# Defect Induced Aging and Breakdown in High-k Dielectrics

by

Runchen Fang

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2018 by the
Graduate Supervisory Committee:

Hugh Barnaby, Co-Chair
Michael Kozicki, Co-Chair
Dragica Vasileska
Trevor Thornton

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

High-k dielectrics have been employed in the metal-oxide semiconductor field effect transistors (MOSFETs) since 45 nm technology node. In this MOSFET industry, Moore's law projects the feature size of MOSFET scales half within every 18 months. Such scaling down theory has not only led to the physical limit of manufacturing but also raised the reliability issues in MOSFETs. After the incorporation of $HfO_2$ based high-k dielectrics, the stacked oxides based gate insulator is facing rather challenging reliability issues due to the vulnerable $HfO_2$ layer, ultra-thin interfacial $SiO_2$ layer, and even messy interface between $SiO_2$ and $HfO_2$. Bias temperature instabilities (BTI), hot channel electrons injections (HCI), stress-induced leakage current (SILC), and time dependent dielectric breakdown (TDDB) are the four most prominent reliability challenges impacting the lifetime of the chips under use.

In order to fully understand the origins that could potentially challenge the reliability of the MOSFETs the defects induced aging and breakdown of the high-k dielectrics have been profoundly investigated here. BTI aging has been investigated to be related to charging effects from the bulk oxide traps and generations of Si-H bonds related interface traps. CVS and RVS induced dielectric breakdown studies have been performed and investigated. The breakdown process is regarded to be related to oxygen vacancies generations triggered by hot hole injections from anode. Post breakdown conduction study in the RRAM devices have shown irreversible characteristics of the dielectrics, although the resistance could be switched into high resistance state.

The dissertation is dedicated to my dearest parents and my past youth.

ACKNOWLEDGMENTS

I would like to sincerely thank my supervisors Dr. Hugh Barnaby and Dr. Michael Kozicki for their invaluable supports and guides throughout my research works. I would also like to express my special gratitude to Dr. Hugh Barnaby for his patient assists and instructions during my toughest time of life. Meanwhile, I'm so grateful to my committee members Dr. Dragica Vasileska for her helpful advices on dielectric breakdown physics and modeling approaches and Dr. Trevor Thornton for his careful suggestions that improved my work.

Most importantly, I would like to thank my parents Huaqian Fang and Jianqiu Wu who provide me boundless love and support. Without their great inspirations I would not get strong motivations to finish the degree. I would also like to thank all my relatives who showed infinite mental support for me to be here today.

Time flies, thinking about the first day I came to US five years ago looks like just happened yesterday. Thus I would like to thank all my previous supervisors Dr. Qingqing Sun and Dr. David Wei Zhang at Fudan University and Dr. Qingquan Liu at Nanjing University of Information Science and Technology who brought me into this area and trained me with all the necessary skills.

During my five-year-study at ASU, I've learnt a lot of new knowledge and gained a lot of experiences from the great people here. So I would also like to thank Dr. Tingyong Chen, Dr. Yu Cao, Dr. Michael Goryll, Dr. Shimeng Yu, Dr. Nathan Newmann, Dr. Yonghang Zhang and Dr. Keith Holbert for help in the research, share of experiences and advices for experiments. I would like to thank the CSSER and their staff for the assistance and advice on the device fabrication process. I would like to thank all the lab mate Dr. Weijie Yu who instructed me with cleanroom works, Dr. Yago Gonzalez-Velo and Dr. Aymeric Privat who helped me with device testing and physical model construction. I would like to thank all the lab mates for

helping on the experiments and having funs in the life. I would like to thank Dr. Jiyong Woo for kind experimental advises and Dr. Zihan Xu for living instructions. I would like to thank all the classmates, cleanroom buddies and office colleagues at ASU for kind advices and helps in course works and research works.

Between May 2017 and Aug. 2017, I did my first internship at Micron Technology Inc. where I made a lot of friends. I would like to thank my manager Jim Fulford who provided me the intern opportunity to work at Micron. I would also like to thank Dr. Peng Cheng who offered me lots of suggestions for job hunting and shared a bunch of life experiences with me.

Besides, I would like to thank my three good friends Laidong Wang, Haoxiang Xu, and Gejian Zhao at ASU also whoever provided me sincere advices and generous helps in study, research and living. Meanwhile, I would like to thank all the relatives in my family and friends back in China who are always giving me supports and wishes. Moreover, I would like to show my appreciation for all the funding agencies who supported my research works including DARPA, DTRA, DOD, AFRL, NSF, NASAJPL, Raytheon, and ASU. Without their supports, I would not be able to finish all the works.

Last but never the least, life would not only have positive sides but also the negative sides. However, I would like to thank all the negative sides which gave me lessons for my future life and taught me to grow up.

Thanks All!

Runchen Fang

Apr. 2018

@ Tempe AZ

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

# CHAPTER 1

# INTRODUCTION

## 1.1. CMOS Scaling

Since the invention of the complimentary-metal-oxide semiconductor (CMOS) integrated circuit (IC) in 1958 [1-3], people's life styles have undergone a huge changes. With continuous advances of CMOS technologies into the 21$^{st}$ century, we are building faster systems which consume less power. According to Moore's law, in order to achieve faster and denser core computing components, we reduce the feature size of the current cutting-edge CMOS transistor by 30% every other year [3-6]. To achieve this goal, various scaling rules such as constant electric field scaling [6] and generalized scaling [7] have been proposed and used in the semiconductor industry. With scaling rules, we have successfully fabricated smaller transistors in each technology generation (node) but we are now facing the physical limit of scaling.

In the early technology nodes, $SiO_2$ was the gate oxide material of choice due to its excellent and clean interface with Si and its very large bandgap (~9 eV). By 1990s, the transistor size had scaled down to 130 nm and the $SiO_2$ thickness scaled to about 3 nm. At this thickness both the direct tunneling-induced gate leakage current and the boron dopant penetration had severely impacted the continuing of scaling predicted by Moore's Law [8]. Fortunately, by incorporating nitrogen into the $SiO_2$ gate dielectric, scaling rules could continue through the rest of the decade. One advantage to the use of nitrogen was that it increased the dielectric constant (k) of the gate oxide and enabled the effective gate oxide thickness to be scaled down to around 1 nm [9]. Silicon oxynitride (SiON) has been in mass production for around a decade as it satisfied both the demanding transistor design

and the reliability concerns. However, continued scaling over the past decade only increased the direct tunneling threat and stopped the scaling down of the gate oxide when the thickness of oxynitride got below 1 nm [10]. Meanwhile, the depletion of the poly-silicon gate electrode has also limited the transistor operation speed by adding the parasitic capacitator on the gate [11]. Consequently, dielectrics with high k dielectrics were developed for the gate oxide accompanied by different gate electrode materials.

At the 45 nm technology node, hafnium oxide high-k dielectrics plus metallic gate electrodes (HK/MG) were introduced by Intel [12]. The higher dielectric constant ($k$~15-20) and thus larger permittivity of the new $HfO_2$ layer enabled the oxide to be made thicker while maintaining the same oxide capacitance per unit area as an $SiO_2$ gate capacitor for the targeted node. Using this high k material, the effective oxide thickness (EOT) could be made thinner than 1 nm, thereby increasing transconductance (and speed) while keeping a thicker physical oxide thickness, which reduced the leakage threat. The equation for calculating EOT is:

$$\mathrm{EOT} = \frac{k_{SiO_2}}{k_{HfO_2}} t_{HfO_2}. \tag{1.1}$$

In addition to gate oxide scaling continued reductions in gate length for planar transistors also posed challenges. At the 22 nm technology node for Intel, 16 nm technology node for TSMC and 14 nm technology node for Globalfoundries, non-planar FinFET transistors with a 3D structures were finally introduced , continuing adherence to Moore's law [13-15]. It has been claimed that 3 nm or some would say 1 nm gate lengths will be the final technology node for transistors [16] and that starting at the 5 nm node, gate-all-around (GAA) or nanowire technologies would be employed [17]. Thus, devices

with novel concepts, novel materials and even novel computing architectures have been an area of very active research and development in recent years [18-20].

## 1.2. High-k Dielectrics

The introduction of the high-k dielectrics had its origins from the early adoption of nitride in the gate dielectrics. With size scaling down, higher-k dielectrics were proposed to replace the old SiON technology. Today most of the foundries use $HfO_2$ gate dielectrics even in the latest 10 nm FinFET technology. The real gate stacks actually consists of two layers, a high-k dielectric layer and an interfacial $SiO_2$ layer. So the Equation 1.1 actually should be revised as:

$$\mathrm{EOT} = t_{SiO_2} + \frac{k_{SiO_2}}{k_{HfO_2}} t_{HfO_2}. \qquad (1.2)$$

### 1.2.1. High-k Dielectrics Candidates

In addition to $HfO_2$, the most common high-k dielectrics candidates are $Ta_2O_5$ [21-23], $SrTiO_3$ [24-25], and $Al_2O_3$ [26-27], which have dielectric constants ranging from 10 to 80. For $Al_2O_3$, it has a large bandgap of 8.8 eV as well as a large conduction band (CB) and valence band (VB) offsets. However its permittivity is not large enough to last for too many device generations [26]. $Ta_2O_5$ has been used in dynamic random access memory (DRAM) applications. It forms an excellent interface with the native interface oxide since it could have interface state density ($D_{it}$) comparable to the older technology transistors that use only $SiO_2$ gate dielectrics. Though the high permittivity (20) of $Ta_2O_5$ could sustain several device generations, its small bandgap (4.5 eV) and small conduction band offset with the Si conduction band (0.35 eV) led to considerable direct tunneling current [21]. Perovskites

used in DRAM like $SrTiO_3$ require control of the channel interface for the dielectric deposition achievable through molecular beam epitaxy (MBE) to grow the dielectric layer by layer. The permittivity of the $SrTiO_3$ deposited by such methods can be as high as 300. However, MBE is a poor throughput method for CMOS fabrication processes which prevents its widespread use in large scale semiconductor manufacturing [24].

Deeper exploration of the high-k dielectrics identified the primary tradeoff as being between the band gap energy and dielectric constant, both of which are desired to be large. Fig. 1.2.1 plots the bandgap vs. dielectric constants relationship for many of the candidate gate materials. Optimal candidates should be dielectrics with a k value of 25-30 with corresponding bandgaps greater than 4 eV (preferably 5 eV). Other than this, thermodynamic and kinetic stabilities should also be considered. These two stabilities require the dielectrics not to react aggressively with Si and should be compatible with the existing CMOS manufacturing processes. Meanwhile, the band offset with the Si conduction is also an important consideration. As shown in Table 1.1, the best candidates are $ZrO_2$, $HfO_2$, $La_2O_3$ and other lanthanides like $Pr_2O_3$, $Gd_2O_3$ and $Lu_2O_3$. Unfortunately $ZrO_2$ forms a silicide with Si and creates $ZrSi_2$ [28] while lanthanides are all hygroscopic [29, 30]. So $HfO_2$ has become the top choice for manufacturers.

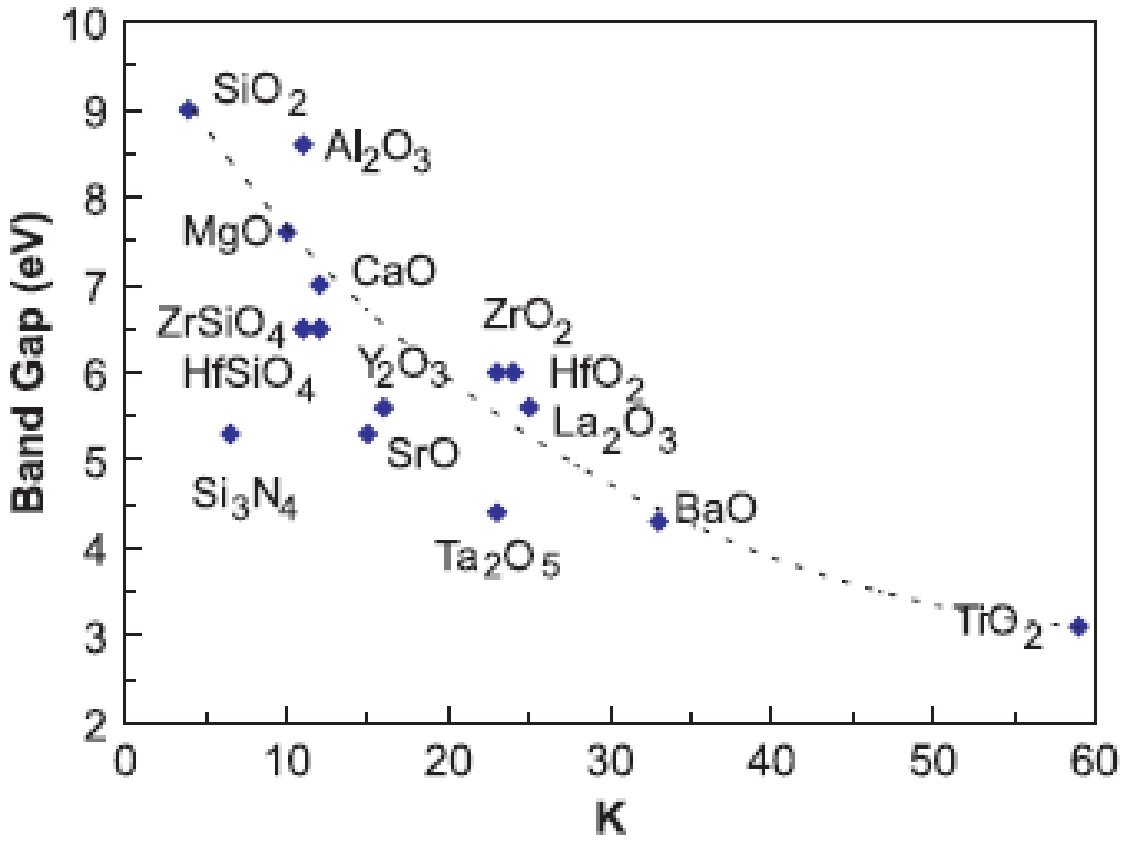


Fig. 1.2.1. Static dielectric constant vs. band gap for candidate gate oxides. Reprinted from [31] with permission of EDP sciences.

Table 1.1.

Static dielectric constant (K), experimental band gap and (consensus) conduction band offset on Si of the candidate gate dielectrics. Reproduced from [31] with permission of EDP sciences.

| | K | Band Gap (eV) | Si CB offset (eV) |
|---|---|---|---|
| $SiO_2$ | 3.9 | 9 | 3.2 |
| $Al_2O_3$ | 9 | 8.8 | 2.8 |
| $Ta_2O_5$ | 22 | 4.4 | 0.35 |
| $TiO_2$ | 80 | 3.5 | 0 |
| $SrTiO_3$ | 2000 | 3.2 | 0 |
| $ZrO_2$ | 25 | 5.8 | 1.5 |
| $HfO_2$ | 25 | 5.8 | 1.4 |
| $La_2O_3$ | 30 | 6 | 2.3 |
| $Y_2O_3$ | 15 | 6 | 2.3 |
| $HfSiO_4$ | 11 | 6.5 | 1.8 |

### 1.2.2. High-k Dielectric Implementations

Currently, $HfO_2$ is deposited by an atomic layer deposition (ALD) method which is a modification of the chemical vapor deposition (CVD) process. ALD controls the growth speed precisely to Å level or monolayer atom level which leads to very excellent step coverage and conformal deposition [32]. Thus the deposited films are very continuous and pin-hole free. Though the processing speed is slow, it is extendible for multiple substrates processing in parallel [33]. Besides, its self-limiting nature of surface reactions will

produce non-stochastic films with amorphous crystalline structures which are preferred for the dielectrics and diffusion barriers [34].

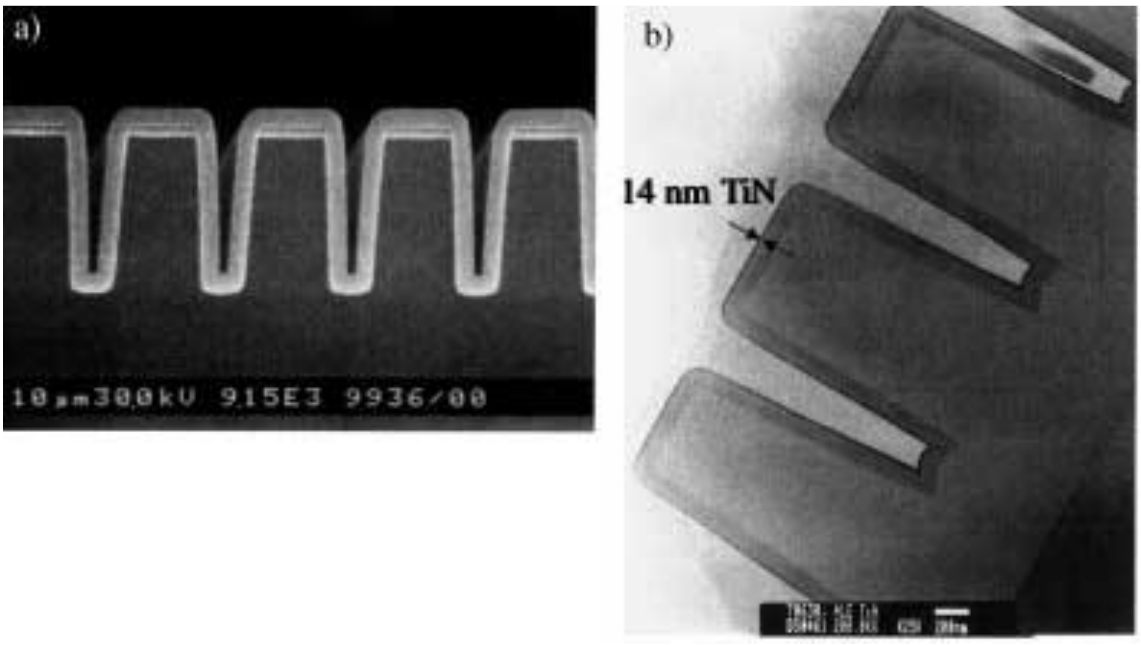


Fig. 1.2.2. Cross-sectional SEM images for a 300-nm $Al_2O_3$ film (a) and a 14-nm TiN film (b) deposited on a patterned silicon substrate. Reprinted by permission from John Wiley and Sons: Angewandte Chemie International Edition, Ref [35].

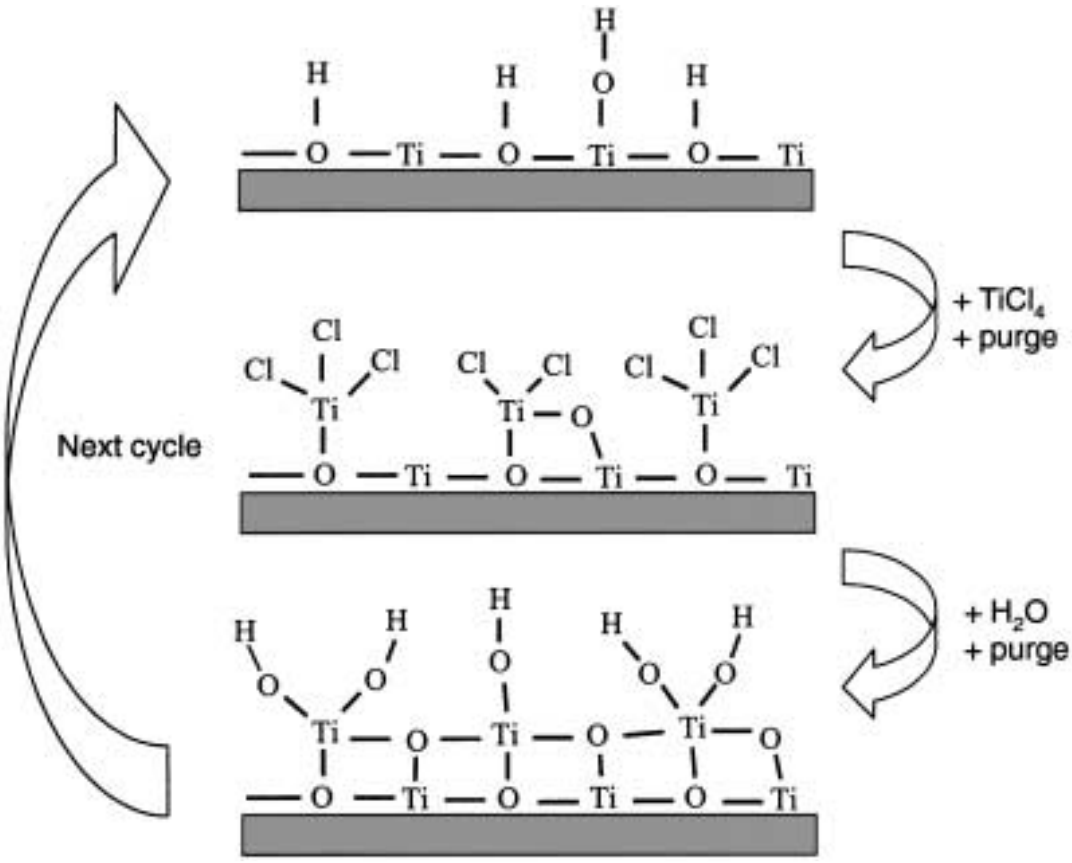


Fig. 1.2.3. An ALD deposition cycle showing schematically the deposition of $TiO_2$ films from $TiCl_4$ and $H_2O$. Reprinted by permission from John Wiley and Sons: Angewandte Chemie International Edition, Ref [35].

The cross-section SEM images showing the films deposited by ALD are shown in Fig. 1.2.2 [35]. As indicated by the images, ALD has excellent conformity and reproducibility for the large area processing. The basic principles of ALD are shown in Fig. 1.2.3 which briefly illustrates $TiO_2$ film deposition from $TiCl_4$ and $H_2O$ [35]. ALD deposits films in a cyclic manner which contains (1) exposure of the first precursor, (2) purge of the reaction chamber, (3) exposure of the second precursor, and (4) a second purge of the reaction chamber. These four steps constitutes one cycle of the ALD process. By repeating such

cycle, films are deposited on the substrate with theoretically each cycle one layer of the targeted film.

Other than ALD methods, there are other deposition methods for high-k dielectrics such as reactive sputtering [36], pulsed laser deposition (PLD) [37], and chemical vapor deposition (CVD) [38]. Reactive sputtering is a physical type of film deposition that employs ionized argon gas bombarding the metal target to eject the metal atoms out, forming vapor clouds. Meanwhile oxygen is blended with argon in the chamber, which creates metal vapor clouds in the chamber that form the molecules of the targeted material films and deposit them on the substrate. PLD is also a physical method. However, unlike sputtering the metal atoms are not ejected by argon bombarding but instead by pulsed laser heating. CVD is a chemical technique for film deposition which is quite similar to ALD. Instead of having different precursors flowing into the chamber at different steps, CVD has all the precursors flowing in at the same time and then one time purge will be performed to clean all the byproducts.

According to the summarized comparison between ALD, CVD and PVD [36, 38] in Table 1.2, it is easy to conclude that ALD is the best option for high-k dielectrics used in thin gate oxides.

Table 1.2.

Technology comparisons between ALD, CVD and PVD [36, 38]. AR stands for aspect ratio.

| Index | ALD | CVD | PVD |
|---|---|---|---|
| Thickness Range | < 200 Å | > 200 Å | > 200 Å |
| Uniformity | Å range | 10 Å range | 50 Å range |
| Film Quality | Excellent | Excellent | Limited |
| Conformity | 100% step coverage in AR 60:1 | 100% step coverage in AR 10:1 | 50% step coverage in AR 10:1 |
| Extendibility | No Limit | 90-65 nm | 100 nm |
| Vacuum Requirement | Medium | Medium | High |

## 1.3. Reliability Issues in HK/MG Based MOSFETs

While $HfO_2$ have been the primary choice for HK/MG MOSFETs there are some important reliability issues with these material systems. First, the $HfO_2$/Si interface is not as clean as $SiO_2$/Si interface and the reaction between $HfO_2$ and Si will create a layer of $HfSiO_x$ [39]. The reaction between dielectrics and the Si will generate undesired disorders at the interfacial region which can degrade channel mobility and create large densities of interface traps and oxide defects. To mitigate such problems, as can be seen in Fig. 1.3.1, an interfacial layer (IL) of $SiO_2$ can be deposited on Si by CVD or thermal oxidation [40]

to push the disordered region into the dielectric layer. Still, the disordered region in the dielectrics will lead to a large density of bulk oxide traps whose densities are denoted as $N_{ot}$.

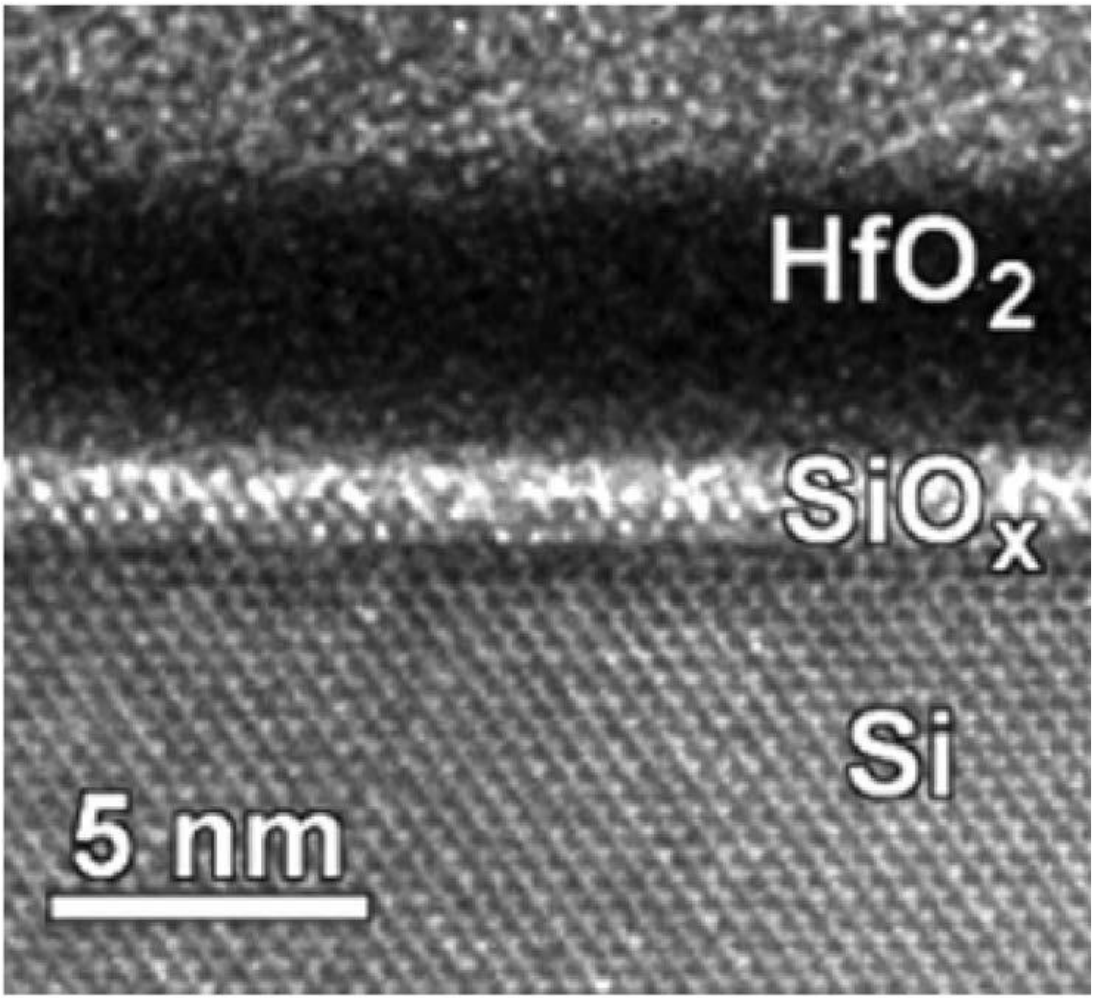


Fig. 1.3.1. TEM image of the 3.4 nm $HfO_2$ on Si (001) with 1 nm interfacial $SiO_2$ Reprinted with permission from [40]. Copyright 2006, American Institute of Physics.

In most advanced CMOS technologies today the thickness of the IL is less than 1 nm. Since the IL thickness will decide how close the bulk oxide traps are to the Si channel if it

is too thin these close to channel bulk oxide traps will be able to capture carriers from the channel through tunneling processes and act as charged centers in gate dielectrics. Second, though the high-k stack/Si interface have been preserved to be as good as $SiO_2$/Si interface, it naturally contains some dangling bonds with hydrogen as well [41]. These bonds can be broken under electrical operations at room temperature or at elevated temperature forming charged states that affect the channel surface potential. These dangling bonds are also called interface states. Removal of interface traps requires high external energies (e.g. high temperature cycling) or long anneal times [42]. Interface traps are commonly is denoted as $N_{it}$ with a unit of density per area or as $D_{it}$ with a unit of density per energy per area.

During operation, these oxide defects can be either charged/discharged or created/annealed causing the electrical characteristic of the MOSFET to degrade. One type of operation or stress-induced degradation is called bias temperature instability (BTI) [43]. For a fresh (un-stressed) MOS stack tunneling current should be very low. But if the transistors are stressed by a constant voltage higher than $V_{dd}$ or even at some elevated temperature, leakage current through the gate stack can gradually increase with time which is also called stress-induced leakage current (SILC) [44]. It is believed SILC mostly is due to the creation of the new defects either at the interface or inside the oxide dielectrics. When stress time increases, the leakage will further increase and eventually cause the dielectric breakdown. If the dielectric breakdown is partially recoverable, then it is called soft breakdown (SBD) [45], otherwise, the gate stack will permanently lose its ability to hold charge by experiencing a hard breakdown (HBD) [46], which is also called time-dependent dielectric breakdown (TDDB).

There are other types of reliability issues in the MOSFETs such as hot carrier injection (HCI) [47], random dopant fluctuations (RDF) [48], line edge roughness (LER) [49], metal-gate granularity (MGG) [50], and so on that might be related to either fabrication process randomness or long time operational stress. However, these issues are not the focus of this work.

## 1.4. Resistive Switching Phenomenon in High-k Dielectrics

The phenomenon of resistive switching in the oxides was originally discovered in the 1960s when SiO was under investigation [51]. At that period, the phenomenon was also called voltage controlled negative resistance (VCNR) which presents a hysteretic current-voltage (I-V) curve as shown in Fig. 1.4.1 [52].

The resistive switching phenomenon has been extensively investigated in various material systems including oxide systems [53-55], nitride systems [56], chalcogenide systems [57-59] and polymer systems [60-62]. Among all these systems, resistive switching, is dependent on the growth and dissolution of a conductive filament (CF) inside the insulator [63].

The form of these filaments can be classified in two basic categories. In one type the filament is made up of vacancy defects in the dielectric [64]. In the second type, the filament is made up of reduced metal ions [65]. The first type has been developed into device called oxide random access memory (OXRRAM) [66] and the second type has been developed into device called conductive-bridge random access memory (CBRAM) or programmable metallization cell (PMC) [67]. Both OXRRAM and CBRAM are types of

RRAM, which is regarded as one of the most promising candidates for the future generation nonvolatile memory [68].

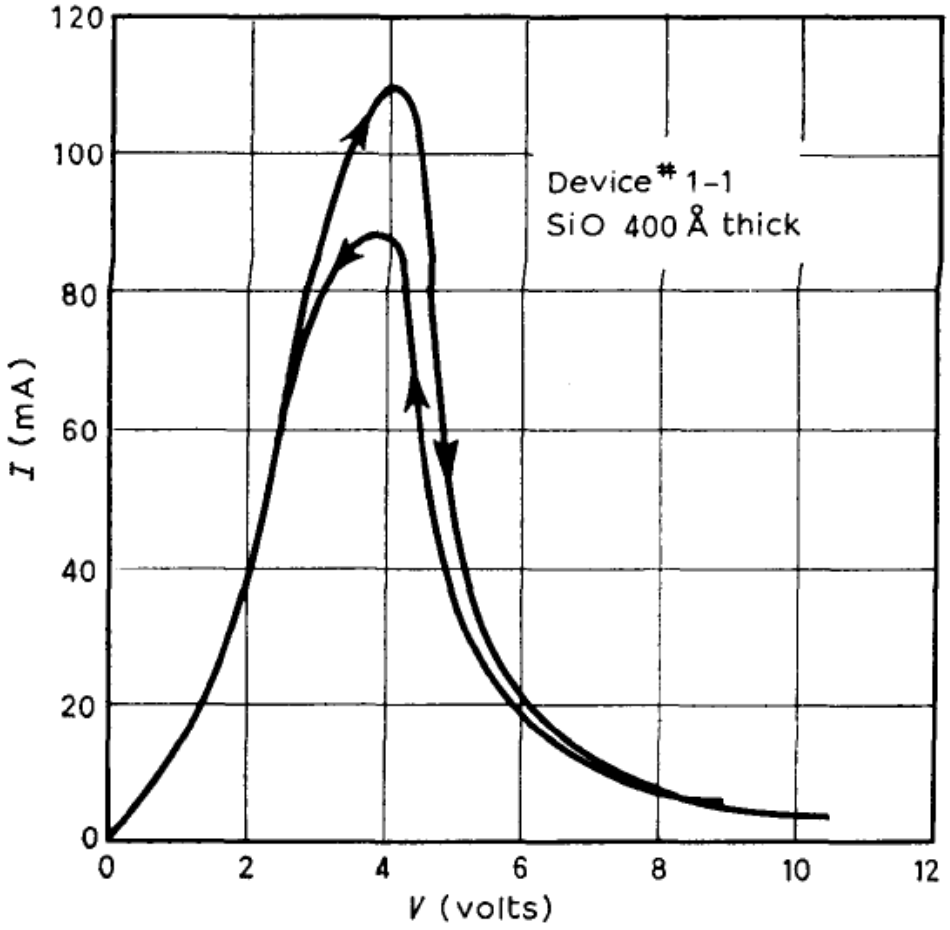


Fig. 1.4.1. Current-voltage characteristics of 400 Å thick insulator device. © 1967 IEEE. Reprinted with permission from Ref. [52].

In OXRRAM, the resistive switching phenomenon seems to share similarities with the dielectric breakdown process especially in gate oxides. Both of them have the insulating states in the very beginning and start to conduct as bias is applied. Such fundamental similarities suggest OXRRAM operation and breakdown in gate oxides share a common

physics. An investigation of these commonalities is one of the primary goals of this dissertation.

### 1.4.1. RRAM Device Structure and Operation

A typical RRAM cell is a two-terminal structure with either metal-insulator-metal (MIM) or metal-insulator-semiconductor (MIS) which has been widely used in the commercial DRAM cells. Shown in Fig. 1.4.2 are the typical structures used in the research presented in this manuscript. The devices use either a (a) common ground or (b) cross-point structure. The switching filament is located between the top electrode (TE) and the bottom electrode (BE) thus the cells would not interfere with each other. $MeO_x$ denotes the transition metal oxide in which the filament is formed (or dissolved).

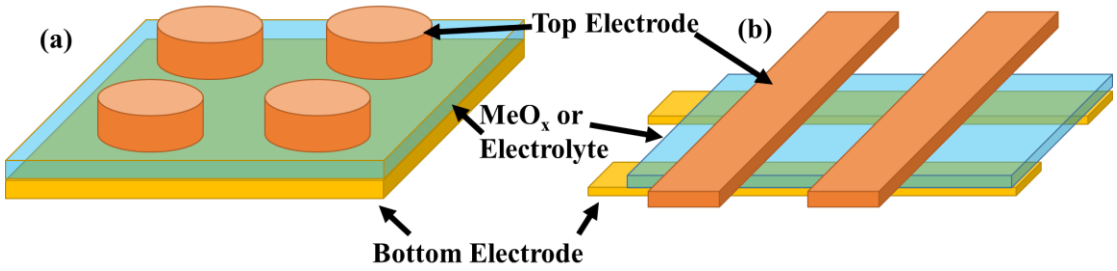


Fig. 1.4.2. The typical structures of RRAM cells used in this work.

Un-programmed (fresh) devices usually need a DC voltage sweep from 0 V to a high voltage (like 8 V or 10 V) to perform the Forming process (OAA') as shown in Fig. 1.4.3 [69]. After that, as indicated by the arrow, the devices would be able to switch between two different resistance states, i.e., HRS and LRS which stand for high and low resistance state, respectively. The two states can be encoded as '1' or '0' to store information in a

binary way. The process OBB'O to switch the device from HRS to LRS is called SET, and the reversed process OCDC'O is called RESET. These operations could also be performed by pulsed voltage operations. The combination of a SET and RESET operation in sequence is called a cycle. If a cycle is performed with DC bias, then it is called DC cycling. Similarly, if a cycle is performed with transient pulses, then it is called pulse cycling. For the evaluation of the potential memory application, distributions of LRS and HRS over many pulse cycles are needed. For memory applications it is required that the resistance states should be retained for long time. To evaluate this characteristic, two different methods can be performed. One method is called retention testing which is performed by reading the current at certain resistance state with a constant low voltage bias but under elevated temperature. The other method is called read disturbance which is performed by reading the current at certain resistance state with a constant high voltage bias at room temperature. The two methods are evaluating either temperature dependent or voltage dependent reliability decay.

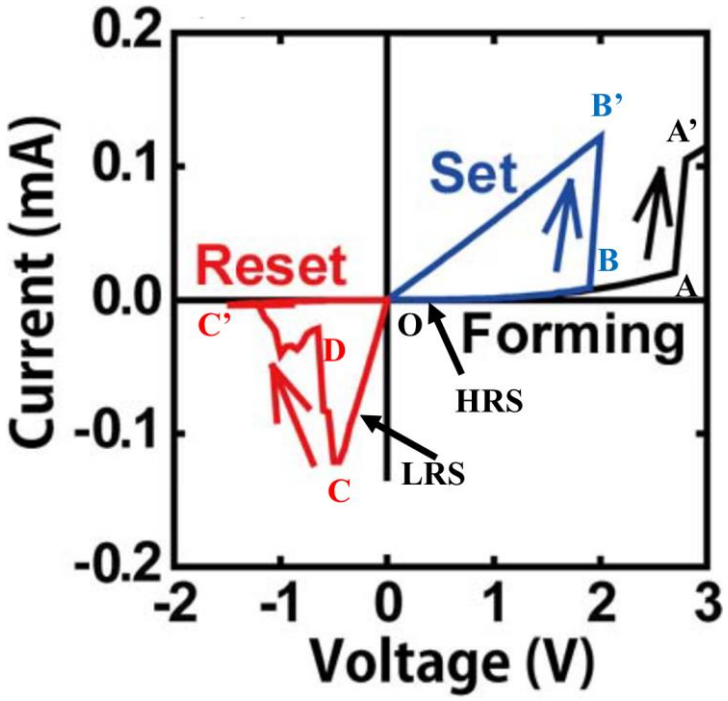


Fig. 1.4.3. Operations of the two-terminal RRAM devices. The voltage sweep is done by the top electrode drive. © 2010 IEEE. Reprinted with permission from Ref. [69].

1.4.2. Popular OxRRAM Candidates

Several binary metal oxides have exhibited the resistive switching phenomenon. The most common transition metals used in OXRRAM are identified on the periodic table in Fig. 1.4.4. Thus the candidate oxide material for OXRRAM are $HfO_x$ [70-72], $TaO_x$ [73, 74], and $TiO_x$ [75-77]. The $HfO_x$ OXRRAM devices has been extensively studied [70-72, 78-85]. In this dissertation, $HfO_2$ based OXRRAM devices will be the primary focus.

