

## ABSTRACT

SARKAR, BIPLAB. Atomic Layer Deposition (ALD) Enabled Techniques for Novel Memory Applications. (Under the direction of Dr. Veena Misra.)

Moore's law has resulted in the aggressive scaling of both logic and memory devices. Although the memory performance has been greatly improved as a result of scaling, but the volatile memory and non-volatile memory still take two different areas in a circuit. Traditionally, permanent data such as codes, look-up tables, etc. are stored in the non-volatile memory segment of the processor. On the other hand, volatile memories are used to perform logical operations. A significant advantage in circuit area can be achieved if the two memory technologies are combined in the same device. Such a device can potentially provide higher processing speed, and require simpler circuit architecture for accessing the memory elements. The device can also be potentially attractive for applications like instant ON computing, neuromorphic systems, sensor systems, etc.

In this dissertation, novel concepts of unified memory devices have been discussed which can combine the volatile and non-volatile operation in a single transistor. A dual floating gate memory MOSFET has been demonstrated to show simultaneous volatile and non-volatile operation. The device structure is very similar to a flash memory, but have an additional floating gate. The gate stack of the device consists of a thin dielectric separating two floating gates sandwiched between a tunnel dielectric and inter-poly dielectric. Depending upon the bias applied at the gate of the transistor, the device can be programmed into volatile mode, non-volatile mode and a concurrent mode. Non-volatile operation of the device is similar to the conventional flash memories, whereas the volatile operation can be realized by the charge transfer between the floating gates. The device is also capable of showing concurrent mode of operation where volatile and non-volatile memory operations can be realized simultaneously by choosing an appropriate pulse scheme.

Flash memories need higher voltage for their operation, and hence they require higher power than some of the emerging memory devices. To provide low power unified memory solutions, resistive switching memories (or RRAMs) have also been explored in this dissertation.  $\text{HfO}_2$  and  $\text{Al}_2\text{O}_3$  based RRAMs were investigated to understand the non-volatile RRAM operation.  $\text{HfO}_2$  RRAM has been widely explored by many researchers, and is considered as one of the mature RRAM candidate. However, a systematic study of  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM revealed the advantages of bilayer dielectric RRAM over the  $\text{HfO}_2$  RRAM.  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM was observed to be advantageous over the  $\text{HfO}_2$  RRAM providing a lower

forming voltage and higher  $I_{on}/I_{off}$  ratio. Similarly,  $TiO_2$  RRAM was explored to understand the low power non-filamentary RRAM operation which can be used as DRAM replacement. Oxygen vacancies in the  $TiO_2$  RRAM was modulated with the application of electric field thereby resulting in a volatile memory transition.

RRAM is also a promising candidate to mimic the synapse of a neuromorphic system. RRAM can behave like an analog memory with infinite number of memory states, and the device can be modulated into those states depending upon the applied synaptic weight. However, one key requirement of RRAM to be applicable for synaptic applications is the feasibility of gradual reset mechanism. In this work, a systematic study leading to attain gradual reset in  $Al_2O_3$  and  $HfO_2$  based RRAM has also been performed. A vacancy modulation technique was developed to convert the abrupt reset into gradual reset. The gradual reset was also shown to have excellent endurance and retention behavior applicable for multi-bit storage. Finally, the synaptic learning was performed on both the RRAMs to confirm the applicability of the RRAM devices to the neuromorphic systems.

Coming back to the unified memory device,  $HfO_2$  and  $Al_2O_3$  RRAM showed filamentary operation with larger memory window, but also needed a higher power for their operation. Thus, the filamentary RRAM can be used to store the non-volatile data. The  $TiO_2$  non-filamentary RRAM showed a low power operation, hence can be used for logical operations. Finally, few new device architectures of unified memory have been proposed in this dissertation. All these devices can be potential candidate for the future embedded memory applications.

© Copyright 2015 by Biplab Sarkar

All Rights Reserved



Atomic Layer Deposition (ALD) Enabled Techniques for Novel Memory Applications

by  
Biplab Sarkar

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the Degree of  
Doctor of Philosophy

Electrical Engineering

Raleigh, North Carolina

2015

APPROVED BY:

---

Dr. Paul Franzon

---

Dr. John Muth

---

Dr. Michael Dickey

---

Dr. Veena Misra  
Chair of Advisory Committee

ProQuest Number: 10585439

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10585439

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 – 1346

## **DEDICATION**

To my father  
Bijan Behari Sarkar

my mother  
Rina Rani Sarkar

and my wife  
Tapasi Dey

## BIOGRAPHY

Biplab was born in a small town of Tripura, India. After completing his high school, he went to North-Eastern Regional Institute of Science and Technology (NERIST), Itanagar, India for his undergraduate studies. He obtained bachelors degree in Electronics and Communication Engineering in the year 2010. During his bachelor studies, he went to Indian Institute of Technology Guwahati for a short internship where he worked under the supervision of Dr. Amit kumar Misra. He worked on apodization techniques applied to signal processing. After completing the bachelors degree, he joined the Indian Institute of Technology (IIT) Bombay, Mumbai, India for his masters program. He completed his masters with a specialty in microelectronics and VLSI, under the department of Electrical Engineering, IIT Bombay. During masters, he worked on spintronic devices based on AMR, GMR and TMR effect. Alongside, he also worked on spin-injection in GaN using ferromagnetic metal contacts. He was also a teaching assistant for several undergraduate and postgraduate courses in IIT Bombay.

After completing the masters program, he joined the research group of Dr. Veena Misra at North Carolina State University for his doctoral studies. During his doctoral studies, he worked extensively on novel memories such as flash memories, RRAM, 1T-0C DRAM etc. He has demonstrated for the first time, a dual floating gate memory MOSFET capable of showing simultaneous volatile and non-volatile operation. This memory was a novel concept of realizing unified memory which can have tremendous impact on instant-ON computing systems, sensor systems, neuromorphic systems, etc. He presented this idea of dual floating gate flash memory in Non-volatile memory technology Symposium (NMVTS, 2013) conference held in Minneapolis, USA. Apart from flash memories, he also worked on RRAM devices to understand the oxygen vacancy induced memory operation in binary oxides. He has worked on  $\text{Al}_2\text{O}_3$  and  $\text{HfO}_2$  based RRAM devices to show synaptic learning behavior applicable for neuromorphic systems. During his Ph.D, he went for an internship at Intel Corporation, Boise, Idaho, USA, where he worked on reliability analysis of NAND flash memories. He worked with the research groups of Intel Corporation and Micron Technology to understand the failure concerns of flash memories highly used in flash drives and solid state drives (SSDs). He successfully defended his PhD on 11th August 2015.

## ACKNOWLEDGEMENTS

I am deeply thankful to my advisor Dr. Veena Misra, for providing me the golden opportunity to work on this exciting project. She is a great mentor. She has always provided valuable time to listen to my troubles, and offered her valuable guidance. She is certainly my role model of a dedicated scientist and a great human being. No word can explain the zeal of working under her supervision.

Special thanks to Dr. Paul Franzon. The Ph.D work could have been impossible without his support and mentorship. Being the PI of the project, he has been organizing the smooth work culture for the project.

I would like to also thank Dr. John Muth and Dr. Michael Dickey for serving on my committee and offering critical comments and feedback.

Special thanks to Dr. Bongmook Lee for introducing me to the world of memory devices and providing all the training I needed for device fabrication and characterization. Dr. Lee has always been my first go-to person whenever I had trouble in my research, and he has been always available. More interestingly, he always had an answer to every research problem I faced in the lab.

Special thanks to Dr. Narayanan Ramanan, my mentor during PhD. Without his constant interaction, training and mentorship, the journey would have been difficult. I will always cherish the "tea time" with Dr. Narayanan and Amir Hassani. Thanks guyz for tolerating me through this time.

I express my sincere appreciation to the staff members of the NCSU Nanofabrication Facility (NNF), especially Marcio, Henry, Bruce, Nicole. I salute their dedicated efforts in keeping the cleanroom facilities functional. I would also like to acknowledge the help and support of my past and present group members, Dr. Srikant Jayanti, Dr. Xiangyu Yang, Dr. Casey Kirkpatrick, Steven Mills, Michael Lim, Akhilesh Taneeru, Faisal Azam, Fransisco Suarez, Simon Wang. A lot of my work is inspired from the discussions we had on device physics, processes and characterization techniques. I am also thankful for their responsible maintenance of all our lab equipments with sincere dedication.

I express my sincere gratitude to Hanmant Belgal for giving me the opportunity of working with him as an intern at Intel Corporation, Boise, ID. I only wish I had more time to spend

under his mentorship. I learnt a lot about the memory industry during my short time at Intel, but there is still so much more to learn from him.

There is no way I can fully repay the unconditional love and affection showered by parents. I am also blessed with caring cousins, uncles, aunts and inlaws who always encouraged all my endeavors. I hope I have done them justice with my research and this dissertation.

I have made a lot of friends in Raleigh and cherished some of my best moments with them here. Most notably, Steve Edwards, Cody Burnet, Issac Kumar, Brittany Jago, John and Sharyn, Andy and Cheryl, Karl and Jo-Ann, Siji and Mathew, Chase and Laura, Tiffany and Jay, Sammy and Debby, Anderson family, Ricky and Errica, Satyankar, Anna and Vasek, Tom George, Danielle, Shilpa, Neelesh Salian, Nilesh Rajule and Shital Shinde Rajule. I can't imagine my 3 years without you folks. And I hope we continue our good run in the future. I have a lot of fond memories of the times with my closest friends here, Steve, Cody, Neelesh and Issac. I really have learned a lot from them.

Finally, I would like to express my gratitude to National Science Foundation (Grant: CNS 1065458) for funding this work.

## TABLE OF CONTENTS

<b>LIST OF TABLES . . . . .</b>	<b>ix</b>
<b>LIST OF FIGURES . . . . .</b>	<b>x</b>
<b>CHAPTER 1 Introduction . . . . .</b>	<b>1</b>
1.1 Technology roadmap . . . . .	1
1.2 DRAM . . . . .	3
1.3 SRAM . . . . .	4
1.4 Flash memories . . . . .	6
1.4.1 NAND flash . . . . .	7
1.4.2 NOR flash . . . . .	7
1.5 Resistive RAM (or RRAM) . . . . .	9
1.6 Phase Change Memory (or PCM) . . . . .	10
1.7 Spin Transfer Torque Magnetic RAM (or STT-MRAM) . . . . .	11
1.8 Other emerging memory devices . . . . .	12
1.9 Comparison of different memory technology . . . . .	13
1.10 Dissertation overview . . . . .	13
1.11 Novel contributions . . . . .	15
1.12 Summary . . . . .	15
<b>REFERENCES . . . . .</b>	<b>16</b>
<b>CHAPTER 2 Experimental Methods . . . . .</b>	<b>20</b>
2.1 Primary fabrication tools used . . . . .	20
2.1.1 Atomic layer deposition . . . . .	20
2.1.2 UHV RF magnetron sputtering . . . . .	22
2.2 Analytical characterization methods . . . . .	24
2.2.1 X-ray photoelectron spectroscopy . . . . .	24
2.2.2 X-ray diffraction analysis . . . . .	26
2.3 Electrical characterization methods . . . . .	27
2.3.1 Capacitance-Voltage measurements . . . . .	27
2.3.2 Current-Voltage measurements of DFGFET . . . . .	28
2.3.3 Current-Voltage measurements of RRAM . . . . .	30
2.4 Summary . . . . .	30
<b>REFERENCES . . . . .</b>	<b>31</b>
<b>CHAPTER 3 Dual Floating Gate Memory MOSFET . . . . .</b>	<b>33</b>
3.1 Theoretical understanding of the dual floating gate memory MOSFET . . . . .	35
3.1.1 Device operation . . . . .	35
3.1.2 Theoretical formulation . . . . .	36
3.2 CV analysis of the DFGFET device . . . . .	41
3.3 Volatile operation of the DFGFET device . . . . .	42

3.4	Non-volatile operation of the DFGFET device . . . . .	43
3.4.1	Reliability analysis of the non-volatile operation in DFGFET. . . . .	45
3.5	Concurrent mode of operation in DFGFET . . . . .	46
3.6	Challenges of DFGFET . . . . .	46
3.7	Summary . . . . .	47
	<b>REFERENCES . . . . .</b>	<b>48</b>
	<b>CHAPTER 4 Engineering the Forming and Reset Mechanism in RRAM . . . . .</b>	<b>51</b>
4.1	Introduction to the memory operations in RRAM . . . . .	52
4.1.1	Forming . . . . .	52
4.1.2	Reset . . . . .	52
4.1.3	Set . . . . .	53
4.2	Understanding the forming process in $\text{Al}_2\text{O}_3/\text{HfO}_2$ bilayer dielectric RRAM . . . . .	54
4.2.1	Contribution of electric field inside the individual layer of bilayer dielectric RRAM . . . . .	55
4.2.2	Poole-Frenkel analysis of the forming currents . . . . .	56
4.2.3	Electric field enhancement in bilayer dielectric RRAM . . . . .	57
4.3	Comparison of reset in $\text{Al}_2\text{O}_3/\text{HfO}_2$ RRAM with $\text{HfO}_2$ RRAM . . . . .	59
4.4	Summary . . . . .	60
	<b>REFERENCES . . . . .</b>	<b>61</b>
	<b>CHAPTER 5 Synaptic Learning in Binary Oxide RRAM . . . . .</b>	<b>66</b>
5.1	Necessity of neuromorphic computation system . . . . .	66
5.1.1	Spike time dependent plasticity . . . . .	67
5.1.2	Why RRAM ? . . . . .	68
5.2	Motivation of work . . . . .	69
5.3	Understanding the gradual reset in RRAM . . . . .	69
5.4	Realizing the gradual reset in RRAM . . . . .	72
5.5	Comparison of conductance modulation in $\text{Al}_2\text{O}_3$ and $\text{HfO}_2$ RRAM . . . . .	73
5.6	Reliability analysis of the gradual reset . . . . .	75
5.7	Implementation of synaptic learning . . . . .	76
5.8	Summary . . . . .	77
	<b>REFERENCES . . . . .</b>	<b>79</b>
	<b>CHAPTER 6 Volatile Memory Operation Using RRAM . . . . .</b>	<b>83</b>
6.1	Introduction to the non-filamentary RRAM . . . . .	83
6.2	Volatile memory operation in $\text{Ni}/\text{TiO}_2/\text{Ni}$ RRAM . . . . .	85
6.3	Concept of novel 1T-0C ZRAM based on non-filamentary RRAM . . . . .	88
6.3.1	Theoretical understanding of the proposed 1T-0C ZRAM . . . . .	88
6.3.2	Volatile memory transition in $\text{SiO}_2/\text{TiO}_2$ MOS capacitor . . . . .	90
6.4	Summary . . . . .	92
	<b>REFERENCES . . . . .</b>	<b>93</b>

<b>CHAPTER 7 Conclusion and Future Work</b>	<b>96</b>
7.1 Conclusion	96
7.2 Future work	97
7.2.1 Hybrid flash-RRAM unified memory	97
7.2.2 1R-1R crossbar unified memory	100
7.3 Summary of the future work	102
<b>APPENDICES</b>	<b>103</b>
APPENDIX A Fabrication procedures	104
A.1 Fabrication procedures of DFGFET	104
A.2 Fabrication procedures of RRAM	107
APPENDIX B Finding the zero-field activation energy of Al <sub>2</sub> O <sub>3</sub> and HfO <sub>2</sub> dielectric	110

## LIST OF TABLES

Table 1.1	Comparison of different emerging memory technology candidates.	13
Table 2.1	Different ALD recipe.	22
Table 4.1	Parameters of RRAM dielectric for vacancy simulation.	56

## LIST OF FIGURES

Figure 1.1	(a) Categorical representation of different memory devices, and (b) memory hierarchy diagram taken from [8]. . . . .	2
Figure 1.2	Operation of DRAM: (a) program operation, and (b) erase operation. . . . .	3
Figure 1.3	(a) Schematic representation of SRAM cell, and (b) circuit representation of 4T-SRAM cell. . . . .	4
Figure 1.4	Operation of flash memory: (a) program operation, (b) erase operation, and (c) schematic of I-V curves of the memory after program and erase operation. . . . .	5
Figure 1.5	Schematic of flash arrangement in (a) NAND configuration, and (b) NOR configuration [15]. . . . .	6
Figure 1.6	Trend of (a) NAND scaling, and (b) DRAM scaling [20]. . . . .	8
Figure 1.7	Schematic of RRAM and its operation; (a) device structure of RRAM, (b) forming operation, (c) reset operation and (d) set operation. . . . .	9
Figure 1.8	Schematic of (a) device structure of PCM, and (b) memory operation in PCM taken from [26]. . . . .	10
Figure 1.9	(a) Device structure of STT-MRAM, (b) HRS state of STT-MRAM and (c) LRS state of STT-MRAM [27]. . . . .	11
Figure 2.1	Schematic of ALD system. . . . .	21
Figure 2.2	Schematic of UHV-RF sputter system. . . . .	23
Figure 2.3	Schematic of the X-ray photoelectron spectroscopy method. . . . .	25
Figure 2.4	Schematic of the X-ray diffraction method. . . . .	26
Figure 2.5	(a) Band diagram of MOS capacitor and (b) CV analysis of MOS capacitor. . . . .	28
Figure 2.6	Typical $I_D$ - $V_G$ characteristics of MOSFET and extraction of $V_T$ . . . . .	29
Figure 3.1	Schematic representation of the dual floating gate memory MOSFET. . . . .	34
Figure 3.2	Band diagram representation of the operation modes in DFGFET. . . . .	36
Figure 3.3	Simulated $V_{FB}$ shift because of volatile operation in DFGFET [19] (considering EOT of the entire gate stack to be $T_{ox}$ , EOT of the TO to be $t$ , EOT of the IPD to be $t_i$ and EOT of the IMD to be $\Delta t$ ). . . . .	38
Figure 3.4	CV hysteresis characteristics of the DFGFET memory device. . . . .	39
Figure 3.5	Variation in $V_{FB}$ during forward and reverse sweep in DFGFET memory device [20]. . . . .	40
Figure 3.6	Volatile retention measurement performed in DFGFET: (a) Pulse CV scheme, and (b) capacitance at different sample points. . . . .	42
Figure 3.7	Non-volatile program/erase window of the DFGFET. . . . .	43
Figure 3.8	Non-volatile (a) retention and (b) endurance characteristics of the DFGFET. . . . .	44
Figure 3.9	Demonstration of concurrent mode with (a) IV analysis, and (b) CV analysis. . . . .	45

Figure 4.1	(a) Forming characteristics of 5.3 nm HfO <sub>2</sub> and 5.3 nm HfO <sub>2</sub> + 1.1 nm Al <sub>2</sub> O <sub>3</sub> RRAM; simulated (b) electric field and (c) vacancy generation rate inside individual layers of Al <sub>2</sub> O <sub>3</sub> /HfO <sub>2</sub> bilayer dielectric RRAM; (d) forming voltage uniformity in HfO <sub>2</sub> RRAM and Al <sub>2</sub> O <sub>3</sub> /HfO <sub>2</sub> bilayer dielectric RRAM . . . . .	54
Figure 4.2	Trap assisted Poole-Frenkel analysis of the forming currents shown in Figure 4.1 . . . . .	57
Figure 4.3	Electric field distribution inside the dielectric at 2 V; (a) 5.3 nm HfO <sub>2</sub> RRAM with a filament formed near the bottom electrode; (b) filament formed in 5.3 nm HfO <sub>2</sub> + 1.1 nm Al <sub>2</sub> O <sub>3</sub> RRAM near the bottom electrode; and (c) electric field along the vertical axis at the center of the filaments . . . . .	58
Figure 4.4	(a) Reset comparison between HfO <sub>2</sub> RRAM and Al <sub>2</sub> O <sub>3</sub> /HfO <sub>2</sub> bilayer dielectric RRAM; (b) band diagram showing the advantage of Al <sub>2</sub> O <sub>3</sub> /HfO <sub>2</sub> bilayer dielectric RRAM over HfO <sub>2</sub> RRAM. . . . .	59
Figure 5.1	Illustration of a synaptic system and an electronic device that can mimic the synapse. . . . .	67
Figure 5.2	(a) Forming, (b) reset after forming, and (c) two distinct set-reset behavior observed in Al <sub>2</sub> O <sub>3</sub> RRAM. . . . .	70
Figure 5.3	(a) Multi-step forming in Al <sub>2</sub> O <sub>3</sub> RRAM; (b) possible vacancy profile after individual forming steps. . . . .	71
Figure 5.4	Reset after the multi-step forming in Al <sub>2</sub> O <sub>3</sub> RRAM. . . . .	72
Figure 5.5	(a) Multi-step forming in HfO <sub>2</sub> RRAM; (b) corresponding reset. . . . .	73
Figure 5.6	Multi-state memory window achieved with gradual reset in (a) Al <sub>2</sub> O <sub>3</sub> RRAM, and (b) HfO <sub>2</sub> RRAM. . . . .	74
Figure 5.7	Comparison of gradual set-reset behavior in (a) Al <sub>2</sub> O <sub>3</sub> RRAM, and (b) HfO <sub>2</sub> RRAM. . . . .	75
Figure 5.8	Conductance modulation comparison of (a) Al <sub>2</sub> O <sub>3</sub> RRAM and (b) HfO <sub>2</sub> RRAM using DC/AC reset techniques. . . . .	76
Figure 5.9	(a) and (b) Multi-state retention behavior of Al <sub>2</sub> O <sub>3</sub> and HfO <sub>2</sub> RRAM respectively; (c) and (d) Endurance behavior of Al <sub>2</sub> O <sub>3</sub> and HfO <sub>2</sub> RRAM respectively . . . . .	77
Figure 5.10	Synaptic learning behavior of Al <sub>2</sub> O <sub>3</sub> and HfO <sub>2</sub> RRAM. . . . .	78
Figure 6.1	Schematic of RRAM circuit: (a) 1D-1R configuration, and (b) 1T-1R configuration. . . . .	84
Figure 6.2	Material characterization of TiO <sub>2</sub> dielectric; XPS analysis showing (a) Ti 2P spectra and (b) O 1s spectra; and (c) XRD analysis of TiO <sub>2</sub> showing the anatase phase. . . . .	86
Figure 6.3	Volatile memory operation obtained in Ni/TiO <sub>2</sub> /Ni RRAM using IV measurement. . . . .	87
Figure 6.4	(a) Schematic of the proposed non-filamentary RRAM based 1T-0C ZRAM device; (b) circuit equivalent of the 1T-0C ZRAM device. . . . .	88
Figure 6.5	Comparison of HFCV characteristics between SiO <sub>2</sub> MOS capacitor and SiO <sub>2</sub> /TiO <sub>2</sub> MOS capacitor. . . . .	90

Figure 6.6	(a) Pulse CV measurement scheme to analyze volatile memory operation in $\text{SiO}_2/\text{TiO}_2$ MOS capacitor; (b) volatile retention in $\text{SiO}_2/\text{TiO}_2$ MOS capacitor analyzed with the pulsed CV measurement. . . . .	91
Figure 7.1	(a) Schematic of a hybrid flash-RRAM unified memory; (b) volatile operation, (c) non-volatile operation and (d) concurrent operation of the hybrid flash-RRAM unified memory. . . . .	98
Figure 7.2	(a) Schematic of a crossbar unified memory; (b) circuit arrangement of the crossbar structure and (c) different memory states corresponding to the resistance state of the non-volatile RRAM and the vacancy modulation oxide RRAM. . . . .	100
Figure A.1	Process flow for fabricating DFGFET (step-1 to step-6) . . . . .	105
Figure A.2	Process flow for fabricating DFGFET (step-7 to step-10) . . . . .	106
Figure A.3	Process flow for fabricating DFGFET (step-11 to step-12) . . . . .	106
Figure A.4	Final structure of DFGFET. Transistor had channel length of $10 \mu\text{m}$ , S/D doping density of $2 \times 10^{21} \text{ cm}^{-3}$ . Stack thinkesses are; TO: 7 nm, BFG: 4 nm, IMD: 4.5 nm, TFG: 4 nm, IPD: 33 nm, CG: 100 nm. . . . .	107
Figure A.5	Process flow for the fabrication of RRAM devices. . . . .	108
Figure B.1	(a)-(b) Weibull distribution of TDDB performed on $\text{Al}_2\text{O}_3$ and $HfO_2$ dielectric RRAM respectively. . . . .	111
Figure B.2	(a)-(b) TDDB vs $1/KT$ plot giving rise to a slope of $E_{a,eff}$ at different fields for $\text{Al}_2\text{O}_3$ and $HfO_2$ dielectric RRAM respectively. . . . .	112
Figure B.3	$E_{a,eff}$ vs electric field plot for $\text{Al}_2\text{O}_3$ and $HfO_2$ RRAM. . . . .	112

# Chapter 1

## Introduction

### 1.1 Technology roadmap

Semiconductor devices have been following the Moore's law for the last four decades resulting in a continuous scaling of devices to improve the performance and lower the cost per transistor [1]. To meet the aggressive scaling requirements, several technologies have been invented and added to the conventional CMOS logic [2]. For example, strained Si was introduced in the 90 nm node [3], high-k/metal gate was introduced in the 45 nm node [4] and FINFET was introduced at the 22nm node [5]. All these innovations have resulted in a faster computing capability in modern smart electronic gadgets. Another important component of the smart electronic gadgets is the memory. Similar to the CMOS logic technology, memory technology has also gone through an aggressive scaling over several decades to improve the storage capacity, latency, circuit area, etc [6, 7]. Thus, proposing novel ideas to meet the scaling requirements of the future memory technology devices is of great importance.

Figure 1.1 (a) shows the categorical arrangement of different memory devices. Broadly speaking, the memory devices can be categorized as volatile memory and non-volatile memory [8]. Volatile memories are capable of fast operation, but loses data as soon as the device gets turned off. On the other hand, non-volatile memories show relatively slower speed of operation than the volatile memories, but retains data even after the device is turned off. Thus, there is a trade-off, fast memories store temporary data, whereas permanent memories are slower in speed. Both the kind of memories are heavily used in smart electronic devices such as smartphones, laptops, tablets, etc. These devices use the volatile memories for implementing logic operation, whereas non-volatile memories are used to store operating system codes, multimedia data, etc. Among volatile memories, dynamic random access memories (or DRAMs) are made up of 1 transistor - 1 capacitor (1T-1C), whereas static random access memories (or SRAMs)

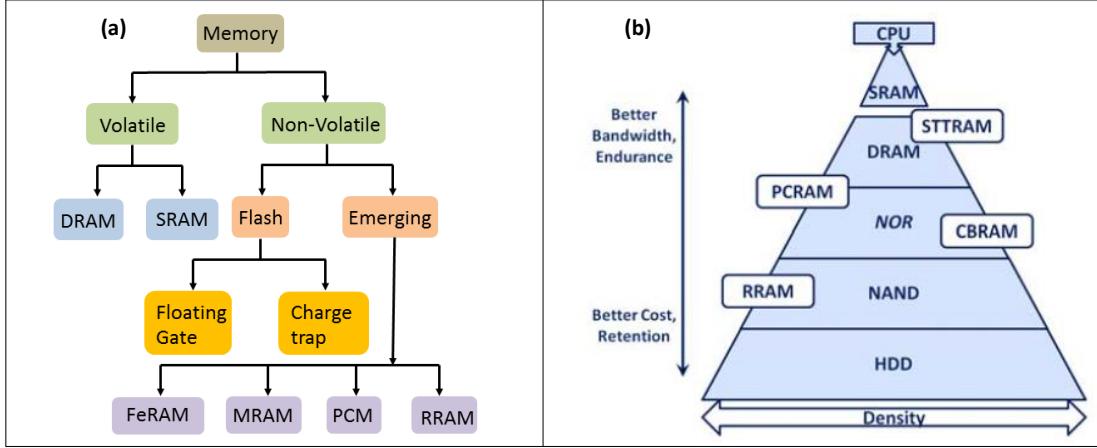


Figure 1.1: (a) Categorical representation of different memory devices, and (b) memory hierarchy diagram taken from [8].

are traditionally made with 6 transistors. Thus, DRAMs are relatively simpler to fabricate. However, DRAMs need refresh in regular interval for retaining the data. SRAMs, on the other hand, are fast accessible volatile storage which doesn't require regular refresh. However, SRAMs occupy a larger area on the chip. Going to the non-volatile storage, flash memories have been the core memory element of non-volatile memories used in solid state drives (SSDs), flash drives, etc. They are broadly categorized into NAND flash and NOR flash depending upon the configuration they are arranged on the circuit. Both these configurations are used to store permanent data, but they lack in processing speed when compared with volatile memories. Note that NOR flash is faster than NAND flash.

Figure 1.1 (b) shows the memory hierarchy diagram. Hard disc drives (HDDs) which has been widely used so far are made up of magnetic cores. They provide highest memory density, but they show very slow operation speed. With advancement in SSD technology, a big share of the permanent storage got acquired by the NAND flash technology. Similarly, NOR provides faster operation than NAND, but provides lower memory density. Finally, volatile memories viz. DRAM and SRAM are placed on the top of the diagram because of their very fast operation. All these memories are used in different applications depending upon factors like speed, endurance, retention, etc. The hierarchy diagram also shows the possible replacement technology for each memory type. A detailed description of all these memories is mentioned in the successive sections.

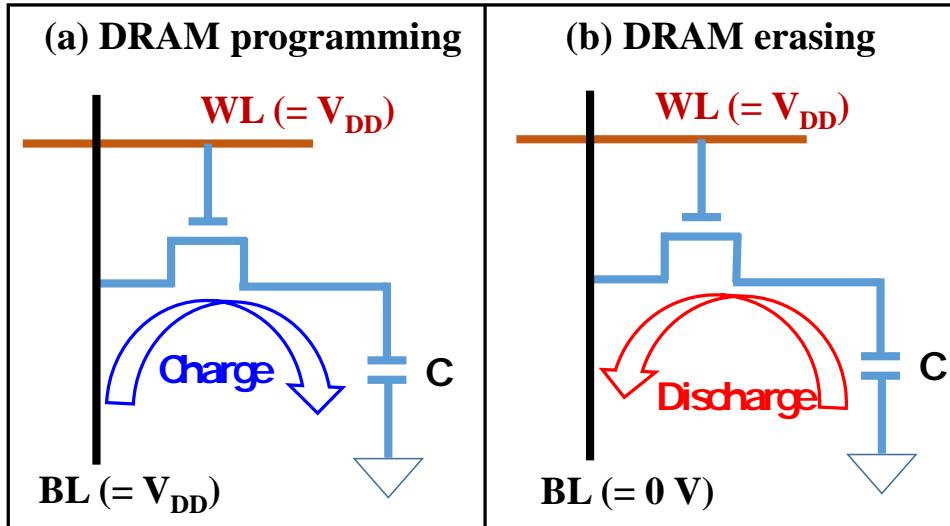


Figure 1.2: Operation of DRAM: (a) program operation, and (b) erase operation.

## 1.2 DRAM

DRAM constitutes the volatile memory element of almost all modern electronic devices. A typical DRAM cell is shown in Figure 1.2. The device consists of a transistor and a capacitor (1T-1C configuration) with a wordline (WL) connected to the gate of the transistor and a bitline (BL) connected to the drain of the transistor. Source of the transistor is connected to the capacitor. Charge content of the capacitor determines the memory states. During programming, WL and BL are biased at  $V_{DD}$ . WL voltage turns on the transistor, and the BL voltage results in charging the capacitor upto  $V_{DD} - V_T$ . Similarly, during erase operation, WL is again biased at  $V_{DD}$  to turn on the transistor, whereas BL is kept at 0 V in order to discharge the capacitor.

The memory read operation is performed by precharging the BL to  $V_{DD}/2$ , and then applying  $V_{DD}$  bias at the WL. If the DRAM is at programmed state, the capacitor will discharge through the BL resulting in an increase in the BL voltage. On the other hand, if the capacitor was in the erased state, BL will gradually charge the capacitor and hence the BL voltage will drop below  $V_{DD}/2$ . A sense amplifier can be used to sense the increase or decrease in the BL voltage thereby reading the memory state. Note that the read disturbs the charge content of the capacitor, hence it is called as a destructive read. For this reason, the memory state of the DRAM is restored to its original state after every read. A detailed review on modern day DRAM technology can be found in [9].

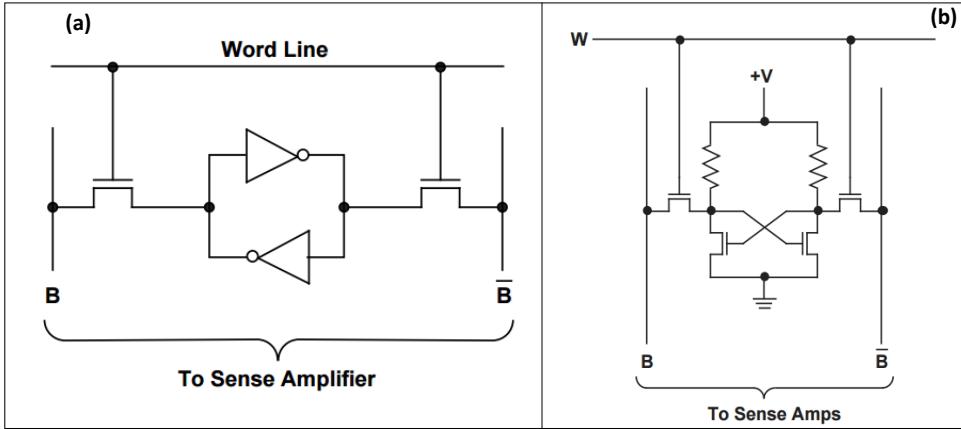


Figure 1.3: (a) Schematic representation of SRAM cell, and (b) circuit representation of 4T-SRAM cell.

Also, as mentioned earlier, current transistor technology suffers from numerous short channel effects. This results in a conduction of transistor even in the off state, and the capacitor charge gets a discharge path through the transistor. In modern day DRAM, the circuit requires a refresh after every 64 ms. Currently, a strong emphasis has been under progress to replace DRAM with one of the non-charge based emerging memory technologies. Spin Transfer Torque Magnetic RAM is among the favorable choice because of their very fast switching and unlimited endurance properties [10]. Spin based memories, however, are not compatible with the standard CMOS process, hence other emerging memories such as Phase Change Memories [11] and Resistive switching memories [12] are attracting interest for potential DRAM replacement.

### 1.3 SRAM

Figure 1.3 shows the schematic representation of a typical SRAM cell. SRAM cells are back to back connection of inverters with two access transistors [13] (as shown in Figure 1.3 (a)). SRAM provides a direct interface with the central processing unit (or CPU) at a higher speed than DRAM. For this reason, they serve as the cache memory of the CPU. Figure 1.3 (b) shows the circuit representation of the 4T-SRAM cell. More popular is the 6T-SRAM configuration where the inverters are realized using 4 transistors (resistors are replaced by pull-up transistors), and they behave like a bi-stable flip-flop. Similar to DRAM, SRAMs also retain data as long as they are connected to the power supply. Unlike DRAM, the data do not leak away in SRAM. For this reason, SRAM do not require periodic refresh making them favorable to replace DRAM in systems that require very low power consumption.

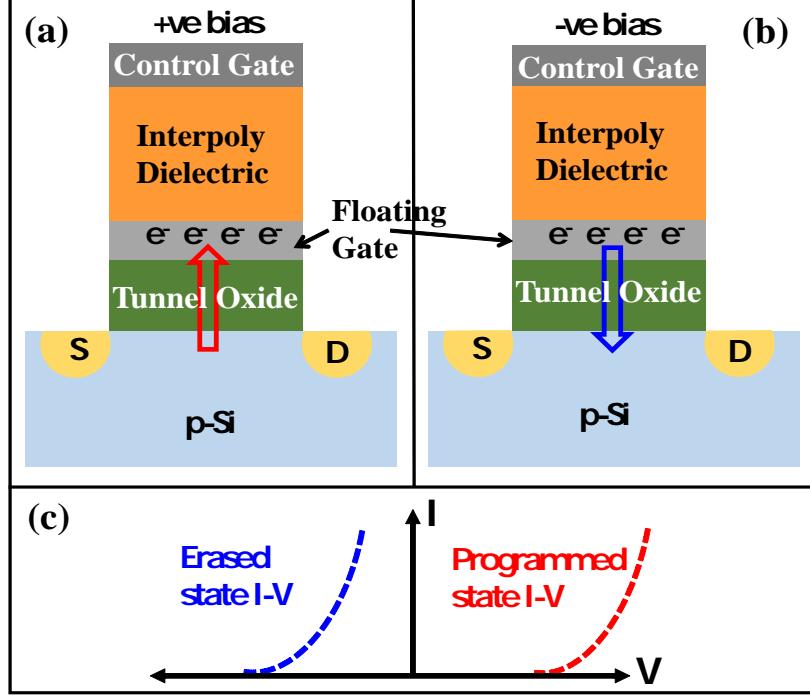


Figure 1.4: Operation of flash memory: (a) program operation, (b) erase operation, and (c) schematic of I-V curves of the memory after program and erase operation.

To read an SRAM cell, the bitlines ( $B$  and  $\bar{B}$ ) are initially pre-charged to a high voltage level (or  $V_{DD}$ ). The word-line is then pulsed to  $V_{DD}$  thereby turning ON the access transistors. One of the bitline gets discharged through the inverter node of the cell storing logical 0. As a result, a differential voltage develops between the bitlines. This differential voltage is then sensed by a sense amplifier to detect the state of the cell. To write data in SRAM cell, the bitlines are initially driven to complementary voltage levels. The access transistors are then turned ON to write the data by providing bias at the wordline. The bitline which is kept at logical "0" allows the corresponding inverter node to discharge through the access transistor. Similarly, the bitline kept at  $V_{DD}$  charges-up the corresponding inverter node.

The primary concern of SRAM technology is the requirement of larger cell area. Although emerging memory devices are predicted to show very fast operation, but a candidate that can show high speed and low power operation like SRAM cell is still under research. More details on SRAM circuit and their scaling challenges can be found in [14].

