Tunable Volatile and Multi-bit Non-volatile Resistive Memory for Neuromorphic Computing Applications

by

Rabiul Islam

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2025

© Rabiul Islam 2025

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Dr. Sambandamurthy Ganapathy Professor, Dept. of Physics University at Buffalo
Supervisor(s):	Dr. Guo-Xing Miao Associate Professor, Dept. of ECE University of Waterloo Dr. Manoj Sachdev Professor, Dept. of ECE University of Waterloo
Internal Member:	Dr. Youngki Yoon Associate Professor, Dept. of ECE University of Waterloo Dr. Bo Cui Professor, Dept. of ECE University of Waterloo
Internal-External Member:	Dr. Mustafa Yavuz Professor, Dept. of MME University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This thesis consists of Background (Chapter 2), Experimental Methodology (Chapter 3), and two research chapters that are written in a manuscript format (Chapter 4 & 5). All manuscripts have been modified to fit into the thesis style. I am the first author and Dr. Guo-Xing Miao is the corresponding author of all manuscripts. Several co-authors contributed to the research papers, and their contributions are described below.

Chapter 3

Rabiul Islam, Bo Cui, Guo-Xing Miao; "Dry etching strategy of spin-transfer-torque magnetic random access memory: A review." ***J. Vac. Sci. Technol. B*** September 2020; 38 (5): 050801. https://doi.org/10.1116/6.0000205

Contributions: Guo-Xing Miao and Bo Cui guided and edited the manuscript of the review article. I conceptualized the idea and wrote the initial draft.

Chapter 4

Rabiul Islam, Yu Shi, Gabriel Vinicius de Oliveira Silva, Manoj Sachdev, and Guo-Xing Miao; "Volatile and Non-volatile Programmable Iontronic Memristor with Lithium Imbued TiO x for Neuromorphic Computing Applications." ***ACS nano*** August 2024; 18 (33) (2024): 22045-22054. https://doi.org/10.1021/acsnano.4c05137

Contributions: Guo-Xing Miao and Manoj Sachdev designed and directed the research. Me and Yu Shi, PhD candidate under Prof. Go-Xing Miao, jointly optimized the materials growth and device fabrication. I performed the electrical measurements and transport measurements for the devices used in this work. Yu Shi performed the RC simulation and wrote the associated portion. Gabriel Vinicius de Oliveira Silva performed the DFT calculation and wrote the associated portion. All authors analyzed the data and wrote the manuscript.

Chapter 5

Rabiul Islam, Peng Li, Marijan Beg, Manoj Sachdev, Guo-Xing Miao; "Helimagnet-based non-volatile multi-bit memory units." ***Appl. Phys. Lett.*** April 2023; 122 (15): 152407. https://doi.org/10.1063/5.0147066

Contributions: Guo-Xing Miao conceptualized the research idea. Peng Li performed the fabrication and measurement of the device. Marijan Beg helped me to learn the simulation software. All authors analyzed the data and wrote the manuscript.

Abstract

In the existing computing paradigm, to compute a function, data from memory needs to be transferred to the processor and upon processing the data, results need to move back to memory again to complete the computation. This data movement process consumes a significant amount of energy. Moreover, the computing process is limited by the bandwidth of the data bus and increased latency. Therefore, alternative non-von Neumann computing architectures, especially neuromorphic computing and in-memory computing architectures, have been extensively investigated by researchers. In neuromorphic computing (NC) architecture, which is inspired by the energy-efficient biological brain, the processing elements and memory are colocated, thereby eliminating the energy cost and latency of data shuffling. CMOS-circuits-based neuromorphic computing architectures, such as Loihi and TrueNorth, have been extensively studied to emulate the synapses and neuron-like behaviour; however, they require tens of CMOS gates to imitate single neuron behaviour, which poses a challenge when large-scale neural network is needed. A single memristors-based device, on the other hand, can easily emulate the synaptic plasticity. Therefore, memristors-based neuromorphic hardware enables high integration density with low energy consumption.

In this thesis, a lithium (Li) imbued TiO_x iontronic device that exhibits both volatile and non-volatile memory characteristics has been demonstrated. A solid-state electrolyte lithium phosphorus oxynitride (LiOPN) behaves as the ion source, and the embedding and releasing of Li ions inside the cathodic like TiO_x renders volatile conductance responses from the device and offers a natural platform for hardware simulating neuron functionalities. In addition, the device unambiguously emulates synapse-like short-term plasticity (STP) behaviour without requiring a forming process beforehand or a compliance current during switching. Besides, these devices possess high uniformity and great endurance as no conductive filaments are present. Different short-term pulse-based phenomena, including paired-pulse facilitation (PPF), post-tetanic potentiation (PTP), and spike rate-dependent plasticity (SRDP), were observed with unique self-relaxation characteristics. Based on the voltage excitation period, the timescale of the volatile memory can be tuned. Temperature measurement reveals the ion displacement-induced conductance channels become frozen below 220 K. In addition, the same volatile analog devices can be configured into non-volatile memory units with multibit storage capabilities after an electroforming process, which is useful for in-memory computing (IMC) applications. Therefore, on the same platform, we can configure volatile units as nonlinear dynamic reservoirs for performing neuromorphic training and the non-volatile units as the weight storage layer. We proceed to use voice recognition as an example with the tunable time constant relationship and

obtain 94.4% accuracy with a minimal training dataset. Thus, this iontronic platform can effectively process and update temporal information for reservoir and neuromorphic computing paradigms.

Moreover, the same lithium-imbued TiO_x device was integrated on top of the CMOS tape-out chip. Due to the incompatibility of Lithium-based materials with the existing CMOS production facility, CMOS integration with Lithium-imbued TiO_x devices is almost impossible. This thesis proposes a BEOL method to integrate the Li-imbued TiO_x with foundry CMOS chips using an academic fabrication facility. This process will allow the other CMOS process-incompatible devices to be integrated with CMOS.

Furthermore, this thesis also proposed a Helimagnet-based Non-volatile multi-bit memory device. The device consists of a Helimagnet layer sandwiched in between two ferromagnetic layers. The writing of the proposed memory was done using an external magnetic field applied on the top free layer. The memory state can be read by using resistance measurement after each stable configuration. Consequently, this device can be used for in-memory computing (IMC) applications due to its multi-bit, non-volatile nature.

Acknowledgements

First of all, I would like to thank almighty Allah, the creator, the lord of the universe, the sustainer, and the cherisher, for bestowing me with constant blessings and providing me with the ability to wrap up the PhD thesis.

I am grateful to my supervisor, Dr. Guo-Xing Miao, for his unwavering support throughout my PhD journey. Whenever I approached him for a discussion, he was always pleased to take his time out. Also, from the very beginning of PhD, he laid out clear research topics for my research development. Although I had no previous experience in semiconductor device fabrication and characterization tools, Dr. Miao accepted me to join his group and thoroughly trained these aspects during my PhD. I will really miss the weekly group meeting, where I observe his insights. I am really indebted to get training from such a great person.

In addition, I would like to thank my co-supervisor, Dr. Manoj Sachdev, with whom we met almost weekly basis to update our research progress. I learnt from him how to think out of the box so that applications can be more impactful. Moreover, I appreciate his time and dedication to reading and providing feedback on my thesis in his busy schedule.

I also want to thank all my groupmates, including Yu Shi, Gabriel, He Ren, Peng Li, and Lin Li, for their support during my stay. All of them are hard-working and genius, and I learned a lot from their company. A special thanks to Yu Shi, from whom I learned a lot, starting from basic fabrication to characterization skills. He carried out the reservoir computing simulation and tape-out circuit design tasks described in this thesis.

I would like to especially thank all my committee members, Dr. Youngki Yoon, Dr. Bo Cui, and Dr. Mustafa Yavuz, for providing me with their consent to evaluate my thesis even in their busy schedules.

I am pleased to thank the team members of The Quantum-Nano Fabrication and Characterization Facility (QNFCF) of the University of Waterloo. All members, including Dr. Nathan Nelson Fitz-Patrick, Dr. Greg Holloway, Dr. Sandra Gibson, and Nicki Shaw, have trained me to achieve my goal in the fabrication and characterization of my devices.

Finally, I would like to acknowledge the sacrifice of my family members. My stay would be impossible without their patience and sacrifice. Special thanks to my wife, who always supported me during this difficult journey. She sacrificed and dedicated her time throughout my PhD journey. I am always grateful to you for your tireless support throughout my PhD journey.

Dedication

This thesis is dedicated to the apple of my eye.

Table of Contents

Examining Committee ii

Author's Declaration iii

Statement of Contributions iv

Abstract v

Acknowledgements vii

Dedication viii

List of Figures xiii

List of Tables xx

List of Abbreviations xxi

List of Symbols xxiv

1 Introduction 1

1.1 Motivation 1

1.2 Thesis Organization 6

2 Background **8**
2.1 Memristive Device . . . 8
2.1.1 Device Layout and Materials . . . 9
2.1.2 Resistive Switching Types . . . 10
2.2 Resistive Switching Mechanisms . . . 14
2.2.1 Filament-type RS . . . 14
2.2.2 Interface-type RS . . . 17
2.3 Conduction Mechanism . . . 18
2.4 Battery-like Memristors . . . 19
2.4.1 Ionic Conductivity and LiPON . . . 20
2.4.2 **TiO_x** and **LiPON** Switching Layer . . . 22
2.5 Memristive Synapses . . . 23
2.5.1 Memristive STP . . . 24
2.5.2 Memristive LTP . . . 26
2.6 Emerging Computing Architecture . . . 28
2.6.1 Reservoir Computing using Memristors . . . 28
2.6.2 Multi-bit-Memory-based In-memory Computing . . . 31

3 Experimental Methodology **34**
3.1 Thin-film Deposition Processes . . . 34
3.1.1 Physical Sputtering . . . 35
3.1.2 Plasma-Enhanced Chemical Vapor Deposition . . . 36
3.2 Etching Processes . . . 37
3.2.1 Ion Milling . . . 37
3.2.2 Reactive Ion Etching . . . 39
3.3 Lithography . . . 40
3.3.1 Photolithography . . . 40
3.3.2 Electron-Beam Lithography . . . 41
3.4 Device Fabrication . . . 42
3.5 BEOL Integration Processes . . . 44

4 Volatile and Non-volatile Iontronic Memory **47**
4.1 Introduction . . . 48
4.2 Experimental Methods . . . 49
4.2.1 Device Fabrication and Characterization Steps . . . 49
4.2.2 Electrical and Temperature Measurement Steps . . . 50
4.2.3 DFT Calculations . . . 50
4.2.4 Voice Recognition Configuration . . . 51
4.3 Results and Discussion . . . 51
4.3.1 Device Material and Structure Characterizations . . . 51
4.3.2 DC Characteristics . . . 52
4.3.3 Pulse Characteristics . . . 56
4.3.4 Temperature Dependent Characteristics . . . 59
4.3.5 Non-volatile Characteristics . . . 63
4.3.6 Memristor-based Reservoir Computing . . . 65
4.4 Conclusions . . . 68

5 Helimagnet-based Non-volatile Multi-bit Memory **70**
5.1 Introduction . . . 71
5.2 Simulation Details . . . 72
5.3 Results and Discussions . . . 74
5.3.1 Competition between Exchange and DMI . . . 74
5.3.2 Writing Mechanism . . . 76
5.3.3 Application of Proposed Writing Method in HMRAM . . . 80
5.3.4 Reading of the Memory states . . . 80
5.3.5 HMRAM crossbar array for IMC . . . 81
5.4 Conclusion . . . 82

6 Conclusion and Future Work **83**

6.1 Conclusion 83

6.2 Future Work 85

References **87**

List of Figures

1.1 (a) Current computing power demands and future computing demands over the past forty years with different applications domain. The computing demand before 2012 doubled every two years, whereas current computing demand doubled every two months. (b) An exponential increase in the cost of AI hardware training would be unsustainable in the long run [1]. Copyright © 2022 Springer Nature. 2

1.2 Constraints of the existing computing schemes, including heat wall, memory wall, and scaling wall. Alternating memristor-based non-von Neumann computing paradigms, which include bio-inspired computing, in-memory computing, and hybrid memory and logic integration, could overcome those barriers [2]. Copyright © 2018 Springer Nature. 3

2.1 (a) Memristive crossbar array with bit lines that are used to read/write data and word lines that are used to select a memristive device. The inset shows a unit memristive device consisting of two metal electrodes and a metal oxide switching layer. (b) Two different stable resistance states of a single memory cell. 9

2.2 Current-Voltage (IV) characteristics of (a) an unipolar type resistive switching (RS). (b) a bipolar type RS. 10

2.3 Volatile resistive switching types: (a) analog-type switching and (b) digital-type switching. (c) Conductance decay in analog-type volatile resistive switching ($Pd/WO_x/W$) after the removal of voltage excitation [3]. Copyright © 2015 John Wiley and Sons. (d) Conductance decay in threshold-type resistive switching device (Au/SiO_xN_y:Ag/Au) [4]. Copyright © 2016 Springer Nature. 12

2.4 Electrical characteristics comparison between bipolar digital and analog-type non-volatile memory. (a) DC I–V characteristics and (b) Change in conductance upon the applications of pulse of a bipolar digital non-volatile memory device. The bipolar digital device exhibits a sharp change in conductance. (c) DC I–V characteristics and (d) pulse operation of an analog-type non-volatile resistive memory device. The analog-type non-volatile memory device exhibits gradual switching [5]. Copyright © 2020 IEEE. . . 13

2.5 Schematic illustration of a conductive bridge resistive random-access memory [6]. Copyright © 2011 IOP Publishing. 15

2.6 Schematic illustration of an oxygen vacancy based RS [7]. 16

2.7 The switching mechanism of interface-type memristors. (a) During the LRS transition, vacancies accumulate near the metal-oxide interface, decreasing the depletion width. (b) During the HRS transition, the depletion width increases due to fewer defects at the interface [8]. 18

2.8 Probable conduction mechanism picture in a metal-insulator-metal system [9]. 19

2.9 Schematic representation of (a) various defects, (b) **Li**-ion migration pathway (left) with calculated energy along the migration path (right), (c) activation barrier of **Li** ion as a function of lattice volume, and (d) migration mechanism in solid-state ceramic electrolytes [10, 11]. Copyright © 2018 ELSEVIER and 2015 Springer Nature. 21

2.10 Crystal structure of different oxide electrolytes used in solid-state batteries. Most of them can be amorphous or crystalline structures except **LiPON**, which is amorphous-type only. Yellow spheres represent the lithium-ion position in the lattice [12]. Copyright © 2021 Springer Nature. 22

2.11 Ternary phase diagram of Lithium-Titanium-Oxygen system with lithiation path (dashed line) [13]. Copyright © 2019 MDPI. 23

2.12 Schematic of a biological synapse connecting a presynaptic neuron with a postsynaptic neuron. Upon the generation of action potential in the presynaptic neuron, voltage-gated calcium ion channels open, and calcium influx causes the neurotransmitter to be released near the synaptic cleft. These neurotransmitters then bind with receptor molecules and alter the connection strength of the postsynaptic neurons [14]. Copyright © 2018 John Wiley and Sons. 25

2.13 (a) Two consecutive presynaptic spikes with a 200ms delay applied in a memristive device and corresponding excitatory post-synaptic current (EPSC) from the device. Here, A_1 and A_2 are the current amplitudes of the first and second EPSCs, respectively. (b) PPF index as a function of the interval between two presynaptic spikes [15]. Copyright © 2014 Springer Nature. (c) Memristive devices mimic short-term plasticity, including PPF and PPD, when consecutive pulses with different pulse delays are applied to the device. (d) Memristive devices with PPD followed by PPF behaviour [4]. Copyright © 2017 Springer Nature. 26

2.14 Spike timing dependent plasticity (STDP) of a biological synapse. If the presynaptic input precedes the postsynaptic one, the conductance increases, known as long-term potentiation (LTP). In the opposite case, the conductance decreases, known as long-term depression (LTP). The closer the timing, the larger the relative conductance change, which is defined as $\Delta G = (G_a - G_p)/ G_p$, where G_p is the conductance value before the synaptic input and G_a is the conductance value after the synaptic input [16, 17]. Copyright © 1998 Society for Neuroscience & 2011 IEEE. 27

2.15 Reservoir computing and reservoir computing architecture with memristive devices. a) A schematic illustration of a typical RC system that shows the reservoir with internal connections and a readout layer. It is interesting to note that only the weight matrices Θ are trained. b) Illustration of a Memristive RC system with encoded pulse streams as the inputs, the reservoir based on memristors, and a readout layer, which is a fully connected neural network. In an example of a digit recognition task that consists of 5×4 pixels images, and the reservoir has 5 volatile memristors [18]. Copyright © 2017 Springer Nature. 29

2.16 Reservoir computing hardware using volatile dynamic memristor(DM) and multi-bit non-volatile memristors. Hysteresis current-voltage curves, an optical image of 24 DMs, and a cross-sectional TEM image of the DM (inset) with vertically stacked Ti/TiO_x/Pd device structure are shown on the left. A photograph of the PCB hardware that integrated the power module with an MCU, reservoir module, and readout module (consisting of four non-volatile memory chips) is shown in the middle. Current-voltage characteristics of multi-bit capable non-volatile memory devices with TiN/TaO_x/$HfAlO_y$/TiN material stack and corresponding cross-sectional TEM image are shown on the right [19]. Copyright © 2022 Springer Nature. 30

2.17 a) Schematics of a typical neural network. b) The illustration of a digital and analog matrix-vector multiplication (MVM) scheme. In digital MVM implementation, multiplier digital blocks are used to multiply two digits followed by adders blocks to add the results. The analog VMM can be realized by adding currents from M lines and repeating the same in N columns of an RS-based crossbar. c) The hardware implementation of MVM using a memristive crossbar, in which each intersection contains a memory cell. The bottom electrodes, where input vectors are applied, act as a word line. The devices' conductance acts as a weight matrix. The resultant current from the top electrode, which also acts as a bit line, represents the MVM result. d) Three-dimensional image of a back-end-of-the-line (BEOL) integrated memristive crossbar [20]. Copyright © 2020 Wiley-VCH GmbH. 32

3.1 Schematic illustration of an RF Magnetron Sputtering system with substrate rotation capability. 35

3.2 Illustration of a typical ion beam etching system. 38

3.3 Simple diagram of different parts of an ICP-RIE etching system. 39

3.4 Memory device fabrication steps along with a brief description of the process under the steps. Both the top view and side view are given to clarify the fabrication steps. 42

3.5 Fabrication steps of BEOL integration process along with a brief description of the processes under the steps. Both the top view and side view are given to clarify the fabrication steps. 44

3.6 (a) AFM image of M6 layer opening showing surface roughness and depth of the hole. (b) The depth profile of the transistor. 45

3.7 SEM images of CMOS chip after bottom electrode deposition, followed by passivation with SiO_2 and RIE etching. The white box indicates the bottom electrode area and the Transistor and RRAM opening. 46

4.1 Material characterizations of the memory device. a) Schematic of the device structure. b) SEM image of the device lamella profile prepared by FIB showing all layers on the sidewall profile. c) Cross-sectional STEM micrograph of the device with a zoomed-in TEM image of the device showing several layers present on the device. d-i) EELS elemental mapping. 52

4.2 DC characteristics of the Li-infused TiO_x memory device. (a) A gradual change in I-V characteristics of the device upon sweeping five full voltage loops, (b) Change in conductivity as a function of increasing voltage, (c) Analogy between biological synapse and Li-ion imbued TiO_x synapse, where Li-induces oxygen vacancies are created due to the migration of Li-ion into TiO_x layer, which then forms a bond with an oxygen atom, leaving the TiO_x more oxygen deficient (reducing the local x), (d) I-V characteristics of ten 4 μm $\times$ 4 μm and ten 10 μm $\times$ 10 μm devices and conductance scales with the area, (e) Self-relaxation and voltage-assisted relaxation from the device (absolute value of the conductance was plotted for negative voltage), (f) I-V characteristics comparison of different control devices before forming, where the switching layers have the same thickness. 53

4.3 (a) Conductance changes as a function of sweep delay between applied voltage. (b) Gradual increase in conductance upon the applications of five consecutive positive sweeps (0 V - (+2 V) – 0 V) and decrease in conductance upon the applications of five consecutive negative sweeps (0 V - (-2V) – 0 V). (c) Current-voltage characteristics of 50 volatile devices distributed across a 1cm-by-1cm area. (d) Endurance test of the device with asymmetric voltage sweep. 54

4.4 Device behaviour upon the application of pulse voltage. (a) The device's response upon varying pulse amplitude and inset depicts the key parameters: V_a, pulse amplitude; t_p, pulse duration; t_d, pulse delay. (b) Device current upon varying pulse delay. (c) Continuous bias and its corresponding conductance. Inset shows the conductance decay at different excitation durations. 57

4.5 (a) Potentiation and Depression characteristics from the device upon the application of positive and negative voltage pulses. (b) Partial conductance decay when the pulse delay is 15ms. (c) Power-law fitting result of the conductance decay as a function of excitation duration. 58

4.6 (a) Temperature-dependent current-voltage characteristics (b) Resistance value at +1V for two different sweep conditions: one at up sweep and another is down sweep. (c) constant-voltage stress at different temperatures with an applied voltage of +1V. 60

4.7 (a) Relative values of the current changes at different temperatures @+1 V. (b) Slope above transition temperature. 61

4.8 (a) Li migration pathway in the O-defective structure of anatase TiO_2 using ionic radii style. (b) The energy barrier profile for Li migration obtained from DFT-NEB (Density Function Theory–Nudged Elastic Band) calculations shows a barrier of 0.4 eV. 61

4.9 The slope below the transition temperature. 62

4.10 (a) Forming process comparison of TiO_x and Li-imbued TiO_x memory devices, (b) Multi-state capability of the proposed device when stopped at different RESET stop voltages, which is denoted by V_s, (c) Retention behaviour of the 16 states from the memory device. 63

4.11 (a) Retention characteristics of the device after applying an incremental reset voltage sweep. With an increment of 0.01V, we can set 32 states easily with a margin of approximate 0.5 $\$mu$A. (b) Twenty Li-imbued TiOx devices IV characteristics, showing the device-to-device variations. 64

4.12 Voice recognition task using proposed memory. (a) Process flow of RC for voice recognition. (b) Comparison of classification accuracy between quantized RC model, RC model and model without RC processing. (c) Schematic of proposed 2T2R CIM array. 65

4.13 (a) Cochleagram of voice data, binarized data, and RC processed data. (b) Voltage input of channel–30 and corresponding current of the volatile device at time scale. 66

5.1 Relaxed spin configurations with (a) Exchange energy term only, (b) DMI energy only, and (c) both Exchange and DMI energy [This figure is constructed by using Mayavi [21]]. 75

5.2 Simulated spin configuration of a magnetic system with a fixed exchange energy and varying Dzyaloshinskii–Moriya interaction (DMI) energy (units are in mJm^{-2}). 76

5.3 Writing scheme on our proposed memory with (a) sense path for reading and programming path for writing (following the scheme of Ref. [22]), (b) Anisotropy Axis along with Write Line axis, and (c) Current pulse diagram for writing a complete cycle. [inset shows the proposed memory structure] 77

5.4 Spin configurations of our proposed memory device (a) when sweeping the magnetic field in clockwise (CW) and counterclockwise (CCW) direction and (b) when each corresponding configuration is relaxed without the magnetic field. (c) Schematic illustrations of the resistance states that correspond to the spin configurations without external magnetic field. 79

5.5 HMRAM crossbar array to realize im-memory computing architecture. . . 81

List of Tables

4.1 Performance parameter comparison between the existing literature and proposed device structure. 68

5.1 List of parameters values to realize the HMRAM 73

List of Abbreviations

1T-1R one transistor and one resistor 5, 33

2T-2R two transistor and two resistor 5, 33, 84

AI artificial intelligence 1

ANN artificial neural network 28, 48

BEOL backend of the line 6, 44, 46, 83, 84, 86

CMOS complementary metal-oxide semiconductor 4, 6, 10, 44, 83, 84

CVD chemical vapor deposition 34

DL deep learning 1

DMI Dzyaloshinskii–Moriya interaction xviii, 7, 71–76

DRAM dynamic random-access memory 3, 28, 83

EBL electron-beam lithography 41, 42, 45, 46

EELS electron energy loss spectroscopy 51

FFNN feedforward neural networks 28

FIBSEM Focused ion beam scanning electron microscopy 51

FPGA field programmable gate arrays 4, 31

GPU graphics processing unit 1

HMRAM helimagnet-based random-access memory 76, 81, 82, 85

HRS High resistance state 10, 12, 17, 52

IMC in-memory computing 4, 5, 28, 31, 82–85

IoT Internet of Things 1

IRDS International Roadmap for Devices and Systems 1

LRS Low resistance state 10, 12, 17, 18, 52

LTD long-term depression 26, 27

LTP long-term plasticity 24, 26, 27

MAC multiply-accumulate 3, 31

ML machine learning 1, 5

MLA Maskless Aligner 40–43

MNIST Modified National Institute of Standards and Technology 30

MVM Matrix-vector multiplication 5, 13, 31, 33

NC neuromorphic computing 4, 83

NN neural network 1

nT-nR n transistor and n resistor 5, 33

PECVD plasma-enhanced chemical vapor deposition 36, 43, 45

PPF paired-pulse facilitation 6, 24, 25, 57, 69, 84

PTP post-tetanic potentiation 6, 24, 25, 57, 69, 84

PVD physical vapor deposition 34

RC Reservoir computing 4, 28, 30, 31, 48, 49, 51, 65, 67, 84

RIE reactive ion etching 39, 40, 43, 46

RNN recurrent neural networks 28

RRAM resistive random-access memory 9

RS resistive switching xiii, xiv, 6, 10, 11, 14, 16, 17

SEM scanning electron microscope 46

SIMS secondary ion mass spectroscopy 43

SRAM static random-access memory 28

SRDP spike-rate dependent plasticity 6, 24, 59, 69, 84

STDP spike-timing-dependent plasticity 27

STP short-term plasticity 6, 24, 26

TEM transmission electron microscope 51, 85

List of Symbols

CO_2 Carbon Dioxide 1

E_a Activation Energy 20

e Electron Charge 41

h Planck's Constant 41

k_B Boltzmann Constant 20

λ Wavelength 41

LiPON Lithium Phosphorus Oxynitride x, xiv, 6, 22

Li Lithium xiv, 6, 19–21, 83, 84

m_0 Electron Mass 41

σ Conductivity 20

τ Decay Constant 59

TiO_x Titanium Oxide x, 6, 22, 23, 83, 84, 86

T Temperature 20

V Acceleration Voltage 41

Chapter 1

Introduction

1.1 Motivation

The recent report published by the International Roadmap for Devices and Systems (IRDS) highlights the upcoming application pulls, which include big data, Internet of Things (IoT), deep learning (DL), machine learning (ML), artificial intelligence (AI), exascale supercomputing, and robotics and autonomous Systems [23]. These emerging applications increase the demand for computing power dramatically, and the demand is projected to outperform Moore's law. As can be seen in Figure 1.1(a), the computing demand is now predicted to rise two-fold every two months due to the constant improvement of graphics processing unit (GPU) architectures, though with an expense of an order of magnitude more power [1]. With the projected pace of computing demand, it is highly unlikely that conventional computing solutions could alone be sufficient. Moreover, the cost associated with the training of the current AI hardware is shockingly high, as shown in Figure 1.1(b). The training of the AI hardware requires burning a lot of fossil fuel-based energy that increases carbon dioxide ($\mathbf{CO_2}$) emissions to the atmosphere, causing global warming. It has been reported that training a single neural network (NN) model with 213 million parameters could produce as much $\mathbf{CO_2}$ as five cars produce in their entire lifetimes [24]. Bottlemecked by high cost and environmental pollution, the existing computing paradigm is in dire need of an alternative computing approach.