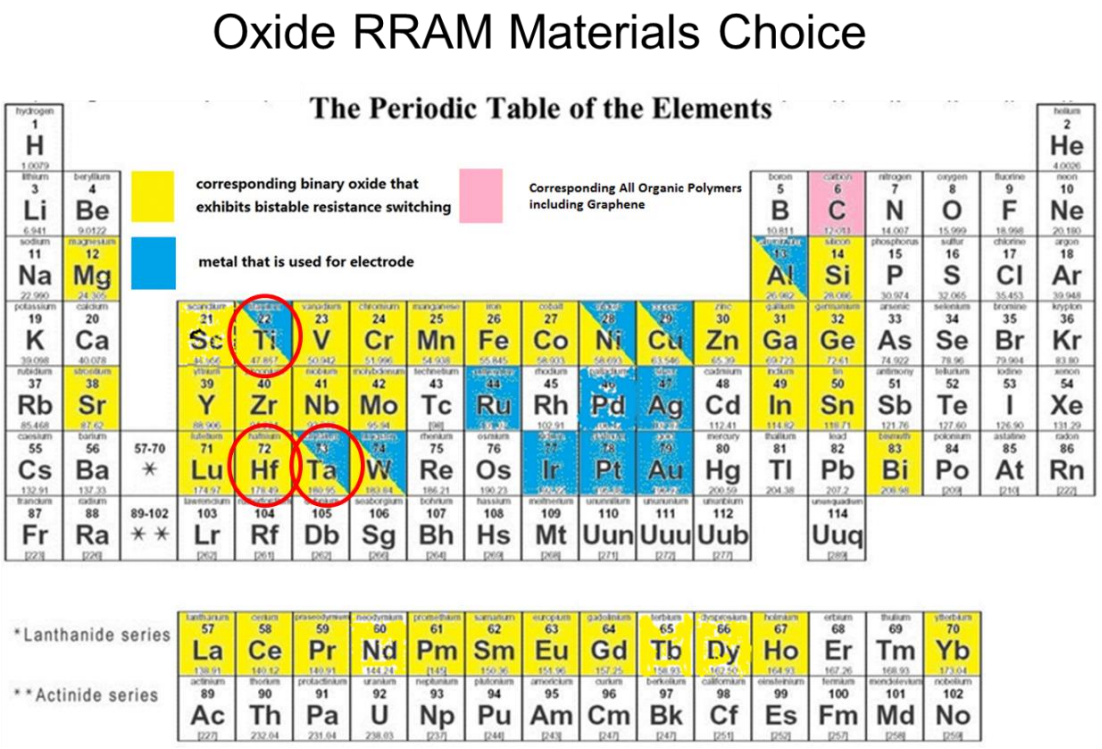


Fig. 1.4.4. Binary oxide RRAM Materials choice summarized in the Periodic Table. Yellow color denotes the elements used for metal oxide, blue color denotes the elements used for electrodes and the pink color denotes all the polymers/graphene based resistive memory.

### 1.4.3. Merits and Concerns of OxRRAM

According to the published works, OxRRAM devices have shown excellent performance, suggesting that the technology is a promising replacement for either FLASH memory or DRAM in the future. In view of operation energy and speed, under programming voltages from 1 V to 3 V, RRAM can be switched within very short times ranging from 300 ps to 10 ns with the programming current below 100 μA [78-80]. Thus the total energy consumption per write operation can be as low as 0.1 pJ. From the aspect

of long time usage, the number of cycles that can be performed can range between $10^6$ times to $10^{10}$ cycles without using error-correction code (ECC) and with each state retainable for more than 10 years at 85 $^oC$ [81-83]. This is much better than the current FLASH technology. With respect to the technology, RRAM cells have already been fabricated with an 8 nm * 8 nm feature size and still retain the ability to further scale down [84, 85]. For the memory on chip technique, RRAM can be fabricated between the fourth or fifth interconnect metal layer with a cell area less than 4 $F^2$ if stacked [86] where F denotes the minimum feature size in manufacturing. Last but not least, the desired amorphous oxide materials in the RRAM devices usually are fabricated by ALD or sputtering which are all relatively low temperature (<300 $^oC$) CMOS process compatible techniques [87, 88].

Though RRAM has these merits, there are critical drawbacks to the technology that have set back its commercialization. The foremost issue are resistance state variations which originate from the instabilities in the conductive filament(s), CF(s), and the complicated environment inside the oxide matrix [89-91]. Such variations lead to wide resistance state variations and can even lead to flipping of states (i.e., bit errors). Recent works indicate that this often happens in the intermediate states and would have greater impact on multi-level cell (MLC) architectures. In order to avoid this, it is suggested to make the CF(s) more conductive in LRS [91] and make resistance even higher in HRS [90], which means a RRAM with very low LRS and very high HRS. Developing RRAM with wide separation between LRS and HRS thus reduces errors and improves retention but makes MLC devices difficult to realize and can also impact cycling ability. To clarify, if many cycles are required, each state should be flipped fairly easily. In order to replace DRAM in the

computer system, the cycling capability might need more than $10^{12}$ times rather the present $10^{10}$ times. If long time retention is required, each state should be difficult to be flipped. In addition to the inherent tradeoff between endurance and retention, in order to maintain low switching power target, current draw during RRAM write operations should be reduced to micro-amps or below. Cell-to-cell and the uniformity in RRAM also needs to be improved.

### 1.4.4. Other Emerging Memories

In addition to RRAM, there are several other emerging memory technologies. The growth in research on the emerging memories is due to the broadening gap between memory and processor performance in which the memory speed now lag far behind the processor speeds [92]. Other NVM technologies widely studied in the last decade include spin-torque transfer magnetic random access memory (STT-MRAM) [93], phase change random access memory (PCRAM) [94], ferroelectric random access memory (FRAM) [95], and ferroelectric field effect transistor (FeFET) memory [96]. The above list represents the most popular candidates to date which are showing excellent characteristics in one or more aspects. Table 1.4.1, which is summarized from Ref. [97] and Ref. [98], presents the metrics comparison of these popular emerging memory candidates with the conventional memories used in the current computer systems. As can be seen from this table, STT-MRAM is quite close to SRAM and DRAM in performance, while PCRAM and RRAM compare much better wiht FLASH memory. Therefore, it has been proposed to replace SRAM and DRAM with STT-MRAM as cache and main memory and PCRAM or RRAM with FLASH as storage class memory [97].

Aside from the popular candidates already fabricated in chips, there are also many other candidates under research including Mott memory based on the Mott transition [99], Macromolecular-based memory which has shown great potential in flexible electronics [100], Molecular-based memory [101], and Carbon-based memory [102].

Table 1.3. Benchmark of Conventional and Emerging Memory Technologies. © 2016 IEEE. Reprinted with permission from Ref. [97, 98].

| Index | SRAM | DRAM | NAND-FLASH | RRAM | PCRAM | STT-MRAM | FeRAM |
|---|---|---|---|---|---|---|---|
| **Cell Size ($F^2$, SLC)** | 140~280 | 6 | 4 | 4~12 | 4~30 | 6~50 | 15~35 |
| **Operation Voltage (V)** | 0.6~1.1 | 0.6~1.1 | ~ 20 | < 3 | 1.5~1.8 | 0.8~1.8 | 1.8 |
| **Write Current (A)** | $10^{-5}$ | $10^{-4}$ | $10^{-7}$ | $10^{-4}$ | $10^{-4}$ | $10^{-5}$ | $10^{-6}$ |
| **Write Time** | ≤ 1 ns | ~ 10 ns | ~ 1 ms | < 10 ns | ~ 100 ns | < 10 ns | < 10 ns |
| **Read Time** | ≤ 1 ns | ~ 10 ns | ~ 100 μs | < 10 ns | < 5 ns | < 5 ns | < 5 ns |
| **Retention** | volatile | ~ 64 ms | > 10 yrs | > 10 yrs | > 10 yrs | > 10 yrs | > 10 yrs |
| **Endurance** | $10^{16}$ | $10^{16}$ | $> 10^4$ | $10^6$~$10^{12}$ | $10^9$~$10^{12}$ | $10^{15}$ | $10^{13}$ |

Fig. 1.4.5 [103] has provides the comparison of the six emerging memories still being investigated extensively based on the critical review conducted by ERD using eight criterial [104]. As indicated in Fig. 1.4.5, RRAM has the best scalability compared to the other candidates and also has moderate abilities in the other aspects. Molecular based memory has very poor operational reliability and on/off ratio though it contains competitive scalability as RRAM. FeFET memory, Carbon-based memory and Mott memory all have moderate performances in all aspects. While Macromolecular-based

memory has very poor on/off ratio as Molecular-based memory. Most of the memory technologies still need more systematic and deep investigations so that there is still a long road ahead.

Bridging the gap between memory and processor is not an easy task that cannot be achieved by improving the performances of the memory alone. In this new era of big data, the old computation architecture probably is not efficient enough since there would be more data than instructions in the real applications. A novel architecture based on data-driven computations would be an even better solution.

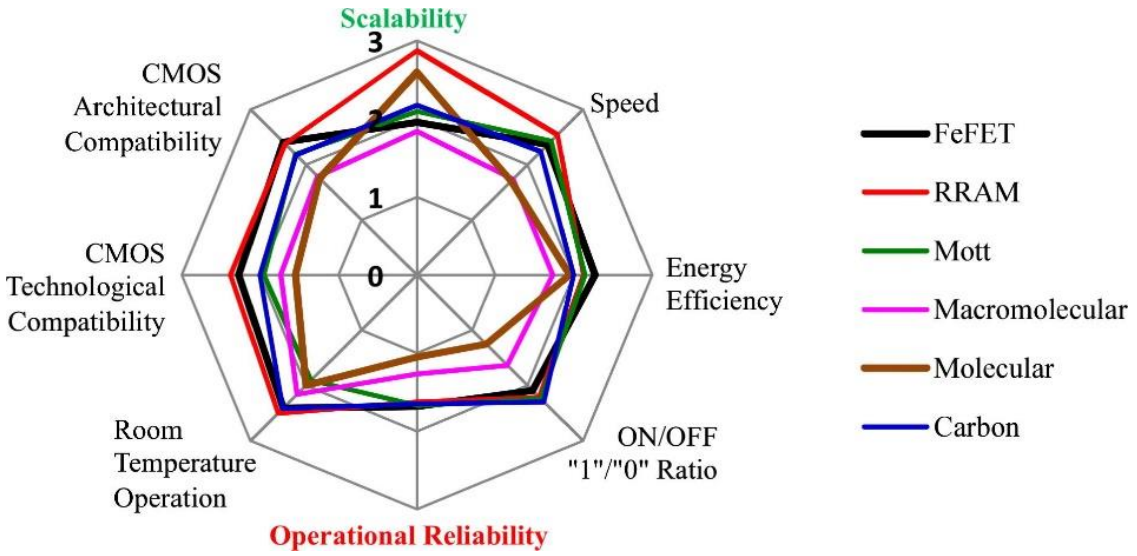


Fig. 1.4.5. 2013 ITRS ERD critical review of emerging memories based on eight criteria [104], where “3” and “1” represent the best and the worst assessment scores, respectively. Reprint from [103] 2016, with permission from Elsevier.

## 1.5. Motivation of the High-k Dielectrics Study

Size scaling has led to increasing reliability issues in the CMOS transistors of advanced technologies. The introduction of the high-k dielectrics has mitigated the reliability problems in the first several new generations of HK/MG transistors. However, for feature sizes below 28 nm, the reliability issues have become prominent again and solving them has become very complicated. Reliability studies on HK/MG advanced technologies such as BTI, SILC and TDDB have become necessary and urgent. Studies on BTI aging effects in HK/MG transistors mostly focus on the charging, discharging, and annealing of defect precursors in the oxide or at interfaces. SILC and TDDB effects share similar mechanisms with BTI effects at early stages in the stress history. However at later stages growth in defects cascades seems to bare a greater resemblance to the filamentary growth processes in RRAM. Since the MOS stack and MIM stack have many similarities, the SILC and TDDB theories developed for MOS stack may be applicable for RRAM MIM stack and vice versa. Such a theoretical extension would be able to further improve and complete the theories for resistive switching phenomenon in the oxide based RRAM devices.

## 1.6. Organization of the Dissertation

Chapter 1 has provided background for the work in this thesis, including the motivation of the reliability study in advanced technologies, the basics and terminologies of different reliability issues and resistive switching phenomenon in oxides, the current status and challenges of RRAM technology development as well as other emerging memories.

Chapter 2 discusses the BTI aging effects and its modeling in HK/MG based MOSFETs. The BTI effects are due to both the charging of the oxide traps and the creation of the

interface traps in the MOS stack. According to the stochastic bivariate defect model, dynamic defect potential can be calculated with given stress time, recover time, stress voltage and temperature. To model the BTI effects with more physical accuracy, a non-iterative surface potential model has been incorporated to precisely calculate the defect potentials caused by the various defects. Devices based on HK/MG technologies have been stressed/recovered for model validations. The circuit level aging effects due to BTI aging have been simulated with the verified model.

Chapter 3 reports the wearing out and breakdown theories in the HK/MG based MOSFETs. Different carrier transport and conduction mechanisms will be briefly discussed. Starting from the established SILC and TDDB theories of poly-Si/$SiO_2$ (SiON)/Si stack followed by a review of works pertaining to SILC on $HfO_2$ MIM structures, new theories for MG/HK/Si stack are developed. Stress results on fabricated MOS capacitors are analyzed to identify degradation mechanism studies in high-k systems. In addition, synergies between TDDB and RRAM device Forming processes are discussed to modify the current percolation path theory and further serve for building up a circuit-level TDDB compact model similar to the framework presented for BTI modeling.

Chapter 4 explores the $HfO_2$ based RRAM devices with both experiments and new compact modeling approaches. The RRAM cell structure and fabrication flow are presented. The basic electrical operations including the radiation robustness and low temperature performances are also discussed. By presenting all the potential cell structures for RRAM devices and their applications, device designs for high non-linearity has been identified to support the further development of RRAM in storage class memory arrays. The analogue resistance tuning effects and random variations of the RRAM devices would

also achieve potential applications in the novel computation architectures and cybersecurity applications,. Finally, by extending RRAM switching theories and laws from thermodynamics and thermochemistry, a new RRAM compact model concerning the multiple physical effects and entropy sources will be presented.

Chapter 5 summarizes the results and contribution of this thesis. Future work is also proposed in this chapter.

# CHAPTER 2

# BTI AGING EFFECTS IN HK/MG BASED MOSFETS

## 2.1. BTI Impacts

As mentioned in Chapter 1, BTI effects are due to the effects of near interfacial oxide trapped charge plus the interface trap creation. The major electrical embodiments of the two kinds of defects are the threshold shifts which could be observed from the $I_{ds}$-$V_{gs}$ curves. As presented in Fig. 2.1.1, the interface trap ($D_{it}$ only) aged device has an observed slope change in the subthreshold region while oxide trap charge ($N_{ot}$ only) degraded devices would exhibit a parallel shift in the subthreshold as well as strong inversion regions. After BTI stress, there would be impacts from both types of defects which cause the devices to degrade electrically, i.e., alteration to subthreshold slope, threshold voltage, etc..

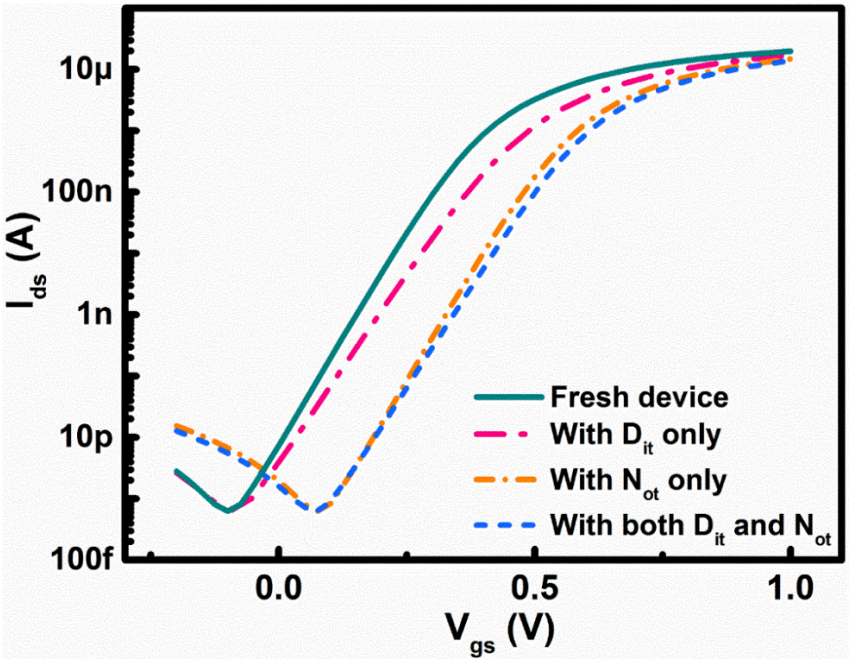


Fig. 2.1.1. Typical NMOS Ids-Vgs curves from fresh device, only interface trap ($D_{it}$) affected, only oxide trap ($N_{ot}$) affected, and both traps affected.

### 2.1.1. BTI Mechanisms

Since BTI stress are normally at bias value close to $V_{dd}$ for the corresponding technology node and under 200 $^{o}$C, the external energies injected to the gate stack of the transistor are not high enough to trigger the bulk trap generation process. Thus, for most BTI aging effects, they are attributed to the charging process of the bulk trap precursors and the creations of the interface traps at the Si/oxide interface.

The oxide traps are usually related to the oxygen atoms in the format of oxygen vacancies or oxygen interstitials. Since there is a wide energy distribution of such defects in the gate dielectrics, the oxide traps could be amphoteric which means they could capture either holes or electrons during BTI stress [105]. So this process has to be dependent on the available carriers in the channel and the physical locations of the traps [106]. Usually, the effective oxide traps should locate within 10 Å from the interface but may vary [107]. In the $HfO_2$ system, the frequently observable defects are neutral vacancies which are potential shallow donor (very close to fermi level) for Si [108]. According to the Arrhenius equation, the low activation energies will correspond to short time frame, so the oxide traps are also called fast traps or recoverable traps in the BTI aging [109].

On the contrary, the interface traps are related to the dangling bonds at the interface of $Si/SiO_2$. As discussed in Chapter 1, even for the high-k dielectrics, most technologies tend to have a $SiO_2$ IL between high-k and Si channel. The reason to have such layer is to achieve a better interface with Si whose degree of disorder will largely mitigate the surface scattering of the electrons in the channel and ensure high enough surface mobility [110]. However, even with a nearly perfect interface, there is always be some degree of disorder. At the $Si/SiO_2$ interface, there are also many natural non-bridging oxygens and Si-H bonds

which are weak and easily broken. During BTI stress, carriers with high energies can break these weak bonds, either directly or indirectly via hydrogen release in the dielectric bulk, to form interface states or dangling bonds at the interface [111]. These interface states could further capture the carriers passing by to form charged defect centers and affect the channel surface potentials [111].

Recently, Grasser *et al.* reported that both types of defects can play a role in BTI which suggests that both types of defects should be considered during the BTI stress [112].

### 2.1.2. NBTI and PBTI

According to the different polarities of $V_{gs}$ used in NMOS and PMOS, BTI effects are categorized as PBTI and NBTI. PBTI happens in NMOS when a positive bias voltage is applied while NBTI happens in PMOS when a negative bias voltage is applied. For NBTI, it is attributed to the hole trapping at oxide trapping precursors inside the dielectrics and interface trap creation caused by reactions with injected holes that lead to hydrogen de-passivation [113]. PBTI is attributed to electron trapping process and similar, although less likely, de-passivation mechanisms [114].Both types of BTI aging will cause the increase of the absolute values of the threshold voltages for which the threshold voltage shifts ($\Delta V_{th}$) has been employed as a metric to describe how much deterioration has been experience by the transistor. This means that the value of the surface potential in a channel would be gradually lowered with stress [114]. In the $SiO_2$ and SiON dielectrics , since the electron traps are typically considered to be uniformly distributed in the oxide but the hole traps are clustered at the interface, negatively charged centers in PBTI are regarded as lower compared to the positively charged centers created during NBTI [113]. With the

technology moving on to the generation when high-k materials are added into gate dielectrics, PBTI effects have significantly degraded the performance of NMOS to the extent that PBTI induced degradation is comparable with NBTI induced damage [114]. As exhibited in Fig. 2.1.2, according to measurements of $\Delta V_{th}$, for 32 nm and 28 nm technologies, PBTI induced $\Delta V_{th}$ is about 30%~40% of the NBTI induced $\Delta V_{th}$. For 20 nm technology which is the last planar technology, the $\Delta V_{th}$ caused by PBTI and NBTI are mostly comparable with each other. Moving on to 14 nm FinFET Technology, the PBTI effects have been mitigated while NBTI effects are more significant than those in the previous technologies.

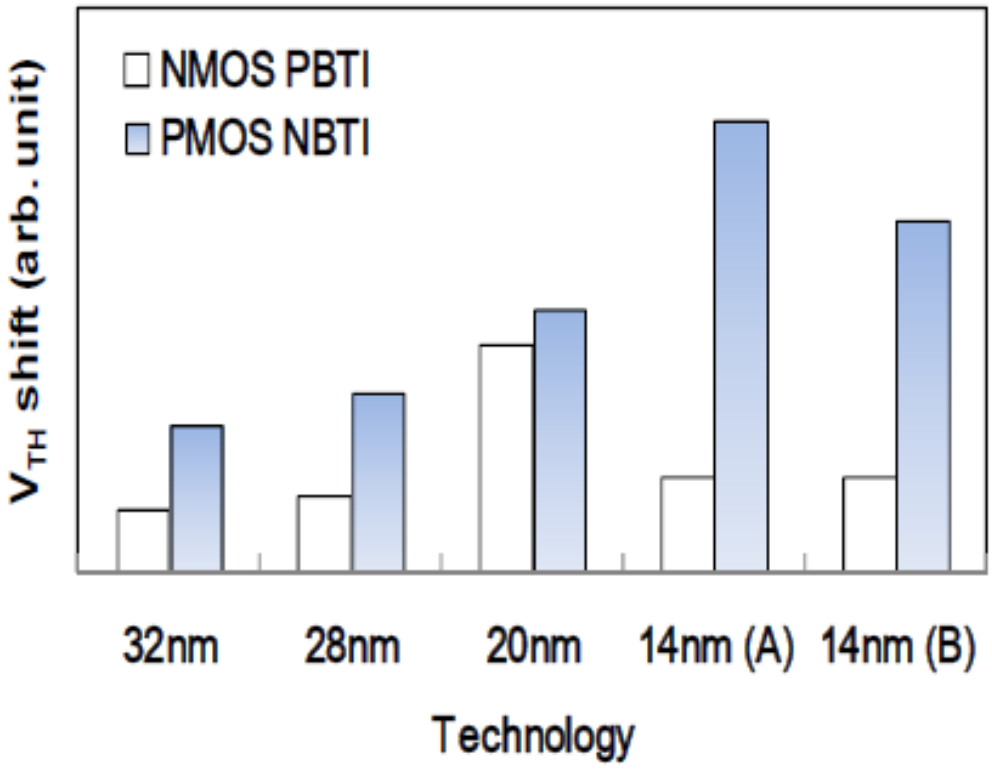


Fig. 2.1.2. The HK/MG Transistor BTI Technology Trends: 32 nm to 14nm generations at 5yrs DC, T=125°C. 14 nm BTI shown in this figure includes both early (A) and more recent (B) data. © 2013 IEEE. Reprinted with permission from Ref. [114].

### 2.1.3. Process Dependence

As mentioned in the previous section, the fabrication process flow has a large effects on device performance, so it is meaningful to study the process dependent reliability issues. According to [114], 32 nm planar transistors, 28 nm planar transistors and 14 nm FinFET transistors are fabricated by replacement metal gate (RMG) HK/MG process, while 20 nm technology planar transistor uses a gate first (GF) HK/MG process. Fig. 2.1.3 has presented the GF HK/MG process flow developed by IMEC [115, 116]. Such a process flow can significantly reduce the numbers of process steps but also causes thermal instabilities in high-k dielectrics and re-growth in the gate stack, which might enhance vulnerabilities to BTI stress [117]. With the EOT further scaling down, such process might not be applicable for the demanding high performance transistors, but rather for low cost applications.

At the same time, Intel has developed the Gate Last HK/MG process flow or RMG HK/MG process flow which has more process steps [118]. The full process flow is illustrated in Fig. 2.1.4. Starting with a dummy poly gate on high-k dielectrics, after the source/drain formation and contact formation, the dummy poly gate will be removed and replaced by metal gates which could be different for NMOS and PMOS. By adopting such process flow, both NBTI and PBTI effects could be effectively suppressed. Recent works have reported that by adding other process flows such as metal capping layer insertion or high-k dielectrics nitridation by plasma or annealing in the $NH_3$ reliability issues can be further mitigated [119].

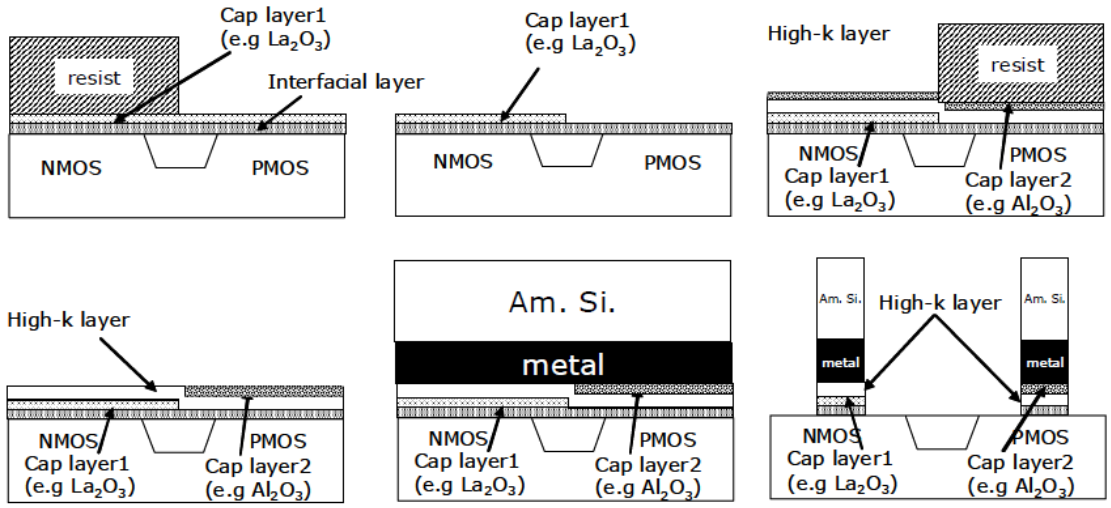


Fig. 2.1.3. Gate First HK/MG CMOS process flow developed by IMEC. © 2008 IEEE. Reprinted with permission from Ref. [115].


Fig. 2.1.4. RMG HK/MG CMOS process flow developed by Intel. © 2008 IEEE. Reprinted with permission from Ref. [118].

## 2.2. BTI Characterization

As mentioned in the previous sections, the major impact from the BTI aging is the shift of the threshold voltage $\Delta V_{th}$ which is the primary metric for damage assessment. Other than $\Delta V_{th}$, surface mobility degradation is another metric, extracted from the electrical response, specifically the saturation region of the Id-Vg curves. Thus to extract these two metrics, Id-Vg curves obtained from DC sweeps are necessary. Other transient test methods have been developed to measure/stress/measure (MSM) in the time frame of BTI response. Indeed, DC sweeps interrupt the BTI stress for relatively log times (i.e., a few seconds) during which BTI degradation typically recovers substantially [120]. To address this issue, ultra-fast measurement (UF-MSM) has been developed in which a transient sweep time has been minimized to tens of microseconds. Though during the short duration of stress interruption BTI degradation might recover, the UF-MSM ensures the fast annealing effect is characterized.

Another characterization technique is called "on-the-fly (OTF)" measurement [121]. The OTF technique has the advantage that it does not reduce gate bias and interrupt the stress state. Instead the Id is measured with a spot voltage $V_{ds}$ that puts the transistor in the linear region [121]. Since there exists a time-zero delay ($t_0$) between the starting point of the application of stress voltage and the first data point measurement, the extracted saturation current degradations will totally rely on this $t_0$. With the conventional DC technique, $t_0$ is about 1 ms, while with fast technique (UF-OTF), $t_0$ has been reduced to 1 μs [122]. Other than such issue, in order to extract the metric $\Delta V_{th}$ from the measured Ids, either approximations or empirical formalism would be adopted [123]. In order to assess

the recovery of BTI aging effects, a gate bias lower than stress voltage is required. Since such bias level is quite low, the mobility degradation would not be easily characterized.

In all, MSM and OTF characterization methods have their own advantages and disadvantages, so the choice of characterization method depends on what one needs to know.

## 2.3. Predictive Modeling of BTI Aging Effects

The response to electrical stress on the gate stack is a function of complicated physics processes. This complexity makes the development of physics-based model constructions difficult. We assume that there are two critical types of physics processes that occur during stress: 1) defect charging and 2) defect generation. Defect charging process are related to the defect energy levels, external biasing situations and ambient temperature. Defect generation processes require high energy carriers to break the bonds either inside the dielectrics or at the interface. In this work, BTI aging effects have been confined to the precursor oxide trap charging processes and interface trap generation/charging process. Once interface traps are created, they can capture carriers from the channel immediately, so there is essentially no time constant related to charging. For interface traps, defect degeneration, not trapping, is the time dependent process. As discussed before, interface traps may be broken Si-O bonds or Si-H bonds. In this work we focus only on Si-H bonds. The charge contributions from these two defect types are $Q_{ot} = \pm qN_{ot}$ and $Q_{it} = qN_{it} = \pm qD_{it}(\psi_s - \phi_b)$, respectively, where $N_{ot}$ is the net density of the charged oxide traps with units [$cm^2$ ], $\phi_b$ is the bulk potential, $D_{it}$ is the interface trap density with units [$cm^2\,eV^{-1}$]

and $\psi_s$ is the surface potential. Note the sign convention (±) for $Q_{ot}$ and $Q_{it}$ is typically considered positive for p-channel transistors and negative for n-channel transistors.

### 2.3.1. BTI Model Framework

In order to model BTI in circuit simulators, we use a voltage-controlled voltage source (VCVS) applied in series with the external gate stimulus (G) of an affected transistor, as shown in Fig. 2.3.1. Such an implementation creates a BTI-aware transistor model without modifying the internal parameters of the standard MOSFET model supplied by the technology manufacturer. The VCVS element is encoded as a Verilog-A function which incorporates technology-specific parameters (e.g., transistor channel doping, $N_a$, and gate oxide thickness, $t_{ox}$) and operational variables such as stress time ($t_s$), recovery time ($t_r$), stress voltage ($V_s$), and temperature ($T$). The means ($\mu$) and standard deviations (σ) of activation energies for defect formation and annihilation extracted from BTI data also provide critical inputs to the model. As discussed below, these inputs can be extracted from data obtained from experiments on devices fabricated in similar CMOS technologies. Once fully parameterized the function performs a self-consistent non-iterative surface potential calculation of the defect potential ($\phi_{nt}$) which sets the value of the VCVS. Using this modeling approach, transistor level circuit simulations can be performed that accurately predict BTI aging effects in ICs.

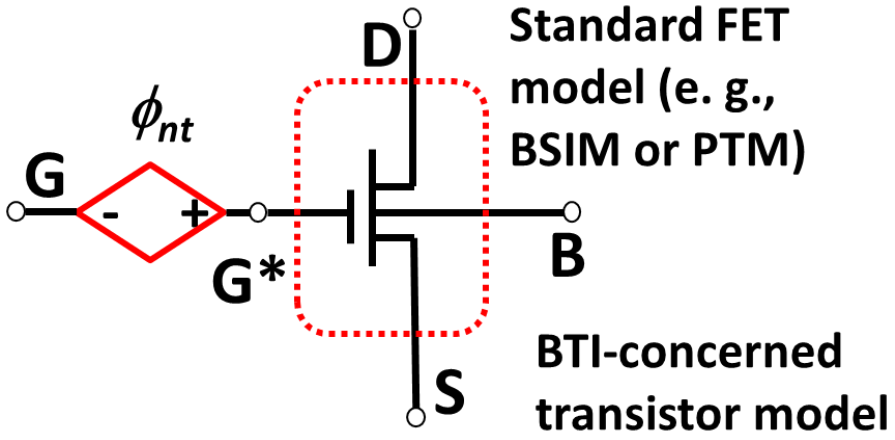


Fig. 2.3.1. Sub-circuit schematic implemented in SPICE as a BTI-enabled transistor model with $\phi_{nt} = f(V_{gb}, V_{sb}, N_{ot}, D_{it})$.

2.3.2. Non-Iterative Surface Potential Equations Solver

For bulk CMOS transistors, surface potential at specific voltage and temperature conditions can be solved with the implicit surface potential equation expressed as [124, 125]

$$(V_{gb} - \Phi_{MS} + \phi_{nt} - \psi_s)^2 = \gamma^2 \phi_t H(\psi_s/\phi_t), \tag{2.1}$$

where

$$H(\psi_s/\phi_t) = e^{-\psi_s/\phi_t} + \psi_s/\phi_t - 1 + e^{\frac{-(2\phi_b+\phi_n)}{\phi_t}}(e^{\frac{\psi_s}{\phi_t}} - \psi_s/\phi_t - 1). \tag{2.2}$$

In equation 2.1 and 2.2, $V_{gb}$ is the gate-to-body voltage, $\Phi_{MS}$ is the metal-to-semiconductor work function difference, $\gamma$ is the body factor, $\phi_t$ is the thermal voltage, $\phi_n$ is the split in the quasi-Fermi levels, $\phi_b$ is the bulk potential, and $\phi_{nt}$ is the defect potential expressed as

$$\phi_{nt}(\psi_s) = -\frac{q}{C_{ox}}[N_{ot} + D_{it}(\psi_s - \phi_b)]. \tag{2.3}$$

In equation 2.3, $q$ is the magnitude of electronic charge and $C_{ox}$ is the gate oxide capacitance per unit area. It should be noted that equations 2.1-2.3 are specific to n-channel

MOSFETs. P-channel MOSFET expressions are essentially the same except for differences in signs. Details regarding the non-iterative $\psi_s$ solving method can be found in [124, 126].

### 2.3.3. Stochastic defect distribution functions

To dynamically set the defect potential for each MOSFET in an IC, the density of bulk oxide trapped charge and interface traps are calculated as functions of the stress and recovery times applied to the transistor during circuit operation. These functions are expressed as

$$N_{ot}(t_s, t_r) = A_{ot}F_{ot}(a, b, \rho), \tag{2.4}$$

$$D_{it}(t_s, t_r) = A_{it}F_{it}(a, b, \rho), \tag{2.5}$$

where the voltage dependent pre-factors $A_{ot}$ and $A_{it}$ have units of [$cm^{-2}$] and [$cm^{-2}eV^{-1}$] and represent the maximum defect densities that can be built up during stress. Pre-factors are dependent on the magnitude of stress voltage and have the general form of $A = (V_S/V_{s0})^m$ where $V_{s0}$ and $m$ are fitting constants [112]. As originally presented by Grasser *et al.* in [127], the probability functions $F_{ot}$ and $F_{it}$ for the two defect types are double integral functions of Gaussian distributions expressed as

$$F(a, b, \rho) = \int_{-\infty}^{b} \phi(y)dy \int_{a}^{\infty} \phi\left(\frac{x-\rho y}{\sqrt{1-\rho^2}}\right) dx. \tag{2.6}$$

The variables $x$ and $y$ in equation 2.6 are functions of the mean ($\mu$) and standard deviation ($\sigma$) of the activation energy distributions for defect capture ($E_c$) and emission ($E_e$), i.e.,

$$x(E_e) = \frac{E_e-\mu_e}{\sigma_e} \text{ and } y(E_c) = \frac{E_c-\mu_c}{\sigma_c}. \tag{2.7}$$

The parameters $\mu$ and $\sigma$ are assumed to be linearly dependent on stress bias and can be extracted from BTI measurements on MOSFETs [112]. Thus for a given stress voltage and temperature, the distribution functions $\phi(x)$ and $\phi(y)$ can be analytically derived for $N_{ot}$ and $D_{it}$. Also in equation 2.6, $\rho=\sigma_c/\sigma_e$ and the non-infinite integration limits are $a=x(k_BTlog(t_r/\tau_0))$ and $b=y(k_BTlog(t_s/\tau_0))$, where $\tau_0$ is the effective time constant and $k_B$ is Boltzmann's constant. With all the parameters determined, it is possible to calculate the density for both defects as a function of $t_s$ and $t_r$.

For constant DC stress, since recovery does not occur, the integration limit $a\rightarrow$-$\infty$ and the limit $b$ is the stress time. For AC stress having a regular pulse pattern with a period of $T_s$ and duty cycle of $\gamma$, equation 2.6 would be simplified as

$$F_\gamma(a,b_1,b_2,\rho) = \int_{-b_1}^{b_2}\phi(y)dy\int_{-a\sqrt{1-\rho^2}+\rho y}^{\infty}\phi\left(\frac{x-\rho y}{\sqrt{1-\rho^2}}\right)dx, \qquad (2.8)$$

where $b_1= x(k_BT * log(\gamma * T_s/\tau_0))$ and $b_2 = x(k_BT * log(\gamma * t_s/\tau_0))$. If the stress/recovery pattern is arbitrary then the model must perform the integration step by step according to equation 2.6.

### 2.3.4. Model Parameterization and Implementation

In the next section, our BTI-concerned transistor model (Fig. 2.3.1) is employed in transistor-level simulations on various digital circuits. All circuits are designed in the same 28 nm bulk foundry technology. In order to supply defect information to the functions we have used previously published data on NBTI-induced threshold voltage shifts ($\Delta V_{th}$) obtained on p-channel MOSFETs from a similar high-k/metal-gate technology [128]. The stress and recovery experiments in [128] were conducted under various bias and temperature conditions using an ultra-fast measurement-stress-measurement (UF-MSM)

method. The published results from experiments on p-channel transistors are plotted as open symbols in Figs. 2.3.2(a-d) [128].

Assuming $\Delta V_{th} = |\phi_{nt}(\psi_s = 2\phi_b)|$, the data in Fig. 2.3.2 can be fit with the appropriate choice of parameters in equations 2.3-2.6. These fits are plotted as solid lines in Figs. 2.3.2(a-d). As can be observed, the analytical calculations fit well to the experimental data under various bias and temperature conditions. The extracted defect densities for p-channel MOSFETs corresponding to the stress/recovery $V_{th}$ shifts are shown in Figs. 2.3.3(a) – 2.3.3(d). Though the stress performed in experiments only contain data points from $10^{-5}$ s to $10^{3}$ s, the simulated stress has been extended to range from $10^{-7}$ s to $10^{8}$ s.

