

During the actual programming cycles, it is difficult to use CC to precisely control R_{on}. Fig. 3.12 shows an example based on a Ti/YSZ device and a single pulse programming scheme. In the measurement V_{SET} is 2V and V_{RESET} is set to -2.5V. In addition, in the test the CC is fixed at 0.5mA.

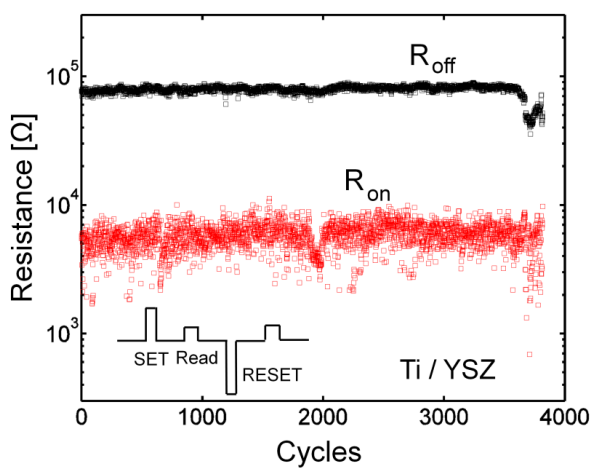


Figure 3.12. The variation of R_{on} due to the capacitance charging effect

As can be seen, there is a large spread of R_{on}. This is mainly due to the capacitance charging effect [17]. In practice, there is always parasitic capacitance present in the system. It arises from both the device itself and the external circuit. During the SET process, the parasitic capacitance can generate a large discharging current which results in a current spike whose

value is much larger than the CC during a very short time period. This current spike can significantly reduce R_{on} and sometimes may even break the device permanently. In addition, due to the stochastic nature of RRAM, the local topography of the filament varies from cycle to cycle and because its self capacitance directly relates to the filament topography, the self capacitance varies. This can also contribute to the spread of R_{on}.

In addition, as can be shown in Fig. 3.13 (a), the R_{on} of Ti/YSZ becomes less voltage dependent and approaches ohmic conduction when its conductance increases by applying larger CC. In addition, its conductance increases as the voltage increases. This is a clear indication of an oxygen vacancy based electron hopping conduction mechanism. The applied voltage not only lowers the potential barriers but also raises the Fermi level and thus enhances the filament conductance. We further notice that the R_{on} of Ti/YSZ is independent of the polarity of applied voltage as shown in Fig. 3.7(a). Such a perfect symmetry may indicate the limiting potential barrier exists not at the metal/dielectric interface but inside the dielectric itself, where a symmetric barrier is present.

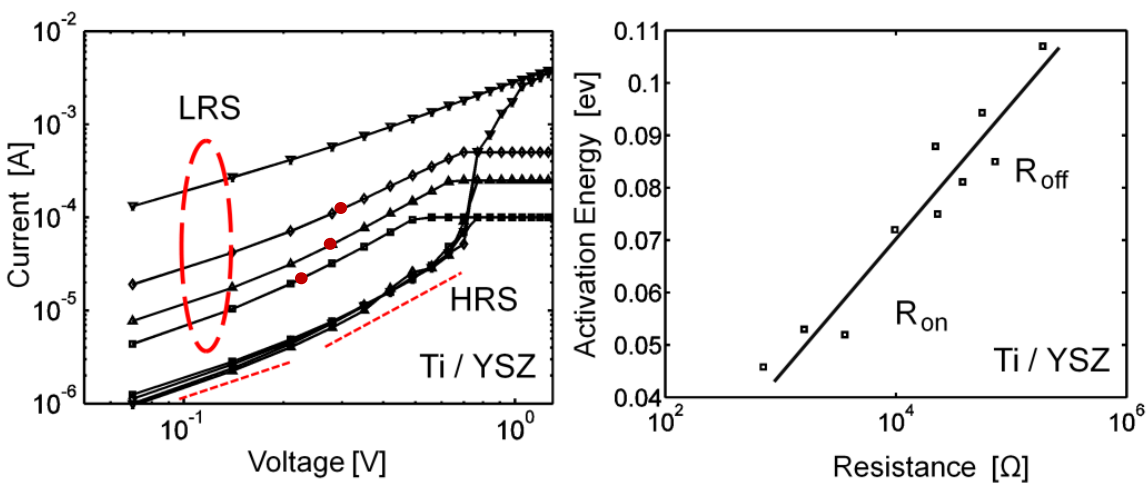


Figure 3.13. (a) The SET stage I-V of Ti/YSZ . The dashed lines indicate slopes of 1 and 2 respectively (b) Extracted activation energy at different resistance levels

Next, we plotted the I-V data of Ti/YSZ using a log-log scale as shown in Fig. 3.13 (a). At both LRS and HRS states, ohmic conduction is observed at the low voltage bias regime. As the voltage increases, the off state current gradually becomes $\propto V^2$. The transition points where the current transits from ohmic to V^2 are labeled by the red solid dots. It can be seen that as the R_{on} of the LRS decreases, the transition point become more positive. This is expected, since the current is more ohmic when the R_{on} is smaller. The V^2 type current may be an indication of

space charge limited conduction (SCLC) [18, 19]. According to SCLC theory, without any traps, the current density is described by the Mott-Gurney Law:

$$j = \frac{9}{8}\mu\varepsilon\varepsilon_0 \frac{V^2}{L^3} \tag{6}$$

where μ is the mobility, ε is the dielectric constant, V is the applied voltage and L is the device length. If there is a single trap level at energy E below the conduction band the above equation is modified and becomes

$$J = \frac{9}{8}\mu\varepsilon\varepsilon_0 \frac{V^2}{L^3}\theta_0 \tag{7}$$

where θ_0 is the ratio between free charge density ρ_f and total charge which is the sum of ρ_f and trapped charge ρ_t

$$\frac{\rho_f}{\rho_f + \rho_t} = \theta_0 = \frac{N_c}{N_t}\exp\left(-\frac{E}{k_B T}\right) \tag{8}$$

where N_c is the effective density of states in conduction band and N_t is the trap density.

At LRS, depending on the current compliance, θ_0 varies. When the current compliance is large, more traps are generated. However, it would be wrong to conclude that in this case θ_0 would decrease instead of increasing. The key point is that when the trap density goes very high, they form a trap sub band, and the barrier between traps gets lowered. At the same time, the free electron concentration increase significantly and most of the trap sites are occupied by electrons. As a result, the θ_0 increases instead of decreasing, which results in larger device conductance. When the current compliance is small, a smaller number of traps are generated and those traps tend to be isolated from each other. Electrons therefore get localized. Therefore, the overall effect is a decrease of θ_0 and a decrease in device conductance.

In order to investigate the current conduction mechanism, we further conducted temperature-dependent I-V measurements. We found that when the temperature increases, the current increases in all cases. Further, generally the current in the ohmic regime follows a simple relationship [16]:

$$J \propto E\exp(-E_{ac}/k_B T) \tag{9}$$

where E_{ac} is the activation energy for electron. It turns out the extracted barrier height depends on the resistance of the device as shown in Fig. 3.13 (b). It can be seen that, as stated earlier, the barrier increases as the device resistance increases. This discloses the fact that the barrier height directly relates to the concentration of oxygen vacancies. When the density of the defects is small, there is a larger hopping barrier for the electron to hop from one site to another. On the other hand, when a larger amount of vacancies are created by increasing CC,

the barrier gets reduced and thus the electrons gradually get delocalized and the transport within the filament approaches ohmic conduction.

So far, the above evidence seems to support the fact that various resistance states are mainly due to Schottky thermal emission and space charge based conduction mechanism. However, the electron tunneling model cannot be totally excluded. Even though to a first order approximation, both direct and F-N tunneling mechanisms are all temperature independent [16], a detailed theoretical analysis discloses that they do depend on temperature [20]. To the best of author' knowledge, an explicit mathematical formulism to describe the temperature dependence of the tunneling current is still lacking. Thus, further study is still needed to confirm the conduction mechanisms.

3.4.2: Endurance, Retention and Uniformity of Ti/YSZ and Cu/YSZ

In this section, the endurance, retention and device uniformity of Ti/YSZ and Cu/YSZ are compared. It is found that Cu/YSZ is inferior to Ti/YSZ in all three aspects. The detailed physics are also explained.

The endurance test is done by adopting an incremental step pulse programming (ISPP) scheme as shown in figure 3.14. During the ISPP measurement, both the V_{SET} and V_{RESET} pulses increase or decrease in steps with interspersed verification reads between every pulse. The targeted R_{on} and R_{off} are predefined. If the targeted resistance is achieved and verified to be stable, the programming voltage and the resistance are recorded. Otherwise, the voltage pulse will keep increasing or decreasing. A device failure occurs when the required resistance level is not achieved by a predefined voltage limit. As will be discussed in later sections, such a programming methodology significantly improves the device reliability compared with a single pulse scheme.

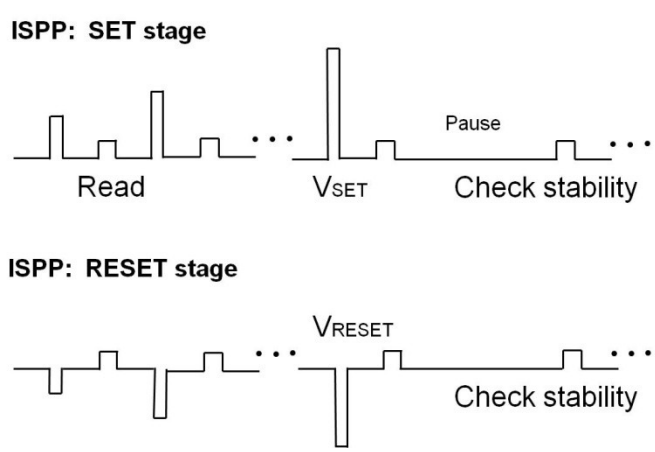


Figure 3.14. The incremental step pulse programming methodology

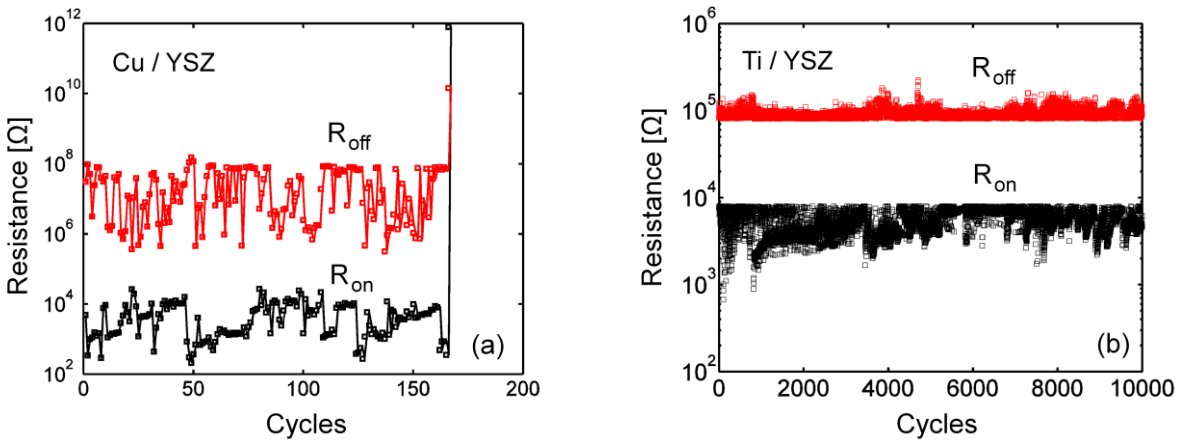


Figure 3.15. Endurance comparison between Cu/YSZ and Ti/YSZ. The ISPP methodology is used in the measurement. (a) Endurance of Cu/YSZ (b) Endurance of Ti/YSZ

