Characterisation of Novel Resistive Switching Memory Devices

Zheng Chai

A thesis submitted in partial fulfilment of the

requirements of Liverpool John Moores University for

the degree of Doctor of Philosophy

May 2017

Abstract

Resistive random access memory (RRAM) is widely considered as a disruptive technology that will revolutionize not only non-volatile data storage, but also potentially digital logic and neuromorphic computing. The resistive switching mechanism is generally conceived as the rupture/restoration of defect-formed conductive filament (CF) or defect profile modulation, for filamentary and non-filamentary devices respectively. However, details of the underlying microscopic behaviour of the resistive switching in RRAM are still largely missing. In this thesis, a defect probing technique based on the random telegraph noise (RTN) is developed for both filamentary and non-filamentary devices, which can reveal the resistive switching mechanism at defect level and can also be used to analyse the device performance issues.

HfO_2 is one of the most matured metal-oxide materials in semiconductor industry and HfO_2 RRAM shows promising potential in practical application. An RTN-based defect extraction technique is developed for the HfO_2 devices to detect individual defect movement and provide statistical information of CF modification during normal operations. A critical filament region (CFR) is observed and further verified by defect movement tracking. Both defect movements and CFR modification are correlated with operation conditions, endurance failure and recovery.

Non-filamentary devices have areal switching characteristics, and are promising in overcoming the drawbacks of filamentary devices that mainly come from the stochastic nature of the CF. a-VMCO is an outstanding non-filamentary device with a set of

unique characteristics, but its resistive switching mechanism has not been clearly understood yet. By utilizing the RTN-based defect profiling technique, defect profile modulation in the switching layer is identified and correlated with digital and analogue switching behaviours, for the first time. State instability is analysed and a stable resistance window of 10 for $>10^6$ cycles is restored through combining optimizations of device structure and operation conditions, paving the way for its practical application.

TaO_x-based RRAM has shown fast switching in the sub-nanosecond regime, good CMOS compatibility and record endurance of more than 10^{12} cycles. Several inconsistent models have been proposed for the Ta_2O_5/TaO_x bilayered structure, and it is difficult to quantify and optimize the performance, largely due to the lack of microscopic description of resistive switching based on experimental results. An indepth analysis of the $TiN/Ta_2O_5/TaO_x/TiN$ structured RRAM is carried out with the RTN-based defect probing technique, for both bipolar and unipolar switching modes. Significant differences in defect profile have been observed and explanations have been provided.

Acknowledgements

I would like to thank my supervisor, Prof. Weidong Zhang for his help and continuous guidance, encouragement and fruitful discussions every week throughout the project. Without his investment of time and efforts, the project would never reach the fruition. Truthfully, I appreciate his efforts in my PhD. I also deeply appreciate the advice and support of my 2nd supervisor, Prof. Jian Fu Zhang, and my 3rd supervisor, Dr. Zhigang Ji, throughout the project.

Many thanks are given to other (ex-)members in the Microelectronic reliability group at LJMU, Dr. Jigang Ma, Dr. Meng Duan, Mr. Rui Gao, Dr. Benbakhti Brahim, Mr. Steven Duffy, Mr. Azrif Manut and Mr. Hui Cao. I would like to express my special gratitude to Dr. Jigang Ma for helping me going through the tough days throughout the project.

I also owe a lot of gratitude to my daily advisor at IMEC, Dr. Bogdan Gorvoreanu in the MDD group, who is so kind to support me and answer any questions. Many thanks to the former and present leadership of the memory group: Dr. Malgorzata Jurczak, Dr. Ludovic Goux, Dr. Gouri Sankar Kar and Dr. Furnemont Arnaud, for their support and understanding.

I am also very thankful for the help and support from my colleagues at IMEC, Belgium. Dr. Robin Degraeve, Mr. Michael Chen, Dr. Attilio Belmonte, Ms. Avasarala Naga Sruti, Dr. Takuya Futase, Dr. Eddy Simoen, Dr. Cor Claeys, Dr. Sergiu Clima, Dr. Yang Yin Chen, especially the SRC team members: Dr. Leqi Zhang, and Ms. Subhali Subhechha.

I would like to acknowledge Prof. Yue Hao and Prof. Xuefeng Zheng for recommending me to pursue my PhD study at LJMU.

Despite of the hard work during my PhD studies, there were still lots of fun with you guys, Dr. Yan Li, Dr. Yi Li, Dr. Bo Wang, Dr. Sijia Jiang, Mr. Xuanwu Kang, Dr. Jie Hu, Dr. Ming Zhao, Mr. Xiangdong Li, Mr. Jingdong Yan, Mr. Ziyang Liu, Dr. Wen Fang, Dr. Zhe Chen, Mr. Xiangyu Wu, Ms. Mengting Jiang, Mr. Linghui Kong, Ms. Xiao Liu, Dr. Lianggong Wen, Dr. Chen Wu, Dr. Weiming Qiu, Dr. Shuzhen You, Dr. Tian-Li Wu, Dr. Tung Huei Ke, Dr. Chang Chen, Mr. Hao Yu, Dr. Yue Pan, Dr. Xubin Chen, Dr. Shengping Mao, Dr. Weiming Guo, Dr. Hu Liang, Dr. Mingxu Liu, Dr. Riqing Zhang, Dr. Yan Cai, Dr. Yang-shun Fan and Dr. Liang Zhao. I would like to thank Dr. Baojun Tang in particular for his kind help in the beginning days of my life in Leuven.

I would like to express my deep gratitude to my girlfriend, Ms. Zhaoying Song, Thank you for waiting for me in our hometown for almost 3 years. With the separation of two places which are thousands of miles apart, you made huge sacrifices to support my studies. Without your love, trust and understanding, I am not able to accomplish this work.

Finally, I would thank my parents (Mr. Dongsheng Chai 柴东生先生 and Mrs. Cuilan Han 韩翠兰女士 for their great support and encouragement during all these years.

游子吟

孟郊(751–814)

慈 母 手 中 线， 游 子 身 上 衣。
临 行 密 密 缝， 意 恐 迟 迟 归。
谁 言 寸 草 心， 报 得 三 春 晖。

Song of the Parting Son

Meng Jiao (751-814)

From the threads a mother's hands weaves,
A gown for parting son is made.
Sown stitch by stitch before he leaves,
For fear his return be delayed.
Such kindness as young grass receives
From the warm sun can't be repaid.

List of abbreviations

Abbreviation	***Signification***
1T1R	One-Transistor One-Resistor
ACF	Average Current Fluctuation
AES	Auger Electron Spectroscopy
AFM	Atomic Force Microscope
ALD	Atomic Layer Deposition
ALWG	Arbitrary Linear Waveform Generation
ASIC	Application Specific Integrated Circuits
BD	Breakdown
BE	Bottom Electrode
BEOL	Back End Of Line
BiCS	Bit Cost Scalable
BL	Bit Line
C-AFM	Conducting-Atomic Force Microscope
CAGR	Compound Annual Growth Rate
CB	Connection Block
CBRAM	Conductive Bridge Random Access Memory
CC	Compliance Current
CDF	Cumulative Distribution Function
CF	Conductive Filament
CFR	Critical Filament Region
CLB	Configurable Logic Block
CMOS	Complementary Metal Oxide Semiconductor
CPU	Central Process Unit
CVS	Constant Voltage Stress
DDR	Defect Deactivated Region
DRAM	Dynamic Random Access Memory
DUT	Device Under Test
E^2PROM	Electrically Erasable Programmable ROM
EDS	Energy Dispersive Spectroscopy
EPSC	Excitatory Postsynaptic Current
FEOL	Front End Of Line
FIB	Focused Ion Beam
FPGA	Field Programmable Gate Array
GPU	Graphics Processing Unit
HBM	Homogeneous Barrier Modulation
HDD	Hard Disk Drive
HMM	Hidden Markov Model
HRS	High Resistance State
IL	Insulating Layer
IoT	Internet of Things
ISPP	Incremental Step-Pulse Programming
LB	Logic Block
LRS	Low Resistance State
LTD	Long Term Depression
LTP	Long Term Potentiation

LUT	Look-Up Table
MIM	Metal Insulator Metal
MLC	Multi-Level Cell
MLH	Maximum Likelihood
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
MTJ	Magnetic Tunnel Junction
MUX	Multiplexer
NBTI	Negative-Bias Temperature Instability
NVM	Non Volatile Memory
OA	Optical Absorption
OTP	One-Time Programmable
PBTI	Positive-Bias Temperature Instability
PCRAM (PCM)	Phase Change RAM
PMCM	Pressure-Modulated Conductance Microscopy
PSD	Power Spectral Density
PVD	Physical Vapour Deposition
QPC	Quantum Point Contact
RAM	Random Access Memory
RDT	Defect Tracking Technique
ROM	Read Only Memory
RRAM	Resistive Random Access Memory
RTN	Random Telegraph Noise
RW	Resistance Window
SB	Switch Block
SE	Spectroscopic Ellipsometry
SEM	Scanning Electron Microscopy
SL	Switching Layer
SMU	Source Measure Unit
SPGU	Semiconductor Pulse Generator Unit
SRAM	Static Random Access Memory
SRC	Self-Rectifying Cell
SSD	Solid State Drive
STDP	Spike-Timing-Dependent Plasticity
STM	Scanning Tunnelling Microscopy
STT-MRAM	Spin Transfer Torque Magneto RAM
TAT	Trap Assistant Tunnelling
TB	Terabyte
TCR	Tunnelling Conduction Region
TE	Top Electrode
TEM	Transmission Electron Microscopy
TLP	Time Lag Plot
TMO	Transition Metal Oxide
UHV	Ultra-High Vacuum
VMCO	Vacancy Modulate Conductive Oxide
Vo	Oxygen Vacancy
WGFMU	Waveform Generator/Fast Measurement Unit
WL	Word Line
XPS	X-ray Photoelectron Spectroscopy
X-TEM	Cross-sectional TEM

List of symbols

Symbol	***Description***	***Unit***
A	Device area	nm^2
$E_{C,ox}$	Conduction band edge of oxide	eV
E_F	Fermi energy	eV
E_T	Trap energy	eV
I_{cc}	Compliance current	A
I_D	Drain current	A
ΔI	RTN amplitude	A
$I_{R,max}$	Maximum reset current	A
I_{read}	Read-out current	A
k_BT	Boltzmann constant	eV
N_C	Number of defects in constriction	--
q	Elementary charge	C
R_{HRS}	HRS resistance	Ω
R_a	Access resistance	Ω
R_{LRS}	LRS Resistance	Ω
T_{ox}	Oxide thickness	m
t_{reset}	Reset pulse width	s
t_{set}	Set pulse width	s
V_G	Gate voltage	V
$V_{forming}$	Forming voltage	V
V_{read}	Read-out voltage	V
V_{reset}	Reset voltage	V
V_{set}	Set voltage	V
V_{th}	Threshold voltage	V
$V_{recovery}$	Recovery voltage	V
V_{TE}	Top electrode voltage	V
X_C	Constriction position	--
X_T	Trap location	--
ε_{ox}	Dielectric constant of oxide	--
t_{high}	High current time	s
t_{low}	Low current time	s
τ_c	Capture time	s
τ_e	Emission time	s
τ_0	Characteristic time constant	s
f_c	Corner frequency	Hz

Table of contents

1 A Review of the Resistive Random Access Memory Devices

1.1 Conventional and emerging memory

1.1.1 Charge-storage based memory

The von Neumann architecture is composed of three distinct sub-systems: a central processing unit (CPU), input/output (I/O) devices and memory [1] as shown in **Figure 1-1**. This architecture was proposed in 1945 and is still being widely used in many forms of computers and many other electronic devices. Nowadays, the functionality and performance of the von Neumann architecture computing system are increasingly dependent on the characteristics of the memory sub-system.

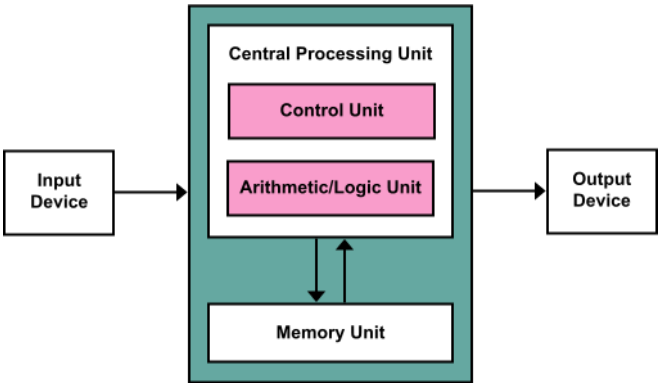


Figure 1-1 The von Neumann architecture scheme [1], consisting of a central processing unit (CPU), input/output (I/O) devices and memory.

Semiconductor memories can be typically categorized into volatile and non-volatile, depending on whether the stored information eventually fades or not when the power supply is turned-off [2]. The typical volatile memories are static random-access memory (SRAM) and dynamic random access memory (DRAM), while the typical non-volatile memories (NVMs) are hard disk drive (HDD) and Flash memory. According to their

capacities, costs and access times, volatile and non-volatile memories have different positions in the von Neumann architecture. For volatile memories, SRAMs are used for caches because of higher speed and DRAMs are used for main memory because of higher density [3, 4]. For NVMs, HDD stores data in the form of magnetic fields and had been the main power horse for high-capacity storage due to its large capacity and low costs in the past decades, but suffers from low speed, bulky size, complex and fragile mechanical components. HDD is being aggressively replaced by the Flash memory in which data is stored in the threshold voltage (V_{th}) of a MOSFET [5]. Compared to HDD, Flash memory features higher read/write speed, smaller size and lower power consumption, and are less prone to damage [6]. SRAM, DRAM and Flash memory are all based on the charge storage mechanism, as is summarized in **Figure 1-2** [7]: SRAM stores the charge at the storage nodes of the cross-coupled inverters; DRAM stores the charge at the cell capacitor; Flash memory stores the charge at the floating gate of the transistor.

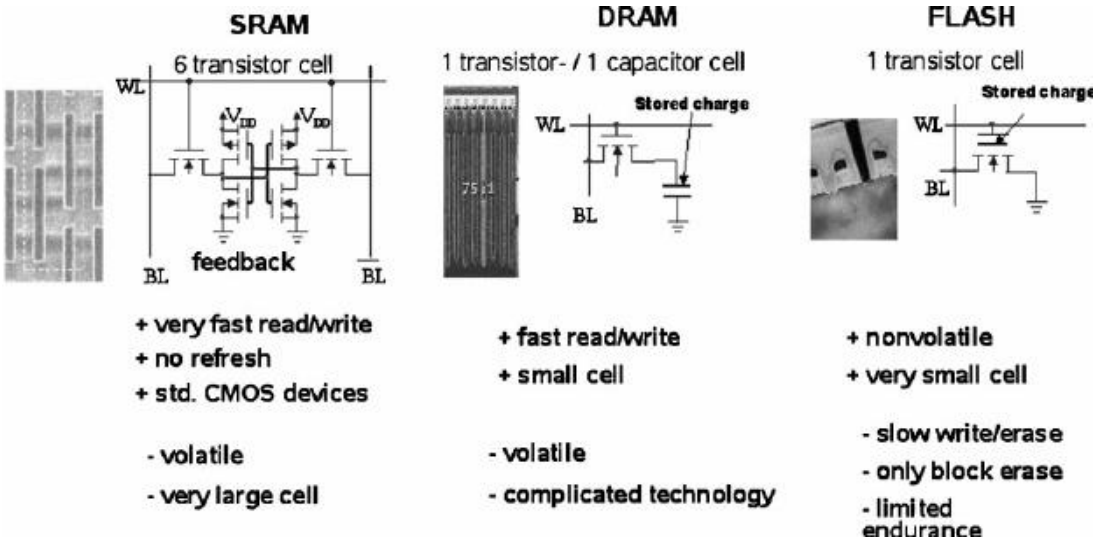


Figure 1-2 Structure and feature of SRAM, DRAM and Flash memory [7]. All these technologies are based on a charge storage mechanism.

All these charge storage based memories are facing challenges to be scaled down to the

10 nm node or smaller. The easy loss of the stored charges at nanoscale results in the degradation of performance, reliability, and noise margin. For NVMs, Flash memory is approaching the integration limit due to the key issues such as charge storage reduction, reliability degradation, controller complexity, write performance degradation and bit-cost-scalable (BiCS) difficulty as shown in **Figure 1-3** [8]. In recent years, many efforts such as the 3D Flash memory have been made to improve the performance of Flash memories while keeping the BiCS, but in the end, they will inevitably hit the physical limitation of charge storage based memories and suffer an intolerable increase in manufacturing cost [9-11].

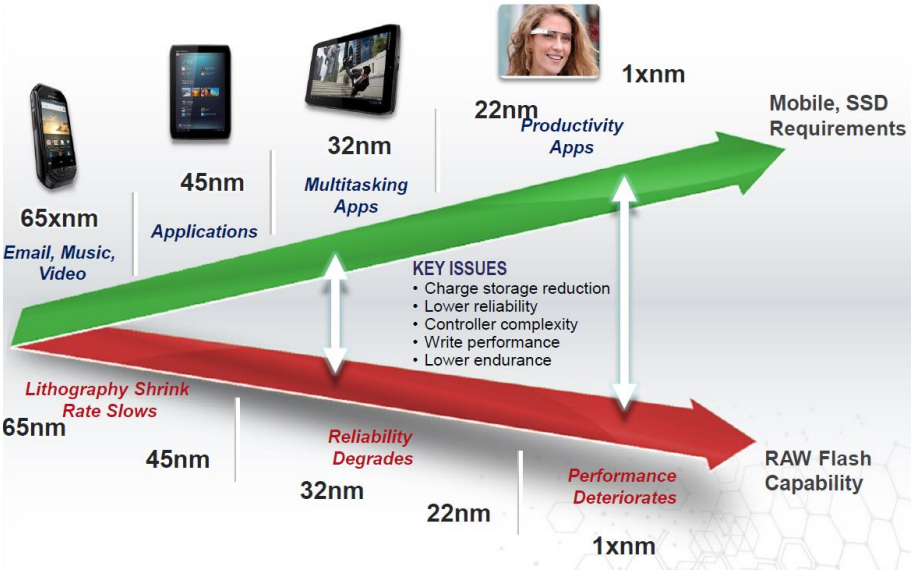


Figure 1-3 The charge storage memories are facing challenges to be scaled down with key issues such as charge reduction, lower reliability, controller complexity, write performance and lower endurance [8].

1.1.2 Emerging memories

The semiconductor industry has been looking for emerging memory technologies, due to the rising demand for highly scalable, low power consumption, cost effective, and fast speed memory solutions. The global emerging memory market has been valued at

USD ~8 billion by 2022 growing with ~45% of compound annual growth rate (CAGR) during forecast period 2017 to 2022 [12].

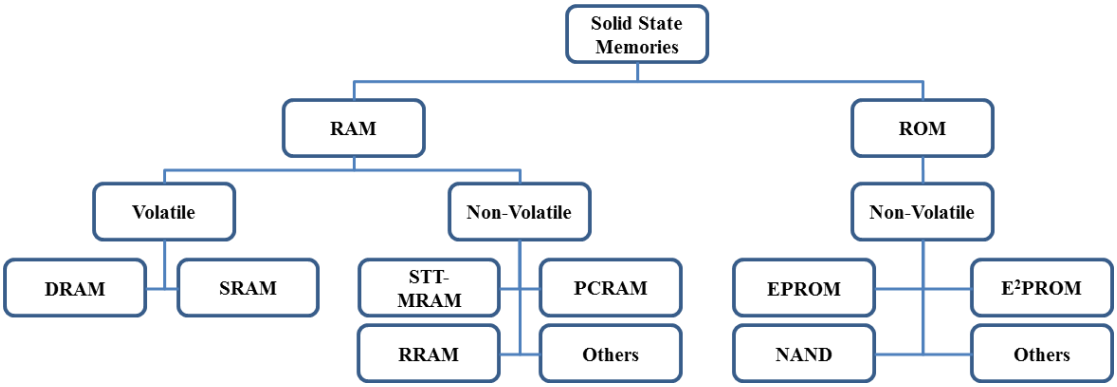


Figure 1-4 Classification of semiconductor solid state memories, including the charge storage memories and emerging memories [13].

Driven by those demands, the last two decades have witnessed the widespread emergence of novel memory technologies, including spin-transfer-torque magneto-resistive RAM (STT-MRAM) [14, 15], phase change RAM (PCRAM) [16, 17] and resistive RAM (RRAM) [18-21], as summarized in **Figure 1-4**. These emerging memory technologies are all two-terminal non-volatile memory devices whose resistance can be changed by electrical input. The data storage mechanisms of emerging memory technologies are not directly based on the number of charges. Despite those similarities, their detailed switching mechanisms are quite different:

- STT-MRAM employs a magnetic tunnel junction (MTJ) as the memory element [22], which consists of two ferromagnetic materials separated by a thin insulating tunnel layer. The resistance difference between the parallel configuration and the anti-parallel configuration of the two ferromagnetic layers in the MTJ determines the resistance state. The magnitude of the tunneling current can be read to indicate whether a logical 1 or 0 is stored.

- PCRAM devices stores data in the resistivity difference between the amorphous and the crystalline states of chalcogenide glass [16] which is sandwiched between a top electrode (TE) and a bottom electrode (BE). The phase change write operation consists of reset and set. During reset the chalcogenide glass is momentarily melted by a short electric pulse and then quickly quenched into amorphous HRS. During set a pulse with lower amplitude but longer time anneals the amorphous phase into crystalline LRS.

- RRAM can be categorized into oxide-RAM (OxRAM) and conductive-bridge RAM (CBRAM). The difference is that OxRAM's filament consists of oxygen vacancies in the oxide layer, while CBRAM's filament consists of metal atoms, formed by fast-diffusive Ag or Au ions migrating into the solid-electrolyte. Despite different underlying physics, these two types of RRAMs share many common device characteristics. This thesis focuses on the Vo-based RRAM. The Vo-based RRAM can be categorized into two types: filamentary RRAM and non-filamentary RRAM. In filamentary RRAM the resistance change is caused by the rupture (corresponding to HRS) and restoration (corresponding to LRS) of a conductive filament (CF) in the insulator between two electrodes [23]. In non-filamentary RRAM, the resistance change is caused by the areal modulation of defect distribution profile in the device [24].

Due to those different underlying switching mechanisms, the device characteristics are also different among emerging memory technologies [25-29] as summarized in **Table 1-1**:

Table 1-1 Comparison of conventional and emerging memories [25-29]. The device characteristics are different among those technologies.

	Volatile memory		***Non-volatile memory***		***Emerging non-volatile memory***		
Type	SRAM	DRAM	NOR Flash	NAND Flash	STT-MRAM	PCM	RRAM
Cell elements	6T	1T1C	1T	1T	1(2)T1R	1T1C	1T1R/1D1R
Mechanism	Voltage latching	Charge storage in trench capacitor	Floating gate/ charge tap	Floating gate/ charge tap	Magneto resistance	Phase change	Localized conduction/interface modification
Min. cell size	140 F^2	6 F^2	10 F^2	5 F^2	20 F^2	4 F^2	4 F^2
Write/Erase time	0.3 ns/0.3ns	<10ns/<10ns	1ms/10ms	1ms/0.1 ms	10ns/10ns	10 ns/50ns	5 ns/5ns
Endurance	> 3×10^{16}	> 3×10^{16}	> 3×10^{4}	> 3×10^{4}	> 3×10^{16}	10^{8}	>10^{12} (bit level)
Retention	Volatile	Volatile	10 year	10 year	Volatile	10 year	10 year
Energy/bit	pJ	30 fJ/ms	10-100 pJ/page	10-100 pJ/page	pJ	10 pJ	<0.1pJ
Application	Cache	Main memory	Storage	Storage	Storage/main memory	Storage	Storage/main memory

Different devices may have different application position due to their unique characteristics, and face different challenges from different aspects.

- Compared to SRAM, STT-MRAM has a smaller cell area while maintaining low programming voltage, fast write/read speed and long endurance, making it attractive as a replacement for embedded memories in the last cache [22]. However, multiple layers of exotic ferromagnetic materials are used in the MTJ stack, which result in poor Si-CMOS process compatibility. Also, to avoid the formation of dead layers/regions of the complicated MTJ stack, precise deposition and etching are required which will bring a significant cost barrier [30].

- PCRAM offers a large resistance window (RW) between its crystalline and amorphous states. This wide RW makes multi-level cells (MLC) possible, which is

of great importance for neuromorphic computing [31]. Also PCRAM has generally good Si-CMOS process compatibility. The key challenges for PCRAM cell design are the relative poor access latency and energy consumption [32, 33].

- Among the emerging memories, RRAM has attracted increasing interest due to its simple structure, low power consumption, fast speed, prolonged endurance and full Si-CMOS process compatibility. The key challenge of RRAM is the variability of the switching parameters owing to the stochastic nature of resistive switching, i.e. device-to-device and cycle-to-cycle variability. The history, category and recent progress of RRAM will be briefly introduced in the next sub-section, and the details of RRAM will be introduced in later sections.

1.1.3 Introduction to RRAM

The first report of resistive switching phenomena in oxides dates back to the 1960s [34] [35-37], but those early observed resistance switching phenomena were not robust enough for memory applications. The revival of interests in resistive switching from both academia and industry came in the late 1990s, when resistive switching was reported in complex metal oxides such as the perovskite oxides of $SrTiO_3$ [38], $SrZrO_3$ [18] and in binary metal oxides such as NiO [39] and TiO_2 [40]. In 2004 Samsung demonstrated the NiO memories integrated with conventional 0.18 μm CMOS in a one-transistor-one-resistor (1T1R) structure [39] and since then, research activity began to intensify.

So far in most RRAM devices it is conceived that the resistive switching mechanism is the rupture/restoration of CF, while in recent years a number of devices have been reported showing non-filamentary areal switching based on physical mechanisms such

as the migration of oxygen vacancies and trapping/de-trapping of charge carriers [24, 41-43]. Nevertheless, RRAM devices, both filamentary and non-filamentary, generally share the following characteristics which make them outstanding among emerging memories.

- Simple structure: the simple 2-terminal Metal-Insulator-Metal (MIM) structure allows highly geometrical scalability and easy 3D integration.

- Good Manufacturability: RRAM devices use fully CMOS compatible materials and can be fabricated using fab-friendly processes.

- Excellent scalability: functional devices have been demonstrated down to $10\times10nm^2$ size for filamentary devices [19], while for non-filamentary devices, the areal resistive switching makes it promising to scale further below the filament size (<10nm) [24].

- Low cost per bit: dense crossbar arrays make the smallest cell footprint possible, i.e. $4F^2$, (F: feature size). Moreover, non-filamentary resistive memories are compatible with Bit-Cost Scalable (BiCS) 3D architecture allowing further cost reduction [24].

- MLC: RRAM can provide large RW, which potentially enables MLC operation. Moreover, analogue switching behaviour has been demonstrated in some non-filamentary devices with excellent device-to-device variation, showing promising MLC application potential [42].

- Fast P/E speed: Program and Erase of resistive memory cells take much less time (>10 ns) than that of Flash memory (>μs).

The development of RRAM has progressed rapidly in the past decade. For the filamentary RRAM, devices have been demonstrated with either one or several of the following features: size down to 10 nm or below [19], programming current of a few μA [44], programming speed in the order of a few ns [19], programming endurance cycles larger than 10^6 with a record up to 10^{12} [45], retention time >3,000 h at 150 °C and extrapolated more than 10 years at 85 °C [46], or even with forming-free characteristics [47]. Most of these characteristics were reported in HfO_x or TaO_x systems. 2-bit and 3-bit multi-level operation has been demonstrated [48]. Chip-level RRAM array macro from 4 Mb to 32 Gb capacity with peripheral circuitry have been demonstrated by industry as well [49]. For the non-filamentary RRAM, although only a few devices have been reported so far, they have attracted intense interest due to their areal switching behaviour which results in excellent device to device operation uniformity and promising scaling capability. Devices with $>10^2$ RW, tuneable μA-range switching current and excellent variability have been demonstrated [43]. The details of filamentary and non-filamentary RRAM devices will be introduced in Section 1.2 and Section 1.3, respectively.

1.2 Filamentary RRAM

Most RRAM devices reported are the filamentary RRAM in which resistive switching is controlled by the restoration and rupture of a CF inside the oxide layer. Tens of binary oxides have been found to exhibit resistive switching behaviour. Most of them are transition metal oxides, and some are lanthanide series metal oxides. The materials for the resistive switching oxide layer and the electrodes reported in literature are summarized in **Table 1-2** and **Table 1-3** [29]. Besides metals, conductive nitrides, e.g., TiN, TaN, are also commonly used as electrode materials.

Table 1-2 Summary of oxide materials reported showing filamentary type of resistive switching [29].

Oxides	***Switching mode***	***Switching media***
NiO	Unipolar/Bipolar	Ni cation/oxygen vacancy
TiO_2, Nb_2O_5, ZrO_2 HfO_2, Ta_2O_5, MgO, Al_2O_3, CoO, ZnO, SnO_2 CeO, WO	Unipolar/Bipolar	Oxygen vacancy
CuO	Unipolar/Bipolar	Cu cation/oxygen vacancy
FeO	Bipolar	Oxygen vacancy
SiO_2	Unipolar/Bipolar	Oxygen vacancy/buffer layer for Cu, Ag cation

Table 1-3 Summary of electrode materials showing resistive switching with the oxide material of HfO_2 [29].

Electrode materials	***Switching mode***	***CMOS compatibility***
Pt	Unipolar/Bipolar	NOT FEOL/BEOL compatible
Ag	Bipolar	
Au	Unipolar	
Cu, Ti, Rr, Hf, Ta, TiN	Bipolar	BEOL compatible
Ru, La	Unipolar	Compatible but with integration difficulty
Ni	Unipolar/bipolar	
Al	Bipolar	Compatible but with pattern difficulty

1.2.1 Switching operation in filamentary RRAM

The filamentary RRAM devices typically include three switching operations: forming, set, and reset, as shown in **Figure 1-5**:

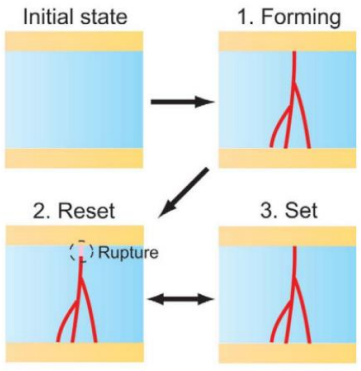


Figure 1-5 Schematic of the CF at initial fresh state and after (1) forming (2) reset and (3) set process [23].

- Set: switching from HRS to LRS.

- Reset: switching from LRS to HRS.

- Forming: usually for the fresh devices, a voltage larger than the V_{set} is needed to trigger on the subsequent resistive switching.

According to the polarity of V_{set} and V_{reset}, the switching modes of filamentary RRAM can be broadly classified into unipolar and bipolar [23] as shown in **Figure 1-6**:

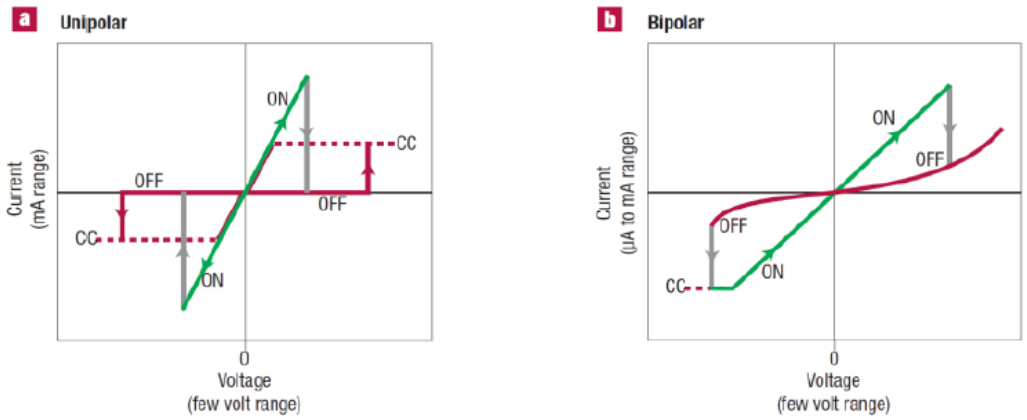


Figure 1-6 Schematic of DC I-V characteristics of resistive memory in (a) unipolar and (b) bipolar switching modes. ON/OFF refers to LRS and HRS respectively. CC stands for current compliance [23].

- Unipolar mode: the switching polarity depends solely on the amplitude but not the polarity of the applied voltage. Thus, set/reset can occur at the same polarity. If the unipolar mode symmetrically occurs at both positive and negative voltages, it is also referred as a nonpolar switching mode.

- Bipolar mode: the switching direction depends on the polarity of the applied voltage. Thus, set can only occur at one polarity and reset can only occur at the reverse polarity.

To avoid a permanent dielectric breakdown in the set process, in both polarity modes, a compliance setup is necessary. For single RRAM devices usually it is provided by the

semiconductor parameter analyser. In practical application, it is often provided by a memory cell selection transistor/diode or a series resistor. To read the resistance from the device, a small V_{read} (e.g. 0.1V) is applied which does not affect the state of the device to detect whether the cell is in HRS or LRS [23].

1.2.2 Switching mechanism in filamentary RRAM

Various filamentary RRAM devices may have different dominant conduction mechanism depending on the dielectric properties, the fabrication process conditions, and the properties of the interface between the oxides and the electrodes. The details of the physical mechanism for a resistance switching phenomenon in filamentary metal-oxide memory are still an active research area. Here a broad overview for simple metal-oxide RRAM is given.

- Forming/set:

  The forming/set process is interpreted to be a dielectric soft breakdown [50]. The forming process is not a spontaneous process at some critical voltage, but an upsurge process resulting from stress-induced defects [51]. Under the high electric field, the oxygen atoms are knocked out of the lattice, and drift toward the TE. Simultaneously, defects in the bulk oxide are generated. The localized deficiency of oxygen leads to the formation of CFs with either oxygen vacancies [52] or metal precipitates [53]. The localized CFs paths were observed in various metal oxide RRAM devices by physical characterisation techniques, confirming the filamentary conduction mechanism [54-57], which will be introduced in detail in Chapter 2. Usually the as-deposited RRAM oxide thin films are amorphous or poly-crystalline, and the CFs are preferentially generated along the grain boundaries.

In fresh samples, usually there are so few intrinsic defects that a high $V_{forming}$ is needed to initiate the switching. After the forming, sufficient defects are generated. In the subsequent set/reset cycles, only a portion of the defects can be recovered during the reset. That is why the $V_{forming}$ is larger than the V_{set}, and the resistance of fresh state is much larger than the R_{HRS} in the subsequent cycles. Often, the remaining defect-rich region at HRS is referred as the residual filament.

A large $V_{forming}$ is not desirable in practical applications, and significant efforts have been made to achieve the so-called forming-free devices. It is found that the $V_{forming}$ is linearly dependent on the thickness of the oxide film [58-60], so a thinner oxide film is effective for reducing the $V_{forming}$. It is demonstrated that TiN/Ti/HfO_2/TiN memory with a 3 nm thick HfO_2 film can be forming-free [47]. Forming is also a strong function of film deposition conditions [61]. It is found that controlling the annealing ambient during deposition is also helpful in reducing the $V_{forming}$, possibly due to the introduction of defects to make the films oxygen deficient [62-64].

- Unipolar/Bipolar Reset

Although the forming/set mechanism appears to have a consensus as discussed above, the reset mechanism for different switching modes is controversial. The thermal dissolution model and the ionic migration model can explain parts of the unipolar and bipolar switching characteristics [65, 66], respectively. However, a full physical description of the two switching modes that can explain all the experimental observations is still missing. Experimental observations by various materials characterisation techniques reveal that the oxygen migration is present in the switching process and plays an important role in both modes. Electro-thermal calculations suggest that the local temperature around the CFs would rise by several

hundred Kelvin due to the large current flow [51-53], which may enhance the oxygen migration. In addition, as summarized in **Table 1-3** the electrode materials have a significant effect on the switching modes of the metal–oxide memory. Even with the same oxide material but with different electrode materials, the switching modes can be different. Therefore, it is inferred that the switching mode is not an intrinsic property of the oxide itself but a property of both the oxide material and electrode/oxide interfaces. In most cases, the unipolar mode is obtained with noble metals such as Pt or Ru as both TE and BE [13]. With one of the electrodes replaced by oxidizable materials such as Ti or TiN, the bipolar mode is obtained [19].

1.3 Non-filamentary RRAM

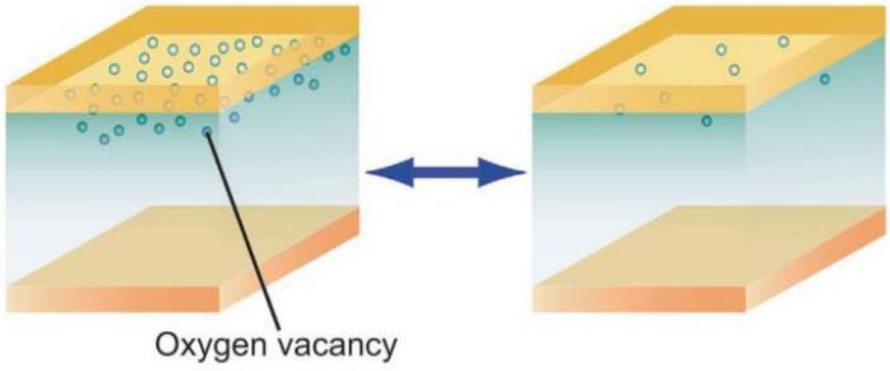


Figure 1-7 Schematic of non-filamentary switching. Change of oxygen vacancy density at the switching interface causes resistive switching[21].

Different from the filamentary RRAM, where the conductive path is formed locally within a small portion of the oxide, non-filamentary resistive switching takes place across most of the device volume, or area, as shown in **Figure 1-7**. For example, resistive switching happens at the interface between the metal electrode and oxide [67],

or inner-interfacial region in multiple layer devices [43]. Those devices are referred to as the non-filamentary resistive switching memories. The resistance of a non-filamentary resistive switching device is inversely proportional to the device area after programming [24], while it is almost area independent for the filamentary switching devices at LRS [23].

A number of devices have been reported showing non-filamentary switching, based on physical mechanisms such as migration of oxygen vacancies and trapping/de-trapping of charge carriers. In 2008, Baikalov et al reported the hysteretic and reversible polarity-dependent resistive switching driven by electric pulses in both $Ag/Pr_{0.7}Ca_{0.3}MnO_3/YBa_2Cu_3O_7$ sandwiches and single-layer $Pr_{0.7}Ca_{0.3}MnO_3$ strips in which the switching takes place at the $Ag\text{-}Pr_{0.7}Ca_{0.3}MnO_3$ interface [41]. In 2013, Wei et al reported $Pt/CuO_x/Si/Pt$ devices prepared by RF sputtering technique at room temperature which show a non-filamentary switching effect [68]. In 2015 Wang et al proposed a $Ta/TaO_x/TiO_2/Ti$ device with analogue synaptic features, where the resistance change is determined by homogeneous barrier modulation (HBM), making it promising for future hardware-based neuromorphic computing application [69]. Since 2013, Govoreanu et al has reported 3 generations of non-filamentary, double-layer structure RRAM devices, namely the vacancy-modulated conductive oxide resistive RAM (VMCO-RRAM). The 1st generation is of $TiN/TiO_2/Al_2O_3/TiN$ structure and features self-compliant, self-rectifying, μA-level operation and tight resistance distribution [24]. The 2nd generation was reported in 2015 with TiN/TiO_2/a-Si/TiN structure and features μA-range switching current levels with forming-free characteristic, improved nonlinearity and operation uniformity [42]. Based on that, the 3rd generation was reported in 2016 with reduced switching currents, increased resistance window, excellent wafer-level uniformity and improved reliability thanks to

careful inner-interface engineering [43].

In general, the resistive switching in non-filamentary devices is attributed to the defects (e.g. oxygen vacancies) moving back and forth with the polarity of the applied voltage, causing a change of barrier (at electrode/oxide interface) seen by the tunnelling carriers. Thus, the change of resistance is due to the modulation of the defect profile and external current compliance is not a must as there is no excess defect generation during forming/set. Moreover, sub-μA current filament RRAM switching suffers from high intrinsic variability due to low number of switching species, and no current reduction is expected before cell size is approaching the filament size. This, however, is not the case for the non-filamentary resistive switching devices. Its areal switching is expected to easily yield below 1 μA operating current, while maintaining wide operation window and having better immunity to low-number switching species variation [24]. Some non-filamentary devices even show perfect analogue switching behaviour and ×100 resistance window [43], making them promising candidates for MLC application and neuromorphic computing.

1.4 Novel application areas of RRAM

RRAM is not only a promising candidate to replace the Flash memories, but also has great potential to result in revolutionary developments in application areas such as FPGA and neuromorphic computing. Its simple structure and non-volatile feature can save the chip area and energy in FPGA significantly. RRAM devices have also emerged as the leading candidate to realize the synapse and neuron functions due to the extra-low energy loss per spike.

Leon Chua envisioned the fourth non-linear passive two terminal electrical component

called memristor, in addition to R, L and C, in 1971 [70]. It relates electric charge and magnetic flux linkage for a particular time interval. In 2008, researchers at Hewlett Packard (HP) Labs reported that the memristor was realized physically using thin film of TiO_2 nanoscale device [71]. Basically the memristor is a resistance with memory. Its resistance changes when a voltage is applied to this element and remains constant on that particular value when the applied voltage is removed. The main difference between the memristor (M) and the three passive elements (R, L, C) is its nonlinear input-output characteristics. RRAM is a device that can switch between one or more resistances under the application of appropriate voltages, or may have a continuously variable resistance as is in the case of a-VMCO non-filamentary RRAM [42]. It shows certain memristive behaviour, and can be even thought of as a specific type of memristor [71]. Whatever the case, it is important that the change in resistance is governed by the past history of the device, that is, by the previous voltage applied, or the previous current that has flowed through the device. Memristor is a strong candidate for future memories because of its non-volatile property and high packing density in a crossbar array. The data is retained in the memory even when the power is turned off for a predefined time such as 10 years.

1.4.1 RRAM for novel FPGA Architectures

The explosive growth of mobile devices, especially in the wake of the forthcoming Internet of Things (IoT) era, pushes up the demand for larger and more powerful datacentres, as the ability to handle vast volumes of streaming data are more important now than ever before. FPGAs, due to its advantages in energy efficiency, scalability and flexibility, are emerging as a suitable accelerator solution used together with CPUs. It is suitable for handling the ever-growing range of tasks such as in cloud and high

performance computing and in deep machine learning, in addition to its many existing applications, particularly in storage and networking. While the combination of FPGAs and CPUs can offer great potential for performance improvements, it and other potential solutions such as GPUs and custom ASICs are based on CMOS technology, which are all facing the near future where Moore's Law is fast approaching its physical limits. Novel solutions beyond CMOS are currently under intensive research world-wide.

Field Programmable Gate Arrays (FPGAs) are semiconductor devices that are based on a matrix of configurable logic blocks (CLBs) connected via programmable interconnects. FPGAs can be programmed and re-programmed to different sets of desired functionalities after being manufactured in the foundry. This feature distinguishes FPGAs from Application Specific Integrated Circuits (ASICs) which are custom manufactured for specific design tasks. Although one-time programmable (OTP) FPGAs are available, the dominant types are SRAM-based which can be reprogrammed as the design evolves. The FPGAs' properties such as reconfiguration and defect tolerance are attractive to the electronic industry, especially when CMOS technology scales down to the nano-metre range and the fabrication yield of IC components hardly ever approaches 100%. It is predicted that the world-wide market of FPGA chips will increase to USD 7.23 billion by 2022, and USD 14.2 billion by 2024 [72]. Much of this revenue comes from the significant increase of high-end FPGA chips with a large amount of memory components.

The FPGA programmable interconnects usually contain three types of components: SRAM-based storage of configuration bits, MUX-based routing switches, and buffers. None of these components are trivial parts in FPGAs. Although optimisation for CMOS-based programmable interconnects in FPGAs have been made and is still on-going, there are fundamental limitations in all three components of programmable

interconnects [73].

- Most FPGAs use SRAMs to store programming bits [74]. Each SRAM cell uses more than six transistors to store only one bit. Modern FPGAs enable serial bit-stream programming by storing configuration bits in latch nodes embedded in a shift register structure, with an area overhead no less than the SRAM-based storage. Moreover, the volatile nature of the storage media causes excessive power consumption during standby.

- MUX-based routing switches in FPGAs are implemented in tree structures with serial pass transistors for less SRAM-based storage of select bits. Each pass transistor has to be wide enough to provide sufficient drive, and this leads to a large footprint.

- In most cases, routing buffers in FPGAs exceed the buffering demand of a given application. The quantity and positions of routing buffers placed in each track of programmable interconnects need to be optimized to meet the worst-case demand of the track.

The non-volatile Flash memory devices as configuration devices enable the FPGA to retain its configured state when the power is off but at the cost of additional processing, low speed, high programme voltage, and the limitation of scaling down to the 10 nm technology nodes [75]. As introduced in Section 1.1, emerging memories have been widely explored to replace Flash memory. The transistor-free feature enables future scaling of high-density memory. Based on the recent developments of emerging memory technologies, it has been suggested recently that they have the potential to revolutionize the reconfigurable computing through a number of novel FPGA

architectures [76] as shown in **Figure 1-8**. Emerging memories can be used to replace SRAMs, as programmable switches, or integrate with nanowire crossbars.

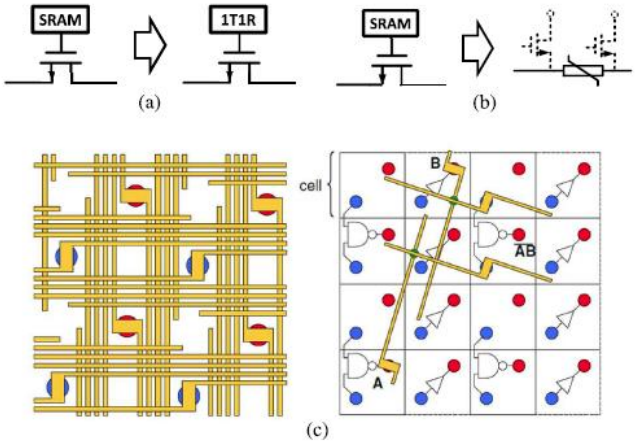


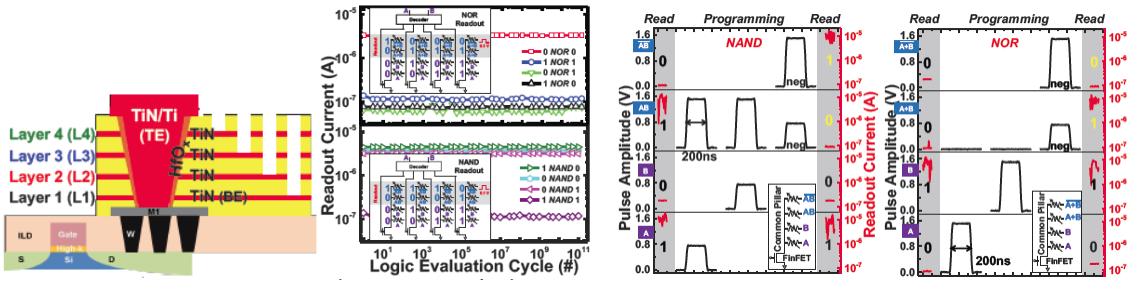
Figure 1-8 (a) Replace SRAMs with emerging memories; (b) Use emerging memories as programmable switches (c) Integration of CMOS, emerging memories and nanowire in field-programmable nanowire interconnect [76].

Among the emerging memories RRAM is outstanding for FPGA application due to its simple structure, low power consumption and good Si-CMOS compatibility. Its 3D cross-bar network structure at the BEOL can potentially solve the density, connectivity and memory bandwidth limitations of conventional hardware.

The non-volatility of RRAM devices is an outstanding feature for its FPGA application. The static power consumption is elevated due to the high ratio of volatile SRAM utilization in the FPGA. During standby, a large amount of leakage power is consumed in FPGAs just to preserve the configuration bit-stream for logic functionality and routing allocation. Moreover, in many applications previously computed data are required for new calculations, as most of the information is contained in the evolution of monitored data. Therefore, if non-volatility is included in the FPGA, it enables us to power-down/restore the whole FPGA without losing any configuration data while reducing the leakage power consumption to zero in standby and to minimize external

communication by saving the context in the FPGA that significantly decreases the overall delay and total power consumption.

Non-volatility can be introduced in the FPGA using RRAM based circuit elements. Recent studies show that RRAM devices provide numerous advantages as they are compatible with CMOS BEOL process and are capable of operating at high speeds with a relatively small area. RRAM integration in the FPGAs is demonstrated to have reduced area and higher performance [76-79]. RRAM-based switch blocks designed in different technologies are assessed for use in FPGAs [80, 81]. It was recently shown that RRAM devices guarantee reliable operation with a considerably high endurance (up to 10^{12} cycles [82]). Therefore, at a circuit level, the integration of RRAM results in a reduced total area, critical path delay, and power consumption with stable functionality. It has been estimated that in theory by using RRAM to replace the SRAM based switches matrix in FPGA, a 96% smaller footprint, 55% higher performance, and 79% lower power consumption can be achieved comparing to existing FPGA counterparts [73].



(a) (b) (c)

Figure 1-9 (a) 4-layer 3D 1T4R RRAM can be programed to (b) implement NAND/AND and NOR/OR logic functions; Deeper logic states and multi-inputs are possible with more layer stacks. (c) Dynamically reconfigurable for multi-stage logic functions [83].

Furthermore, 3D RRAM array may also be used to replace the existing SRAM-based LUTs in FPGAs, as shown in **Figure 1-9** [83], as it can operate as truly non-volatile

"logic in memory" devices with dynamically reconfigurable multi-stage logic functions, without consuming much silicon surface, leading towards disruptive developments in reconfigurable computing. Despite of these preliminary developments, a number of critical challenges need to be overcome from device to system design level, before RRAM's practical implementation in reconfigurable computing becomes possible.

1.4.2 RRAM-based neuromorphic application

Conventional computer systems are based on the von Neumann model of computation [1], in which instructions are fetched, decoded and executed in the sequential way, resulting in less energy-efficient machines due to the increasing complexity. Although the current digital computers can process at very high speed to emulate, to certain extent, the brain functionality of animals like a spider, mouse, and cat [84], the associated energy dissipation in the system grows exponentially along the hierarchy of animal intelligence [85-87]. For example, to achieve cortical simulations at the cat scale, the super computer, Blue Gene/P (BG/P) with 147,456 CPUs and 144 TB of total memory has to be used [84].

On the other hand, human brains are the most power-efficient computational engines, where many complex computations including cognition, action and thoughts are conducted while consuming less than 20 W [88]. Our brain consists of approximately 10^{11} neurons, each of which creates action potentials (voltage spikes) at an average rate of 10 Hz [89]. These communication signals are then sent to other neurons through synapses; the typical fan-out of a neuron is in the range of 5,000–10,000, resulting in a total of about 10^{15} synapses. The tokens of information processing in the brain are the action potentials, whereas the strength of communication between neurons is encoded in

the effective conductance of the synapse. The communication strength of synapses can change with activity, and it is believed that our ability to learn and form new memories is based on this synaptic plasticity [90, 91].

Machine learning based on neuromorphic computing has shown great potential to lead the next wave of ICT revolution. Engineers have been motivated to mimic the key algorithmic and computational features of the brain in both software and hardware [92] [93]. On the software part, neuromorphic computation aims to mimic the key operating principles, algorithms, and architecture of the brain and holds promise to deliver the next generation of systems capable of tackling a wide variety of unstructured computational problems, and significant progress has been made in the past decades [94-98]. On the hardware part, efforts have been made for large neuromorphic systems development to interpret fuzzy and noisy inputs and make intelligent decisions, but primarily based on CMOS technology [99-101]. Bio-inspired machine learning has sparked enormous interests recently, e.g. the success of Google's AlphaGo games, which use supercomputers based on CMOS technology with a large quantity of CPUs and GPUs. It requires hundreds of kWs power consumption to support the huge amount of computing power, prohibiting its large-scale implementation, especially in the forthcoming IoT era, where local smart devices with low power consumption are urgently needed to overcome network bandwidth bottleneck for transferring vast amount of data.

It has recently been demonstrated that the RRAM devices can be used as electronic synapses [95] and neurons [102], and become strong potential candidates to revolutionise the neuromorphic computing. It provides simple nanometre-sized two-terminal MIM structure with ion-controlled analogue, programmable, multi-level resistance, similar to that in the biological synapse/neuron, as shown in **Figure 1-10(a)**

[71, 103-105]. RRAM devices can be arranged in passive crossbar arrays of minimum feature size or potentially integrated in 3D synaptic arrays. This BEOL 3D crossbar network structure can solve the density, connectivity and memory bandwidth limitations of conventional CMOS neuromorphic hardware [106,107]. These outstanding properties of RRAM enable its application in energy-efficient neuromorphic hardware systems.

RRAM can also serve as reconfigurable wiring through its crossbar array on top of the CMOS-based neurons [108]. The tuneable resistive state of RRAM is used as a synaptic weight, which can be adapted by suitable methods for changing RRAM resistance in analogy with synaptic plasticity rules. By updating the weight of the synapses, the electrical connection between a presynaptic neuron and a postsynaptic neuron changes, thus enabling the possibility of implementing a variety of learning models for pattern recognition and memory storage. As shown in **Figure 1-10(b) and (c)**, the weight of the synapses is typically updated following the learning rule called STDP, where the conductance change of the synapse depends on the relative timing of the electrical pulses delivered from the presynaptic neuron and the postsynaptic neuron [90, 109, 110].