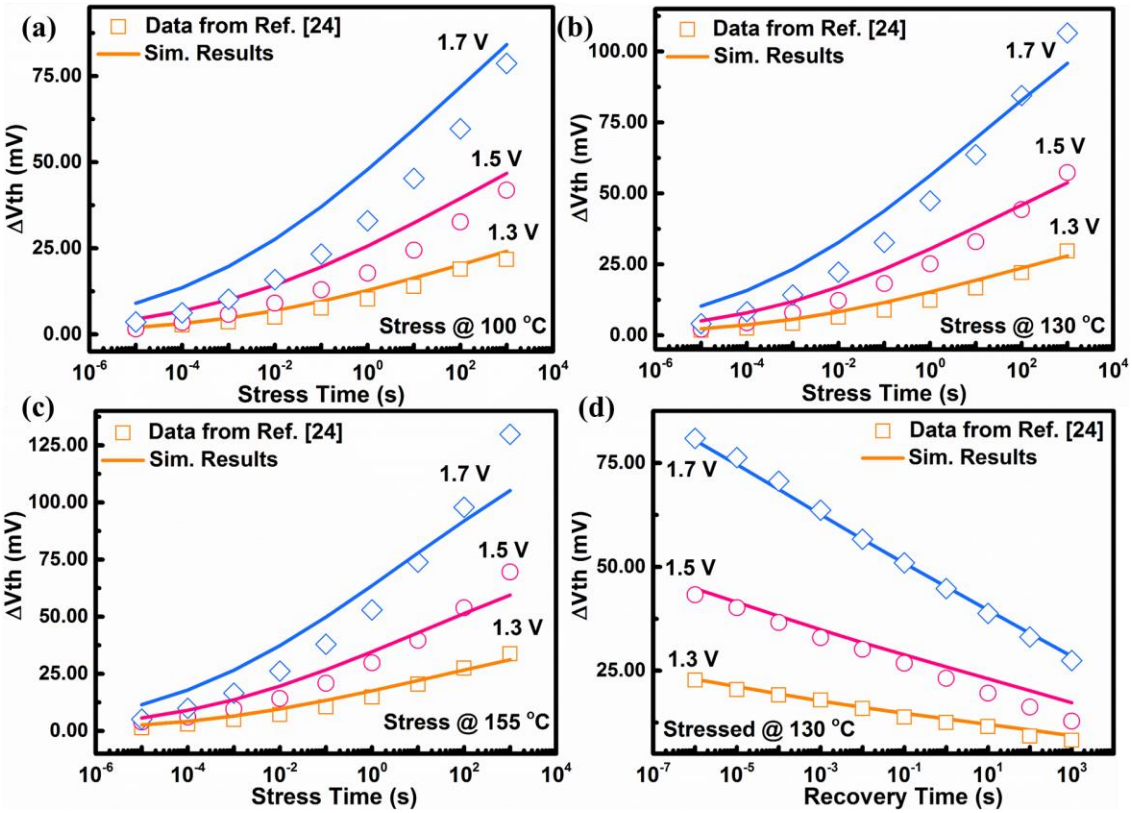


Fig. 2.3.2. The experimental data (open symbol) from [128] and simulated results (solid line) used the BTI-enabled compact model with (a) stress at 100 $^{o}$C under different bias, (b) stress at 130 $^{o}$C under different bias, (c) stress at 155 $^{o}$C under different bias, and then (d) recovery after stress at 130 $^{o}$C under different bias.

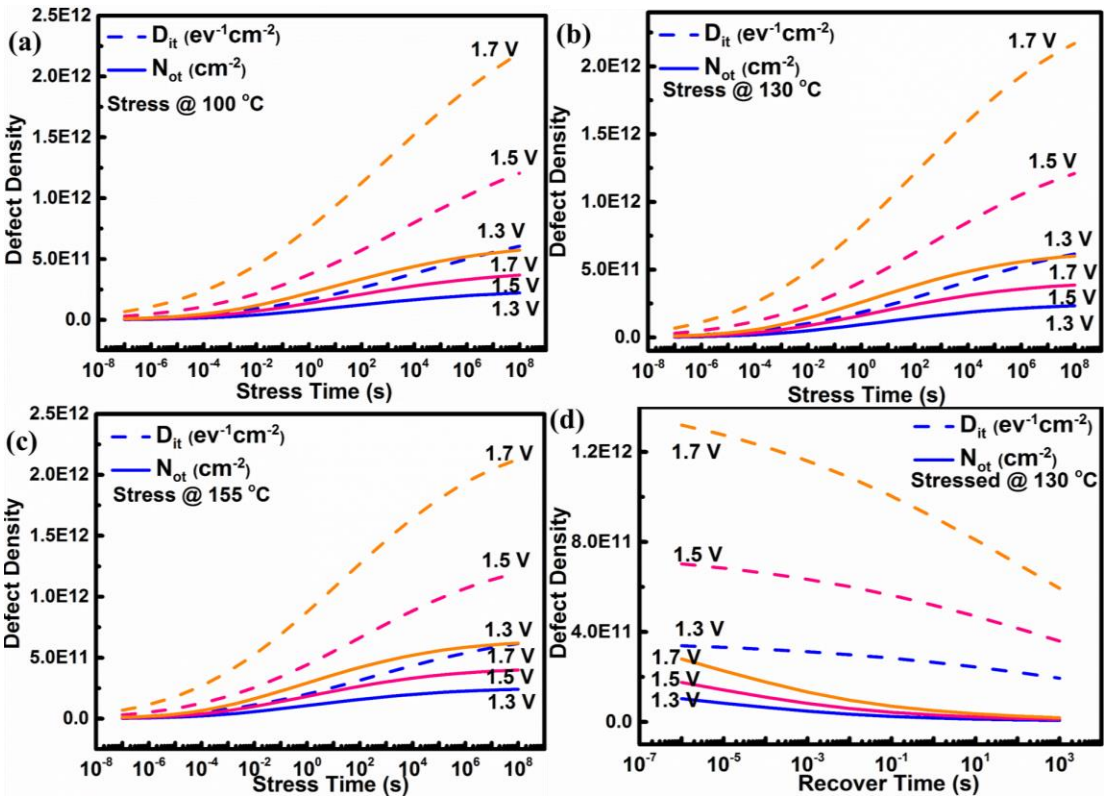


Fig. 2.3.3. Extracted $N_{ot}$ and $D_{it}$ with (a) stress at 100 °C under different bias, (b) stress at 130 °C under different bias, (c) stress at 155 °C under different bias, and then (d) recovery after stress at 130 °C under different bias.

### 2.3.5. Defect Maps

The defect distributions in equations 2.4 and 2.5 obtained from optimized fitting to BTI data (Fig. 2.3.2) can be visualized with capture/emission energy (CEE) maps [112]. Since energy distributions are bias and thermally dependent, CEE maps are unique to stress voltage and temperature [112]. CEE map can be converted from energy domain into time domain through the Arrhenius law ($\tau=\tau_0\exp(E/k_BT)$) to produce capture/emission time (CET) maps [112]. In Figs. 2.3.4 (a) and (b) CEE and CET maps extracted from the BTI stress and recovery data in Fig. 2.3.2 [128] are shown as contour plots. Here the transistor

is stressed under -1.3V gate bias at 130 °C. The recovery condition is 0 V gate bias at room temperature. Both maps have been normalized by $\log_{10}(g/g_{max})$ and show two peaks in the contour. In the CET map (Fig. 2.3.4(b)) the two peaks are particularly distinct: one shorter peak at a capture time of ~10s and emission time of ~$10^{-3}$s and a second higher peak at a capture time of ~$10^{3}$s and emission time of ~$10^{7}$ s.

The relatively low emission time for the distribution around the shorter peak corresponds to the recoverable component (*R*) in the BTI signature. The distribution around the higher peak has a much larger emission time and is thus said to correspond to the permanent component (*P*) of BTI. The *R* component is typically associated with $N_{ot}$ due to the relatively fast emission time of trapped carriers from oxide vacancy defects near the interface [129]. The permanent component is associated with $D_{it}$. Interface traps anneal much more slowly [130] since these defects require hydrogen to re-passivated dangling bonds.

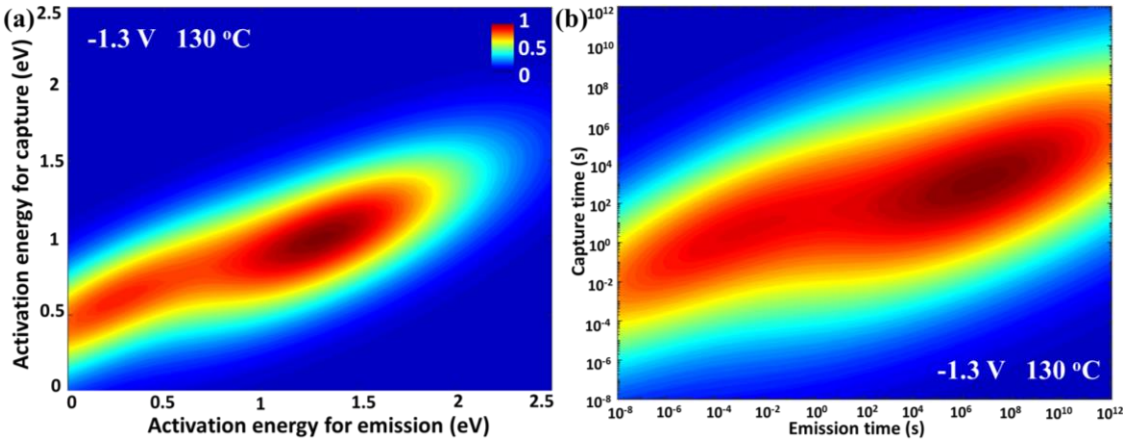


Fig. 2.3.4. Logarithmic normalized CEE and CET maps extracted from results in [128] with device stressed at -1.3 V 130 °C and recovered at 0 V room temperature.

Once the $N_{ot}$ and $D_{it}$ functions, equations 2.4 and 2.5, are parameterized for the technology, CEE and CET maps can be generated over of range of stress conditions. For example, Fig. 2.3.5 plots the contour for a gate bias stress of -0.45 V and at 27 $^{o}$C and recover at 0 V at 27 $^{o}$C. Under these conditions the defect distribution model reveals more distinct peaks compared to Fig. 2.3.4 and predicts a higher concentration of the *R* component relative to the *P* component.

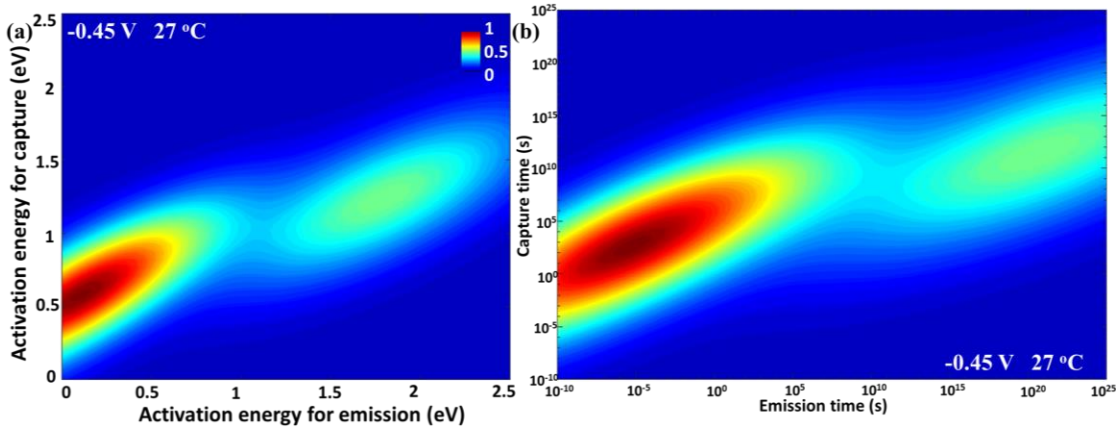


Fig. 2.3.5. Logarithmic normalized CEE and CET maps with device at -0.45 V bias at 27 $^{o}$C.

When comparing the CET maps in Fig. 2.3.4 (b) and Fig. 2.3.5 (b) it is interesting to observe that under the reduced bias and temperature stress conditions (-0.45 V and at 27$^{o}$C for stress and 0 V 27 $^{o}$C for recover) the capture time for the peak *R* ($N_{ot}$) has not changed significantly while the emission time for peak *R* has dropped a two order of magnitude, i.e., to ~$10^{-5}$ s. By contrast, under the reduced stress both the capture and emission time for peak *P* ($D_{it}$) has increased many orders of magnitude, to ~$10^{11}$ s for capture and ~$10^{20}$ s for emission. Such a long "capture/emission" time would indicate that under lower stress

interface traps are very hard to create and essentially impossible to anneal. This agrees previous work published by Fleetwood *et al.* in [130].

At this point we have parameterized a (N)BTI-enabled p-channel MOSFET model. However in order to simulate BTI effects in CMOS circuits a (P)BTI-enabled n-channel model must be constructed. Due to lack of PBTI data on n-channel MOSFETs on a similar technology, we have made the following assumptions based on previous studies. As reported in [131-133] at the 28 nm and 20 nm technology nodes (P)BTI effects on n-channel MOSFETs are comparable with (N)BTI effects in p-channel MOSFETs. Thus we have assumed that all defect distribution parameters for both device types are the same except for the maximum values of $N_{ot}$ and $D_{it}$, i.e., $A_{ot}$ and $A_{it}$ in equations 2.4 and 2.5. Based on results reported in [133] both $A_{ot}$ and $A_{it}$ induced by (P)BTI are assumed to be 40% of the densities induced by (N)BTI under same stress bias and temperature.

## 2.4. Circuit BTI Aging Effects Predictions

Analyses of aging effects using the BTI-enhanced model are performed through simulations on digital circuits constructed with standard n- and p-channel transistors provided in a 28 nm bulk CMOS technology design kit. These circuits include a single stage inverter (INV), 2-input NAND (NAND2), 2-input NOR (NOR2), and a 3-stage inverter (INV3). Stress conditions are chosen to be same as the supply voltages ($V_{dd}$) of 1.0 V, 1.2 V and 1.4 V. The stress temperatures simulated are 25ºC, 55ºC, and 85 ºC. Both DC and AC stress conditions are modeled. Only BTI effects are considered and other aging effects (e.g., HCI, SILC, and TDDB) have been ignored.

2.4.1. BTI aging caused by DC stress

For DC stress, the stress bias is not removed and therefore no recovery is modeled. The simulation time ranges from $10^{-7}$ s to $10^{8}$ s. In Figs. 2.4.1, the time, temperature and stress voltage dependent propagation delay degradation rate (PDDR) of a single stage inverter is plotted in the 3D surface plots. The PDDR is calculated as the change rate of the maximum propagation delay, i.e. $(t_{phl}+t_{plh})/(t_{phl0}+t_{plh0})$ where in $(t_{phl0}+t_{plh0})$ is the propagation delay before stress. In Fig. 2.4.1, the results after $10^{3}$ s (~20 mins) and $10^{8}$ s (~3 years) stress are presented where only p-FET (NBTI) or only n-FET (PBTI) was stressed. In Figs. 2.4.1 (a) and (c), both NBTI and PBTI induced degradations after $10^{3}$ s stress have observed monotonic increases with greater voltage and temperature. However, in Figs. 2.4.1 (b) and (d), the degradation rate change after $10^{8}$ s stress have observed differences between NBTI and PBTI. For NBTI (Fig. 2.4.1 (b)), the degradation profile has a similar shape to the profile after $10^{3}$ s stress which increases monotonically with higher voltage and temperature. But in Fig. 2.4.1 (d) for PBTI, a flat plateau has been observed in the higher voltage and higher temperature stress region indicating a saturation trend. According to the simulation results, the degraded PDDR by PBTI effect is only 10% of the NBTI degradations no matter how long the stress are performed. In addition, the PDDR profile after $10^{7}$ s stress (not shown here) appears almost similar with the profile after $10^{8}$ s stress for NBTI effects and PBTI effects, respectively. Such similarity could indicate the BTI degradations considered in this work would eventually saturate with time. However, the saturation time for PBTI effects would be much earlier than NBTI effects. This is in accordance with the most studies that conclude NBTI causes more damage to circuits than PBTI.

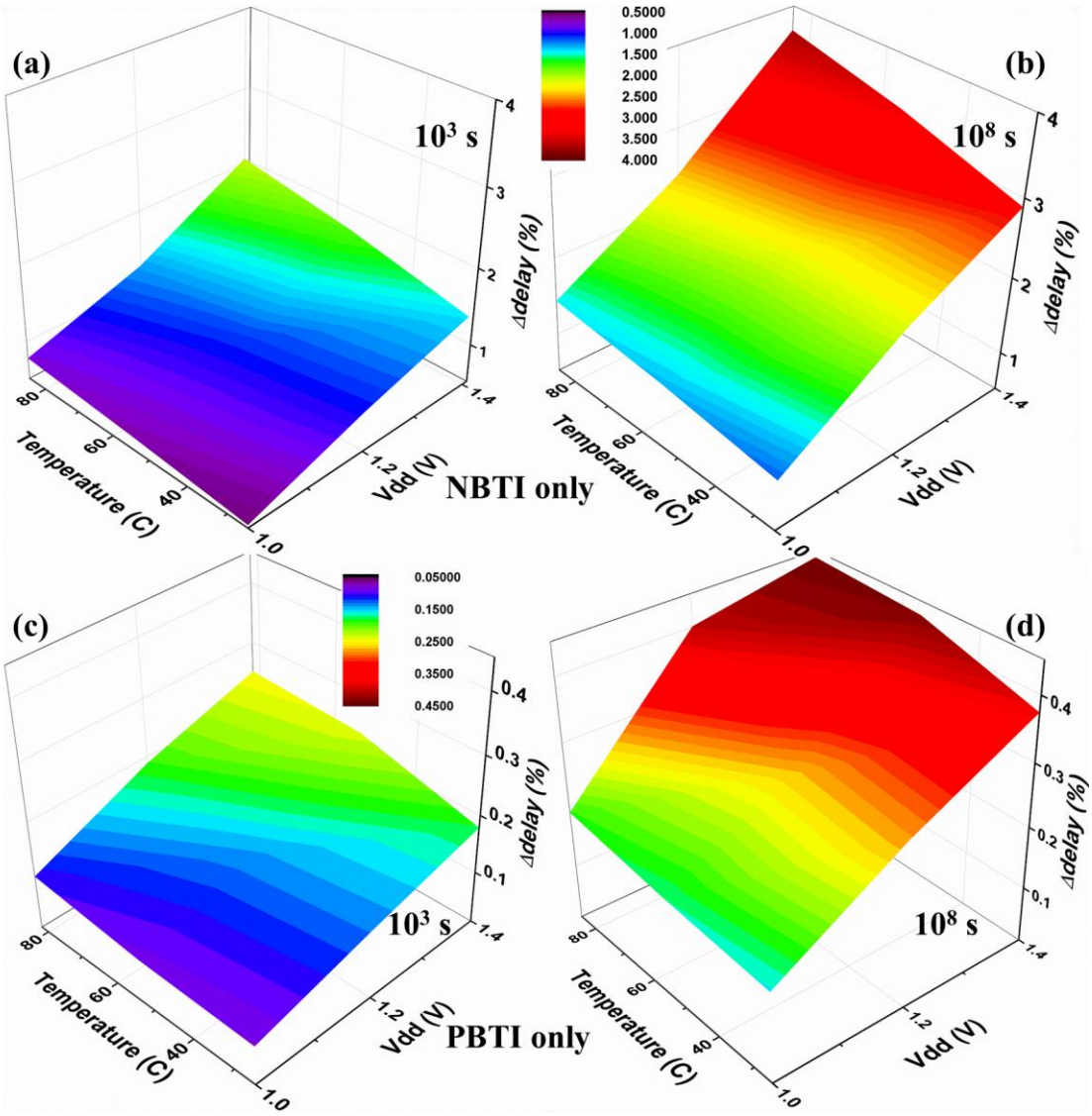


Fig. 2.4.1. Propagation delay degradation of the single stage inverter with different DC stress voltages and temperatures with solely stress on PMOS (NBTI only) for (a) $10^3$ s and (b) $10^8$ s and solely stress on NMOS (PBTI only) for (c) $10^3$ s and (d) $10^8$ s. (a) and (b) shares a same scale bar shown in (b), while (c) and (d) shares a same scale bar shown in (d).

Since for many technologies PBTI effects are much smaller that NBTI effects, NBTI alone have been simulated on the NAND2 and NOR2 circuits. The simulated PDDR results caused by degradation after a $10^8$ s (3 year) DC stress with different $V_{dd}$ at 55 $^o$C and 85 $^o$C are compared with the INV results simulated before and shown in Table 2.1. Comparing the results shown in Table 2.1 we can see that NBTI degradation are more sensitive to stress voltages than temperature. At first, according to the results, more than 3X degradations have been observed when $V_{dd}$ is increased from 1.0V to 1.4V for all the three circuits. But the temperature increase from 55 $^o$C to 85 $^o$C caused less than a 20% increase in PDDR. When $V_{dd}$ is not larger than 1.2 V, the NAND2 has shown least NBTI degradations comparing with INV and NOR2. Such results indicate that a pull-up network with multiple parallel p-FETs would potentially mitigate the NBTI degradations. However, when $V_{dd}$ is beyond the nominal value 1.2 V, the degradations in NAND2 becomes even worse than those in INV. These results also suggest that $V_{dd}$ fluctuations in real life applications would be the dominant source accelerated degradation and which should demand more attention during circuit design. It should also be noted that the NOR2 configuration shows the worst cased degradation in PDDR for all bias conditions. The greater susceptibility of the NOR2 gate may be due to the higher NBTI vulnerability of series p-channel pull-up network.

Table 2.1. The PDDP of the INV, NAND2 and NOR2 with $10^8$ s DC stress under different temperatures.

| **Δdelay(%) @ 55 °C** | | | |
|---|---|---|---|
| | 1.0 V $V_{dd}$ | 1.2 V $V_{dd}$ | 1.4 V $V_{dd}$ |
| INV | 1.32 | 2.15 | 3.35 |
| NAND2 | 1.03 | 2.13 | 3.74 |
| NOR2 | 1.44 | 2.54 | 4.32 |

| **Δdelay(%) @ 85 °C** | | | |
|---|---|---|---|
| | 1.0 V $V_{dd}$ | 1.2 V $V_{dd}$ | 1.4 V $V_{dd}$ |
| INV | 1.44 | 2.4 | 3.7 |
| NAND2 | 1.12 | 2.09 | 3.95 |
| NOR2 | 1.56 | 2.76 | 4.64 |

2.4.2 BTI aging caused by AC stress

Although useful in predicting maximum degradation, combinational logic typically does not operate for long durations under DC stress. In most cases the gate stimulus on the MOSFETs in digital gate will be time varying (AC) signals. During AC stress recovery can occur, thereby reversing deleterious effects of stress-induced defect buildup. Since the bias cycling during AC stress alternates the stress/recovery sequence experienced by the pull-up and pull-down networks two factors regarding AC stress signals are critical: 1) the frequency of the signal and 2) the duty cycle (the fraction of a period in which the signal is in high-level voltage). In Figs. 2.4.2 (a) and (b), the simulated results on the three-stage inverter chain (INV3) are plotted for AC stress conditions sampled after $10^3$ s and $10^8$ s. The simulation results obtained for different duty cycles and switching frequencies are shown in Figs. 2.4.2 (a) and (b), respectively. The results indicate that higher stress

frequency and lower duty cycle will lead to greater BTI-induced shift in the propagation delay. In order to understand this frequency behavior consider equation 2.8 above. A higher frequency means lower period, $T_s$, which reduces the limit of integration variable, $b_1$. This leads to higher occupancy rates for defects (i.e., less recovery), increasing threshold voltages and propagation delay. For AC stress at duty cycles less than 50%, the more sensitive p-MOSFETs will be exposed to more stress, leading to greater overall damage to the inverter chain delay. It should be noted that this duty cycle dependence is primarily an artifact of our modeled assumption that maximum defect buildup is greater in p-channel devices. For technologies that exhibit greater PBTI sensitivity in n-channel MOSFETs, the model would show an opposite trend in the duty cycle dependence.

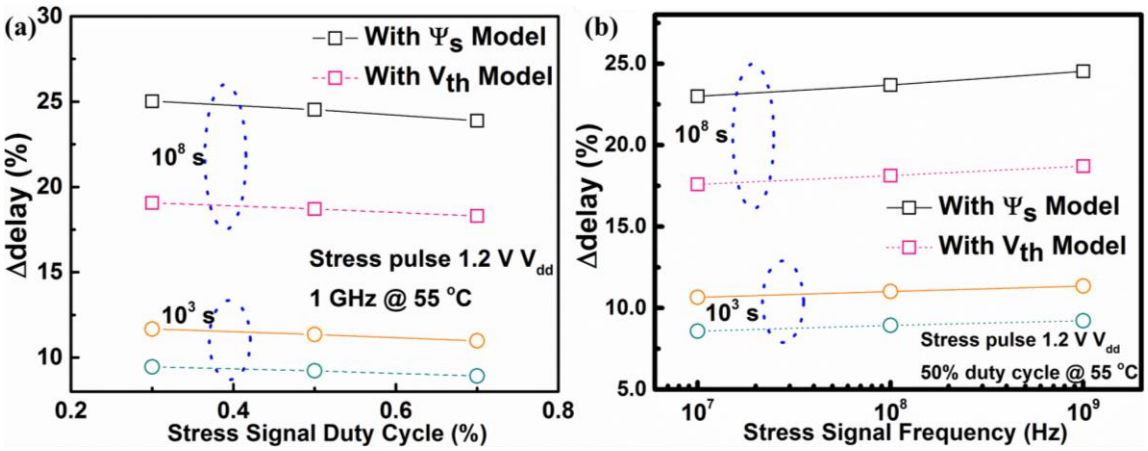


Fig. 2.4.2. The PDDP of the INV3 after $10^3$ s and $10^8$ s AC stress at 1.2 V and 55 °C with (a) different stress signal duty cycles and (b) with different stress signal frequencies. The solid line plot is for the $\psi_s$ model and the dash line is for the $V_{th}$ model.

Figs. 2.4.2 (a) and (b) also provide simulation results obtained by modeling aging effects with the traditional $V_{th}$ model for BTI. The plots show that compared to our $\psi_s$ model, the $V_{th}$ model predicts less degradation in delay. Although the reasons for discrepancy between the two models is still being studied, it is may be due to inherent physical inaccuracies in the traditional model. The $V_{th}$ model does not account for the effects of varying surface potential, therefore the bias dependence of the interface trap. Through its use of the equation 2.8, the $\psi_s$ model correctly captures the charge state of interface traps over the full range of gate biases applied to the MOSFETs in a circuit, i.e., between 0 V and $V_{dd}$. This is particularly important during AC operation where dynamic biasing continuously shifts surface potentials and thus the charge contributed to the system by the stress-induced oxide defects.

### 2.4.3. Discussions

The BTI-concerned transistor model proposed in this work has been demonstrated to simulate both DC and AC stress induced BTI effects. It has also shown to work across different technologies by incorporating the distributions of defect energies extracted from experiments. By including surface potential dependent defect potential, the model provides more accurate predictions than the conventional threshold voltage based modeling approach. Thus the circuit-level simulation results with the model can provide more insights into circuit reliability concerns for designers in specific technologies.

There are some drawbacks in this model that need to be improved in the future versions. First, the model does not working well with a 0 V bias. This is because in equations 2.4 and 2.5, the format of the pre-factor A is power law dependent with the stress voltage.

When stress voltage is 0 V, A will be 0. So with 0 V bias applied on the gate, the defect densities will be 0 and the defect maps are totally blank. Methods to deal with 0 V bias should be proposed. Secondly, we have restricted the mechanisms of BTI degradation to only charging process of the precursor oxide traps and the creation of the Si-H bonds related to interface traps during the stress. However, due to the specific focus on these processes, the predictions of the degradations caused by real electrical stress might miss other effect and therefore might not be fully accurate. Other stress-induced failures are discussed and analyzed more deeply in the next chapter.

## 2.5. Summary of the Chapter

In this chapter, the BTI aging effects and a proposed modeling method for HK/MG based MOSFETs have been discussed. A BTI-aware transistor model has been constructed by wrapping a commercial transistor model together with parameterized BTI degradation elements. The proposed BTI degradation elements can simulate stress/recovery induced threshold voltage shifts by solving the non-iterative surface potential equations and stochastic bivariate defect potential equations. With the defect energies extracted from the published results and incorporated into the defect potential equations, circuit level simulations with BTI aging effects are presented. The simulated results have been compared with a conventional threshold voltage based modeling method. The results of the comparison suggest that BTI-concerned transistor model is more physically accurate. Finally, the drawback in the new model are discussed. The models would be improved in the future concerning more other possible aging effects.

# CHAPTER 3

# HIGH-K DIELECTRIC BREAKDOWN

Last Chapter has discussed the MOSFET instabilities caused by the precursor bulk traps and Si-H related interface traps. These instabilities could be recovered once the devices are left in annealing environment after certain time. In this chapter, the aging and breakdown of the high-k dielectrics would be investigated. Usually, the level of leakage current through the dielectrics would be the one of the metrics that measures the dielectric breakdown. High quality thin films usually show very insulating characteristics before aging or breakdown whose current density could be as low as 1 $\mu A/cm^2$. The dielectric breakdown happens when some of the atomic bonds are broken and these bonds are usually oxygen related. Thus the bond breaking process would generate oxygen vacancies which would conduct electrons. Either constant voltage stress (CVS) or ramp voltage stress (RVS) could trigger the dielectric breakdown leaving oxide conductive. But with different current compliances, we might get either soft breakdown (SBD) or hard breakdown (HBD). After HBD, the resistance across the oxide would be very low and it is not possible to recover it back to insulator. However, after SBD, the resistance across the oxide actually could be tunable which are showing a resistance switching phenomenon. Such phenomenon is also called resistive switching effects.

## 3.1. Constant Voltage Stress Study

### 3.1.1. Time Frame of the Constant Voltage Stress

In Fig. 3.1.1, the stages of dielectric degradations under stress have been illustrated by a gate current-Time plot of a MOS capacitor under stress [131]. Here the stress denotes the

constant voltage stress (CVS) in which stress voltage does not change with time. Once the stress is on, the charge trapping and trap generation processes are both starting. The charge trapping effects in different types of MOSFETs could be different due to the different spatial distributions of traps created by fabrication process. Such effects actually should be negligible comparing with the whole stress period because the total density of the fabrication-induced traps is low. Usually, this stage is also related to BTI effects which has already been extensively discussed. Once the stress bias is increased, the time for such stage would be exponentially decreased. In the later on stress stage, new traps would be generated which typically has a power law dependent in time [132]. In a dielectric thinner than 50 Å, these newly generated traps would increase the gate current (SILC) gradually by offering available sites for carriers to tunnel through which is also called trap assisted tunneling (TAT) process [132]. While in thicker films, carriers might be possibly trapped in those generated defects permanently. If the energy level of these traps are shallow, the trapped carriers might tunnel into the electrode under high electric fields, but if the energy level is deep enough, they will form charged centers affect the function of the transistor and increase the leakage through the gate dielectrics [132]. However, SILC would be only prominent under low bias where the carriers are under direct tunneling (DT) processes. Once under high bias voltage, the SILC would be shaded by the Fowler Nordheim Tunneling (F-NT) current [132].

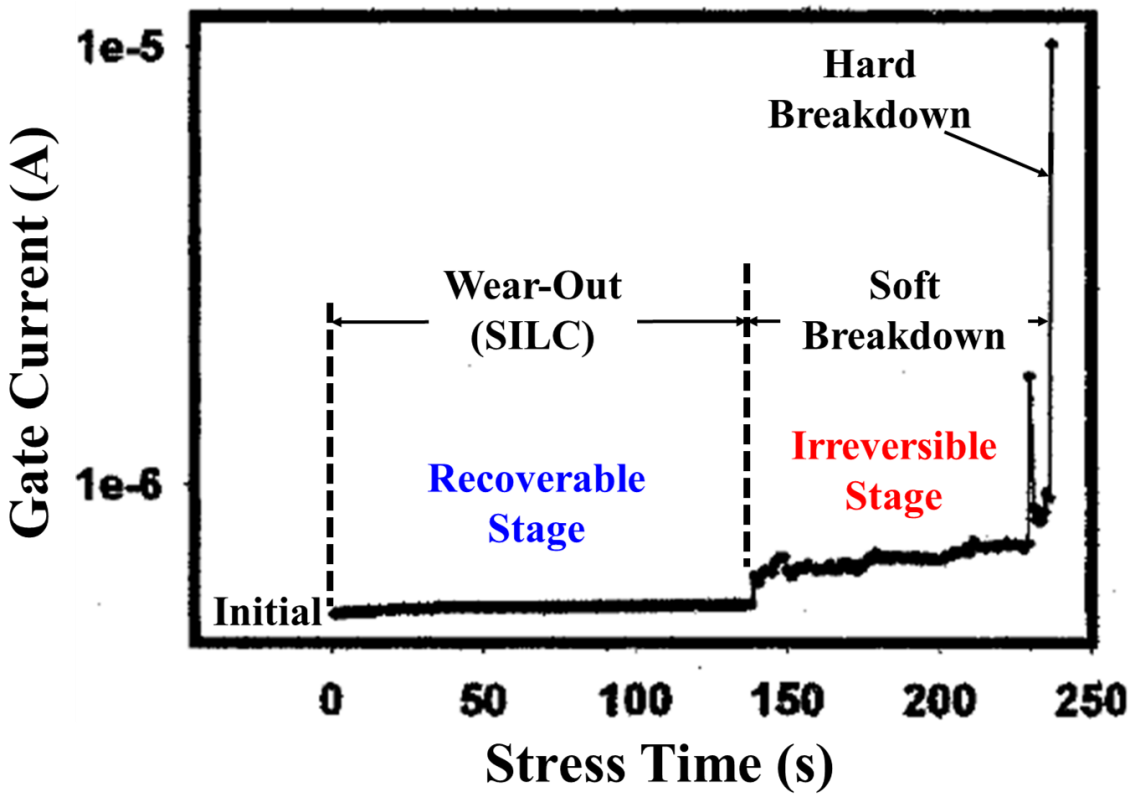


Fig. 3.1.1. Stages of dielectric degradations under stress. © 2002 IEEE. Reprinted with permission from Ref. [131].

In thin dielectrics below 50 Å, stress at high bias voltage will lead to soft breakdown (SBD) of the device during which a conductive path would be formed to connect the anode and the cathode [133]. Such conductive path is composed of large amount of stress-generated traps whose density will exceed the critical value to form an energy level allowing carriers to pass through. For the carrier transport post SBD in the dielectrics, several mechanisms have been proposed to describe the non-linear current conduction such as variable range hopping [134], quantum point contact (QPC) [135], percolation through a nonlinear conductor network [136], space charge limited current (SCLC) [137], and the electron co-tunneling through a coulomb blockade [138]. Usually SBD denotes the 1$^{st}$ time

breakdown of the gate dielectrics after which the devices might still be able to operate and are usually possible to get partially recovered. But from this stage on the damages are not fully restorable. Catastrophic hard breakdown (HBD) of the device would occur after the SBD or sometimes together with SBD but would rely on the electrical stress environment such as the current compliance or stress temperature [139]. After HBD, the current through the gate dielectrics would show an Ohmic I-V with a post breakdown resistance less than 10 kΩ. Therefore HBD might seem like dielectric burnout or thermal breakdown.

The term time-dependent dielectric breakdown (TDDB) occurred in the CMOS transistors are mostly related to the first time breakdown which could be either SBD or HBD. In the old time when supply voltage was beyond 3 V, the 1$^{st}$ time breakdown of the transistor might happen with HBD. But nowadays the supply voltage has drastically decreased to about 1 V or even lower, then the 1$^{st}$ time breakdown of the transistor might have rare possibility to be HBD. However, if the stress continues after SBD, then the conductive paths would grow stronger and finally lead to HBD or burnout. Consequently, when we study the breakdown process, both time and energy should be accounted in order to get a complete explanation of certain process.

### 3.1.2. Electron Transport and Energy

In the last section, the time stages of dielectric degradations under stress have been briefly discussed. Regarding the stress induced degradations or damages, not only stress time but also stress energy should be concerned in order to completely understand the degradation process. For devices under stress at certain temperature, the injected energy could be represented by the effective energies of the carriers injected into the whole system.

Thus the transport process and the energy of the carrier under certain transport mechanism should be concerned. In this work, since HKMG stacks are under investigation, the majority carrier transport through the stack could only be electron. Thus most of the discussions below would only consider electron transport process. Usually, with the electrode applied with positive voltage would be called anode where electrons are collected and the other electrode called cathode where electrons are sent out. In the MOSFET stack or simply the MOS-cap stack under stress, starting from cathode there are potentially two types of transport processes for electrons: 1) move from cathode to anode directly; 2) move from cathode to anode with going through some of the defects inside the dielectrics. The first type transport process is called electrode-limited conduction mechanism while the second type is the bulk-limited transport mechanism [140].

#### 3.1.2.1. Electron Energy during Transport in Dielectrics

Before discussing all the transport processes, the energy for each process would be defined here. During the elastic process, the electron does not have kinetic or total energy change until after it enters anode. During ballistic process, there are no collisions (no change in total energy) until after the electron reaches the anode dielectric interface. During inelastic process, both kinetic and total energy of the electron changes. By correlating the transport process with electron energies inside the dielectrics, it would be helpful to understand various energy injection induced degradation processes.

3.1.2.2. Electrode-Limited Transport

The electrode-limited transport mechanisms depend mainly on the electrical properties at the electrode-dielectric contact [140], including direct tunneling (DT), Fowler-Nordheim tunneling (FNT) and Schottky emission or thermionic emission depending on the bias level applied across the dielectrics. This type of conduction mechanism mainly depends on the barrier height at the electrode-dielectric interface and the effective mass of the conduction carriers in the dielectrics [140]. The band diagrams of the three major electrode-limited transport mechanisms are illustrated in Figs. 3.1.2.

Direct tunneling (DT) usually occurs when the bias voltage is lower than the Si-$SiO_2$ (part of the high-k stack) barrier height so that electron will see a trapezoidal barrier as indicated by the black dashed box shown in Fig. 3.1.2 (a). The gate current in such case would be entirely from the direct tunnels of electrons through the entire gate dielectrics. At the end of the DT process when electrons just arrives the anode-dielectric interface, its energy distribution is nearly mono-energetic. Thus DT is an elastic process during which electron energy alone would not be sufficient to breaking bonds and creating new defects inside the dielectrics. But this does not exclude the damages at the anode-dielectric and cathode-dielectric interfaces.

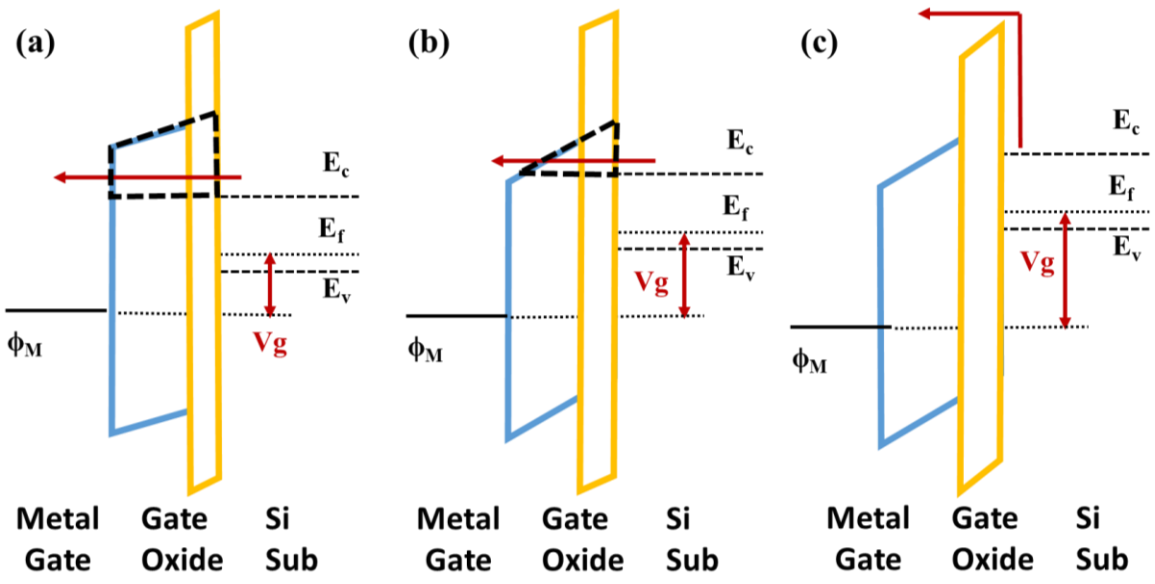


Fig. 3.1.2. Schematic energy band diagram of the three electrode-limited transport mechanisms in metal-oxide-semiconductor structure with (a) direct tunneling, (b) F-N tunneling, and (c) Schottky emission.