Furthermore, to realize a computing operation in existing computing architecture, data needs to be fetched from memory to the processor and upon completion, output data is sent back to memory. The typical addition and multiplication operation does not consume much energy, although fetching data to and from external memory consumes a lot of energy.

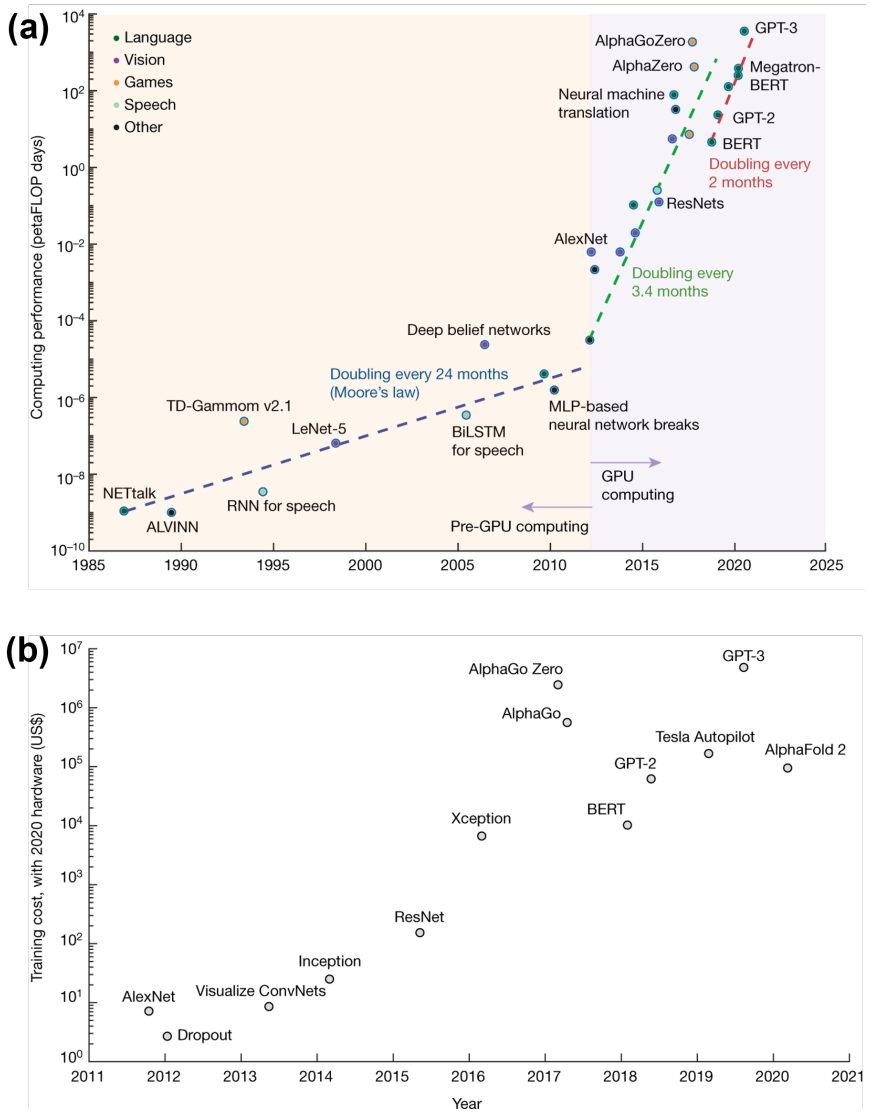


Figure 1.1: (a) Current computing power demands and future computing demands over the past forty years with different applications domain. The computing demand before 2012 doubled every two years, whereas current computing demand doubled every two months. (b) An exponential increase in the cost of AI hardware training would be unsustainable in the long run [1]. Copyright © 2022 Springer Nature.

It has been found that dynamic random-access memory (DRAM) access consumes three orders of magnitude energy compared to multiply-accumulate (MAC) operation energy in a hardware accelerator using conventional von Neumann architecture [25]. This problem is known as the Memory wall. According to Moore's law, the scaling of the transistor should be doubled every two years. However, scaling reaches its fundamental physical limits, and further scaling down is way more expensive, with its characteristics unreliable, putting another obstacle in computing known as the scaling wall. Although transistor size scales down at a fast pace, the operating voltage does not scale down as proportionally, which eventually increases power density and results in a power wall or heat wall in the road of future computing. Figure 1.2 illustrates the existing computing system's constraints, including memory wall, heat wall, and scaling wall [2]. With all the above-mentioned constraints, searching for an alternative energy-efficient computing approach is necessary.

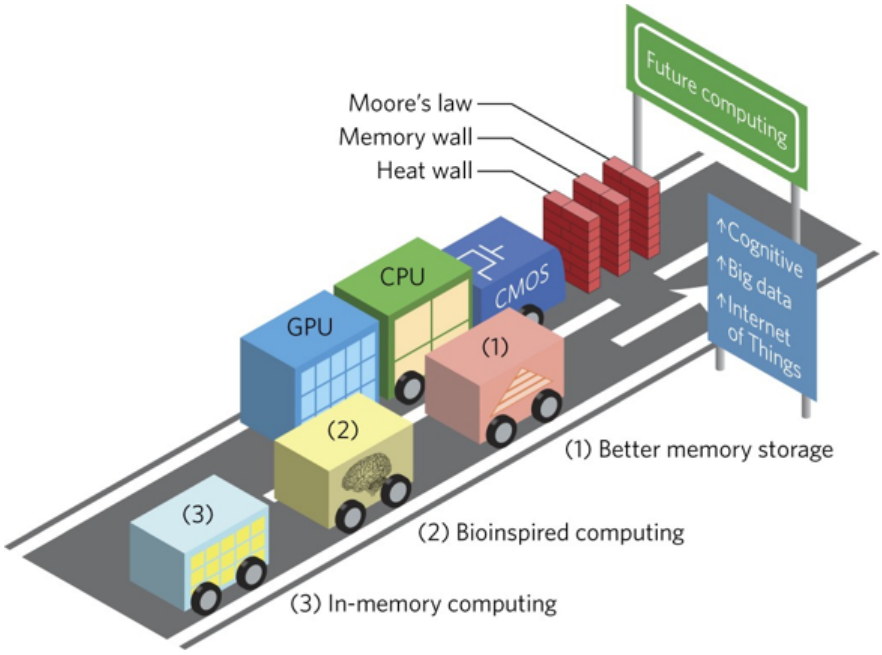


Figure 1.2: Constraints of the existing computing schemes, including heat wall, memory wall, and scaling wall. Alternating memristor-based non-von Neumann computing paradigms, which include bio-inspired computing, in-memory computing, and hybrid memory and logic integration, could overcome those barriers [2]. Copyright © 2018 Springer Nature.

Alternative computing architectures, including bioinspired or neuromorphic computing (NC) and in-memory computing (IMC), have been investigated by researchers, as can be seen from Figure 1.2. As computation and memory are co-localized in those architectures, they eliminate the data movement, resulting in highly energy-efficient and fast-speed computing. In NC, the computation happens in a network of synapses and neurons, similar to the connection of the biological brain. Each neuron is connected with thousands of other neurons via synapses, and when thousands of synaptic weights arrive at the neuron terminal, the neuron activates and fires the action potentials to the next neurons. Recently, Intel and IBM developed Loihi and TrueNorth technology that implements NC at the hardware level. The function of the synapses and neurons is designed using complementary metal-oxide semiconductor (CMOS) circuits. To mimic a single neuron with thousands of synapses, the neuromorphic chips require about sixteen thousand transistors, rendering the feasibility of large-scale systems difficult and less efficient [26].

In contrast, Memristors-based devices could mimic the synaptic plasticity in a single two-terminal-like device structure, allowing them to be a suitable candidate for NC. Memristive devices are defined as devices whose conductance can be tuned to two or more stable states upon the application of electrical stresses. Also, based on the retention measurement results, memristor resistance states can be volatile or non-volatile. The volatile and non-volatile natures of memristors resemble the short-term and long-term memory of the biological brain. Reservoir computing (RC), which is a type of NC architecture, consists of a reservoir layer and an output layer. In RC architecture, the training updates only at the output layer, thus bypassing the interior complexity and reducing the number of training parameters. The volatile memristors can act as a reservoir of the RC system, whereas the output layer can benefit from non-volatile elements to retain the training weights.

Previous implementations of the RC scheme were realized by fabricating separate devices for the reservoir and output layer, respectively, as the reservoir layer needs a highly nonlinear and dynamic system, for which volatile memory was utilized, and the output layer needs to retain training weights for a long time, for which non-volatile memory was used. As the devices' materials stack are different, the fabrication cost of the devices was expensive, and they required larger areas and/or additional metal layers. Therefore, a universal memory device that can be configured as either volatile or non-volatile is of utmost importance to reduce footprints and latency and to realize low-cost RC systems with customizable layouts on a universal platform. The customization can be tailored/optimized towards specific applications, which is similar to the concepts of field programmable gate arrays (FPGA) but geared towards non-digital realizations. Therefore, the first objective focuses on the material stacks that can be configured to volatile or non-volatile memory.

Another path to overcome the barrier of von Neumann architecture is IMC architec-

ture, in which computation occurs inside the memory device, obviating the data transfer delay and energy needed to transfer the data between memory and processing elements. Memristor-based IMC architecture is preferable because of its multi-bit capability, energy efficiency, and non-volatile nature, meaning the weight can be remembered even without a power supply. Matrix-vector multiplication (MVM) is considered the most fundamental mathematical operation in numerous applications, including ML algorithms, digital signal processing, and solving non-linear equations. The prime advantage of memristive IMC is the massive parallelism of executing MVM using a memristive crossbar array. There are several memristive devices that can be used to encode multiple bits of information in a non-volatile manner, which include resistive random-access memory (RRAM), phase change memory (PCM), magnetic random-access memory (MRAM), and ferroelectric random-access memory (FeRAM). While RRAM, PCM, and MRAM rely on the resistance changes due to the resistive switching (RRAM) or phase changing of a material itself (PCM) or spin-dependent transport across a material stack (MRAM), FeRAM depends on the polarization change instead of resistance inside the ferroelectric materials. The charge can be found by integrating the current with respect to the voltage sweep. In order to change the resistance of the ferroelectric material, a ferroelectric field-effect transistor (FeFET), consisting of three terminals, configuration is needed [27]. In FeFET, the resistance change is caused by the change in dielectric polarization inside the channel; therefore, it can be used in IMC architecture [28]. However, the three-terminal FeFET structure is the major bottleneck. In contrast, RRAM-based IMC is two-terminal and easy to integrate with CMOS to scale up the technology.

To represent the synaptic weight appropriately, memristor devices need to be capable of storing multi-bit information with non-volatility. Typically, based on how many bits a memristive device can store, different crossbar architectures have been proposed. Based on the number of states and transistor configurations, one transistor and one resistor (1T-1R), two transistor and two resistor (2T-2R), and n transistor and n resistor (nT-nR) structures have been proposed to accelerate the MVM operation. The challenge is to encode as many states as possible in memristive devices. In the case of limited conductance states, memristor devices with multiple transistors can enhance the resolution. This thesis uses a 2T-2R structure that can efficiently store multi-bit and accelerate the MVM operation. Also, helimagnet-based resistive memory that is capable of storing multi-bit information is proposed in this research.

1.2 Thesis Organization

A brief summary of the rest of the thesis chapters is provided below.

Chapter 2 presents some of the fundamental concepts related to memristive devices and emerging computing architecture using memristors. This chapter briefly explains the memristive device structure and materials, RS types, including bipolar-type and unipolar-type switching, and volatile and non-volatile memristive devices. An explanation of two types of mechanisms, namely filamentary and interfacial resistive switching, is provided afterwards. Next, conduction mechanisms are explained. A brief overview of solid-state electrolytes that can be used to controllably transport lithium ions in and out of the electrode will be described. Moreover, how synapses, a crucial element of the human brain, can be emulated using memristors will be explained. Finally, a brief description of memristive emerging computing architectures, such as reservoir computing and in-memory computing, is presented.

Chapter 3 briefly describes three basic components of micro- and nano-fabrication, including lithography, deposition, and etching processes. At first, thin-film deposition techniques, such as sputtering and plasma-enhanced chemical vapor deposition, are presented. Next, dry etching techniques, such as ion milling and reactive ion etching, are covered. The most important lithography techniques, for example, photolithography using MLA and e-beam lithography techniques, are briefly explained thereafter. Moreover, a description of the whole fabrication process of standalone memory devices and the backend of the line (BEOL) integration process of the memory device on top of tape-out CMOS chips will be presented.

Chapter 4 demonstrates a lithium (**Li**) imbued $\mathbf{TiO_x}$ iontronic device that exhibits synapse-like short-term plasticity (STP) behaviour. A solid-state electrolyte lithium phosphorus oxynitride (**LiPON**) behaves as the ion source, and the embedding and releasing of **Li** ions inside the cathodic like $\mathbf{TiO_x}$ renders volatile conductance responses from the device. Besides, these devices possess high uniformity and great endurance as no conductive filaments are present. Different short-term pulse-based phenomena, including paired-pulse facilitation (PPF), post-tetanic potentiation (PTP), and spike-rate dependent plasticity (SRDP), were observed with self-relaxation characteristics. Based on the voltage excitation period, the timescale of the volatile memory can be tuned. Temperature measurement reveals the ion displacement-induced conductance channels become frozen below 220 K. In addition, the volatile analog devices can be configured into non-volatile memory units

with multibit storage capabilities after an electroforming process. Therefore, on the same platform, we can configure tunable volatile units as nonlinear dynamic reservoirs for performing neuromorphic training and the non-volatile units as the weight storage layer. We proceed to use voice recognition as an example with the tunable time constant relationship and obtain 94.4% accuracy with a minimal training dataset. Thus, this iontronic platform can effectively process and update temporal information for reservoir and neuromorphic computing paradigms.

Chapter 5 proposes a design of a helimagnet-based emerging memory device that is capable of storing multiple bits of information per device. The device consists of a helimagnet layer placed between two ferromagnetic layers, which allows us to lock in specific spin configurations. We begin by finding the relaxed spin structure, which is the result of the competition between the DMI and exchange energy and is referred to as the equilibrium state ("0"). The writing of a memory state is simulated by applying an in-plane field that rotates and transforms the spin configurations of the memory device. Our results indicate that stable configurations can be achieved at rotations of an integer multiple of 180 degrees (corresponding to states "-2", "-1", "1", "2", etc.), where the anisotropy stabilizes the free layer and thus the exchange coupled helimagnet. These states are separated by magnetic energy barriers, and intermediate, unstable spin configurations tend to revert to their adjacent states. By simply changing the direction of the field, we can achieve multibit data storage per unit memory cell. The maximum number of bits is reached when the anisotropy energy barriers cannot withstand the strong DMI energy. Reading can be done by evaluating the different resistance states due to the twisted spin texture.

Chapter 6 summarizes the contributions of this thesis and outlines the future directions of the research.

Chapter 2

Background

Overview

This chapter describes some of the fundamental concepts related to memristive devices and emerging computing architecture using memristors. Starting from memristive device structure and materials, this chapter briefly explains the resistive switching types, including bipolar-type and unipolar-type switching. Based on the memory retention capabilities, RS can fall into either volatile or non-volatile memristive devices. An explanation of two types of mechanisms, namely filamentary and interfacial resistive switching, which are widely accepted as switching mechanisms, is provided afterwards. Next, conduction mechanisms are also explained. A brief overview of solid-state electrolytes that can be used to controllably transport lithium ions in and out of the electrode will be described. Moreover, how synapses, a crucial element of the human brain, can be emulated using memristors will be explained. Finally, a brief description of memristive emerging computing architectures, such as reservoir computing and in-memory computing, is presented.

2.1 Memristive Device

A memristive device, also known as a memristor, is a device that has both finite resistance and intrinsic memory. The memristor device theory was first proposed by Prof. Leon Chua in 1971 [29], and the experimental evidence was realized in 2008 by the Hewlett Packard (HP) research group [30]. Since then, researchers have extensively investigated memristors due to their simple two-terminal structure, fast operational speed, ultra-low power

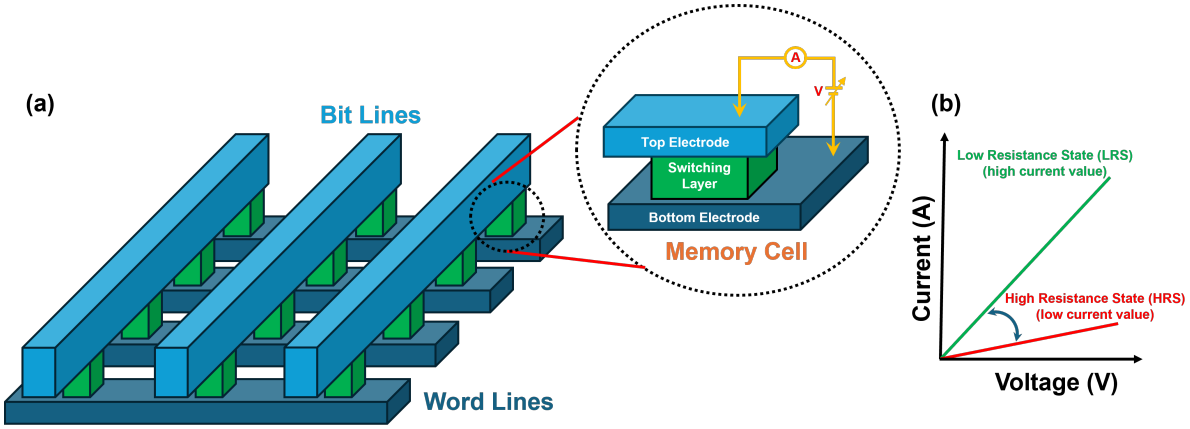


Figure 2.1: (a) Memristive crossbar array with bit lines that are used to read/write data and word lines that are used to select a memristive device. The inset shows a unit memristive device consisting of two metal electrodes and a metal oxide switching layer. (b) Two different stable resistance states of a single memory cell.

consumption, and synapses and neuron emulation capability for neuromorphic computing applications.

2.1.1 Device Layout and Materials

A memristive device typically consists of a switching layer (usually metal oxide) sandwiched between two metal electrodes, as shown in the inset of Figure 2.1(a). Due to its simple two-terminal-like structure, a large scalable memristive crossbar array can be easily constructed, as shown in Figure 2.1(a). This array is promising for building a large neural network and neuromorphic computing applications [31]. This crossbar structure also allows one to randomly select any cell within the crossbar, which is why this crossbar is also known as resistive random-access memory (RRAM).

The electrode in memristive devices is not just a mere contact; it plays a critical role in the resistive switching process of memristive devices. The electrode materials also govern the switching mechanism. The lists of electrode materials include metal electrodes (Ag, Au, Cu, Pt, W), nitrides (TiN), graphene, conductive oxides (ITO and $SrRuO_3$), and p- and n-type Si. Similar to the electrodes, switching layer materials, typically an insulator or a semiconductor, are also crucial as switching takes place in this layer. Metal oxides such as TiO_x, TaO_x, HfO_x, and WO_x or chalcogenide materials are typically used as switching

layers in the cell. Recently switching layers, including nanowires, two-dimensional (2D) materials, transition metal chalcogenides (TMDs), and nanoparticles, have attracted a great deal of attention from researchers. This thesis will focus on the nanoscale thin-film type switching layer because of its fast switching speed, great retention and endurance characteristics, and, more importantly, CMOS compatibility.

2.1.2 Resistive Switching Types

A typical digital memristive cell contains two resistance states: (a) High resistance state (HRS) and (b) Low resistance state (LRS). Figure 2.1(b) represents these two resistance states with two linear slopes within a small voltage window. Switching between these two resistance states is the most crucial step in resistive memory. By applying an external electrical bias, the resistance of the memory device can be altered between HRS and LRS. The SET process, also known as "write", is defined as the transition from HRS to LRS, whereas the RESET process, also known as "erase", is described as the transition from LRS to HRS. Based on the switching dynamics, memristive devices can be divided into two categories: (a) digital-type and (b) analog-type. In digital memristors, conductance changes occur abruptly, whereas in analog memristors, conductance changes are gradual. The former is useful for storage applications, and the latter is suitable for artificial synaptic devices.

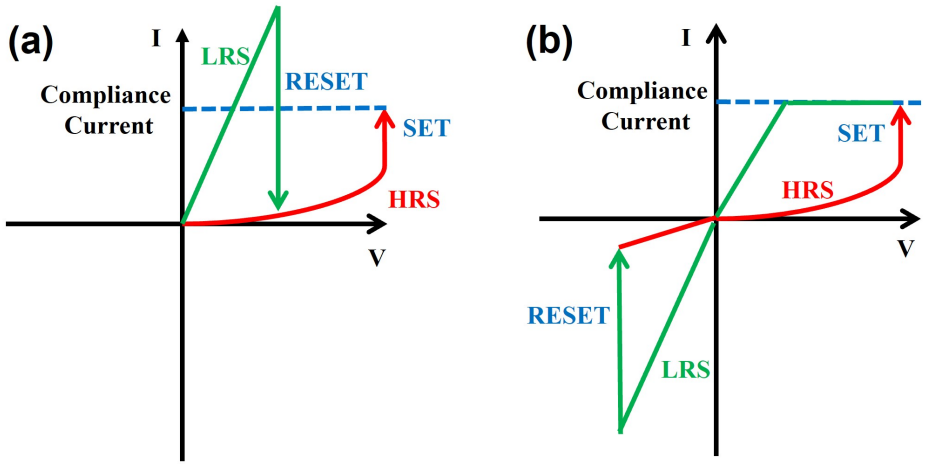


Figure 2.2: Current-Voltage (IV) characteristics of (a) an unipolar type RS. (b) a bipolar type RS.

Based on SET and RESET's voltage polarity, RS can be divided into two categories: (a) Unipolar and (b) Bipolar. In unipolar type RS, resistive switching is independent of the voltage polarity. So, both Set and Reset occur at the same voltage polarity. In bipolar resistive switching, the voltage polarity of the RESET is opposite from that of the SET process. The current-voltage characteristics of a unipolar switching mode are illustrated in Figure 2.2(a), and those of a bipolar switching mode are shown in Figure 2.2(b). For both Unipolar and Bipolar operation modes, compliance current is employed in the SET process to eliminate the permanent breakdown of the resistive memory cell. A small biasing voltage (about 0.1 V or 0.2 V) is applied without affecting the memory state for a typical read operation to check the current state of the memory, which is known as a memory read operation. A continuous read for a certain time interval is known as a "retention test".

Volatile Memory Device

A memory device that forgets its state when the external power supply is removed is known as a volatile memory device. In a memristive volatile device, after the voltage sweep (0 V - [$> V_{th}$] - 0 V), the retention measurement from the device exhibits conductance decay. Based on the IV characteristics, volatile devices can be divided into two categories: (a) Analog-type volatile memory and (b) Threshold-type volatile memory. Typical current-voltage characteristics of analog-type and threshold-type are shown in Figure 2.3(a) and Figure 2.3(b), respectively. Both characteristics are suitable to emulate the behaviour of artificial synapses and neurons. Various switching materials, including WO_x, Nb-$SrTiO_3$, and $Pr_{0.7}Ca_{0.3}MnO_3$ (PCMO), showed spontaneous conductance decay. The decay characteristics are more important to realize the short-term and long-term plasticity of biological synapses.

Figure 2.3(c) shows the retention measurement after voltage sweep in an analog-type memristor device, Pd/WO_x/W [3]. As can be seen from the figure, the conductance decay has two different slopes, indicating two different time scales. Initially, the decay is faster and later, the decay is slower. After fitting using two stretched exponential functions, the time constant of fast decay was found to be around 52.5 ms, and that of slow decay was found to be around 92.5 ms.

Similarly, threshold switching characteristics show volatile resistive switching. Wang *et al.* demonstrated a diffusive threshold memory based on SiO_xN_y:Ag, shown in Figure 2.3(d) [4]. The pulse test revealed that conductance can increase when a voltage pulse is applied, and upon removal of the voltage pulse, the conductance relaxes back to its initial conductance. As can be seen from the figure, after a certain delay, the current starts rising, which is known as delay time. Similarly, the current does not instantly reach

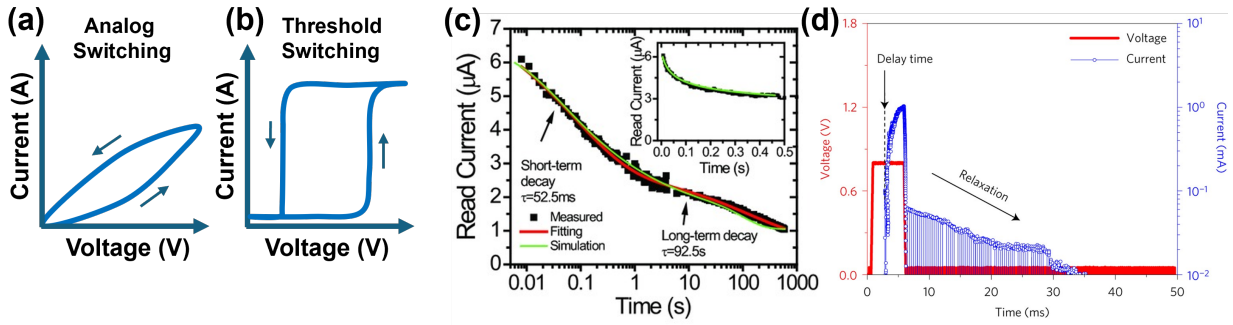


Figure 2.3: Volatile resistive switching types: (a) analog-type switching and (b) digital-type switching. (c) Conductance decay in analog-type volatile resistive switching ($Pd/WO_x/W$) after the removal of voltage excitation [3]. Copyright © 2015 John Wiley and Sons. (d) Conductance decay in threshold-type resistive switching device (Au/SiO_xN_y:Ag/Au) [4]. Copyright © 2016 Springer Nature.

the initial conductance value (before applying bias). The time required to reach the initial conductance value is known as relaxation time. Interestingly, the non-linear volatile current-voltage characteristics can be utilized as a reservoir for neuromorphic computing applications.

