Understanding of oxide based resistive random access memory devices with multi-level resistance states and application

Wenbo Chen

University of Toledo

Follow this and additional works at: http://utdr.utoledo.edu/theses-dissertations

Recommended Citation

Chen, Wenbo, "Understanding of oxide based resistive random access memory devices with multi-level resistance states and application" (2017). Theses and Dissertations. 2172.
http://utdr.utoledo.edu/theses-dissertations/2172

This Dissertation is brought to you for free and open access by The University of Toledo Digital Repository. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of The University of Toledo Digital Repository. For more information, please see the repository's About page.
A Dissertation

entitled

Understanding of Oxide Based Resistive Random Access Memory Devices with Multi-level Resistance States and Application

by

Wenbo Chen

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the

Doctor of Philosophy Degree in Engineering

_________________________________________
Dr. Rashmi Jha, Committee Chair

_________________________________________
Dr. Mansoor Alam, Committee Member

_________________________________________
Dr. Christopher Melkonian, Committee Member

_________________________________________
Dr. Richard G. Molyet, Committee Member

_________________________________________
Dr. Hong Wang, Committee Member

_________________________________________
Dr. Amanda Bryant-Friedrich, Dean
College of Graduate Studies

The University of Toledo

August 2016
Copyright 2016, Wenbo Chen

This document is copyrighted material. Under copyright law, no parts of this document may be reproduced without the expressed permission of the author.
An Abstract of

Understanding of Oxide Based Resistive Random Access Memory Devices with Multi-level Resistance States and Application

by

Wenbo Chen

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the Doctor of Philosophy Degree in Engineering

The University of Toledo

August 2016

Non-volatile memory (NVM) are broadly used in removable media storage, smart phones, solid-state drive, etc. and flash memory technology has been dominating NVM market for over 20 years. However, further scaling of flash memory beyond sub 20 nm node brings tremendous challenges in device performance. For further scaling of Non-Volatile Memory technology beyond flash memory devices, Resistive Random Access Memory (ReRAM) devices have been proposed as a promising candidate due to its superior device performance and CMOS compatible process flow.

This dissertation focuses on three aspects of ReRAM research, multi-level cell (MLC) storage capacity, ReRAM crossbar array integration, and ReRAM practical application. In MLC part, multi-step forming technology was developed to substitute traditional one step forming in which current overshoot was suppressed and four stable resistance states were obtained. A comprehensive electrical characterization was conducted and trade-offs among different states were studied. In the section detailing crossbar integration, switchable diode based ZnO ReRAM device was identified as one of the promising candidates. Ru/ZnO/TiN/W stack demonstrated forming-free, self
compliance-current controlled, non-volatile, switchable diode type ReRAM characteristics in crossbar arrays which establishes potential application as high-density non-volatile memory. Finally, a complete voice cognitive system was built utilizing MLC ReRAM crossbar array to demonstrate functionality at the simulation level. With these contributions, this work has established solid guidance for future ReRAM fabrication and characterization as well as hardware system realization.
To my parents and best friends
Acknowledgements

How time flies and it is almost the end of my PhD studying at the University of Toledo. I would like to thank all of people who help and company me during these four years. This dissertation work would not have been possible without the help of theirs.

First of all, I would like to thank my advisor Dr. Rashmi Jha who started a great journey in my life at the University of Toledo and I really appreciated this opportunity. I personally benefit a lot from her research ideas and ambitiousness. This work would not have been possible without her consistent help. I am so proud to be her student.

Secondly, I would like to thank Tom Jacob and Dr. Christopher Melkonian. Tom provided a lot of help managing semiconductor research lab with his expertise in electronics and mechanics. While Dr. Melkonian and all engineers at Midwest MicroDevices offered tremendous assistance for device fabrications and problem solutions. This dissertation benefits a lot from their professionals and expertise.

Thirdly, I would like to thank my group mate and best friend, Wenchao Lu for valuable research discussion and his great suggestions.

Finally, I would like to thank the University of Toledo and Electrical Engineering and Computer Science Department for their financial support.
Table of Contents

An Abstract of .................................................................................................................. iii
Acknowledgements ........................................................................................................ vi
Table of Contents .......................................................................................................... vii
List of Tables ................................................................................................................ xi
List of Figures ................................................................................................................ xii
List of Abbreviations .................................................................................................... xx

Chapter 1. Introduction on Background ........................................................................ 1
  1.1. Research Background ............................................................................................ 1
  1.2. Emerging NVM technology and limitations ......................................................... 3
      1.2.1 MRAM ........................................................................................................... 4
      1.2.2 STT-MRAM .................................................................................................. 4
      1.2.3 FeRAM ......................................................................................................... 4
      1.2.4 PCRAM ........................................................................................................... 5
  1.3. Memristor .............................................................................................................. 6
  1.4. Dissertation Outline ............................................................................................... 7

Chapter 2. Literature Review .......................................................................................... 8
  2.1. Memristive Device ................................................................................................. 8
  2.2. Analog & Digital Memristor .................................................................................. 11
  2.3. Basic Operation of ReRAM Devices .................................................................... 12
Chapter 4. Performance of ReRAM Arrays

4.6. Recognition Rate ................................................................................................................. 50
4.7. Summary ............................................................................................................................... 52

Chapter 5. Ru/ZnO/TiN ReRAM Devices ..................................................................................... 54

5.1. Fabrication and testing setup ............................................................................................... 54
5.2. IRS Initial I-V and scalability ............................................................................................. 55
5.3. Switchable diode based resistive switching ......................................................................... 57
5.4. Switchable diode model ....................................................................................................... 59
5.5. IRS, LRS, HRS Schottky Emission fitting ........................................................................... 61
5.6. Endurance test and device degradation ............................................................................... 63
5.7. MLC Switching .................................................................................................................... 67
5.8. MLC read endurance .......................................................................................................... 68
5.9. Sneak current path in crossbar array ................................................................................... 68
5.10. Read out margin in N×N crossbar .................................................................................... 71
5.11. Pulse based set/reset ........................................................................................................ 71
5.12. Summary .......................................................................................................................... 72

Chapter 6. Conclusions and Future Work ................................................................................ 73

6.1. Conclusions ......................................................................................................................... 73
6.2. Future work ........................................................................................................................ 75

References ................................................................................................................................ 77

Appendix A. Image Recognition using Memristor ...................................................................... 93

A.1. Neurons ............................................................................................................................... 94
A.2. STDP Learning .................................................................................................................... 94
A.3. Crossbar Array ................................................................................................................... 97
A.4. STDP Implementation Using Memristive Synapse........................................98
A.5. STDP Based Learning Simulation............................................................101
  A.5.1 APS as Input Image Sensor ...............................................................101
  A.5.2 Architecture ...................................................................................103
  A.5.3 Learning Simulation .......................................................................104
  A.5.4 Image Recognition .......................................................................108
A.6. Summary .............................................................................................109

Appendix B. Ru/HfO2/TiN ReRAM Devices ..................................................110
  B.1. Device Fabrication and Testing ............................................................110
  B.2. Electrical Characteristics .................................................................111
  B.3. Summary ..........................................................................................118

Appendix C. Crossbar Mask Design and Process Flow .................................120

Appendix D. Ru/ZnO/TiN Fabrication Process Flow ......................................141
List of Tables

1.1: Next generation NVM tech. comparison with best projection in 2024 ...................... 7
3.1: Summary of Read/Write/Erase energy in MLC ReRAM ........................................ 36
4.1: Weight Value from Endurance Test ................................................................. 47
5.1: Schottky Barrier Height at Top and Bottom for IRS, LRS and HRS. ...................... 63
B.1: Endurance Pulse Parameter ........................................................................... 118
List of Figures

1-1: Worldwide NAND flash revenue forecast from 2011 to 2016. Adapted from [3]........ 2
1-2: Adversely impacts of scaling node on device endurance and raw bit error rate.
   Adapted from [4]................................................................. 2
1-3: MRAM device structure and switching mechanism........................................ 3
1-4: FeRAM device structure and switching mechanism....................................... 5
1-5: PCRAM device structure and switching mechanism....................................... 6
2-1: The relationship between four basic electrical quantities............................... 9
2-2: The missing “memristor” found............................................................... 9
2-3: A typical voltage driven memristor electrical characterization....................... 11
2-4: Switching mechanism for (a) analog memristor, and (b) digital memristor.......... 12
2-5: (a) Typical forming process by DC voltage. (b) After forming, conductive filament
   are formed....................................................................................... 13
2-6: (a) Unipolar switching mode and (b) bipolar switching mode.......................... 14
2-7: Filament shape after reset................................................................................ 15
2-8: Current overshoot caused by parasitic capacitance. Picture From [].................. 16
2-9: Dielectric materials and electrode materials used in modern ReRAM devices. Picture
   of periodic table from []........................................................................ 19
2-10: (a) Sneak path in ReRAM crossbar. (b) 1D1R structure to eliminate sneak path.. 22
2-11: Current-rectifying device I-V characteristics............................................... 23
2-12: Sneak path problem can be mitigated if current-rectifying ReRAM device is
employed. ........................................................................................................23

3-1: Devices schematics and initial I-V. (a) 1 μm ×1 μm W/Zr/HfO2/TiN structure. (b)
Initial I-V for multiple VRS devices, demonstrating repeatability. Inset shows
experimental setup..........................................................................................26

3-2: One-step forming and multi-step forming comparison. (a) One-step forming with
1T1R configuration and its resistive switching, significant current overshoot in the
initial reset (cycle 1) was observed. (b) Multi-step forming is consisting of multiple
one-step forming process. Compliance current increasing from 300 nA to 50 μA
were used. Inset shows device conductance level after each one-step forming. (c)
MLC ReRAM resistive switching after multi-step forming, the current overshoot in
the initial reset was effectively reduced. .................................................................. 28

3-3: Multi-level cell resistive switching and uniformity. (a) Multi-level cell resistive
switching was achieved by resetting the device with different reset stop voltages. (b)
Device conductance level of each resistance state. (c) Resistance distribution for
state 00, state 01, state 10 and state 11. .............................................................. 29

3-4: MLC write endurance. Triangular pulse of (a) trise and tfall = 100 μs, (b) trise and
tfall = 10 μs and (c) trise and tfall = 1 us were applied with set/reset voltages of
1.5V/-1.5V for switching between states 00 and 11, 1.5V/-1.2V for switching
between states 01 and 11, and 1.5V/-1V for switching between states 10 and 11,
respectively. Resistances were measured at 0.1V. .............................................. 31
3-5: MLC read endurance at 85°C. The devices were stressed with a train of read pulses of amplitude (a) 0.1 V and (b) -0.1 V with trise and tfall of 20 ns and pulse width of 20 ns at 10 MHz. The resistance was measured at 0.1 V.......................... 34

3-6: CVS test of each resistance state was conducted for three times. (a) 0.1 V CVS. (b) -0.1 V CVS. State10 was identified as the most unstable state. ......................... 36

3-7: (a) Filament shape and length for (a) state 11, (b) state 10, (c) state 01 and (d) state 00. Different HRS states correspond to different dielectric gap d between filaments and BE, (b) F-P curve fitting for I-V obtained in three HRS states 00, 01, and 10. 39

4-1: Neuromorphic voice recognition system conceptual process flow. 42

4-2: Training process architecture. The voice signal is filtered by a group of band-pass filters. Then signal in each channel is integrated by voltage integrator. The integration results distribution decides crossbar array weight distribution.......... 42

4-3: Voltage integration distribution and crossbar weight distribution. (a), (c), (e), (g) and (i) shows the voltage integration results for five vowels, while (b), (d), (f), (h) and (j) shows the crossbar weight distribution for five vowels. ................................. 45

4-4: nanoelectronic crossbar array mimic neural network................................................. 45

4-5: Recognition process architecture. Filtered and sampled signal fed into the crossbar array simultaneously and accumulated at the post-synaptic current summing neuron. Winner-take-all mechanism implements the highest input current as the best match ................................................................. 48

4-6: Dendrite summing current evolution over time. (a), (c), (e), (g) and (i) shows the evolution result under typical weight value, while (b), (d), (f), (h) and (j) shows the evolution result under random weight value. ................................................................. 52
4-7: The simulated recognition rate for recognizing ‘a’, ‘e’, ‘i’, ‘o’ and ‘u’ based on 100 voice samples (20 samples for each vowel). ................................................................. 53

5-1: (a) Devices I-V characteristics in initial resistance state (IRS). 20 µm × 20 µm, 30 µm × 30 µm, 50 µm × 50 µm, 70 µm × 70 µm and 90 µm × 90 µm were characterized. Multiple devices were tested and show repeatable initial I-V for different size, but we only show one device I-V of each size here for clarity. (b) Schematic representation of fabricated Ru/ZnO/TiN/W devices in crossbar structure. Device area is varying from 400 µm² to 8100 µm² with crossbar size ranging from 1 × 1 to 16 × 16. Bias was applied to top electrode, whereas the bottom electrode was grounded during electrical characterization. (c) Fabricated crossbar wafer....................................................................................................................... 56

5-2: (a) Switching I-V characteristics of Ru/ZnO/TiN ReRAM devices of area of 30 µm × 30 µm with ultra-low self-compliance 100 µA. The self-rectifying effect in both LRS and HRS were also observed. (b) Typical resistive switching I-V data across all device areas. DC voltage operation order is 0 → 7V → 0V → -7V → 0V. Similar switching I-V characteristics were observed for all devices. Devices show self-rectifying and forming free effects................................................................. 58

5-3: Maximum current during set and reset and resistance ratio extrapolation for different device areas. Smaller device shows lower operation current and higher resistance ratio HRS/LRS. .............................................................................................................. 58

5-4: I-V data of (a) LRS and (b) HRS by using DC voltage sweeping from -2V to 2V. Both HRS and LRS devices show current rectifying effect................................. 60
Schematic illustration of switching mechanism based on switchable diodes for our devices. (a) IRS device. (b) LRS after set. Positive bias is applied on top during set.
(c) HRS after reset. Negative bias is applied on top during reset. 60
CV analysis to support switching diode model. 61
(a), (c) and (e) show Schottky emission (SE) fitting for IRS, LRS and HRS. (b), (d) and (f) show band diagram for IRS, LRS and HRS devices. 63
DC endurance data using DC voltage with amplitude of 7V/-7V. (a) Resistance read at 2V. (b) Resistance read at -2V. 65
Forward and reverse current distribution obtained from 100 cycles DC endurance for (a) LRS and (b) HRS. 65
Pulse endurance under triangular pulse with rise and fall time 50 ms and amplitude of 7.2 V/-7.2 V. (a) Resistance read at 2 V. (b) Resistance read at -2 V. 66
Multiple resistance states can be achieved by using different set voltages or using different reset voltages. (a) Set voltages of 5.6V, 6V, 6.4V, 6.8V and 7V are used to set device to different LRS. (b) Reset voltages of -6V, -6.4V, -6.8V and -7V are used to reset device to different HRS. 67
Read endurance for obtained eleven resistance states from Fig. 13. (a) Read endurance for resistance states obtained by setting the device with different set voltages and read at 2 V. (b) Read endurance for resistance states obtained by resetting the device with different reset voltages and read at -2 V. 68
Sneak path in 1R crossbar structure in worst case and (b) device with self-rectifying effect can suppress sneak current to improve integration density. 70
5-14: Simulation of the read-out margin in $N \times N$ crossbar array for the worst case based on 1R setup. Experimental data matches projected results as indicated by ‘X’ mark in the figure. ................................................................. 70

5-15: Lower resistance can be obtained by longer pulse width. Device is set (a) or reset (b) by pulse with pulse width 30 ms, 50 ms and 90 ms. ............................................. 71

A-1: A typical neuron cell. Picture by Dr. C. George Boeree, webspace.ship.edu. ............ 95

A-2: Experimental STDP data adapted from Bi and Poo’s paper in 1998. ................. 96

A-3: Nanoelectronic crossbar array mimic neural network. ........................................ 98

A-4: Memristor moving wall model. In this model, the wall position $m$, which is dependent on bias voltage, separates two different resistance regions in series. ..... 98

A-5: (a) Ideal neuron spike shape used for simulation, (b) resulting STDP memristance update function and (c) resulting STDP memductance update function. Similarly, (c), (d), (e) corresponds to adapted neuron spike, its resulting STDP memristance and memductance update function, respectively............................................. 101

A-6: (a) Active pixel sensor (APS) schematic. PD is modeled with a current source ($I_{ph}$) and a capacitor ($C_{ph}$). (b) The dependence of APS circuit output current $I_{out}$ on image gray value shows that $I_{out}$ is proportional to gray value/light intensity. Note that for each training period (10 s), the first 1 s is always the system reset cycle. 102

A-7: Memristor neural network architecture using CMOS neurons. Memristor is connected in the crossbar array as shown in the inset. $G_{ij}$ refers the synapse conductance located at the $i$th row and the $j$th column. For a $N\times M$ crossbar array, it can store $M$ images of $N \times N$ pixels................................................................. 103
A-8: A 3x3 grayscale image with pixel numbered from 1 ~ 9 for neural network learning, and the corresponding gray value of each pixel is shown in the brackets............ 105

A-9: (a) Synaptic weights evolution for image (Figure A-8) training. (b) The resulting synaptic weights value (in the brackets) after training. G1 ~ G9 are the corresponding synapse of pixel 1~9 on the selected column.............................. 105

A-10: Binary image input for neural network training. They are named letter X, letter I and letter L from the left to the right, respectively................................. 106

A-11: Synapse training result for Figure 2-12. It clearly shows that all the input images has been memorized to the 1st ~ 5th crossbar column........................................ 107