Approximately the electron kinetic energy after tunneling is about q|Vg| which has a maximum value of q|$\Phi_M$-$E_F$|. However, DT is a dielectric thickness dependent process at given filed, and when dielectric thickness is below 40 Å, DT current would be dominant in the total gate current [141]. With the Wentzel–Kramers–Brillouin (WKB) approximation, direct tunneling current would be expressed as [142]:

$$J_{DT} = AE_{ox}^2 / \left[1 - \left(\frac{\phi_{BA}+qV_{ox}}{\phi_B}\right)^{\frac{1}{2}}\right]^2 \exp[(B/E_{ox}\phi_{BA}^{\frac{3}{2}})(\phi_{BA}^{3/2} - (\phi_B - qV_{ox})^{3/2})], \quad (3.1)$$

in which $\Phi_{BA}$ is the barrier height between anode and dielectric, $E_{ox}$ and $V_{ox}$ are the electric filed and voltage across the oxide, and A and B are separately expressed as [141]:

$$A = \frac{q^3}{16\pi^2\hbar\phi_{BA}}, \quad (3.2)$$

$$B = \frac{3}{4}\frac{(2m_{ox})^{1/2}}{q\hbar}\phi_{BA}^{3/2}, \quad (3.3)$$

where $m_{ox}$ is the effective mass of the electron in the dielectrics, ħ is the reduced Planck constant, and q is the elementary charge.

Once the electrical bias is greater than the anode/dielectric interface barrier height $\Phi_{BA}$, FNT will start to be dominant during which electrons would see a triangle barrier as indicated by the black box shown in Fig. 3.1.2 (b). With such shape of barrier, electrons will tunnel into the conduction band of the dielectrics and only see partial thickness of the dielectrics. After entering into/beyond the conduction band of the dielectrics electrons would continue moving until reaches anode. Depending on whether scattered by oxide phonons on the conduction band of the dielectrics before entering anode, FN tunneling would further be subdivided into ballistic FN tunneling (BFN) process and steady-state FN tunneling (ssFN) process.

BFN tunneling process is an inelastic process because electrons gain kinetic energies in the oxide conduction band when they traverse the dielectrics [142]. When BFN occurs, there would be interference between incident and reflected electron waves propagation in the oxide conduction band which could lead to the observation of the oscillations in the current-voltage (I-V) curves [143]. Since FNT is a thickness dependent transport process, such oscillating effect could only be clearly observed when oxide thickness is below 70 Å [143].

Steady-state FN tunneling is also an inelastic process but electrons will lose energies to oxide phonons by scattering during transport. When electric filed is below 5 MV/cm, the primary scattering mechanism is the longitudinal optical (LO) phonon emission with an energy loss of 0.15 eV [144]. While for higher electric field, electrons usually would be

stabilized by acoustic phonon scattering with average energy loss below 6 eV [145]. The ssFN tunneling usually occurs when the oxide thickness is greater than tunneling distance under certain electric field plus the heat up distance for the carrier kinetic energy which would be at least 66 Å [146].

With the parameters specified above, the FN tunneling current for both cases could be similarly expressed as [140]:

$$J_{FN} = \frac{q^3 E_{ox}^2}{8\pi h q \phi_{BA}} \exp[\frac{-8\pi(2qm_{ox})^{1/2}}{3hE_{ox}} \phi_{BA}^{3/2}]. \quad (3.4)$$

As have been discussed so far, FN tunneling process generally could be divided into two steps: 1) tunneling through the dielectric within the band gap into the conduction band of the dielectric, 2) transport on the conduction band of the dielectric until reach the anode. Step 1) is a shortened DT process during which the electrons are not losing energy. Step 2) is an accelerating process for electrons during which electrons would gain kinetic energies before reach anode. Obviously, step 2) is very critical because extra energies would be gained by electrons during transport. Thus it is possible that by the end of this step damages of dielectrics or anode would occur.

If the electrical bias is further increased to exceed the cathode-dielectric interface barrier $\Phi_{BC}$, electrons would be able to gain enough energy provided by thermal activation and overcome such barrier to go to the conduction band of the dielectric directly. Such transport process would be often observed at relatively high temperature under large bias. The current expression for Schottky emission is [140]

$$J_{sc} = A^* T^2 \exp[\frac{-q(\phi_{BC} - \sqrt{qE_{ox}/4\pi\varepsilon_r\varepsilon_0})}{kT}], \quad (3.5)$$

where T is the ambient temperature, k is the Boltzmann's constant, $\varepsilon_0$ is the permittivity in vacuum, $\varepsilon_r$ is the optical dielectric constant (i.e., the dynamic dielectric constant), and $A^*$ is the effective Richard constant which could be expressed as [140]

$$A^* = \frac{4\pi q k^2 m_{ox}}{h^3} = \frac{120 m_{ox}}{m_0}, \tag{3.6}$$

in which $m_0$ is the free electron mass and the other notations are the same defined before.

It is worth mentioning here that if the electron transit time from the interface to the barrier maximum position is shorter than the dielectric relaxation time, the dielectric would not have enough time to get polarized leaving only a few polarization mechanisms contributing to the total polarization [147]. Consequently, the optical dielectric constant is smaller than the static dielectric constant. However, in our case of MOS cap or MOSFET stressing, the thermionic emission is a rare situation unless devices are under great voltage stress at high temperature. With so much external energies injected, the later electron on transport process on the conduction band in the dielectric would be ballistic.

3.1.2.2. Bulk-Limited Transport

The bulk-limited transport mechanisms mainly depend on the electrical properties of the dielectric itself. The most critical parameter in such type of transport mechanism is the defect energy level in the dielectric [140]. Major mechanisms of this type of transport would be considered in this work include Pool-Frenkel (P-F) emission, hopping conduction, and Ohmic conduction. There are also some other transport processes belonging to the same type of transport mechanism such as space-charge-limited conduction (SCLC), ionic conduction, and grain-boundary-limited conduction (GBLC) but would not be discussed in

this work. The energy band diagram of the three transport mechanism has been illustrated in Figs. 3.1.3.

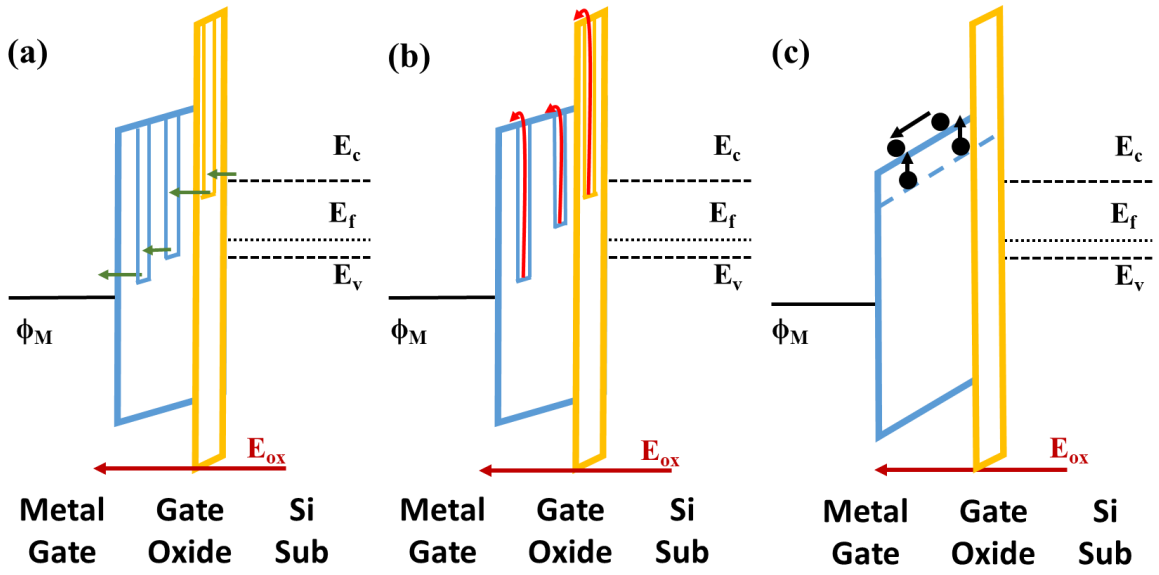


Fig. 3.1.3. Schematic energy band diagram of the three bulk-limited transport mechanisms in metal-oxide-semiconductor structure with (a) hopping conduction, (b) P-F emission, and (c) Ohmic conduction.

For hopping and P-F emission, they depends on the transport through the discrete energy levels created by the defect inside the dielectrics. During electron transport along the electric field, these defects would capture/emit electrons. Depending how much energy could be obtained during emission, the electrons could either tunnel to the next trap site/anode or be emitted into the conduction band of the dielectric. All these processes related to the traps or defects in the bulk of dielectrics could be generally called trap-assisted-tunneling (TAT) effects where traps act as a helper for electrons to transit from

cathode to anode. But detailed transport processes of TAT might be very complicated with at least one trap involving in the transport process.

For Ohmic conduction, the shallow defect levels either intrinsic or extrinsic close to the conduction band might have a density exceeding the critical values forming an energy levels inside the forbidden band and acting as dopants inside the "wide bandgap insulator". These shallow donors would donate electrons once get excited under the field and make the conduction inside the dielectric Ohmic.

Hopping conduction is much likely a reduced version of direct tunneling process between defect sites or between electrode and defect sites. The expression of hopping conduction is [148]

$$J = qan\upsilon \exp[\frac{qaE_{ox}}{kT} - \frac{E_a}{kT}], \qquad (3.7)$$

where a is the hopping distance, n is the electron concentration in the conduction band of the dielectric, ν is the thermal vibration frequency of electrons at trap sites, and $E_a$ is activation energy; the other terms are as defined before. In hopping conduction, electrons does not gain high enough energy to get emitted into the conduction band of the dielectric, so that they transit through defect sites using tunneling mechanism. Since the hopping distances between different trapping sites are not long enough and there contains the capture processes of electrons by trapping sites, the hopping conduction should be an inelastic process where electrons are transferring the energies it gained to traps. Thus electrons under such transport mechanism do not contain any damaging ability and this process usually takes place when electrical bias is low.

P-F emission is very similar to the thermionic emission where electrons would gain enough energy to overcome the potential barrier and enter the conduction band of the

dielectric. But thermionic emission is from electrode and P-F emission is from trap level inside the dielectric so it is sometimes called internal thermionic emission. The current density due to the P-F emission is expressed as [149]

$$J = q\mu N_c E_{ox} \exp\left[\frac{-q(\phi_T - \sqrt{qE_{ox}/\pi\varepsilon_r\varepsilon_0})}{kT}\right], \quad (3.8)$$

where μ is the electronic drift mobility, $N_C$ is the density of states in conduction band, $\Phi_T$ is the defect energy level; the other terms are as defined before. Similar to thermionic emission, after entering the conduction band in the dielectric, the electron transport should also be ballistic process.

Ohmic conduction is caused by the movement of the mobile electrons in the conduction band and holes in the valence band. Usually this conduction mechanism gives a linear I-V relationship. The electrons excited by the thermal energy from valence band or shallow donors could move freely in the conduction band in the dielectric. The current density of this type of conduction mechanism can be expressed as [150]

$$J = \sigma E_{ox} = nq\mu E_{ox}, \quad n = N_c \exp\left[\frac{-(E_c - E_F)}{kT}\right], \quad (3.9)$$

where σ is the electrical conductivity, n is the number of electrons in the conduction band, μ is electron mobility and the other terms are as defined before.

By far, most common transport mechanisms have been discussed above. But during real electrical stress, since electrons might have a distribution of energy while transmitting through the dielectrics, several transport mechanisms might need to be considered together. Though one type of mechanisms would be dominating the whole electron transport process at a time, the wide distribution of electron energies make it very complicated to fit the I-V curves with only one type of transport equation.

### 3.1.3. Defects in Gate Stack

After realizing the transport mechanism and electron energies, potential types of damages for the dielectrics could be determined. These damaging events might include charge trapping, trap generation, impact ionization and even species releasing. According to level of electron energies in dielectrics, different events would happen. Charge trapping occurs when electron has quite low energy and get trapped by the existing defects during transport and this has already been discussed. Trap generation process could be further divide into interface trap generation and bulk trap generation. The former one tends to happen when the electron energy is not high enough and the latter one happens when hot electrons are generated by F-N tunneling injection or substrate injection. Impact ionization happens when kinetic energy of the electron is too high so that it knocks the electrons out from the internal orbit generating electron-hole pairs and these energetic electron-hole pairs continue creating more pairs by multiplication. Along with impact ionization process, there could be trapping of both types of charges occurring and further changing the band diagram of the original system. The species releasing event takes place when all the bonds of certain species are broken by the high energy breakdown leaving the species, or sometimes ions diffusing or drifting inside the dielectrics. The typical example could be found in the hydrogen releasing after substrate hot electron injection or substrate hot hole injection.

Thus most damaging process will happen with either defect generations or defect reaction. To understand the stress induced degradations and breakdown in dielectrics, a better understanding of the potential defects in the gate stack is necessary. Last chapter has discussed as-grown bulk traps would trap charge and cause instability. These as-grown traps will certainly involve in the stress induced degradations by various ways such as

acting as helping centers for electrons to tunnel through (trap-assisted tunneling), or trapping positive charges and pulling down the barrier for electrons, or even trapping electrons and changing its energy levels. With the improvement of fabrication process, the density of as-grown defects are lowered to a negligible level as indicated in the last chapter. Thus the newly generated defects by stress would be more effectively involving the damaging of the dielectrics.

3.1.3.1. Stress Induced Defects in poly-Si/$SiO_2$ (SiON)/Si Gate Stack

In the early CMOS technology when poly-Si/$SiO_2$/Si gate stack was in use, the stress induced degradations and breakdowns are often due to the defects generated either at the $SiO_2$/Si interface (interface traps) or inside the $SiO_2$ (bulk traps). The defects at interface could be either hydrogen related interface states or Si-O bonds breaking at interface. The former type needs lower energy to occur while the latter type requires either long time or higher energy to trigger.

The bulk traps in $SiO_2$ generated by stress would be neutral centers which might capture either electron or hole [151,152]. Both electron and hole traps could be generated by either electron or hole injections. For stress generated electron traps, there could be three different types. One type of the generated electron traps is called high field trap which stay occupied at high electric field. The other type is called low field trap which has almost zero occupancy at high electric field [153]. This phenomenon probably is due to the different generation kinetics and shallower energy level of the low field trap. The third type electron traps would be related to water-related centers so it is related to the processing methods. All these electron traps were found spatially uniformly distributed in the oxide with

centroid at middle of the oxide [154] whose build-up would further trigger dielectric breakdown. For the stress generated hole traps, there could be two different types. One is known as the anti-neutralization positive charge (ANPC) [155] and the other one is referred to anomalous positive charge (APC) or cyclical positive charge (CPC) [155, 156]. The ANPC center could be recharged without hole injection but to discharge it high field F-N injection of electrons are needed [155]. Possible explanation might be that the trap energy level moves beyond the conduction band once get charged which renders it improbable to discharge the charged traps by filling electrons from the conduction band directly [155]. While the other type of defect APC or CPC has a fixed energy level inside the oxide band gap very close to the conduction band of silicon. It could exchange charges with silicon and could be charged and discharged without hole injection [156]. Quite different from the electron traps, the hole traps prefer to cluster close to the oxide interface [156].

#### 3.1.3.2. Intrinsic Breakdown in $SiO_2$

The breakdown of $SiO_2$ could be categorized into many different mechanisms depending on the applied electrical bias across the $SiO_2$ and the thickness of the $SiO_2$. In a case of oxide thickness over 200 Å under an electric field higher than 7 MV/cm when FN injection is happening, the main breakdown mechanism has been attributed to the impact ionization and hole trapping in the oxide [155]. Though the threshold energy for impact ionization energy is about 9 eV (band gap of SiO2) which is way above the average kinetic energy of electrons in $SiO_2$, the tail of energy distribution of electron could be above 15 eV [147]. Once impact ionization was triggered, large density of electron-hole pairs would be created. The electrons would be swept away by the electric field while holes would be

trapped by the as-grown traps discussed in the last section forming positive charge centers and pulling conduction band down which would further increase the electron transport length beyond conduction band (which equals to increase the electron kinetic energy). This process has formed a positive feedback to boost itself until a conductive path is formed connecting the anode and cathode [133]. However in the case of thin oxide under a low field where FN injection has been off, positive charging effects were still observed to trigger the oxide breakdown [157]. Aforementioned in the last section, electron direct tunneling in thin oxide is an elastic process which gains electron kinetic energies while entering the gate electrode. Such high kinetic energy could trigger anode impact ionization. Once the impact ionization happens in the Si (gate), a large amount of electron-hole pairs would be generated and holes swept into the oxide [157]. After the injection of holes into the oxide, same breakdown phenomenon could be observed. Depending on kinetic energy of the electrons, there could be majority ionization or minority ionization. Usually majority ionization has a higher rate than minority ionization while minority ionization has a lower trigger energy (about 3 eV) than majority ionization (about 6 eV) [158].

Recent works in ultra-thin oxide breakdown under low bias (lower than 3 V) has ruled out most of the breakdown mechanisms mentioned above [159]. The breakdown mechanism has been attributed to hydrogen release induced defect generations. Interface states generations have been regarded as the critical part for oxide breakdown in this case. Meanwhile, both bulk trap and interface trap is required for breakdown to occur.

3.1.3.3. Defects in $HfO_2$

In the high-k dielectrics based CMOS technologies, since $HfO_2$ will react with Si and bring very low quality interface layer oxides, most techniques are introducing another ultra-thin layer of high quality $SiO_2$ between them to maintain the gate performance. In recent years, there are works proposing the quality of $HfO_2/SiO_2$ interfacial layer is the critical portion of gate stack reliability [161] while many works are attributing the aging and breakdown problem to the trap generations in $SiO_2$ rather than in $HfO_2$ [165, 166]. Conversely, some other works are pointing the vulnerability of $HfO_2$ layer which gets firstly degraded first during stress [167, 168]. As the interfacial layer between $SiO_2$ and $HfO_2$ has complicated the study of the wearing out in the MOS stack, it is not simple to reach a conclusion which one dominates the degradation unless separate studies could be performed. Recently, some works on defects in $HfO_2$ alone have been done by either stressing metal/$HfO_2$/metal stack [160] or ab-initio calculations of defect states in $HfO_2$ [108] in order to offer clues of the stress impacts on $HfO_2$.

As shown in Fig. 3.3.1, potential defect energy levels based on ab-initio calculations have been presented here. The $HfO_2$ used in the CMOS technology usually has a dielectric constant between 16-20 and an amorphous crystalline structure (a-$HfO_2$). The amorphous structure is very close to monoclinic crystalline (m-$HfO_2$) which naturally contains two types of neutral oxygen vacancies, one is the $V_O3$ vacancy which connects with 3 Hf around it, and the other one is the $V_O4$ vacancy which connects with 4 Hf around it. Recent works indicate the natural defect density is more than $10^{20}$ $cm^{-3}$ by the extraction from the trap-assisted tunneling equations [162, 163]. These natural vacancies have an energy level about 1.1 eV-1.8 eV below the conduction band of $HfO_2$ [164]. The variation of the energy level

probably is due to the different results of $HfO_2$ band gap which brings different conduction band affinity. In addition, the charged vacancy tend to move its energy level deep into the $HfO_2$ band gap as indicated in Fig. 3.3.1. At the same time, if possible, there could also be neutral oxygen atom as interstitial or even $O_2$ molecule inside the $HfO_2$ band gap forming energy levels very close to $HfO_2$ valence band. These defect levels could act as hole trap centers during stress. Unlike the vacancy based traps, once captured electrons the oxygen species based traps tend to get closer to $HfO_2$ valence band.

Previous work by Cédric et al. on electrical stress on metal/$HfO_2$/metal stack has provided some useful information for the as-grown and stress induced defects in high-k dielectrics [108]. The cyclic electric stress/relaxation experiments showed that absorption current only occurs at the first time stress and the relaxation currents are same for all relaxation cycles but the leakage current under stress was increasing when more stress cycles have been performed. The origin of such phenomenon has been attributed to the extinct nature of the as-grown defects which only affects the threshold voltage but not involve in the stress-induced leakage current (SILC). This behavior is similar to the ANPC discussed in the last section which would change its energy level above conduction band of the oxide once charged. Those ANPC-like centers might locate at the grain boundaries (GBs) in the $HfO_2$. Thus they concluded that SILC is only due to the stress generated oxygen vacancies via Fowler-Nordheim electron injection (F-N injection) from cathode. Thus the work function difference of the two metal electrodes is very critical. Moreover, the temperature dependent breakdown behavior also suggests that the conductive paths formed inside the oxide would frequently rupture/rebuild by the thermally agitated defects. Thus the recombination of the $V_O$ and the oxygen ions probably is the SILC recovery origin.

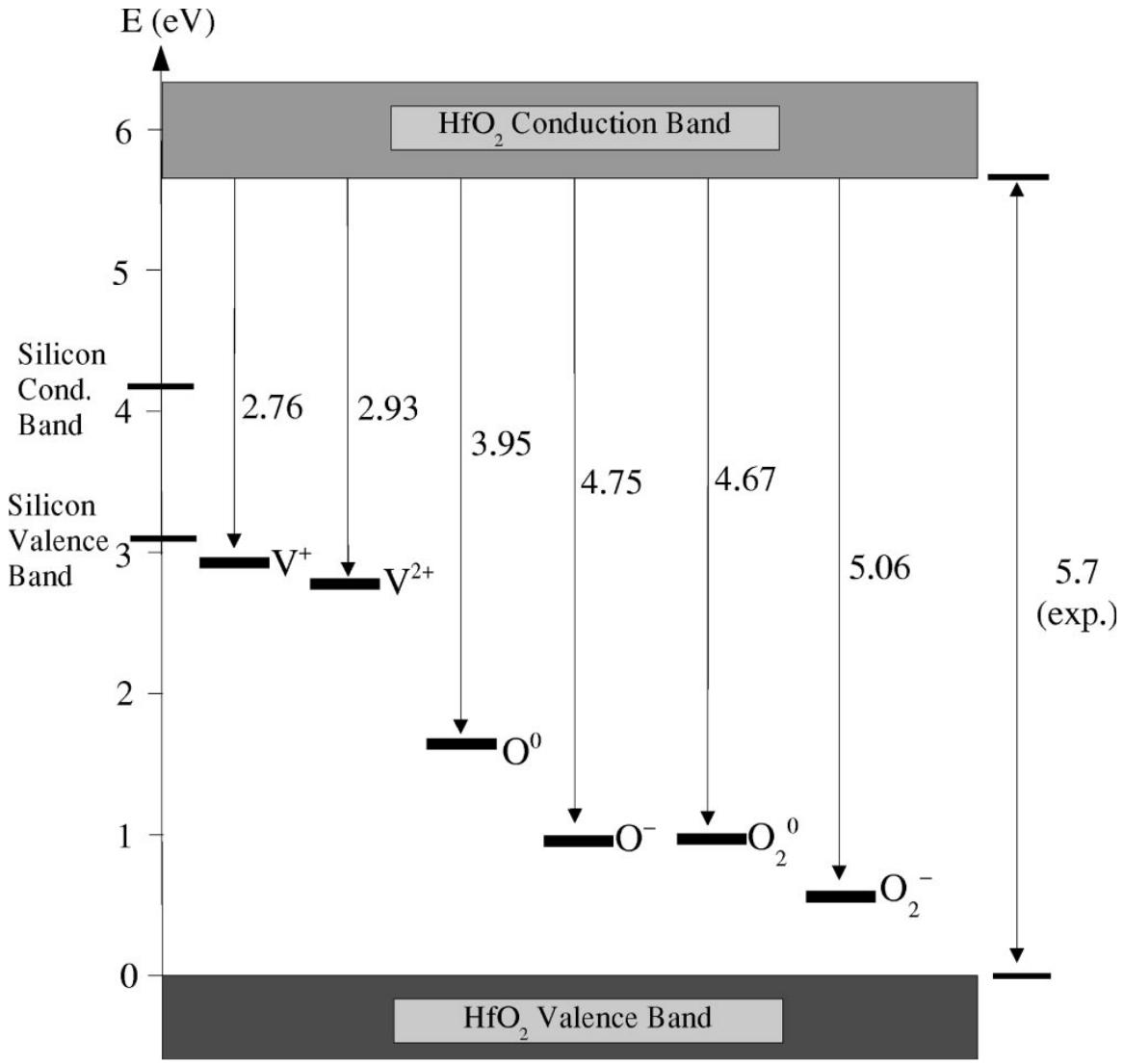


Fig. 3.1.4. Energy level diagram showing the electron affinities for various defects in monoclinic hafnia. All energies are in eV. Reprinted with permission from [108] Copyright (2002) by the American Physical Society.

Accordingly once the HKMG based MOSFET is under stress, the degradations would first happen at either high-k/metal interface or high-k/$SiO_2$ interface depending on the type of transistors used. For p-FET in which substrate is made of n-type silicon, the normal stress bias is negative (gate injection). When the negative bias is at the critical point where

F-N tunneling just starts and DT not turns off yet, the $HfO_2/SiO_2$ interface will first see the injection of electrons into the conduction band of $HfO_2$ from the gate. Since then oxygen vacancies and oxygen ions pairs would be generated at the $HfO_2/SiO_2$ interface. Oxygen ions would propagate into $SiO_2$ layer under the electric field where further damages take places. For n-FET which has a normally positive stress bias (substrate injection), the metal/$HfO_2$ interface will first see the injected electrons from substrate and suffer from damaging. With positive bias increasing, there could be impact ionization happening at interface and followed by the positive charge induced dielectric breakdown from $HfO_2$ all the way until reaching Si. Otherwise, the materials used in the gate stack would also affect the degradations. The materials characteristics such as semiconductor-metal work function differences, band gap of the high-k dielectrics used in the gate stack and the band offset between metal work function and oxide conduction band and so on will all determine the reliability of the transistor built upon.

### 3.1.4. Constant Voltage Stress on TiN/$HfO_2/SiO_2/n^+$-Si Gate Stack

By far, the damaging scenarios that could happen in the HKMG based gate stacks have been fully discussed. Potential damaging effects have also been forecasted. Stress experiments on metal/$HfO_2/SiO_2/n^+$-Si based MOS stack would be presented in this section with further discussions.

#### 3.1.4.1. Device Fabrication

Heavily doped n-type prime grade silicon wafer with a resistivity range of 0.001-0.005 Ω·cm was used for the MOS capacitor fabrication. After soaking in acetone and DI-water

in sequence for cleaning, silicon wafer was dried and baked at 100 $^{o}$C for 30 s. Later on, the wafer was dipped in 49% HF for one minute to remove the natural oxide. Immediately after HF dipping, wafer was placed in the ALD chamber under 150 $^{o}$C for about 2 minutes in order to warm up the substrate as well as getting a very thin layer of $SiO_2$. Then a layer of 30 Å $HfO_2$ was deposited at same temperature with TDMAH and $H_2O$ as the deposition precursors whose deposition rate is about 1 Å/cycle. After ALD, gate electrode patterns were then formed by optical lithography on the photoresistor which was spin coated on the wafer. Followed by the lifting-off process in acetone, a layer of 50 nm TiN was deposited on the $HfO_2$ by reactive sputtering. The power and pressure during sputter was about 200 W and 4 mTorr, respectively. No thermal annealing was performed on the samples.

3.1.4.2. Electrical Stress and Degradation Mechanisms

The electrical bias was applied on the TiN top electrode with back side of the wafer grounded. All the measurements in this section were performed with Agilent 4155B and Micromanipulator probe station. All devices employed for stress have same active area of $100\times100\ \mu m^2$ and were stressed under different polarities and different levels of biases. Before electrical stress, the current-voltage (I-V) curves of the MOS capacitor have been measured by a voltage ramping method with voltage step of 20 mV from 0 to 2 V or -2.5 V. In Fig. 3.1.5, the swept I-V curves on both directions have been shown and fitted with F-N tunneling equations which is simplified from equation 3.4. It could be observed that the overlapping parts between the experimental and theoretical are different for different polarities of bias. For negative bias, the fitted line starts to overlap with the experimental data from around -1.8 V. While for positive bias, it is close to 1.4 V. The EOT for such

gate stack was estimated by the current leakage at inversion region |V|= 1 V which gives a value of 1.5 nm [169]. This value would be used in the following field calculations.

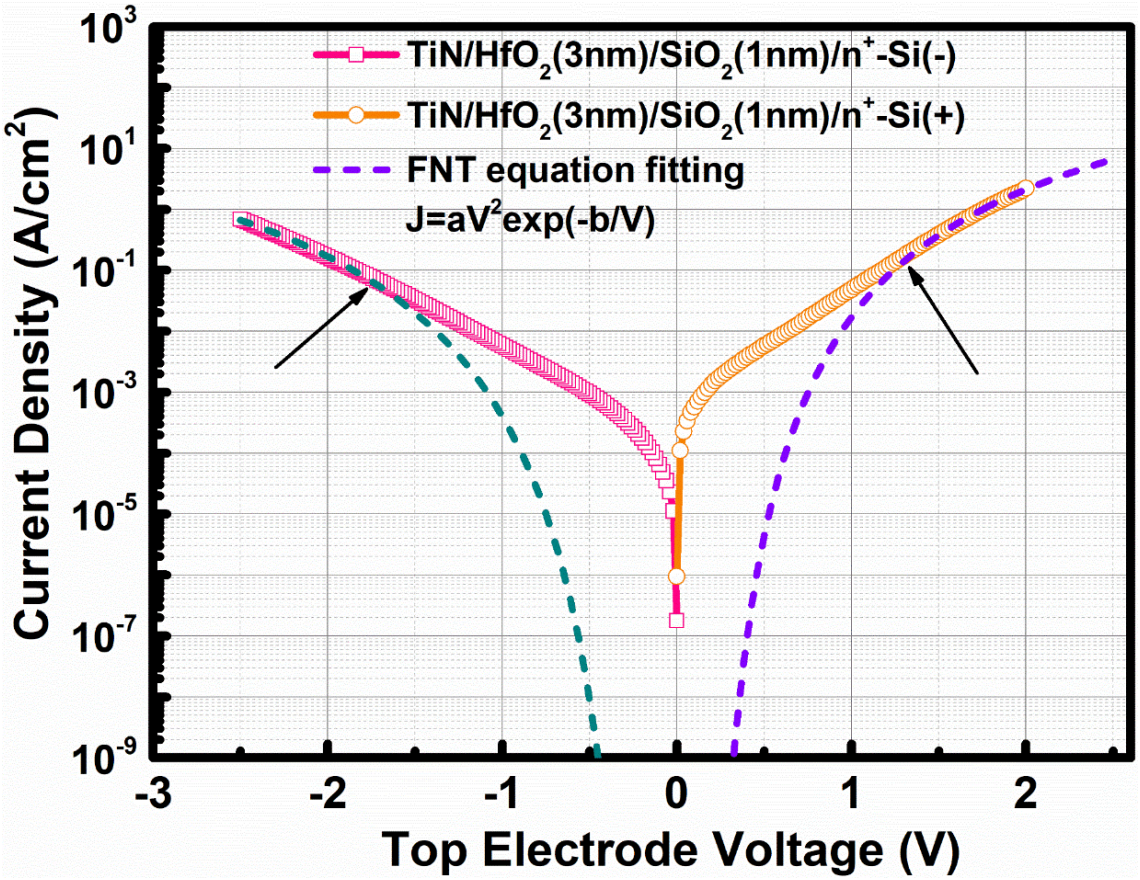


Fig. 3.1.5. Measured DC I-V curves of MOS capacitor with different polarities from -2.5 V to 2 V.

Then multiple devices of same active area were stressed under constant voltage (CVS) with sampling of the current at each time step. Sampling time interval during different levels of stress has been set differently in order to capture as many data points as possible before breakdown. The devices under very high stress such as ±3 V have shown immediate breakdown once bias is on indicating a time-to-failure (TTF) shorter than 5 ms which is

out of the measurement range. Hence, no available plots for these stresses were made and plotted. In Figs. 3.1.6, the typical current-time (I-t) plots for different levels of stress are shown. The current values in all plots have been normalized by dividing each time-0 current ($I_0$) when stress just turned on.

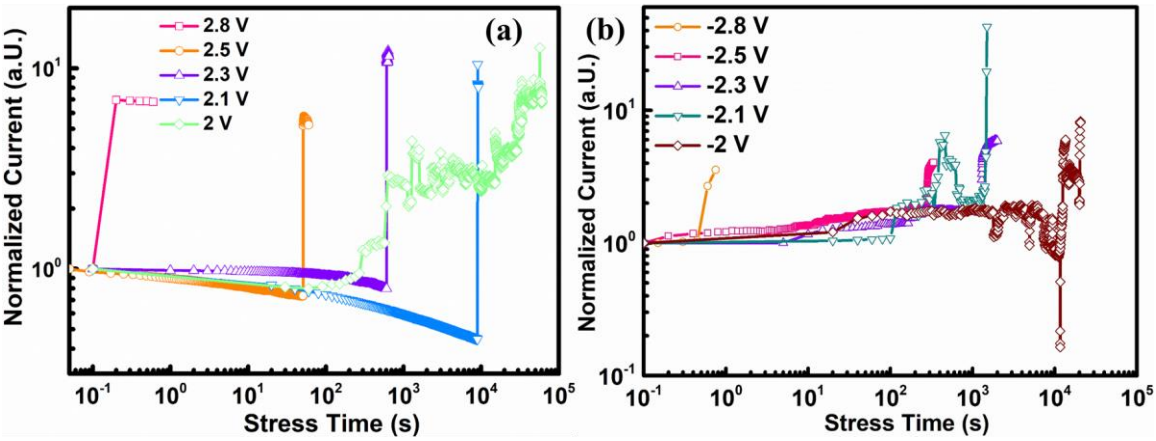


Fig. 3.1.6. Dual-log scale of I-t characteristics recorded for TiN/$HfO_2$/$SiO_2$/$n^+$-Si under stress with different levels of bias in (a) negative polarity and (b) positive polarity. All devices are stressed at room temperature.

As shown in Fig. 3.1.6 (a), I-t plots under different positive bias stresses have been presented. During positive bias stress, the currents under different levels of stress decreases at first stage of stress until a sharp current increase occurs together with dielectric breakdown. The absorption current behavior is similar to the $HfO_2$ based MIM structure SILC study discussed in the last section [108]. However, for most positive stress except for 2 V one, only a sharp current increase was observed after certain time of absorption current instead of the gradual current increase (SILC) reported in the aforementioned MIM stress study. With 2 V bias, gradual current increase has been observed during stress as well as

current fluctuations. The reason to observe different response current under different stress levels might be from different degradation mechanisms. In our devices, the ALD is performed at 150 $^{o}$C, but in [108] the $HfO_2$ is deposited at 300 $^{o}$C with various thermal annealing studied. This indicates that the absorption current induced by the ANPC like defect centers are natural characteristics of amorphous $HfO_2$ by which $HfO_2$ degradations under stress could be implied. For the higher bias situations, first breakdown spot could happen at $TiN/HfO_2$ interface triggered by the substrate injected electrons. These substrate injected electrons would gain sufficient energies during the tunneling process in the dielectrics and then lose their energies at $TiN/HfO_2$ interface to excite electron-hole pairs by surface plasmon oscillations [170]. The generated high energy holes (hot holes) could be swept into the dielectrics by electric field while those with lower energies would recombine with electrons at the interface. After swept into the dielectrics, hot holes would either get captured by the existing traps at the interface/inside the bulk of the gate dielectrics or tunnel deep into the bulk of the gate dielectrics (ultra-hot holes). During the tunneling in the gate dielectrics, these ultra-hot holes would create more defects (oxygen vacancies) in the gate stack by either transfer their kinetic energies to the lattice bond or recombine with electrons and emitting large amounts of energies. Since this breakdown is an instant process within micro-seconds, no gradual change of current could be seen. But this type of breakdown process requires high energy electrons to trigger for which the 2 V stress case might not satisfy. Thus for the 2 V stress, there must be other breakdown mechanisms taking places degrading the dielectrics. According to the estimated energy band diagram Fig. 3.1.7, it could be inferred that below 2.1 V, not only the entire thickness of the $SiO_2$ barrier but also some thickness of $HfO_2$ barrier are seen by the substrate injected

electrons. While with bias above 2.1 V, the substrate injected electrons would only see the entire $SiO_2$ thickness as barrier. Accordingly, the degradations with stress under 2.1 V would start from the $SiO_2/HfO_2$ interface and then propagate into the $HfO_2$ and finally $SiO_2$.

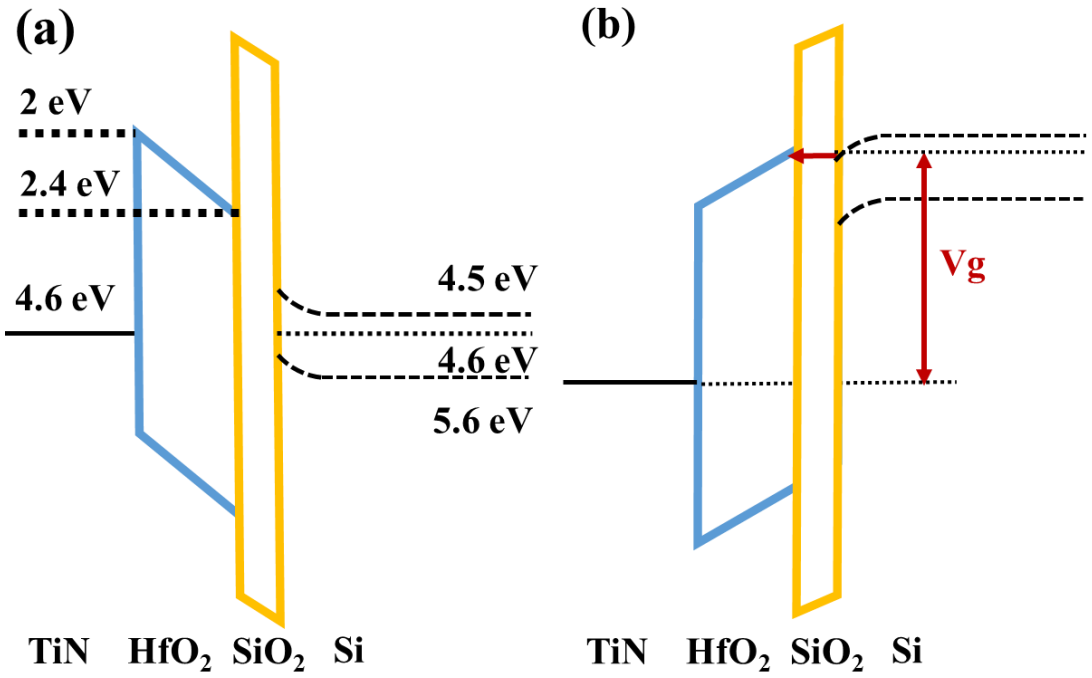


Fig. 3.1.7. The estimated band diagram (a) at equilibrium and (b) under positive bias.

In Fig. 3.1.6 (b), the I-t plots under different negative stress levels are shown. It could be observed that the current under all levels of stress kept increasing until first time breakdown happens with normalized current increasing to 3-6 times of $I_0$. Such behavior could also be called progressive dielectric breakdown with the first time breakdown happening in the $SiO_2$ from either $HfO_2/SiO_2$ or $SiO_2/Si$ interface degradations depending on the energies of the gate injected electrons. These gate injected electrons after FN tunneling would first cause damages (creating oxygen vacancies) at the $HfO_2/SiO_2$

interfaces where high degrees of disorders are expected. These damages would further propagate into the $SiO_2$ leading to the breakdown of the entire $SiO_2$. Once the energies of the injected electrons are sufficiently high, it would also trigger impact ionization (II), anode hole injection (AHI) or anode hydrogen release (AHR) in substrate where positive charges would be generated. Under the electric field, these positive charges would be swept into the $SiO_2$ causing further breakdown of the entire $SiO_2$. Though these two degradation mechanisms are from different directions, their final targets are same. Interesting behavior of the fluctuations of current under stress were observed until an even sharper breakdown current takes place. Since the filament is made up from oxygen vacancies which are easily thermally agitated, the electron transport through these defect sites definitely would transfer energies to them and excite them to recombine and generate frequently. The following breakdowns would be due to the entire dielectric stack breakdown after long time stress. Once the entire gate stack breakdown happens, the filament would get stronger in the $SiO_2$ layer during continued stress but still weak inside the $HfO_2$ layer. The robustness of the filament mainly differentiates by the density of defects or the number of discrete conductive filaments.