Non-volatile Memory Device

A non-volatile memory is a type of memory device that can remember its state even after the removal of an external power supply. In non-volatile memory, the retention measurement from the device demonstrates almost no change in conductance value over time. Similar to volatile memory, non-volatile memory can be classified into bipolar digital memory and analog-type non-volatile memory. In bipolar digital memory, the device conductance abruptly changes into LRS from HRS once the voltage reaches the SET threshold value and HRS from LRS when the RESET value reaches the threshold value, as shown in Figure 2.4(a). One of the conditions for bipolar digital memory is that the value of SET and RESET threshold voltages should be opposite. Pulse characteristics of bipolar digital device, shown in Figure 2.4(b), also show abrupt changes in conductance even for short-duration pulses. This type of resistance switching is preferable in data storage technology because of its long retention, fast switching speed, and high on/off ratio. Recent demonstrations suggest that digital switching can be used to construct artificial synapses. The lack of tunability in multiple conductance states and nonlinear conductance or weight

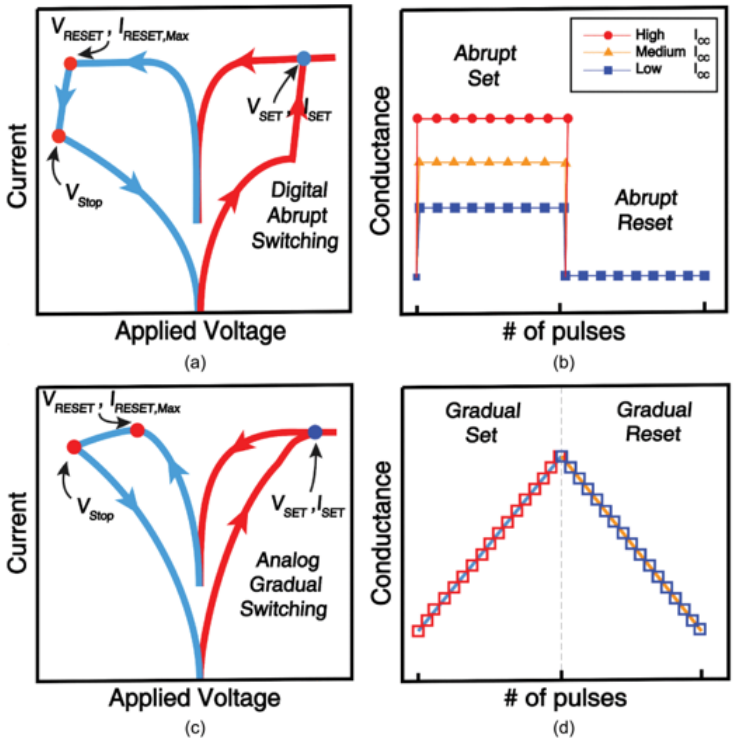


Figure 2.4: Electrical characteristics comparison between bipolar digital and analog-type non-volatile memory. (a) DC I–V characteristics and (b) Change in conductance upon the applications of pulse of a bipolar digital non-volatile memory device. The bipolar digital device exhibits a sharp change in conductance. (c) DC I–V characteristics and (d) pulse operation of an analog-type non-volatile resistive memory device. The analog-type non-volatile memory device exhibits gradual switching [5]. Copyright © 2020 IEEE.

updates of bipolar digital memory, however, hinder its application in neuromorphic computing.

In contrast to bipolar digital memory, analog-type non-volatile memory exhibits a gradual change in conduction for both DC I–V characteristics and pulse I–V characteristics, as shown in Figure 2.4(c) & Figure 2.4(d). Analog–type non-volatile memory has the capability to tune the conductance into multiple states and the gradually increasing conductance update capability, making them suitable for hardware accelerators for neural networks and brain-inspired neuromorphic computing applications. The core of the artificial neural network is the MVM operation, which can be realized in a crossbar of analog non-volatile

resistive memory, where Ohm's law is used to multiply, and Kirchhoff's current law is used to accumulate the sum of the product. The details of this operation are described in the memristive in-memory computing section 2.6.2.

2.2 Resistive Switching Mechanisms

After HP labs' experimental breakthrough, the mechanism behind resistive switching has been meticulously investigated to obtain excellent resistive switching performance that could potentially be commercialized into a product. In applications like artificial synapses and neurons, better control over conductance modulation and quick response to input pulse stimuli are essential. As a result, a deeper understanding of the resistive switching mechanisms becomes crucial to generate more breakthroughs. Based on the current knowledge of the memristor device, the switching mechanism can be classified into two categories: filament-type RS and interface-type RS.

2.2.1 Filament-type RS

In filament-type resistive switching, conduction happens through the filament, which is formed during the first resistive switching process (also known as the forming process). The ions inside the switching layers start migrating to the opposite polarity electrode due to a large electric field concentrated on a thin film. Once the filament is formed, the forming process is done. A partial rupture of the filament causes RESET of the device. This partial rupture can be compensated again by the SET process, but now SET voltage is much smaller as only smaller portions need to form. One way to characterize whether the switching is filamentary or not is to check the area-dependent conductance. The conductance is less dependent on the area in the case of filamentary RS. There are two types of filamentary RS devices: anion-type and cation-type filamentary RS. Details of both devices are given below.

Cation-type Filamentary RS

Cation-type filamentary RS is known as conductive bridge resistive random-access memory (CBRAM). This memory system depends on the electrochemical metallic filament formation of an active electrode. The top electrode needs to be made of electrically active materials such as Ag, Cu or Ni, whereas the bottom electrode must be made of electrically

inert materials, such as Pt, Ir, or W. The insulator or electrolyte can be oxide, halide or chalcogenide material. Upon the application of a positive voltage at the top electrode like Ag or Cu, it dissolves in the electrolyte due to the high electrical field and is converted into Ag^+ ion, and then accumulated near the bottom electrode, as shown in Figure 2.5. Thus, gradually a metallic channel is created, and resistance starts decreasing while the current is increasing. This channel will not rupture until a very high reverse polarity voltage is applied. Upon the application of a negative voltage, the metallic filament is dissolute, and the channel is broken, leaving the device in a high resistance state. Figure 2.5 illustrates the whole process of filament formation and rupture of CBRAM.

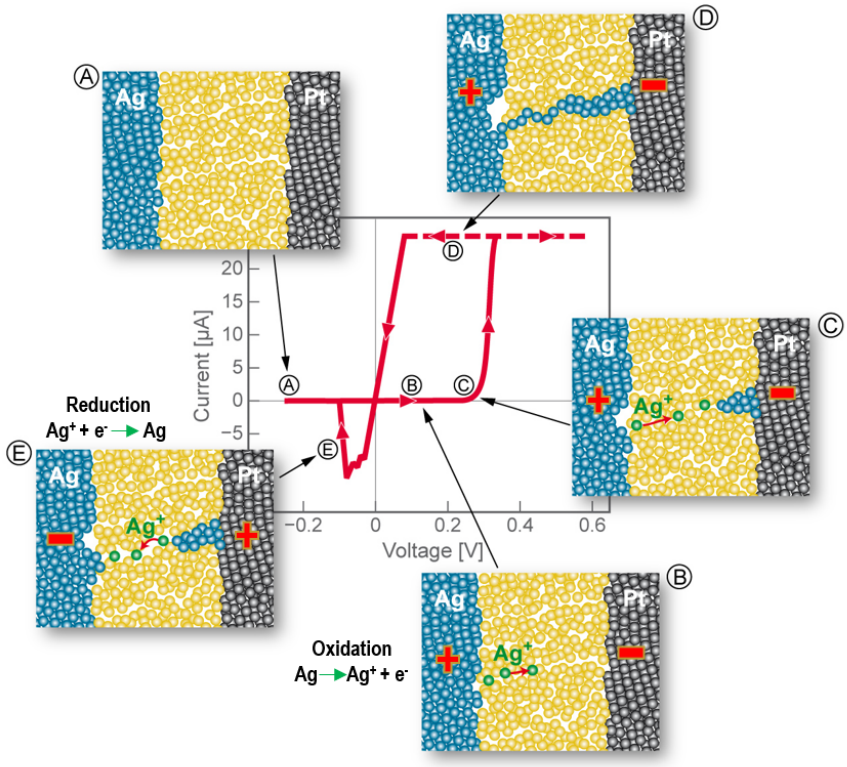


Figure 2.5: Schematic illustration of a conductive bridge resistive random-access memory [6]. Copyright © 2011 IOP Publishing.

Anion-type Filamentary RS

Unlike metal ions in cation-type filamentary RS, anions, usually oxygen ions, participate in anion-type filamentary RS. The switching layer is usually transition metal oxides and electrodes do not need to be electrochemically active. The switching mechanism is governed by the creation and annihilation of oxygen vacancies.

Figure 2.6 shows the filament formation and rupture process of Anion-type Filamentary RS memory. At first, as a result of a large applied voltage on a thin oxide layer at the top electrode, a highly localized field is created, causing the localized bound oxygen atom to escape and form an oxygen ion while leaving behind an oxygen vacancy. The process is very similar to electron-hole pair generation; instead, we have oxygen ion and oxygen

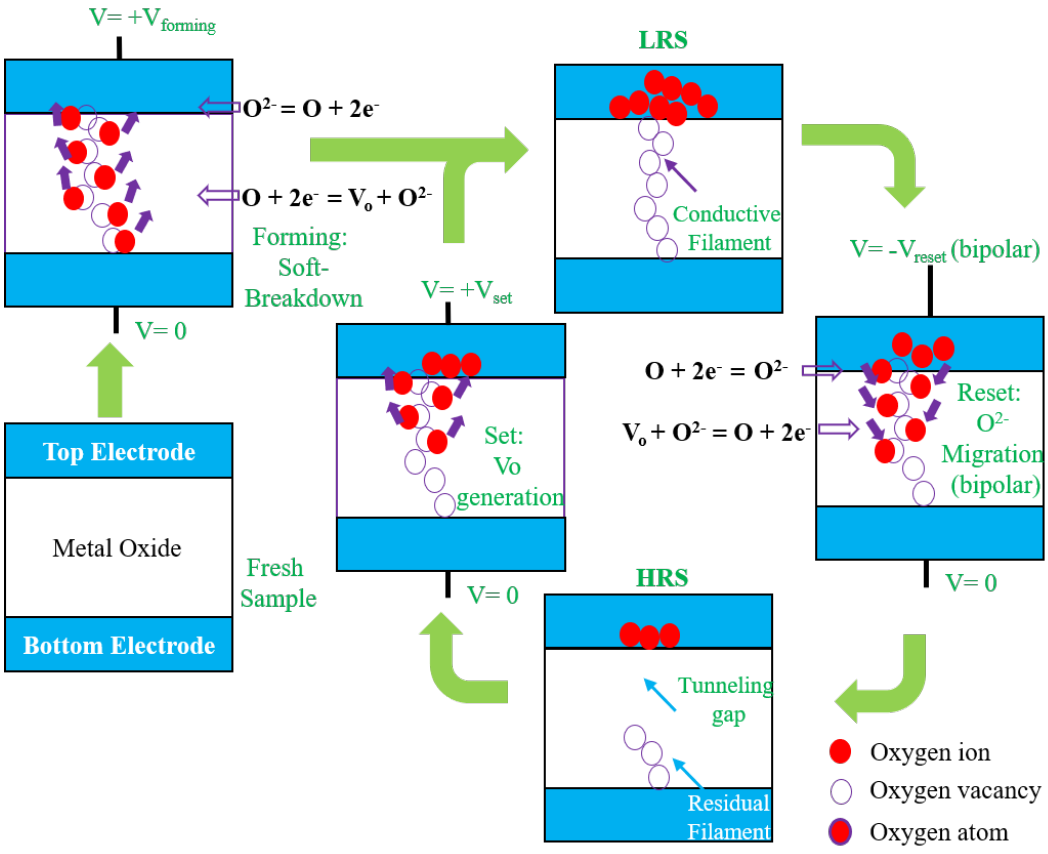


Figure 2.6: Schematic illustration of an oxygen vacancy based RS [7].

vacancy in this type of RS. Since the ion is a charged particle, the oxygen ion can move with the applied bias. Upon implementing positive bias at the top electrode, oxygen ions (since oxygen ion, O^{2^-}, is negative) will migrate under the top electrode and be discharged at the interface of the electrode and oxide layer. Simultaneously, immobile oxygen vacancy acts as a conductive path for the electron, eventually constructing the filament formation process. Now, current can flow through the conductive filament since electrons can hop through these vacancies, which are very similar to traps. Thus, we obtain the LRS because the current can quickly go through the filaments. For the reset process or HRS, we need to rupture the filament. By applying a negative voltage at the top electrode, the accumulated oxygen vacancies at the electrode and oxide layers are again driven back to the oxide layer. These ions then recombine with the vacancy and the filament path ruptures. Ion migration results in a gap between oxide and metal, and due to the low probability of tunnelling, the current will be significantly less, and we achieve HRS. In the next cycle, we can set the resistance state at LRS by applying a smaller voltage since we do not need to form the filament at the whole oxide length.

2.2.2 Interface-type RS

In interface-type resistive switching, migration of oxygen vacancies causes a change in the Schottky barrier height. The conductance is proportional to the area of the electrode, indicating that the entire area of the electrode is participating in the conduction. Therefore, this is forming-free and suitable for multilevel conduction. Since no forming is needed, devices could easily achieve device-to-device uniformity and excellent reproducibility. The interface-type RS was first demonstrated by Sawa [32]. Rui Wang *et al.* demonstrate an interface-type RS device with Pt/WO_{3-x}/Pt, where oxygen vacancies concentration changes the conductivity of WO_{3-x} layer [33]. Similarly, oxygen vacancies change the conductivity in amorphous InGaZnO [34] and Nb-doped $SrTiO_3$ [35].

A schematic of the switching mechanism in interface-type RS is shown in Figure 2.7, where a metal and n-type semiconductor were used as an example. When a negative bias is applied at the metal layer (Figure 2.7(a)), oxygen vacancies are attracted towards the metal electrode and narrow the depletion width, resulting in the LRS of the device. Now, a positive bias in the metal electrode would repel the vacancies away from the interface, resulting in a wider depletion width, as shown in Figure 2.7(b). Thus, the device switches back to HRS.

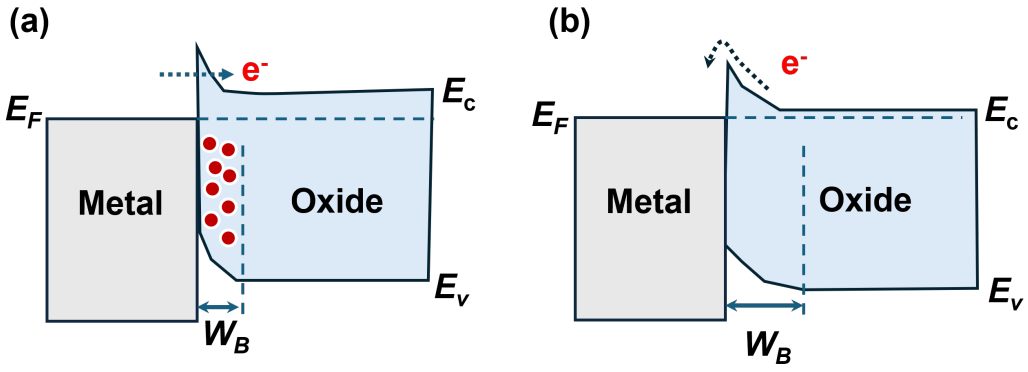


Figure 2.7: The switching mechanism of interface-type memristors. (a) During the LRS transition, vacancies accumulate near the metal-oxide interface, decreasing the depletion width. (b) During the HRS transition, the depletion width increases due to fewer defects at the interface [8].

2.3 Conduction Mechanism

An image of various conduction mechanisms from cathode to anode is shown in Figure 2.8. Conduction mechanisms can be divided into two parts: electrode-limited and bulk-limited. The conduction mechanism changes by the insulator properties (trap or bandgap energy level), the device fabrication process (annealing temperature or ambient condition during annealing), and the interface properties between oxides and contact electrodes (barrier height). After the set process, the conduction process is mostly ohmic for large operating current(200 μA). A larger current in LRS indicates the presence of many defects in the filaments, which ultimately help to achieve tighter conductance distribution and longer retention. In contrast, scaling of LRS switching current results in fewer defects in the filaments, which eventually broaden the conductance distribution and lower the retention [36].

Typically, the internal conduction mechanisms can be inferred by fitting the experimental current-voltage characteristics data. In addition, from the basics of the tunnelling phenomenon, one can neglect direct tunnelling if the oxide thickness is greater than 5 nm. FN tunnelling is negligible at low bias to medium bias, while FN tunnelling will have a more significant influence at high bias.

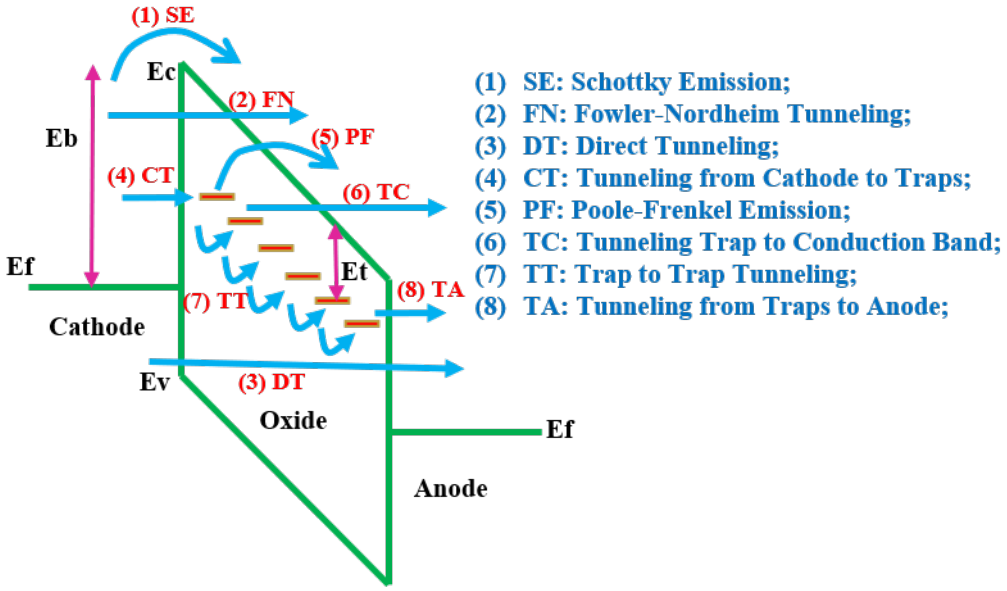


Figure 2.8: Probable conduction mechanism picture in a metal-insulator-metal system [9].

2.4 Battery-like Memristors

So far, memristors based on a metal oxide switching layer have been discussed. Those devices show either volatile or non-volatile memory characteristics. It is important to note that controllable doping in the switching layer can drastically improve the device's performance. One of the ways of controllable doping can stem from all thin-film solid-state battery-like structures, where controllable lithium ion intercalation and deintercalation cause the charging and discharging of the battery. Ion intercalation/deintercalation can be defined as the insertion/extraction of a foreign ion, for example, lithium or silver, within a host material. Guofei *et al.* demonstrated a "battery-like" tunnel junction, where reversible Li-ion motion changes the conductivity [37]. Besides, solid-state electrolytes can readily overcome two barriers posed by existing organic or liquid electrolytes: flammability and limited electrochemical stability. Among the different solid-state electrolytes, lithium-ion-based solid-state electrolytes are especially promising because of their smaller size with high ionic conductivity. Therefore, the integration of lithium-based solid-state electrolytes with the existing metal oxide switching layer defined the battery-like memristors, and this thesis addresses how the battery-like memristors will perform.

2.4.1 Ionic Conductivity and LiPON

Lithium-ion transport mechanism in solid-state electrolytes is governed by defects. There are several types of defects, such as point defects, planar defects, line defects, electron defects and volume defects. Compared to any other kinds of defects, point defects greatly influence the ion transport in crystals. In 1926, Frenkel first introduced the idea of point defects. The schematics in Figure 2.9(a) illustrate different point defect models, including Schottky and Frenkel defects. In Schottky defects, an anion vacancy is accompanied by a cation vacancy, whereas in Frenkel defects, a vacancy is accompanied by an interstitial ion. Ions are incapable of escaping their host sites in perfectly ordered crystals. Ionic current or transport occurs when the point defects are in motion in a crystal. Ionic conductivity is also governed by the diffusion pathway, as shown in Figure 2.9(b) and (c). According to the figure, the migration barrier or activation energy is different in different closed-pack frameworks.

In solid-state electrolytes, there are three dominant diffusion mechanisms, including vacancy-mediated, interstitial-mediated, and interstitial–substitutional exchange-mediated, as shown in Figure 2.9(d). The vacancy mechanism typically depends on Schottky defects, creating numerous vacancies and facilitating ion hopping within the crystal. Once a **Li**-ion finishes hopping, a new vacancy is created in the original site, enabling lithium ions to diffuse and perpetuate the mechanism. In interstitial-mediated transport, interstitial ions travel through the Frenkel defects. Lithium ions migrate in the interstices between molecular structures by continually substituting Li ions at nearby available sites. The Li-ion transport in a ceramic-like solid-electrolyte is governed by three factors: the type of the carrier, ionic diffusion pathway, and type of diffusion. In ceramic crystal structures, point defects govern both the charge carrier types and carrier concentrations that directly influence the ionic conductivity. Moreover, during transport, lithium ions interact with each other and their surroundings, which significantly impacts ionic conductivity.

Ionic conductivity also depends on the activation energy, meaning how easily an ion can migrate to a neighbouring site. Because of the exponential relationship between conductivity and activation energy, understanding of activation energy is of paramount importance. The activation energies can be deduced from Arrhenius's expression, expressed as,

$$\sigma = \frac{A}{T} * exp(-\frac{E_a}{k_B T}) \tag{2.1}$$

where σ is the conductivity at temperature, T (K), E_a is the activation energy, k_B is the Boltzmann's constant, A is the pre-exponential constant. The activation energy, E_a, can be calculated from the slope of $ln(\sigma T)$ versus $1/T$ plot.

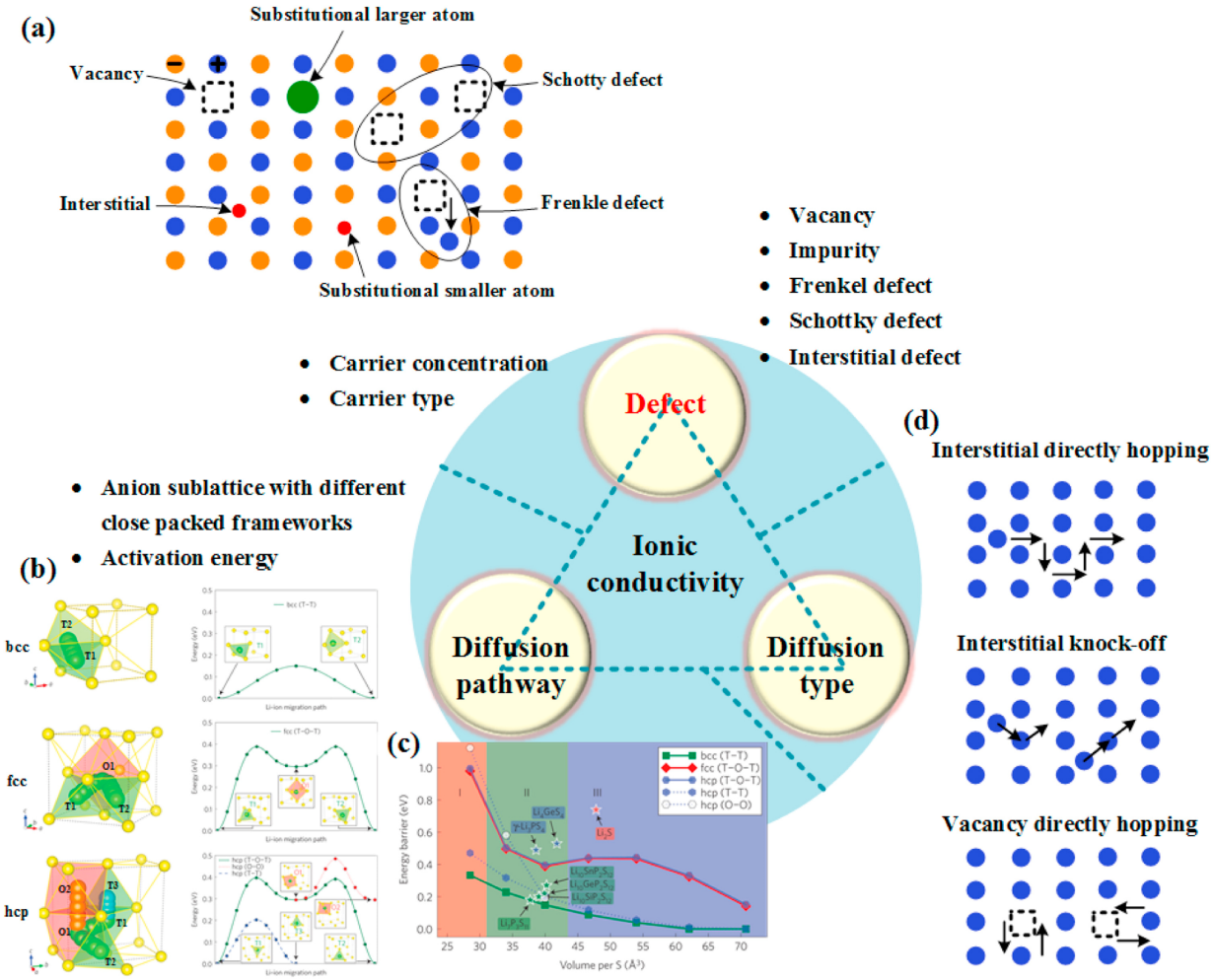


Figure 2.9: Schematic representation of (a) various defects, (b) Li-ion migration pathway (left) with calculated energy along the migration path (right), (c) activation barrier of Li ion as a function of lattice volume, and (d) migration mechanism in solid-state ceramic electrolytes [10, 11]. Copyright © 2018 ELSEVIER and 2015 Springer Nature.

There are some electrolytes that pose extremely high ionic conductivity of about 0.01 S/cm and are usually known as superionic conductors. In a typical solid, ion conduction is mediated by the migration of a single ion among sites in the crystal structure, whereas in superionic conductors, ion conduction happens through the diffusion of multiple ions because of the low migration barriers. This type of migration is a result of disordered ions

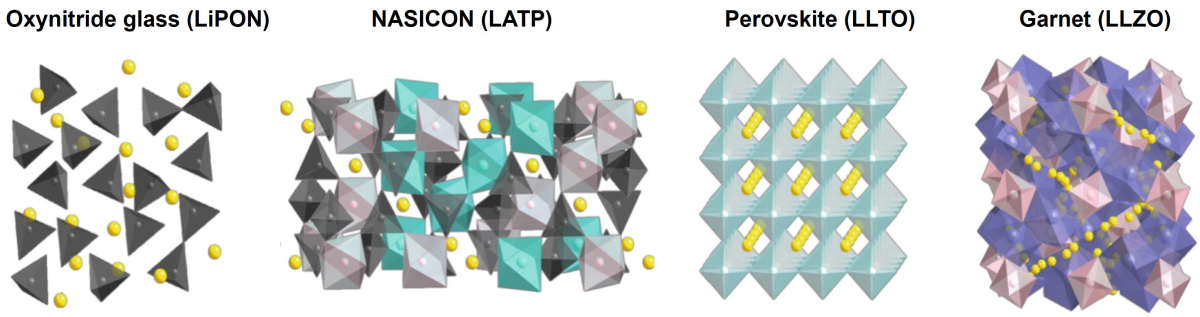


Figure 2.10: Crystal structure of different oxide electrolytes used in solid-state batteries. Most of them can be amorphous or crystalline structures except **LiPON**, which is amorphous-type only. Yellow spheres represent the lithium-ion position in the lattice [12]. Copyright © 2021 Springer Nature.

in the lattice caused by geometric frustration. Lithium-ion-based electrolytes are promising candidates for superionic conductors. There are three types of solid-state electrolytes: Polymer electrolytes, Oxide electrolytes, and Sulfide electrolytes. Polymer electrode suffers from low ionic conductivity and less energy density, whereas Sulfide electrolytes have poor stability and are environmentally hazardous. Oxide electrolytes, on the other hand, have high ionic conductivity. Figure 2.10 shows different crystal structures of oxide-type solid-state electrolytes. Among different lithium-ion-based superionic electrolytes, Lithium phosphorous oxynitride (**LiPON**) has been used as a superionic electrolyte in thin-film solid-state batteries (SSB) since 1990. Besides having high ionic conductivity, **LiPON** is a good electrical insulator. This is why **LiPON** will be used in this research.