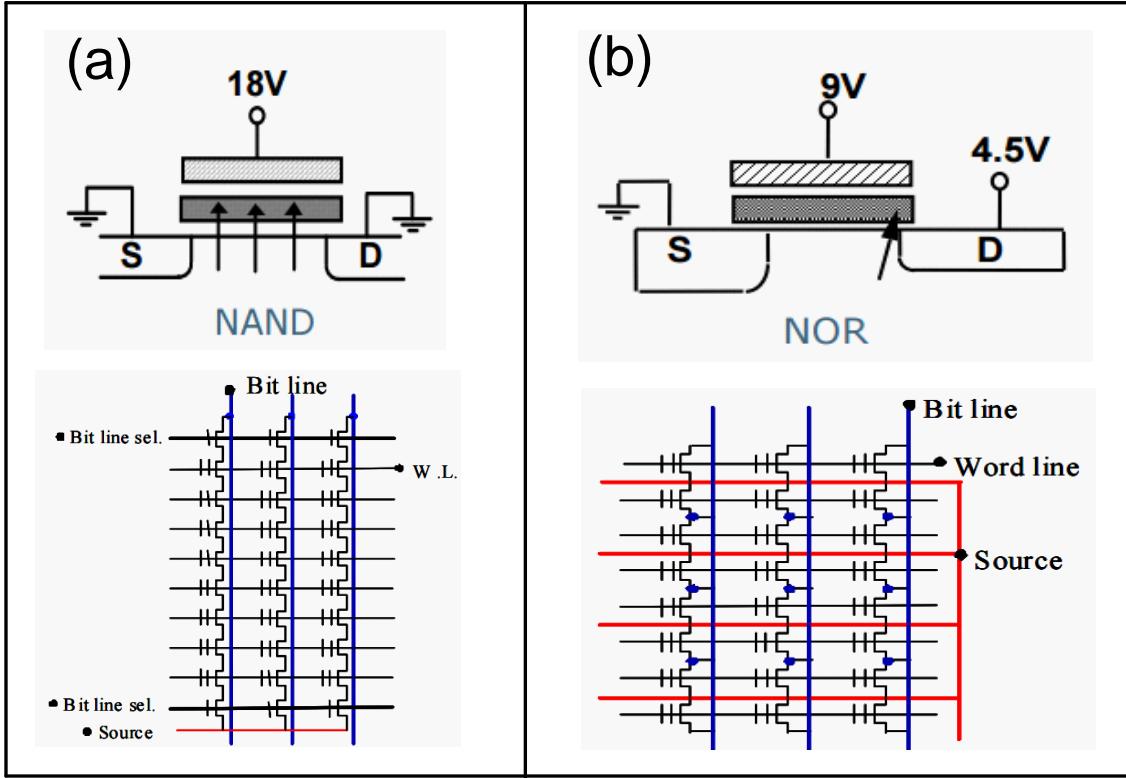


Figure 1.5: Schematic of flash arrangement in (a) NAND configuration, and (b) NOR configuration [15].

## 1.4 Flash memories

Flash memories are the core element of modern day SSDs [16]. These memories operate on the principle of storing electrons in a floating gate or charge-trap layer thereby realizing the programmed state of the device. These electrons can be taken back to the substrate with an erase operation. Figure 1.4 shows the device operation mechanisms of a typical floating gate type flash memory. The device consists of a transistor with a tunnel oxide, a floating gate serving as the charge storage layer, an interpoly dielectric (IPD) and a control gate. Instead of floating gate, a charge trap layer such as Nitride can also be used to store the electrons.

During programming, electrons are transferred from the substrate to the floating gate or nitride layer thereby increasing the threshold voltage of the transistor (shown in Figure 1.4 (a)). Similarly, an erase operation tunnels the electrons back to the substrate, thereby reducing the threshold voltage or  $V_T$  (shown in Figure 1.4 (b)). Figure 1.4 (c) the schematic of drain current vs gate voltage plot after programming and erasing operation. The programmed state can be

considered to be binary-1 state, whereas erased state can be considered to be binary-0 state. With adequate control the number of electrons tunneled into the floating gate, intermediate states can be achieved, resulting in a multi-bit memory device. Once the charge is transferred into the floating gate, it should remain at the floating gate for 10 years, hence the tunnel oxide is chosen thick enough to prevent the charge loss after program operation. Similarly, to avoid the charge transfer between the control gate and the floating gate, very thick IPD are chosen. The flash cells can be arranged in either NAND or NOR configuration.

#### 1.4.1 NAND flash

In NAND architecture, flash cells are connected in series, as shown in Figure 1.5 (a). During programming, a high +ve bias is applied to the wordline keeping the bitline grounded. The high bias at wordline (or at the control gate) results in a tunneling of electrons from the substrate to the floating gate. During erase, a high -ve bias is applied to the control gate which results in tunneling the electrons back to the substrate, as shown in Figure 1.4 (b). This operation reduces the  $V_T$  of the transistor. Following are the key points about NAND flash:

- Program and erase operations are achieved with Fowler-Nordheim (FN) tunneling.
- Provides higher bit density than NOR.
- cells are read sequentially, thus access time is much higher than NOR.  
item Block erase: since all the cells share the same substrate, erase is achieved by applying bias at the substrate. NAND erase is faster than NOR.
- Ideal usage: PC flash, secure digital media, flash drives, SSDs, etc.

Current NAND technology is based on 3D architecture where several layers of flash cells are arranged in a vertical trench [17]. Although this architecture provides significant area scaling advantage, but the scaling of flash memories rises several concerns. Figure 1.6 (a) show the NAND memory density trend [18]. The multi-level operation is achieved by setting different  $V_T$  levels with appropriate control of number of electrons in the floating gate. With scaling, the number of electrons involved in flash operation has reduced significantly. Below 22 nm node, loss of just few electrons from the floating gate can change the  $V_T$  by more than 100 mV. Thus, loss of just few electrons can completely destroy the stored data.

#### 1.4.2 NOR flash

In NOR arrangement, the cells are connected parallel to each other. Source of every flash cells are connected to ground, and the drains are connected to bitlines. Thus, NOR flash cells can be

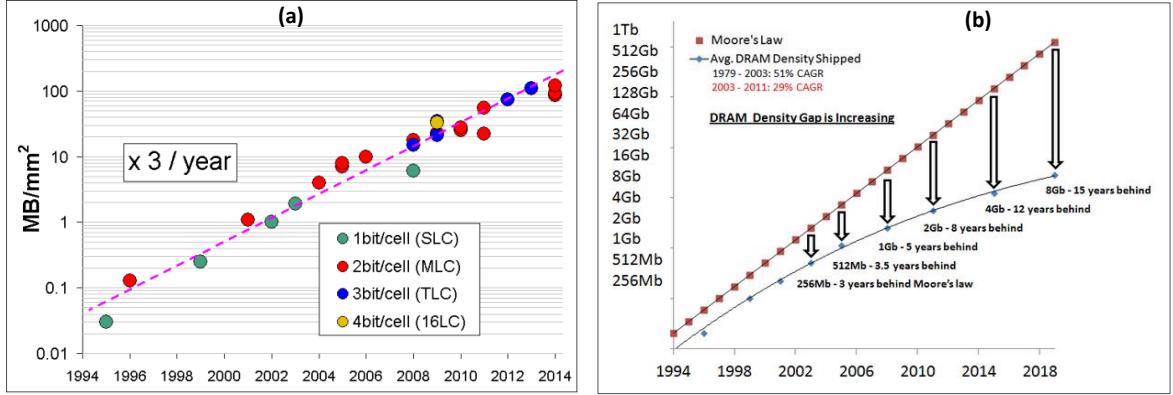


Figure 1.6: Trend of (a) NAND scaling, and (b) DRAM scaling [20].

accessed randomly making them attractive for fast memory applications. During programming, bias is applied to both bitline and wordline resulting in generation of hot electrons which gets injected into the floating gate. For this reason, the voltage requirement of NOR flash is lower than NAND cell. Schematic of NOR cell arrangement is shown in Figure 1.5 (b). Some of the key points of NOR flash are:

- Programming: Channel Hot Electron (CHE) injection is used. Erasing: achieved with Fowler-Nordheim (FN) tunneling.
- Provides lower bit density because of metal connectivity required to source and drain of every transistor.
- Provides random access, thus access time is lower than NAND.
- Ideal usage: Code storage in PDAs and cell phones, embedded systems, etc.

NOR technology is still in 2D architecture. Because of CHE injection which causes higher damage to the tunnel oxide compared to FN tunneling, NOR flash has been facing severe scaling issues below 45 nm node. A review on flash storage prospects and challenges can be found in [2, 18]. Since NAND and NOR have pros and cons, a storage class memory is in great demand which can provide high bit density, faster access and program/erase time, and should show excellent endurance and retention properties. Compatibility with the conventional CMOS process is added advantage. Several emerging memory candidates are attracting great research interest to implement storage class memory [35]. Some of these emerging memory devices are discussed in successive sections.

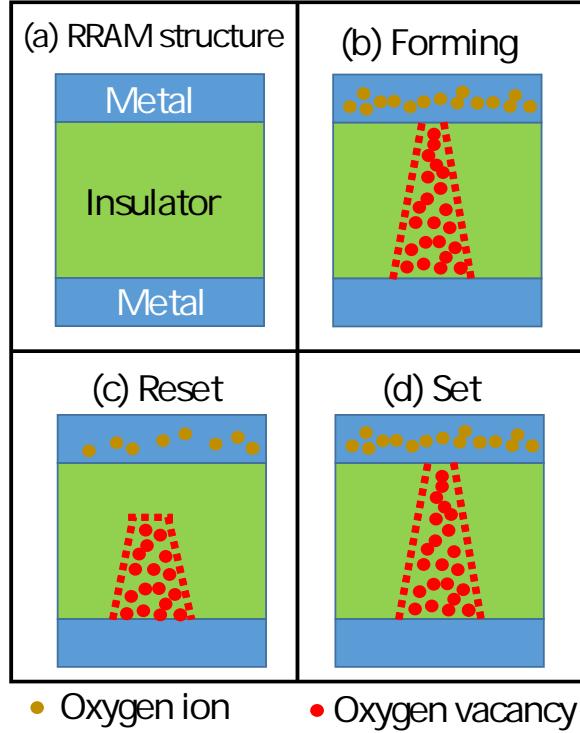


Figure 1.7: Schematic of RRAM and its operation; (a) device structure of RRAM, (b) forming operation, (c) reset operation and (d) set operation.

## 1.5 Resistive RAM (or RRAM)

RRAM has been attracting research interest for its tremendous potential for replacing the NAND cells for future memory technology. With the configuration of cross-bar array, they can achieve cell size lower than  $4F^2$  thereby allowing aggressive area scaling possibilities. However, physical operation principles of these memories are still under research phase. Several models of operation has been proposed [21, 22, 23], but an accurate model to describe different RRAM operations observed in different RRAM devices is still missing.

RRAM devices are simple metal-insulator-metal capacitors, as shown in Figure 1.7 (a). The device needs a forming step to create a conductive filament in the insulator. Filament is created by applying bias at one of the electrode thereby generating oxygen vacancies in the insulator. At certain bias voltage, these vacancies forms a filament between the electrode resulting in a low resistance state (or LRS) of the device, shown in Figure 1.7 (b). This filament can be ruptured by applying bias of either same polarity (corresponds to the unipolar operation [24]) or opposite polarity (corresponds to the bipolar operation [25]). this process is called the reset

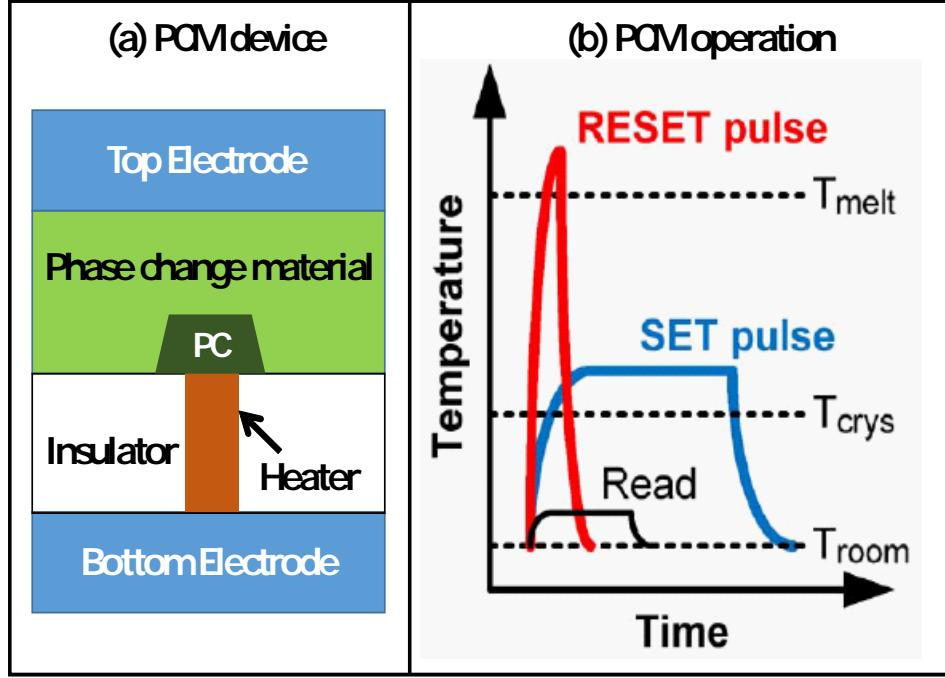


Figure 1.8: Schematic of (a) device structure of PCM, and (b) memory operation in PCM taken from [26].

operation which brings the insulator into high resistance state (HRS), shown in Figure 1.7 (c). Similarly, the ruptured part of the filament can be formed again with a set operation thereby bringing the LRS, shown in Figure 1.7 (d). Mode details of the RRAM and its operation are mentioned in Chapter-4 of the Dissertation.

## 1.6 Phase Change Memory (or PCM)

Phase change memory is a matured emerging memory candidate that is expected to replace flash memories in NAND and NOR architecture [26]. These memories work on the principle of phase transformation of the chalcogenide materials. Figure 1.8 (a) shows a typical configuration of PCM cell. A heater is connected to one of the electrodes connecting it to the phase change material. As deposited PCM cell shows the LRS behavior because of the crystalline phase of the chalcogenide material. The operation scheme of PCM is shown in Figure 1.8 (b). To bring the cell into HRS, a voltage pulse is applied so that the temperature goes higher than 600 C followed a rapid cooling. The HRS state is achieved with a narrow reset pulse of higher voltage amplitude. This results in the amorphous state of the chalcogenide material with high resistance. Consequently, the device can be brought back to the LRS state by applying a longer

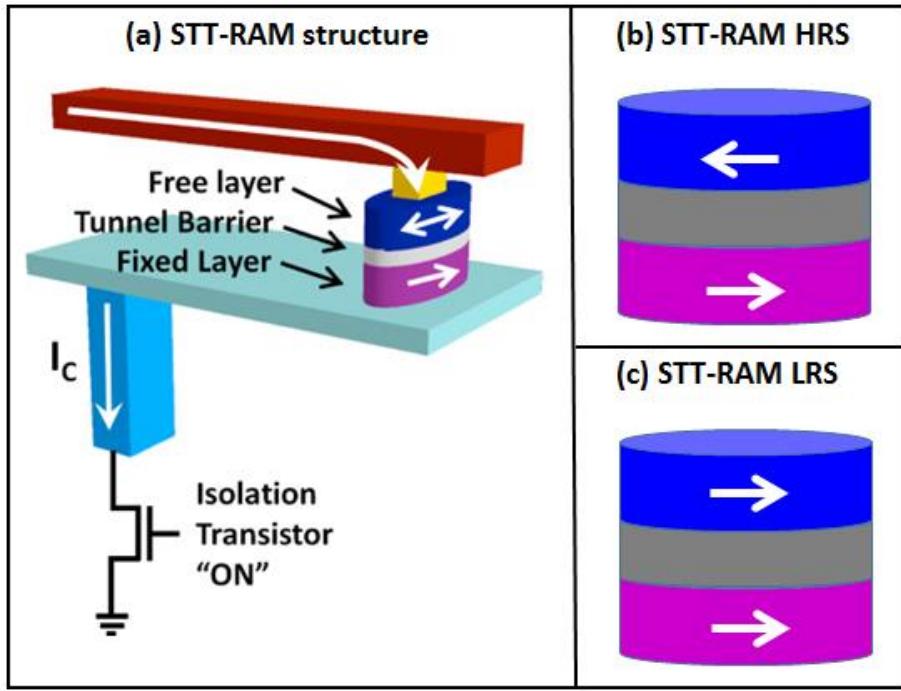


Figure 1.9: (a) Device structure of STT-MRAM, (b) HRS state of STT-MRAM and (c) LRS state of STT-MRAM [27].

set pulse allowing the chalcogenide to attain its crystalline state. Successive set and reset pulses are used to operate the PCM cell.

PCM has already been proposed to replace the flash in NOR architecture. Fast switching speed and excellent endurance behavior of PCM are also attractive for potential DRAM replacement. However, the major concerns for PCM devices are higher cell size and inability to integrate in 3D stacks. More details of PCM devices can be found in [26].

## 1.7 Spin Transfer Torque Magnetic RAM (or STT-MRAM)

STT-MRAM has been a topic of research for several decades. These memories work on the principle of changing the magnetization state of magnetic material with a high current pulse. A typical STT-MRAM device structure and its operation is shown in Figure 1.9. The device consists of two magnetic layers separated by a non-magnetic material serving as the tunnel barrier, shown in Figure 1.9 (a). Magnetization of one of the layer is kept fixed, hence called as fixed layer. The magnetization state of the other layer can be altered by the application

of a high current pulse. The high current passed from fixed layer into the free layer carries spin momentum which can switch the magnetization state of the free layer. Similarly, opposite polarity current pulse can reverse the magnetization state of the free layer. When two layers have similar magnetization states, the device attains its LRS state. On the other hand, if the layers have opposite magnetization states, this result in the HRS state of the device. Schematic representation of the HRS and LRS state of STT-MRAM are shown in Figure 1.9 (b) and Figure 1.9 (c).

Note that the materials used in STT-MRAM are ferromagnetic materials which are not compatible to the standard CMOS process. Moreover, a complete STT-MRAM device requires various metal layers to improve the ferromagnetic coupling and to fix the magnetization of the fixed layer. These are the prime concerns of the STT-MRAM. Another concern of STT-MRAM is the requirement of very high current to switch the magnetization of the free layer. Thus, STT-MRAMs show high power memory operation. However, STT-MRAMs can be programmed/erased within few tens of nanosecond, and they offer unlimited endurance (more than  $10^{16}$ ). These two crucial properties of STT-MRAM makes them the most suitable candidate for DRAM replacement at present. More details of STT-MRAM can be found in [27].

## 1.8 Other emerging memory devices

The demand of meeting scaling laws in memory technology has resulted in several device concepts, both for DRAM and NAND replacement. Future research of all these emerging memory technologies is expected to result in a memory candidate that can satisfy all the criteria of a storage class memory [36].

Conductive Bridge RAM (or CBRAM) is very similar to the RRAM with an active electrolyte serving as the resistive switching layer, and the filamentary operation is controlled by an electrochemically active electrode and an inert electrode. One example is Pt/Ag<sub>2</sub>S/Ag where the filament is formed in the Ag<sub>2</sub>S electrolyte [21]. Similarly, molecular electronics have also been explored to make them applicable in memory technology [28]. Ferroelectric non-volatile memories are already in production [29]. However, ferroelectric memories show destructive read property which is a critical challenge. Some of the transparent oxides have also been observed to show memory like behavior making them attractive for transparent flexible electronic applications [30]. Likewise, mechanical switch as memory element has also been explored by researchers [31]. An overview of different emerging memories can be found in [8].

## 1.9 Comparison of different memory technology

Emerging memory candidates seem to be promising in resolving some of the challenges faced by current flash and DRAM devices. PCM has been predicted to replace the flash memories in NOR cell below 45 nm node and below. RRAM provides very fast switching and cell size can be lower than  $4F^2$ . Thus, RRAM is expected to be a great competitor to NAND technology for non-volatile storage applications. STT-MRAM shows excellent endurance behavior with reasonable access time making them suitable for DRAM replacement. A comparison between all the promising memory technologies are presented in table 1.1 [33, 34, 35, 8].

Table 1.1: Comparison of different emerging memory technology candidates.

Parameter	DRAM	NAND	RRAM	PCM	STT-MRAM
Cell Size ( $F^2$ )	6-8	8	4	8-16	14-64
Write time	10-50 ns	0.1 ms	5-30 ns	20-30 ns	10 ns
Write energy	0.005 pJ	100 pJ	1 pJ	6 pJ	4 pJ
Endurance	$10^{16}$	$10^5$	$10^{10}$	$10^8$	$10^{15}$
Retention	64 ms	10 years	10 years	10 years	weeks
3D capability	via TSV	Yes	Yes	No	No
Process Complexity	High	High	Low	Low	High

## 1.10 Dissertation overview

There has been continuous scaling of memory technology, both for volatile and non-volatile applications resulting in newer devices which can meet the scaling requirements. Volatile memories, which are used for processing and executing the logical operations, needs to be fast and low power. On the other hand, non-volatile memories are used to store permanent data such as execution codes, look-up tables, etc. Current processing system have volatile and non-volatile memories taking different area of the chip. Hence, a significant time is elapsed during execution in order to access the codes from the non-volatile storage and then communicate the data to the volatile memories for logical operations. To improve performance, embedded memories have been proposed where volatile memories and non-volatile memories can be stacked together with the help of through silicon via (TSV). But a better performance can be achieved by realizing a unified memory which can show simultaneous volatile and non-volatile memory operation. Such a memory device can be very applicable in applications like instant ON computing, neuromor-

phic systems, sensor systems. Unified memory can also result in significant reduction in circuit area (mitigating the requirement of TSV), as well as provide better security.

The primary requirement of a unified memory is to provide two basic needs, (i) a fast non-volatile storage element that can be used to store codes, and (ii) a low power, very fast memory that can behave like a DRAM for executing the logical operations. In this Dissertation, novel concepts of unified memory has been discussed and demonstrated.

Chapter-3 of the Dissertation discusses the concept of dual floating gate flash memory which can resemble a simultaneous volatile and non-volatile operation. The addition of an extra floating gate into the flash device has been observed to provide additional volatile memory functionality. A complete description of the theoretical insight of the memory operation has been discussed in this chapter. Unified memory operation from a fabricated device is also discussed in chapter-3. However, emerging memories such as RRAM provides lower cell size and faster switching speed compared to NAND technology. Thus, achieving a unified memory operation with RRAM devices can provide better performance than flash memories. For this purpose, this dissertation also covers unified memory operation that can be achieved with RRAM technology.

Since RRAM is a relatively new technology, a significant emphasis has been put to understand RRAM device operation principles. This consists of understanding the non-volatile and volatile memory behavior of different RRAM candidates. Chapter-4 and chapter- 5 deals with the understanding of RRAM non-volatile operation. Understanding the forming operation of RRAM devices has been carried out in chapter-4 of the dissertation. A technique to reduce the forming voltage has also been explored in this chapter. Understanding of set and reset operation has been carried out in chapter-5 of the dissertation. A relation between abrupt and gradual reset with their corresponding set was observed which was further used to control the conductance change in the dielectric with a novel technique. This technique was further used to realize analog memory behavior in binary oxide RRAMs applicable for neuromorphic computation systems.

The RRAMs studied in chapter-3 and chapter-4 are filamentary RRAM which requires a large current for set and reset operations. However, the DRAM replacement technology needs to show low power memory operation. To achieve low power operation with RRAM, non-filamentary RRAM devices has been explored in chapter-6 of the dissertation. A novel 1T-0C ZRAM has also been explored in chapter-6. Following this, new unified memory device architectures based on flash and RRAM devices have been proposed in chapter-7 of the dissertation.

## **1.11 Novel contributions**

Chapter-3 of this dissertation describes about a dual floating gate unified memory MOSFET. This was the first demonstration of a transistor which can show simultaneous volatile and non-volatile operation. In the RRAM research, the technique of gradual filament dissolution has been adopted to implement synaptic learning. Such a gradual dissolution is a promising aspect of analog memory. The volatile memory behavior in RRAM has been used to realize a proof-of-concept 1T-0C ZRAM MOS capacitor device. In future work, the realization of 1T-0C ZRAM transistor can be a potential candidate for DRAM replacement.

## **1.12 Summary**

Hence, this dissertation gives an insight of a novel emerging memory technology devices that can be used to realize simultaneous volatile and non-volatile operation. The dual floating gate memory MOSFET is a novel unified memory device made with flash transistor. Similarly, storage class memory candidates such as RRAM can also be used to realize the unified memory operation. With further device engineering, these memory candidates can be very attractive to applications like embedded memory, neuromorphic system, sensor system, etc.

## REFERENCES

- [1] International technology roadmap for semiconductors: Emerging Research Devices 2013, available online at [www.itrs.net](http://www.itrs.net).
- [2] H. Iwai, "Technology roadmap for 22nm and beyond," *International Workshop on Electron Devices and Semiconductor Technology (IEDST)*, pp. 1-4, 2009.
- [3] S. E. Thompson, M. Armstrong, C. Auth, M. Alavi, M. Buehler, R. Chau, S. Cea, T. Ghani, G. Glass, T. Hoffman, C. H. Jan, C. Kenyon, J. Klaus, K. Kuhn, Z. Ma, B. Mcintyre, K. Mistry, A. Murthy, B. Obradovic, R. Nagisetty, P. Nguyen, S. Sivakumar, R. Shaheed, L. Shifren, B. Tufts, S. Tyagi, M. Bohr and Y. El-Mansy, "A 90-nm Logic Technology Featuring Strained-Silicon," *IEEE Transactions on Electron Devices*, vol. 51, no. 11, pp. 1790-1797, 2004.
- [4] C. Auth, A. Cappellani, J. S. Chun, A. Dalis, A. Davis, T. Ghani, G. Glass, T. Glassman, M. Harper, M. Hattendorf, P. Hentges, S. Jaloviar, S. Joshi, J. Klaus, K. Kuhn, D. Lavric, M. Lu, H. Mariappan, K. Mistry, B. Norris, N. Rahhal-orabi, P. Ranade, J. Sandford, L. Shifren, V. Souw, K. Tone, F. Tambwe, A. Thompson, D. Towner, T. Troeger, P. Vandervoorn, C. Wallace, J. Wiedemer and C Wiegand, "45nm High-k + metal gate strain-enhanced transistors," *IEEE Symposium on VLSI Technology* pp.128-129, 2008.
- [5] C. H. Jan, U. Bhattacharya, R. Brain, S. J. Choi, G. Curello, G. Gupta, W. Hafez, M. Jang, M. Kang, K. Komeyli, T. Leo, N. Nidhi, L. Pan, J. Park, K. Phoa, A. Rahman, C. Staus, H. Tashiro, C. Tsai, P. Vandervoorn, L. Yang, J. Y. Yeh and P. Bai, "A 22nm SoC Platform Technology Featuring 3-D Tri-Gate and High-k/Metal Gate, Optimized for Ultra Low Power, High Performance and High Density SoC Applications," *IEEE International Electron Device Meeting (IEDM)*, pp. 44-47, 2012.
- [6] J. H. Yoon, H. C. Hunter and G. A. Tressler, "Flash & DRAM Si Scaling Challenges, Emerging Non-Volatile Memory Technology Enablement - Implications to Enterprise Storage and

Server Compute systems,” *Flash Memory Summit*, 2013.

- [7] Y. Nishi, “Advances in Non-volatile Memory and Storage Technology,” *Woodhead Publishing Series in Electronic and Optical Materials: Number 64*, 2014.
- [8] G. S. Sandhu, “Emerging Memories Technology Landscape,” *IEEE Non-Volatile Memory Technology Symposium (NVMTS)*, pp. 1-4, 2013.
- [9] M. Koyanagi, “The Stacked Capacitor DRAM Cell and Three-Dimensional Memory,” *IEEE Solid-State Circuits Society Newsletter*, vol. 13, no. 1, pp. 37-41, 2008.
- [10] E. Kultursay, M. Kandemir, A. Sivasubramaniam and O. Mutlu, “Evaluating STT-RAM as an Energy-Efficient Main Memory Alternative”, *International Symposium on Performance Analysis of Systems and Software (ISPASS)*, pp. 256-267, 2013.
- [11] B.C. Lee, E. Ipek, O. Mutlu and D. Burger, “Architecting phase change memory as a scalable DRAM alternative,” *International Symposium on Computer Architecture (ISCA)*, pp. 2-13, 2009.
- [12] A. Makarov, V. Sverdlov and S Selberherr, “New Trends in Microelectronics: Towards an Ultimate Memory Concept,” *International Caribbean Conference on Devices, Circuits and Systems (ICCDCS)*, pp. 1-4, 2012.
- [13] J. Singh, S. P. Mohanty and D. Pradhan, “Robust SRAM Designs and Analysis,” *Springer Publications*, 2013.
- [14] C. Shin, “Advanced MOSFET Designs and Implications for SRAM Scaling”, *Ph.D Thesis*, University of California Berkeley, 2011.
- [15] R. Micheloni, G. Campardo and P. Olivo, “Memories in Wireless Systems,” *Springer Publications*, 2008.
- [16] P. Pavan, R. Bez, P. Olivo, and E. Zanoni, “Flash Memory CellsAn Overview,” *Proceedings of the IEEE*, vol. 85, no. 8, pp. 1248-1271, 1997.

- [17] A. Goda, “Opportunities and challenges of 3D NAND scaling,” *International Symposium on VLSI Technology, Systems, and Applications (VLSI-TSA)*, pp. 1-2, 2013.
- [18] ISSCC Technology Trend, *International Solid-State Circuits Conference (ISSCC)*, 2014.
- [19] J. H. Yoon and G. A. Tressler, “Advanced flash technology status, scaling trends and implications to enterprise SSD technology enablement,” *Flash memory summit*, 2012.
- [20] S. K. Park, “Technology Scaling Challenge and Future Prospects of DRAM and NAND Flash Memory,” *IEEE International Memory Workshop (IMW)*, pp. 1-4, 2015.
- [21] K. M. Kim, D. S. Jeong and C. S. Hwang, “Nanofilamentary resistive switching in binary oxide system; a review on the present status and outlook,” *Nanotechnology*, vol. 22, 254002, 2011.
- [22] J. J. Yang, D.B. Strukov and D. R. Stewart, “Memristive Devices for Computing,” *Nature Nanotechnology*, vol. 8, pp. 13-24, 2013.
- [23] R. Waser, “Resistive non-volatile memory devices (Invited Paper),” *Microelectronic Engineering*, vol. 86, pp. 19251928 , 2009.
- [24] C. B. Lee, B. S. Kang, A. Benayad, M. J. Lee and S. E. Ahn, “Effects of metal electrodes on the resistive memory switching property of NiO thin films,” *Applied Physics Letters*, vol. 93, 042115, 2008.
- [25] A. Prakash, J. Park, J. Song, J. Woo, E. J. Cha and H. Hwang, “Demonstration of Low Power 3-bit Multilevel Cell Characteristics in a TaOx-Based RRAM by Stack Engineering,” *IEEE Electron Device Letters*, vol. 36, no. 1, pp. 32-34, 2015.
- [26] H. S. P. Wong, S. Raoux, S. Kim, J. Liang, J. P. Reifenberg, B. Rajendran, M. Asheghi and K. E. Goodson, “Phase Change Memory,” *Proceedings of the IEEE*, vol. 98, no. 12, pp. 2201-2227, 2010.

- [27] K. L. Wang, J. G. Alzate and P. K. Amiri, “Low-power non-volatile spintronic memory: STT-RAM and beyond,” *Journal of Physics D: Applied Physics*, vol. 46, 074003, 2013.
- [28] B. Radha, A. A. Sagade and G. U. Kulkarni, “Metal-organic molecular device for non-volatile memory storage,” *Applied Physics Letters*, vol. 105, 083103, 2014.
- [29] C. M. Lin, W. Shih, I. Y. Chang, P. C. Juan and J. Y. Lee, “Metal-ferroelectric ( BiFeO<sub>3</sub> ) -insulator (Y<sub>2</sub>O<sub>3</sub>) -semiconductor capacitors and field effect transistors for nonvolatile memory applications,” *Applied Physics Letters*, vol. 94, 142905, 2009.
- [30] J. W. Seo, J. W. Park, K. S. Lim, S. J. Kang, Y. H. Hong, J. H. Yang, L. Fang, G. Y. Sang and H. K. Kim, “Transparent flexible resistive random access memory fabricated at room temperature,” *Applied Physics Letters*, vol. 95, 133508, 2009.
- [31] W. W. Jang, J. O. Lee, H. H. Yang, and J. B. Yoon, “Mechanically Operated Random Access Memory (MORAM) Based on an Electrostatic Microswitch for Nonvolatile Memory Applications,” *IEEE Transactions on Electron Devices*, vol. 55, no. 10, pp. 2785-2789, 2008.
- [32] J. S. Meena, S. M. Sze, U. Chand and T. Y. Tseng, “Overview of emerging nonvolatile memory technologies,” *Nanoscale Research Letters*, vol. 9, 526, 2014.
- [33] M. Marinella, “The Future of Memory,” *IEEE Aerospace Conference*, pp. 1-11, 2013.
- [34] D. S. Jeong, R. Thomas, R. S. Katiyar, J. F. Scott, H. Kohlstedt, A. Petraru and C. S. Hwang, “Emerging memories: resistive switching mechanisms and current status,” *Reports On Progress in Physics*, vol. 75, 076502, 2012.
- [35] A. Chen, J. Hutchby, V. Zhirnov and G. Bourianoff, “Emerging Nanoelectronic Devices,” *John Wiley and Sons Ltd*, 2014.
- [36] J. B. Kwon, “Exploiting Storage Class Memory for Future Computer Systems: A Review,” *IETE Technical Review*, vol. 32, no. 3, 2015.

# Chapter 2

## Experimental Methods

This dissertation accounts for fabrication and characterization of DFGFET and RRAM devices. The fabrication process flow of these devices is mentioned in Appendix-A. Standard CMOS process such as deposition, etching, lithography, etc, were used for fabrication of the devices. Details of the individual unit process can be found in [1]. Characterization of the devices was performed by either electrical characterization systems or physical characterization systems. Electrical characterization was performed with keithley 4200 SCS parameter analyzer [2], HP 4155B semiconductor parameter analyzer [3] and Agilent E-4980A LCR meter [4]. Physical characterization was performed using systems like XRD and XPS.

### 2.1 Primary fabrication tools used

Description of some of the tools primarily used for device fabrication is mentioned below.

#### 2.1.1 Atomic layer deposition

Atomic layer deposition (or ALD) is highly used to deposit dielectric and metals with precise thickness control [5]. In this work, ALD process has been used to deposit dielectric films. For this purpose, a commercial Savannah S100 ALD system from Cambridge Nanotech was used [6]. Schematic of the ALD system is shown in Figure 2.1. The tool is equipped with high speed pneumatic valves in order to precisely control the pulse time (in a range of ms) of precursors. Two heater lines were connected to the substrate holder supporting temperature range up to 400 °C.

ALD is a self-limiting process which allows layer by layer control on the material (such as metals and oxides) to be deposited. Passivation in successive cycle enables the precise thickness and uniformity control of film growth which is essential to deposit ultrathin dielectric films with

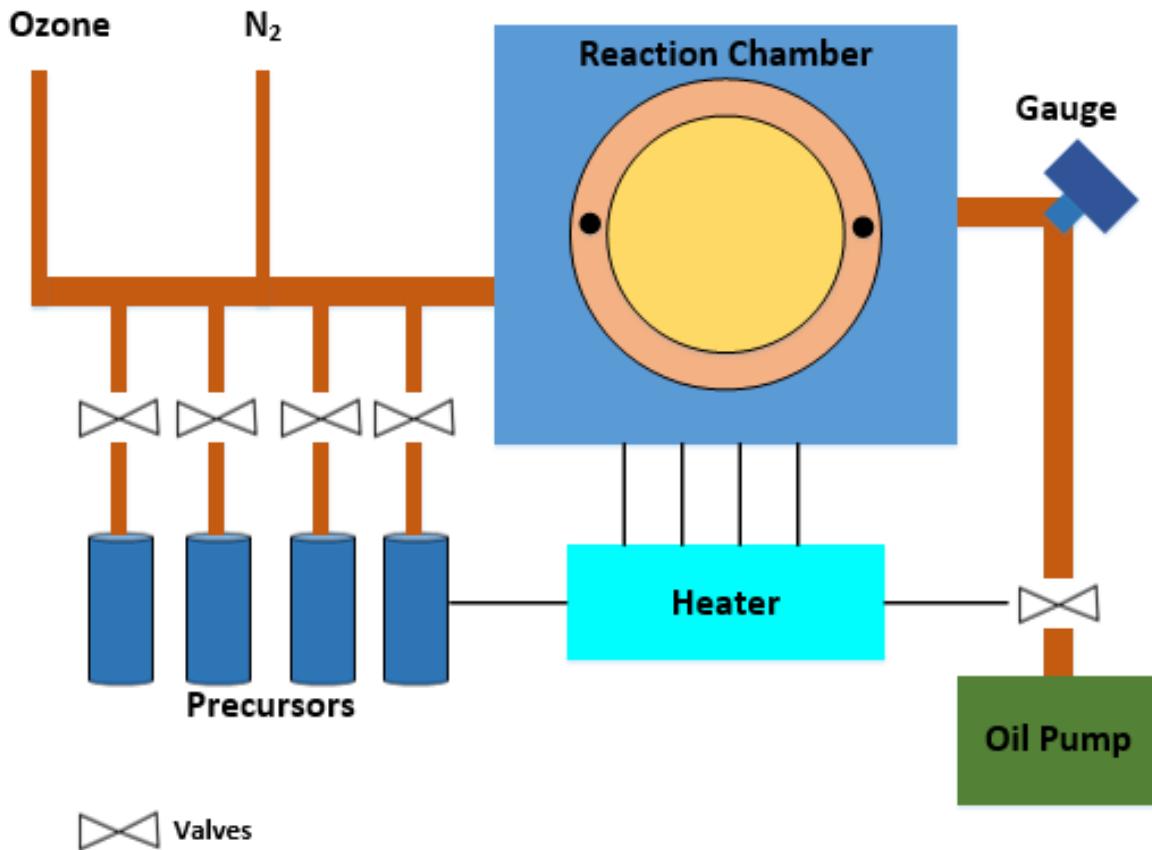


Figure 2.1: Schematic of ALD system.

great conformality. Another advantage of ALD is the low temperature deposition process which is highly used in flexible/transparent electronics. However, the deposition temperature should be kept at higher than the minimum required temperature for the reaction to occur on the substrate surface. Too low deposition temperature results in a thermally activated chemisorption as well as decrease in the deposition rate. On the other hand, if the deposition temperature is too high, chemical bonding cannot be sustained or the density of chemically reactive sites is reduced.

Figure 2.1 shows the schematic of the ALD chamber used in this dissertation work. The samples are kept in the reaction chamber which is connected to several precursor cylinders as well as carrier gas N<sub>2</sub>. All the precursors have high speed valves which can be opened and closed in fraction of seconds. This control allows the precise control on the flow time of the precursors. A heater is connected to the chamber whose temperature can be adjusted accordingly. The exhaust is connected to an oil pump via a gauge. Details on the ALD layer by layer deposition operation can be found in [14].

The dielectric deposition process involves the repeated cyclic execution of four critical steps:

- The metalorganic precursor is served in the ALD chamber by an inert carrier gas like N<sub>2</sub>. Precursor molecules get chemisorbed on the hydroxylated/oxidized surfaces of the substrate. ALD is a self-limited reaction which stops once the entire surface is saturated by the precursor molecules.
- The unreacted precursor molecules are purged out of the chamber through a pump.
- Next, the oxidizing precursor such as water or ozone is sent into the chamber to oxidize the new surface and eliminate the organic ligand as a gaseous by-product.
- The unreacted precursor and the gaseous by-products are then purged out of the chamber through the pump.

Table 2.1: Different ALD recipe.