A-12: Post-synaptic neuron output spike waveform monitored for 10 s during image recognition stage. (a) ~ (c) corresponds to the 1st ~ 3rd crossbar column and total spikes counted are 57, 18 and 36, respectively. ........................................ 107

B-1: (a) Typical NF and PF forming. (b) VRS initial IV, FRS initial IV after NF and FRS initial IV after PF. VRS refers to Virgin Resistance State. FRS refers Formed Resistance State. (c) Process flow................................................................. 113

B-2: Switching characteristics for (a) NFURS and (b) PFBRS................................... 114

B-3: UFBRS device I-V switching characteristics. (b) NFURS, PFBRS and UFBRS HRS/LRS distribution. (c) NFURS, PFBRS and UFBRS reset current distribution. ........................................................................................................... 116

B-4: Endurance for (a) PFBRS, (b) UFBRS and (c) NFURS................................. 118

C-1: 5” × 5” crossbar mask design template. All mask design in the following discussion is developed by this template. .............................................................. 121
C-2: Global marks using four letters “U”, “T”, “C” and “B” to roughly align two masks in fabrication process................................................................. 122
C-3: Precise mark design on wafer for etching mask to precisely align two masks...... 123
C-4: Precise mark design on mask for etching mask to precisely align two masks. ...... 124
C-5: Precise mark on wafer for lift-off mask to precisely align two masks. .............. 125
C-6: Precise mark on mask for lift-off mask to precisely align two masks. ............. 126
C-7: Mask #1 for patterning BE or bottom crossbar.............................................. 127
C-8: Mask #2 for patterning stack and TE............................................................. 128
C-9: Mask #3 for patterning insulator openings. ................................................. 129
C-10: Mask #4 for patterning top crossbar. ....................................................... 130
C-11: Overview of mask design by overlapping four masks together...................... 131
C-12: Mask layout showing placement of different device size areas.................... 132
C-13: BE patterned by using Mask #1............................................................... 134
C-14: Stack and TE patterned by using mask #2............................................... 134
C-15: Insulator openings patterned by using mask #3........................................ 134
C-16: Top crossbar patterned by using Mask #4. .............................................. 135
C-17: Design dimensions for 1 × 1 crossbar...................................................... 136
C-18: Design dimensions for 2 × 2 crossbar...................................................... 137
C-19: Design dimensions for 4 × 4 crossbar..................................................... 138
C-20: Design dimensions for 16 × 16 crossbar................................................. 139
C-21: Design dimensions for crossbar DIP package. ........................................ 140
List of Abbreviations

1D1R .......................... One Diode One ReRAM
1R .............................. One ReRAM
1T1R .......................... One Transistor One ReRAM

BE .............................. Bottom Electrode
BL .............................. Bit Line
BRS ............................ Bipolar Resistive Switching

CC .............................. Compliance Current
CF .............................. Conductive Filament
CMOS ........................ Complementary Metal Oxide Semiconductor
CVS ............................ Constant Voltage Stress

FeRAM ........................ Ferroelectric Random Access Memory
FRS ............................ Formed Resistance State

HRS ............................ High Resistance State
IRS ............................ Initial Resistance State

LRS ............................ Low Resistance State

MIM ............................ Metal-Insulator-Metal
MLC ............................ Multi-level Cell
MRAM .......................... Magnetic Random Access Memory

NF .............................. Negative Forming
NFURS ........................ Negative Forming Unipolar Resistive Switching
NVM ............................ Non-volatile Memory

PCRAM ........................ Phase Change Random Access Memory
PF .............................. Positive Forming
PFBRS .......................... Positive Forming Bipolar Resistive Switching

RAM ............................ Random Access Memory
ReRAM ........................ Resistive Random Access Memory
RS .............................. Resistive Switching
SC..............................Self-compliance
SLC..............................Single-level Cell
STT-MRAM......................Spin-transfer Torque Magnetic Random Access Memory
TE..............................Top Electrode
TMO..............................Transition Metal Oxide
UFBRS.........................Unipolar Forming Bipolar Resistive Switching
WL..............................Word Line
Chapter 1

Introduction on Background

1.1. Research Background

The Flash memory device has been dominating the non-volatile memory (NVM) application market since it was commercially introduced in the early 1990s [1-2]. During the past several decades, flash has been following Moore’s law and scaling rules to accommodate ever-increasing demands for large storage, fast accessing rate, and low power dissipation in a NVM storage medium, such as memory used in removable media storage devices (USB flash drive, SD card used in cameras, music players, etc.), solid-state drive (SSD), tablets, and smart phones. Worldwide NAND flash memory revenue forecast, reported by iSuppli in 2013 [3], is shown in Figure 1-1, and this trend was estimated to increase for the next couple of years.
Figure 1-1: Worldwide NAND flash revenue forecast from 2011 to 2016. Adapted from [3].

Figure 1-2: Adversely impacts of scaling node on device endurance and raw bit error rate. Adapted from [4].
However, further scaling of flash memory brings numerous problems in device fabrication and performance. It is generally accepted that 25 nm node is the limitation for flash memory. The impacts of scaling node on device endurance and raw bit error rate (BER) is shown in Figure 1-2 [4]. As technology node scales from 90 nm to 20 nm, the endurance data degrades from 10,000 to below 2,000, and, at the same time, BER increased from $10^{-9}$ to $10^{-2}$. Thus, the next generation of NVM memory devices are urged by the global market.

1.2. Emerging NVM technology and limitations

Much emerging NVM memory technology has been proposed to replace flash. The main types of NVM memory technology include Magnetic RAM (MRAM), Spin-transfer Torque Magnetic RAM (STT-MRAM), Ferroelectric RAM (FeRAM), and Phase Change RAM (PCRAM). One common point in these devices is that information (e.g. digital ‘0’ or digital ‘1’) is stored by changing resistance states (i.e. high resistance state or low resistance state) or polarization state of the device and stored information can be retrieved by measuring the electrical current. However, switching mechanisms are different among devices.

![MRAM device structure and switching mechanism.](image)

Figure 1-3: MRAM device structure and switching mechanism.
1.2.1 MRAM

MRAM was first discovered by IBM research group in 1970s [5]. The basic device structure for MRAM contains a magnetic tunnel junction (MTJ) which includes free layer, tunnel layer, and fixed layer for data storage, as shown in Figure 1-3. The resistance change is through electrical bias. Low resistance states (LRS) happen when magnetic moments in the free layer are parallel with the magnetic moments in the fixed layer, and high resistance states (HRS) happen when they are anti-parallel. In terms of endurance (~ $10^{14}$) and access speed (< 20 ns), MRAM has advantages over flash memory devices (endurance: ~ $10^4$, accessing speed: ~ ms). However, limitations in MRAM are large device size (>> $10^{-2}$) and high unit price [6].

1.2.2 STT-MRAM

STT-MRAM is often regarded as one type of MRAM as both MRAM and STT-MRAM shared similar device structure design (Figure 1-3), and they are all magnetic memories. However, STT-MRAM revives the promises of MRAM with fast read time (< 10 ns) and write time (< 1 ns), and small cell sizes (best projected for 8 $F^2$) since switching in STT-MRAM is determined by switching current density rather than switching the current. Despite improvements to the STT-MRAM device over the MRAM device, thermal stability and high power consumption (~ $10^4$ times higher than flash) are still key challenges for STT-MRAM development [7].

1.2.3 FeRAM

FeRAM structure is consisting of one capacitor and one transistor structure. The transistor accesses ferroelectric state of the FeRAM to read information stored in the cell. When an electric field is applied across dielectric or ferroelectric materials, the FeRAM
polarizes, and while that field is removed, it depolarizes. However, the ferroelectric material exhibits hysteresis in a plot of polarization versus electric field, and it retains its polarization state. FeRAM has also achieved a level of commercial success with first devices released in 1993, but the high unit price, large device area (22 F^2), and high power consumption (~ 100 times higher than flash) are still big concerns in FeRAM technology [8].

Figure 1-4: FeRAM device structure and switching mechanism.

1.2.4 PCRAM

PCRAM is one type of NVM based on a class of materials called chalcogenide glasses that exist in two phase states, i.e. crystalline and amorphous. The device shows LRS when it is switched into the crystalline or ordered phase, and HRS when it is switched into an amorphous or disordered state. The main problem in PCRAM is that the joule heating causes power consumption (~ 10^4 times higher than flash). Another concern
is that phase changes depend on surface to volume ratio (S/V), whether or not the phase change can be uniformly controlled as the device further scales with high S/V ratio [9].

Figure 1-5: PCRAM device structure and switching mechanism.

1.3. Memristor

Unfortunately, none of the above devices seem ideal for next generation memory device applications. Here, we will focus on a new concept of memory device which is called memristive device or memristor. Memristor has a simple, two-terminal metal-insulator-metal (MIM) structure, which is considered to be the most promising candidate for its faster speed than PCRAM as well as having a simpler structure than magnetic memories, i.e. MRAM and STT-MRAM. In addition, memristor can be integrated to crossbar structure to achieve the highest density and with theoretical potential to further scale down to 5 nm nodes without device performance degeneration. The best performance and device parameters projection of different memory tech. in 2024 is shown in Table 1 from ITRS [10].
Table 1.1: Next generation NVM tech. comparison with best projection in 2024

<table>
<thead>
<tr>
<th>Feature Size (F)</th>
<th>NAND-Flash</th>
<th>MRAM</th>
<th>STT-MRAM</th>
<th>FeRAM</th>
<th>PCRAM</th>
<th>Memristor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Area</td>
<td>4 F²</td>
<td>10 F²</td>
<td>8 F²</td>
<td>12 F²</td>
<td>4 F²</td>
<td>4 F²</td>
</tr>
<tr>
<td>Read Time</td>
<td>0.1 ms</td>
<td>&lt; 0.5 ms</td>
<td>&lt; 10 ns</td>
<td>&lt; 20 ns</td>
<td>&lt; 10 ns</td>
<td>&lt; 10 ns</td>
</tr>
<tr>
<td>W/E Time</td>
<td>1 / 0.1 ms</td>
<td>&lt; 0.5 ms</td>
<td>&lt; 1 ns</td>
<td>&lt; 10 ns</td>
<td>&lt; 50 ns</td>
<td>&lt; 1 ns</td>
</tr>
<tr>
<td>Retention</td>
<td>10 years</td>
<td>&gt; 10 years</td>
<td>&gt; 10 years</td>
<td>30 years</td>
<td>&gt; 10 years</td>
<td>&gt; 10 years</td>
</tr>
<tr>
<td>Endurance</td>
<td>5000</td>
<td>&gt; 1E16</td>
<td>&gt; 1E15</td>
<td>&gt; 1E15</td>
<td>&gt; 1E9</td>
<td>&gt; 1E16</td>
</tr>
<tr>
<td>Write Volts.</td>
<td>15 V</td>
<td>&lt; 1.5 V</td>
<td>&lt; 1 V</td>
<td>0.7 - 1.5 V</td>
<td>&lt; 3 V</td>
<td>&lt; 0.5 V</td>
</tr>
<tr>
<td>Read Volts.</td>
<td>1 V</td>
<td>&lt; 1.8 V</td>
<td>&lt; 1 V</td>
<td>0.7 - 1.5 V</td>
<td>&lt; 1 V</td>
<td>&lt; 0.2 V</td>
</tr>
<tr>
<td>Write Energy</td>
<td>&gt; 2E-17 J/bit</td>
<td>1.5E-13 J/bit</td>
<td>1.5E-13 J/bit</td>
<td>7E-15 J/bit</td>
<td>1E-15 J/bit</td>
<td>1E-17 J/bit</td>
</tr>
</tbody>
</table>

1.4. Dissertation Outline

The remainder of the dissertation will be organized as follows. Chapter 2 will do a comprehensive literature review of memristor development, implementations, and issues. Integrating memristive devices into crossbar arrays and solving issues to further increase crossbar integration density will be the focus of the entire dissertation. Multi-level cell (MLC) memristor devices with current-rectifying effect is regarded as the most efficient way to achieve this objective. Chapter 3 will comprehensively study TiN/HfOx/Zr MLC memristor devices. In Chapter 4, an application of speech cognitive system using MLC memristor crossbar array is introduced. To further increase integration density in crossbar array, Ru/ZnO/TiN devices with current-rectifying effects are fabricated in crossbar structure and discussed in Chapter 5. Chapter 6 presents conclusions and future work.
Chapter 2

Literature Review

2.1. Memristive Device

In 1973, Leon Chua postulated a new theoretical device model that connects flux $\varphi$ and charge $q$ for the first time [11]. As we all know, there are four fundamental electrical quantities, i.e. voltage $v$, current $i$, charge $q$ and flux $\varphi$. The relationship between four basic electrical quantities is shown in Figure 2-1. In Figure 2-1, flowing charge $q$ is integral current overtime; flux $\varphi$ is integral voltage over time. In addition, a resistor $R$ shows a static relationship between voltage $v$ and current $i$. A capacitor shows a static relationship between charge $q$ and voltage $v$. An inductor shows a static relationship between its current $i$ and flux $\varphi$. However, the relationship between flux $\varphi$ and charge $q$ is obviously missing. Mathematically, Chua proved that there should be a fourth fundamental circuit element existing, which is called memristor.
Figure 2-1: The relationship between four basic electrical quantities.

Figure 2-2: The missing “memristor” found.
Chua discussed two types of memristor, i.e. charge or current controlled and flux or voltage controlled. For charge controlled or current controlled memristor, the relationship between flux $\phi$ and charge $q$ is given by equation (2-1),

$$d\phi = M(q(t))dq$$

(1)

in which $M$ is the device memresistance. Analogously, the flux controlled or voltage controlled memristor relationship can be given as equation (2),

$$dq = W(\phi(t))d\phi$$

(2)

in which $W$ is the device memductance. As shown Figure 2-2, memristor completes the relationship in Figure 2-1 and shows the static relationship between flux $\phi$ and charge $q$.

Memristor is actually a type of resistor that the resistance can be changed by different voltage or current and the resistance after changing can be memorized in a non-volatile fashion [12]. In a broadly defined memristive system, memristive device can refer to any device has resistive behavior that the resistance can be changed by basic electrical quantities and at the same time exhibits memory for that resistance. A typical voltage driven memristor electrical characterization is shown in Figure 2-3. Triangular pulses were applied, with positive bias first, followed by negative bias second. Rise and fall times were kept at 10 ms and the pulse amplitudes were 1 V and -1 V. Device begin with a high resistance, and as the voltage increases, the current slowly increases until the maximum voltage reached, and this resistance is memorized by the device as can be seen by the return voltage sweep.
2.2. Analog & Digital Memristor

There are two types of memristor, analog memristor (i.e. memristor) and digital memristor (i.e. Resistance Random Access Memory or ReRAM or RRAM), as shown in Figure 2-4(a) and Figure 2-4(b) [13]. Although the switching mechanism are still unknown for memristor, here two most prevalent mechanisms are introduced to illustrate analog memristor and digital memristor switching behavior. The switching mechanism for analog memory is based on moving wall model [14], while the switching mechanism for digital memory is filamentary model [15]. The difference in switching mechanism model makes the difference in the resistance changing process for two types of memristor. For analog memristor, the moving-wall between the low resistance region and the high resistance region can be precisely controlled by an applied voltage or current so that memristance can be changed gradually or continuously, as shown in Figure 2-4(a). Ideally, any electrical resistance value can be obtained and thus this type memristor is called analog memristor. However, currently the fabricated analog memristor device are faraway to be applicable as the accuracy achieved in tuning memresistance is too low. On
the contrary, most memristors that people successfully fabricated and characterized are digital memristor which possibly based on filamentary switching mechanism. In filamentary switching mechanism, filament can be disrupted and re-formed in the resistive switching layer by electrical voltage, thus a high resistance state (HRS) or low resistance state (LRS) are stored. Digital memristor provides higher tolerance for resistance variability.

![Diagram of switching mechanism](image)

**Figure 2-4:** Switching mechanism for (a) analog memristor [14], and (b) digital memristor [15].

### 2.3. Basic Operation of ReRAM Devices

#### 2.3.1 Forming

ReRAM device is usually required to be formed prior to any other operation. Forming is considered to be the most critical step in achieving stable ReRAM switching characteristics and device performance, which was interpreted as a dielectric soft breakdown along the grain boundaries since grain boundaries have large number of defects and therefore are considered as the first locations to breakdown. However, the formation and the structure of filament are still unknown. Most accepted explanations are that they are grown along the grain boundaries. In the forming process, electrical voltage or current is stressed on the ReRAM device until current drastically changes at some
point and dielectric soft breakdown happens. To prevent device hard breakdown (or permanent break down), a compliance current is usually implemented to protect device. After forming, device is switched to formed resistance state (FRS). In FRS, a conductive filament which connects Top Electrode (TE) and Bottom Electrode (BE) is formed in the insulator layer. Figure 2-5(a) shows typical forming process by sweeping DC voltage and result in filament shown in Figure 2-5(b) is formed using current compliance 50 µA. After forming, device can be switched between low resistance state (LRS) and high resistance state (HRS) either in unipolar mode or in bipolar mode.

2.3.2 Unipolar and Bipolar Switching

Unipolar or bipolar switching depends on set and reset voltage polarity. If the set and reset voltages have the same polarity, then device is classified as unipolar, as shown in Figure 2-6(a). Otherwise, if device has opposite set and reset voltage polarity, then the device is considered to be bipolar, as shown in Figure 2-6(b). Since voltage polarity are
opposite for bipolar resistive switching, thus bipolar switching devices are more preferable over unipolar devices due to better noise margin between set and reset voltages.

Figure 2-6: (a) Unipolar switching mode and (b) bipolar switching mode.