LiPON thin film can be grown in several deposition techniques, including radio frequency magnetron sputtering, pulsed laser deposition (PLD), ion beam-assisted deposition (IBAD), and atomic layer deposition (ALD). The most popular method is radio frequency magnetron sputtering, in which Li_3PO_4 target is used in a nitrogen plasma environment to grow amorphous glass-like electrolytes with different stoichiometry of $Li_xPO_yN_z$ ($2.6 \leq x \leq 3.5$, $1.9 \leq y \leq 3.8$, $0.1 \leq z \leq 1.3$) [38, 39, 40].

2.4.2 TiO_x and LiPON Switching Layer

To realize a battery-like memristor device, this research incorporates **TiO_x**, a metal oxide switching layer, and **LiPON**, a superionic solid-state conductor. It has been well established that the conductivity of **TiO_x** is determined by the concentration of oxygen

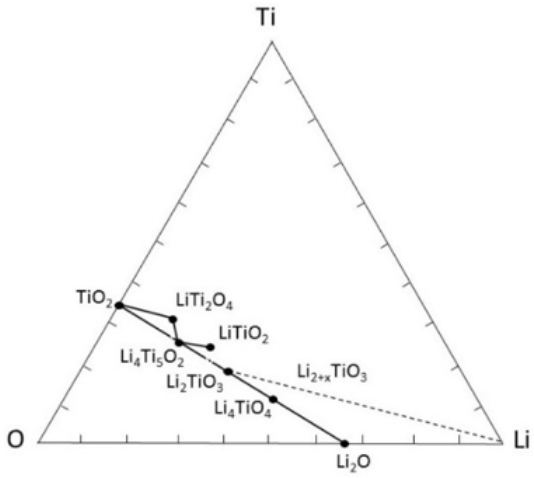


Figure 2.11: Ternary phase diagram of Lithium-Titanium-Oxygen system with lithiation path (dashed line) [13]. Copyright © 2019 MDPI.

vacancies, which is also the basis of $\mathbf{TiO_x}$-based memristors. $\mathbf{TiO_x}$ is also known as a promising electrode material in rechargeable batteries in which Li ions can travel in and out of the nanostructures, easily forming intercalated ion storage. If the device thickness is a few nanometers in the thickness direction, Li insertions and extractions happen readily. A ternary phase diagram of Li-Ti-O compounds is used to understand more about how lithium intercalation and deintercalation change conductivity. The ternary Li-Ti-O system can assume various forms, with the oxygen-defected ones highly conducting, serving well their roles in batteries. As shown in Figure 2.11, phases on the line connecting TiO_2 and Li_2O are stoichiometric and insulating. Phases above this line contain oxygen vacancies and are highly conductive, therefore commonly used as battery electrodes. If the Li intercalated $\mathbf{TiO_x}$ is in this region, Li tends to bond with O and free up electrons from Ti, making the conductivity go up as the Li concentration increases. The more detailed results of this system will be described in the next chapter.

2.5 Memristive Synapses

In a biological brain, there are about 10^{11} neurons and about 10^{15} synapses. The functionality of our brain is determined by a complex neural network of synapses and neurons. The neuron functionality includes the generation and transmission of action potential to

neighbouring neurons, whereas the synapses basically connect those neurons and build the communication. A schematic of biological synapses, in which a postsynaptic neuron is connected with a postsynaptic neuron, is shown in Figure 2.12. Communication between two neurons is developed via a synapse. At first, neurotransmitters are synthesized and stored in the synaptic vesicles. When an action potential is generated at the presynaptic neuron terminal or axon, voltage-gated Calcium channels trigger and Calcium ions (Ca^{2+}) start flowing into the pre-synaptic neuron and synaptic vesicles shift towards the synaptic cleft and release neurotransmitters, which then bind to receptor proteins across the gap and cause Na or K ions to conduct through the post-synaptic neuron, and this process strengthens or weakens the connection strength between synapses, known as synaptic plasticity.

Based on the retention time period, synaptic plasticity can be divided into two categories: (1) Short-term plasticity (STP) and (2) long-term plasticity (LTP). While STP is useful for quick response and information filtering, LTP is crucial for the development of synapses and the regulation of neural circuits. Nanoscale memristive devices can efficiently emulate different brain-like synaptic functions. To date, much research has been conducted to demonstrate the STP and LTP in memristive devices.

2.5.1 Memristive STP

Synaptic plasticity forms that are in the range of milliseconds to minutes are known as STP, which includes PPF, PTP, synaptic facilitation/depression, and augmentation. As described earlier, whenever neurotransmitters migrate to the postsynaptic receptors, they temporarily modify the synaptic efficacy. If there is another identical action potential generated before the recovery time of the calcium ions, the response from the postsynaptic neuron becomes higher compared to its first one. This effect is known as PPF. Likewise, a train of consecutive identical pulses of a certain duration results in a gradual augmentation of synaptic plasticity, which is known as PTP. In addition, synaptic plasticity is a function of the spike rate as well. Spike rate modifies the number of neurotransmitters from the presynaptic neuron, which modifies the strength of synapses; this process is known as SRDP.

Figure 2.13(a) shows the paired pulsed facilitation when the applied presynaptic spikes are close. The PPF index, which is defined as the ratio of A_1 and A_2, as a function of timing between applied presynaptic spikes, is shown in Figure 2.13(b). If the time interval between two presynaptic events is large, the PPF indexing becomes small because of the ion relaxation phenomenon associated with the excitation. Therefore, the PPF index versus

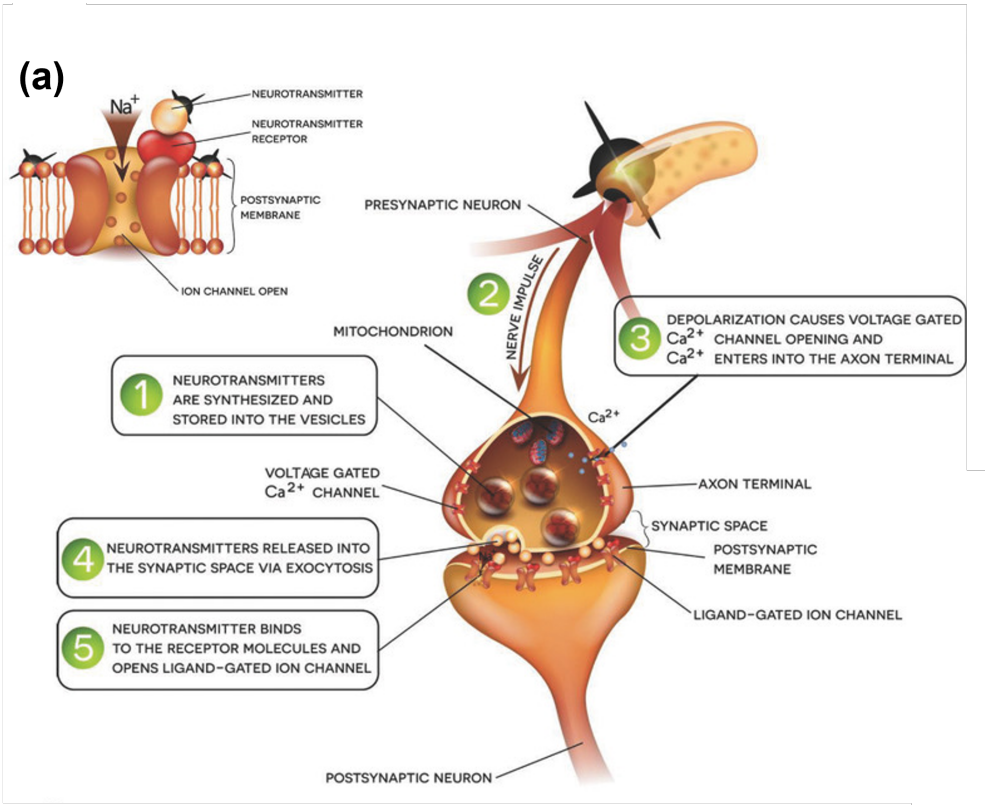


Figure 2.12: Schematic of a biological synapse connecting a presynaptic neuron with a postsynaptic neuron. Upon the generation of action potential in the presynaptic neuron, voltage-gated calcium ion channels open, and calcium influx causes the neurotransmitter to be released near the synaptic cleft. These neurotransmitters then bind with receptor molecules and alter the connection strength of the postsynaptic neurons [14]. Copyright © 2018 John Wiley and Sons.

presynaptic time interval shows an exponential decay-like behaviour. Moreover, a train of similar pulse shows PTP from the device, and when the pulse interval is too high, about two orders of magnitude, it causes paired-pulse depression (PPD) characteristics, as shown in Figure 2.13(d) . The pulse characteristics also confirm the usual PPF and PPD characteristics from a memristive device. Therefore, memristive devices are capable of emulating short-term plasticity.

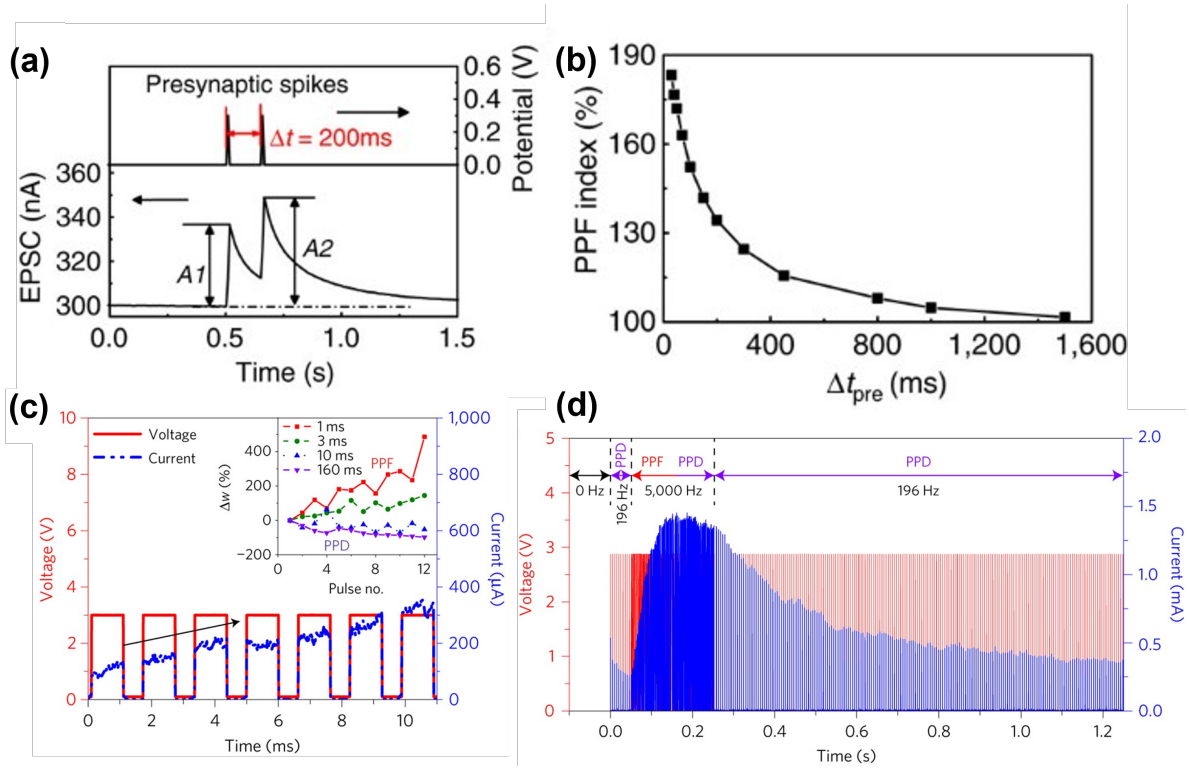


Figure 2.13: (a) Two consecutive presynaptic spikes with a 200ms delay applied in a memristive device and corresponding excitatory post-synaptic current (EPSC) from the device. Here, A_1 and A_2 are the current amplitudes of the first and second EPSCs, respectively. (b) PPF index as a function of the interval between two presynaptic spikes [15]. Copyright © 2014 Springer Nature. (c) Memristive devices mimic short-term plasticity, including PPF and PPD, when consecutive pulses with different pulse delays are applied to the device. (d) Memristive devices with PPD followed by PPF behaviour [4]. Copyright © 2017 Springer Nature.

2.5.2 Memristive LTP

In contrast to STP, long-term synaptic plasticity, including LTP and long-term depression (LTD), prevails for several hours to days. Although compared to STP, LTP lasts longer, it also decays over time. The conductance decay rate depends on the pulse duration and amplitude. Raymond *et al.* studied LTP response for varied input conditions. When the input stimulus is a single pulse train of 100 Hz, the conductance goes back to the initial

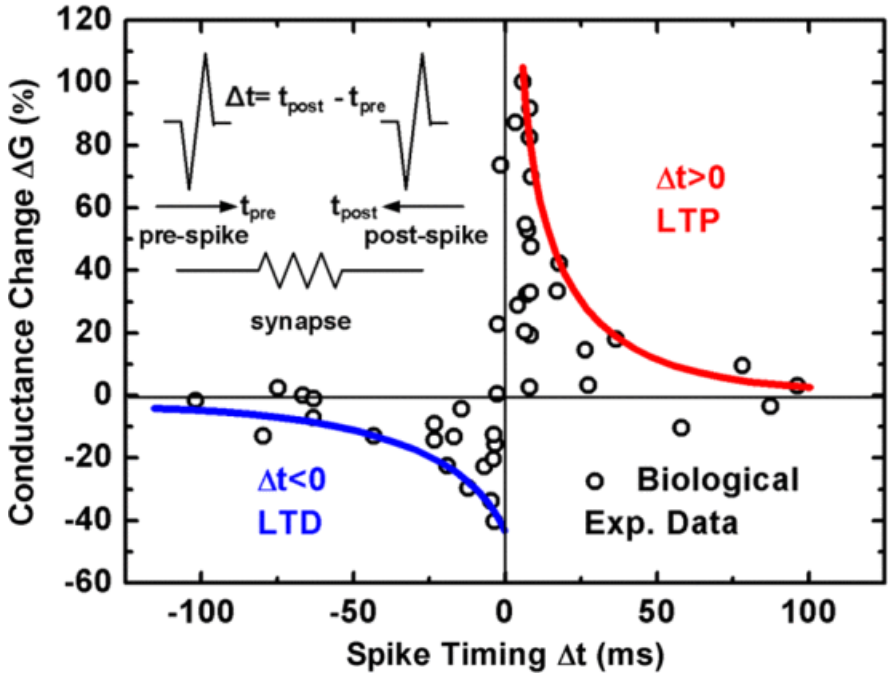


Figure 2.14: Spike timing dependent plasticity (STDP) of a biological synapse. If the presynaptic input precedes the postsynaptic one, the conductance increases, known as long-term potentiation (LTP). In the opposite case, the conductance decreases, known as long-term depression (LTP). The closer the timing, the larger the relative conductance change, which is defined as $\Delta G = (G_a - G_p)/ G_p$, where G_p is the conductance value before the synaptic input and G_a is the conductance value after the synaptic input [16, 17]. Copyright © 1998 Society for Neuroscience & 2011 IEEE.

value within two hours. The EPSC decays back slowly to the initial EPSC after four pulse trains and decays at a much slower speed when eight pulse train stimuli are applied.

The synaptic plasticity in LTP relies on the relative timing difference between presynaptic and postsynaptic neurons, which is known as spike-timing-dependent plasticity (STDP) or Hebbian-like learning. In STDP, if the presynaptic spike input arrives before a postsynaptic spike input within a short interval in a memristive device, the synaptic weight increases; similarly, if the postsynaptic input arrives before the presynaptic input, the synaptic strength decreases. If the relative timings are large, there will be no change in the synaptic weight. Figure 2.14 shows STDP like LTP and LTD in a memristive device.

2.6 Emerging Computing Architecture

Existing computing architecture is plagued by higher energy consumption and latency, as the memory and processor are physically segregated. That's why neuromorphic computing, which is inspired by the human brain, is proposed where the memory and processing elements are co-located, thereby eliminating the energy cost and latency of data shuffling. There are several emerging computing architectures; however, in this thesis, RC and IMC will be discussed.

2.6.1 Reservoir Computing using Memristors

Among various artificial neural network (ANN) architectures based on static random-access memory (SRAM), DRAM, and memristor, the memristor-based ANN has been extensively studied due to its simple fabrication process and low integration cost. Furthermore, memristors can have volatile or non-volatile natures, resembling the short-term and long-term memory of the biological brain. In terms of ANN architectures, feedforward neural networks (FFNN) are particularly useful for static spatial patterns, such as image recognition, while recurrent neural networks (RNN) are more practical for dynamic data, including voice recognition and forecasting. RNN requires less number of training parameters compared to the FFNN network, thus reducing complexity. However, the training of RNN suffers from a vanishing or exploding gradient problem. In a recurrent neural network, weight matrices are multiplied both in forward and backward propagation. When the error is multiplied and back propagated, there is a possibility that weights are becoming weaker, known as the vanishing gradient problem or weights are becoming larger, known as the exploding gradient. This phenomenon is more pronounced when the time sequence is large as the error is multiplied again and again, which reduces the value close to zero or infinite, resulting in the incapability to decide which direction to move to minimize the loss function. In contrast, in reservoir computing schemes such feedback is avoided. To solve this issue, a RC method was introduced where training updates only at the output layer, thus bypassing the interior complexity and reducing the number of training parameters. There are two main parts in the RC scheme: a reservoir layer and an output layer. The volatile memristors can act as a reservoir of the RC system, whereas the output layer can benefit from non-volatile elements to retain the training weights.

A typical RC system is depicted in Figure 2.15(a), consisting of an input layer connected to a dynamic reservoir and an output layer. The purpose of the dynamic reservoir is to transform the input into a high-dimensional space representation that can be used to

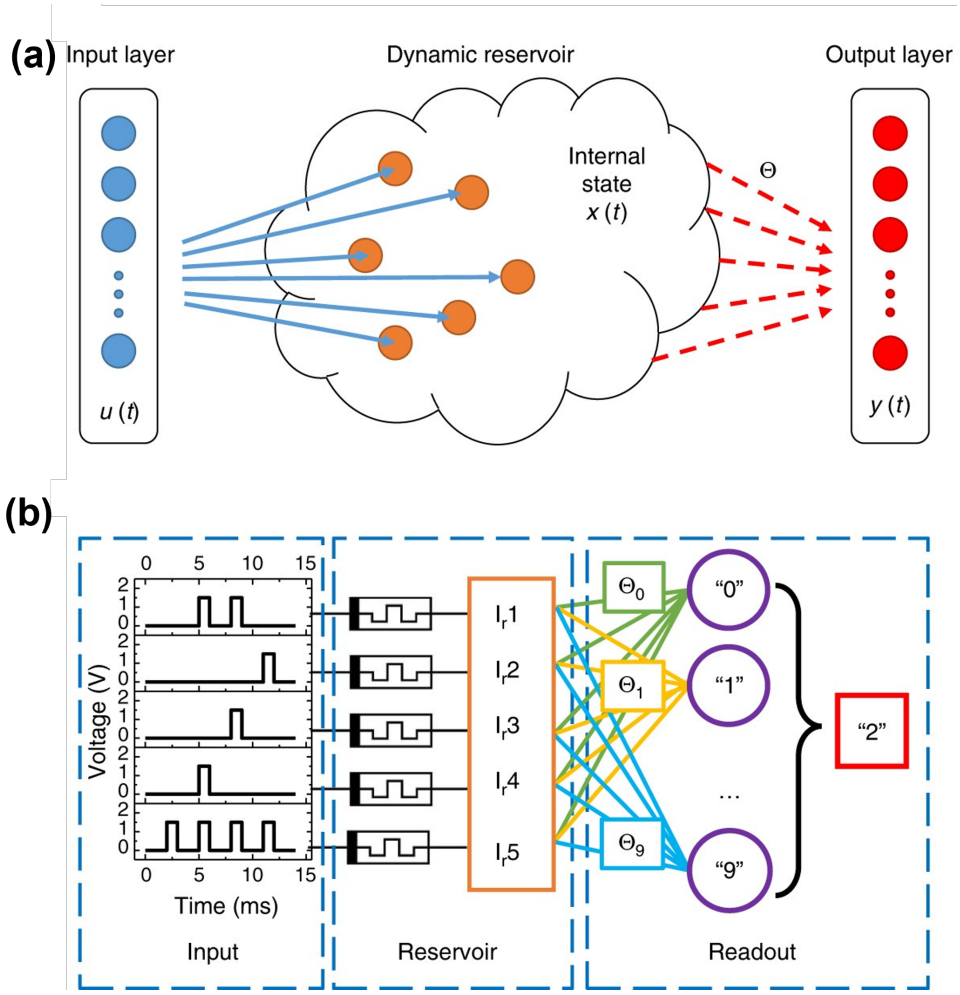


Figure 2.15: Reservoir computing and reservoir computing architecture with memristive devices. a) A schematic illustration of a typical RC system that shows the reservoir with internal connections and a readout layer. It is interesting to note that only the weight matrices Θ are trained. b) Illustration of a Memristive RC system with encoded pulse streams as the inputs, the reservoir based on memristors, and a readout layer, which is a fully connected neural network. In an example of a digit recognition task that consists of 5×4 pixels images, and the reservoir has 5 volatile memristors [18]. Copyright © 2017 Springer Nature.

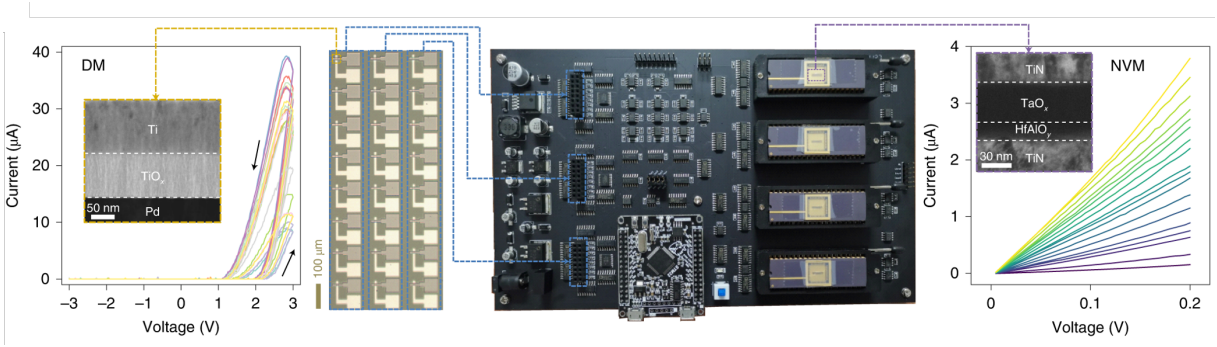


Figure 2.16: Reservoir computing hardware using volatile dynamic memristor(DM) and multi-bit non-volatile memristors. Hysteresis current-voltage curves, an optical image of 24 DMs, and a cross-sectional TEM image of the DM (inset) with vertically stacked $Ti/TiO_x/Pd$ device structure are shown on the left. A photograph of the PCB hardware that integrated the power module with an MCU, reservoir module, and readout module (consisting of four non-volatile memory chips) is shown in the middle. Current-voltage characteristics of multi-bit capable non-volatile memory devices with $TiN/TaO_x/HfAlO_y/TiN$ material stack and corresponding cross-sectional TEM image are shown on the right [19]. Copyright © 2022 Springer Nature.

linearly separate the input and the output layer is a fully connected neural network. It is interesting to note that only the weight matrices Θ are trained. Figure 2.15(b) shows the memristive device-based reservoir computing architecture. The input layer is a stream of pulses, which is a resultant of 5×4 pixels encoded as a basis of black and white image values. The memristors are used as a reservoir due to the volatile non-linear current-voltage characteristics and produce the specific output that will connect via a fully-connected neural network.

There are several studies investigating the feasibility of memristive reservoir computing architecture. Most of the implementation was carried out using volatile memristors that act as a reservoir and non-volatile memristors on the output layer to store the synaptic weights. Rivu Midya *et al.* have demonstrated an RC system to classify the Modified National Institute of Standards and Technology (MNIST) handwritten digits database with diffusive memristors(Pd/SiO_2:Ag/Pt) and drift memristors($Pd/Ta_2O_5/Ta$) [41], even though the accuracy of the implementation was 83%. John Moon *et al.* performed classification and forecasting of temporal data using $W/WO_x/Pd/Au$ volatile memristor [42], and a parallel dynamic RC system was investigated by Yanan Zhong *et al.* using $Ti/TaO_y/TiO_x/Pt$ volatile

memristors [43]. In the above-mentioned works, the output layer was trained separately or using a different set of memristors [19]. Figure 2.16 shows memristive reservoir computing hardware architecture implementation using dynamic memristors(Ti/TiO_x/Pd) as reservoir layer and multi-bit capable non-volatile memristors (TiN/TaO_x/$HfAlO_y$/TiN) to store the weight of fully connected neural network. Previous implementations were realized by fabricating separate devices for the reservoir and output layer, respectively, as the reservoir layer needs a highly nonlinear and dynamic system, for which volatile memory was utilized, and the output layer needs to retain training weights for a long time, for which non-volatile memory was used. As the devices' materials stack are different, the fabrication cost of the devices was expensive, and they required larger areas and/or additional metal layers. Therefore, a universal memory device that can be configured as either volatile or non-volatile is of utmost importance to reduce footprints and latency and to realize low-cost RC systems with customizable layouts on a universal platform. The customization can be tailored/optimized towards specific applications, which is similar to the concepts of FPGA but geared towards non-digital realizations. Therefore, the first objective focuses on the material stacks that can be configured to volatile or non-volatile memory.