Fig. 3.15 shows the endurance comparison between Cu/YSZ and Ti/YSZ. For both devices, the targeted R_{on} is $10\,k\Omega$ and the targeted R_{off} is around $100\,k\Omega$. Surprisingly, the best performance device of Cu/YSZ only shows an endurance of around 180 cycles, while for almost all the Ti/YSZ devices the endurance is beyond 10^4 cycles, which is limited only by the measurement time. In addition, the uniformity of the V_{SET} and V_{RESET} for Cu/YSZ is very bad. As can be seen in Fig. 3.16 (a) Cu/YSZ has a very large V_{SET} variation from cycle to cycle. On the other hand, Ti/YSZ shows a much smaller variation.

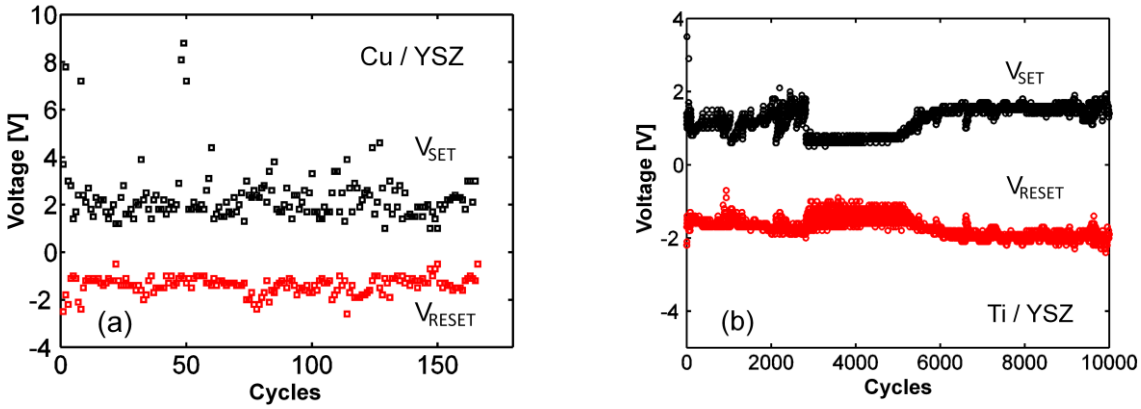


Figure 3.16. SET and RESET voltage distribution for Cu/YSZ (a) and Ti/YSZ device (b)

The physical reason for such poor endurance and uniformity of Cu/YSZ is not very clear yet. Similar phenomena are also observed in Pt/NiO/Pt [21] and Ru/HfOx/TiN [22] stacks. From the available literature results, it seems the interface between the electrode and the dielectric layer has an important impact on uniformity and reliability. First all of, as mentioned above, Cu/YSZ is an ECM based RRAM. In such systems, the metal crystallization is the key process to form the conductive filament. Thus, the requirement for a high-quality interface is very stringent. A 'bad' interface (for example, significant difference in crystal orientation between the resistive layer and the electrodes or the unclean electrode surface) may have a detrimental effect on device performance. This argument is supported by systems like Pt/NiO/Pt [21] as shown in Fig. 3.17.

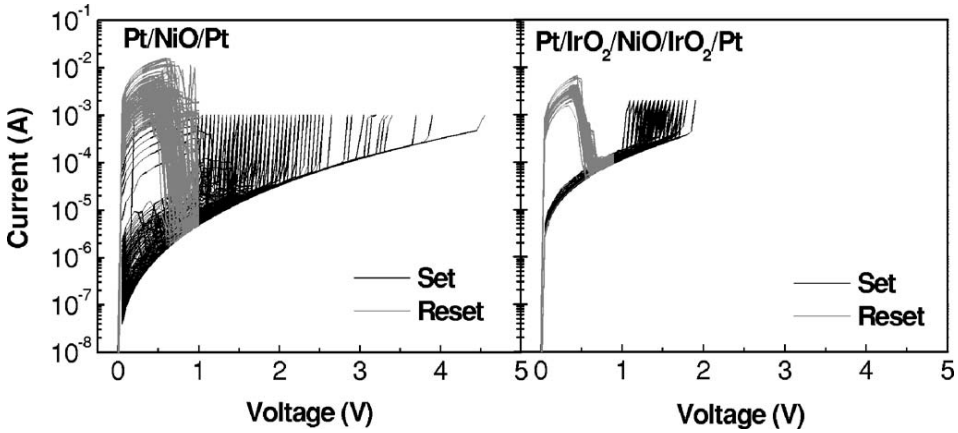


Figure 3.17. I-V characteristics of Pt/NiO/Pt stack (left) and Pt/IrO_2/NiO/IrO_2/Pt stack [21].

In this system, researchers find that there is a large V_{SET} variation as shown in Fig. 3.17 (a). In addition it also has a very poor endurance of less than 200 cycles. Both of those phenomena are quite similar to what is observed in the Cu/YSZ/Au system studied herein. The authors then added a very thin layer of conductive oxide IrO_2 on both sides of NiO layer. The resulting device performance gets significantly improved as shown in Fig. 3.17 (b), where the V_{SET} variation is dramatically reduced. Through Transmission Electron Microscopy (TEM), the authors found that IrO_2 layer enhances the crystallinity of NiO at the NiO-IrO2 interfaces. Thus, the improvement of the device performance is attributed to the change of the crystal structure at the interface.

Besides the microstructure of the film at the interface, the electrode surface cleanness is also another factor. T. Bakhishev [23] shows that the organic contamination presented on the inert electrode side can have a significant negative impact on the device endurance since those contaminants can prevent the crystallization process on the inert electrode and make the device on-state unstable.

Even though the ECM based RRAMs have the aforementioned limitations, it is improper to assume all such systems have poor device performance. In fact, literature results [24] shows

that high endurance of around 10^4 cycles can be obtain in Cu / amophousTa_2O_5. Therefore a relatively good ECM device is still possible if very good process control is adopted. Nevertheless, generally speaking the VO based RRAM tend to have better device performance than ECM based RRAM in terms of endurance and uniformity.

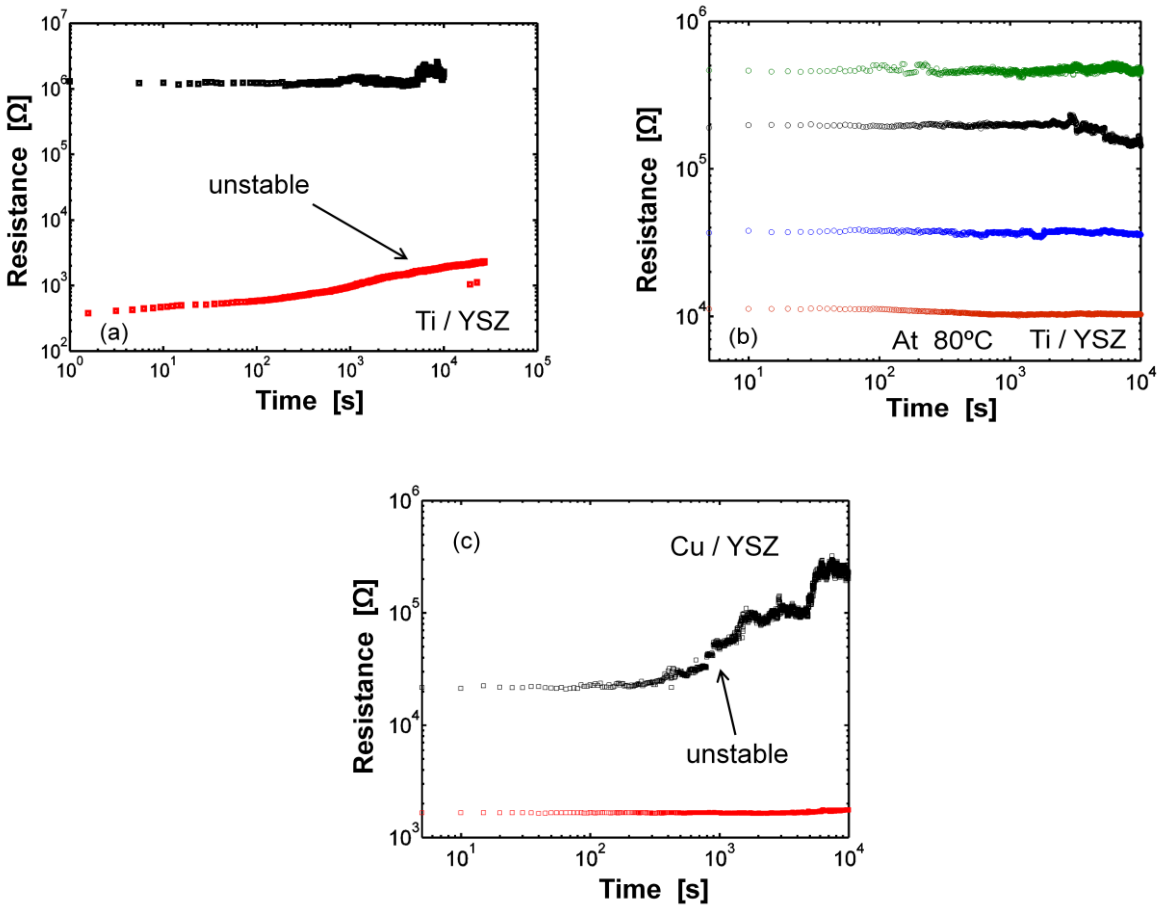


Figure 3.18. The retention properties of Cu/YSZ and Ti/YSZ at different resistance states

In addition, the retention properties of Cu/YSZ and Ti/YSZ have also been studied. The measurement results are shown in Fig. 3.18. Typically, Ti/YSZ shows a stable R_{off} ranging from 20kΩ up to 1MΩ (tested on separate devices). In the SET stage, its R_{on} is stable close to $10\,k\Omega$, yet is very unstable for R_{on} around 1kΩ. Interestingly, Cu/YSZ shows a different trend as shown in Fig. 3.18 (c). It is stable at very low resistance state when R_{on} is around 1kΩ, yet it becomes very unstable when its resistance reaches 20kΩ.