2.3.3 Reset

If we assume filamentary model is correct, then the filament will be ruptured in the reset process. It is still not understood the main cause for rupturing filament. But most
are in agreement are that it relates with the movement of oxygen in TMO materials, since most switching materials reported are TMOs. For bipolar switching, oxygen movement is mainly caused by electric field direction, whereas for unipolar switching, joule heating plays a more important role [15]. As shown in Figure 2-6(a) and Figure 2-6(b), in reset process, the device starts at FRS or LRS, when electrical voltage is high enough, device will be switched to HRS. The highest voltage point in the reset process is called reset point and the corresponding voltage is called reset voltage. After reset, the filament shape is shown in Figure 2-7.

![Diagram of filament shape after reset](image)

**Figure 2-7: Filament shape after reset.**

### 2.3.4 Set

The set process is similar to the forming process with a partially formed filament at the beginning. After set, the conductive filament is formed again and thus device is switched to the LRS. Owing to residual filament from the last reset cycle, set voltage, which is the breakdown point in the set process, is usually lower than that in the forming process.
2.3.5 Current Overshoot

A stringent control over the maximum current during the forming/set process (the compliance current, $I_{cc}$) is necessary for achieving a filament of desirable resistivity. This is typically achieved by having a transistor in series with the ReRAM, as shown in Figure 2-8, and applying appropriate gate voltage on the transistor to control the current through the ReRAM. However, as was reported [16-18], if the transistor is externally connected to the ReRAM, an overshoot in the forming current ($I_{ov}$) over the pre-defined ($I_{cc}$) can occur due to the charging of parasitic capacitance ($C_p$) at the transistor-ReRAM junction, as shown in Figure 2-8. The magnitude of $I_{ov}$ is given as:

$$I_{ov} = C_p \frac{dV}{dt}$$

(3)

where, $dV/dt$ is the rate of voltage change across the $C_p$. Due to this transient charging, the total current flowing through ReRAM during forming exceeds $I_{cc}$, which leads to a formation of a much larger filament. As a consequence, devices experience higher reset current, and in the worst case, may fail to reset.

Figure 2-8: Current overshoot caused by parasitic capacitance.
2.4. Implementations and Issues

2.4.1 Memristor

In 2008, research group in Hewlett-Packard (HP) for the first time claimed they have found the missing memristor based on bilayer TiO$_2$/TiO$_{2-x}$ active layer [14]. In their model, a moving wall moving between conductive layer (i.e. TiO$_{2-x}$) and resistive layer (TiO$_2$) can be driven by electrical current and thus resistance is changed. However, analog switching, as claimed in the simulations, was not shown.

In 2009, the research group at the University of Michigan claimed they have fabricated memristor by co-sputtering of Ag and Si as an active layer with a properly designed Ag/Si mixture ratio gradient that leads to the formation of a Ag-rich (high conductivity) region and a Ag-poor (low conductivity) region, thus a moving wall, which is the interface between Ag-rich region and Ag-poor region, is formed [19]. The moving wall can be changed between Ag-rich area and Ag-poor area driven by biased voltage and thus memristance was changed accordingly. However, the device programming variations, device to device variations, and device immunity to read voltage needed further investigations [19].

After 2009, several other groups also demonstrated their memristor devices structure and performance. For example, in 2012, Alibart et al. at the University of California Santa Barbara reported Pt/TiO$_2$/Pt memristive device [20-21]. In 2013, research group at University of Toledo proved analog device reconfigurability in Mn doped HfO$_x$ [22-24]. Overall speaking, the variations in the memristor device is still a major concern, since even a small amount of memristance variation can lead to fatal system error in memristor based analog systems. Thus, high accuracy memristor device
employing new potential materials and new device stack is urged and still an active research area.

Another research interest lies in integrating memristor devices into crossbar structure to mimic neuromorphic system learning process in the human brain. To do so, an ideal memristor is usually employed to explore analog-memristor-based neuromorphic system functionality [14, 25-26]. With spike timing dependent plasticity (STDP) learning rule found in real neural network of human brain, memristor is employed to simulate the function of neuron synapse due to its analog resistance nature. Various neuromorphic recognition system is built up, such as handwriting recognition system, image recognition system, and voice recognition system [27-30]. However, the learning algorithms and recognition algorithms are still ambiguous, and there is no unified algorithms that can be generally accepted.

2.4.2 ReRAM

The first ReRAM device was reported in 1962 by T. K. Hickmott at General Electric Research Lab in metal-insulator-metal (MIM) structure of Al/Al2O3/Al, and they found the resistive switching was controlled by applied electric field [31]. In 1967, switching behavior between LRS and HRS in SiO2 was observed by D. R. Lamb and P. C. Rundle [32]. In 1968, a quaternary semiconductor consisting of tellurium, arsenic, silicon and germanium was reported and filamentary switching mechanism was first provided to support the device model [33]. ReRAM research started to take off since 1990s, during which various materials including transition-metal oxides (TMO), perovskite-type magnets and titanates, silicon oxides, and organic materials were explored. Dielectric materials and electrode materials that was mostly used is summed up
in Figure 2-9 [34-38]. Through years of development, ReRAM devices has shown exciting result and promises for next generation NVM technology, such as can be scaled down to 5 nm, fast switching speed (< 1ns), extremely low write/erase voltage (0.6 V/-0.2 V), etc. Besides further improving memory device important parameters, such as write/read voltage, switching speed, endurance and write/read energy, the main issues in the ReRAM devices are described in the next paragraphs.

Figure 2-9: Dielectric materials and electrode materials used in modern ReRAM devices.

Physical switching mechanism for ReRAM is still unclear. The popular mechanism can be classified into two groups: thermal effect and ionic effect [39]. Thermal effect refers to the filament dissolution is caused by joule-heating. This process is more like a fuse broken process in household but in nanoscale. The ionic process refers to migration of oxygen ions under electric field to form or dissolve filament in the switching materials. So far, parts of unipolar switching characteristics can be explained by thermal effect, and ionic migration model can address most of the bipolar switching
characteristics. Also, some other switching mechanisms were offered to explain resistive switching such as electronic effect and so on [40]. However, there was no single switching mechanism to explain all switching phenomenon and this topic is still under heated debate.

On the other hand, current overshoot caused by abrupt resistance change during the device forming/set process is considered to be main reason that results in the device performance degradation and limited reliability of devices [41-43]. It imposes major challenges for ReRAM development. One effort to address this issue lies in optimizing device structure to reduce forming/set voltage to minimize parasitic capacitance effect [44-46]. For example, doping N or metal (e.g. Mg, Hf, etc.) into resistive layer or by inserting an interfacial layer (e.g. Zr, Ti, etc.) can efficiently reduce current overshoot during forming/set due to significantly reduced forming/set voltage. Another effort lies in changing characterization method during the device forming/set process to reduce current overshoot, including multi-step forming, hot-forming and constant voltage forming [47-49]. Recently, a new concept of self-compliance (SC) was proposed to solve such current overshoot problems [50-53]. Self-compliance refers to device can apply compliance current by itself through depositing one layer of internal resistor layer or utilizing oxygen ions migration to form an internal resistor layer during device switching. The self-compliance can solve current overshoot problem radically as it eliminates parasitic capacitance issue caused by series-connected compliance force unit (e.g. transistor), and it is a promising for future ReRAM devices development.

Meanwhile, integration memory cell into crossbar structure is inevitable trend for achieving the highest density memory storage. In this fashion, the density of $4F^2$ cell area
can be achieved. However, the issue in the crossbar structure is the so-called sneak path problem. As shown in Figure 2-10 (a), the path of interest is to read off-state cell, but multiple sneak path exists as shown in red. In this way, the sneak path can result in false read [54-55]. To solve such problems, one common technique is integrating one transistor or one diode to constitute one transistor one ReRAM (1T1R) or one diode one ReRAM (1D1R) device stack in the crossbar array as access device in each cell. For 1T1R structure, it is obviously not desired due to scalability issue due to large device area of transistor, which do not offer competitive advantages compared to conventional floating-gate based memories (i.e. flash memory). Thus, 1D1R is preferred over 1T1R in which an ideal diode would be bidirectional, with low threshold voltage in both forward and reverse bias. Series-connected diode eliminates the sneak current as shown in Figure 2-10(b), but this diode should be specifically engineered for ReRAM devices. Instead of relying on an external diode as an access device, a more ideal approach is to take advantage of the inherent characteristics obtained in some ReRAM devices to break the sneak current path [56-58]. For example, several ReRAM devices were reported on inherently rectifying I-V characteristics at on resistance state. The current rectifying devices exhibits I-V characteristics which blocks current from flowing at the negative bias at on-state, as shown in Figure 2-11. Figure 2-12 shows the schematic of sneak path solution by employing such devices. In Figure 2-10(a), we can clearly observe that sneak path always include at least one memory cell at on-state which is in reversed bias, and if ReRAM device exhibits current-rectifying effect, the sneak path could be greatly mitigated, as shown in Figure 2-12. Eliminating external diode significantly simplifies the crossbar array complexity and also reduce the whole system power dissipation.
Figure 2-10: (a) Sneak path in ReRAM crossbar. (b) 1D1R structure to eliminate sneak path.
Figure 2-11: Current-rectifying device I-V characteristics.

Figure 2-12: Sneak path problem can be mitigated if current-rectifying ReRAM device is employed.

To further increase the scalability of ReRAM technology, multi-level cell (MLC) ReRAM devices and 3D stackable crossbar structure are two possible methodologies. MLC memory device refers to a memory cell that can store 2 bits per cell which is an
approach to exploit the layout area of a memory device towards high density memory application. Many materials, such as CuOx [59], TiOx [60], TaOx [61] and WOx [62], have demonstrated potential for MLC switching. However, majority of these studies have mainly focused on device fabrication and basic electrical testing to demonstrate multiple resistive states, verification algorithm programming and storage system scheme design, but less on the comprehensive electrical characterization of the access energies, reliability, and stability of each state and associated trade-offs. 3D crossbar structure can further reduce device area without occupying additional silicon area. For example, device area can be reduced to $4 \, \text{F}^2$ to $1 \, \text{F}^2$ with four crossbar layers. Also, the stackable structure was also proved by many research groups, including University of Michigan [63], Stanford University [64], etc.
Chapter 3

TiN/HfO2/Zr ReRAM Devices

Resistive random access memory (ReRAM) devices can be reconfigured into multiple resistive states that provides opportunities for achieving a multi-level cell (MLC), and hence, low-cost/gigabyte non-volatile data storage solutions [65-66]. In this chapter, firstly, we demonstrate ReRAM devices that can be programmed with low set/reset current and voltages. A low set/reset current could be achieved by implementing a multi-step forming technique that results in a controlled filament formation even without an ‘on-chip’ transistor in series to control maximum currents. Secondly, we have demonstrated MLC using these devices and characterized the endurance, reliability, and write/read/erase energy for each state. Read access energy depends significantly on the resistive state of the device as expected. However, reset or set energy shows a weak dependence on the resistive state. Reliability studies indicate that all states are not equally stable. Intermediate states obtained due to partial re-oxidation of the filament can drift over time.
Figure 3-1: Devices schematics and initial I-V. (a) 1 μm ×1 μm W/Zr/HfO2/TiN structure. (b) Initial I-V for multiple VRS devices, demonstrating repeatability. Inset shows experimental setup.
3.1. Fabrication and Testing

Resistive random access memory devices consisting of a TiN (bottom electrode, BE)/5nm-ALD HfO$_2$/Zr (interfacial layer)/W (contact layer, TE) stack with an area of 1 $\mu$m$^2$ (Figure 3-1(a)). Electrical characterization was performed using a LakeShore cryogenic probe station (under vacuum pressure of $1 \times 10^{-4}$ Torr) with a Keithley 4200 SCS and 4225 Ultra-fast I-V Pulse Modulation Unit (PMU). All testing was done using a 1T1R configuration: ReRAM device connected to an external NMOS transistor defining the maximum current (a so called current compliance limit) through the ReRAM devices, as shown in the inset of Figure 3-1(b)). Figure 3-1(b) shows the initial I-V plot for five devices in the as-fabricated state showing excellent uniformity across the wafer. A slight asymmetry in I-V between positive and negative bias can be attributed to a difference in the work function of Zr (~ 4.1 eV) vs. TiN (~ 4.6 eV). For device operation, reset is termed a “write operation” and set is termed an “erase operation”.

3.2. Multi-Step forming vs. one step forming

A forming step, which results in the formation of a conductive filament through the HfO$_x$ connecting the top and bottom electrodes, is considered to be one of the most critical steps for achieving repeatable switching. One-step forming implemented in the device and its corresponding resistive switching is shown in the Figure 3-2(a). Even though $I_{cc}$ of 50 $\mu$A was used during forming, the reset current ($I_{reset}$), which is the maximum current in the reset process, was almost 2 mA. Such high reset current indicates that the device suffered from current overshoot in the forming process ($I_{reset} = I_{ov}$). As a result of this current overshoot, sufficient reset of the filament could not be
achieved, which resulted in poor switching characteristics, evident from the DC switching curves in this plot.

Figure 3-2: One-step forming and multi-step forming comparison. (a) One-step forming with 1T1R configuration and its resistive switching, significant current overshoot in the initial reset (cycle 1) was observed. (b) Multi-step forming is consisting of multiple one-step forming process. Compliance current increasing from 300 nA to 50 μA were used. Inset shows device conductance level after each one-step forming. (c) MLC ReRAM resistive switching after multi-step forming, the current overshoot in the initial reset was effectively reduced.
Figure 3-3: Multi-level cell resistive switching and uniformity. (a) Multi-level cell resistive switching was achieved by resetting the device with different reset stop voltages. (b) Device conductance level of each resistance state. (c) Resistance distribution for state 00, state 01, state 10 and state 11.

To address the overshoot in $I_{cc}$ and achieve a more controllable filament formation, several other forming techniques have been studied, such as, constant voltage forming (CV-F), hot forming (H-F) or current sweep forming. In the CV-F technique, the device is stressed with a constant voltage until it is formed. Constant voltage forming can provide better forming results if lower stress voltage is used. However, using small stress voltage can take several hours for forming, which is not desired. In H-F technique, devices are formed using a standard DC sweep, or CV-F technique, at elevated
temperature (\(\sim 150^\circ C\)). Hot forming can efficiently suppress current overshoot by reducing the forming voltage at higher temperature. In addition, the forming time could be lower too compared to that in the CV-F forming technique. However, this process needs to be performed at higher temperature complicating forming set-up and detrimentally affecting device reliability. Current sweep forming is a forming process as opposite to voltage sweep which continuously sweep current from 0 to an expected current value to form filament and had been proved an effective forming method for HfO\(_x\) based ReRAM devices in terms of improving device performance [67]. Inspired by current sweep forming method but sticking with voltage sweep fashion while maintaining a good control over the forming process and resistive switching, a multi-step forming (MS-F) technique was implemented in this study. In this approach, the compliance current (CC) was step-wise increased from some smaller value up to the intended level. For instance, if the intended forming CC was 50 \(\mu A\), then the CC was gradually increased from 300 nA \(\rightarrow\) 1 \(\mu A\) \(\rightarrow\) 2 \(\mu A\) \(\rightarrow\) 50 \(\mu A\) through a series of I-V sweeps in a 1T1R configuration, as shown in Figure 3-2(b). The inset shows how conductance of the device increases after each sweep indicating a gradual formation of the filament. The gradual filament formation reduces the rate of the voltage change (dV/dt) across the device and hence the \(C_p\), which helps in limiting \(I_{ov}\) caused by the charging of the parasitic capacitance, \(C_p\). After the MS-F, the device was switched between low resistive state (LRS) and high resistance state (HRS) using CC of 50 \(\mu A\) during the set operation, with a reset stop voltage of -1.25 V, as shown by the DC switching curves in Figure 3-2(c). The CC was controlled using the external transistor during the set operation. Interestingly, a much lower reset current, which was almost equal to the forming and set current of 50
μA, was achieved using the MS-F technique as shown in Figure 3-2(c). When compared to Figure 3-2(a), a reduction in $I_{ov}$ and controlled filament formation is evident in Figure 3-2(c) when the MS-F technique is used.

![Figure 3-4: MLC write endurance. Triangular pulse of (a) trise and tfall = 100 μs, (b) trise and tfall = 10 μs and (c) trise and tfall = 1 us were applied with set/reset voltages of 1.5V/1.5V for switching between states 00 and 11, 1.5V/1.2V for switching between states 01 and 11, and 1.5V/1V for switching between states 10 and 11, respectively. Resistances were measured at 0.1V.](image)

### 3.3. MLC Switching using different reset stop voltages

After identifying a reliable forming technique, we focused on achieving MLC in these devices when formed using MS-F. Multi-level cell has been widely demonstrated in
ReRAM devices fabricated using various dielectric stacks. However, a critical factor is identifying the performance benefits and trade-offs of each of the states. In particular, the differences in endurance, reliability, and read/write/erase energies between the states were not yet studied. An in-depth understanding of these characteristics can provide guidance for designing energy-efficient architectures and data storage schemes [68]. Since our devices demonstrated switching with low set/reset currents and voltages when formed using the MS-F technique, these studies are particularly important and technologically relevant. As shown in Figure 3-3(a), multiple high resistance states were able to be achieved by resetting the ReRAM devices with different reset stop voltages from the same LRS (achieved using 50 µA CC, referred to as state 11), indicating the possibility for MLC. The three HRS achieved with reset stop voltage of -1.25 V, -1 V, and -0.65 V are referred to as states 00, 01, and 10 respectively. Figure 3-3(b) shows I-V for four resistance states. To understand the cycle to cycle resistance variation of these states, devices were switched from LRS to states 00, 01, and 10 for 20 cycles using a DC sweep. The resistance values obtained for different states after DC switching for 20 cycles is shown in Figure 3-3(c). The four clearly distinct resistance states with a tight distribution of resistance values is apparent in this figure. When comparing the variability among different states in Figure 3-3(c), one can observe that state 11 demonstrated the least variability followed by states 01, 00, and 10. This trend signifies that the least variable resistive state of the device corresponds to the condition when the filament is fully formed (i.e. state 11). This is because this state is strictly governed by the current compliance which can be tightly controlled by having a transistor. On the other hand, the states obtained after the retraction of filament (i.e. states 00, 01, 10) can demonstrate
variability to different extents based on the re-oxidation dynamics of the filament which is not completely understood.