3.1.4.3. Lifetime Predictions

In Fig. 3.1.8, the TTF-oxide field relationships have been plotted according to the results shown in Fig. 3.1.6. Then a linear fitting on the available TTF data has been performed and shown together with the relationship plots. Since 1.5 nm EOT is expected to be used below 32 nm technologies in which 1.0-1.5 V $V_{dd}$ are used, the TTF under such voltages should be paid more attentions. From the fitting and extrapolation, the TTF is more than 10 years

for positive bias while 3 years for negative bias. The slope for positive is around 0.73 and is around 0.96 for negative bias. These results might indicate that the degradation mechanisms at different polarities could be related to the reason why PMOS is more vulnerable than NMOS under stress. However, the TTF extracted here is not from statistics of large volume of stress data so it could only stand for a rough estimation.

As we discussed before, in order to predict the lifetime of the MOSFET under stress, it is necessary to consider many aspects such as material systems and their processing methods, device geometries, electrical energies injected, temperatures and accumulated time. Each impact factor might contribute a reliability barrier for the total reliability lifetime. Therefore, in the real life application we need to individualize the impact from all these factors and put them back to the prediction methods in a reasonable manner. Usually, single degradation mechanism could be described by the Arrhenius equation which has an exponential relation with temperature, electric field and activation energy for each barrier. Nevertheless, these barriers might not be in a simple serial relation instead could be correlating with each other in a complicated way. Taking the high-k stack under stress as an example, if the breakdown of each layer was considered separately, then the whole stack might now breakdown under certain bias where only one layer is undergoing breakdown. However, if we consider the situation after one layer breakdown that most of the field would be applied across the rest layers, then probably under certain situation, the whole stacks would suffer from breakdown or none of the layer would breakdown.

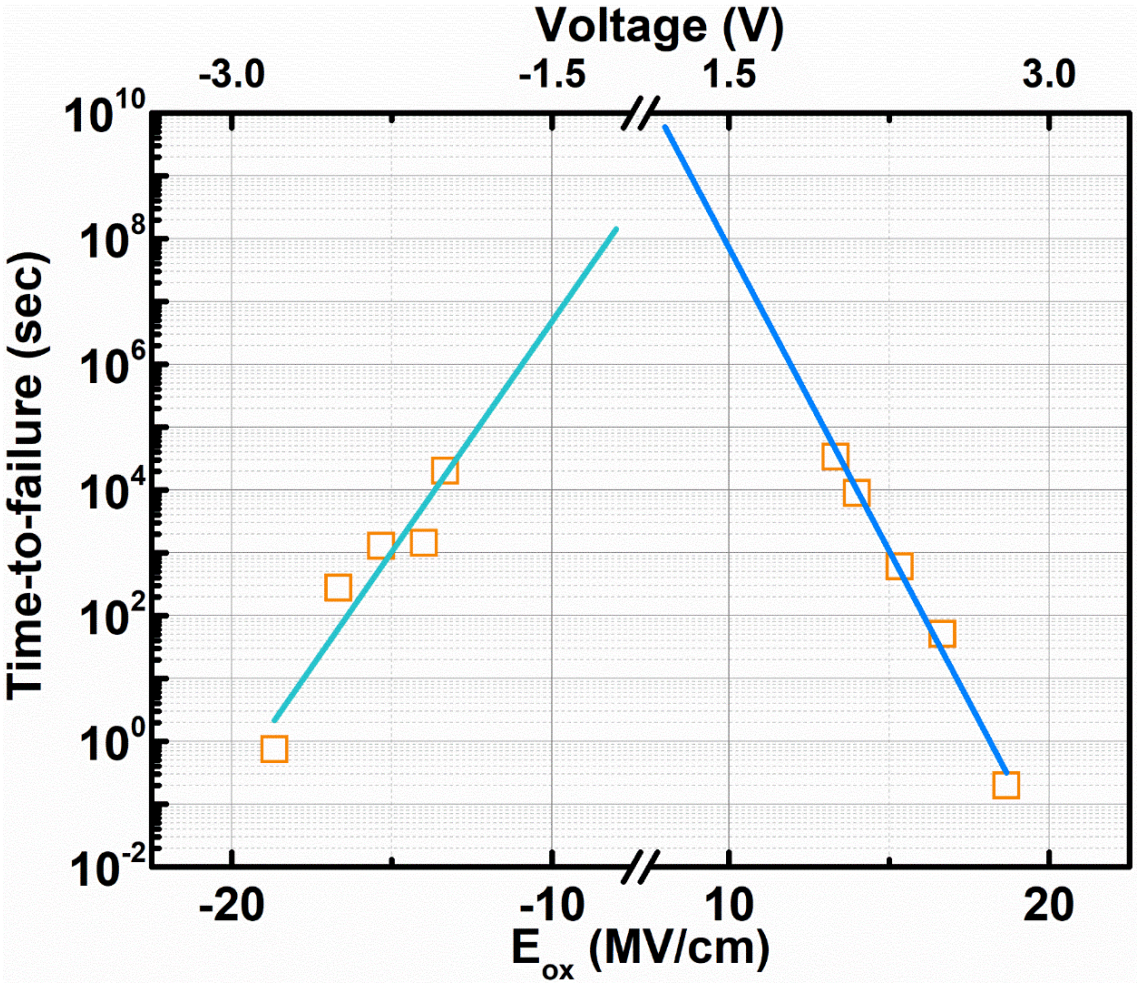


Fig. 3.1.8. TTF-Electric field in oxide plot extracted from the results shown in Fig. 3.1.6 with linear fitting with negative stress and positive stress and extrapolation to ±1.2 V, respectively. All the stress is performed at room temperature.

3.2. Ramp Voltage Stress Study

Since MOSFET is the critical component for computation networks, most of the degradation studies of high-k dielectrics were performed with the MOS structure. However, the fast development of computation networks have led to the giant speed gaps between processors and memory. Though recent development of dynamic random access memory

(DRAM) has come to DDR5 or DDR6 which have drastic improvement in speed, the density of DRAM is still the bottle neck limiting the whole computing speed. Basically, DRAM contains a metal/insulator/metal (MIM) structure based capacitor and a control MOSFET. To enlarge the density of DRAM, it is to make the capacitor as dense as possible while still maintain its stability. This requires the capacitor unit in DRAM to be as large as possible to reduce the refresh time which requires either a larger capacitor area or a thin enough dielectric or even both. As discussed in the last section, large area plus thin dielectrics layer indicate fast risks of breakdown and aging of the dielectrics. Recently, it has been discovered that the MIM structure based RRAM devices could get Forming free with very thin dielectric layers [171]. The Forming process in RRAM usually requires a voltage sweep from 0 to a large value (like 6 V) to cause the breakdown of the dielectrics. Such breakdown process is very similar to the substrate injection induced dielectric breakdown process discussed before. But in MIM structure, there is no substrate silicon, instead, as the injection is mostly from cathode electrode, so we would call it "Cathode Injection." A Forming free specification for RRAM devices means its Forming stop voltage is very close or same to the switching voltage (SET and RESET) which is usually below 2 V. Thus investigation of the I-V curves from the RRAM Forming process would potentially provide very useful information for the thin dielectric reliability studies. In this section, the RVS induced dielectric breakdown in $HfO_2$ would be investigated with MIM structures.

### 3.2.1. MIM Structure Fabrication and Characterization

#### 3.2.1.1. TiN/$HfO_x$/Pt Based MIM Structure

In the first chapter, the three popular oxide candidates for RRAM devices are mentioned that they are $HfO_x$, $TiO_x$ and $TaO_x$ in which $HfO_x$ has been regarded as the most promising one to be used in the next generation nonvolatile memory cells. Thus $HfO_x$ based RRAM devices were designed with a cross-point (X-point) like structure as shown in Fig. 1.4.2 (b). The active cell area is only between the overlapping part of the top and bottom electrode since dielectric breakdown is a localized phenomenon. TiN and Pt electrodes are used to sandwich $HfO_x$ in between due to the excellent performances ever reported in such combination [180]. The devices are fabricated in the ASU Nano Fab with a 4 inch wafer where a set of 5 inch masks are designed and ordered and shown in Figs. 3.2.1. There are 52 repeated dies in such set of masks with each die containing 4 different areas where different functions are reserved as shown in Fig 3.2.1 (b). In the upper part of the mask, it has been designed as the 12×12 X-point RRAM arrays with 1×1 $\mu m^2$, 2×2 $\mu m^2$, 5×5 $\mu m^2$, and 10×10 $\mu m^2$ feature areas. Since the active area is a square, the devices would be named with feature size such as 1 μm device. Under the individual devices, there is a large square is reserved for materials characterization or metrology purposes. Right of the large square is the functional area for 3D X-point RRAM arrays where X-point structures are vertically designed. However, this part might need different process flows comparing with the other three parts.

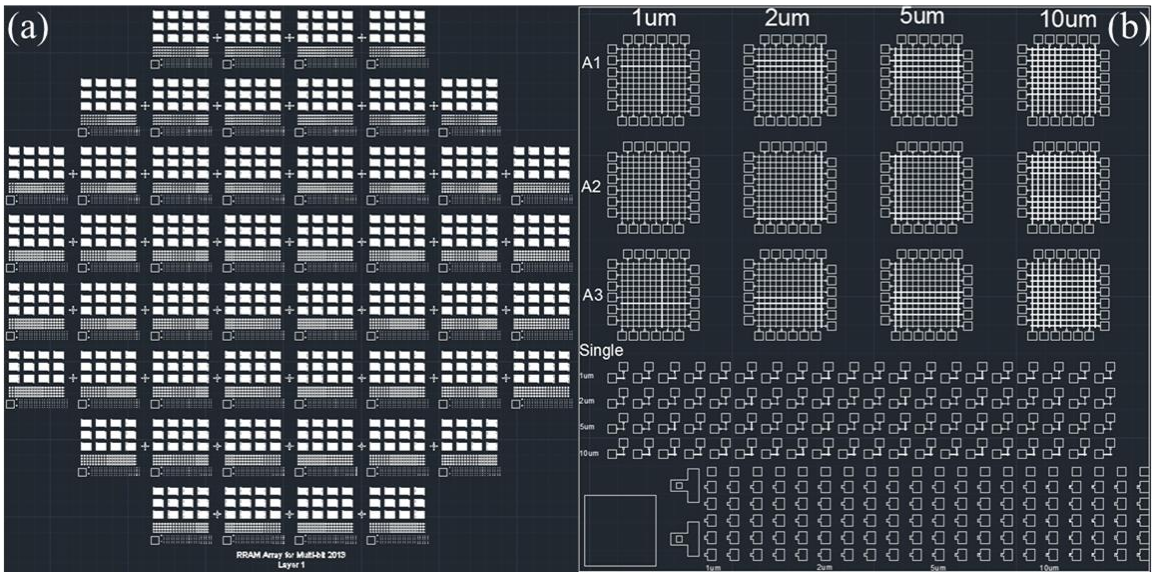


Fig. 3.2.1. The (a) whole view of masks and (b) one die in the mask for the RRAM devices.

3.2.1.2. Fabrications of the TiN/$HfO_x$/Pt Structure

By using the mask shown in Fig. 3.2.1 (a), the RRAM devices are fabricated except for the 3D X-point RRAMs. As shown in Figs 3.2.2, the fabricated 12×12 X-point RRAM arrays and individual RRAM devices are shown with a 1 μm feature size. In Figs. 3.2.2, the electrode with yellow or bronze color is TiN which is underneath as the bottom electrodes. While the white metal electrodes are made of Pt acting as the top electrodes. The $HfO_x$ oxides are everywhere and was etched out above the TiN pads for electrical measurements.

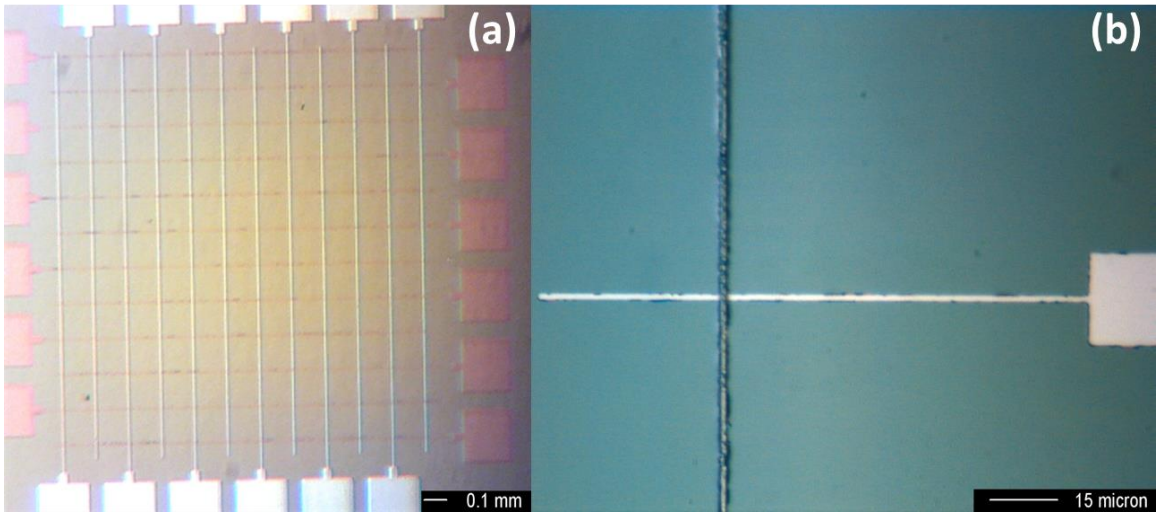


Fig. 3.2.2. (a) The fabricated X-point RRAM array and (b) individual RRAM device with a feature size of $1\times1$ $\mu m^2$ and under different scales.

The fabrication flow of the X-point RRAM devices are shown in Fig. 3.2.3. In the X-point RRAM device process, it starts from a 4-inch wafer with 100 nm oxide and finishes with 3 times lithography. The typical bottom electrode thickness is about 20-30 nm and the oxide thickness is about several nanometers (6 nm in this study). In order to prevent the discontinuity problem at the cross-point location, the typical top electrode thickness is around 50-65 nm. The TiN bottom electrode was formed by lift-off process with optical lithography on the $SiO_2$/Si substrate. The $HfO_x$ film was deposited by atomic layer deposition (ALD) at 150 $^oC$ using TDMAH and $H_2O$ as precursors with a deposition rate of 1 Å/cycle. After the DC sputtering and lifting off of Pt top electrode, the bottom electrode was exposed by wet etching the $HfO_x$ layer with buffered oxide etch (BOE) ultra-etchant. The photoresistor used here is AZ 3312 for all the cases with different exposing times for different purposes. No thermal annealing are required or performed. The

fabrication flows for the 3D X-point RRAM devices would not be presented here and are beyond this work.

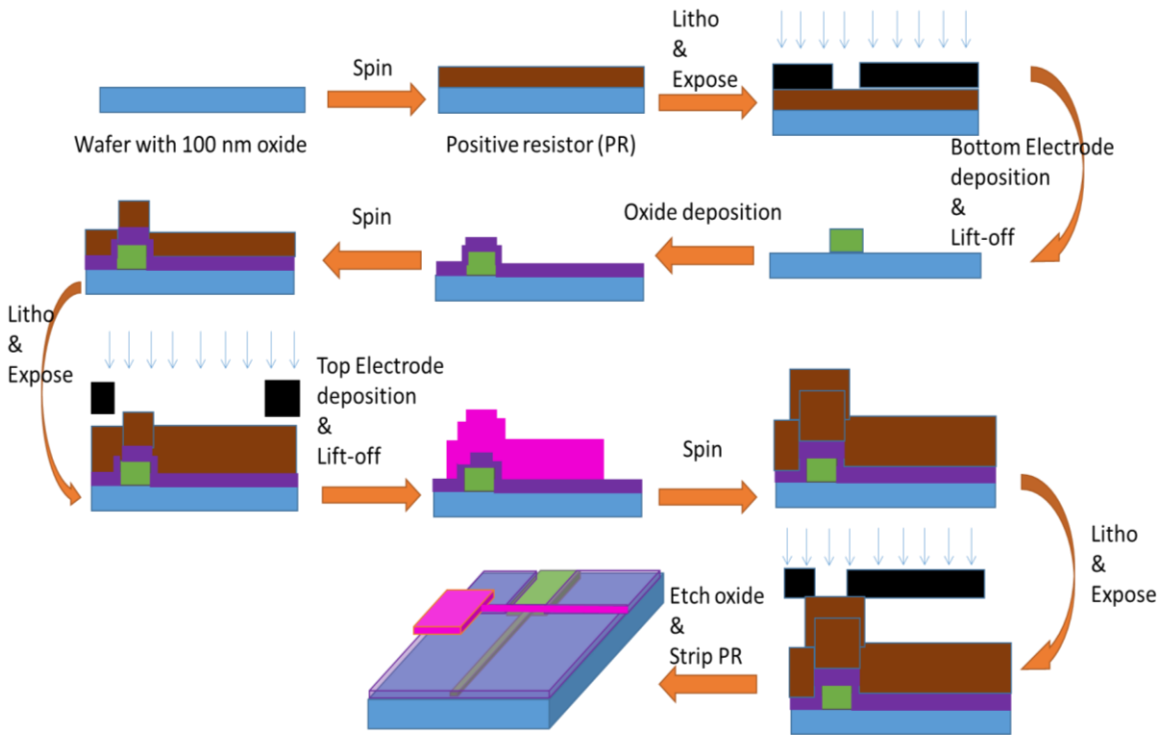


Fig. 3.2.3. The Fabrication flow of X-point RRAM devices applicable for individual device and array.

3.2.1.3. The Measurement System for TiN/$HfO_x$/Pt Cells

Devices are ready for electrical testing once all process flows finished. To test the individual devices, bias would be applied on TiN electrodes with Pt grounded except for extra statements. Wafers are placed on the bench stage of Cascade Summit-1200 probe station with probe needles of various sizes connecting to the electrodes of the RRAM devices. DC characteristics are measured by Keithley 4200 semiconductor parameter

analyzer (SPA), the pulse testing is using Agilent 81160A arbitrary waveform generator (AWG) with a Keithley 707B switch matrix.

Meanwhile, in Fig. 3.2.4, the testing systems are shown for measuring 12×12 X-point RRAM arrays. To test the 12×12 X-point RRAM arrays, 24-pin probe cards are designed according to the masks for aligning with the pad locations on wafer. Operation commands are sent out by Keithley 4200 through GPIB cables to control all the equipment shown here with all the commands programmed in C language. Testing data were collected directly from Keithley 4200 SPA. The details about the array testing works are beyond this work so would not be extensively discussed.

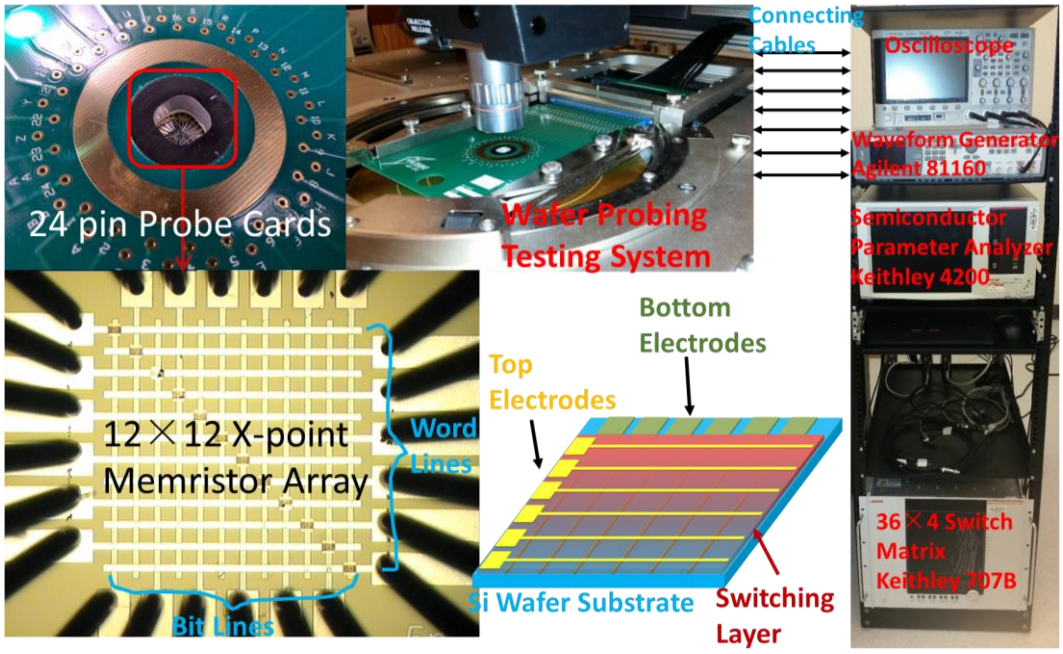


Fig. 3.2.4. The testing systems for measuring 12×12 X-point RRAM arrays.

3.2.2. Breakdown I-V Curves from RVS

In Fig. 3.2.5 (a), the typical semi-log scale forming curves for different $HfO_2$ thickness based RRAM devices are shown with the structures listed in a top-down sequence. The Pt electrodes (cathode) are grounded for all samples with bias applied on the anode. The devices are measured by a Keithley 4200 semiconductor parameter analyzer in voltage sweep mode from 0 V and stopped at different voltages for devices of different oxide thickness with 50 mV voltage step. Current compliance was set to be 1 μA for all devices. Forming voltage polarity was chosen to be positive though negative bias is also able to Form the devices. To better clarify the conduction mechanisms, current density vs electric field plots are made from the I-V plots and shown in Fig. 3.2.5 (b).

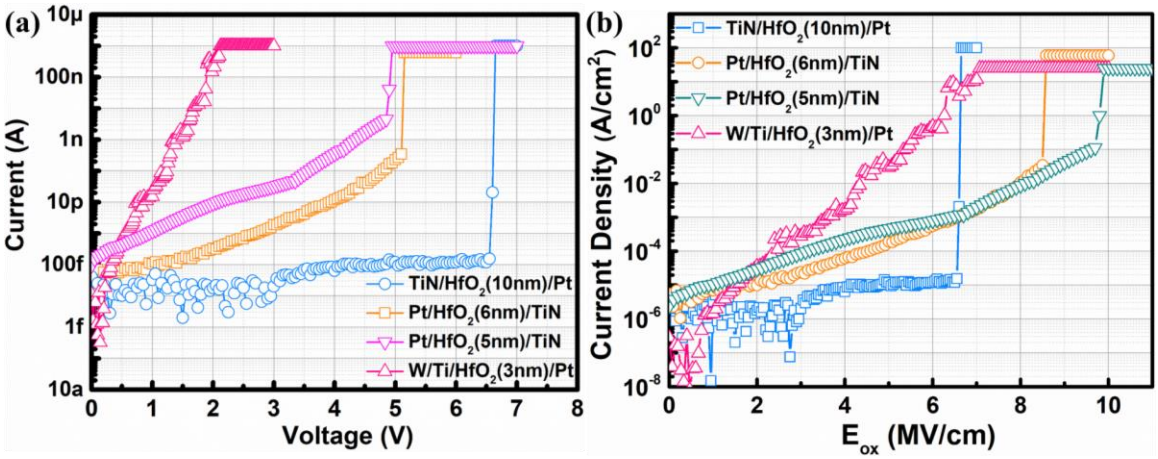


Fig. 3.2.5. The semi-log scale forming (a) I-V curves and (b) J-E curves from RRAM devices with different $HfO_2$ thickness.

An obvious reduced Forming voltage as well as the increased leakage current before Forming while thickness is decreasing. For device with 10 nm $HfO_2$, the current before breakdown is only at 100 fA level. But devices with 6 nm $HfO_2$ has the leakage increased

to several hundred-pA and with 3 nm $HfO_2$ even increased to hundred-nA. For each Forming I-V curve, the slope change positions are different from others. For 10 nm device, the current below 3 MV/cm is very noise which is supposed to be direct tunneling current. When field goes beyond 3 MV/cm, the current starts to increase with voltage. Theoretically, the anode-$HfO_2$ barrier is about 2.6 eV with TiN and 2.3 eV with Ti. Such transition at 3 V would indicate the F-N tunneling regime but the current level is too small to be regarded as. Therefore, I-V curve fittings have been executed and shown in Fig. 3.2.6 with F-N tunneling regime and trap-assisted tunneling regime. For 10 nm device under the field from 3 MV/cm to 7 MV/cm, the curve is not seeing slopes for F-N tunneling but fits well with trap-assisted tunneling. To understand this, a rough band and thickness calculation would be presented here. Aforementioned in the previous section, the F-N tunneling is made up from two steps: 1) direct tunneling through the triangular barrier of $HfO_2$ and 2) moving freely beyond the conduction band. In order to tunnel into the 10 nm $HfO_2$ conduction band, the triangular barrier width should be within 40 Å where the voltage drop on the oxide should be at least 8.1 V. Since the Forming voltage is around 6 V where the thickness for step 1) is still larger than 40 Å, the conduction mechanism would not be F-N tunneling current. Instead, it might be trap-assisted tunneling (TAT).

Devices with 6 nm and 5 nm $HfO_2$ have shown very similar I-V curves with 5 nm device Forming at higher electric field but lower voltage and conducting more leakage current before Forming. As could be observed in Fig. 3.2.5 (b), both of them have I-V slopes changes before Forming. In Fig. 3.2.6 (a), the F-N tunneling equation fitted linearly with 5 nm and 6 nm devices between 7 MV/cm to 10 MV/cm. Meanwhile, from the curve fitting shown in Fig. 3.2.6 (b), TAT conduction was found valid for 5 nm device with electric

field between 3 MV/cm and 7 MV/cm while it is valid between 2 MV/cm to 4 MV/cm for 6 nm device. Between 4 MV/cm to 7 MV/cm for 6 nm device looks quite like a transition region where the current could be constituted by both types of conduction mechanisms or possibly by the P-F emission since the traps involved in the conduction might emit electrons into the conduction band directly. For device with 3 nm HfO2, it has a perfect TAT current fitting between 2 MV/cm and 4 MV/cm. In addition, when fields are higher than 4 MV/cm, the transport mechanism has changed to F-N tunneling.

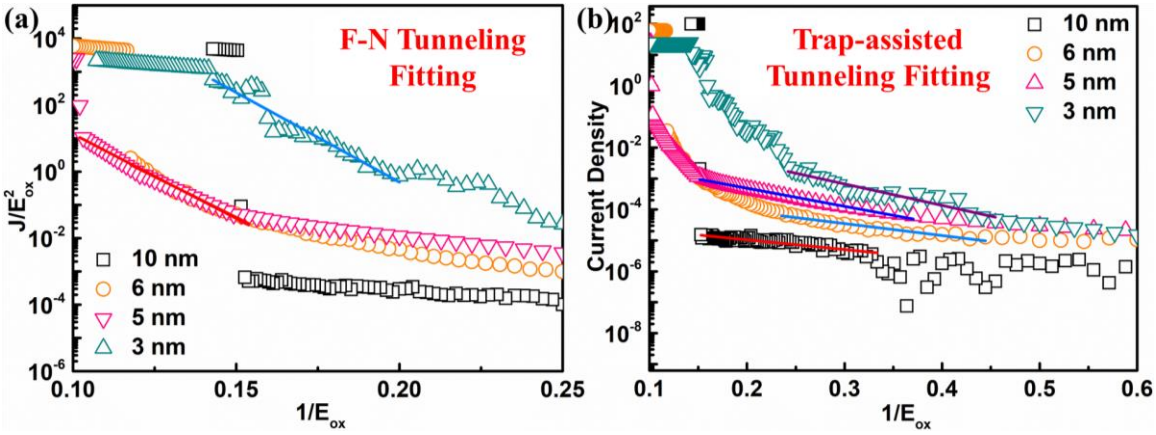


Fig. 3.2.6. Fitted semi-log scale I-V curves for RRAM devices with different $HfO_2$ thickness by (a) F-N tunneling shown in equation 3.4 and (b) trap-assisted tunneling shown in equation 3.12.

The devices with thin dielectrics all have seen F-N tunneling while the thick one (10 nm) only see TAT current indicating that the thickness of the dielectrics would influence the main transport mechanism that further impacts the breakdown or Forming process for the dielectrics. It is worth mentioning here that the traps involved in the TAT conduction are only left from fabrication processes [172]. Since most of the TAT transport is an

inelastic process, no new defects would be generated during TAT [173]. In addition, though the electron-phonon reaction could result in the excitation of the intrinsic traps, the electron energy here is not high enough to trigger that. Then it comes out a question that since device with 10 nm $HfO_2$ only saw TAT current before breakdown, does that mean such TAT current is elastic one? We preserve the possibility that such TAT current is elastic and generated new traps or triggered breakdown. However, if the trap-assisted transport process in the 10 nm device is recalled (shown in Fig. 3.2.7), then it is possible to find that the trap actually could only function as a bridge to help the electrons to tunnel through without generating high energy carriers. For such scenario, TAT is the dominant current but breakdown is trigger by the indirectly injected cathode electrons.

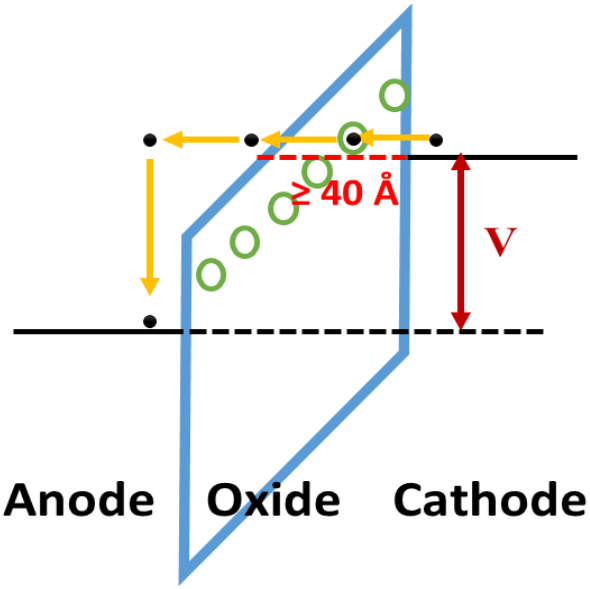


Fig. 3.2.7. The illustrated energy band diagram with applied voltage larger than anode barrier but current was dominated by trap-assisted tunneling current.

The conduction current before dielectric breakdown for 3 nm $HfO_2$ based devices are very different from the other three. As could be seen from Fig. 3.2.5 (b), the current density

for 3 nm device below 1 MV/cm is much lower than the other three devices and the linear J-E relationships have been observed from 1 MV/cm. Though I-V fitting in Figs. 3.2.6 has shown that the current potentially could be dominated by nonlinear conduction mechanism, its direct appearances are much like Ohmic conduction. Such behavior probably is due to the Ti layer scavenging effects [174] that a thin interface layer of $TiO_x$ would be generated by reducing the $HfO_2$. Such redox process does not have to happen during process but could be undergoing during the early stage of the Forming sweep. Once the interface layer has been formed, the barrier between Anode-oxide would be shortened to about 0.4 eV [175] which is exactly at 1 MV/cm in field as shown in Fig. 3.2.5 (b).

Before discussing about the detailed breakdown physics, statistical data of the Forming data has been presented in Fig. 3.2.8 with Weibull distributions. Linear fittings have been performed on the Weibull distributions with Weibull slope extracted. At least 20 devices were performed with Forming sweeps for each thickness based devices with all the devices functioning well after Forming. The electric field was then calculated and normalized with natural log scale. Weibull percentiles were calculated by dual natural log of 1/(1-F) where F is the cumulative percentage for each x-value. It could be seen from the Weibull plot that the Forming process complies with bimodal distributions one steep Weibull slope and one gradual Weibull slope. For the steep slope values, 10 nm devices are quite close to 3 nm devices while 5 nm devices are quite close to 6 nm devices. This suggests that 3 nm and 10 nm is dominated by same breakdown mechanism while 5 nm and 6 nm is determined by another. Except for 6 nm one, all have seen a small value for the gradual slope where the numbers are very close for devices with different thickness. These gradual slopes locate at higher field with a broader distribution than the steep ones indicating the third type of

breakdown mechanism. The steep slope stands for tighter distributions and might mean uniformed energies while the gradual slope stands for broader distributions and might signify a wide distribution of energies.

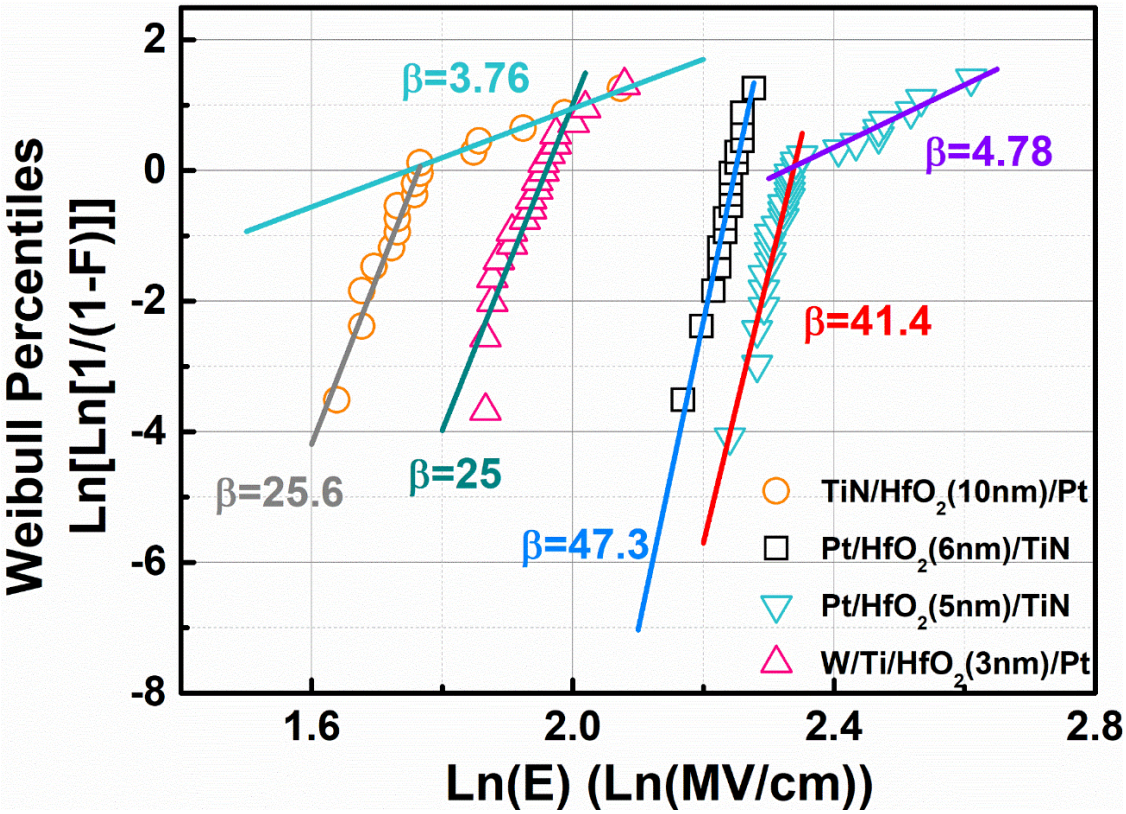


Fig. 3.2.8. The Weibull plot of the electric field during Forming in natural log scale and linearly fitted for Weibull slope extraction.

3.2.3. Synergies between Cathode Injection and Substrate Injection

Within the distribution for certain thickness, the steep slopes (> 20) in the Weibull plots (shown in Fig. 3.2.8) all locate at relatively lower field while the gradual ones (<10) all locate at high field. As aforementioned, gradual slope indicates broader distribution of energies at high field which most probably are due to the direct band gap impact ionization

in $HfO_2$ layer. Bandgap impact ionization (BII) usually requires an energy close to bandgap of the material that could only get satisfied by the tail distribution of the tunneling electrons injected from the cathode [158]. For the breakdown under relatively lower field, it could be cathode injection induced surface plasma oscillation (SPO) by which electron-pairs would be generated at the interface [170]. Though both BII process and SPO process generate electron-hole pairs, their behaviors are quite different with SPO requiring lower energy but retaining higher efficiency. Then if the 10 nm and 3 nm devices are grouped into Group 1 and the rest grouped into Group 2, we could find the different SPO slopes between Group 1 and 2. The reason could trace back to the fabrication sequence of the devices. For Group 1, Pt was deposited first and TiN last while Group 2 has the reversed stack. It could be observed in Fig. 3.2.8 Group 1 has distributions all locate at lower field even for thicker oxide. This must have something to do with reactions between TiN and $HfO_2$. As reported before [108], TiN first based stacks have 0.3 eV lower barrier between TiN and $HfO_2$ than the TiN last based stacks. In [108], the authors claimed that the origin of such barrier lowering is the interface dipoles. From the works performed here, it could be concluded that such interface dipoles though pulled down the barrier but could have offered another strengths to block the SPO at surface by suppressing the generation of the holes or recombine the holes with low energies. To trigger stable SPO at the interface, higher energy is required to eliminate such dipole first. The energy to break such dipole might already exceed the energy to trigger SPO and generate hot holes. Thus, the breakdown process becomes quite uniform once with such interface dipole.

With the experiments and discussions aforementioned, it is possible to reach a conclusion that for substrate injection induced dielectric breakdown in MOS capacitors and

the cathode injection induced Forming in RRAM devices or dielectric breakdown in MIM capacitors, the breakdown theories and physics are same. This conclusion could only be valid when same processing method and same material systems are used. Though the intrinsic breakdown of $HfO_2$ (BII in $HfO_2$) in MOS capacitors has not been profoundly studied with high level DC bias stress, this conclusion is still valid since the most effective way to induce electron-hole pairs is SPO which requires fewer energies than BII. After the generations of electron-hole pairs, the AHI based dielectric breakdown process is a very fast one and exactly the same for RRAM Forming and MOS capacitor breakdown. In summary, the high-k dielectric breakdown process in the scenarios under our investigation could be described as: 1) electron-hole generations by either medium energy electrons induced SPO at metal/$HfO_2$ interface or high energy electrons induced BII in $HfO_2$ followed by 2) AHI effects leading to holes moving and trapping in the $HfO_2$; 3) meanwhile hot holes would create oxygen vacancies from anode to cathode forming a conductive path between electrodes which leads to current sharp increase. Process 1) is an initiation, 2) is a must process for lowering the barrier and 3) is the final breakdown of dielectrics.

### 3.2.4. Percolation Path Theory

Many works have been discussing the existence of the percolation path between the two electrodes once dielectric breakdown happens where the path is made up from defects such as oxygen vacancies [131, 133, 176-177]. As discussed before, the oxygen vacancies or defects start to generate once the injected electrons contain sufficiently high energies. The ways to gain sufficiently high energy might include direct cathode injection (DCI) and indirect cathode injection (ICI). DCI is to serve the thin dielectric or high bias across the

oxide where electrons are in F-N tunneling regime. ICI works when large amounts of intrinsic traps inside the dielectrics are activated and the dielectrics is thick enough and bias is sufficient to trigger either BII or SPO. Here, the generation of electron-hole pairs is only part of the breakdown process which takes place less than nano-seconds and the actual percolation path formation process could sustain from the generation of electron-hole pairs until the end of electrical bias.