<b>Dielectric</b>	<b>Precursors</b>	<b>Temperature</b>
Al <sub>2</sub> O <sub>3</sub>	Tri-methyl Aluminium (TMA) + H <sub>2</sub> O	200 °C
HfO <sub>2</sub>	Tetrakis-dimethyl-amino-hafnium (TDMAH) + H <sub>2</sub> O	200 °C
SiO <sub>2</sub>	3-amino-propyl-triethoxy-silane (APTS) + H <sub>2</sub> O + O <sub>3</sub>	150 °C
TiO <sub>2</sub>	tetrakis(dimethylamino) titanium (TDMAT) + H <sub>2</sub> O	200 °C

Table 2.1 shows the typical recipe for ALD used for different parts of the dissertation work [7]. In many cases, multiple oxidizing chemistries may be required to ensure completeness of the oxidizing step (e.g. both water and ozone are required for SiO<sub>2</sub> deposition). Also, an optimization of the ALD process recipe is required to ensure sufficient precursor quantity and reaction time for each precursor step. This may also require heating the precursor cylinder to increase the vapor pressure. Details of the specific precursors and the recipe conditions used for different ALD dielectrics are provided in Table 2.1. Note that compound dielectric like HfAlO was deposited with repeated cycles of HfO<sub>2</sub> followed Al<sub>2</sub>O<sub>3</sub> cycle.

### 2.1.2 UHV RF magnetron sputtering

Sputtering is widely used to deposit metals and ceramic thin films [8]. The sputtering process is done in a plasma ambient where active plasma ejects out the target material and accelerates them to deposit on the substrate. The target material is kept at negative potential, whereas the

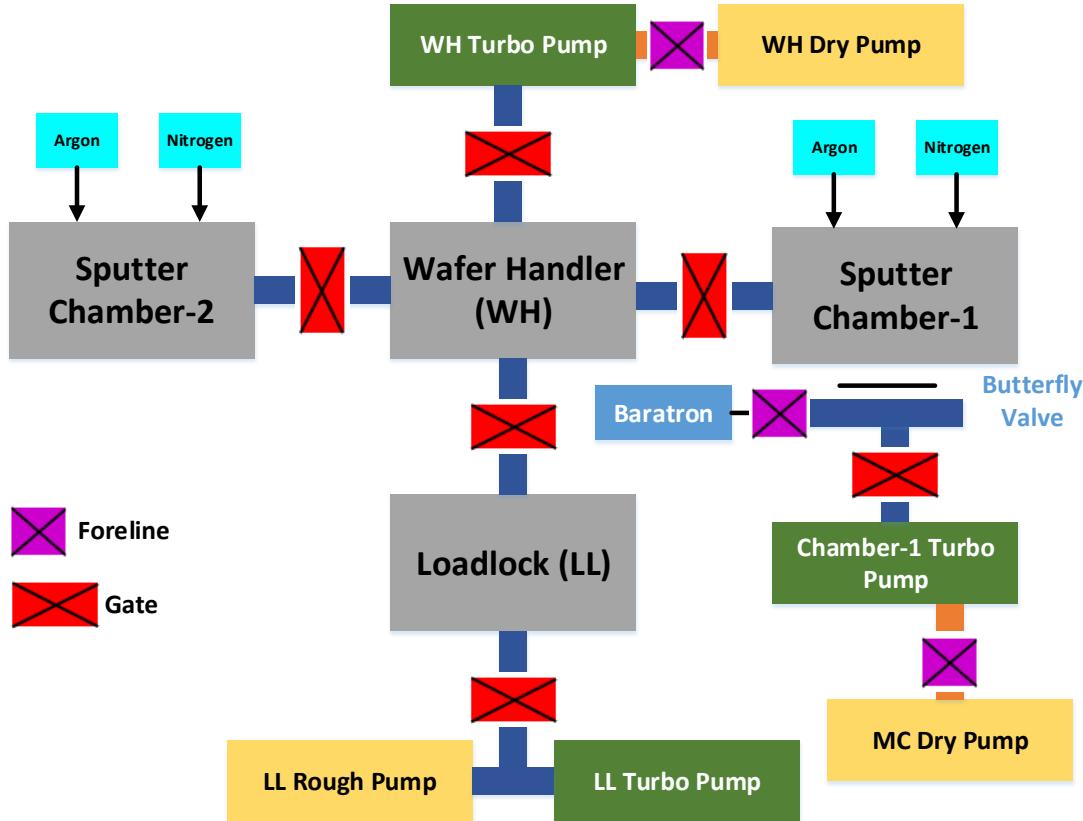


Figure 2.2: Schematic of UHV-RF sputter system.

substrate is kept at ground potential. When the plasma of inert gas like Argon is introduced in the chamber, positively charged Ar ions accelerate to the negatively charged target material. Since these energetic positively charged ions eject target atoms through the energy transfer. The ejected atoms then get deposited on the substrate.

A magnetic field can be used to enhance sputtering yield. The magnetic field is applied by using a magnetron behind the target material. The magnetron creates a magnetic field that increases the density of plasma near the target resulting in a higher deposition rate at a lower input power. Radio-Frequency (RF) sputtering is also used to deposit dielectric materials. During RF sputtering, electron oscillation with ac field results in no charge accumulation at the target surface.

Thus, RF magnetron sputtering provides several advantages:

- Provides high deposition rate.
- Provides capability to deposit complex alloys.
- Allows deposition of most of the metals as well as dielectrics.

All the metal electrodes used in this dissertation work were deposited using UHV RF magnetron sputtering system. Schematic of the UHV-RF sputtering system is shown in Figure 2.2. Details about RF sputtering mechanism can be found in [9]. The base pressure of main chamber was kept lower than  $1 \times 10^{-7}$  Torr in order to minimize the level of impurities in the chamber. Three RF-magnetron guns are attached on top of the each sputtering chamber. Thus, a total of six different materials or metal alloys can be easily deposited without breaking the vacuum. For reactive sputtering of metal like TaN, deposition was done by flowing Ar and  $N_2$  simultaneously. The deposition pressure was set to 5 mTorr after the gas flow in order to prevent gas-phase collision and optimized deposition rate. More details on the sputtering system can be found in [14].

## 2.2 Analytical characterization methods

Analytical methods are useful in order to understand the physical properties of different materials. Some of the essential physical properties are atomic structure of the material, crystallinity of the material etc. Following techniques were used to characterize different dielectric used in this dissertation.

### 2.2.1 X-ray photoelectron spectroscopy

X-Ray Photoelectron Spectroscopy (XPS) is highly used to analyze the chemical information in ultrathin high-k oxides or at the surface [10]. After depositing thin films, XPS is used to find the presence of the desired material, along with traces of contaminants as well as defects like oxygen vacancies. During XPS measurement, x-rays are illuminated on the substrate, and electrons ejected from the substrate material is analyzed to find out chemical information.

Figure 2.3 shows a schematic diagram of the XPS electron emission process and the typical measurement system. The sample under test is irradiated with monochromatic Mg or Al X-ray with the help of an X-ray source. The high energy x-rays then eject electrons present in the outermost orbital of the substrate material. The kinetic energy ( $KE$ ) of electrons ejected from the sample is the measured and analyzed using an electron analyzer.

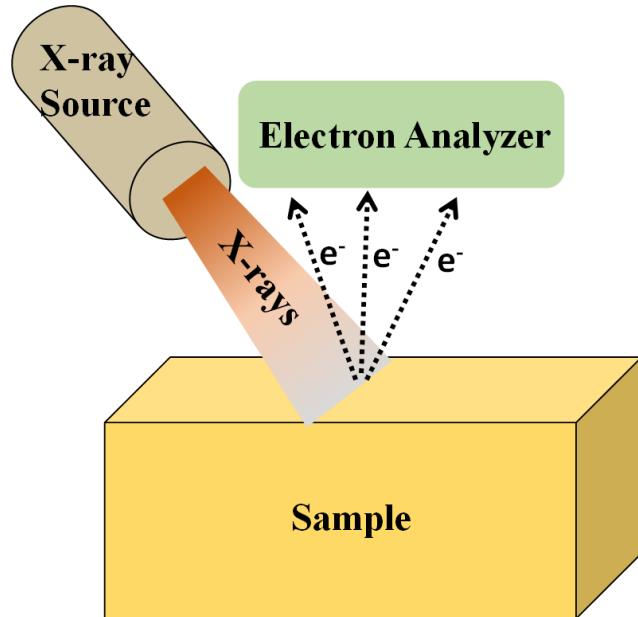


Figure 2.3: Schematic of the X-ray photoelectron spectroscopy method.

Consider the incident x-ray have an energy of  $h\nu$ , then the energy relation is expressed as:

$$h\nu = KE + BE \quad (2.1)$$

where  $BE$  is the binding energy of the electron. Since  $\nu$  of the x-rays are a known quantity, and the quantity  $h\nu$  is the photon energy. Subtracting the  $KE$  of the ejected electrons from the photon energy results in the binding energy of the material under observation. This binding energy can reveal many important information about the material. If the binding energy is low, it means the electrons in the orbital are far away from the nucleus, and vice versa. The exact binding energy can be used to determine the orbital of the electrons thereby estimating the entire electronic configuration. This technique can also be used to complex materials such as dielectrics, alloys etc. Details about XPS analysis can be found in [11].

In this dissertation, XPS analysis has been used to determine the electronic configuration of  $TiO_2$  dielectric, mentioned in chapter-6. The volatile operation using RRAM requires the presence of inherent oxygen vacancies. Thus, the XPS measurement was performed to confirm two basic requirements, (i) presence of fully oxidized dielectric without any contaminants, and (ii) presence of oxygen vacancies which can be modulated to obtain a volatile RRAM operation.

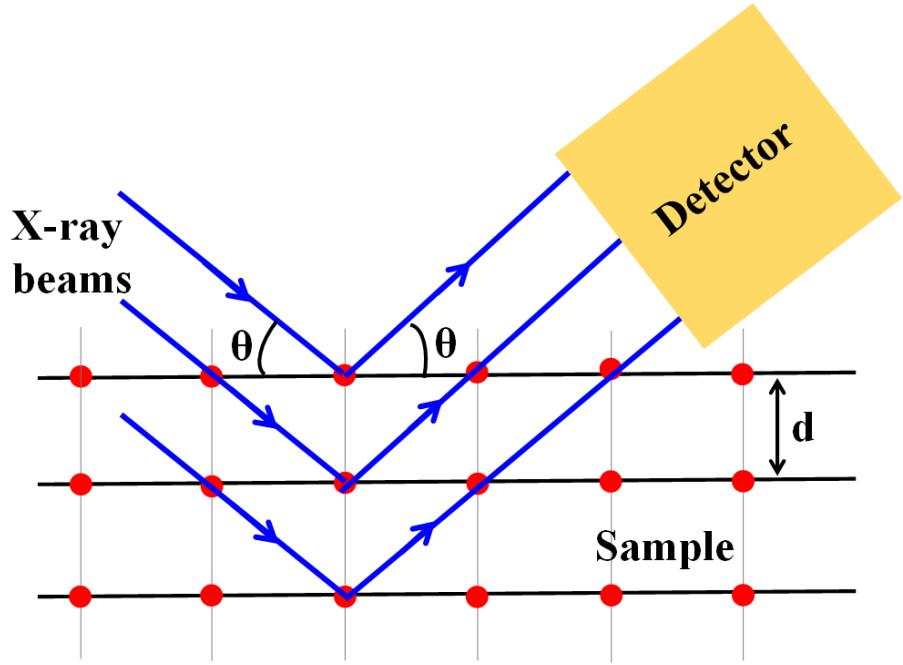


Figure 2.4: Schematic of the X-ray diffraction method.

### 2.2.2 X-ray diffraction analysis

X-ray diffraction (or XRD) analysis is performed to analyze if a material is in crystalline, polycrystalline or amorphous form [12]. During XRD, x-rays at different incident angle are injected into the material and the reflected x-ray is analyzed by a detector, as shown in Figure 2.4. Only short wavelength x-rays (in the range of a few angstroms) are used so that the wavelength of x-rays is comparable to the size of atoms in the sample under observation. Commonly, Cu and Mo are used as the x-ray source which emits 8 keV and 14 keV x-rays with corresponding wavelengths of 1.54 and 0.8 respectively. The energetic x-rays can penetrate into the materials under observation, and provide desired information about the material.

The x-ray have known characteristic wavelength (say  $\lambda$ ). When they are thrown on a sample, they get reflected and the reflected spectra is analyzed by a detector to determine the crystalline state of the material. Consider the rays fall on the sample at an angle  $\theta$ , then the Bragg's law can be written as:

$$2ds\sin\theta = n\lambda \quad (2.2)$$

where  $d$  is the inter-atomic distance in the material. If the material is crystalline, and the distance between the atoms are integral multiple of  $\lambda$ , then the x-ray received at the detector will have a constructive interference at certain angle  $\theta$  thereby resulting in a peak in the received signal. On the other hand, amorphous materials will have only the destructive interference of the x-rays at the detectors, hence no peaks will be observed. Depending upon the location of peaks along the  $\theta$  axis (usually  $2\theta$  is used for convenience), one can determine the materials in the sample as well as its crystal information. Similarly, in poly-crystalline materials, several peaks are observed in the XRD analysis. Hence XRD is a useful tool to extract the crystallographic information of the material.

In this dissertation, XRD has been used to determine crystallinity of  $\text{TiO}_2$  dielectric deposited using ALD system. To understand the volatile operation in non-filamentary RRAM, it is important to analyze the crystallographic properties of the RRAM dielectric. Results of  $\text{TiO}_2$  XRD measurement is mentioned in chapter-6 of the dissertation.

## 2.3 Electrical characterization methods

Electrical characterization methods has been extensively used in this dissertation work. Electrical characterization provides the information of device behavior such as current-voltage relationship, capacitance-voltage relationship, etc. Electrical characterization of semiconductor devices is a vast area, with several techniques used to understand different device behavior. A detailed analysis of most of the electrical characterization techniques used in modern day semiconductor electronics can be found in [15].

### 2.3.1 Capacitance-Voltage measurements

Capacitance-Voltage (or CV) measurements are highly used to characterize MOS capacitors in order to find parameters like dielectric constant, flatband voltage, threshold voltage etc. Figure 2.5 shows the typical CV characteristics of a MOS capacitor fabricated on p-Si substrate [13]. At -ve bias voltages, the holes at the substrate accumulates near the interface, and this mode is called the accumulation mode. At this mode, the capacitance is referred at  $C_{max}$ , which is the capacitance corresponding to the equivalent oxide thickness.

As the voltage goes toward +ve bias, holes are repelled from the Si/ $\text{SiO}_2$  interface gradually making the depletion mode. At certain bias, all the bands are flat, and the voltage at which this condition is met is termed as flatband voltage ( $V_{FB}$ ). A maximum width of depletion width is reached after which the inversion charge starts appears at the substrate only for low

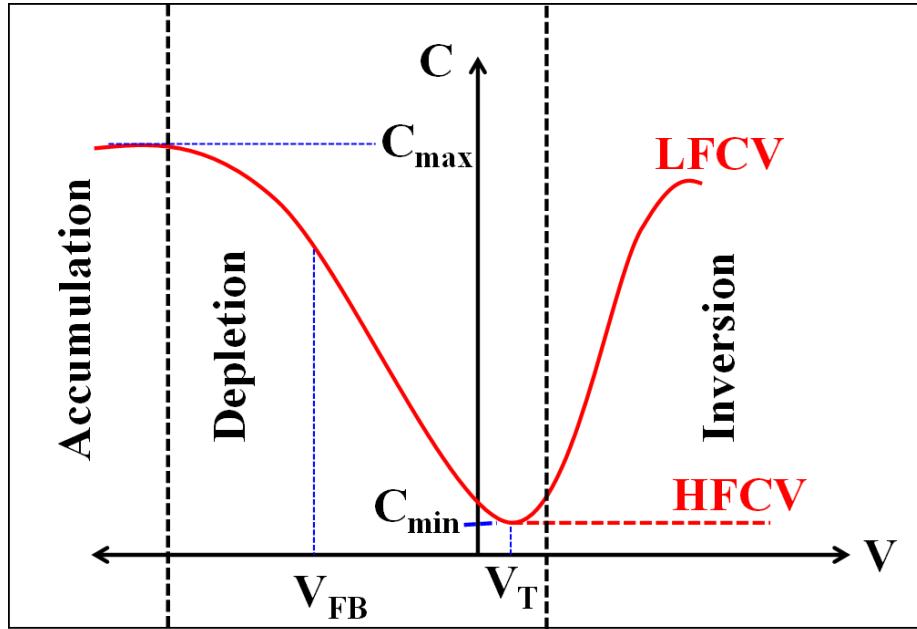


Figure 2.5: (a) Band diagram of MOS capacitor and (b) CV analysis of MOS capacitor.

frequency CV (or LFCV) analysis. This is called threshold point where the inversion charge starts appearing in the LFCV curve of the MOS capacitor. The capacitance just before inversion is  $C_{min}$ , and it goes back to  $C_{max}$  after the inversion point. At high frequency (HFCV), the carriers don't get enough time to respond to the signal, hence capacitance remains at  $C_{min}$  in the inversion region.

### 2.3.2 Current-Voltage measurements of DFGFET

Current-voltage (or IV) measurements are used to understand the transistor behavior and several current conduction mechanisms. For understanding the dual floating gate memory MOSS-FET device, 4-terminal IV measurements were performed to find the  $V_T$  of the transistor. Figure 2.6 shows the typical linear  $I_D$ - $V_G$  characteristics of MOSFET. For extracting the  $V_T$ , transconductance is calculated from the  $I_D$ - $V_G$  curve and plotted against the voltage. At the peak of the transconductance point, a line is drawn which intersects the  $I_D$ - $V_G$  curve. At the intersection point, a tangent is drawn and intersection point at the X-axis is used to determine the  $V_T$  of the transistor [14].

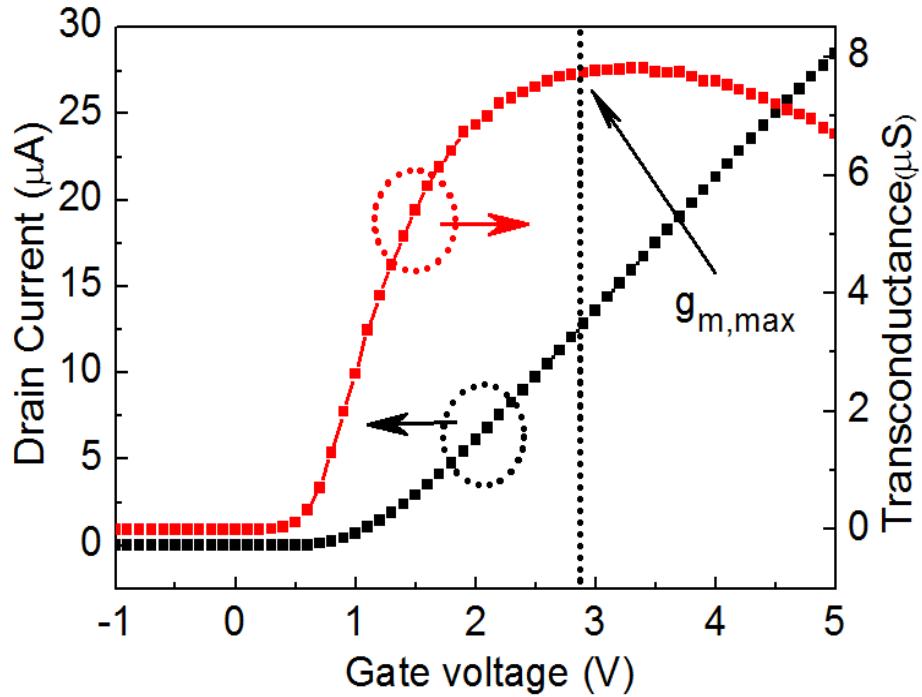


Figure 2.6: Typical  $I_D$ - $V_G$  characteristics of MOSFET and extraction of  $V_T$ .

All the non-volatile performance analysis of the dual floating gate memory MOSFET was carried out with  $V_T$  analysis of the transistor. Key points of electrical characterization of dual floating gate MOSFET are mentioned below:

- To measure the non-volatile program and erase window:  $V_T$  of the MOSFET was measured to analyze the programming or erasing behavior. During program window analysis,  $V_T$  of selected virgin devices were measured, followed by  $V_T$  measurement after every programming pulse applied. While during erase, the devices were initially taken into program state followed by successive erasing operations.
- Non-volatile retention measurement: Devices were taken into program and erase state, and  $V_T$  was monitored for different time intervals.
- Non-volatile endurance measurement: A single device was initially selected, and  $V_T$  of the device was measured after program and erase operation. Repeated program and erase cycles were then applied to the device, and the  $V_T$  of the device was measured after specific different cycle numbers.
- Volatile retention measurement: The volatile retention was performed with pulse-CV

scheme. In this measurement, the program and read pulses were applied, and the capacitance of the gate stack was monitored. The details are mentioned in chapter-3.

### 2.3.3 Current-Voltage measurements of RRAM

Unlike floating gate memory devices, RRAMs are two terminal devices. Hence the 2-terminal IV measurement for resistor was used to study the RRAMs. In this measurement setup, one terminal of the resistor is grounded, and bias is applied to the other terminal. Key points of measuring RRAM behavior are mentioned below:

- Forming operation: Forming operations were performed by applying bias at one of the selected electrode keeping the other electrode grounded. To prevent the devices going into hard breakdown, the current through the terminals were limited to a compliance limit.
- Reset operation: All the RRAM devices mentioned in this thesis were observed to show bipolar RRAM operation. During reset, negative bias was applied to the desired electrode keeping the other electrode grounded. No compliance limit is required during reset.
- Set operation: set operation is very similar to the forming operation. The only difference between set operation and forming operation is the voltage window. Since forming refers to the creation of the filament in the dielectric which can connect two electrode, a high bias is desired during forming operation. Set operation, on the other hand, requires to just re-create the broken filament. Hence the set voltage is lower than forming voltage.
- Synaptic learning: synaptic learning operation mentioned in chapter-5 of the thesis were performed with set and reset pulses. After every set and reset pulses, read pulses were applied to measure the conductance of the dielectric.
- Volatile RRAM operation: Since the volatile RRAM operation was measured on a  $\text{SiO}_2/\text{TiO}_2$  MOS capacitor, pulsed-CV scheme was used to monitor the capacitance. Program pulses were applied to the gate of the MOS capacitor, and the capacitance change was measured with successive read pulses.

## 2.4 Summary

This chapter reviewed the fabrication and characterization methods used for this dissertation work. Successive chapters will explain more about the results obtained from dual floating gate memory MOSFET and different RRAM devices.

## REFERENCES

- [1] J. D. Plummer, M. D. Deal and P. B. Griffin, "Silicon VLSI Technology: Fundamentals, Practice and Modeling," *Prentice Hall publication*, 2000.
- [2] Keithley 4200-SCS Parameter Analyzer,  
[www.keithley.com/products/semiconductor/parametricanalyzer/dcac/currentvoltage/4200scs](http://www.keithley.com/products/semiconductor/parametricanalyzer/dcac/currentvoltage/4200scs)
- [3] 4155B / 4156B Semiconductor Parameter Analyzers, [www.keysight.com/en/pd-542813-pn-4155B/semiconductor-parameter-analyzers](http://www.keysight.com/en/pd-542813-pn-4155B/semiconductor-parameter-analyzers)
- [4] E4980A Precision LCR Meter, [www.keysight.com/en/pd-715495-pn-E4980A/precision-lcr-meter-20-hz-to-2-mhz](http://www.keysight.com/en/pd-715495-pn-E4980A/precision-lcr-meter-20-hz-to-2-mhz)
- [5] N. Pinna and M. Knez, "Atomic Layer Deposition of Nanostructured Materials," *Wiley-VCH publication*, 2011.
- [6] Ultratech/Cambridge NanoTech Simply ALD, <http://www.cambridgenanotechald.com/atomic-layer-deposition.shtml>.
- [7] N. Ramanan, "Investigation of ALD Dielectrics for Improved Threshold Voltage Stability and Current Collapse Suppression in AlGaN/GaN MOS-HFETs," *Ph.D Thesis, North Carolina State University*, 2014.
- [8] S. Franssila, "Introduction to Microfabrication, 2nd edition" *Wiley publication*, 2010.
- [9] Y. W. Chung, "Introduction to Materials Science and Engineering," *CRC Press*, 2006.
- [10] S. Iwata and A. Ishizaka, "Electron spectroscopic analysis of the Si/SiO<sub>2</sub> system and correlation with metal-oxide-semiconductor device characteristics," *Journal of Applied Physics*, vol. 79, pp. 6653-6709, 1996.
- [11] P. van der Heide, "X-ray Photoelectron Spectroscopy: An introduction to Principles and Practices," *Wiley publication*, 2011.

- [12] Y. Waseda, E. Matsubara and K. Shinoda, “X-Ray Diffraction Crystallography,” *Springer Berlin Heidelberg publication*, 2011.
- [13] Y. Taur and T. H. Ning, “Fundamentals of Modern VLSI Devices,” *Cambridge University Press*, 2009.
- [14] B. Lee, “A Study of Group III Elements (La, Gd, Eu, and Al) Incorporation on Metal Gate / Highk Stacks for Advanced CMOS Applications,” *Ph.D Thesis, North Carolina State University*, 2010.
- [15] D. K. Schroder, “Semiconductor Material and Device Characterization, 3rd Edition,” *Wiley-IEEE Press*, 2015.

## Chapter 3

# Dual Floating Gate Memory MOSFET

Modern smart electronic devices rely extensively on volatile and non-volatile memories. While codes and look-up tables are stored in the non-volatile memory, the logical operations are performed using the volatile memories [1]. The processing unit has separate area on chip dedicated to the non-volatile storage and the volatile memories. Random Operating Memory (ROM) and non-volatile memories serve as the storage section of the processor, whereas volatile memories like DRAM and SRAM are highly used for execution along with other logic circuits. Flash memories serve as the core device of NAND and NOR non-volatile memory array [2].

Among NAND and NOR, NOR is usually preferred to store codes because of their faster access time [3]. Traditionally, nitride based charge trap layer is used to store the charge in the oxide. Charge trap flash memories (like SONOS and TANOS flash memories) have been studied extensively to meet P/E window requirements, good retention and endurance characteristics [4, 5, 6, 7]. All these devices have their own advantages and disadvantages. A thorough understanding of these devices have resulted in tremendous memory capacity in solid state drives and flash drives. Similarly, DRAM and SRAM have also been a major components of electronic processor. DRAMs are necessary to process data and make logical operations, whereas SRAMs are used as cache memory. 1T-1C DRAM has been the preferred DRAM choice, but several new DRAM technologies which employs 1T-0C configuration have been proposed to improve DRAM performance [8, 9]. Similarly, different SRAM architectures have also been proposed to meet the scaling needs [10].

The future course of memory needs strong research on candidates that can perform storage and execution in the same area on the chip [11]. This will eventually improve the execution

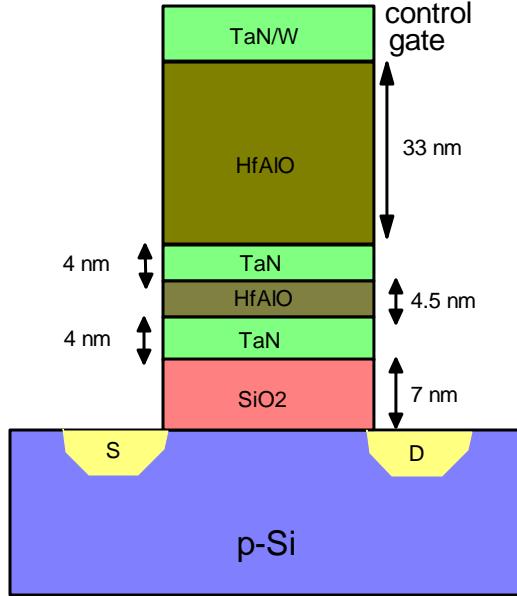


Figure 3.1: Schematic representation of the dual floating gate memory MOSFET.

speed, provide better security and area efficient solutions. 3D vertical stacking using through silicon via (TSV) technology seems to be a promising idea to meet this requirement [12, 13]. Although independent improvement of DRAM and flash memory can meet future power requirements, a significant improvement in power consumption can be obtained if both technologies are combined into a single transistor [14]. A unified memory which can simultaneously operate in volatile and non-volatile regime can also help to create a simpler computational architecture that occupies a lesser circuit area [18].

In this chapter, we discuss about a novel device capable simultaneous volatile and non-volatile operation. The device is very similar to a conventional flash memory. Instead of using single floating gates, this device has two floating gates separated by a thin dielectric. The addition of a floating gate provides additional functionality to the flash memory. A volatile memory behavior can be obtained by modulating charge between the floating gates. This memory raises the possibility of combining flash and DRAM functionality in the same transistor. The unified memory can also be used for applications like instant-ON computing, fast in-situ check-pointing for rollback/recovery leading to improved resiliency, neuromorphic systems, sensor systems, etc.

### **3.1 Theoretical understanding of the dual floating gate memory MOSFET**

The schematic representation of the dual floating gate memory MOSFET (or DFGFET) is shown in Figure 3.1. The transistor consists of following gate stack: tunnel oxide (TO), bottom floating gate (BFG), intermediate dielectric (IMD) which separates the floating gates, top floating gate (TFG), inter-poly dielectric (IPD) and control gate (CG). Detailed fabrication process flow can be found in Appendix-A. The purpose of keeping TO, BFG and IPD are similar to that of flash memory. However, an additional floating gate (TFG) and a thin dielectric (IMD) provides additional functionality to the flash memory. The device operation principles are discussed in subsequent sections.

#### **3.1.1 Device operation**

Figure 3.2 shows the band-diagram description of the device. Non-volatile operation is realized with charge transfer between the channel and the floating gates, whereas the volatile operation is performed with a charge transfer between the floating gates. A very high voltage at CG will result in charge transfer between the floating gates and the substrate with the help of Fowler-Nordheim (FN) Tunneling. This transition results in a significant increase in the threshold voltage (or  $V_T$ ). Similarly, a very high negative bias at the CG will move back the electrons from the floating gates to the substrate thereby decreasing the  $V_T$  significantly. These two operations are very similar to the non-volatile operations in conventional flash memories. The increased  $V_T$  state is referred as the non-volatile programmed state, whereas the decreased  $V_T$  state is referred as non-volatile erase state.

However, with a relatively smaller bias at the CG, charge can be moved within the floating gates with the help of direct tunneling thereby making a relatively smaller threshold voltage (or  $V_T$ ) shift in the device. To enable this criteria, the EOT of the dielectric (IMD) separating the floating gates is required to be very low (less than 1 nm), and the physical thickness should be more than 4 nm. Note that the conduction band of the IMD after the charge transfer will not be flat, resulting in a built-in electric field. This built-in electric field will result in charge balance between the floating gates in order to make the band flat. Hence, once the charge is moved between the floating gates, the inbuilt electric field at the IMD will assist them to tunnel back to their equilibrium position thereby enabling the volatile mode operation of the device. Thus, by engineering the IMD and band-offset between the electrode, a reasonable volatile retention time can be achieved.

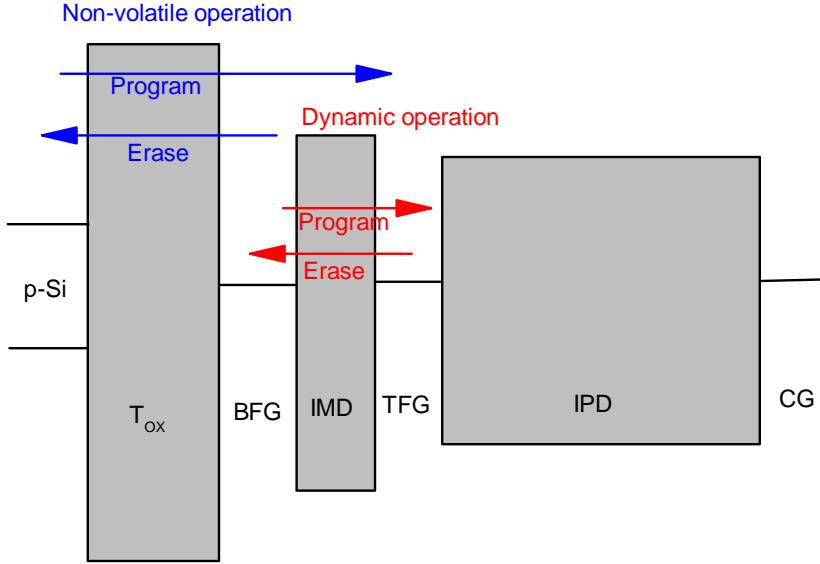


Figure 3.2: Band diagram representation of the operation modes in DFGFET.

Most attractive feature of the device is the concurrent mode of operation where the device can be used simultaneously for volatile and non-volatile mode operation. Let's assume the non-volatile operation with a high positive bias at CG brought  $K$  electrons into the floating gates from the substrate. After a reasonable time, these charges will balance themselves between the floating gates, say  $K/2$  electrons are stored in BFG and  $K/2$  electrons in TFG. The  $V_T$  at this condition corresponds to the non-volatile programmed state. Now a relatively smaller bias at CG can move  $K/4$  electrons from BFG into TFG, leaving behind  $K/4$  electrons at the BFG and  $3K/4$  electrons in the TFG.  $V_T$  at this condition will be different from the non-volatile  $V_T$ . This corresponds to the concurrent mode of operation (non-volatile programmed state, volatile programmed state). A similar operation can be performed after non-volatile erase operation. Hence the device can potentially store 2-bits at a  $V_T$  location, one bit corresponding to the non-volatile operation (programmed or erased), and the other bit corresponding to the volatile bit.

### 3.1.2 Theoretical formulation

A better understanding of the device working can be obtained by formulating the device operations. Consider the structure shown in Figure 3.1. It is known that the flat-band voltage ( $V_{FB}$ ) and  $V_T$  of a transistor is highly dependent on the trapped charge in the dielectric. Floating gates also have inherent charge which is equivalent of trapped charge in the dielectric. The pres-

ence of trapped charge tends to increase or decrease the  $V_{FB}$  depending upon the polarity of charge. The presence of positive charge in the dielectric is known to increase the  $V_T$  for n-MOS (having p-Si substrate). Similarly, the  $V_T$  decreases due to presence of negative trapped charge for n-MOS. Overall, the contribution to change in  $V_{FB}$  due to the presence of trapped charges in the dielectric is given as [15]

$$\Delta V_{FB} = -\frac{1}{\epsilon_{ox}} \int_0^{T_{ox}} x \rho(x) dx \quad (3.1)$$

where  $T_{ox}$  is the equivalent oxide thickness of the entire gate stack and  $\rho_x$  is the volume charge density at distance  $x$  from the channel. The floating gate thicknesses are assumed to be zero since they are made of metal which supports no potential drop across it. The DFGFET device has two floating gates acting as charge storage locations. Consider the BFG is located at a distance  $t$  from Si-SiO<sub>2</sub> interface,  $\Delta t$  be the thickness of IMD. Thus, the TFG is located at a distance  $t + \delta t$  from the Si-SiO<sub>2</sub> interface. At equilibrium, let  $Q_1$  ( $C\ cm^{-2}$ ) be the charge at BFG and  $Q_2$  ( $C\ cm^{-2}$ ) be the charge at TFG. Solving equation 3.1 for these values, the equation can be re-written as:

$$\Delta V_{FB} = -\frac{Q_1}{\epsilon_{ox}}(T_{ox} - t) - \frac{Q_2}{\epsilon_{ox}}[T_{ox} - (t + \Delta t)] \quad (3.2)$$

Equation 3.2 describes the possibility of creating a separate volatile bit storage capability in dual floating gate memory transistor. The change in  $V_{FB}$  is affected by charge stores in individual floating gates, as well as the thickness of IMD. A small positive voltage at the CG can tunnel some electrons from the BFG into the TFG via direct tunneling resulting in different values of Q1 and Q2 which in turn can shift the  $V_{FB}$ . Interestingly, a higher change in  $V_{FB}$  can be obtained by increasing the thickness of IMD. However, the thickness of IMD needs to be restricted in order to maintain the charge coupling ratio (also called as gate coupling ratio). The charge coupling ratio is defined as the ratio of the IPD capacitance to the total gate capacitance. A reduction in charge coupling ratio can result in slower program/erase operation thereby degrading the flash performance.

Now, the volatile mode is obtained by making a charge transfer between the floating gates. With the application of a positive bias at the CG, electrons can be tunneled from BFG to TFG. This transition will change the value of  $V_{FB}$  as the number of electrons in BFG is different compared to the situation before tunneling. Considering  $Q$  be the charge moved from BFG to TFG and incorporating the values in 3.2, change in  $V_{FB}$  due to this transition can be given as:

$$\Delta V_{FBd} = \Delta V_{FB} \pm \frac{|\Delta Q| \Delta t}{\epsilon_{ox}} \quad (3.3)$$

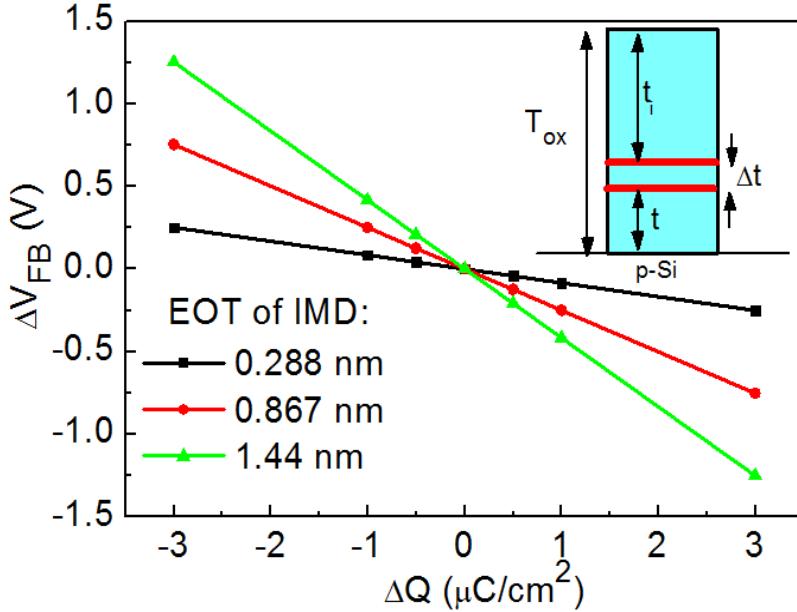


Figure 3.3: Simulated  $V_{FB}$  shift because of volatile operation in DFGFET [19] (considering EOT of the entire gate stack to be  $T_{ox}$ , EOT of the TO to be  $t$ , EOT of the IPD to be  $t_i$  and EOT of the IMD to be  $\Delta t$ ).

Thus,  $V_{FB}$  will decrease if electrons are tunneled from BFG to TFG, and vice-versa. The tunneling of electrons also causes an imbalance of electron concentration in the floating gates. Moreover, there is an in-built electric field dropped at the IMD once the charge imbalance situation is created. This built-in field will result in tunnel back of electrons to attain the equilibrium. With the use of an engineered IMD candidate having a band offset and leakage characteristics appropriate to retain these redistributed charges for more than 64 ms [16], this mode of operation can potentially be made volatile like a DRAM.

Note that if positive bias is given to the CG, electrons will tunnel from BFG to TFG thereby reducing the  $V_{FB}$ , and vice-versa. Figure 3.3 shows the change in  $V_{FB}$  because of charge transfer between the floating gates. A CV sweep performed at DFGFET with bias at CG will hence result in an anti-clockwise hysteresis because of this volatile mode operation.