2.6.2 Multi-bit-Memory-based In-memory Computing

A typical neural network consists of an input vector, weight matrix, and output vector, as shown in Figure 2.17(a). The core of the neural network is to accomplish MVM, which can be done in a digital-like von Neumann architecture or in an analog-like memristive crossbar architecture. In existing von Neumann-like computing processors such as CPUs and GPUs, MVM operations are typically accomplished in dedicated hardware units known as MAC units, as shown in Figure 2.17(b). A MAC unit generally consists of a multiplier unit, an adder unit, and a register unit that stores the result of the MVM operation. At each clock operation, the data that are retrieved from the external memories are first multiplied, followed by accumulating the output in a register. This process, nonetheless, involves large latency, register operations as well as data access delay based on the matrix size.

In contrast to digital MVM architecture, the memristive crossbar-based MVM architecture features a fully interconnected topology, which enables massive parallelism of MVM processing, as depicted in Figure 2.17(c). Within the memristive crossbar array, the input vector (multiplier) is applied to the bottom electrode of the crossbar array, and the memory devices' conductance acts as a weight matrix (multiplicand). Ohm's law is used to multiply the input with weight matrices, and Kirchhoff's current law is used to sum up all the multiplied results. The output current from the bit line (top electrode) represents the resultant MVM operation. With massive parallelism capability, memristive IMC has

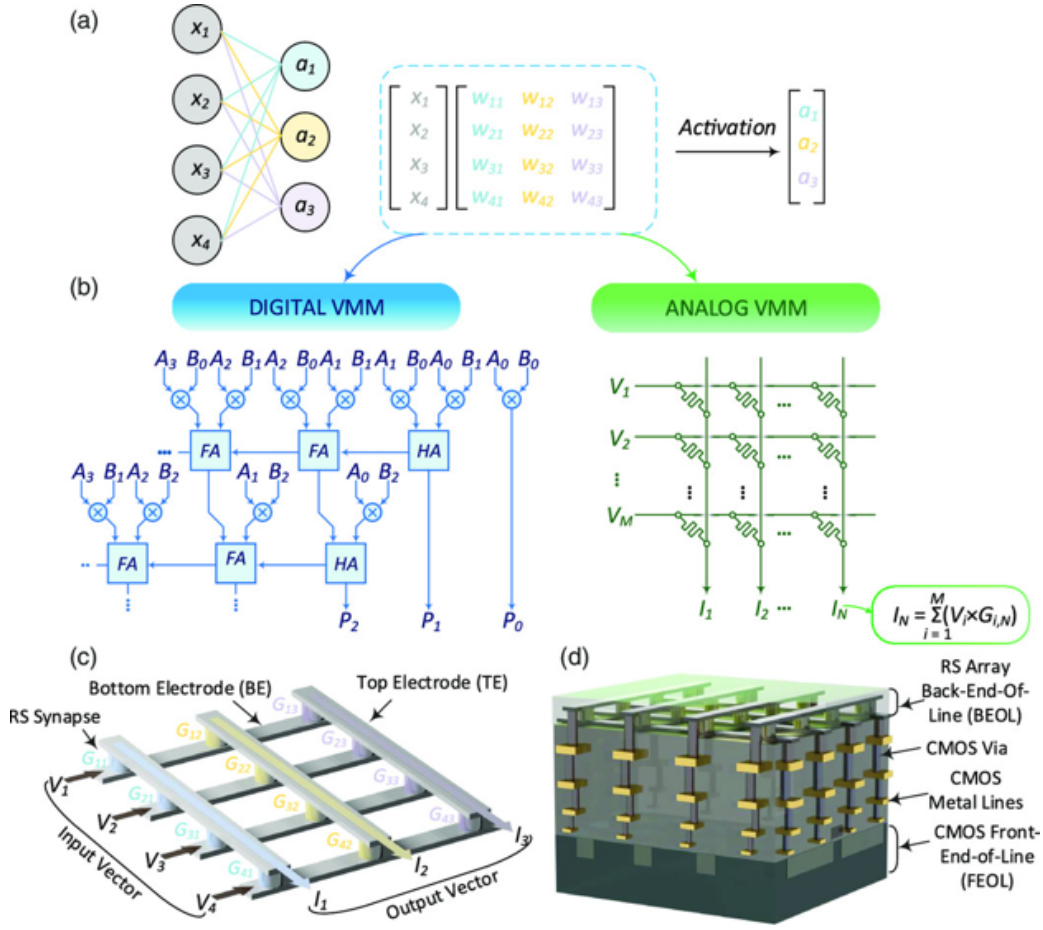


Figure 2.17: a) Schematics of a typical neural network. b) The illustration of a digital and analog matrix-vector multiplication (MVM) scheme. In digital MVM implementation, multiplier digital blocks are used to multiply two digits followed by adders blocks to add the results. The analog VMM can be realized by adding currents from M lines and repeating the same in N columns of an RS-based crossbar. c) The hardware implementation of MVM using a memristive crossbar, in which each intersection contains a memory cell. The bottom electrodes, where input vectors are applied, act as a word line. The devices' conductance acts as a weight matrix. The resultant current from the top electrode, which also acts as a bit line, represents the MVM result. d) Three-dimensional image of a back-end-of-the-line (BEOL) integrated memristive crossbar [20]. Copyright © 2020 Wiley-VCH GmbH.

emerged as a promising alternative for efficiently accelerating MVM operations and related applications.

Memristive devices can be used to represent as synaptic weights of a memory cell. Memristive crossbar architecture is a promising candidate for highly energy-efficient and area-efficient implementation due to non-volatility, meaning no voltage is needed to store the data, and the unit crossbar area required by each device is $4F^2$. However, passive memristive crossbar arrays suffer from the sneak-path current problem, which is defined as an undesired current path due to the simple resistive structure, resulting in failure to read accurate data value in the cells. The common method to suppress sneak path current is to use a transistor, diode, or a selector in series with the memristors. To select the specific cell, the transistor connecting to that cell will be turned on, and all others will be turned off, which reduces the sneak path current. The transistor connected with a memristive array is shown in Figure 2.17(d).

To represent the synaptic weight appropriately, memristor devices need to be capable of storing multi-bit information with non-volatility. Typically, based on how many bits a memristive device can store, different crossbar architectures have been proposed. Based on the number of states and transistor configurations, 1T-1R, 2T-2R, and nT-nR structures have been proposed to accelerate the MVM operation. The challenge is to encode as many states as possible in memristive devices. In the case of limited conductance states, memristor devices with multiple transistors can enhance the resolution. This thesis proposed a 2T-2R structure that can efficiently store multi-bit and accelerate the MVM operation. Also, helimagnet-based resistive memory that is capable of storing multi-bit information is proposed in this research.

Chapter 3

Experimental Methodology

Overview

In this section, three basic components of micro- and nano-fabrication, including lithography, deposition, and etching processes, will be elaborated. At first, thin-film deposition techniques, such as sputtering and plasma-enhanced chemical vapor deposition, are presented. Next, dry etching techniques, such as ion milling and reactive ion etching, are covered. The most important lithography techniques, for example, photolithography using MLA and e-beam lithography techniques, are briefly explained afterwards. Moreover, a description of the whole fabrication process of standalone memory devices and CMOS-integrated memory devices is presented.

3.1 Thin-film Deposition Processes

Thin-film deposition is a crucial step in any micro-fabrication or nano-fabrication process, which facilitates the creation of active layers on the substrates for various applications in photonics, electronics, MEMS, and other state-of-the-art technologies. The deposited layer can be conductors, insulators, semiconductors, or any other type. At present, there are two main thin-film deposition techniques: physical vapor deposition (PVD), and chemical vapor deposition (CVD). The choice of a deposition process relies on the preferred film properties, uniformity, thickness, and compatibility with underlying substrate materials. In this thesis, sputtering as PVD and plasma-enhanced chemical vapor deposition as PVD were employed to fabricate the memristor devices.

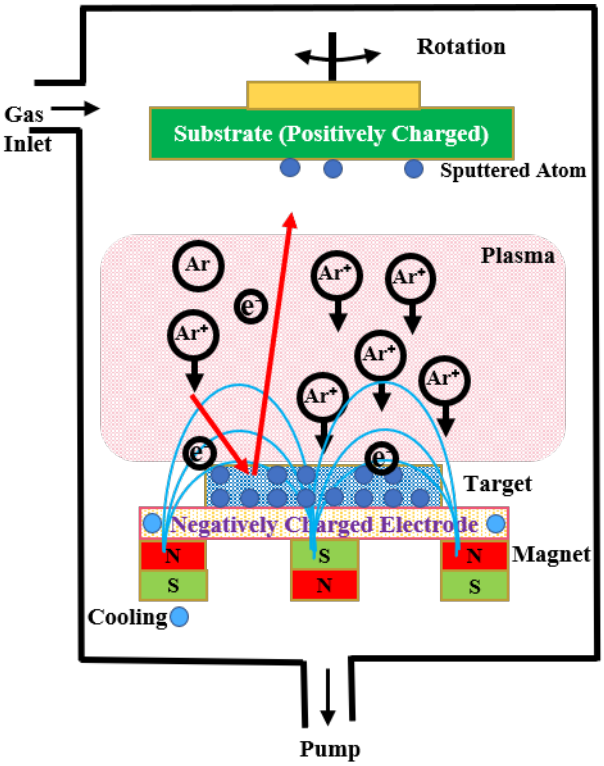


Figure 3.1: Schematic illustration of an RF Magnetron Sputtering system with substrate rotation capability.

3.1.1 Physical Sputtering

Sputtering is a type of physical vapour deposition (PVD). In this technique, we need a target, a thick layer of desired materials, and energetic ions typically come from electron collision with inert Argon gas. When a bias voltage is applied, these energetic ions are accelerated and collide with the target surface atoms, ejecting an atom from the target and travelling to the substrate where it condenses. With continuous atoms coalescing, these atoms eventually form a molecular bond and result in an atomic layer.

One of the factors that make the sputtering system more popular is the better "step coverage". Besides, minor damage due to radiation and high-quality compounding of compounds and alloy layers give good film quality to the deposited layer. Figure 3.1 illustrates an RF Magnetron Sputtering system with substrate rotation capability.

The sputtering rate is directly proportional to the sputtering yield, which is defined as the number of atoms ejected per incident ion. Several factors that influence the sputtering yields are incident ion energy, target atoms and ions masses, atom's binding energy in the target, and ion incident angle. Although sputtering yield is not dependent on the target's temperature, very high temperature leads to an increase in deposition rate due to the material's thermal excitation.

There are four widely used sputtering techniques: direct current (DC) sputtering, radio frequency (RF) sputtering, Magnetron sputtering, and Reactive sputtering. DC sputtering is only suitable for conductive materials deposition, such as metals. RF sputtering can be used for both conductive and insulating materials. If the material deposition rate and lower chamber pressure are of concern, Magnetron sputtering is the right choice. The reactive sputtering method is best for depositing compounds.

3.1.2 Plasma-Enhanced Chemical Vapor Deposition

In micro- and nano-fabrication, plasma-enhanced chemical vapor deposition (PECVD) is a popular technique for depositing thin films such as silicon dioxide (SiO_2). This process requires a lower temperature compared to the chemical vapor deposition process due to the presence of plasma during deposition. The PECVD process is especially useful where thermal budgets are constrained to avoid damage to the underlying layer. Several intertwined processes occur during the deposition of thin films, which are summarized below.

- Creation of highly reactive species;
- Transport of highly reactive species onto the wafer/ surface;
- Adsorption onto the wafer surface;
- Reaction onto the wafer surface;
- Desorption of volatile byproducts;
- Pump out volatile byproducts from the chamber;

To passivate the device structure, a SiO_2 layer was used, deposited using the PECVD process, where Silane (SiH_4) and nitrous oxide (N_2O) were used as precursor gases. The chamber temperature was kept at 330°C. The deposition time controls the final thickness.

3.2 Etching Processes

The process of removing materials from the parent structure's surface is popularly known as etching. It is commonly used in semiconductor device fabrication processes. In principle, one can use two general classes of techniques for etching: wet etching and dry etching. In wet etching, the solution-based chemical method is used for delivering high concentrations of etchant to the substrate, which allows efficient use of chemicals for achieving high etching rates. The main advantages of this method are low cost, high etching rate, and high selectivity (for some materials). Due to the usual isotropic nature of wet etching, it cannot be utilized for patterning complex structures in nanometer dimensions. Moreover, isotropic etching often results in an undercut, which reduces the overall integrity of the devices. In dry plasma etching, instead of liquid chemical etchants, gas phase etchants are excited into plasma, and ions and reactive species (mostly free radicals) are generated. These extremely reactive species chemically react with the materials to be etched and form volatile by-products, which will be pumped away. Dry etching is much more precise and controllable with the capability of defining structures less than 100 or even 10 nm. In this thesis, ion milling and reactive ion etching processes were used and described in detail in the following sections.

3.2.1 Ion Milling

In physical ion milling, heavy inert gases (typically Ar, because of its high sputtering yield) are used to etch solid surfaces. Since its etching principle is based on ion bombardment, it is a physical process without involving chemical reactions. This process is also known as ion beam etching. For generating gaseous ions, Argon gas, which is chemically inert, is introduced, and plasma or electron bombardment is used to ionize it. After extracting the ion beam, they are accelerated onto a substrate with a patterned mask situated in a vacuum chamber. Argon ion energy leads to various types of physical interaction. If the ion energy is low ($\sim$ 10 eV), accelerated ions can damage the sample surface up to a few atomic layers. On the other hand, ions with higher energy ($\sim$ 10 keV) can go deeper into the sample and become implanted (ion implantation) [44]. As a result, the typical ion energy for ion milling is kept between hundreds to thousands of eV.

Both the ion density (determined by the plasma source) and ion energy (determined by DC bias voltage) can be independently controlled in ion beam etching. Besides, the electric field is set up with independent electrodes, allowing the wafer to remain neutral. The wafer can be tilted and rotated relative to the direction of the ions to ensure that the

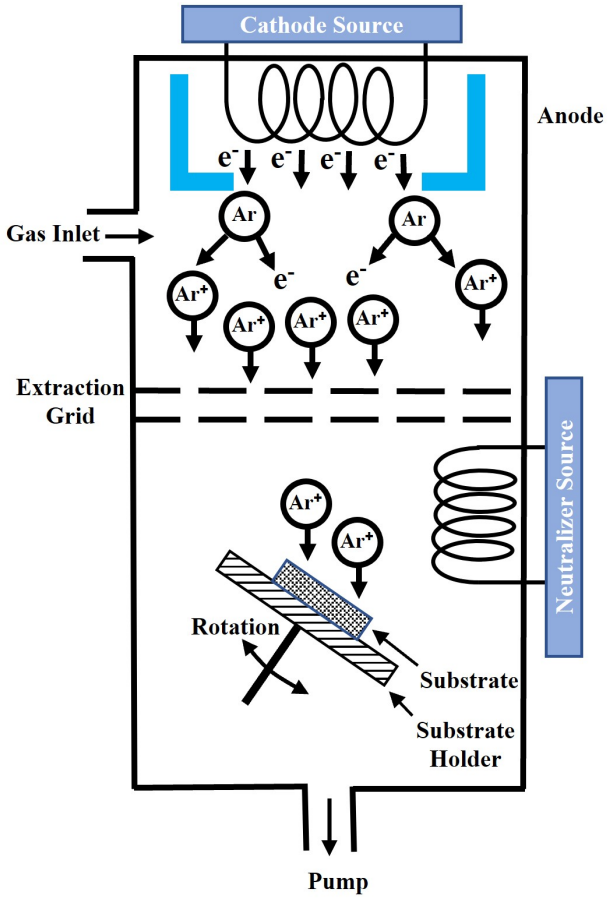


Figure 3.2: Illustration of a typical ion beam etching system.

etch process can remove tapered sections of the stack. Ion milling is mostly used when RIE can't produce volatile by-products because not all the materials can react chemically to form a volatile compound. Sidewall profile control, tilted feature, and optimization of radial uniformity can be enabled by ion milling. Due to the anisotropic nature of ion milling, it can easily be used to transfer nanometer structures effectively.

The schematic of a typical ion beam etching system is represented in Figure 3.2. In this system, the ion source is separated from the sputtering chamber. Ions can be generated by an RF source or DC source. Figure 3.2 is a DC source-based system using electron bombardment on gas molecules. Electrons gain enough energy when there is a potential difference between the thermionic cathode and anode. Ions are formed when highly energetic electrons collide with the gas molecules. After that, the ionized gases are extracted

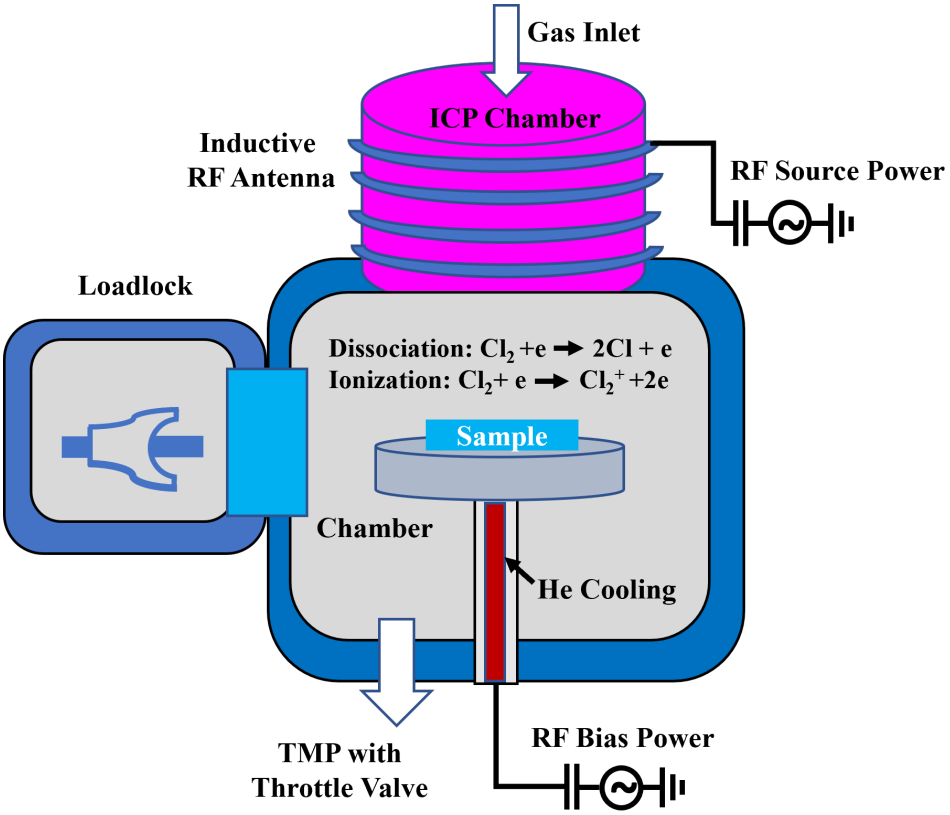


Figure 3.3: Simple diagram of different parts of an ICP-RIE etching system.

as a broad beam toward the sample's surface. An optional electron gun, located at the bottom part of the chamber, is used to neutralize the positive charges that accumulate on the surface of the sample (usually for insulating materials). At Ar ion energy of 1 keV and current density of 1 mA/cm^2, the typical etching rate is 10–300 nm/min.

3.2.2 Reactive Ion Etching

Coburn et al. first showed that incorporation of a chemically reactive gas such as Cl_2(g) into the argon ion beam chamber can increase the etching rate of silicon from 2 Å/min to 10 Å/min [45]. This was the first demonstration of reactive ion etching (RIE). So far, it is one of the most popular forms of ion-enhanced etching, and the process relies on both chemical reactions and physical bombardment by ions on the wafer. Also, ions are accelerated by

an electric field between an electrode in the etch chamber and the wafer itself, thus making the process highly directional. Another motivation for plasma etching is the reduction of chemical residue [46]. RIE with parallel plate setup is a low-density plasma system (ions $10^8 - 10^{10}$ /cm^3), which implies a low-density of free radicals, thus low etch rate. In brief, the reactive ion etching process takes place in three steps. At first, reactions happen on the sample surface. The reaction with the free radicals can be spontaneous, i.e., without the need of ion bombardment, or ion-assisted, in which the physical ion bombardment can greatly enhance the reaction rate by mechanisms such as damaging or weakening the sample surface or boosting the reaction product desorption [47]. After that, reaction products are formed on the material's surface to be etched. Finally, desorption of these products occurs, and the etching continues.

To enhance the etching rate, we need highly dense plasma, which can easily be obtained by an Inductively Coupled Plasma (ICP). The system runs by using two independent power controls, namely the ICP coil power and the bias power, and operates at low pressure. In ICP-RIE etching, the ion flux and ion bombardment energy can be controlled by varying respectively the ICP and bias RF power. Figure 3.3 shows components of an ICP-RIE etching system. Ionization is much more efficient, and it can operate at lower pressure, which leads to fewer ion collisions. Thus, the etching is more directional or anisotropic with a vertical profile. In ICP plasma, a circular electric field can be generated by using the AC magnetic field, which accelerates electrons to boost plasma density. Also, no top plate is actually needed, in contrast to the case with parallel plate RIE.

3.3 Lithography

The lithography process is one of the crucial steps in semiconductor device fabrication. It is used to transfer a pattern into a resist, which will later be removed after the etching step (known as the stripping process) or after the deposition of the desired material (known as the lift-off process). There are two types of lithography processes: photolithography and electron-beam lithography (EBL). While photolithography is typically used to transfer microsize features, EBL is used to transfer nanometer-sized features. The basics of both lithography systems are described in the following sections.

3.3.1 Photolithography

In photolithography, a light source is used to transfer the design into a photoresist by using a mask or direct write method, such as Maskless Aligner (MLA). The photoresist

on which the pattern is transferred can be positive or negative. In positive photoresists, UV light exposure breaks the photoactive compound bonds, which makes it readily soluble in developer solutions, such as TMAH. Therefore, the exposed area will be removed after the development process. In contrast, in a negative photoresist, exposure results in cross-linking or bonding of neighbouring polymer chains, which increases the molecular weight that becomes insoluble. Thus, unexposed areas have lower molecular weights, which will dissolve in aromatic developer solutions. The minimum feature size that can be achieved by photolithography is directly related to the wavelength of the light. The diffraction limit is the prime bottleneck of this technology, which can be described as,

$$d = \frac{\lambda}{2 * n * sin(\theta)} \tag{3.1}$$

where d denotes the feature size, λ represents the light wavelength, n is the refractive index of the medium through which the light is travelling, and θ is the half angle from the aperture to the focal spot.

The MLA is used in this thesis to write a pattern on the resist. The MLA approach is cost-effective because it obviates the expensive photomask fabrication process. Moreover, the patterns can be edited in the software rather than in the physical mask. The advantage of MLA is the fast writing process. Compared to e-beam lithography, MLA is faster because of its wider field laser exposure instead of pixel-by-pixel scanning. However, the resolution is limited in micrometer range.

3.3.2 Electron-Beam Lithography

Instead of lights in photolithography, electron-beam lithography (EBL) utilizes electrons to transfer the patterns. The feature size can be achieved as small as tens of nanometers due to the electron's short wavelength. The wavelength of an electron, λ can be written as,

$$\lambda = \frac{h}{\sqrt{2 * m_0 * e * V}} = \frac{1.226}{\sqrt{V}}(nm) \tag{3.2}$$

where, h is the Planck's constant, m_0 is the mass of the free electron, e is the electron charge, V is the relativistic corrected acceleration voltage. For example, the associated wavelength of an electron with a kinetic energy of 1 eV, is 1.23nm, which is about three orders of magnitude smaller compared to a photon. To write precise and accurate patterns, an electron beam needs to be well-focused with a circular shape, which can be achieved

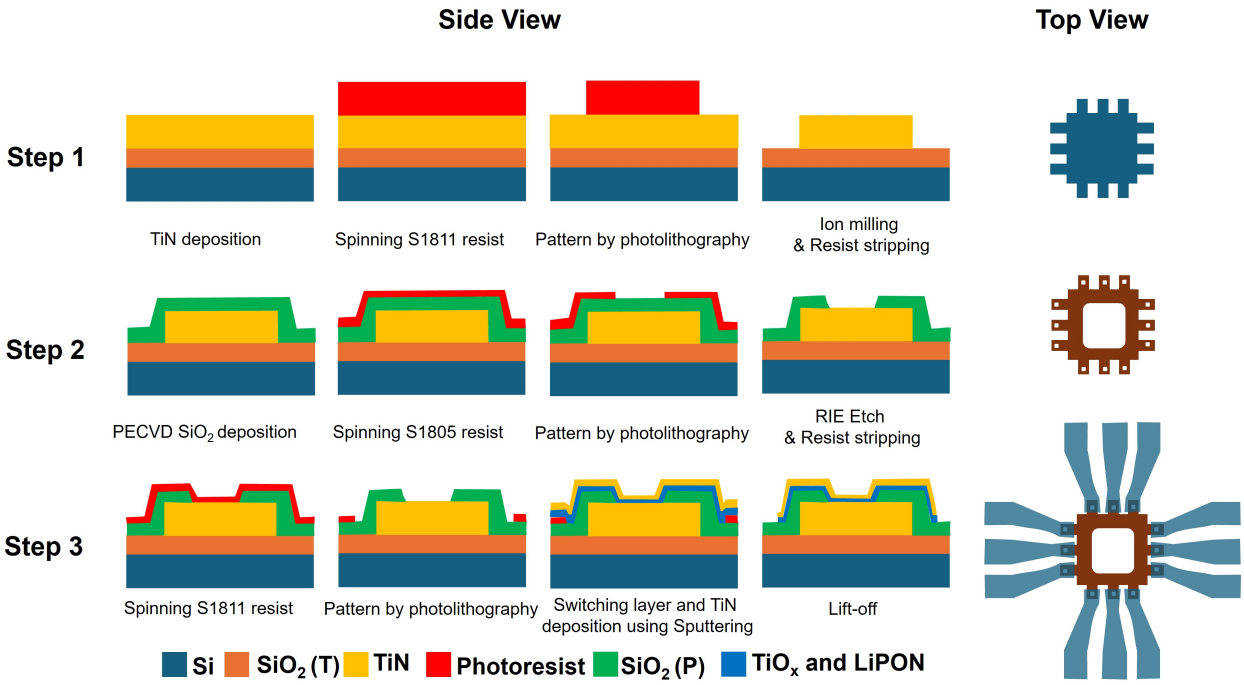


Figure 3.4: Memory device fabrication steps along with a brief description of the process under the steps. Both the top view and side view are given to clarify the fabrication steps.

by changing the astigmatism of the beam. Once the beam is corrected, the next step is to perform the writefield alignment that would eliminate the stitching errors. Proximity effects and charging effects need to be considered to realize a desired pattern. The EBL system requires a long time to expose a large wafer because of its pixel-by-pixel exposure nature, and tools are expensive compared to the MLA system.

3.4 Device Fabrication

This section covers the standard microfabrication process of memory devices. The fabrication steps were performed at the University of Waterloo's Quantum-Nano Fabrication and Characterization Facility (QNFCF) and Research Advancement Center (RAC2).