The underlying physics of those phenomena are investigated. It is found that R_{on} of Ti/YSZ is typically independent of the device area. Yet, R_{off} is inversely proportional to it. Therefore, it is reasonable to conclude that the on-state of Ti/YSZ devices is based on localized filament conduction. In the off state, on the other hand, the device is based on bulk conduction. Bulk YSZ

is thermally and chemically stable, which results in a stable R_{off}. However, for the on state, since yttrium cations tend to have very strong bonding with oxygen anion, VO formed during the SET process tends to recombine with O^{2-} to form a chemically stable phase. Therefore, R_{on} more likely increases as the recombination process goes on. In addition, this degradation process may be accelerated due to poor device encapsulation, which facilitates absorption of oxygen or moisture from the environment and results in a degradation of on state retention. This is shown in the figure above, where the R_{on} of Ti /YSZ starts from 400Ω gradually increases to 3kΩ in 10^4 seconds.

The Cu/YSZ device, however, shows the opposite trend. First of all, as will be discussed in the next section, unlike Ti/YSZ , the R_{off} of Cu/YSZ typically is a constant. In addition, as stated above, its R_{on} can be tuned by applying different CC in the SET operation. Generally, R_{off} of Cu/YSZ is very stable. However, the stability of R_{on} depends on the resistance level. For R_{on} less than 1kΩ, it usually has a very stable retention property (opposite to Ti/YSZ). However, the retention becomes very poor when R_{on} becomes around $20\,k\Omega$ as shown in Fig. 3.18 (c). This can be attributed to the filament width. When R_{on} is small, the width of the copper filament is large. Thus, it gives rise to a stable retention behavior. On the other hand, when the R_{on} increases, the filament width decreases accordingly. Thus, the retention becomes poor since a small amount of lateral diffusion of copper atoms will make the effective filament width decrease significantly and thus make the device resistance go up. As will be shown in next chapter, because of the retention difference between Cu/YSZ and Ti/YSZ, it is more desirable to use Ti/YSZ to achieve multilevel programming capability.

3.4.3: Noise Analysis for Ti/YSZ and Cu/YSZ

One of the reliability issues associated with RRAMs is noise. Due to the filament conduction nature, RRAMs are vulnerable to random motion or fluctuation of surrounding atoms, ions and traps. Noise such as random telegraph noise (RTN) and 1/f noise flicker is frequently observed in RRAM systems [25]. Those noise phenomena can increase the reading error rate and thus should be maximally suppressed. In addition, it is also important to study the dynamic evolution of the system noise behavior during the cycling process since a typical memory device should be stable after many programming cycles. In this section, the basic physics of the RTN and its relationship with 1/f flicker noise is reviewed. After that a noise study and comparison are conducted for Ti/YSZ and Cu/YSZ devices. Finally, the cycling dependence of the noise behavior on the Ti/YSZ system is investigated.

The noise in a system is generally described by a stochastic process. In addition, most of the noise we encounter is a stationary process whose joint probability distribution doesn't change with time and position. If the random process is denoted as x(t) then its autocorrelation c(t) can be written as [26]

$$c(\tau) = < x(t)x(t+\tau) > = \lim_{T\to\infty}\int_{-T/2}^{T/2} x(t)x(t+\tau)dt \qquad (10)$$

Furthermore, its power spectral density can be written as

$$S(\omega) = 4\int_0^{\infty} c(t)\cos\omega t dt \tag{11}$$

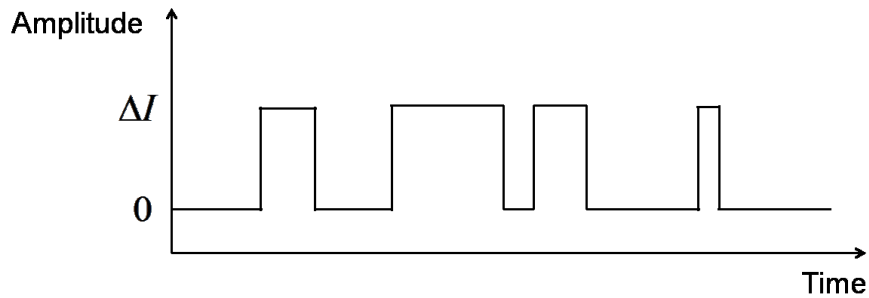


Figure 3.19. Illustration of Random Telegraph Noise

The typical feature of RTN is shown in Fig. 3.19. The system toggles between two states x_0 and x_1 and it is convenient to define $x_0 = 0$ and $x_1 = \Delta I$. In addition, if the transition rate from state 1 to state 0 is given by $1/\tau_1$ and from state 0 to state 1 is $1/\tau_2$ (τ is the characteristic time a system stays in one state before transition to the other state), it can be shown that [26]

$$c(\tau) = (\Delta I)^2 \frac{\tau_1}{\tau_0 + \tau_1}\left\{\frac{\tau_1}{\tau_0 + \tau_1} + \frac{\tau_0}{\tau_0 + \tau_1}\exp\left[-\left(\frac{1}{\tau_0} + \frac{1}{\tau_1}\right)t\right]\right\} \tag{12}$$

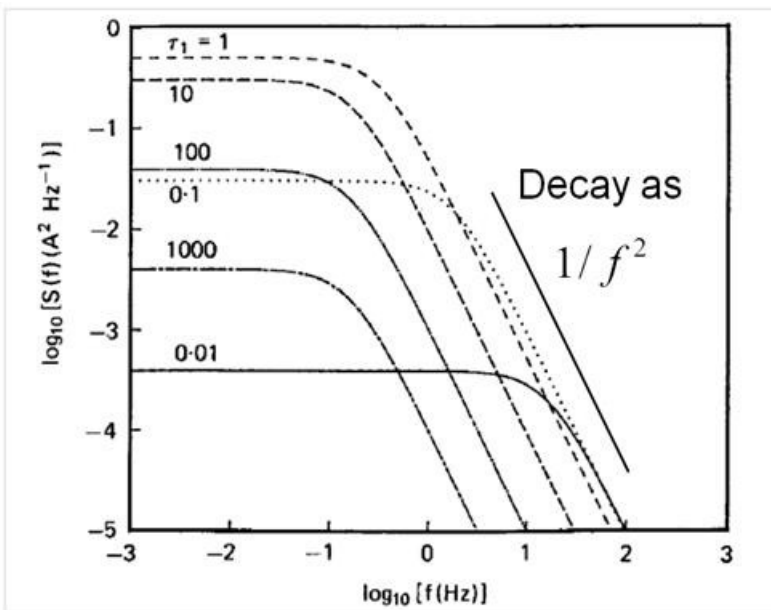


Figure 3.20. Lorentizian power spectra calculated based on Equation 13 [26]

And

$$S(f) = \frac{4(\Delta I)^2}{(\tau_0 + \tau_1)[(1/\tau_0 + 1/\tau_1)^2 + (2\pi f)^2]} \tag{13}$$

Figure 3.20 [26] plots the PSD based on Equation 13 where, ΔI = 1A, τ_0 = 1s and τ_1 varies from 0.01 to 1000s. One of the features of the RTN PSD is that its amplitude is constant at low frequency, yet beyond the corner frequency defined as

$$f_{RTS} = \frac{1}{\tau_0} + \frac{1}{\tau_1} \tag{14}$$

the spectrum decays with frequency squared. For the symmetric case, where $\tau_0 = \tau_1 = \tau$, Equation 13 reduces to

$$S(f) = \frac{2(\Delta I)^2 \tau}{4 + (2\pi f \tau)^2} \tag{15}$$

In the above discussion, only one trap center is assumed. In reality, there is a wide spread of trap states. It can be shown [27] that in a system if the traps are uniformly distributed, the characteristic time τ decreases exponentially with an increase in the trap depth assuming all the traps are independent. Then the PSD of a system can be described by

$$S(f) \propto \int \frac{1}{\tau} \frac{2\tau}{1 + (2\pi f \tau)^2} d\tau \propto \frac{1}{f} \tag{16}$$

In the other words, a superposition of lots of independent $1/f^2$ noise source will give rise to 1/f noise. Therefore, typically 1/f type noise is observed when the device size is large where lots of noise sources are available.

In the rest of this section, the noise measurement is performed by using an HP35670A dynamic signal analyzer along with a BTA9812B low noise transimpedance amplifier. During the measurement, the bottom electrode Au electrode is always grounded.

We first study the relationship between PSD and V_{Read}. For this purpose, a Ti/YSZ device is programmed into different resistance states and the noise is measured at different voltage levels. Fig. 3.21 shows one of the devices based on Ti/YSZ. For this particular sample the device resistance is set to 2.1 $k\Omega$. While Fig. 3.21 (a) shows that the absolute noise level increases as the V_{Read} increases, Fig. 3.21 (b) shows that the normalized PSD (PSD / V^2) is almost independent of the V_{Read}. In reality, only the normalized PSD is important, since it is this term that determines the signal to noise ratio (SNR) and therefore the device error rate. The independency of normalized PSD on voltage is expected since it can be shown that [28] if we assume the number of traps N_{dev} inside a device is large enough to ensure a wide range distribution of time constants then:

$$\frac{S_I(f)}{I^2} = \frac{S_V(f)}{V^2} \propto \frac{1}{N_{dev}} \tag{17}$$

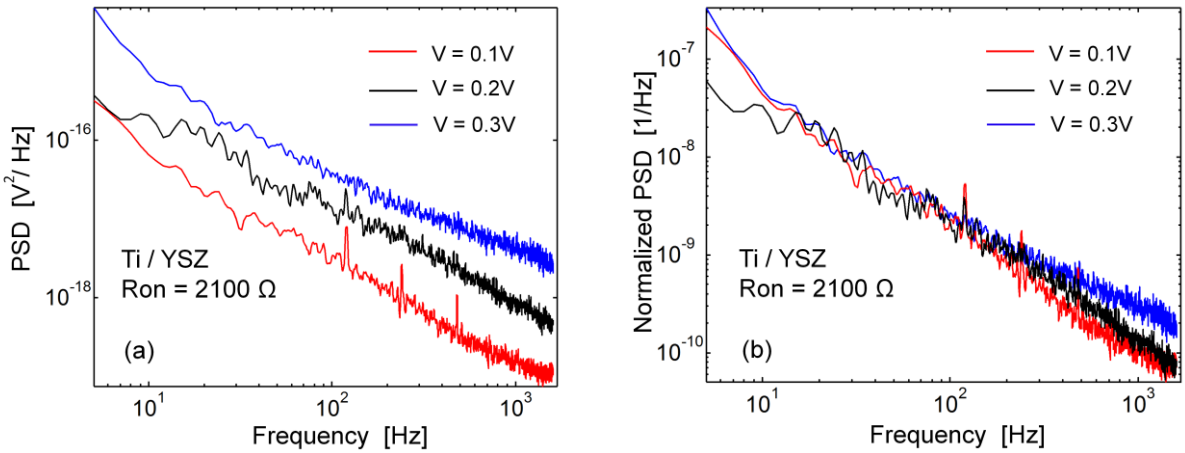


Figure 3.21. (a) The noise power spectrum density and (b) normalized PSD of Ti/YSZ at the resistance state of 2100Ω.

where $S_I(f)$ is the normalized current PSD and $S_V(f)$ is the normalized voltage PSD. From the above equation we can see that the normalized PSD is indeed independent of reading voltage. In addition, the noise spectrum in Fig. 3.21 shows a 1/f type distribution. This is also expected since in LRS the trap density is large and there is a wide distribution of the characteristic time τ for a wide range noise sources. Each of them gives rise to a $1/f^2$ type spectrum and the summation of all those noise sources become 1/f -like flicker noise.