3.4. MLC write endurance and device degradation mechanism

Next, we studied the endurance of these states. Figure 3-4 (a)-(c) shows MLC write endurance (i.e., switching of devices from state 11→00→11, 11→01→11, 11→10→11) with a series of symmetric triangular voltage pulses with rise time ($t_{\text{rise}}$) and fall time ($t_{\text{fall}}$) equal to 100 µs, 10 µs and 1 µs. Switching of the devices from LRS 11 to HRS 10, 01, or 00 is referred as a “write operation” while switching of devices from HRS 10, 01, or 00 to LRS 11 is referred as an “erase operation”. A clear dependency of the device endurance on the pulse time at constant voltage peak height is evident. $10^6$ cycles endurance for all states using $t_{\text{rise}}$ and $t_{\text{fall}}$ of 100 µs was observed. However, endurance degraded as the $t_{\text{rise}}$ and $t_{\text{fall}}$ were reduced. Interestingly, most of the failure was associated with the write operation where devices failed to reset to the desired state. The origin of this failure for faster pulses (e.g. 1 µs) is understood based on the “voltage-time trade-off” when the reset voltage needs to be increased as the switching time decreases to provide a sufficient amount of oxygen to re-oxidize the filament [69]. A question of the origin of the device failure to reset after a certain number of cycles (with 1 µs or 10 µs pulse) is of significant importance. The LRS resistance is defined by the CC (controlled by transistor), and does not indicate any obvious drift. Thus, a possible explanation of the insufficient reset for smaller pulse width at constant pulse peak voltage may be associated with the time required to allow oxygen ions to migrate towards the filament and sufficient electrical field to assist oxygen ion movements. Similar LRS for identical CC is expected even for different filament forming times, however, the oxygen ion distribution near the
filament is affected by different pulse time and/or electric field during filament formation or reformation and shorter pulse conditions may lead to more oxygen ions post filament forming or reforming in nearer proximity to the filament. These nearby oxygen ions can recombine with the filament leading to filament instability [70]. A longer pulse at constant peak voltage is needed to drive sufficient oxygen ions either towards the filament (during reset) or away from the filament (during set) as the number of switching cycles increase to maintain filament stability. It is noted that state 00 shows better write endurance than state 10 and 01, which can be correlated to the reliability of the device measured using Constant Voltage Stress (CVS) discussed below.

Figure 3-5: MLC read endurance at 85°C. The devices were stressed with a train of read pulses of amplitude (a) 0.1 V and (b) -0.1 V with trise and tfall of 20 ns and pulse width of 20 ns at 10 MHz. The resistance was measured at 0.1 V.

3.5. MLC read endurance test

Multi-level cell read endurance was measured using read pulses of ±0.1 V/60 ns at 10 MHz to understand a dependency on the polarity of the read pulse. To perform this test, first, four devices were switched to four different resistance states (11, 10, 01, and 00) using a DC sweep. Then the resistance corresponding to each state was measured by applying a series of read voltage pulse of 0.1 V/60 ns at 10MHz, shown in Figure 3-5 (a),
and -0.1 V/60 ns at 10 MHz, shown in Figure 3-5 (b). Figure 3-5 (a) and Figure 3-5 (b) shows MLC read endurance up to $10^9$ cycles for all states under these test conditions at 85°C. Excellent read endurance can be observed without any obvious signs of failure for all states.

3.6. **MLC cell access energy**

Next, we focused on characterizing the write/read/erase energy of the different states. These were calculated by integration of the voltage-current product over time. The write energy of state 00, 01, and 10 was calculated as the energy needed to switch the device from state 11 to state 00, state 11 to state 01, and state 11 to state 10, respectively. A triangular write pulse of $t_{\text{rise}}$ and $t_{\text{fall}}$ = 1 µs was used for each case while amplitude was -1.5 V to write state 00, -1.2 V to write 01, and -1.0 V to write state 10. The erase energy was calculated as the energy needed to switch the device from any of the states 00, 01 and 10 to state 11 using triangular pulse of $t_{\text{rise}}$ and $t_{\text{fall}}$ = 1 µs and amplitude of 1.5 V. The read energy was calculated using the read pulse of $t_{\text{rise}}$ and $t_{\text{fall}}$ = 1 µs and amplitude of 0.1 V. The write, erase and read energy of different states are summarized in Table 3.1. A high level of asymmetry (i.e., highest for state 11 and lowest for 00) was observed in read energy, as expected, due to difference in resistance values. Unlike read energy, significantly less skew was observed in write/erase energy across the states. Such intrinsic asymmetry in cell access behavior can have significant circuit/micro-architecture level implications [68]. It is worth noting that the device write energy/bit (~ pJ) is 4 orders of magnitude higher than the corresponding value for flash (~ $10^{-4}$ pJ) but comparable to Spin-Torque RAM (0.1 pJ) and other reported ReRAM devices (~ pJ) based on the International Technology Roadmap for Semiconductors (ITRS) reports [10].
Table 3.1: Summary of Read/Write/Erase energy in MLC ReRAM

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>92.3e-12</td>
<td>65.4e-12</td>
<td>3.5e-15</td>
</tr>
<tr>
<td>01</td>
<td>80.2e-12</td>
<td>71.2e-12</td>
<td>66.4e-15</td>
</tr>
<tr>
<td>10</td>
<td>54.8e-12</td>
<td>84.6e-12</td>
<td>149.3e-15</td>
</tr>
<tr>
<td>11</td>
<td>N/A</td>
<td>N/A</td>
<td>509.8e-15</td>
</tr>
</tbody>
</table>

Figure 3-6: CVS test of each resistance state was conducted for three times. (a) 0.1 V CVS. (b) -0.1 V CVS. State10 was identified as the most unstable state.

3.7. MLC reliability utilizing CVS

To understand the reliability of the different states, the CVS test was employed to accelerate device failure [71]. All resistance states were stressed with a constant voltage 0.1 V and -0.1 V at Room Temperature (19 °C), as shown in Figure 3-6 (a) and Figure 3-6 (b). The CVS test lasted for 3000 s unless a device failed. Multiple devices were tested to collect failure statistics. Interestingly, state 10 was identified as the most unstable one, which drifts towards the state 11 under positive stress voltage and state 00 under negative voltage. States 00, 01 and 11 showed good stability.

The origin of the observed endurance and instability of different states can be explained based on the prevalent theory of filament formation and its partial re-oxidation.
During forming, the Hf-O bonds were broken and O$^{2-}$ moved away from the filament region, generally towards TE, leading to the formation of the Hf-rich filament defining the LRS (i.e. state 11). During a reset operation, some of the oxygen ions located in the interstitial positions in the surrounding oxide are pushed back towards the BE, re-oxidize a portion of the Hf-rich filament next to the TE leading to the filament disruption. Under a set operation, the Hf-O bond in the re-oxidized section of the filament were broken again, and O$^{2-}$ moves away from the forming filament in the HfO$_x$. The length $d$ of the disrupted portion during reset can be controlled by controlling the reset-stop voltages for states 10, 01, and 00. If $d_{10}$, $d_{01}$, and $d_{00}$ represent the lengths of the disrupted filament from TE for states 10, 01, and 00 then $d_{10} < d_{01} < d_{00}$, as schematically shown in Figure 3-7 (a). To investigate this further, the I-V curves obtained for different HRS states (10, 01, and 00) were analyzed using different current transport models. The best fitting was obtained with Frenkel-Poole (F-P) model for all three HRS (10, 01, and 00), shown in Figure 3-7(b) which indicated that the current in HRS is dominated by F-P conduction of electrons through the oxide. According to F-P model, current density ($J$) vs. electric field (E) relation can be represented as [72]:

$$J = qn_0\mu E e^{\left(\frac{q}{kT}\left(\Phi_B - \frac{qE}{\mu}\right)\right)}$$

(4)

where, $q$ is the unit electronic charge, $n_0$ is the density of states in conduction band, $\mu$ is the electron mobility, $k$ is Boltzmann’s constant, $T$ is temperature, $\varepsilon_i$ is dielectric constant, and $\Phi_B$ is the trap energy level. With assumption that the F-P current in HRS is caused by the conduction between the filament and the BE via the oxide gap ($d$), equation (4) can be written in terms of applied voltage ($V$) and $d$ as:
or,

\[ \ln \left( \frac{I}{V} \right) = -\ln \left( \frac{d}{q n_0 \mu A} \right) - \frac{q \Phi_B}{kT} + \frac{q}{kT} \sqrt{\frac{q}{\pi \varepsilon_r d}} \sqrt{V} \]  

(6)

where \( A \) is the device area. From equation (6), one can observe that the slope (m) of \( \ln(I/V) vs. \sqrt{V} \) should be proportional to \( \sqrt{1/d} \). From fitting data in Figure 3-7(b) slopes \( m_{00} \) (corresponding to state 00), \( m_{01} \) (corresponding to state 01), and \( m_{10} \) (corresponding to state 10) were obtained as 1.74, 2.32, and 2.51 respectively. Clearly, \( m_{00} < m_{01} < m_{10} \) which indicates that \( d_{00} > d_{01} > d_{10} \). While this study provides a qualitative relation between \( d_{00}, d_{01}, \) and \( d_{10} \) corresponding to three HRS, a more detailed device model studies are needed to obtain the accurate values of \( d_{00}, d_{01}, \) and \( d_{10} \), which will be addressed in our future work.

Based on these studies, during CVS with 0.1 V, the drift of the state 10 towards 11 can be attributed to generation of oxygen vacancies in the d10 region due to a high voltage drop across this thin dielectric barrier. On the contrary, the drift of the state 10 towards state 01 under -0.1 V stress voltage could be due to a slow oxidation of partially disrupted filament over time due to a diffusion of the \( O^{2-} \) located in the filament vicinity in the surrounding HfO\(_x\). Since d10 is smallest, it experiences a higher electric field, which makes it the most unstable state. However, further studies are necessary to experimentally support this hypothesis. Nevertheless, from this study, it is clear that even though MLC using sequential reset voltages is possible in ReRAM devices, one has to carefully design the reset voltages to achieve multiple stable resistive states.
3.8. Summary

In this work, we demonstrate a multi-step forming technique, which provides a better control over the forming process against the overshoot phenomenon. Using this forming technique, the $I_{ov}$ was significantly reduced, which allowed repeatable ultra-low voltage/current set (0.5 V/50 µA) and reset (-1.25 V/50 µA) operations. By controlling
the reset stop voltages, multiple distinct resistive states were achieved. All states showed
excellent write (reset) endurance of $10^6$ cycles with the switching pulses of 100 $\mu$s. For
shorter pulses, the endurance was found to degrade. Excellent read endurance at 85°C
was observed for all states. A skew in the read energy of different states is observed
while write and erase energies demonstrated weaker dependency on states. The reliability
studies indicated that ‘intermediate’ HRS states (i.e. 10) obtained by applying
‘insufficient’ reset pulses were prone to failure over time due to additional
growing/reduction processes in the filament.
Chapter 4

Nanoelectronic Speech Cognitive Platform

A neuromorphic voice recognition system based on multi-level cell (MLC) ReRAM crossbar structure was studied for recognizing five vowels ‘a’, ‘e’, ‘i’, ‘o’ and ‘u’. Training process decides weight matrix distribution and encodes the trained patterns into nanoscale crossbar neural network. The weight matrix distribution or trained pattern depends on voice input frequencies components obtained by 64-channel band-pass filter bank and voltage integrator module. Recognition process implements winner-take-all (WTA) mechanism to compare the input with the stored training patterns and defines the best match. Both typical synapse weight value and weight variations in device programming were considered in the simulation. The average recognition rate was estimated to be 83% tested on 100 voice input samples.

4.1. Recognition process overview

In the proposed system, training is required before any kind of recognition operation. A unique sequential code is generated for each vowel during speech training. Then the obtained training patterns are orderly programmed to crossbar array. In recognition phase, five vowels are randomly fed into the system as shown in Figure 4-1. Similar to human cochlea, speech signal is firstly divided into 64 channels based on voice
frequencies by 64 band-pass filters ranging from 20 Hz ~ 2 KHz. Next, the filtered signal enters crossbar array module where input voice is compared with trained patterns of ‘a’, ‘e’, ‘i’, ‘o’ and ‘u’ five vowels. The best matched pattern is output as recognition result to the next level of the circuits. Finally, 100 voice samples are fed into the system and 83% average recognition rate is achieved.

Figure 4-1: Neuromorphic voice recognition system conceptual process flow.

Figure 4-2: Training process architecture. The voice signal is filtered by a group of band-pass filters. Then signal in each channel is integrated by voltage integrator. The integration results distribution decides crossbar array weight distribution.

4.2. Training

Simulation framework was constructed as shown in Figure 4-2. Prior to any recognition process, ReRAM crossbar array needs to be trained. The purpose of training is to obtain crossbar weight distribution. To do so, speech signal is filtered by 64 band-pass filters into 64 channels in filter bank with frequencies ranging from 20 Hz ~ 2 KHz.
The filtered and sampled signal in each channel is integrated by voltage integrator. The voltage integration decides crossbar matrix weight distribution. This training process is explained in detail in the following sections.

A voice was sampled at 44.1 KHz for 2 s. The sampled voice \( V(t) \) contains multiple frequency components \( V(f) \). To extract this speech frequency information, a band-pass filter bank was designed to mimic human cochlea filtering process. Voice frequencies from 20 Hz ~ 2 KHz was equally split into 64 frequency bands by 64 band-pass filters (BPFs). Then in each channel, the filtered and sampled signal \( V(t) \) was integrated by a voltage integrator during the whole voice input period. Different voice inputs have their own respective distinct frequency distribution over frequency spectrum, therefore, the voltage integration results \( \sum V(f) \) distribution differs after training. Voltage integration output for each channel was normalized to account for any difference in the integration circuits. The normalization was achieved by dividing the voltage output corresponding to each channel by maximum output voltage for each vowel, as shown in figure 4-3(a), figure 4-3(c), figure 4-3(e), figure 4-3(g) and figure 4-3(i).
Figure 4-3: Voltage integration distribution and crossbar weight distribution. (a), (c), (e), (g) and (i) shows the voltage integration results for five vowels, while (b), (d), (f), (h) and (j) shows the crossbar weight distribution for five vowels.

Figure 4-4: nanoelectronic crossbar array mimic neural network.

4.3. Crossbar Array

Learning is believed to be an intricate process in our brain which is achieved by the reconfiguration of synapse-neuron networks in response to an input signal. This reconfiguration is achieved by changing the plasticity of synapses connecting the two neurons [73]. For example, electrical pulses from pre-synaptic neuron pass through
synaptic connection and transfer the signal to post-synaptic neuron. The synaptic strength defines how much pre-synaptic signal is transferred to the next. In this way, multiple pre-synaptic neurons connect multiple post-synaptic neurons. By changing the connection strength, i.e. synaptic plasticity, human brain could learn and memorize information. Similarly, nanoelectronic crossbar array is employed to mimic the synapse-neuron network. As shown in Figure 4-4, an N × N crossbar array is organized as an N × N crosspoint matrix to connect N input ports to N output ports. In this fashion, crossbar array could simultaneously translate multiple inputs to multiple outputs ports to realize parallel neuromorphic computing. Each cross point of input and output ports has one MLC ReRAM device, and its weight (or conductance) value determines connection strength between its input and output port. Synaptic weight w is considered to be analog in nature. Multi-level cell can offer more resistance states than single-level cell (SLC) ReRAM and less programming variations than memristor device, thus we use MLC ReRAM as synapse rather than SLC ReRAM device or memristor device. If we employ a crossbar as a neural network, the MLC ReRAM cell can function as a biological synapse connecting pre-synaptic neuron (i.e. the crossbar matrix input port) and post-synaptic neuron (i.e. the crossbar matrix output port) [73]. The connection strength, defined by synapse weight, determines the amount of pre-synaptic neuron input that contributes to the post-synaptic neuron output, which plays an important role in the speech training and recognition process. It should be noted that the row of crossbar array corresponds to neuron axon, while the column of crossbar array corresponds to neuron dendrite.
Now, we will define crossbar weight distribution as follows. At the beginning, all synapses are reset to the highest resistance state, i.e. state 00. Next, five vowels, ‘a’, ‘e’, ‘i’, ‘o’ and ‘u’, was sequentially encoded into the 1st ~ 5th column of the crossbar array. The synapse crossing at the 1st ~ 64th row of each column respectively corresponds to the 1st ~ 64th frequency channel in filter bank. We evenly divided the normalized voltage integration range (0 ~ 1) into four regions, i.e. 0 ~ 0.25, 0.25 ~ 0.5, 0.5 ~ 0.75 and 0.75 ~ 1. Synapse weight followed the rule as defined by expression (7),

\[
\begin{align*}
    w_{ij}^\text{state00} &= w_{ij}, & \text{if } 0 \leq (\sum V_i) / \max(\sum V_i) < 0.25 \\
    w_{ij}^\text{state01} &= w_{ij}, & \text{if } 0.25 \leq (\sum V_i) / \max(\sum V_i) < 0.5 \\
    w_{ij}^\text{state10} &= w_{ij}, & \text{if } 0.5 \leq (\sum V_i) / \max(\sum V_i) < 0.75 \\
    w_{ij}^\text{state11} &= w_{ij}, & \text{if } 0.75 \leq (\sum V_i) / \max(\sum V_i) \leq 1
\end{align*}
\]

in which i and j represent the row number and column number of the crossbar array, respectively. \(\sum V_i\) is voltage integration of the ith channel. Following expression (7), the weight distribution for letter ‘a’, ‘e’, ‘i’, ‘o’ and ‘u’ was obtained by substituting the typical weight value \(w\) from device endurance test result from Figure 3-4 in Chapter 3, as shown in Table 4.1. Figure 4-3(b), (d), (f), (h) and (j) shows crossbar weight distribution.