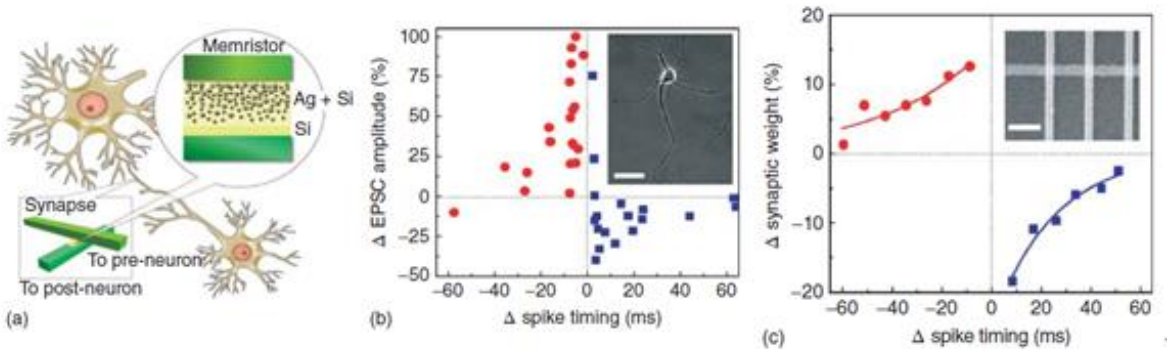


Figure 1-10 (a) Schematic description of the role of the RRAM as a synapse between two neurons [111]. (b) Experimental change of excitatory postsynaptic current (EPSC) of rat hippocampal neurons as a function of the relative spike timing [110]. (c) Experimental RRAM STDP curve. The exponential behaviour is similar to biological measurements in (b) [97].

Experimental evidence of the feasibility of the STDP learning curve with RRAM devices has been reported [97]. The experimentally measured STDP, namely the conductance update as a function of inter spike delay. The experimental demonstration of memristive STDP in **Figure 1-11** has also been achieved [112]. According to this scheme, a weight update requires the presence of a postsynaptic spike and a presynaptic spike, as the conductance is a function of the timing difference between Pre-neuron-spike and Post-neuron-spike.

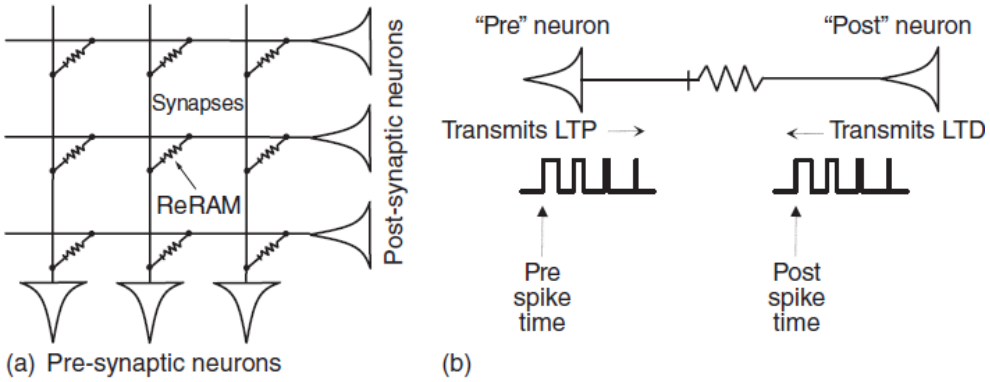


Figure 1-11 (a) RRAM-based synaptic array and (b) spiking scheme to either induce long term potentiation (LTP) or long-term depression (LTD) in the synapse [112].

Non-filamentary devices, due to its areal switching characteristic, have also shown strong potential for neuromorphic applications. For example, the non-filamentary a-VMCO RRAM in which the resistive switching is controlled by defect profile modulation, shows analogue switching characteristics, i.e. its resistance can be arbitrarily determined with DC stop voltage or AC pulse amplitude/width [42]. This makes a-VMCO RRAM devices ideal for neuromorphic application.

Despite the progress, a clear understanding of the requirements and targets for operating conditions of RRAM devices in neuromorphic application has been lacking and also, there's still a gap between the ability to build semiconductor devices and the ability to

design systems, which limits the implementation of fully functional large-scale neuromorphic systems to mimic brain functions [113, 114].

1.5 Key device performance metrics

1.5.1 Variability

Variability, which includes cycle-to-cycle variation and device-to-device variation, is a major concern for the large-scale manufacturing of RRAM, as it imposes constraints on memory cell/array design for both filamentary and non-filamentary RRAM devices. Hence, a deeper understanding of these issues is essential for RRAM's applications in memory architecture [115].

For filamentary devices, the origin of the variation has been attributed to the stochastic nature of the oxygen vacancies/ions processes [23]. The LRS variation comes from the variation of the number or the size of CFs, while the HRS variation comes from the variation of the length of the ruptured CFs, i.e., the gap [116]. The HRS variation is generally larger than the LRS variation because of the exponential dependence of the tunnelling current on the gap distance. Variability can be improved or mitigated with various methodologies, which includes material engineering that aims to decrease the device resistance variation by utilizing enhanced material or interfaces [23]. Another method is to use verification techniques such as incremental step pulse programming (ISPP) similar to that used in the NAND Flash memories to constrain the distribution of device resistance [117, 118]. The variation effect can also be suppressed by circuit design innovations inside the RRAM memory array or its CMOS peripherals [119].

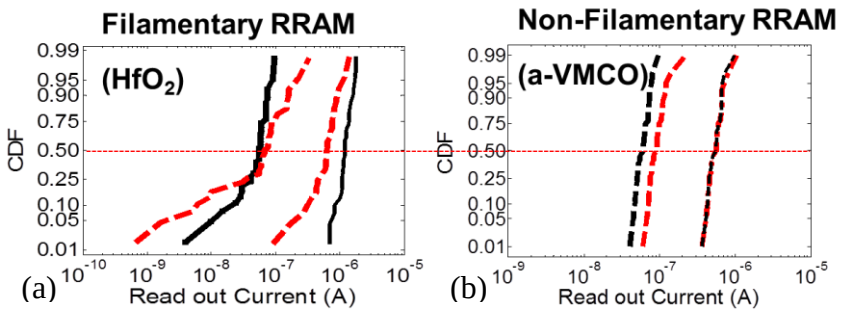


Figure 1-12 Read-out current before (black) and after (red) the retention test, for filamentary HfO_2 RRAM (a) and non-filamentary a-VMCO RRAM (b). Non-filamentary RRAM shows much better device-to-device variation at both HRS and LRS, compared with filamentary HfO_2 RRAM [42].

In contrast to the filamentary counterpart, current transportation in non-filamentary is dominated by interfacial barrier modulation, which leads to much better device-to-device variation [42] as shown in **Figure 1-12**. However, resistance drift at both HRS and LRS can be significant in these devices, which causes large cycle-to-cycle variation. It will be of great importance if new techniques can be developed to investigate these issues at defect level experimentally, which will help to understand the variation problem. The cycle-to-cycle variation may also be improved with careful engineering of the interface where resistance switching takes place, and may also be reduced with verification techniques as mentioned above.

1.5.2 Endurance

Endurance is one of the essential criteria for the universal memory application. During the cycling operations, the resistance value may gradually drift, and the device could eventually become stuck at either HRS or LRS and cannot be switched anymore. The endurance performance of RRAM depends on a variety of factors: material, processing, device structure, operation scheme [23]. Both the stuck-at-HRS and stuck-at-LRS cases may be mitigated by the ISPP verification operation programme. A lot of efforts have

been made for the endurance failure mechanism, but few provide insight analysis at the defect level, i.e. look into how defect movement happens, or how defect profile changes when the device starts to fail during cycling. Therefore, new techniques are needed urgently for the endurance evaluation and optimization at microscopic level.

1.5.3 Retention

Data retention refers to how long the memory states can be maintained. A data retention time longer than ten years is expected and it must be maintained at thermal stress up to 85 °C (operating temperature) and under electrical stress such as a constant stream of read pulses [23]. Data retention at both room temperature and elevated temperature is mandatory for meeting specifications for embedded memory and/or automotive applications [120]. A simple linear extrapolation method is commonly used to bake the devices at a high temperature and monitor the device resistance by applying read pulses at certain time intervals, and extrapolate the resistance evolution to the 10-year point [121]. The most accurate method is to record the time-to-failure at each temperature, draw the Arrhenius ($1/k_BT$) plot to extract the activation energy and then extrapolate down to the operating temperature [121]. In filamentary RRAM devices, one of the critical reliability issues is the trade-off between retention and endurance [122]. Long endurance can be obtained with highly reactive cap materials, while less reactive metal cap material results in longer retention time. In HfO_2 devices, both HRS and LRS resistance increase with the backing time. Cycled devices show worse retention compared with fresh devices, so retention test should be combined with endurance test [123]. Both endurance and retention may be improved by using the high-pressure hydrogen annealing [124]. Despite the results so far, there's still a lack of microscopic explanation for the retention failure at defect level based on experimental results,

similar to the case of endurance.

1.5.4 Multi-level operation

RRAM devices with MLC capability are highly desirable as it allows one memory device to realize more than one bit of digital data per cell in a much smaller footprint [23]. For filamentary device, the LRS resistance can be changed by the set I_{cc} possibly due to the modulation of the diameter or number of CFs, while the HRS resistance can be controlled by the V_{reset} possibly due to the modulation of the ruptured CF length [58]. Multilevel resistance can be achieved with either exponentially increasing the programming pulse width or linearly increasing the programming pulse amplitude [125]. Non-filamentary RRAM devices have also shown outstanding MLC characteristic thanks to their large RW and analogue behaviour. For example, in the a-VMCO device, the resistance can be arbitrarily controlled with switching conditions and the total resistance window can be as large as 100, leaving enough room for MLC operation [42]. For MLC operation, the resistance window should be large enough; the endurance and immunity to V_{read} disturbance should be good enough for each state, as well as the thermal stability [126], which remains as a challenging research topic.

1.5.5 Scaling

The potential scalability continuously drives the development of RRAM technology. For filamentary devices, the resistance of LRS has a weak dependence on the cell area while that of HRS increases inversely with the cell area. Recently the HfO_x memory device size has been aggressively scaled down to 10nm×10nm while retaining the good performances such as speed, operation voltage and switching energy [19]. Besides

the conventional top-down fabrication approach, the self-assembly grown metal–oxide nanowires also exhibit the resistive switching behaviour, further illustrating the scalability of RRAM in the nanometre regime [127]. However, no LRS current decrease is expected in filamentary devices before device size is approaching the filament size (<10nm). In non-filamentary devices where areal switching takes place, both LRS and HRS resistance increase inversely with the cell area, and low operation current can be achieved while maintaining operation window with better variation immunity [24].

1.5.6 Maximum reset current

In filamentary RRAM devices, the maximum current during the reset process is an important parameter of concern because it dominates the peak power consumption. Most reports in the literature show a typical reset current in the order of mA or hundreds of μA for a single memory cell, and it reduces only slightly when the devices are scaled down, thus leading to a remarkable increase of the current density required for reset. This presents a significant challenge for the memory cell selection devices, which need to provide a very large current density for ultra-scaled cells, e.g., 10^7 A/cm^2 even for a 100 nm×100 nm cell. This problem, however, can be alleviated by using a smaller set compliance current during the set process, because the reset current is almost linearly proportional to the set compliance current. This current has been shown to be related to the effective cross-sections of the created conductive filaments (CF) through the dielectric, as higher CF cross-sections translate to higher $I_{R,max}$ values, indicative of higher power consumption during memory cell operation. K. Kinoshita et al. reported the reduction in the reset current by reducing a parasitic capacitance [128], and Y. S. Fan et al demonstrated suppression of the overshoot current effect by using a test structure with an integrated series access resistor [129].

1.5.7 RRAM arrays

High-density crossbar RRAM array can be implemented with the 2-terminal MIM RRAM devices. A crossbar array consists of parallel interconnects at upper and lower planes perpendicular to each other with RRAM devices at each crossing point. An effective cell area will be $4F^2$ assuming the width of both lines and spaces is F, achieving the smallest single layer cell footprint. Moreover, multiple 2D layers can be stacked into a 3D configuration, thus reducing the minimal feature size to $4F^2/n$ with n the number of layers, as shown in **Figure 1-13**. Therefore stacked crossbar arrays are expected to be a promising architecture for high density and large capacity memory application.

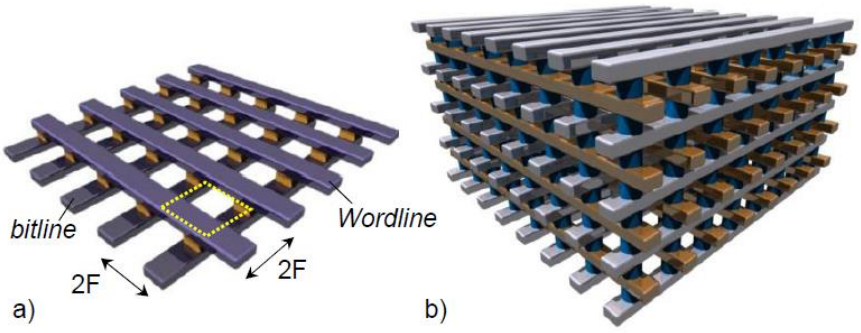


Figure 1-13 Schematic of (a) a 2D crossbar memory array. (b) stacked 3D crossbar array. Multiple 2D layers can be stacked into a 3D configuration, thus reducing the minimal feature size to $4F^2/n$ with n the number of layers.

One of the main issues that hinder the successful commercialization of RRAM crossbar array is the sneak current caused by the poor non-linearity of devices. Sneak current is the current in a passive crossbar memory array in which multiple parallel conduction paths exist between a pair of TE and BE as shown in **Figure 1-14** [130]. Sneak current significantly degrade the accessibility to the target cell and raise power consumption.

Parallel leakage paths can be suppressed by either increasing the non-linearity using a separate, non-linear device serially connected with each RRAM device in a one-selector one-resistor (1S1R) configuration, or by directly introducing non-linearity into the memory cell itself, which is called self-rectifying cell (SRC) [24].

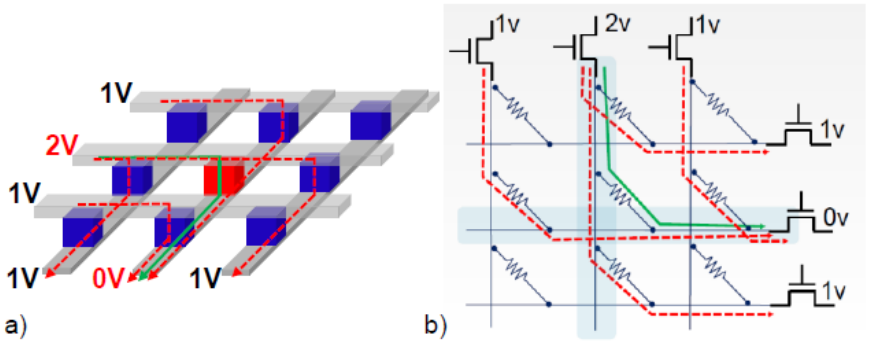


Figure 1-14 (a) Illustration of sneak current in crossbar array with (nearly) linear resistive switching memory cells during read operation. (b) Circuit schematic of crossbar array. Red dash lines: sneak current paths. Green solid line: actual readout signal from the selected element [13].

1.6 Thesis content overview

1.6.1 Thesis objective

RRAM devices are widely considered as a disruptive technology in the areas such as digital memory, memristor architecture, FPGA and neuromorphic computing. RRAM devices can be categorized into filamentary devices and non-filamentary devices, depending whether the resistive switching is caused by the rupture and restoration of CF, or areal change in the defect profile. Various models have been proposed to explain the resistive switching mechanisms, but all of them are at best phenomenological descriptions. There are disputes about how and where the CF is switched on/off in filamentary devices, and how the profile of defects is modulated in non-filamentary

devices. Further understanding of the underlying physics of RRAM, especially the switching parameter variations, is required for industrial large-scale manufacturing. The lack of characterisation techniques for key defect parameters plays a major role in holding back the development of RRAM. It is necessary to design and develop new characterisation techniques, specifically for new RRAM memory applications, to understand the mechanisms responsible for resistive switching, and provide guidance for the device design and performance improvement.

The first objective of this thesis is to develop an electrical characterisation technique to probe the defect profile in the bipolar HfO_2 RRAM devices, which shows promising potential in industry application. HfO_2 RRAM devices are filamentary, therefore it is of great importance to observe the region where the filament ruptures and restores, either in the relative or absolute location inside the oxide. It is widely accepted that resistive switching is caused by defect movement driven by field, so it will be good if this technique can also track the defect movement during switching, and link the microscopic defect movement to resistance change of device. Based on all above, it should be possible to explain why and how the defect fails, e.g. stuck at one of the resistance states. All those will provide insight to the resistive switching mechanism with direct experimental evidence with this technique.

The second objective of this thesis is regarding the non-filamentary RRAM devices, which shows promising application in not only digital memory, but analogue memory and neuromorphic computing, due to its areal switching mechanism and other characteristics. A novel technique is needed to provide insight evidence for the defect distribution profile especially near the interface which is critical in resistive switching. With this technique it should also be possible to explain the device degradation and offer solutions for device performance improvement.

The third objective of this thesis is to study the unipolar and bipolar modes that co-exist in the TaO_x-based device. TaO_x-based RRAM has shown fast switching in the sub-nanosecond regime, good CMOS compatibility and record endurance of more than 10^{12} cycles, making it one of the competitive candidates because such a long endurance capability enables it to be used in embedded memory applications and potentially can make a change of the memory hierarchy. Unipolar switching mode brings great facility for the development of large-scale RRAM array, but will bring instability if it co-exists with bipolar switching in the same device. Many models have been proposed but there's still a lack of experimental results for this issue. Therefore it is necessary to look for a different explanation for the switching mechanism of unipolar operation. Under this guidance, the instability of switching of Ta_2O_5/TaOx devices can be improved.

1.6.2 Thesis outline

This thesis is arranged with the following chapters.

Chapter 2: RTN-based defect extraction technique

In this chapter, the characterisation and measurement techniques used in the literature and in this thesis are reviewed. The switching operations of RRAM devices are briefly presented first, followed by the review of existing physical/electrical characterisation techniques. The physical characterisation techniques include the conductive-atomic force microscopy (C-AFM), scanning tunnelling microscopy (STM) and Transmission Electron Microscopy (TEM). The electrical characterisation techniques include the DC and AC I-V measurement and simulation, and Monte Carlo simulation technique. Advantages and limitations of those techniques for understanding the resistive switching mechanism are discussed. An RTN-based technique is used and further

developed in this thesis, as it has become an important topic in nano-scale devices. It provides a promising tool for defect analysis, which would help depict the microscopic picture of defect behaviour inside the device and improve the understanding of resistive switching of RRAM devices. The background knowledge of RTN is introduced, and advanced data analysis tools for RTN are reviewed. The results of previous studies on RTN in RRAM devices are reviewed. The basic defect extraction model based on the RTN in the MIM-structure device is presented at the end of this chapter.

Chapter 3: characterisation of HfO_2 RRAM

In this chapter, an RTN-based defect tracking (RDT) technique has been developed that can detect individual oxygen vacancy (Vo) movement and provide statistical information of CF modification during normal operations of nanoscale RRAM devices without device destruction. This chapter is organized as follows: After a description of the devices and measurement setup in Section 3.2, the results and discussions are in Section 3.3, which includes observation of the critical filament region (CFR), verification of CFR by defect movement tracking, correlation of CFR modification with operation conditions, endurance failure and recovery.

Chapter 4: Characterisation of a-VMCO RRAM

In this chapter, by utilizing the recently developed defect profiling technique based on RTN, in-depth analysis is carried out in a-VMCO RRAM devices. Defect profile modulation in the TiO_2 switching layer of a-VMCO RRAM is correlated with the analogue switching. It is also evidenced that the gradual growth of a defect-deactivation-region (DDR) in the TiO_2 layer near its interface with the a-Si barrier layer causes the LRS endurance instability, while tr_{eset}/ts_{et} ratio is found critical

for the HRS instability. Under this guidance, a stable ×10 window for $>10^6$ cycles is restored through combining optimizations of device structure and set/reset conditions, paving the way for a-VMCO's practical applications. The RTN-based defect profiling technique proves to be a useful tool for the non-filamentary RRAM studies.

Chapter 5: Characterisation of Ta_2O_5 RRAM

In this chapter an overall picture of the resistive switching mechanism is provided for TiN/Ta_2O_5/TaO_x/TiN structured RRAM device. Bipolar and unipolar resistive switching modes have been observed co-existing in the same device under different operation conditions. For the bipolar mode, defects movement into/out of the CFR is responsible for the resistance switching, similar to that in HfO_2 RRAM. The location of CFR has been identified, which is located at the middle of the TCR, different to that near the BE in HfO_2 RRAM. This may be caused by the weaker scavenging capability of the TaO_x layer and lower oxygen affinity of Ta compared with Hf. For the unipolar mode observed at negative TE biases, it is observed that the effective voltage for set is much stronger than that for reset, and time for set is much shorter than reset. The unipolar switching mechanism, whether it is defect horizontal movement caused by thermal Joule heating, or defect energy alternation, and is still under investigation.

Chapter 6: Conclusion and outlook

This chapter summarize the main results and contributions of this thesis to the research field of RRAM. Finally, suggestions for future works and outlook of the RRAM technology are briefly discussed.

1.7 Novelty and publications of the research

Chapter 3:

For the first time, an RTN-based defect extraction technique is developed for the HfO_2 devices to detect individual defect moment and provide statistical information; a critical filament region (CFR) is observed and further verified by this technique; new explanation for endurance failure is also proposed based on this technique.

1. **Z. Chai**, J. Ma, W. Zhang, B. Govoreanu, E. Simoen, J. F. Zhang, Z. Ji, R. Gao, G. Groeseneken, M. Jurczak, *RTN-based defect tracking technique: experimentally probing the spatial and energy profile of the critical filament region and its correlation with HfO_2 RRAM switching operation and failure mechanism*, Digest of Technical Papers - **Symposium on VLSI Technology, IEEE 2016 Symposia on VLSI Technology and Circuits (Orally presented in Honolulu, HI, US, June 2016)**

2. W. Zhang, **Z. Chai**, J. Ma, J. F. Zhang, Z. Ji *Analysis of RTN signals in Resistive-Switching RAM device and its correlation with device operations*, **ICSICT 2016**

3. C. Claeys, M. G. C. de Andrade, **Z. Chai**, W. Fang, B. Govoreanu, B. Kaczer, W. Zhang, E. Simoen, *Random Telegraph Signal Noise in Advanced High Performance and Memory Devices*, SBMicro 2016 (Belo Horizonte, Brazil)

4. **Z. Chai**, J. Ma, W. Zhang, B. Govoreanu, J. F. Zhang, Z. Ji, G. Groeseneken, M. Jurczak, *Probing the Critical Region of Conductive Filament in Nanoscale HfO_2 Resistive-Switching Memory by Random Telegraph Signals*, **Trans. Electron Devices** (submitted)

Chapter 4:

For the first time, by utilizing the RTN-based defect profiling technique, defect profile modulation in the switching layer is identified and correlated with digital and analogue

switching behaviours; state instability is analysed and a stable resistance window of 10 for 10^6 cycles is restored through combining optimizations of device structure and operation conditions.

1. J. Ma, **Z. Chai**, W. Zhang, B. Govoreanu, J. F. Zhang, Z. Ji, B. Benbakhti, G. Groeseneken and M. Jurczak, *Identify the critical regions and switching/failure mechanisms in non-filamentary RRAM (a-VMCO) by RTN and CVS techniques for memory window improvement*, Technical Digest - **International Electron Devices Meeting**. Institute of Electrical and Electronics Engineers (IEEE 2016)
2. C. Claeys, M.G.C. de Andrade, **Z. Chai**, W. Fang, B. Govoreanu, B. Kaczer, W. Zhang, E. Simoen, *Random Telegraph Signal Noise in Advanced High Performance and Memory Devices*, SBMicro 2016 (Belo Horizonte, Brazil)
3. **Z. Chai**, J. Ma, W. Zhang, B. Govoreanu, H. Cao, J. F. Zhang, Z. Ji, L. Goux, A. Belmonte, R. Degraeve, L. Di Piazza, G. Kar, *Understanding defect profile in non-filamentary RRAM (a-VMCO) for analogue switching and correlation with different endurance behaviour in practical memory applications* **IEEE Trans. Electron Devices** (submitted)
4. J. Ma, **Z. Chai**, W. Zhang, B. Govoreanu, J. F. Zhang, Z. Ji, B. Benbakhti, G. Groeseneken and G. Kar, *Investigation of cell failure in a-VMCO RRAM: identify critical regions and weak paths through CVS/RTN analysis*, **IEEE Trans. Electron Devices** (in preparation)

Chapter 5:

For the first time, an in-depth analysis is carried out with the RTN-based defect profiling technique, for both bipolar and unipolar switching modes that co-exist in the same TiN/Ta_2O_5/TaO_x/TiN device; significant differences in defect profile have been observed and explanations have been proposed.

1. **Z. Chai**, Jigang Ma, Weidong Zhang, Bogdan Govoreanu, Jian Fu Zhang, Zhigang Ji, Guido Groeseneken, Gouri S. Kar *Understanding the switching model of TaOx-Based Unipolar and*

Bipolar Resistive Random Access Memory based on RTN technique, **IEEE Trans. Electron Devices** (in preparation)

2 Characterisation and measurement techniques for RRAM

2.1 Introduction

In this chapter, the characterisation and measurement techniques for RRAM that are used in the literature and in this thesis will be reviewed. The working operations of RRAM devices will be briefly presented, followed by the review of existing physical characterisation techniques and electrical characterisation techniques. The physical characterisation techniques include the conductive-atomic force microscopy (C-AFM), scanning tunnelling microscopy (STM) and Transmission Electron Microscopy (TEM). The electrical characterisation techniques include the DC and AC I-V measurement and simulation, and Monte Carlo simulation technique. Advantages and limitations of those techniques for understanding the resistive switching mechanism will be discussed. An RTN-based technique will be used and further developed in this thesis, as it has become an important topic in nano-scale devices. It provides a promising tool for defect analysis, which will help depict the microscopic picture of defect behaviour inside the device and improve the understanding of resistive switching of RRAM devices. The background knowledge of RTN will be introduced, and advanced data analysis tools for RTN will be reviewed. The results of previous studies on RTN in RRAM devices will be reviewed. The basic defect extraction model based on the RTN in the MIM-structure device will be presented at the end of this chapter.

2.2 Measurement System

In this section the measurement systems and instrumentation are introduced.

In this thesis, two measurement systems, namely Measurement System (I) and (II), are employed for the characterisation of RRAM.

As shown in **Figure 2-1(a)**, the Measurement System (I) is based on the Keysight 1500A Semiconductor Device Parameter Analyser. Besides 4 Source Measurement Units (SMUs), this system is also equipped with 4 B1530A Waveform Generator/Fast Measurement Units (WGFMUs) and 2 B1525A Semiconductor Pulse Generator Units (SPGUs). Keysight B1500A features current-voltage (IV) measurement capabilities of spot, sweep, sampling and pulse measurement in the range of 0.1 fA - 1 A / 0.5 μV - 200 V, AC capacitance measurement in multi frequency from 1 kHz to 5 MHz and Quasi-Static Capacitance-Voltage (QS-CV) measurement capabilities, advanced pulsed IV and ultra-fast IV measurement capability from minimum 5 ns sampling interval (200 MSa/s), and up to 40 V high voltage pulse forcing for non-volatile memory evaluation. This system can be controlled with either the embedded controlling software EasyExpert, or a GPIB card with an IEEE 488 port. Connected to the Analyser is a Signatone probe station.

As shown in **Figure 2-1(b)**, the Measurement System (II) is an all-in-one system designed for RRAM measurement. This system consisting of two Agilent 81110A dual channel Pulse generators, two K2602A Keithley dual channel source meters, one Keithley S46 Microwave switch system and one LeCroy WavePro 735Zi Oscilloscope. It integrates DC characterisation, AC characterisation and real-time monitoring. It allows measuring all typical test structures, including 2-terminal device (1R and 1D1R), 3-terminal device (1T1R). It is compatible with both unipolar and bipolar operation modes. Connected to this system is a Cascade 300 nm probe station which is controlled by an automatic wafer-level measurement programme.

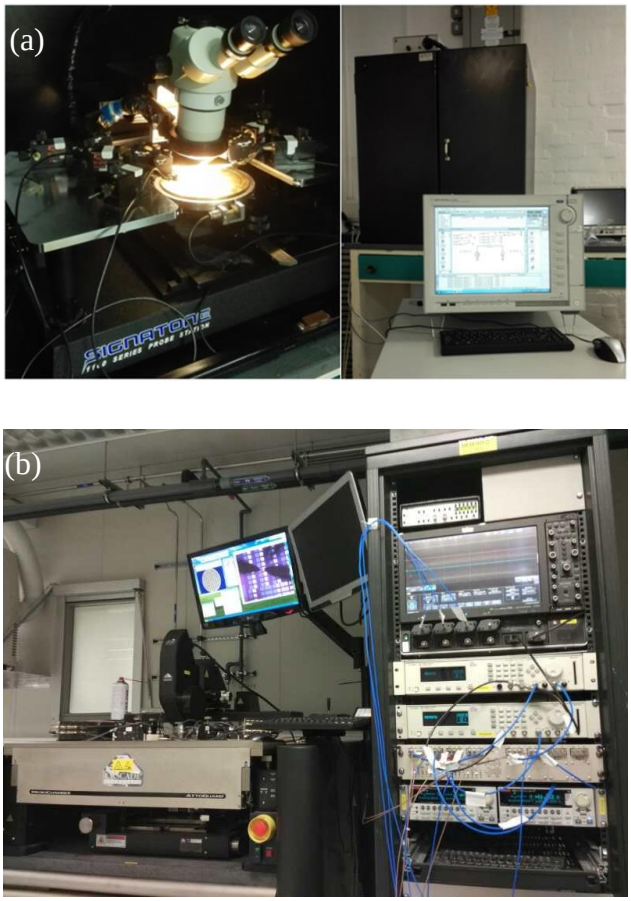


Figure 2-1 Measurement System (I) and (II). (I) is based on the Keysight 1500A and (II) is an all-in-one system.

2.2.1 DC Measurement System

As the Keysight B1500A includes 4 Source Measure Units (SMUs) which allows accurate DC measurement down to 10^{-12} A, if not specified otherwise, all the DC tests in this thesis are carried out using Measurement System (I) with the embedded software EasyExpert, including the RTN measurement.

2.2.2 AC measurement System

The AC switching can be done with either Measurement System (I) or (II).

In (I), the AC function is based on the B1530A WGFMU and the B1525A SPGU. The measurement capabilities of WGFMU include DC output and arbitrary waveform generation with 10 ns programmable resolution, high-speed voltage/current measurement, and wide current range available by dynamic ranging capability in arbitrary waveform output and dual channel output [131]. The advanced capabilities of SPGU include two-level and three-level pulse capability supported by each channel, flexible arbitrary waveform generation with 10ns resolution supported by arbitrary linear waveform generation (ALWG) function, and voltage monitoring capability with minimum 5μs resolution [132].

In (II), the AC function is based on the Agilent 81110A dual channel Pulse generators and the Keithley S46 Microwave switch system. The shortest pulse can be generated with 2 ns fall/rise times and 3 ns width, which makes it possible to deliver pulse width down to 5 ns to the device with good integrity, which is much shorter than (I). If not specifically stated, all the AC tests in this thesis are carried out with Measurement System (II).

2.3 Conventional characterisation techniques

In this section, the conventional characterisation techniques are presented, including the basic switching operations, physical characterisation techniques and electrical characterisation techniques for the understanding of resistive switching in RRAM devices.

2.3.1 Basic switching operations

As mentioned in Chapter 1, RRAM devices can be categorized into filamentary and

non-filamentary. Their corresponding operation setups are introduced as follows.

- Filamentary RRAM devices

As introduced in Chapter 1, reset/set cycling for filamentary RRAM devices can be carried out after an initial forming process (except for some forming-free devices). Depending on the polarity of reset/set voltage, the switching modes can be categorized into unipolar and bipolar, as shown in **Figure 2-2**.

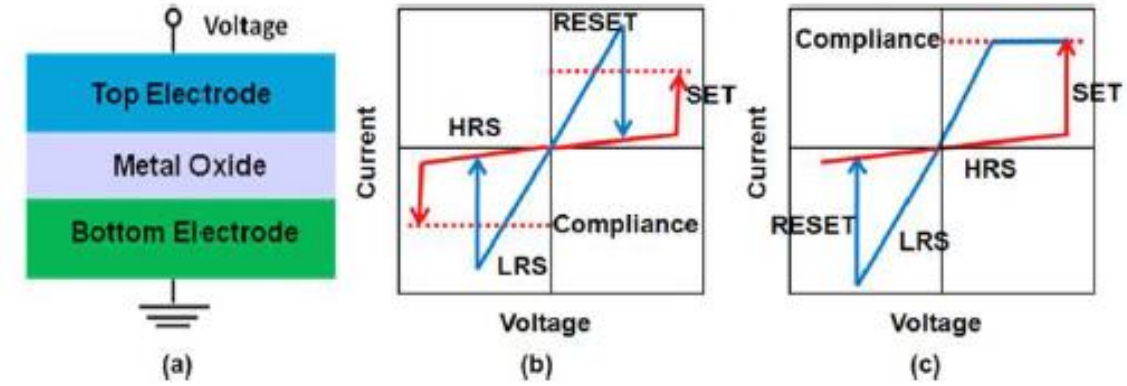


Figure 2-2 Schematic of a filamentary RRAM device (a) and schematic of I-V curves showing unipolar (b) and bipolar (c) mode [23].

For single RRAM devices the current compliance is usually provided by the semiconductor parameter analyser, for example, the compliance current setup in the DC measurement units. For 1T1R structure, the compliance current is provided by a selection transistor [23] as shown in **Figure 2-3**. During forming/set/read-out, a positive voltage is applied on the BL with SL grounded, and the current is limited by the voltage on WL. During reset, a positive voltage is applied on the SL with BL grounded, and the current compliance set by V_{WL} can be relaxed.

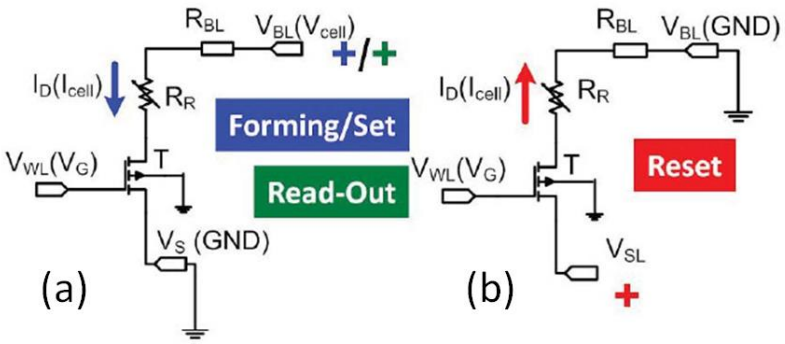


Figure 2-3 (a) Bias scheme for forming/set and read-out on the 1T1R device. (b) Bias scheme for reset on the 1T1R device [123].

- Non-filamentary devices

The switching operation of non-filamentary RRAM devices can be quite different from the filamentary ones. Take the a-VMCO RRAM as an example. It is forming-free and self-compliant so the initial forming process and external compliance setup is no longer necessary [42] as shown in **Figure 2-4**. This non-filamentary device requires a positive V_{TE} for reset and a negative one for set, which is opposite to most of its filamentary counterparts. The initial reset (black curve) is slightly different from the subsequent reset curves.

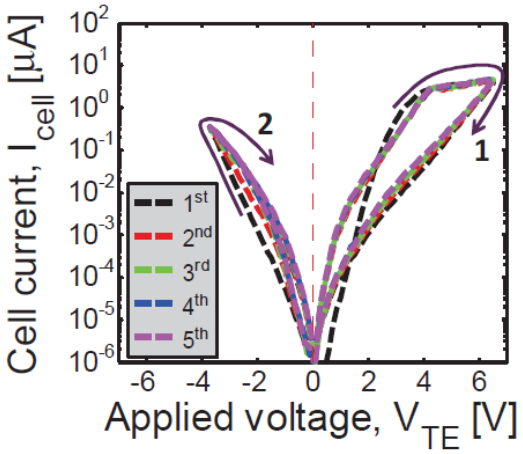


Figure 2-4 First 5 DC reset/set cycles for 40 nm a-VMCO RRAM devices. 1: Positive sweep for reset; 2:

Negative sweep for set [42].

2.3.2 Physical characterisation techniques

To effectively control the switching characteristics of RRAM devices, a thorough understanding of the switching behaviour is essential [21,104,133,134]. A lot of results have been achieved by tremendous efforts in the past decade for the underlying microscopic picture of the switching process. High-resolution physical analysis techniques, such as C-AFM, TEM and STM have been applied to probe the CF and a review will be given as follows.

- C-AFM

AFM is a popular high-resolution technique used for surface characterisation. The basic principle of AFM is whereby a fine tip is scanned across the surface of the sample to measure surface morphology and properties to construct a 3D image of the surface. Celano et al developed an AFM-based tomography technique, of which a diamond tip is used as a scalpel to remove material in a controlled way and at the same time, the spatial variation in conductivity is collected during the entire process [135] as shown in **Figure 2-5**. At different heights, two-dimensional (2D) conductivity profiles are collected in this way. A full tomogram of the CF is obtained by combining those profiles into one 3D-representation using computer software. High vacuum environment is required to minimize the possible re-oxidation of the CF. It is demonstrated that the switching occurs through the formation of a single CF. The CFs exhibit sizes below 10 nm and present a constriction near the oxygen-inert electrode. Different atomic-size contacts are observed as a function of the programming current, providing evidence for the CF's nature as a defects modulated quantum contact. Unfortunately, at HRS, this method

does not have the required sensitivity to probe the CF due to the reduced conductivity [136].

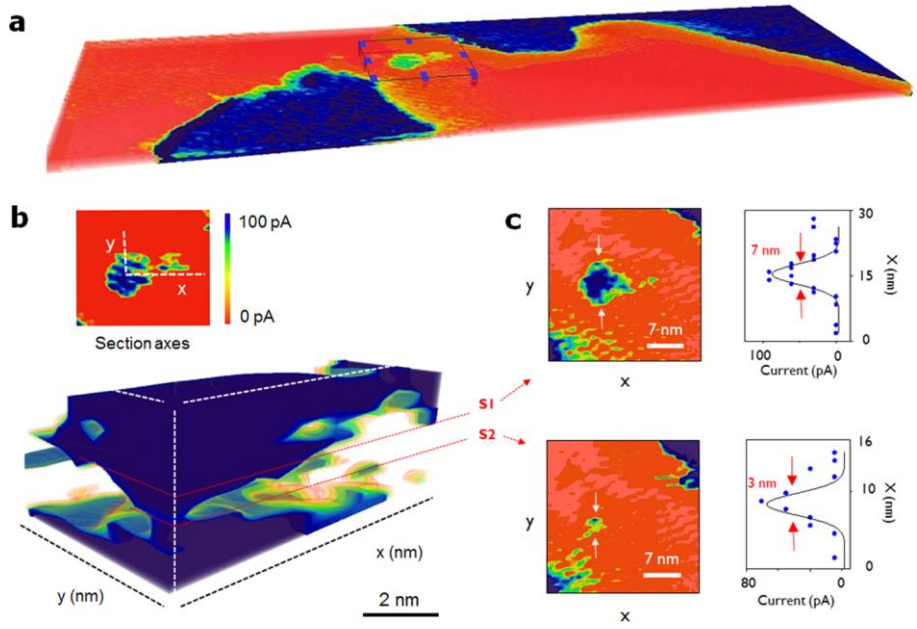


Figure 2-5 Three-dimensional observation of the CF (a) Tomographic reconstruction of the cross-bar memory cell visualized by volume rendering after TE removal (scan size 800×200 nm^2). (b) A 2D zoom into the region containing the CF and observation by volume rendering and iso-surface at fixed threshold (blue shape) for the CF under investigation in 5 nm thick HfO_2. The CF is shown in double cross-section. The low current contribution in the tomogram is suppressed to enhance the contrast of the highly conductive features. (c) A 2D observation of the CF section-planes (left panels) and C-AFM spectra (right panels) to determine effective CF size [136].

- TEM

TEM is a microscopy technique in which a beam of electrons is transmitted through an ultra-thin specimen, interacting with the specimen as it passes through it. Based on TEM, Miao et al managed to find and characterize the active conduction channel applying the following steps as shown in **Figure 2-6**: (1) precisely locating the active switching region(s) through pressure-modulated conductance microscopy (PMCM) on a functioning device; (2) cross-sectioning the active switching region(s) by focused ion

beam (FIB) milling; and (3) examining the structure and composition through high-resolution cross-sectional transmission electron microscopy (X-TEM) [137]. With this technique, a nano-scale conducting channel consisting of an amorphous Ta(O) solid solution surrounded by nearly stoichiometric Ta_2O_5 in Ta_2O_5-based RRAM devices is observed. This precisely designed technique successfully locates the key switching component in Ta-based RRAM, but the high complexity makes it time consuming and less capable to collect statistical results.

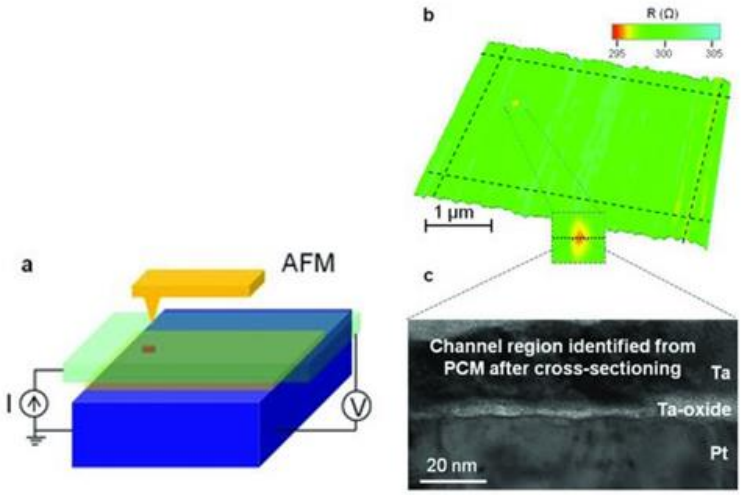


Figure 2-6 Identification and visualization of the conduction channel. (a) Schematic illustration of PMCM, for which a non-conducting AFM tip applied pressure to the TE while the resistance of the device was monitored, yielding a resistance map as a function of tip position. (b) The resistance map of a TaOx-based memristor, where the red dot (resistance decrease), highlighted by the dashed square in the magnified inset, corresponds to the conduction channel. The colour scale represents the measured resistance values. The conduction channel was cross- sectioned by FIB across the centre (indicated by the black dashed line in the inset). (c) TEM image of the conduction channel region identified from PMCM [137].

- STM

STM, which is based on the concept of quantum tunnelling, provides the ultimate lateral resolution and a noncontact option for electrical manipulation of resistive switching

materials. It facilities high-resolution imaging of topographic features and enables detection of the early stage of the resistive switching process [138]. Furthermore, STM in the spectroscopy mode (STS) uncovers the electronic structure of devices, characterized by tunnelling I-V measurements. Moors et al used combined STM/STS to study the switching mechanisms in RRAM devices as shown in **Figure 2-7**. Local resistive switching of TaO_x and $SrRuO_3$ thin film is demonstrated with an STM tip without necessitating a TE in physical contact with the surface. Resistive switching in TaO_x strongly depends on the non-stoichiometry that is affected by the sample history while $SrRuO_3$ shows resistive switching via oxidation/reduction and migration of oxygen ions [139]. This technique is limited to probing the surface of bare material.

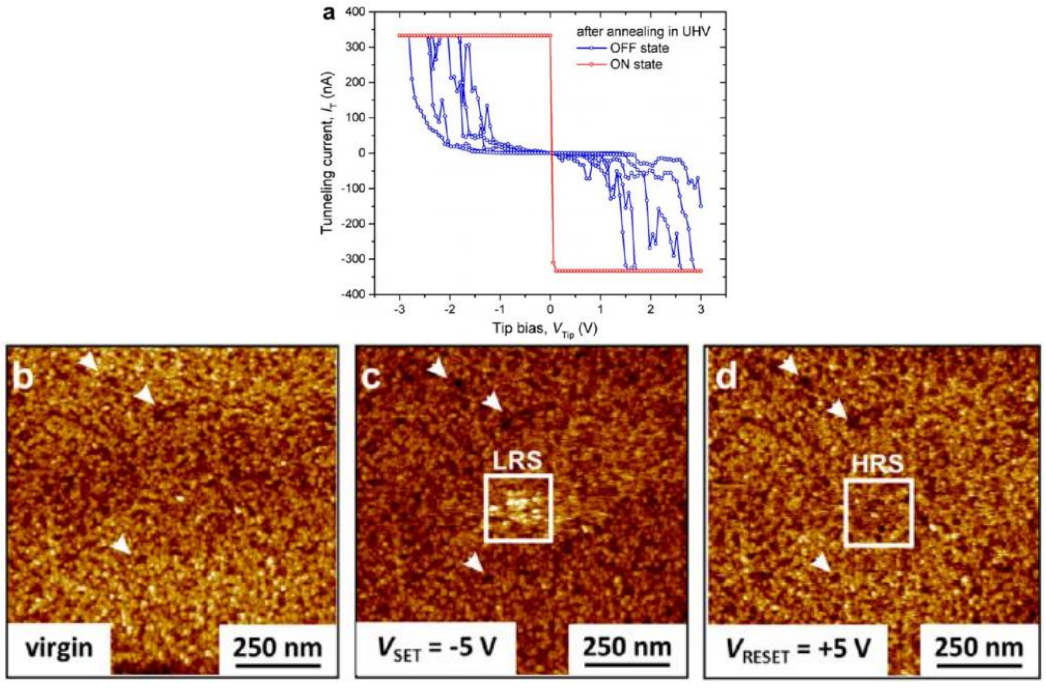


Figure 2-7 (a) I−V curves on bright regions corresponding to LRS and on unmodified regions corresponding to HRS on TaO_x thin films after annealing in ultra-high vacuum (UHV). (b−d) STM images (1000 nm×1000 nm, taken with V_{Tip}=−3.0 V and I_T=0.5 nA) of the UHV annealed TaO_x film (b) before and (c) after scanning the marked area at the centre with V_{Tip}=−5.0 V; (d) after additionally scanning the marked area in (c) with V_{Tip}=+5.0 V. The arrows in (b), (c), and (d) show the features on the surface that are the same and, thus, used as fiducial markers in the consecutive images [139].

In summary, those physical characterisation methods could, to certain extent, provide evidence for the existence of CF in filamentary RRAM devices and help understand resistive switching, but each of them has its own limitations. C-AFM cannot be used for HRS, TEM is too complex and STM is limited to the surface of bare material. In general, those physical characterisation techniques require expensive instruments, ultra-clean environment, and precise sample preparation, which make statistical characterisations unrealistic. Also, devices are destructed during characterisation. Compared with those physical characterisation techniques, electrical characterisation techniques are generally quicker, more repeatable and normally non-destructive. Various electrical characterisation techniques have been developed for the study of resistive switching in RRAM devices and are reviewed here.

2.3.3 Electrical characterisation techniques

- AC conductance

In impedance spectroscopy, or AC conductance measurement, the impedance of a sample is measured over a wide range of frequency. Different regions of the material are characterized according to their electrical relaxation times or time constants [140]. It is a relatively simple technique and can be applied to a wide variety of materials, e.g. the resistive switching in metal oxides such as $Pr_{0.7}Ca_{0.3}MnO_3$, NiO and TiO_2 [141-144]. The AC conductance can be extracted as a function of applied small signal frequency to investigate the conduction and switching mechanism [145]. Yu et al shows that LRS conductance is independent of frequency, while HRS conductance has a corner frequency beyond which the conductance rises [146] as shown in **Figure 2-8**. The rise of AC conductance above a certain corner frequency is caused by the electron hopping

between the nearest neighbour traps in the conduction filament. The higher the HRS resistance is, the lower the corner frequency is, corresponding to a larger tunnelling gap formed in the ruptured CF region.

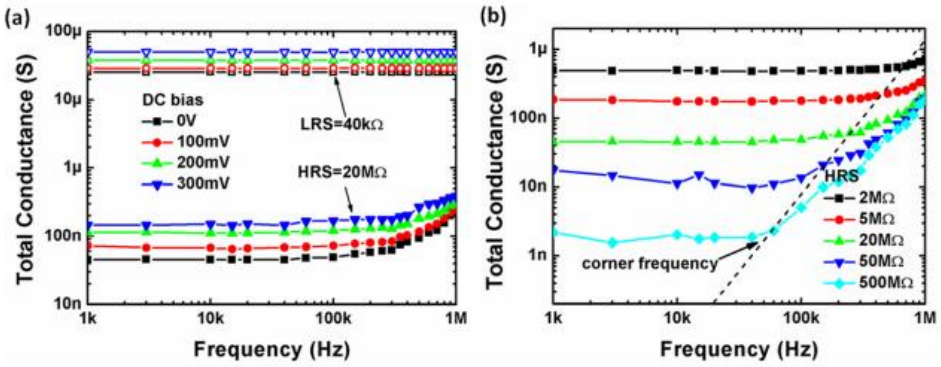


Figure 2-8: Device total conductance versus the applied AC small signal frequency for (a) typical LRS and HRS under different DC bias voltages and for (b) different HRS resistance levels. LRS conductance remains its DC value, while HRS conductance exceeds its DC value when passing a particular corner frequency. The higher the resistance is, the lower the corner frequency is [146].

- Low frequency noise (LFN)

LFN is not only one of the important issues in semiconductor devices, but also a useful tool whose characteristics can be utilized to analyse the internal physics of devices at defect level. Efforts have been made to investigate the LFN in RRAM devices in recent years as the explanations for resistive switching mechanism remain controversial. Yu et al investigated the LFN properties in RRAM devices with TiN/HfO_x/AlO_x/Pt structure as shown in **Figure 2-9** [146]. For both HRS and LRS the measured LFN is fitted to the $1/f^{\alpha}$ power law, with α~1 for LRS and α~2 for HRS, and a cut-off frequency in this transition, suggesting that CFs are ruptured and a tunnelling gap is formed during the reset process [146].

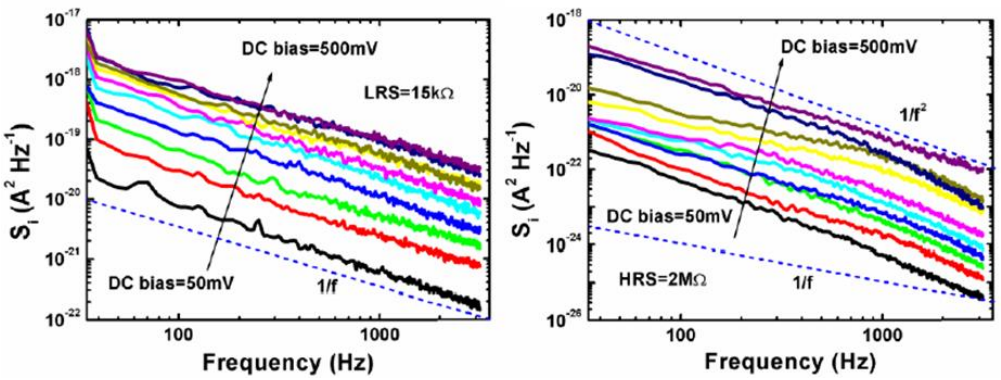


Figure 2-9: The PSD of LFN in (a) a typical LRS state (~15 kΩ) and (b) a HRS state (~2 MΩ) at different DC bias voltages, respectively. The larger the bias is, the larger the PSD is. $1/f$ and $1/f^2$ spectrums are shown respectively for LRS and HRS [146].

Fang et al studied the LFN characteristics in TiN/HfO_x/Pt structured RRAM devices with different sizes, and confirmed that current conduction is localized without the area dependence at LRS, whereas at HRS, there is a uniform leakage current throughout the whole device area [147], as shown in **Figure 2-10**.

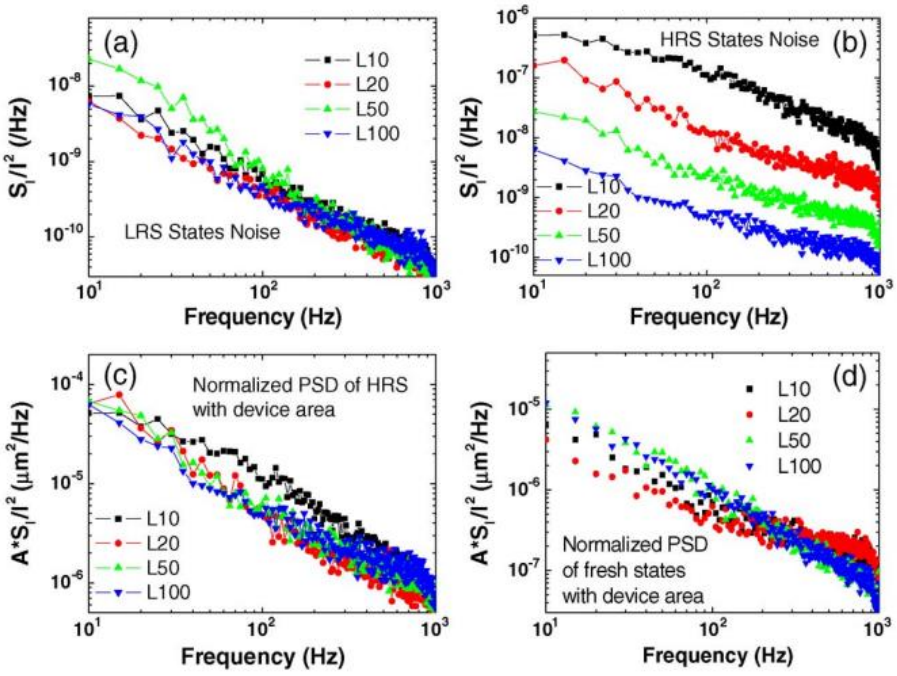


Figure 2-10 LFN PSD of (a) LRS (b) HRS (c) normalized LFN PSD with device area of the HRS, and (d) fresh resistance [147].