A relatively high voltage bias at CG would bend the conduction band of tunnel oxide allowing FN tunneling of electrons from the channel into both the floating gates. This condition is the non-volatile programming operation which increases the  $V_{FB}$  significantly. Similarly, applying a large negative bias at the CG will result in the non-volatile erase operation thereby

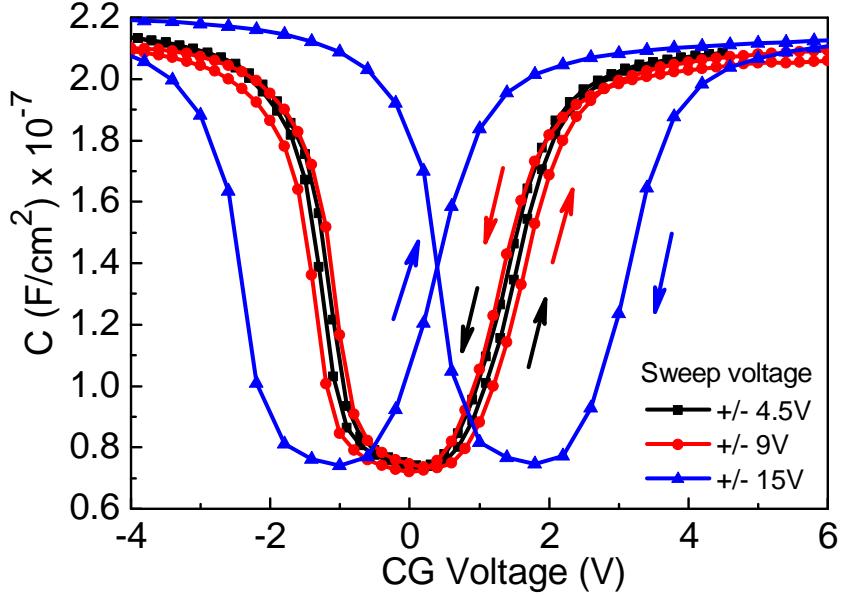


Figure 3.4: CV hysteresis characteristics of the DFGFET memory device.

a significant reduction in  $V_{FB}$ . If  $k_1$  and  $k_2$  are the extra charge added/subtracted at BFG and TFG respectively because of non-volatile programming/erasing, the  $\Delta V_{FB}$  can be written as:

$$\Delta V_{FBn} = \Delta V_{FB} \pm \frac{K_1}{\epsilon_{ox}}(T_{ox} - t) \pm \frac{K_2}{\epsilon_{ox}}[T_{ox} - (t + \Delta t)] \quad (3.4)$$

2nd and 3rd term in the right side of equation 3.4 is positive in case of non-volatile programming, whereas they are negative in case of non-volatile erase. The use of a high quality IPD and TO that helps retain these charges over 10 years can help use this mode of operation in a non-volatile mode like the current flash device. An oxide which has high bandgap and lower trap charge density is a suitable TO candidate and can help in excellent non-volatile retention. Traditionally, thermally grown thick  $\text{SiO}_2$  (around 8 nm) is used as TO. Similarly, IPD should be a high-k material which can offer thicker oxide maintaining a good coupling ratio and less tunneling of electrons from floating gates into the CG, and vice-versa. Degradation of TO and IPD can cause charge leakage after programming thereby creating retention issues, as well as lower the non-volatile endurance. Since non-volatile programming increases the  $V_{FB}$  and non-volatile erase reduces the  $V_{FB}$ , a CV sweep with a gate bias will result in a clockwise hysteresis in case of non-volatile transitions. The width of hysteresis loop will be much higher than the volatile hysteresis width.

Even after making a non-volatile transition, the application of the small positive gate voltage

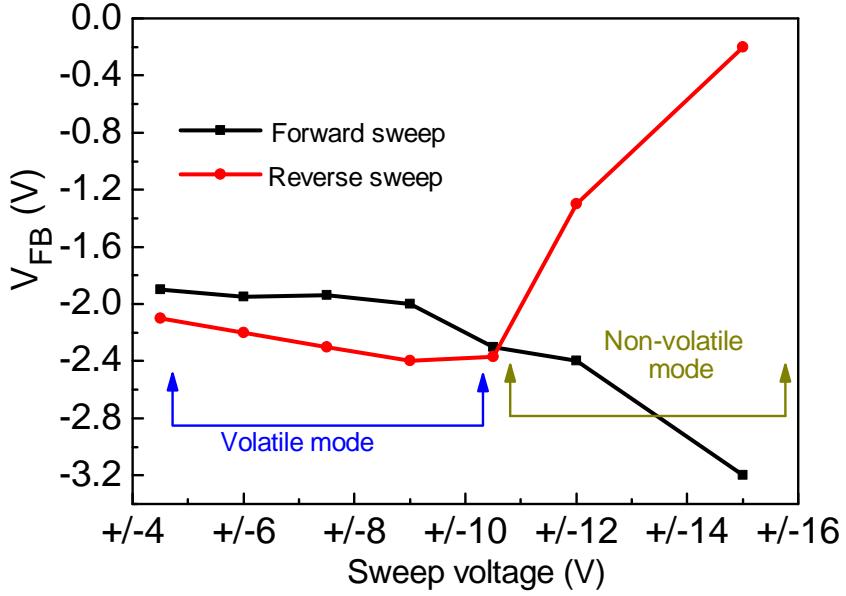


Figure 3.5: Variation in  $V_{FB}$  during forward and reverse sweep in DFGFET memory device [20].

can redistribute the charge between BFG and TFG, thereby creating a volatile  $V_{FB}$  shift. This constitutes the concurrent mode of operation where both volatile and non-volatile modes exist simultaneously. If  $\Delta k$  is the charge transferred from BFG to TFG with a volatile programming applied after non-volatile transition, then  $\Delta V_{FB}$  is given as:

$$\Delta V_{FBc} = \Delta V_{FBn} \pm \frac{|\Delta K| \Delta t}{\epsilon_{ox}} \quad (3.5)$$

Equation 3.5 is the governing equation of DFGFET device. The first term in the right hand side is the change in  $V_{FB}$  after non-volatile operation, whereas the second term is the change in  $V_{FB}$  after the volatile operation. These two operations can be controlled by allowing reasonable operation voltage window at the CG. Thus, depending upon the voltage applied at the CG, two bit memory can be realized in dual floating gate transistor where 1 bit corresponds to a non-volatile operation whereas the other bit corresponds to the volatile operation.

After formulating all the theoretical behavior of the device, a transistor was fabricated to experimentally verify the idea. Subsequent sections explain the experimental results obtained from the fabricated device.

### 3.2 CV analysis of the DFGFET device

Electrical characterization using IV and CV measurement tools was performed to verify the DFGFET device operation. Figure 3.4 shows the CV hysteresis measurements performed on the DFGFET device. The frequency of measurement was kept at 1 MHz. The CV measurements were performed using a forward sweep where voltage is varied from negative to positive values, followed by a reverse sweep where the voltage is varied from positive to negative values. The transistor CV shows an increase in capacitance at the inversion region because of charge supplied from the source and drain. This behavior was observed to be present in the DFGFET device.

The CV characteristics showed two distinct features: (a) at lower voltage sweep, the hysteresis was observed to be anti-clockwise; (b) Whereas at high sweep voltages, the hysteresis was observed to be clockwise. As mentioned previously, lower sweep voltages results in a charge movement within the floating gates resulting in a volatile  $V_{FB}$  shift. During the lower voltage forward sweep, electrons from the BFG gets tunneled into TFG thereby reducing the  $V_{FB}$ ; whereas during negative sweep, electrons gets tunneled into BFG from TFG thereby increasing the  $V_{FB}$ . Hence at lower sweep voltages, the hysteresis was observed to be anti-clockwise.

However, during the forward sweep at very high positive CG bias, electrons gets tunneled from the channel into floating gates with the help of FN tunneling mechanism. This results in a significant increase in  $V_{FB}$  because of higher charge content at the floating gates. Similarly, during the negative sweep with a high negative bias at the CG, electrons are tunneled back to the substrate thereby reducing the  $V_{FB}$  significantly. Hence the non-volatile operations were evident where hysteresis was observed to be clockwise higher voltage sweeps. Note that the width of the hysteresis loop is also high in case of the non-volatile operation, which matched the prediction.

Figure 3.5 shows the variation of  $V_{FB}$  during forward and reverse sweep in the DFGFET memory device. The window between the forward and reverse sweep continually increases as the CG bias is swept up to 9 V. The increase in the window indicates a higher charge gets transferred from BFG into TFG with an increase in the CG bias voltage. After 9 V, the window starts decreasing, which signifies a gradual tunneling of electrons from the channel into the floating gates. The change in  $V_{FB}$  because of non-volatile operation starts dominating over the volatile  $V_{FB}$  shift. After 10.5 V, the position of  $V_{FB}$  during forward and reverse sweep interchange their position, which means the charge injected from the channel into the floating gates, due to FN tunneling, is significantly higher than the charge interchanging between the floating gates. This transition indicates the onset of non-volatile mode of operation. Thus, two distinct hysteresis

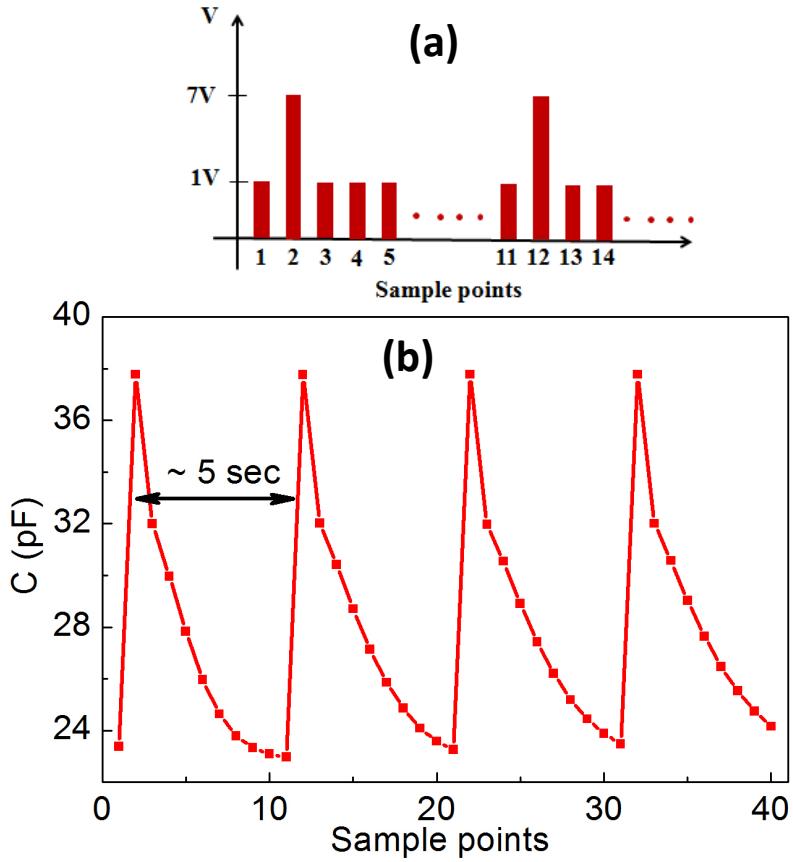


Figure 3.6: Volatile retention measurement performed in DFGFET: (a) Pulse CV scheme, and (b) capacitance at different sample points.

behavior at different sweep voltage confirms the presence of two different modes of operation.

CV analysis also gives the idea of all the operation voltages. The volatile operation is evident up to 9 V, whereas the completely non-volatile is evident after 10.5 V. These results were used to determine the programming and erasing voltage window for different modes of operation.

### 3.3 Volatile operation of the DFGFET device

As observed from the CV analysis, the volatile operation can be obtained with a relatively low voltage bias applied to the CG. Pulsed capacitance measurement was performed to study the volatile characteristics of the device. The volatile characteristic was evaluated with programming voltage was kept at 7 V, and the read was performed at 1 V, as shown in Figure 3.6 (a). Figure 3.6

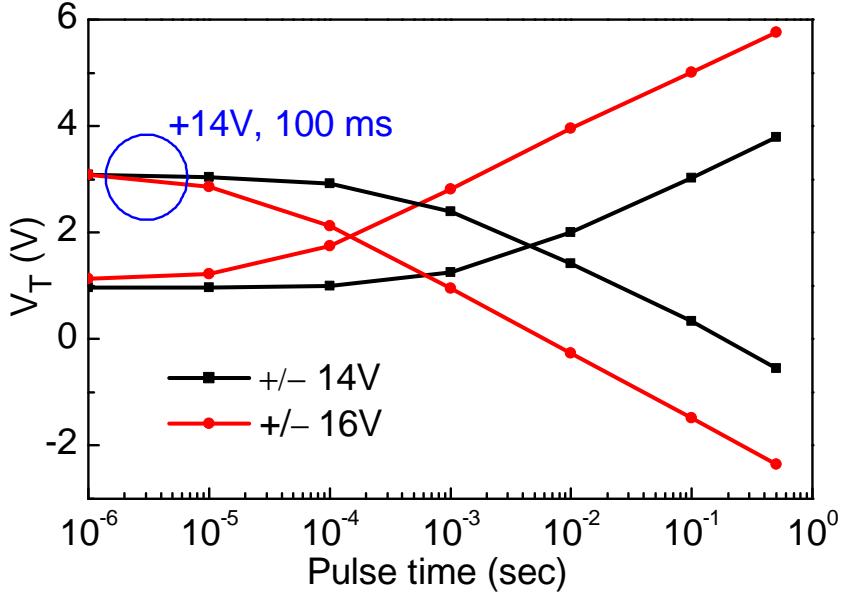


Figure 3.7: Non-volatile program/erase window of the DFGFET.

(b) shows the capacitance variation in the DFGFET after applying narrow programming and read pulses.

The pulsed CV measurement revealed the volatile operation in the DFGFET device. The capacitance was observed to increase after application of the 7 V pulse, and was observed to decay with a retention time of approximately 5 sec. This retention time is significantly higher than the ITRS roadmap for DRAM refresh time requirement forecast [16]. Note that the transistor received an anneal for mitigating sputter induced damage, and making S/D silicide contact. Such an anneal is known to degrade the dielectric quality thereby allowing higher charge leakage [17]. A similar device (MOS capacitor only) where the device didn't receive the post deposition anneal showed the volatile retention time of around 20 sec [18]. Current 1T1C DRAM technology suffers from charge leakage requiring refresh within 64 ms. Thus, the idea of controlling the volatile behavior using IMD can be attractive to increase the retention time.

### 3.4 Non-volatile operation of the DFGFET device

After analyzing the device with CV measurements, IV measurements were performed to study the non-volatile mode of operation. Program and erase voltage was kept more than 12 V in order to make assure the  $V_T$  shift caused by non-volatile operation is significantly higher than the volatile operation. Virgin devices were tested for observing the programming window for +14 V

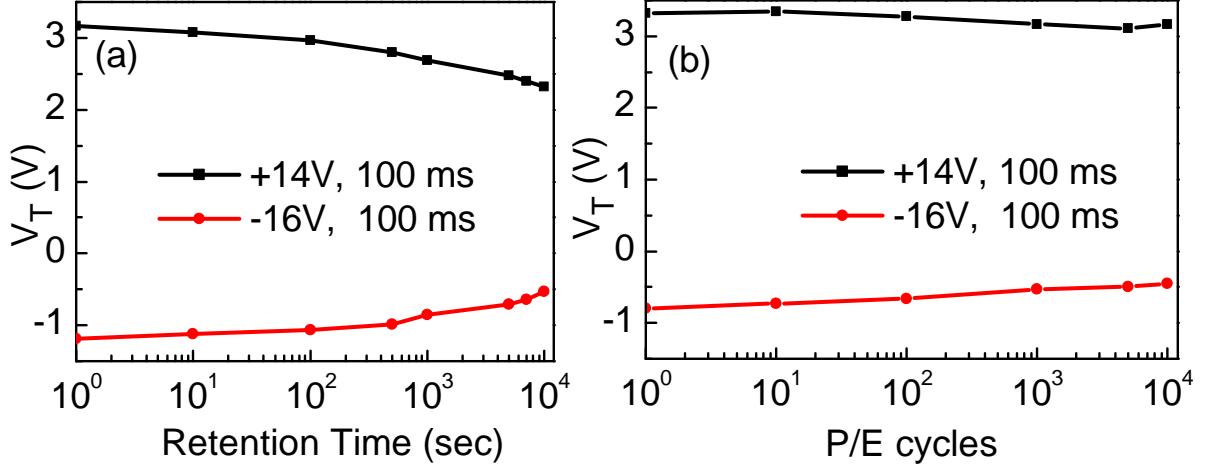


Figure 3.8: Non-volatile (a) retention and (b) endurance characteristics of the DFGFET.

and +16 V bias voltages at CG. During measurement,  $V_T$  of the transistor was measured before and after the application of the programming pulse. The programming window was obtained by varying the program pulse time. For evaluating the erase window, the devices were initially taken into non-volatile programmed state followed by erasing operation. Similarly, the  $V_T$  of the transistor was measured before and after application of the erase pulse, and erase window was obtained by varying the erase pulse time. Figure 3.7 shows the programming and erasing window of the device during non-volatile operation.

A rise in the  $V_T$  was observed with the programming pulse time. During erase, devices were initially programmed with a +14 V, 100 ms pulse and then the erase operations were performed. A reduction of  $V_T$  was observed during the erase operation. The non-volatile operation was observed to be similar to that of traditional flash memory. During programming, a significantly higher number of electrons are tunneled into the floating gates which in turn increase the  $V_T$  of the transistor. The charge at the floating gates balance themselves so that the band at the IMD is flat. Similarly, during erase, electrons are taken back into the substrate thereby reducing the  $V_T$  of the transistor.

Note that the programming slope is higher than erasing slope, which is a very common feature of flash devices [21, 22]. It has been mentioned that tunneling electrons from the substrate to the floating gate is easier compared to tunneling back from the floating gates. A detailed understanding of the non-volatile operation of the flash transistor can be found in the following Ph.D thesis [17, 23]. Several methods have been employed to address this issue. One of the easiest way to tackle this issue is to incorporate poly-Si floating gate which provides a higher

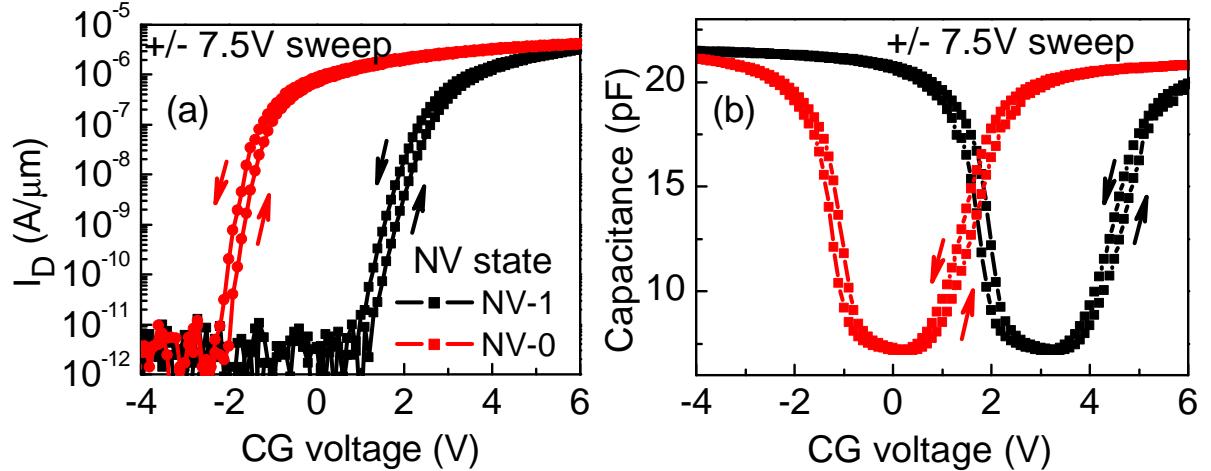


Figure 3.9: Demonstration of concurrent mode with (a) IV analysis, and (b) CV analysis.

density of state for holes. Thus, during erase operation, not only electrons get tunneled back to the substrate, holes gets tunneled into the floating gate thereby making a higher  $V_T$  shift. There are trade-off of different received improvement schemes (a detailed analysis of different tradeoffs can be found in [23]).

### 3.4.1 Reliability analysis of the non-volatile operation in DFGFET.

The reliability analysis of the non-volatile operation was carried out to understand device variability with successive operation and charge retention characteristics. Figure 3.8 (a) shows the non-volatile retention behavior of the device at room temperature. A decrease in the  $V_T$  window was observed signifying the charge loss from the floating gates with time elapsed. Non-volatile memories need a retention time of more than 10 years. This charge loss in DFGFET can be further improved using a better quality thermal oxide as TO candidate. The device under experiment has ALD SiO<sub>2</sub> which is known to be of poorer quality compared to the thermally grown oxide.

Next, the non-volatile endurance analysis was performed. During endurance measurement, the device was programmed with +14 V, 100 ms pulse; and erased with -16 V, 100 ms pulse.  $V_T$  of the transistor was measured at the beginning, after  $10^1$  cycles,  $10^2$  cycles,  $10^3$  cycles,  $5 \times 10^3$  cycles and  $10^4$  cycles. Figure 3.8 (b) shows the non-volatile endurance behavior of the device. Excellent  $V_T$  was observed even after  $10^4$  cycles of programming and erasing. This signifies excellent flash performance of the DFGFET device.

### 3.5 Concurrent mode of operation in DFGFET

As mentioned earlier, the most promising feature of this device is the concurrent mode of operation where the device can show simultaneous non-volatile and volatile mode operation. This feature was analyzed using both IV and CV measurements. Figure 3.9 (a) shows the concurrent mode demonstration using CV measurement. The device was initially taken into non-volatile programmed or erased state, a low voltage sweep ( $\pm 7.5$  V) was performed to study the hysteresis. Non-volatile programmed state, represented as NV-1 state, was achieved with a pulse of +14 V, 100 ms duration at the CG. Similarly non-volatile erased state, represented as NV-0 state, was achieved with a pulse of -16 V, 100 ms duration at the CG. Anticlockwise hysteresis was observed for  $\pm 7.5$ V sweep at both non-volatile program and erased state. This confirms the presence of volatile mode of the DFGFET irrespective of the non-volatile programmed or erased state.

Similarly, IV measurements also showed the existence of concurrent mode of operation, shown in Figure 3.9 (b). All these observations confirm the simultaneous volatile and non-volatile operation of the dual floating gate memory MOSFET.

### 3.6 Challenges of DFGFET

A Major challenge of DFGFET is the loss of the volatile data after every non-volatile operation. During the non-volatile operation, a high voltage bias is applied at the CG. This high voltage bias can disturb the charge stored at the floating gates corresponding to the volatile operation. Thus, the information of volatile memory data is completely lost after a high voltage operation. However, the controller can be programmed to restore the volatile data after every non-volatile operation. The details of volatile bit restoratoin scheme can be found in *Daniel*. Moreover, DFGFET volatile mode operation is carried out by re-distributing the charge between the floating gates with direct tunneling mechanism. Tunneling is known to be a slow process, thus the volatile operation of DFGFET is expected to be much slower than that of conventional 1T-1C DRAM. Some of the key challenges of DFGFET are mentioned below.

- DFGFET uses tunneling mechanism for the volatile operation, and such mechanism is known to be slow. Hence a candidate is required which can switch at nanosecond speed.
- During non-volatile operations, the dielectric stack of DFGFET gets sufficiently higher electric field which can easily damage the thin IMD separating the floating gates. Hence the charge transfer between the floating gate will be balanced in no time thereby resulting in a complete loss of volatile operation.

- It is well known that programming/erasing disturb in NAND/NOR can result in a  $V_T$  shift by more than a volt in modern device architectures. Hence a very high-k dielectric will be required as the volatile transition element in order to avoid the fringing effect of nearby cell.

However, more studies can be carried out to address the challenges of DFGFET. This chapter provides a proof-of-concept idea of a unified memory device capable of simultaneous volatile and non-volatile operation. Successive chapters of the thesis will focus on resolving some of the issues faced by DFGFET. Thus, this chapter provides a valuable insight of a unified memory candidate that can be attractive for future embedded memory applications.

### **3.7 Summary**

A dual floating gate flash memory transistor was fabricated and characterized and was shown to have three independent modes of operation. The volatile memory behavior relies on choice of IMD as well as its thickness. Non-volatile memory behavior is similar to conventional flash memory. The third state of operation is the concurrent mode of operation, where volatile memory and non-volatile memory can be realized simultaneously within the same transistor. This mode of operation makes this device a promising candidate for future low power, area efficient memory architectures. However, some of the challenges of the device asks for further device improvement. Subsequent chapters discusses the feasibility of using RRAM and integrating it into the flash for unified memory operation.

## REFERENCES

- [1] V. Eijkhout, "Introduction to High Performance Scientific Computing," available online at <http://pages.tacc.utexas.edu/~eijkhout/istc/istc.html>.
- [2] W. Kwon, "Novel Technologies for Next Generation Memory," *Ph.D Thesis, University of California Berkeley*, 2013.
- [3] C. Zitlaw, "The Future of NOR Flash Memory," *Memory Designline. UBM Media*, 2011.
- [4] C. C. Lin, Y. H. Wu, Y. S. Lin, M. L. Wu and L. L. Chen, "Electrical Characteristics for Flash Memory With Germanium Nitride as the Charge-Trapping Layer, C. C. Lin, IEEE Transactions on Nanotechnology, vol. 12, no. 3, pp. 436-441, 2013.
- [5] K. Ramkumar, I. Kouznetsov, V. Prabhakar, K. Shakeri, X. Yu, Y. Yang, L. Hinh, S. Lee, S. Samanta, H. M. Shih, S. Geha, P. C. Shih, C. C. Huang, H. C. Lee, S. H. Wu, J. H. Gau and Y. K. Sheu, "A scalable, low voltage, low cost SONOS memory technology for embedded NVM applications," *IEEE International Memory Workshop (IMW)*, pp. 199-202, 2013.
- [6] G. V. Bosch, A. Arreghini, L. Breuil, A. Cacciato, T. Schram, A. Suhane, M. B. Zahid, M. Jurczak and J. Van Houdt, "Understanding the impact of metal gate on TANOS performance and retention," *IEEE International Memory Workshop (IMW)*, pp. 1-4, 2010.
- [7] Z. Xu, Z. Huo, C. Zhu, J. Liu and M. Liu, "Performance optimization for TANOS by using pre-treatment of plasma oxygenic ions," *IEEE Non-Volatile Memory Technology Symposium (NVMTS)*, pp. 1-4, 2011.
- [8] M. G. Ertosun, K. Y. Lim, C. Park, J. Oh, P. I. Kirsch and K. C. Saraswat, "Novel Capacitorless Single-Transistor Charge-Trap DRAM (1T CT DRAM) Utilizing Electrons," *IEEE Electron Device Letters*, vol. 31, no. 5, pp. 405-407, 2010.

- [9] J. T. Lin, P. H. Lin, Y. C. Eng and Y. R. Chen, "Novel Vertical SOI-Based 1T-DRAM With Trench Body Structure," *IEEE Transactions on Electron Devices*, vol. 60, no. 6, pp. 1872-1877, 2013.
- [10] V. Asthana, M. Kar, J. Jimenez, J. P. Noel, S. Haendler and P. Galy, "Circuit optimization of 4T, 6T, 8T, 10T SRAM bitcells in 28nm UTBB FD-SOI technology using back-gate bias control," *European Solid-State Circuits Conference (ESSCIRC)*, pp.415-418, 2013.
- [11] V. Alves and K. Souri, "In-Situ Processing Bringing Intelligence to Storage," *Flash Memory Summit*, 2014.
- [12] B. Prince, "Vertical 3D Memory Technologies," *John Wiley and Sons Ltd*, 2014.
- [13] Q. Zhu, B. Akin, H. E. Sumbul, F. Sadi, J. C. Hoe, L. Pileggi and F. Franchetti, "A 3D-stacked logic-in-memory accelerator for application-specific data intensive computing," *IEEE 3D Systems Integration Conference (3DIC)*, pp.1-7, 2013.
- [14] D. Schinke, N. Di Spigna, M. Shiveshwarkar and P. Franzon, "Computing with novel floating-gate devices," *Computer*, vol. 44, no. 2, pp. 29-36, 2011.
- [15] Y. Taur and T. H.Ning, "Fundamentals of Modern VLSI Devices," *Cambridge University Press*, 2009.
- [16] International technology roadmap for semiconductors (ITRS) 2011 edition, available online at <http://www.itrs.net/>.
- [17] S. Jayanti, "Investigation of High-k Dielectrics and Metal Gate Electrodes for Non-volatile Memory Applications," *Ph.D Thesis, North Carolina State University*, 2012.
- [18] N. Di Spigna, D. Schinke, S. Jayanti, V. Misra and P. Franzon, "A novel double floating-gate unified memory device, N. Di Spigna, IEEE/IFIP International Conference on VLSI and System-on-Chip (VLSI-SoC)," pp. 53-58, 2012.

- [19] B. Sarkar, S. Jayanti, N. Di Spigna, B. Lee, V. Misra and P. Franzon, “Investigation of intermediate dielectric for dual floating gate MOSFET,” *IEEE Non-Volatile Memory Technology Symposium (NVMTS)*, pp. 1-4, 2013.
- [20] B. Sarkar, N. Ramanan, S. Jayanti, N. Di Spigna, B. Lee, P. Franzon, and V. Misra, “Dual floating gate unified MOSFET with simultaneous volatile and non-volatile operation,” *IEEE Electron Device Letters*, vol. 35, no. 1, pp. 48-50, 2014.
- [21] S. Jayanti, X. Yang and V. Misra, “Investigation of Thermal Stability of High-k Interpoly Dielectrics in TaN Metal Floating Gate Memory Structures,” *IEEE International Memory Workshop (IMW)*, pp. 1-4, 2011.
- [22] S. Jayanti, X. Yang, R. Suri and V. Misra, “Ultimate scalability of TaN metal floating gate with incorporation of high-k blocking dielectrics for flash memory applications,” *IEEE International Electron Device Meeting (IEDM)*, pp. 106-109, 2010.
- [23] S. Raghunathan, “Scaled Planar Floating-Gate NAND Flash Memory Technology: Challenges and Novel Solutions,” *Ph.D Thesis, Stanford University*, 2010.
- [24] D. J. Schinke, “Computing with Novel Floating Gate Devices,” *Ph.D Thesis, North Carolina State University*, 2011.

## Chapter 4

# Engineering the Forming and Reset Mechanism in RRAM

Flash memories have been scaled aggressively for last several decades to improve the performance. Recently a transition from planar-NAND to 3D NAND architecture has been demonstrated resulting in a tremendous advantage on integration over area advantage [1]. However, there are certain limitations which need to be addressed before going to the sub 14 nm technology nodes. One prime concern is the number of electrons transferred to the floating gate during program operation, and vice-versa [2]. Below 20 nm technology, a 100 mV  $V_T$  shift is caused by just few electrons. Hence the charge loss is a critical issue where losing a few electron results in a significant shift in  $V_T$  window. Apart from the number of electrons, process complexity is another issue with scaling of NAND flash memories.

Similarly, NOR flash has been facing scaling issues beyond 45 nm node [3]. One important scaling concerns of NOR is the low punch-through voltage due to gate length scaling. During programming operation or NOR cell, channel hot electrons are injected into floating gate. The electron has to gain enough energy to cross the Si/SiO<sub>2</sub> barrier. With scaling, the voltage windows are going down, making the electrons harder to overcome the barrier [4, 5]. Another important issue of NOR cell is the random telegraph noise (RTS)  $V_T$  variation in flash cells due to trapping or detrapping of electrons from the Si/SiO<sub>2</sub> interface [6].

All these issues have resulted in great interest of research in emerging memory technology devices that can be used as replacement of flash memories in NAND as well as NOR architecture [8]. Among different emerging memory candidates, RRAM is the only candidate which is completely CMOS compatible and 3D stackable. They also offer very high speed switching and very high endurance, making them an attractive candidate for the future DRAM replacement

[7, 9]. However, several models have been proposed for RRAM operation, thus the complete understanding of the memory operations in RRAM is still under research [10]. Another issue is the device to device variability and poor switching uniformity [11, 12]. More importantly, most of these memories need an activation step known as forming operation before they can be used in memory circuit [13]. Mitigating all these issues requires a solid understanding of the RRAM devices.

In this dissertation, the primary goal in studying RRAM devices is to find ways which can enable incorporation of RRAM into flash memories resulting in a superior unified memory compared to the DFGFET. To understand the RRAM operations, two different experimental methods were designed. This chapter is focused to understand the forming process of the RRAM and applying novel techniques to improve the forming voltage requirement. Next chapter is then focused to understand the reset operation of RRAM. Lets start with an introduction to RRAM devices.

## 4.1 Introduction to the memory operations in RRAM

As mentioned in the introduction chapter, RRAM devices are simple metal-insulator-metal capacitors which can show resistive switching properties. By applying bias in one of the electrode, oxygen vacancies are generated in the dielectric [13]. At certain bias, these vacancies forms a complete conducting filament between the electrodes thereby resulting in a low resistance state (LRS). This filament can be broken to realize the high resistance state (HRS). The memory operations are realized with three main steps, namely forming, reset and set operation.

### 4.1.1 Forming

Forming operation refers to the creation of the filament for the first time in the RRAM device [14]. Most of the CMOS compatible dielectrics used in RRAM generate vacancies with application of bias in the electrode. At sufficient bias voltage, vacancies form a complete conductance path between the electrodes thereby attaining the LRS state. Interestingly, some of the high-k dielectrics don't need a forming operation [15, 16]. Hence a complete understanding of the forming process is still under the research phase.

### 4.1.2 Reset

Once the dielectric reach its LRS, the high resistance state (HRS) can be obtained with a reset process which ruptures the filament [13]. It is believed that the oxygen ions recombine with the

vacancies thereby breaking the filament and leading to HRS state. The reset process has been observed to be unipolar and bipolar in different dielectric.

- In unipolar RRAM, the reset bias polarity is same as that of set or forming operation. Some of the oxides which shows unipolar operation are NiO, HfO<sub>2</sub>, ZnO, etc [14].
- In bipolar RRAM, opposite polarity bias is applied during reset compared to the forming operation. Most of the high-k dielectric shows bipolar operation [14].

#### 4.1.3 Set

The set operation refers to bringing back the LRS state of the dielectric [20]. This is achieved with a much lower bias compared to the forming process. It is well known that all the vacancies generated during forming operation are not annihilated during the reset operation, hence set operation requires much lower bias voltage to generate additional vacancies thereby realizing the LRS state. Successive memory operations in RRAM are then carried out with set and reset operations.

Although HfO<sub>2</sub> is the most explored RRAM candidate, recent reports on bilayer dielectric RRAM has been observed to show superior performance over the single layer dielectric RRAM. For example, Al<sub>2</sub>O<sub>3</sub>/HfO<sub>2</sub> bilayer dielectric RRAM is attracting the research interest because of several advantages over HfO<sub>2</sub> RRAM, like better switching uniformity [17], larger memory window [18], better read disturb immunity [19], lower random fluctuation of conducting filament formation process [20], etc. Interestingly, Al<sub>2</sub>O<sub>3</sub> have lower dielectric constant than HfO<sub>2</sub>, thus a higher electric field will drop across the Al<sub>2</sub>O<sub>3</sub> layer resulting in several unique properties of bilayer dielectric RRAM.

In this chapter, we tried to understand the implications of adding thin layer of Al<sub>2</sub>O<sub>3</sub> into HfO<sub>2</sub> RRAM thereby making a bilayer dielectric RRAM. We observed a significant improvement in forming voltage in Al<sub>2</sub>O<sub>3</sub>/HfO<sub>2</sub> bilayer dielectric RRAM compared to HfO<sub>2</sub> RRAM. A higher  $I_{on}/I_{off}$  ratio was also observed in bilayer dielectric RRAM compared to HfO<sub>2</sub> RRAM. A thorough study of individual dielectric was performed to understand these advantages, mentioned in the following sections.

In the experimental side, HfO<sub>2</sub> and Al<sub>2</sub>O<sub>3</sub>/HfO<sub>2</sub> bilayer dielectric RRAMs were fabricated with tungsten (W) electrodes. 100 nm of W was deposited on a Si/SiO<sub>2</sub> wafer with 10 nm Ti adhesion layer using UHV RF-sputtering system. All the dielectrics were deposited using the ALD technique at 200 °C. 100 nm of W serving as top electrode was deposited using UHV RF-sputtering system and was patterned using conventional lithography technique. Schematic

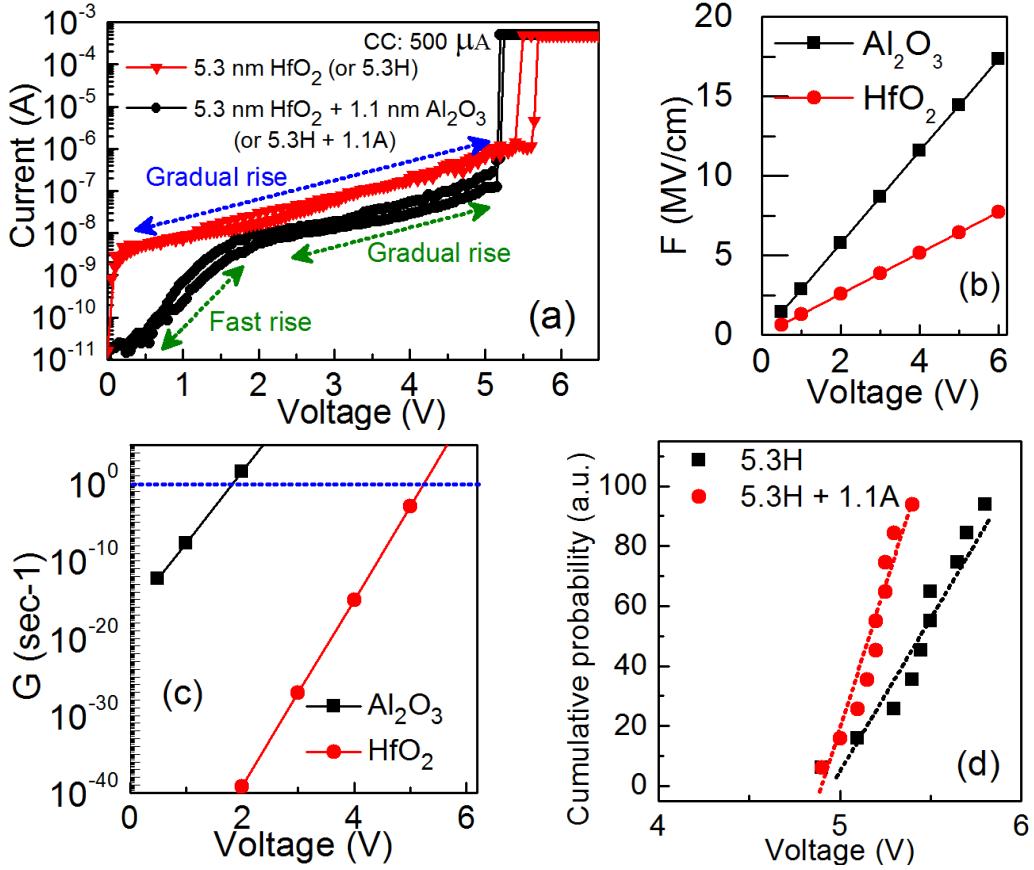


Figure 4.1: (a) Forming characteristics of 5.3 nm HfO<sub>2</sub> and 5.3 nm HfO<sub>2</sub> + 1.1 nm Al<sub>2</sub>O<sub>3</sub> RRAM; simulated (b) electric field and (c) vacancy generation rate inside individual layers of Al<sub>2</sub>O<sub>3</sub>/HfO<sub>2</sub> bilayer dielectric RRAM; (d) forming voltage uniformity in HfO<sub>2</sub> RRAM and Al<sub>2</sub>O<sub>3</sub>/HfO<sub>2</sub> bilayer dielectric RRAM

of the fabrication process flow can be found in Appendix-A. All devices had dimensions of 100 x 100 m<sup>2</sup>. The measurements were performed using HP 4155b semiconductor parameter analyzer and the simulations were performed using Sentaurus-TCAD.