The reliability of Ti/YSZ and Cu/YSZ is further compared. First the devices were driven to difference resistance states. Then the resistance sampling measurement was performed. Fig. 3.22 (a) shows the time sampling result for both Cu/YSZ and Ti/YSZ. RTN is clearly seen in the Cu/YSZ device. In addition, as the resistance of the Cu/YSZ increases, the noise level increases. However, at the same resistance level, Ti/YSZ shows much less variation. Next, a low frequency noise measurement was conducted. Fig. 3.22 (b) shows the normalized PSD at different resistance states for both Cu/YSZ and Ti/YSZ. As expected, at the same resistance state, the Cu/YSZ has much larger noise than the Ti/YSZ device. For both Ti/YSZ and Cu/YSZ, the noise increases as the resistance increases and this is consistent with previous reports [29]. The reason that Ti/YSZ has much lower noise than Cu/YSZ at the same resistance level can be explained by the fact that the conductivity of a Cu filament is much larger than the conductivity of a VO filament; thus, for a given resistance state, the width of a metallic filament is much narrower than the width of a VO based filament. Therefore, the Cu/YSZ is more vulnerable to

the atomic motion of the Cu atoms at the edge of the filament. The noise goes worse when the resistance increases since the filament becomes narrower. This makes the Cu/YSZ unfavorable for using in multilevel cells.

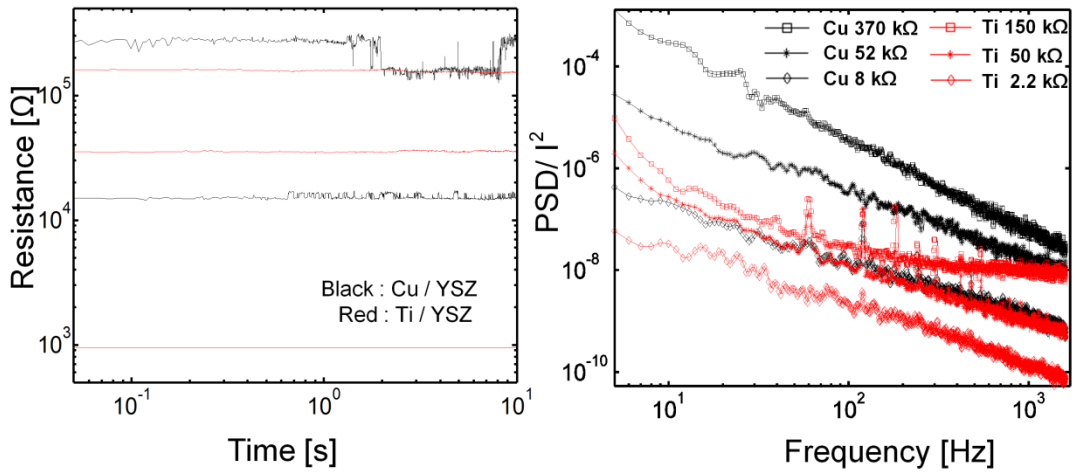


Figure. 3.22. (a) Resistance sampling for both Cu/YSZ and Ti/YSZ devices at different resistance states. (b) Normalized noise power spectrum density for both types of devices at different resistance states.

Next, cycle-dependent noise behavior of Ti/YSZ in the RESET state was further investigated. Typically, as shown in Fig. 3.23 the noise PSD increases as the cycling increases. This can be due to the fact that during every switching cycle, there is additional damage caused or additional traps generated in the active region of the filament. Those traps act as additional noise sources. This can be further demonstrated by the fact that initially the noise is 1/f type. Gradually, at 25 cycles the noise becomes Lorentzian type and this is an indication that a dominant trapping center is eventually formed [26]. However, this trapping center can be annihilated when sufficient inactive time is provided. At 26 cycles, after a 10 minutes break, the spectrum gets back to low noise level and become 1/f type again. Such a self-healing process makes the OV based device attractive for MP applications.

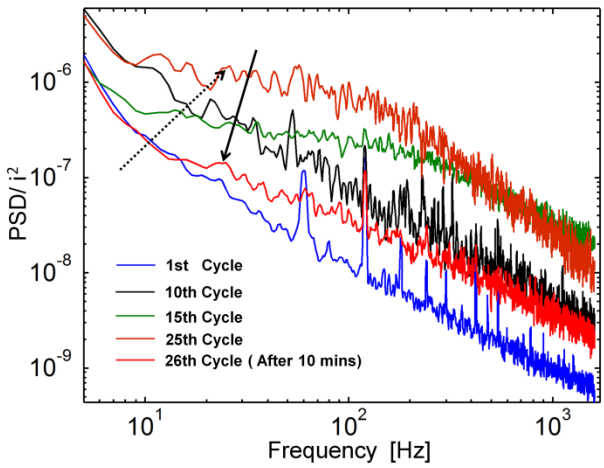


Figure 3.23. The dependence of normalized PSD on the switching cycles in the RESET stage for Ti / YSZ.

3.5: Summary

In summary, in this chapter the resistive switching characteristics of different top electrodes on amorphous YSZ dielectric film are systematically studied. It is shown that only Cu/YSZ and Ti/YSZ have reliable resistance switching characteristics. Further, it is shown that Cu/YSZ shows a metallic filament conduction feature, while Ti/YSZ has an electron hopping based conduction characteristic. In addition, it is found that Ti/YSZ has much better performance than Cu/YSZ in terms of endurance, retention and reliability. For almost every device, Ti/YSZ has endurance of at least 10^4 cycles. In comparison, an endurance of only 100 cycles is achieved for Cu/YSZ devices. It is possible that the endurance of Cu/YSZ relates to the quality of the interface between the inert electrode and dielectric. In addition, Ti/YSZ shows better retention property in HRS, while Cu/YSZ is only stable when the on state resistance is very small. This difference in retention can be due to the physical size difference between metallic and VO based filaments. Finally, noise measurements were conducted that confirm that Ti/YSZ is more reliable than Cu/YSZ.

3.6: References

[1] J.D. Ballard, J. Davenport, C. Lewis, W. Nelson, R.H. Doremus, and L.S. Schadler"Phase Stability of Thermal Barrier Coatings Made From 8 wt.% Yttria Stabilized Zirconia: A Technical Note" *Journal of Thermal Spray Technology*, Volume 12, 1, March 2003.

[2] R. Ruh, K.S. Mazdigasni, P.G. Valentine, and H.O. Bielstein: "Phase Relations in the System ZrO2-Y2O3 at LowY2O3 Contents" *Journal of the American Ceramic Society* 67(9), 1984, pp. 190-192.

[3] http://electronicstructure.wikidot.com/predicting-the-ionic-conductivity-of-ysz-from-ab-initio-calc
[4] R. Pornprasertsuk, P Ramanarayanan, C.B. Musgrave, Fritz B. Prinz "Predicting ionic conductivity of solid oxide fuel cell electrolyte from first principles" *Journal of applied physics* 98, 103513, 2005.
[5] Carlos López-Gándara, Francisco M. Ramos, and Albert Cirera, "YSZ-Based Oxygen Sensors and the Use of Nanomaterials: A Review from Classical Models to Current Trends," *Journal of Sensors*, vol. 2009, Article ID 258489, 15 pages, 2009.
[6] C.S. Kim, et.al,"Influence of the Oxygen Concentration of YSZ Gate Dielectric Layer on the Low Voltage Operating Pentacene Thin Film Transistor" *Electrochemcial and solid-state Letters*, 9 (3) G96-G99, 2006.
[7] S. Horita, M. Watanabe, S. Umemoto and A. Masuda "Material properties of heteroepitaxial yttria-stabilized zirconia films with controlled yttria contents on Si prepared by reactive sputtering" *Mater. Sci. Eng., B, Solid-State Mater. Adv. Technol*. 54 (1998) 79.
[8] S.H.Jeong et. al, "Physical and electrical properties of ZrO_2 and YSZ high-k gate dielectric thin films grown by RF magnetron sputtering" *Thin Solid Films,* 475 (2005) 354– 358
[9] James D. Plummer, Michael D. Deal, Peter B. Griffin "Silicon VLSI technology Fundamentals, practice and modeling" Prentice Hall; 1 edition, July 24, 2000.
[10] Process in SOI structures and devices operating at extreme conditions Nato Science Series, 2002, Vol. 58, Part 1, page 11-15.
[11] O. Jongprateep, V. Petrovsky, F. Dogan "Effects of Yttria concentration and microstructure on electric breakdown of Yttria Stabilized Zirconia" *Journal of metals, materials and minerals*. Vol. 18 No. 1 page 9-14, 2008
[12] C. Lin. C. Wu, C.Wu, T. Lee, F. Yang, C, Hu, T, Tesng " Effect of top Electrode material on resistive switching properties of ZrO2 film memory devices" *IEEE Electron Device Letters* Vol.2, No.5, May 2007
[13] X.P. Wang, Y.Y. Chen, L. Pantisano, L. Goux, M.Jurczak, G. Groeseneken, D.J. Wouters, "Effect of anodic interface layers on the unipolar switching of HfO2-based resistive RAM," *VLSI Technology Systems and Applications (VLSI-TSA),* 2010 International Symposium on , vol., no., pp.140-141, 26-28 April 2010
[14] I. Fisher, M. Eizenberg "Copper ion diffusion in porous and nonporous SiO2-based dielectrics using bias thermal stress and thermal stress tests" *Thin Solid Films* 516 (2008) 4111-4121
[15] S.M. Sze "Current transport and maximum dielectric strength of silicon nitride films" *Journal of Applied Physics*, Vol. 38, Number 7, June 1967
[16] S.M. Sze "Physics of semiconductor devices second edition" Wiley-Interscience, 1981.
[17] K. Kinoshita, K. Tsunoda, Y. Sato, H. Noshiro, S. Yagaki, M. Aoki, and Y. Sugiyama, "Reduction in the reset current in a resistive random access memory consisting of NiO*x* brought about by reducing a parasitic capacitance" *APPLIED PHYSICS LETTERS* 93, 033506 , 2008
[18] Mott. N. F and Gurney. R.W " Electronics Processes in Ionic Crystals" Chap. 5, Clarendon Press, 1948
[19] P. N. Murgatroyd "Theory of space-charge-limited current enhanced by Frenkel effect" *Journal of Physics D: Applied Physics* 3, 151, 1970

[20] L.F. Mao "Modeling of temperature dependence of the leakage current through a hafnium silicate gate dielectric in a MOS device" *Semiconductor science and technology* 22, page 1203-1208, 2007

[21] D. C. Kim, M. J. Lee, S. E. Ahn, S. Seo, J. C. Park, I. K. Yoo, I. G. Baek, H. J. Kim, E. K. Yim, J. E. Lee, S. O. Park, H. S. Kim, U. I. Chung, J. T. Moon, and B. I. Ryu, "Improvement of resistive memory switching in NiO using IrO2", *Applied Physics Letters*, vol. 88, 232106, Jun. 2006.