In order to describe the conductive percolation path theory in mathematical way, we need input parameters such as dielectric thickness, bandgap and affinity of the oxides, anode-cathode work functions, electrical bias, bias time and some parameters extracted from experiments. To calculate the conduction current under certain bias and temperature, transmission probabilities under different transport regimes would be calculated to determine among DT, F-N T, and TAT, so that current densities could be figured out. Once the energy threshold was reached, the percolation path would begin growing by calculating Mott-Gurney ion transport model [178] in both vertical and parallel direction. Once the conductive path are formed to connect anode and cathode, the growth in vertical direction would be stopped but the parallel growth would continue until the end of the bias. Details about this mathematical description would be further discussed in chapter 4 where RRAM modeling would be introduced. Here only physical concepts would be discussed.

Most of the related transports have been introduced in the second section in this chapter except for the TAT conduction since it is a quite complicated process. Here a simplified trap-assisted tunneling current density function would be supplemented in equation 3.12 [179]:

$$J_{TAT} = J_0 \exp\left(-\frac{4\sqrt{2qm_{ox}}}{3\hbar E_{ox}}\Phi_t^{\frac{3}{2}}\right), \quad (3.12)$$

where $\Phi_t$ the trap-conduction band energy barrier, $J_0$ is a fitting parameter and the other notations are the same defined before.

## 3.3. Post Breakdown Resistive Switching Effects Study

### 3.3.1. Basic Electric Behaviors of Resistive Switching

#### 3.3.1.1. Resistive Switching I-V curves

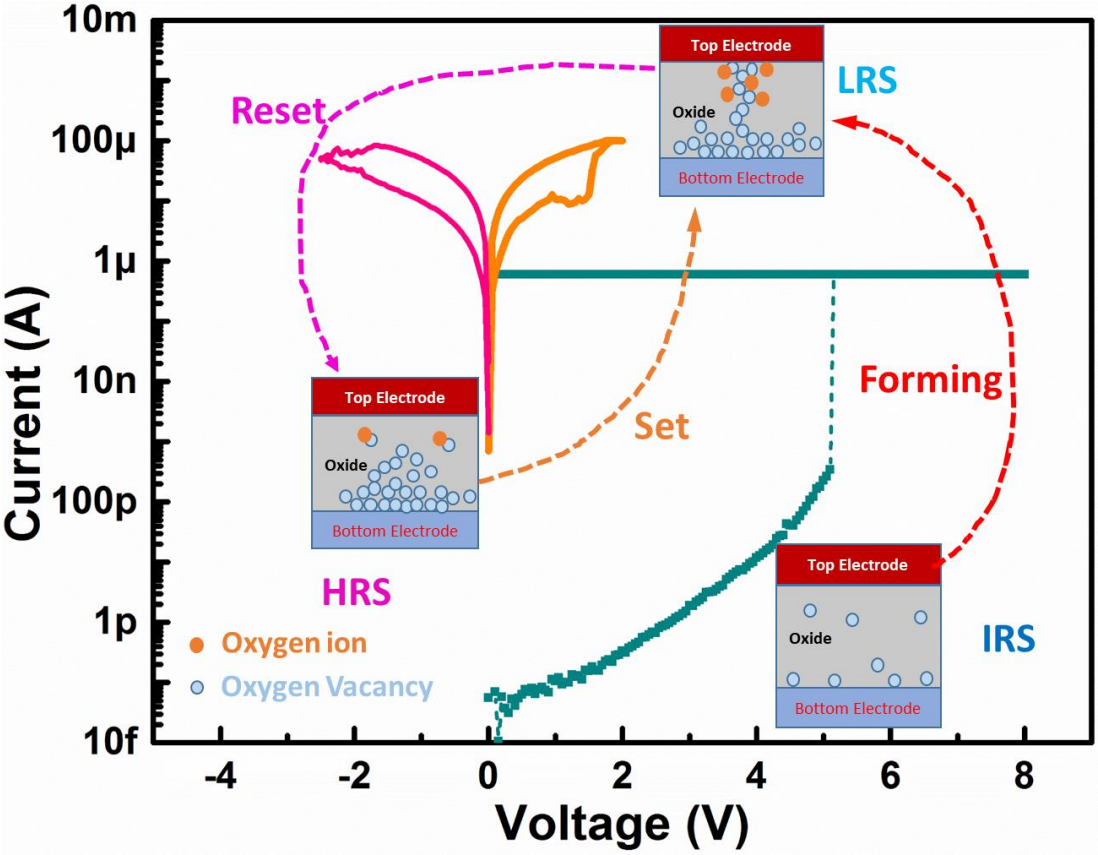


Fig. 3.3.1. The typical operation schematics for TiN/HfO$_x$/Pt based RRAM

In Fig. 3.3.1, the resistive switching curves together with I-V curves in sweep mode are shown. According to the figure, Forming (Breakdown) and SET are achieved by positive biases while RESET by negative biases on the TiN electrode. Since the breakdown process has been discussed in the previous section, here would start discussing after that. After Forming, there are some conditioning switches (not shown here) to stabilize the switching characteristics. During conditioning switches, the SET and RESET voltage values are decreasing until repeatable switches are achieved. The typical I-V characteristics shown in Fig. 3.3.1 are the repeatable switching cycles with SET sweeping from 0 V to 2 V and RESET sweeping from 0 V to -2.5 V. Since the switching of the resistances are performed using different polarities of biases, this type of RRAM is also called bipolar RRAM. The different states (LRS and HRS) can represent "1" and "0" in the binary system, respectively. All the DC sweeps are accompanied with back sweeps in order to check if the switching is successful. Current compliance in SET has been set higher than that in Forming so that a robust resistance state could be assured.

3.3.1.2. DC Cyclings

To inspect the stability of the oxygen vacancy density modulations effects, 100 DC sweeps have been performed on each devices. Devices with different areas are employed as shown in Fig. 3.3.2. The cycling was performed with SET sweep from 0 to 2 V and RESET sweep from 0 to -2 V. In the extraction process, SET voltage is defined as the first voltage reaching current compliance and RESET voltage is defined as the voltage corresponding to the peak current during RESET. Resistances after switching were sensed by a 0.1 V short voltage pulse. Tight distributions of either operation voltages or resistance

states after operation has indicated a stable modulation effects though only one device were used in each device size. In addition, decreases for both operating voltages and resistances with device size increasing have been observed. Devices with larger areas have shown smaller operation voltages and smaller resistance states as well as larger variations. With such relationships, it is possible to predict the performances of the devices with scaled feature sizes [85, 87].

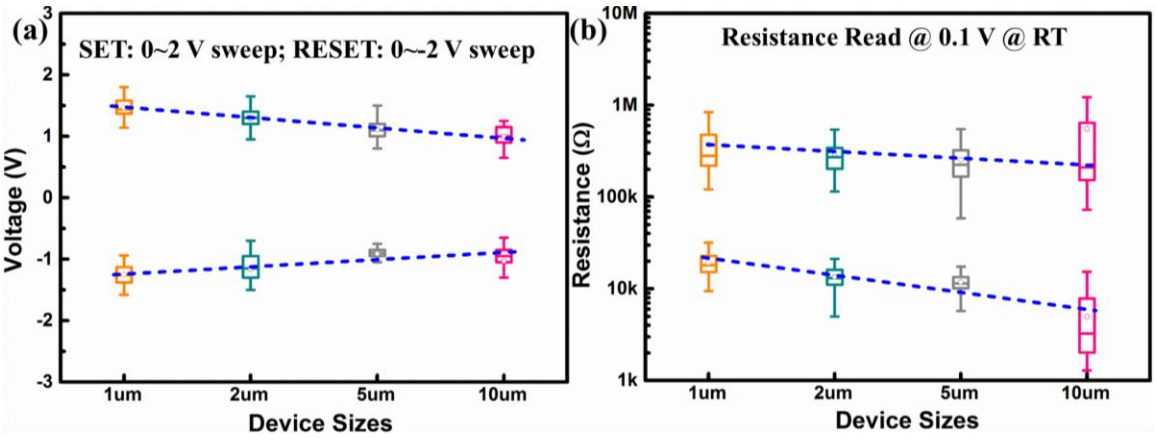


Fig. 3.3.2. Statistics of (a) operation voltages and (b) LRS and HRS values from 100 DC switching cycles of RRAM devices with different active areas at room temperature.

3.3.1.3. Transports in Filament

The hypothesized simplified switching mechanisms are also shown in the inset in Fig. 3.3.1. In LRS, at least one conductive filament (made up of oxygen vacancies) would be existing in $HfO_x$ connecting anode and cathode. While in HRS, part of the filament(s) would be ruptured resulting in a restricted conduction through the residues of the filament. In a word, RRAM is achieving the resistance states switch by electrical bias. For SET, it is quite similar to FORMING process in which oxygen vacancies and oxygen ions are created

by electrical bias. Oxygen vacancies would get neutralized by the electron fluence while oxygen ions would move towards anode or grain boundaries (GB) where they get neutralized. To demonstrate such switching mechanism, the switching I-Vs from RRAM devices at different temperatures from 4 K to 300 K are investigated and shown in Figs. 3.3.3. Multiple RRAM devices were employed for multi-cycles switching and the statistics of SET/RESET voltages slightly increase with lowering temperature. As temperature cools down, the thermal energies of oxygen vacancies and ions are reduced. Therefore, more electric field is needed to form or rupture the conductive filament. This signifies that thermal energy might be one part of the switching dynamics.

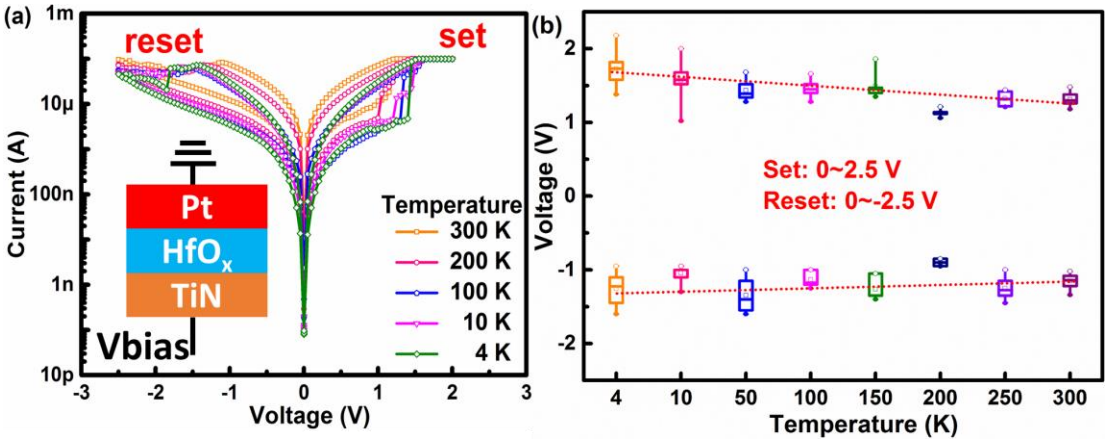


Fig. 3.3.3. (a) The switching I-V curves of the TiN/HfO$_x$/Pt RRAM devices from 300 K to 4 K and (b) the statistics of the set and reset voltages of the devices in different temperatures.

To further understand the role of thermal energy in switching, the I-V sweeps of the devices across a wide range of temperatures have been presented in Figs. 3.3.4 (a) and (b) in a voltage range where no switching happens. It could be observed that the current for both LRS and HRS are all decreasing when temperature is decreasing. The log scale plots

have shown slope changes around ±0.3 V for both LRS and HRS indicating the conduction mechanism change around this voltage. I-V curve fittings then have been segmented into two ranges: low field (< ±0.3 V) and high field (> ±0.3 V).

Im Fig. 3.3.4 (a), perfect low field fittings have been presented by using hopping conduction equations expressed in equation 3.7. In addition, the hopping distances and hopping activation energies have been extracted and shown for HRS and LRS, respectively. For LRS, the hopping distance at room temperature is around 1.8 Å, and yet it is about 3-5 Å for HRS. The activation energy at room temperature for LRS is about 40 meV, but it is close to 80 meV for HRS. As all the values are within the reasonable range, it is rational to conclude that hopping is the dominant transport at low field. At the same time, all the extracted values have manifested temperature dependences which further suggests that the degree of excitement of the hopping sites (oxygen vacancies) in the oxide has some kind of relationships with thermal energy. Accordingly, it is reasonable to infer variable range hopping (VRH) when temperature is below critical value such as Debye temperature (around 100 K) or the trap density at the weakest point is below critical value. As discussed in chapter 3, hopping mechanism somehow resembles direct tunneling process in which the transmission probability of electrons from one defect site to another is very important. Under low field with either low temperature or sparse trap sites, the electron hopping probability would be relatively low so that the current would also be proportionally low. Since oxygen vacancy densities in the oxides are higher in LRS than in HRS, it could still allow electrons to hopping to the nearest sites (NNH) at room temperature when traps are all activated. But once temperature cools down, the hopping probabilities for both states would get close but still different. These changes lead to differences in LRS and HRS.

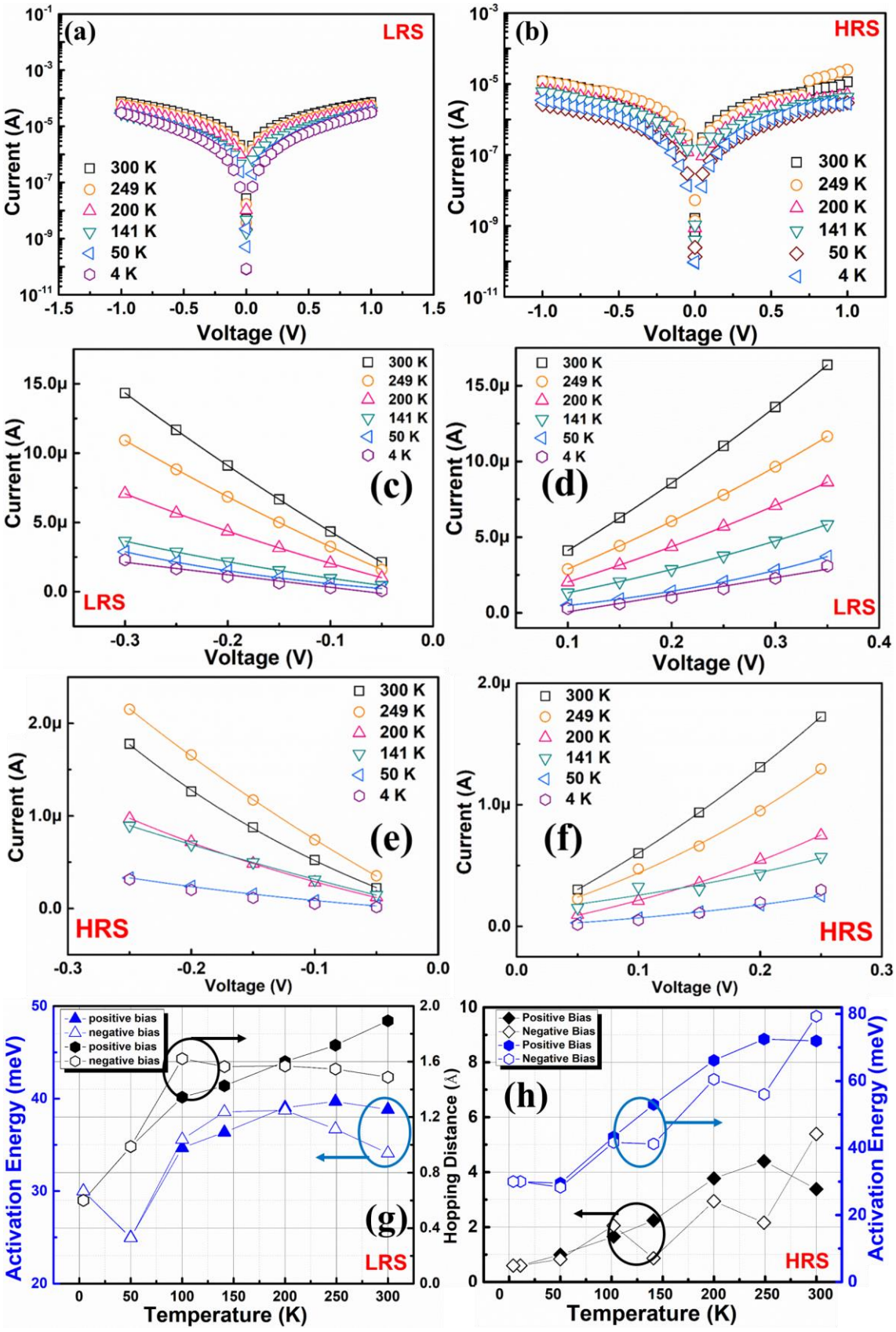
(a)
LRS
Current (A)
Voltage (V)
300 K
249 K
200 K
141 K
50 K
4 K
(b)
HRS
(c)
LRS
(d)
LRS
(e)
HRS
(f)
HRS
(g)
LRS
positive bias
negative bias
positive bias
negative bias
Activation Energy (meV)
Hopping Distance (Å)
Temperature (K)
(h)
HRS
Positive Bias
Negative Bias
Positive Bias
Negative Bias
Activation Energy (meV)
Temperature (K)

Fig. 3.3.4. Log scale I-V sweeps of (a) LRS and (b) HRS taken from the switching curves under different temperatures from 4 K to 300 K. Low field I-V fitting with hopping conduction of (c) LRS in negative bias, (d) LRS in positive bias, (e) HRS in negative bias and (f) HRS in positive bias with extracted temperature dependent activation energies and hopping distances for (g) LRS and (h) HRS.

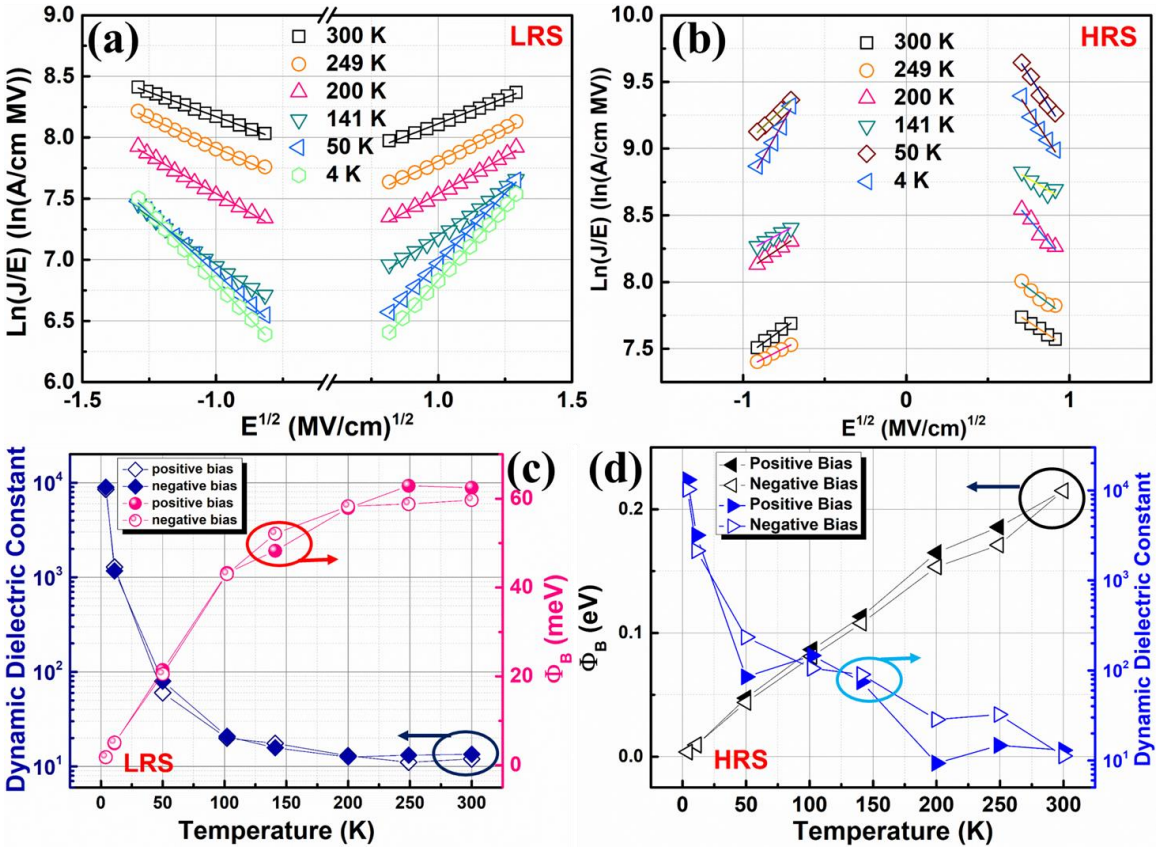


Fig. 3.3.5. High field I-V fittings by P-F emissions and extracted parameters with (a) and (c) for LRS and (b) and (d) for HRS, respectively.

Similarly, the high field I-V fittings are performed and shown in Figs. 3.3.5. Under high field, LRS and HRS I-V curves are fitting well with P-F emission mechanisms across the

wide range of temperatures. Meanwhile, dynamic dielectric constant and energy barriers have all been extracted and shown in Figs. 3.3.5 (c) and (d). The extracted dynamic dielectric constants for LRS and HRS at room temperature are all about 12 which is very close to the value reported for amorphous $HfO_2$. However, there is a gigantic differences for energy barriers between LRS and HRS. In LRS, the barrier is only 50 meV, which is quite close to its hopping activation energy. But in HRS, the barrier has increased to 0.2 eV which is almost 4 times higher than LRS. This difference indicating that the oxygen vacancy density differences are significant when resistance switched. The interesting behavior happened when dynamic constant increases exponentially with temperature decreases. Reason to observe such exponential temperature dependence most probably is because of the different numbers of polarization mechanisms involved at different temperatures. As we all know, the dielectric constants are calculated by one plus electrical susceptibility which is related to the dielectric polarization capabilities. At higher temperatures, the agitated vacancies as well as the GBs inside the amorphous oxide would suppress the polarization effects. Yet, at lower temperature, these defect sites are all deactivated, thus more polarization mechanisms would get involved.

Besides, it is also sound to observe the shift from hopping at low field to P-F emission at high field. At low field, electrons could only obtain limited energies from the field so that they could only hop between oxygen vacancies. But when the field is high enough, the energies gained under high field would be sufficiently high and together with the help of thermal energy, the electrons could then be emitted into either conduction band or the newly formed sub-band inside the oxide bandgap.

To summarize the transport mechanisms, a new comparison plots have been made and shown in Figs. 3.3.6. In Fig. 3.3.6 (a) and (b), the conduction scenarios for HRS and LRS have been presented with low field hopping between vacancy sites and high field emission into either conduction band or newly formed sub-band in the bandgap of the oxide [187]. In Fig. 3.3.6 (c) and (d), the extracted parameters have been plotted together for comparing the differences between HRS and LRS. During switching process, there might be localized density modulations happening in the oxide. However, the stability of such density modulation process are unknown.

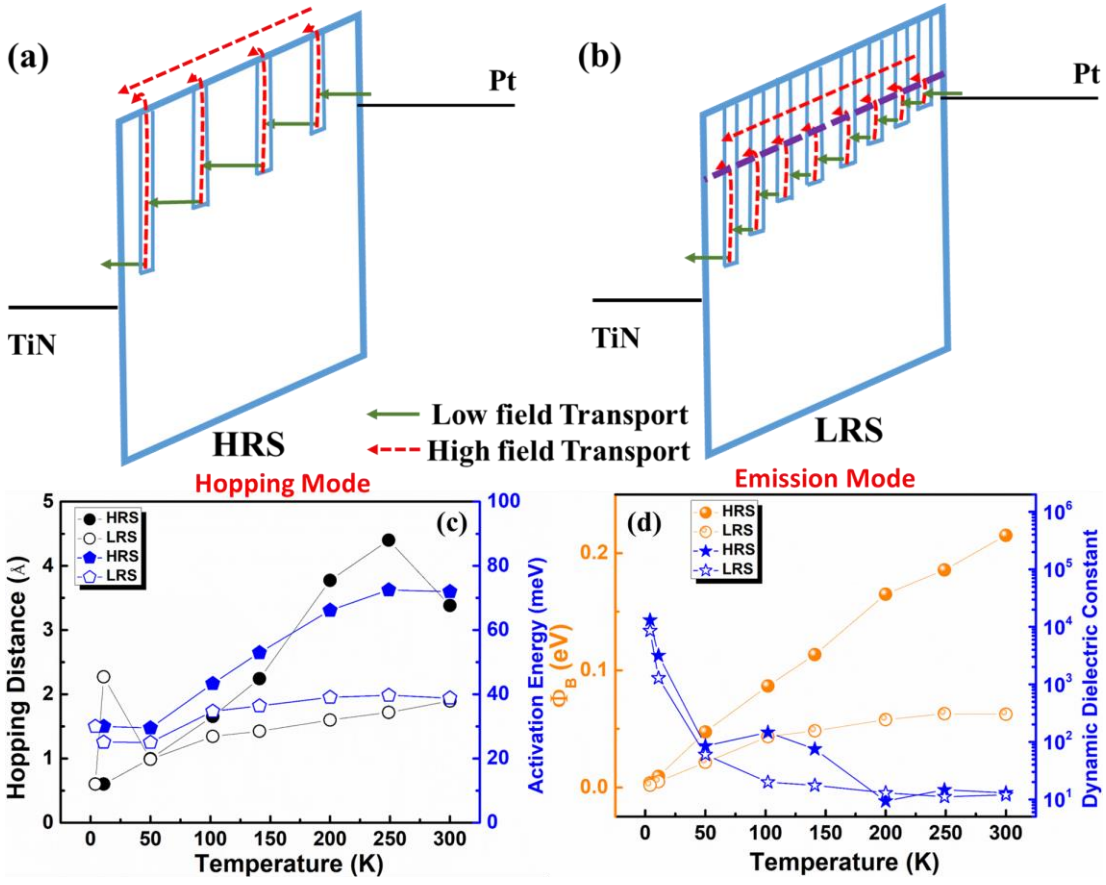


Fig. 3.3.6. Proposed conduction mechanisms for (a) HRS and (b) LRS and the extracted parameters in LRS and HRS for (c) hopping mode and (d) emission mode from fitting.

3.3.2. Resistance States Stability

The retention of each states would be one of the very critical metrics to decide whether specific technology is suitable for memory applications or not. In this part, the filament stability has been investigated with various means including retention studies, low temperature response [187] and γ-ray Irradiation response [188].

3.3.2.1. Resistance States Retentions

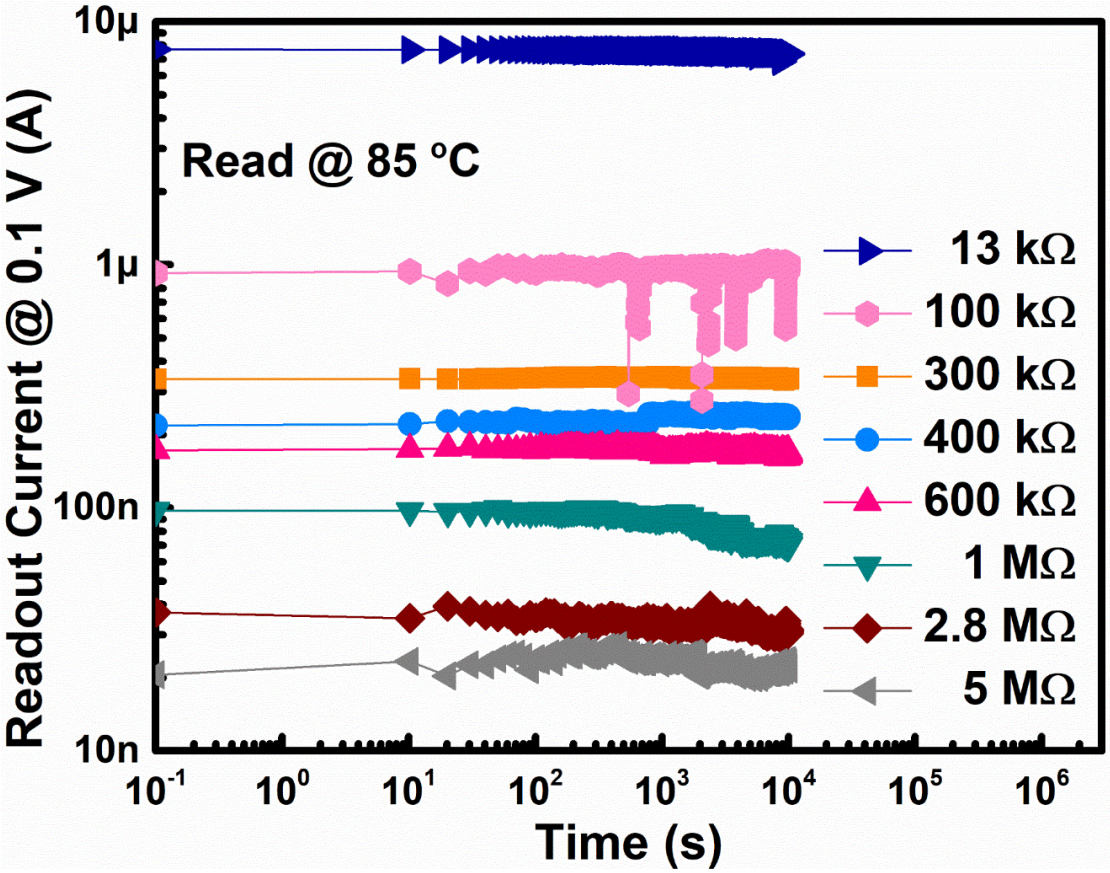


Fig. 3.3.7. The current-time (I-t) plot for 85 °C retention studies from various resistance states of RRAM devices with current sensed by a short 0.1 V pulse at each time step.

In Fig. 3.3.7, $10^4$ s retentions of the RRAM devices with different levels of resistance states have been assessed at 85 $^oC$. Various resistance states ranging from 13 kΩ to 5 MΩ have been deliberately programmed by different RESET stop voltages. According to the results, LRS has shown robust retention due to the strong filament. Meanwhile, the stabilities for very high resistance states such as 2.8 MΩ and 5 MΩ are also excellent though fluctuations have been observed. In addition, excellent stabilities have been observed in 300 kΩ, 400 kΩ and 600 kΩ resistance states together with some variations. However, for states of 100 kΩ and 1 MΩ, the retentions are poor. State with 100 kΩ has seen considerable fluctuations after 500 s. For 1 MΩ resistance state, the resistance kept very well in the early stage but starts to degrade towards even higher states after 2000 s. Though the other medium resistance states have not observed considerable fluctuations, they might not be suitable to be used in the real memory applications which weakens its capability in multi-level cell (MLC).

3.3.2.2. Low Temperature Response

Multiple devices were intentionally programmed into four different states including failure state, low resistance state, intermediate resistance state and high resistance state. The failure state means no longer able to switch after undergoing hard breakdown showing a resistance of 3 kΩ. Batches of devices with programmed to such four states were cooled down to 4 K and get resistance tracking during cooling. As shown in Figs. 3.3.8, the resistance of the device in failure state decreases with lowering temperature, indicating a metallic behavior. The metallic conduction has a linear dependence on temperature as expressed,

$$R(T) = R_0[1 + \beta(T - T_0)], \quad (4.1)$$

where $R_0$ is the resistance at temperature $T_0$ (300 K) and β is the temperature coefficient. Fitting were done with failure state in Fig. 3.3.8 (a). The result shows that β is about $4.2\times10^{-4}$ $K^{-1}$ which is smaller than the value ~$8\times10^{-4}$ $K^{-1}$ for sputtered Hf metal films [181]. This can be explained by the fact that the conductive filament is a highly reduced sub-oxide with a high density of oxygen vacancies created during the breakdown process, and electrons' extended waveform may overlap and form a metallic sub-band in the oxide thereby possibly forming metallic nanowire. It is noted that the resistance does not further decrease below 50 K, which is common for metals [182]. This effect is due to that the electron mean free path affected by static lattice imperfections becomes constant at very low temperature [182].

For the normal LRS (~20 kΩ), IRS (~100 kΩ) and HRS (~1 MΩ), the conductions all show a semiconducting behavior [Fig. 3.3.8 (b)-(d)]. All the resistance states were read 10 times with a 0.1 V DC voltage and the values shown here are the average values. In the log (R) vs. 1/kT relationship plots, resistances increase with lowering temperature (referred to as linear region) and tend to saturate below 77 K (referred to as saturation region). Considering the traps are mostly at mid-gap in the band diagram of the $HfO_x$ system ($x<2$), it is very difficult to excite electrons to the conduction band, thus the most probable conduction path is to hop via traps [184]. The linear region indicates a nearest neighboring hopping (NNH) [148] in which electrons could obtain sufficient thermal energy to hop to the nearest trap (as shown in the path 1 in Fig. 3.3.9). In the saturation region at very low temperature, the thermal energy is so small that electrons could only seek for longer hopping path to the farther traps, but these traps have lower energy barrier height (as shown

in path 2 of Fig. 3.3.9). Thus the conduction transits to the variable range hopping (VRH) [148]. This is in consistent with the small polaron hopping [185] in $HfO_x$ which shows a slope change in 1/kT plot at a transition temperature around half of the Debye temperature ~100 K [186]. All the reported phenomenon in this work is very similar with $TaO_x$ based RRAM devices [183].

An interesting observation in Fig. 3.3.8 (b)-(d) is that in the linear region, two distinct different slopes (or different activation energies) were found from the 1/kT fittings even for the same resistance level. The larger slope (or activation energy) is about 40 meV and the smaller one is about 4 meV. The two different slopes could be found at the same resistance level in the same device after different programming cycles. To avoid artifacts in the measurement, we checked the functionality of every device after the temperature is heated up to room temperature, and found that all the resistance values almost unchanged as compared to the values before the cooling down. This ensured that no resistance state-flipping or drifting during the cooling process. In addition, we repeated the cooling experiments several times. Every time before the cooling down, we preprogramed the resistances of the devices randomly to different states. It turned out that the same device with the same resistance state may exhibit different slopes, while different devices in different resistance states may exhibit similar slope. Such measurement protocols verified that the observed phenomenon is not the device-specific but the cycle-dependent.

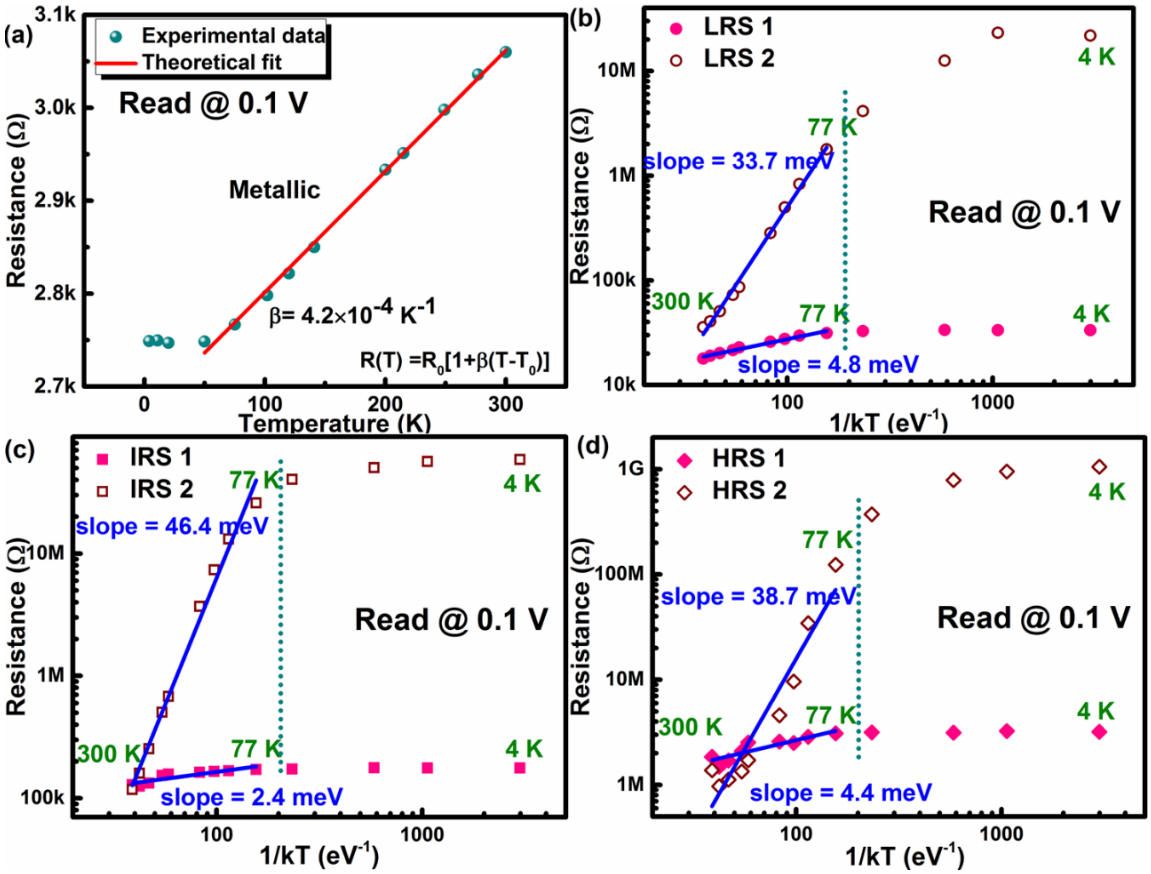


Fig. 3.3.8. (a) The temperature dependence of the failure state of a breakdown device and (b)-(d) the temperature dependence of resistance for the preprogrammed LRS, IRS and HRS of the normal devices. Different slopes were observed for the same states in different programming cycles (cycle 1 and 2), indicating the cycle-dependent variations of the activation energies of hopping barrier heights.

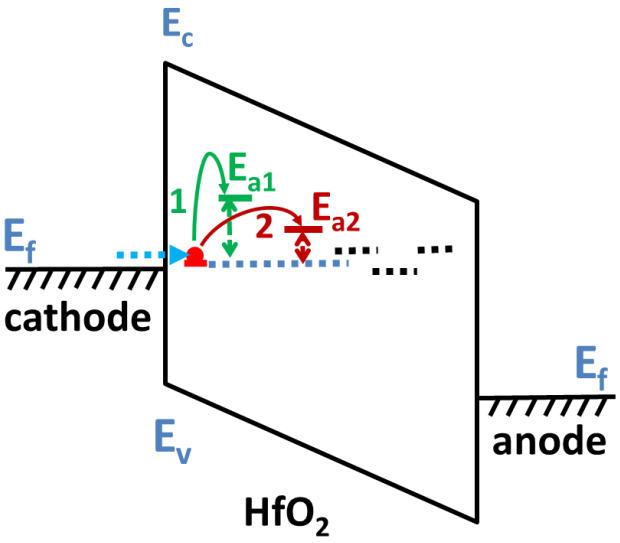


Fig. 3.3.9. Schematic of the possible hopping modes in Pt/HfOx/TiN stack and distribution of activation energies (path 1 indicates the nearest-neighboring hopping (NNH) from 300 K to 77 K and path 2 indicates the variable-range hopping (VRH) below 77 K).

Table 3.1.

The estimated hopping barrier ($E_a$)

| States | Et (meV) | qaE (meV) | Ea (meV) |
|---|---|---|---|
| LRS 1 | 4.8 | 1.3 | 6.1 |
| LRS 2 | 33.7 | 1.9 | 35.5 |
| IRS 1 | 2.5 | 6.3 | 8.7 |
| IRS 2 | 46.4 | 6.3 | 52.6 |
| HRS 1 | 4.4 | 15.0 | 19.4 |
| HRS 2 | 38.7 | 15.0 | 53.7 |

Note: $E_t$ is the extracted slope from Fig. 3.3.8 (b)-(d), qaE is the barrier lowering term due to the electric field, $E_a$ is the barrier height between two traps.