## 4.2 Understanding the forming process in Al<sub>2</sub>O<sub>3</sub>/HfO<sub>2</sub> bilayer dielectric RRAM

As mentioned earlier, forming process is an essential step of RRAM with which the conductive filament is created. Figure 4.1 (a) shows the forming characteristics of 5.3 nm HfO<sub>2</sub> RRAM and 5.3 nm HfO<sub>2</sub> + 1.1 nm Al<sub>2</sub>O<sub>3</sub> bilayer dielectric RRAM. A gradual rise in the forming

current leading to breakdown at around 5.5 V was observed in case of HfO<sub>2</sub> RRAM. On the other hand, a two-step rise in current as well as an early breakdown at around 5.1 V was observed in case of Al<sub>2</sub>O<sub>3</sub>/HfO<sub>2</sub> bilayer dielectric RRAM. A Similar two-step rise has been observed in forming characteristics of other bilayer RRAMs [21]. The reduction in forming voltage with bilayer strategy has also been observed for other dielectric combinations [37]. Thus, the bilayer dielectric RRAM not only showed a different forming characteristics, it also resulted in a decrease in forming voltage even after an increase in the physical thickness (or even EOT) of the dielectric.

In this work, a thorough understanding of this behavior has been carried out to understand the contribution of individual layer of bilayer dielectric in reducing the forming voltage. This reduction in forming voltage is assisted by two main factors, viz. (i) Contribution of electric field inside the individual layer of bilayer dielectric RRAM, and (ii) contribution of electric field near the oxygen vacancies.

#### 4.2.1 Contribution of electric field inside the individual layer of bilayer dielectric RRAM

Since the dielectric constant of Al<sub>2</sub>O<sub>3</sub> is lower than that of HfO<sub>2</sub>, this results in a significantly higher electric field inside the Al<sub>2</sub>O<sub>3</sub> layer of the bilayer dielectric RRAM. The relation between the electric field and dielectric constant are given as [38]

$$\frac{k_{HfO_2}}{K_{Al_2O_3}} = \frac{F_{Al_2O_3}}{F_{HfO_2}} \quad (4.1)$$

Where  $k$  is the dielectric constant and  $F$  is the electric field inside the dielectric. The values of  $F_{Al_2O_3}$  and  $F_{HfO_2}$  for a bilayer dielectric RRAM can be obtained by solving the Gauss law. The two-step forming behavior in Al<sub>2</sub>O<sub>3</sub>/HfO<sub>2</sub> bilayer dielectric RRAM can possibly be explained by investigating the vacancy generation process in the individual layer of the bilayer dielectric RRAM. The vacancy generation rate in a dielectric due to applied bias at the electrode is given by the equation [22]

$$G = G_0 \exp\left(-\frac{E_a - b.F}{KT}\right) \quad (4.2)$$

where  $G$  represents the vacancy generation rate,  $G_0$  represents the effective vibration frequency between metal and oxygen ion,  $b$  is the bond polarization factor. The evaluation of  $E_a$  for Al<sub>2</sub>O<sub>3</sub> and HfO<sub>2</sub> were obtained experimentally and observed to be 1.8 eV and 4.6 eV respectively [23], and both the values are similar to the previously reported values [24]. Experimental steps for determining the values are mentioned in Appendix-B of this dissertation.

Table 4.1: Parameters of RRAM dielectric for vacancy simulation.

Parameter	$\text{HfO}_2$	$\text{Al}_2\text{O}_3$
$k$	18	8
$E_a$ (eV)	4.6	1.8
$b$ (e cm)	$5.6 \times 10^{-7}$	$1.9 \times 10^{-7}$
$G_0$ ( $\text{sec}^{-1}$ )	$7 \times 10^{13}$ [26]	$2 \times 10^{13}$ [25]

Table 4.1 shows all the parameters of individual layer of the bilayer dielectric RRAM. A lower dielectric constant of  $\text{Al}_2\text{O}_3$  than  $\text{HfO}_2$  results in a higher electric field across the  $\text{Al}_2\text{O}_3$  layer of the bilayer dielectric, as shown in Figure 4.1 (b). This results in an early vacancy generation inside  $\text{Al}_2\text{O}_3$  layer of the  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM. A lower  $E_a$  in  $\text{Al}_2\text{O}_3$  also contributed to attain in a higher value of  $G$  in the  $\text{Al}_2\text{O}_3$  layer at lower bias, shown in Figure 4.3 (c). Hence the vacancies start originating in the  $\text{Al}_2\text{O}_3$  layer at very low applied bias. Simulation shows that the filament gets completely formed (when  $G$  approaches 0 [22]) in the  $\text{Al}_2\text{O}_3$  layer at around 2 V. After the filament formation in  $\text{Al}_2\text{O}_3$  layer, the vacancies starts originating at  $\text{HfO}_2$  layer. This is similar to a double-forming process where one layer of the bilayer dielectric RRAM gets the filament followed by the filament formation in the second layer [27]. Apart from reduction in forming voltage, the forming voltage uniformity was also observed to be improved in bilayer dielectric RRAM, shown in Figure 4.1 (d).

Note that the existing vacancies helps in creation of newer vacancies inside the dielectric due of higher electric field around them [22]. Thus, once the filament gets formed in the  $\text{Al}_2\text{O}_3$  layer at around 2 V, the vacancies help in generating additional vacancies in the  $\text{HfO}_2$  layer and the filament progresses toward the  $\text{HfO}_2$  layer. Comparing this with single layer 5.3 nm  $\text{HfO}_2$  dielectric RRAM, 2 V corresponds to an electric field of 3.77 MV/cm and a generation rate of around  $10^{-29}$  vacancies per second, which is significantly low to generate additional vacancies inside the dielectric. Hence it is unlikely that  $\text{HfO}_2$  RRAM generates additional vacancies at lower applied bias, whereas  $\text{Al}_2\text{O}_3$  layer of the  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM gets the filament formed at the same bias thereby providing a lower forming voltage in  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM.

#### 4.2.2 Poole-Frenkel analysis of the forming currents

In order to reconfirm the vacancy generation hypothesis in the forming characteristics of Figure 4.1, trap-assisted Poole-Frenkel (PF) analysis of the forming currents was performed, shown in Figure 4.2. It is known that trap-assisted tunneling is the dominant current conduction mech-

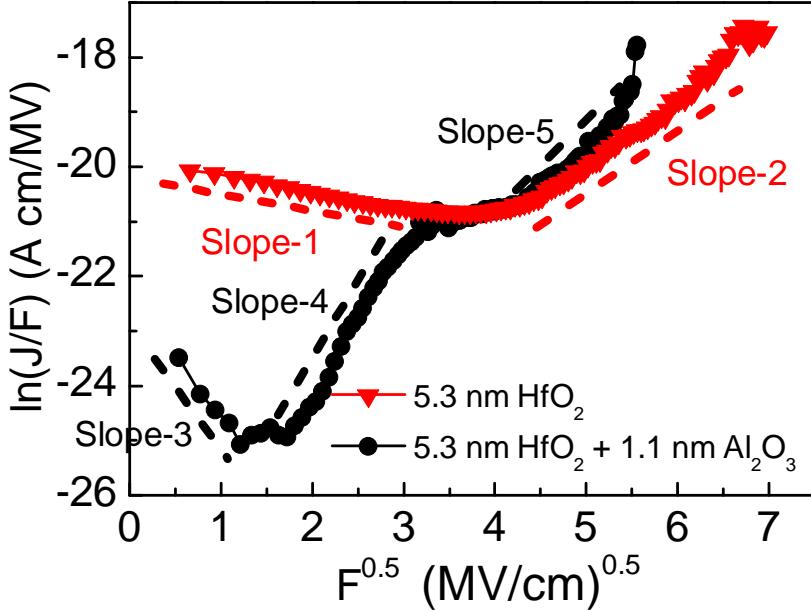


Figure 4.2: Trap assisted Poole-Frenkel analysis of the forming currents shown in Figure 4.1.

anism in RRAM because of existing traps inside the dielectric and additional vacancy generation during the voltage sweep [28]. The  $\text{HfO}_2$  RRAM PF-analysis showed two slopes (slope-1 and slope-2), whereas the  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM PF-analysis showed three slopes (slope-3, slope-4 and slope-5). The initial slope for both the RRAMs (slope-1 and slope-3) were observed to be negative, which is possibly due to tunneling through the existing traps [29]. Slope-2 of  $\text{HfO}_2$  RRAM was observed to be positive and this possibly reflects the vacancy generation inside the dielectric which enhances the current conduction.

However,  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM showed two positive slopes, a positive slope (slope-4) at lower bias followed by a second positive slope (slope-5) at higher bias. Comparing these slopes with the vacancy generation rate mentioned in Figure 4.3 (b), slope-4 possibly corresponds to the vacancy generation in  $\text{Al}_2\text{O}_3$  layer, whereas slope-5 corresponds to the vacancy generation in  $\text{HfO}_2$  layer of the  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM. Also, slope-5 was observed to be higher than slope-2, which signifies that the vacancy generation in  $\text{HfO}_2$  layer is faster in case of  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM compared to  $\text{HfO}_2$  RRAM.

#### 4.2.3 Electric field enhancement in bilayer dielectric RRAM

To understand the difference of slopes in  $\text{HfO}_2$  dielectric, filament of the same size and position were simulated for 5.3 nm  $\text{HfO}_2$  RRAM and 1.1 nm  $\text{Al}_2\text{O}_3 + 1.1$  nm  $\text{HfO}_2$  bilayer dielectric

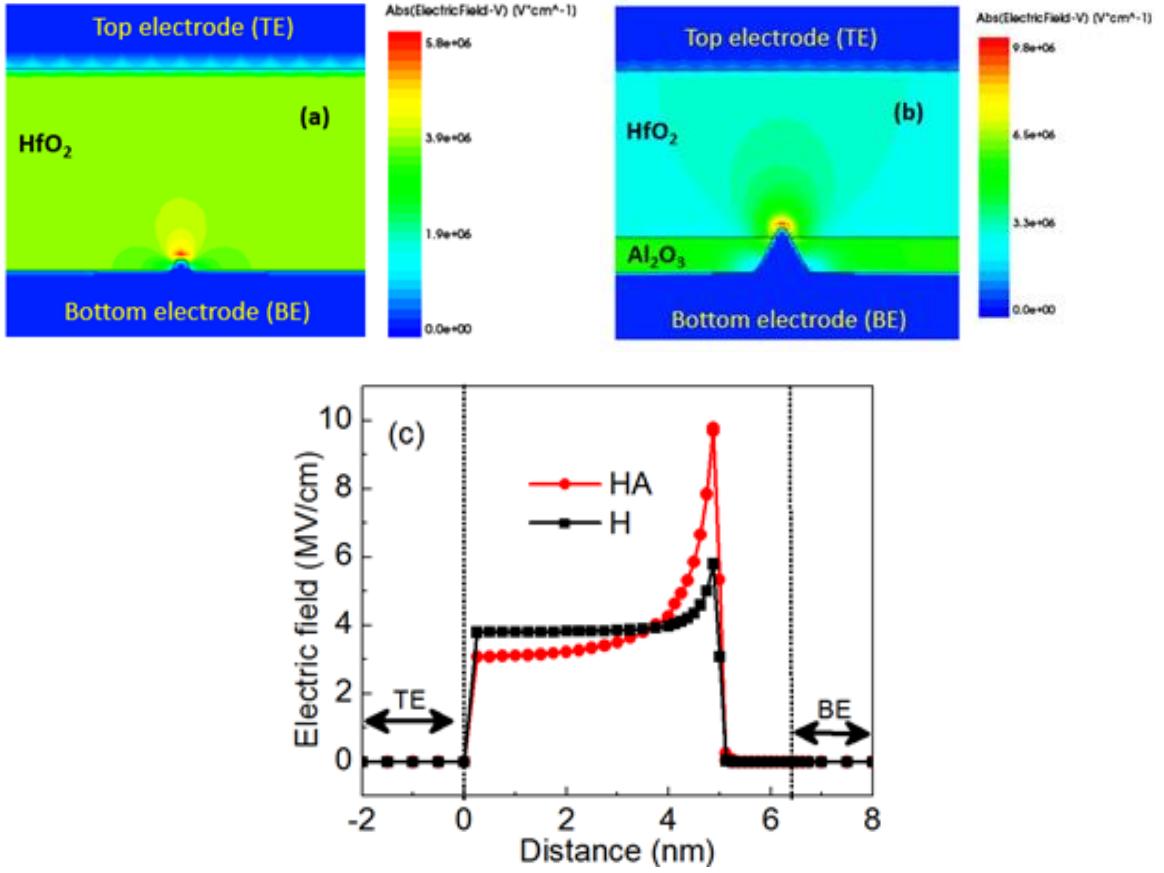


Figure 4.3: Electric field distribution inside the dielectric at 2 V; (a) 5.3 nm HfO<sub>2</sub> RRAM with a filament formed near the bottom electrode; (b) filament formed in 5.3 nm HfO<sub>2</sub> + 1.1 nm Al<sub>2</sub>O<sub>3</sub> RRAM near the bottom electrode; and (c) electric field along the vertical axis at the center of the filaments.

RRAM, shown in Figure 4.3. It should be noted that every vacancy generates a positive charge when an oxygen atom is kicked off. However, the charge imbalance is countered by charge in the electrode, and electrons from bottom electrode can easily move into the dielectric thereby maintaining the charge neutrality. Hence all the simulations were performed maintaining the charge neutrality. In Figure 4.3 (a), a bilayer dielectric was simulated and the filament was considered to have formed in the Al<sub>2</sub>O<sub>3</sub> layer, and started forming in the HfO<sub>2</sub> layer. Whereas in Figure 4.3 (b), a similar filament shape to that of Figure 4.3 (a) was considered for single layer HfO<sub>2</sub> RRAM. Both the cases were tested for 2 V bias. Figure 4.3 (c) shows the electric field along the vertical-axis at the center of filament for both the cases.

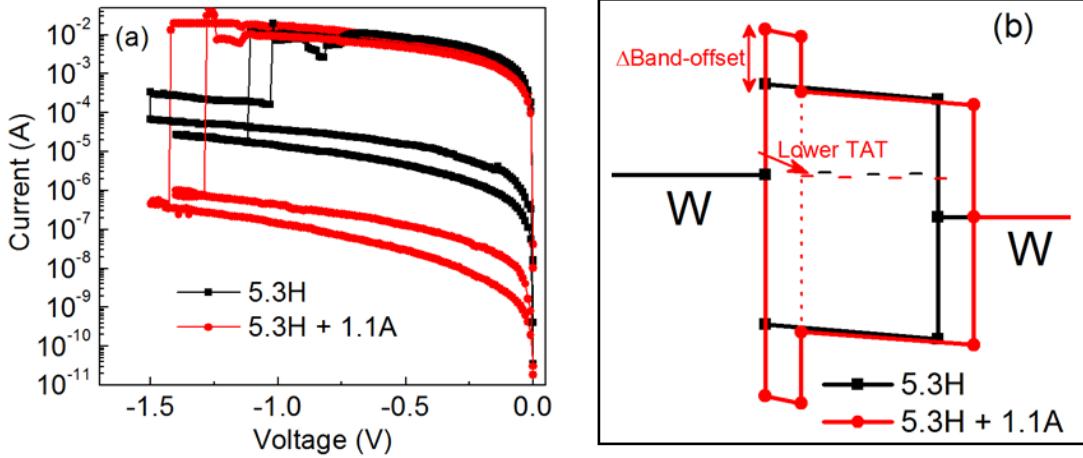


Figure 4.4: (a) Reset comparison between  $\text{HfO}_2$  RRAM and  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM; (b) band diagram showing the advantage of  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM over  $\text{HfO}_2$  RRAM.

It was observed that the electric field near the vacancy edge in  $\text{HfO}_2$  is significantly higher in the  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric as compared to single layer  $\text{HfO}_2$ , indicating that the vacancy generation near the existing vacancies is faster in case of  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM as compared to  $\text{HfO}_2$  RRAM. This in turn provides a higher slope in PF-analysis of  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM as compared to  $\text{HfO}_2$  RRAM, hence slope-5 was observed to be higher than slope-2 in Figure 4.1. Thus, bilayer dielectric RRAM allows to control the electric field distribution inside the individual dielectric which can result in modulating the vacancy creation process thereby reducing the forming voltage requirement.

### 4.3 Comparison of reset in $\text{Al}_2\text{O}_3/\text{HfO}_2$ RRAM with $\text{HfO}_2$ RRAM

Finally, the reset characteristics of  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM were compared with  $\text{HfO}_2$  RRAM. The devices were initially set to  $500 \mu\text{A}$  CC limit.  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM also offers a higher  $I_{on}/I_{off}$  ratio after reset, shown in Figure 4.4 (a). This can be attributed to the higher band offset offered by the  $\text{Al}_2\text{O}_3$  layer when kept near the electron injection electrode [18]. Figure 4.4 (b) shows the band diagram of  $\text{HfO}_2$  RRAM and  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM. During reset, vacancies in both  $\text{HfO}_2$  and  $\text{Al}_2\text{O}_3$  layer recombines with oxygen ions, and a higher electric field in  $\text{Al}_2\text{O}_3$  layer results in higher vacancy recombination in  $\text{Al}_2\text{O}_3$  layer of  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RAM [20]. Since the trap-assisted tunneling (TAT) current is exponentially dependent upon the offset between trap energy and

conduction band edge [39], the additional band offset provided by the  $\text{Al}_2\text{O}_3$  layer was observed to be advantageous in providing a lower  $I_{off}$  after reset. Thus, the higher electric field in the  $\text{Al}_2\text{O}_3$  layer of  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM was also observed to provide efficient reset resulting in a higher  $I_{on}/I_{off}$  ratio.

Hence,  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM provides an efficient reset due to higher electric field in  $\text{Al}_2\text{O}_3$  layer thereby providing a better  $I_{on}/I_{off}$  ratio over  $\text{HfO}_2$  RRAM. Hence  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM is a promising RRAM candidate offering a lower forming voltage, better forming voltage uniformity and a higher  $I_{on}/I_{off}$  ratio over  $\text{HfO}_2$  RRAM.

#### 4.4 Summary

A systematic study on forming and reset characteristics of  $\text{HfO}_2$  RRAM and  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric was carried out to understand the advantages of  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM over  $\text{HfO}_2$  RRAM. It was observed that  $\text{Al}_2\text{O}_3$  layer of the bilayer dielectric RRAM helps in early vacancy generation thereby lowering the forming voltage of the bilayer dielectric RRAM compared to  $\text{HfO}_2$  RRAM. An efficient reset at  $\text{Al}_2\text{O}_3$  layer lowers the current after reset thereby providing a higher  $I_{on}/I_{off}$  ratio in the  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM. Hence  $\text{Al}_2\text{O}_3/\text{HfO}_2$  bilayer dielectric RRAM was observed to be advantageous over  $\text{HfO}_2$  RRAM providing a lower forming voltage and higher  $I_{on}/I_{off}$  ratio.

## REFERENCES

- [1] J. H. Yoon, “3D NAND Technology Implications to Enterprise Storage Applications,” *Flash memory summit*, 2014.
- [2] J. H. Yoon and G. A. Tressler, “Advanced flash technology status, scaling trends and implications to enterprise SSD technology enablement,” *Flash memory summit*, 2012.
- [3] W. H. Kwon, Y. H. Song, Y. Cai and S. P. Shim, “A Novel Structure for Beyond-45 nm NOR Flash Technology Featuring Short Channel Effect Immunity and Low Random Telegraph Signal Noise,” *Japanese Journal of Applied Physics*, vol. 47, no. 12, pp. 88028804, 2008.
- [4] W. J. Tsai, T. F. Ou, J. S. Huang, C. H. Cheng, C. Y. Lu, T. Wang, K. F. Chen, T. T. Han, T. C. Lu, K. C. Chen and C. Y. Lu, “A highly punchthrough-immune operation method for an ultra-short-channel hot-carrier-injection type non-volatile memory cell,” *IEEE International Electron Device Meeting (IEDM)*, pp. 1-4, 2008.
- [5] W. J. Tsai, T. F. Ou, H. I. Kao, E. K. Lai, J. S. Huang, L. H. Chong, Y. Y. Liao, S. P. Hong, M. T. Wu, S. C. Tsai, C. H. Leng, F. H. Hsu, S. Y. Wang, C. M. Cheng, T. Luoh, Y. T. Hung, S. A. Luo, C. H. Huang, C. Y. Lu, T. Yang, K. C. Chen and C. Y. Lu “A Novel Trapping-Nitride-Storage Non-Volatile Memory Cell Using a Gated-Diode Structure With an Ultra-Thin Dielectric Dopant Diffusion Barrier,” *IEEE Transactions on Electron Devices*, vol. 55, no. 8, pp. 2202 - 2211, 2008.
- [6] X. Yang, Z. Zheng, Y. Wang, Z. Huo, L. Jin, D. Jiang, Z. Wang, S. Chiu, H. Wu and M. Liu, “Gate Bias Dependence of Complex Random Telegraph Noise Behavior in 65-nm NOR Flash Memory,” *IEEE Electron Device Letters*, vol. 36, no. 1, pp. 26-28, 2015.
- [7] Y. Xie, “Modeling, architecture, and applications for emerging memory technologies,” *IEEE Design Test of Computers*, vol. 28, no. 1, pp. 44-51, 2011.
- [8] J. S. Meena, S. M. Sze, U. Chand and T. Y. Tseng, “Overview of emerging nonvolatile memory technologies,” *Nanoscale Research Letters*, vol. 9, 526, 2014.

- [9] R. Waser, R. Dittmann, G. Staikov and K. Szot, “Redox-Based Resistive Switching Memories Nanoionic Mechanisms, Prospects, and Challenges,” *Advanced Materials*, vol. 21, pp. 26322663, 2009.
- [10] A. Prakash, D. Jana and S. Maikap, “TaOx-based resistive switching memories: prospective and challenges,” *Nanoscale Research Letters*, vol. 8, 418, 2013.
- [11] X. Guan, S. Yu and H. S. P. Wong, IEEE, “On the Switching Parameter Variation of Metal-Oxide RRAMPart I: Physical Modeling and Simulation Methodology,” *IEEE Transactions on Electron Devices*, vol. 59, no. 4, pp. 1172-1182, 2012.
- [12] S. Yu, X. Guan and H. S. P. Wong, IEEE, “On the Switching Parameter Variation of Metal Oxide RRAMPart II: Model Corroboration and Device Design Strategy,” *IEEE Transactions on Electron Devices*, vol. 59, no. 4, pp. 1183-1188, 2012.
- [13] K. M. Kim, D. S. Jeong and C. S. Hwang, “Nanofilamentary resistive switching in binary oxide system; a review on the present status and outlook,” *Nanotechnology*, vol. 22, 254002, 2011.
- [14] H. S. P. Wong, H. Y. Lee, S. Yu, Y. S. Chen, Y. Wu, P. S. Chen, B. Lee, F. T. Chen and M. J. Tsai, “MetalOxide RRAM,” *Proceedings of the IEEE*, vol. 100, no. 6, pp. 1951-1970, 2012.
- [15] T. M. Pan and C. H. Lu, “Structural properties and electroforming-free resistive switching characteristics of GdO<sub>x</sub>, TbO<sub>x</sub>, and HoO<sub>x</sub> memory devices,” *Materials Chemistry and Physics*, vol. 139, pp. 437-442, 2014.
- [16] F. M. Lee, Y. Y. Lin, W. C. Chien, E. K. Lai, D. Y. Lee, C. C. Yu, H. H. Hsu, M. H. Lee, H. L. Lung, K. Y. Hsieh and C. Y. Lu, “A Low-cost, Forming-free WO<sub>x</sub> ReRAM Using Novel Self-aligned Photo-induced Oxidation,” *IEEE International Electron Devices Meeting (IEDM)*, pp. 546-549, 2013.

- [17] X.A. Tran, B. Gao, J. F. Kang, L. Wu, Z. R. Wang, Z. Fang, K. L. Pey, Y. C. Yeo, A. Y. Du, B. Y. Nguyen, M. F. Li and H. Y. Yu, "High Performance Unipolar AlOy/HfO<sub>x</sub>/Ni based RRAM Compatible with Si Diodes for 3D Application," *Symposium on VLSI Technology (VLSIT)*, pp. 44-45, 2011.
- [18] L. Goux, A. Fantini, G. Kar, Y. Y. Chen, N. Jossart, R. Degraeve, S. Clima, B. Govoreanu, G. Lorenzo, G. Pourtois, D. J. Wouters, J. A. Kittl, L. Altimime and M. Jurczak, "Ultralow sub 500nA operating current high-performance TiN/Al<sub>2</sub>O<sub>3</sub>/HfO<sub>2</sub>/Hf/TiN bipolar RRAM achieved through understanding-based stack-engineering," *Symposium on VLSI Technology (VLSIT)*, pp. 159-160, 2012.
- [19] Y. S. Chen, H. Y. Lee, P. S. Chen, P. Y. Gu, C. W. Chen, W. P. Lin, W. H. Liu, Y. Y. Hsu, S. S. Sheu, P. C. Chiang, W. S. Chen, F. T. Chen, C. H. Lien and M. J. Tsai, "Highly scalable hafnium oxide memory with improvements of resistive distribution and read disturb immunity," *IEEE International Electron Device Meeting (IEDM)*, pp. 105-108, 2009.
- [20] L. Chen, Y. W. Dai, Q. Q. Sun, J. J. Guo, P. Zhou and D. W. Zhang, "Al<sub>2</sub>O<sub>3</sub>/HfO<sub>2</sub> functional stack films based resistive switching memories with controlled SET and RESET voltages," *Solid State Ionics*, vol. 273, pp. 66-69, 2015.
- [21] A. Prakash, S. Maikap, W. Banerjee, D. Jana and C. S. Lai, "Impact of electrically formed interfacial layer and improved memory characteristics of IrO<sub>x</sub>/high-k<sub>x</sub>/W structures containing AlO<sub>x</sub>, GdO<sub>x</sub>, HfO<sub>x</sub>, and TaO<sub>x</sub> switching materials," *Nanoscale Research Letters*, vol. 8, 379, 2013.
- [22] L. Vandelli, A. Padovani, L. Larcher, and G. Bersuker, "Microscopic Modeling of Electrical Stress-Induced Breakdown in Poly-Crystalline Hafnium Oxide Dielectrics," *IEEE Transactions on Electron Devices*, vol. 60, no. 5, pp. 1754-1762, 2013.
- [23] B. Sarkar, B. Lee and V. Misra, "Implications of Lower Zero Field Activation Energy of Dielectric in Al<sub>2</sub>O<sub>3</sub>/HfO<sub>2</sub> Bilayer Dielectric RRAM Forming Process," *ECS Transactions*, vol. 64, no. 14, pp. 43-48, 2014.

- [24] J. McPherson, J. Y. Kim, A. Shanware and H. Mogul, "Thermochemical description of dielectric breakdown in high dielectric constant materials," *Applied Physics Letters*, vol. 82, pp. 2121-2123, 2003.
- [25] A. Padovani, L. Larcher, G. Bersuker and P. Pavan, "Charge Transport and Degradation in HfO<sub>2</sub> and HfO<sub>x</sub> Dielectrics," *IEEE Electron Device Letters*, vol. 34, no. 5, pp. 680-682, 2013.
- [26] S. Davis and G. Gutierrez, "Structural, elastic, vibrational and electronic properties of amorphous Al<sub>2</sub>O<sub>3</sub> from ab initio calculations," *Journal of Physics: Condense Matter*, vol. 23, 495401, 2011.
- [27] C. Y. Huang, C. Y. Huang, T. L. Tsai, C. A. Lin and T. Y. Tseng, "Switching mechanism of double forming process phenomenon in ZrO<sub>x</sub>/HfO<sub>y</sub> bilayer resistive switching memory structure with large endurance," *Applied Physics Letters*, vol. 104, 062901, 2014.
- [28] M. Specht, M. Stadele, S. Jakschik and U. Schroder, "Transport mechanisms in atomic-layer-deposited Al<sub>2</sub>O<sub>3</sub> dielectrics," *Applied Physics Letters*, vol. 84, no. 16, pp. 3076-3078, 2004.
- [29] J. H. Kim, V. A. Ignatova, J. Heitmann and L. Oberbeck, "Deposition temperature effect on electrical properties and interface of high-k ZrO<sub>2</sub> capacitor," *Journal of Physics D: Applied Physics*, vol. 41, 172005, 2008.
- [30] J. Robertson, "Band offsets of wide-band-gap oxides and implications for future electronic devices," *Journal of Vacuum Science & Technology B*, vol. 18, no. 3, pp. 1785-1791, 2000.
- [31] Y. S. Chen, H. Y. Lee, P. S. Chen, C. H. Tsai, P. Y. Gu, T. Y. Wu, K. H. Tsai, S. S. Sheu, W. P. Lin, C. H. Lin, P. F. Chiu, W. S. Chen, F. T. Chen, C. Lien and M. J. Tsai, "Challenges and opportunities for HfO<sub>x</sub> based resistive random access memory," *IEEE International Electron Devices Meeting (IEDM)*, pp. 31.3.1-31.3.4, 2011.

- [32] Y. Wu, B. Lee and H. S. P. Wong, "Al<sub>2</sub>O<sub>3</sub>-Based RRAM Using Atomic Layer Deposition (ALD) With 1-A RESET Current," *IEEE Electron Device Letters*, vol. 31, no. 12, pp. 1449-1451, 2010.
- [33] K. M. Kim, B. J. Choi, B. W. Koo, S. Choi, D. S. Jeong and C. S. Hwang, "Resistive Switching in Pt/Al<sub>2</sub>O<sub>3</sub>/TiO<sub>2</sub>/Ru Stacked Structures," *Electrochemical and Solid-State Letters*, vol. 9, no. 12, pp. G343-G346, 2006.
- [34] S. Yu, Y. Wu, Y. Chai, J. Provine and H. S. P. Wong, "Characterization of switching parameters and multilevel capability in HfO<sub>x</sub>/AlO<sub>x</sub> Bi-layer RRAM devices," *International Symposium on VLSI Technology, Systems and Applications (VLSI-TSA)*, pp. 1-2, 2011.
- [35] C. Y. Huang, J. H. Jieng, W. Y. Jang, C. H. Lin and T. Y. Tseng, "Improved Resistive Switching characteristics by Al<sub>2</sub>O<sub>3</sub> Layers Inclusion in HfO<sub>2</sub>-Based RRAM Devices," *ECS Solid State Letters*, vol. 2, no. 8, pp. 63-65, 2013.
- [36] M. Kimura, "Field and temperature acceleration model for time dependent dielectric breakdown," *IEEE Transactions on Electron Devices*, vol. 46, no. 1, pp. 220-229, 1999.
- [37] H. Zhang, N. Aslam, M. Reiners, R. Waser and S. Hoffmann-Eifert, "Atomic Layer Deposition of TiO<sub>x</sub>/Al<sub>2</sub>O<sub>3</sub> Bilayer Structures for Resistive Switching Memory Applications," *Chemical Vapor Deposition*, vol. 20, pp. 282-290, 2014.
- [38] Z. Chen, F. Zhang, B. Chen, Y. Zheng, B. Gao, L. Liu, X. Liu and J. Kang, "High performance HfO<sub>x</sub>/AlO<sub>y</sub> based resistive switching memory cross-point array fabricated by atomic layer deposition," *Nanoscale Research Letters*, vol. 10, 70, 2015.
- [39] M. Houssa, R. Degraeve, P. W. Mertens, M. M. Heyns, J. S. Jeon, A. Halliyal and B. Ogle, "Electrical properties of thin SiON/Ta<sub>2</sub>O<sub>5</sub> gate dielectric stacks," *Journal of Applied Physics*, vol. 86, no. 11, pp. 6462, 1999.

## Chapter 5

# Synaptic Learning in Binary Oxide RRAM

In the previous chapter, we looked into the non-volatile operation of RRAM devices. The forming and reset was observed to be abrupt, which can be used to realize one bit memory with LRS and HRS memory states. Interestingly, RRAMs also show analog memory behavior with which multi-bit operation per cell can be realized [1, 2]. This feature makes the RRAM a promising candidate to implement brain inspired computing systems.

### 5.1 Necessity of neuromorphic computation system

Neuromorphic computing is expected to emulate brain functions in the near future enabling the massive parallel computing and less complex architecture [3]. All the sensory operations of the human brain are performed by approximately  $10^{11}$  neurons and  $10^{15}$  synapses. In brains synaptic communication, neurons generate spikes which are then transmitted to the next neuron through the synapses. Each neuron is connected to other neurons with a complex network of  $10^3$  to  $10^4$  number of synapses. The responsibility of synapse is to carry those spike pulses by changing their connection strength as a result of neuronal activity, which is known as synaptic plasticity. The receiver neuron then analyses the synapses and information is passed on to that neuron. A similar operation has been realized with the help of electronic circuitry. However, all these circuits are made from Von-Neumann neural network architectures which poses huge requirement of processors, memory, cooling capacity, etc. For example, IBM Blue gene supercomputer can simulate Cats brain function, however, it has 147456 processors, 144 TB memory consuming a power of 1.5 MW [4]. One primary criteria for an electronic device to mimic the human brain is to show the spike time dependent plasticity behavior [5]. The details are mentioned below:

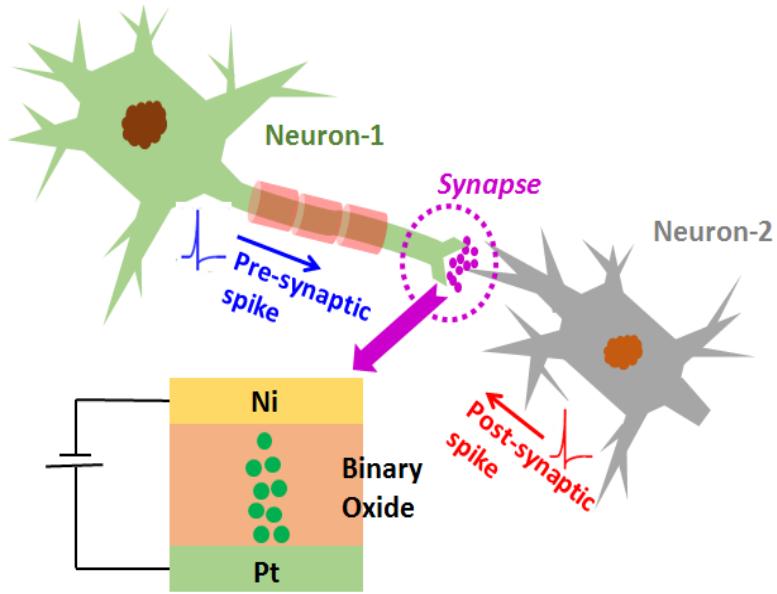


Figure 5.1: Illustration of a synaptic system and an electronic device that can mimic the synapse.

### 5.1.1 Spike time dependent plasticity

In biological brain, the connectivity strength between different synapses are adjusted depending upon the neuron interaction. Spike timing dependent plasticity (or STDP) is a process by which the synaptic strength gets adjusted with precise timing of spikes. STDP is believed to be the main mechanism responsible for learning and memory storage of biological brain [3]. STDP is achieved via long-term potentiation (LTP) and long-term depression (LTD) mechanisms. LTP refers to the increase in synaptic weight if a pre-synaptic spike, on average, precedes a post-synaptic spike. Similarly, LTD refers to the decrease in synaptic weight if a post-synaptic spike, on average, precedes a pre-synaptic spike. The increase or decrease in the synaptic strength is determined by the value of timing differences between the pre-neuron and the post-neurons spikes. A smaller time difference between the pre-synaptic spike and the post-synaptic spike results in a larger change in the synaptic weight. In the potentiation process, the inputs that might be the cause of the post-synaptic spike are made more likely to contribute in the future, whereas the inputs that are not the cause of the post-synaptic spike are made less likely to contribute in the future.

Synaptic plasticity can also be obtained with modern nanoelectronic devices with proper

spike firing techniques. The primary requirement of a synaptic device to emulate plasticity are; 1) should have analog-like memory transition between the different conductance states, and 2) should have a very low energy consumption for every synaptic event. Recent advancement in semiconductor memory devices such as RRAM, PCM, CBRAM, etc. showed to fulfill both the criteria thereby attracting dramatic research interest toward brain inspired computation [6]. Figure 5.1 shows the schematic of a typical synaptic communication between two neurons with the help of synapse. Neuron-1 is responsible for generating pre-synaptic spike, whereas the neuron-2 is responsible for generating post-synaptic spike. During learning, depending upon the position of the spikes, the synaptic weight at the synapse gets adjusted. As mentioned previously, a potentiation will occur when the pre-synaptic spike precedes the post-synaptic spike, resulting in an increase in the synaptic weight. Similar mechanism can be realized with a nanoelectronic memory device (RRAM in this case, as shown in the inset of Figure 5.1) by modulating the conductance of the memory.

Traditionally, Von-Neumann architecture has been preferred for implementing such neuromorphic systems because of their extreme maturity [9, 10]. However, recent technology flash memories show poor performance with further scaling thereby demanding a potential replacement. On the other hand, most of the emerging memory candidates have been shown to have multiple conductance states [7]. This feature of the memory enables them to program to a desired conductance state depending upon the synaptic weight applied with the programming pulse strength. Going to the second criteria, the electronic synapse should consume low energy per synaptic operation. Flash memories work at a programming voltage of more than 12 V (for erase operation, the voltage requirement is even higher). On the other hand, emerging memories show switching with applied bias within few volts. Some of the emerging memories have been reported to switch within a volt, taking current in nA range [7]. All these interesting functionalities in the emerging memory devices have attracted feasibility of realizing neural network systems with much improved performance compared to the existing Von-Neumann architecture.

### 5.1.2 Why RRAM ?

Several non-volatile memory candidates such as phase change RAM (PCRAM) [6], conductive bridge RAM (CBRAM)[9], Resistive RAM (RRAM) [10, 11, 12] etc. have been proposed as synapse element. Among them, RRAM is a promising candidate because of its CMOS compatibility and relatively simple to fabricate [10]. RRAMs can be arranged in the form of cross-bar array thereby enabling 3D high-density memory architecture [14]. Emerging memories are known for their fast switching, excellent retention and endurance behavior, RRAM also have all these properties. Interestingly, recent reports in RRAM show memory transition in sub-

$\mu\text{A}$  current range. Thus a very low energy switching is a great possibility using RRAM. As mentioned earlier, The RRAM memory operation is achieved with a filament formation in the dielectric with a forming or set operation (low resistance state or LRS), followed by the rupture of filament for a reset operation ((high resistance state or HRS). Gradual conductance change in the RRAM can be achieved by controlling the filament properties using set and reset operations [8]. During set, the CC limit can be kept at different values which in turn decides the conductivity of the filament. Similarly, during reset, the rupturing process of the filament can be controlled by specified reset stop voltage resulting in multi-state analog memory behavior. Thus, set process can be effectively used to gradually increase the conductance of the dielectric (similar to the potentiation process), and reset control can be used to decrease the conductance of the RRAM (similar to the depression process).