[22] H.Y. Lee, P.S. Chen, T.Y. Wu, C.C. Wang, P.J. Tzeng, C.H. Lin, F. Chen, M.J. Tsai, and C. Lien, Electrical evidence of unstable anodic interface in Ru/HfOx/TiN unipolar resistive memory, *Applied Physics Letters*. 92 (2008), p. 142911.

[23] T.T. Bakhishev, "Solution-Processed Inorganic Electronics" PhD dissertation, University of California, Berkeley, May, 2011.

[24] T. Sakamoto, N, Banno, N, Iguchi, H, Kawaura, H, Sunamura, S, Fujieda, K, Terabe, T, Hasegawa and M, Aono "A Ta_2O_5 solid-electrolyte switch with improved reliability," *Symposium on VLSI Tech. Digest*, 2007, page. 38–39.

[25] Masayuki Terai, Yukihiro Sakotsubo, Yukihiro Saito, Setsu Kotsuji, and Hiromitsu Hada "Memory-State Dependence of Random Telegraph Noise of Ta2O5/TiO2 Stack ReRAM" *ELECTRON DEVICE LETTERS*, Vol. 31, No. 11, NOVEMBER 2010

[26] M.J. Kirton, M.J. Uren "Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency (1/*f*) noise" *Advances in Physics*, 38:4, 367-468

[27] http://www.eng.auburn.edu/~wilambm/pap/2011/K10147_C011.pdf

[28] McWhorter A. L. 1/*f* noise and germanium surface prosperities. In *Semiconductor Surface Physics*. R. H. Kingdton (Ed.), University of Pennsylvania Press, Philadelphia, PA, 1957, pp. 207–228.

[29] M. Terai , Y. Sakotsubo , Y. Saito , S. Kotsuji and H. Hada "Effect of bottom electrode of ReRAM with Ta2O5/TiO2 stack on RTN and retention", IEDM Tech. Dig., p.775 , 2009.

Chapter 4: Multi-level Programming and Impact of Programming Methodology on Device Reliability

4.1: Introduction

One key property for RRAMs to be feasible for mass production is the possibility of multi-level programming (MLP) [1, 2] since it significantly increases the effective data storage density. So far, several materials such as WO_x [1], HfO_2 [2], Ta_2O_5 [3] and TiO_2 [4] have shown MLP capability. Generally, there are two major ways to achieve MLP. One is to obtain MLP during the SET stage and the other way is to do so in the RESET process.

During the SET process, by controlling the maximum current flowing through the device, different resistance levels can be obtained. In fact, as shown in Fig. 3.9 both Ti/YSZ and Cu/YSZ can achieve MLP through this approach. However, as stated in Chapter 1 and 3, one problem with this approach is the capacitive charging effect. The discharging current can cause large variations in R_{on}. Even though a 1T1R structure can effectively eliminate this effect, it significantly reduces the RRAM array density. On the other hand, for diode based selector structures, the capacitive discharging effect still exists. In addition, as demonstrated in Chapter 1, the history dependent R_{on} instability problem may also hinder this approach.

The second way to achieve MLP is through the RESET process. In this case, it would be ideal that the I-V characteristic in the RESET stage is smooth enough that R_{off} can gradually increase with the applied voltage. Through this approach, there is no capacitive charging effect involved and it is also a voltage-controlled operation and easier for circuit implementation.

The physics of MLP is also not very clear yet [5]. Researchers [5], from a pure electronic point of view, suggest that only a dielectric with a medium bandgap where oxygen vacancies acts as electron donors, has an MLP capability. Yet this theory may not be able to explain all the experimentally observed phenomena. For some systems such as Cu_xO [6] and NiO [7], which meet all the above mentioned criteria, still show very sharp transitions in the RESET stage.

In this chapter, a resistive switch based on Ti/YSZ with MLP capability is demonstrated and a brief physical explanation is provided. In addition, it is also demonstrated that the use of incremental step pulse programming (ISPP) can realize RRAM cells with good reliability and achieve optimized tradeoffs among resistance window, endurance and programming speed.

4.2: Multi-level Programming of Ti/YSZ in RESET Stage

For both Cu/YSZ and Ti/YSZ, as shown in Fig. 4.1, RRAM cells have sharp transitions in the SET stage. In the RESET process, Cu/YSZ still shows an abrupt transition from LRS to HRS, which is quite similar to the situation in SET stage. In addition, its conductance typically drops to a minimum which is the only stable state for Cu/YSZ in HRS. This feature makes MLP impossible for Cu/YSZ during the RESET process. In contrast to Cu/YSZ, Ti/YSZ shows a very smooth transition during the RESET stage. As indicated by the arrow in Fig. 4.1 (b), the I-V curve is very flat during this process, and the device conductance decreases gradually. This is further demonstrated in Fig. 4.2, where ISPP technique is adopted and the duration of the programming pulse is set to 1µs. It can be seen that after a V_{RESET} threshold, the R_{off} of Ti/YSZ exponentially increases with applied voltage, whereas a steep transition occurs for Cu/YSZ and after that its resistance remains constant. In addition, this figure also shows the good achievable resistance window for Ti/YSZ. For a fresh device, this on/off window can be more than 10^4 which is good for MLP.

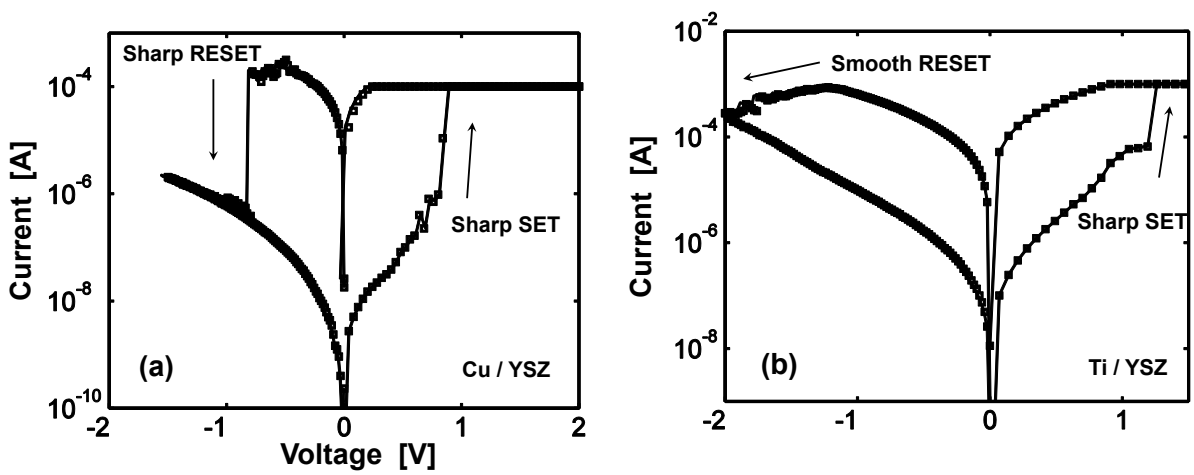


Figure 4.1. Typical I-V characteristics for (a) Cu/YSZ and (b) Ti/YSZ.

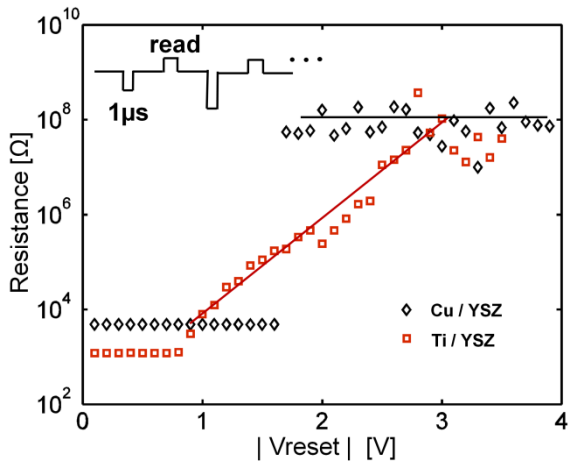


Figure 4.2. Off state resistance versus the different V_{RESET} pulses with duration 1µs. Inset shows the increment pulse programming technique

Even though RRAMs typically are classified into metal cation and VO based devices, as stated in Chapter 1 there are no essential differences between them. They are two different ways to describe the same physical and chemical processes inside the metal oxide cells. In the SET stage, this is a locally increasing metal element concentration and at the same time, a metal cation reduction process such that a metal rich region (or suboxide) can be formed. In the RESET stage, this is a locally decreasing metal element concentration and metallic species or cations oxidation process such that more insulating oxide can be formed locally. The metal rich region formed in the SET stage can have metallic or semiconducting features depending on the detailed chemical composition. The conductivity of the features varies. Typically, it is easier to form a more conductive metal rich region if the metal cations are easier to reduce and precipitate as in CuO and NiO, which are more likely to form a more metallic-like phase filament.

For a smooth RESET, it is better to have a semiconducting phase (or electron hopping based system) instead of a metallic-like phase, since by doing so, at given resistance level, the filament would have much larger cross section area and also, at the same RESET current level, the heat density is smaller inside such filament. Both of the two factors help realize a gradual change of the conductance. Another important consideration is that the filament dissolution process for more metallic-like species is a self acceleration process [8] and is more sensitive to joule heating due to its positive temperature resistance coefficient. Therefore, weak spots within the filament are easily created. On the other hand, the corresponding process is a negative feedback for a more semiconducting-like or hopping based system and thus it is difficult to form such weak spots within the filament. This feature also helps to realize a more gradual change in RESET.