<table>
<thead>
<tr>
<th>Resistance State</th>
<th>Typical Weight (w) ((\mu)S)</th>
<th>Variation Range ((\mu)S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘00’</td>
<td>0.97</td>
<td>0.7 ~ 1.38</td>
</tr>
<tr>
<td>‘01’</td>
<td>3.93</td>
<td>2.94 ~ 5.29</td>
</tr>
<tr>
<td>‘10’</td>
<td>21.6</td>
<td>20.1 ~ 24.3</td>
</tr>
<tr>
<td>‘11’</td>
<td>93.37</td>
<td>72 ~ 130</td>
</tr>
</tbody>
</table>

4.4. Weight Distribution

Table 4.1: Weight Value from Endurance Test. Data obtained from figure 3-4.
across 64 crossbar rows for five vowels. As shown in expression (7), synapse weight is actually proportional to the voltage integration results of input frequency channel. This is reasonable because the frequency band resulting higher voltage after integration dominates in representing the input voice. Therefore, synapse corresponding to this band is expected to contribute more towards the column (or dendrite) output current and dominate during the recognition in winner-take-all circuits, as discussed later in the recognition part. After weight distribution is accomplished, digital programmer (e.g. microcontroller) programs ReRAM device one by one either by VDD/3 or VDD/2 scheme to eliminate sneak path of the array [74].

![Recognition process architecture](image)

Figure 4-5: Recognition process architecture. Filtered and sampled signal fed into the crossbar array simultaneously and accumulated at the post-synaptic current summing neuron. Winner-take-all mechanism implements the highest input current as the best match.

4.5. Recognition

In the recognition phase, one of the five vowels (‘a’, ‘e’, ‘i’, ‘o’ and ‘u’) was input to the system again. Similar to the training phase, speech signal was separated into 64
frequency channels and simultaneously feeding into the crossbar array. A current
summing post-synaptic neuron at each end of crossbar column (i.e. dendrite) was used to
integrate dendrite current over time. The output current from integrating neuron I_a, I_e, I_i,
I_o and I_u was compared with each other in the WTA circuit [75] to decide which column
(i.e. dendrite) is the best match to the speech input. Winner-take-all mechanism
implements the highest input current as the best match and output this result to the next
level of the circuit. The recognition phase architecture is shown in Figure 4-5.

To test this approach, speech signal ‘a’, ‘e’, ‘i’, ‘o’ and ‘u’ was fed to the system
sequentially. Typical weight values, w, were taken for calculating synapse weight from
Table 4.1. Figure 4-6 (a), (c), (e), (g) and (i) shows the dendrite summing current I_a, I_e, I_i,
I_o and I_u evolution over time which can be given by equation (8) as:

\[ I_j = \sum_{i=1}^{i=64} w_{ij} V_i; \quad j = a, e, i, o, u \]  

From Figure 4-6 one can clearly observe that the system can successfully differentiate
voice input by using the proposed framework. We also considered weight value
variations in device programming. In this case, synapse weight was chosen as a random
number within variation range shown in Table 4.1. Figure 4-6 (b), (d), (f), (h) and (j)
shows recognition evolution results. The time window of speech sampling for each vowel
was 2 s. During this 2 s, one can speak at any time, and current at each dendrite was
integrated from 0A when speech was first input to the system. This integration process
lasts for several hundred milliseconds until speech was over. Since five dendrites (j = a, e,
i, o, u) had their own unique synapse weight distribution, integration current were
different for the same voice input. From equation (7), it is expected that the dendrite with
matching vowel results in maximum integration current because weight was assigned proportional to the voltage integration results corresponding to the channel frequencies for each dendrite during training. One interesting study lies in understanding the robustness of this scheme with noise or uncertainty in the input signal during training and recognition, which will be addressed in our future work.

4.6. Recognition Rate

100 voice samples were input the crossbar based neuromorphic voice recognition system in which 20 samples for each vowel were simulated with typical synapse weight value shown in Table 4.1. Figure 4-7 shows the simulation results for the recognition rate. The average recognition rate is around 83%. The highest recognition rate 90% is for vowel ‘u’ while the lowest recognition rate 75% for vowel ‘e’. In the proposed work, the system was provided only one time training for each voice sample which could be insufficient and result in the false recognition. The system can be made more immune to the recognition mistakes by multiple trainings. However, 83% recognition rate is still acceptable for most applications considering one time training is time-efficient.
4.7. Summary

In this work, a voice training and recognition scheme is proposed which can be implemented using ReRAM devices in crossbar network. The validity of the proposed scheme was tested via simulations where experimental device data for MLC ReRAM was used. In particular, we demonstrated that system can successfully recognize five vowels. This simulation work paves groundwork for future voice processing circuit designs based on ReRAM crossbar arrays.
Figure 4-7: The simulated recognition rate for recognizing ‘a’, ‘e’, ‘i’, ‘o’ and ‘u’ based on 100 voice samples (20 samples for each vowel).
Chapter 5

Ru/ZnO/TiN ReRAM Devices

In Chapter 3, HfO$_x$ based ReRAM device presents excellent characteristics, however, it is not suitable for crossbar array integration due to linear I-V in LRS unless 1T1R or 1D1R cell was employed, otherwise sneak path could lead to mistaken read, as explained later. In this chapter, Ru/ZnO/TiN resistive memory device based on switchable diodes effect is studied. Devices were fabricated in crossbar array with simple MIM structure. Forming-free and current compliance-free characteristics were observed in resistive switching. Meanwhile, this device shows great potential for crossbar integration as unit cell owing to device self-rectifying effect to suppress sneaking current. The switching mechanism was also studied based on switchable diode model. Eleven distinct resistance states were obtained by controlling set/reset stop voltages, and all eleven states showed excellent endurance.

5.1. Fabrication and testing setup

ReRAM devices consisting of a Ru/ZnO/TiN/W stacks were fabricated in crossbar structure on a p-type Si substrate by RF magnetron reactive sputtering, as shown in Figure 5-1(b). 1000 A SiO2 was first grown on Si wafer followed by depositing 20 nm Ru as bottom electrode (BE). 60 nm ZnO was then fabricated using Zn target at
temperature 100 °C in a gas mixture of Argon and Oxygen with gas ratio of 9.6:2.4. Top electrode (TE) TiN was then deposited using Ti target at temperature 300 °C in a gas mixture of Ar and N₂ with gas ratio of 6:6. Finally, devices were capped with 50 nm W as hard contact layer. Different device sizes of 20 µm × 20 µm (400 µm²), 30 µm × 30 µm (900 µm²), 50 µm × 50 µm (2500 µm²), 70 µm × 70 µm (4900 µm²) and 90 µm × 90 µm (8100 µm²) were fabricated. Crossbar size varies from 1 × 1 to 16 × 16. Electrical characterization was performed using LakeShore cryogenic probe station (under vacuum pressure of 1×10⁻⁴ Torr) with Keithley 4200 SCS and 4225 Ultra-fast I-V Pulse Modulation Unit (PMU). For all testing, voltage was biased on TE with BE grounded. All the test was performed on 30 µm × 30 µm devices in 1 × 1 crossbar array unless otherwise mentioned.

5.2. IRS Initial I-V and scalability

Figure 5-1(a) shows I-V characteristics of initial resistance state (IRS) devices. In this test, current was measured by sweeping DC voltage from -2 V to 2 V with voltage step 0.2 V. Multiple devices with different cell area ranging from 400 µm² (i.e. 20 µm × 20 µm) to 8100 µm² (i.e. 90 µm × 90 µm) were tested. We can clearly see device conductivity is scalable with device area, i.e. highest for 8100 µm² devices and lowest for 400 µm² devices. A slight asymmetry in I-V between positive and negative bias can be attributed to the difference in the work function of TE TiN vs. BE Ru.
Figure 5-1: (a) Devices I-V characteristics in initial resistance state (IRS). 20 µm × 20 µm, 30 µm × 30 µm, 50 µm × 50 µm, 70 µm × 70 µm and 90 µm × 90 µm were characterized. Multiple devices were tested and show repeatable initial I-V for different size, but we only show one device I-V of each size here for clarity. (b) Schematic representation of fabricated Ru/ZnO/TiN/W devices in crossbar structure. Device area is varying from 400 µm² to 8100 µm² with crossbar size ranging from 1 × 1 to 16 × 16. Bias was applied to top electrode, whereas the bottom electrode was grounded during electrical characterization. (c) Fabricated crossbar wafer
5.3. **Switchable diode based resistive switching**

Figure 5-2(a) shows bipolar RS I-V characteristics. DC voltage was first swept from 0 V to 7 V as shown by curve #1 in Figure 5-2(a). Current increased abruptly after 5.5 V, indicating the change in resistance state from HRS to LRS. Device switched to LRS which can be seen from the return loop as indicated by #2. To reset the device back to HRS, DC voltage was swept from 0 V to -7 V and then back to 0 V as shown by #3 and #4. Resistance transition happened at -5.5 V till to -7 V after which the device was switched back to HRS as indicated by #4. Thereafter, device can be switched between HRS and LRS when following the order 1 → 2 → 3 → 4. Electrical current at 7 V (i.e. maximum current in set) was ~100 µA and at -7 V (i.e. maximum current in reset) was ~10 µA. No compliance current/forming is required in device switching, thus it was forming-free and current compliance-free. Here, it should be noted that HRS and LRS are both defined by resistance read at 2 V. Figure 5-2(b) shows similar resistive switching (RS) behavior of the ReRAM devices with different cell area varying from 400 µm² to 8100 µm². In this test, five different size devices were taken and swept a same series of DC voltage 0V → 7V → 0 → -7V → 0V. All devices presented bi-stable switching behaviors. The extracted max current during set/reset and resistance ratio HRS/LRS (read@2V) across five size of devices are shown in Figure 5-3. It clearly indicates that the extracted parameters can be scalable as cell area scales down. In addition, ~1 nA max current in set, ~2 pA max current in reset, and ~10³ resistance ratio can be extrapolated for 1 µm² (1 µm × 1 µm) device.
Figure 5-2: (a) Switching I-V characteristics of Ru/ZnO/TiN ReRAM devices of area of 30 µm × 30 µm with ultra-low self-compliance 100 µA. The self-rectifying effect in both LRS and HRS were also observed. (b) Typical resistive switching I-V data across all device areas. DC voltage operation order is 0 → 7V → 0V → -7V → 0V. Similar switching I-V characteristics were observed for all devices. Devices show self-rectifying and forming free effects.

Figure 5-3: Maximum current during set and reset and resistance ratio extrapolation for different device areas. Smaller device shows lower operation current and higher resistance ratio HRS/LRS.
5.4. **Switchable diode model**

Figure 5-4(a) shows a pronounced current rectifying effect after device set process at LRS. Current in the positive direction (i.e. forward current, $I_{\text{forward}}$) was $\sim 220 \times$ larger than negative part (i.e. reverse current, $I_{\text{reverse}}$) when read voltage was given at $\pm 2$ V. Also, as shown in Figure 5-4(b), rectifying effect was observed for HRS. HRS $I_{\text{forward}}$ was $\sim 51$ times larger than $I_{\text{reverse}}$. Self-rectifying effect existed in both LRS and HRS can be qualitatively elucidated by switchable diode model [76-78]. ZnO can be considered as n-type semiconductor [79]. In Figure 5-5(a), two back-to-back Schottky diode at top and bottom result in slight asymmetry I-V characteristics in devices across all cell areas due to different work function of TE TiN and BE Ru, as shown in Figure 5-1(a). In IRS, oxygen vacancies ($V_o$) created by fabrication are randomly distributed in the ZnO [80-82]. When positive bias is applied on TE, more $V_o$ are generated under voltage bias, and both original and generated $V_o$ move towards cathode [83], as shown in Figure 5-5(b). Thus, Schottky barrier height at cathode decreases due to heavily doped $V_o$. Since Schottky barrier at cathode reduces whereas the Schottky contact at anode still remain high, I-V characteristics, as shown in Figure 5-4(a), exhibits a typical diode I-V characteristics due to Schottky effect at top interface. Similarly, top Schottky contact reduces with sufficiently high negative voltage bias on top, as shown in Figure 5-5(c). Consequently, only Schottky contact at bottom interface participates in device current transport, which is also in good agreement with I-V data shown in Figure 5-4(b). Moreover, device C-V analysis was conducted to support our proposed model, as shown in Figure 5-6. Initially, IRS capacitance is almost constant for both forward and reverse bias due to two back-to-back diodes connected. For both LRS and HRS, we can clearly
see capacitance in forward bias region, i.e. junction capacitance $C_j$, is much lower than that in the reverse bias region, i.e. diffusion capacitance $C_d$, indicating only one-sided diode comes into play after device is switched.

Figure 5-4: I-V data of (a) LRS and (b) HRS by using DC voltage sweeping from -2V to 2V. Both HRS and LRS devices show current rectifying effect.

Figure 5-5: Schematic illustration of switching mechanism based on switchable diodes for our devices. (a) IRS device. (b) LRS after set. Positive bias is applied on top during set. (c) HRS after reset. Negative bias is applied on top during reset.
5.5. **IRS, LRS, HRS Schottky Emission fitting**

To further investigate the current transport model of IRS, LRS and HRS, the obtained I-V data, i.e. IRS I-V shown in Figure 5-1(a), LRS I-V shown in Figure 5-4(a) and HRS I-V shown in Figure 5-4(b), were analyzed and we found it was well fitted with Schottky emission (SE) model, as shown in Figure 5-7(a), Figure 5-7(c) and Figure 5-7(e) of SE fitting plot. The SE conduction of electrons can be expressed as equation (1) [83].

$$\ln I = \ln(A^*T^2) - \frac{q(\phi_B - \sqrt{qV/4\pi\epsilon})}{kT}$$  \hspace{1cm} (9)

where I is the electrical current, V is the bias voltage, q is the electric charge, $\phi_B$ is the barrier height at interface, $\epsilon$ is the dielectric constant, k is Boltzmann’s constant, T is the temperature, and $A^*$ is the effective Richardson constant. For IRS, current under negative bias lower than that under positive one signifies that the Schottky contact at top have a
higher barrier, as shown in Figure 5-7(b). With the same analysis, higher Schottky barrier can be concluded at top for LRS and at bottom for HRS from Figure 5-7(d) and Figure 5-7(f), respectively. Therefore, the SE I-V fitting results for IRS, LRS and HRS are well matched to the proposed switchable diodes model showing that Schottky barrier height at both top and bottom can be tunable by doping density of $V_o$, leading to bi-stable RS behavior.
Figure 5-7: (a), (c) and (e) show Schottky emission (SE) fitting for IRS, LRS and HRS. (b), (d) and (f) show band diagram for IRS, LRS and HRS devices.

Table 5.1: Schottky Barrier Height at Top and Bottom for IRS, LRS and HRS.

<table>
<thead>
<tr>
<th>Resistance State</th>
<th>Top Schottky Barrier (eV)</th>
<th>Bottom Schottky Barrier (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRS</td>
<td>0.40</td>
<td>0.37</td>
</tr>
<tr>
<td>LRS</td>
<td>0.39</td>
<td>0.38</td>
</tr>
<tr>
<td>HRS</td>
<td>0.34</td>
<td>0.40</td>
</tr>
</tbody>
</table>

5.6. Endurance test and device degradation

To evaluate switching reliability of ReRAM devices, device endurance were studied. DC endurance was conducted using +7 V/-7 V DC voltage to switch device between LRS and HRS for 100 cycles. Device resistance read at 2 V/-2 V, respectively. The results are shown in Figure 5-8(a) and Figure 5-8(b). LRS (read@-2V) gradually decreases. The 1st, 50th and 100th cycle I-V data are also shown in Figure 5-2(a). One can observe that I_{reverse} of LRS become more conductive but I_{reverse} of HRS still remains same during 100 switching cycles. DC endurance indicates that Schottky barrier at top contact TiN/ZnO reduces as switching cycles increase, whereas the barrier height at bottom Ru/ZnO shows much less change. This difference can be illustrated as follows.
Under high electric field, more V$_o$ can be generated due to reaction of O$_o$ → V$_o^{++}$ + 2e + $\frac{1}{2}$O$_2$ [84]. At top interface of TiN/ZnO, since Gibb's free energy $\Delta G$ of Ti (-883.29 kJ/mol) is much lower than that of Zn (-320.51 kJ/mol), O$_2$ is more reactive with Ti [85], whereas at bottom interface of Ru/ZnO, $\Delta G$ of Ru (-152.2 kJ/mol) is much higher than that of Zn (-320.51 kJ/mol), Ru hardly reacting with O$_2$ than Zn [85]. Therefore, V$_o$ generation rate must be higher at top TiN/ZnO interface than that of bottom one. After certain cycles, more and more newly generated V$_o$ at top cannot fully drift to cathode, thus top Schottky barrier reduces, thus $I_{\text{reverse}}$ of LRS increases. Furthermore, the degradation caused by the residual V$_o$ near TiN leads to an increase in $I_{\text{reverse}}$ of LRS resulting in reduced $I_{\text{forward}}/I_{\text{reverse}}$ of LRS. As shown in Figure 5-9(a), LRS $I_{\text{forward}}/I_{\text{reverse}}$ is ~ 40 at 50% cumulative probability during 100 DC switching cycles. When comparing with the ratio of ~220 at initial cycle, the device asymmetry level is reduced by 4 times. On the contrary, the HRS $I_{\text{forward}}/I_{\text{reverse}}$ (~50 at 50%) is tightly distributed shown in Figure 5-9(b). Pulse Based endurance was also conducted as shown in Figure 5-10(a) and Figure 5-10(b). Triangular pulses were used with rise and fall time of 50 ms and pulse amplitude 7.2 V/-7.2 V. The endurance was up to $10^3$ cycles and the same trend of degradation as DC endurance was observed.
Figure 5-8: DC endurance data using DC voltage with amplitude of 7V/-7V. (a) Resistance read at 2V. (b) Resistance read at -2V.