## 5.2 Motivation of work

Even though some binary oxide RRAMs has been reported showing gradual reset thereby resembling the analog memory behavior, but in many cases, these oxides typically shows abrupt reset thereby eliminating the possibility of realizing synaptic learning RRAM52. This chapter of the dissertation is dedicated to understand the gradual resistive switching properties of  $\text{Al}_2\text{O}_3$  and  $\text{HfO}_2$  RRAM. In bipolar RRAM reset process, two main mechanisms has been proposed as (i) drift of oxygen ions by electric field and (ii) diffusion of oxygen ions caused by concentration gradient of oxygen ions stored in the electrode [16]. It should be noted that the temperature of the dielectric caused by Joule heating also plays a major role in the reset process [15]. In many cases, the reset in RRAM has found to be instantaneous [17], as well as gradual [11]. As mentioned earlier, one of the prime requirements of RRAM to be applicable for synaptic applications is the feasibility of gradual reset characteristics [3]. Hence a detailed study on the gradual and abrupt carried in this part of the dissertation.

## 5.3 Understanding the gradual reset in RRAM

There can be different mechanisms driving the gradual reset behavior. There is a dependence of reset on electrode choice. A recent report on reset in  $\text{HfO}_2$  RRAM was observed to be dependent on the concentration of oxygen ions stored in the ITO electrode [19]. A higher concentration of oxygen in the electrode was observed to induce a gradual type reset with multiple intermediate states. On the other hand, lower concentration of oxygen was observed to result in abrupt reset where current drops instantaneously. This behavior can be correlated with other reports where gradual reset was observed in bipolar RRAM [20] and unipolar RRAM [21] RRAM. In both the cases, the situation where the filament comprised of higher number of oxygen vacancies

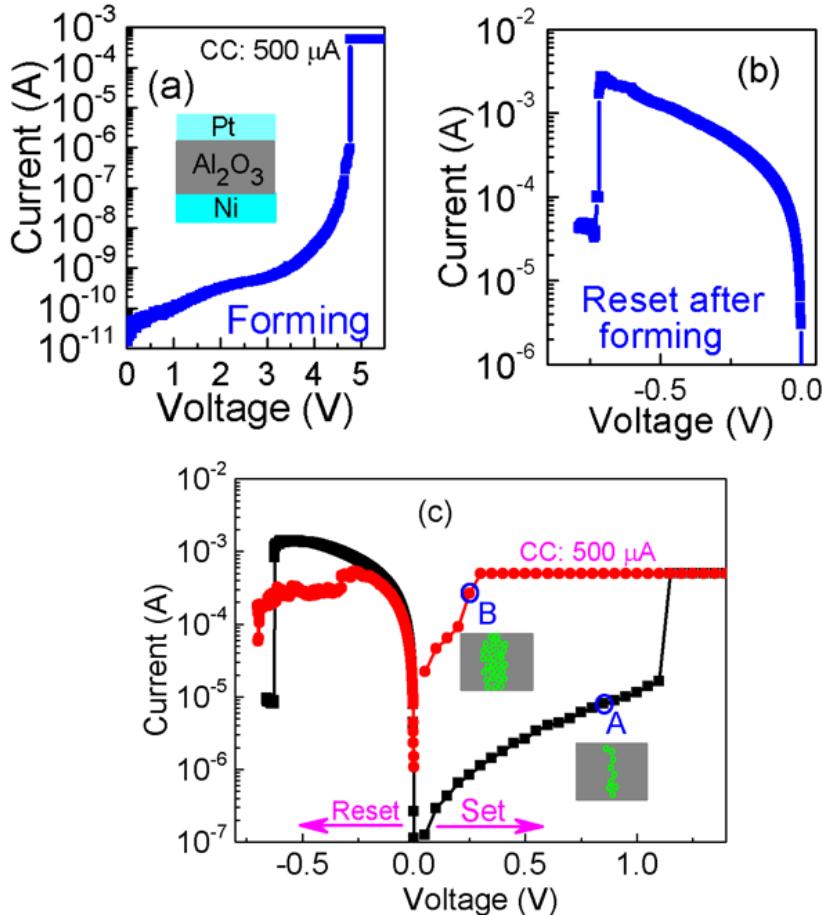


Figure 5.2: (a) Forming, (b) reset after forming, and (c) two distinct set-reset behavior observed in Al<sub>2</sub>O<sub>3</sub> RRAM.

(which in turn stores higher number of oxygen ions in the electrode) resulted in a gradual reset. Thus, the gradual reset is possibly because of oxygen ion concentration gradient diffusion dominated reset process which ruptures the filament gradually. The other reason might be because of gradual filament dissolution during reset when the filament is comprised to higher concentration of oxygen vacancies.

Figure 5.2 shows the forming, reset after forming and set-reset behavior of Pt/Al<sub>2</sub>O<sub>3</sub>/Ni RRAM. CC was kept at 500  $\mu$ A during forming and set operation for all the cases in this chapter of the dissertation. Forming was observed to have an abrupt rise in current before reaching the CC, shown in Figure 5.2 (a). The reset after the forming was observed to be abrupt, shown in Figure 5.2 (b). Two different set-reset behavior was observed during successive set-reset cycles,

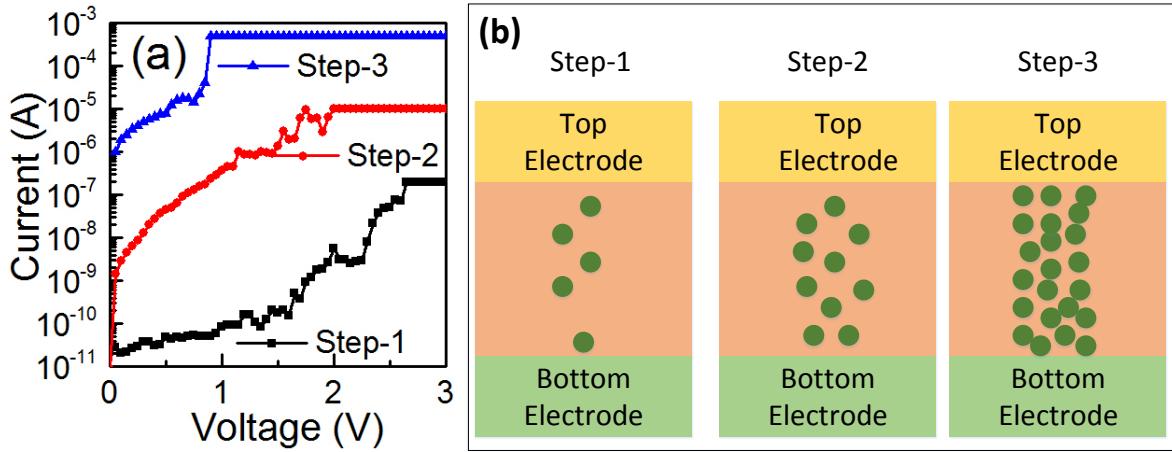


Figure 5.3: (a) Multi-step forming in  $\text{Al}_2\text{O}_3$  RRAM; (b) possible vacancy profile after individual forming steps.

are shown in Figure 5.2 (c). These two behaviors can be categorized as case-(A): a set operation where abrupt rise in current leads to reach CC, the corresponding reset was also observed to be abrupt. Whereas in case-(B): a set where a gradual rise in current leads to reach CC, and the corresponding reset was also gradual. Forming was observed to be similar to case-(A), and the reset after forming was also similar to case-(A) reset.

Considering point A and point B in the set curves before the CC is reached in case of abrupt set (or case-(A)) and gradual set (or case-(B)) respectively, point A has lower current at higher voltage signifying the lesser number of vacancies present in the dielectric. Whereas in the case of point B of gradual set, higher number of vacancies are present in the dielectric resulting in a higher current at lower voltage. This indicates the gradual set leads to generation of significantly higher number of oxygen vacancies and hence higher number of oxygen ions are stored in the Pt electrode in case of gradual set as compared to abrupt set. Higher number of vacancies inside the dielectric is known to create higher temperature gradient in the dielectric near the filament [18, 27]. Thus oxygen ions can recombine with the vacancies at lower reset currents because of concentration gradient dominated diffusion of oxygen ions in case of gradual reset [16]. Hence the reset in case-(B) was possibly because of a diffusion dominated filament dissolution where vacancies are ruptured gradually. Also note that gradual reset was observed to be occurring at lower current compared to the abrupt reset case, which is another advantage of obtaining wider filament.

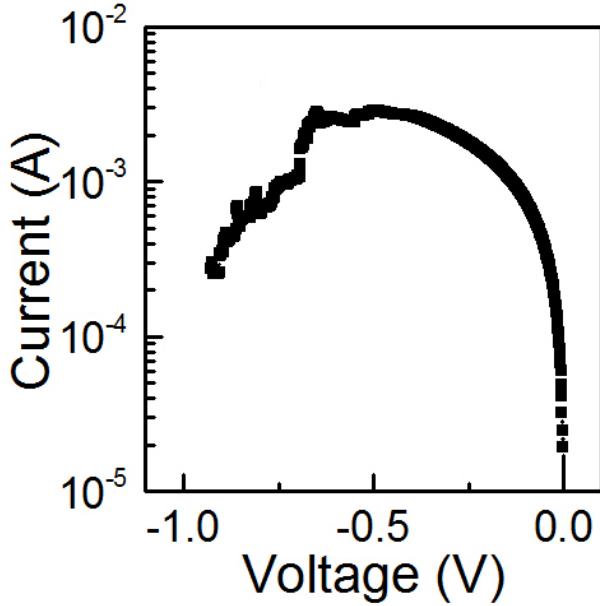


Figure 5.4: Reset after the multi-step forming in  $\text{Al}_2\text{O}_3$  RRAM.

#### 5.4 Realizing the gradual reset in RRAM

A better control on the filament diameter (or resistance) can be achieved with multi-step forming [20] and multi-step set techniques [24, 25]. This technique also prevents the current overshoot during forming or set event. To reconfirm the gradual filament dissolution hypothesis, a virgin device was analyzed and forming process was broken into three parts in order to generate a higher number of vacancies in the dielectric before the device reach the final CC, shown in Figure 5.3. During forming, CC was kept at 200 nA, 10  $\mu$ A and 500  $\mu$ A respectively during successive steps, shown in Figure 5.3 (a). Successive forming steps were observed to start at higher current values signifying an increase in the concentration of oxygen vacancies in the dielectric. Possible vacancy profile after successive forming step is illustrated in Figure 5.3 (b). The forming current at step-3 was observed to be similar to set at case-(B) of Figure 5.2 (c), which indicates a higher number of vacancies are present in the dielectric during step-3 of the multi-step forming compared to abrupt forming of fig Figure 5.2 (a). The corresponding reset after the multi-step forming is shown in Figure 5.4. The reset after multi-step forming was also observed to be gradual which ascertains the hypothesis of the gradual filament dissolution mechanism.

After obtaining a gradual reset in  $\text{Al}_2\text{O}_3$  RRAM, multi-step forming method was applied

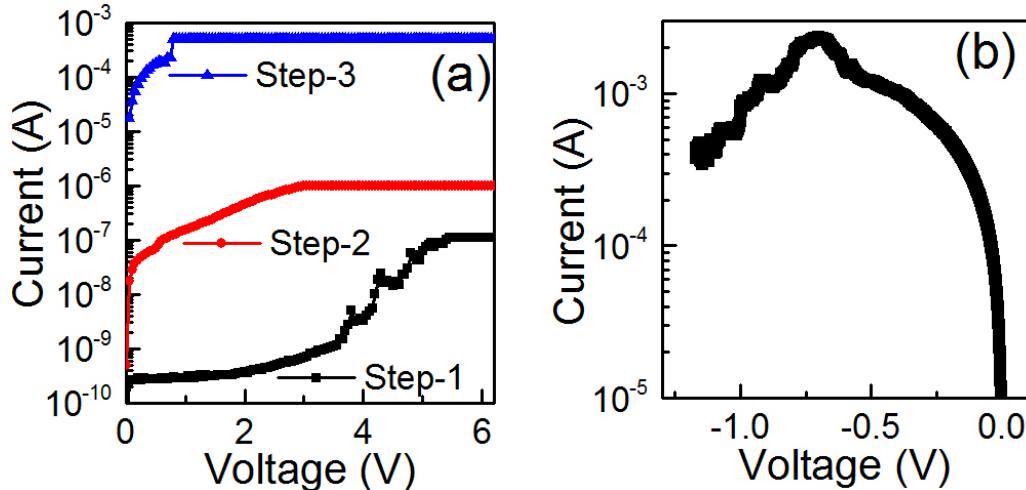


Figure 5.5: (a) Multi-step forming in  $\text{HfO}_2$  RRAM; (b) corresponding reset.

to  $\text{HfO}_2$  RRAM. Similar to the  $\text{Al}_2\text{O}_3$  RRAM,  $\text{HfO}_2$  RRAM also showed a gradual reset after multi-step forming process, as shown in Figure 5.5. Thus, the dependence of the reset mechanism with the filament diameter was observed to be consistent for both the RRAMs. This gradual reset is the enabler of analog type memory behavior. The reset window can be controlled thereby resulting in a controllable  $I_{on}/I_{off}$  ratio, as shown in Figure 5.6. Both  $\text{Al}_2\text{O}_3$  and  $\text{HfO}_2$  RRAM showed controllable gradual reset by controlling the reset stop voltage. For  $\text{Al}_2\text{O}_3$  RRAM (shown in Figure 5.6 (a)), the reset was observed to appear at around -0.4 V; whereas for  $\text{HfO}_2$  RRAM (shown in Figure 5.6 (b)), the reset was observed to appear at around -0.6 V. The multi-state memory operation in  $\text{Al}_2\text{O}_3$  RRAM was obtained by choosing the reset stop voltage as -0.45 V, -0.5 V and -0.6 V.. Similarly, controllable gradual reset was obtained in  $\text{HfO}_2$  RRAM by choosing reset stop voltage as -0.7 V, -0.8 V and -0.9 V. Interestingly,  $\text{Al}_2\text{O}_3$  RRAM shows a larger  $I_{on}/I_{off}$  ratio with gradual reset compared to  $\text{HfO}_2$  RRAM. The reason behind such observation is mentioned in the next section.

## 5.5 Comparison of conductance modulation in $\text{Al}_2\text{O}_3$ and $\text{HfO}_2$ RRAM

Although  $\text{Al}_2\text{O}_3$  has higher band-offset with the electrode which should limit the trap assisted tunneling (TAT) current during the set process, the set current rise in  $\text{Al}_2\text{O}_3$  RRAM was observed to be similar to that of  $\text{HfO}_2$  RRAM, shown in Figure 5.7. In the previous chapter, we observed the  $\text{Al}_2\text{O}_3$  having lower zero-field activation energy ( $E_a$ ) which is a key parameter in

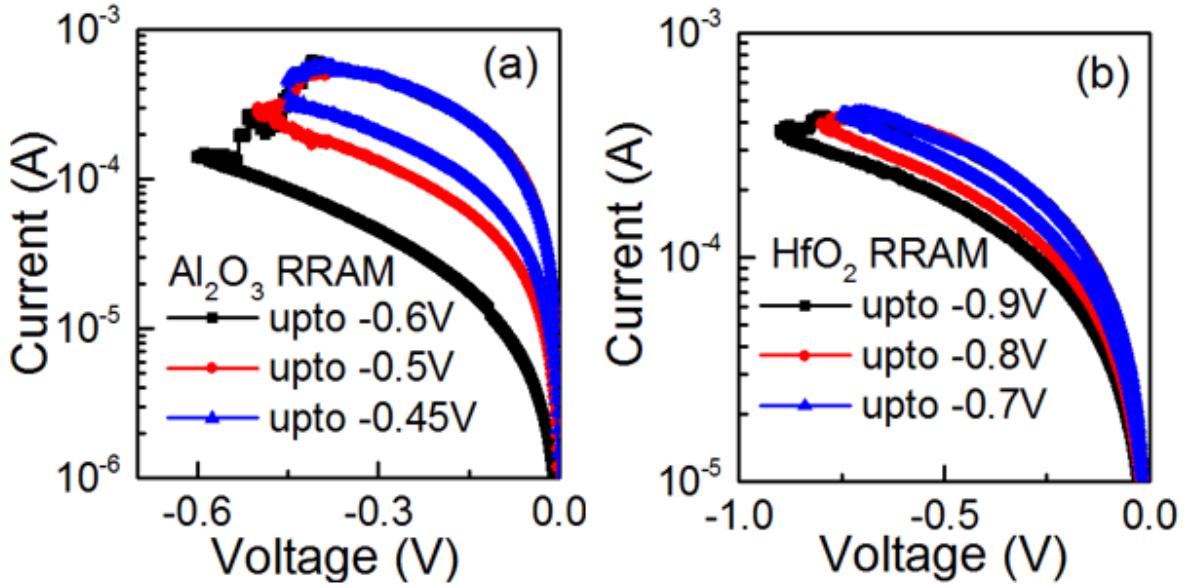


Figure 5.6: Multi-state memory window achieved with gradual reset in (a)  $\text{Al}_2\text{O}_3$  RRAM, and (b)  $\text{HfO}_2$  RRAM.

the generation of oxygen vacancies. This low  $E_a$  in  $\text{Al}_2\text{O}_3$  results in a higher vacancy generation rate inside the dielectric during the set process. Thus, it is possible that the  $\text{Al}_2\text{O}_3$  RRAM gets a higher number of oxygen vacancies compared to  $\text{HfO}_2$  RRAM during the set process, which in turn provides a higher density of trap states thereby raising the TAT current (as shown in Figure 5.7 (a)). During reset, the vacancies get annihilated gradually, and a higher band-offset with the electrode is advantageous to limit the TAT current. Since  $\text{Al}_2\text{O}_3$  have a higher band offset with the electrode, the current after reset is expected to be lower in  $\text{Al}_2\text{O}_3$  RRAM compared to  $\text{HfO}_2$  RRAM. For this reason, a higher reset window was observed in  $\text{Al}_2\text{O}_3$  RRAM, as shown in Figure 5.7 (b)). Hence  $\text{Al}_2\text{O}_3$  was observed to show higher conductance change compared to  $\text{HfO}_2$  because of two main reasons: lower  $E_a$  of  $\text{Al}_2\text{O}_3$  helps in generating a higher number of oxygen vacancies in the dielectric during the set process, and a higher band-offset restricting the TAT current thereby proving a larger window during reset.

Next, the gradual reset characteristics were further used to study the conductance modulation of the  $\text{Al}_2\text{O}_3$  and  $\text{HfO}_2$  RRAM using DC reset and AC pulse reset technique, shown in Figure 5.8. Set operation was performed initially in the RRAMs keeping the CC at  $500 \mu\text{A}$ , followed by the reset operation performed with either DC sweep or AC pulse. Reset stop voltage was kept same as mentioned in the previous section. Figure 5.8 (a) shows the comparison of conductance change for  $\text{Al}_2\text{O}_3$  and  $\text{HfO}_2$  RRAM with DC reset technique. As observed earlier,

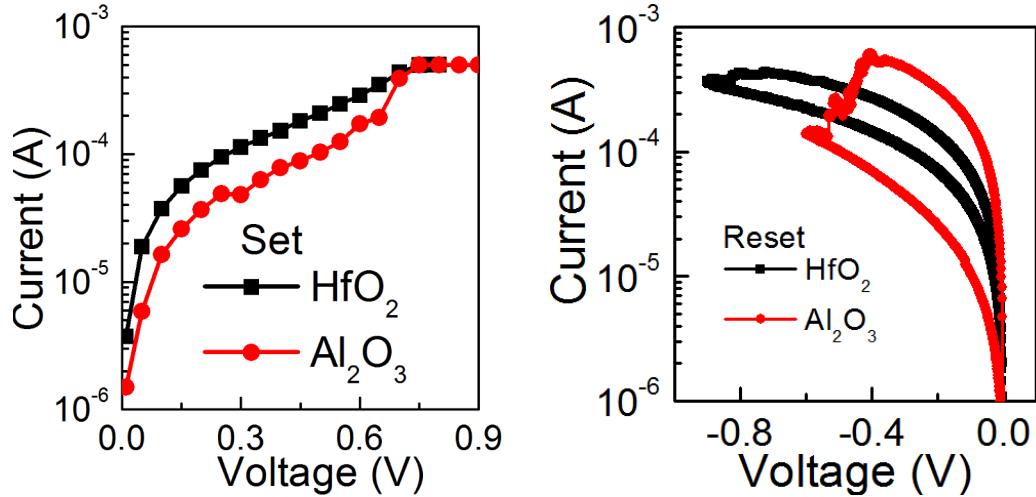


Figure 5.7: Comparison of gradual set-reset behavior in (a)  $\text{Al}_2\text{O}_3$  RRAM, and (b)  $\text{HfO}_2$  RRAM.

$\text{Al}_2\text{O}_3$  RRAM shows a significantly higher conductance change compared to  $\text{HfO}_2$  RRAM. Similarly, Figure 5.8 (b) shows the comparison of conductance change for  $\text{Al}_2\text{O}_3$  and  $\text{HfO}_2$  RRAM with AC reset pulses. In a real neuromorphic system, the set and reset needs to be implemented using AC pulses, hence it is mandatory to achieve conductance change with AC reset method. For  $\text{Al}_2\text{O}_3$  RRAM, reset pulse voltage was fixed at -0.5 V, whereas for  $\text{HfO}_2$  RRAM, the pulse voltage was fixed at -0.8 V. Both the RRAM showed the dependence on the reset pulse time. However, the change in conductance was observed to be much lower with AC reset compared to the DC reset technique.

## 5.6 Reliability analysis of the gradual reset

After attaining the multiple memory states in RRAMs, retention and endurance characteristics of the RRAMs were carried out to evaluate their reliability. Figure 5.9 (a) and Figure 5.9 (b) shows the retention of  $\text{Al}_2\text{O}_3$  and  $\text{HfO}_2$  RRAM respectively at multiple resistance state. The devices were initially set to a CC limit of  $500 \mu\text{A}$ , followed by a gradual reset with voltage ranges mentioned in the previous section. Both  $\text{Al}_2\text{O}_3$  and  $\text{HfO}_2$  RRAM showed excellent retention behavior. No degradation was observed after  $10^3$  sec, which is far better than the retention characteristics observed in DFGFET. Similarly, both the RRAMs were tested for DC endurance after gradual reset, shown in Figure 5.9 (c) and Figure 5.9 (d) respectively. Successive set-reset endurance cycles after obtaining gradual reset showed very good uniformity. Unlike the abrupt reset case where the  $I_{on}/I_{off}$  ratio varies randomly with successive endurance

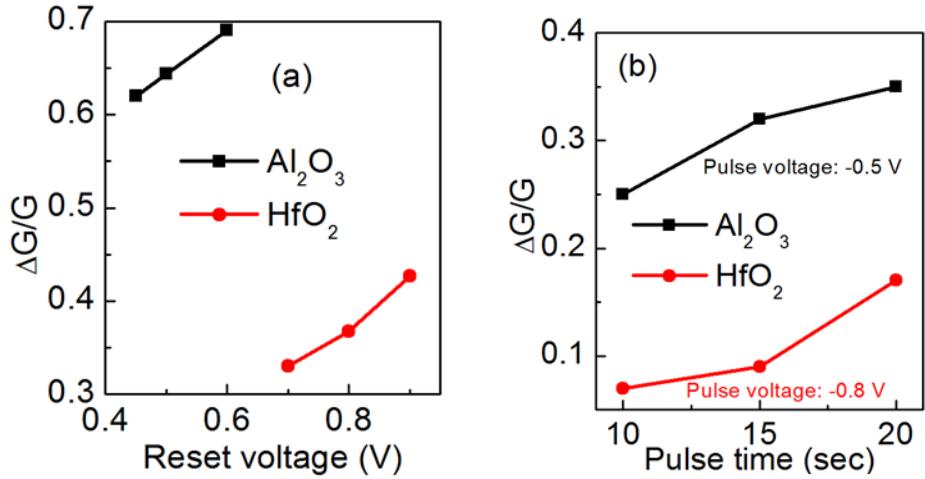


Figure 5.8: Conductance modulation comparison of (a)  $\text{Al}_2\text{O}_3$  RRAM and (b)  $\text{HfO}_2$  RRAM using DC/AC reset techniques.

cycles, the gradual reset after vacancy diffusion showed almost no degradation even after 100 DC cycles. The set-reset uniformity was also observed to be greatly improved. This kind of controllable gradual reset is advantageous over abrupt reset, and are of particular interest to analog memory applications where RRAMs can resemble multi-bit operation with excellent retention and endurance. The gradual reset was observed to be present even after 100 set-reset DC sweep cycles. Hence it can be concluded that the RRAMs are capable of showing excellent retention and endurance characteristics with gradual reset technique. Such a uniform behavior provides the added advantage of realizing synaptic systems where device a very low variation during successive operation is desired.

## 5.7 Implementation of synaptic learning

Finally, the synaptic learning operation was carried out in both the RRAMs. The learning consists of either potentiation or depression. Potentiation can be obtained with gradual set process. Similarly, depression can be realized with the help of controllable reset pulses. The synaptic learning was carried out in both the RRAM where the conductance of the dielectric was changed gradually using AC set and reset pulses, shown in Figure 5.10. The potentiation in RRAMs were achieved with 60 steps of set by gradually varying the CC during the set. Similarly, the depression was achieved with 70 reset steps where the pulse voltage was increased gradually during the reset. During set, CC was varied from  $100 \mu\text{A}$  to  $500 \mu\text{A}$  keeping the pulse voltage fixed at  $+1.5 \text{ V}$ ,  $5 \text{ ms}$  for both the device. During reset, the reset pulse voltage was varied

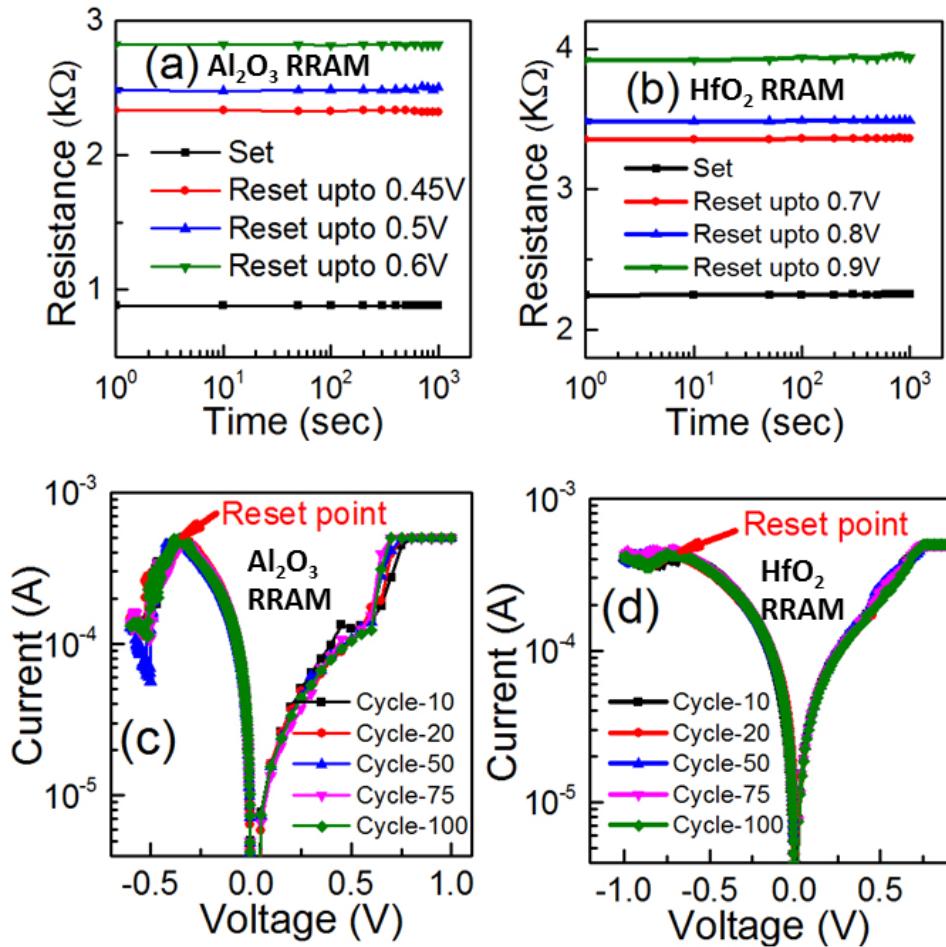


Figure 5.9: (a) and (b) Multi-state retention behavior of  $\text{Al}_2\text{O}_3$  and  $\text{HfO}_2$  RRAM respectively; (c) and (d) Endurance behavior of  $\text{Al}_2\text{O}_3$  and  $\text{HfO}_2$  RRAM respectively

gradually from -0.4 V to -0.58 V for  $\text{Al}_2\text{O}_3$  RRAM, and -0.7 V to -1 V for  $\text{HfO}_2$  RRAM. A controllable synaptic learning was observed in both the RRAMs. The synaptic learning confirms the applicability of using such RRAMs in neuromorphic systems after obtaining the gradual reset. Depending upon the position of pre-synaptic and post-synaptic pulse, the conductance of the RRAM can be either increased or decreased. The transitions were observed to occur within operation voltage of 1 V, which is significantly better than the flash memories.

## 5.8 Summary

RRAM is one of the potential candidates for implementing neuromorphic systems. However, some of the RRAM shows the gradual reset behavior which eliminates the possibility of creating

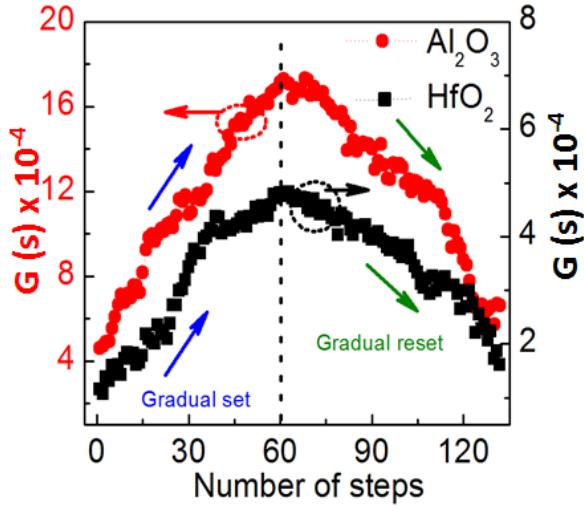


Figure 5.10: Synaptic learning behavior of  $\text{Al}_2\text{O}_3$  and  $\text{HfO}_2$  RRAM.

analog memory type behavior. A better understanding of the filament creation process allowed to create a gradual reset operation in RRAMs. It was observed that a higher number of vacancies created during forming or set event results in a gradual filament dissolution process during the reset. In order to generate a higher number of vacancies in the dielectric, forming operation was broken into three steps which resulted in a gradual reset in  $\text{Al}_2\text{O}_3$  and  $\text{HfO}_2$  RRAMs. Such a gradual reset resulted in a multi-bit analog memory characteristics in the RRAMs where the conductance of the dielectric can be changed gradually. The property of controllable conductance change was then used to demonstrate synaptic learning behavior in the RRAMs. Hence, by generating a higher number of vacancies inside the dielectric during forming or set process, RRAMs were demonstrated to have gradual reset characteristics suitable for multi-bit analog memory and neuromorphic computation applications.

## REFERENCES

- [1] J. Park, K. P. Biju, S. Jung, W. Lee, J. Lee, S. Kim, S. Park, J. Shin and H. Hwang, "Multibit Operation of TiO<sub>x</sub>-Based ReRAM by Schottky Barrier Height Engineering," *IEEE Electron Device Letters*, vol. 32, no. 4, pp. 476-478, 2011.
- [2] B. Sarkar, B. Lee and V. Misra, "Understanding the gradual reset in Pt/Al<sub>2</sub>O<sub>3</sub>/Ni RRAM for synaptic applications," *Semiconductor Science and Technology*, vol. 30, 105014, 2015.
- [3] D. Kuzum, S. Yu and H. S. P. Wong, "Synaptic Electronics: materials, devices and applications," *Nanotechnology*, vol. 24, 382001, 2013.
- [4] S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder and W. Lu, "Nanoscale Memristor Device as Synapse in Neuromorphic Systems," *Nano Letters*, vol. 10, no. 4, pp. 12971301, 2010.
- [5] L. B. Bernabe, S. G. Teresa, C. M. Luis A., P. C. Jose A., Z. R. Carlos and M. Timothee, "On Spike-Timing-Dependent-Plasticity, Memristive Devices, and building a Self-Learning Visual Cortex," *Frontiers in Neuroscience*, vol. 5, no. 00026, 2011.
- [6] M. Suri, O. Bichler, D. Querlioz, O. Cueto, L. Perniola, V. Sousa, D. Vuillaume, C. Gamrat and B. DeSalvo, "Phase change memory as synapse for ultra-dense neuromorphic systems: Application to complex visual pattern extraction," *IEEE International Electron Devices Meeting (IEDM)*, pp. 4.4.1-4.4.4 , 2011.
- [7] J. S. Meena, S. M.Sze, U. Chand and T. Y. Tseng, "Overview of emerging nonvolatile memory technologies," *Nanoscale Research Letters*, vol. 9, 526, 2014.
- [8] A. Prakash, J. Park, J. Song, J. Woo, E. J. Cha and H. Hwang, "Demonstration of Low Power 3-bit Multilevel Cell Characteristics in a TaO<sub>x</sub>-Based RRAM by Stack Engineering," *IEEE Electron Device Letters*, vol. 36, no. 1, pp. 32-34, 2015.

- [9] M. Suri, O. Bichler, D. Querlioz, G. Palma, E. Vianello, D. Vuillaume, C. Gamrat and B. DeSalvo, "CBRAM devices as binary synapses for low-power stochastic neuromorphic systems: Auditory (Cochlea) and visual (Retina) cognitive processing applications," *IEEE International Electron Devices Meeting (IEDM)*, pp. 10.3.1-10.3.4, 2012.
- [10] Y. Wu, S. Yu, H. S. P. Wong, Y. S. Chen, H. Y. Lee, S. M. Wang, P. Y. Gu, F. Chen and M. J. Tsai, "AlOx-Based Resistive Switching Device with Gradual Resistance Modulation for Neuromorphic Device Application," *IEEE International Memory Workshop (IMW)*, pp. 1-4, 2012.
- [11] S. Ambrogio, S. Balatti, F. Nardi, S. Facchinetto and D. Ielmini, "Spike-timing dependent plasticity in a transistor-selected resistive switching memory," *Nanotechnology*, vol. 24, 384012, 2013.
- [12] S. Yu , B. Gao , Z. Fang , H. Yu , J. Kang and H. S. P. Wong, "A Low Energy Oxide-Based Electronic Synaptic Device for Neuromorphic Visual Systems with Tolerance to Device Variation," *Advanced Materials*, vol. 25, pp. 17741779, 2013.
- [13] Y. Wu, S. Yu, B. Lee and H. S. P. Wong, "Low-power TiN/Al<sub>2</sub>O<sub>3</sub>/Pt resistive switching device with sub-20 A switching current and gradual resistance modulation," *Journal of Applied Physics*, vol. 110, 094104, 2011.
- [14] C. Cagli, J. Buckley, V. Jousseaume, T. Cabout, A. Salaun, H. Grampeix, J. F. Nodin, H. Feldis, A. Persico, J. Cluzel, P. Lorenzi, L. Massari, R. Rao, F. Irrera, F. Aussénac, C. Carabasse, M. Coué, P. Calka, E. Martinez, L. Perniola, P. Blaise, Z. Fang, Y. H. Yu, G. Ghibaudo, D. Deleruyelle, M. Bocquet, C. Muller, A. Padovani, O. Pirrotta, L. Vandelli, L. Larcher, G. Reimbold and B. de Salvo, "Experimental and Theoretical Study of Electrode Effects in HfO<sub>2</sub> based RRAM," *IEEE International Electron Devices Meeting (IEDM)*, pp. 28.7.1-28.7.4, 2011.
- [15] L. Vandelli, A. Padovani, L. Larcher, G. Broglia, G. Ori, M. Montorsi, G. Bersuker and P. Pavan "Comprehensive physical modeling of forming and switching operations in HfO<sub>2</sub>

RRAM devices," *IEEE International Electron Devices Meeting (IEDM)*, pp. 17.5.1-17.5.4, 2011.

- [16] S. Yu and H. S. P. Wong, "A Phenomenological Model for the Reset Mechanism of Metal Oxide RRAM," *IEEE Electron Device Letters*, vol. 31, no. 12, pp. 1455-1457, 2010.
- [17] X.i Yang, S. Long, K. Zhang, X. Liu, G. Wang, X. Lian, Q. Liu, H. Lv, M. Wang and H. Xie, "Investigation on the RESET switching mechanism of bipolar Cu/HfO<sub>2</sub>/Pt RRAM devices with a statistical methodology," *Journal of Physics D: Applied Physics*, vol. 46, 245107, 2013.
- [18] L. Vandelli, A. Padovani, L. Larcher and G. Bersuker, "Microscopic Modeling of Electrical Stress-Induced Breakdown in Poly-Crystalline Hafnium Oxide Dielectrics," *IEEE Transactions on Electron Devices*, vol. 60, no. 5, pp. 1754-1762, 2013.
- [19] C. W. Zhong, W. H. Tzeng, K. C. Liu, H. C. Lin, K. M. Chang, Y. C. Chan, C. C. Kuo, P. S. Chen, H. Y. Lee, F. Chen and M. J. Tsai, "Effect of ITO electrode with different oxygen contents on the electrical characteristics of HfO<sub>x</sub> RRAM devices," *Surface and Coatings Technology*, vol. 231, pp. 563-566, 2013.
- [20] N. Raghavan, R. Degraeve, A. Fantini, L. Goux, D. J. Wouters, G. Groeseneken and M. Jurczak, "Stochastic variability of vacancy filament configuration in ultra-thin dielectric RRAM and its impact on OFF-state reliability," *IEEE International Electron Devices Meeting (IEDM)*, pp. 21.1.1-21.1.4 , 2013.
- [21] S. Long, L. Perniola, C. Cagli, J. Buckley, X. Lian, E. Miranda, F. Pan, M. Liu and J. Sune, "Voltage and Power-Controlled Regimes in the Progressive Unipolar RESET Transition of HfO<sub>2</sub>-Based RRAM," *Scientific Reports*, vol. 3, 2929, 2013.
- [22] H. Lv, X. Xu, H. Liu, R. Liu, Q. Liu, W. Banerjee, H. Sun, S. Long, L. Li and M. Liu, "Evolution of conductive filament and its impact on reliability issues in oxide-electrolyte based resistive random access memory," *Scientific Reports*, vol. 5, 7764, 2015.