- I-V measurement and simulation

Degraeve et al proposed an analytical hour-glass model which utilizes the quantum point contact (QPC) conduction theory based on I-V measurement and simulation in a TiN/HfO_2/Hf/TiN filamentary RRAM device [148]. The CF is described as an hour-glass shaped sub-stoichiometric HfO_{2-x} region where fast ion drift-diffusion occurs, formed by the initial electroforming. The non-linear I-V curves in the low current range are modelled by the QPC approach. The close-range interaction of Vos inside the filament forms a conduction band and in the narrowest point of the filament, i.e. the constriction, current-controlling energy barriers exist. The subsequent switching is determined by the changes in the number of Vo in its constriction. Based on that knowledge, a filament schematic consists of a top reservoir (TR) and bottom reservoir (BR), connected by a constriction (C) with variable width as shown in **Figure 2-11(c)** [149]. This resembles an 'hour-glass' with a variable nozzle size and 'sand' (i.e. Vo) that move from TR to BR or vice versa depending on the applied polarity. Based on this model, the reset is described as a dynamic equilibrium process as shown in **Figure 2-12** and the set as a constriction growth limited by ion mobility and current compliance [149], as shown in **Figure 2-13**.

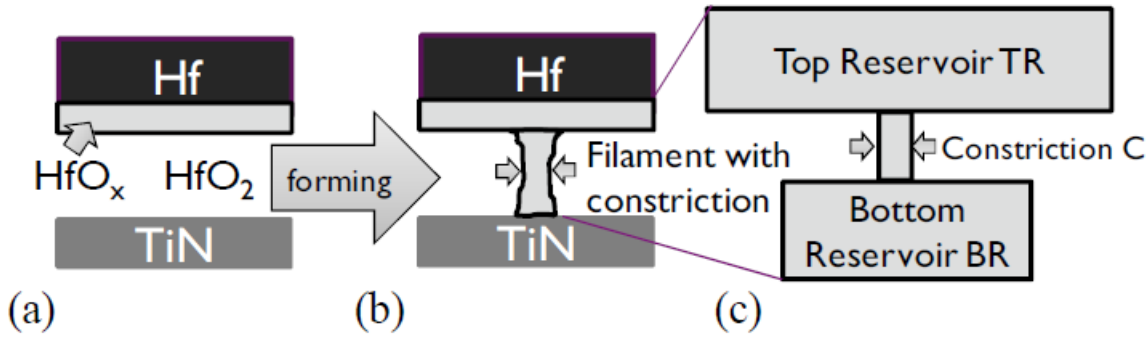


Figure 2-11 (a) Stoichiometric HfO_x is present between Hf electrode and HfO_2 dielectric. (b) Forming extends the HfO_x region to the filament. (c) Filament is modelled as a container with a top and a bottom Vo reservoir (TR and BR), connected by a constriction C with variable width. [149].

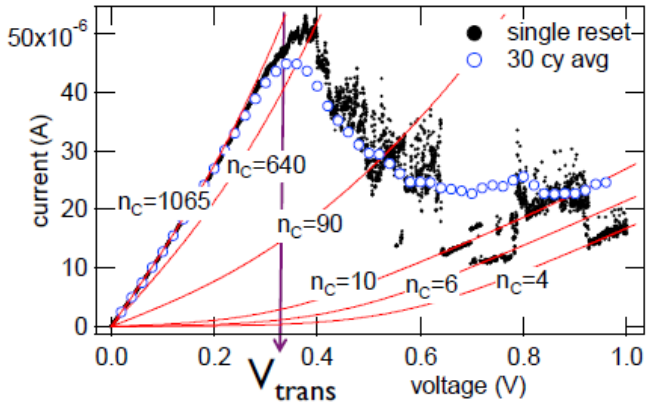


Figure 2-12 Example of a slow DC reset (0.1 V/min). The current jumps abruptly to different levels. The smooth curves are QPC calculations for integer number of defects n_C in the constriction [149].

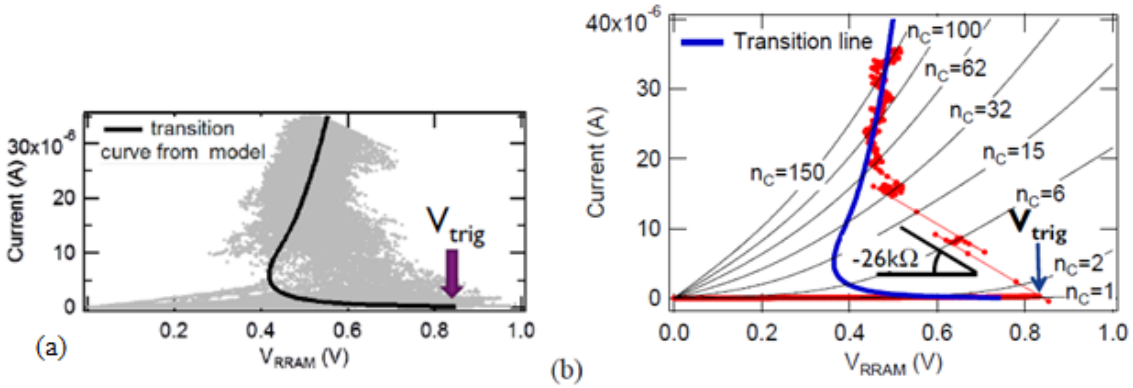


Figure 2-13 Set behaviour with high resistive load and small V_{step}. (a) shows 100 set curves on a single RRAM device. Systematically, a V_{trig} needs to be overcome, followed by a rapid 'snap back' to a lower voltage. The voltage transition line calculated with the model is shown. (b) shows an example of a single set curve. QPC I-V curves for corresponding integer number of constriction defects n_C are added. The resistive load was 26 kΩ. Vo start to move at V_{trig}. The constriction grows rapidly until the transition line is again reached [149].

- Monte Carlo simulation

Huang et al proposed the conical-shaped filament model based on Monte Carlo simulations. By using Monte Carlo simulation, it is observed that the evolutions of filament geometry during set are quite different from reset corresponding to the same

resistance level [150, 151]. In contrast to the hour-glass model in which the resistive switching is caused by the changes of Vo number in the constriction, in the conical-shaped filament model, the whole CF is ruptured firstly at the TE and then extended into the interior as voltage increases during reset as shown in **Figure 2-14**. The CF is formed firstly in the ruptured region and then gradually extended along the radius direction as current increases as shown in **Figure 2-15**. The mechanism has been attributed to Vo-O pair generation and recombination [152].

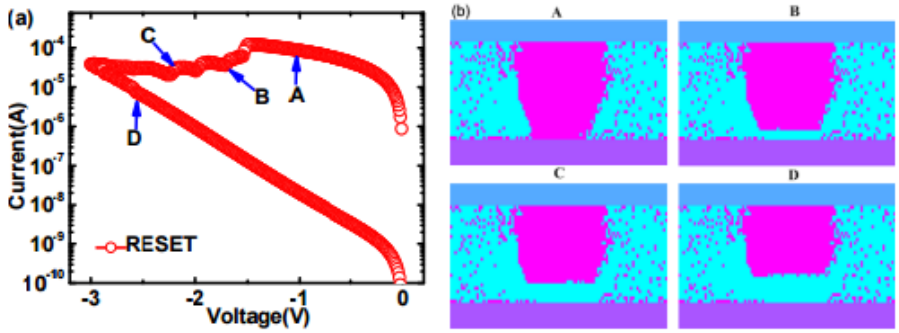


Figure 2-14 (a) I-V curve in reset process under DC sweep. (b) CF geometry at different point in (a) [151].

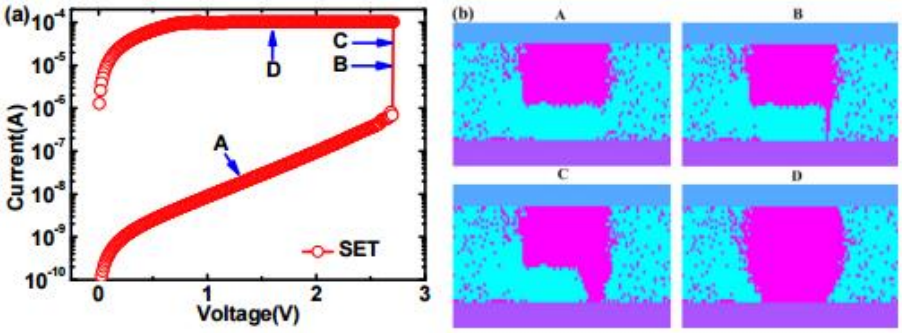


Figure 2-15 (a) I-V curve in set process under DC sweep. (b) CF geometry at different point in (a) [151].

There are clear controversies, i.e. between the hour-glass model and the conical-shape model, regarding the filament shape and resistive switching mechanism, which should be clarified. It also should be noted that those techniques are largely based on

simulation and there's a lack of direct experimental result to support either of them. Therefore it is of great importance to develop experimental techniques and link the theoretical models with microscopic evidence at the defect level for a comprehensive understanding of resistive switching.

It should be noted that all these techniques and models are developed for the filamentary RRAM devices only. This is because filamentary RRAM devices have a longer development history and represent the mainstream RRAM technology. However, as mentioned in Chapter 1, novel non-filamentary RRAM devices are attracting increasing attentions because of their unique switching mechanisms and promising features such as highly-scalable, self-compliance, forming-free and self-rectifying. Non-filamentary RRAM devices have been developed in recent years with different materials and structures, but there's still a lack of in-depth explanation of their switching mechanism based on experimental results [42] [67].

2.4 Introduction to RTN

RTN, the random current fluctuation between discrete levels, has been known since the days of vacuum electronics [153]. Later in the 1950s and 1960s, it was observed in the current flowing through bipolar semiconductor devices. The first observation of RTN in MOSFETs was reported in 1984 [154]. Since then, RTN has been found in many different semiconductor devices including MOSFETs, LEDs, Flash memories and emerging memories such as RRAM devices [155, 156].

It is commonly accepted that RTN is caused by the random capture and emission of charge carriers into or out of traps in semiconductor devices. RTN has become a serious reliability issue in the aggressively scaled CMOS technology, because the device size is

so small that trapping/de-trapping of one individual defect has greater impact on the device performance [153]. RRAM devices with an area of less than 10nm×10nm have been reported [19] and RTN in RRAM has been reported to be very significant at HRS [157]. RTN technique will be further developed in this thesis as an important defect detecting tool for the understanding of resistive switching mechanism, for both filamentary and non-filamentary RRAM devices.

2.4.1 RTN time distributions

The RTN with two discrete current levels is characterized by three parameters: the high time (t_{high}), the low time (t_{low}) and the amplitude (ΔI). The high-current state of the RTN is taken as state 1 and the low-current state as state 0. It is assumed that the probability (per unit time) of a transition from state 1 to state 0 is given by $1/\tau_1$, with $1/\tau_0$ being the corresponding probability from 0 to 1. The transitions are instantaneous. Transition between state 1 and 0 is a Poisson process and therefore, the times in state 0 and 1 are exponentially distributed [158] as shown in **Figure 2-16**.

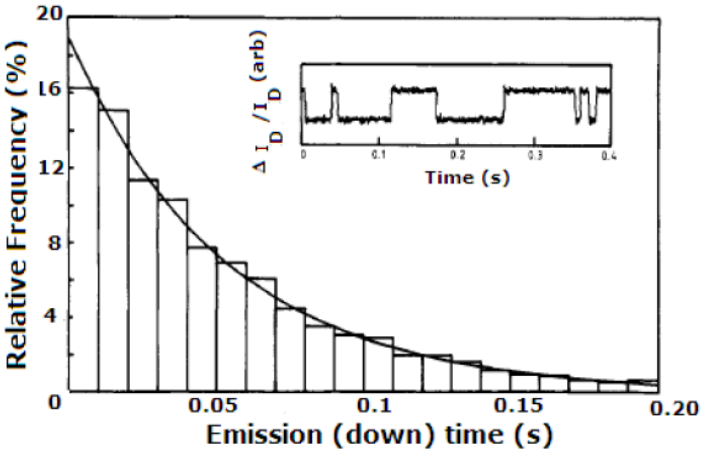


Figure 2-16 Distribution of 4425 emission times at 95 K and V_{GS}=1.15 V, showing that the time is distributed exponentially. τ_e=0.0528 s, standard deviation = 0.0505 s. The inset shows a portion of the I_D-t characteristic; the down time corresponds to emission [159].

Let p1(t)dt be the probability that state 1 will not make a transition for time t, then will make one between t and t+dt. Thus

$$p_1(t) = A(t)/\bar{\tau_1} \tag{2-1}$$

where A(t) is the probability that after time t state 1 will not have made a transition and $1/\bar{\tau_1}$ is the probability (per unit time) of making a transition to state 0 at time t. However,

$$A(t + dt) = A(t)(1 - dt/\bar{\tau_1}) \tag{2-2}$$

That is, the probability of not making a transition at time t+dt is equal to the product of the probability of not having made a transition at time t and the probability of not making a transition during the interval from t to t+dt. We can rearrange to give

$$\frac{dA(t)}{dt} = -\frac{A(t)}{\bar{\tau_1}} \tag{2-3}$$

Integrating both sides of **Equation 2-3** we find

$$A(t) = exp(-t/\bar{\tau_1}) \tag{2-4}$$

such that A(0)=1. Thus

$$p_1(t) = \frac{1}{\overline{\tau_1}} exp(-\frac{t}{\overline{\tau_1}}) \tag{2-5}$$

p1(t) is correctly normalized such that

$$\int_0^\infty p_1(t)dt = 1 \tag{2-6}$$

The corresponding expression for $p_0(t)$ is

$$p_0(t) = \frac{1}{\overline{\tau_0}} exp(-\frac{t}{\overline{\tau_0}}) \tag{2-7}$$

Hence, on the assumption that the high and low times are characterized by single attempt rates, we expect the times to be exponentially distributed. The mean time spent in state 1 is given by

$$\int_0^\alpha t p_1(t)dt = \overline{\tau_1} \tag{2-8}$$

and the standard deviation is

$$\left[\int_0^\alpha t^2 p_1(t)dt - \overline{\tau_1}^2 = \overline{\tau_1}\right]^{1/2} = \overline{\tau_1} \tag{2-9}$$

Equivalent expressions hold for the down state. Thus the standard deviation is equal to the mean time spent in the state.

2.4.2 RTN spectrum

RTN can be studied either in the time domain by investigating the time trace directly or in the frequency domain by applying a Fourier transform of the fluctuating two-level signal [154,158,159] as shown in **Figure 2-17**. The noise spectrum of RTN is a Lorentzian which is described by [158]

$$S_I = \frac{2(\Delta I)^2 \tau_0}{[4 + (\omega\tau_0)^2]} \tag{2-10}$$

with ΔI the RTN amplitude and τ_0 the characteristic time constant; the radial frequency is given by $\omega=2\pi f$. In other words, this particular spectrum is characterized by a plateau at relatively low frequency, a knee at a corner frequency and then a negative slope ~$1/f^2$ [153]. The characteristic time constant τ_0 is derived from the corner frequency fc of the Lorentzian, corresponding with half the plateau amplitude or with the maximum in $f \times S_I$, giving rise to:

$$2\pi f_c = \frac{1}{\tau_0} = \frac{1}{\tau_e} + \frac{1}{\tau_c} \tag{2-11}$$

Although the time-domain analysis of RTN can be potentially more complex and time constant extraction can be very time-consuming, it is more informative compared with the frequency-domain analysis, as the time-domain RTN analysis avoids the fitting problem caused by the degree of freedom when the spectrum is analysed.

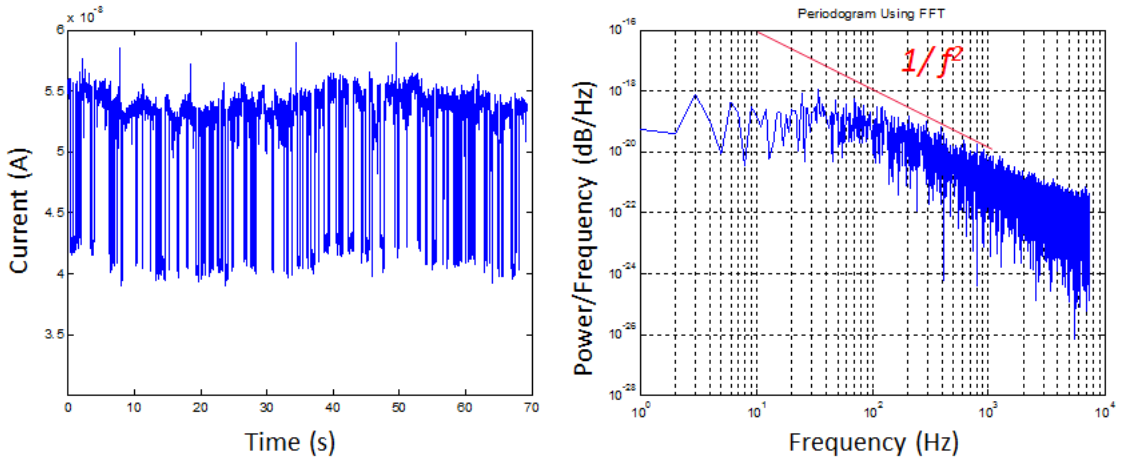


Figure 2-17 RTN time trace and corresponding noise spectrum in frequency domain after Fourier transform.

As device size keeps scaling down, it was observed that larger area devices give the ubiquitous 1/*f* spectrum, while small area devices exhibited RTN. The RTN is attributed to individual carrier trapping events. It was demonstrated that the 1/*f* noise in large devices is the result of a superposition of RTN signals as shown in **Figure 2-18**.

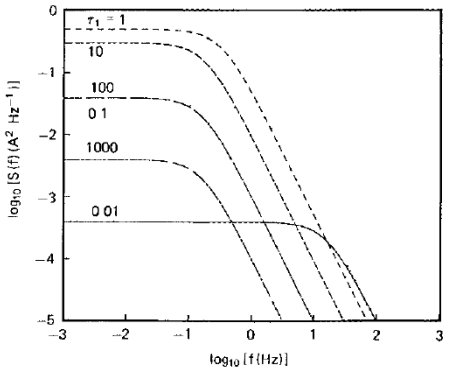


Figure 2-18 Lorentzian power spectra of multiple RTN signals: 1/*f* noise is the result of a superposition of RTN signals [159].

It is now clear that carrier trapping into states in the oxide drives the 1/*f* noise process [158]. Furthermore, as discussed in the introduction, RTN can be used as a novel approach to help clarify the noise generating mechanism in other physical systems.

2.5 RTN analysis tools

This section reviews several RTN analysis tools that are used for parameter extraction.

2.5.1 Time lag plot

The complexity of RTN is a major obstacle for understanding its behaviour, and time lag plot (TLP) is often used to overcome this issue [160]. TLP checks whether a time series is random or not. Random data should not exhibit any identifiable structure, while non-random structure in the TLP indicates that the underlying data are not random. TLP can be drawn by simply plotting points in an x-y plane, where x is set to the i^{th} sampled data, and y to the next $(i+1)^{th}$ sampled data as shown in **Figure 2-19**. For multi-level RTN, each level will be clearly shown as a cluster of points along the diagonal line in TLP, and their autocorrelation can be easily identified. Therefore, TLP makes it easy to count the number of multiple traps, and even makes it possible to recognize the existence of traps that are faster than the SMU sampling rate [160]. An example of the TLP for 2-level RTN is shown in **Figure 2-20**.

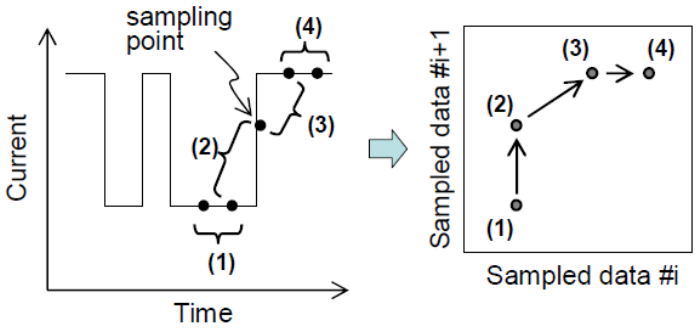


Figure 2-19 TLP visualizes autocorrelation of RTN levels in the time series [160]. In the TLP, x is set to the i^{th} sampled data, and y to the next $(i+1)^{th}$ sampled data.

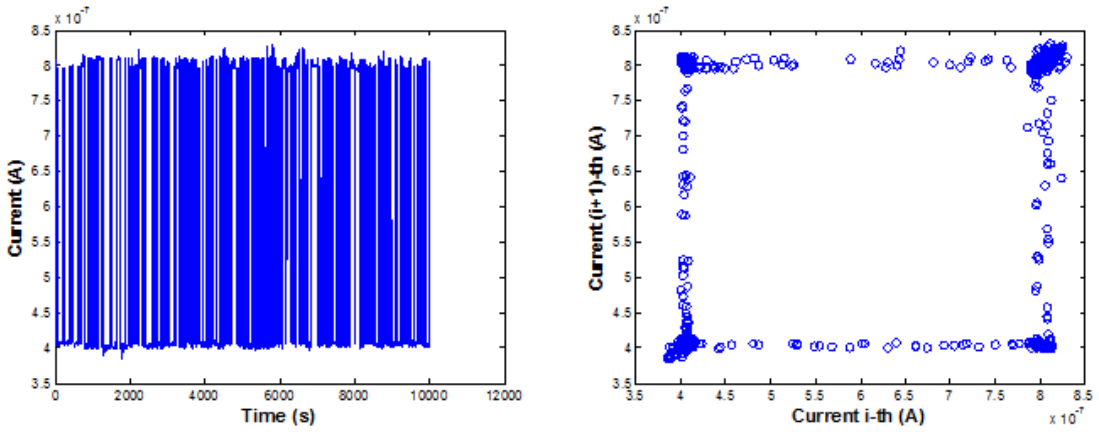


Figure 2-20 2-level RTN and its TLP. The 2 clusters in the bottom-left and top-right corner correspond to the low and high current level in the time trace.

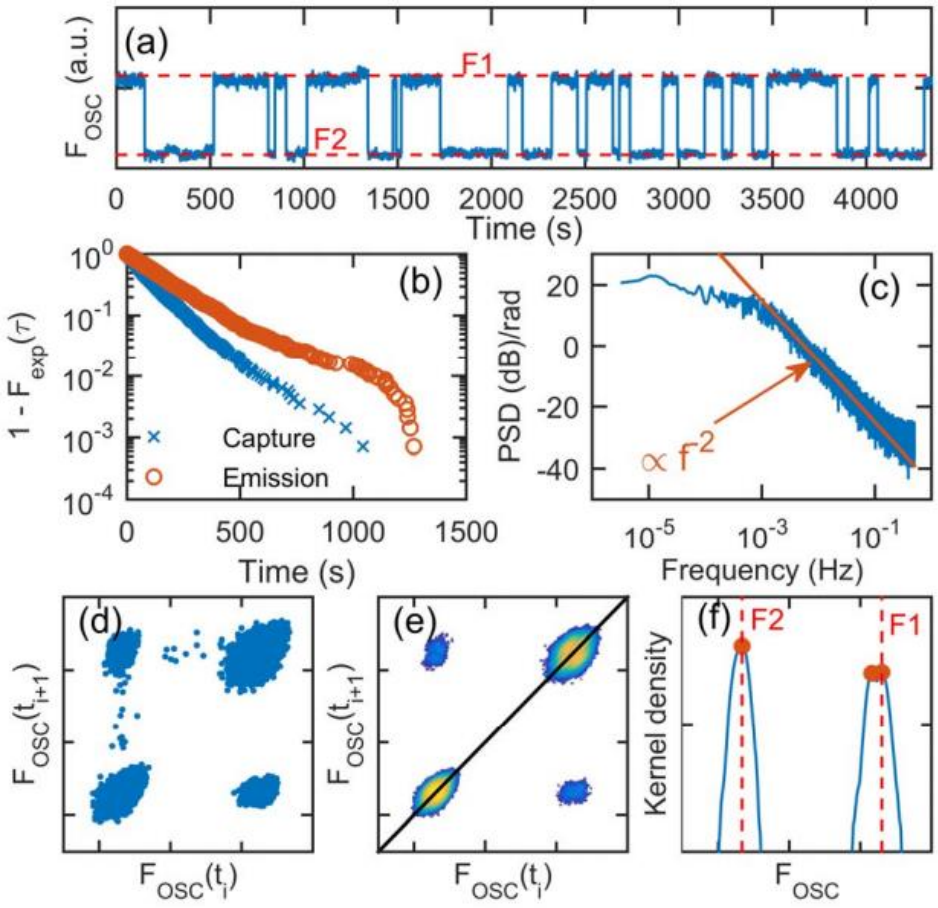


Figure 2-21 (a) Typical frequency vs. time for one of the DUTs. (b) Capture and emission times showing exponential distribution. (c) PSD of the signal showing expected Lorentzian shape. RTN level extraction scheme, starting with the lag plot (d), corresponding 2D kernel density (e), and peaks on the diagonal (f). The extracted levels (F1, F2) are shown on the trace in (a). [161].

The TLP can be improved by estimating its 2D kernel density and taking a diagonal cut. The location of the peaks on this diagonal are calculated to find the RTN levels in the time trace. This method effectively discards the irregular base fluctuations which appear like RTN in the conventional TLP. This method was developed by Dongaonkar et al for the RTN reflected as V_{th} fluctuation in 14nm transistors [161] as shown in **Figure 2-21**. Note in this work RTN in individual transistors is measured using a modified ring oscillator circuit so that the V_{th} fluctuation is reflected as the oscillation frequency.

2.5.2 Maximum likelihood

Conventionally RTN is measured in the sampling mode of semiconductor analysers under constant voltage stress (CVS). However, as RTN is a random process whose probability of occurrence is determined by many factors such as device, voltage, sampling rate and total measurement time, it is very difficult to do statistical RTN measurement for the in-depth analysis of time constants. Toledano-Luque et al developed a maximum likelihood (MLH) method considering the exponential dependence of time constants with voltage through the coupling factors (a_c and a_e) and first principles kinetics for the occupancy probability [162] as shown in **Figure 2-22**. This method was firstly used for the RTN fluctuation in I_D under different V_G in 3D Flash memory [162]. The good agreement between the time constants extracted from the conventional CVS technique at fixed V_G and this technique supports the validity of the MLH method. This method scans a large V_G window at once, increasing the probability to find RTN events, which shows promising potential for industry application while maintains a reasonable accuracy. In MOSFETs there are relatively fewer defects in the oxide and the RTN signal is normally quite clear. However, in RRAM this MLH technique can be difficult to apply, as there are a large number of

mobile defects in the oxide and RTN signal can be very complex. Optimizing the fitting parameters takes a lot of time and effort and thus the efficiency can be extremely low.

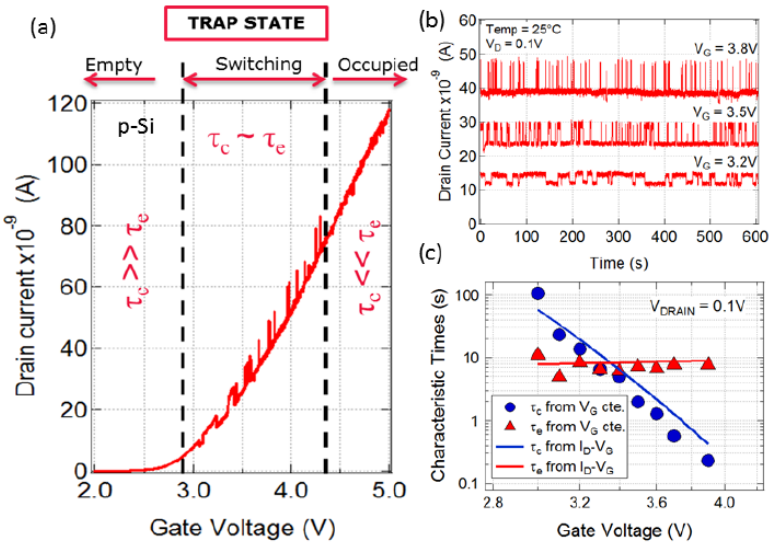


Figure 2-22 (a) RTN is clearly observable during the I_D-V_G sweeping. (b) Conventionally τ_c and τ_e are obtained by measuring the I_D at fixed V_G under CVS test. (c) The characteristic times obtained from the I_D-V_G curve and from the conventional CVS test agree very well [162].

2.5.3 Hidden Markov Model

For the extraction of RTN time constants in the time domain, several methods have been discussed in Section 2.4. RTN can be fitted accurately and efficiently if its levels are well separated. However, in RRAM, that is not always the case, as the signal obtained from CVS measurement is normally combined with a lot of other noises and current fluctuations, making RTN less clear. TLP can hardly deal with those signals and it costs a lot of time and efforts for parameter optimization if MLH is used. Therefore, a more robust method is needed for the RTN time constants extraction. In the following, the Hidden Markov Model (HMM) used in this thesis will be discussed in detail.

RTN can be described by an HMM, i.e. a Markov (memory-less) process with

unobserved (hidden) states. HMM analysis is a power tool commonly used in signal processing and pattern recognition [163, 164]. An HMM can be presented as the simplest dynamic Bayesian network. The mathematics behind the HMM was developed by Baum et al [165-169]. Whereas in simple Markov models the state of the system at each instant of time is directly visible to the observer, in HMM the output of the system is directly visible at each instant of time while the state of the system is hidden, even though the output strictly depends on the state [170]. Each state is characterized by a probability distribution over all the possible values assumed by the output, statistically linking the sequence of observations (output) to the sequence of hidden states. Moreover, each state is associated to a set of transition probabilities (one per each state) defining how likely it is for the system, being in a given instant of time, to switch to another of the possible states (including the same state) at the successive instant of time.

In HMM, a sequence of observations {Yt} t = 1…T is modeled by specifying a probabilistic relation between the observations and a set of hidden states St through a Markov transition structure linking the states. In this framework the state is represented by a random variable assuming one out of N values at each instant of time. The HMM approach relies on two conditional independence assumptions:

- S_t only depends on S_{t-1} (known as the first-order Markov or "memory-less" property)
- Y_t is independent of all other observations $Y_1,\ldots,Y_{t-1}, Y_{t+1},\ldots, Y_T$ given S_t.
- The joint probability for the state sequence and observations can be formalized as:

$$P(S_t|y_t) = P_{init} \cdot \prod_{t=2}^{T} P(S_t|S_t - 1) \cdot P(Y_t|S_t) \qquad (2\text{-}12)$$

$$P_{init} = P(S_1) \cdot P(Y_1|S_1) \quad (2\text{-}13)$$

A schematic representation of the HMM is given in **Figure 2-26** where the Markov property is evidenced. According to the formalism used by Rabiner et al [170], an HMM is completely defined as a 5-tuple (N; M; A; B; π). N is the number of hidden states, S, in the model (i.e. the number of discrete current levels to be found in RTN); since observations assume discrete values, M is defined as the number of distinct observable symbols (i.e. the possible current values assumed by RTN). A is an N-by-N matrix defining the transition probabilities among states and B is a N-by-M matrix defining the observation probability of each observable symbol in each hidden state; π is a vector defining the initial state probability distribution. The inference problem in this model consists in finding the most likely set of probability of hidden states given the observations. This is achieved through a maximum likelihood estimate of the HMM parameters given the observations using the forward-backward algorithm. Then the most likely sequence of hidden states representing the dynamics of the observations can be achieved via the "Viterbi" algorithm, a dynamic programming paradigm. As a result, HMM analysis can efficiently estimate the discrete current levels and the best sequence of states representing RTN data.

In the Matlab Statistic Toolbox™, functions related to HMM are shown in **Table 2-1** [171]. For RTN analysis, since we need the estimates of transition and emission probabilities of states, and based on that, the most probable state path is needed for the fitting, for which the hmmtrain and hmmviterbi functions are used.

Table 2-1: Matlab functions related to HMM, part of which the RTN analysis may rely on to get the most probable state.

Name	***Function***
hmmgenerate	Generates a sequence of states and emissions from a Markov model
hmmestimate	Calculates maximum likelihood estimates of transition and emission probabilities from a sequence of emissions and a known sequence of states
hmmtrain	Calculates maximum likelihood estimates of transition and emission probabilities from a sequence of emissions
hmmviterbi	Calculates the most probable state path for a hidden Markov model
hmmdecode	Calculates the posterior state probabilities of a sequence of emissions

In practice, the extraction process can be described as follows:

- RTN is measured in time trace as shown in Figure 2-23.

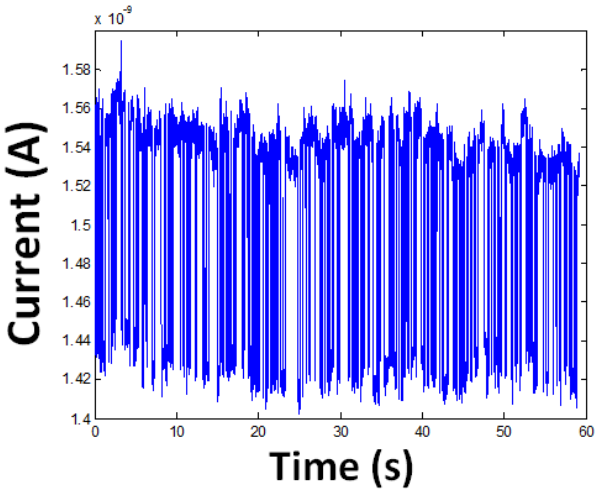


Figure 2-23 Time trace of RTN. This is a typical 2-level RTN despite of slight fluctuation in the base current.

- With Gaussian fitting on the histogram, we can get the initial estimates of the transition and emission probability matrices as shown in **Figure 2-24:**

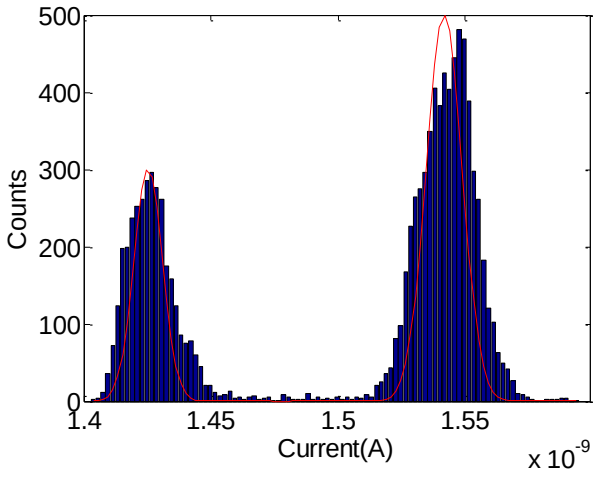


Figure 2-24 Gaussian fitting of RTN histogram. The peak of each Gaussian distribution corresponds to one RTN level.

- Output sequence is calculated with the HMM as shown in **Figure 2-25**.

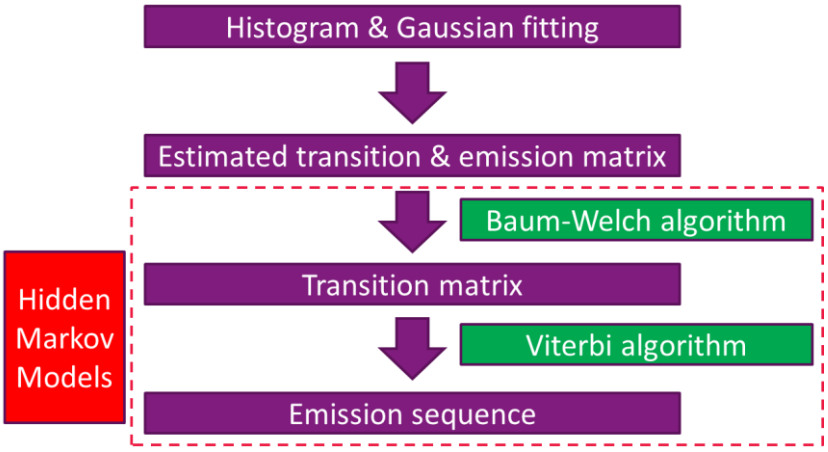


Figure 2-25 Flow chart of RTN extraction. RTN analysis is carried out with HMM using Baum-Welch and Viterbi algorithms.

- Schematic of HMM calculation: relation between states, possible observations, state transition probabilities and emission (output matrix) as shown in **Figure 2-26**.

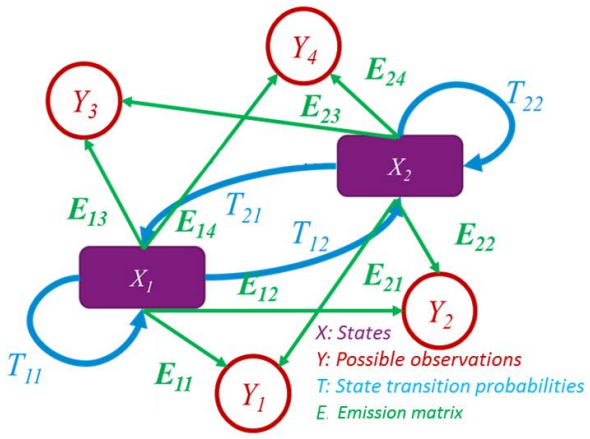


Figure 2-26 Schematic of HMM. X is the hidden states; Y is the possible observations, T is the state transition probabilities and E is the emission matrix.

- RTN is perfectly fitted. Time constants: τ_c=0.2182 s, τ_e=0.1157 s, as shown in **Figure 2-27**.

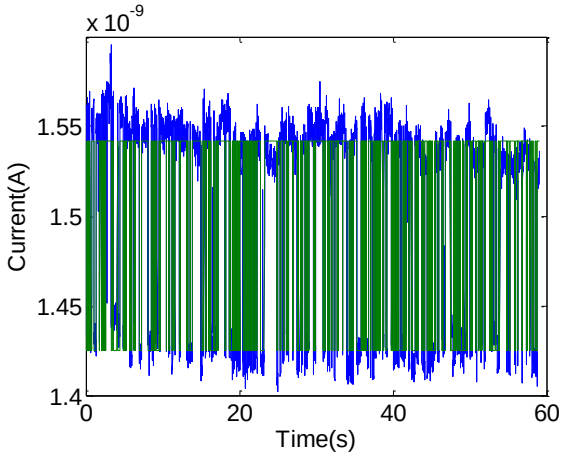


Figure 2-27 Extraction result. The raw RTN time trace (blue) is a good fit (green) with HMM and time constants can be extracted.

2.6 RTN analysis in RRAM

The early interest in RTN in memory devices was triggered by the observation of variable retention times in DRAMs caused by single defects in the depletion region [172] [173]. RTN started drawing the attention of the nonvolatile memory community only in

the last few years, but quickly became one of the main issues to be considered in the development of Flash and emerging memories. In Flash memories, the RTN-induced modulation of cell conduction can result in severe threshold-voltage (V_T) fluctuations, which could eventually affect the memory application [174].

Recently RTN in RRAM is attracting more and more attentions. RTN is more observed in HRS while 1/*f* noise prevails in LRS [157]. The statistical spread in RTN amplitude at HRS is decreasing with the devices area as the total number of traps reduces. Numerical models have been developed for RTN in RRAM. Puglisi et al identified two possible mechanisms responsible for the RTN current fluctuations, i.e. the Coulomb blockade due to electron trapping in adjacent defect sites, and the existence of more than one defect configuration for oxygen vacancies assisting the trap-assisted-tunneling transport [175]. Ambrogio et al studied RTN in RRAM with a 3D numerical model where the two-level fluctuation is explained by the change of charge state in a bi-stable defect close to the CF [176]. Lee et al investigated RTN properties in the α-TiO_x based bipolar RRAM devices and extracted the vertical location and energy of the trap that causes RTN [177]. Huang et al found the different behavior of the RTN currents generated by sweep and pulse operation modes respectively, and more soft-break-down paths are created by the latter mode [178]. Pan et al decoupled the discrete RTN and continuous average current fluctuation (ACF) in TaOx-based RRAM based on a noise power spectral density analysis at room temperature and at ultra-low temperature of 25K [179]. Chung et al found that 2-level RTN in HRS exhibits a large amplitude distribution tail as compared to LRS, which is caused by traps in the rupture region of HfO_2 mostly as is confirmed by the correlation between trap location and RTN amplitude [157]. Veksler et al demonstrated that read instability can be described by a Gaussian distribution, which can be used to estimate the number of RTN-related read failures in a memory array, thus

determining the intrinsic limit on the array size [180]. Raghavan et al linked the RTN phenomenon with the QPC model. RTN in RRAM is categorized into the e-RTN and Vo-RTN [181]. The e-RTN is caused by the standard electron-trap/detrap, while the Vo-RTN by the field-assisted Vo transport. RTN is measured with I-V sweep at a very slow ramp rate. The I-V curve is superposed on the QPC model based I-V prediction for a given set of $\{\omega_X, \omega_Y\}$ [148, 149] as shown in **Figure 2-28**. If the current starts to jump between two different levels (N_C, N_{C+1}), the voltage is identified as the disturb voltage which triggers Vo-RTN and the underlying Vo transport in/out of the constriction. The e-RTNs are ruled out because their amplitude does not fit the QPC estimated current levels. It is found that it is difficult to fit the deep reset case, probably due to the nucleation of a tunnel barrier (TUN) where QPC does not applies.

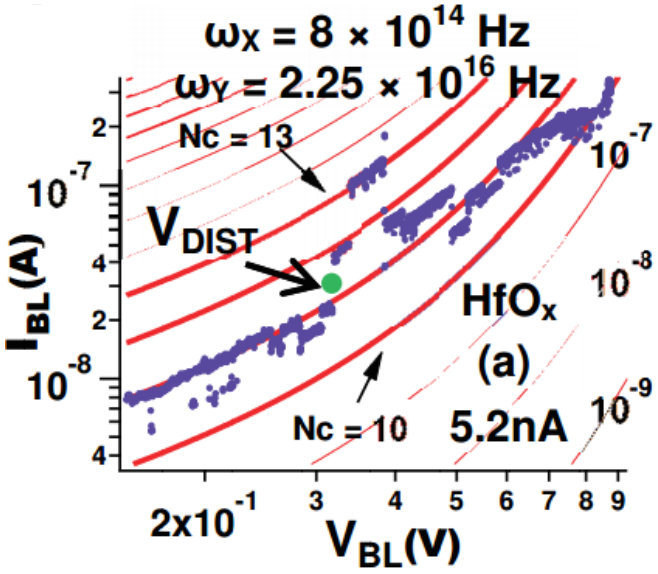


Figure 2-28 Methodology to identify V_{DIST} from slow I-V sweep in HRS by superimposing the data on theoretical I-V curves based on the QPC model, for integral number of defects in the constriction (Nc), given the parameters $\{\omega_X, \omega_Y\}$. All RTN fluctuations with current jumps lower than QPC-estimated current step are classified as electron trap/detrap events. [148, 149]

2.7 RTN-based defect extraction model in RRAM

2.7.1 Definition of time constants

The following qualitative explanation of the model is perhaps best understood with the aid of **Figure 2-29** [182, 183]. An electron which tunnels from one electrode to the trap can be captured to the excited state with a time constant τ_{tc}, in which case it tunnels very rapidly (~ps) out to either of the electrodes with a time constant τ_{te}. The current remains high for as long as this process continues. Alternatively, the electron may be captured to the ground state with a time constant τ_c, thereby closing the tunnel path through the excited state, and the current becomes normal (=low) [182]. From the ground state, the electron must be emitted thermally with a time constant τ_e before it can tunnel to either of the electrodes, or alternatively, to the top of the valence band. This is a much slower process than the two-step tunnelling across the excited state. After the electron emission, the trap relaxes and becomes ready to accept a new electron, which is captured either to the ground state again or to the excited state [182].

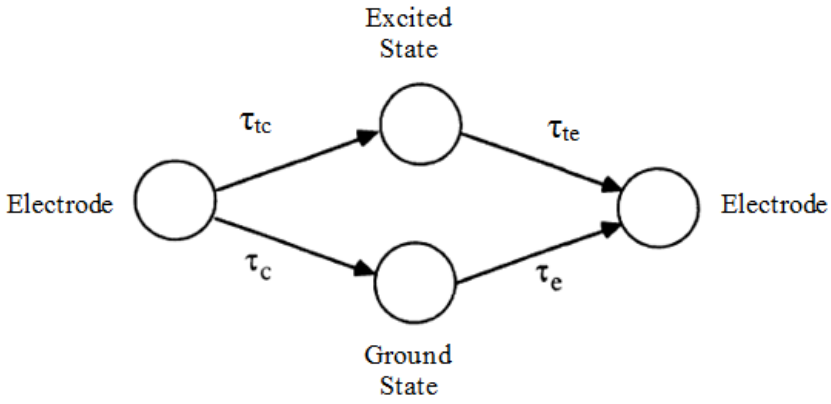


Figure 2-29: The essential electron transport paths of the model illustrated schematically. RTN is caused by traps at the ground state [182].

2.7.2 Defect extraction model for MIM structure

As introduced in Chapter 1, the fundamental RRAM structure is the MIM structure. A basic model can be easily developed based on this simple structure to extract the spatial and energy location of single RTN-responsible defects inside the oxide without considering any details in practice at this stage, and act as the fundamental model for the defect extraction in RRAM devices, which will be explained in detail in the following chapters.

In RRAM devices, the positive-charged Vos act as the current conducting defects. As introduced in the last sub-section, RTN is caused by the ground-state traps that temporarily close the tunnel path until the electron is emitted again and the trap come back to the excited state. This process can take μs, ms, seconds, days or even years. Therefore, the mean times of high and low current levels correspond to the capture and emission times, respectively. It should be noted that the term capture time means the mean time it takes for a capture to happen and the same for emission time.

It should also be mentioned that RTN shows only a small fraction of existing defects, because it is limited by the time resolution and the maximum measurement time, i.e. tester's patience. Kirton suggested that to obtain a reasonable estimate (10% error) of the up and down times of the RTN requires that a time record containing more than 200 transitions is stored. Similarly in order not to miss the transitions, the sampling rate must be at least 100× the average up and down time, implying a minimum time record length of 20 000 points [158]. Kapila et al confirms that confidence in extracted RTN emission and capture times can only be obtained through an assessment of the parameter's sampling frequency dependence. Incorrect choice of sampling rate can lead to orders of magnitude in error in extracted RTN τ_c and τ_e [184].

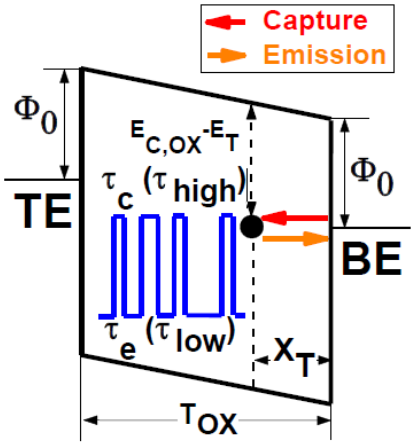


Figure 2-30: The energy band diagram of the metal-insulator-metal MIM structure considering the trap location and energy location, E_T and X_T, respectively.

The energy band diagram of the metal-insulator-metal MIM structure considering the trap energy level E_T and depth X_T is shown in **Figure 2-30**. The fractional occupancy of the trap is governed by:

$$\frac{\tau_c}{\tau_e} = exp\left(\frac{E_T - E_F}{k_B T}\right) \tag{2-14}$$

where k_B is the Boltzmann constant, T is the absolute temperature, τ_c and τ_e are the mean capture and emission time constants, respectively, E_T is the trap energy level, and E_F is the Fermi level. From the energy band diagram of the **Figure 2-30**, the expression for the capture and emission times in terms of the position of the trap can be derived as follow:

$$k_B T ln\frac{\tau_c}{\tau_e} = \Phi_0 - [(E_{Cox} - E_T) + E_x] \tag{2-15}$$

$$E_x = \left| q \frac{x_T}{T_{ox}} V_{ox} \right| \tag{2-16}$$

where Φ_0 is the difference between the work function of metal electrode and electron affinity of oxide, E_{Cox} is the conduction band edge of the oxide, q is the elementary charge, T_{ox} is oxide thickness, x_T is the position of the trap in the oxide from TE, and V_{ox} is the voltage applied on the oxide. **Equations 2-15** and **2-16** can be used to find the trap energy $E_{Cox}-E_T$. Also, by differentiating **Equations 2-15** with respect to the applied bias, the X_T is derived as

$$x_T = \frac{k_B T}{q} T_{ox} \frac{\partial}{\partial V} [ln(\tau_c/\tau_e)] \tag{2-17}$$

The bias-dependence of the τ_c and τ_e was investigated to characterize the spatial location of the trap in the oxide. **Figure 2-31** schematics the bias-dependence of the τ_c and τ_e measured at room temperature.

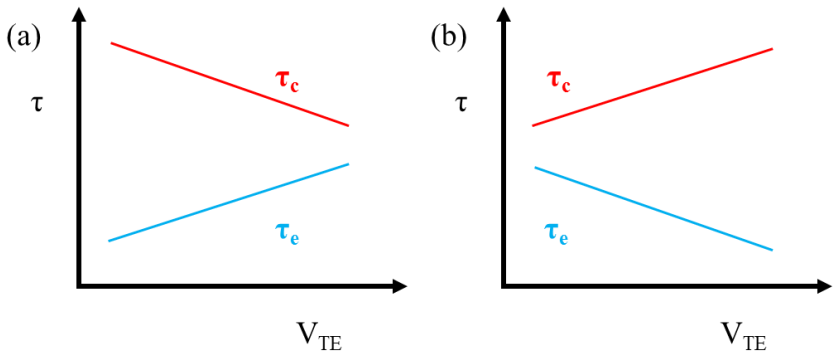


Figure 2-31: The bias-dependence of the τ_c and τ_e. In this example, as V increase, τ_c decrease and τ_e increase, which indicates the defect is interacting with the BE.

As V_{TE} increase, in **Figure 2-31(a)** τ_c decreases and τ_e increases, indicating that the trap is getting easier and easier to capture an electron from one electrode, and more and

more difficult to emit one back to the same electrode. Meanwhile, in the energy band diagram in **Figure 2-30**, if V_{TE} increase, E_T will be relatively lower and lower than $E_{F,BE}$ but relatively higher and higher than $E_{F,TE}$. Therefore, it can be derived that in **Figure 2-31(a)**, the trap is interacting with the BE, and on the contrary in **Figure 2-31(b)**, interacting with the TE. From the polarity of the $\tau_{c,e} \sim V$ slope we can distinguish which electrode the defect is interacting with, and based on this, choose the proper model for the extraction of defect spatial and energy location.

In the simple MIM structure with only one oxide layer, the extraction model is quite straightforward. The polarity of $d\ln(\tau_c/\tau_e)/dV$ determines whether the trap is interacting with the TE or BE, and the location can be extracted from its value. After which, the energy can be extracted from the interception of $\ln(\tau_c/\tau_e) \sim V$ curve against on the y-axis. This model is sufficient for the conventional MIM structured filamentary device. It should be noted that this is the simplest extraction based on the MIM structure. For its practical application in the RRAM devices, many other issues should be taken into consideration. For example, the RRAM device at HRS cannot simply be taken as shown in **Figure 2-30** without any modification because only a small part of the filament ruptures and metallic conduction may still exist in the other parts.

Many novel RRAM device structures consist of two or more layers, e.g. the a-VMCO RRAM has one barrier layer and one switching layer. The corresponding extraction formula should take the thickness and dielectric constants into consideration. RTN has been used for characterisation of oxide traps in high-k and metal gate MOSFETs [185] [186], which consists of a high-k layer and a SiO_2 interfacial layer has been used. During the extraction, similar to the single layer structure, the interacting electrode can be determined from the polarity of $d\ln(\tau_c/\tau_e)/dV$. The value of $d\ln(\tau_c/\tau_e)/dV$, i.e. time

constant variation rate, determines whether a trap is located in the high-k dielectric layer or the insulating layer. The time constant variation rate at the interface between the high-k dielectric and the insulating layer can be predicted and acts as the standard to justify if the defect locates in one layer or the other. This model can be easily applied into the dual-layer structure for RRAM, and even extended into the multiple-quantum-well (MQW) light emitting diodes (LED) [187].

The detailed implementation of the above models in RRAM devices will be presented in the following chapters in this thesis, i.e. in Chapter 3 for the single-layer model in filamentary HfO_2 devices, and in Chapter 4 for the double-layer model in non-filamentary a-VMCO devices.

2.8 Summary

In this chapter, the experimental systems and instrument for characterizing RRAM are presented and their principles are reviewed. Conventional physical and electrical characterisation techniques and their advantages and limitations are discussed. Based on the fundamental knowledge of RTN, recent progress in the analysis of RTN in RRAM is discussed, followed by the introduction of conventional RTN analysis techniques. The defect extraction model based on RTN in the MIM structure is introduced, laying the foundation for its further development in the remaining chapters of this thesis. In the following chapters, it will be shown that how this technique is applied onto RRAM devices with different materials and structures, even with different mechanisms. It will be shown that, the RTN technique is a useful tool for the in-depth analysis of the resistive switching mechanism at defect level without device destruction, and is also highly repeatable for statistical analysis.