Figure 5-9: Forward and reverse current distribution obtained from 100 cycles DC endurance for (a) LRS and (b) HRS.
Figure 5-10: Pulse endurance under triangular pulse with rise and fall time 50 ms and amplitude of 7.2 V/-7.2 V. (a) Resistance read at 2 V. (b) Resistance read at -2 V.
Figure 5-11: Multiple resistance states can be achieved by using different set voltages or using different reset voltages. (a) Set voltages of 5.6V, 6V, 6.4V, 6.8V and 7V are used to set device to different LRS. (b) Reset voltages of -6V, -6.4V, -6.8V and -7V are used to reset device to different HRS.

5.7. MLC Switching

Next, we focused on achieving multiple resistance states using these devices. Multiple resistance states can be achieved by setting the device with different set voltages from the same HRS, as shown in Figure 5-11(a), or by resetting the devices with different reset voltages from the same LRS, as shown in Figure 5-11(b). In Figure 5-11(a), five LRS states can be achieved using set voltages of 5.6 V, 6 V, 6.4 V, 6.8 V and 7 V, respectively, while keeping the same reset voltage -7 V for all states. One can observe that the I\(_{\text{forward}}\) of LRS becomes more conductive as set voltage is increased from 5.6 V to 7 V. Similarly in Figure 5-11(b), four HRS states can be achieved by using the same set voltage 7 V and resetting the device with different reset voltages of -6 V, -6.4 V, -6.8 V and -7 V, respectively.
Figure 5-12: Read endurance for obtained eleven resistance states from Fig. 13. (a) Read endurance for resistance states obtained by setting the device with different set voltages and read at 2 V. (b) Read endurance for resistance states obtained by resetting the device with different reset voltages and read at -2 V.

5.8. MLC read endurance

To estimate read reliability of eleven resistance states obtained in MLC switching, read endurance was measured using read pulses of ±2 V/60 ns at 10 MHz. To perform this test, at first, eleven devices were switched to eleven different resistance states by DC sweeps. Then resistance corresponding to each state was measured by applying a series of read voltage pulse of 2 V/60 ns at 10 MHz, as shown in Figure 5-12(a), and -2 V/60 ns at 10 MHz, as shown in Figure 5-12(b). Figure 5-12(a) and figure 5-12(b) shows excellent read endurance up to $10^9$ cycles for all states under these test conditions without any obvious signs of failure.

5.9. Sneak current path in crossbar array

The device with such current-rectifying effect can efficiently suppress crosstalk effect in crossbar array. For instance, the worst case of sneak current in $2 \times 2$ crossbar array is shown in Figure 5-13(a), which the device of interest, selected by word line (WL)
and bit line (BL), is in off-state (high resistance) whereas all the other cells are in on-state (low resistance). When $V_{\text{read}}$ is applied on WL with BL grounded to read targeted cell, sneak current as indicated by red path influences read result of target cell, and at worst cast, it can lead to misread [86-88]. However, one can observe that in each sneak path, there must be at least one cell in reversed bias (cell #3) [89-91]. As shown in Figure 5-11(b), if we employ reset/set stop voltages -6 V/7 V, the switching current at negative voltage part can be negligible when comparing with positive part. Ideally, device with such current-rectifying effect can shut off current flowing under reversed bias, and sneak current can thus be eliminated as shown in Figure 5-13(b). In this regard, this device shows excellent potential for suppressing sneak current due to device’s self-rectifying characteristics. By eliminating external rectifying diode as an access device in the 1D1R structure, device with self-rectifying effect can further simplify fabrication complexity and reduce unit cost [92-95].
Figure 5-13: (a) Sneak path in 1R crossbar structure in worst case and (b) device with self-rectifying effect can suppress sneak current to improve integration density.

Figure 5-14: Simulation of the read-out margin in $N \times N$ crossbar array for the worst case based on 1R setup.
5.10. Read out margin in $N \times N$ crossbar

Considering there are $(M - 1) \times (N - 1)$ sneak paths at worst case for $M \times N$ crossbar array, Figure 5-14 shows the read-out margin in $N \times N$ crossbar array for the worst case based on 1R setup [96-98]. In the simulation, read voltage 2 V was biased on the selected WL with selected BL grounded, while all the other unselected WLs and BLs are unconnected. As word line $N$ increases, HRS approaches LRS due to ever-increasing sneak current which influences the read results [99-101]. At $N = 5$ (i.e. $5 \times 5$ crossbar), HRS/LRS ratio shrinks from 60 times to 8 times. This indicates that at any given resistance state, the non-linearity in current should be further improved.

5.11. Pulse based set/reset

Figure 15(a) and figure 15(b) shows device resistance can be gradually modulated by pulse width. Trapezoidal pulses with pulse amplitude 7.2 V/-7.2 V and pulse width 30 ms, 50 ms and 90 ms were respectively biased on six fresh devices. Figure 5-15(a) and figure 5-15(b) clearly show that one with longer pulse width would get lower resistance.

Figure 5-15: Lower resistance can be obtained by longer pulse width. Device is set (a) or reset (b) by pulse with pulse width 30 ms, 50 ms and 90 ms.
5.12. Summary

In this chapter, we presented a simple structured MIM switchable diodes as resistive memory devices. Device can be operating at low switching voltages and current with excellent device switching endurance. Schottky emission was found to be the electron transportation model. By tuning Schottky barrier height using different set/reset stop voltages, multi-level cell (MLC) switching can be achieved. Read endurance of each state was up $10^9$ without any obvious sign of failure. Meanwhile, sneak current in crossbar array suppressed by self-rectifying effect eliminates employing external diode as an access device in 1D1R structure.
Chapter 6

Conclusions and Future Work

6.1. Conclusions

ReRAM has been recognized as one promising candidate to succeed flash for the next generation of NVM memory. However, as described in Chapter 2, the main concerns in ReRAM research are concluded as follows. Current overshoot caused by abrupt resistance change during device forming/set process in 1T1R testing setup is considered to be one of the main reasons that results in device variability and failure. It hinders ReRAM development. Second, cross talk effect in crossbar/3D crossbar structure impedes higher crossbar array integration density. Meanwhile, there is still lack of comprehensive studies on electrical characterization of access energies, reliability, and stability as well as associated trade-offs for MLC ReRAM. The studies performed in this research explore solutions to solve these problems.

Comprehensive MLC characteristics were studied for TiN/HfO$_x$/Zr memristor devices. A novel multi-step forming technique was implemented to efficiently suppress current overshoot that allowed device switching at a very low set/reset voltage and current without the need for ‘on-chip’ current limiters. Four distinct resistance states were achieved by controlling reset stop voltage and showed excellent endurance.
Write/read/erase energy for different states was also calculated. Among four MLC states, it was found that the lowest resistance value of three distinct High-resistance states (HRS) was prone to fail over time under constant voltage stress (CVS).

Although there are excellent MLC switching characteristics of such HfO$_x$ based memory devices, as linear I-V characteristics in LRS and abrupt resistance change at forming/set point, ‘on chip’ current limiters, i.e. transistor in 1T1R, or additional diodes to limit sneak current is necessary for crossbar integration. The extra transistor or diode not only increases the entire system’s power consumption, but also needs to be specifically engineered for distinct ReRAM devices. To avoid using such current limiters, a more ideal approach was proposed utilizing device self-current rectifying effect in Chapter 5. Ru/ZnO/TiN memory devices were fabricated in crossbar structure. Due to switchable diode effect behind resistive switching, forming-free and current compliance-free could be obtained; thus, current overshoot issues resulting from an external current limiter could be effectively solved. Essentially, self-rectifying effect of the device can efficiently block sneak current, illustrating great potential for achieving high integration density. The switching mechanism was also studied based on a switchable diode model. Meanwhile, MLC switching was also conducted for this device. By tuning Schottky barrier height, eleven distinct resistance states were obtained using different set/reset stop voltages, and all eleven states showed excellent endurance.

To demonstrate application of memristor devices, a voice recognition system was constructed and demonstrated in Chapter 4 by using multi-level cell (MLC) memristor crossbar structure for recognizing five vowels ‘a,’ ‘e,’ ‘i,’ ‘o,’ and ‘u.’ The training process decides weight matrix distribution and encodes trained patterns into a nanoscale
crossbar neural network. The weight matrix distribution or trained pattern depends on voice input frequencies components obtained by 64-channel band-pass filter bank and voltage integrator module. Recognition process implements a winner-take-all (WTA) mechanism to compare input with stored training patterns and defines the best match. Both typical synapse weight value and weight variations in device programming were considered in the simulation. The average recognition rate was estimated to be 83% tested on 100 voice input samples.

6.2. Future work

Since tremendous exciting characteristics were observed in Ru/ZnO/TiN ReRAM devices, such as forming-free, current compliance-free, self-current rectifying, etc., the ZnO based ReRAM device should be further investigated. Material analysis should be conducted to further research switching mechanisms of ZnO based ReRAM devices. The physical mechanism behind device resistive switching was critical for understanding switching behavior in order to propose a better stack. Despite relatively low operation voltage (±7 V) as compared to other published switchable diode research, ±7 V is still high and unacceptable for practical application of memory devices. To decrease switching voltage, an efficient way is to reduce thickness of ZnO (currently it is 60 nm). Several samples with different thickness need to be fabricated; doping metal, e.g. Mg, Zn, etc., into ZnO active layer is another good way to achieve this goal. Moreover, since electrical characterization of ZnO crossbar devices were performed at room temperature, a more comprehensive device characterization and performance analysis at different temperatures should be implemented. Improved switching characteristics can be achieved by further scaling down the device area, so smaller devices should also be studied.
Additionally, CMOS compatible materials, such as Ti, etc., should be used as electrode to replace Ru in the future.

For crossbar level research, in Chapter 4, we demonstrated speech recognition algorithm using HfO$_2$ based MLC ReRAM devices shown in Chapter 3. In order to relate the proposed speech recognition algorithm with fabricated ZnO crossbar devices, ZnO crossbar device and its multi-level resistance states need to be integrated into the proposed speech recognition system and simulation should be repeated. Furthermore, complete hardware level circuits should be also designed to program and read the crossbar array, which is necessary and critical for ReRAM crossbar hardware system realization in the future.
References


[66] Xu, Cong, Dimin Niu, Naveen Muralimanohar, Norman P. Jouppi, and Yuan Xie. "Understanding the trade-offs in multi-level cell reram memory design." In Design


Appendix A

Image Recognition using Memristor

Memristors have been identified as one of the most promising candidates for application as analog memory components, used as synapse, in the modern neuromorphic computing architecture. Integrating memristors into the crossbar structure to perform image learning and recognition, imitating the human virtual cortex, has been researched and specific system level implementation is proposed in this chapter. We will show a complete scheme from image learning to image recognition. In particular, we have utilized a novel scheme of converting the image, incident on the APS module, into spike-patterns. In this APS module, grayscale image was encoded into sensing current that was fed into leaky integrate-and-fire (LIF) input neurons. These input neurons were connected to all output neurons via a memristor based crossbar array, which formed a neural network. The neural network, using fully asynchronous feedback-forward architecture, can store different images into different post-synaptic neuron dendrites according to Spike-Timing Dependent Plasticity (STDP) learning rule. The whole learning process was fully unsupervised. At last, in the image recognition process, we simply implemented a pulse counter circuit to count post-synaptic output spikes to locate matched output neuron.
A.1. Neurons

Nerve cells are called neurons. There are about one billion neurons in the human brain. A typical neuron cell is shown in Figure A-1. In simple terms, a neuron is a cell specialized to generate and conduct electrical impulses and to carry information from one part to another [102]. The electrical impulses is also called action potential (or AP). A neuron cell consists of the following parts: soma, dendrite, axon and axon terminals. Soma is the cell body as it contains cell nucleus. For neuron cell, it collects signals from other neurons through a structure called dendrite. Dendrite have many branches so that one neuron cell can communicate with many other neurons at the same time. Neuron sends out electrical/chemical signal by a long structure called axon. Similar to dendrite, axon branches at the axon terminal to communicate with multiple target neuron cells. One neuron cell only has one axon. At the end of each axon branch, a structure called synapse connects two neurons. Synaptic plasticity, also called synaptic strength or synaptic weight, defines how much signal can be passed to the next neuron and it can be modified by neuron internal mechanism, such as spike timing dependent plasticity (STDP). Learning occurs by modifying synaptic strength under such mechanisms so that the influence of one neuron on another changes.

A.2. STDP Learning

STDP is a type of learning mechanism observed in neural systems of human brain in which synaptic plasticity is changed based on relative timing of pre-synaptic and post-synaptic spikes [103]. In 1998, the experimental evidence of STDP was first reported by Bi and Poo in hippocampal cultures, pointing out synaptic plasticity or synaptic weight depends on the relatively timing of pre-synaptic and post-synaptic neuron spikes [104].
The pre-synaptic neuron sends pre-synaptic spike $V_{\text{pre}}(t)$ to post-synaptic neuron through axon to the neuron synapse. The cumulative effect of pre-synaptic spikes at the post-synaptic neuron dendrite eventually triggers a post-synaptic neuron spike $V_{\text{post}}(t)$ when threshold is reached. This triggered post-synaptic spike $V_{\text{post}}(t)$ feedback to the synapse. The synaptic weight can be modified by relative timing of $V_{\text{pre}}(t)$ and $V_{\text{post}}(t)$ so that learning occurs. Figure A-2 shows the original data presented in the Bi and Poo’s paper. If the pre-synaptic spike precedes post-synaptic spike ($\Delta T > 0$), the synaptic weight increases and the synapse gets stronger (potentiation). On the contrary, if a post-synaptic spike precedes pre-synaptic spike ($\Delta T < 0$), the synaptic weight decreases, which means the synapse get weaker (depression).
In human brain, a dendrite is connected to several pre-synaptic axons. Similarly, an axon also connects multiple post-synaptic dendrites. In this way, a neural network is formed between neurons through synapses. Learning happens when synaptic plasticity is modified based on STDP learning rule. Different pre-synaptic action potentials (AP) are summed at the dendrite. The contribution of each pre-synaptic AP is determined by the strength of the corresponding synapse. Once there is a post-synaptic depolarization at the dendrite, an AP is generated at the dendrite which then travels through the post-synaptic neuron. The strength of the synapse changes according to the relative timing ($\Delta T$) of the incoming pre-synaptic AP and the outgoing post-synaptic AP. If the pre-synaptic spike precedes the post-synaptic spike, the weight of the corresponding synapse is increased, and it contributes more towards the post-synaptic neurons firing; otherwise, if the post-synaptic spike precedes the pre-synaptic spike, the strength of the corresponding synapse is decreased, and it contributes less towards the post-synaptic neurons firing. Synaptic plasticity is greater when the relative pre-synaptic and post-synaptic spike timing is
smaller. This STDP learning process takes some time to finish and knowledge is thus memorized after training.

A.3. Crossbar Array

Nanoelectronic crossbar array is employed to mimic synapse-neuron network. As shown in Figure A-3, \( N \times N \) crossbar array is organized as an \( N \times N \) crosspoint matrix to connect \( N \) input ports to \( N \) output ports. In this fashion, crossbar array could simultaneously translate multiple inputs to multiple outputs ports to realize parallel neuromorphic computing. Each cross point of input and output ports has one memristor device as synapse, and its weight (or conductance) value determines connection strength between its input and output port. Synaptic weight \( w \) is considered to be analog in nature. Conductance change in memristor device is in analog as well, thus we use memristor as synapse. If we employ a crossbar as a neural network, the memristor cell can function as a biological synapse connecting pre-synaptic neuron (i.e. the crossbar matrix input port) and post-synaptic neuron (i.e. the crossbar matrix output port). The connection strength, defined by synapse weight, determines how much of pre-synaptic neuron input that contributes to the post-synaptic neuron output, which plays an important role in the neuromorphic computing system which will be discussed in this chapter. It should be noted that the row of crossbar array corresponds to neuron axon, while the column of crossbar array corresponds to neuron dendrite.
Figure A-3: Nanoelectronic crossbar array mimic neural network.

Figure A-4: Memristor moving wall model. In this model, the wall position $m$, which is dependent on bias voltage, separates two different resistance regions in series.