- [23] P. Huang, B. Chen, Y. J. Wang, F. F. Zhang, L. Shen, R. Liu, L. Zeng, G. Du, X. Zhang, B. Gao, J.F. Kang, X.Y. Liu, X. P. Wang, B. B. Weng, Y. Z. Tang, G. Q. Lo and D. L. Kwong, "Analytical model of endurance degradation and its practical applications for operation scheme optimization in metal oxide based RRAM," *IEEE Electron Device Meeting (IEDM)*, pp. 22.5.1-22.5.4, 2013.
- [24] Y. Z. Tang, Z. Fang, X. P. Wang, B. B. Weng, Z. X. Chen and G. Q. Lo, "A Novel RRAM Stack With TaOX/HfOy Double-Switching-Layer Configuration Showing Low Operation Current Through Complimentary Switching of Back-to-Back Connected Subcells," *IEEE Electron Device Letters*, vol. 35, no. 6, pp. 627-629, 2014.
- [25] B. Long, Y. Li and R. Jha, "Switching Characteristics of Ru/HfO<sub>2</sub>/TiO<sub>2-x</sub>/Ru RRAM Devices for Digital and Analog Nonvolatile Memory Applications," *IEEE Electron Device Letters*, vol. 33, no. 5, pp. 706-708, 2012.
- [26] J. Sun, Q. Liu, H. Xie, X. Wu, F. Xu, T. Xu, S. Long, H. Lv, Y. Li, L. Sun and M. Liu, "In situ observation of nickel as an oxidizable electrode material for the solid-electrolyte-based resistive random access memory," *Applied Physics Letters*, vol. 102, 053502, 2013.
- [27] D. Ielmini, F. Nardi and C. Cagli, "Universal Reset Characteristics of Unipolar and Bipolar Metal-Oxide RRAM," *IEEE Transactions on Electron Devices*, vol. 58, no. 10, pp. 3246-3253, 2011.
- [28] B. Sarkar, B. Lee, and V. Misra, "Understanding the influence of Ea and band-offset toward the conductance modulation in Al<sub>2</sub>O<sub>3</sub> and HfO<sub>2</sub> synaptic RRAM, *Device Research Conference (DRC)*, pp.149-150, 2015.

## Chapter 6

# Volatile Memory Operation Using RRAM

Previous chapters on RRAM were focused to understand the non-volatile memory operation using RRAM. Although RRAM shows very fast operation like DRAM, however, the filamentary operation requires a very high current [1]. Thus, the power requirement of RRAM is much higher than that of DRAM. Hence, the filamentary RRAM devices will require further research in order to replace the DRAM. This chapter is dedicated to understanding the low power non-filamentary RRAM operation which may be a potential candidate for DRAM replacement. The chapter also proposes a novel 1T-0C ZRAM device.

### 6.1 Introduction to the non-filamentary RRAM

Most of the RRAM studied so far are based on the creation and rupture of a conducting filament in the RRAM dielectric. As mentioned previously, these memories can potentially replace NAND, and few studies suggest their feasibility to replace DRAM as well. This brings the concept of storage class memory which refers to a memory candidate which can have very fast access, excellent retention and endurance characteristics, and can be arranged in 3D to enable very dense memory architecture [2, 3]. Storage class memory is expected to replace NAND as well as DRAM in the near future. Although the filamentary RRAM seems suitable to replace NAND within a short time, but more research has to be done in order to replace the DRAM. A critical requirement of a DRAM replacement should have (i) very low access time, (ii) program/erase speed in tens of nanoseconds, (iii) endurance of  $10^9$  -  $10^{12}$  or even more, and (iv) low power operation.

The non-volatility is required in storage class memory to provide efficient solutions for per-

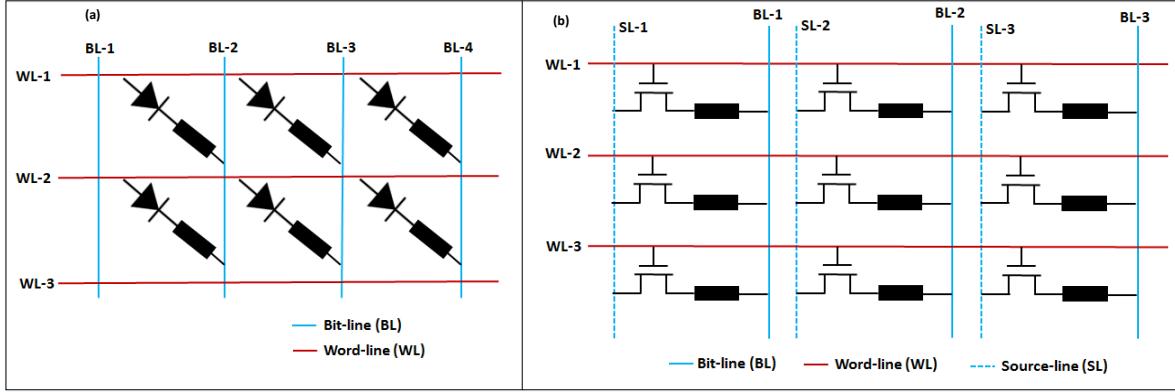


Figure 6.1: Schematic of RRAM circuit: (a) 1D-1R configuration, and (b) 1T-1R configuration.

manent storage. Although 3D-NAND is getting scaled to 14nm and beyond, but the access time of 3D-NAND remains a challenge. Similarly, conventional 1T-1C DRAM is facing scaling issues beyond 20 nm technology node [4]. With Scaling, the trench capacitor present in the DRAM poses several process challenges [5]. Current DRAM technology also suffers from high standby power due to excessive leakage and memory refresh cycles [6, 7]. This problem has resulted in several alternative technologies which can replace the DRAM [8, 9]. One such alternative is 1T-0C DRAM, also known as zero-capacitance RAM (or ZRAM) where the memory operation is realized without using the trench capacitor [10, 11]. However, these devices have challenges like memory refresh and higher operating voltage making them unsuitable to replace the traditional 1T-1C DRAM [9, 12].

Among several emerging memory candidates, RRAM technology is promising because of its favorable properties like excellent endurance, CMOS compatibility, very fast switching, etc. The feasibility of cross-bar arrangement providing tremendous area scaling is advantageous which is currently lacking in DRAM technology. Several DRAM circuits using RRAM non-volatile memory have been proposed to meet the future DRAM requirements [13]. However, the RRAM operation based on conductive filament poses a variability challenge, and the power requirement for such operation is typically high.

Recently, a new type RRAM known as non-filamentary RRAM (also known as vacancy modulation conductive oxide RRAM, or VMCO-RRAM), has been observed to show memory behavior by modulating the inherent oxygen vacancies inside the dielectric [14]. These memories are forming free, and memory operation can be both volatile and non-volatile [15].

Dielectrics such as  $\text{TiO}_2$  [14],  $\text{TaO}_X$  [16], WOX [17], etc. have been observed to show non-filamentary RRAM operation. Correspondingly, models have also been proposed to address the non-filamentary RRAM operation [15]. Alongside, the non-filamentary RRAM has also been used to realize variable resistor, variable inductor and variable capacitor behavior [19]. Unlike filamentary RRAM, memory operations in non-filamentary RRAM doesn't require a current flow through the dielectric, thereby making them a promising low power memory candidate [14].

Similar to filamentary RRAM, non-filamentary RRAM also requires a selector device to prevent the current sneak path. The selector device can be either diode or a transistor. Schematic of typical RRAM array configuration is shown in Figure 6.1. The array made with diode as the selector device is referred as 1D-1R configuration (shown in Figure 6.1 (a)), whereas the array made with transistor as the selector device is referred as 1T-1R configuration (shown in Figure 6.1 (b)). Moreover, the ratio of conductance change in non-filamentary RRAM has also been observed to be low in many cases [15]. In order to realize a DRAM type operation, one needs a memory element which shows low power operation, excellent endurance, and avoid the trench capacitor requirement.

In this chapter, we investigated a novel 1T-0C ZRAM concept using  $\text{SiO}_2$ /non-filamentary RRAM as the MOSFET gate dielectric. Instead of keeping the RRAM element in the bit-line of 1T-1R array, we propose to keep the RRAM in the worldline of the array. The capacitance change property of the non-filamentary RRAM is expected to change the drain current enabling the possibility of obtaining a memory operation. For experimental validation,  $\text{TiO}_2$  was chosen as the non-filamentary RRAM dielectric because its mem-capacitive effect which has been reported recently [14].  $\text{TiO}_2$  is a low bandgap oxide, and contains a significant concentration of inherent oxygen vacancies. Before going there, let's understand the volatile memory operation in  $\text{TiO}_2$  RRAM.

## 6.2 Volatile memory operation in $\text{Ni}/\text{TiO}_2/\text{Ni}$ RRAM

Initially, to reconfirm the presence of inherent oxygen vacancies inside the  $\text{TiO}_2$  dielectric, XPS and XRD analysis of  $\text{TiO}_2$  was performed, shown in Figure 6.2. A  $\text{TiO}_2$  sample was prepared for the analytical characterization. 13 nm  $\text{TiO}_2$  was deposited on  $\text{Si}/\text{SiO}_2$  wafer using ALD system.  $\text{TiO}_2$  ALD recipe can be found in chapter-2. Figure 6.2 (a) shows the Ti-2p XPS spectra of the  $\text{TiO}_2$ . Two peaks near 458 eV and 464 eV confirms that the  $\text{TiO}_2$  is completely oxidized. The O1S XPS spectra showed an asymmetric peak near 530 eV, as shown in Figure 6.2 (b). This asymmetry can be attributed to the presence of oxygen vacancies inside the  $\text{TiO}_2$  dielectric

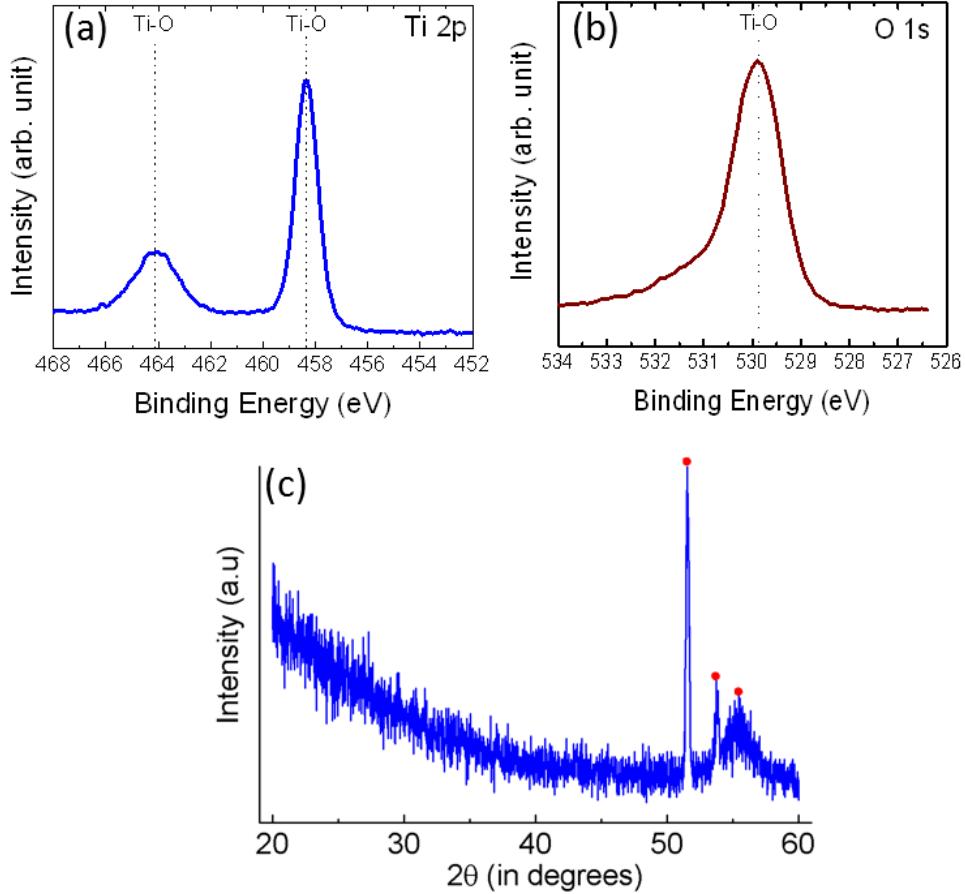


Figure 6.2: Material characterization of  $\text{TiO}_2$  dielectric; XPS analysis showing (a) Ti 2P spectra and (b) O 1s spectra; and (c) XRD analysis of  $\text{TiO}_2$  showing the anatase phase.

[22]. The XRD spectra shown in Figure 6.2 (c) revealed the anatase phase of the  $\text{TiO}_2$ , which is a stable phase of  $\text{TiO}_2$  and can be obtained with very low processing temperature. Thus, presence of oxygen vacancies was confirmed in the  $\text{TiO}_2$  dielectric.

Next, to understand the volatile memory operation in  $\text{TiO}_2$  RRAM, metal-insulator-metal capacitors were fabricated with Ni as the metal electrodes, and  $\text{TiO}_2$  as the dielectric. 13 nm thick  $\text{TiO}_2$  was deposited using ALD, and Ni electrodes were deposited using UHV-RF sputtering system. Fabrication steps are similar to that of RRAMs fabricated in the previous chapters. Figure 6.3 shows the volatile memory operation obtained with  $\text{TiO}_2$  RRAM. The inset of Figure 6.3 shows the schematic of the fabricated device. To measure the IV hysteresis, a forward sweep was applied at the Ni top electrode (TE) where voltage was varied from 0 V to the desired voltage, followed by an immediate reverse sweep taking the voltage back to 0 V.

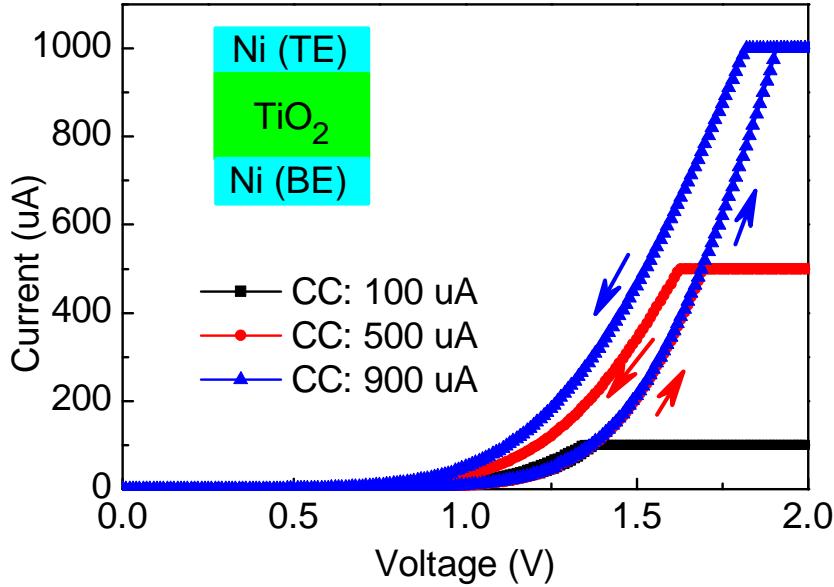


Figure 6.3: Volatile memory operation obtained in Ni/TiO<sub>2</sub>/Ni RRAM using IV measurement.

Bottom electrode (BE) was kept grounded.

A hysteresis was observed in the IV curve of TiO<sub>2</sub> RRAM. The current was observed to increase during the reverse sweep, signifying an increase in the conductance of the TiO<sub>2</sub> with a higher voltage applied at one of the electrodes. The successive forward sweeps were observed to follow the same curve, signifying the conductance of the dielectric reaches its original value after the sweep. Thus, a volatile memory transition was confirmed in TiO<sub>2</sub> RRAM. The volatile memory retention time will be discussed in the next section. This also means the voltage sweep did not generate additional vacancy inside the dielectric. Thus, a non-filamentary operation was also confirmed with this experiment.

Note that the memory window increases with increase in compliance current limit. When the device current reaches compliance limit, the electric field across the device remains the same at the compliance point even after the applied voltage goes high. This signifies the dependence of conductance on the electric field across the TiO<sub>2</sub> dielectric. A higher electric field increases the conductance of the TiO<sub>2</sub> dielectric with a volatile transition. Hence a volatile memory operation was observed in TiO<sub>2</sub> RRAM where a higher electric field resulted in a higher conductance change.

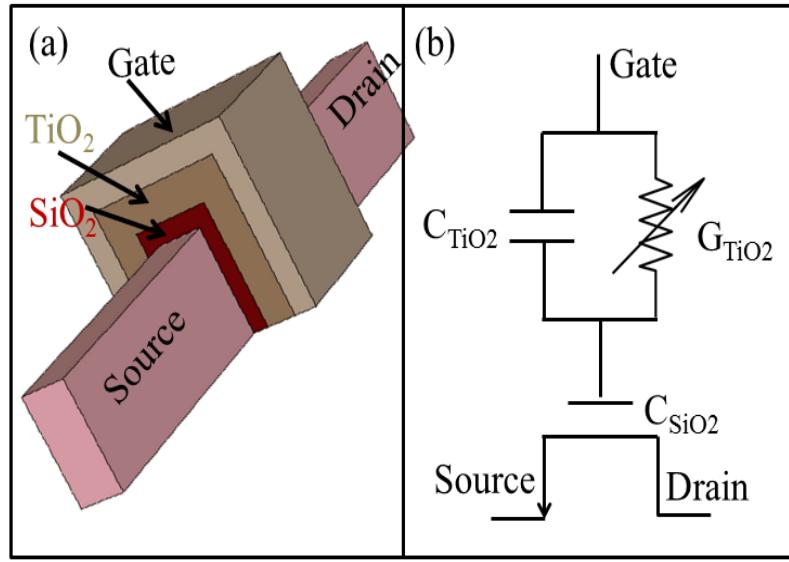


Figure 6.4: (a) Schematic of the proposed non-filamentary RRAM based 1T-0C ZRAM device; (b) circuit equivalent of the 1T-0C ZRAM device.

### 6.3 Concept of novel 1T-0C ZRAM based on non-filamentary RRAM

The volatile operation observed in  $\text{TiO}_2$  can be used to realize 1D-1R or 1T-1R RRAM array. However, integration of diode with the resistor in 1D-1R array results in several process and device related challenges [20]. For these reasons, 1T-1R RRAM configuration is preferred [21]. A significant advantage in the fabrication flow as well device operation can be obtained by realizing a 1T-0R RRAM array. This device can also mimic the 1T-0C ZRAM devices. This section discusses a novel 1T-0C device using the non-filamentary RRAM operation.

#### 6.3.1 Theoretical understanding of the proposed 1T-0C ZRAM

Schematic of the proposed 1T-0C ZRAM device is shown in Figure 6.4 (a). The device employs the non-filamentary RRAM ( $\text{TiO}_2$  in this case) at the gate of the transistor.  $\text{SiO}_2$  is used as a tunnel barrier between the substrate and the gate in order to restrict the gate leakage current.  $\text{TiO}_2$  is kept between the  $\text{SiO}_2$  and the gate electrode. The change in capacitance property of  $\text{TiO}_2$  RRAM can be used to change the drain current of the transistor. Since the change in gate capacitance results in a decrease in threshold voltage of the transistor, a larger memory

window can be achieved by sensing the drain current of the transistor. Note that the memory operation of the device doesn't require a current flow through the RRAM. Thus, this device can be an attractive candidate for low power memory applications.

Figure 6.4 (b) shows the circuit representation of the 1T-0C ZRAM.  $\text{TiO}_2$  can be equivalently represented as a parallel connection of a capacitor ( $C_{\text{TiO}_2}$ ) and a variable conductor ( $G_{\text{TiO}_2}$ ). The equivalent capacitance of the  $\text{TiO}_2$  RRAM is a complex quantity, but can be varied by modulating the oxygen vacancies inside the dielectric. Now, drain current ( $I_D$ ) in the saturation regime of transistor is governed by the equation [23]

$$I_D = \frac{1}{2} \mu C_{ox} \frac{W}{L} (V_{GS} - V_T)^2 \quad (6.1)$$

Where  $\mu$  is the channel carrier mobility,  $C_{ox}$  is the oxide capacitance,  $W/L$  is the channel width to channel length ratio,  $V_{GS}$  is the gate to source voltage and  $V_T$  is the threshold voltage of the transistor.  $V_T$  can be written as:

$$V_T = V_{FB} + 2\psi_s + \frac{\sqrt{(4\epsilon_s q N_a \phi_B)}}{C_{ox}} \quad (6.2)$$

Where  $V_{FB}$  is the flatband voltage,  $\psi_s$  is the surface potential,  $N_a$  is the substrate doping and  $\phi_B$  is the difference between intrinsic fermi level and the substrate fermi level, and  $\epsilon_s$  is the relative permittivity of the substrate. For the device shown in Figure 6.4 (a),  $C_{ox}$  would be the serial connection of the capacitance of  $\text{SiO}_2$  and  $\text{TiO}_2$  dielectric. With the application of external bias, the vacancies inside  $\text{TiO}_2$  can be modulated to change the capacitance of the dielectric. Comparing with equation (6.1), an increase in  $C_{ox}$  contributes to the increases the  $I_D$  by the same factor. Alongside,  $V_T$  is inversely proportional to the change in  $C_{ox}$ . Thus an increase in  $C_{ox}$  can result in an increase in  $I_D$  by a larger factor because of a square dependency of  $I_D$  to the  $V_T$  of the transistor. Assuming an increase of gate capacitance by  $\Delta C$ , the decrease in the  $V_T$  can be written as:

$$\Delta V_T = \Delta V_{FB} + \frac{\Delta C \sqrt{(4\epsilon_s q N_a \phi_B)}}{(C_{ox} + \Delta C)} \quad (6.3)$$

Interestingly,  $V_{FB}$  also decreases with the increase in gate capacitance [23]. Consequently, the modified drain current in the MOSFET after the programming operation can be re-written as:

$$I_{DP} = \frac{1}{2} \mu (C_{ox} + \Delta C) \frac{W}{L} (V_{GS} - (V_T - \Delta V_T))^2 \quad (6.4)$$

Where  $I_{DP}$  refers to the drain current after an increase in the gate capacitance. Equation

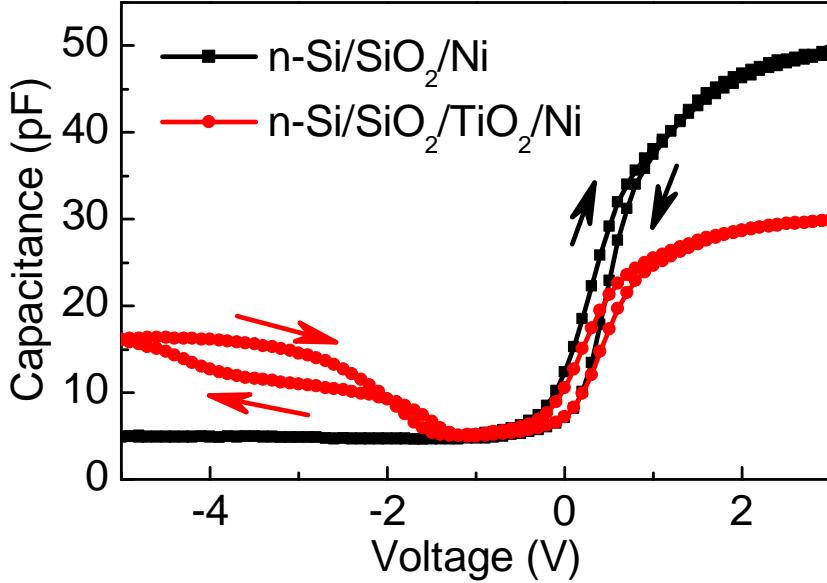


Figure 6.5: Comparison of HFCV characteristics between  $\text{SiO}_2$  MOS capacitor and  $\text{SiO}_2/\text{TiO}_2$  MOS capacitor.

(6.4) describes the possibility of creating a memory operation in the proposed 1T-0C ZRAM by using a non-filamentary RRAM dielectric in the gate of the transistor. By applying a program operation, the capacitance can be increased thereby increasing the drain current of the transistor. A sensor can be employed to sense the change in the drain current thereby realizing the memory operation.

### 6.3.2 Volatile memory transition in $\text{SiO}_2/\text{TiO}_2$ MOS capacitor

As mentioned previously, the modulation of oxygen vacancies inside the  $\text{TiO}_2$  dielectric can result in both volatile and non-volatile operation [15]. These vacancies can be physically moved to nearby conductance wells with the application of bias voltage thereby making a non-volatile conductance change in the dielectric. However, a moderate bias takes the vacancy at higher energy levels within the same conductance well, which in turn shows a volatile transition. If the bias is removed, the vacancy tries to attain at its equilibrium energy state. This transition time can effectively be used as a DRAM element.

After having a basic understanding of the non-filamentary RRAM operation, MOS capacitors were fabricated to demonstrate the feasibility of the proposed 1T-0C ZRAM. The MOS capacitors had  $\text{SiO}_2/\text{TiO}_2$  as the gate oxide. The MOS capacitors were fabricated on n-Si substrate with doping density of  $10^{16} \text{ cm}^{-3}$ . After wafer cleaning, 7 nm  $\text{SiO}_2$  dry oxide was grown

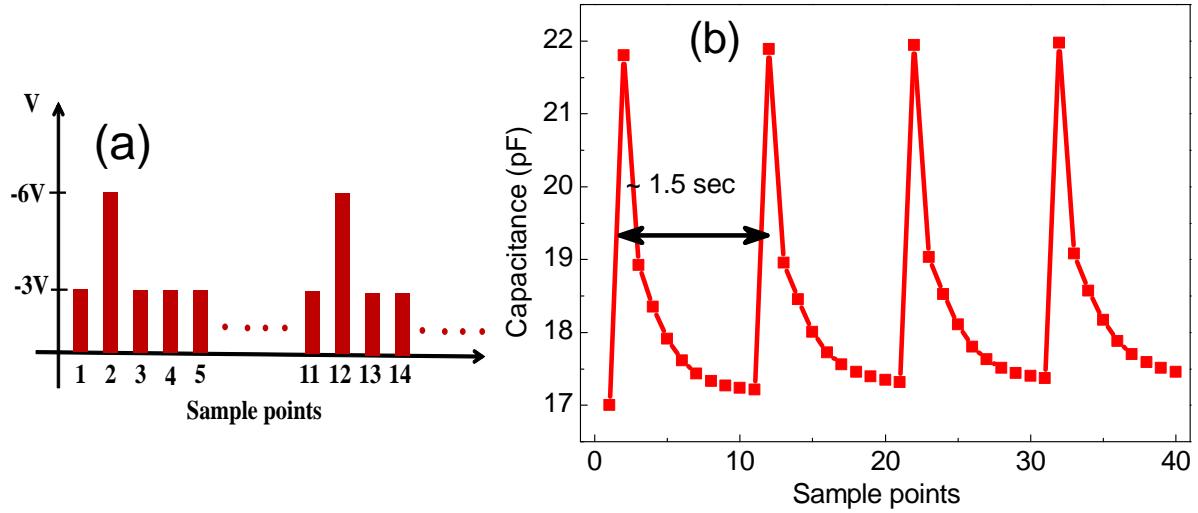


Figure 6.6: (a) Pulse CV measurement scheme to analyze volatile memory operation in  $\text{SiO}_2/\text{TiO}_2$  MOS capacitor; (b) volatile retention in  $\text{SiO}_2/\text{TiO}_2$  MOS capacitor analyzed with the pulsed CV measurement.

on the substrate, followed by 13 nm  $\text{TiO}_2$  deposited using ALD system. 90 nm Ni gate electrode was deposited using UHV-RF sputtering system, and was patterned using conventional lithography. The area of capacitors under observation was chosen to be 100  $\mu\text{m} \times 100 \mu\text{m}$ . For comparison, a control sample of  $\text{SiO}_2$  MOS capacitor was also fabricated with similar process steps. HFCV analysis was initially performed to study the effect of the presence of  $\text{TiO}_2$  in the gate stack of MOS capacitor.

$\text{SiO}_2$  MOS capacitor showed the typical HFCV characteristics at 1 MHz with a hysteresis because of charge trapping at the dielectric, as shown in Figure 6.5. On the other hand,  $\text{SiO}_2/\text{TiO}_2$  MOS capacitor showed a rise in capacitance at inversion, along with a hysteresis at the inversion regime. The reason behind the small rise in capacitance at the inversion regime of  $\text{SiO}_2/\text{TiO}_2$  MOS capacitor is unknown, but the presence of hysteresis loop is a good indication of memory operation possible with  $\text{SiO}_2/\text{TiO}_2$  gate dielectric. Thus, a change in capacitance of the gate oxide at the inversion regime of the MOS capacitor was confirmed. The hysteresis so observed at the inversion regime is possibly because of the volatile memory transition of inherent vacancies present inside the  $\text{TiO}_2$  layer.

Next, the MOS capacitor was tested for capacitance change using the pulse measurement technique, shown in Figure 6.6. The measurements were performed with negative voltages to assure the operations are carried at the inversion region. The pulse CV scheme is shown in

Figure 6.6 (a). Programming and read pulses were generated by the measurement system at different sample points. The programming pulse voltage was kept at -6 V, whereas the read voltage was kept at -3 V. Figure 6.6 (b) shows the measured capacitance at various sample points. A volatile increase in capacitance was observed in the  $\text{SiO}_2/\text{TiO}_2$  the MOS capacitor, and the capacitance was observed to decay within 1.5 sec. Similar volatile change has also been observed in other filamentary RRAM, and the retention time can vary from a few second to hundreds of seconds [17, 18]. Interestingly, other reports also confirm the non-volatile memory transition in the  $\text{TiO}_2$  based non-filamentary RRAM, which is even more promising for 1T-0C ZRAM. Keeping a non-volatile type mem-capacitive element in the proposed 1T-0C ZRAM will mitigate the memory refresh issue. Both the memory operations in non-filamentary RRAM doesn't require a current flow, and vacancies are modulated with the application of electric field. Thus, the incorporation of non-filamentary RRAM element in the gate oxide of a transistor can enable a promising 1T-0C ZRAM device.

## 6.4 Summary

Thus, non-filamentary RRAM was studied in this chapter of the thesis. Volatile memory operation using non-filamentary RRAM was successfully realized in  $\text{Ni}/\text{TiO}_2/\text{Ni}$  RRAM device. This volatile memory operation was further realized in  $\text{SiO}_2/\text{TiO}_2$  MOS capacitor with a retention time of around 1.5 seconds. With further research, a 1T-0C ZRAM device can be realized using the similar gate stack. Thus, this chapter provides the concept of a novel 1T-0C ZRAM along using non-filamentary RRAM.

## REFERENCES

- [1] M. J. Lee, C. B. Lee, D. Lee, S. R. Lee, M. Chang, J. H. Hur, Y. B. Kim, C. J. Kim, D. H. Seo, S. Seo, U. Chung, I. K. Yoo and K. Kim, "A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta<sub>2</sub>O<sub>5x</sub>/TaO<sub>2x</sub> bilayer structures," *Nature Materials*, vol. 10, pp. 625-630, 2011.
- [2] R. F. Freitas and W. W. Wilcke, "Storage-Class Memory: The Next Storage System Technology, *IBM Journal of Research & Development*, vol. 52, pp. 439-447, 2008.
- [3] J. B. Kwon, "Exploiting Storage Class Memory for Future Computer Systems: A Review," *IETE Technical Review*, vol. 32, no. 3, 2015.
- [4] S. K. Park, Technology Scaling Challenge and Future Prospects of DRAM and NAND Flash Memory, *IEEE International Memory Workshop (IMW)*, pp. 1-4, 2015.
- [5] A. Chen, J. Hutchby, V. Zhirnov, G. Bourianoff, Emerging Nanoelectronic Devices, John Wiley and Sons Ltd, 2014.
- [6] J. Liu, B. Jaiyen, Y. Kim, C. Wilkerson and O. Mutlu, An experimental study of data retention behavior in modern DRAM devices: Implications for retention time profiling mechanisms, *Proceedings of the 40th Annual International Symposium on Computer Architecture*, pp. 60-71, 2013.
- [7] M. Gajek, J. J. Nowak, J. Z. Sun, P. L. Trouilloud, E. J. OSullivan, D. W. Abraham, M. C. Gaidis, G. Hu, S. Brown, Y. Zhu, R. P. Robertazzi, W. J. Gallagher and D. C. Worledge, Spin torque switching of 20nm magnetic tunnel junctions with perpendicular anisotropy, *Applied Physics Letters*, vol. 100, 132408, 2012.
- [8] J. Li, Enabling phase-change memory for data-centric computing: Technology, circuit and system, *IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 21-24, 2015.

- [9] N. Rodriguez, F. Gamiz, and S. Cristoloveanu, A-RAM Memory Cell: Concept and Operation, *IEEE Electron Device Letters*, vol. 31, no. 9, pp. 972-974, 2010.
- [10] S. Okhonin, M. Nagoga, E. Carman, R. Beffa and E. Faraoni, New Generation of Z-RAM,, S. Okhonin, *IEEE International Electron Devices Meeting (IEDM)*, pp. 925-928, 2007.
- [11] C. Kuo, T. J. King and C. Hu, A capacitorless double-gate DRAM cell, *IEEE Electron Device Letters*, vol. 23, no. 6, pp. 345-347, 2002.
- [12] A. Makarov, V. Sverdlov, and S. Selberherr, “New Trends in Microelectronics: Towards an Ultimate Memory Concept,” *International Caribbean Conference on Devices, Circuits and Systems (ICCDSCS)*, pp. 12, 2012,
- [13] W. Wei, K. Namba and F. Lombardi, Extending Non-Volatile Operation to DRAM Cells, *IEEE Access*, vol. 1, pp. 758-769, 2013.
- [14] 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, *IEEE International Electron Device Meeting (IEDM)*, pp. 256-259, 2013.
- [15] R. Berdan, C. Lim, A. Khiat, C. Papavassiliou and T. Prodromakis, A Memristor SPICE Model Accounting for Volatile Characteristics of Practical ReRAM, *IEEE Electron Device Letters*, vol. 35, no. 1, pp. 135-137, 2014.
- [16] Y. F. Wang, Y. C. Lin, I. Wang, T. P. Lin and T. H. Hou, “Characterization and Modeling of Nonfilamentary Ta/TaO<sub>x</sub>/TiO<sub>2</sub>/Ti Analog Synaptic Device,” *Scientific Reports*, vol. 5, 10150, 2015.

- [17] R. Yang, K. Terabe, G. Liu, T. Tsuruoka, T.i Hasegawa, J. K. Gimzewski and M. Aono, On-Demand Nanodevice with Electrical and Neuromorphic Multifunction Realized by Local Ion Migration, *ACS Nano*, vol. 6, no. 11, pp. 9515-9521, 2012.
- [18] T. Chang, S. H. Jo and W. Lu, Short-Term Memory to Long-Term Memory Transition in a Nanoscale Memristor, *ACS Nano*, vol. 5, no. 9, pp. 7669-7676, 2011.
- [19] L. Qingjiang, A. Khiat, I. Salaoru, C. Papavassiliou, X. Hui and T. Prodromakis, Memory Impedance in TiO<sub>2</sub> based Metal-Insulator-Metal Devices, *Scientific Reports*, vol. 4, 4522, 2014.
- [20] C. T. Chou, B. Hudec, C. W. Hsu, W. L. Lai, C. C. Chang and T. H. Hou, “Cross-bar array of selector-less TaO<sub>x</sub>/TiO<sub>2</sub> bilayer RRAM,” *Microelectronics Reliability*, 2015. (doi:10.1016/j.microrel.2015.04.002)
- [21] M. Zangeneh and A. Joshi, “Design and Optimization of Nonvolatile Multibit 1T1R Resistive RAM,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 8, pp. 1815-1828, 2014.
- [22] S. Mills, B. Lee and V. Misra, “Atomic Layer Deposited TiO<sub>2</sub> Thin Films for Environmental Gas Sensing,” *IEEE Sensors*, pp. 1-4, 2013.
- [23] Y. Taur and T. H.Ning, “Fundamentals of Modern VLSI Devices,” *Cambridge University Press*, 2009.

## Chapter 7

# Conclusion and Future Work

### 7.1 Conclusion

In conclusion, this dissertation provides a broad overview of different emerging memory candidates in realizing the unified memory operation. Such a memory can be a potential candidate for applications like instant-ON computing, neuromorphic system, sensor system, etc. Flash is currently a mature technology, and has been under investigation for several decades. A novel flash memory was shown to have simultaneous volatile and non-volatile operation. Similarly, RRAM, which is currently attracting great research interest, can also be used to realize the unified memory operation. Independent non-volatile and volatile operation using different RRAM candidates has been demonstrated in this dissertation. Some of the key points to summarize the dissertation are mentioned below:

- Volatile memories (DRAM and SRAM) are highly used for logical operations, whereas non-volatile memories (flash) are used to store permanent data such as codes, media files, etc. Current computing systems have DRAM and flash located in different areas of the circuit. This in turn results in a slow processing speed and inefficient area scalability. Thus, the future memory technology devices are required to have storage and computation capability in the same area of the circuit.
- To achieve this goal, stacked flash and DRAM embedded memory has been proposed. A better scalability and performance can be obtained if volatile memory and non-volatile memory are realized in the same device. However, the stacking is achieved using TSV technology which poses several process and device level challenges. Thus, a better candidate can be a stand-alone device showing simultaneous volatile and non-volatile operation.
- Dual floating gate flash memory has been demonstrated to show simultaneous volatile and non-volatile operation. The non-volatile operation was observed to be similar to that

of traditional flash memory. The addition of extra floating gate provides efficient charge control between the floating gates thereby enabling a volatile memory operation.

- HfO<sub>2</sub> and Al<sub>2</sub>O<sub>3</sub> dielectric RRAM was observed to show non-volatile operation. The Al<sub>2</sub>O<sub>3</sub>/HfO<sub>2</sub> bilayer dielectric RRAM was observed to be advantageous than the HfO<sub>2</sub> RRAM in terms of improvement in forming voltage and I<sub>on</sub>/I<sub>off</sub> ratio. The non-volatile HfO<sub>2</sub> and Al<sub>2</sub>O<sub>3</sub> RRAM was used to demonstrate the analog memory behavior. The non-volatile analog memory behavior of RRAM was then further used to realize synaptic learning.
- TiO<sub>2</sub> RRAM was observed to show non-filamentary volatile memory behavior. The non-filamentary RRAM operation was used to propose a novel 1T-0C ZRAM device which can be attractive for future DRAM replacement.
- The future research in unified memory will consist of realizing an RRAM device capable of simultaneous filamentary and non-filamentary operation. The filamentary RRAM operation requires a high power, but provides a larger memory window than non-filamentary RRAM. For this reason, the filamentary operation can be used to store the permanent data. On the other hand, non-filamentary RRAM operation requires lower power than filamentary RRAM. This can be used to perform the logical operation requiring a low power memory.

Hence, this dissertation provides valuable ideas of realizing unified memory operation. With advancement in research, the ideas proposed in this dissertation can help in meeting the scaling needs of future memory technology, as well as improving the computing performance of future smart electronic devices.