3 Characterisation of HfO_2 RRAM

3.1 Introduction

There are a large number of materials that exhibit resistive switching behaviour, such as NiO, TiO_2, ZrO_2, HfO_2, Ta_2O_5, Al_2O_3 and SiO_2 [19, 39, 57, 71, 188-197]. A suitable RRAM oxide should have characteristics such as good CMOS compatibility, uniformity and thermal stability [29]. Among the above mentioned materials, HfO_2 has the advantages of having been employed as a high-k dielectric for the gate insulator for high-performance MOSFETs [198, 199], and defect-rich HfO_x is also a superior RRAM material [200] as it fulfils the above requirements. The choice of metal electrodes may also have impact on the resistive switching. RRAM devices with different electrode materials, such as Pt, Cu, and TiN [191, 201, 202], have been reported with HfO_2 as the dielectric layer. Metals such as Pt cannot be integrated in the CMOS process while the use of Cu faces risk of contamination. TiN is a metal commonly used in CMOS process and has been widely used for electrical contact in RRAM stacks. Govoreanu et al has reported excellent performances for a 10×10 nm^2 TiN/Hf/HfOx/TiN RRAM stack, which employs Hf as the capping layer [19]. The CMOS compatibility of both HfO_2 and TiN allows the TiN/Hf/HfO_x/TiN structured device to be stacked in 3D crossbar architectures with an effective memory cell area of $4F^2/n$, where n is the number of 3D stacked memory layers [23]. At present, TiN/Hf/HfO_x/TiN is one of the most matured RRAM structures that have been explored.

As mentioned in Chapter 2, details of the underlying microscopic picture of the switching process in HfO_2 RRAM devices are still largely missing. Furthermore, due to the lack of a clear picture of the CF at HRS, different CF profiles have been assumed in modelling and simulation, i.e. either with a conical shape or an hour-glass shape [71,149] as illustrated in **Figure 3-1**. In the two models, the CF modulation has been attributed either to the Vo-O pair generation/recombination at the interface of the BE, or to Vo movements along the CF, respectively. Direct experimental evidence of CF modification at Vo level is needed to provide insight to the CF rupture/restore process to improve our understanding of the switching mechanism and to advance the HfO_2 RRAM technology.

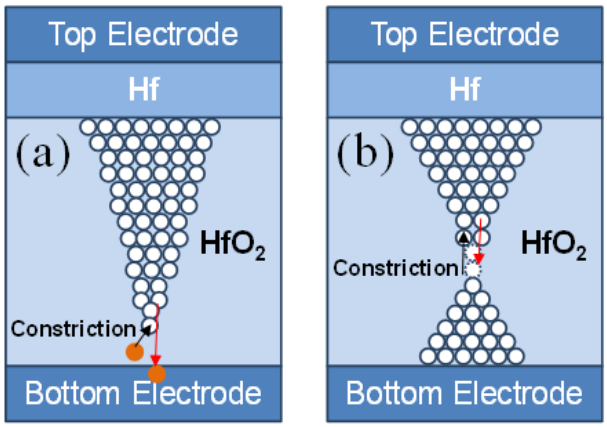


Figure 3-1: Structure of the HfO_2/Hf memory element and its resistive switching characteristics. Either (a) a conical shaped or (b) an hour-glass shaped CF has been proposed.

In this chapter, an RTN-based defect tracking technique (RDT) has been developed that can detect individual oxygen vacancy (Vo) movement and provide statistical information of CF modification during normal operations of nanoscale RRAM devices, without device destruction. This chapter is organized as follows: After a description of the devices and measurement setup in Section 3.2, the results and discussions are in Section 3.3, which includes observation of the critical filament region (CFR), verification of CFR by defect movement tracking, correlation of CFR modification with

operation conditions, endurance failure and recovery.

3.2 Devices and experiments

3.2.1 Devices

The RRAM device under investigation is a TiN/Hf/HfO_2/TiN memory cell patterned into a crossbar shape to achieve the smallest area while minimizing the impact of parasitic elements [19]. The device is formed in the overlap region between the bottom and top TiN electrodes, achieving 40 nm × 40 nm cell size, as shown in **Figure 3-2**.

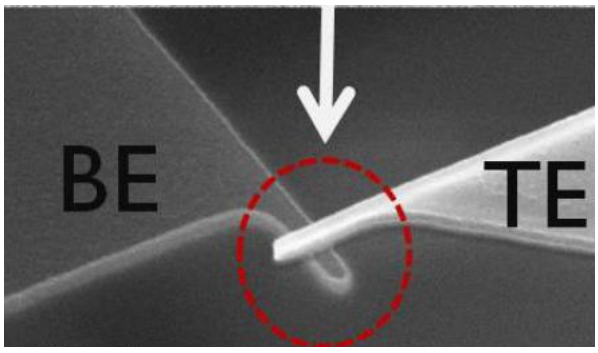


Figure 3-2 Top-view SEM picture of the Hf/HfO_2 crossbar memory device, which is formed at the intersection between the bottom and top TiN electrodes, achieving 40 nm × 40 nm cell size.

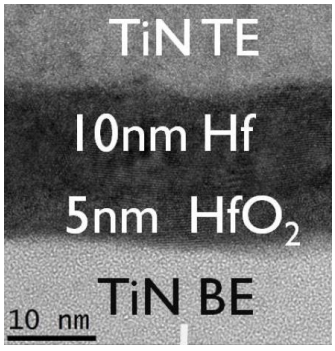


Figure 3-3 TEM of the RRAM. From bottom to top: TiN BE, 5 nm ALD HfO_2, 10 nm PVD Hf, TiN TE. The switching layer appears brighter due to the relatively lower concentration of Hf metal [203].

The 5 nm thick HfO_2 dielectric, deposited on a planarized TiN surface by atomic layer

deposition (ALD) $HfCl_4$ as a precursor and H_2O as oxidizing agent, is mainly amorphous. On top of the Hf film, a 10nm thick Hf metal-cap was deposited by physical vapour deposition (PVD) acting as an oxygen reservoir during the device operations, allows an increase in the concentration of mobile defects (Vo) in HfO_2, obtaining O-deficient, substoichiometric HfO_{2-x}, thus enabling the ionic and electronic behaviour required for resistive switching as is shown in **Figure 3-3**. After an initial electroforming process to form the HfO_{2-x} filament, typical current-voltage characteristics show a stable bipolar resistive switching behaviour: set to a LRS at less than 1 V, and reset to a HRS at about -1.5 V as is shown in **Figure 3-4**. The maximum current flowing through the device during the set is limited at a predefined compliance current, I_{cc}, between 50 μA and 150 μA to avoid breakdown. N-channel MOS transistors of 0.13 μm channel length were processed in a standard 65 nm process technology as to provide embedded compliance current. The resistive switching stack was fabricated afterwards with a back-end-compatible thermal budget, not exceeding 400 °C, forming the 1T1R structure [123]. The 1T1R RRAM device is used in the AC pulse switching test where an external compliance setup is not available as is shown in **Figure 3-5**.

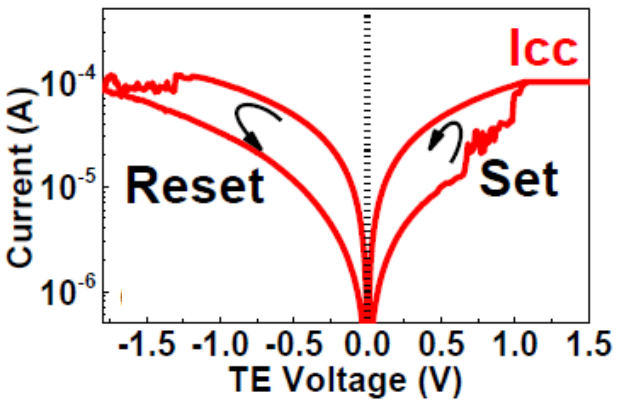


Figure 3-4 The bipolar resistive switching characteristics of a 40 nm HfO_2 RRAM cell. V_{set}=1.5 V, V_{reset}=-2.0 V, I_{cc}=100 μA.

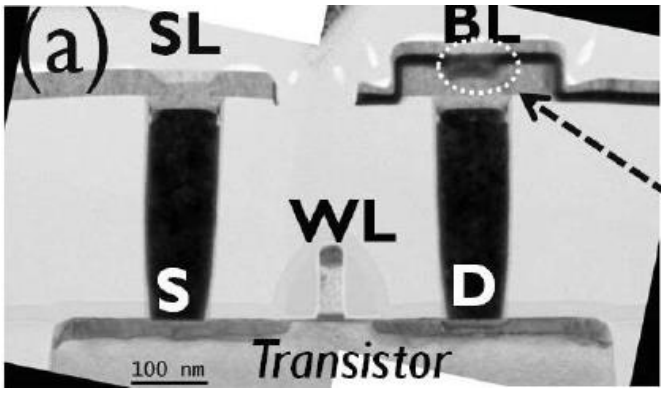


Figure 3-5 TEM cross-section picture of Hf/HfO_2 1T1R RRAM device. The Hf/HfO_2 RRAM stack is standing above the drain of the transistor [123].

3.2.2 Experiment methodology

As introduced in Chapter 2, RTN can be useful in providing the information on mobile defects (Vos) in RRAM devices and it is employed as the electrical measurement tool in this chapter for investigating the V_o that form the CF without destructing the device [159, 186]. RTN can be easily observed in nano-scaled devices. The size of our device under test (DUT) is only 40 nm×40 nm, in which the random trapping/detrapping of a single oxide defect has significant impact on the current conduction [153]. Also it can be observed that the high and low times depend on the TE biases applied, as introduced in Chapter 2 and shown in **Figure 3-6**.

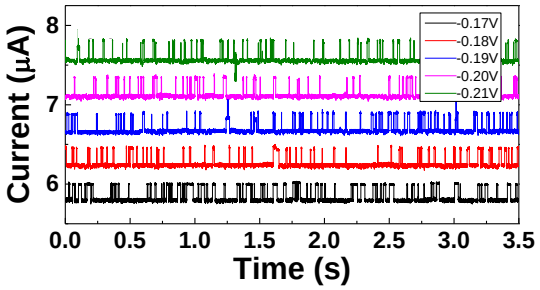


Figure 3-6 Typical RTN under various TE biases. The high and low time constants are dependent on V_{TE}.

As is mentioned in Chapter 2, the traps in RRAM devices are mostly positive charged oxygen vacancies. The mean time constants for high (τ_{high}) and low (τ_{low}) current levels in RTN are the capture (τ_c) and emission times (τ_e), respectively [182] [186]. The bias dependence of τ_c and τ_e allows not only for the extraction of defect location and energy level, but also the investigation of the mechanisms of electron transferring between the electrode and a single oxide trap, as shown in **Figure 3-7**.

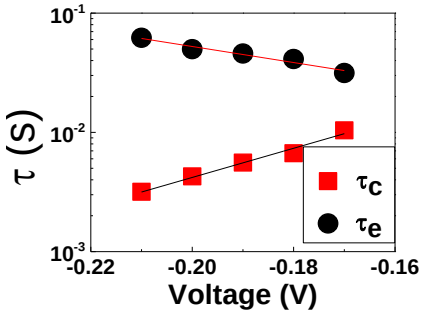


Figure 3-7 The mean capture and emission time constants (τ_c and τ_e) measured during the voltage sweep in **Figure 3-6**, from which defect's spatial and energy location (X_T , E_T) is extracted.

Based on **Equation 2-14** to **2-17**, the location and energy of a defect can be extracted with the experimental time constants. In particular, E_F is the Fermi level of the TiN electrode. Φ_0 is the difference between the work function of TiN and electron affinity of HfO_{2-x}, and is set to 1.4 eV [204].

Although many efforts have been made to investigate the RTN in RRAM devices, not many focus on the Vo extraction, mainly because complex signals are often observed [181, 204], to which the above equations are often not applicable. For example, defect movement (Vo) and electron trapping/detrapping can co-exist [181], both of which can result in current fluctuation and lead to complex signals. Moreover, **Equations 2-14** to **2-17** are only applicable to the cases that an electron tunnels from an electrode to an oxide defect, and then back to the same electrode, in which τ_c and τ_e should have the opposite polarity of bias dependence.

However, due to the nanoscale oxide thickness in RRAM devices, RTN can also result from electron tunnelling through the defect to the opposite electrode or to other defects, in which cases τ_c and τ_e show the same polarity of bias dependence. There also might be a large quantity of defects near each electrode, which leads to metallic-like local conduction. To avoid these complications, we only consider RTN that clearly follows the classical theory, as shown in **Figure 3-8**. All RTN signals are evaluated individually and we only use signals for which τ_c and τ_e have the opposite polarity of bias dependence. This ensures that the defects we examined are located in the electron tunnelling region of the oxide, and the RTN caused by electrons tunnelling through the defects are excluded. We also consider the metallic local conduction regions as part of the metal electrodes, thus T_{ox} in **Equations 2-14** to **2-17** is modified to $T_{ox,TCR}$, and $X_T/T_{ox,TCR}$ is the defect's relative spatial location within the electron tunnelling conduction region (TCR), as shown in **Figure 3-9**.

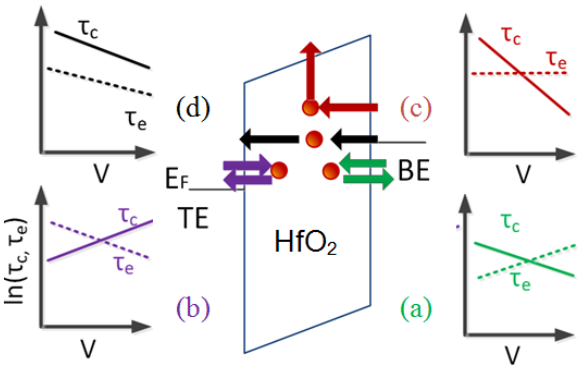


Figure 3-8: Energy band diagram illustrating various RTNs examined in this work caused by electron tapping and detrapping. (a) Defect interacts with BE; (b) Defect interacts with TE; (c) Electron is captured from one electrode, and then thermally emitted into the conduction band; (d) Electron tunnelling from one electrode (or trap) to the other. Defect's spatial and energy location (X_T, E_T) can be extracted from the τ-V_{TE} dependence for the RTNs in (a)-(c) by using **Equations 2-14 to 2-17** [158, 186]. These equations are not applicable for (d) where RTNs are caused by electron tunnelling through the defects to the opposite electrode, in which case τ_c and τ_e have the same polarity of bias dependence.

Based on this methodology, the CF in HfO_2 RRAM devices will be probed and Vo movement will be correlated to device resistance change, providing insight into the resistive switching mechanism.

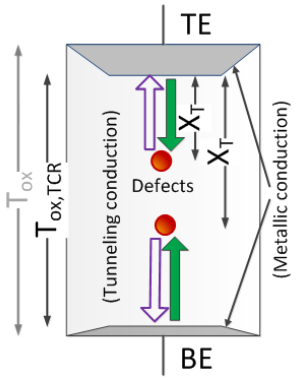


Figure 3-9 Illustration of the metallic conduction region near both electrodes, in which **Equations 2-14** to **2-17** are not applicable. Defect's relative location within the electron tunnelling conduction region TCR, X_T/T_{ox_TCR}, can be extracted from **Equation (3-1)** even though the absolute value of $T_{ox,TCR}$ is unknown.

3.2.3 Experiment procedure

An experiment procedure is developed for the RTN measurement on HfO_2 RRAM devices. The devices are switched on and off repeatedly between the LRS and HRS for a number of cycles as shown in **Figure 3-10**. During each cycle, the RTN measurement was carried out at HRS to extract the Vo locations under different biases, as shown in **Figure 3-11**. Metallic conduction prevails at LRS, where RTN is negligible. A large number of Vos can be collected during the cycling operation. To link defect movement with device resistance change, read-out is carried out at 0.1V after each stepping bias of RTN measurement.

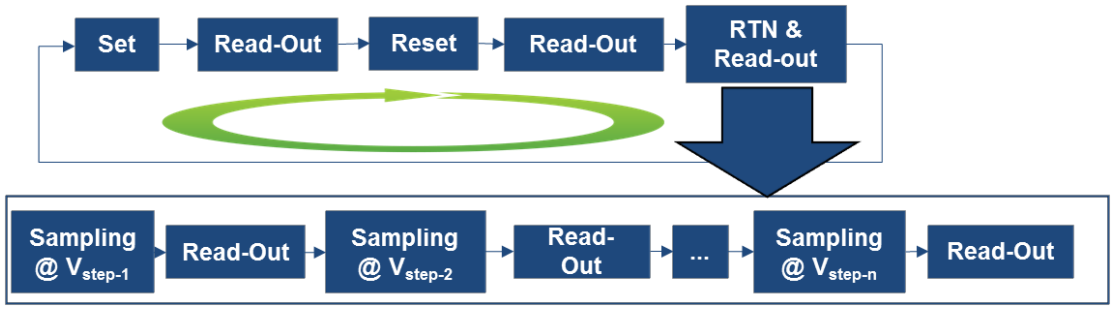


Figure 3-10 The RRAM devices are switched on and off repeatedly for a number of cycles. During each cycle, the RTN measurement was carried out at HRS to extract the Vo locations under different biases. To link defect movement with device resistance change, read-out is carried out at 0.1V after each stepping bias of RTN measurement.

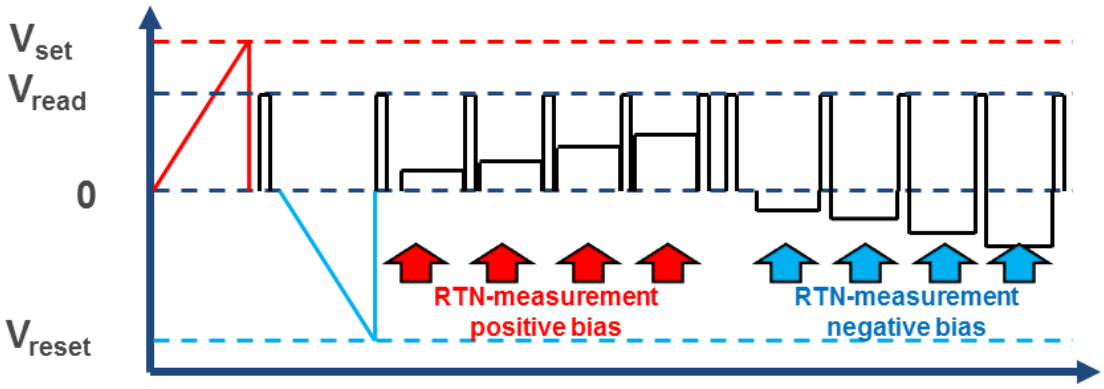


Figure 3-11 Schematic of voltage waveform. RTN measurement is carried out under both positive and negative biases.

3.3 Results and discussions

3.3.1 Observation of the critical filament region

The first key observation in this work, the constriction in the CF, was achieved by statistically analysing the defect spatial location and energy level during normal device operation cycles. As mentioned earlier, the switching behaviour of filamentary RRAM devices has been described by various controversial models. The analytical hour-glass

model utilizes the quantum point contact (QPC) conduction theory [149] in which the CF is described as an hour-glass shaped substoichiometric HfO_{2-x} region where fast ion drift-diffusion occurs, formed by the initial electroforming. The subsequent switching between LRS and HRS are determined by the changes in the number of Vo in its constriction. On the other hand, other Monte Carlo simulations and physics-based HSpice compact models [71] [152] assumed a conical-shaped CF instead, which is ruptured from the BE during reset and is restored back to the BE during set, and the mechanism has been attributed to Vo-O pair generation and recombination [152].

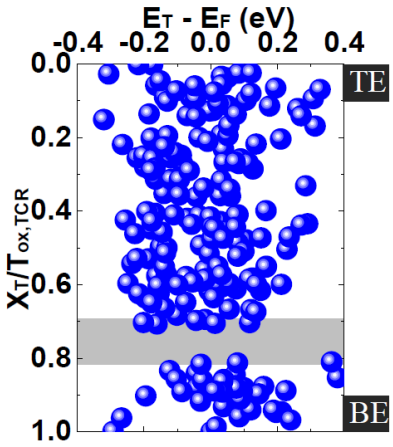


Figure 3-12 Defects detected at HRS during normal DC switching on/off cycles, which clearly shows a constriction (grey region) with the least number of defects near, but not at, the BE. DC V_{set}=1V, V_{reset}=-1.6 V, I_{cc}=150 μA.

In order to clarify this controversy, in this work, we switched on and off the devices repeatedly between the LRS and HRS for a number of cycles, during each cycle the RTN measurement was carried out at HRS to extract the defects locations. A large number of defects can therefore be collected during the cycling operation. A summary of the measured defects profile at HRS is shown in **Figure 3-12**, and a region (grey) with the least defects near, but not at, the BE can be clearly observed. This is the first time that the constriction region is experimentally observed entirely based on the

analysis of electrical measurement data, confirming the hour-glass shaped CF. In particular, the large number of defects observed nearby the BE does not support the conical model, which assumes that at HRS the filament is ruptured from the BE interface where the least defects should occur. The constriction is not located in the centre of the CF either, which is a consequence of asymmetric oxygen vacancy reservoirs. Since the 10nm Hf metal layer is intentionally inserted at the TE interface as the Vo reservoir to increase the Vo concentration in HfO_2 near TE, it is therefore understandable that the constriction is located near BE, agreeing with the results of hour-glass modelling [149, 181]. The width of the observed constriction is at nanometre scale, as the total HfO_{2-x} thickness is 5nm.

3.3.2 Verification of CFR by defect movement tracking

Since the constriction region has the least defects and hence the largest local resistance, the movement of defects into and out of the constriction should have the most significant impact on CF resistance, leading to CF modification and resistive switching. To verify this, we further developed the RTN technique to enable the tracking of individual defect movements, so that the missing information on defect drift-diffusion process in the HfO_{2-x} can be obtained experimentally and correlated directly with CF constriction modification and resistive switching.

As mentioned previously during RTN measurement read-out is carried out after each stepping RTN bias to monitor the device resistance at the same time. The voltage waveform is illustrated in **Figure 3-13(a)**. If there's no defect movement, the read-out result of device resistance will be constant throughout the RTN measurement and defects can be extracted with fixed location/energy from the RTN, as shown in **Figure**

3-13(b). Device resistance change and defect movement will happen simultaneously if there is defect movement involved during the RTN measurement, as shown in **Figure 3-13(c)**. Defect movement information is obtained by comparing the defect location/energy extracted before and after the movement, i.e. the abrupt jump in current and in device resistance.

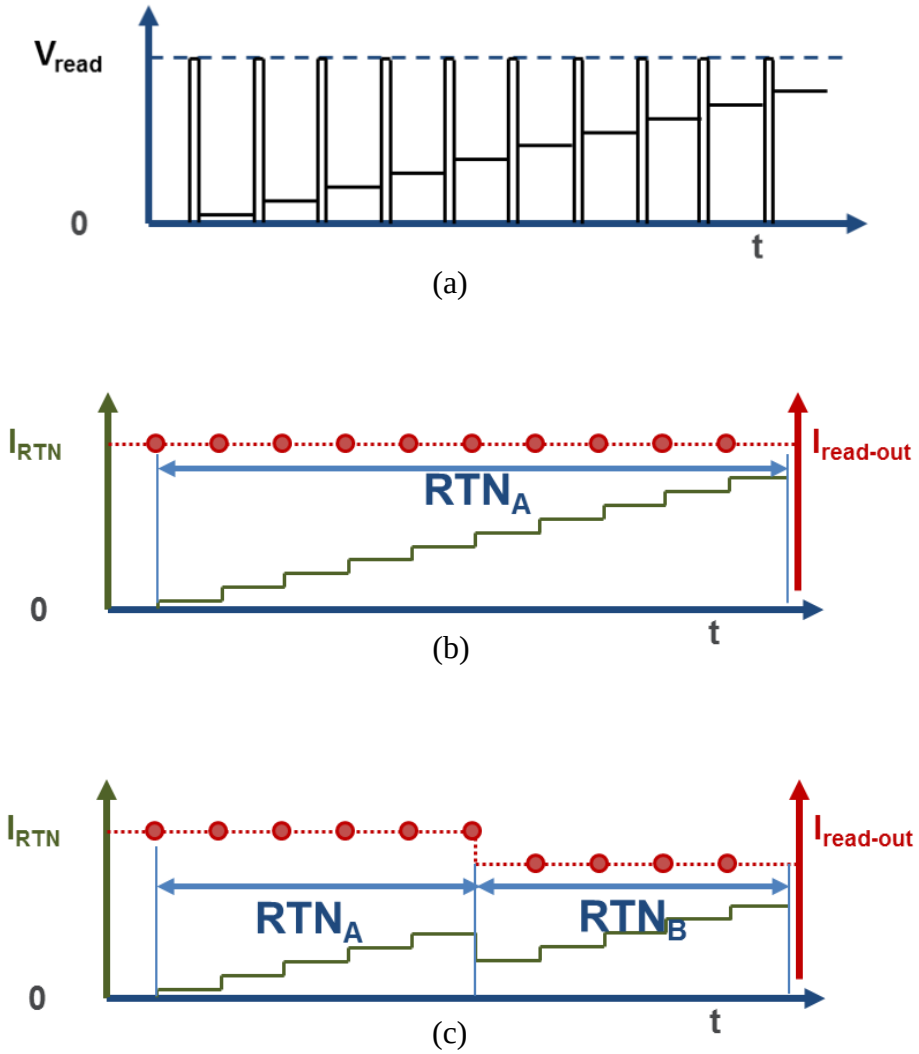


Figure 3-13 Figure Schematic of defect movement tracking method. (a) Voltage pattern: stepping biases for RTN and read-out for device resistance (b) No defect movement: constant device resistance; defects can be extracted with fixed location/energy from RTN (c) Defect movement: device resistance change and defect movement happen simultaneously.

To support that the current jumps are caused by Vo movement induced filament alteration, a second repetitive RTN measurement is carried out following the first one in which a current jump is observed without any switching operation in between, as shown in **Figure 3-14(a)**. It is observed that both defect location/energy and device resistance retain their altered values when being re-measured in the second RTN sweep across the bias range, and the defect at the previous location A can no longer be detected, as it has moved away to location B and B', as shown in **Figure 3-14(b)**. Location B and B' are so close to each other, that they can be regarded as the same location at least in close proximity taking the measurement and extraction error into consideration.

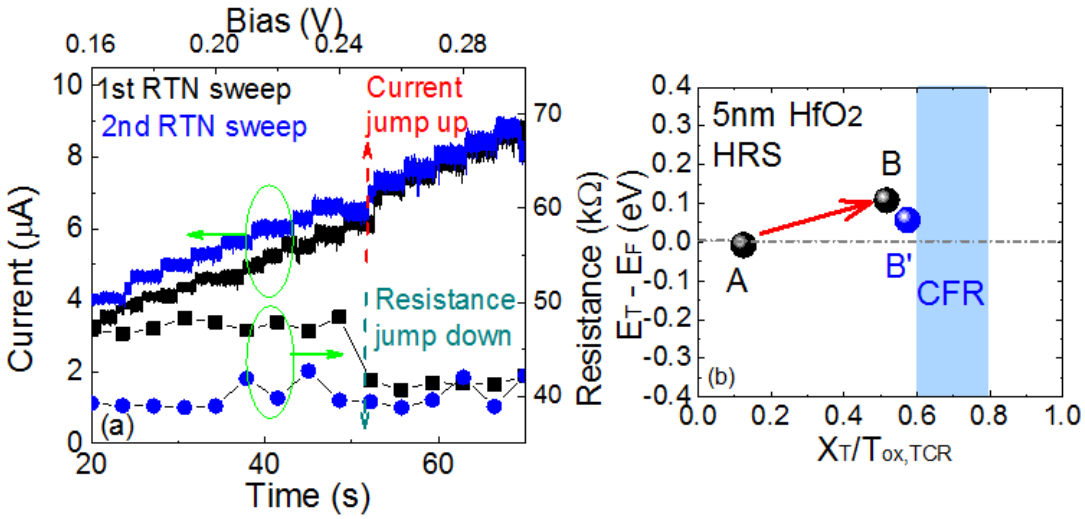


Figure 3-14 (a) Current jump-up during 1st RTN test is associated with resistance jump-down. Both current and resistance retain the altered values during 2nd RTN test, because (b) the defect has moved from A to B during 1st test and stay at B' during 2nd test (≈B). B and B' can be taken as the same location or at least in close proximity.

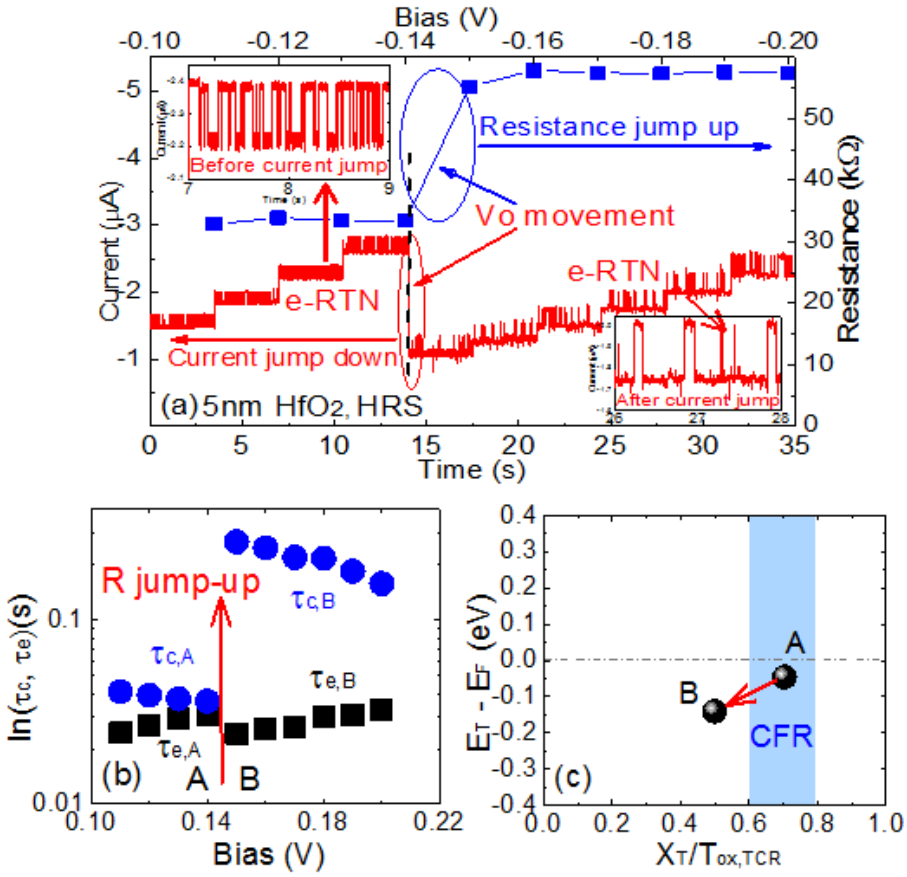


Figure 3-15 RTN based defect tracking (a) RTN signals consist of both Vo movement and e-RTN. Bias is incremented from ±0.1 V to ±0.35 V (-0.1 V~-0.2 V is shown), step =±0.01 V, time=3.5 s/step. R, measured at 0.1V at each step, changes simultaneously with Vo movement. DC V_{set}=1 V, V_{reset}= -1.6 V, I_{cc}=150 μA. (b) τ_c and τ_e measured before/after the R jump, from which (c) defect movement in (E_T , E_T) is extracted.

Figure 3-15 shows an example of the typical defect-tracking test results. During the RTN measurement at HRS, current jumps, either upwards or downwards, can be observed, which is associated with a simultaneous change in device resistance. In this particular example, the current jumps downwards during the bias sweeping, and the resistance measured at the read voltage of 0.1V increases simultaneously, supporting that the current jumps are caused by the modification of CF, as shown in **Figure 3-15(a)**. The RTN and the bias dependence of the RTN time constants extracted before and after the current jumps also changes simultaneously as shown in the inset of **Figure**

3-15(b), resulting in a change of the extracted defect location, as shown in **Figure 3-15(c)**, i.e. defect moves out of the constriction, leading to the CF modification and an increase of resistance.

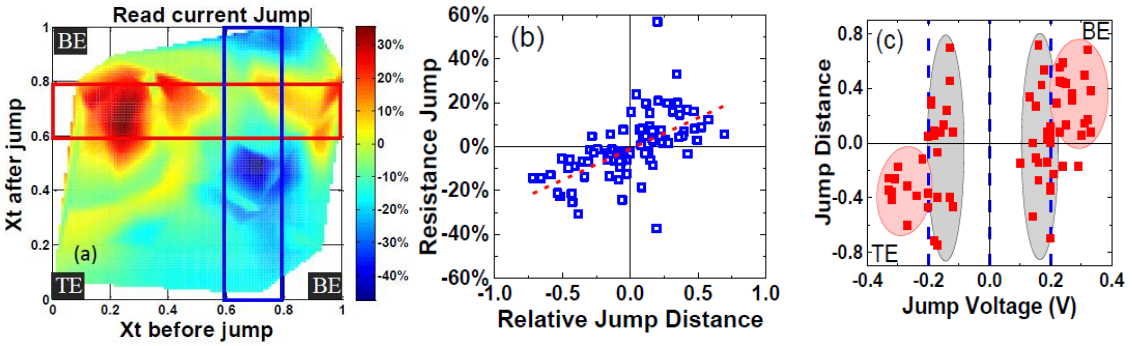


Figure 3-16 Correlation between defect movements and filament modification. a) Statistical analysis of the relative read current change ΔI/I vs X_T before and after Vo movement. (b) Positive correlation between ΔR/R vs relative movement against constriction position X_c: $|X_{t,after}-X_c|-|X_{t,before}-X_c|$. (c) Vo moves towards BE ($\Delta X_T>0$) at V>0.2V, and moves towards TE ($\Delta X_T<0$) at V<-0.2V. Vo moves in either direction at -0.2V<V<0.2V

To further demonstrate the correlation between defect movement and CF modification, the relative change of the read current amplitudes, ΔI/I, are analysed statistically against the defect locations before and after Vo movements. **Figure 3-16(a)** shows that the increases of ΔI/I are correlated with Vo moving into the highlighted CF constriction region (red rectangle), either from the TE direction or the BE direction, and the decreases of ΔI/I are correlated with Vo moving out of the CF constriction region (blue rectangle), either towards the TE direction or the BE direction. The largest amplitudes of ΔI/I are clearly observed when the CF constriction is either the destination or the origin of defect movements. These results highlight the importance of the start/stop location of defect movements. ΔI/I is small when defect movement is not involved with the constriction, contributing less to resistance switching. A broad correlation can also

be obtained between the amplitude of ΔR/R and the relative distance of the defect movement regarding the constriction position, X_c, as shown in **Figure 3-16(b)**. The farther away relatively from the constriction the movement is, the larger change in resistance can be statistically observed. The above correlations provide strong statistical evidence supporting that our technique can indeed track defect movement, because it is impossible for two defects randomly detected in the oxides before and after the jump to have such a correlation with resistance switching.

Further details of the defect movement mechanism can also be revealed by the Vo movement dependence on the external bias and its polarity. As shown in **Figure 3-16(c)**, Vo moves mainly towards BE at V>0.2 V, and mainly towards TE at V<-0.2 V, agreeing with that defects drift according to the direction of electric field at a large bias, underpinning the mechanism of the set/reset operations. At a weaker Eox when -0.2V<V<0.2V, which is often encountered during the read operation and write/read disturbance, Vo can move in either directions, suggesting that defect movement is controlled by internal stochastic thermal/chemical processes instead. The above results provide a clear insight that the defect movement is dominated by oxygen vacancy movements in O-deficient HfO_{2-x} with the scavenging Hf layer, as it requires very little energy in such a structure [205, 206]. In contrast, interstitial O formation is unlikely the responsible mechanism because it requires more than 6 eV to form the oxygen vacancy which cannot be provided by a bias as small as 0.2 V [206].

3.3.3 CFR modification vs. operation conditions and oxide thickness

Further correlations between the modification of the CF constriction at HRS and different operation conditions can be clearly observed by applying the RTN technique,

for the first time. As shown in **Figure 3-17(a)**, increasing V_{reset} from -1.2 V to -1.8 V widens the constriction mainly towards to the inner bulk of the oxide, thus increases the resistance at HRS from 12.5 kΩ to 50 kΩ. Reducing the maximum compliance current during the set operation, I_{cc}, from 150 µA to 80 µA, as shown in **Figure 3-17(b)**, widens the defect energy distribution due to the weaker regulation power caused by the lower set current. In **Figure 3-18(a)**, in order to observe the difference in constriction between HRS and LRS, set operation is terminated at a lower voltage of 0.6 V to result in a partial set where electron tunnelling in the CF still dominates. This is because switching to a full-scale LRS at high V_{set} will lead to a full metallic conduction path where the RTN technique is no longer applicable. More defects in the CFR can indeed be detected at the partial LRS, resulting in a weaker constriction and thus lower resistance. This result provides the experimental evidence of the difference in defect profiles at HRS and LRS, and reveals that the defect number in the constriction indeed determines the device resistance state. The constriction can also be observed in devices with the HfO_2 thickness scaled further down to 3 nm, as shown in **Figure 3-18(b)**, confirming its applicability in devices further scaling into the lower nanometre regime.

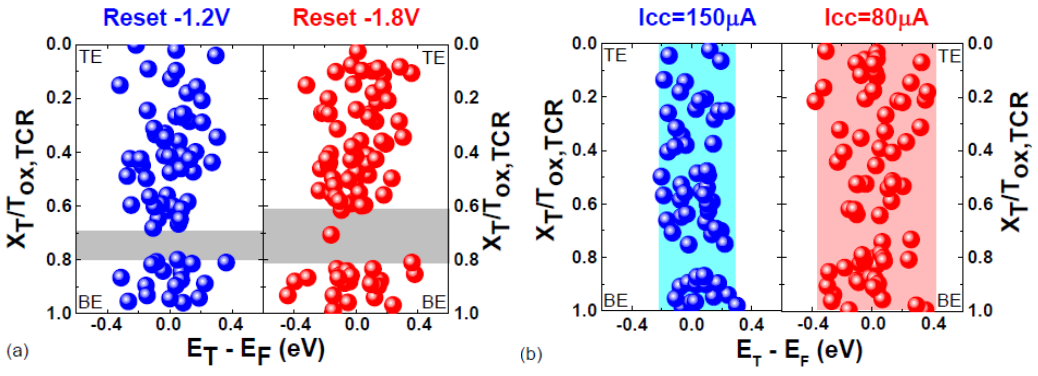


Figure 3-17 (a) Constriction is widened at a higher V_{reset}, and R at HRS increases from 12.5kΩ to 50kΩ . V_{set}=1V. (b) Defects have a wider energy distribution at a lower compliance I_{cc}, probably due to weaker regulation during set.

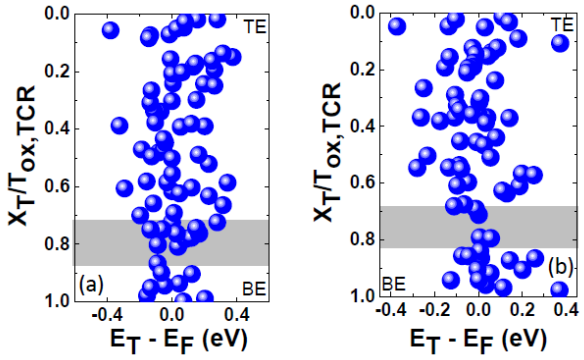


Figure 3-18 (a) At a partial LRS where electron tunnelling dominates, more defects can be detected within the constriction. Set: 0.6 V, Reset: -1.4 V, I_{cc}=100 μA. (b) Constriction is also observed near the BE in devices with 3 nm HfO_2 layer. Set: 0.65 V, reset: -1.4 V, I_{cc}=100 μA.

Since the results demonstrate that the defect movements start to follow the direction of E_{ox} at low electric field and lead to resistance change, it is therefore confirmed that the positively charged oxygen vacancies are acting as the electron traps to produce the conductive path. During set operation from HRS towards LRS, the modification of CF is initiated by the positive external electric field, which pushes Vo towards the constriction mainly from TE direction. Thermal simulation confirmed our experimental results as temperature has a marginal role in the initiation of the HRS to LRS set process [207]. During the reset, defect movements from the constriction towards TE are dominant, resulting in a wider constriction and higher resistance. Our results also confirms that defects movement along the CF direction plays a dominant role in controlling the constriction modification, opposing to the cases of unipolar switching in which the defects move into and out of CF in the direction perpendicular to CF in previous works [208]. Combined with previous results of physical characterisation with c-AFM [136] and modelling [149] for the same devices, unambiguous evidence shows that the CF has an hour-glass shape at HRS with the constriction located closer to the more oxygen-inert BE and the length of the constriction is less than 1 nm.

Further evidence of the correlation between filament modification and the device operation can be obtained from the impact of defect movement on RTN amplitude. **Figure 3-19** compares the measured RTN amplitude distributions at HRS, weak LRS and LRS. The distribution becomes wider at higher resistance, and a significant tail is observed at large RTN amplitude, as compared to LRS, agreeing with previous results [157]. The large RTN amplitude at HRS suggests that the trapping/detrapping into/from the defects within CFR may lead to large noise during device read operations.

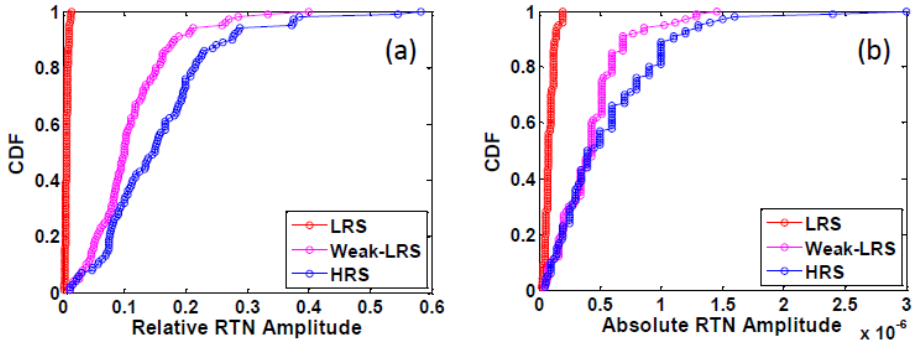


Figure 3-19 Distributions of (a) relative RTN amplitude $(\Delta I/I)_{RTN}$ and (b) absolute RTN amplitude ΔI_{RTN}, at LRS, weak LRS, and HRS. The distribution becomes wider at higher resistance and a significant tail at large RTN amplitude is observed for both relative and absolute RTN amplitude.

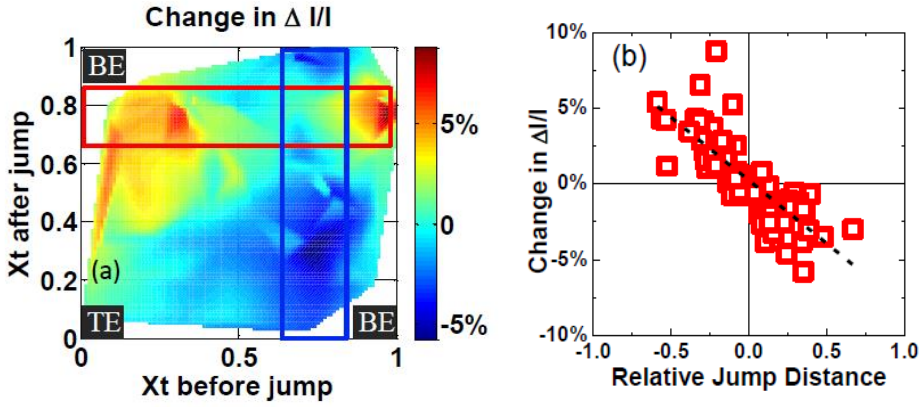


Figure 3-20 Correlation between defect movements and relative RTN amplitude. (a) Statistical analysis of the change in relative RTN amplitudes $(\Delta I/I)_{RTN}$ vs X_T before and after Vo movement. (b) Negative correlation between $(\Delta/I)_{RTN}$ vs relative movement distance against the constriction position X_c: $|X_{t,after}-X_c|-|X_{t,before}-X_c|$.

A statistical analysis of the RTN amplitude before and after the defect movements at HRS confirms this point. **Figure 3-20(a)** shows that the increases of $(\Delta I/I)_{RTN}$ are correlated with Vo moving into the highlighted CF constriction region and the decreases of $(\Delta I/I)_{RTN}$ are correlated with Vo moving out of the CF constriction region. The largest changes in $(\Delta I/I)_{RTN}$, i.e. $(\Delta I/I)_{after\ jump}$-$(\Delta I/I)_{before\ jump}$ are clearly observed when the CF constriction is either the destination or the origin of defect movements. These results are in agreement with the resistance change observed in **Figure 3-16(a)**. The change in $(\Delta I/I)_{RTN}$ is small when defect movement is not involved with the constriction. A broad correlation can also be obtained between the change in $(\Delta I/I)_{RTN}$ and the relative distance of the defect movement regarding the constriction position, Xc, as shown in **Figure 3-20(b)**. The farther away relatively from the constriction the movement is, the larger change in $\Delta I_{RTN}/I_{RTN}$ can be statistically observed, agreeing with the observation in **Figure 3-5(b)**. A single charge trapping/detrapping in the CFR region may significantly alter a tunnelling path and induce a large variation in tunnelling current and thus a large RTN induced read noise. On the contrary, the constriction region is much less significant at LRS than at HRS, the impact of single trapping/detrapping becomes much weaker and thus RTN amplitudes are relatively small and much narrowly distributed.

3.3.4 Endurance failure and recovery

As is introduced in Chapter 1, endurance is one of the essential criteria for the universal memory application. During the cycling, in general, the device could eventually be stuck at either HRS or LRS and cannot be switched anymore under the normal condition. Efforts have been made for the failure mechanism analysis but there's still a lack of microscopic insight of evidence at the defect level. In this sub-section, this

RTN-based method was applied for the endurance failure (stuck-at-HRS) analysis.

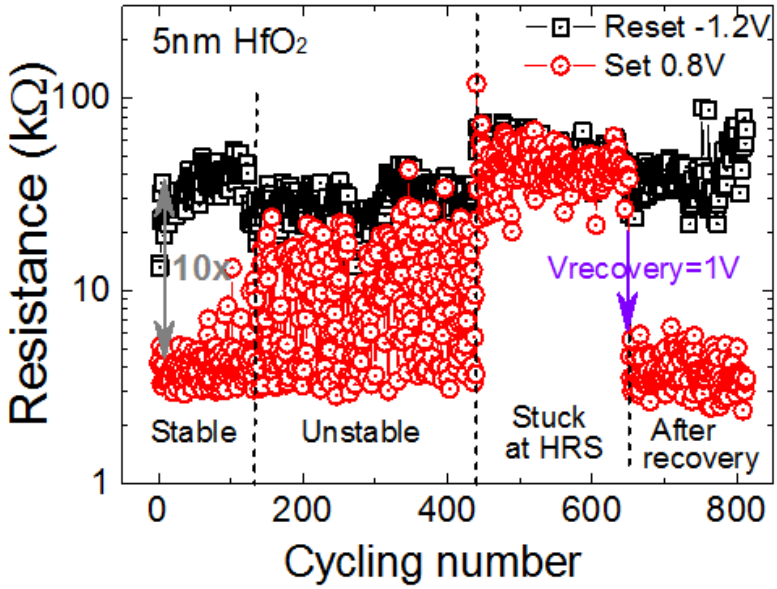


Figure 3-21 Typical DC endurance test results, where cycling underwent four phases: stable, unstable, stuck at HRS failure, and after recovery.

A typical endurance test result is shown in **Figure 3-21**, in which the cycling underwent four phases: stable, unstable, and stuck at HRS failure which can then be recovered by applying a higher $V_{recovery}$ [19]. In the stable phase, the device is switched between HRS and LRS by alternating reset and set normally. In the unstable phase, the device can be reset to the normal HRS but sometimes cannot be set normally to LRS and a lot of abnormal LRS appears, i.e. set failure. In the stuck-at-HRS phase, the device totally lost the ability to be set to normal LRS. It got stuck at HRS entirely and cannot be switched back to LRS anymore, unless a stronger set is applied when the device can be recovered.

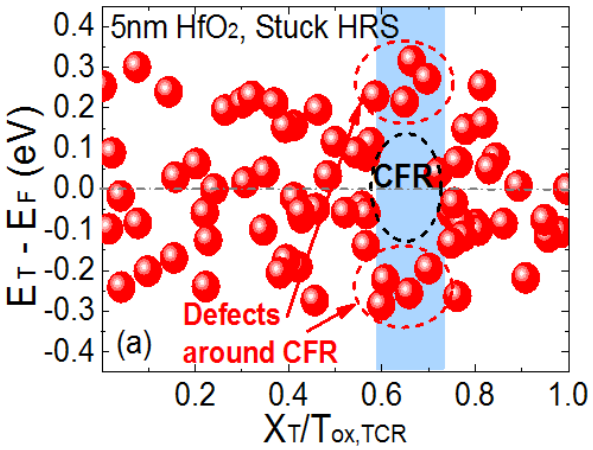


Figure 3-22 Defects detected during cycling after device stuck at HRS. CFR after the failure is surrounded by defects at lower or higher energy levels, causing set failure.

Comparing the defect distribution before and after the failure, it is observed that after the failure, CFR at around the region where $E_T = E_F$ is surrounded by defects at lower or higher energy levels, as is shown in **Figure 3-22** and compared with **Figure 3-12**. Those defects never appear in the normal HRS where there are the least number of defects in the CFR with any energy level, nor in the normal LRS where defects are assumed to be in the CFR but with energy close to the Fermi level, as is indicated by the weak LRS in **Figure 3-18(a)**. Those defects may be responsible for the stuck-at-HRS failure if they accumulate surrounding CFR and the amount reaches a critical level. There are two possible mechanisms that could explain the failure: One is that these surrounding defects will repel other defects from moving into CFR because of their location during the set process, as shown in **Figure 3-23**; The other possibility is that as the V_{set} is around 1 V, those defects may be aligned to the Fermi level during set and form a shunt current path at V_{set}, reducing the bias across CFR so that a normal set could not happen, as illustrated in **Figure 3-24**.

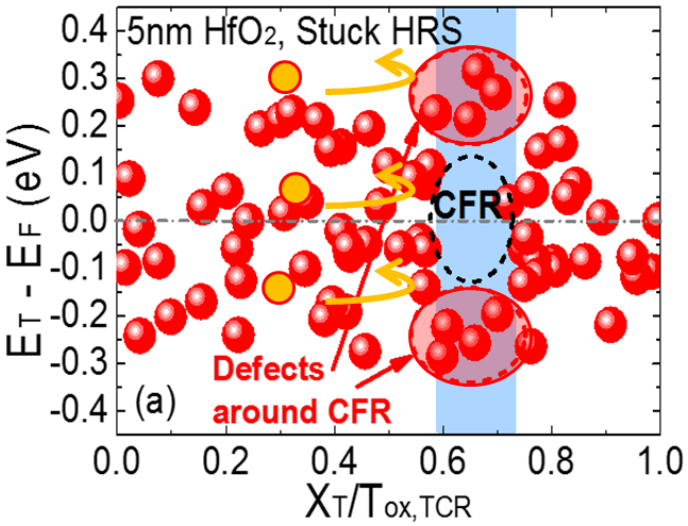


Figure 3-23 Stuck-at-HRS: during set the defects surrounding CFR may repel other defects from moving into CFR and cause a set failure.

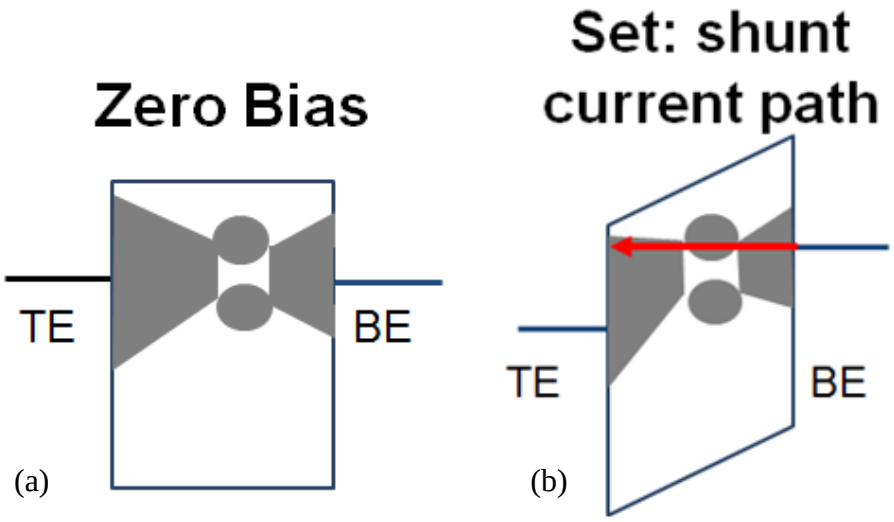


Figure 3-24 (a) Without V_{TE} or under small $V_{TE}=V_{read}$, those defects surrounding CFR does not help conduction. (b)Stuck at HRS: during set those defects surrounding CFR may form a shunt current path, reducing the bias across CFR and assisting the failure.

The defects surrounding CFR hardly contribute to the conduction at V_{read} due to their energy level misalignment with electrodes and/or longer electron tunnelling path, since the V_{read} is only 0.1V, as shown in **Figure 3-25**. This leads to HRS at read-out condition after the failure. At the unstable phase, defects can either move around or into CFR, causing unreliable set.