A 3-inch silicon wafer with 100 nm thermally grown SiO_2 was selected as a substrate. The substrate was cleaned using Acetone, IPA, and DI water to remove the surface con-

taminants. The substrate is then loaded into a loadlock. Once the vacuum reaches 1e-7 torr, the sample is transferred to the Sputtering chamber to grow a TiN thin film, and during growth, the flow rate of Argon and Oxygen gas sources were kept at 30 and 2.5 sccm, respectively. The sample is then ready for the first lithography step, denoted as Step 1 in Figure 3.4. In Step 1, the sample is spin-coated with Shipley 1811 photoresist with the following recipe.

- Step–1 — Speed: 500 rpm, Ramp rate: 500 rpm, Duration: 5 sec
- Step–2 — Speed: 5000 rpm, Ramp rate: 1000 rpm, Duration: 60 sec
- Step–3 — Speed: 0 rpm, Ramp rate: 1000 rpm, Duration: 0 sec
- Baking Recipe: 120°C for 90 seconds.

Next, the pattern transfer was accomplished using MLA photolithography (Dose: 105, Defocus: -2) and developed using MF-319 for about 45 seconds, followed by cleaning using DI water. The developed wafer is ready to load in AJA ion milling to etch the TiN layer. During Ar-ion milling, the ion sources were ejected at an angle of 90°, and secondary ion mass spectroscopy (SIMS) was used during ion milling to determine the etching end-point. Finally, the stripping of the resist was done by dipping the wafer in acetone. If needed, an overnight PG rover was suggested to ease the stripping process. After stripping, a cleaning process was performed using IPA and DI water to remove the residue.

In Step 2, the wafer is loaded in the PECVD chamber to grow a 60 nm SiO_2 layer. After being passivated with SiO_2, the wafer is spin-coated with the same recipe (previous step) except for the photoresist. Shipley 1805 was used instead of Shipley 1811 as it can result in better resolution and easy stripping. The coated wafer is then exposed and developed, making it ready to load in the RIE chamber. The SiO_2 layer was slightly over-etched to eliminate the residue. Similar to the previous step, stripping completes the second step.

The wafer is ready to begin the final step, where the switching layer will be deposited. First, it is again spin-coated with the same recipe as Step–1 and then patterned using photolithography. After the development process, the wafer is ready to grow the switching layer (TiO_x and LiPON) and top electrode (TiN/Ti/Al) using sputtering. A lift-off process finishes the device fabrication steps.

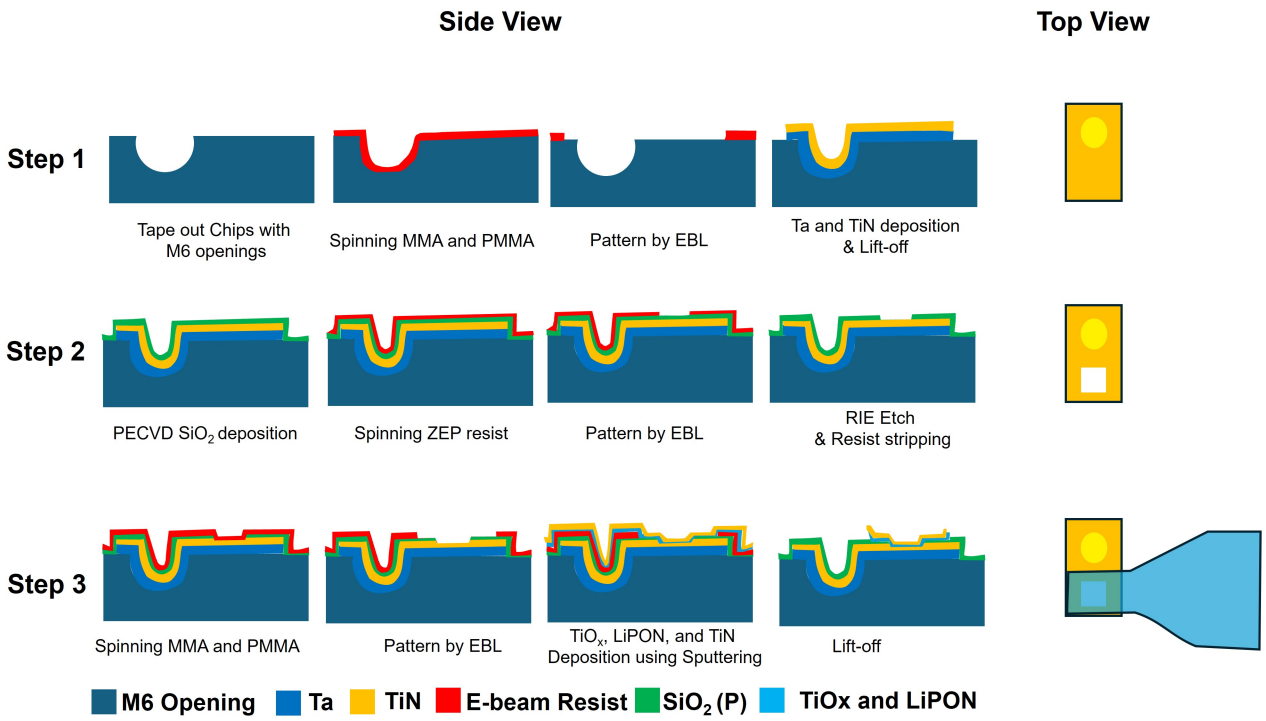


Figure 3.5: Fabrication steps of BEOL integration process along with a brief description of the processes under the steps. Both the top view and side view are given to clarify the fabrication steps.

3.5 BEOL Integration Processes

Backend-of-the-line (BEOL) integration of Lithium-imbued memristor devices on foundry CMOS (TSMC) chips will be described in this section. The integration is crucial to realize the 1T-1R configurations that will allow us to realize computing in memory without sneakpath current issues. However, the existing CMOS foundry production system is not compatible with Li-based materials, rendering the lithium-imbued device to be integrated with CMOS. Therefore, a BEOL integration process that will integrate the lithium-imbued TiOx devices with tape-out CMOS chips is developed using a university cleanroom fabrication facility. This method can be used to integrate other foundry-forbidden materials with CMOS. Like standalone devices, the fabrication steps were performed at the University of Waterloo's Quantum-Nano Fabrication and Characterization Facility (QNFCF) and

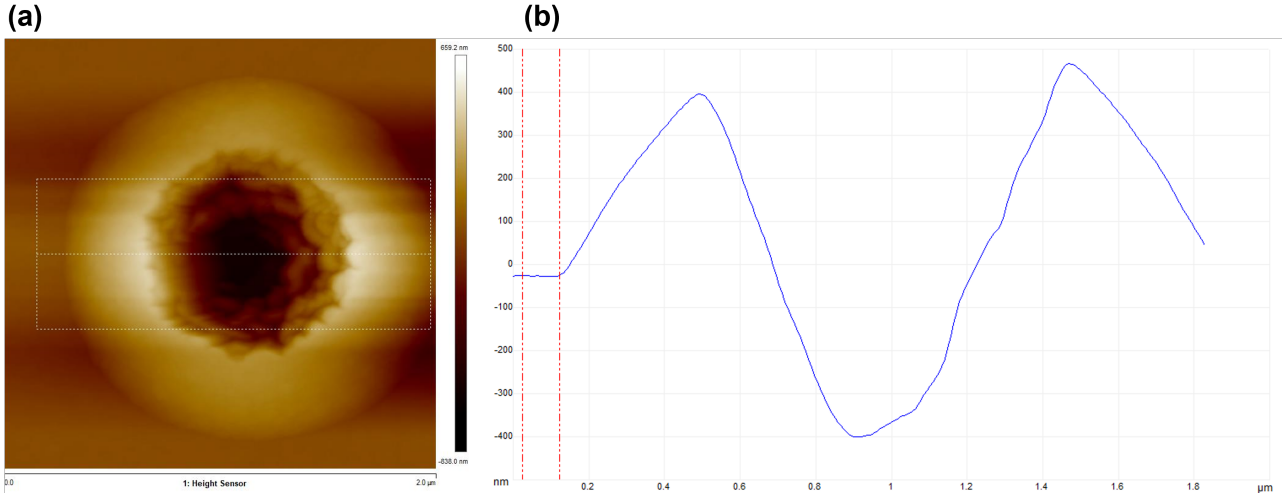


Figure 3.6: (a) AFM image of M6 layer opening showing surface roughness and depth of the hole. (b) The depth profile of the transistor.

Research Advancement Center (RAC2).

The size of the tape-out chip was 3 mm × 4 mm. The spin coater machine can't handle this tiny piece to spin coat the resist. Instead, the chip is mounted on a carrier wafer with PMMA e-beam resist that acts as an adhesion layer (Baking was done at 150° C for 4 minutes). Upon the completion of spin coating, the chip is removed from the carrier wafer by using gentle external force. The chip was spin-coated with MMA and PMMA e-beam resist so that the lift-off became easier. The chip is then loaded into the RAITH EBL chamber to finish the patterning. To maintain continuity, a thick layer of Ta (200 nm) and TiN (30 nm) was deposited using sputtering techniques so that the hole (M6 opening) was covered properly. With the lift-off process, the first step of the fabrication is completed, as shown in Figure 3.5.

Next, the chip is loaded in the PECVD chamber to grow a 60 nm SiO_2 layer. After being passivated with SiO_2, the chip is spin-coated with the help of a carrier wafer as described in previous steps. This time, ZEP 520A e-beam resist was used for spin coating as it can result in better resolution and easy stripping. The chip is then exposed using RAITH EBL and developed. Figure 3.6 shows the AFM measurement and corresponding average depth profile of the M6 opening of the chip. As can be seen from the figure, the surface depth is not suitable for any device fabrication on top of the transistor, so the patterning of the device opening was done close to the opening, where the surface is close to flat. The chip

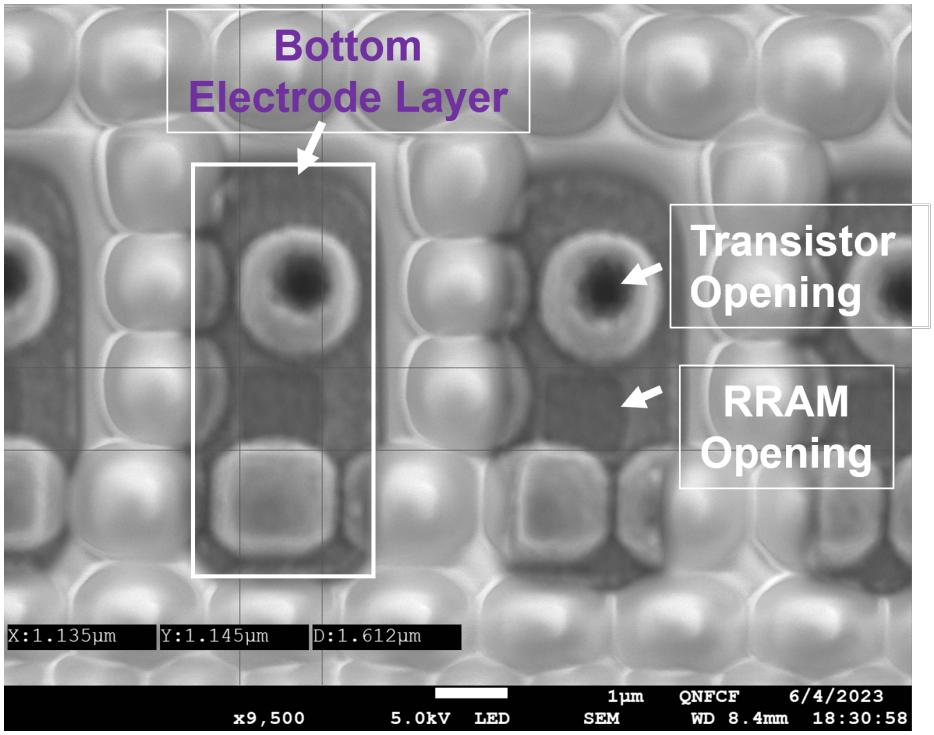


Figure 3.7: SEM images of CMOS chip after bottom electrode deposition, followed by passivation with SiO_2 and RIE etching. The white box indicates the bottom electrode area and the Transistor and RRAM opening.

is loaded in the RIE chamber to etch the SiO_2. The SiO_2 layer was slightly over-etched to eliminate the residue. The stripping of the e-beam resist completes the second step. The scanning electron microscope (SEM) image of the device with the bottom electrode and openings for the transistor and memory device is shown in Figure 3.7.

In the final step, the chip is again spin-coated with MMA and PMMA e-beam resist with a carrier wafer. Upon patterning with EBL and developing the resist, the switching layer (TiO_x and LiPON) and top electrode (TiN/Ti/Al) will be deposited using a sputtering system. A lift-off process finishes the fabrication of the BEOL integration processes.

Chapter 4

Volatile and Non-volatile Iontronic Memory

Overview

This chapter demonstrates a lithium (Li) imbued TiO_x iontronic device that exhibits synapse-like short-term plasticity (STP) behaviour without requiring a forming process beforehand or a compliance current during switching. A solid-state electrolyte lithium phosphorus oxynitride (LiOPN) behaves as the ion source, and the embedding and releasing of Li ions inside the cathodic like TiO_x renders volatile conductance responses from the device and offers a natural platform for hardware simulating neuron functionalities. Besides, these devices possess high uniformity and great endurance as no conductive filaments are present. Different short-term pulse-based phenomena, including paired-pulse facilitation (PPF), post-tetanic potentiation (PTP), and spike rate-dependent plasticity (SRDP), were observed with self-relaxation characteristics. Based on the voltage excitation period, the timescale of the volatile memory can be tuned. Temperature measurement reveals the ion displacement-induced conductance channels become frozen below 220 K. In addition, volatile analog devices can be configured into non-volatile memory units with multibit capabilities after an electroforming process. Therefore, on the same platform, we can configure volatile units as nonlinear dynamic reservoirs for performing neuromorphic training and the non-volatile units as the weight storage layer. We proceed to use voice recognition as an example with the tunable time constant relationship and obtain 94.4% accuracy with a minimal training dataset. Thus, this iontronic platform can effectively process and update temporal information for reservoir and neuromorphic computing.

4.1 Introduction

As a result of the significant advancements in artificial intelligence (AI), the Internet of Things (IoT), and big data technologies, a variety of intricate challenges including image and voice recognition, autonomous driving, and natural language processing, have been addressed through the utilization of the prevailing von Neumann architecture [48, 49, 50, 51]. However, this architecture is plagued by higher energy consumption and latency, as the memory and processor are physically segregated. To address this issue, researchers have proposed the adoption of neuromorphic computing (NC) architectures inspired by the human brain, in which the memory and processing elements are co-located, thereby eliminating the energy cost and latency of data shuffling [52, 42, 53, 54].

Among various artificial neural network (ANN) architectures based on SRAM, DRAM, and memristor, the memristor-based ANN has been extensively studied due to its simple fabrication process and low integration cost [55, 56, 57, 58, 59], and memristors can have volatile or non-volatile natures, resembling the short-term and long-term memory of the biological brain. In terms of ANN architectures, feedforward neural networks (FFNN) are particularly useful for static spatial patterns, such as image recognition, while recurrent neural networks (RNN) are more practical for dynamic data, including voice recognition and forecasting. ANN requires less number of training parameters compared to the FFNN network, thus reducing complexity [60]. However, the training of ANN suffers from a vanishing gradient problem. To solve this issue, a Reservoir Computing (RC) method was introduced where training updates only at the output layer, thus bypassing the interior complexity and reducing the number of training parameters. There are two main parts in the RC scheme: a reservoir layer and an output layer. The volatile memristors can act as a reservoir of the RC system, whereas the output layer can benefit from non-volatile elements to retain the training weights.

Some of the existing volatile memories are digital, meaning they set a certain threshold voltage to trigger the conductance change, and they cannot emulate the analog brain functionalities, though they are still good in selector applications and artificial nociceptors [61, 62, 63, 64, 65, 66, 67]. Recently, some researchers have shown analog-type memristor devices, where the conductance can be gradually tuned, thus emulating biological synapse [68, 69, 70].

Rivu Midya *et al.* have demonstrated an RC system to classify the Modified National Institute of Standards and Technology (MNIST) handwritten digits database with diffusive memristors(Pd/SiO_2:Ag/Pt) and drift memristors($Pd/Ta_2O_5/Ta$) [41], even though the accuracy of the implementation was 83%. John Moon *et al.* performed classification and

forecasting of temporal data using W/WO_x/Pd/Au volatile memristor [42], and a parallel dynamic RC system was investigated by Yanan Zhong *et al.* using Ti/TaO_y/TiO_x/Pt volatile memristors [43]. In the above-mentioned works, the output layer was trained separately or using a different set of memristors [19]. In light of these, a universal memristor that can be configured as either volatile or non-volatile is of utmost importance to reduce footprints and latency in the memristor-based RC system. This would also allow one to realize low-cost RC systems with customizable layouts on the same platform of memristors.

Inspired by natural neurons where the transport of Na and K ions forms the basic neural impulses, we seek solutions in an alkaline ion-enabled solid-state system that is also CMOS compatible. In this research, we report an iontronic device consisting of TiO_x, which is known as a good cathode and anode material, and Lithium Phosphorous Oxynitrate (LiPON), a good solid-state Li electrolyte, together forming a device capable of controllable ionic motion under applied voltages, and configurable to either volatile or non-volatile behaviour. Volatile memory characteristics can be obtained without needing device forming or compliance currents, thus obviating complex setups. The pulse-dependent study revealed Short-Term Plasticity (STP), where conductance gradually varies with the application of short-duration pulses. The timescale of the volatile memory devices is a function of the voltage excitation pulse widths. With a temperature-dependent conduction study, there was a sharp change in conductance above a threshold temperature of 220 K. In contrast, below the threshold temperature, the ion displacement-mediated conductance gives way to weakly temperature-dependent variable range hopping conduction. After electroforming, i.e., creating conductive filaments, these devices can also be configured for non-volatile memory applications with good retention and endurance. Thus, both volatile and non-volatile properties can be obtained on the same memristor-based neuromorphic computing chips, a desirable iontronic platform for creating non-von-Neumann architectures. The tunable timing characteristics of devices were applied to a model task of voice recognition, showing excellent accuracy of about 94.4% with a minimal amount of training data input.

4.2 Experimental Methods

4.2.1 Device Fabrication and Characterization Steps

Three-inch Si wafers with thermally grown 100 nm SiO_2 were used as the substrates. After finishing the cleaning process, the TiN bottom electrode was deposited using DC sputtering. It is then patterned with standard photolithography processes and etched using ion

milling with a secondary ion mass spectroscopy (SIMS) to detect the end-point. Next, a 60 nm SiO_2 passivation layer was deposited using the plasma-enhanced chemical vapor deposition (PECVD) process. Then the sample was again patterned using photolithography and the SiO_2 layer was etched using reactive ion etching (RIE) to open active areas of the devices. Another lithography and liftoff step defined the switching layer and the top electrode (TiN/Al) and formed junctions with the bottom electrode through the active openings. The LiPON and TiO_x were reactively sputtered by properly adding N_2 and O_2 gases, respectively. Focused ion beam scanning electron microscopy (FIBSEM) and transmission electron microscope (TEM) were done using ZEISS SEMFIB and JEOL TEM.

4.2.2 Electrical and Temperature Measurement Steps

All electrical measurements, including DC and Pulse, were done with a Keithley 4200A semiconductor parameter analyzer, equipped with source measure units (SMUs) for DC measurement and pulse-measurement units (PMUs) and remote preamplifier/switch modules (RPM) for pulse measurements. Temperature-dependent measurement was performed using a Janis probe station.

4.2.3 DFT Calculations

The density functional theory (DFT) study was conducted using the Vienna Ab-initio Simulation Package (VASP) to optimize the structure of anatase TiO_2 to its ground state, with the ultimate goal of determining the energy barrier for Li migration [71]. Initially, the structure of anatase TiO_2 underwent relaxation with a kinetic energy cutoff of 450 eV for the plane wave. Brillouin zone integration was conducted using the Monkhorst-Pack scheme, employing a k-point mesh of 5 x 5 x 5, with the exchange-correlation functional approximated using Perdew-Burke-Ernzerhof (PBE) within the generalized gradient approximation (GGA) [72]. Under the same conditions, a four times larger supercell was used to find the local minima for the Li-doped TiO_2 O-defective before the climbing image nudge elastic band (CI-NEB) calculation [73]. The CI-NEB calculation was carried out using high-performance computers equipped with graphics processing units (GPUs) [74, 75]. To strike a practical balance between convergence and accuracy, the migration path consisted of five images.

4.2.4 Voice Recognition Configuration

The voice data was converted into a cochleagram (frequency channel vs time) using Lyon's passive ear model. After conversion, the processed data consists of 64 channels, each of which would be the input vector applied to the reservoir. To reduce the training data budget, input vectors are binarized first with a threshold value. Each data point of the binarized input vector would be the voltage applied to the physical reservoir, with "1" representing 2 V and "0" representing 0 V (Figure 4.13(b)). After RC processing, we again shrink the data size by evenly sampling eight data points on the time scale and fed into a fully connected layer for classification. The fully connected layer is pre-trained and optimized using the RMS-prop method, and softmax is used to convert the output of the fully connected layer to the classification result. 10-fold cross-validation is used to train (450 samples) and validate (50 samples) the RC system.

4.3 Results and Discussion

4.3.1 Device Material and Structure Characterizations

A schematic of the device structure, consisting of two TiN layers as top and bottom electrodes, and a TiO_x and LiPON bilayer as the active switching layer, is shown in Figure 4.1(a). The top electrode is connected to the source, whereas the bottom electrode is grounded. Figure 4.1(b) shows a lamella with the cross-sectional Focused ion beam scanning electron microscopy (FIBSEM) overview of the device layers, which will later be processed for transmission electron microscope (TEM) analysis. The region of interest is shown with a green rectangular box, and the active region (thinner part in the middle) is surrounded by side walls of SiO_2 passivation.

In the TEM images, shown in Figure 4.1(c), the different material layers are contrasted (from bottom to top, SiO_2 on the substrate, TiN of 50 nm, TiO_x of 8 nm, LiPON of 4 nm, TiN of 45 nm, Al of 70 nm). Both the top and bottom TiN electrodes are polycrystalline, whereas TiO_x and LiPON don't show any visible crystallinity, reflecting an amorphous nature. Next, we performed electron energy loss spectroscopy (EELS), shown in Figure 4.1(d-i), and the different elements, including Titanium, Oxygen, Nitrogen, Phosphorous, and Silicon, are precisely traced in their corresponding layers.

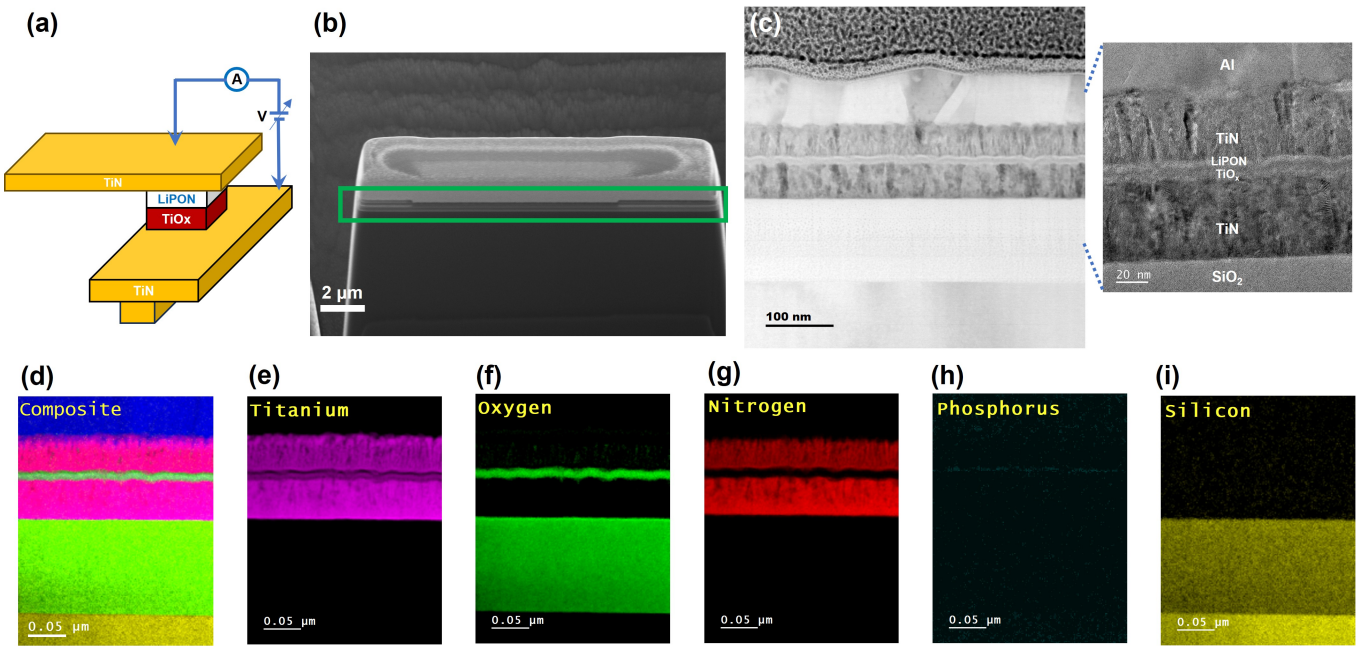


Figure 4.1: Material characterizations of the memory device. a) Schematic of the device structure. b) SEM image of the device lamella profile prepared by FIB showing all layers on the sidewall profile. c) Cross-sectional STEM micrograph of the device with a zoomed-in TEM image of the device showing several layers present on the device. d-i) EELS elemental mapping.

4.3.2 DC Characteristics

Upon the application of linearly sweeping voltages, the device shows a hysteresis loop, meaning that there is a volatile resistive switching from a HRS to a LRS at a positive polarity and from LRS to HRS at opposite polarity, as shown in Figure 4.2(a). The conductance modulation is also a function of sweep speed, as demonstrated in Figure 4.3(a), where a positive current-voltage sweep is plotted with different sweep delays, ranging from 10 ms to 0.5 seconds delay between voltage sweeps. As the speed decreases, the ratio (*i.e.* at a read voltage of 0.1 V) of the conductance modulation becomes smaller due to self-relaxation, indicating more relaxation towards quasi-static states. The reversed rotation in opposite polarities is a clear indication that the volatile conductance change is not from trivial mechanisms such as parasitic capacitance. The inset of Figure 4.2(a) shows that repeatedly sweeping full voltage loops increases the overall conductance.