One possible way to describe the ability for a metal cation to trap electrons is standard electrode potential (SEP). The SEP for Y is -2.37V, -1.53V for Zr and -0.25V for Ni and finally it is

0.34V for Cu [9, 10]. A more positive number indicates the material more easily captures an electron to form the reduced metallic state. As can be seen, for both Y and Zr, it is more difficult to capture electrons while materials like Ni and Cu exhibit easier electron capture. Therefore, most devices based on Cu_xO and NiO show a metallic conduction feature and have abrupt transitions in RESET stage.

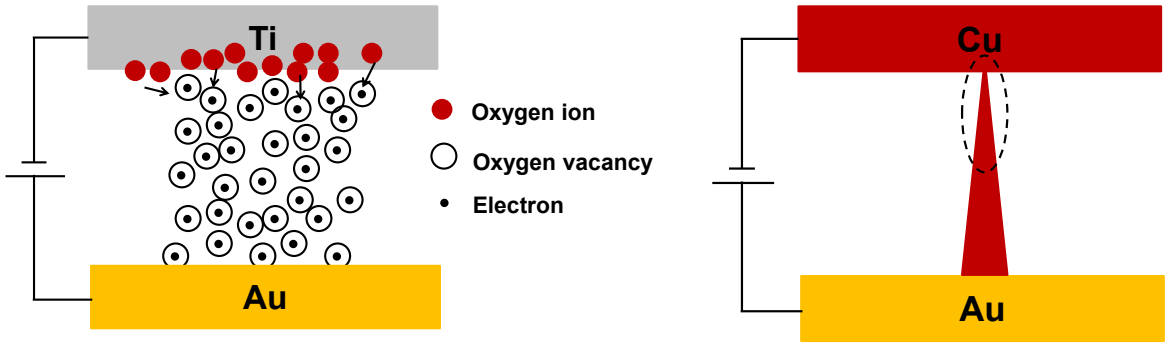


Figure 4.3. The filament width comparison between Ti/YSZ and Cu/YSZ in RESET stage.

The above discussions can explain the different V_{RESET} pulse response between Ti/YSZ and Cu/YSZ shown in Fig. 4.2. Generally, the conductivity of a filament in Cu/YSZ is much larger than that of that of Ti/YSZ (This is supported by the experimental results: even when R_{on} is less than 1kΩ, devices like Ti/YSZ still have an activation energy barrier around 0.02ev. No such barrier is found in Cu/YSZ) as illustrated in Fig. 4.3. Therefore, for a given R_{on}, the effective metallic-like filament area is much narrower than that of Ti/YSZ devices. In addition, since a filament in Cu/YSZ is more sensitive to joule heating, a weak spot can be quickly created when current is flowing through, thus eventually a very small change on the filament tip can cut off the entire path leading to an abrupt transition.

Fig. 4.4 demonstrates the excellent MLP capability of Ti/YSZ devices. During the measurement, the V_{SET}/V_{RESET} increases/ decreases in steps with interspersed verification reads. If the targeted resistance is achieved and verified to be stable, the programming voltage is recorded. Otherwise, the voltage pulse will keep increasing / decreasing. A device failure occurs when the required resistance level is not achieved by a predefined voltage limit. As can be seen, the device has a low resistance state of ~8 $k\Omega$ and three distinct high resistance states up to ~2MΩ.

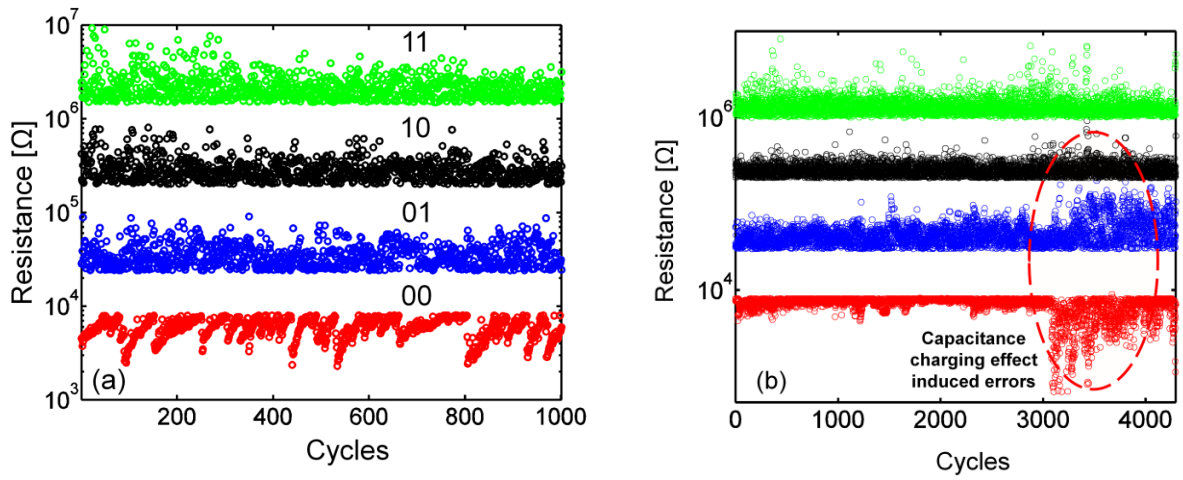


Figure 4.4. Four levels MLP of Ti/YSZ. (a) Device one. (b) Device two

In the MLP process, the parasitic capacitance should be minimized, since it can cause programming errors as shown in Fig. 4.4 (b). For this particular device, it can be successfully programmed for the first three thousand cycles. After that, a capacitance charging effect occurs and the R_{on} (indicated by the red dots) significantly decreases below $1\,k\Omega$, as a result, the first level of R_{off} (shown in blue dots) increases and some of them mix with the second level R_{off} (shown in black dots). The underlying physics is that when R_{on} significantly reduces there is more joule heating generated at the same RESET voltage level. The excessive heat can weaken the Ti-O bonding at the Ti and YSZ interface so that excessive O^{2-} can be generated in a very short time under the electric field. Those excessive O^{2-} species combine with a VO and lead to a significant increase of resistance. In addition, this excessive heat can also help YSZ dielectric to overcome the phase transition barrier from LRS to HRS. Thus, the overall effect is that instead of gradual increase of R_{off}, a sharp increase occurs and causes the programming errors. Therefore, based on above discussion, the capacitance charging effect should be minimized for MLP purposes.

4.3: Impact of Programming Methodology on Device Reliability

One of the biggest challenges in RRAMs is device reliability issue [11]. The device switching failure typically occurs when it gets stuck either at a low resistance state (LRS) or at a high resistance state (HRS). The traditional single pulse programming techniques let the ratio between off-state resistance (R_{off}) and on-state resistance (R_{on}) arbitrarily scatter every cycle. If

the magnitude of the SET or RESET voltage (V_{SET}, V_{RESET}) is small, the SET or RESET process will become less effective and on the other hand if the pulse magnitude is too large, the device will breakdown quickly and lifetime seriously reduces. In this section, it is demonstrated that by adopting an incremental step pulse programming (ISPP) scheme similar to that which is commonly used in Flash memory, an excellent balance between device endurance, resistance on-off ratio can be achieved and the device lifetime substantially increases.

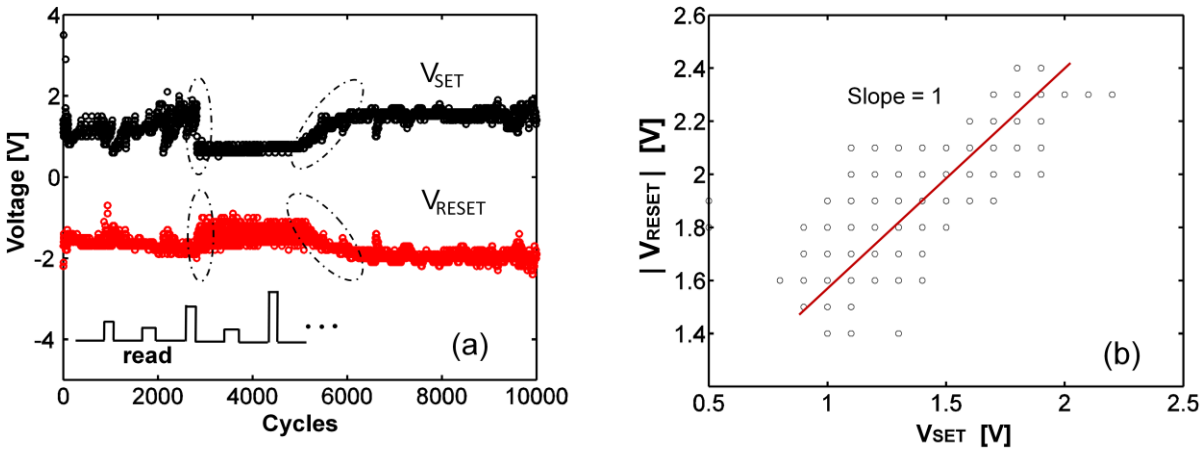


Figure 4.5. (a) The Evolution of V_{SET} and V_{RESET}. The dash line shows both a sudden and a gradual change in V_{SET} and V_{RESET} (b) The relationship between V_{SET} and V_{RESET} for an endurance test of 10^4 cycles.

For an ideal RRAM device, V_{SET} and V_{RESET} should be invariant with cycling. During the SET process, the spatial location of all newly created VOs should be the same as VOs created in the previous SET cycle. On the other hand, in the RESET process, all the created VOs during the SET process should be recombined with the oxygen ions released from either the cathode or from the dielectric itself. In reality, however, due to the stochastic nature of RRAMs, unexpectedly large variations occasionally occur. For example, in the HRS, additional oxygen vacancies may be thermally generated in the gap between the filament and electrode; additionally, during RESET, some VOs at the tip of filament may escape annihilation. These residual VOs will lower the local dielectric strength of the oxide and lead to a sudden decrease in V_{SET}. This is illustrated in Fig. 4.5 (a).

During the measurement, an ISPP scheme is adopted and the targeted R_{on} and R_{off} are 8kΩ and 80kΩ respectively. As can be seen, V_{SET} shows a sudden drop of ~1V around 3000 cycles. Interestingly, V_{RESET} at the same time is also reduced accordingly. This is due to a capacitive charging effect [12], coming from both the device itself and external measurement circuitry. When a smaller V_{SET} is used to turn on a device, there is less capacitive discharging current flowing through the device and thus less filament overgrowth; this in turn makes the RESET

operation easier and hence a smaller V_{RESET} is needed. This nearly linear correlation between V_{SET} and V_{RESET} generally holds true for all conditions as shown in Fig. 4.5 (b), which shows the correlation between V_{SET} and V_{RESET} across 10^4 cycles. The observed slope is ~1.