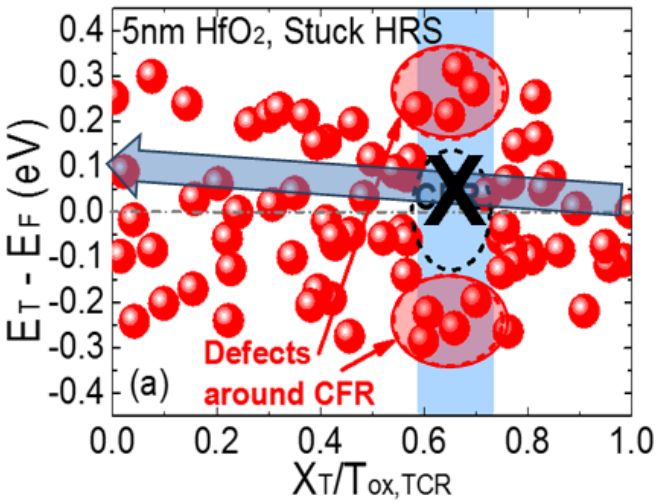


Figure 3-25 At read-out condition: those defects surrounding CFR hardly contribute to the conduction at Vread.

Moreover, these surrounding defects are removed after the recovery, strongly supporting their roles in causing failure, as shown in **Figure 3-22** and **Figure 3-26(b)**.

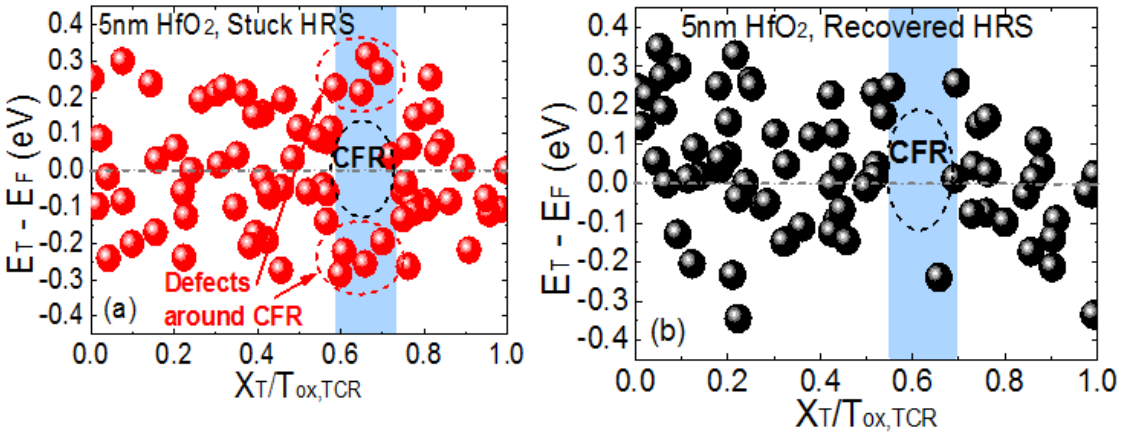


Figure 3-26 Defects detected during cycling. (a) Stuck at HRS failure phase. CFR after the failure is surrounded by defects at lower or higher energy levels, causing set failure. (b) After recovery, the defect profile also recovered, providing strong support for the role of defects surrounding CFR causing device set failure.

Similar failure mode has also been observed under AC operations in which the recovery is done by a strong DC sweep. Defects surrounding CFR cannot be observed during normal/stable cycles, but can be observed in failed device stuck at HRS, (t_{set}=20 ns,

t_{reset}=20 ns, V_{set}=1.4 V, V_{reset}=-2 V) as shown in **Figure 3-28(a)** and **(b)**, respectively. It should be noted that under AC operations the conventional 1R setup cannot provide the compliance current and the 1T1R setup has to be employed as is schematically shown in **Figure 3-27**. In order to exclude the transistor as an RTN source, the BE terminal is grounded during RTN measurement so that RTN current only flows through the RRAM, as shown in **Figure 3-27(c)**.

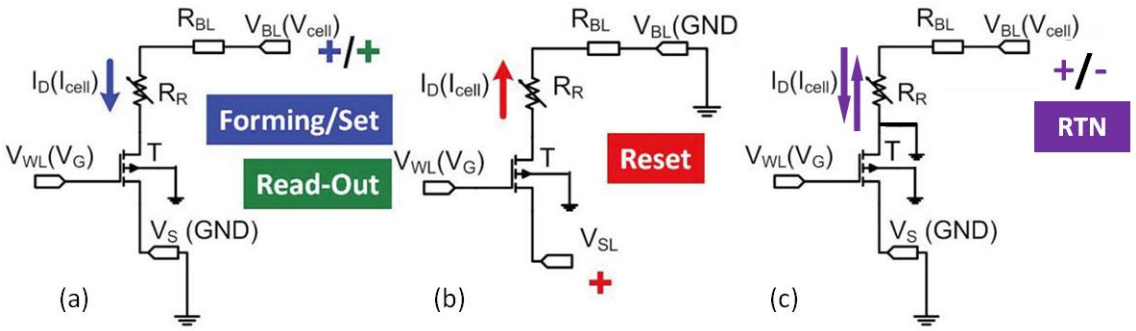


Figure 3-27 AC cycling setup with RTN measurement (a) Bias scheme for forming/set and read-out on the 1T1R device (b) Bias scheme for reset on the 1T1R device (c) Bias scheme for RTN measurement on the 1T1R device, To exclude the transistor as an RTN origin the BE of RRAM is grounded directly.

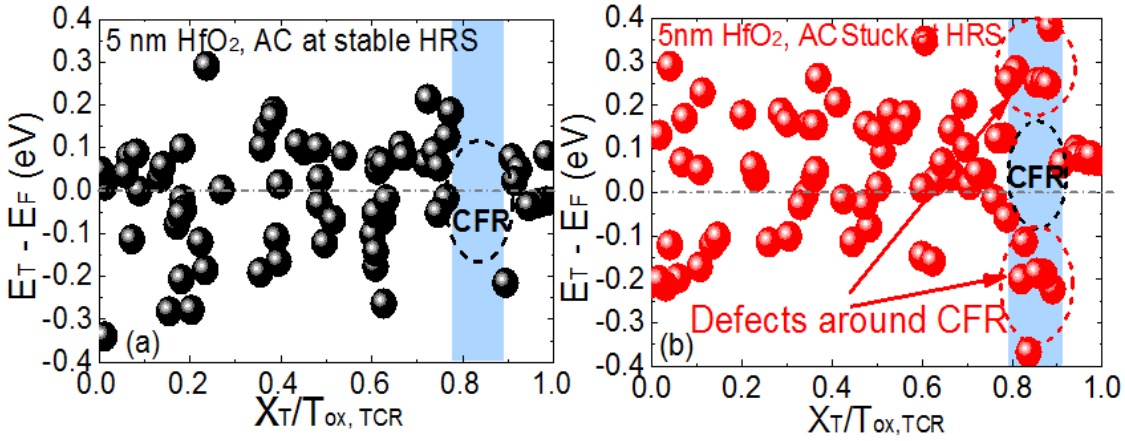


Figure 3-28 Similar failure mode has also been observed under AC operations, where (a) defects surrounding CFR cannot be observed during normal/stable cycles, (b) but can be observed in failed device stuck at HRS. t_{set}/t_{reset}=20 ns, V_{set}=1.4 V, V_{reset}=-2 V.

3.4 Conclusions

An overall picture of the switching mechanism in bipolar HfO_2-based RRAM has been provided in this chapter, showing the repeatable modification of the local constriction. The CF is formed by the generation of defects in the bulk of the oxide during electroforming, resulting in a percolation current path. Under the applied forming electric field, the oxygen ions drift toward the anode leaving behind the Vo forming a local conductive path in the HfO_{2-x}. The CF connects the top oxygen reservoir with the BE at LRS. During the resistive switching cycle the number of Vo in the constriction and the length of the constriction of the hour-glass shaped CF, where it has the least defect number and is closer to the BE, are changed by the defect movement according to the electric field at a bias as low as ±0.2V, inducing the resistance change and eventually leading to the switching of resistance states. The stuck-at-HRS failure is attributed to defects surrounding CFR which can be removed after the recovery. Our RTN based CF characterisation technique provides the physical insights of oxide based filamentary RRAM, improves the understanding of the CF and its constriction at defect level, reveals a new endurance failure mechanism and thus provides a useful tool for RRAM technology development. Next, this technique will be applied onto the novel non-filamentary RRAM devices which feature promising characteristics but the switching mechanism is not clearly understood yet.

4 Characterisation of a-VMCO RRAM

4.1 Introduction

Despite the promising progress that has been achieved in the past decade for the understanding of filamentary resistive switching mechanism and its applications, there are still issues related to the device performance, such as scaling, variability, endurance, retention, etc. [23]. Many of those issues come from the stochastic nature of the CF in the oxide, for example, the LRS depends only slightly on the cell area unless the device size approaches the filament size [23, 24], and HRS resistance suffers from the significant variation that comes from the small number of defects in the ruptured critical filament region [116, 181, 209].

Non-filamentary RRAM devices, which show areal switching behaviour and of which the resistive switching mechanism is generally conceived as the uniform modulation of defect profile at the interface with the electrode and/or between two dielectric layers [24] [67], are promising candidates in overcoming those drawbacks. Hou et al developed a $Ta/TaO_x/TiO_2/Ti$ double-layer 3D vertical RRAM which demonstrates a substantial self-rectifying ratio of 10^3 with sub-μA operating current and extrapolated endurance of over 10^{15} cycles at 100 °C [67]. Since 2013, Govoreanu et al has reported 3 generations of non-filamentary, double-layer RRAM devices, namely the vacancy modulated conductive oxide resistive RAM (VMCO-RRAM), which feature characteristics such as self-rectifying, compliance-free, back end of line (BEOL) CMOS-compatible, and analogue switching. The 1st generation was a $TiN/TiO_2/Al_2O_3/TiN$ structure and showed μA level operation and tight resistance distribution [24]. The 2nd generation, namely

analogue VMCO (a-VMCO), was reported in 2015 with a TiN/TiO_2/a-Si/TiN structure and added analogue switching behaviour, improved nonlinearity and operation uniformity [42]. Based on that, the 3rd generation, namely advanced a-VMCO, was reported in 2016 with reduced switching currents, increased resistance window, excellent wafer-level uniformity and improved reliability, thanks to the inner-interface engineering [43]. Its MLC operation and small variability make it promising to serve as synapses in the neuromorphic computing hardware systems as introduced in Chapter 1 [111].

Non-filamentary RRAM is a relatively novel topic and only a few devices have been reported in recent years [24, 42, 43, 67]. Their resistive switching mechanisms are not clearly understood yet. For example, for the a-VMCO RRAM devices, there is no consensus on the mechanisms for its analogue resistive switching behaviour yet. Several physical characterisation techniques, such as high resolution scanning transmission electron microscopy (HR-STEM), energy dispersive spectroscopy (EDS) and spectroscopic ellipsometry (SE) have been employed to provide detailed information of the stacks [43]. However, as discussed in Chapter 3, they are time-consuming, destructive and statistical-unfriendly. Neither can they provide microscopic description of the defect profile modulation based on direct experimental results. Furthermore, those techniques can hardly help with the device performance improvement, such as the state drift in the endurance test under constant voltage stress (CVS).

In this chapter, by utilizing the recently developed defect profiling technique based on RTN, an in-depth analysis is carried out in a-VMCO RRAM devices. For the first time, defect profile modulation in the TiO_2 switching layer of a-VMCO RRAM is correlated

with the analogue switching. It is also evidenced that the gradual growth of a defect-deactivation-region (DDR) in the TiO_2 layer near its interface with the a-Si barrier layer causes the LRS endurance instability, while t_{reset}/t_{set} ratio is found critical for the HRS instability. Under this guidance, a stable resistance window of 10 for $>10^6$ cycles is restored through combining optimizations of device structure and set/reset conditions, paving the way for a-VMCO's practical applications. The RTN-based defect profiling technique proves to be a useful tool for the non-filamentary RRAM studies.

4.2 Devices and Experiments

4.2.1 Devices

The a-VMCO RRAM devices are crossbar devices integrated in a CMOS-compatible process. The a-VMCO RRAM device has a TiN BE width down to 30 nm and a TiN TE width down to 40 nm, as is shown in **Figure 4-1(a)** with SEM. The reference active stack consists of 8 nm PVD amorphous-Si (a-Si) and 8 nm ALD TiO_2 crystallized in anatase phase, as is shown in **Figure 4-1(b)** with TEM and in **Figure 4-2(a)** with HR-STEM. The layer thickness may vary for optimization and the device structure used in this chapter is summarized in the end of this sub-section. Atomic concentration profiles by EDS line scans across the a-VMCO stack is shown in **Figure 4-2(b)**. Ti profile rises up with a z-offset compared to that of O, suggesting formation of a SiO_x layer, attributed to the darker contrast "line" between a-Si and TiO_2, as observed by HR-STEM in **Figure 4-2(a)**. The passivation process has a thermal budget of 370 °C. Both the LRS and HRS resistances depend on the device area with a 1:1 slope, indicating non-filamentary switching [24], with a switching current density of ~

0.3MA/cm^2, as is shown in **Figure 4-3**.

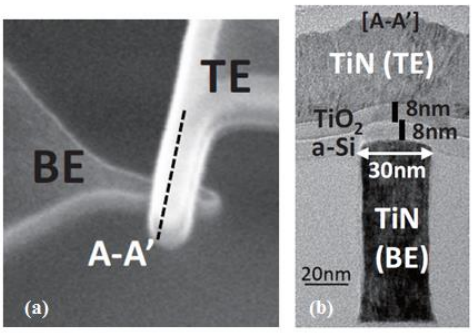


Figure 4-1 Top-view SEM and cross-section TEM along the AA' direction of an a-VMCO device structure of nominal size [42].

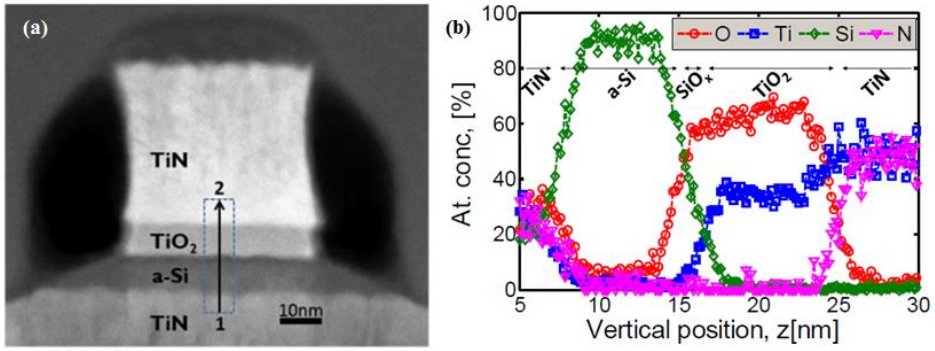


Figure 4-2 (a) HR-STEM cross section image of a reference a-VMCO crossbar cell, of 40 nm nominal size. (b) Atomic concentration profiles by EDS line scans across the a-VMCO stack. Ti profile rises up with a z-offset compared to that of O, suggesting formation of a SiO_x layer, attributed to the darker contrast "line" between a-Si and TiO_2 [43].

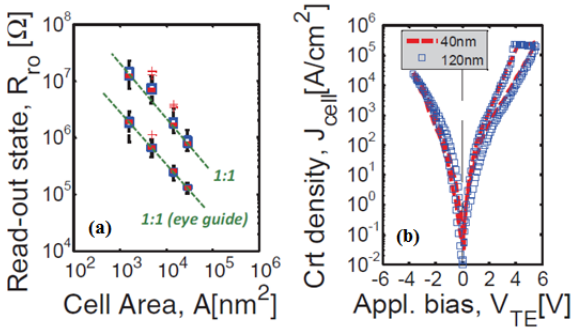


Figure 4-3 (a) LRS and HRS resistances depend on device area with a 1:1 slope, with good separation between the two states. The data are collected over an area range A_{max}/A_{min}>20. (b) Current density–voltage (JV) sweeps, for cells of 2 different areas (A_1, A_2, A_2/A_1=9) show that current scales with cell size at any V_{TE} [42].

The a-VMCO RRAM is a low switching current memory device in which the conductance is electrically modulated by defect distribution in the switching layer [210]. This enables self-rectifying and self-compliant operations, and also greatly simplifies the process integration by eliminating the additional selector element. In the a-VMCO structure, the a-Si layer serves a dual role: it acts as a barrier layer controlling the current through the stack, and as a scavenging layer inducing the as-grown defect profile in the switching layer [210].

The a-VMCO RRAM devices are forming-free and initially at LRS. This is an advantage over filamentary RRAM devices, because forming normally requires a higher voltage than the normal set operation. To reset the devices, a positive V_{TE} is applied, while set requires a negative V_{TE} [42], as shown in **Figure 4-3b** and **Figure 4-4**. Both set and reset switching processes are progressive. A read-out at ~+3V is performed to maximize the resistance window, which is in the order of ×10 or larger.

The extraction of the optical bandgap for a-Si and TiO_2 from measurements of the optical absorption (OA) coefficient by spectroscopic ellipsometry (SE) is shown in **Figure 4-5(a)**. The OA coefficients were calculated using harmonic oscillators models yielding best fit. Idealized band diagrams reconstructed at “flat band” for a dual-layer VMCO stack, using XPS, SE and electrical measurements, emphasizing a built-in potential ($\Delta\Phi_b$=0.95 eV) as is shown in **Figure 4-5(b)** [211, 212]. Ideal dielectrics have been assumed. At thermal equilibrium (no external bias), Fermi levels in the electrodes line up leading to the formation of a built-in field, with consequences for the asymmetry of the set/reset device operation voltages as shown in **Figure 4-5(c)**. This energy band diagram of a-VMCO RRAM will be used in the trap extraction technique based on

RTN.

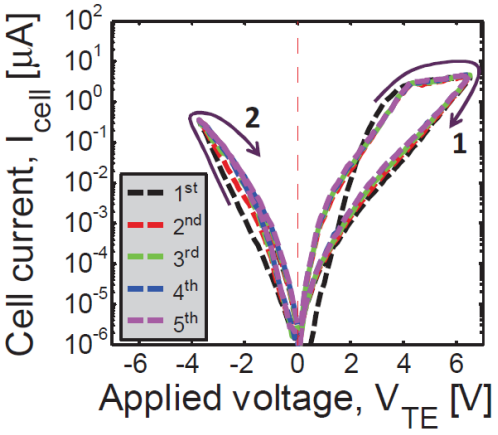


Figure 4-4 First 5 DC reset/set cycles for 40 nm cells showing good cycle to cycle uniformity. The devices are forming free and are initially at LRS. The very first sweep may undergo a low-bias initialization but it affects neither switching I-V curves nor the read-out window. Black curves correspond to the first reset/set cycle. Each sweep is averaged over the whole set of tested devices. No external current compliance is used. DC sweeps conditions: V_{reset}=6.5 V, V_{set}=-3.8 V. [42]

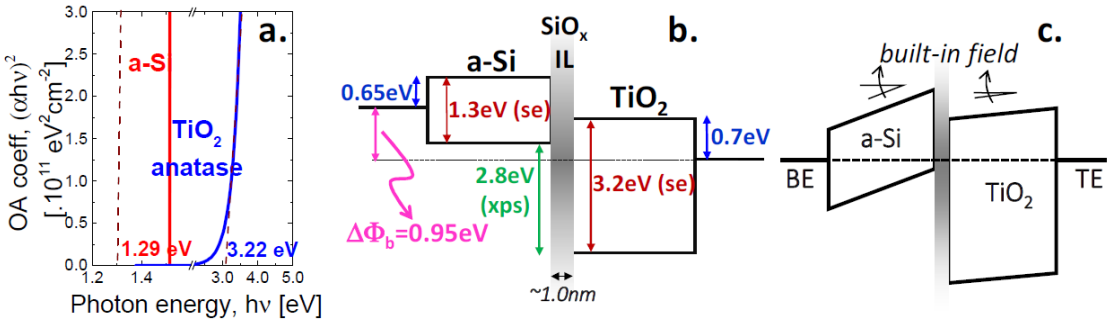


Figure 4-5 (a) Extraction of the optical bandgap for a-Si and TiO_2 from measurements of the optical absorption (OA) coefficient by spectroscopic ellipsometry (SE). The OA coefficients were calculated using harmonic oscillators models yielding best fit. (b) Idealized band diagrams reconstructed at "flat-band" for a dual-layer a-VMCO stack, using XPS, SE and electrical measurements, emphasizing a built-in potential ($\Delta\Phi_b$=0.95 eV). Ideal dielectrics have been assumed. (c) At thermal equilibrium (no external bias), Fermi levels in the electrodes line up leading to the formation of a built-in field, with consequences for the asymmetry of the set/reset device operation voltages [43].

Based on the a-VMCO RRAM, the advanced a-VMCO has been proposed with inner interface engineering for low current, wide resistance window, improved variability and reliability margin, and tuneable resistance [43]. The inner interface engineering is achieved by inserting an additional thin, controlled interfacial layer following the a-Si barrier layer deposition and prior to the TiO_2 deposition. Cross-section through the a-a' plane of a modified-IL a-VMCO device in **Figure 4-6.** (x) shows the crystallographic orientation of anatase-phase TiO_2, with a grain extending across the whole device width (70nm size was used for this TEM).

The DC I-V sweeps of the new, engineered IL a-VMCO device, of nominal size (red) vs. reference a-VMCO (black) is shown in **Figure 4-7(a)**. Compared with the reference a-VMCO, a switching current of ~1μA and a resistance window >100 are achieved, at the expense of only 1V increase of the V_{reset}. The I_{read} scales with the device area for both HRS and LRS, consistent with the non-filamentary switching of the modified-IL a-VMCO device, as is shown in **Figure 4-7(b)**.

In this chapter, both a-VMCO and advanced a-VMCO RRAM devices are analysed and the impact of IL on the endurance performance is discussed. The details of the device that are used in this chapter are summarised in **Table 4-1**. W1 is a-VMCO with 8 nm TiO_2/8 nm a-Si, 40×40 nm^2 structure, W2 is advanced a-VMCO with 8 nm TiO_2/1 nm ILA/8 nm a-Si 40×40 nm^2 structure and ILA is the inserted inner interfacial layer. W3 is also advanced a-VMCO with 8 nm TiO_2/1 nm ILA/5 nm a-Si, 40×40 nm^2 structure; note that the a-Si layer thickness is reduced to 5 nm.

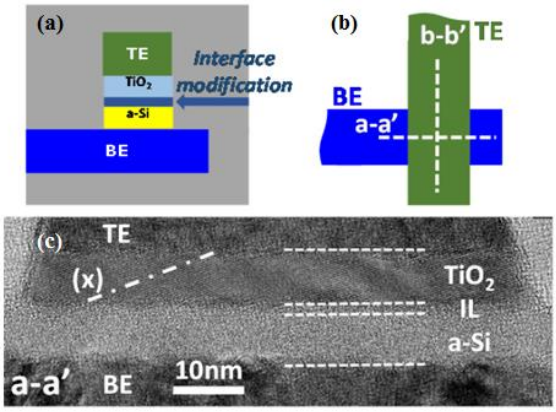


Figure 4-6 Cross-section through the a-a' plane of a modified-IL a-VMCO device. (x) shows the crystallographic orientation of anatase-phase TiO_2, with a grain extending across the whole device width [43].

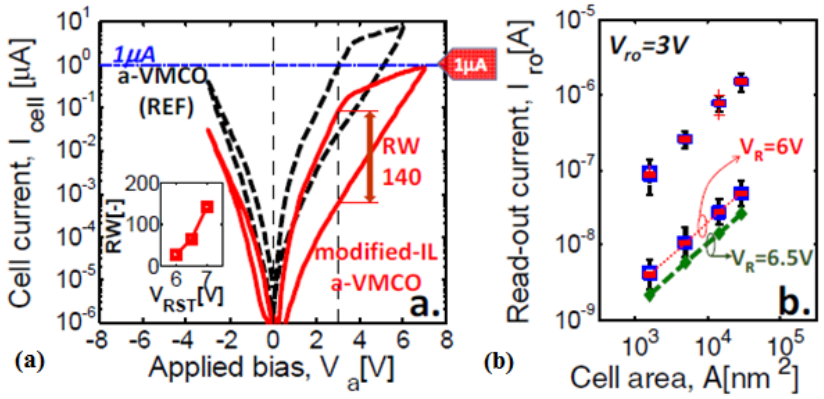


Figure 4-7 (a) DC I-V sweeps of the new, engineered IL a-VMCO device, of nominal size (red) vs. reference a-VMCO (black). A switching current of ~1 μA and a resistance window >100 are achieved, at the expense of only 1 V increase of the V_{reset}. (b) The I_{read} scale with the device area for both HRS and LRS, consistent with the non-filamentary switching of the modified-IL a-VMCO device [43].

Table 4-1 a-VMCO devices used in this chapter: W1 is the 2L structure. W2 and W3 are the 3L structure.

Sample	***IL***	***Process***
W1	--	8nm TiO_2/8nm a-Si 40x40nm^2
W2	ILA	8nmTiO_2/1nm ILA/8nm a-Si 40x40nm^2
W3	ILA	8nmTiO_2/1nm ILA/5nm a-Si 40x40nm^2

4.2.2 Experiments

As is mentioned in Chapter 2, as the defects in RRAM devices are mostly positive charged Vos. The mean time constants for high (τ_{high}) and low (τ_{low}) current levels in RTN are the capture (τ_c) and emission times (τ_e), respectively [182, 183], as is shown in the time trace of RTN in **Figure 4-8**. Based on that, RTN is employed to measure the defect profile from its mean capture/emission time constants dependence on the applied TE bias, in both HfO_2 RRAM, and a-VMCO RRAM, based on **Equation 4-1** and **4-2**, adapted from [158, 186]. The details of **Equation 4-1** and **4-2** are explained in Chapter 2 (Page 75, 76). Note that ε_{oxS} and ε_{oxB} are the relative dielectric constants of switching and barrier layer, respectively, and T_{oxS} and T_{oxB} are the thicknesses of switching and barrier layer, respectively.

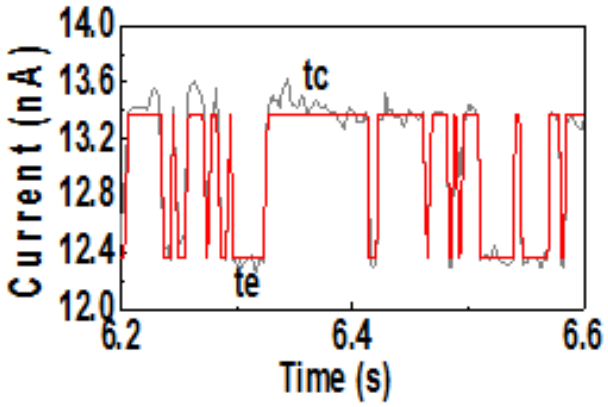


Figure 4-8 Example of RTN in a-VMCO. The mean time constants for high (τ_{high}) and low (τ_{low}) current levels in RTN are the capture (τ_c) and emission times (τ_e), respectively.

$$X_T = \left[1 - \frac{k_B T}{q}\frac{dln(\tau_c/\tau_e)}{dV}(1 + \frac{\varepsilon_{oxS}}{\varepsilon_{oxB}}\frac{T_{oxB}}{T_{oxS}})\right]T_{oxS} + T_{oxB} \quad (4\text{-}1)$$

$$X_T = -\frac{k_B T}{q}\frac{d\ln(\tau_c/\tau_e)}{dV}(T_{oxB} + \frac{\varepsilon_{oxB}}{\varepsilon_{oxS}}T_{oxS})T_{oxS} \quad (4\text{-}2)$$

The RTN-based defect profiling test is carried out as follows: the a-VMCO RRAM device is switched alternatively between LRS and HRS for a number of cycles by either DC sweeps or AC pulses. After each switch operation, either set or reset, an RTN measurement follows. RTN is measured with stepping bias ranging from 2.5V to 3V with V_{step}=0.05V. The applicable bias range is limited by the measurement resolution of instrument: As is shown in **Figure 4-7(a)**, a-VMCO features good non-linearity. If the RTN measurement bias is lower than 2.5V, the current will be lower than ~nA. Although that's still within the analyser's current measurement range as introduced in Chapter 2, the sampling time has to be increased to ensure an acceptable resolution which is essential for noise recognition. Meanwhile, measurement time efficiency is also a major concern if statistical result is required. Therefore a trade-off has to be made for the lower bias limit. The upper limit of bias is set by the V_{read}. The bias range of RTN measurement should be lower than the read-out voltage to avoid the risk of impact on the device resistance, since RTN measurement itself is a CVS process. The sampling time is 2 ms per point. RTN measurement is carried out at both LRS and HRS. In contrast to the filamentary RRAM devices of which LRS is dominant by metallic conduction where RTN can hardly be observed, in a-VMCO RRAM devices the conduction mechanism is still considered as electron tunnelling and indeed RTN can be observed at LRS. The current conduction mode of LRS and HRS should be the same despite the difference in defect profile.

It has been reported that defects in the oxide of MOSFET devices can only be scanned

through a "window" areas in the X_T vs. E_T plane by RTN, because the equipotential line sweeps only within these windows by varying the applied gate voltage [213]. Based on the defect extraction **Equation 4-1** and **4-2**, the shadowed region in **Figure 4-9** is the scan-able region in a-VMCO, where the RTN resembles a "search light" in the X_T vs. E_T plane. Its "rotation angle" is determined by the bias and the "beam width" is determined by the maximum measurable(τ_c/τ_e). Under positive RTN measurement bias, the region covers the entire thickness of two layers while under negative bias range only part of the thickness is covered as shown in **Figure 4-9(b)**. Positive RTN bias is preferable to profile the defects location distribution across the whole thickness of both layers.

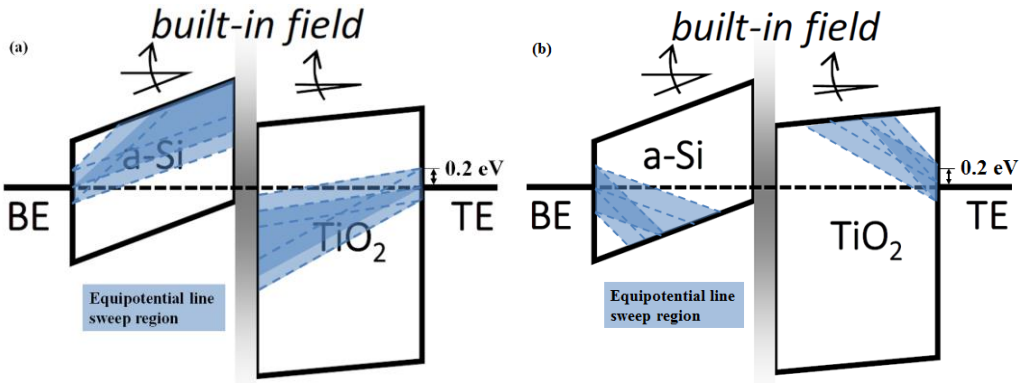


Figure 4-9 Schematic of the defect scannable region in a-VMCO under (a) positive and (b) biases with a margin of 0.2 eV. Defect distribution throughout the oxide thickness under positive bias.

As introduced in Chapter 2, RTN shows only a small fraction of existing defects, limited by the time resolution and the maximum measurement time. If one of the time constants is too large, given the fixed number of sampling points, the total fluctuation number will be too few to be statistically reliable. The difference between τ_c and τ_e should be within 3 orders, i.e.$\left|log\left(\frac{\tau_c}{\tau_e}\right)\right| < 3$, therefore, by which defects within a margin of 0.2 eV from

the equipotential line can be scanned.

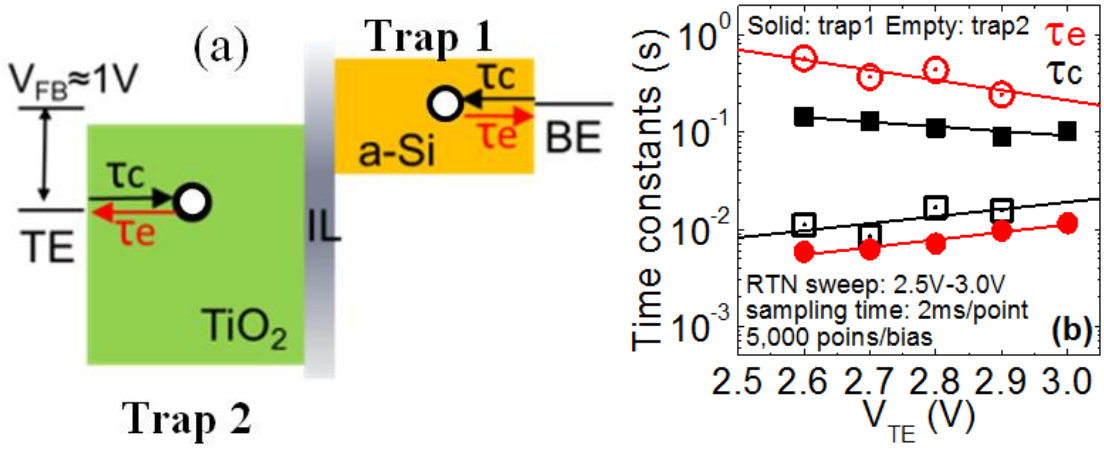


Figure 4-10 Dependence of time constants on TE bias. As V_{TE} increases, for trap 1, τ_C decrease while τ_C increase, corresponding to a trap interacting with the bottom electrode, and for trap 2 the other way round.

The double-layer structure of a-VMCO RRAM devices should be taken into consideration when developing the defect extraction model based on RTN, which is similar to those in MOSFETs with a high-k dielectric layer and a SiO_2 interfacial layer [185] [186]. Given a set of RTN measured under different TE bias as shown in **Figure 4-10**, the first step is to determine whether this trap interacts with the TE or BE. This information can be obtained from the polarity of $d\left(ln\left(\frac{\tau_c}{\tau_e}\right)\right)/dV$, similar to the case of single layer RRAM [185]. If the polarity is negative, the defect interacts with the BE, and vice versa. The next step is to determine which layer the defect locates in as a-VMCO RRAM has two layers. The value of $d\left(ln\left(\frac{\tau_c}{\tau_e}\right)\right)/dV$, i.e. time constant variation rate, determines whether a trap is located in the switching layer or the barrier layer. For the a-VMCO RRAM, if a trap locates exactly at the interface between the switching layer and the barrier layer, its time constant variation rate is calculated to be

12.82, using the parameters as shown in **Figure 4-11** (same as **Figure 4-5(b)**). Therefore, a trap is found to be located in the barrier layer if the time constant variation rate is over 12.82, and is located in the switching layer if it is below 12.82. In theory it is possible for a trap in the switching layer to interact with BE, or a trap in the barrier layer to interact with TE. However, since the oxide thickness of a single layer in all devices is at least 5 nm thick, the probability of electron tunnelling from one electrode through the total thickness of one layer and finally to a trap in the other layer is very small in reality. Therefore, it is observed that traps in the switching layer interact mostly with TE and traps in the barrier layer interact with BE.

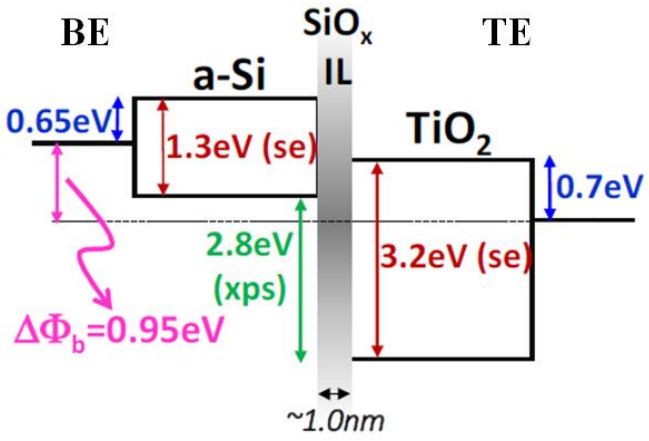


Figure 4-11 Energy band diagram of a-VMCO RRAM (same as **Figure 4-5(b)**). Idealized band diagrams reconstructed at "flat-band" for a dual-layer a-VMCO stack, using XPS, SE and electrical measurements, emphasizing a built-in potential ($\Delta\Phi_b$=0.95 eV). Ideal dielectrics have been assumed.

4.3 Resistive switching mechanism

Using the above defect extraction model based on RTN in a-VMCO, the defect profile modulation in a-VMCO will be correlated with the resistive switching.

4.3.1 Two-level operation

Defect profiles are extracted at both HRS and LRS from the RTN in a-VMCO, as shown in **Figure 4-12**. As can be seen in **Figure 4-12(a)**, the distribution of extracted defects is limited within the scannable region. At HRS, there is a defect-less region in the TiO_2 side of TiO_2/a-Si interfacial layer, which does not exist at LRS, suggesting that defect profile modulation occurs predominantly at TiO_2 side of IL. The correlation between defect distribution and device is better illustrated by using a X_T vs. I_{RO} plane in which the I_{RO} denotes the I_{read} under V_{TE}=3.0 V, as shown in **Figure 4-12(b)**. The defectless region becomes wider at HRS and narrower at LRS, confirming that defect profile in TiO_2 near the IL is responsible for the resistive switching. This result provides direct experimental evidence for the resistive switching mechanism. Note that this defect profile modulation is caused by the movement of pre-existing defects, which have uniform spatial distribution in the lateral direction. This is supported by the forming-free and area-dependent non-filamentary switching characteristics [42, 43, 210].

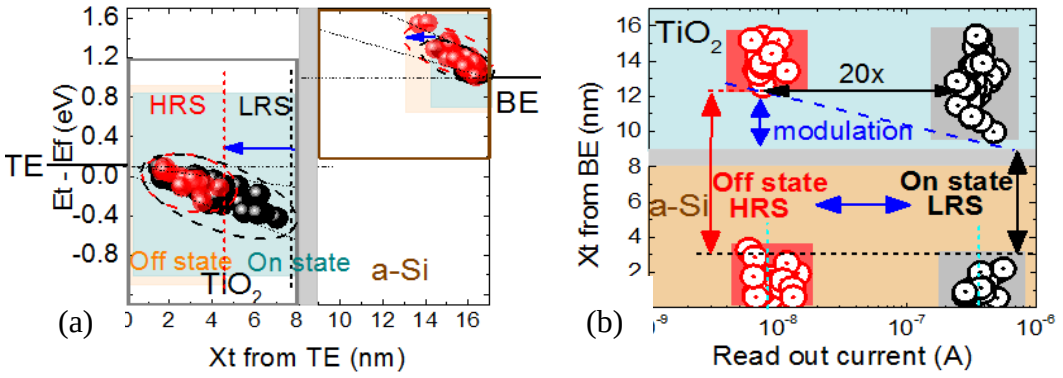


Figure 4-12 5 Extracted defects profile at LRS (‘black ●’) and HRS (‘red ●’) (a) vs. the distance from TE and (b) vs I_{read} at V_{TE}=3V. Defect profile modulation predominantly occurs in a defect-‘less’ region in TiO_2 near IL. Note that this defect profile does not provide information on the actual defect density.

To investigate the impact of defect profile modulation on device operation, the statistical distribution of the relative amplitude of RTN signals in unstressed a-VMCO cells is further examined, because large RTN amplitude could lead to large noise during the read operation, and reduce the resistance window. As shown in **Figure 4-13(a)**, RTN amplitude at LRS is relatively small, with a median value of ~4%, and a relatively tight distribution with a maximum value of ~8%. At HRS, the median increases to ~8% and the distribution is shifted almost in parallel except the lower 10% percentiles. In comparison, RTN amplitude distribution in filamentary RRAM, such as in HfO_2 based RRAM devices, is significantly different, especially at HRS. Its RTN at LRS is much smaller and tightly distributed, while at HRS the distribution is much wider with a long tail at larger relative RTN amplitudes reaching 50%, as shown in **Figure 4-13(b)**.

The differences in RTN amplitude distribution support that the conduction and switching mechanisms in filamentary and non-filamentary are different. In filamentary RRAM, conduction at LRS is metallic-like, with electrons hopping through a filament formed by oxygen vacancies, leading to very small RTN signals [157]. At HRS, its resistance is controlled by the critical constriction region where the least number of defects exist. Trapping and de-trapping of individual defect within and near the constriction has significant impact on the current conduction, hence the much larger RTN amplitude and its much wider distribution [181]. In contrast, the similarity between RTN distributions at LRS and HRS in non-filamentary RRAM suggests a similar conduction mechanism. The slightly larger RTN amplitude at HRS can be attributed to the relatively larger impact by trapping/detrapping since the defect profile is narrower in TiO_2 at HRS. The RTN amplitude at HRS is much smaller and tightly distributed than in filamentary RRAM, supporting that the impact of individual defect is

significantly reduced due to the uniform defect distribution in a-VMCO device.

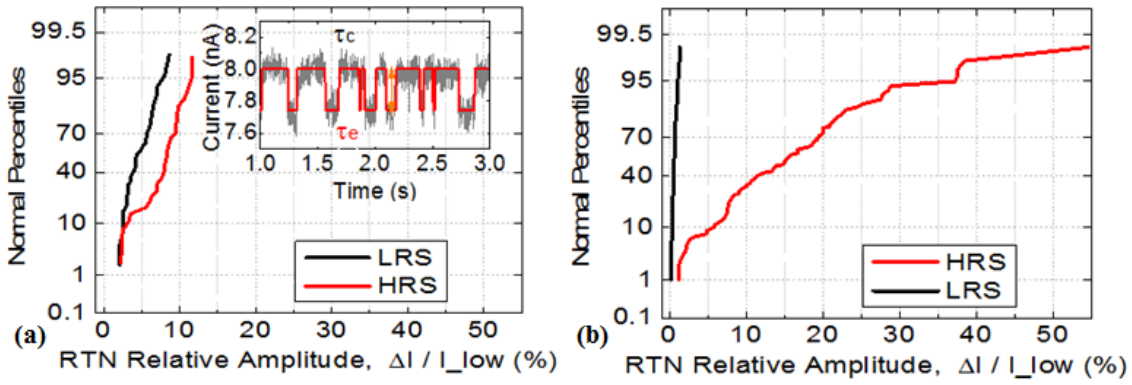


Figure 4-13 Statistical distribution of the relative RTN amplitude at LRS and HRS, in (a) a-VMCO RRAM device and (b) HfO_2-based filamentary device.

4.3.2 Analogue switching

Analogue switching means the resistance level can be arbitrarily controlled by the operation conditions. In a-VMCO an arbitrary chosen HRS can be controlled by the DC sweep stop voltage or AC pulse width/amplitude, and so it is for set. As shown in **Figure 4-14**, consecutive I-V double-sweep traces show reset to an arbitrary chosen HRS, controlled by the sweep stop voltage [42].

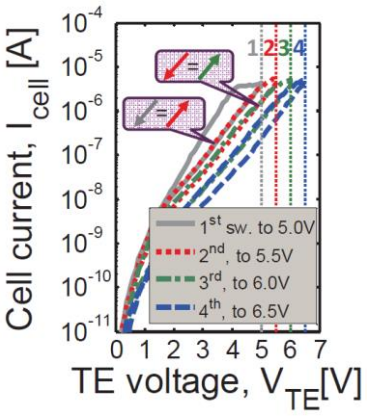


Figure 4-14 Consecutive reset double-sweep traces to an arbitrary chosen HRS, controlled by V_{reset}

To investigate the correlation between defect profile and analog/multilevel switching, I-V and defect profiles at various V_{reset} are compared in **Figure 4-15**. Here the a-VMCO device can be reset to a target analogue resistance level, controlled by the DC sweep stop V_{reset}. The stronger the V_{reset}, the wider the defect-'less' region, corresponding to the higher the R_{HRS}, as the defect profile in TiO_2 is modulated farther away from IL. This result reveals that the analogue resistive switching is caused by the modulation of defect profile in the TiO_2, the extent of which can be arbitrarily controlled by the operation condition within the resistive switching range.

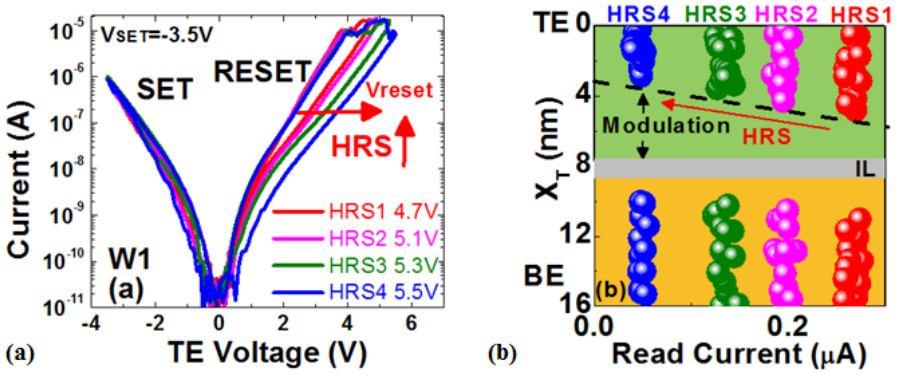


Figure 4-15 a-VMCO can be reset to an analogue resistance value by the DC sweep stop V_{reset}. The stronger the reset, the higher the R_{HRS}, and wider the defect-'less' region, as the defect profile in TiO_2 is modulated farther away from IL.

The analogue resistance level can also be controlled by changing set/reset conditions and the same analogue resistance level can be achieved by either reset or set that lead to the same defect profile as shown in **Figure 4-16**. LRS1 and HRS2 are similar in resistance but LRS1 is achieved by set from HRS1, and HRS2 is achieved by reset from LRS2. It is observed that although LRS1 and HRS2 are achieved from different directions, their defect profiles are quite similar. This strong correlation between defect profiles and analogue resistance levels shows a-VMCO's potential for analogue

applications, and enables an in-depth investigation of a-VMCO's endurance stability, as follows.

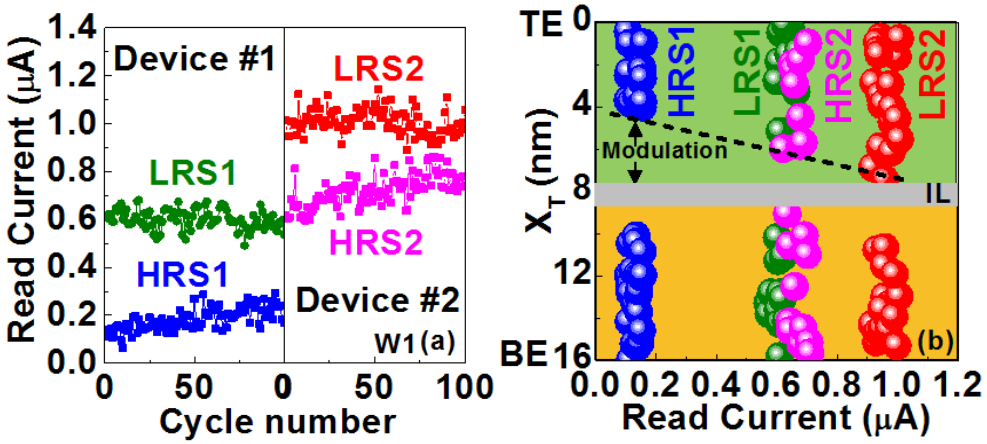


Figure 4-16 (a) The similar analogue resistance level, LRS1 and HRS2, can be achieved by either reset or set under different conditions, (b) which leads to the similar defect profile.

4.4 Endurance performance

- DC endurance

As introduced in Chapter 1, endurance is one of the essential criteria for the universal memory application. During the cycling, in general, the filamentary RRAM device could eventually get stuck at either HRS or LRS and cannot be switched anymore under the normal condition. For non-filamentary RRAM devices such as the a-VMCO, the endurance issue is more often a degradation process that happens gradually during cycling, in which R_{HRS} and R_{LRS} change slightly cycle-by-cycle but this degradation become significant by accumulation. Endurance degradation is found in both DC and AC cycling tests.

In the DC endurance test, the a-VMCO RRAM is switched for 300 cycles as is shown in **Figure 4-17(a)**. Read-out is carried out after each reset/set. V_{reset}=5.5 V and V_{set}=-3.5 V. V_{read}=3.0 V. It is observed that R_{HRS} and R_{LRS} gradually decrease and increase, respectively, and the resistance window decreases. In order to depict the defect profile during cycling, RTN measurement is carried out after each reset/set, at both HRS and LRS. The blue and red spheres in **Figure 4-17(b)** correspond to HRS and LRS in the endurance test. It is observed that similar to the previous results, the defect-'less' region is much wider at HRS than at LRS. During cycling, as R_{HRS} keeps decreasing and R_{LRS} keeps increasing, the width of their defect-'less' region also changes correspondingly.

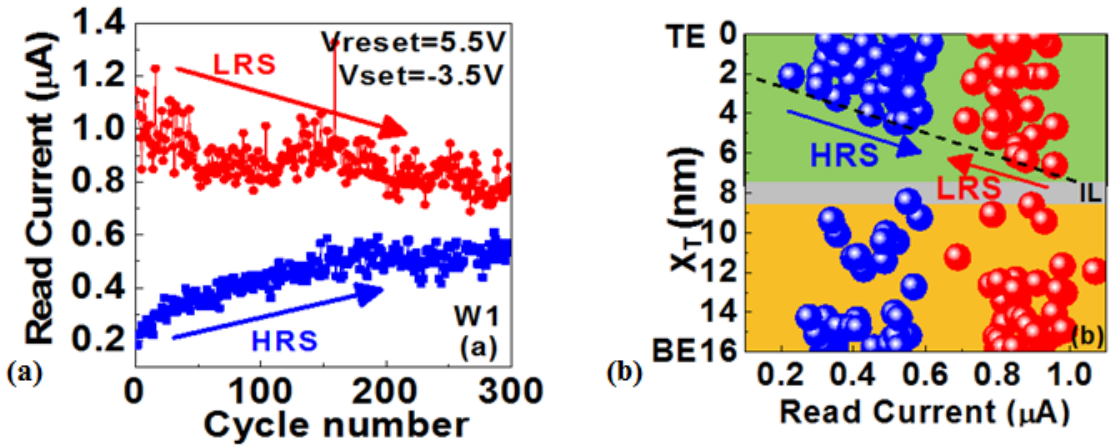


Figure 4-17 (a) A typical DC endurance cycling test. HRS & LRS merge gradually towards the middle (b) A strong correlation between device resistance and defect profiles in the TiO_2 layer during cycling.

- AC condition optimization

As mentioned in Section 4.1, the resistance of a-VMCO RRAM can be arbitrarily controlled with switching operations. Since in practical applications the switching is performed with AC pulses, it is important to optimize the operation conditions on both pulse amplitude and width for the largest resistance window.

To find the optimal AC set/reset conditions, pulses with different amplitudes and widths have been applied on multiple fresh a-VMCO RRAM devices. The pulse time, t, is incremented in a fresh device at each bias condition. It is observed in **Figure 4-18** that, in a fresh device, increasing V_{reset} and V_{set} accelerates switching, i.e. giving a higher R_{HRS} or lower R_{LRS}, respectively, which agrees with the DC result. It should be noted that increasing pulse amplitude soon causes breakdown, which sets the upper limit of pulse amplitude. The optimized V_{reset} and V_{set} are 5.6V and -4.4V, respectively. Set requires a much shorter time than reset. Increasing t_{reset} and t_{set} will give a larger resistance window, but this effect will be saturated after the pulse widths reach a certain level. t_{set} is 10^3 times shorter than t_{reset} for maximizing the resistance window. A t_{reset}/t_{set} ratio of 10^4 leads to only slightly larger resistance window. In summary, the optimized condition is: V_{reset}=5.6 V, t_{reset}=10 ms or 1 ms (corresponding to the pulse width ratio of 10^4 or 10^3); V_{set}=-4.4 V, t_{set}=1 µs. Under this condition, the maximum resistance window in the early cycles is achieved.

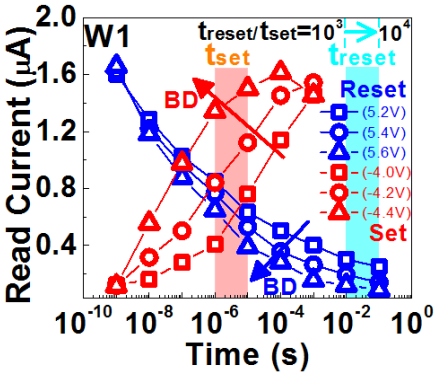


Figure 4-18 Increasing pulse amplitude or width will result in larger RW. Too large pulse amplitude soon causes breakdown. The effect of pulse width will saturate. t_{set} is much shorter than t_{reset}. At V_{reset}/V_{set}=5.6V/-4.4V & t_{set}=10 µs, t_{reset}/t_{set} of 10^3 leads to slightly smaller resistance window than 10^4. t is incremented in a fresh device at each bias condition.

- AC endurance test

In the AC endurance test, reset and set operations are done alternatively. For reset, a positive pulse is applied on the TE and the BE is grounded. For set, a positive pulse is applied on the BE and the TE is grounded. Different from the DC endurance test, read-out during AC cycling is done only 3 times per decade, i.e. after the 1^{st}, 2^{nd}, 5^{th}, 10^{th}, 20^{th}, 50^{th},…, $(10^6)^{th}$ cycle.

The optimized condition is used as the guidance for the cycling test. The endurance results of two different pulse width ratio are compared, i.e. $t_{reset}/t_{set}=10^3$ or 10^4. It is observed in **Figure 4-19** that in the very early cycles the R_{HRS} and R_{LRS} are similar respectively under the two set of conditions. But after about 10 cycles the trends starts to be different, especially for HRS.

When $t_{reset}/t_{set}=10^3$, R_{HRS} and R_{LRS} gradually decrease and increase during cycling respectively, and therefore this degradation mode is named as "window closure". In this mode, R_{HRS} and R_{LRS} finally merge in the middle after about 10^3 cycles, similar to the DC endurance test. The defect profiles during cycling merge in the middle, causing early endurance failure as is shown in **Figure 4-20(a)**. Increasing t_{reset}/t_{set} to 10^4 has a significant impact on the endurance performance, as the ratio of R_{HRS} to R_{LRS} keeps almost constant. Unfortunately, both states drift towards the higher resistance during cycling, and thus this mode is named as "state drift". After a certain number of cycles, the R_{LRS} will become even higher than the initial R_{HRS}, causing problems for practical application. It is observed from the defect profile that by applying a relatively longer t_{reset}, the increase in R_{HRS} is raised by widening the defect-"less" region, as shown in **Figure 4-20(b)**. 10^6 unverified cycles are achieved without window closure. It should be noted that the change in LRS is hardly affected by different t_{reset}/t_{set} ratios.