The device's conductance also depends on the amplitude of the applied biases. Fig-

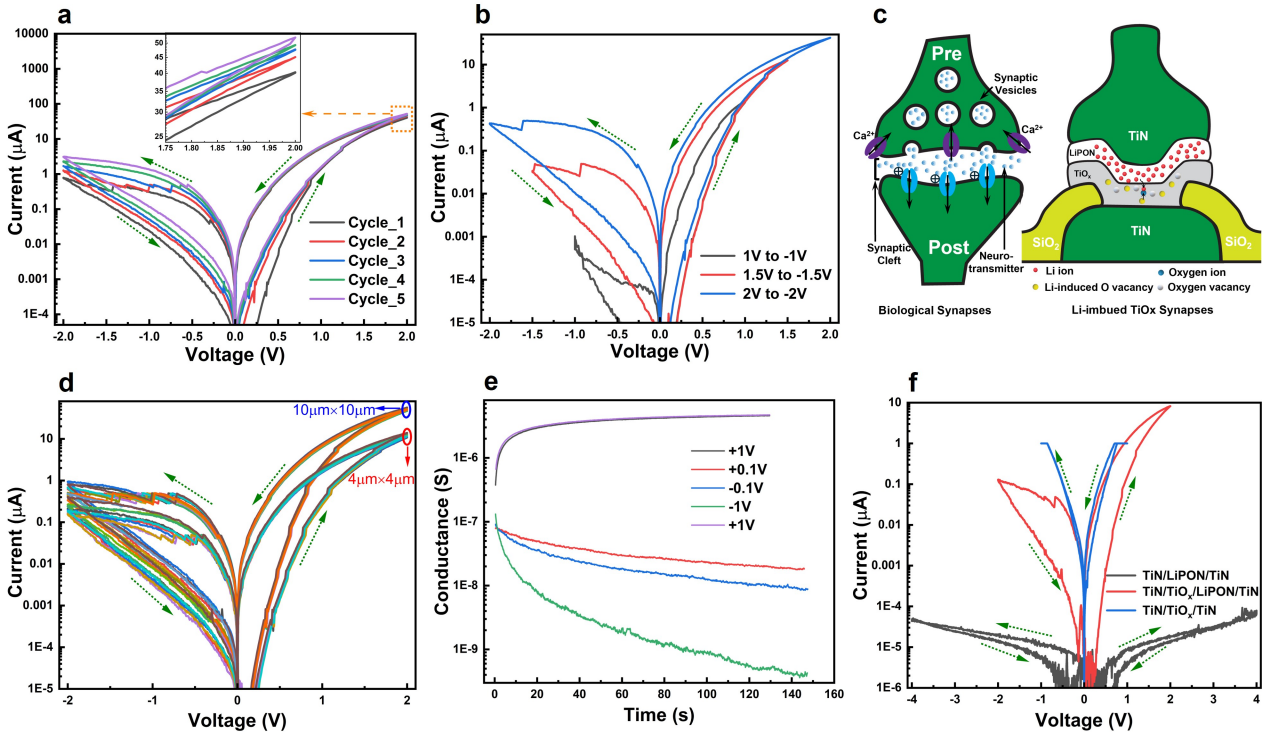


Figure 4.2: DC characteristics of the Li-infused TiO_x memory device. (a) A gradual change in I-V characteristics of the device upon sweeping five full voltage loops, (b) Change in conductivity as a function of increasing voltage, (c) Analogy between biological synapse and Li-ion imbued TiO_x synapse, where Li-induces oxygen vacancies are created due to the migration of Li-ion into TiO_x layer, which then forms a bond with an oxygen atom, leaving the TiO_x more oxygen deficient (reducing the local x), (d) I-V characteristics of ten 4 μm $\times$ 4 μm and ten 10 μm $\times$ 10 μm devices and conductance scales with the area, (e) Self-relaxation and voltage-assisted relaxation from the device (absolute value of the conductance was plotted for negative voltage), (f) I-V characteristics comparison of different control devices before forming, where the switching layers have the same thickness.

ure 4.2(b) illustrates the change in the conductance with different amplitudes of voltage biases. This indicates that with a higher applied voltage, more ions can enter and depart the TiO_x layer and alter the conductance more. Figure 4.3(b) shows the I-V characteristics of five consecutive positive sweeps (0 V - (+2 V) - 0 V) and five consecutive negative sweeps (0 V - (-2 V) - 0 V). When the voltages sweep only in one polarity (half loops), the devices' conductance increases after consecutive positive sweeps and decreases after

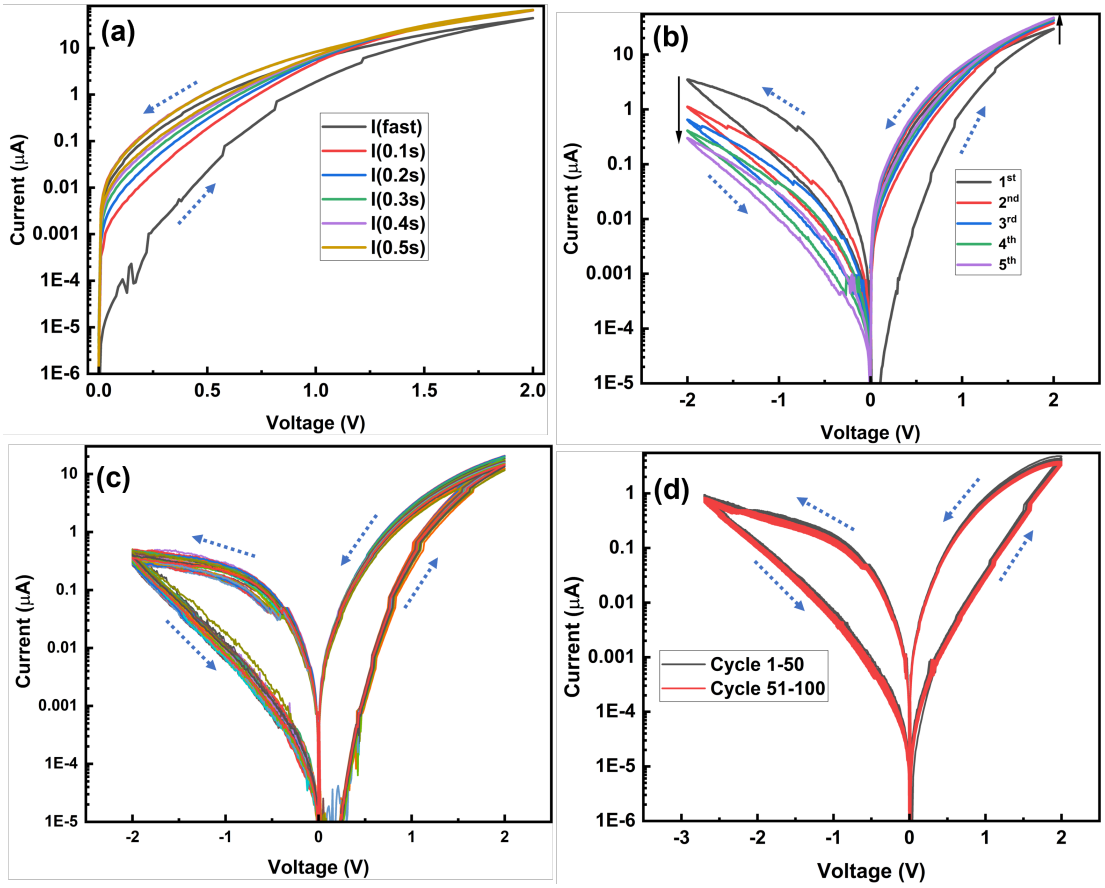


Figure 4.3: (a) Conductance changes as a function of sweep delay between applied voltage. (b) Gradual increase in conductance upon the applications of five consecutive positive sweeps (0 V - (+2 V) – 0 V) and decrease in conductance upon the applications of five consecutive negative sweeps (0 V - (-2V) – 0 V). (c) Current-voltage characteristics of 50 volatile devices distributed across a 1cm-by-1cm area. (d) Endurance test of the device with asymmetric voltage sweep.

consecutive negative voltage sweeps, as shown in Figure 4.3(b). With repeated positive potentials, more Li-ion can migrate into the TiO_x layer and create defects, causing the increase in conductance; and repeated negative potentials do the opposite by depleting ions. This gradual change in conductance is analogous to the efficacy modulation in biological synapses upon the application of the repeated stimulus. Thus, Li-ion modulated TiO_x, before forming, exhibits gradual and volatile conductance modulation and can be used for artificial synapses.

When an action potential is generated at the presynaptic neuron terminal or axon, voltage-gated Calcium channels trigger and Calcium ions (Ca^{2+}) start flowing into the pre-synaptic neuron and synaptic vesicles shift towards the synaptic cleft and release neurotransmitters, which then bind to receptor proteins across the gap and cause Na or K ions to conduct through the post-synaptic neuron, and this process strengthens or weakens the connection strength between synapses. Similarly, in Li-imbued TiO_x iontronic devices, under positive applied bias, lithium ions (red spheres) can migrate into the TiO_x layer as shown in Figure 4.2(c). Here, the yellow spheres represent Li-ion-induced vacancies, and the gray ones are native oxygen vacancies already in TiO_x. Li ions preferably combine with oxygen and chemically reduce Ti, which has similar effects as forming oxygen defects in TiO_x (reducing x). Therefore the Li-ion insertion/extraction increases/decreases the device conductivity accordingly. Li migration is similar to oxygen migration, but it is much easier and happens readily at room temperature and low current densities, i.e., no set/reset voltages/currents or current induced hotspots are needed. The current amplitude can vary well over an order of magnitude in these processes, depending on the history. The devices, therefore, possess history-dependent volatile memory suitable for neuromorphic applications.

The conduction of the device scales well with the device area, which indicates a non-filamentary type of conduction. Figure 4.2(d) shows a total of twenty devices' IV characteristics with two different device sizes: 10 μm×10 μm and 4 μm×4 μm. Also, Figure 4.3(c) shows 50 more (6 μm×6 μm) devices' IV characteristics, which are distributed across a 1cm-by-1cm area. As can be seen from those figures, there is little device-to-device variation as no forming or filament is needed, depicting the good uniformity of the device's volatile characteristics. Moreover, the conductance is proportionally larger for larger area devices. Figure 4.3(d) represents the I-V characteristics of the volatile devices for 100 cycles, meaning the devices were SET followed by RESET 100 times. The devices can sustain repeated cycling without much degradation, representing the robustness of the devices.

Another important phenomenon observed from the devices is their self-relaxation behaviour. To understand this, several systematic tests were performed. First, we applied a very small read voltage of -0.1 V to check that there was no conductance change from the fresh device. Once we applied +1 V to the device, the conductance increased (changed from HRS to LRS), indicated by the black color curve in Figure 4.2(e). Next, a small read voltage of -0.1 V (blue color) and +0.1 V (red color) were applied to capture the self-relaxation towards lower conductance, confirming the volatility of the memory device. A higher negative bias of -1.0 V shows similar but voltage-assisted relaxation to a much lower conductance state (green color curve), which largely accelerates the relaxation process and can be a way to speed up the system for practical applications. After changing the polarity

back, the conductance returns to the original behaviour, represented by the violet color curve. It is worth noting that due to the cathodic nature of TiO_x, the center voltage of the ion insertion/extraction is not at 0 V, which leads to the monotonic conductance drift between cycles depending on the voltage sweep window. As an example, when the negative voltage is not enough to fully extract the ions, the corresponding conductance drifts up appreciably from cycle to cycle (Figure 4.2(a)). When sweeping asymmetrically instead (-2.7 V to 2 V, Figure 4.3(d)), the cycle-to-cycle drift becomes negligible. Therefore the thermodynamic center of the system is approximately at -0.35 V, indicating a spontaneous redistribution of Li ions due to chemical potentials. This is also the origin of the system exhibiting faster conductance change at 1 V than at -1 V (Figure 4.2(e)).

The strong volatile behaviour in these devices is a combined result of the electrolyte LiPON and the cathodic TiO_x. To verify this, we fabricated control devices with only TiO_x and only LiPON of the same 12 nm thickness, and the characteristics are shown in Figure 4.2(f). The LiPON-only devices are extremely resistive and exhibit different hysteresis changes (dashed arrows, curve rotation is different from the volatile changes) due to lagged charging and discharging of their capacitance with a large time constant. Whereas, the TiO_x-only devices before forming have small hysteretic behaviour because the TiO_x is populated with some oxygen vacancies that can also migrate, especially under electric fields, and more so in hot spots (such as during the forming process). The LiPON/TiO_x hybrid devices have intermediate conductance as expected, and the much stronger volatile change can be attributed to their mutual influence. When LiPON was grown with reactive sputtering from the Lithium phosphate target, some Li ions were already incorporated inside the TiO_x layer and created ionic defects ready for voltage manipulation. LiPON functions as a Li source and reservoir and can absorb or release ions at applied biases. TiO_x functions as the cathode and its conductivity varies with the concentration of oxygen vacancies [76, 77]. Also, TiO_x is often used as electrodes in rechargeable batteries where Li ions can travel in and out of the nanostructures, easily forming intercalated ion storage and reducing TiO_x [78, 79, 80, 81]. With the ion intercalation and modified TiO_x stoichiometry, the conductivity of our devices increases or decreases accordingly, analogous to the synapse systems in its stimulated conductivity modulations.

4.3.3 Pulse Characteristics

To emulate the brain-like synaptic functions, pulse-dependent characteristics from the devices are essential as most of the signals in the biological brain are similar to short pulses. Synaptic plasticity, which is defined as the proficiency of neurons to adjust the level of connection strength between them, is one of the most important aspects of neuromorphic

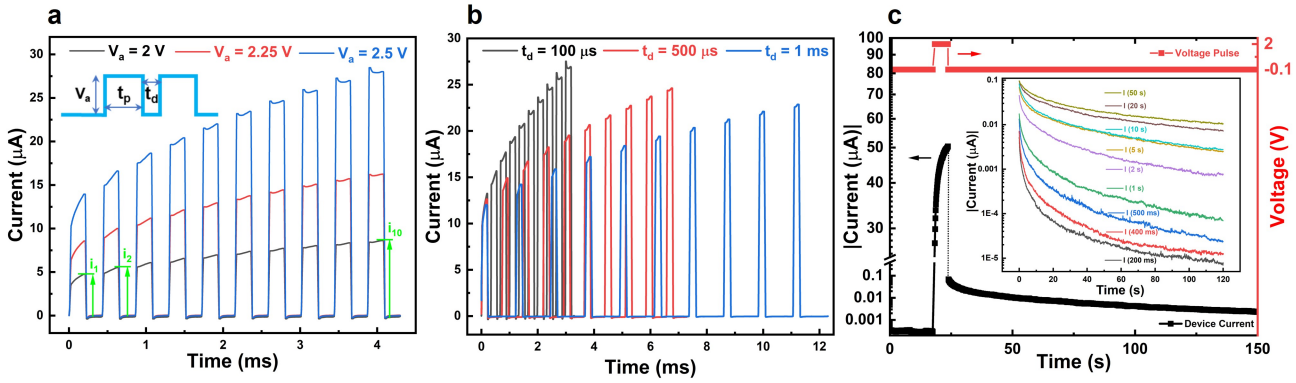


Figure 4.4: Device behaviour upon the application of pulse voltage. (a) The device's response upon varying pulse amplitude and inset depicts the key parameters: V_a, pulse amplitude; t_p, pulse duration; t_d, pulse delay. (b) Device current upon varying pulse delay. (c) Continuous bias and its corresponding conductance. Inset shows the conductance decay at different excitation durations.

computing [82]. In biological synapses, once an action potential is generated at the axon, it induces a Calcium ion influx into the presynaptic neuron, causing the synaptic vesicles to move toward the synaptic cleft. Upon reaching the synaptic cleft, the synaptic cleft releases neurotransmitters, which then migrate toward the postsynaptic receptors, temporarily modifying the synaptic efficacy. If there is another identical action potential generated before the recovery time of the calcium ions, the response from the postsynaptic neuron becomes higher compared to its first one. This effect is known as paired-pulse facilitation (PPF) [83, 84, 85]. Likewise, a train of consecutive identical pulses of a certain duration results in a gradual augmentation of synaptic plasticity, which is known as post-tetanic potentiation (PTP) [86].

The inset of Figure 4.4(a) shows the key parameters of the applied pulses on the devices, including pulse amplitude V_a, pulse duration t_p, and pulse delay t_d. The indexing of the PPF is defined as the ratio of the current value of the second pulse relative to the current value of the first pulse, whereas that of PTP is represented by the ratio of the current value of the 10^{th} pulse to the current value of the first pulse. We first vary the pulse amplitudes to observe their effects on device conduction. As can be seen from Figure 4.4(a), when the applied potentials are higher in amplitude, both the PPF and PTP increase much faster. This fact can be explained by the amount of ions participating inside the device is higher when the electric field is higher, resulting in a much higher conductance increase. The potentiation and depression characteristics are shown in Figure 4.5(a), where the pulse

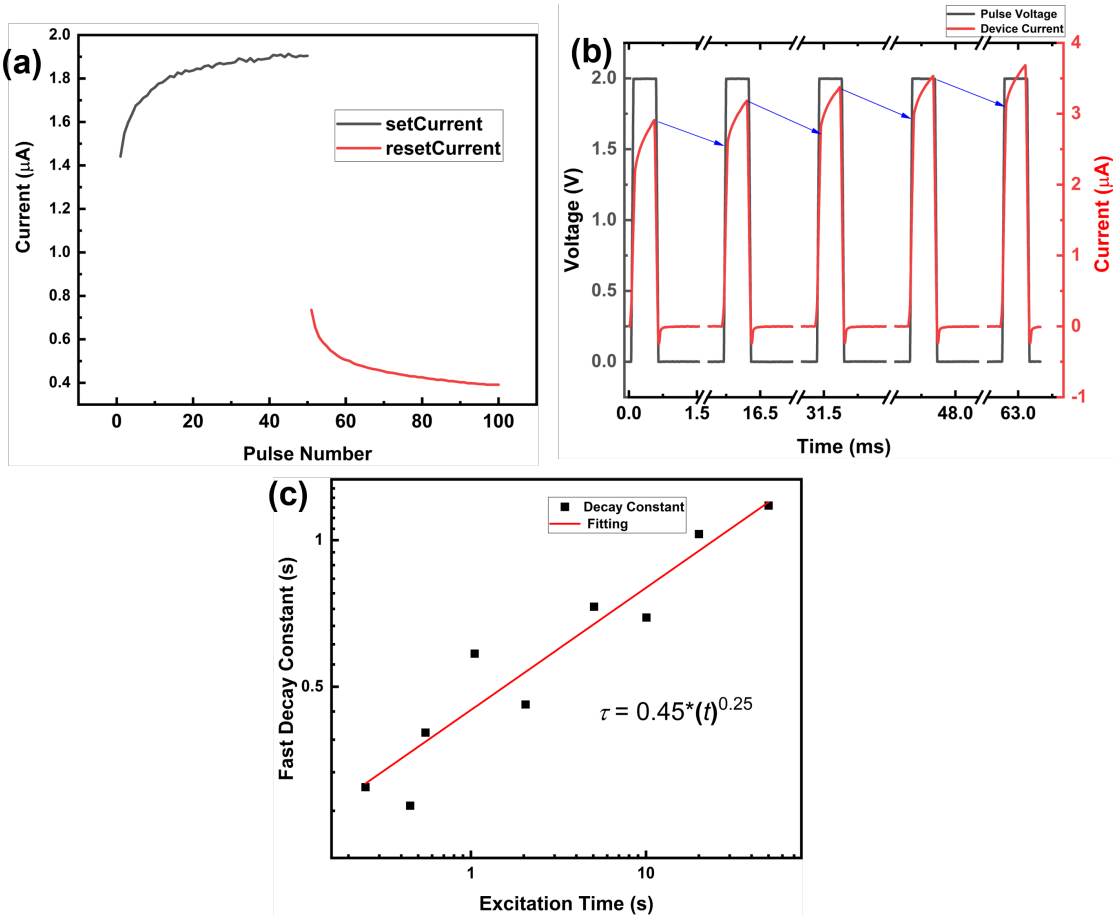


Figure 4.5: (a) Potentiation and Depression characteristics from the device upon the application of positive and negative voltage pulses. (b) Partial conductance decay when the pulse delay is 15ms. (c) Power-law fitting result of the conductance decay as a function of excitation duration.

amplitude, pulse width, and pulse interval are kept the same. The first fifty pulses are positive and result in potentiation (conductance increase) from the device and the next fifty pulses are negative and result in depression (conductance decrease) characteristics. The input voltage for potentiation is 2.5 V with a pulse width of 0.5 ms and for depression is -3.5 V with a pulse width of 0.5 ms. Like DC Sweep, consecutive positive pulse excitation (SET pulses) results in augmentation in conductance, known as potentiation, and consecutive negative pulse excitation (RESET pulses) results in a decrease in conductance, known as depression.

Synaptic plasticity is a function of the spike rate as well. Spike rate modifies the number of neurotransmitters from the presynaptic neuron, which modifies the strength of synapses; this process is known as spike-rate dependant plasticity (SRDP). Analogous to biological synapses, the conduction also depends on the delay between pulses as the competing self-relaxation behaviour exists. To test this assumption, consecutive voltage pulses with a certain delay are applied to the device and corresponding currents are measured and plotted in Figure 4.5(b). The applied pulse voltage is 2 V, and the pulse width is 0.5 ms, whereas the pulse interval is 15 ms. With the high pulse interval value, some ions relax back, and the conductance increase is slower. The arrow in Figure 4.5(b) indicates the change in the conductance values. This is why a longer delay results in a smaller final current despite the accumulated pulse time remaining the same, which is depicted in Figure 4.4(b). Thus, Li-ion imbued TiO_x shows the SRDP of a biological brain and can be used for artificial synaptic devices.

A continuous biasing approach was performed to determine the timescale of the volatile memory device. To capture the decay after excitation, we incorporated continuous bias of different time intervals and amplitudes (for example, 2 V for 5 seconds was used to capture the potentiation; and to capture the decay, a small negative voltage, -0.1 V, was used). Figure 4.4(c) shows the overall voltage pulse and corresponding current. Also, the inset of Figure 4.4(c) shows that a larger duration of voltage stress results in slower conductance decay, which can be attributed to a large number of Lithium-ion migrations. Next, the current decay (durations of 1 s to capture only the relevant, fast relaxation) was fitted with an exponential decay and extracted the time scales. The extracted decay constant is then plotted as a function of the excitation pulse widths. As can be seen from Figure 4.5(c), the fitted results can be approximated with a power-law relationship. The curve is best fitted to a general behaviour of Equation 4.1, where t is the excitation time and τ is the approximate decay constant.

$$\tau = 0.45 * (t)^{0.25} \tag{4.1}$$

Therefore, the short-term memory can be tuned with the excitation duration, which was used in the later neuromorphic simulations to better represent the system's true behaviour. Thus, our device shows tunable volatile memory feature.

4.3.4 Temperature Dependent Characteristics

The participation of ions can be distinctly identified by varying the device temperature, as ions tend to freeze out easily at reduced temperatures. The temperature was varied over a wide range, from 40 K to 350 K, while the voltage sweep was fixed from 0 V to

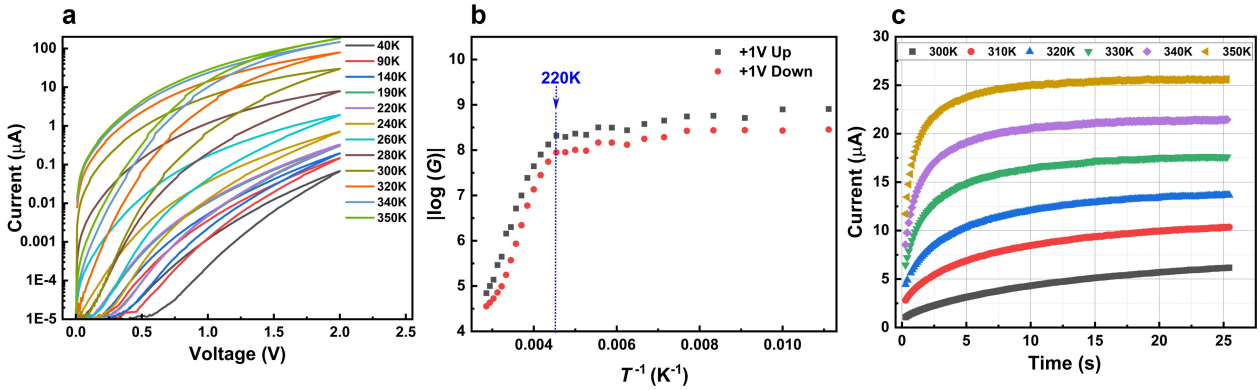


Figure 4.6: (a) Temperature-dependent current-voltage characteristics (b) Resistance value at +1V for two different sweep conditions: one at up sweep and another is down sweep. (c) constant-voltage stress at different temperatures with an applied voltage of +1V.

2 V (up sweep) and then 2 V to 0 V (down sweep). Figure 4.6(a) shows that before the temperature reached a certain threshold value, the IV loop openings were small and device behaviour did not change much, but changes were more pronounced above the threshold temperature, meaning some thermally activated conduction channels are now participating. To determine the threshold value, we extract the discrete current data value at +1 V at both up sweep and down sweep and plot it as a function of the inverse of Temperature, i.e. 1/T. There is a clear slope change at 220 K as shown in Figure 4.6(b). This transition can be attributed to lithium ions becoming frozen below this threshold temperature, which is close to the reported freezing temperature of Li ions. Therefore, only thermally promoted Li ions are able to migrate. On the other hand, the relative change under sweeping voltage is a result of the applied electric fields driving the already promoted ions, therefore not as temperature dependent. Similarly, Figure 4.7(a) shows that the absolute relative change of current value between up sweep (I_{up}) and down sweep (I_{down}) at +1 V is increasing above the transition temperature. Above the transition temperature (220 K), the volatile hysteresis is more pronounced.

The slope of the semilog plot of the conductance with respect to 1/kT, which is known as the activation energy barrier, was calculated from Figure 4.7(b), where the conductance values are recorded above the transition temperature. The activation barrier was found to be around 0.20 eV, which approximately corresponds to the density functional theory (DFT) calculated value (a low energy barrier of around 0.4 eV). In the O-defective TiO_2 supercell shown in Figure 4.8(a), the green spheres along the path represent lithium (Li)

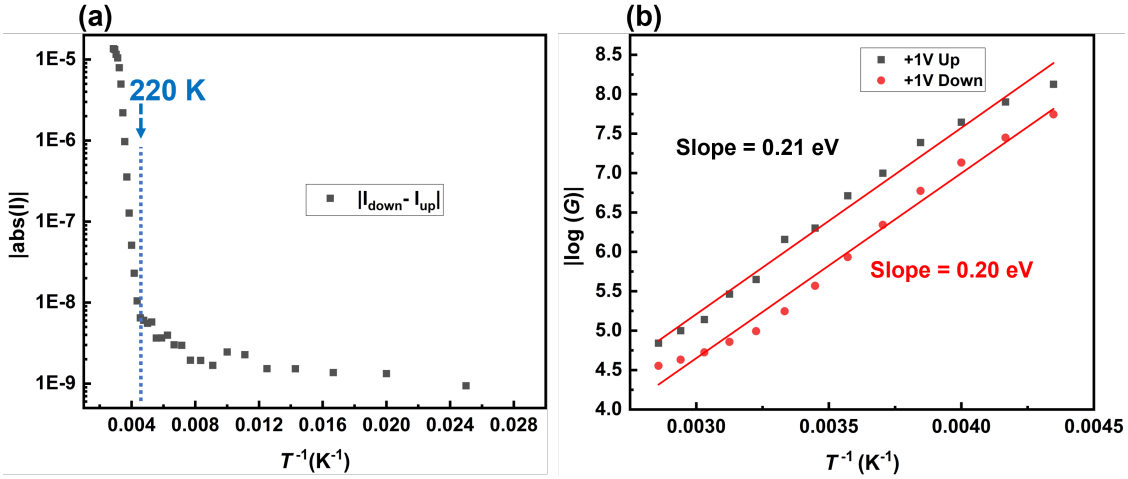


Figure 4.7: (a) Relative values of the current changes at different temperatures @+1 V. (b) Slope above transition temperature.