A.4. STDP Implementation Using Memristive Synapse

We used two-terminal voltage driven memristor as synapses to realize this STDP learning rule. Here, we will consider one particular subset of memristor model developed in [103]: moving wall model, in Figure A-4. The wall position $m$, which is bound to $[0, L]$, separates two different resistance regions in series, and moves depending on the bias voltage,

$$\Delta m (\Delta T) = \int f(v_{MR}(t, \Delta T)) \, dt = \xi(\Delta T)$$

(10)
where the change of wall position $\Delta m$ or $\xi$ is the result of function $f()$ integration over time; in memristive nanoscale devices, function $f()$ describe ionic drift under electric field; $\Delta T$ is the relative timing of AP between pre-synaptic spike and post-synaptic spike, i.e. $\Delta T = t_{\text{pos}} - t_{\text{pre}}$; $v_{MR}$ is the voltage drop across the memristor device. As observed in Figure A-3, the pre-synaptic neuron is connected to memristor top electrode (TE), and the post-synaptic neuron is connected to the bottom electrode (BE), therefore, $v_{MR} = v_{\text{pre}} - v_{\text{post}}$, in which $v_{\text{pre}}$ and $v_{\text{post}}$ are the AP of the pre-synaptic neuron and AP of the post-synaptic neuron. Function $f()$ has the following form,

$$
\begin{align*}
    f(v_{MR}) &= I_0 \cdot \text{sign}(v_{MR}) \cdot [e^{v_{MR}/v_0} - e^{v_{th}/v_0}], \quad \text{if } |v_{MR}| > v_{th} \\
    f(v_{MR}) &= 0, \quad \text{otherwise}
\end{align*}
$$

(11)

where $I_0$ and $v_0$ are constant parameters. Now, we have achieved the wall position update function $\Delta m$ (or $\xi$). However, as we will see in the next section, when memristor is used in the neuron circuits as synapse, synaptic strength (or weight) is actually the conductance of the memristor. According to the model shown in Figure A-3, the instantaneous resistance of the memristor is linear with $\Delta m$, consequently, $\Delta R (m) = k \cdot \Delta m$. Therefore, we finally get

$$
\Delta w = \Delta \left( \frac{1}{R} \right) = -\frac{\Delta R}{R^2} = -w^2 k \xi (\Delta T)
$$

(12)

where $w$ is the conductance level of the memristive device. It should be noted that the change in conductance $\Delta w$ is quadratically proportional to device conductance $w$, which is a positive feedback for synapse weights separation to improve image learning rate.

Here, we first simulate the AP as shown in Figure A-5(a). The resulting STDP memristance update function is shown in Figure A-5(b) and the resulting memductance
update function is shown in Figure A-5 (c). This relation is obtained by substituting \( f() \) from equation (2) into equation (1) and plotting the normalized \( \xi \) vs. \( \Delta T \), in which \( v_{th} = 1 \) V, \( v_0 = 1/7 \) V, \( I_0 = 1 \) mA, \( k=1 = 1.8 \) µA. But the spike shape shown in Figure A-5 (a) is hard to obtain and the related circuits is relatively complex. Therefore, in our simulation, we employed the rectangular-triangular neuron spike as shown in Figure A-5(d) and its resulting memristor update function is shown in Figure A-5(e) and Figure A-5 (f).
Figure A-5: (a) Ideal neuron spike shape used for simulation, (b) resulting STDP memristance update function and (c) resulting STDP memductance update function. Similarly, (c), (d), (e) corresponds to adapted neuron spike, its resulting STDP memristance and memductance update function, respectively.

A.5. STDP Based Learning Simulation

A.5.1 APS as Input Image Sensor

This section discusses the spike-coding scheme for grayscale images using an array of APS and LIF neurons. A grayscale image is actually a digital image in which each pixel only carries the intensity information which have 256 discrete intensity levels (between 0 and 255). When the grayscale image is illuminated onto active pixel sensor (APS) array, APS module serves as a photoelectric conversion module converting light intensity to the APS circuit’s output current. The 3T-APS schematic is shown in Figure A-6(a). Initially, before receiving light illumination, the photodiode (PD) voltage is reset by the reset transistor to $V_{dd}$ as the reference voltage $V_1$ which is read out by the readout transistor. By turning off the reset transistor, the capacitance of PD discharges with a time constant proportional to the light intensity during integration time $T$. After integration time, signal voltage is again read out by the readout transistor, referenced as $V_2$. Then, the APS output voltage $\Delta V=V_2-V_1$ can be measured using correlated double sampling...
(CDS) circuits [105]. In order to input this light intensity dependent signal, ΔV, to the neural network input layer, 3T-APS output voltage ΔV from CDS is directly fed into a voltage controlled current source (VCCS) with current gain G. Thus, the output current I_{out} of VCCS is actually proportional to light intensity (or gray value of the input image). AMI 0.6 um n-III CMOS technology was taken for APS module simulation, and other simulation parameters are listed in Table A.1 [106]. I_{ph} is the PD current change range, and C_{ph} is the PD capacitance. If we assume maximum gray value 255 corresponds to highest I_{ph} of 300 pA, and minimum gray value 0 corresponds to lowest I_{ph} of 50 pA, then I_{out} dependence on image gray value over time is shown in Figure A-6(b).

Figure A-6: (a) Active pixel sensor (APS) schematic. PD is modeled with a current source (I_{ph}) and a capacitor (C_{ph}). (b) The dependence of APS circuit output current I_{out} on image gray value shows that I_{out} is proportional to gray value/light intensity. Note that for each training period (10 s), the first 1 s is always the system reset cycle.

Table A.1: APS Simulation Parameters

<table>
<thead>
<tr>
<th>Para_{\text{tech}}</th>
<th>I_{ph}^{\circ}</th>
<th>C_{ph}^{\circ}</th>
<th>V_{dd}^{\circ}</th>
<th>G^{\circ}</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMI 0.6 (\mu)m Tech</td>
<td>300 pA ~ 50 pA</td>
<td>1 pF</td>
<td>5 V</td>
<td>(228 e^{-6}) S</td>
</tr>
</tbody>
</table>
A.5.2 Architecture

Figure A-7 shows the overall architecture of our simulation framework. Based on discussions presented in section II, when an $N \times N$ grayscale image is projected onto the APS array, gray values of input image are encoded into different sensor output current $I_{out}$. The $I_{out}$ corresponding to each gray value is directly fed to the $N \times N$ pre-synaptic neurons. For the whole learning/memory stage, $I_{out}$ remains constant for each pixel. The learning stage was performed by a crossbar array with a memristive device at each cross-point, shown in Figure A-7. The pre-synaptic and post-synaptic CMOS neurons were connected in a feedback-forward manner, using a simple LIF scheme [107]. The capacitor of LIF would integrate $I_{out}$ from APS array and fire a spike when integrated potential reached threshold voltage. To prevent two neurons from spiking at the same time, we employ lateral inhibitory connections on both input and output layer to implement winner-take-all (WTA) mechanism, such that as soon as one neuron fires, it
strongly inhibits its neighbors during inhibition time. For pre-synaptic neurons, WTA ensures image gray value and synapse weight value proportionality after training. For post-synaptic neurons, WTA prevents two neurons from learning the same image in one training epoch [108].

This scheme allows different images to be learned in different crossbar column. In the image recognition stage, we input an image, which has already been learned, and count the output spikes at the end of each post-synaptic neuron. The post-synaptic neuron with the highest firing rate was regarded as storing the matched image. Following sections discuss the details of each step.

A.5.3 Learning Simulation

The learning process refers to the process that neural network memorizes images by training. After training, different images will be stored into different crossbar columns. Therefore, for a $N^2 \times M$ crossbar array, it can store $M$ images of $N \times N$ pixels at most. In order to memorize one image into one selected column without influencing other columns, during one training epoch, only the selected post-synaptic neuron is active, while all the other neurons are inactive. In other words, only the column of synapses connected to the active column has a feedback spike to perform synaptic weight update, while all the other synapses will keep its initial value. This inhibition can be achieved by using WTA circuits at the neural network output layer.
Figure A-8: A $3 \times 3$ grayscale image with pixel numbered from 1 to 9 for neural network learning, and the corresponding gray value of each pixel is shown in the brackets.

Figure A-9: (a) Synaptic weights evolution for image (Figure A-8) training. (b) The resulting synaptic weights value (in the brackets) after training. G1 ~ G9 are the corresponding synapse of pixel 1~9 on the selected column.

A $3 \times 3$ grayscale image, as shown in Figure A-8, is fed into the APS array. After the APS array, the gray value of each pixel is encoded into the APS output current $I_{\text{out}}$. $I_{\text{out}}$ is proportional to the input light intensity/gray value, and held constant during training. First, the pre-synaptic neurons sum $I_{\text{out}}$, using simple LIF circuits, and fire an output spike across the pre-synaptic axon when the summed potential reaches to 1 V. The pre-synaptic spikes are then multiplied with the synapse conductance and the resulting current is summed using the post-synaptic integrator circuits of the selected crossbar column. When the summed potential reaches to 1 V, a post-synaptic spike is
fired in both the forward direction and the reverse direction. This reverse feedback spike will be responsible for synapse weight update according to STDP rule. For this numerical simulation, we used the following parameters: $I_0 = 1$ mA, $v_0 = 1/7$ V, $v_{th} = 1$ V, $k_{-1} = 1.8$ µA, pre-synaptic capacitance $C_1 = 2$ µF and post-synaptic capacitance $C_2 = 10$ pF. All the synapses have the same initial weight value ($G_{initial} = 10$ nS) and weight update range was from 10 nS to 100 nS. The entire training takes 10 s and the training result is shown in Figure A-9. The synaptic weight after training is proportional to the input pixel gray value, indicating that the image has been stored. Actually, in this case, this proportionality results from WTA mechanism of input layer. Once one pre-synaptic neuron fires, it strongly inhibits others (WTA) during the synapse update period. In other words, in one synapse update period, WTA ensures that only one synapse is updated, while all the others keep previous value. That is, the neurons with higher pixel gray value (higher $I_{out}$) is updated more frequently, thus larger synaptic weight value is achieved if we properly adjust the integrator parameters and training time.

![Binary image input for neural network training. They are named letter X, letter I and letter L from the left to the right, respectively.](image)

Figure A-10: Binary image input for neural network training. They are named letter X, letter I and letter L from the left to the right, respectively.
Figure A-11: Synapse training result for Figure 2-12. It clearly shows that all the input images has been memorized to the 1st ~ 5th crossbar column.

Figure A-12: Post-synaptic neuron output spike waveform monitored for 10 s during image recognition stage. (a) ~ (c) corresponds to the 1st ~ 3rd crossbar column and total spikes counted are 57, 18 and 36, respectively.
A.5.4 Image Recognition

Image recognition is the process of identifying or detecting an input target image based on learning. Here, we only show how to perform binary image recognition. To realize binary image recognition, first, the neural network was trained to memorize three $3 \times 3$ images (see Figure A-10), in which the gray value at each white color pixel is 255 and at each black color is 0. After 3 training epoch, the resulting synaptic weights are shown in Figure A-11. All the white color pixel corresponding synapses were trained to the maximum conductance $G_{\text{max}}$, while all the black color corresponding synapses were still kept in its initial conductance $G_{\text{min}}/G_{\text{initial}}$.

In the recognition stage, an unknown image, which is one of the learned image (in Figure A-10), was fed into the neural network. The summing current of the post-synaptic neuron at the matched column should be always higher than that of the other unmatched columns, because for the target input image, the white color pixel always correspond to $G_{\text{max}}$ synapse at the matched column, while it may correspond to $G_{\text{min}}$ at the unmatched column. Therefore, the matched column has the highest post-synaptic firing rate. In order to prove this assumption, we input the first binary image letter X (in Figure A-10) to the network, and design a pulse counter (with 0.5 V threshold voltage) to count firing spikes for 10 s at each post-synaptic neuron. The result is shown in Figure A-12. The 1st dendrite which is the matched column has the highest firing times of 57 in 10 s, while the firing times for the 2nd, 3rd dendrite were 18 and 36, respectively.

Note that during the image recognition stage, post-synaptic neuron output spikes only transmit in the forward direction to prevent it from affecting neuronal synapse
status. Meanwhile, all the post-synaptic neurons should be active, differing from the learning stage where only the selected post-synaptic neuron was active.

A.6. Summary

In this work, we demonstrated image learning and recognition using memristor based neural network combined with APS array. APS module was integrated with integrate-and-fire neurons array which can efficiently provide the spike coding of the input image. System level simulation results indicate that the network can memorize multiple images using one crossbar array and perform binary image recognition. This design work could be a groundwork for image processing circuits based on memristor crossbar arrays.
Appendix B

Ru/HfO₂/TiN ReRAM Devices

In this section, the fabrication and characterization of Ru/HfOₓ/TiN device was an attempt at achieving working ReRAM devices. Both unipolar and bipolar switching characteristics observed in Ru/HfOₓ/TiN are reported. This work could be a groundwork and guidance for future device fabrication and characterization.

B.1. Device Fabrication and Testing

ReRAM devices consisting of a Ti/Ru/HfO₂/TiN/W stack were fabricated on a p-Si substrate. All layers were deposited using RF magnetron sputtering. A 3 nm Ti layer as an adhesion layer followed by the 100 nm of Ru was deposited. The substrate temperature was increased to 300°C, after which 15 nm thick HfOₓ was deposited by reactive sputtering of Hf in 20% oxygen. 20 nm of TiN capped with 70 nm of W were deposited to form the top electrode (TE). The process flow is shown in Figure B-1(c). 30 µm × 30 µm device was employed for all the tests. Electrical characterization was taken in the LakeShore cryogenic probe station under 1×10⁻⁴ Torr and using Keithley 4200 Semiconductor Characterization System. All bias was applied to the top electrode TiN with bottom electrode grounded.
B.2. Electrical Characteristics

The initial I-V of device in Virgin Resistance State (VRS) is shown in Figure B-1(b). Then electroforming process was taken using both negative and positive DC voltage sweep on two different devices with the same VRS state. For Negative Forming (NF) device, the DC voltage was swept from 0 to -7 V with compliance current (CC) 1 mA and for Positive Forming (PF) device, the voltage sweep was from 0 to 7 V with CC also 1 mA. The device was formed at -4.5 V for NF and 6 V for PF, respectively (Figure B-1(a)). After forming, the device state was switched to Forming Resistance State (FRS).

Figure B-2(a) and Figure B-2(b) show unipolar resistive switching (URS) and bipolar resistive switching (BRS) of the device. Before any further discussion, we should make clear some abbreviations. NFURS device refers to URS switching device using NF forming; and PFBRS device refers to BRS switching device using PF forming. In Figure B-2(a), the NFURS device was formed using NF forming with the CC 1 mA (Figure B-2(a)). The device was thus changed to Low Resistance State (LRS) after forming (or FRS state). Initial reset was then performed by sweeping DC voltage from 0 V to -2.5 V. The device was sharply to be reset at -1.1 V and reset current was 5 mA (Reset current was the maximum current in the reset process, which indicates the current overshoot in the previous forming or set process [109]). The device was continuously reset until to -2.5 V and then the device was switched back to High Resistance State (HRS). By sweeping the voltage from 0 to -4 V, the LRS state can be again achieved by set process. NFURS device can get $10^4$ HRS/LRS resistance ratio but with high reset current ~ 5 mA.
(a) Log-log plot of current (A) vs voltage (V) showing the performance of different materials: W, TiN, HfO2, Ru. The plot compares high and low performance (PF and NF) conditions.

(b) Semi-log plot of current (A) vs voltage (V) showing the resistance states (VRS, FRS) with high and low performance (PF and NF) conditions.
Figure B-1: (a) Typical NF and PF forming. (b) VRS initial IV, FRS initial IV after NF and FRS initial IV after PF. VRS refers to Virgin Resistance State. FRS refers Formed Resistance State. (c) Process flow.
Figure B-2: Switching characteristics for (a) NFURS and (b) PFBRS.

In Figure B-2 (b), after PF forming as shown in Figure B-1 (a), a negative polarity voltage sweep biased to reset device. The reset started at -1 V and continued to -2.5 V and the device was changed to HRS state. Then device was set by DC sweep at around 1 V. In the PFBRS mode, the reset current was decreased below 1 mA, but HRS/LRS resistance ratio is only ~10.

Given that NFURS device has a higher resistance ratio HRS/LRS ($10^4$) and PFBRS device has relatively lower reset current (< 1 mA), we propose a new UFBRS technique, which is Unipolar Forming for Bipolar Resistive Switching technique, to combine the advantages of both NFURS and PFBRS to keep high resistance ratio device with low reset current simultaneously. We use the same forming and initial reset process as in the NFURS, and the same bipolar resistive switching process as in the PFBRS, which is shown in Figure B-3(a). For NFURS, PFBRS, UFBRS device, we extract resistance value (read at -0.1 V) and reset current to compare for 5 DC switching cycles,
shown in Figure B-3 (b) and Figure B-3 (c), from which we can see by UFBRS the device can still keep high HRS/LRS ratio (~1000) with low efficiently reduced reset current (~ 1 mA).
Figure B-3: UFBRS device I-V switching characteristics. (b) NFURS, PFBRS and UFBRS HRS/LRS distribution. (c) NFURS, PFBRS and UFBRS reset current distribution.

Endurance data is shown in the Figure B-4 (a), (b) and (c) for NFURS, PFBRS and UFBRS device respectively. Endurance test was performed by triangular switching pulse with compliance current 1 mA, and 5 devices were tested for each mode. The endurance of PFBRS device was ranging from $10^{3}$ and $10^{5}$ cycles. After Unipolar forming was used for BRS, BRS endurance was improved to ~$10^{6}$. NFURS has the best write endurance, which was larger than $10^{6}$ cycles. The endurance data shows that UFBRS endurance lies between PFBRS and NFURS. Interestingly, the endurance data of PFBRS showed NFURS endurance characteristics (High resistance ratio) in the first 100 cycles, while it showed PFBRS endurance characteristics (low resistance ratio) in the following cycles. This results also proves that UFBRS combine the characteristics of PFBRS and NFURS.
B.3. Summary

In conclusion, switching characteristics of NFURS and PFBRS based on Ru/HfOx/TiN stack device was studied. Based on this study, device exhibit low reset current (i.e. low power consumption in reset process) in the bipolar mode, while HRS/LRS ratio and high endurance is obtained in unipolar mode. Combining the advantages of NFURS and PFBRS, we come up with a new UFBRS technique. Experimental data shows the UFBRS can efficiently decrease the reset current of NFURS and increase the HRS/LRS ratio of PFBRS. This study is the first attempt at achieving
ReRAM devices and paves a solid foundation for future ReRAM fabrication and characterization.
Appendix C

Crossbar Mask Design and Process Flow

In Chapter 5, we have presented electrical characterization data of Ru/ZnO/TiN ReRAM device in crossbar structure. In this section, we will focus on complete process flow of fabricating crossbar devices.