It is speculated that different programming cycles may result in different conductive filament configurations consisting of traps with different barrier heights. In the next, we extract the hopping barrier distribution from the data in Fig. 3.3.8 (b)-(d). The hopping conduction equation exploited was expressed in equation 3.7. For LRS1, LRS2, IRS and HRS, the hopping distance a in equation 4.2 is assumed to be 0.1 nm, 0.15 nm, 0.5 nm, and 1.2 nm respectively [184]. By this means, we could estimate a continuous distribution of the trap barriers ($E_a$) ranging from ~6 meV to ~54 meV as shown in Table 4.1.

The low temperature studies of the filament verified the hypothesized variable range hopping discussed in the I-V fittings across the wide range of temperatures. This consistency have indicated the trap configurations could be a source of variations during electrical bias switching. Since this variation source is intrinsic, it could also be called switching entropy which is temperature dependent.

3.3.2.3. γ-ray Irradiation Effects

Since filament are composed of oxygen vacancies which are neutral in the oxides, methods based on ionizing the vacancies with external energy sources have been proposed. The most common way to inject ionizing energies is to perform a total ionizing dose radiations in order to disturb the filament. As the best charge based storage memory, FLASH technology can only sustain a total ionizing dose (TID) up to 75 krad (Si) [189]. For the $HfO_x$-based RRAM, there were reports on X-ray [190] and the heavy ion [191] based TID effects but no reports on the γ-ray TID effect before this work. A batch of $HfO_x$ RRAM devices were exposed to $^{60}Co$ γ-rays (1.25 MeV) at a dose-rate of 557 rad($HfO_2$)/min in a Gammacell 220 irradiator to observe the step-radiation and total

ionizing dose effects. RRAM devices with a structure of TiN(50 nm)/$HfO_x$(10 nm)/Pt(50 nm) were fabricated at Stanford Nanofabrication Facility with three different device sizes and tested at ASU. X-ray photoelectron spectroscopy (XPS) was performed on pre- and post- irradiation samples with a Vacuum Generator, Escalab 220i-XL.

The typical DC sweep responses and pulse cycling of the $HfO_x$ based RRAM devices are shown in Figs. 3.3.10 with an active area of 5×5 $\mu m^2$. A positive pulse of 2 V/50 ns was applied to SET the devices to the LRS and a negative pulse of -2.7 V/50 ns was applied to RESET the devices to the HRS.

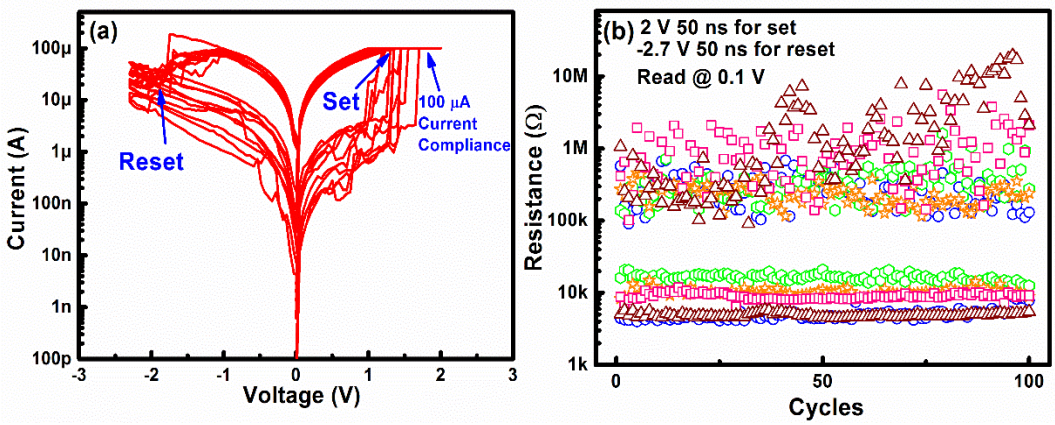


Fig. 3.3.10. (a) Typical I-V curves of the resistive switching behaviors of the TiN/$HfO_x$/Pt RRAM devices and (b) pulse cycling of five different devices with an area of 5×5 $\mu m^2$.

In order to examine the effects of radiation on the data retention, the resistance evolution was tracked on 60 samples with different active areas, 0.5×0.5 $\mu m^2$, 1×1 $\mu m^2$, and 5×5 $\mu m^2$ (20 samples for each size). Before irradiation, different values of resistances covering a full range from the LRS to the HRS (~4 kΩ to ~100 MΩ) were deliberately programmed to include different filament shapes or gaps (where the material is deficient in oxygen

vacancies). Fig. 3.3.11 (a) shows the evolution of resistance states of five different devices with an active area of 5×5 $\mu m^2$ along with increasing γ-ray dose. The samples were taken out of the radiation source to perform the resistance value check by applying a 0.1 V read voltage immediately (within 1.5 hours) after exposure. The devices were then put back in the radiation source. The resistance values remained almost constant after the 5.2 Mrad ($HfO_2$) dose except the high resistance values decreased slightly, nevertheless, the memory window is still well maintained. Fig. 3.3.11 (b) shows the resistance comparison of all the 60 samples between pre-irradiation and post-irradiation. It is seen that the resistance values before and after the exposure are almost the same (as indicated by a line with a slope of 1 in this graph), suggesting that the γ-ray irradiation does not affect the memory states, and thus the filament shape does not change significantly.

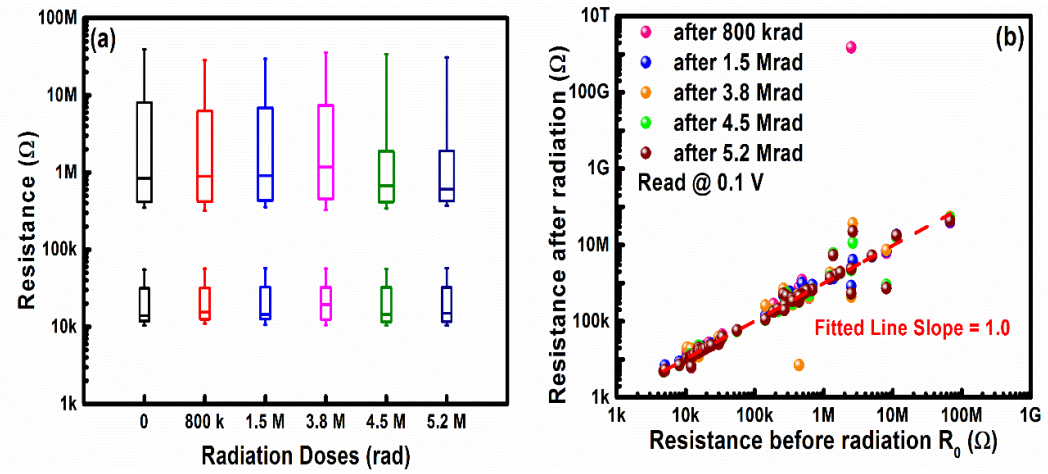


Fig. 3.3.11. (a) Resistance states evolution with different γ-ray doses for 20 samples of 5×5 $\mu m^2$ (the upper edge of the box shows the 75% of the distribution and lower edge shows the 25% of the distribution) (b) Resistance comparison before and after irradiation for all the 60 samples.

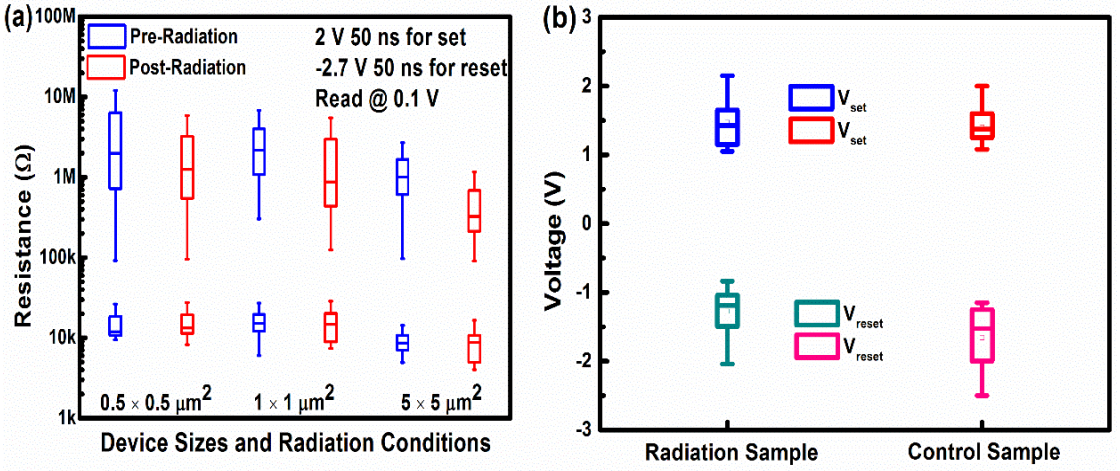


Fig. 3.3.12. (a) Statistics of resistance distribution by pulse cycling of the control samples and the irradiated samples. (b) The distribution of set voltage and reset voltage before and after irradiation. The statistics were obtained from devices with three different areas and each area includes 5 different devices (the upper edge of the box shows the 75% of the distribution, and lower edge shows the 25% of the distribution, and the square shows the mean value of the distribution).

After the 5.2 Mrad ($HfO_2$) γ-ray dose, the pulse cycling of the irradiated samples and the control samples was tested using the same pulsing condition (2 V/50 ns for SET and -2.7 V/50 ns for RESET). Statistics of the resistance distribution obtained for 100 pulse cycles for pre- and post-irradiation samples with three different active areas (5 cells in each size) are shown in Fig. 3.3.12 (a). If we compare the pre- and post-irradiation data for each size, the average HRS value decreases slightly after the exposure and the LRS value remains almost the same, indicating that more oxygen vacancies are generated in the non-filament region in the irradiated samples which may contribute to more leakage current in the HRS. If we compare the pre- and post-irradiation data for each size, the average HRS

value decreases slightly after the exposure and the LRS value remains almost the same, indicating that more oxygen vacancies are generated in the non-filament region in the irradiated samples which may contribute to more leakage current in the HRS. Moreover, we examined the DC performance of 15 devices after irradiation including the initial state, forming voltage, set voltage, and reset voltage distributions. The distribution of set voltage and reset voltage before and after irradiation is shown in Fig. 3.3.12 (b). The initial state, FORMING voltage slightly decrease, and SET voltage almost keeps the same after irradiation. However, RESET voltage decreased about 0.4 V after irradiation.

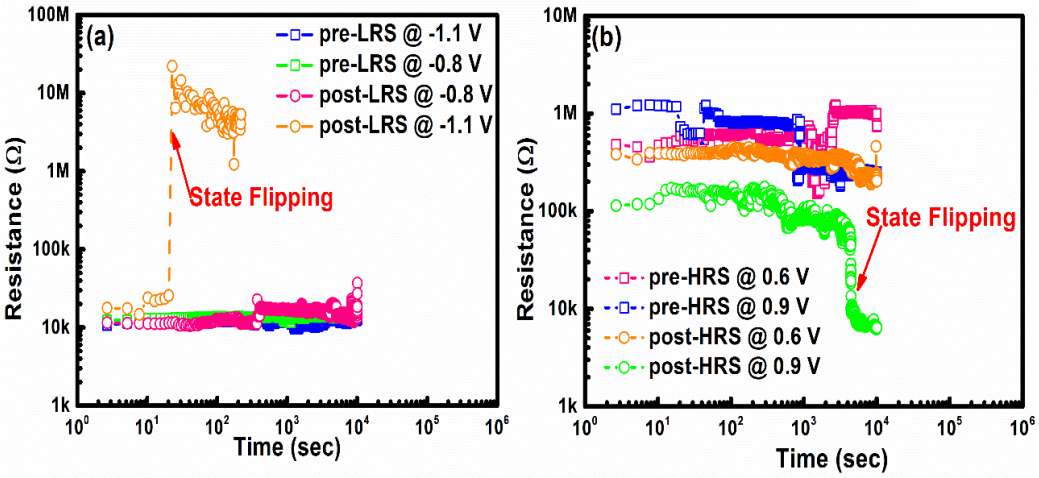


Fig. 3.3.13. The $10^4$ s electrical stress on the control samples and the radiation samples with constant bias of (a) -0.8 V and -1.1 V on the LRS and (b) 0.6 V and 0.9 V on the HRS.

In addition, the electrical stress response was measured and the result is shown in Fig. 3.3.13. For the control samples, the LRS can withstand 104 s with the highest constant bias of -1.1 V and the HRS can withstand 104 s with the highest constant bias of 0.9 V. However, for the irradiated samples, the HRS can be maintained for several thousand seconds only

under the constant bias of 0.9 V and then it switches to the LRS. The LRS could not even sustain several tens of seconds under the constant bias of -1.1 V until it switches to the HRS. This suggests that the filament stability is weakened after the radiation. This observation of unstable filament (especially in the LRS) at high voltage stress after γ-ray irradiation is consistent with the reset voltage reduction after γ-ray irradiation in Fig. 3.3.12 (b).

In order to gain more insight into the physical mechanism of γ-ray radiation effects on the $HfO_x$ RRAM devices, XPS was performed on the control and irradiated samples. The XPS depth sensitivity is about 5 nm, thus the information from the sample surface down to 5 nm deep region was collected. In Fig. 3.3.14 (a), the O 1s peaks show that the O-Hf bonds are located around 530 eV and other non-bridging oxygen bonds are located around 532 eV [192]. The area ratio of O-Hf bonds to non-bridging oxygen bonds changed from 7:1 to 2.6:1 which indicates that the amount of non-bridging oxygen at the surface of the $HfO_x$ has increased after irradiation. Fig. 3.3.14 (b) shows the Hf 4f peaks from the XPS. An obvious shift of 0.6 eV of the two Hf 4f half peaks 4f5/2 and 4f7/2 to higher energy level is observed, indicating the reduction of the valence of the Hf element [193]. The reduction in valence of the Hf element also agrees with the fact of O-Hf bond breaking.

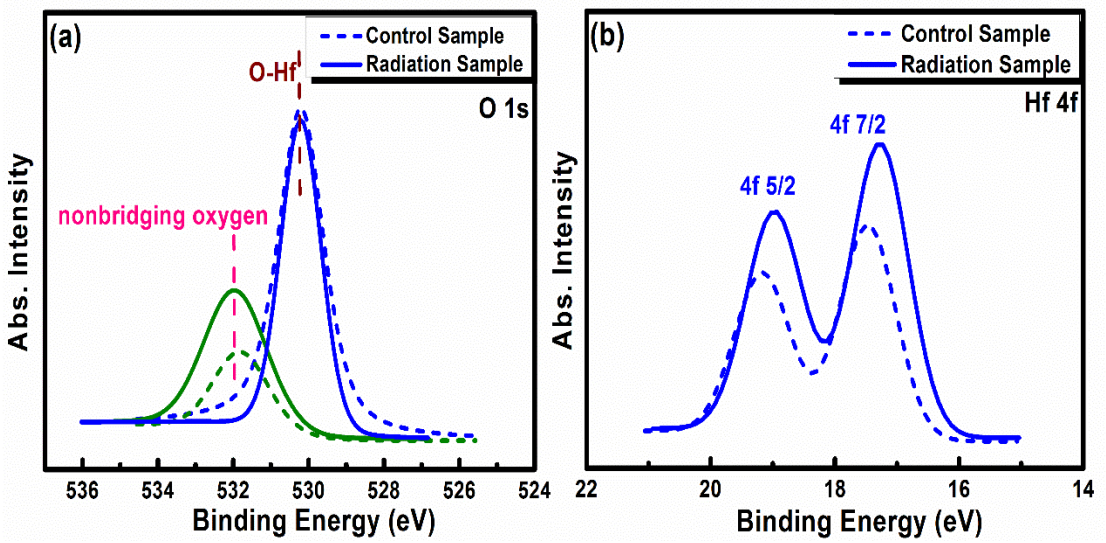


Fig. 3.3.14. The XPS spectrum of (a) O 1s and (b) Hf 4f peaks of the control and irradiated samples.

The key observations from the above experiments include that 1) the HRS and LRS resistance values do not change significantly after the γ-ray radiation, indicating the filament shape may not change as well; 2) the HRS resistance in the pulse cycling decreases slightly after the γ-ray irradiation, indicating more oxygen vacancies generated in the non-filament region; 3) the maximum voltage stress that the device can sustain and the reset voltage decrease after irradiation, indicating the stability of the filament becomes weakened; 4) more non-bridging oxygen bonds show up after irradiation, indicating that high energy γ-rays may break up the Hf-O bonds.

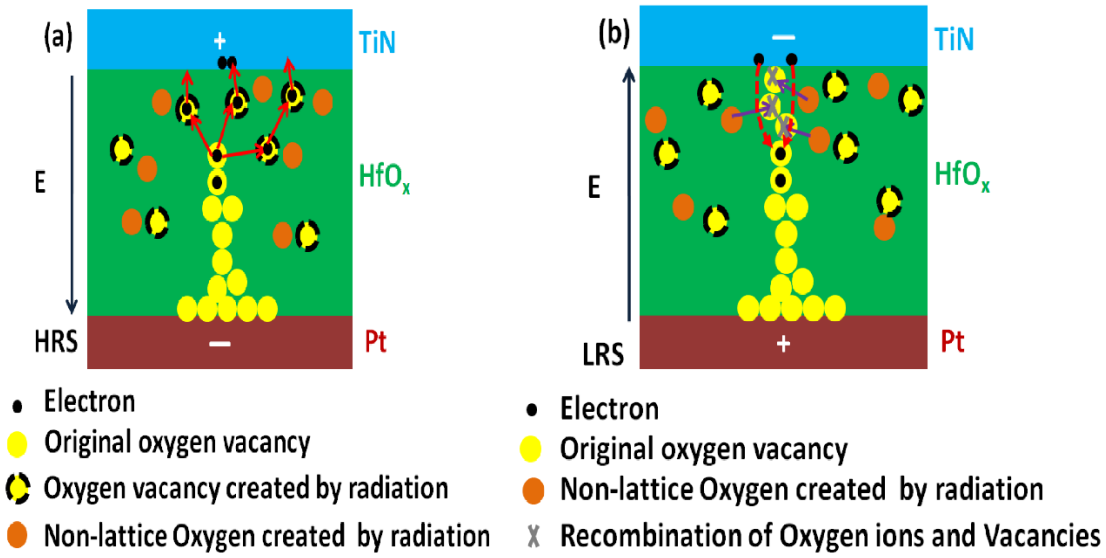


Fig. 3.3.15. Schematic of the physical mechanism that illustrates the γ-ray radiation effect on the RRAM cell in (a) the HRS and (b) the LRS. (Red line means the electron transport path and purple line means the oxygen migration path).

Based on these observations, we propose (in Fig. 3.3.15) a physical picture to illustrate how γ-rays exposures may affect $HfO_x$ RRAM characteristics. That is, high energy γ-rays break up some of the Hf-O bonds, which may result in the oxygen vacancies and non-lattice oxygen in interstitial sites. The shape of the filament in the LRS or the residual filament in the HRS is not affected noticeably as the amount of the density of newly generated oxygen vacancies is too low as compared to those in the filament region. Therefore, the LRS and HRS resistance states do not change significantly after radiation exposure. However, in the subsequent pulse cycling or DC sweep operations, there are more oxygen vacancies present in the bulk, thus the average HRS resistance decreases slightly as those oxygen vacancies far away from the filament region are not recovered during the reset process. Although the shape of the filament does not change substantially,

the stability of the filament in the LRS or the residual filament in the HRS is weakened. In the HRS, in the gap region, there are more oxygen vacancies that help to reconnect the filament under the set operation electrical stress (Fig. 3.3.15 (a)). In the LRS, there are even more non-lattice oxygen around the filament region that can migrate towards the filament to rupture the filament easily under the reset operation electrical stress (Fig. 3.3.15 (b)).

By far, the stability issues have been observed with both low temperature responses and γ-ray TID effects. Through the detailed examining of these issues, we have found the instabilities could be from either the internal trap configurations or the external high energy injections. The already formed oxygen vacancies might be activated under high temperature and deactivated under Debye temperature. These oxygen vacancies could not be eliminated by themselves alone but by recombining with oxygen ions nearby. Such recombination process requires very high energy injections like RESET process or high energy particles injections. Thus the stability of the filament is moderately guaranteed but the variability definitely would be the critical issues. In order to stabilize these filaments, we need to localize the switching spot and strengthen the other residues.

### 3.3.3. RRAM Switching Mechanism

The detailed switching mechanisms or physics need to start from the Forming process discussed in the last chapter. As discussed before, the Forming process is a dielectric breakdown process. To our knowledge, there could be many types of dielectric breakdown mechanisms. The mechanisms applicable here might be limited to only three: 1) intrinsic or electrical breakdown; 2) thermal breakdown; 3) electro-mechanical breakdown [207, 208]. The reason to limit the breakdown mechanisms within these three is because of their

time frames. All these three breakdown mechanisms would complete within millisecond and intrinsic breakdown would even within microsecond or nanoseconds [207]. Thus we attribute either the aforementioned bandgap impact ionization (BII) or the surface plasma oscillation (SPO) into the intrinsic breakdown scope which should be within nanoseconds. However, intrinsic breakdown process would not have the ability to form a robust percolation path between electrodes within nanoseconds. Thus many works attribute the growing processes of the percolation paths to the thermal breakdown in the TDDB works [155-158]. Certainly the external electrical and thermal environment during breakdown would also influence the robustness of the paths formed, but that could not be the reason to exclude the intrinsic breakdown. Basically, the intrinsic breakdown process is supposed to supply positive charges as mentioned in the TDDB works for the following thermal breakdown processes [176, 177]. Since the supply is from anode to cathode, the filament or percolation paths should be formed in the same direction. During this process, Hf-O bonds would be broken with oxygen ions released and oxygen vacancies generated. But once the filament are fully formed, there would be a fluence of electrons flowing through the breakdown region when all the defects, including oxygen ions and oxygen vacancies, would get neutralized. Such neutralization process would be counted into the dielectric relaxation processes since there must be polarization during Forming.

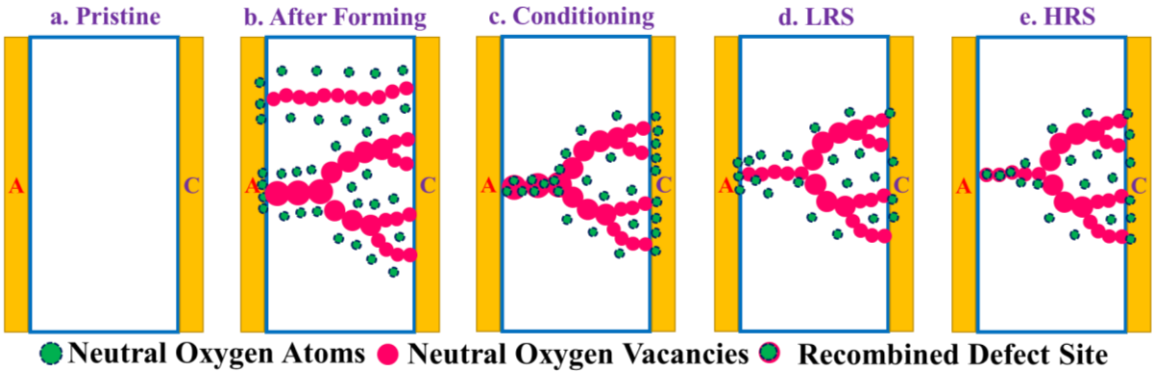


Fig. 3.3.16. The switching model and physics of RRAM devices from pristine state to steady switching.

As indicated in Fig. 3.3.16 (b), after Forming, the percolation paths might be too strong to achieve steady switching. The paths either require high voltage to rupture or high voltage to form the new paths. Thus conditioning processes are imperative to gradually stabilize the filament and localize the switching spot little by little. During conditioning switchings, the operation voltages for SET and RESET are reduced manually until stabilized or localized. In Fig. 3.3.16 (c), the strong filaments close to anode have been ruptured or recombined by the oxygens nearby after the conditioning processes while the residual filaments are remaining the same. It is worth noticing here that the potential location that neutral oxygen atoms could stay inside the dielectrics probably is the grain boundaries (GB) that is very close to the filaments [209]. But most of the recombination processes are not happening inside the oxides but very close to the anode/oxide interface. The reason to claim this is because RESET process is also a positive feedback process.

According to the oxygen atoms configurations in Fig. 3.3.16 (d), since there is a conductive filament connecting the electrodes, the voltage across the insulating regions are limited so is the electric field. At the anode/oxide interface, the oxygen atoms would be

ionized first because during RESET, anode is applied negative bias where electrons are injected. Once ionized, oxygen ions would move back to oxides to recombine with the vacancies nearby in a very short time. Meanwhile, since the ion drift speed equals to field times mobility and the field at this time is not high enough, the oxygen ions inside the oxides is moving very slowly. But once the interface vacancies get recombined, the field would increase immediately as well as the drift speed. This is why the gradual reset current was observed first when interface vacancies are getting recombined and then sharp decrease was observed when bulk vacancies are recombined. Theoretically, the dielectrics after breakdown could be fully recovered if the cathode is not involving in the reaction processes.

As could be seen in Fig. 3.3.16 (c)-(e), there are also some oxygen atoms in the cathode/oxide interface which is due to the pulling out of the oxygen ions very close to cathode by the electric field. Thus, SET and RESET process is like a competition process. That is also the reason usually the cathode is an inert metal [171, 210]. For SET process, it is much like the Forming process, but it is a compact version of Forming.

### 3.3.4. RRAM Architectures and Applications

The studies above have profoundly discussed the performances of the RRAM cells in various ways. Now in this section, various high-level architectures would be briefly discussed in this section. The cell architectures are supposed to be repeated to form the memory arrays or pages. These include one RRAM (1R) [194], one diode one RRAM (1D1R) [195], one transistor several RRAM (1TXR) [196], one selector one RRAM (1S1R)

[197], complementary resistive switching (CRS) [198], and three-dimensional RRAM (3D RRAM) [199] cells.

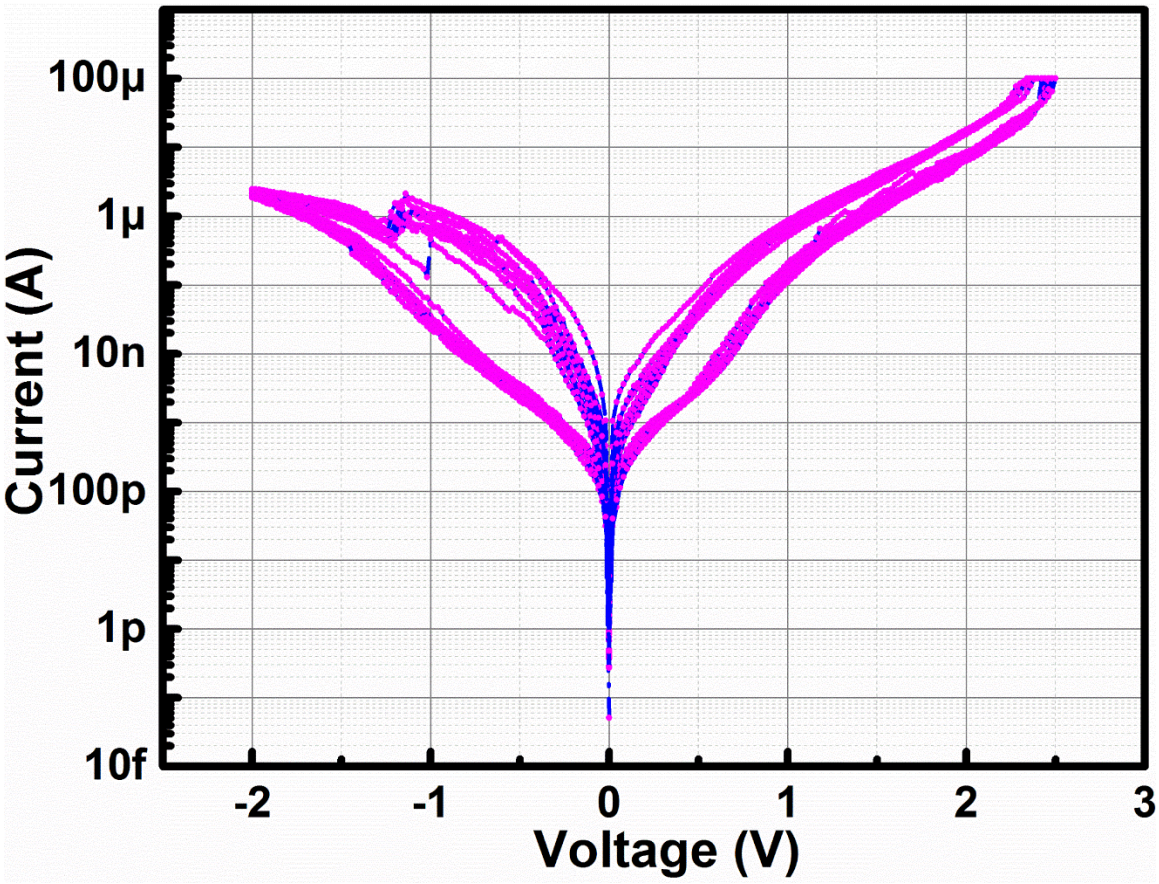


Fig. 3.3.17. I-V sweeps of Pt/$NbO_x$/Pt based RRAM cells with high non-linearity.

Among all these structures, 1TXR has been used by most of the companies in the production. But this type of cell would consume large areas so the volume of the storages would not be large enough. Meanwhile, 1D1R has shown its limitations for bipolar RRAM, CRS has problems with probabilities of which cell get switched, 3D RRAM has faced fabrication challenges, and 1R might face either energy consumption problems or sneak path problem. To solve the sneak path problem [200], 1S1R cells have been proposed in

which selector is a bi-directional diode. Typically, there are two critical voltages for selector devices. One is the turn on voltage ($V_t$) meaning at which the selector could be turned on. The other one is the hold voltage ($V_h$) meaning below which the selector would be turned off. To turn on the selector, a voltage of $V_t$ is necessary, but only $V_h$ is required to keep selector in on-state when it is already on. Therefore, the trade-offs for selector appears, whether a large $V_h$ is good or a small value is preferred. For the Write/Erase operations, we need $V_h$ to be as small as possible. Once selector is on, we need as much voltage as possible to be across RRAM. However, if we would like to eliminate the sneak path problem, $V_h$ has to be as large as possible. There might be a global optima for the $V_h$ value we believe, but it would raise a material level question again. Thus the questions have been pushed back to use only 1R cell which could show very high non-linearity or self-rectifying characteristics as shown in Fig. 3.3.17.

Other than memory applications, some other ways to exploit RRAM devices have also been proposed by research scholars such as neuromorphic computation components such neural network hardware [201], physically unclonable functions [202], and random number generator [203]. But all these proposals are still under investigations.

## 3.4. Summary

In this chapter, dielectric breakdown has been profoundly discussed. At first, the CVS time frame have been discussed and presented with current response. Different current conduction mechanisms including both bulk-limited ones and electrode-limited ones have been briefly discussed with many of the quantitive equations given. After that, the stress studies on $SiO_2$ and $HfO_2$ have been presented separately with potential defects in the gate

stacks summarized. With our own stress studies on the MOS capacitors (SILC and TDDB), degradation mechanisms in the gate stacks have been discussed according to the bias regime. Substrate injection and gate injections are further discussed in the view of energies and band diagrams. Lifetime predictions have provided views on why NMOS is more robust than PMOS. Then the RVS induced dielectric breakdown studies by the RRAM device Forming process have been performed and discussed. Different thickness of $HfO_2$ have been used to form the RRAM devices and the Forming sweep measurements. By analyzing the Forming I-Vs and their Weibull statistics, synergies between RRAM Forming and MOS capacitor breakdown have been discovered. With RRAM Forming process originated from the Cathode Injections induced dielectric breakdown process, substrate injection has been compared with Cathode Injection. The conclusion then was reached that cathode injection is almost same as substrate injection which are all triggered by either bandgap impact ionization in the $HfO_2$ (high energy electrons) or surface plasma oscillations at the anode/$HfO_2$ interface (medium energy electrons) that are followed by anode hole injection (AHI) like effects leading to the breakdown of the entire dielectrics via the moving and trapping process of the hot holes. Finally, the post breakdown resistance switching behaviors have been presented by measuring the RRAM device performances after Forming. Basic electrical behaviors of resistance switching in TiN/$HfO_x$/Pt RRAM devices are presented with typical I-V curves, DC cycling and studies on LRS and HRS. Variable range hopping was observed in the TiN/$HfO_x$/Pt based RRAM devices in LRS and HRS. Then the stability of the cell has also been investigated by retention study, low temperature response and radiation response. All the studies have shown that the configuration of traps and external energies (including electrical bias, temperature and high

energy particles) would be the threats to the resistance stabilities. After briefly discussion on the potential applications of RRAM devices, the switching mechanisms and modeling approaches have been presented. RRAM devices have been studied for decades where many companies have joined in developing storage or memory products based on RRAM. However, by far, none of them could achieve sufficiently reliable performances from these products thus leaving fewer and fewer players in this area. This probably is due to the instability of the filaments or cells that have been observe in our studies. Hopefully, one day, such issue could be addressed with some new technologies to commercialize RRAM products.

# CHAPTER 4

# CONCLUSION AND OUTLOOK

## 4.1. Summary of Contributions

This dissertation has discussed the defects in the high-k dielectrics including the origins, the generations and their own behaviors. Additionally, by investigations of the phenomenon due to these defects, both CMOS reliabilities and RRAM devices have been studied. The contributions of this dissertation include:

1. Via modeling the charge trapping behaviors in high-k dielectrics and employing the non-iterative surface potential equations, the predictive and physical model for BTI effects have been successfully implemented in circuit-level for circuit lifetime prediction. Although the BTI effects in circuits have been discovered negligible through our circuit level simulations, the modeling work is quite meaningful and solid. Meanwhile we first time pointed out the question that whether the accelerated testing methods could be used to predict BTI lifetime, since more external energies might trigger other degradation mechanisms.
2. Starting from the transport behavior of the carriers in MOS capacitors and summarizing the potential defects in gate stacks, the physical scenario and theory on high-k dielectric breakdown has been envisioned and developed, respectively. By stressing the MOS capacitors with ultra-thin dielectrics fabricated at ASU Nano Fab, two different degradation and breakdown mechanisms have been proposed. With gate injection, the $HfO_2/SiO_2$ interface will degrade and breakdown first. But with substrate injection, metal/HfO2 interface will degrade and breakdown first. These two mechanisms are all related with the as-grown defects in the gate stack.

3. Investigation on Forming of RRAM devices, the dielectric breakdown theory in MIM structures have been put forward. The breakdown mechanism is somewhat similar with the substrate injection induced breakdown but amended by cathode injection induced breakdown. According to the fittings and statistics of the Forming I-V, we proposed either the bandgap impact ionization or surface plasma oscillation would trigger the generations of electron-hole pairs in the dielectrics. The further breakdown of the whole dielectrics, in which a conductive filament will be formed from anode to cathode, are directed by the movement and trapping of the holes inside the oxide.
4. Works and analyses on the performances of the TiN/$HfO_x$/Pt based RRAM devices in various ways such as filament transport, filament stability and potential applications have improved the understanding of the resistive switching phenomenon. With the low temperature studies of either transport behavior or switching studies, we first time disclosed the transport in filament is with trap energy influence which should depend on the trap density, electrical field and ambient temperatures. The observed differences in trap configurations between LRS and HRS has rendered us to attribute the RRAM switching mechanisms to be electrical modulation of trap configurations.

## 4.2. Future Works and Outlook

This dissertation presents a comprehensive study which provides solid foundation for more exciting future work in several directions.

With respect to the percolation path model, the circuit design or prediction of the RRAM devices could be performed by turning on the entropic effects in the model as well as tweaking the parameters. To achieve this goal, the model has to be implemented in Verilog-A form in SPICE simulator to enable circuit level simulations.

With respect to the dielectric breakdown theories, the temperature dependencies are needed to make the model more solid.

With the noise and variations in the RRAM devices, the model to consider random telegraph noise (RTN) are imperative. More improvement works are needed to obtain high performance Forming free RRAM devices with high non-linearity. This might request more efforts in the materials explorations.

With more efforts spent on high-k dielectrics, it might be possible to construct a unified and physic model to predict most of the reliability effects happened in the MOSFETs as well as capturing the RRAM devices behaviors with different material systems.

REFERENCES


[1] R. Courtland, "Moore's law's next step: 10 nanometers," *IEEE Spectrum*, vol. 54, pp. 52-53, 2017.

[2] J. G. Pellerin, "Dimensions of innovation to enable the next era of intelligent systems," *2017 IEEE Electron Devices Technology and Manufacturing Conference (EDTM)*, pp. 1, 2017.

[3] E. Track, N. Forbes and G. Strawn, "The End of Moore's Law," *Computing in Science & Engineering*, vol. 19, pp. 4-6, 2017.

[4] T. N. Theis and H. S. P. Wong, "The End of Moore's Law: A New Beginning for Information Technology," *Computing in Science & Engineering*, vol. 19, pp. 41-50, 2017.

[5] E. P. DeBenedictis, "It's Time to Redefine Moore's Law Again," *Compute*r, vol. 50, pp. 72-75, 2017.

[6] M. T. Bohr and I. A. Young, "CMOS Scaling Trends and Beyond," *IEEE Micro*, vol. 37, pp. 20-29, 2017.

[7] G. Baccarani, M. R. Wordeman and R. H. Dennard, "Generalized scaling theory and its application to a ¼ micrometer MOSFET design," *IEEE Transactions on Electron Devices*, vol. 31, pp. 452-462, 1984.

[8] S. V. Hattangady, H. Niimi, and G. Lucovsky, “Nitrogen engineering of ultrathin oxynitrides by a thermal NO/$O_2$/NO process,” *Appl. Phys. Lett.*, vol. 66, pp. 3495, 1995.

[9] G. Lucovsky and J. L. Whitten, “High Dielectric Constant Materials: VLSI MOSFET Applications,” vol. 11, pp. 315, Springer-Verlag, 2005.

[10] A. Chatterjee et al., "CMOS metal replacement gate transistors using tantalum pentoxide gate insulator," *International Electron Devices Meeting 1998. Technical Digest*, pp. 777-780, 1998.

[11] S. J. Wang, I. C. Chen and H. L. Tigelaar, "Effects of poly depletion on the estimate of thin dielectric lifetime," *IEEE Electron Device Letters*, vol. 12, pp. 617-619, Nov. 1991.

[12] D. James, "From Strain to High-K/Metal Gate - the 65 - 45 nm Transition," *2008 IEEE/SEMI Advanced Semiconductor Manufacturing Conferenc*e, pp. 76-81, 2008.

[13] K. J. Kuhn, "CMOS scaling for the 22 nm node and beyond: Device physics and technology," *Proceedings of 2011 International Symposium on VLSI Technology, Systems and Applications*, pp. 1-2, 2011.

[14] E. Karl et al., "A 0.6 V 1.5 GHz 84 Mb SRAM design in 14 nm FinFET CMOS technology," *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, pp. 1-3, 2015.

[15] L. C. Cho et al., "A 4 GHz clock distribution architecture using subharmonically injection-locked coupled oscillators with clock skew calibration in 16 nm CMOS," *2017 Symposium on VLSI Circuits*, pp. C130-C131, 2017.

[16] Sujay B. Desai et al., "$MoS_2$ transistors with 1-nanometer gate lengths," *Science*, vol. 354, pp. 99-102, 2016.

[17] M. De Marchi et al., "Top–Down Fabrication of Gate-All-Around Vertically Stacked Silicon Nanowire FETs with Controllable Polarity," *IEEE Transactions on Nanotechnology*, vol. 13, no. 6, pp. 1029-1038, Nov. 2014.