Fig. 4.5 (a) also shows a regime around 5000 cycles where V_{SET} and V_{RESET} gradually increase again. Under those conditions, if the applied voltage is not adjusted accordingly, it may either exert too much electrical stress, resulting in device degradation, or the voltage may be too small and makes the switching less effective. This is a key disadvantage of the traditional single pulse technique. In fact, due to the unavoidable device-to-device variation, it is almost impossible to find an optimum pulse magnitude for both V_{SET} and V_{RESET} for all devices in a large memory array. Further, even for a single device, as mentioned above, the optimum pulse magnitude depends on the operational history and is a function of time. Thus, the single pulse technique has serious limitations.

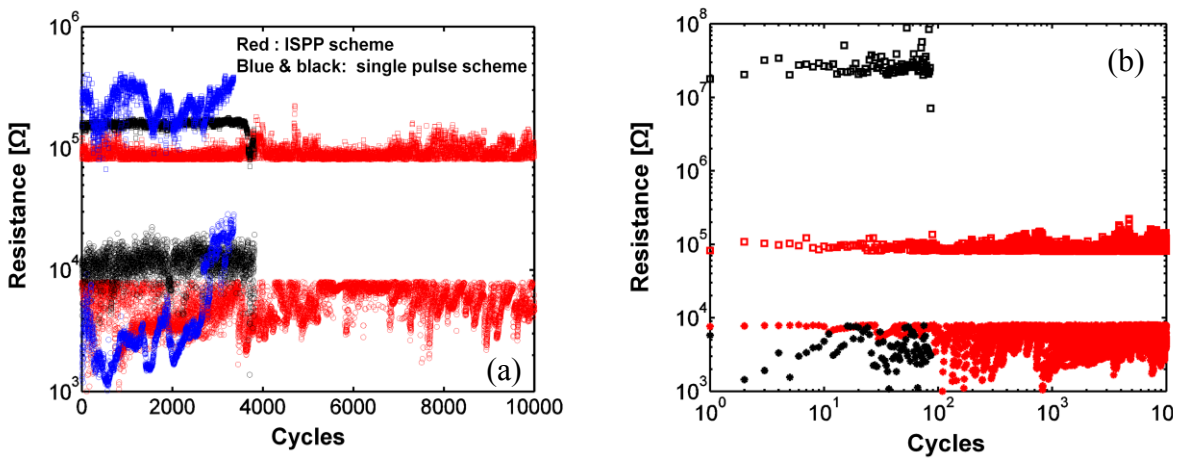


Figure 4.6. (a) Comparison between the single pulse scheme and ISPP. (b) Illustration of tradeoff between the resistance programming window and device endurance. ISPP allows for optimal balancing between these properties.

By using ISPP, however, those problems can be solved easily. ISPP can tailor an optimum V_{SET} and V_{RESET} not only for every device but also for every operational cycle. This is clearly demonstrated in Fig. 4.6 (a), which compares device cycling under ISSP and a standard single pulse scheme. Both of the aforementioned failures are observed for the single pulse scheme. The black curve corresponds to a device that becomes stuck in LRS, likely due to extra electrical stress asserted during SET. On the other hand, the blue curve corresponds to a device that becomes stuck at the HRS due to an inefficient V_{RESET}. By utilizing ISPP, the device endurance can reach beyond 10^4 cycles (shown in red) and neither type of failure is observed during the entire process. And this improvement is consistent with previous research result [13].

ISPP also allows for optimal balance of the resistance window with device endurance. Generally, a larger resistance window between R_{on} and R_{off} degrades the device endurance. Fig. 4.6 (b) shows this effect. As can be seen, when the resistance window is beyond 3000, some devices only endure 100 cycles. However, almost all the devices tested have endurance beyond 10^4 (limited by measurement time) when a window of 10 is adopted. In addition, in the devices tested, we mainly see the degradation of HRS (R_{off} drops with cycling). R_{on}, on the other hand, is set to $10\,k\Omega$ which is quite large and easily to be achieved. The reason of degradation of R_{off} can be described as following: In order to get a large R_{off} for a large resistance window, a large V_{RESET} has to be applied in order to extract more O^{2-} from Ti electrode and deliver them to the VO sites. This process does not degrade the device only when there is enough O^{2-} available locally such that a larger V_{RESET} can make trap density decrease and increase the dielectric strength. However due to the random nature of RRAM, occasionally local O^{2-} deficiency may occur. At this point, the amount of O^{2-} delivered to the VOs saturates and a larger V_{RESET} doesn't increase the O^{2-} flow anymore. Even worse, under this situation a large V_{RESET} actually makes the dielectric bonding weak and generates more VO (a process like in SET stage). Therefore, the HRS gradually degrades.

One problem with ISPP, however, is that it is generally more time consuming to program a memory cell. However, this can be partially alleviated by increasing the voltage difference between two adjacent pulses. For example, instead of increasing the pulse magnitude by 0.1V every step, 0.3V or 0.5V can be adopted. Hence, a good tradeoff between the endurance and programming speed can be obtained.

4.4: Summary

In this chapter, a reliable multi-level programming cell based on Ti/YSZ is demonstrated. It is shown that this cell has large resistance programming window with good endurance. And by using ISPP methods, four programming levels are successfully implemented. It is furthermore found that compared with Cu/YSZ, the MLP capability of Ti/YSZ is due to the fact that both yttrium and zirconium are difficult to precipitate and the filament in Ti/YSZ shows a hopping based conduction mechanism so that a weak spot is not easily formed in such filament.

In addition, it is found that in order to reduce MLP error rate, the capacitance charging effect has to be minimized. When the capacitance charging effect is present, excessive heat will be generated in the RESET stage, which in turn makes excessive O^{2-} form at the Ti electrodes. As a result, a large amount of VOs get recombined with O^{2-} in a very short time which leads to a dramatic increase of R_{off} and causes the error.

Further, a single pulse program scheme was compared with ISPP. It is found that ISPP can tailor an optimum V_{SET} and V_{RESET} for every device and every operation cycle which significantly

improves the device endurance and reliability. It is also found that failure of Ti/YSZ devices typically occurs at HRS, where there is not enough O^{2-} from the Ti electrode to recombine with VOs.

Finally, by using ISPP, tradeoffs between resistance window and endurance; programming speed and device lifetime can be obtained. A larger resistance window typically degrades the endurance. A finer step pulse in ISPP improves device endurance but increases programming time.

4.5: References

[1] W.C. Chien, M.H Lee, F.M. Lee, Y.Y Lin, H.L Lung, K.Y. Hsieh, and C.Y. Lu "Multi-level 40nm WO_x resistive memory with excellent reliability" *Electron Devices Meeting, 2011 IEEE International* page 31.5.1-31.5.4

[2] S.S Sheu, M.F. Chang, K.F. Lin, C.W. Wu, Y.S. Chen, P.F. Chiu, C.C. Kuo, Y.S. Yang, P.C. Chiang, W.P. Lin, C.H. Lin, H.Y. Lee, P.Y. Gu, S.M. Wang, Chen, F.T., K.L. Su, C.H. Lien, K.H. Cheng, H.T. Wu, T.K. Ku, M.J. Kao, M.J. Tsai , "A 4Mb embedded SLC resistive-RAM macro with 7.2ns read-write random-access time and 160ns MLC-access capability," *Solid-State Circuits Conference Digest of Technical Papers* (ISSCC), 2011 IEEE International , vol., no., pp.200-202, 20-24 Feb. 2011

[3] M. Terai, Y. Sakotsubo, S. Kotsuji, and H. Hada "Resistance Controllability of Ta2O5/TiO2 Stack ReRAM for Low-Voltage and Multilevel Operation" *IEEE ELECTRON DEVICE LETTERS*, VOL. 31, NO. 3, MARCH 2010

[4] Chikako Yoshida, Kohji Tsunoda, Hideyuki Noshiro, and Yoshihiro Sugiyama"High speed resistive switching in Pt/TiO2 /TiN film for nonvolatile memory application" *Applied Physics Letters*, **91**, 223510, 2007

[5] Hur J.H, Kim K.M, Chang M, Lee S.R, Lee D, Lee C.B, Lee M.J, Kim Y.B, Kim C.J, Chung U.I, "Modeling for multilevel switching in oxide-based bipolar resistive memory" *Nanotechnology* 23 (2012) 225702

[6] H. B. Lv, M. Yin, X. F. Fu, Y. L. Song, L. Tang, P. Zhou, C. H. Zhao, T. A. Tang, B. A. Chen, and Y. Y. Lin, "Resistive memory switching of Cu_xO films for nonvolatile memory application," *IEEE Electron Device Letters*., vol. 29, no. 4, pp. 309–311, Apr. 2008.

[7] J.-B. Yun, S. Kim, S. Seo, M.-J. Lee, D.-C. Kim, S.-E. Ahn, Y. Park, J. Kim, H. Shin, "Random and localized resistive switching observation in Pt/NiO/Pt" *Phys. Status Solidi* (RRL) 2007, 1, 280

[8] R. Waser, R. Dittmann, G. Staikov and K. Szot , " Redox-Based Resistive Switching Memories – Nanoionic Mechanisms, Prospects and Challenges", *Advanced Materials*. 2009, 21, page 2632.

[9] Peter Atkins, "Physical Chemistry", 6th edition, W.H. Freeman and Company, New York. 1997.

[10] Vanýsek, Petr "Electrochemical Series", in *Handbook of Chemistry and Physics: 88th Edition*, Chemical Rubber Company, 2007
[11] B. Chen, Y. Lu, B. Gao, Y.H. Fu, F.F. Zhang, P. Huang, Y.S. Chen, L.F. Liu, X.Y. Liu, J.F. Kang, Y.Y. Wang, Z. Fang, H.Y. Yu, X. Li, X.P. Wang, N. Singh, G. Q. Lo, D. L. Kwong "Physical mechanisms of endurance of degradation in TMO-RRAM" *IEEE International Electron Devices Meetings*, 2011 Washington DC, USA.
[12] A. Chen "Switching control of resistive switching devices" *Applied Physics Letters* 97, 263505 (2010)
[13] M. Yin, P. Zhou, H. B. Lv, J. Xu, Y. L. Song, X. F. Fu, T. A. Tang, B. A. Chen, and Y. Y. Lin, "Improvement of resistive switching in CuxO using new RESET mode", *IEEE Electron Device Letters*, vol. 29, no. 7, July, 2008.

Chapter 5: Conclusions and Future Work

5.1: Conclusion

5.1.1 KMC Simulation for ECM based RRAM

This work first conducted comprehensive KMC simulation studies on ECM based RRAMs. Inside the simulator, physical and chemical processes such as oxidation, reduction, and surface/bulk diffusion are taken into account. The transition rate is simulated based on transition state theory. Major simulation conclusions are listed below [1-4].