## 7.2 Future work

### 7.2.1 Hybrid flash-RRAM unified memory

Dual floating gate memory MOSFET has been observed to be a potential candidate for unified memory applications. However, there are some challenges which need to be addressed in order to realize a mature unified memory candidate. For example, modern day flash memories suffers from several disturb mechanisms which can alter the  $V_T$  by more than a volt. The volatile operation of the DFGFET has a window of less than 1 V, hence a disturb mechanism can deteriorate the volatile data. Going back to some of the major challenges of DFGFET mentioned in chapter-3, (i) we require a replacement in IMD where we can realize volatile operation at

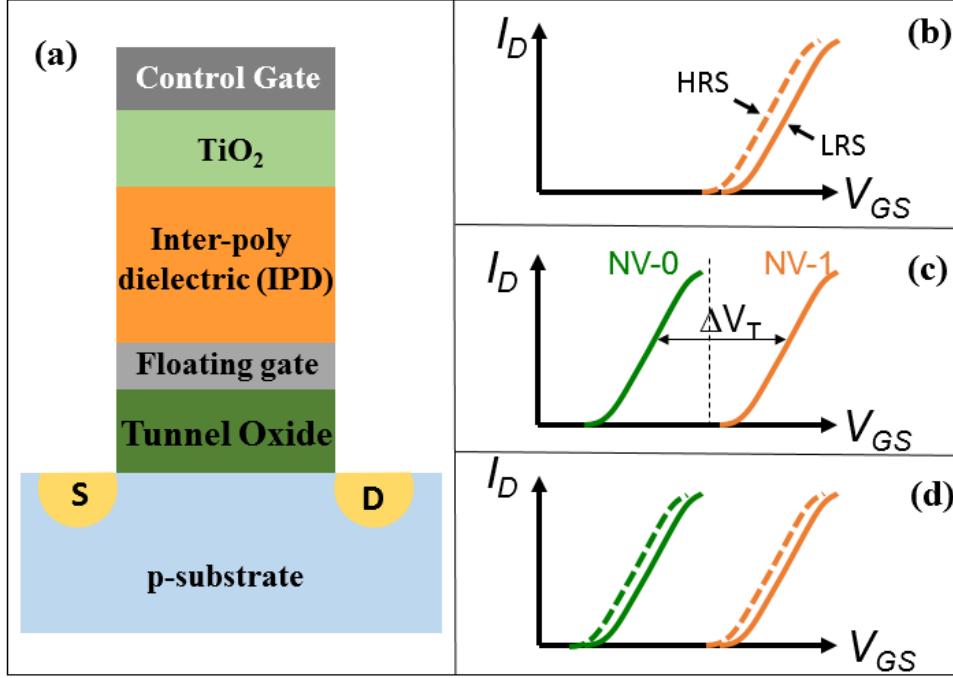


Figure 7.1: (a) Schematic of a hybrid flash-RRAM unified memory; (b) volatile operation, (c) non-volatile operation and (d) concurrent operation of the hybrid flash-RRAM unified memory.

nanosecond speed and (ii) we require a very high-k dielectric to prevent the fringing field effect from disrupting the volatile data.

Potential solutions to overcome these issues of DFGFET are:

- RRAMs have been demonstrated to show very fast switching like DRAM. Thus, an efficient solution is to achieve volatile memory operation using RRAM dielectric in the gate stack of the flash.
- Some of the RRAM dielectrics offer very high-k dielectric constant. Incorporation of such a high-k dielectric can reduce the fringing field effect from the nearby cells. This will help in improving the volatile retention time.
- Memory operation in non-filamentary RRAM doesn't require a current flow through the dielectric. Thus, placing a non-filamentary RRAM in the gate stack of the flash can be useful in realizing a low power memory operation.

Hence the next part of the work would be to incorporate the RRAM into the unified flash memory architecture. For this purpose, non-filamentary RRAM (such as TiO<sub>2</sub>) can be used in

the gate stack of the transistor. The device can be connected in NOR configuration to enable random access of the memory elements. The volatile transition in  $\text{TiO}_2$  can be effectively used to replace the IMD in DFGFET device. The requirement of additional floating gate can also be resolved making the device simpler in fabrication. Figure 7.1 shows the device structure of  $\text{TiO}_2$  incorporated hybrid flash-RRAM unified memory. The device looks very similar to a traditional flash memory, except the gate stack have  $\text{TiO}_2$  near the control gate. The new device architecture can have following advantages over the DFGFET:

- The device will have single floating gate. This will be advantageous in terms of process complexity.
- Non-filamentary RRAM dielectric will be used to store the volatile bit. This in turn will provide low power memory with faster program/erase operation.
- Incorporation of dielectrics like  $\text{TiO}_2$  which have very high dielectric constant will help in mitigating the disturbance created by fringing field.

Similar to DFGFET, the device can also have three modes of operation viz. volatile mode, non-volatile mode and concurrent mode. Illustration of the different memory states in hybrid flash-RRAM unified memory device are shown in Figure 7.1 (a)-(c).

- **Volatile mode:** The volatile mode can be achieved by modulating the oxygen vacancies inside  $\text{TiO}_2$  RRAM with a lower control gate bias. Note that the volatile operation will need a refresh after every 1.5 sec. However, as mentioned previously, some of the non-filamentary RRAM also shows non-volatile operation. Integrating such a non-filamentary RRAM instead of  $\text{TiO}_2$  can be advantageous in solving the memory refresh criteria.
- **Non-volatile mode:** The non-volatile mode of operation is similar to the conventional flash memories. Charge transfer between the floating gate and the substrate will determine the non-volatile program/erase operation. A higher control gate bias will be required to transfer the charge from the floating gate to the substrate, and vice-versa.
- **Concurrent mode:** Concurrent mode of operation can be achieved by selecting proper control gate bias. Similar to DFGFET, the volatile operation will need to be restored in the device after every non-volatile operation. For this purpose, the pulse scheme mentioned in chapter-3 can be followed.

Thus, the incorporation of RRAM into the DFGFET structure can be advantageous in many terms. The flash can be used to store the permanent data. With appropriate control of charge flow between the substrate and the floating gate, the flash can be allowed to store

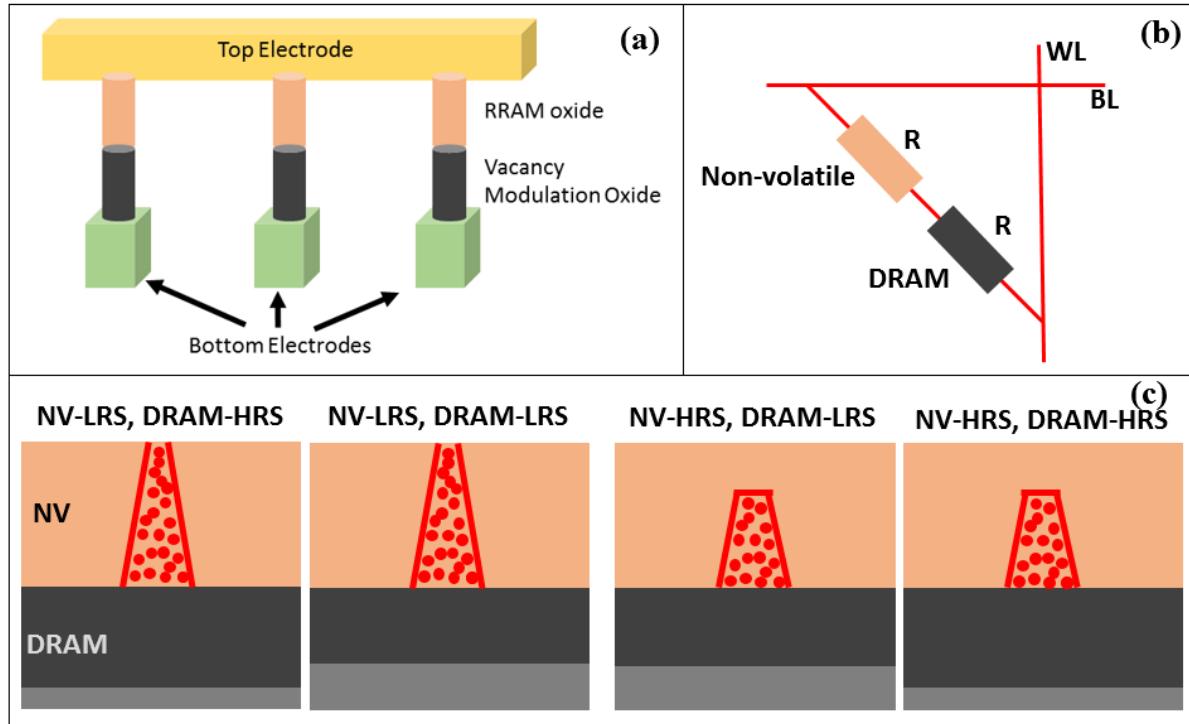


Figure 7.2: (a) Schematic of a crossbar unified memory; (b) circuit arrangement of the crossbar structure and (c) different memory states corresponding to the resistance state of the non-volatile RRAM and the vacancy modulation oxide RRAM.

multi-bit permanent data. Low power memory operation using non-filamentary RRAM can be used for logical computations. Hence, the hybrid flash-RRAM unified memory device can be an attractive candidate for future embedded memory applications.

### 7.2.2 1R-1R crossbar unified memory

In previous chapters, independent volatile and non-volatile operation has been demonstrated using RRAM. Thus, the next goal of RRAM research would be to realize the unified memory using RRAM array. The Schematic of the device and its operation is shown in Figure 7.2. The device consists of two memory resistors connected in series, one resistor corresponding to the filamentary RRAM (or RRAM oxide) and the other corresponding to the non-filamentary RRAM (or vacancy modulation oxide), as shown in Figure 7.2 (a). The circuit arrangement of the 1R-1R RRAM memory is shown Figure 7.2 (b). HfO<sub>2</sub> or Al<sub>2</sub>O<sub>3</sub> RRAM can be used as the filamentary RRAM candidate (Non-volatile resistor), whereas TiO<sub>2</sub> can be used as the non-filamentary RRAM candidate (low power memory like DRAM). Each 1R-1R element is connected between WL and BL represented as metal electrodes. Comparing with the hybrid

flash-RRAM unified memory device, 1R-1R unified memory device have a much simpler architecture. The 1R-1R RRAM unified memory can also be configured in crossbar array providing the lowest cell size and excellent scalability.

The 1R-1R crossbar unified memory will have four memory states depending upon the resistance status of the filamentary and the non-filamentary RRAM. Illustration of the different memory states in 1R-1R RRAM unified memory are shown in Figure 7.2 (c). The filament in the filamentary RRAM can be either in the set state (or LRS) or in the ruptured state (or HRS). Similarly, the volatile operation can be realized in non-filamentary RRAM by modulating inherent oxygen vacancies inside the dielectric. As observed in chapter-6, the conductance of the  $\text{TiO}_2$  can be changed with the application of bias voltage. The application of positive bias in  $\text{Ni}/\text{TiO}_2/\text{Ni}$  RRAM was observed to increase the conductance. Considering the higher conductance state of  $\text{TiO}_2$  to be LRS, and lower conductance state to be HRS, different modes of memory operation in 1R-1R crossbar unified memory can be described as:

- **Volatile: HRS, Non-volatile: LRS** → Presence of a filament in the filamentary RRAM will result in the LRS state of the non-volatile bit. At this point, a lower conductance of the volatile RRAM dielectric will result in the HRS state of the volatile bit.
- **Volatile:LRS, Non-volatile: LRS** → Similar to the above condition, presence of a filament in the filamentary RRAM will result in the LRS state of the non-volatile bit. However, a higher conductance of the volatile RRAM dielectric will result in the LRS state of the volatile bit.
- **Volatile: HRS, Non-volatile: HRS** → If the Filament in the filamentary RRAM is ruptured, this will result in the HRS state of the non-volatile bit. At this point, a lower conductance of the volatile RRAM dielectric will result in the HRS state of the volatile bit.
- **Volatile: LRS, Non-volatile: HRS** → Similar to the above condition, rupture of the Filament in the filamentary RRAM will result in the HRS state of the non-volatile bit. Similarly, a higher conductance of the volatile RRAM dielectric will result in the LRS state of the volatile bit.

Hence the 1R-1R crossbar unified memory can have at least four different states of operation realized. Since the filamentary RRAM provides a higher memory window, multi-bit non-volatile storage can be achieved with the filamentary RRAM. This structure is crossbar compatible thereby providing excellent scaling advantages. The non-filamentary memory operation doesn't require current flow through the dielectric, hence the volatile memory operation is a low power

operation. Note that the  $\text{TiO}_2$  used in chapter-6 showed a volatile operation with a retention of around 1.5 seconds. Further research is required to find a better candidate for the non-filamentary RRAM in order to resolve the refresh criteria.

### 7.3 Summary of the future work

The next phase of the unified memory research will consist of two devices, (i) hybrid flash-RRAM unified memory and (ii) 1R-1R RRAM crossbar array. The hybrid flash unified memory provides ease in the fabrication process flow as well several device level advantages over the DFGFET. The hybrid flash unified memory can be incorporated in NOR architecture to attain the fast random access property. Charge transfer between the floating gate and the substrate can be used to store permanent data. A non-filamentary RRAM candidate can be used to provide the low power memory required for logical operations. Similarly, 1R-1R can also be a potential unified memory candidate providing simultaneous volatile and non-volatile operation. A filamentary RRAM can be used for permanent data storage, whereas a non-filamentary RRAM can be used as low power memory for logical operations. 1R-1R unified memory can potentially be a cost effective unified memory candidate providing excellent scalability and very fast operation. Finally, more research is needed in the field of unified memory to provide cost effective solutions for future memory for storage and computation requirements.

## APPENDICES

# Appendix A

## Fabrication procedures

### A.1 Fabrication procedures of DFGFET

As mentioned in Chapter-3, The dual floating gate transistor consists of following gate stack: tunnel oxide (TO), bottom floating gate (BFG), intermediate dielectric (IMD), top floating gate (TFG), inter-poly dielectric (IPD) and control gate (CG). Figure A.1, Figure A.2 and Figure A.3 shows the process flow for fabricating the DFGFET device.

DFGFET was fabricated on p-type Si substrate with doping density of  $3 \times 10^{17} \text{ cm}^{-3}$ . Step by step pocess flow for fabricating DFGFET are mentioned below:

- Around 300 nm wet  $\text{SiO}_2$  serving as field oxide was grown on the wafer.
- Source/Drain active area was patterned on the wafer, and the oxide of the source/drain area was etched. The wafer was dipped into 6:1 BOE for 8 minutes. (etch rate of  $\text{SiO}_2$  in 6:1 BOE was observed to be around 53 nm/min).
- Source/Drain implantation of dopants was performed with energy of 20 keV and dose of  $5 \times 10^{15} \text{ cm}^{-2}$ .
- Dopant activation was done with 950 °C, 60 sec RTA at  $\text{N}_2$  ambient.
- Gate oxide area was then patterned for removal of sacrificial gate oxide. After patterning, the wafer was again dipped into 6:1 BOE for 8 minutes.
- Around 7 nm  $\text{SiO}_2$  serving as tunnel oxide was then deposited using ALD system.
- Around 4 nm TaN serving as the BFG was then deposited and patterned using conventional lithography technique. TaN was deposited using UHV-RF sputtering system.
- Around 4.5 nm of  $\text{HfAlO}$  serving as the IMD was then deposited using ALD system.

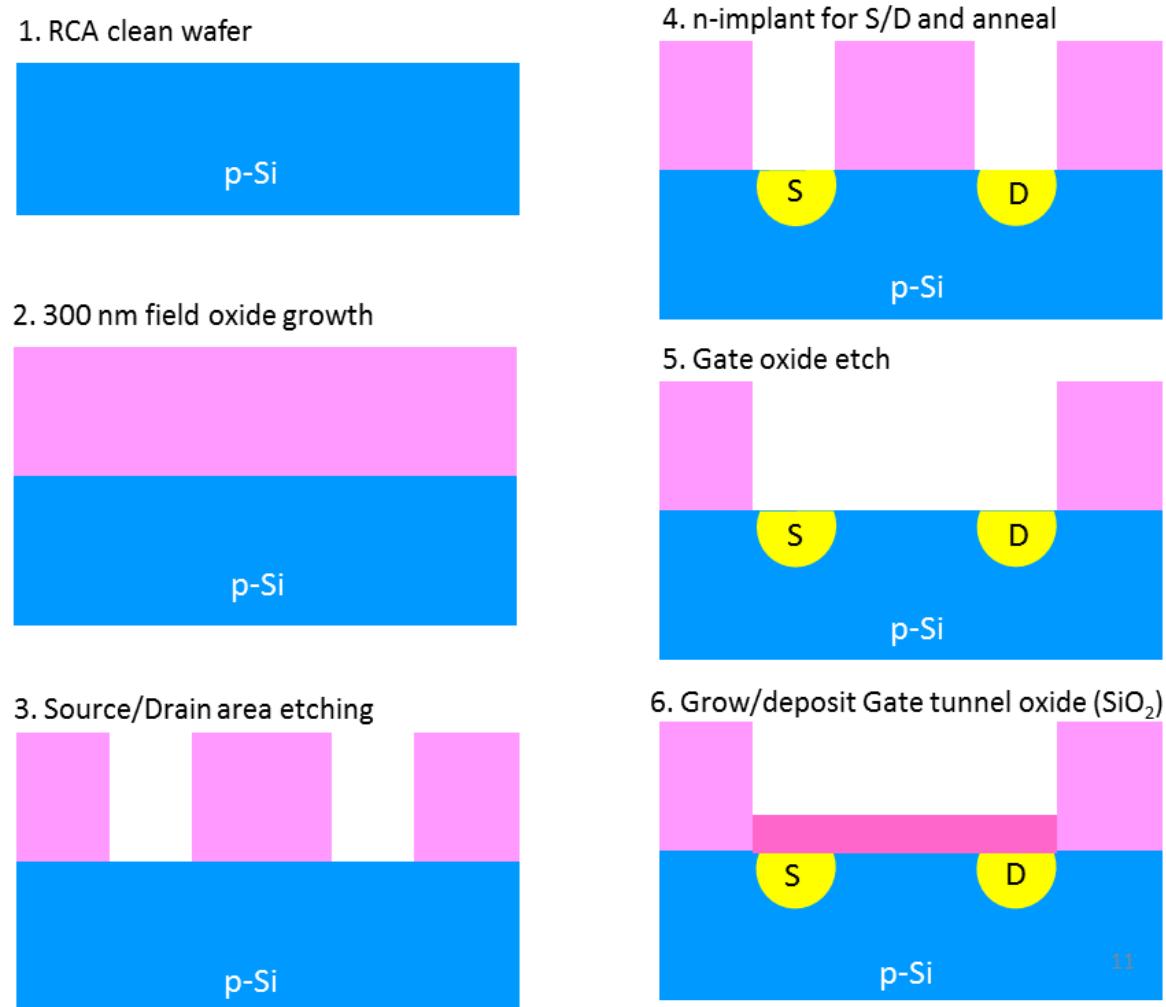


Figure A.1: Process flow for fabricating DFGFET (step-1 to step-6)

- Around 4 nm TaN serving as the TFG was then deposited and patterned using conventional lithography technique.
- Around 33 nm of HfAlO serving as the IPD was then deposited using ALD system. Interpoly dielectric was chosen to be 33 nm HfAlO with higher concentration of  $\text{HfO}_2$  in order to realize a physical thicker oxide without sacrificing the EOT.
- Via holes were then patterned and oxide layers were etched to open contact holes for source and drain.
- TaN capped with W serving as the CG as well as source/drain metal was then deposited and patterned.

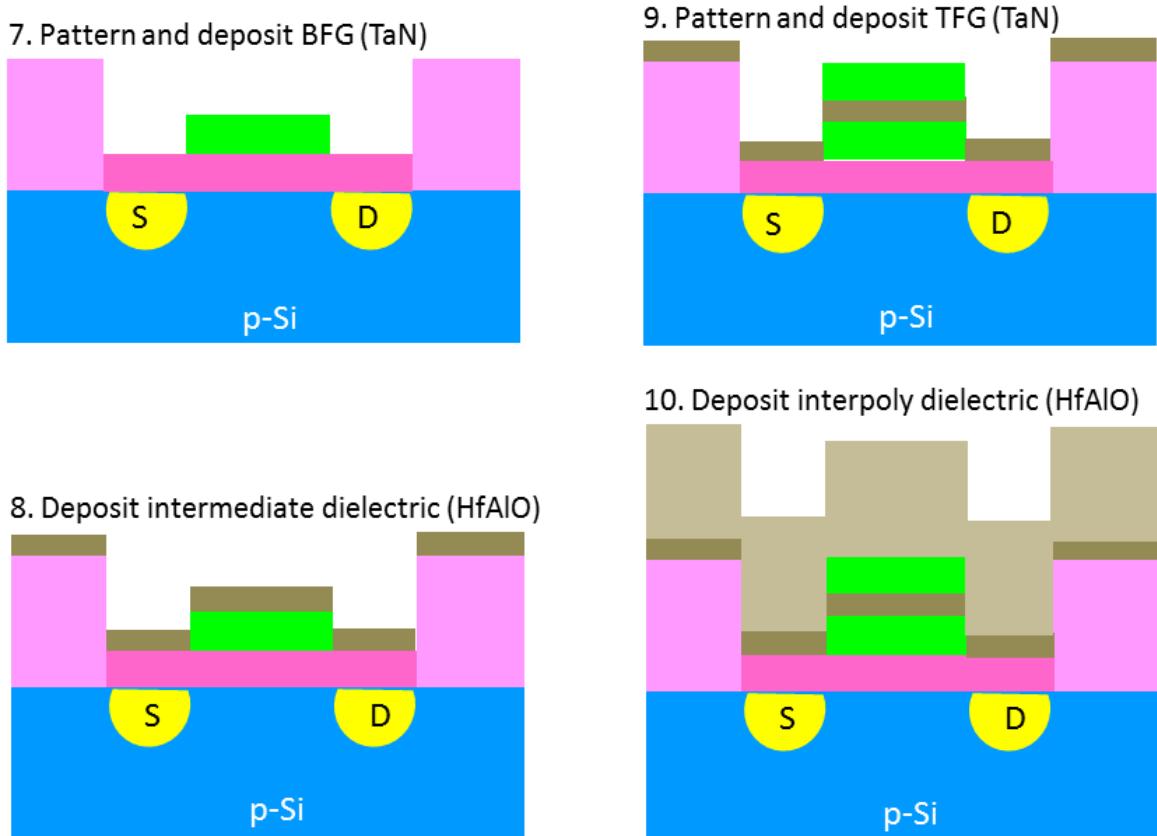


Figure A.2: Process flow for fabricating DFGFET (step-7 to step-10)

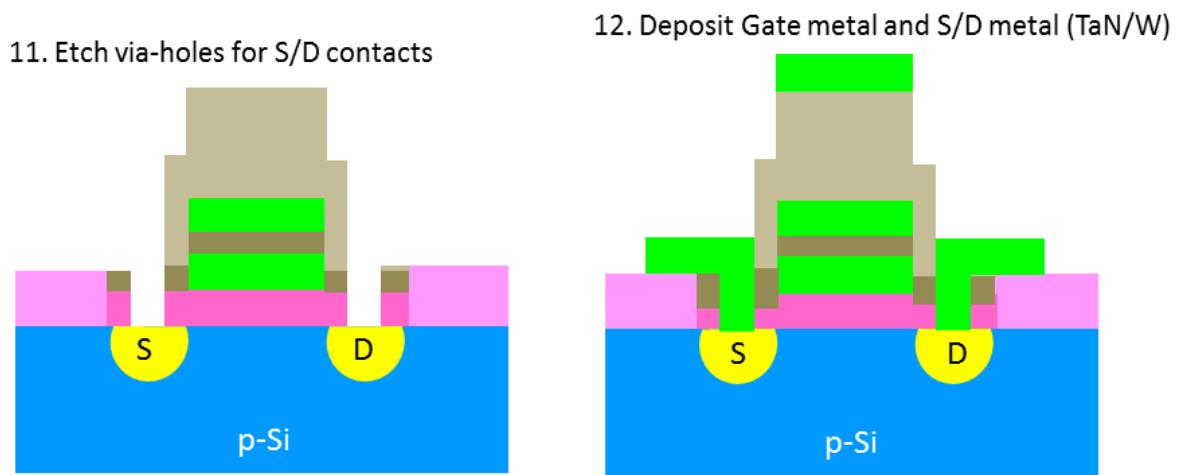


Figure A.3: Process flow for fabricating DFGFET (step-11 to step-12)

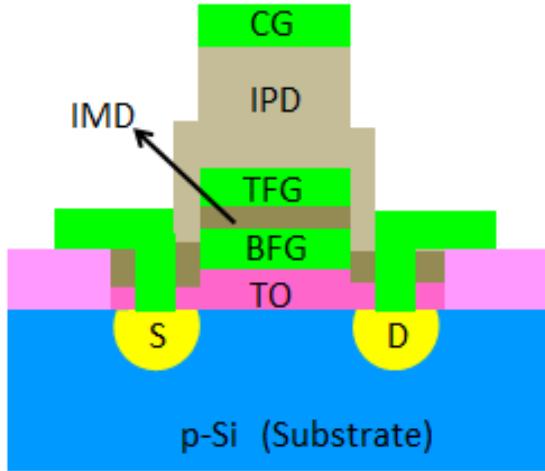


Figure A.4: Final structure of DFGFET. Transistor had channel length of  $10 \mu\text{m}$ , S/D doping density of  $2 \times 10^{21} \text{ cm}^{-3}$ . Stack thicknesses are; TO: 7 nm, BFG: 4 nm, IMD: 4.5 nm, TFG: 4 nm, IPD: 33 nm, CG: 100 nm.

- A post metal anneal was performed with  $750^\circ\text{C}$ , 60 sec RTA at  $\text{N}_2$  ambient to remove sputter induced defects and make silicide source/drain contact.

Figure A.4 shows the schematic of final device structure of the fabricated DFGFET. During all the experiments in chapter-3, source and substrate were grounded. Bias was applied at CG with drain grounded for every program and erase operation. During  $V_T$  or  $I_D$  measurements, bias were applied at CG and drain.

## A.2 Fabrication procedures of RRAM

RRAM has a relatively simpler device architecture. As mentioned in the introduction section, RRAM devices are just metal-insulator-metal capacitors. Figure A.5 shows the fabrication process flow for the RRAM devices mentioned in chapter-4, chapter-5 and chapter-6 respectively. Step by step process flow for fabricated RRAM are mentioned below:

- Around 400 nm wet  $\text{SiO}_2$  was grown on n-Si wafer.
- Bottom electrode was then deposited using UHV-RF sputtering system.
- The RRAM dielectric was then deposited on top of the bottom electrode. All the dielectrics were deposited using ALD system. In bi-layer dielectric RRAM mentioned in chapter-4,  $\text{Al}_2\text{O}_3$  was kept near the bottom electrode.

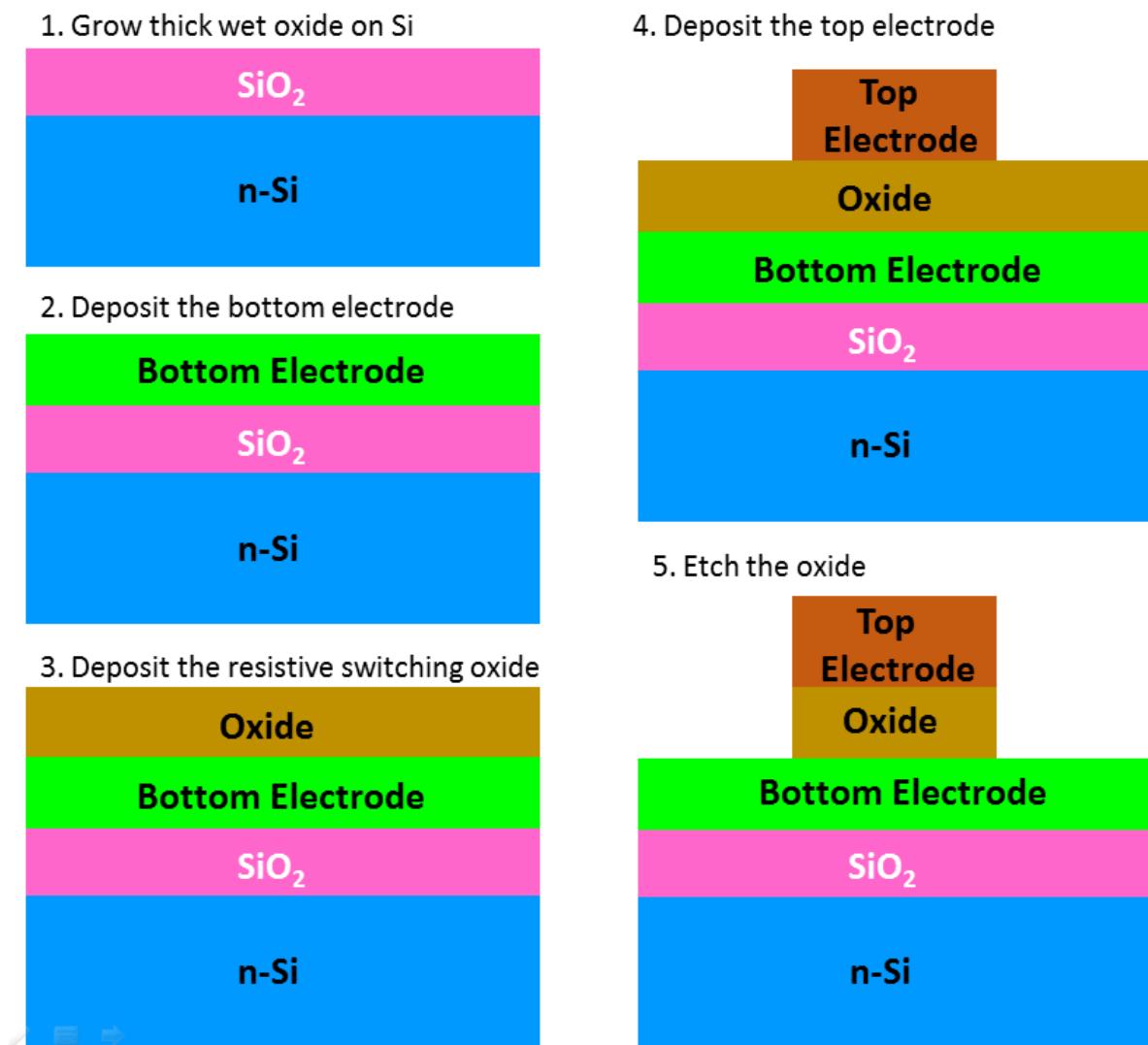


Figure A.5: Process flow for the fabrication of RRAM devices.

- Top electrode was then deposited using RF-UHV sputtering system, and was patterned using conventional photolithography system.
- wafers were then dipped into 1% HF for exposing the bottom electrode.

All the devices mentioned in chapter4, chapter-5 and chapter-6 had an area of  $100 \mu\text{m} \times 100 \mu\text{m}$  (unless specified otherwiese). The values of compliance current and the bias conditions are mentioned in every chapter accordingly.

## Appendix B

# Finding the zero-field activation energy of Al<sub>2</sub>O<sub>3</sub> and HfO<sub>2</sub> dielectric

Zero-field activation energy of the dielectric plays an important role in determining the breakdown characteristics of the dielectric. It has been shown that the origin of the forming process for RRAM device is attributed to the generation of oxygen vacancies inside the dielectric when electrical field is applied. Generation of oxygen vacancies are described by the equation:

$$G = G_0 \exp\left(-\frac{E_a - bF}{KT}\right) \quad (\text{B.1})$$

Where  $G$  represents the vacancy generation rate,  $G_0$  represents the effective vibration frequency between metal and oxygen ion,  $E_a$  is the zero-field activation energy,  $b$  is the bond polarization factor and  $F$  is the effective electric field. It has also been observed that time dependent dielectric breakdown (TDDB) is inversely proportional to  $G$ , which gives:

$$-KT \ln\left(\frac{1}{TDDB \cdot G_0}\right) = E_{a,eff} = E_a - bF \quad (\text{B.2})$$

Where  $E_{a,eff}$  refers to the effective activation energy of the dielectric at certain applied electric field. Hence a plot between  $E_{a,eff}$  and  $F$  for a dielectric will result in  $b$  (slope) and  $E_a$  (intercept) values. In order to determine values of  $E_{a,eff}$  at different electric field, TDDB measurements needs to be performed. Relationship between TDDB and temperature can be written as:

$$\ln(TDDB) = \frac{E_{a,eff}}{KT} - \gamma F \quad (\text{B.3})$$

Hence the slope of line plotted between  $\ln(TDDB)$  and  $1/KT$  will determine the values of  $E_{a,eff}$  which in turn will determine the  $E_a$  of the dielectric. To evaluate the  $E_a$  of individual

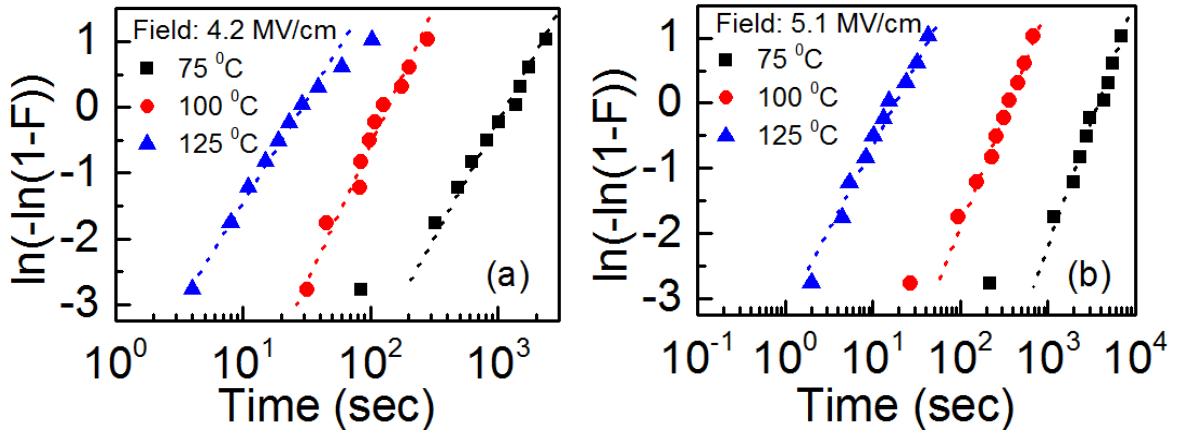


Figure B.1: (a)-(b) Weibull distribution of TDDB performed on Al<sub>2</sub>O<sub>3</sub> and HfO<sub>2</sub> dielectric RRAM respectively.

dielectric, TDDB measurements performed on single layer dielectric MIM capacitors with 6.4 nm of Al<sub>2</sub>O<sub>3</sub> and 5.3 nm HfO<sub>2</sub> as dielectric respectively, shown in Figure B.1. Thickness of Al<sub>2</sub>O<sub>3</sub> and HfO<sub>2</sub> were so chosen in order to have similar forming voltage. TDDB for both the dielectric were measured for three different temperatures and three different electric fields.

From the TDDB analysis, mean breakdown time was found out using Weibull distribution plot. Figure B.1 (a) and shows the Weibull distribution of mean breakdown time for Al<sub>2</sub>O<sub>3</sub> at 4.2 MV/cm for different temperatures. Similarly Figure B.1 (b) shows the mean breakdown time for HfO<sub>2</sub> dielectric at 5.1 MV/cm for different temperatures. Mean time is then recorded as 0-point of the Weibull plot, which corresponds to mean breakdown time for 63% devices. Similarly, mean breakdown time was also observed for both the dielectrics at three different applied electric field. All the mean breakdown times and corresponding electric fields were then plotted. Figure B.1 (c) and Figure B.1 (d) shows the mean breakdown time vs  $1/KT$  plot for Al<sub>2</sub>O<sub>3</sub> and HfO<sub>2</sub> dielectric respectively. Slopes of lines at Figure B.1 (c)-(d) represents the  $E_{a,eff}$  at different electric field. Accordingly, all the  $E_{a,eff}$  were then plotted vs corresponding electric field, shown in Figure B.3. Slope of lines at Figure B.3 results in bond polarization factor ( $b$ ) of the dielectric and the intercept results in zero-field activation energy ( $E_a$ ) of the dielectric. Details of the analysis and corresponding references can be found in B. Sarkar, *et. al.* ECS Transactions, 2014.

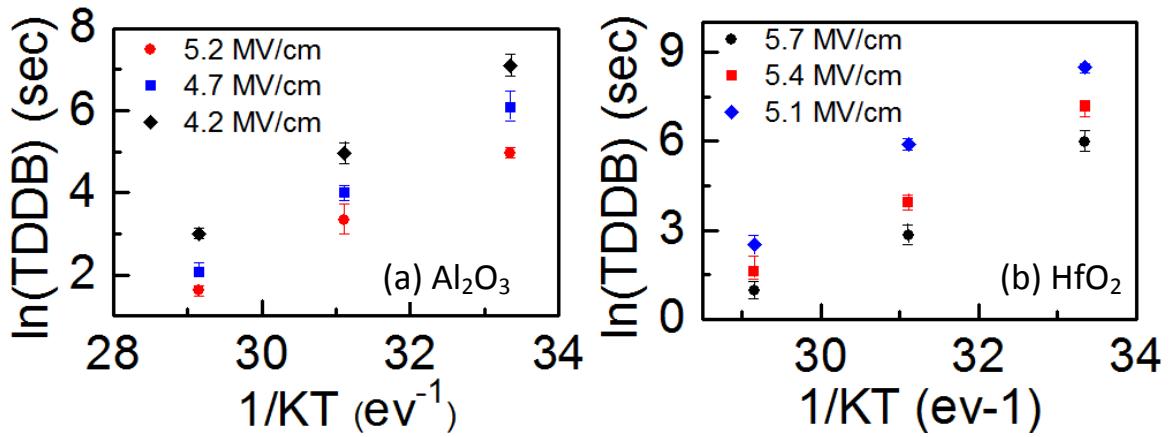


Figure B.2: (a)-(b) TDDB vs  $1/KT$  plot giving rise to a slope of  $E_{a,\text{eff}}$  at different fields for  $\text{Al}_2\text{O}_3$  and  $\text{HfO}_2$  dielectric RRAM respectively.

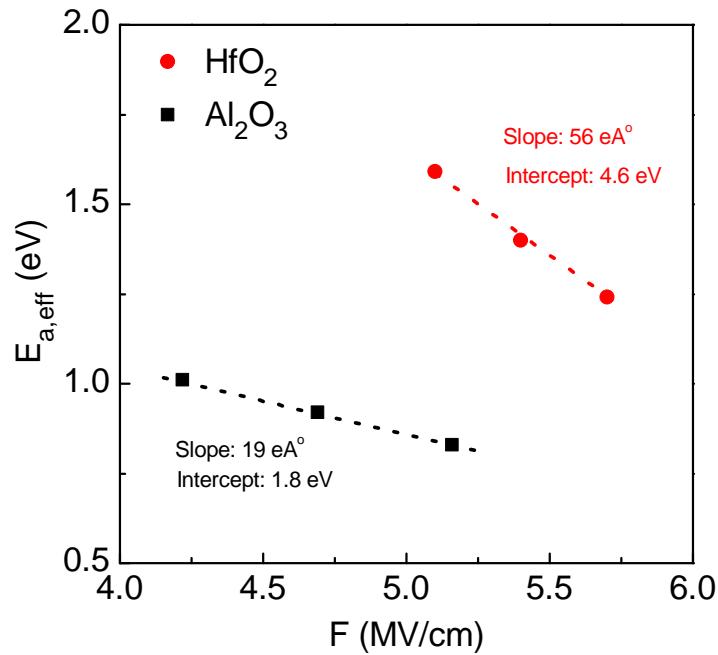


Figure B.3:  $E_{a,\text{eff}}$  vs electric field plot for  $\text{Al}_2\text{O}_3$  and  $\text{HfO}_2$  RRAM.