[18] J. P. Colinge, "3D transistors," *2013 International Symposium on VLSI Technology, Systems and Application (VLSI-TSA)*, pp. 1-2, 2013.

[19] J. J. Gu et al., "III-V gate-all-around nanowire MOSFET process technology: From 3D to 4D," *2012 International Electron Devices Meeting*, pp. 23.7.1-23.7.4, 2012.

[20] Y. Liu et al., "A 65 nm ReRAM-enabled nonvolatile processor with 6x reduction in restore time and 4x higher clock frequency using adaptive data retention and self-write-termination nonvolatile logic," *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, pp. 84-86, 2016.

[21] J. r. Shih, et al., "A new phenomenon of retention time evolution in embedded DRAM technology with high-k dielectrics ($Ta_2O_5$) MIM capacitor after HTOL test," *2006 IEEE International Reliability Physics Symposium Proceedings*, pp. 703-704, 2006.

[22] R. f. Tsui et al., "Reliability assessment of the embedded DRAM technology with PMOSFET transfer transistor and high-k dielectrics ($Ta_2O_5$) MIM capacitor," *2006 International Symposium on VLSI Technology, Systems, and Applications*, pp. 1-2, 2006.

[23] Kee-Won Kwon et al., "Thermally robust $Ta_2O_5$ capacitor for the 256-Mbit DRAM," *IEEE Transactions on Electron Devices*, vol. 43, pp. 919-923, 1996.

[24] M. A. Pawlak et al., "Towards 1x DRAM: Improved leakage 0.4 nm EOT STO-based MIMcap and explanation of leakage reduction mechanism showing further potential," *2011 Symposium on VLSI Technology - Digest of Technical Papers*, pp. 168-169, 2011.

[25] G. Giusi et al., "Impact of Electrode Composition and Processing on the Low-Frequency Noise in $SrTiO_3$ MIM Capacitors," *IEEE Electron Device Letters*, vol. 35, pp. 942-944, 2014.

[26] A. Chin et al., "Fast write time and long retention 1T memory," *Device Research Conference. Conference Digest*, pp. 18-19, 2001.

[27] T. Boscke et al., "Investigation of the high temperature stability of TiN-$Al_2O_3$-TiN capacitors for sub 50 nm deep trench DRAM," *2006 European Solid-State Device Research Conference*, pp. 391-394, 2006.

[28] S. Q. Wang and J. W. Mayer, "Reactions of Zr thin films with $SiO_2$ substrates," *J. Appl. Phys.*, Vol. 64, pp. 4711-4713, 1988.

[29] S. Jeon, H. Hwang, "Effect of hygroscopic nature on the electrical characteristics of lanthanide oxides ($Pr_2O_3$, $Sm_2O_3$, $Gd_2O_3$, and $Dy_2O_3$)," *J. Appl. Phys.*, Vol. 93, pp. 6393-6395, 2003.

[30] Y. Zhao, "Design of Higher-*k* and More Stable Rare Earth Oxides as Gate Dielectrics for Advanced CMOS Devices," *Materials*, vol. 5, pp. 1413–1438, 2013.

[31] J. Robertson, "High dielectric constant oxides," *Eur. Phys. J. Appl. Phys.*, vol. 28, pp. 265-291, 2004.

[32] M. Leskelä, M. Ritala, "Atomic layer deposition (ALD): from precursors to thin film structures," *Thin Solid Films*, Vol. 409, pp. 138-146, 2002.

[33] Steven M. George, "Atomic Layer Deposition: An Overview," *Chemical Reviews*, vol. 110, pp. 111-131, 2010.

[34] C. Marichy, M. Bechelany, and N. Pinna, "Atomic Layer Deposition of Nanostructured Materials for Energy and Environmental Applications," *Adv. Mater.*, vol. 24, pp. 1017–1032, 2012.

[35] M. Leskelä, and M. Ritala, "Atomic Layer Deposition Chemistry: Recent Developments and Future Challenges," *Angewandte Chemie International Edition*, vol. 42, pp. 5548–5554, 2003.

[36] C. T. Kuo, R. Kwor, K.M. Jones, "Study of sputtered HfO2 thin films on silicon," *Thin Solid Films*, vol. 213, pp. 257-264, 1992.

[37] M. A. Sahiner et al., "Pulsed laser deposition and characterization of Hf-based high-k dielectric thin films," *Thin Solid Films*, vol. 515, pp. 6548-6551, 2007.

[38] Erli Chen, Thin Film Deposition, Slides for Applied Physics 298r in Harvard University, Available: http://www.mrsec.harvard.edu/education/ap298r2004/Erli%20chenFabrication%20II%20-%20Deposition-1.pdf, 2004.

[39] S. J. Lee et al., "High quality ultra-thin CVD $HfO_2$ gate stack with poly-Si gate electrode," *International Electron Devices Meeting 2000. Technical Digest. IEDM*, pp. 31-34, 2000.

[40] G. D. Wilk, R. M. Wallace, J. M. Anthony, "High-k gate dielectrics: Current status and materials properties considerations," *J. Appl. Phys.*, Vol. 89, pp. 5243- 5275, 2001.

[41] J. F. Zhang, S. Taylor, and W. Eccleston, "A quantitative investigation of electron detrapping in $SiO_2$ under Fowler–Nordheim stress," *J. Appl. Phys.*, Vol. 71, pp. 5989- 5996, 1992.

[42] K. Joshi et al., "HKMG process impact on N, P BTI: Role of thermal IL scaling, IL/HK integration and post HK nitridation," *2013 IEEE International Reliability Physics Symposium (IRPS)*, pp. 4C.2.1-4C.2.10, 2013.

[43] C. H. Wu, et al., "Bias Temperature Instabilities for High-κ $HfO_2$ LTPS-TFTs with Dual Plasma Treatment," *Journal of Display Technology*, vol. 12, pp. 232-239, 2016.

[44] E. Rosenbaum and L. F. Register, "Mechanism of stress-induced leakage current in MOS capacitors," *IEEE Transactions on Electron Devices*, vol. 44, pp. 317-323, 1997.

[45] W. Liu et al., "Analysis of Correlated Gate and Drain Random Telegraph Noise in Post-Soft Breakdown TiN/HfLaO/$SiO_x$ nMOSFETs," *IEEE Electron Device Letters*, vol. 35, pp. 157-159, 2014.

[46] K. B. Yeap et al., "A Realistic Method for Time-Dependent Dielectric Breakdown Reliability Analysis for Advanced Technology Node," *IEEE Transactions on Electron Devices*, vol. 63, pp. 755-759, 2016.

[47] M. S. Liang et al., "Hot carriers induced degradation in thin gate oxide MOSFETs," *1983 International Electron Devices Meeting*, pp. 186-189, 1983.

[48] S. Dwivedi and G. Saini, "Effect of random dopant fluctuation in nanoscale junctionless FinFET using low and high-k spacers," *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1-4, 2017.

[49] X. Jiang et al., "Investigations on the correlation between line-edge-roughness (LER) and line-width-roughness (LWR) in nanoscale CMOS technology," *2012 IEEE 11th International Conference on Solid-State and Integrated Circuit Technology (ICSICT)*, pp. 1-3, 2012.

[50] M. Bajaj et al., "Effect of Metal Gate Granularity Induced Random Fluctuations on Si Gate-All-Around Nanowire MOSFET 6-T SRAM Cell Stability," *IEEE Transactions on Nanotechnology*, vol. 15, pp. 243-247, 2016.

[51] J. G. Simmons and R. R. Verderber, "New conduction and reversible memory phenomena in thin insulating films," *Proc. R. Soc. Lond. A*, vol. 301, pp. 77-98, 1967.

[52] J. G. Simmons and R. R. Verderber, "New thin-film resistive memory," *Radio and Electronic Engineer*, vol. 34, pp. 81-89, 1967.

[53] X. A. Tran et al., "A High-Yield $HfO_x$-Based Unipolar Resistive RAM Employing Ni Electrode Compatible With Si-Diode Selector for Crossbar Integration," *IEEE Electron Device Letters*, vol. 32, pp. 396-398, 2011.

[54] A. Prakash et al., "Demonstration of Low Power 3-bit Multilevel Cell Characteristics in a $TaO_x$-Based RRAM by Stack Engineering," *IEEE Electron Device Letters*, vol. 36, pp. 32-34, 2015.

[55] J. Park et al., "New Set/Reset Scheme for Excellent Uniformity in Bipolar Resistive Memory," *IEEE Electron Device Letters*, vol. 32, pp. 228-230, 2011.

[56] B.J. Choi et al., "Nitride memristors," *Applied Physics A*, vol. 109, pp. 1-4, 2012.

[57] Y.-H. Huang et al., "Forming-free, bi-directional polarity conductive-bridge memory devices with $Ge_2Sb_2Te_5$ solid-state electrolyte and Ag active electrode," *J. Appl. Phys.*, Vol. 117, pp. 014505, 2015.

[58] I. Valov et al., "Electrochemical metallization memories—fundamentals, applications, prospects," *Nanotechnology*, vol. 22, pp.25403, 2011.

[59] M. N. Kozicki, Mira Park and M. Mitkova, "Nanoscale memory elements based on solid-state electrolytes," *IEEE Transactions on Nanotechnology*, vol. 4, pp. 331-338, 2005.

[60] U. S. Bhansali et al., "Metal-Free, Single-Polymer Device Exhibits Resistive Memory Effect," *ACS Nano*, vol. 7, pp.10518-10524, 2013.

[61] D. I. Son et al., "Flexible Organic Bistable Devices Based on Graphene Embedded in an Insulating Poly(methyl methacrylate) Polymer Layer," *Nano Letters*, vol. 7, pp.2441-2447, 2010.

[62] K. Krishnan et al., "Mechanism for Conducting Filament Growth in Self-Assembled Polymer Thin Films for Redox-Based Atomic Switches," *Adv. Mater.*, vol. 28, pp.640–648, 2016.

[63] H. S. P. Wong et al., "Metal–Oxide RRAM," *Proceedings of the IEEE*, vol. 100, pp. 1951-1970, 2012.

[64] R. Waser and M. Aono, "Nanoionics-based resistive switching memories," *Nature Materials*, vol. 6, pp. 833, 2007.

[65] T. Hasegawa et al., "Atomic Switch: Atom/Ion Movement Controlled Devices for Beyond Von-Neumann Computers," *Adv. Mater.*, vol. 24, pp. 252–267, 2012.

[66] W. Lu and C. Lieber, "Nanoelectronics from the bottom up," *Nature Materials*, vol. 6, pp. 841, 2007.

[67] M. N Kozicki and H. J Barnaby, "Conductive bridging random access memory—materials, devices and applications," *Semiconductor Science and Technology*, vol. 31, pp.113001, 2016.

[68] D. Ielmini et al., "Resistive switching memories based on metal oxides: mechanisms, reliability and scaling," *Semiconductor Science and Technology*, vol. 31, pp.063002, 2016.

[69] H. Akinaga and H. Shima, "Resistive Random Access Memory (ReRAM) Based on Metal Oxides," *Proceedings of the IEEE*, vol. 98, pp. 2237-2251, 2010.

[70] Y. Hou et al., "Sub-10 nm low current resistive switching behavior in hafnium oxide stack," *Appl. Phys. Lett.*, vol. 108, pp. 123106, 2016.

[71] F. De Stefano et al., "Semiconducting-like filament formation in TiN/$HfO_2$/TiN resistive switching random access memories," *Appl. Phys. Lett.*, vol. 100, pp.142102, 2012.

[72] J. Woo et al., "Optimized Programming Scheme Enabling Linear Potentiation in Filamentary $HfO_2$ RRAM Synapse for Neuromorphic Systems," *IEEE Transactions on Electron Devices*, vol. 63, pp. 5064-5067, 2016.

[73] T. W. Lee and J. H. Nickel, "Memristor Resistance Modulation for Analog Applications," *IEEE Electron Device Letters*, vol. 33, pp. 1456-1458, 2012.

[74] M.-J. Lee et al., "A fast, high-endurance and scalable non-volatile memory device made from asymmetric $Ta_2O_{5-x}$/$TaO_{2-x}$ bilayer structures," *Nature Materials*, vol. 10, pp. 625, 2011.

[75] Y. Du, "Symmetrical Negative Differential Resistance Behavior of a Resistive Switching Device," *ACS Nano*, vol. 3, pp. 2517-2523, 2012.

[76] I. Salaoru et al., "Resistive switching of oxygen enhanced $TiO_2$ thin-film devices," *Appl. Phys. Lett.*, vol. 102, pp. 013506, 2013.

[77] R. Sun et al., "Atomistic mechanisms of nonstoichiometry-induced twin boundary structural transformation in titanium dioxide," *Nature Communications*, vol.6, pp.7120, 2015.

[78] H. Y. Lee et al., "Evidence and solution of over-RESET problem for $HfO_x$ based resistive memory with sub-ns switching speed and high endurance," *2010 International Electron Devices Meeting*, pp. 19.7.1-19.7.4, 2010.

[79] U. Chand et al., "Investigation of thermal stability and reliability of HfO2 based resistive random access memory devices with cross-bar structure," *J. Appl. Phys.*, Vol. 117, pp. 184105, 2015.

[80] C. Wang et al., "Ultrafast RESET Analysis of $HfO_x$-Based RRAM by Sub-Nanosecond Pulses," *Adv. Electron. Mater.*, vol. 3, pp. 1700263, 2017.

[81] B. Govoreanu et al., "$10\times10nm^2$ Hf/HfOx crossbar resistive RAM with excellent performance, reliability and low-energy operation," *2011 International Electron Devices Meeting*, pp. 31.6.1-31.6.4, 2011.

[82] Y. Y. Chen et al., "Balancing SET/RESET pulse for $>10^{10}$ endurance in $HfO_2$/Hf 1T1R bipolar RRAM," *IEEE Transactions on Electron Devices*, vol. 59, pp. 3243-3249, 2012.

[83] Y. S. Chen et al., "Challenges and opportunities for $HfO_x$ based resistive random access memory," *2011 International Electron Devices Meeting*, pp. 31.3.1-31.3.4, 2011.

[84] C. Li et al., "Three-dimensional crossbar arrays of self-rectifying Si/$SiO_2$/Si memristors," *Nat. Com.*, vol. 8, pp. 15666, 2017.

[85] Shuang Pi, Peng Lin, and Qiangfei Xia, "Cross point arrays of $8\,nm \times 8\,nm$ memristive devices fabricated with nanoimprint lithography," *J. of Vac. Sc. & Tec. B*, vol. 31, pp. 06FA02, 2013.

[86] A. Chen, "A comprehensive crossbar array model with solutions for line resistance and nonlinear device characteristics," *IEEE Transactions on Electron Devices*, vol. 60, pp. 1318-1326, 2013.

[87] H. Y. Lee et al., "Low-power and nanosecond switching in robust hafnium oxide resistive memory with a thin Ti Cap," *IEEE Electron Device Letters*, vol. 31, pp. 44-46, 2010.

[88] C. H. Lien et al., "Overview and high density application of $HfO_x$ based RRAM," *2014 12th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT)*, pp. 1-4, 2014.

[89] M. B. Gonzalez et al., "Investigation of filamentary current fluctuations features in the high-resistance state of Ni/$HfO_2$-based RRAM," *IEEE Transactions on Electron Devices*, vol. 63, pp. 3116-3122, 2016.

[90] C. Wang et al., "Relaxation effect in RRAM arrays: demonstration and characteristics," *IEEE Electron Device Letters*, vol. 37, pp. 182-185, 2016.

[91] B. Traoré et al., "On the origin of low-resistance state retention failure in $HfO_2$-based RRAM and impact of doping/alloying," *IEEE Transactions on Electron Devices*, vol. 62, pp. 4029-4036, 2015.

[92] C. C. Liu et al., "Bridging the processor-memory performance gap with 3D IC technology," *IEEE Design & Test of Computers*, vol. 22, pp. 556-564, 2005.

[93] X. Fong et al., "Spin-transfer torque memories: devices, circuits, and systems," *Proceedings of the IEEE*, vol. 104, pp. 1449-1488, 2016.

[94] H. S. P. Wong et al., "Phase change memory," *Proceedings of the IEEE*, vol. 98, pp. 2201-2227, 2010.

[95] A. Sheikholeslami and P. G. Gulak, "A survey of circuit innovations in ferroelectric random-access memories," *Proceedings of the IEEE*, vol. 88, pp. 667-689, 2000.

[96] M. Kobayashi and T. Hiramoto, "Device design guideline for steep slope ferroelectric FET using negative capacitance in sub-0.2V operation: Operation speed, material requirement and energy efficiency," *2015 Symposium on VLSI Technology (VLSI Technology)*, pp. T212-T213, 2015.

[97] T. Endoh et al., "An overview of nonvolatile emerging memories— spintronics for working memories," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 6, pp. 109-119, 2016.

[98] S. Yu and P. Y. Chen, "Emerging memory technologies: recent trends and prospects," *IEEE Solid-State Circuits Magazine*, vol. 8, pp. 43-56, 2016.

[99] Z. Yang et al., "Oxide electronics utilizing ultrafast metal-insulator transitions," *Annu. Rev. Mater. Res.*, vol. 41, pp. 337-367, 2011.

[100] S. H. Liu et al., "High-performance polyimide-based ReRAM for nonvolatile memory application," *IEEE Electron Device Letters*, vol. 34, pp. 123-125, 2013.

[101] H. Song et al., "Single molecule electronic devices," *Adv. Mater.*, vol. 23, pp.1583–1608, 2011.

[102] F. Kreupl et al., "Carbon-based resistive memory," *2008 IEEE International Electron Devices Meeting*, pp. 1-4, 2008.

[103] An Chen, "A review of emerging non-volatile memory (NVM) technologies and applications," *Solid-State Electronics*, vol. 125, pp. 25-38, 2016.

[104] ITRS Emerging Research Devices Chapter, 2013.

[105] D. M. Fleetwood et al., "Effects of oxide traps, interface traps, and "border traps" on metal-oxide-semiconductor devices," *J. Appl. Phys.*, Vol. 73, pp. 5058, 1993.

[106] F. P. Heiman and G. Warfield, "The effects of oxide traps on the MOS capacitance," *IEEE Transactions on Electron Devices*, vol. 12, pp. 167-178, 1965.

[107] P. V. Sushko et al., "Oxygen vacancies in amorphous silica: structure and distribution of properties," *Microelectronic Engineering*, vol. 80, pp. 292-295, 2005.

[108] A. S. Foster et al., "Vacancy and interstitial defects in hafnia," *Phys. Rev. B*, vol. 65, pp. 174117, 2002.

[109] P. Weckx et al., "Defect-based compact modeling for RTN and BTI variability," *2017 IEEE International Reliability Physics Symposium (IRPS)*, pp. CR-7.1-CR-7.6, 2017.

[110] S. Mahapatra et al., "On the generation and recovery of interface traps in MOSFETs subjected to NBTI, FN, and HCI stress," *IEEE Transactions on Electron Devices*, vol. 53, pp. 1583-1592, 2006.

[111] W. Wang et al., "Compact Modeling and Simulation of Circuit Reliability for 65-nm CMOS Technology," *IEEE Transactions on Device and Materials Reliability*, vol. 7, pp. 509-517, 2007.

[112] T. Grasser, "The capture/emission time map approach to bias temperature instability," Bias Temperature Instability for Devices and Circuits, pp. 447-481, *Springer Science + Business Media*, 2014.

[113] G. Pobegen et al., "Impact of gate poly doping and oxide thickness on the N- and PBTI in MOSFETs," *Microelectronics Reliability*, vol. 51, pp. 1530-1541, 2011.

[114] K. T. Lee et al., "Technology scaling on high-K & metal-gate FinFET BTI reliability," *2013 IEEE International Reliability Physics Symposium (IRPS)*, pp. 2D.1.1-2D.1.4, 2013.

[115] T. Schram et al., "Novel process to pattern selectively dual dielectric capping layers using soft-mask only," *2008 Symposium on VLSI Technology*, pp. 44-45, 2008.

[116] M. Chudzik et al., "High-performance high-κ/metal gates for 45nm CMOS and beyond with gate-first processing," *2007 IEEE Symposium on VLSI Technology*, pp. 194-195, 2007.

[117] L. Å. Ragnarsson et al., "Ultra low-EOT (5 Å) gate-first and gate-last high performance CMOS achieved by gate-electrode optimization," *2009 IEEE International Electron Devices Meeting (IEDM)*, pp. 1-4, 2009.

[118] K. Mistry et al., "A 45nm logic technology with high-k+metal gate transistors, strained Silicon, 9 Cu interconnect layers, 193nm Dry Patterning, and 100% Pb-free Packaging," *2007 IEEE International Electron Devices Meeting*, pp. 247-250, 2007.

[119] K. Joshi et al., "A consistent physical framework for N and P BTI in HKMG MOSFETs," *2012 IEEE International Reliability Physics Symposium (IRPS)*, pp. 5A.3.1-5A.3.10, 2012.

[120] N. Kimizuka et al., "The impact of bias temperature instability for direct-tunneling ultra-thin gate oxide on MOSFET scaling," *1999 Symposium on VLSI Technology. Digest of Technical Papers*, pp. 73-74, 1999.

[121] M. Denais et al., "On-the-fly characterization of nbti in ultra-thin gate oxide pmosfeT's," *2004 Proc. Intl. Electron Devices Meeting (IEDM)*, pp. 109–112, 2004.

[122] V. D. Maheta et al., "Development of an ultrafast on-the-fly IDLIN technique to study NBTI in plasma and thermal oxynitride p-MOSFETs," *IEEE Transactions on Electron Devices*, vol. 55, pp. 2614, 2008.

[123] A. Chaudhary and S. Mahapatra, "A physical and SPICE mobility degradation analysis for NBTI," *IEEE Trans. Electron Devices*, vol. 60, pp. 2096, 2013.

[124] I. S. Esqueda and H. J. Barnaby, "A defect-based compact modeling approach for the reliability of CMOS devices and integrated circuits," *Solid-State Electronics*, vol. 91, pp. 81-86, 2014.

[125] I. S. Esqueda and H. J. Barnaby, "Surface-potential-based compact modeling of BTI," *2016 IEEE International Reliability Physics Symposium (IRPS)*, pp. XT-06-1-XT-06-6, 2016.

[126] G. Gildenblat (Ed), "Compact modeling – principles, techniques and applications," *Springer*, 2010.

[127] T. Grasser et al., "Analytic modeling of the bias temperature instability using capture/emission time maps," *2011 International Electron Devices Meeting*, pp. 27.4.1-27.4.4, 2011.

[128] N. Goel et al., "A comprehensive DC/AC model for ultra-fast NBTI in deep EOT scaled HKMG p-MOSFETs," *2014 IEEE International Reliability Physics Symposium*, pp. 6A.4.1-6A.4.12, 2014.

[129] V. Huard, "Two independent components modeling for negative bias temperature instability," *2010 IEEE International Reliability Physics Symposium*, pp. 33-42, 2010.

[130] D. M. Fleetwood et al., "Estimating oxide-trap, interface-trap, and border-trap charge densities in metal oxide-semiconductor transistors," *Appl. Phys. Lett.*, vol. 64, pp.1965, 1994.

[131] J. S. Suehle, "Ultrathin gate oxide reliability: physical models, statistics, and characterization," *IEEE Transactions on Electron Devices*, vol. 49, pp. 958-971, 2002.

[132] D. J. DiMaria and E. Cartier, "Mechanism for stress-induced leakage currents in thin silicon dioxide films," *J. Appl. Phys.*, Vol. 78, pp. 3883, 1995.

[133] K.R. Farmer, R. Saletti, and R.A. Buhrman, "Current fluctuations and silicon oxide wear out in metal-oxide-semiconductor tunnel diodes," *Appl. Phys. Lett.*, vol. 52, pp. 1749, 1988.

[134] K. Okada and K. Taniguchi, "Electrical stress induced variable range hopping conduction in ultra-thin silicon dioxides," *Appl. Phys. Lett.*, vol. 70, pp. 351, 1987.

[135] R. Degraeve et al., "A consistent model for the thickness dependence of intrinsic breakdown in ultra-thin oxides," *IEDM Technical Digest*, pp. 863, 1995.

[136] M. Houssa et al., "Model for the current-voltage characteristics of ultra-thin gate oxides after soft breakdown," *J. Appl. Phys.*, vol. 73, pp. 490, 1998.

[137] E. Miranda et al., “Soft breakdown fluctuation events in ultra-thin $SiO_2$ layers”, *Appl. Phys. Lett.*, vol. 73, pp. 490, 1998.

[138] T. Nigam, S. Martin, and D. Abusch-Magder, "Temperature dependence and conduction mechanism after analog soft breakdown," *Proceedings of the IRPS*, 417, 2003.

[139] B.P. Linder et al., "Gate oxide breakdown under current limited constant voltage stress," *2000 VLSI Technical Digest*, pp. 214, 2000.

[140] Fu-Chien Chiu, "A review on conduction mechanisms in dielectric films," *Advances in Materials Science and Engineering*, vol. 2014, pp. 1-18, 2014.

[141] M. Depas et al., "Determination of tunneling parameters in ultra-thin oxide layer poly-Si/$SiO_2$/Si structures," *Solid-State Electronics*, vol. 38, pp. 1465, 1995.

[142] D. J. DiMaria, "Explanation for the polarity dependence of breakdown in ultra-thin silicon dioxide films," *Appl. Phys. Lett.*, vol. 68, pp. 3004, 1996.

[143] J. Maserjian, "Tunneling in thin MOS structures," *Journal of Vacuum Science and Technology*, vol. 11, pp. 996, 1974.

[144] S. Zafar, Q. Liu, and E.A. Irene, "Study of tunneling current oscillation dependence on sio$_2$ thickness and Si roughness at the Si/SiO$_2$ interface," *Journal of Vacuum Science and Technology*, vol. 13, pp. 47, 1995.

[145] M.V. Fischetti et al., "Theory of high-field electron transport in silicon dioxide," *Physical Review B*, vol. 31, pp. 8124, 1985.

[146] B. K. Ridley, "Mechanism of electrical breakdown in SiO$_2$ films," *J. Appl. Phys.*, vol. 43, pp. 3274, 1972.

[147] D. J. DiMaria et al., "Electron heating in silicon dioxide and off-stoichiometric silicon dioxide films," *J. Appl. Phys.*, vol. 57, pp. 1214, 1985.

[148] N. F. Mott and E. A. Davis, "Electronic processes in non-crystalline materials," *Oxford University Press*, 1979.

[149] Y.-C. Yeo, T.-J. King, and C. Hu, "Direct tunneling leakage current and scalability of alternative gate dielectrics," *Appl. Phys. Lett.*, vol. 81, pp. 2091, 2002.

[150] F.-C. Chiu, P.-W. Li and W.-Y. Chang, "Reliability characteristics and conduction mechanisms in resistive switching memory devices using ZnO thin films," *Nanoscale Research Letters*, vol. 7, pp. 178, 2012.

[151] W. D. Zhang et al., "Two types of neutral electron traps generated in the gate silicon dioxide," *IEEE Transactions on Electron Devices*, vol. 49, pp. 1868, 2002.

[152] J. J. Tzou, J. Y. Sun, and C. T. Sah, "Field dependence of two large hole capture cross sections in thermal oxide on silicon," *Appl. Phys. Lett.*, vol. 43, pp. 861, 1983.

[153] S. Ogawa, N. Shiono, and M. Shimaya, "Neutral electron trap generation in SiO$_2$ by hot holes," *Appl. Phys. Lett.*, vol. 56, pp. 1329, 1990.

[154] W. D. Zhang et al., "Electrical signature of the defect associated with gate oxide breakdown," *IEEE Electron Device Letters*, vol. 27, pp. 393-395, 2006.

[155] S. D. Brorson et al., "Direct measurement of the energy distribution of hot electrons in silicon dioxide," *J. Appl. Phys.*, vol. 58, pp. 1302, 1985.

[156] J. F. Zhang et al., "Hole traps in silicon dioxides - Part I: Properties," *IEEE Transactions on Electron Devices*, vol. 51, pp. 1267, 2004.

[157] D. J. DiMaria, E. Cartier, and D. Arnold, "Impact ionization, trap creation, degradation, and breakdown in silicon dioxide films on silicon," *J. Appl. Phys.*, vol. 73, pp. 3367, 1993.

[158] D. J. DiMaria, D. Arnold, and E. Cartier, "Impact ionization and positive charge formation in silicon dioxide films on silicon," *Appl. Phys. Lett.*, vol. 60, pp. 2118, 1992.

[159] P. E. Nicollian et al., "The traps that cause breakdown in deeply scaled SiON dielectrics," *International Electron Devices Meeting*, pp. 1-4, 2006.

[160] C. Mannequin et al., "Stress-induced leakage current and trap generation in HfO2 thin films," *J. Appl. Phys.*, vol. 112, pp. 074103, 2012.

[161] B. H. Lee, "Unified TDDB model for stacked high-k dielectrics," *IEEE International Conference on IC Design and Technology*, pp. 83-87, 2009.

[162] O. Pirrotta et al., "Leakage current through the poly-crystalline $HfO_2$: Trap densities at grains and grain boundaries," *J. Appl. Phys.*, vol. 114, pp. 134503, 2013.

[163] C. Couso et al., "Local defect density in polycrystalline high-k dielectrics: CAFM-based evaluation methodology and impact on MOSFET variability," *IEEE Electron Device Letters*, vol. 38, pp. 637-640, 2017.

[164] A. Kerber and E. A. Cartier, "Reliability challenges for CMOS technology qualifications with hafnium oxide/titanium nitride gate stacks," *IEEE Transactions on Device and Materials Reliability*, vol. 9, pp. 147-162, 2009.

[165] C. D. Young et al., "Electron trap generation in high-k/ gate stacks by constant voltage stress," *IEEE Transactions on Device and Materials Reliability*, vol. 6, pp. 123-131, 2006.

[166] G. Bersuker et al., "Progressive breakdown characteristics of high-k/metal gate stacks," *IEEE International Reliability Physics Symposium Proceedings. 45th Annual*, pp. 49-54, 2007.

[167] P. Samanta and T. Y. Man, "Direct tunneling stress-induced leakage current in ultrathin $HfO_2/SiO_2$ gate dielectric stacks," *Appl. Phys. Lett.*, vol. 100, pp. 094507, 2006.

[168] E. Cartier and A. Kerber, "Stress-induced leakage current and defect generation in nFETs with $HfO_2$/TiN gate stacks during positive-bias temperature stress," *IEEE International Reliability Physics Symposium*, pp. 486-492, 2009.

[169] Y.-C. Yeo, T.-J. King, and C. Hu, "Direct tunneling leakage current and scalability of alternative gate dielectrics," *Appl. Phys. Lett.*, vol. 81, pp. 2091, 2002.

[170] M. V. Fischetti, Z. A. Weinberg, and J. A. Calise, "The effect of gate metal and $SiO_2$ thickness on the generation of donor states at the Si-$SiO_2$ interface," *J. Appl. Phys.*, vol. 57, pp. 418, 1985.

[171] Y. S. Chen et al., "Highly scalable hafnium oxide memory with improvements of resistive distribution and read disturb immunity," *IEEE International Electron Devices Meeting*, pp. 1-4, 2009.

[172] J. Wu, L. F. Register and E. Rosenbaum, "Trap-assisted tunneling current through ultra-thin oxide," *1999 IEEE International Reliability Physics Symposium Proceedings. 37th Annual*, pp. 389-395, 1999.

[173] G. Jegert et al., "Modeling of leakage currents in high-κ dielectrics: Three-dimensional approach via kinetic Monte Carlo," *Appl. Phys. Lett.*, vol. 96, pp. 062113, 2010.

[174] H. Y. Lee et al., "Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust $HfO_2$ based RRAM," *2008 IEEE International Electron Devices Meeting*, pp. 1-4, 2008.

[175] J. Shin et al., "$TiO_2$-based metal-insulator-metal selection device for bipolar resistive random access memory cross-point application," *J. Appl. Phys.*, vol. 101, pp. 033712, 2011.

[176] A. K. Jonscher, "Physical basis of dielectric breakdown," *Journal of Physics D: Applied Physics*, vol. 13, pp. 143, 1980.

[177] J. F. Verweij, J. H. Klootwijk, "Dielectric breakdown I: A review of oxide breakdown," *Microelectronics Journal*, vol. 27, pp. 611-622, 1996.

[178] N. F. Mott and R. W. Gurney, "Electronic processes in ionic crystals," *J. Phys. Chem.*, vol. 42, pp. 1142, 1941.

[179] O. Blank et al., "A model for multistep trap-assisted tunneling in thin high-k dielectrics," *J. Appl. Phys.*, vol. 97, pp. 044107, 2005.

[180] K.-M. Chang et al., "Modulation of the electrical characteristics of $HfO_x$/TiN RRAM devices through the top electrode metal," *ECS Transactions*, vol. 28, pp. 119-126, 2010.

[181] M. T. Thomas, "Preparation and Properties of Sputtered Hafnium and Anodic $HfO_2$ Films," *J. Electrochem. Soc.*, vol.117, pp.396-403, 1970.

[182] E. H. Sondheimer, "The mean free path of electrons in metals," *Adv. Phys.*, vol. 50, pp. 499-537, 2001.

[183] Y. Zhang et al., "Metallic to hopping conduction transition in $Ta_2O_{5-x}$/$TaO_y$ resistive switching device," *Appl. Phys. Lett.*, vol. 105, pp. 063508, 2014.

[184] S. Yu, X. Guan, and H.-S. P. Wong, "Conduction mechanism of TiN/$HfO_x$/Pt resistive switching memory: a trap-assisted-tunneling model," *Appl. Phys. Lett.*, vol. 99, pp. 063507, 2011.

[185] J. Schnakenberg, "Polaronic Impurity Hopping Conduction," *Phys. Status Solidi B*, vol. 28, pp.623-633, 1968.

[186] Z. Wang et al., "Transport properties of $HfO_{2-x}$ based resistive-switching memories," *Phys. Rev. B*, vol.85, pp. 195322, 2012.

[187] R. Fang et al., "Low-Temperature Characteristics of $HfO_x$-Based Resistive Random Access Memory," *IEEE Electron Device Letters*, vol. 36, pp. 567-569, 2015.

[188] R. Fang et al., "Total ionizing dose effect of γ-ray radiation on the switching characteristics and filament stability of $HfO_x$ resistive random access memory," *Appl. Phys. Lett.*, vol. 104, pp. 183507, 2014.

[189] S. Gerardin et al., "Radiation Effects in Flash Memories," *IEEE Trans. on Nucl. Sci.*, vol. 60, pp. 1953-1969, 2013.

[190] J. S. Bi et al., "The Impact of X-Ray and Proton Irradiation on $HfO_2$/Hf-Based Bipolar Resistive Memories," *IEEE Trans. Nucl. Sci.*, vol. 60, pp. 4540-4546, 2013.

[191] X. He, and R. E. Geer, "Heavy Ion Radiation Effects on TiN/$HfO_2$/W Resistive Random Access Memory," *IEEE Aerospace Conference*, pp. 1-7, 2013.

[192] J.-C. Dupin, "Systematic XPS studies of metal oxides, hydroxides and peroxides," *Phys. Chem. Chem. Phys.*, vol. 2, pp. 1319-1324, 2000.

[193] Y. S. Lin et al., Resistive switching mechanisms relating to oxygen vacancies migration in both interfaces in Ti/$HfO_x$/Pt memory devices, *J. Appl. Phys.*, vol. 113, pp. 064510, 2013.

[194] T. Y. Liu et al., "A 130.7mm2 2-layer 32Gb ReRAM memory device in 24nm technology," *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, pp. 210-211, 2003.

[195] G. Wang et al., "High-Performance and Low-Power Rewritable $SiO_x$ 1 kbit One Diode–One Resistor Crossbar Memory Array," *Adv. Mater.*, vol. 25, pp. 4789–4793, 2013.

[196] R. Fackenthal et al., "A 16Gb ReRAM with 200MB/s write and 1GB/s read in 27nm technology," *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 338-339, 2014.

[197] J. Woo et al., "Multi-layer tunnel barrier ($Ta_2O_5/TaO_x/TiO_2$) engineering for bipolar RRAM selector applications," *Symposium on VLSI Technology*, pp. T168-T169, 2013.

[198] A. Schönhals et al., "3-bit Resistive RAM Write-Read Scheme Based on Complementary Switching Mechanism," *IEEE Electron Device Letters*, vol. 38, pp. 449-452, 2017.

[199] M. C. Hsieh et al., "Ultra high density 3D via RRAM in pure 28nm CMOS process," *IEEE International Electron Devices Meeting*, pp. 10.3.1-10.3.4, 2013.

[200] R. Waser et al., "Redox-Based Resistive Switching Memories – Nanoionic Mechanisms, Prospects, and Challenges," *Adv. Mater.*, vol. 21, pp. 2632–2663, 2009.

[201] D. Garbin et al., "$HfO_2$-Based OxRRAM Devices as Synapses for Convolutional Neural Networks," *IEEE Transactions on Electron Devices*, vol. 62, pp. 2494-2501, 2015.

[202] P. Y. Chen et al., "Exploiting resistive cross-point array for compact design of physical unclonable function," *IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*, pp. 26-31, 2015.

[203] S. Balatti et al., "True Random Number Generation by Variability of Resistive Switching in Oxide-Based Devices," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 5, pp. 214-221, 2015.

[204] Z. Jiang et al., "A Compact model for metal–oxide resistive random access memory with experiment verification," *IEEE Transactions on Electron Devices*, vol. 63, pp. 1884-1892, 2016.

[205] P. Y. Chen and S. Yu, "Compact modeling of RRAM devices and its applications in 1T1R and 1S1R array design," *IEEE Transactions on Electron Devices*, vol. 62, pp. 4022-4028, 2015.

[206] P. Huang et al., "A physics-based compact model of metal-oxide-based RRAM DC and AC operations," *IEEE Transactions on Electron Devices*, vol. 60, pp. 4090-4097, 2013.

[207] S. Kasap, "Principles of Electronic Materials and Devices," *McGraw-Hill Inc.*, New York, NY, 2006.

[208] Y. M. Lu et al., "Impact of Joule heating on the microstructure of nanoscale TiO2 resistive switching devices," *J. Appl. Phys.*, vol. 113, pp. 163703, 2013.

[209] M. Lanza et al., "Grain boundaries as preferential sites for resistive switching in the $HfO_2$ resistive random access memory structures," *Appl. Phys. Lett.*, vol. 100, pp. 063507, 2012.

[210] B. Govoreanu et al., "10×10 $nm^2$ $Hf/HfO_x$ crossbar resistive RAM with excellent performance, reliability and low-energy operation," *International Electron Devices Meeting*, pp. 31.6.1-31.6.4, 2011.

# APPENDIX A

## COPYRIGHT

This dissertation includes some of my published works. The detailed copyright permission is showing as below.

Section 3.3.2.3 of Chapter 3 is reproduced from [188] with the permission of AIP Publishing:
Runchen Fang, Yago Gonzalez Velo, Wenhao Chen, Keith E. Holbert, Michael N. Kozicki, Hugh Barnaby, and Shimeng Yu, Total ionizing dose effect of γ-ray radiation on the switching characteristics and filament stability of $HfO_x$ resistive random access memory, Applied Physics Letters 2014 104:18, 183507, https://doi.org/10.1063/1.4875748.

Section 3.3.1.2 and Section 3.3.2.2 of Chapter 4 is reproduced from [187] with granted copy right:
© 2015 IEEE. Reprinted, with permission, from Runchen Fang, Wenhao Chen, Ligang Gao, Weijie Yu, and Shimeng Yu, Low-Temperature Characteristics of $HfO_x$-Based Resistive Random Access Memory, IEEE Electron Device Letters, vol. 36, no. 6, pp. 567-569, June 2015. DOI: 10.1109/LED.2015.2420665.