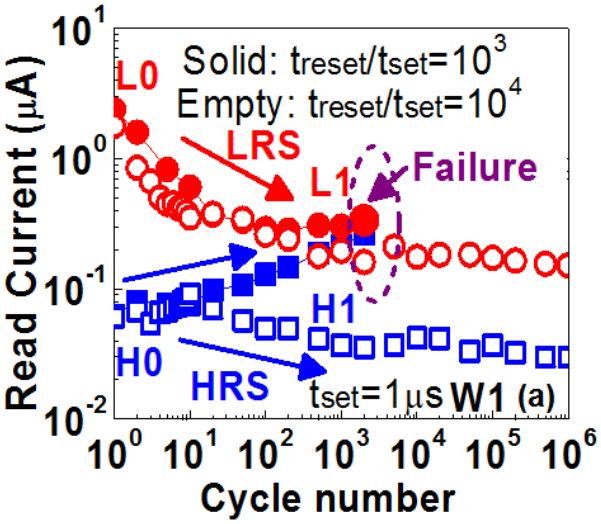


Figure 4-19 AC endurance instability at $t_{reset}/t_{set}=10^3$ (LRS and HRS merge in the middle) and 10^4 (both LRS and HRS shift downwards; the window remains open for 10^6 unverified cycles).

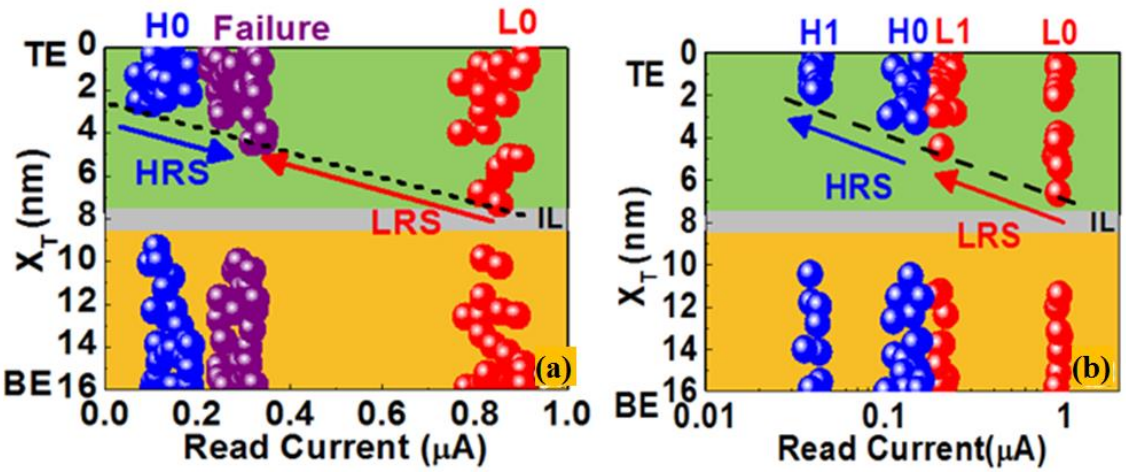


Figure 4-20 (a) resistance and defect profiles merge in the middle when $t_{reset}/t_{set}=10^3$. (b) Increasing t_{reset}/t_{set} to 10^4 leads to higher HRS by modulating the defect profile farther away from IL and avoids window closure. Change in LRS is hardly affected by different t_{reset}/t_{set}. V_{reset}=5.6V, and V_{set}=-4.4V.

For LRS, it appears that the defects in a region in TiO_2 near IL are gradually deactivated during cycling and cannot be re-activated under the same set conditions, as is shown in

Figure 4-21(a), leading to a gradually increased R_{LRS}. This region is named as the defect deactivated region (DDR). This is confirmed in **Figure 4-21(b)** where R_{LRS} is correlated to the width of the increasing DDR and cycles. The impact of different t_{reset}/t_{set} ratio can be explained by that defect profile cannot be modulated farther enough by a small t_{reset}/t_{set}, and LRS cannot be kept constant because of the DDR growth near IL.

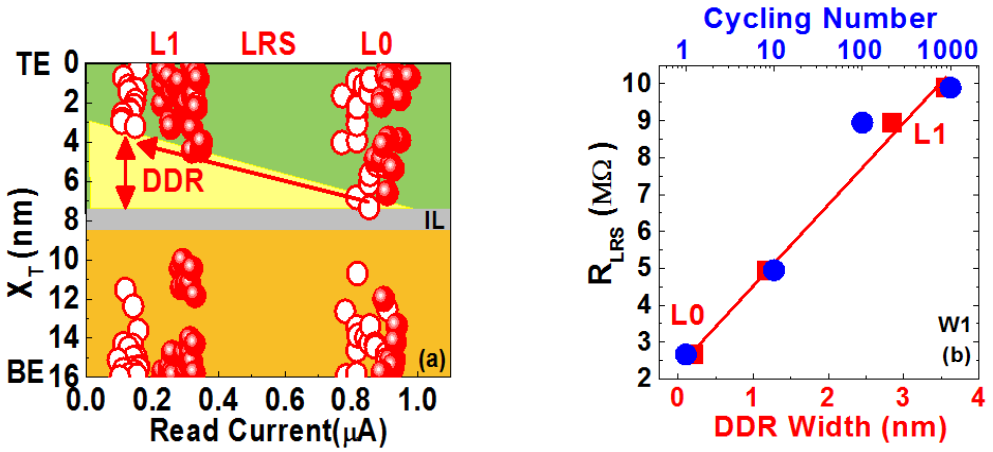


Figure 4-21 Corresponding to **Figure 4-20**, degradation of LRS is hardly affected by treset/tset ratio, either 10^3 (●) or 10^4 (o), because defects are gradually deactivated during cycling in a region in TiO_2 near IL (DDR). (b) R_{LRS} correlates with the width of DDR and the cycling number, confirming the correlation between the width of defect-"less" region and the resistance.

The asymmetric I-V and large differences in set/reset times & voltages in a-VMCO RRAM indicate that defect profile modulation may involve different de-activation and re-activation processes, characterized by different energy [206] and/or switching kinetics [214]. It is speculated that the DDR growth could be caused by the imbalance between the set and reset processes, gradually accumulating through endurance cycling, which requires higher energy to restore.

4.4.1 Improvement on operation condition

It has been introduced in Chapter 3 that HfO_2 RRAM devices that are stuck-at-HRS can be recovered by a stronger set operation through the removal of defects surrounding the CFR. A similar approached can be applied to help with the degradation in a-VMCO RRAM devices. After the LRS degrades to a certain extent, as shown in the blue dots in **Figure 4-22**, it can be recovered to its fresh level by applying a higher $|V_{set}|$, because the DDR, which prevents the defect profile from returning towards IL, has been re-activated.

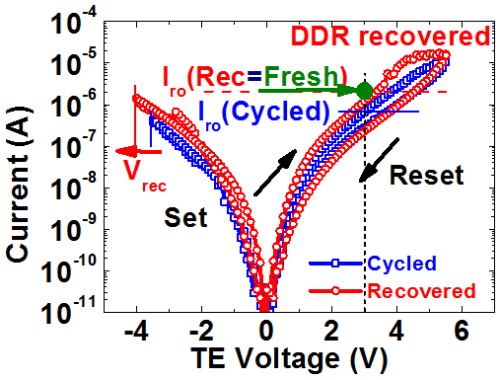


Figure 4-22 After an a-VMCO RRAM degrades, a higher positive V_{set} (V_{rec}) can re-activate the DDR and recover the LRS ($I_{ro,rec}$) to the fresh level ($I_{ro,Fresh}$).

Based on the above understanding, two programing strategies are attempted to improve the endurance: The first is to insert a recovery voltage, V_{rec}, after a certain number of AC cycles to re-activate the DDR, as shown in **Figure 4-23**. A recover set operation with Vset=-4.7 V is applied after each 20 cycles. This mitigates the instability to a certain extent, but cannot stop DDR from re-growing. The second is to apply the incremental step pulse programming (ISPP) with incremental t_{reset}/t_{set} ratio or V_{set} & V_{reset}. As shown in **Figure 4-24**, it is observed that both LRS and HRS can be stabilized

for a limited number of cycles, and then the device failed because either V_{reset} becomes too high and causes breakdown, or t_{reset} becomes too long for practical memory operation. Since the endurance issues cannot be solved by the optimization of operation conditions, device structure optimization is explored for further improvement next.

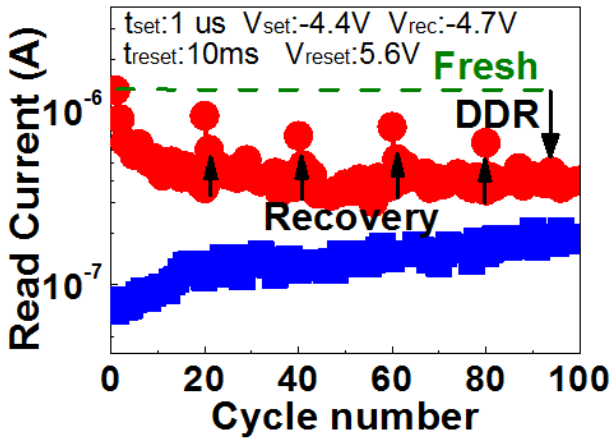


Figure 4-23 A higher recovery voltage, V_{rec}, is inserted after a certain number of AC cycles to re-activate DDR. DDR re-growing cannot be stopped, causing window reduction.

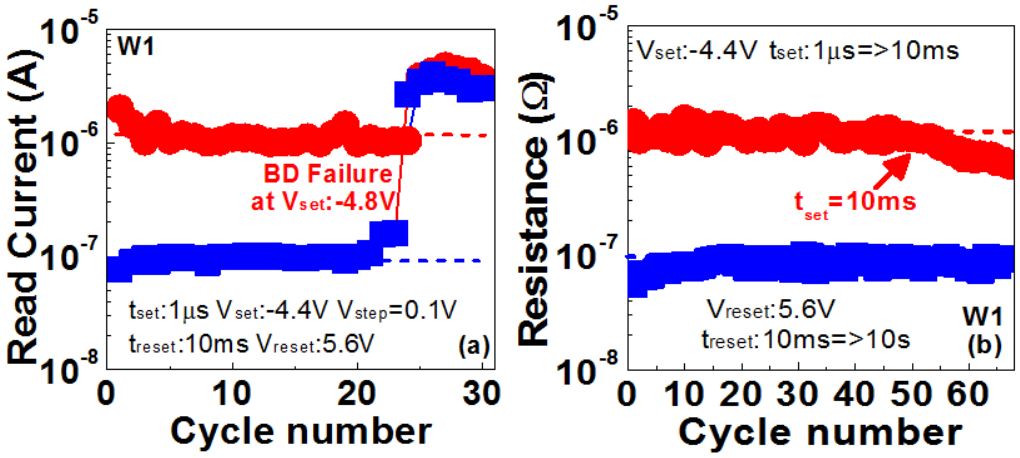


Figure 4-24 ISPP with (a) incremental V_{set} & V_{reset} and (b) incremental t_{reset}/t_{set} ratio applied to retain the full resistance window. Device failed because either (a) V_{reset} becomes too large and causes breakdown, or (b) t_{reset} becomes too long for practical memory operation.

4.4.2 Improvement on device structure

In order to suppress the DDR growth, an additional interfacial layer (ILA) is inserted between the TiO_2 switching layer and the a-Si barrier layer. The optimal conditions are explored, similar to the a-VMCO RRAM devices without IL, as shown in **Figure 4-25**. It is observed that, in a fresh device, increasing V_{reset} and V_{set} accelerates switching but soon causes breakdown.

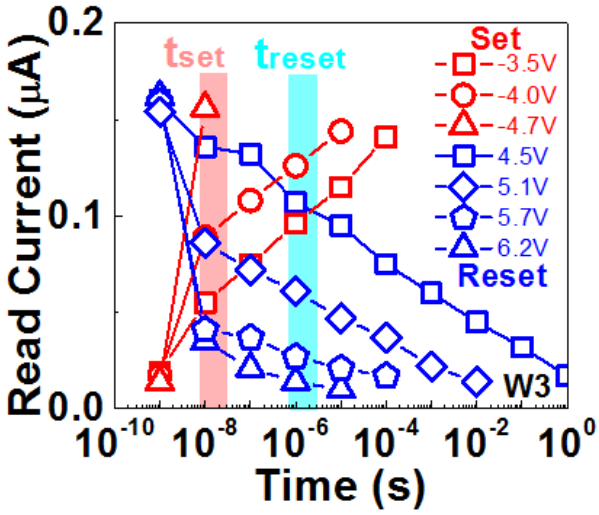


Figure 4-25 AC set/reset conditions in W3. t_{reset} and t_{set} become much shorter than that in W1 (**Figure 4-19**) after device structure optimization by the insertion of ILA and the thinner a-Si layer in W3.

Typical endurance result without V_{rec} and verification is shown in **Figure 4-26** (W2). The window keeps almost constant during the initial 10^3 cycles and then reduces gradually. The resistance window disappears by stuck-at-HRS after about 10^6 cycles. Defect profiles after the failure reveal that the DDR growth and LRS degradation can only be postponed, but not avoided. The device structure is then further optimized by reducing the a-Si thickness to 5 nm. The window closure is avoided and the state drift is reduced at 10^6 unverified cycles, confirmed by the defect profiles. Thinning the barrier layer leads to a net increase of the electric field throughout the full stack, presumably

increasing the modulation strength.

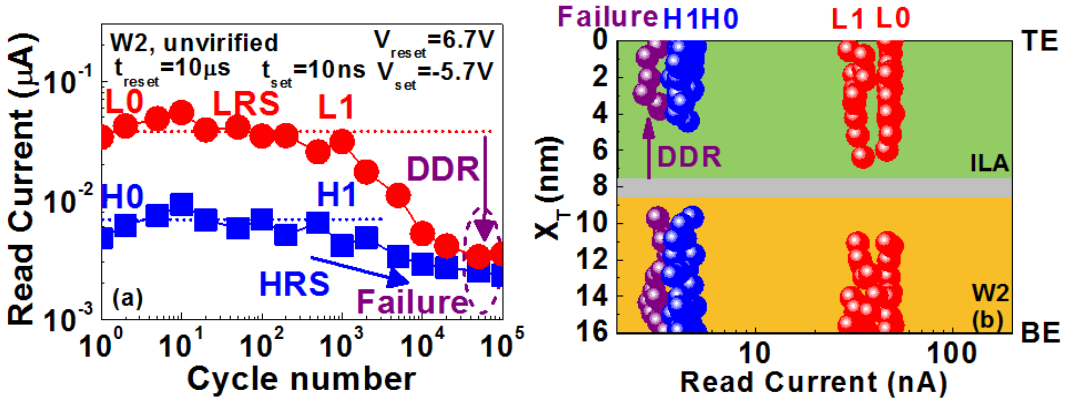


Figure 4-26 An additional interfacial layer (ILA) is inserted between TiO_2 and a-Si. The window shift is improved within 10^3 unverified cycles under optimal set/reset condition. (b) DDR growth is indeed postponed, but still cannot be avoided, leading to the failure at 10^4 cycles.

Inserting the ILA could modulate the defect's structure, improve its stability, and increase the reset/set speed significantly. Further evidence is needed to verify these speculations.

It is desirable that the window closure or state drift can be recovered in practical operations, so an ISPP-based verification algorithm is developed. As mentioned previously, in the AC endurance test read-out is done 3 times per decade. After each read-out, the resistance is verified against a target value. The conventional ISPP method employs either incremental pulse amplitude or width for verification, which will easily breakdown the device or the verification will takes too long, respectively. A new algorithm combines the two methods together as is shown in **Figure 4-27**: the verification is firstly done with incremental pulse widths in the inner loop. If the resistance is not reaching the target value after a certain number of attempts, the pulse width will be incremented in the outer loop. This algorithm improves the verification

efficiency and lowers the risk of breakdown.

By applying this verification algorithms with incremental V_{set} & t_{set} at certain cycling intervals (3 times per decade), a stable resistance window of 10 can be restored for 10^6 cycles as is shown in **Figure 4-28(a)**, showing excellent endurance performance and paving the way for a-VMCO's practical applications. The corresponding defect profile modulation is confirmed by RTN in **Figure 4-28(b)**.

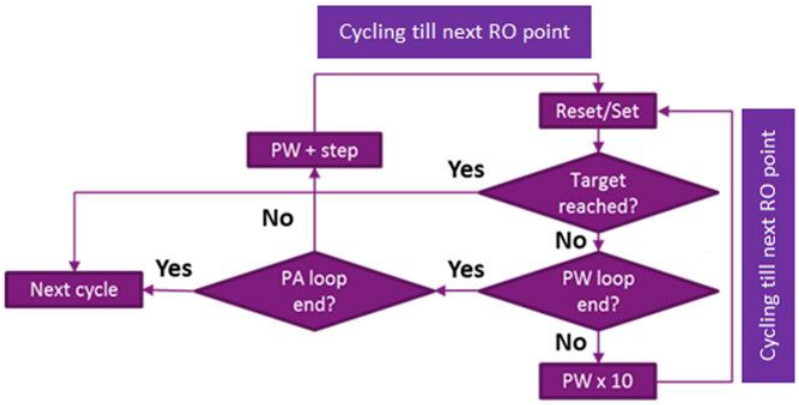


Figure 4-27 Verification algorithm with incremental width (inner loop) and amplitude (outer loop).

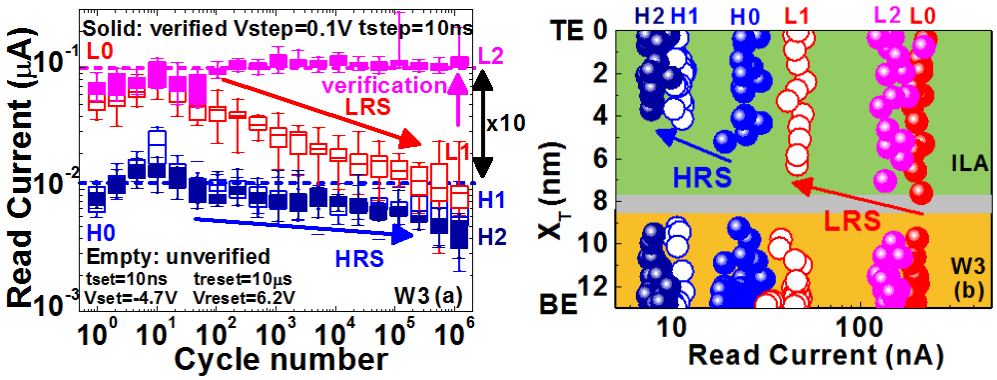


Figure 4-28 (a) Reducing the a-Si thickness to 5nm improves the unverified resistance window up to 1M cycles (empty symbols). A stable resistance window of 10 for 10^6 cycles (solid symbols) can be restored and the endurance drift at LRS can be overcome by applying verification at certain cycles with incremental V_{set}/t_{set}. Both have good reproducibility (see error bars). (b) Corresponding defect profiles without and with verification confirm the results.

4.5 CVS degradation

In this section, the defects in non-filamentary a-VMCO RRAM have been investigated with the focus on their impact on stress-induced RTN distribution. Defects profiles will be extracted during different stages of constant voltage stress (CVS), based on the RTN technique. Statistical analysis will be carried out to characterize the RTN amplitude distribution. This will provide insights into the operation and degradation of a-VMCO device, assisting its further improvement.

RTN measurements are also inserted during the CVS stress to extract the defects induced by the stress, as shown in **Figure 4-29**. RTN amplitude distribution is analysed and evaluated statistically in both fresh and degraded devices.

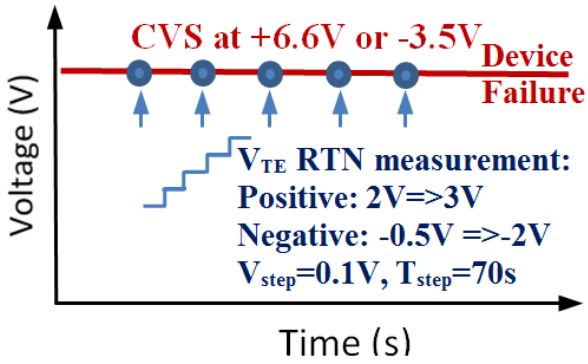


Figure 4-29 Test procedure for combined CVS and RTN measurements to extract the defects during the stress. V_{TE} during the RTN sweeping has the same polarity as the CVS stress.

4.5.1 Degradation and RTN in stressed a-VMCO device

As shown in **Figure 4-30**, a higher V_{reset}, can enhance the on/off resistance window by a factor of ×3, which is desirable for both its digital and analogue applications. The higher

program voltage, however, may cause degradation in the memory cell and lead to earlier device failure. To investigate the degradation mechanism, a CVS voltage, V_{stress}, is applied to the TE at either negative or positive polarity.

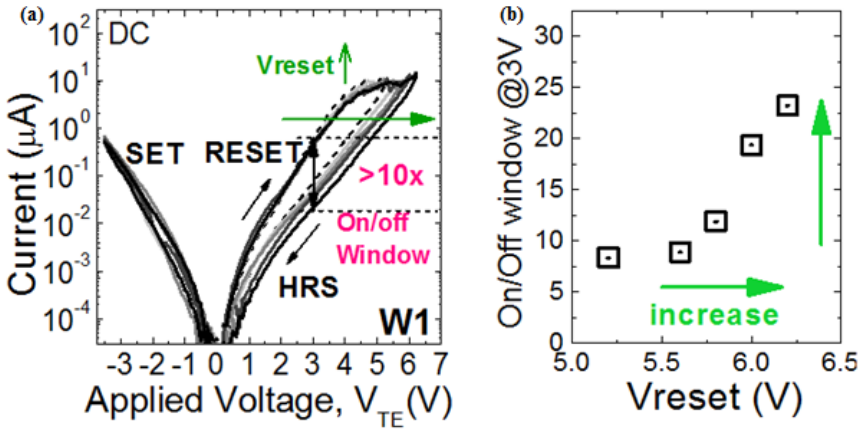


Figure 4-30 (a) DC I-Vs of device switching at increased V_{reset} (b) Larger resistance window can be achieved by increasing the V_{reset}.

A typical I-t characteristics during the negative CVS is shown in **Figure 4-31(a)**, where V_{stress}=-3.5V is applied on a fresh device, which leads to the memory cell set at LRS. The device exhibits a two-stage cell degradation process. The CVS current is stable in the 1st stage, and there is a large current fluctuation in the 2nd stage, before the device reaches the final failure. As shown in the inset of **Figure 4-31(a)**, the stress induced CVS current fluctuation can be as large as 3~8 times of the initial current, which does not exist in the fresh device. This is a strong indication of stress-induced defects generation.

Figure 4-31(b) shows that under a positive CVS, V_{stress}=6.6V, which leads to the memory cell reset at HRS, the two-stage cell degradation process can also be observed. The slight CVS current reduction in the 1st stage is caused by the further reset process,

and, the large CVS current fluctuation in the 2nd stage also indicates defects generation during the cell degradation.

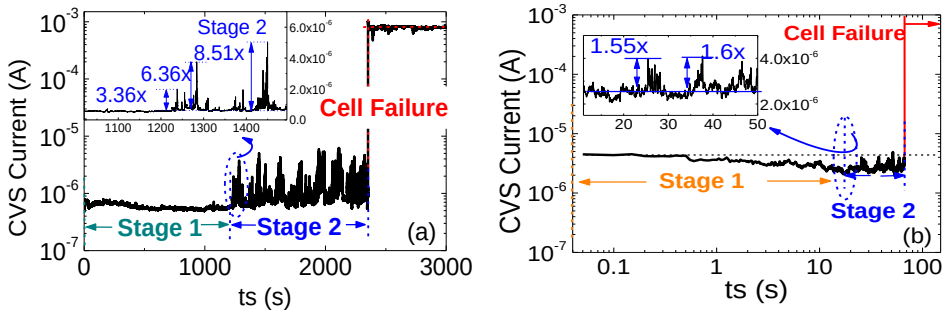


Figure 4-31 Typical I-t characteristics in a-VMCO RRAM. (a) Negative CVS (V_{stress}=-3.5V) is performed on a fresh device at LRS. (b) Positive CVS (V_{stress}=6.6V) is performed on a device at HRS. The inset is the zoom-in of I-t when large fluctuations start to occur, as marked out by dotted circle. Current fluctuation of 50% is used as a criterion to define the onset of the 2nd stage.

Figure 4-32 shows the measured RTN signals during the negative CVS. RTN amplitude in the 1st stage is small at ~4%, as shown in **Figure 4-32(b)**. This is consistent with the results in devices at fresh LRS state shown in **Figure 4-13(a)**. In the 2nd stage, however, RTN signals with both small and large amplitudes are observed, at ~4% and ~20%, respectively, as shown in **Figure 4-32(c)** and **(d)**. Statistical analysis of the RTN amplitude against the measurement bias during the 1st and 2nd stages of negative CVS are shown in **Figure 4-33(a)** and **(b)**, respectively. It confirms that a larger range of relative RTN amplitudes, including both small and large RTNs varying from 2% to 40% are observed in the 2nd stage, while only small RTNs less than 10% are observed in the 1st stage. Similar observations are also found during the positive CVS, as shown in **Figure 4-33(c)** and **(d)**.

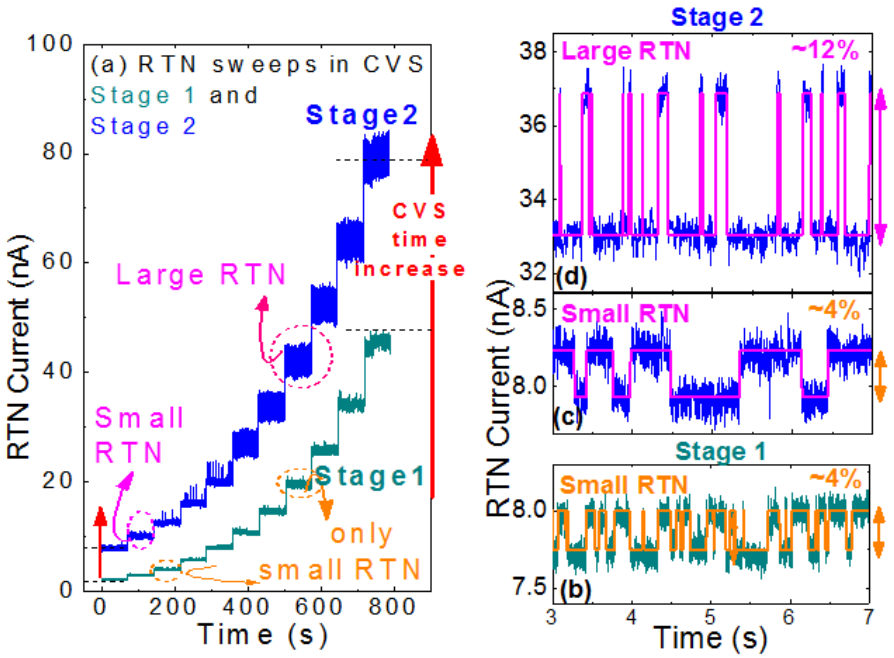


Figure 4-32 (a) RTN signals during the incremental bias sweep measurements inserted during the negative CVS. (b) RTN signal during stage 1: only RTNs with small amplitude are observed. Both small (c) and large (d) RTNs are observed in stage 2.

It has been observed in MOSFETs that defects are generated in gate dielectrics under the NBTI/PBTI stress following exponential law kinetics. It is likely that in a-VMCO devices the large CVS current fluctuation, as shown in **Figure 4-31**, and the large RTN amplitudes, as shown in **Figure 4-33(b) & (d)**, in the 2nd stage of CVS are correlated and both are caused by progressive defect generation. In the 1st stage, the defect generation is less important and the small noises are mainly originated from the trapping/detrapping of the pre-existing defects. In the 2nd stage, as the number of generated defects increase, local percolation conduction paths are gradually formed, and the electron trapping/detrapping via those paths leads to RTN with large amplitude, alongside with RTN with small amplitude resulted from those defects not located in the percolation paths. This speculation is supported in **Figure 4-34**, where the CDF distributions of RTN amplitude have long tails at large RTN amplitude in the 2nd stage,

similar to that in the filamentary devices at HRS shown in **Figure 4-13(b)**, indicating formation of percolation conduction path. There also exist small RTN amplitudes overlapping with that in the 1st stage.

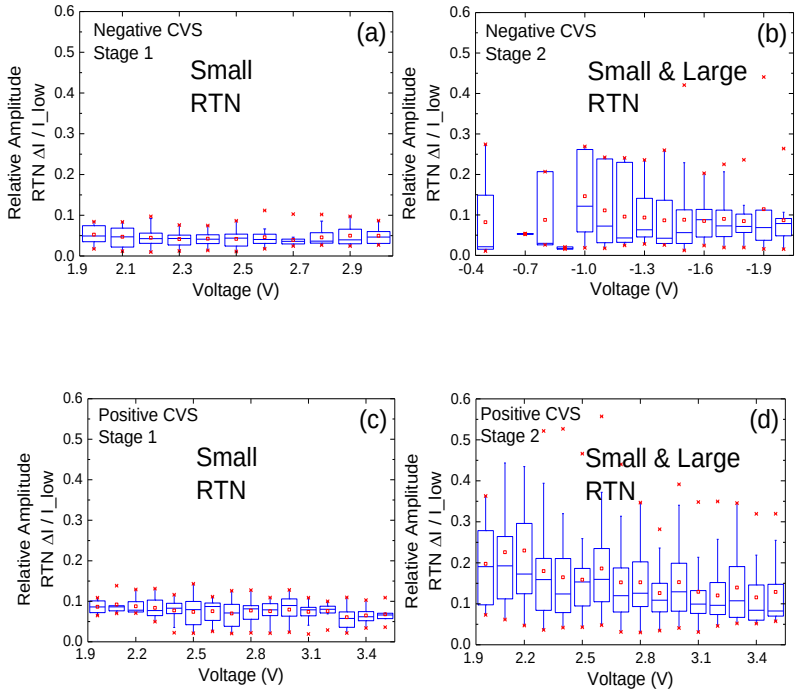


Figure 4-33 Statistical analysis of the relative RTN amplitude dependence on the measurement bias during (a) & (b) negative and (c) & (d) positive CVS in a-VMCO during the 1st and 2nd stage.

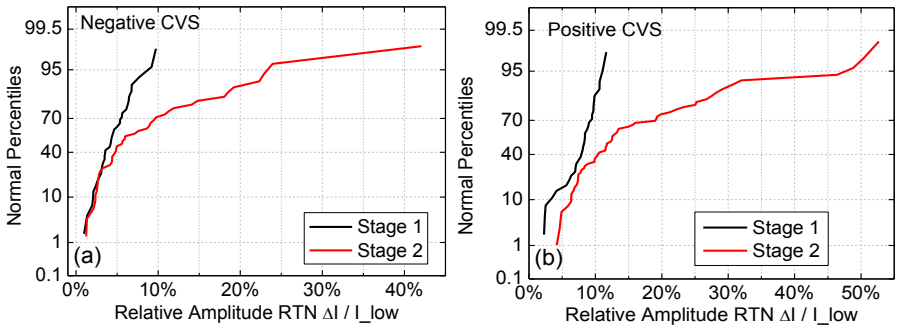


Figure 4-34 CDF plot of the relative RTN amplitude distribution during the 1st and 2nd stage of (a) Negative and (b) Positive CVS.

4.5.2 Generated defects and percolation path in 2nd CVS stage

In order to provide direct experimental evidence to confirm the percolation current path formation in the 2nd stage during CVS, induced by defect generation as we speculated in the last sub-section, defect locations are extracted from the small RTNs in the 1st stage and from the large RTNs in the 2nd stage, respectively, during both the positive CVS, as shown in **Figure 4-35(a)**, and the negative CVS, as shown in **Figure 4-35(b)**.

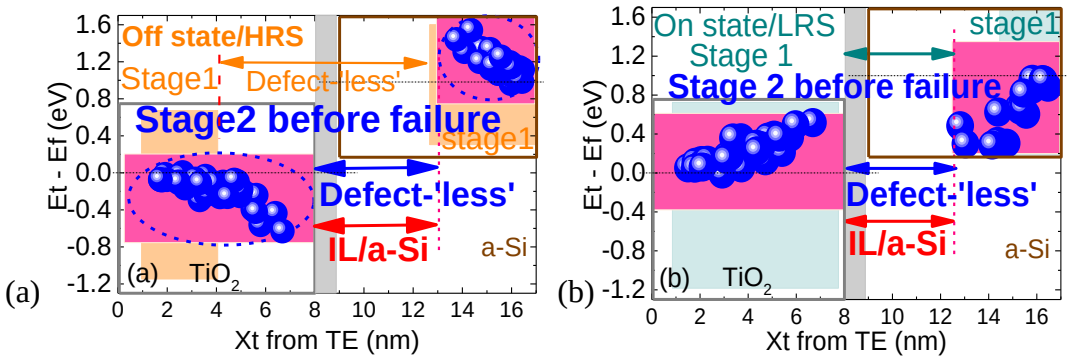


Figure 4-35 Percolation path formation by defects generation in the 2nd stage of (a) positive and (b) negative CVS, in addition to the pre-existing defects observed during the 1st stage. The defect-'less' region in a-Si is the most robust region. Large noises observed in the 2nd stage are originated from the trapping/detrapping of defects along the percolation path, especially within and near the defect-'less' region, similar to those within and near the constriction region of the filamentary devices.

During the positive CVS, the a-VMCO device is reset at HRS, i.e. the off state. The profile of the defects extracted from the small RTNs in the 1st stage is represented by the orange rectangle in the background of **Figure 4-35(a)**. Since there is a wide defect-'less' region near the IL at HRS, there is no percolation current path formed across the stack in this stage. During the 2nd stage, however, the defects extracted from the large RTNs show a different profile, which not only overlaps with that in the 1st stage, but also approaches the IL region across the TiO_2 layer, as shown in **Figure 4-35(a)**. Before the final failure occurs, defects extracted from the large RTNs can also be observed in

the a-Si layer near the IL, which completes the percolation path across the entire stack and leads to the hard breakdown. This suggests that the 'defect-less' region in a-Si is the last stronghold of the dielectric stack before the device's final failure.

During the negative CVS, the a-VMCO device is set at LRS, i.e. the on state. The profile of the defects extracted from the small RTNs in the 1st stage is represented by the light blue rectangle in the background of **Figure 4-35(b)**. Although there is no defect-'less' region in TiO_2 near the IL at LRS in the 1st stage, the density of pre-existing defects is not high enough to form a percolation current path across the stack, which can only lead to small noises. During the 2nd stage, the defects extracted from the large RTNs show that the defect are generated across the TiO_2 layer, in addition to the pre-existing defects, as shown in **Figure 4-35(b)**, which forms a percolation current path and leads to large noises. Defects generation in a-Si layer near the IL layer can also be observed at final stress stage, and a complete percolation path is formed across the entire stack, leading to the final hard breakdown. In fact, both pre-existing and generated defects should contribute to the percolation path formation in the 2nd stage, as both the small and large RTNs are observed in the 2nd stage, as shown in **Figure 4-32 & Figure 4-33**.

The striking similarity in the CDF distributions of RTN amplitudes between the unstressed filamentary device at HRS (**Figure 4-13b**) and the stressed non-filamentary devices in the 2nd CVS stage at both HRS and LRS as shown in **Figure 4-33** also provides strong supporting evidence for the percolation formation in a-VMCO. Large noises observed in the 2nd CVS stage in a-VMCO are originated from the trapping/detrapping of defects along the percolation path, especially within and near the

defect-‘less’ region, which is similar to the filament and its constriction region in the filamentary devices. The quality of the defect-‘less’ regions around the interfacial layer region is, therefore, critical to a-VMCO device performance. A poorer a-Si/TiO_2 interface quality could introduce large read noises and lead to earlier failure. According to the results in this work, improvement of the a-Si/TiO_2 interface quality should lead to better a-VMCO device performance.

4.6 Summary

In this chapter, by utilizing the defect profiling technique based on RTN, in-depth analysis is carried out for the resistive switching mechanism, endurance optimization and CVS degradation of a-VMCO RRAM devices. This provides insights and guidance for understanding and optimizing a-VMCO for its practical memory application. It has been demonstrated in this work that the resistive switching in a-VMCO device is controlled by the profile modulation of pre-existing defects near the IL. The a-Si layer acts as an oxygen scavenging layer to provide a profile of defects in the TiO_2 layer in the form of oxygen vacancies. The defect profile can be modulated by the external bias, as the reset occurs in the bias between +5V and +6V, while the set occurs at around -3V. The difference in set and reset bias may be partially due to the 1V flat-band voltage as shown in **Figure 4-5(b)**, and partially due to the asymmetric dual-layer structures. Details of the defect movements in the stack are subject to further investigations. The reduced impact from individual defect of non-filamentary switching leads to much smaller resistance variability and read instability in unstressed a-VMCO device [215].

For the endurance optimization, defect profile modulation in TiO_2 is correlated with the

analogue switching, and the gradual growth of a DDR near its interface with a-Si causes the LRS endurance instability, while t_{reset}/t_{set} ratio is found critical for the HRS instability. Under this guidance, a stable resistance window of 10 is restored for 10^6 cycles through combining optimizations of device structure and set/reset conditions.

The CVS degradation mechanism is identified. Defect generation induced percolation path formation is experimentally observed and correlated with the larger RTN amplitude and wide distribution in the stressed device. The quality of the IL region is found to play a critical role in memory cell performance. The large read instability induced by defect generation and percolation path formation in severely stressed devices could be improved by further material and structure optimization, especially around the interfacial layer region [212, 216-219]. This work provides insight guidance for further process and device structure optimization of a-VMCO device.

5 Characterisation of Ta_2O_5 RRAM

5.1 Introduction

As mentioned in Chapter 2, there are a large number of materials that exhibit resistive switching behaviour. Among these materials, binary oxides exhibit the best switching performance in terms of switching speed and switching endurance. TaO_x-based RRAM has shown fast switching in the sub-nanosecond regime, good CMOS compatibility and record endurance of more than 10^{12} cycles, making it one of the competitive candidates because such a long endurance capability enables it to be used in embedded memory applications and potentially can make a change of the memory hierarchy [82,129,220,221]. X-ray photo electron (XPS) [193] and electron energy-loss spectroscopy [222] reveal that TaO_x usually consists of two phases: One is the stoichiometric Ta_2O_5, a typical crystal structure contains 22 Ta and 55 O atoms in the primitive cell for the low temperature phase. In its atomic structure, some oxygen sites are partially occupied to satisfy the stoichiometric ratio, resulting in its insulating dielectric properties; The other phrase is the substoichiometric TaO_2, which is a conducting phase exhibiting the resistivity three orders smaller than Ta_2O_5 [23]. The outstanding endurance performance is believed to be linked to such dual oxide phase diagram [223].

Both bipolar and unipolar switching behaviours have been observed in TaO_x RRAM. As discussed in Chapter 1, for unipolar switching: the switching polarity depends only on the amplitude of the applied voltage, but not its polarity, so that set/reset can occur at the same polarity. For bipolar switching, the switching direction depends on the polarity of

the applied voltage, so that set can only occur at one polarity and reset can only occur at the reverse polarity. This is illustrated in **Figure 5-1** below.

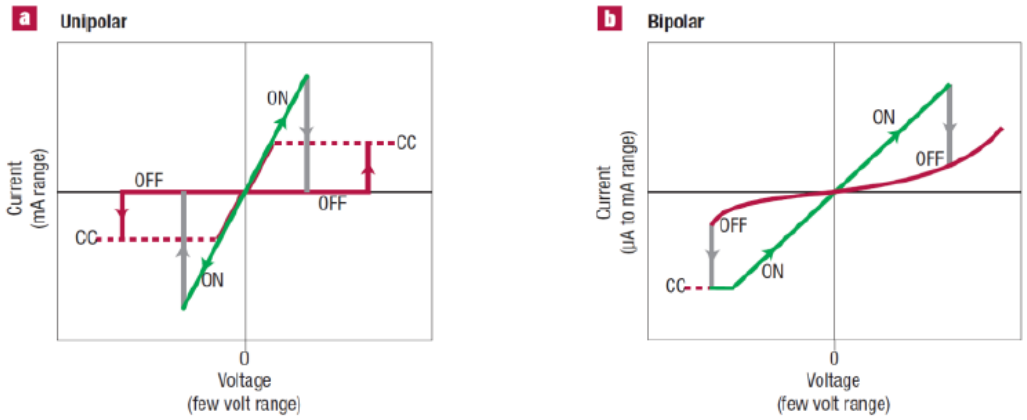


Figure 5-1 Schematic of DC I-V characteristics of resistive memory in (a) unipolar switching mode and (b) bipolar switching mode. ON/OFF refers to low resistance and high resistance state, respectively. CC stands for current compliance [23].

It has been observed in the literature that bipolar and unipolar switching is determined by different materials/electrode combinations as shown in **Table 1-2** and **Table 1-3**. Even with the same oxide material, the switching mode can be different. In most cases, bipolar switching can be achieved by using oxidizable electrode such as Ti, Hf, TiN, etc., while unipolar switching is obtained by using inert electrodes for both sides [13]. In some cases, both unipolar and bipolar switching can be achieved on the same material system, depending on the polarity of the voltages which are applied to the device [191].

It has also been observed that the conversion from unipolar to bipolar can be achieved by inserting a Ta_2O_5 layer into the Pt/TaO_x/Pt structure [224]. The operation mode is an important issue for RRAM devices: from a circuit point of view a unipolar switching is preferable because of simpler selector device. However, unipolar switching normally

requires very high reset current compared with bipolar switching. Anyway, the co-existence of those two operation modes may bring reset-set instability [225] and therefore it is necessary to have a clear understanding of the physical mechanisms during resistive switching, especially for the unipolar mode. In this work, we have observed that bipolar and unipolar modes can co-exist in the same device structure under different operation conditions, as will be discussed in detail in the following sections.

For Ta_2O_5/TaO_x bilayer structured devices different models have been proposed [82, 223, 226] based on physical characterisation or electrical measurement [82, 193, 220, 227-233]. The switching process is attributed to the phase transition effect between Ta_2O_5 and TaO_2 based on the calculation of Gibbs free energy [193, 227]; or the restoration/rupture of the Ta-rich CF inside TaO_x RRAM based on experimental observations [82, 220, 230-232]; or the Vo-formed CF's restoration/rupture caused by Vo movement [232]. Besides, the generation and recombination (G-R) of V_O with O^{2-} in an oxide layer has also been used to account for the switching behaviours of RRAM [66, 234-236]. These inconsistent models make it difficult to quantify and optimize the performances of TaO_x RRAM, largely due to the lack of microscopic description of resistive switching based on experimental results, especially for the unipolar switching mode.

In this chapter, an in-depth analysis of the $TiN/Ta_2O_5/TaO_x/TiN$ structured RRAM is carried out with the RTN-based technique introduced in Chapter 2 and 3, for both bipolar and unipolar switching modes. This chapter is organized as follows: After a description of the devices and measurement setup in Section 5.2, the results and

discussions are in Section 5.3, which includes the observation of the CFR and defect movement in the bipolar mode, and defect energy alteration in the unipolar mode.

5.2 Devices and experiments

5.2.1 Devices

The crossbar RRAM devices with a TiN/Ta_2O_5/TaO_x/ TaN/TiN stack were processed in an integrated process [19], which yielded device sizes of various areas ranging from 3×3 μm^2 down to 40×40 nm^2, as shown in **Figure 5-2**. The TiN BE was sputtered at room temperature and patterned. A 4 nm-thick stoichiometric Ta_2O_5 layer of high quality was deposited by ALD. Furthermore, a nonstoichiometric 20 nm-thick TaO_x film was deposited by reactive DC magnetron sputtering using a Ta target under oxygen ambient. Without breaking the vacuum, a 10nm thick TaN capping layer was sputtered. Finally, a 30 nm-thick TiN film was sputtered. The patterning of the whole TE stack was done so as to have predefined different lengths (L_{TE}) connecting the TE contact pad to the crossbar device, which enabled an integrated access resistance (R_a) [129]. Depending on the wire length, R_a ranges from 50 Ω to 5 kΩ.

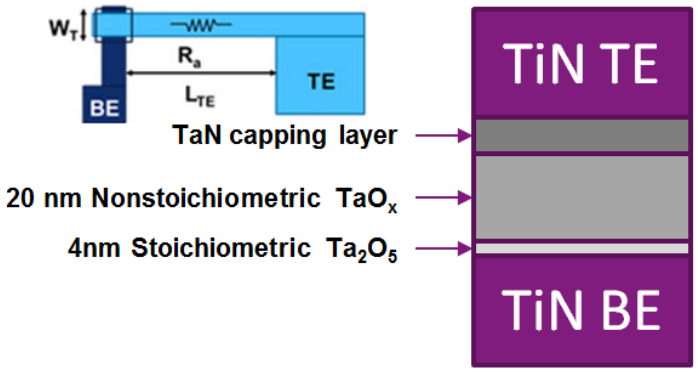


Figure 5-2 Schematic of the TiN/TaOx/Ta_2O_5/TiN structure. The inset shows the schematic top-view structure with access resistance [129].

5.2.2 Experiment procedure

In this chapter, the electrical measurements are conducted by the Measurement System (I) introduced in Chapter 2. The DC measurement programme is similar to that used for HfO_2 in Chapter 3 despite some parameter variations. The AC measurement programme is somehow different. As introduced in Chapter 1, filamentary devices requires a compliance current to prevent the device from hard breakdown during forming/set, which can be achieved either by setting up a compliance current in the analyser, or by using a current-limiting transistor/diode/resistor connected in series to the RRAM device. For the TaO_x device used in this work, the R_a acts as the current-limiting resistor and enables the AC set operation without any other compliance setup. AC pulses can be simply applied on to the TE of the device. The setup for AC operation is illustrated in **Figure 5-3**. AC pulses are generated by the SPGU and applied on the TE. The BE SMU is always grounded. The SMU connected to TE is used for RTN measurement, which is carried out at HRS after AC reset. In this chapter, AC switching is performed in unipolar mode on the devices with a large current limiting R_a. DC switching is performed in bipolar mode on the devices without a large R_a, and a compliance current is set by the analyser.

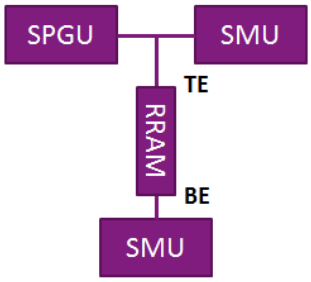


Figure 5-3 Schematic of AC switching setup. The BE SMU is always grounded. AC pulses are generated by SPGU for switching. The SMU connected to TE is used for RTN measurement.

Although the TiN/TaO_x/Ta_2O_5/TiN structure consists of two oxide layers which are similar to the dual-layer a-VMCO device, the single-layer model employed in Chapter 3 can still be employed, because the resistivity of the films increases exponentially with the O/(Ta+O) ratio as shown in **Figure 5-4** [193]. As a slight decrease in the oxygen component would cause a sharp decrease in the resistivity, the stoichiometric Ta_2O_5 layer takes most of the voltage drop across the device and can be regarded as switching layer similar to the HfO_2 layer in Chapter 3. The TaO_x layer, however, is similar to the Hf scavenging layer by acting as an oxygen reservoir and inducing the formation of oxygen vacancies in the adjacent Ta_2O_5 layer. The work function of TiN is 4.5 eV [14] and the electron affinity of Ta_2O_5 is 3.2 eV [15]. Thus the Φ_0 in the extraction model shown in **Figure 2-30** is 1.3 eV.

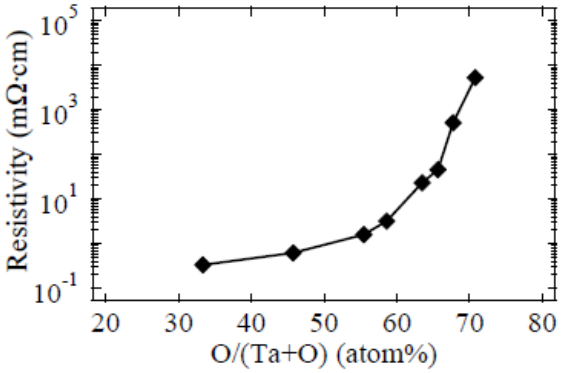


Figure 5-4 Oxygen ratio dependence of the TaO_x resistivity: The resistance increase exponentially with the O/(Ta+O) ratio [193].

5.3 Results and discussions

5.3.1 Bipolar DC switching

The TaO_x device can be switched between HRS and LRS in the same way as for HfO_2,

where V_{set}=1.2 V, V_{reset}=-1.5 V, I_{cc}=20 μA, as shown in **Figure 5-5**.

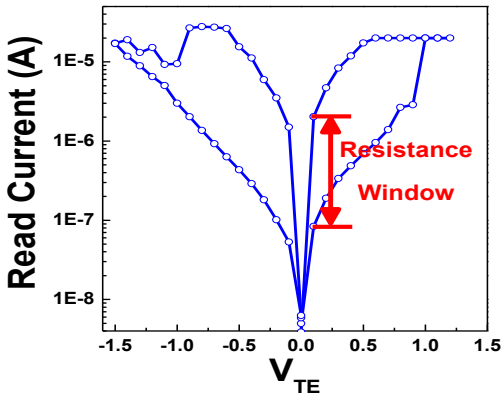


Figure 5-5 Bipolar DC switching of Ta_2O_5 RRAM: V_{set}=1.2 V, V_{reset}=-1.5 V, I_{cc}=20 μA.

By using the RTN extraction method as introduced in Chapter 3, the defect profile in TaO_x device at HRS is shown in **Figure 5-6**. It is observed that, unlike in the HfO_2 device where the CFR locates near BE at 70% ~ 80% of the tunnelling conduction region (TCR), the CFR appears in the middle of Ta_2O_5 device. This could be explained by the difference in device structures. In the TiN/Hf/HfO_2/TiN structured device, because of the existence of the Hf scavenging layer which acts as the oxygen reservoir, a large number of V_Os are generated near the TE during forming and become the strongest part of the CF. During reset the CF ruptures at the weakest point which has to be close to the BE, and in the subsequent set operation, it is where the filament restores. For Ta_2O_5, however, although the nonstoichiometric TaO_x also acts as the scavenging layer, the effect is not as strong as Hf because of the lower metal component and lower oxygen-affinity [237], so that the CFR is in a more balanced position around the middle of the TCR region.

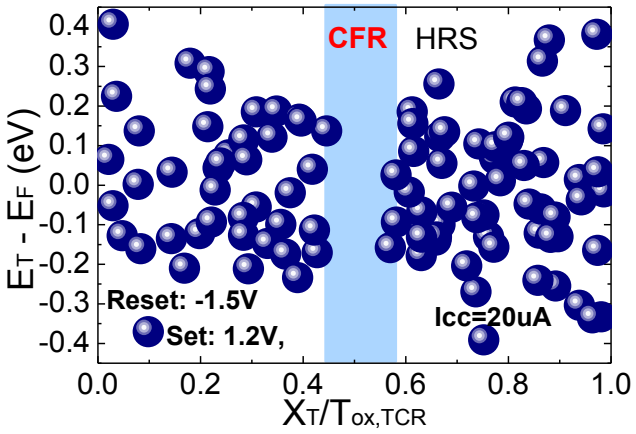


Figure 5-6 Defect profile at HRS in Ta_2O_5 device. The CFR is observed in the middle of the TCR. V_{reset}=-1.5 V

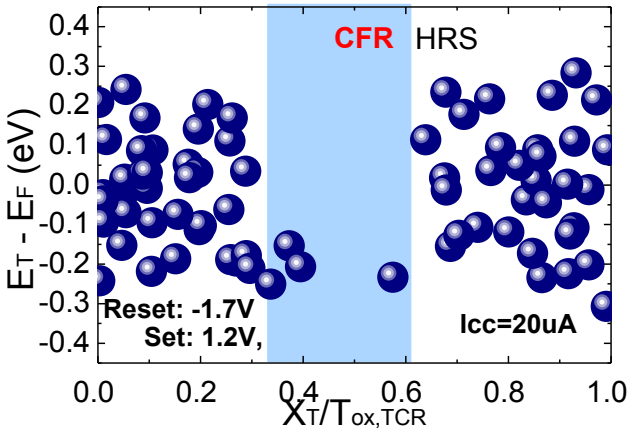
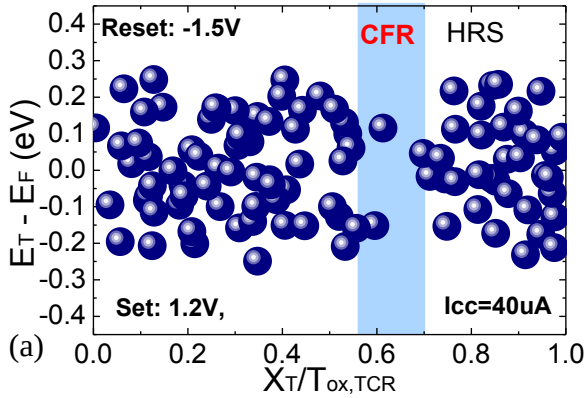


Figure 5-7 Defect distribution under V_{reset}=-1.7 V: larger V_{reset} leads to wider CFR and causes higher R_{HRS}, similar to the bipolar HfO_2 RRAM.