The purpose of crossbar mask design is to demonstrate ReRAM functionality both on single device level and crossbar level. The mask design template is shown in Figure C-1. All masks design were made with this template for 4 inch wafer. Four letters “U”, “T”, “C”, and “B” as global marks were employed on each mask to roughly align any two different masks, as shown in Figure C-2. For precise align marks, we use the design on wafer as shown in Figure C-3 and design on mask as shown in Figure C-4 to align two masks for etching purpose, whereas we use the design on wafer as shown in Figure C-5 and design on mask as shown in Figure C-6 to align two masks for lift-off process. Precise marks were distributed on the both left and right side of mask in the middle line. Follow such rules, four masks identified as Mask #1, Mask #2, Mask #3 and Mask #4 were designed, as shown in Figure C-7 to Figure C-10. Mask #1 is bottom crossbar or BE mask. Mask #2 is stack and TE mask. Mask #3 is insulator openings mask. Mask #4 is
top crossbar mask. Mask #1, #2 and #4 are etching mask, whereas mask #3 is lift-off mask.

Figure C-1: 5” × 5” crossbar mask design template. All mask design in the following discussion is developed by this template.
Figure C-2: Global marks using four letters “U”, “T”, “C” and “B” to roughly align two masks in fabrication process.
Figure C-3: Precise mark design on wafer for etching mask to precisely align two masks.
Figure C-4: Precise mark design on mask for etching mask to precisely align two masks.
Figure C-5: Precise mark on wafer for lift-off mask to precisely align two masks.
Figure C-6: Precise mark on mask for lift-off mask to precisely align two masks.
Figure C-7: Mask #1 for patterning BE or bottom crossbar.
Figure C-8: Mask #2 for patterning stack and TE.
Figure C-9: Mask #3 for patterning insulator openings.
Figure C-10: Mask #4 for patterning top crossbar.
Figure C-11: Overview of mask design by overlapping four masks together.
Figure C-12: Mask layout showing placement of different device size areas.

Overlap Mask #1, Mask #2, Mask #3 and Mask #4 together and the mask design overview is shown in Figure C-11. The whole wafer was divided into twelve 22 mm × 22 mm chips according to device area which arranges from 20 µm × 20 µm to 90 µm × 90 µm as shown in Figure C-12. For each size area, there are 1 × 1, 2 × 2, 4 × 4 and 16 × 16 crossbar were designed.
The complete process follows as described below. Here, a 1 × 1 crossbar is used to illustrate photolithography process. Firstly, a layer of 1000 A SiO$_2$ was deposited on the 4” wafer as insulating layer. 200 A Ru as bottom electrode (BE) was then deposited on wafer. Figure C-13 shows the BE pattern after photolithography using Mask #1. An active layer of 60 nm ZnO and top electrode 10 nm TiN with hard contact layer 50 nm W were then depositing on wafer by RF magnetron reactive sputtering, and Mask #2 was used to define ReRAM device area, as shown in Figure C-14. Then, an insulator layer (e.g. photoresist S1813 or Si$_3$N$_4$) was deposited on the wafer to prevent shorting between top crossbar and bottom crossbar. Insulator openings were patterned using Mask #3, as shown in Figure C-15. Finally, top crossbar layer of 50 nm W was deposited and patterned to form top crossbar, as shown in Figure C-16.

For different crossbar scale, i.e. 1 × 1, 2 × 2, 4 × 4 and 16 × 16, only the stack and TE area patterned by Mask #2 and openings on top of stack and TE are different, and all the other design dimensions are the same. Thus, in the following discussion, we only show the crossbar dimensions for 20 µm × 20 µm device area. Figure C-17 to Figure C-20 show dimensions of 1 × 1, 2 × 2, 4 × 4 and 16 × 16 crossbar, and Figure C-21 shows crossbar dimensions for DIP package. The unit is micro (µm).
Figure C-13: BE patterned by using Mask #1.

Figure C-14: Stack and TE patterned by using mask #2.

Figure C-15: Insulator openings patterned by using mask #3.
Figure C-16: Top crossbar patterned by using Mask #4.
Figure C-17: Design dimensions for $1 \times 1$ crossbar.
Figure C-18: Design dimensions for 2 × 2 crossbar.
Figure C-19: Design dimensions for $4 \times 4$ crossbar.
Figure C-20: Design dimensions for 16 × 16 crossbar.
Figure C-21: Design dimensions for crossbar DIP package.
Appendix D

Ru/ZnO/TiN Fabrication Process Flow

<table>
<thead>
<tr>
<th>Steps</th>
<th>Description</th>
</tr>
</thead>
</table>
| 1     | **Wafer Labeling (Scribing):**  
       | Total Number of Wafers: 1  
       | Wafer ID: CBS_ZnO_4_Mask  
       | Comments: Facility: UT NF  
       | Operator: date: |
| 2     | **Cleaning & Surface Treatment**  
       | Wafer ID: CBS_ZnO_4_Mask  
       | Recipe: Pre-Furnace Cleaning  
       | Comments: Facility: MMD  
       | Operator: date: |
| 3     | **Deposition of SiO2**  
       | Wafer ID: CBS_ZnO_4_Mask  
       | Tool: MMD Dry Oxidation Furnace  
       | Temperature: Thickness: 1000 A  
       | Comments: Facility: MMD  
<pre><code>   | Operator: date: |
</code></pre>
<p>| 4     | <strong>Deposition of Bottom Electrode</strong> |</p>
<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
<th>Wafer ID</th>
<th>Tool</th>
<th>Target</th>
<th>Reactive Gas</th>
<th>Base Pressure</th>
<th>Process &amp; Parameters</th>
<th>Thickness</th>
<th>Time</th>
<th>Comments</th>
<th>Facility</th>
<th>Operator</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sputtering</td>
<td>CBS_ZnO_4_Mask</td>
<td>UT Sputtering</td>
<td>Ru</td>
<td>None</td>
<td></td>
<td>100 W, 5 mTorr, 40 rpm, 10 mm, Room Temperature</td>
<td>200 A</td>
<td>7 minute and 36 seconds</td>
<td></td>
<td>UT NF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Prebake</td>
<td>CBS_ZnO_4_Mask</td>
<td>UT Digital Hotplate</td>
<td></td>
<td></td>
<td></td>
<td>Temperature: 115 C</td>
<td>2 minutes</td>
<td>Cool for 2 minutes</td>
<td></td>
<td>UT NF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Spincoating Photoresist</td>
<td>CBS_ZnO_4_Mask</td>
<td>UT Spinner</td>
<td></td>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>UT NF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Soft Bake</td>
<td>CBS_ZnO_4_Mask</td>
<td>UT Digital Hotplate</td>
<td></td>
<td></td>
<td></td>
<td>Temperature: 115 C</td>
<td>1 minute</td>
<td></td>
<td></td>
<td>UT NF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Lithography to Pattern Bottom Cross-Bars (also Bottom Electrode)</td>
<td>CBS_ZnO_4_Mask</td>
<td>UT Mask Aligner</td>
<td></td>
<td></td>
<td></td>
<td>Photoresist: S1813</td>
<td></td>
<td>Intensity: 25 mW/cm^2</td>
<td></td>
<td>UT NF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step</td>
<td>Description</td>
<td>Details</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Develop Photoresist</td>
<td>Wafer ID: CBS_ZnO_4_Mask, Developer: MF319, Time: 1 minute, Comments:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Metal Etching</td>
<td>Wafer ID: CBS_ZnO_4_Mask, Etchant: Ruthenium Etchant, Ru-44, Time: 40 seconds, Comments:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Photoresist Removal</td>
<td>Wafer ID: CBS_ZnO_4_Mask, Remover: PG Remover in Ultrasonicator, Time: 60 minutes (overnight is better), Comments:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Deposition of ZnO</td>
<td>Wafer ID: CBS_ZnO_4_Mask, Tool: UT Sputtering, Target: Zn, Reactive Gas: Oxygen, Flow: Ar = 9.6 and O = 2.4, Base Pressure: Process &amp; Parameters: Zn: 100 W, 5 mTorr, 40 rpm, 10 mm, 100 °C, Thickness: 600 Å, Time: 1 hour, Comments:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Deposition of Upper Metal Electrode</td>
<td>Wafer ID: CBS_ZnO_4_Mask, Tool: UT Sputtering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Target: Ti  
Reactive Gas: Nitrogen  Flow: Ar = 6 and N = 6  
Base Pressure:  
Process & Parameters: 150 W, 5 mTorr, 40 rpm, 10 mm, 300 °C  
Thickness: 100 A  
Time: 39 minutes and 45 seconds  
Comments: Let Ar:N flow until 50 °C.  
Facility: UT NF  
Operator:  

14  
**Deposition of Capping Layer**  
Wafer ID: CBS_ZnO_4_Mask  
Tool: UT Sputtering  
Target: W  
Reactive Gas: None  
Base Pressure:  
Process & Parameters: 100 W, 5 mTorr, 40 rpm, 10 mm, Room Temperature  
Thickness: 500 A  
Time: 14 minutes and 36 seconds  
Comments:  
Facility: UT NF  
Operator:  

15  
**Prebake**  
Wafer ID: CBS_ZnO_4_Mask  
Tool: UT Digital Hotplate  
Temperature: 115 C  
Time: 2 minutes  
Comments: Cool for 2 minutes  
Facility: UT NF  
Operator:  

16  
**Spincoating Photoresist**  
Wafer ID: CBS_ZnO_4_Mask  
Tool: UT Spinner  
Primer: None  
Comments:  
Facility: UT NF  
Operator:  

17  
**Soft Bake**  
Wafer ID: CBS_ZnO_4_Mask  
Tool: UT Digital Hotplate  
Temperature: 115 C  
Time: 1 minute
<table>
<thead>
<tr>
<th>Comments:</th>
<th>Facility: UT NF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operator:</td>
<td>date:</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>18</th>
<th><strong>Lithography to Pattern Stack</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wafer ID: CBS_ZnO_4_Mask</td>
<td></td>
</tr>
<tr>
<td>Photoresist: S1813</td>
<td></td>
</tr>
<tr>
<td>Tool: UT Mask Aligner</td>
<td></td>
</tr>
<tr>
<td>Mask: CBS - Mask 2</td>
<td></td>
</tr>
<tr>
<td>Intensity: 25 mW/cm^2</td>
<td></td>
</tr>
<tr>
<td>Time: 6 seconds</td>
<td></td>
</tr>
<tr>
<td>Comments:</td>
<td></td>
</tr>
<tr>
<td>Facility: UT NF</td>
<td></td>
</tr>
<tr>
<td>Operator:</td>
<td>date:</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>19</th>
<th><strong>Develop Photoresist</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wafer ID: CBS_ZnO_4_Mask</td>
<td></td>
</tr>
<tr>
<td>Developer: MF319</td>
<td></td>
</tr>
<tr>
<td>Time: 1 minute</td>
<td></td>
</tr>
<tr>
<td>Comments:</td>
<td></td>
</tr>
<tr>
<td>Facility: UT NF</td>
<td></td>
</tr>
<tr>
<td>Operator:</td>
<td>date:</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>20</th>
<th><strong>Metal Etching</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wafer ID: CBS_ZnO_4_Mask</td>
<td></td>
</tr>
<tr>
<td>Etchant: Tungsten Etchant</td>
<td></td>
</tr>
<tr>
<td>Time: 19 seconds</td>
<td></td>
</tr>
<tr>
<td>Comments:</td>
<td></td>
</tr>
<tr>
<td>Facility: UT NF</td>
<td></td>
</tr>
<tr>
<td>Operator:</td>
<td>date:</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>21</th>
<th><strong>Metal Etching</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wafer ID: CBS_ZnO_4_Mask</td>
<td></td>
</tr>
<tr>
<td>Etchant: Titanium Nitride Etchant</td>
<td></td>
</tr>
<tr>
<td>Time: 8 seconds</td>
<td></td>
</tr>
<tr>
<td>Comments:</td>
<td></td>
</tr>
<tr>
<td>Facility: UT NF</td>
<td></td>
</tr>
<tr>
<td>Operator:</td>
<td>date:</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>22</th>
<th><strong>Photoresist Removal</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wafer ID: CBS_ZnO_4_Mask</td>
<td></td>
</tr>
<tr>
<td>Remover: PG Remover</td>
<td></td>
</tr>
<tr>
<td>Time: 1 minute</td>
<td></td>
</tr>
<tr>
<td>Comments:</td>
<td></td>
</tr>
<tr>
<td>Facility: UT NF</td>
<td></td>
</tr>
<tr>
<td>Operator:</td>
<td>date:</td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
</tr>
<tr>
<td>24</td>
<td></td>
</tr>
<tr>
<td><strong>Prebake</strong></td>
<td></td>
</tr>
<tr>
<td>Wafer ID: CBS_ZnO_4_Mask</td>
<td></td>
</tr>
<tr>
<td>Tool: UT Digital Hotplate</td>
<td></td>
</tr>
<tr>
<td>Temperature: 115 C</td>
<td></td>
</tr>
<tr>
<td>Time: 2 minutes</td>
<td></td>
</tr>
<tr>
<td>Comments: Cool for 2 minutes</td>
<td></td>
</tr>
<tr>
<td>Facility: UT NF</td>
<td></td>
</tr>
<tr>
<td>Operator:</td>
<td>date:</td>
</tr>
<tr>
<td>25</td>
<td></td>
</tr>
<tr>
<td><strong>Spincoating Photoresist</strong></td>
<td></td>
</tr>
<tr>
<td>Wafer ID: CBS_ZnO_4_Mask</td>
<td></td>
</tr>
<tr>
<td>Tool: UT Spinner</td>
<td></td>
</tr>
<tr>
<td>Primer: None</td>
<td></td>
</tr>
<tr>
<td>Comments:</td>
<td></td>
</tr>
<tr>
<td>Facility: UT NF</td>
<td></td>
</tr>
<tr>
<td>Operator:</td>
<td>date:</td>
</tr>
<tr>
<td>26</td>
<td></td>
</tr>
<tr>
<td><strong>Soft Bake</strong></td>
<td></td>
</tr>
<tr>
<td>Wafer ID: CBS_ZnO_4_Mask</td>
<td></td>
</tr>
<tr>
<td>Tool: UT Digital Hotplate</td>
<td></td>
</tr>
<tr>
<td>Temperature: 115 C</td>
<td></td>
</tr>
<tr>
<td>Time: 1 minute</td>
<td></td>
</tr>
<tr>
<td>Comments:</td>
<td></td>
</tr>
<tr>
<td>Facility: UT NF</td>
<td></td>
</tr>
<tr>
<td>Operator:</td>
<td>date:</td>
</tr>
<tr>
<td>27</td>
<td></td>
</tr>
<tr>
<td><strong>Lithography to Pattern Photoresist Openings</strong></td>
<td></td>
</tr>
<tr>
<td>Wafer ID: CBS_ZnO_4_Mask</td>
<td></td>
</tr>
<tr>
<td>Photoresist: S1813</td>
<td></td>
</tr>
<tr>
<td>Tool: UT Mask Aligner</td>
<td></td>
</tr>
<tr>
<td>Mask: CBS - Mask 3</td>
<td></td>
</tr>
<tr>
<td>Intensity: 25 mW/cm^2</td>
<td></td>
</tr>
<tr>
<td>Time: 6 seconds</td>
<td></td>
</tr>
<tr>
<td>Comments:</td>
<td></td>
</tr>
<tr>
<td>Facility: UT NF</td>
<td></td>
</tr>
<tr>
<td>Operator:</td>
<td>date:</td>
</tr>
<tr>
<td>28</td>
<td></td>
</tr>
<tr>
<td><strong>Develop Photoresist</strong></td>
<td></td>
</tr>
<tr>
<td>Wafer ID: CBS_ZnO_4_Mask</td>
<td></td>
</tr>
<tr>
<td>Developer: MF319</td>
<td></td>
</tr>
<tr>
<td>Time: 1 minute</td>
<td></td>
</tr>
<tr>
<td>Comments:</td>
<td></td>
</tr>
<tr>
<td>Facility: UT NF</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Event</td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>31</td>
<td><strong>Deposition of Contact Pads</strong></td>
</tr>
<tr>
<td>32</td>
<td><strong>Prebake</strong></td>
</tr>
<tr>
<td>33</td>
<td><strong>Spincoating Photoresist</strong></td>
</tr>
<tr>
<td>34</td>
<td><strong>Soft Bake</strong></td>
</tr>
<tr>
<td>35</td>
<td><strong>Lithography to Pattern Top Cross-Bars</strong></td>
</tr>
<tr>
<td>Process</td>
<td>Wafer ID</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
</tr>
<tr>
<td>36</td>
<td>Develop Photoresist</td>
</tr>
<tr>
<td>37</td>
<td>Metal Etching</td>
</tr>
<tr>
<td>38</td>
<td>Photoresist Removal</td>
</tr>
</tbody>
</table>