- Filament topography study: Simulation results show that at the moment when the filament shorts the two electrodes, the width of the filament is inversely proportional to the applied voltage – the width of the filament formed at high voltage bias is much narrower than the width of the filament formed at low voltage bias. This counterintuitive result is due to the fact that during the metallization process the adsorbed metal ions try to diffuse to more stable step and hole sites through surface diffusion, which makes the filament surface smoother and the width wider. However, soon after that, the filament overgrowth effect occurs and within a very short time span the filament grows much wider. In the final equilibrium state, devices with a larger applied voltage/current have a wider filament formed.

- Through the simulation of the voltage and time relationships, it is found that due to the combined effect of the various physical and chemical processes, there is always a threshold voltage present, below which the device can't be turned on. In addition, when the voltage is large enough, an inversely exponential dependence of the switch time on voltage is observed, which makes fast switching of RRAM possible. In addition, it is found that the required Forming/SET voltage is voltage sweeping rate dependent. When the voltage sweep rate is large, a larger Forming/SET voltage requirement is observed. All the above simulated results are confirmed by experiments.

- The simulator simulated the I-V characteristics and demonstrated the local conduction mechanism for ECM based RRAM devices. In addition, the distribution of the on state resistance, device switching time and the relationship between SET and RESET voltage is obtained. Major conclusions are: first, a larger applied voltage/current results in a more conductive filament with a tighter statistical distribution. Second, the V_{RESET} is proportionally correlated to V_{SET}. A larger V_{SET} typically requires a larger V_{RESET} to turn off the device. This is due to the fact that if a stronger filament is formed in the SET stage, it requires a larger RESET voltage to break it since more atoms have to be oxidized and migrate away. Third, it is found that depending on the lateral electric field, the filament

can break at various positions. When the lateral field is small, filaments may tend to break in the middle of the resistive layer or near the inert electrode surface. On the other hand, when there is a large transverse electric field, the filament typically breaks near the active electrode surface.

- The simulator also simulated the impact of the device size and material properties on the device performance. It is found that the forming/SET voltage is linearly dependent on device thickness and when the device area increases, the forming/SET voltage gradually decreases. Also, it is found that even though, by adopting a more diffusive resistive layer it is possible to increase the switching speed of the device, eventually the switching speed is limited by oxidization and crystallization of metal cations. Finally, it is shown that depending on the value of difference activation energies, filaments with various topographies can be obtained.

5.1.2 YSZ based RRAM

In the second part of the dissertation, an experimental study on RRAM based on amorphous YSZ was carried out. The performance of devices with different top electrodes was investigated. Also noise measurements of the Ti/YSZ and Cu/YSZ systems was carried out. The important observations are listed below.

- It is found that Cu/YSZ, Al/YSZ and Ag/YSZ show ECM based RRAM characteristics. On the other hand, Ti/YSZ shows VO based RRAM's features. In addition, Ni/YSZ and Au/YSZ don't have switching capability at all. It is also found that due to the large diffusivity of Ag, Ag/YSZ has the smallest Forming/SET voltages. Next, the I-V characteristics of Ti/YSZ shows that the off state is dominated by SCLC and electron hopping conduction mechanisms while the on state has ohmic as well as the electron hopping conduction features. Through temperature dependent I-V measurements, it is found that for Ti/YSZ devices, the electron activation energy barrier is positively correlated to the resistance state of the cell.

- It is further found that Ti/YSZ has much better endurance and reliability than Cu/YSZ. In addition, Cu/YSZ has a much larger SET/RESET voltage variation. Both of those can be attributed to the poor quality and uncontrolled microstructure at the interface between YSZ and the Au inert electrode. This the impacts the device because this interface critically affects the metallization process for Cu/YSZ device. However, it doesn't affect the device performance so much for VO based cells.

- In addition, it is found that R_{on} of both Ti/YSZ and Cu/YSZ can be varied by applying different CC in the SET operation. It is also found that R_{off} of Ti/YSZ decreases gradually in the RESET operation whereas R_{off} of Cu/YSZ drops sharply to a constant. Further,

experimental results show that Ti/YSZ is more stable in the off-state. This is perhaps due to the bulk conduction nature of Ti/YSZ in the off state, whereas a localized conduction is observed at the on-state. In addition, even though R_{off} of Cu/YSZ is very stable, the stability of R_{on} depends on the resistance level. For R_{on} less than 1kΩ, Cu/YSZ usually has very stable retention. However, retention becomes very much shorter when R_{on} becomes around 20kΩ. This is because when the cell resistance is small the width of the Cu filament is much wider, thus the conductive bridge is more robust. On the other hand, when R_{on} decreases, the filament becomes narrower, and thus a small amount of lateral diffusion of copper atoms will significantly vary the cell resistance and cause the observed degradation in retention.

- Finally, through noise measurements, it is found that at the same resistance level, Ti/YSZ has much lower noise than Cu/YSZ cells. This can be explained by the fact that the conductivity of a Cu filament is much larger than the conductivity of a VO filament; thus, for a given resistance state, the width of a metallic filament is much narrower than the width of a VO based filament. Therefore, the Cu/YSZ is more vulnerable to the atomic motion of the Cu atoms at the edge of the filament. The noise goes worse when the resistance increases since the filament become narrower.

5.1.3 Multi-level Programming of Ti/YSZ based RRAM

Unlike Cu/YSZ, Ti/YSZ shows a multi-level programming capability in the RESET stage. The main conclusions are listed below.

- Even though RRAMs are classified into metal cation and VO based devices, there are no fundamental differences between them in terms of the impact of the various atomistic processes. In the SET stage, there is a locally increasing metal element concentration and also a metal cation reduction process such that a metal rich region (or suboxide) can be formed. In the RESET stage, there is a locally decreasing metal element concentration and metallic element oxidation process such that more insulating oxide can be formed. The metal rich region formed in the SET stage can have metallic or semiconducting features depending on the detailed chemical composition. The conductivity of the features varies.

- Generally, it is easier to form a more conductive metal rich region if the metal cations are easier to reduce and precipitate. For a smooth RESET, a semiconducting phase (or electron hopping based system) is desired such that at given resistance level, the filament would have much larger cross section area and also at the same RESET current level the heat density is smaller. Both of the two factors help realize a gradual change of the cell conductance. In addition the filament dissolution process for more metallic-like species is a self acceleration process and is more sensitive to joule heating due to its positive temperature resistance coefficient. On the other hand, the corresponding

process is a negative feedback for a more semiconducting-like or hopping based system and thus, it is difficult to form such weak spots within the filament.

- A multi-level programming cell is realized in Ti/YSZ devices. In those devices a low resistance state of ~8kΩ and three distinct high resistance states up to ~2MΩ can be achieved. The device has endurance exceed 10^4 cycles (limited by measurement time). In addition, it is found that in the RESET process, the parasitic capacitance should be also minimized to avoid programming errors. The reason is that when R_{on} significantly reduces due to the capacitance charging effect in the SET stage, there is more current and thus more joule heating generated at RESET. The excessive heat can weaken the Ti-O bonding at the Ti and YSZ interface so that excessive O^{2-} can be generated in a very short time under the applied electric field. Those excess O^{2-} species combine with VO sites and lead to a significant increase in resistance which causes programming errors.

- Finally, Ti/YSZ shows a reduced power requirement, in that the RESET current can be scaled down to 100µA, which is desirable for low power applications.

5.1.4 Impact of Programming Methodology on Device Reliability

In this work, the single pulse programming scheme has been systematically compared with an ISPP scheme. The experimental results show that ISPP is superior to single pulse programming scheme in several aspects [5].

- The experimental results show that ISPP can tailor an optimum V_{SET} and V_{RESET} for every device and every operation cycle, which significantly improve the device endurance and reliability.

- Tradeoff can be made between the device programming window and device endurance. A larger resistance window degrades the device endurance.

- Tradeoff can be made between programming speed and device endurance. In order to have a fast device programming speed, the step height between adjacent voltage pulses in ISPP should be increased, though this can potentially degrade the device endurance.

- It is also found that failure of Ti/YSZ devices typically occurs at HRS, where there is not enough O^{2-} from the Ti electrode to recombine with VOs.

5.2: Future Work

First, in the present KMC simulator, the temperature effect is not included, which has very important impacts on the filament overgrowth and RESET operation. For future study, heat conduction equations should be coupled inside the simulation.

Second, there is also a simulation time limitation in the present simulator. The existing simulator is very time inefficient in simulating certain systems. For example, if the anode atom is very easy to oxidize but difficult to precipitate and if the dielectric diffusion barrier is very large, the atom or ions will just move around at each side of electrode surface without crystallization. Hence, there is no effective filament formation. Under such situation, the simulator wastes tremendous time on those trivial motions of particles. A new algorithm should be developed to make the simulation more time efficient such that it can ignore the trivial motion of the particle and focus more intelligently on the essential processes.

For YSZ based RRAM cells, further studies are also needed. Due to experimental fabrication limitations, in this work E-Beam evaporation was adopted to deposit the YSZ layer. However, this technique offers relatively poor control of the chemical composition of the deposited film. In the future, other more controlled deposition techniques such as sputtering or Atomic layer deposition (ALD) should be used.

In the current work, the performance of YSZ based RRAM is compared with devices made of ZrO_2. The result shows the yield of Ti/YSZ is 20% higher than the yield of device Ti/ZrO_2. There is no definite conclusion on other characteristics such as retention and endurance. To make a detailed and accurate comparison, future experiments with better film deposition technique and process control are desired and also massive measurement has to be done to get the statistical results. In addition, it would be also very interesting to vary the yttrium concentration inside zirconia to see the change in the devices' performance.

5.3: References

[1] F. Pan and V. Subramanian, "A kinetic Monte Carlo study on the dynamic switching properties of electrochemical metallization RRAMs during the SET process," in Proc. Int. Conf. Simulation of Semiconductor Processes and Devices (SISPAD), 2010, pp. 19–22.
[2] F. Pan, V. Subramanian "Kinetic Monte Carlo simulation of resistive switching and filament growth in electrochemical RRAMS" Device Research Conference, 2010, page 255.

[3] F. Pan, S. Yin, V. Subramanian “A Comprehensive Simulation Study on Metal Conducting Filament Formation in Resistive Switching Memories” Memory Workshop (IMW), 2011 3rd IEEE International
[4] F. Pan; S. Yin; V. Subramanian, "A Detailed Study of the Forming Stage of an Electrochemical Resistive Switching Memory by KMC Simulation," Electron Device Letters, IEEE, vol.32, no.7, pp.949-951, July 2011
[5] Feng Pan, Jaewon Jang and Vivek Subramanian “A very reliable multilevel YSZ resistive switching memory” Device Research Conference, 2012, page 217.