Similar to HfO_2, the impact of V_{reset} and I_{cc} on the CFR is analysed. Increasing the V_{reset} from -1.5 V (**Figure 5-6**) to -1.7 V (**Figure 5-7**) widens the CFR, and increasing the compliance current from 20 μA (**Figure 5-6**) to 40 μA (**Figure 5-8**) narrows the defect energy distribution due to the stronger regulation power caused by the larger set current.



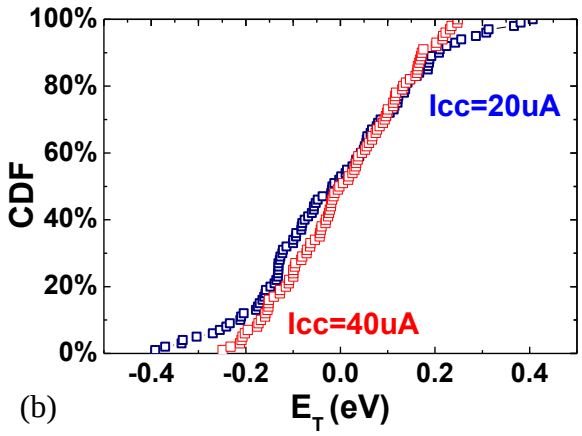


Figure 5-8 Defect distribution under different I_{cc}. (a) Defect distribution at I_{cc}=40 μA. (b) CDF of the defect energy distribution. Larger I_{cc} leads to narrower defect energy distribution.

Similar to the HfO_2, the correlation between defect movement and CF modification is summarized under negative bias in **Figure 5-9(a)** and under positive bias in **Figure 5-9(b)**. The relative change of the read current amplitudes, ΔI/I, are analysed statistically against the defect locations before and after Vo movements, which shows that the increases of ΔI/I are correlated with Vo moving into the highlighted CF constriction region, either from the TE direction or the BE direction, and the decreases of ΔI/I are correlated with Vo moving out of the CF constriction region, either towards the TE direction or the BE direction. The largest amplitudes of ΔI/I are clearly observed when the CF constriction is either the destination or the origin of defect movements.

These results highlight the importance of the start/stop location of defect movements. ΔI/I is small when defect movement is not involved with the constriction, contributing less to resistance switching. Unlike in HfO_2 where defect interaction between the constriction and BE is relatively rare and has smaller impact on the device resistance, interaction between the constriction and BE in Ta_2O_5 is significant and even comparable with the interaction between TE and constriction. This can be explained with the weaker oxygen affinity of Ta [237] and the more symmetric CF, as defect defects can move into/out of the constriction from both sides driven by the field.

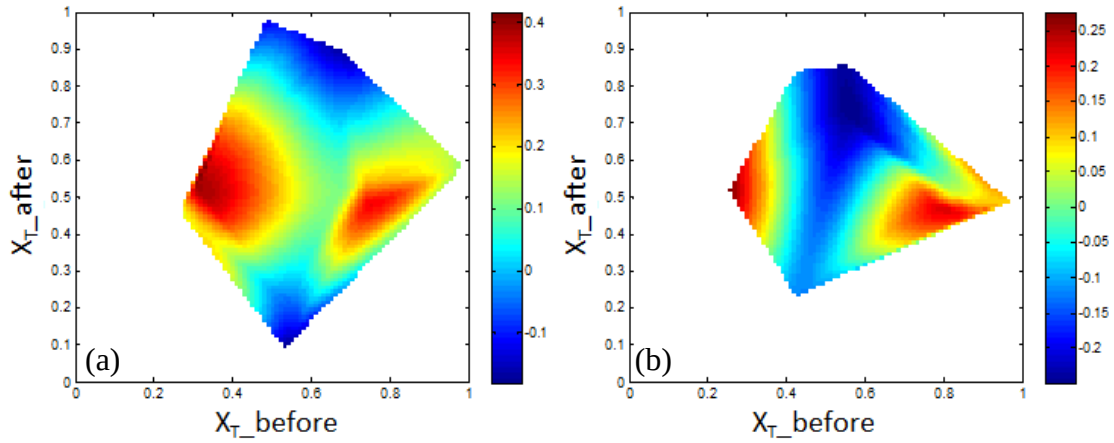


Figure 5-9 Defect movement during normal cycling (a) Under negative V_{TE} ((b) Under positive V_{TE}. The correlation between resistance change and defect movement agrees with bipolar HfO_2 device.

5.3.2 Unipolar switching

From a circuit point of view, a unipolar switching is preferable, since a normal diode can be used as a select device to block snack currents of the other paths in a cross point array. It is unusual that bipolar and unipolar modes could co-exist in the same device, but it can be observed in this work for this type of Ta_2O_5 device. It is firstly observed that if the DC V_{reset} is further increased after the device has been reset, there will be a

U-shaped turn in the reset I-V curve, as shown in **Figure 5-10.** This indicates that if V_{reset} is larger than the critical point, the current will increase as if a "set" process happens.

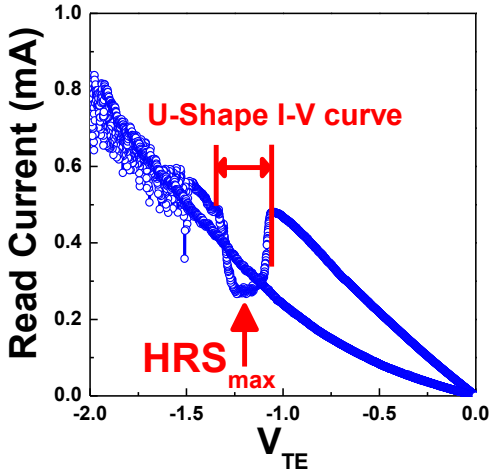


Figure 5-10 DC double sweep for reset with large V_{reset}. The U-shaped turn indicates that a precise controlled negative voltage on the TE might have a set effect on the Ta_2O_5 device

This, however, will cause set-reset instability during switching, as the device cannot be operated normally after this strong reset. The U-shaped turn indicates that a precise controlled negative voltage on the TE might have a set effect on the Ta_2O_5 device. The DC sweep over the critical V_{reset} might be too strong and thus damage the device. This is supported by the following test: a device is firstly reset to HRS by DC sweep, and then negative pulses with different amplitudes are applied to TE. The pulse width is initially fixed at 100 ns. It is observed that the device resistance will decrease if the pulse amplitude exceeds -3V, as shown in **Figure 5-11**. Although the averaged I_{read} increased less than 10%, this is impressive because it is a statistical result. It supports that negative pulse may set the device, showing unipolar switching characteristic.

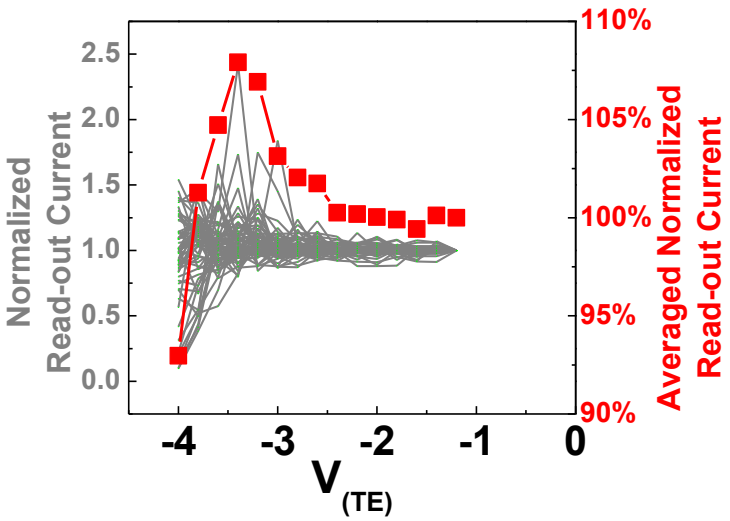


Figure 5-11 Normalised and averaged I_{read} after negative pulses with different amplitude and fixed width are applied onto TE.

The unipolar switching is further proved below. After the optimization of operation conditions, i.e. increasing the set pulse width and reducing the set pulse amplitude, a ×10 window can be achieved, as shown in **Figure 5-12**. Here cycling is done with DC sweep for reset (V_{reset}=-2.0V) and AC pulse for set (V_{set}=-2.0V, t_{set}=10μs), both on TE.

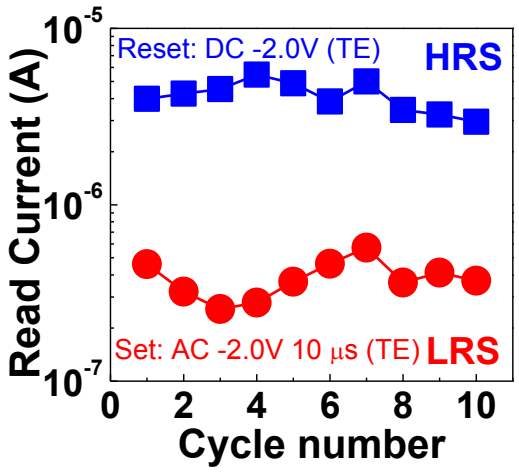


Figure 5-12 Cycling with DC reset and AC set: V_{reset}=-2.0 V, V_{set}=-2.0 V, t_{set}=10 μs. A resistance window of 10 can be achieved.

When unipolar switching by negative DC reset and negative AC set, and bipolar switching by negative DC reset and positive DC set are preformed alternatively, it can be observed that almost the same HRS and LRS levels can be achieved, which means that the same LRS level can be achieved by either AC negative pulse ($V_{set,AC}$=-2.0 V, $t_{set,AC}$=10 μs) or DC positive sweep ($V_{reset,DC}$=-2.0 V), both applied on TE under the reset conditions $V_{set,DC}$=2.0 V, as shown in **Figure 5-13**.

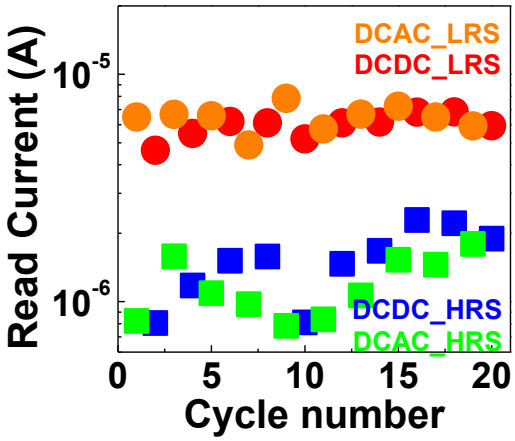


Figure 5-13 Alternating DC reset – AC set ("DCAC") and DC reset – DC set ("DCDC") cycling: the same LRS level can be achieved by either AC negative pulse or DC positive sweep both applied on TE under the conditions of $V_{reset,DC}$=-2.0 V, $V_{set,DC}$=2.0 V, $V_{set,AC}$=-2.0 V, $t_{set,AC}$=10 μs.

Furthermore, if the DC reset is replaced with a negative reset pulse on the TE, it is observed that the Ta_2O_5 device can be successfully switched in the unipolar mode with optimised conditions, V_{reset}=-2.5 V, t_{reset}=100 ms, V_{set}=-2.5 V, t_{set}=10 μs. A ×5 resistance window can be kept for over 100 cycles, as shown in **Figure 5-14**.

It should be noted that a large series resistance is attached to the device which should be taken into account during the operation. The series resistance is estimated to be 5 kΩ [129]. At LRS, the effective voltage drop on the RRAM is 0.25 V and it is 1.25 V at

HRS. The $V_{effective,reset}$ is much smaller than $V_{effective,set}$, which agrees with the observation about unipolar switching in previous reports [224].

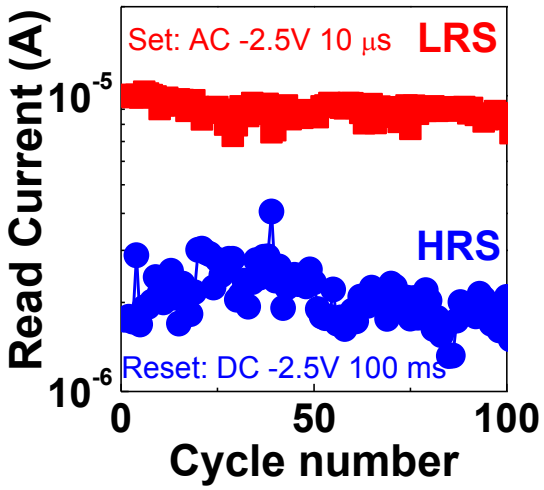


Figure 5-14 Pure AC unipolar cycling: V_{reset}=-2.5 V, t_{reset}=100 ms, V_{set}=-2.5 V, t_{set}=10 μs. Over 100 cycles are achieved with stable LRS and HRS.

In the bipolar mode RTN measurement is carried out under both positive and negative bias, as the RTN measurement process is expected to show how a single defect is disturbed by the electric field of positive/negative polarities and its impact on the device resistance, and how the defects are moved during a stronger set/reset condition. In the unipolar case, this is not applicable anymore, since both set and reset is achieved with negative pulses on the TE. Only negative RTN bias is applied on the TE for extracting defect movement. Defects are extracted at HRS in the unipolar mode, as shown in **Figure 5-15**. Interestingly, there's no clear CFR observable in the TCR. As introduced in Chapter 3, in the bipolar mode, defect movement is closely related to the polarity of bias: under positive TE bias defect tend to move towards BE and under negative TE bias towards TE. In the unipolar mode, the scenario might be different: If defects are still driven by the field, defects have to move in only one direction. After a few cycles the

defects at one end will be exhausted and the device endurance will degrade quickly, which is not the case as shown in **Figure 5-14**. Therefore, the reset process in unipolar mode should be different from the bipolar mode.

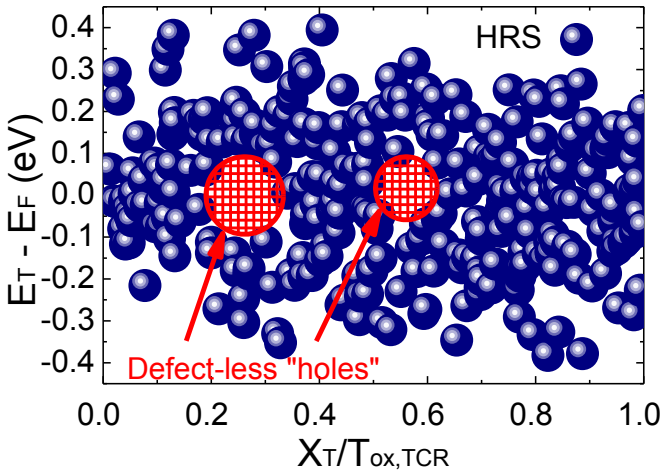


Figure 5-15 Defect distribution at HRS in the unipolar mode. There are several defect-less "holes" along the filament with energy close to the Fermi level.

The reason for unipolar switching is not yet well understood. There are several possible explanations for the unipolar switching mechanism: (1) defect movements are still driven by the field, though not in the same way as in bipolar mode. (2) Defects move in the horizontal direction during set and reset. Defects move away from/closer to the filament makes the CF thinner and thicker, respectively. (3) The defects only experience a change in its energy during switching. Defect energy become closer to the Fermi level during set and away from the fermi level during reset.

If the defect movements are still driven by the polarity of field during set and reset, then there should be a clear correlation between the defect location change and its impact on the device resistance, as is shown in Chapter 3 and previously in this chapter for the

bipolar mode. Defects moving into/out of a critical region will cause significant change in device resistance. However, there is no such clear correlation as shown in **Figure 5-16**. The largest I_{read} increase happens if defects move from 30% to 40% and the largest I_{read} decrease happens if defects move from 30% to 10%. This does not agree with the previous CFR observed in the bipolar mode, and is difficult to explain the switching by defect movement in the vertical direction. The possibility of the assumption (1) that defect movements are along the CF and driven by the field is unlikely, therefore.

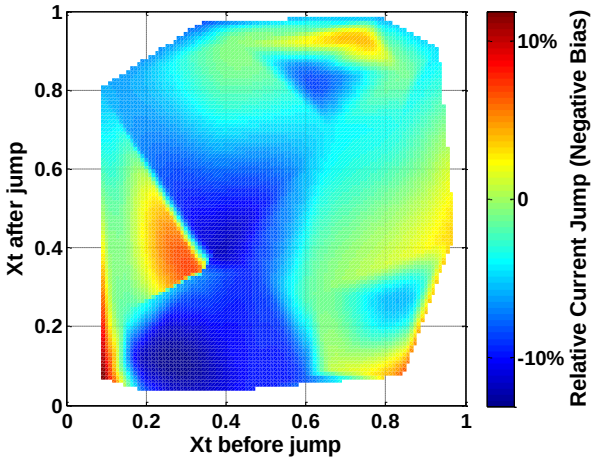


Figure 5-16 Defect movements under negative RTN bias on TE: no clear correlation between defect movement and device resistance change.

For the 2nd possible explanation, Chang et al investigated the switching characteristics of NiO films at different temperatures, and reported that the mechanism is controlled by the Joule heating effect and the stability of filament is governed by the competition between Joule heating and thermal dissipation [238], which caused the defects moving into/out of the CF from the horizontal direction. Since the Joule heating effect does not depend on the polarity of the current, this kind of devices displays unipolar switching behaviour. It is possible that the unipolar switching of Ta_2O_5 device could also be controlled by a similar mechanism. However, exploration of how to detect defect

movement in the horizontal direction is still on-going, as the existing RTN technique can detect only the defect's vertical location and energy level.

The third possible explanation is the alternation of defect energy levels during set and reset. As is introduced in Chapter 3, under the read-out condition, current is mainly controlled by defects with energy aligned to the Fermi level, as the V_{read} is merely 0.1V~0.2V. Defects with energy far from the E_F will contribute less to the I_{read}. If, during set or reset, the energy of defects are somehow changed, i.e., defects are activated or deactivated to different energy levels, this could have an impact on the device resistance.

As shown in **Figure 5-17**, there is a correlation between the absolute value of defect energy before and after jumps with the change in I_{read}. It is observed that defect energy levels change towards far from the Fermi level, either above or below, when the corresponding I_{read} decrease, and vice versa. However, this is just a preliminary observation and there are a lot more clarifications needed to explain the unipolar switching behaviour. For example, there are several "holes" along the filament with energy close to the Fermi level which can be clearly observed in **Figure 5-15**. The lack of defects in those "holes" may reduce the current during read-out and could be one of the possible reasons for HRS. However, by increasing the read-out voltage a little bit, the Fermi level of the electrode would be able to align with the defects in the region outside the "holes" and LRS should be obtained, but this is not observed in the I-V curves. Further investigation is needed to clarify this.

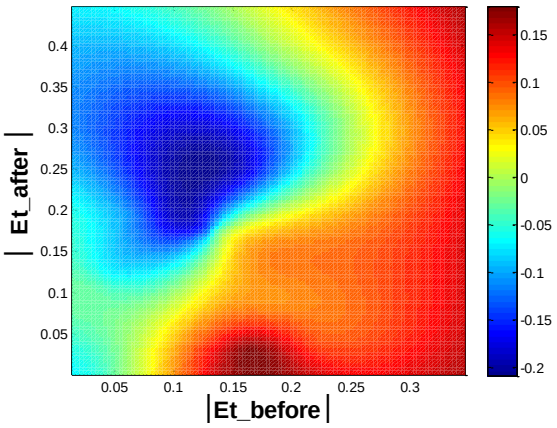


Figure 5-17 Correlation between defect energy change and I_{read}: if defect energy move away from the E_F, I_{read} will decrease, and vice versa.

5.4 Summary

In this chapter an overall picture of the resistive switching mechanism is provided for $TiN/Ta_2O_5/TaO_x/TiN$ structured RRAM device. Bipolar and unipolar resistive switching modes have been observed co-existing in the same device under different operation conditions. For the bipolar mode, defects movement into/out of the CFR is responsible for the resistance switching, similar to that in HfO_2 RRAM. The location of CFR has been identified, which is located at the middle of the TCR, different to that near the BE in HfO_2 RRAM. This may be caused by the weaker scavenging capability of the TaO_x layer and lower oxygen affinity of Ta compared with Hf. For the unipolar mode observed at negative TE biases, it is observed that the effective voltage for set is much stronger than that for reset, and time for set is much shorter than reset. The unipolar switching mechanism, whether it is defect horizontal movement caused by thermal Joule heating, or defect energy alternation, and is still under investigation.

6 Conclusions, future work and outlook

6.1 Conclusions

The characterisation of the novel RRAM devices have been systematically investigated in this thesis. It can be divided into three main parts: (1) the characterisation of HfO_2 filamentary RRAM (2) the characterisation of a-VMCO non-filamentary RRAM and (3) the characterisation of Ta_2O_5 RRAM. Conclusions for each part are given below:

6.1.1 Characterisation of HfO_2 RRAM:

An overall picture of the switching mechanism in bipolar HfO_2-based RRAM has been provided in Chapter 3, showing the repeatable modification of the local constriction. The CF is formed by the generation of defects in the bulk of the oxide during electroforming, resulting in a percolation current path. Under the applied forming electric field, the oxygen ions drift toward the anode leaving behind the Vo forming a local conductive path in the HfO_{2-x}. The CF connects the top oxygen reservoir with the BE at LRS. During the resistive switching cycle the number of Vo in the constriction and the length of the constriction of the hour-glass shaped CF, where it has the least defect number and is closer to the BE, are changed by the defect movement according to the electric field at a bias as low as ±0.2V, inducing the resistance change and eventually leading to the switching of resistance states. The stuck-at-HRS failure is attributed to defects surrounding CFR which can be removed after the recovery. Our RTN based CF characterisation technique provides the physical insights of oxide based filamentary RRAM, improves the understanding of the CF and its constriction at defect level, reveals a new endurance failure mechanism and thus provides a useful tool for

RRAM technology development.

6.1.2 Characterisation of a-VMCO RRAM:

In Chapter 4, by utilizing the defect profiling technique based on RTN, in-depth analysis is carried out for the resistive switching mechanism, endurance optimization and CVS degradation of a-VMCO RRAM devices. This provides insights and guidance for understanding and optimizing a-VMCO for its practical memory application. It has been demonstrated in this work that the resistive switching in a-VMCO device is controlled by the profile modulation of pre-existing defects near the IL. The a-Si layer acts as an oxygen scavenging layer to provide a profile of defects in the TiO_2 layer in the form of oxygen vacancies. The defect profile can be modulated by the external bias, as the reset occurs in the bias between +5 V and +6 V, while the set occurs at around -3 V. The difference in set and reset bias may be partially due to the 1 V flat-band voltage as shown in **Figure 4-5(b)** (Page 108), and partially due to the asymmetric dual-layer structures. Details of the defect movements in the stack are subject to further investigations. The reduced impact from individual defect of non-filamentary switching leads to much smaller resistance variability and read instability in unstressed a-VMCO device.

For the endurance optimization, defect profile modulation in TiO_2 is correlated with analogue switching, and the gradual growth of a DDR near its interface with a-Si causes the LRS endurance instability, while t_{reset}/t_{set} ratio is found critical for the HRS instability. Under this guidance, a stable resistance window of 10 for 10^6 cycles is restored through combining optimizations of device structure and set/reset conditions.

For the CVS degradation, its mechanism under constant-voltage-stress is identified.

Defect generation induced percolation path formation is experimentally observed and correlated with the larger RTN amplitude and wide distribution in the stressed device. The quality of the IL region is found to play a critical role in memory cell performance. The large read instability induced by defect generation and percolation path formation in severely stressed devices could be improved by further material and structure optimization, especially around the interfacial layer region. This work provides insight guidance for further process and device structure optimization of a-VMCO device.

6.1.3 Characterisation of Ta_2O_5 RRAM:

In Chapter 5 an overall picture of the resistive switching mechanism is provided for TiN/Ta_2O_5/TaO_x/TiN structured RRAM device. Bipolar and unipolar resistive switching modes have been observed co-existing in the same device under different operation conditions. For the bipolar mode, defects movement into/out of the CFR is responsible for the resistance switching, similar to that in HfO_2 RRAM. The location of CFR has been identified, which is located at the middle of the TCR, different to that near the BE in HfO_2 RRAM. This may be caused by the weaker scavenging capability of the TaO_x layer and lower oxygen affinity of Ta compared with Hf. For the unipolar mode observed at negative TE biases, it is observed that the effective voltage for set is much stronger than that for reset, and time for set is much shorter than reset. The unipolar switching mechanism, whether it is defect horizontal movement caused by thermal Joule heating, or defect energy alternation, and is still under investigation.

6.2 Future work and Outlook

6.2.1 Future work

Despite the efforts made in this thesis, there are still important issues that remain to be solved in the future. These include, but are not limited to, the followings:

Automatic RTN analysis technique

Automatic analysis of RTN, especially its time constant extraction, is always a difficult task, as RTN signal can be very complex in reality. It often has multiple levels combined with other noise and current fluctuation. The ideal 2-level RTN that can be easily analysed occurs sporadically. HMM can help separating RTN from other sources of noise, but cannot remove the current fluctuation, and it is also time-consuming to optimise the initial matrix for HMM calculation if the other sources of noise are significant. Automatic extraction without considering the RTN quality has the risk of producing unreliable time constants and defect information. In this thesis, all the extraction work is carried out manually to ensure accuracy at the cost of efficiency. On the other hand, since RTN is a stochastic process and induce high variability in scaled devices, to obtain a sufficiently large sample size, a large number of devices need to be measured, and a huge amount of RTN signal needs to be analysed. A novel technique is urgently in need to distinguish the RTN signal and to effectively separate RTN from the other sources of noise and fluctuation. To accomplish this task artificial intelligence may be employed for recognition and evaluation of the RTN signal before deciding how to extract it. A development of this automatic extraction technique will have the potential

to improve the efficiency of RTN analysis for large scale measurement and improve the extraction accuracy, paving the way towards a statistical understanding of RTN at the microscopic level, not only for RRAM but also for other scaled devices such as MOSFETs.

New technique for quantitative defect analysis in RRAM

Although RTN is capable of providing the location and energy information of a single defect in the oxide layer, and helping correlate the defect profile with resistance change in both filamentary and non-filamentary RRAM devices, it should be noted that RTN is not a perfect defect spectroscopy technique. It can only give an indication on trap distribution and generation. Only defects in the scanning region can be detected, which is limited by the narrow bias range applied to the device. In addition, only part of the defects in the scanning region can be detected because one RTN measurement sweep can be used to extract only a few defects at best. The range of time constants is limited by the resolution and the maximum measurement time, which further limits the measurement window. Moreover, RTN cannot be used to estimate the state density since it can only detect whether a defect has once appeared in a particular location, but cannot detect how many defects are in this location. Therefore, a novel defect spectroscopy technique is required to provide quantitative information such as defect density, defect energy distribution, and time constants in a certain spatial/energy range. This type of technique will be particularly important for the analysis of non-filament RRAM in which resistive switching is caused by the modulation of defects and defect density through the layers. A full defect spectrum in addition to those RTN-responsible defects could be depicted through the development of new characterisation technique to extend

the understanding of RRAM devices.

New technique for defect profile probing at LRS in RRAM

The RTN-based defect extraction technique is applicable only when electron tunnelling dominates. In filamentary RRAM, its application is restricted at HRS at which the CF ruptures. In this thesis, the defect profile at LRS can only be speculated from the so-called “weak LRS” where the device is slightly disturbed towards LRS and tunnelling still dominants. Physical characterisation methods can provide some information about the CF at LRS, but they are destructive, time consuming and statistical unfriendly. More importantly, they cannot provide defect information at microscopic level due to the resolution limitation. Therefore, a novel technique is needed, which should depict an overall picture of the defect profile, or the CF, at LRS. This picture could help clarify how defects move out of/into the constriction during set and reset in large numbers, how the defects interact with each other, and how the constriction is modified throughout the switching, and combined with the RTN technique for HRS, will improve the understanding of resistive switching in RRAMs.

Through developing the above mentioned new characterisation techniques in future work, we hope to provide not only the comprehensive information for defect profile modulation at device level, but also to link this information to the circuit operations for RRAM based novel applications such programmable switches and logic circuits, bio-inspired neuromorphic computing for machine learning and artificial intelligence.

6.2.2 Outlook

Memory

The last decade has seen significant progress in emerging NVM technologies (eNVMs). RRAM devices are promising candidates for the applications of both stand-alone and embedded NVM memories, thanks to its low programming voltage, fast speed, small size, non-volatility and 3D compatibility with conventional Si CMOS technology. Until now, the key industry players have all demonstrated Gb-scale capacity in advanced technology nodes, including 1 Gb PCM at 45 nm by Micron [239], 8 Gb PCM at 20 nm by Samsung [240], 32 Gb RRAM at 24 nm by Toshiba/Sandisk [241], 16 Gb conductive bridge (CBRAM, a special type of RRAM) at 27 nm by Micron/Sony [242], and most recently 128 Gb 3D XPoint technology by Micron/Intel [243]. These products are expected to bring revolutionary developments to the computer architecture by augmenting DRAM and Flash in the memory hierarchy, i.e. as the storage class memory (SCM) that has high capacity and economics similar to Flash and low-voltage fast performance similar to DRAM. To achieve this goal, the major remaining challenge is the large variability in device performance and reliability, which is a major barrier for using the RRAM in large memory arrays and multilevel operations, due to the substantial cycle-to-cycle and device-to-device variation of the device characteristics. A more complete understanding of the conduction and resistive switching mechanism are urgently needed to make further progress in this aspect. A combination of materials engineering, device structure optimization and innovations in addressing/readout/program/verification circuitry are key for providing viable solutions. The work in this thesis provides important technique and information in this aspect, and

laid a solid foundation for carrying out further work in order to overcome challenges in terms of cost-per-bit, latency, power, endurance, and retention and to take over the dominant mature market of DRAM or Flash. Moreover, to enable a wide adoption of eNVM technologies, a potentially viable path is to explore non-traditional applications or new paradigms beyond traditional memory applications, as in the programmable logic (FPGA) and neuromorphic computing discussed in detail below.

Programmable Logic

The use of RRAM devices in FPGA as both the interconnects and logic components has been proposed and shown great potential to drastically improve the density, speed and power consumption [73, 83, 244-248]. With the recent development of emerging NVM technologies, a number of novel FPGA architectures have been proposed, and eNVMs have been applied not only to replace the SRAMs in programmable interconnects, both locally and in switching matrix, but also the SRAMs in LBs, including:

- Replace configuration SRAMs with eNVMs: FPGA area is reduced since NVMs have a 5 to 25 times higher density than SRAMs and can be placed over CMOS transistors with the proposed 1T2R or 1T1R structures [244]. The non-volatility of eNVMs also saves the excessive leakage power during standby and avoids the configuration process.

- eNVMs used as programmable switches (HRS and LRS) in place of SRAM-based pass transistors to replace SRAM-based routing multiplexers [73, 245-247]. It can save the area of SRAMs and also the pass transistors that build routing switches. Different configurations, e.g. 2T1R, 2TG1R and 4T1R, have been proposed and up

to 50% improvements has been estimated in area reduction, performance improvement and power saving. Further improvements can be achieved with the architectural enhancements allowing larger multiplexers and less crossbar levels.

- Replace the LUTs with 3D RRAM/CMOS arrays that can not only save the data in nonvolatile memory but also perform NAND/NOR/Multiplication operations [83, 248]. The inherent dot-product capability of the crossbar structure can also be utilized to accelerate matrix multiplication, which is a key computational kernel in a wide range of applications including deep learning, optimization, etc. By augmenting RRAM crossbar design with various digital or analogue periphery circuits, these architectures can realize different accelerator functions that are built atop matrix multiplication and other arithmetic operations.

Different NVMs have their own advantages/shortcomings when used as programmable switches. For example, STT-MRAM usually has an ON/OFF ratio below 10. This ratio is sufficient for a memory application but far from enough for a routing switch. In contrast, one RRAM device can be fabricated within a $4F^2$ area and two metal layers. The cell area of a PCRAM can also be as small as that of an RRAM. However, the programming of PCRAMs relies on temperature and is hard to control. In addition, RRAMs provide more freedom in tuning their device properties than PCRAMs. Different fabrication technologies can be chosen to manufacture RRAMs with a very small write latency (<5 ns) or a very high endurance (>10^{12}) for the memory application. Other fabrication technologies can be chosen to manufacture RRAMs with a very high on/off ratio (~10^6) for the application of routing switches.

The main challenge of using eNVMs is the performance of write operations in terms of

latency, energy, endurance, and variability, although their read operations are competitive with SRAMs. As the programmable switches in FPGAs, one of the key challenges for the two-terminal eNVMs integrated in interconnects is the programmability and interference, as the two terminals are shared between the programming and signal paths. Novel designs in material, device, circuit and system levels are needed to overcome the challenges of area, speed, power, stability, variability and reliability requirements. Other challenges includes virtual memory support to ensure a united address space, memory/cache coherence, fault tolerance, security and privacy, thermal and power constraints, compatibility with modern programming models, etc. All of these will require collaborative efforts between technologies, IC designers and system engineers. Despite the research is still in its infant stage, it is envisaged that the eNVMs will become an enabling technology for revolutionary programmable logic development.

Machine Learning

Despite the great technological achievements in artificial intelligence systems [249], such as IBM Watson, Facebook DeepFace, and Google DeepMind AlphaGo, software-based neuromorphic computing currently faces several fundamental limitations, including scalability in power consumption, form factor, and cost, because such outstanding computing capabilities rely heavily on centralized, expensive data centres. The lack of ideal synaptic and neuron devices in the Si CMOS technology becomes the major obstacle. Neuromorphic computing with RRAM devices has a more efficient computing architecture that mimics biological neural networks, in which memory and logic coexist at the same physical location, overcoming the classical von

Neumann limitations.

Among various electronic synapse candidates, RRAM is one of the most promising [250-253]. Ultralow pJ energy consumption levels per synaptic operation as well as the possibility of realizing 3D crossbar array architecture have been demonstrated recently [250-252] to compute matrix–vector multiplication in parallel in learning algorithms such as convolutional NNs and deep Boltzmann machines. Furthermore, all intermediate values and weights in the network can be stored on chip in dense and compact arrays, which overcomes the input–output communication bottlenecks of large software-based NNs. Numerous neuromorphic applications such as pattern and audio recognition have been proposed [250-252]. Recently, a resistive-switching based neuron was also demonstrated in which the membrane potential is stored in the atomic configuration of a PCRAM device, thereby implementing the integrate-and-fire (IF) of an artificial neuron. An all-resistive-switching neuromorphic architecture using a single spiking neuron and an array of PCRAM synapses has been demonstrated performing unsupervised learning tasks [253].

The devices investigated in this thesis have shown great potential for synaptic/neuron applications. The filamentary RRAM could be used for offline learning in which the network is trained at the start with all the possible input patterns, while the non-filamentary RRAM with smooth set and reset such as a-VMCO are more attractive for online training for self-adaptive unsupervised learning where the non-filamentary RRAM acts as analogue synapses with continuous weight update.

Despite the progress, the characteristics of 3D synaptic and neuron arrays with RRAM devices have yet to be fully evaluated. Other critical challenge of RRAM-based synaptic

devices include the highly nonlinear characteristics of synaptic weight updates and significant change in weight distribution against time and stress, which impair the learning accuracy of RRAM based neural networks. The required bidirectional, linear, and symmetric response in conductance with large dynamic range for RRAM synaptic devices have not been fully demonstrated. The existing works are all based on simplified neuro-synaptic implementation and modelling, which cannot fully represent the inherent variability of RRAM cell. Further work is needed to address the algorithmic/hardware implications of the variability, stochasticity and storage resolution of the RRAM neurons and synapses in practical neural network configurations. Open issues related to the interconnectivity and the integration of the RRAMs in a neuromorphic processor chip also remain to be addressed. A full crossbar-array implementation, including dedicated neuron circuitry for the summation of synaptic weights during both forward propagation and backpropagation through nearly identical high-performance nonlinear selector devices has yet been demonstrated. Further research in novel neural network configurations and evaluations, combined with device and material innovation and characterisation, is much needed, therefore.

Summary

Today's RRAM is in a position similar to the early steam engines in the 18th century which when invented initially was used only for pumping water in the coal mines but soon became the dominant source of power and changed the entire world for more than 200 years. Similarly, RRAM is initially expected to replace the current memory system, but the unique physical properties of RRAM make it possible to re-consider the design of the computer system for revolutionary performance improvement. It has to be

admitted that it may be challenging for RRAM technology to change the current memory system at this stage, not to mention the computer architecture, but the current technology will finally come to its end and there will be strong motivation looking for replacement technologies, and RRAM will surely be among the strongest candidates.

Reference

[1] J. von Neuman, “First draft of a report on the EDVAC,” Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, 1945.

[2] R. Bez, E. Camerlenghi, A. Modelli, and A. Visconti, “Introduction to Flash Memory,” *Proc. IEEE,* vol. 91, no. 4, p. 489–502, 2003.

[3] J. Singh, S. P. Mohanty, D. K. Pradhan, “Introduction to SRAM,” in *Robust SRAM Designs and Analysis*, New York, Springer New York, 2012, pp. 1-29.

[4] R. H. Dennard, “Past Progress and Future Challenges in LSI Technology: From DRAM and Scaling to Ultra-Low-Power CMOS,” *IEEE Solid State Circuits Mag.,* vol. 7, no. 2, pp. 29-38, 2015.

[5] D. Patterson and J. Hennessy, Computer Organization and Design: The Hardware/Software Interface, Elsevier, 1971.

[6] F. Masuoka, M. Momodomi, Y. Iwata, and R. Shirota, “New ultra high density EPROM and flash EEPROM with NAND structure cell,” in *IEDM Tech. Dig.*, 1987.

[7] T. Mikolajick, M. Salinga, M. Kund, and T. Kever, “Nonvolatile Memory Concepts Based on Resistive Switching in Inorganic Materials,” *Adv. Eng. Mater.,* vol. 11, no. 4, pp. 235 - 240, 2009.

[8] S. H. Jo, “Recent Progress in RRAM: Materials and Devices,” SEMICON2015, 2015.

[9] S. Lai, “Non-volatile memory technologies: The quest for ever lower cost,” in *IEDM Tech. Dig.*, 2008.

[10] C. Y. Lu, “Future prospects of nand flash memory technology—the evolution from floating gate to charge trapping to 3D stacking,” *Journal of nanoscience and nanotechnology,* vol. 12, no. 10, p. 7604–7618, 2012.

[11] A. Goda and K. Parat, “Scaling directions for 2D and 3D nand cells,” in *IEDM Tech. Dig.*, 2012.

[12] Orbis Research, “Global Next-Generation Memory Market Research Report- Forecast 2022,” 2017.

[13] L. Zhang, Study of the Selector Element for Resistive Memory, KULeuven: Ph.D. dissertation, 2015.

[14] W. Gallagher and S. Parkin, “Development of the magnetic tunnel junction,” *IBM Journal of Research and Development,* vol. 50, pp. 5-23, 2006.

[15] M. Hosomi, H. Yamagishi, T. Yamamoto, K. Bessho, Y. Higo, K. Yamane,H. Yamada, M. Shoji, H. Hachino, C. Fukumoto, H. Nagao, H. Kano, “A novel nonvolatile memory with spin torque transfer magnetization switching: spin-ram,” in *IEDM Tech. Dig.*, Washington, 2005.

[16] S. Raoux, G. Burr, M. Breitwisch, C. Rettner, Y. Chen, R. Shelby, M. Salinga,D. Krebs, S.H. Chen, H. Lung, and C. Lam, “Phase-change random access memory: A scalable technology,” *IBM Journal of Research and Development,* vol. 52, p. 465–479, 2008.

[17] R. Bez, “Chalcogenide PCM: a memory technology for next decade,” in *IEDM Tech. Dig.*, 2009.

[18] A. Beck, J. Bednorz, C. Gerber, C. Rossel, and D. Widmer, “Reproducible switching effect in thin oxide films for memory applications,” *Appl. Phys.,* vol. 77, no. 1, p. 139–141, 2000.

[19] B. Govoreanu, G. Kar, Y. Chen, V. Paraschiv, S. Kubicek, A. Fantini, I. Radu, L. Goux, S. Clima, and R. Degraeve, “$10\times10nm^2$ Hf/Hfox crossbar resistive ram with excellent performance, reliability and low-energy operation,” in *IEDM Tech. Dig.*, 2011.

[20] M. N. Kozicki, M. Park, and M. Mitkova, “Nanoscale memory elements based on solid-state electrolytes,” *IEEE Trans. Nanotechnol.,* vol. 4, no. 3, pp. 331-338, 2005.

[21] A. Sawa, “Resistive switching in transition metal oxides,” *Materials today,* vol. 11, no. 6, pp. 28-36, 2008.

[22] J. Akerman, “Toward a Universal Memory,” *Science,* vol. 308, no. 5721, pp. 508-510, 2005.

[23] H. S. P. Wong, H. Y. Lee, S. Yu, Y. S. Chen, P. S. Chen, B. Lee, F. T. Chen, and M. J. Tsai, “Metal–Oxide RRAM,” *Proc. IEEE,* vol. 100, no. 6, pp. 1951-1970, 2012.

[24] B. Govoreanu, A. Redolfi, L. Zhang, C. Adelmann, M. Popovici, S. Clima, H. Hody, V. Paraschiv, I. P. Radu, A. Franquet, J. C. Liu, J. Swerts, O. Richard, H. Bender, L. Altimime, and M. Jurczak, “Vacancy-modulated conductive oxide resistive ram (VMCO-RRAM): An area-scalable switching current, self-compliant, highly nonlinear and wide on/off-window resistive switching cell,” in *IEDM Tech. Dig.*, 2013.

[25] D. S. Jeong, R. Thomas, R. Katiyar, J. Scott, H. Kohlstedt, A. Petraru, and C. S. Hwang, "Emerging memories: resistive switching mechanisms and current status," *Reports on Progress in Physics*, vol. 75, no. 7, p. 076502, 2012.

[26] F. T. Chen, H. Y. Lee, Y. S. Chen, Y. Y. Hsu, L. J. Zhang, P. S. Chen, W. S. Chen, P. Y. Gu, W. H. Liu, S. M. Wang, C. H. Tsai, S. S. Sheu, M. J. Tsai, and R. Huang, "Resistance switching for RRAM applications," *Sci. China. Technol. Sc.*, vol. 54, no. 5, p. 1073–1086, 2011.

[27] H. P. Wong, S. Raoux, S. Kim, J. Liang, J. P. Reifenberg, B. Rajendran, M. Asheghi, and K. E. Goodson, "Phase change memory," *Proc. IEEE*, vol. 98, no. 12, p. 2201–2227, 2010.

[28] A. D. Kent and D. C. Worledge, "A new spin on magnetic memories," *Nat. Nanotechnol.*, vol. 10, no. 3, p. 187–191, 2015.

[29] Y. Chen, "Electrical characterisation and functional oxide for resistive random access memory (RRAM) application," KULeuven, Ph.D. dissertation, 2013.

[30] S. Yu and P. Y. Chen, "Emerging memory technologies," *IEEE Solid State Circuits Mag.*, vol. 8, no. 2, pp. 43-56, 2016.

[31] T. Nirschl, J. B. Phipp, T. D. Happ, G. W. Burr, B. Rajendran, M. H. Lee, A. Schrott, M. Yang, M. Breitwisch, C. F. Chen, E. Joseph, M. Lamorey, R. Cheek, S. H. Chen, S. Zaidi, S. Raoux, Y. C. Chen, Y. Zhu, R. Bergmann, H. L. Lung, C. Lam, "Write strategies for 2 and 4-bit multi-level phase-change memory," in *IEDM Tech. Dig.*, 2007.

[32] B. Lee, E. Ipek, O. Mutlu, and D. Burger, in *ISCA*, 2009.

[33] B. C. Lee, P. Zhou, J. Yang, Y. Zhang, B. Zhao, E. Ipek, O. Mutlu and D. Burger, "Phase-Change Technology and the Future of Main Memory," *IEEE Micro.*, vol. 30, no. 1, pp. 131-141, 2010.

[34] T. W. Hickmott, "Low-frequency negative resistance in thin anodic oxide films," *J. Appl. Phys.*, vol. 33, no. 9, p. 2669, 1962.

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

[36] G. Dearnale, A. M. Stoneham, and D. V. Morgan, "Electrical phenomena in amorphous oxide films," *Rep. Progr. Phys.*, vol. 33, no. 3, pp. 1129-1191, 1970.

[37] J. G. Simmons, "Conduction in thin dielectric films," *J. Phys. D, Appl. Phys.*, vol. 4, no. 5, p. 613, 1971.

[38] Y. Watanabe, J. G. Bednorz, A. Bietsch, Ch. Gerber, D. Widmer, A. Beck, and S. J. Wind, "Current-driven insulator-conductor transition and nonvolatile memory in chromium-doped SrTiO3 single crystals," *Appl. Phys. Lett.*, vol. 78, no. 23, p. 3738–3740, 2001.

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

[40] C. Rohde, B. J. Choi, D. S. Jeong, S. Choi, J. S. Zhao, and C. S. Hwang, "Identification of a determining parameter for resistive switching of TiO_2 thin films," *Appl. Phys. Lett.*, vol. 86, no. 26, p. 262907, 2005.

[41] A. Baikalov, Y. Wang, B. Shen, B. Lorenz, S. Tsui, Y. Sun, Y. Xue, and C. Chu,, "Field-driven hysteretic and reversible resistive switch at the Ag-Pr0.7Ca0.3MnO_3 interface," *arXiv:cond-mat/0212464*, 2002.

[42] B. Govoreanu, D. Crotti, S. Subhechha, L. Zhang, Y. Y. Chen, S. Clima, V. Paraschiv, H. Hody, C. Adelmann, M. Popovici, O. Richard, and M. Jurczak, "A-VMCO: A novel forming-free, self-rectifying, analog memory cell with low-current operation, nonfilamentary switching and excellent variability," in *VLSI Symp. Tech. Dig.*, 2015.

[43] B. Govoreanu, L. D. Piazza, J. Ma, T. Conard, A. Vanleenhove, A. Belmonte, D. Radisic, M. Popovici, A. Velea, A. Redolfi, O. Richard, S. Clima, C. Adelmann, H. Bender, and M. Jurczak, "Advanced a-VMCO resistive switching memory through inner interface engineering with wide (> 10^2) on/off window, tunable μA-range switching current and excellent variability," in *VLSI Symp. Tech. Dig.*, 2016.

[44] J. Zhou, F. Cai, Q. Wang, B. Chen, S. Gaba, and W. D. Lu, "Very Low-Programming-Current RRAM With Self-Rectifying Characteristics," *IEEE Electron Device Lett.*, vol. 37, no. 4, pp. 404-407, 2016.

[45] C. W. Hsu, I. T. Wang, C. L. Lo, M. C. Chiang, W. Y. Jang, C. H. Lin, T. H. Hou, "Self-Rectifying Bipolar TaOx/TiO_2 RRAM with Superior Endurance over 10^{12} Cycles for 3D High-Density Storage-Class Memory," in *VLSI Symp. Tech. Dig.*, 2013.

[46] Y. Y. Chen, M. Komura, R. Degraeve, B. Govoreanu, L. Goux, A. Fantini, N. Raghavan, S. Clima, L. Zhang, A. Belmonte, A. Redolfi, G. S. Kar, G. Groeseneken, D. J. Wouters, and M. Jurczak, "Improvement of data retention in HfO_2/Hf 1T1R RRAM cell under low operating current," in *IEDM Tech. Dig.*, 2013.

[47] Y. S. Chen, T. Y. Wu, P. J. Tzeng, P. S. Chen, H. Y. Lee, C. H. Lin, F. Chen, and M. J. Tsai, "Forming-free HfO_2 bipolar RRAM device with improved endurance and high speed operation," in *VLSI Technology, Systems, and Applications, 2009. VLSI-TSA '09. International Symposium on*, 2009.

[48] S. R. Lee, Y. B. Kim, M. Chang, K. M. Kim, C. B. Lee, J. H. Hur, G. S. Park, D. Lee, M. J. Lee, C. J. Kim, U. I. Chung, I. K. Yoo and K. Kim, "Multi-level Switching of Triple-layered TaOx RRAM with Excellent Reliability for Storage Class Memory," in *VLSI Symp. Tech. Dig.*, 2012.

[49] T. Y. Liu, T. H. Yan, R. Scheuerlein, Y. Chen, J. K. Lee, G. Balakrishnan, G. Yee, H. Zhang, A. Yap, J. Ouyang, T. Sasaki, A. Al-Shamma, C. Chen, M. Gupta, G. Hilton, A. Kathuria, V. Lai, M. Matsumoto, A. Nigam, A. Pai, J. Pakhale, C. H. Siau, et al, "A 130.7-mm 2-Layer 32-Gb ReRAM Memory Device in 24-nm Technology," *IEEE J. Solid-State Circuits*, vol. 49, no. 1, pp. 140 - 153, 2014.

[50] N. Xu, L. F. Liu, X. Sun, X. Y. Liu, D. D. Han, Y. Wang, R. Q. Han, J. F. Kang, and B. Yu, "Characteristics and mechanism of conduction/set process in TiN/ZnO/Pt resistance switching random-access memories," *Appl. Phys. Lett.*, vol. 92, p. 232112, 2008.

[51] G. H. Buh, I. Hwang, and B. H. Park, "Time-dependent electroforming in NiO resistive switching devices," *Appl. Phys. Lett.*, vol. 95, p. 142101, 2009.

[52] M. Janousch, G. I. Meijer, U. Staub, B. Delley, S. F. Karg, and B. P. Andreasson, "Role of oxygen vacancies in Cr-doped $SrTiO_3$ for resistance-change memory," *Adv.Mater.*, vol. 19, p. 2232–2235, 2007.

[53] G. S. Park, X. S. Li, D. C. Kim, R. J. Jung, M. J. Lee, and S. Seo, "Observation of electric-field induced Ni filament channels in polycrystalline NiO_x film," *Appl. Phys. Lett.*, vol. 91, p. 222103, 2007.

[54] J. Y. Son and Y. H. Shin, "Direct observation of conducting filaments on resistive switching of NiO thin films," *Appl. Phys. Lett.*, vol. 92, p. 222106, 2008.

[55] K. Szot, W. Speier, G. Bihlmayer, and R. Waser, "Switching the electrical resistance of individual dislocations in single-crystalline $SrTiO_3$," *Nature Mater.*, vol. 5, pp. 312-320, 2006.

[56] D. Lee, D. J. Seong, I. Jo, F. Xiang, R. Dong, S. Oh, and H. Hwang, "Resistance switching of copper doped MoO_x films for nonvolatile memory applications," *Appl. Phys. Lett.*, vol. 90, p. 122104, 2007.

[57] B. J. Choi, D. S. Jeong, S. K. Kim, C. Rohde, S. Choi, J. H. Oh, H. J. Kim, C. S. Hwang, K. Szot, R. Waser, B. Reichenberg, and S. Tiedke, "Resistive switching mechanism of TiO_2 thin films grown by atomic-layer deposition," *J. Appl. Phys.*, vol. 98, no. 3, p. 033715, 2005.

[58] H. Y. Lee, P. S. Chen, T. Y. Wu, Y. S. Chen, C. C. Wang, P. J. Tzeng, C. H. Lin, F. Chen, C. H. Lien, and M. J. Tsai, "Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO_2 based RRAM," in *IEDM Tech. Dig.*, 2008.

[59] I. H. Inoue, S. Yasuda, H. Akinaga, and H. Takagi, "Nonpolar resistance switching of metal/binary-transition-metal oxides/metal sandwiches: Homogeneous/ inhomogeneous transition of current distribution," *Phys. Rev. B*, vol. 77, p. 035105, 2008.

[60] C. Walczyk, C. Wenger, R. Sohal, M. Lukosius, A. Fox, J. Dabrowski, D. Wolansky, B. Tillack, H. J. Mussig, and T. Schroeder,, "Pulse-induced low-power resistive switching in HfO_2 metal-insulator-metal diodes for nonvolatile memory applications," *J. Appl. Phys.*, vol. 105, p. 114103, 2009.

[61] W. Wang, S. Fujita, and S. S. Wong, "Elimination of forming process for TiOx nonvolatile memory devices," *IEEE Electron Device Lett.*, vol. 30, no. 7, p. 763–765, 2009.

[62] W. Y. Chang, Y. T. Ho, T. C. Hsu, F. Chen, M. J. Tsai, and T. B. Wu, "Influence of crystalline constituent on resistive switching properties of TiO2 memory films," *Electrochem. Solid State Lett.*, vol. 12, pp. H135-H137, 2009.

[63] X. Cao, X. M. Li, X. D. Gao, W. D. Yu, X. J. Liu, Y. W. Zhang, L. D. Chen, and X. H. Cheng, "Forming-free colossal resistive switching effect in rare-earth-oxide Gd_2O_3 films for memristor applications," *J. Appl. Phys.*, vol. 106, p. 073723, 2009.

[64] L. Goux, J. G. Lisoni, X. P.Wang, M. Jurczak, and D. J. Wouters, "Optimized Ni oxidation in 80-nm contact holes for integration of forming-free and low-power Ni/NiO/Ni memory cells,"

IEEE Trans. Electron Devices, vol. 56, no. 10, p. 2363–2368, 2009.

[65] U. Russo, D. Ielmini, C. Cagli, A. L. Lacaita, S. Spiga, C. Wiemer, M. Perego, and M. Fanciulli, "Conductive-filament switching analysis and self-accelerated thermal dissolution model for reset in NiO-based RRAM," in *Proc. IEDM*, 2007.

[66] B. Gao, S. Yu, N. Xu, L. F. Liu, B. Sun, X. Y. Liu, R. Q. Han, J. F. Kang, B. Yu, and Y. Y. Wang, "Oxide-based RRAM switching mechanism: A new ion-transport-recombination model," in *IEDM. Tech. Dig.*, 2008.

[67] C. W. Hsu, Y. F. Wang, C. C. Wan, I. T. Wang, C. T. Chou, W. L. Lai, Y. J. Lee, T. H. Hou, "Homogeneous barrier modulation of TaO_x/TiO_2 bilayers for ultra-high endurance three-dimensional storage-class memory," *Nanotechnology,* vol. 25, no. 16, p. 165202, 2014.

[68] L. L. Wei, D. S. Shang, J. R. Sun, S. B. Lee, Z. G. Sun, B. G. Shen, "Non-filamentary memristive switching in Pt/CuOx/Si/Pt systems," *Nanotechnology,* vol. 24, p. 325202, 2013.

[69] Y. F. Wang, Y. C. Lin, I. T. Wang, T. P. Lin and T. H. Hou, "Characterisation and Modeling of Nonfilamentary Ta/TaOx/TiO_2/Ti Analog Synaptic Device," *Scientific Reports,* vol. 5, p. 10150, 2015.

[70] L. O. Chua, "Memristor-The missing circuit element," *IEEE Trans. Circuit Theory,* vol. 18, no. 5, pp. 507-519, 1971.

[71] D. Struckov, G. Snider, D. Stewart, and R. Williams, "The missing memristor found," *Nature,* vol. 453, no. 7191, pp. 80-83, 2008.

[72] Grand New Research, "Global FPGA Market Size To Reach USD 14.2 Billion By 2024" Retrieved on 1st May 2017 from *http://www.grandviewresearch.com/press-release/global-fpga-market.*

[73] J. Cong and B. Xiao, "FPGA-RPI: A novel FPGA architecture with RRAM-based programmable interconnects," *IEEE Trans. VLSI,* vol. 22, no. 4, pp. 864-877, 2014.

[74] I. Kuon and J. Rose, "Measuring the gap between FPGAs and ASICs," *IEEE Trans. Comput. Aided Design Integr. Circuits Syst.,* vol. 26, no. 2, p. 203–215, 2007.

[75] International Technology Roadmap for Semiconductors. (2010), Retrieved on 1st May 2017 from *http://www.itrs.net/*.

[76] J. Cong and B. Xiao, "mrFPGA: a novel FPGA architecture with memristor-based reconfiguration," in *IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH),* 1-8, 2011.

[77] Y. Chen, J. Zhao, and Y. Xie, "3D-NonFAR: three-dimensional non-volatile FPGA architecture using phase change memory," in *2010 ACM/IEEE International Symposium on Low-Power Electronics and Design, ISLPED,* 55–60, 2010.

[78] S. Tanachutiwat, M. Liu, and W. Wang, "FPGA based on integration of CMOS and RRAM," in *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, 2011.

[79] D. Lewis, E. Ahmed, D. Cashman, T. Vanderhoek, C. Lane, A. Lee, and P. Pan, "Architectural enhancements in stratix-IIITM and stratix-IVTM," in *Proceedings of the ACM/SIGDA International Symposium on Field Programmable Gate Arrays*, 2009.

[80] P. E. Gaillardon, M. H. Ben-jamaa, G. B. Beneventi, F. Clermidy, and L. Pemiola, "Emerging memory technologies for reconfigurable routing in FPGA architecture," in *17th IEEE International Conference on Electronics, Circuits, and Systems (ICECS),* 2010.

[81] S. Onkaraiah, P. E. Gaillardon, M. Reyboz, F. Clermidy, J. M. Portal, M. Bocquet, C. Muller, "Using O_xRRAM memories for improving communications of reconfigurable FPGA architectures," in *IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH),* 2011.

[82] M. Lee, C. B. Lee, D. Lee, S. R. Lee, M. Chang, J. H. Hur, Y. Kim, C. Kim, D. Seo, S. Seo, U. Chung, I. Yoo, and K. Kim, "A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta_2O_{5-x}/TaO_{2-x} bilayer structures," *Nature Mater.,* vol. 10, pp. 625-630, 2011.

[83] H. Li, K. S. Li, C. H. Lin, J. L. Hsu, W. C. Chiu, M. C. Chen, T. T. Wu, J. Sohn, S. B. Eryilmaz, J. M. Shieh, W. K. Yeh, and H. S. P. Wong, "Four-layer 3D vertical RRAM integrated with FinFET as a versatile computing unit for brain-inspired cognitive information processing," VLSI Symp. Tech. Dig., 2016.

[84] R. Ananthanarayanan, S. K. Esser, H. D. Simon, and D. S. Modha, "The cat is out of the bag: cortical simulations with 10^9 synapses," in *Proceedings of 2009 IEEE/ACM Conference High*

Performance Networking Computing, 2009.

[85] L. S. Smith, "Handbook of Nature-Inspired and Innovative Computing: Integrating Classical Models with Emerging Technologies," Springer, 2006, pp. 433-475.

[86] E. M. Izhikevich and G. M. Edelman, "Large-scale model of mammalian thalamocortical systems," *Proc. Natl. Acad. Sci.,* vol. 105, no. 9, p. 3593–3598, 2008.

[87] G. Indiveri, E. Chicca, and R. Douglas, "A VLSI Array of Low-Power Spiking Neurons and Bistable Synapses With Spike-Timing Dependent Plasticity," *IEEE Trans. Neural Netw.,* vol. 17, no. 1, pp. 211-221, 2006.

[88] G. J. Siegel, B. W. Agranoff, and R. W. Albers, Basic Neurochemistry: Molecular, Cellular and Medical Aspects, Philadelphia, PA: Lippincott-Raven, 1999.

[89] P. Lennie, "The cost of cortical computation," *Curr. Biol.,* vol. 13, no. 6, p. 493–497, 2003.

[90] L. F. Abbott and S. B. Nelson, "Synaptic plasticity: Taming the beast," *Nat. Neurosci.,* vol. 3, no. 11, p. 1178–1183, 2000.

[91] E. R. Kandel, "Physiology or Medicine," Nobel Lecture, 2000.

[92] C. Mead, "Neuromorphic electronic systems," *Proc. IEEE,* vol. 78, no. 10, p. 1629–1636, 1990.

[93] G. Indiveri and T. K. Horiuchi, "Frontiers in neuromorphic engineering," *Frontiers in Neuroscience,* vol. 5, pp. 1-2, 2011.

[94] A. K. Friesz, A. C. Parker, C. Zhou, K. Ryu, and J. M. Sanders, "A biomimetic carbon nanotube synapse circuit," in *Proc. BMES Annu. Fall Meeting*, 2007.

[95] H. Choi, H. Jung, J. Lee, J. Yoon, J. Park, D. J. Seong, W. Lee, M. Hasan, G. Y. Jung, and H. Hwang, "An electrically modifiable synapse array of resistive switching memory," *Nanotechnology,* vol. 20, no. 34, pp. 345 201-1–345 201-5, 2009.

[96] S. D. Ha and S. Ramanathan, "Adaptive oxide electronics: A review," *J. Appl. Phys.,* vol. 110, no. 7, pp. 071101-1–071101-20, 2011.

[97] S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, "Nanoscale memristor device as synapse in neuromorphic systems," *Nano Lett.,* vol. 10, no. 4, p. 1297–1301, 2010.

[98] T. Hasegawa, T. Ohno, K. Terabe, T. Tsuruoka, T. Nakayama, J. K. Gimzewski, and M. Aono, "Learning abilities achieved by a single solid-state atomic switch," *Adv. Mater.,* vol. 22, no. 16, pp. 1831-1834, 2010.

[99] Defense Advanced Research Project Agency (DARPA), "Systems of Neuromorphic Adaptive Plastic Scalable Electronics-SyNAPSE," Retrieved on 1st May 2017 from *http://www.darpa.mil/program/systems-of-neuromorphic-adaptive-plastic-scalable-electronics*

[100] J. R. Cary, "Framework Application for Core-Edge Transport Simulations Project-FACETS," PSACI, PPPL, 8 Jun 07

[101] "A universal spiking neural network architecture (spinnaker)". Retrieved in Oct, 2011 from *http://apt.cs.man.ac.uk/projects/ SpiNNaker/*

[102] T. Tuma, A. Pantazi, M. L. Gallo, A. Sebastian and E. Eleftheriou, "Stochastic phase-change neurons," *Nat. Nanotechnol.,* vol. 11, p. 693–699, 2016.

[103] G. Indiveri, B. Linares-Barranco, R. Legenstein, G. Deligeorgis, and T. Prodromakis, "Integration of nanoscale memristor synapses in neuromorphic computing architectures," *Nanotechnology,* vol. 24, p. 384 010, 2013.

[104] R. Waser, R. Dittmann, G. Staikov, and K. Szot, "Redox-based resistive switching memories - nanoionic mechanisms , prospects, and challenges," *Adv. Mater.,* Vols. 21 (25-26), p. 2632–2663, 2009.

[105] L. Chua and S. Kang, "Memristive devices and systems," *Proc. IEEE,* vol. 64, no. 2, p. 209–223, 1976.

[106] B. Gao, Y. Bi, H. Y. Chen, R. Liu, P. Huang, B. Chen, L. Liu, X. Liu, S. Yu, H. S. P. Wong, and J. Kang, "Ultra-low-energy three-dimensional oxide-based electronic synapses for implementation of robust high-accuracy neuromorphic computation systems," *ACS Nano,* vol. 8, no. 7, p. 6998–7004, 2014.

[107] M. Prezioso, F. Merrikh-Bayat, B. D. Hoskins, G. C. Adam, K. K. Likharev, and D. B. Strukov, "Training and operation of an integrated neuromorphic network based on metal-oxide memristors," *Nat. Lett.,* vol. 521, pp. 61-64, 2015.

[108] O. Turel and K. Likharev, "Crossnets: possible neuromorphic networks based on nano-scale components," *Int. J. Circuit Theory Appl.,* vol. 31, no. 1, pp. 37-53, 2003.

[109] H. Markram, "The blue brain project," in *ACM/IEEE Conference on Supercomputing*, 2006.

[110] G. Q. Bi, and M. M. Poo, "Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type," *J. Neurosci.,* vol. 24, no. 10 464–10 472, p. 18, 1998.

[111] G. Indiveri, E. Linn, and A. Ambrogio, "ReRAM-Based Neuromorphic Computing," in *Resistive Switching: From Fundamentals of Nanoionic Redox Processes to Memristive Device Applications*, Wiley-VCH Verlag GmbH & Co. KGaA, 2016, pp. 715-735.

[112] G. Snider, "Spike-timing-dependent learning in memristive nanodevices," *NANOARCH,* p. 85–92, 2008.

[113] B. Rajendran,Y. Liu, J. Seo, K. Gopalakrishnan, L. Chang, D. J. Friedman and M. B. Ritter, "Specifications of Nanoscale Devices and Circuits for Neuromorphic Computational Systems," *IEEE Trans. Electron Devices,* vol. 60, no. 1, pp. 246-253, 2013.

[114] S. Park, J. Noh, M. Choo, A. M. Sheri, M. Chang, Y. B. Kim, C. J. Kim, M. Jeon, B. G. Lee, and H. Hwang, "Nanoscale RRAM-based synaptic electronics: toward a neuromorphic computing device," *Nanotechnology,* vol. 24, no. 38, pp. 384009-384015, 2013.

[115] H. Li, P. Huang, B. Gao, X. Liu, J. Kang, and H. S. P. Wong, "Device and Circuit Interaction Analysis of Stochastic Behaviors in Cross-Point RRAM Arrays," *arXiv preprint arXiv:1606.07457.*

[116] S. Yu, X. Guan, and H. S. P. Wong, "On the stochastic nature of resistive switching in metal oxide RRAM: Physical modeling, monte carlo simulation, and experimental characterisation," *Proc IEDM,* p. 17.3.1–17.3.4, 2011.

[117] A. Chen, M. R. Lin, "Variability of Resistive Switching Memories and its Impact on Crossbar Array Performance," *IEEE IPRS,* p. pp.10.1–10.4, 2011.

[118] K. D. Suh, B. H. Suh, Y. H. Lim, J. K. Kim, Y. J. Choi, Y. N. Koh, S. S. Lee, S. C. Kwon, B. S. Choi, J. S. Yum, J. H. Choi, J. R. Kim, and H. K. Lim, "A 3.3V 32Mb NAND Flash memory with incremental step pulse programming scheme," in *ISSCC*, 1995.

[119] P. Pouyan, E. Amat, S. Hamdioui and A. Rubio, "RRAM variability and its mitigation schemes," in *Power and Timing Modeling, Optimization and Simulation (PATMOS), 2016 26th International Workshop on*, 2016.

[120] D. Ielmini, "Resistive switching memories based on metal oxides: mechanisms, reliability and scaling," *Semicond. Sci. Technol.,* vol. 31, p. 063002, 2016.

[121] S. Yu, Resistive Random Access Memory (RRAM), Morgan & Claypool Publishers, 2016.

[122] Y. Y. Chen, L. Goux, S. Clima, B. Govoreanu, R. Degraeve, G.S. Kar, A. Fantini, G. Groeseneken, D. J. Wouters, and M. Jurczak, "Endurance/Retention Trade-off on HfO_2/Metal Cap 1T1R Bipolar RRAM," *IEEE Trans. Electron Devices,* vol. 60, no. 3, pp. 1114-1121, 2013.

[123] Y. Y. Chen, R. Degraeve, B. Govoreanu, S. Clima, L. Goux, A. Fantini, G.S. Kar, D.J. Wouters, G. Groeseneken, and M. Jurczak, "Postcycling LRS Retention Analysis in HfO_2/Hf RRAM 1T1R Device," *IEEE Electron Device Lett.,* vol. 34, no. 5, pp. 626-628, 2013.

[124] J. Song, D. Lee, J. Woo, E. Cha, S. Lee, J. Park, K. Moon, Y. Koo, A. Prakash and H. Hwang, "Improvement in reliability characteristics (retention and endurance) of RRAM by using high-pressure hydrogen annealing," in *Silicon Nanoelectronics Workshop (SNW), 2014 IEEE*, 2014.

[125] S. Yu, Y. Wu, and H. S. P. Wong, "Investigating the switching dynamics and multilevel capability of bipolar metal oxide resistive switching memory," *Appl. Phys. Lett.,* vol. 98, p. 103514, 2011.

[126] L. Zhao, H. Y. Chen, S. C. Wu, Z. Jiang, S. Yu, T. H. Hou, H. S. P. Wong and Y. Nishi, "Improved multi-level control of RRAM using pulse-train programming," in *VLSI Technology, Systems and Application (VLSI-TSA), Proceedings of Technical Program - 2014 International Symposium on*, 2014.

[127] D. Ielmini, C. Cagli, F. Nardi and Y. Zhang, "Nanowire-based resistive switching memories: devices, operation and scaling," *J. Phys. D: Appl. Phys.*, vol. 46, no. 7, p. 074006, 2013.

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

[129] Y. S. Fan, L. Zhang. D. Crotti, T. Witters, M. Jurczak, and B. Govoreanu, "Direct Evidence of the Overshoot Suppression in Ta_2O_5-Based Resistive Switching Memory With an Integrated Access Resistor," *IEEE Electron Device Lett.,* vol. 36, no. 10, pp. 1027-1029, 2015.

[130] S. H. Jo, T. Kumar, S. Narayanan, and H. Nazarian, "Cross-Point Resistive RAM Based on Field-Assisted Superlinear Threshold Selector," *IEEE Trans. Electron Devices,* vol. 62, no. 11, pp. 3477-3481, 2015.

[131] Keysight, "B1530A (B1500A-A30, B1500AU-030) Waveform Generator/Fast Measurement Unit (WGFMU)," Retrieved on 1st May 2017 from *http://www.keysight.com/en/pd-1443698-pn-B1500A-A30re/waveform-generator-fast-measurement-unit-wgfmu-module-for-the-b1500a?nid=-33019.793465&cc=GB&lc=eng*.

[132] Keysight Technologies, "B1525A (B1500A-A25, B1500AU-025) High Voltage Semiconductor Pulse Generator Unit (HV-SPGU)," Retrieved on 1st May 2017 from *http://www.keysight.com/en/pd-2253453-pn-B1500A-A25/high-voltage-semiconductor-pulse-generator-unit-hv-spgu?nid=-33019.1048180&cc=GB&lc=eng.*

[133] M. J. Rozenberg, M. J. Sánchez, R. Weht, C. Acha, F. Gomez-Marlasca, and P. Levy, "Mechanism for bipolar resistive switching in transition-metal oxides," *Phys. Rev. B,* vol. 81, no. 11-15, p. 115101, 2010.

[134] R. Waser and M. Aono, "Nanoionics-based resistive switching memories.," *Nat. Mater.,* vol. 6, p. 833–840, 2007.

[135] U. Celano, L. Goux, A. Belmonte, K. Opsomer, A. Franquet, A. Schulze, C. Detavernier, O. Richard, H. Bender, M. Jurczak, and W. Vandervorst, "Three-dimensional observation of the conductive filament in nanoscaled resistive memory devices," *Nano Lett.,* vol. 14, no. 5, p. 2401−2406, 2014.

[136] U. Celano, L. Goux, R. Degraeve, A. Fantini, O. Richard, H. Bender, M. Jurczak and W. Vandervorst, "Imaging the Three-Dimensional Conductive Channel in Filamentary-Based Oxide Resistive Switching Memory," *Nano Lett.,* vol. 15, no. 12, p. 7970−7975, 2015.

[137] F. Miao, J. P. Strachan, J. J. Yang, M. X. Zhang, I. Goldfarb, A. C. Torrezan, P. Eschbach, R. D. Kelley, G. M. Ribeiro and R. S. Williams, "Anatomy of a Nanoscale Conduction Channel Reveals the Mechanism of a High-Performance Memristor," *Adv. Mater.,* vol. 23, no. 47, p. 5633–5640, 2011.

[138] I. Valov, I. Sapezanskaia, A. Nayak, T. Tsuruoka, and T. Bredow, "Atomically Controlled Electrochemical Nucleation at Superionic Solid Electrolyte Surfaces.," *Nat. Mater.,* vol. 11, pp. 530-535, 2012.

[139] M. Moors, K. K. Adepalli, Q. Lu, A. Wedig, C. Baumer, K. Skaja, B. Arndt, H. L. Tuller, R. Dittmann, R. Waser, B. Yildiz, and I. Valov, "Resistive Switching Mechanisms on TaOx and $SrRuO_3$ Thin-Film Surfaces Probed by Scanning Tunneling Microscopy," *ACS Nano,* vol. 10, no. 1, p. 1481–1492, 2016.

[140] J. T. S. Irvine, D. C. Sinclair, and A. R. West, "Electroceramics: characterisation by Impedance Spectroscopy," *Adv. Mater.,* vol. 2, no. 3, pp. 132-138, 1990.

[141] N. Das, S. Tsui, Y. Y. Xue, Y. Q. Wang, and C. W. Chu, "Electric-field-induced submicrosecond resistive switching," *Phys. Rev. B,* vol. 78, no. 23-15, p. 235418, 2008.

[142] Y. H. You, B. S. So, and J. H. Hwang, "Impedance spectroscopy characterisation of resistance switching NiO thin films prepared through atomic layer deposition," *Appl. Phys. Lett.,* vol. 89, no. 22, p. 222105, 2009.

[143] D. S. Jeong, H. Schroeder, and R. Waser, "Impedance spectroscopy of TiO_2 thin films showing resistive switching," *Appl. Phys. Lett.,* vol. 89, no. 8, p. 082909, 2006.

[144] M. H. Lee, K. M. Kim, G. H. Kim, J. Y. Seok, S. J. Song, J. H. Yoon, and C. S. Hwang, "Study on the electrical conduction mechanism of bipolar resistive switching TiO_2 thin films using impedance spectroscopy," *Appl. Phys. Lett.,* vol. 96, no. 15, p. 152909, 2010.

[145] S. Yu, R. Jeyasingh, Y. Wu, and H. S. P. Wong, "AC conductance measurement and analysis of the conduction processes in HfO_x based resistive switching memory," *Appl. Phys. Lett.,* vol. 99, no. 23, p. 232105, 2011.

[146] S. Yu, R. Jeyasigh, Y. Wu, and H. S. P. Wong, "Understanding the Conduction and Switching Mechanism of Metal Oxide RRAM through Low Frequency Noise and AC Conductance Measurement and Analysis," in *IEDM Tech. Dig.*, 2011.

[147] Z. Fang, H. Y. Yu, J. A. Chroboczek, G. Ghibaudo, J. Buckley, B. De Salvo, X. Li, and D. L. Kwong, "Low-Frequency Noise in Oxide-Based ($TiN/HfO_x/Pt$) Resistive Random Access Memory Cells," *IEEE Trans. Electron Devices,* vol. 59, no. 3, pp. 850-853, 2012.

[148] R. Degraeve, L. Goux, S. Clima, B. Govoreanu, Y.Y. Chen, G.S. Kar, Ph. Roussel, G. Pourtois, D. J. Wouters, L. Altimime, M. Jurczak, G. Groeseneken, and J. A. Kittl, "Modeling and tuning the

filament properties in RRAM metal oxide stacks for optimized stable cycling," in *VLSI Symp. Tech. Dig.*, 2012.

[149] R. Degraeve, A. Fantini, S. Clima, B. Govoreanu, L. Goux, Y. Y. Chen, D. J. Wouters, Ph. Roussel, G. S. Kar, G. Pourtois, S. Cosemans, J.A. Kittl, G. Groeseneken, M. Jurczak, and L. Altimime, "Dynamic 'Hour Glass' Model for SET and RESET in HfO_2 RRAM," in *VLSI Symp. Tech. Dig.*, 2012.

[150] P. Huang, B. Gao, B. Chen, F. F. Zhang, L. F. Liu, G. Du, J. F. Kang, X. Y. Liu, "Stochastic Simulation of Forming, SET and RESET Process for Transition Metal Oxide-based Resistive Switching Memory," in *SISPAD*, 2012.

[151] P. Huang, X.Y. Liu, W.H. Li, Y.X. Deng, B. Chen, Y. Lu, B. Gao, L. Zeng, K.L. Wei, G. Du, X. Zhang, and J.F. Kang, "A Physical Based Analytic Model of RRAM Operation for Circuit Simulation," in *IEDM Tech. Dig.*, 2012.

[152] J. F. Kang, B. Gao, P. Huang, H. T. Li, Y. D. Zhao, Z. Chen, C. Liu, L. F. Liu, and X. Y. Liu, "Oxide-based RRAM: Requirements and Challenges of Modeling and Simulation," in *IEDM Tech. Dig.*, 2015.

[153] E. Simoen, B. Kaczer, M. T. Luque, and C. Claeys, "Random Telegraph Noise: From a Device Physicist's Dream to a Designer's Nightmare," *ECS Trans.*, vol. 39, no. 1, pp. 3-15, 2011.

[154] K. S. Ralls, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth and D. M. Tennant, "Discrete Resistance Switching in Submicrometer Silicon Inversion Layers: Individual Interface Traps and Low-Frequency (1/f) Noise," *Phys. Rev. Lett.*, vol. 52, no. 3-16, p. 228, 1984.

[155] T. Kang, J. Park, J. K. Lee, G. Kim, D. Woo, J. K. Son, J. H. Lee, B. G. Park, H. Shin, "Random telegraph noise in GaN-based light-emitting diodes," *Electron. Lett.*, vol. 47, no. 15, pp. 873-875, 2011.

[156] C. M. Compagnoni, M. Ghidotti, A. L. Lacaita, A. S. Spinelli, A. Visconti, "Random Telegraph Noise Effect on the Programmed Threshold-Voltage Distribution of Flash Memories," *IEEE Electron Device Lett.*, vol. 30, no. 9, pp. 984 - 986, 2009.

[157] Y. T. Chung, Y. H. Liu, P. C. Su, Y. H. Cheng, T. Wang, and M. C. Chen, "Investigation of Random Telegraph Noise Amplitudes in Hafnium Oxide Resistive Memory Devices," in *Proc. IRPS*, 2014.

[158] M. J. Kirton and M. J. Uren, ""Noise in solid state micro structures: A new perspective on individual defects, interface states, and low frequency (1/*f*) noise," *Adv. Phys.*, vol. 38, no. 4, p. 367–468, 1989.

[159] M. J. Kirton, M. J. Uren, S. Collins, M. Schulz, A. Karmann and K. Scheffer, "Individual defects at the $Si:SiO_2$ interface," *Semicond. Sci. Technol.*, vol. 4, no. 12, p. 1116, 1989.

[160] T. Nagumo, K. Takeuchi, S. Yokogawa, K. Imai and Y. Hayashi, "New Analysis Methods for Comprehensive Understanding of Random Telegraph Noise," in *IEDM Tech. Dig.*, 2009.

[161] S. Dongaonkar, M. D. Giles, A. Kornfeld, B. Grossnickle, J. Yoon, "Random Telegraph Noise (RTN) in 14nm Logic Technology: High Volume Data Extraction and Analysis," in *VLSI Symp. Tech. Dig.*, 2016.

[162] M. T. Luque, R. Degraeve, Ph. J. Roussel, V. Luong1, B. Tang, J.G. Lisoni, C. L. Tan, A. Arreghini, G. Van den bosch, G. Groeseneken, J. Van Houdt., "Statistical spectroscopy of switching traps in deeply scaled vertical poly-Si channel for 3D memories," in *IEDM Tech. Dig.*, 2013.

[163] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257-286, 1989.

[164] J. Hornegger, H. Niemann, D. Paulus and G. Schlottke, "Object recognition using hidden Markov models," *Proc. Pattern Recognition in Practice*, vol. IV, pp. 37-44, 1994.

[165] L. E. Baum and T. Petrie, "Statistical Inference for Probabilistic Functions of Finite State Markov Chains," *The Annals of Mathematical Statistics*, vol. 37, no. 6, p. 1554–1563, 1966.

[166] L. E. Baum, J. A. Eagon, "An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology," *Bulletin of the American Mathematical Society*, vol. 73, no. 3, p. 360, 1967.

[167] L. E. Baum, G. R. Sell, "Growth transformations for functions on manifolds," *Pacific Journal of Mathematics*, vol. 27, no. 2, p. 211–227, 1968.

[168] L. E. Baum, T. Petrie, G. Soules, N. Weiss, "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains," *The Annals of Mathematical*

Statistics, vol. 41, p. 164, 1970.

[169] L. E. Baum, "An Inequality and Associated Maximization Technique in Statistical Estimation of Probabilistic Functions of a Markov Process," *Inequalities,* vol. 3, pp. 1–8,, 1972.

[170] L. R. Juang and B. H. Rabiner, "An introduction to hidden Markov models," *IEEE ASSP Mag.*, vol. 3, no. 1, pp. 4-16, 1986.

[171] Mathworks, "Hidden Markov Models (HMM)", Retrieved on 1st May 2017 from *https://uk.mathworks.com/help/stats/hidden-markov-models-hmm.html*.

[172] D. S. Yaney, C. Y. Lu, R. A. Kohler, M. J. Kelly and J. T. Nelson, "A meta-stable leakage phenomenon in DRAM charge storage – variable hold time," in *IEDM Tech. Dig.*, 1987.

[173] P. J. Restle, J. W. Park and B. F. Floyd, "DRAM variable retention time," in *IEDM Tech. Dig.*, 1992.

[174] A. Ghetti, C. M. Compagnoni, A. S. Spinelli, and A. Visconti, "Comprehensive Analysis of Random Telegraph Noise Instability and Its Scaling in Deca–Nanometer Flash Memories," *IEEE Trans. Electron Devices,* vol. 56, no. 8, pp. 1746-1752, 2009.

[175] F. M. Puglisi and P. Pavan, "A Microscopic Physical Description of RTN Current Fluctuations in HfO_x RRAM," in *Proc. IRPS*, 2015.

[176] S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Ramaswamy, and D. Lelmini, "Statistical Fluctuations in HfO_x Resistive-Switching Memory: Part II—Random Telegraph Noise," *IEEE Trans. Electron Devices,* vol. 61, no. 8, pp. 2920 - 2927, 2014.

[177] J. K. Lee, J. W. Lee, J. Park, S. W. Chung, J. S. Roh, S. J. Hong, I. Cho, H. I. Kwon, and J. H. Lee, "Extraction of trap location and energy from," *Appl. Phys. Lett.,* vol. 98, p. 143502, 2011.

[178] Y. J. Huang, S. S. Chung, H. Y. Lee, Y. S. Chen, F. T. Chen, P. Y. Gu and M. J. Tsai, "The physical insights into an abnormal erratic behavior in the resistance random access memory," in *Proc. IRPS*, 2013.

[179] Y. Pan, Y. Cai, Y. Liu, Y. Fang, M. Yu, S. Tan and R. Huang, "Microscopic origin of read current noise in TaO_x-based resistive switching memory by ultra-low temperature measurement," *Appl. Phys. Lett.,* vol. 108, no. 15, p. 153504, 2016.

[180] D. Veksler, G. Bersuker, B. Chakrabarti, E. Vogel, S. Deora, K. Matthews,, "Methodology for the statistical evaluation of the effect of random telegraph noise (RTN) on RRAM characteristics D. C. Gilmer, H. F. Li, S. Gausepohl, P. D. Kiesch," in *IEDM Tech. Dig.*, 2012.

[181] N. Raghavan, R. Degraeve, L. Goux, A. Fantini, D.J. Wouters, G. Groeseneken, M. Jurczak, "RTN Insight to Filamentary Instability and Disturb Immunity in Ultra-Low Power Switching HfOx and AlO_x RRAM," in *VLSI Symp. Tech. Dig.*, 2013.

[182] M. O. Andersson, Z. Xiao, S. Norrman, and O. Engstrom, "Model based on trap-assisted tunneling for two-level current fluctuations in submicrometer metal-silicon-dioxide-silicon diodes," *Phys. Rev. B,* vol. 41, no. 14, p. 9836, 1990.

[183] B. Kacer, M. T. Luque, W. Goes, T. Grasser, G. Groeseneken, "Gate current random telegraph noise and single defect conduction," *Microelectronic Engineering,* vol. 109, no. C, pp. 123-125, 2013.

[184] G. Kapila and V. Reddy, "Impact of Sampling Rate on RTN Time Constant Extraction and Its Implications on Bias Dependence and Trap Spectroscopy," *IEEE Trans. Device Mater. Rel.,* vol. 14, no. 2, pp. 616-622, 2014.

[185] S. Lee, H. J. Cho, Y. Son, D. S. Lee and H. Shin, "Characterisation of Oxide Traps Leading to RTN in High-k and Metal Gate MOSFETs," in *IEDM Tech. Dig.*, 2009.

[186] C. M. Chang, Steve S. Chung, Y. S. Hsieh, L. W. Cheng, C. T. Tsai, G. H. Ma, S. C. Chien, and S. W. Sun, "The Observation of Trapping and Detrapping Effects in High-k Gate Dielectric MOSFETs by a New Gate Current Random Telegraph Noise (I_G-RTN) Approach," in *IEDM. Tech. Dig.*, 2008.

[187] J. Park, D. Kang, J. K. Son and H. Shin, "Extraction of Location and Energy Level of the Trap Causing Random Telegraph Noise at Reverse-Biased Region in GaN-Based Light-Emitting Diodes," *IEEE Trans. Electron Devices,* vol. 59, no. 12, pp. 3495 - 3502, 2012.

[188] L. Goux, J. G. Lisoni, M. Jurczak, D. J. Wouters, L. Courtade, and Ch. Muller, "Coexistence of the bipolar and unipolar resistive-switching modes in NiO cells made by thermal oxidation of Ni layers," *J. Appl. Phys.,* vol. 107, no. 2, p. 024512, 2010.

[189] D. Lee, H. Choi, H. Sim, D. Choi, H. Hwang, M. J. Lee, S. A. Seo and I. K. Yoo, "Resistance Switching of the Nonstoichiometric Zirconium Oxide for Nonvolatile Memory Applications,"

IEEE Electron. Dev. Lett., vol. 26, no. 9, pp. 719-721, 2005.

[190] S. Y. Wang, C. H. Tsai, D. Y. Lee, C. Y. Lin, C. C. Lin, T. Y. Tseng, "Improved resistive switching properties of Ti/ZrO_2/Pt memory device for RRAM application," *Microelectronics Eng.*, vol. 88, pp. 1628-1632, 2011.

[191] L. Goux, Y. Y Chen, L. Pantisano, X. P. Wang, G. Groeseneken, M. Jurczak, and D. J. Wouters, "On the Gradual Unipolar and Bipolar Resistive Switching of TiN\HfO_2\Pt Memory Systems," *Electrochemical and Solid-state Letter,* vol. 13, no. 6, pp. G54-G56, 2010.

[192] L. Zhang, M. Zhu, R. Huang, D. Gao, Y. Kuang, C. Shi and Y. Wang, "Forming-Less Unipolar TaOx-Based RRAM with Large CC-Independence Range for High Density Memory Applications," *ECS Trans.*, vol. 27, no. 1, pp. 3-8, 2010.

[193] Z. Wei, Y. Kanzawa, K. Arita, Y. Katoh, K. Kawai, S. Muraoka, S. Mitani, S. Fujii, K.Katayama, M. Iijima, T. Mikawa, T. Ninomiya, R. Miyanaga, Y. Kawashima, K. Tsuji, A. Himeno, T. Okada, R. Azuma, K. Shimakawa, H. Sugaya, and T. Takagi, R. Yasuhara et al, "Highly Reliable TaO_x ReRAM and Direct Evidence of Redox Reaction Mechanism," in *IEDM Tech. Dig.*, 2008.

[194] K. M. Kim, B. J. Choi, B. W. Koo, S. Choi, D. S. Jeong and C. S. Hwang, "Resistive Switching in Pt/Al_2O_3/TiO_2/Ru Stacked Structures," *Electrochem. Solid-State Lett.*, vol. 9, no. 12, p. G343 – 346, 2006.

[195] Y. Wu, B. Lee and P. H. Wong, "Ultra-low power Al_2O_3-based RRAM with 1μA reset current," in *VLSI-TSA*, 2010.

[196] K. L. Lin, Y. M. Tseng, J. H. Lin, J. Shieh, Y. J. Lee, T. H. Hou and T. F. Lei,, "High-performance Ni/SiO_2/Si Programmable Metallization Cell," in *Nanoelectronics Conference (INEC)*, 2011.

[197] L. Liu, B. Gao, B. Chen, Y. Chen, Y. Wang, J. Kang and R. Han, "On the bipolar and unipolar resistive switching characteristics in Ag/SiO_2/Pt memory cells," in *International Conference on Solid-State and Integrated Circuit Technology - ICSICT*, 2010.

[198] Y. H. Kim, J. C. Lee, "Reliability characteristics of high-k dielectrics," *Microelectron. Reliab.*, vol. 44, no. 2, pp. 183-193, 2004.

[199] R. Chau, S. Datta, M. Doczy, B. Doyle, J. Kavalieros, M. Metz,, "High-k/metal–gate stack and its MOSFET characteristics," *IEEE Electron Device Lett.*, vol. 25, no. 6, pp. 408-410, 2004.

[200] F. M. Puglisi, A Unified Understanding of HfO_2-RRAM Operations: Modeling and Reliability, University of Modena and Reggio Emilia: Doctoral dissertation, 2014.

[201] T. Sakamoto, N. Banno, N. Iguchi, H. Kawaura, H. Sunamura, S. Fujieda, K. Terabe, T. Hasegawa, M. Aono, "A Ta_2O_5 solid-electrolyte switch with improved reliability," in *VLSI Symp. Tech. Dig.*, 2007.

[202] Y. Y. Chen, L. Goux, J. Swerts, M. Toeller, C. Adelmann, J. Kittl, M. Jurczak, G. Groeseneken and D. J. Wouters, "Hydrogen-Induced Resistive Switching in TiN/ALD HfO_2/PEALD TiN RRAM Device," *IEEE Electron Dev. Lett.*, vol. 33, no. 4, pp. 483-485,, 2012.

[203] Y. Y. Chen, R. Degraeve, S. Clima, B. Govoreanu, L. Goux, A. Fantini, G. S. Kar, G. Pourtois, G. Groeseneken, D. J. Wouters and M. Jurczak, "Understanding of the Endurance Failure in Scaled HfO_2-based 1T1R RRAM through Vacancy Mobility Degradation," in *IEDM. Tech. Dig.*, 2012.

[204] S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Ramaswamy, D. Ielmini, "Understanding switching variability and random telegraph noise in resistive RAM," in *IEDM. Tech. Dig.*, 2013.

[205] F. A. Kroger and H. J. Vink, "Relations between the concentrations of imperfections in solids," *J. Phys. Chem.*, vol. 5, no. 3, p. 208, 1958.

[206] Y. Guo and J. Robertson, "Materials selection for oxide-based resistive random access memories," *Applied Phys. Lett.*, vol. 105, no. 22, p. 223516, 2014.

[207] B. Govoreanu, S. Clima, I. P. Radu, Y. Y. Chen, D. J. Wouters, and M. Jurczak, "Complementary Role of Field and Temperature in Triggering ON/OFF Switching Mechanisms in Hf/HfO_2 Resistive RAM Cells," *IEEE Trans. Elect. Dev.*, vol. 60, no. 8, pp. 2471-2478, 2013.

[208] D. B. Strukov, F. Alibart and R. S. Williams, "Thermophoresis/diffusion as a plausible mechanism for unipolar resistive switching in metal–oxide–metal memristors," *Appl. Phys. A*, vol. 107, no. 3, p. 509–518, 2012.

[209] N. Raghavan, R. Degraeve, A. Fantini, L. Goux, S. Strangio, B. Govoreanu, D. Wouters, G. Groeseneken, and M. Jurczak, "Microscopic origin of random telegraph noise fluctuations in aggressively scaled RRAM and its impact on read disturb variability," in *IRPS*, 2013.

[210] S. Subhechha, B. Govoreanu, Y. Chen, S. Clima, K. D. Meyer, J. V. Houdt, and M. Jurczak, "Extensive Reliability Investigation of a-VMCO Nonfilamentary RRAM: Relaxation, Retention

and Key Differences to Filamentary Switching," in *IRPS*, 2016.

[211] M. Perego, G. Seguini, G. Scarel, M. Fanciulli, and F. Wallrapp, "Energy band alignment at TiO_2/$SiTiO_2$/Si interface with various interlayers," *Journal of Applied Physics*, vol. 103, no. 4, p. 043509, 2008.

[212] L. Zhang, B. Govoreanu, A. Redolfi, D. Crotti, H. Hody, V. Paraschiv, S. Cosemans, C. Adelmann, T. Witters, S. Clima, Y. Y. Chen, P. Hendrickx, D. J. Wouters, G. Groeseneken, M. Jurczak, "High-drive current (>1MA/cm^2) and highly nonlinear (>10^3) TiN/amorphous-Silicon/TiN scalable bidirectional selector with excellent reliability and its variability impact on the 1S1R array performance," in *IEDM. Tech. Dig.*, 2014.

[213] T. Nagumo, K. Takeuchi, T. Hase and Y. Hayashi, "Statistical characterisation of Trap Position, Energy, Amplitude and Time Constants by RTN Measurement of Multiple Individual Traps," in *IEDM. Tech. Dig.*, 2010.

[214] K. Baek, S. Park, J. Park, Y. M. Kim, H. Hwang and S. H. Oh, "In situ TEM observation on the interface-type resistive switching by electrochemical redox reactions at a TiN/PCMO interface," *Nanoscale*, vol. 9, no. 2, p. 582–593, 2017.

[215] Y. Chen, and C. Petti, "ReRAM technology evolution for storage class memory application," in *Solid-State Device Research Conference (ESSDERC)*, 2016.

[216] F. Mondon, and S. Blonkowski, "Electrical characterisation and reliability of HfO_2 and Al_2O_3–HfO_2 MIM capacitors," *Microelectronics Reliability*, vol. 43, no. 8, pp. 1259-1266, 2003.

[217] E. Wu, B. Li, J. Stathis, R. Achanta, R. Filippi, and P. McLaughlin, "A time-dependent clustering model for non-uniform dielectric breakdown," in *IEDM. Tech. Dig.*, 2013.

[218] W. C. Luo, J. C. Liu, H. T. Feng, Y. C. Lin, J. J. Huang, K. L. Lin, and T. H. Hou, "RRAM SET speed-disturb dilemma and rapid statistical prediction methodology," in *IEDM. Tech. Dig.*, 2012.

[219] N. Raghavan, D. D. Frey, M. Bosman, and K. L. Pey, "Monte Carlo model of reset stochastics and failure rate estimation of read disturb mechanism in HfO_x RRAM," in *Reliability Physics Symposium (IRPS)*, 2015.

[220] Y. B. Kim, S. R. Lee, D. Lee, C. B. Lee, M. Chang, J. H. Hur, M. J. Lee, G. S. Park, C. J. Kim, U. I.Chung, I. K. Yoo, and K. Kim, "Bi-layered RRAM with unlimited endurance and extremely uniform switching," in *Proc. Symp. VLSI Technol.*, 2011.

[221] A. C. Torrezan, J. P. Strachan, G. Medeiros-Ribeiro and R. S. Williams, "Sub-nanosecond switching of a tantalum oxide memristor," *Nanotechnology*, vol. 22, no. 48, p. 485203, 2011.

[222] Z. Wei, T. Takagi, Y. Kanzawa, Y. Katoh, T. Ninomiya, K. Kawai, S. Muraoka, S.Mitani, K. Katayama, S. Fujii, R. Miyanaga, Y. Kawashima, T. Mikawa, K. Shimakawa, and K. Aono, "Demonstration of high-density ReRAM ensuring 10-year retention at 85°C based on a newly developed reliability model," in *IEDM Tech. Dig.*, 2011.

[223] J. J. Yang, M. X. Zhang, J. P. Strachan, F. Miao, M. D. Pickett, R. D. Kelley, G. M. Ribeiro and R. S. Williams, "High switching endurance in TaO_x memristive devices," *Appl. Phys. Lett.*, vol. 97, no. 23, p. 232102, 2010.

[224] H. K. Yoo, S. B. Lee, J. S. Lee, S. H. Chang, M. J. Yoon, Y. S. Kim, B. S. Kang, M. J. Lee, C. J. Kim, B. Kahng, and T. W. Noh, "Conversion from unipolar to bipolar resistance switching by inserting Ta_2O_5 layer in Pt/TaOx/Pt cells," *Applied Physics Letters*, vol. 98, no. 18, p. 183507, 2011.

[225] D. Ielmini, "Reset–Set Instability in Unipolar Resistive-Switching Memory," *IEEE Electron Device Letters*, vol. 31, no. 6, pp. 552-554, 2010.

[226] J. H. Hur, M. J. Lee, C. B. Lee, Y. B. Kim and C. J. Kim, "Modeling for bipolar resistive memory switching in transition-metal oxides," *Phys. Rev. B*, vol. 82, no. 15, p. 155321, 2010.

[227] K. Kawai, A. Kawahara, R. Yasuhara, S. Muraoka, Z. Wei, R. Azuma, K. Tanabe and K. Shimakawa, "Highly-reliable TaO_x ReRAM technology using automatic forming circuit," in *ICICDT*, 2014.

[228] C. W. Hsu, C. C. Wan, I. T. Wang, M. C. Chen, C. L. Lo, Y. J. Lee, W. Y. Jang, C. H. Lin and T. H. Hou, "3D vertical TaO_x/TiO_2 RRAM with over 10^3 self-rectifying ratio and sub-μA operating current," in *IEDM. Tech. Dig.*, 2013.

[229] C. B. Lee, D. S. Lee, A. Benayad, S. R. Lee, M. Chang, M. J. Lee, J. Hur, Y. B. Kim, C. J. Kim and U. I. Chung, "Highly uniform switching of tantalum embedded amorphous oxide using self-compliance bipolar resistive switching," *IEEE Electron Device Lett.*, vol. 32, no. 3, p. 399–401, 2011.

[230] S. R. Lee, Y. B. Kim, M. Chang, K. M. Kim, C. B. Lee, J. H. Hur, G. S. Park, D. Lee, M. J. Lee, C. J. Kim, U. I. Chung, I. K. Yoo, K. Kim, "Multi-level switching of triple-layered TaO_x RRAM with excellent reliability for storage class memory," in *Proc. Symp. VLSI. Tech.*, 2012.

[231] W. S. Chen, T. Y. Wu, S. Y. Yang, W. H. Liu, H. Y. Lee, Y. S. Chen, C. H. Tsai, P. Y. Gu, K. H. Tsai, P. S. Chen, H. W. Wei et al, "Stabilization of resistive switching with controllable self-compliant Ta_2O_5-based RRAM," in *Proc. IEEE Int. Symp. VLSI-TSA*, 2012.

[232] G. S. Park, Y. B. Kim, S. Y. Park, X. S. Li, S. Heo, M. J. Lee, M. Chang, J. H. Kwon, M. Kim, U. I. Chung, R. Dittmann, R. Waser and K. Kim, "In situ observation of filamentary conducting channels in an asymmetric Ta_2O_{5-x}/TaO_{2-x} bilayer structure," *Nature Communications*, vol. 4, no. 2382, 2013.

[233] L. Goux, A. Fantini, A. Redolfi, C.Y. Chen, F.F. Shi, R. Degraeve, Y.Y. Chen, T. Witters, G. Groeseneken, M. Jurczak, "Role of the Ta scavenger electrode in the excellent switching control and reliability of a scalable low-current operated TiN\Ta_2O_5\Ta RRAM device," in *VLSI. Symp. Tech. Dig.*, 2014.

[234] X. Guan, S. Yu, and H. S. P. Wong, "On the switching parameter variation of metal-oxide RRAM—Part I: Physical modeling and simulation methodology," *IEEE Trans. Electron Devices*, vol. 59, no. 4, p. 1172–1182, 2012.

[235] B. Chen, B. Gao, S. W. Sheng, "A novel operation scheme for oxide-based resistiveswitching memory devices to achieve controlled switching behaviors," *IEEE Electron Device Letters*, vol. 2, no. 3, pp. 282 - 284, 2011.

[236] B. Gao, H. Zhang, B. Chen, L. Liu, X. Liu, R. Han, J. Kang, Z. Fang, H. Yu, B. Yu, and D. L. Kwong, "Modeling of Retention Failure Behavior in Bipolar Oxide-Based Resistive Switching Memory," *IEEE Electron Device Lett.*, vol. 32, no. 3, p. 276–278, 2011.

[237] C. Y. Chen, L. Goux, A. Fantini, A. Redolfi, S. Clima, R. Degraeve, Y. Y. Chen, G. Groeseneken, M. Jurczak, "Understanding the Impact of Programming Pulses and Electrode Materials on the Endurance Properties of Scaled Ta_2O_5 RRAM Cells," in *IEDM. Tech. Dig.*, 2014.

[238] S. H. Chang, J. S. Lee, S. C. Chae, S. B. Lee, C. Liu, B. Kahng, D.-W. Kim, and T. W. Noh, "Occurrence of Both Unipolar Memory and Threshold Resistance Switching in a NiO Film," *Phys. Rev. Lett.*, vol. 102 , no. 2-16, p. 026801, 2009.

[239] C. Villa, D. Mills, G. Barkley, H. Giduturi, S. Schippers, D. Vimercati, "A 45nm 1GB 1.8V phase-change memory," in *ISSCC*, 2010.

[240] Y. Choi, I. Song, M. H. Park, H. Chung, S. Chang, B. Cho, J. Kim, et al, "A 20nm 1.8V 8Gb PRAM with 40MB/s program bandwidth," in *ISSCC*, 2012.

[241] T. Y. Liu, T. H. Yan, R. Scheuerlein, Y. Chen, J. K. Lee, G. Balakrishnan, G. Yee, et al, "A 130.7mm^2 2-layer 32GB ReRAM memory device in 24nm technology," in *ISSCC*, 2013.

[242] M. Adams, "Winter Analyst Conference," Micron Technology, Inc., 2015.

[243] Micron Technology, Inc., "Breakthrough Nonvolatile Memory Technology," Retrieved on 1st May 2017 from *http://www.micron.com/about/innovations/3D-xpoint-technology.*

[244] Y. Chen, J. Zhao, and Y. Xie, "3D-nonFAR: Three-dimensional non-volatile FPGA architecture using phase change memory," in *Proc. Int. Symp. Low Power Electron. Design*, 2010.

[245] P. Gaillardon, D. Sacchetto, G. B. Beneventi, M. H. B. Jamaa, L. Perniola, F. Clermidy, I. O'Connor, G. De Micheli, "Design and Architectural Assessment of 3-D Resistive Memory Technologies in FPGAs," *IEEE Trans. Nanotechnol.*, vol. 12, no. 1, pp. 40-50, 2013.

[246] K. Huang, R. Zhao, W. He and Y. Lian, "High-Density and High-Reliability Nonvolatile Field-Programmable Gate Array With Stacked 1D2R RRAM Array," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 24, no. 1, pp. 139-150, 2016.

[247] X. Tang, G. Kim, P. E. Gaillardon, and G. De Micheli, "A Study on the Programming Structures for RRAM-Based FPGA Architectures," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 63, no. 4, pp. 503-516, 2016.

[248] S. Onkaraiah, O. Turkyilmaz, M. Reyboz, F. Clermidy, E. Vianello, J. M. Portal and C. Muller, "A Hybrid CBRAM/CMOS Look-Up-Table structure for improving performance efficiency of Field-Programmable-Gate-Array," in *ISCAS*, 2013.

[249] G. Indiveri and S. C. Liu, "Memory and Information Processing in Neuromorphic Systems," in *Proceedings of the IEEE*, 2015.

[250] D. Querlioz, O. Bichler, A. F. Vincent and C. Gamrat, "Bioinspired Programming of Memory Devices for Implementing an Inference Engine," *Proc. IEEE*, vol. 103, no. 8, pp. 1398-1416,

2015.

[251] G. W. Burr, R. M. Shelby, D. di Nolfo, J. W. Jang, R. S. Shenoy, P. Narayanan, K. Virwani, E. U. Giacometti, B. Kurdi, H. Hwang, "Experimental Demonstration and Tolerancing of a Large-Scale Neural Network (165 000 Synapses) Using Phase-Change Memory as the Synaptic Weight Element," in *IEDM. Tech. Dig.*, 2015.

[252] I. T. Wang, C. C. Chang, L. W. Chiu, T. Chou and T. H. Hou, "3D Ta/TaO_x/TiO_2/Ti synaptic array and linearity tuning of weight update for hardware neural network applications," *Nanotechnology,* vol. 27, no. 36, p. 365204, 2016.

[253] A. Pantazi, S. Wozniak, T. Tuma and E. Eleftheriou, "All-memristive neuromorphic computing with level-tuned neurons," *Nanotechnology,* vol. 27, no. 35, p. 355205, 2016.

[254] P. Zhou, M. Yin, H .J. Wan, H. B. Lu, T. A. Tang, and Y. Y. Lin, "Role of TaON interface for Cu_xO resistive switching memory based on a combined model," *Appl. Phys. Lett.,* vol. 94, p. 053510, 2009.

Publication List

1. **Z. Chai**, J. Ma, W. Zhang, B. Govoreanu, E. Simoen, J. F. Zhang, Z. Ji, R. Gao, G. Groeseneken, M. Jurczak, *RTN-based defect tracking technique: experimentally probing the spatial and energy profile of the critical filament region and its correlation with HfO_2 RRAM switching operation and failure mechanism*, Digest of Technical Papers - **Symposium on VLSI Technology, IEEE 2016 Symposia on VLSI Technology and Circuits (Orally presented in Honolulu, HI, US, June 2016)**

2. J. Ma, **Z. Chai**, W. Zhang, B. Govoreanu, J. F. Zhang, Z. Ji, B. Benbakhti, G. Groeseneken and M. Jurczak, *Identify the critical regions and switching/failure mechanisms in non-filamentary RRAM (a-VMCO) by RTN and CVS techniques for memory window improvement,* Technical Digest - **International Electron Devices Meeting**. Institute of Electrical and Electronics Engineers (IEEE 2016)

3. W. Zhang, **Z. Chai**, J. Ma, J. F. Zhang, Z. Ji *Analysis of RTN signals in Resistive-Switching RAM device and its correlation with device operations*, **ICSICT 2016**

4. C. Claeys, M.G.C. de Andrade, **Z. Chai**, W. Fang, B. Govoreanu, B. Kaczer, W. Zhang, E. Simoen, *Random Telegraph Signal Noise in Advanced High Performance and Memory Devices*, SBMicro 2016 (Belo Horizonte, Brazil)

5. **Z. Chai**, J. Ma, W. Zhang, B. Govoreanu, J. F. Zhang, Z. Ji, G. Groeseneken, M. Jurczak, *Probing the Critical Region of Conductive Filament in Nanoscale HfO_2 Resistive-Switching Memory by Random Telegraph Signals*, **Trans. Electron Devices** (submitted)

6. **Z. Chai**, J. Ma, W. Zhang, B. Govoreanu, H. Cao, J. F. Zhang, Z. Ji, L. Goux, A. Belmonte, R. Degraeve, L. Di Piazza, G. Kar, *Understanding defect profile in non-filamentary RRAM (a-VMCO) for analogue switching and correlation with different endurance behaviour in practical memory applications* **IEEE Trans. Electron Devices** (submitted)

7. **Z. Chai**, Jigang Ma, Weidong Zhang, Bogdan Govoreanu, Jian Fu Zhang, Zhigang Ji, Guido Groeseneken, Gouri S. Kar *Understanding the switching model of TaOx-Based Unipolar and Bipolar Resistive Random Access Memory based on RTN technique*, **IEEE Trans. Electron Devices** (in preparation)

8. J. Ma, **Z. Chai**, W. Zhang, B. Govoreanu, J. F. Zhang, Z. Ji, B. Benbakhti, G. Groeseneken and G. Kar, *Investigation of cell failure in a-VMCO RRAM: identify critical regions and weak paths through CVS/RTN analysis*, **IEEE Trans. Electron Devices** (in preparation)

9. J. Ma, **Z. Chai**, W. Zhang, B. Govoreanu, J. F. Zhang, Z. Ji, B. Benbakhti, G. Groeseneken and G. Kar, *Investigate the defects in non-filamentary RRAM and their impacts on stressed induced read disturbance by RTN technique*, **IEEE Trans. Electron Devices** (in preparation)