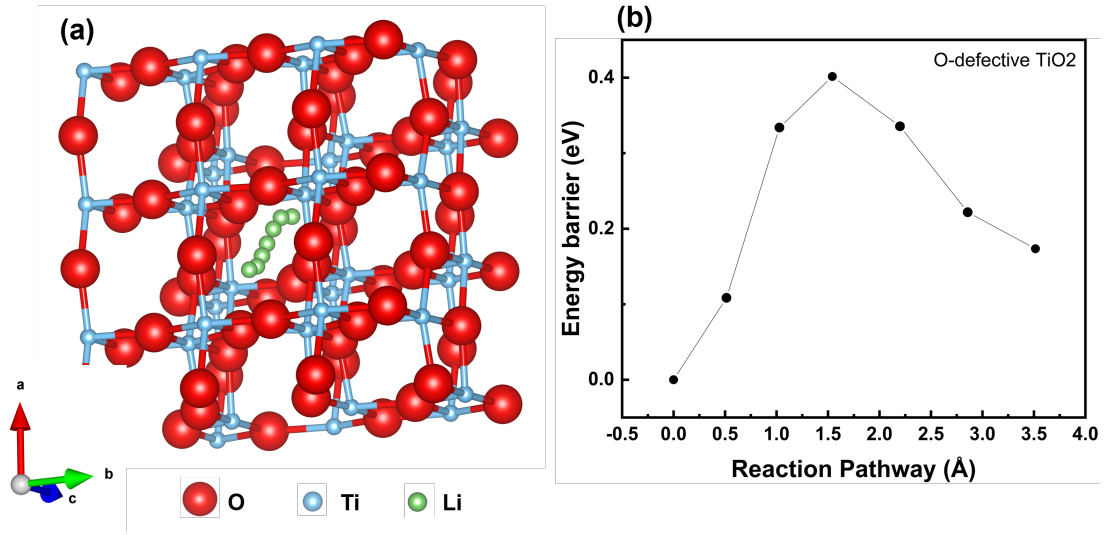


Figure 4.8: (a) Li migration pathway in the O-defective structure of anatase TiO_2 using ionic radii style. (b) The energy barrier profile for Li migration obtained from DFT-NEB (Density Function Theory–Nudged Elastic Band) calculations shows a barrier of 0.4 eV.

atoms at different positions, illustrating their optimized migration path, i.e., the most energetically favorable path for Li-ions. The graph in Figure 4.8(b) shows the energy

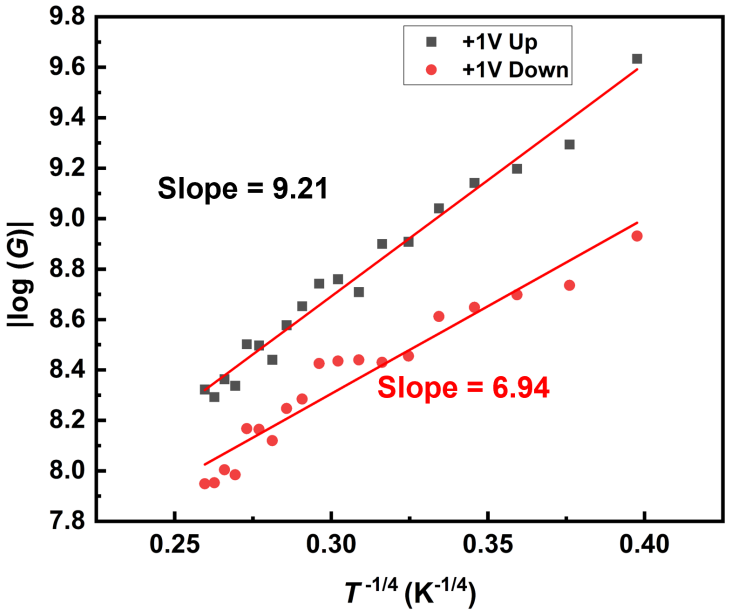


Figure 4.9: The slope below the transition temperature.

barrier for Li migration along the reaction pathway shown in Figure 4.8(a). The energy barrier peak at around 0.4 eV indicates the maximum energy required for Li to migrate between two local minima. This low energy barrier suggests that Li ions can move relatively easily in the O-defective TiO_2 structure, implying high ionic conductivity, which is crucial for low-operating voltage devices. This energy barrier is comparable with the reported energy barrier for pristine anatase TiO_2 structures [87].

Now, below the transition temperature, the Li ions are frozen as there is a minor conductance change. Figure 4.9 illustrates the absolute log value of conductance as a function of the temperature ($T^{-1/4}$). A linear relationship between conductance and the temperature ($T^{-1/4}$) can be seen from Figure 4.9, indicating that the conduction mechanism is mostly dominated by variable-range hopping (VRH), which is much less temperature-dependent. This conduction mechanism is similar to reported mott VRH at low temperatures [88, 89].

The constant voltage stress curves above the threshold temperature are shown in Figure 4.6(c). The applied voltage was +1 V. It is clear that higher temperature leads to higher final conductance and faster conductance change. Faster ionic motion is expected for higher temperatures, therefore the system can reach dynamic equilibrium significantly faster. The temperature is a useful controlling knob for the system if tuning operation speed is necessary.

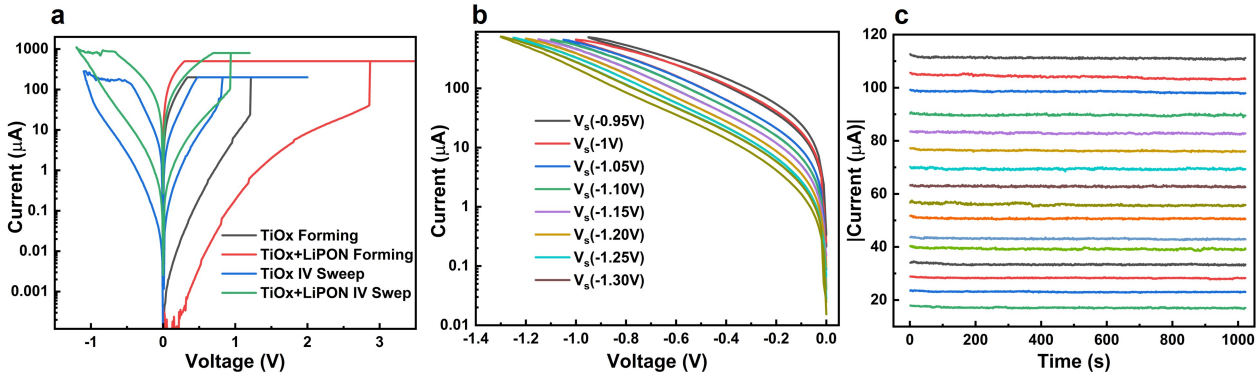


Figure 4.10: (a) Forming process comparison of TiO_x and Li-imbued TiO_x memory devices, (b) Multi-state capability of the proposed device when stopped at different RESET stop voltages, which is denoted by V_s, (c) Retention behaviour of the 16 states from the memory device.

4.3.5 Non-volatile Characteristics

The same volatile memory device can be configured into a non-volatile one by an electro-forming process. The forming process introduces conductive filaments in Li-imbued TiO_x resistive switching devices, and the device is SET to LRS. The device can be RESET to HRS by applying opposite polarity voltages. We compare the electroforming behaviour of hybrid 8nm TiO_x/4nm LiPON devices with pure 8nm TiO_x control devices, shown in Figure 4.10(a), and the former need higher forming voltage and larger compliance current to form appropriate conductive filament width. This is reasonable as LiPON is a good electrical insulator and a large fraction of the overall applied voltage is dropped across this layer instead of TiO_x.

After forming the devices, we perform the RESET and SET operations for both devices. For the reset operation, the voltage was kept at a fixed value rather than a compliance current as current may vary from device to device. As clearly shown in Figure 4.10(a), the Li-infused TiO_x becomes non-volatile after filament formation. Overall, these devices have higher conductance despite the total layer thickness being higher, because TiO_x becomes more defective in the presence of Li ions, and LiPON likely experiences dielectric breakdown and no longer participates.

Interestingly, this non-volatile device can be reset to multiple resistance states without overlapping in between, meaning it is capable of storing multibit information. Followed by progressive RESET sweep (for example: 0 V – (-0.66 V) – 0 V), the retention behavior was

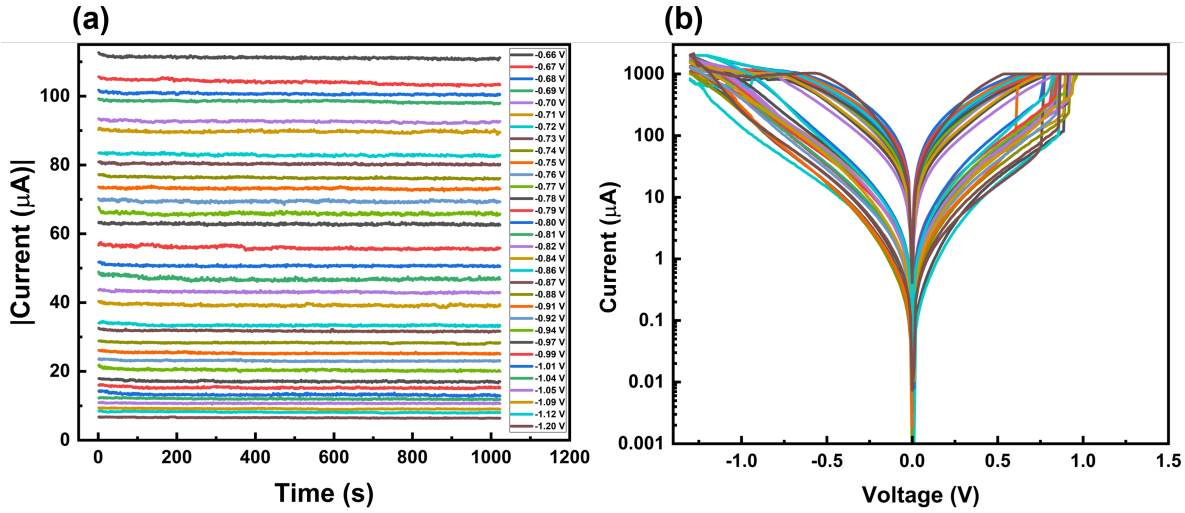


Figure 4.11: (a) Retention characteristics of the device after applying an incremental reset voltage sweep. With an increment of 0.01V, we can set 32 states easily with a margin of approximate 0.5 muA. (b) Twenty Li-imbued TiOx devices IV characteristics, showing the device-to-device variations.

monitored by applying a constant bias of -0.2 V for 1000 seconds. As we stop at different reset voltages, the current from the device is gradually reduced, as shown in Figure 4.10(b). Figure 4.11(a) represents the retention characteristics after each reset test of the device. As can be seen from the figure, the devices retain more than 32 states for thousands of seconds without much noticeable change. This ensures sufficient margins for programming 4-bit (16-state) information on each device as shown in Figure 4.10(c), which will be assumed in the next session for the proposed circuit integration.

Figure 4.11(b) depicts twenty non-volatile Li-imbued TiO_x devices' IV characteristics. There are some device-to-device variations due to their stochastic filamentary nature, which is reasonable for TiO_x-based devices fabricated with in-house facilities. It is worth pointing out that the Li-containing devices' transition from volatile to non-volatile is irreversible, and once the conductive filament is formed, the system behaves very similarly to pure TiO_x devices with non-volatile memristor behaviour.

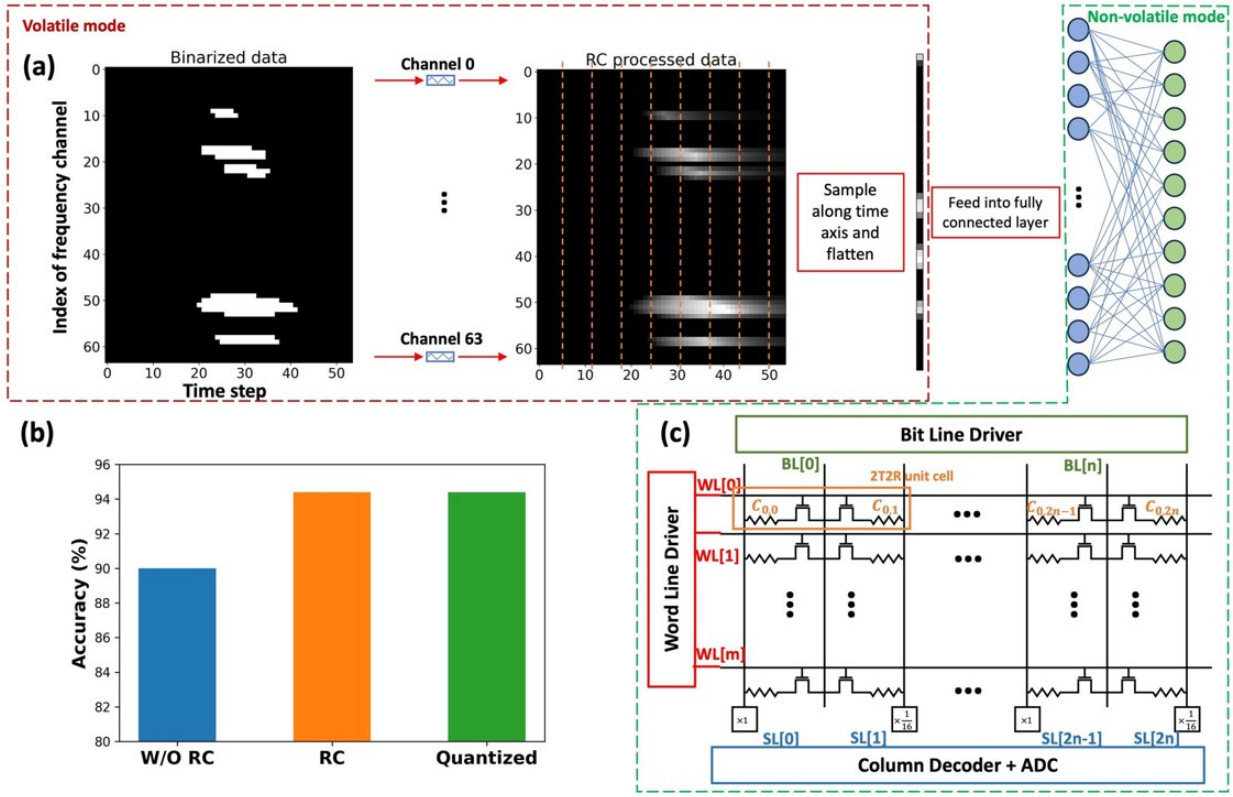


Figure 4.12: Voice recognition task using proposed memory. (a) Process flow of RC for voice recognition. (b) Comparison of classification accuracy between quantized RC model, RC model and model without RC processing. (c) Schematic of proposed 2T2R CIM array.

4.3.6 Memristor-based Reservoir Computing

Reservoir computing (RC) is a computation framework that maps input signals to higher dimensional space, through a non-linear mapping "reservoir". "Reservoir" can be a dynamic system that evolves in time, which can be expressed as f(t+1) = f(t, input). The volatile switching behaviour of our fabricated device, however, shares similar characteristics with a dynamic reservoir: the future resistance state, f(t+1), is dependent on the applied voltage (input) and current resistance state, f(t). To illustrate the neuromorphic capabilities of our devices, a physical reservoir computing system is proposed for voice recognition tasks. The workflow of the RC system is illustrated in Figure 4.12(a). After pre-processing the raw voice data using Lyon's passive ear model, the generated Cochleagram (frequency channel vs time) is normalized and binarized (with a threshold value of

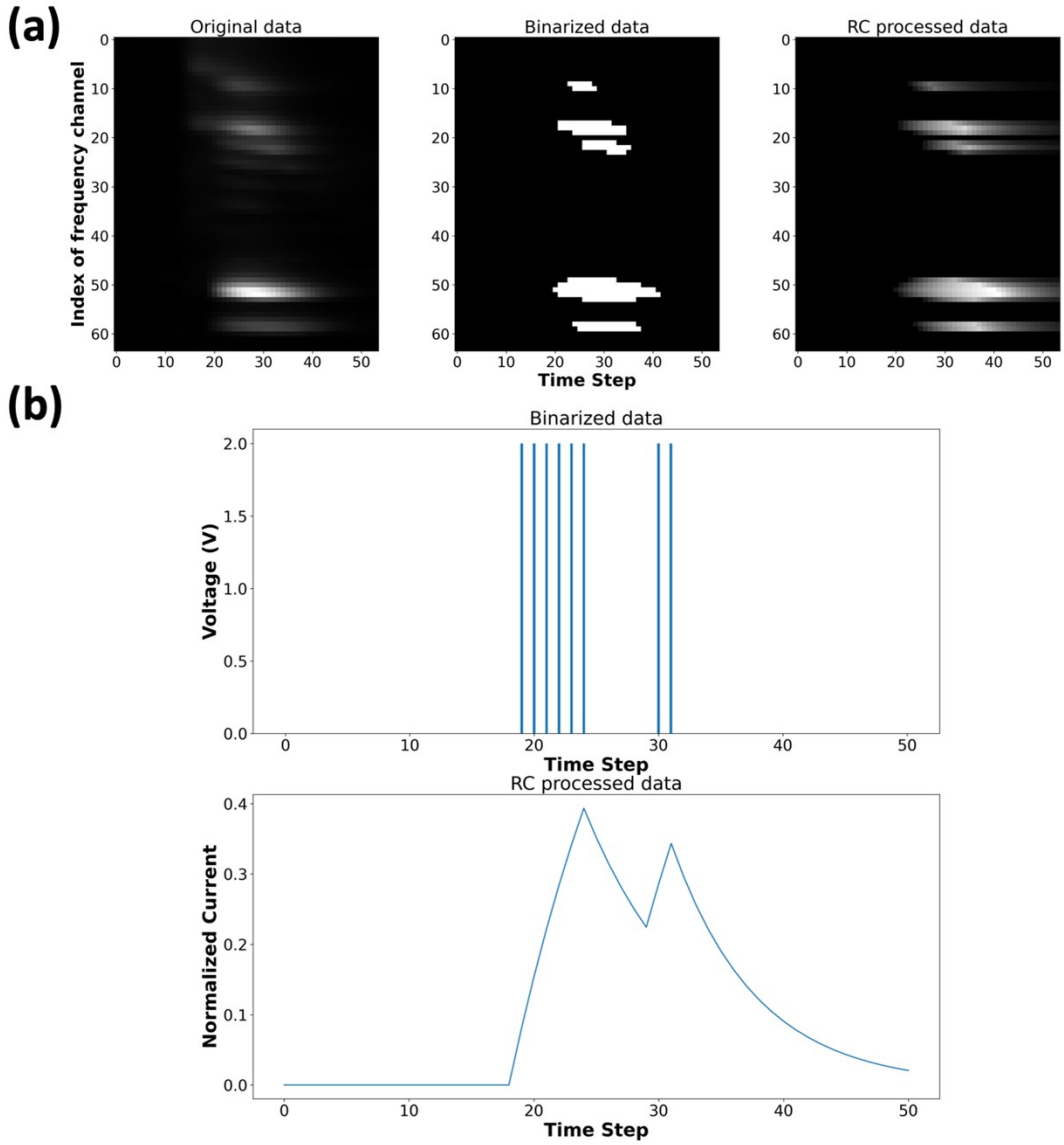


Figure 4.13: (a) Cochleagram of voice data, binarized data, and RC processed data. (b) Voltage input of channel–30 and corresponding current of the volatile device at time scale.

0.2). The x-axis is the time steps of the voice, whereas the y-axis represents the index number of the frequency channels, which are within human audible ranges.

The Cochleogram of the original voice data is shown on the left of Figure 4.13(a). The

pixels are the vector values, which are then binarized with a certain threshold and plotted in the middle of Figure 4.13(a). The data vector of each frequency channel (along the time step axis) is converted into a voltage input vector (with data "0" representing 0 V and data "1" representing 2 V), and the time step of the voltage input vector is 33 ms. The voltage input vector of each channel is fed into the volatile device connected with the channel. The Cochleogram of the RC processed data from binarized input data is shown on the right of Figure 4.13(a), which almost perfectly resembles the original voice data. Figure 4.13(b) represents the voltage input of a frequency channel (30) and the corresponding current from the volatile memory device. As can be seen from the figure, the conductance of the volatile device will change (potentiate or self-relax) based on the voltage input vector. Next, the conductance value of each device is evenly sampled along the time axis (8 data points sampled in our simulation). Sampled conductance values of 64 channels are flattened into a 1D feature vector and fed into the fully connected layer to generate the classification results.

The whole system has been trained for 300 epochs, the classification accuracy of the RC model, defined as the average accuracy of the ten-fold cross-validation, is 94.4%, which outperforms the binary data without RC processing using the same network architecture by 4%, as shown in Figure 4.12(b). An experimentally determined variable time scale, as shown in Figure 4.5(c), was used in our simulation. If we also use the conventional constant time scale assumption, the simulated accuracy can well exceed 97% for this particular task. The accuracy improvement is achieved without computation overhead. A performance comparison between existing literature and our demonstrated device is shown in Table 4.1. It is clear from the table that our devices show tunable volatility and multi-bit non-volatile capability, which is special in a single device. Our platform excels with great device tunability and platform customizability.

The fully connected layer discussed above is based on software simulation, each weight inside the layer is stored in the format of a 32-bit floating point number (float32). Arithmetic operations, as well as storage of floating-point numbers, are resource-intensive and not energy-efficient for hardware implementation. To explore the possibilities of scaling down the representing bits of weights, post-training quantization (PTQ) is used to map the trained weights from float32 to int8. The quantized model is validated using the same validation dataset, and the results show that there is no accuracy drop after quantization. Based on that, a 2T2R compute-in-memory (CIM) array is proposed (shown in Figure 4.12(c)) to accelerate the operation of vector-matrix multiplication, which is used inside the fully connected layer, with our devices working in non-volatile mode.

With each device programmed to store 4-bit data, the 2T2R unit cell consists of two non-volatile devices and two access transistors. Two non-volatile devices inside the 2T2R

Table 4.1: Performance parameter comparison between the existing literature and proposed device structure.

Device	Volatility	Volatile Time Tunability	Non-volatility	Multi-bit Capability	Accuracy	Ref.
Pd/SiO_2:Ag/Pt	Yes	No	No	No	83%	[41]
Pd/Au/WO_x/W	Yes	No	No	No	99.2%	[42]
Ti/(TiO_x/TaO_y)/Pt	Yes	No	No	No	99.6%	[43]
Pt/ SiO_2:Ag/Pt	Yes	Yes	No	No	99.6%	[90]
TiN/TiO_x/LiPON/TiN	Yes	Yes	Yes	Yes	97.2%	[This Work]

unit cell are used to store one 8-bit weight, the left device is used to represent the most significant 4 bits (MSB) and the right is used to represent the least significant 4 bits (LSB). Both devices share the same bit line signal (V_{BL_n}), therefore, current flowing through the MSB source line ($I_{SL_{2n-1}}$) would be

$$I_{SL_{2n-1}} = \sum_{m=0}^{M} V_{BL_{2n}} * C_{2n-1,m}$$

, while the current for LSB line ($I_{SL_{2n}}$) is

$$I_{SL_{2n}} = \sum_{m=0}^{M} V_{BL_n} * C_{2n,m}.$$

I_{LSB} is fed through a 1/16 current mirror so that the sum of I_{MSB} and I_{LSB} can be used to represent a 8-bit value. The current sum, ($I_{MSB} + 1/16 * I_{LSB}$), is sampled by ADC, which would be digitalized and output the 8-bit product results.

4.4 Conclusions

In summary, we have demonstrated a Li-ion imbued TiO_x iontronic device, with both volatile and non-volatile memory capabilities. Before forming, the devices show inherently

uniform gradual conductance change upon the application of external bias due to the controlled ionic motion. A self-relaxation phenomenon confirms the volatile nature of the memory. The Li-ion-induced vacancy creation inside TiO_x resembles Calcium-ion-induced synaptic modulation inside biological synapses. Pulse-dependent results strongly resemble the different brain functionalities such as PPF, PTP, and SRDP characteristics.The timescale of the volatile memory devices can be tuned by changing the voltage pulse widths. In addition, the same volatile device can be transformed into non-volatile memory by an electro-forming process. To illustrate their performance, we applied both tunable volatile and non-volatile versions of the devices to a proposed reservoir computing architecture and performed voice recognition tasks with over 94% accuracy. Therefore, we've demonstrated a general iontronic platform that is customarily configurable to mimic brain functions on a hardware level.

Chapter 5

Helimagnet-based Non-volatile Multi-bit Memory

Overview

In this chapter, a design of a helimagnet-based emerging memory device that is capable of storing multiple bits of information per device is presented. The device consists of a helimagnet layer placed between two ferromagnetic layers, which allows us to lock in specific spin configurations. The bottom pinned layer has high anisotropy energy or stays exchange biased, which keeps its spin configuration fixed on a specific direction, while the top layer is free to rotate under the influence of in-plane fields. We begin by finding the relaxed spin structure, which is the result of the competition between the Dzyaloshinskii–Moriya interaction (DMI) and exchange energy and is referred to as the equilibrium state ("0"). The writing of a memory state is simulated by applying an in-plane field that rotates and transforms the spin configurations of the memory device. Our results indicate that stable configurations can be achieved at rotations of an integer multiple of 180 degrees (corresponding to states "-2", "-1", "1", "2", etc.), where the anisotropy stabilizes the free layer and thus the exchange coupled helimagnet. These states are separated by magnetic energy barriers and intermediate, unstable spin configurations tend to revert to their adjacent states. By simply changing the direction of the field, we can achieve multi-bit data storage per unit memory cell. The maximum number of bits is reached when the anisotropy energy barriers cannot withstand the strong DMI energy. Reading can be done by evaluating the different resistance states due to the twisted spin texture.

5.1 Introduction

Due to the advancement of information technology and computing, the use of electronic devices has been rapidly increasing. As all these devices have integrated memory systems, a substantial increase in their storage capacity has occurred to fill the growing need. Different memory devices, including existing and emerging memory, have been used to fulfill this demand. Existing memory devices, which include static random-access memory (SRAM), dynamic random-access memory (DRAM), and Flash memory, have several drawbacks when considering low power consumption and high-density issues; for example, larger footprints are required by SRAM as the single-bit cell area is equivalent to 150 to $300F^2$, where F is the lithography feature size [91, 92], periodic refresh is needed by DRAM [93], and limited cycles can be withstood by Flash memory [94, 95]. On the other hand, emerging spin-based memory devices that include spin-transfer torque magnetic random-access memory (STT-MRAM) [96, 97, 98], and domain wall motion (DWM)-based racetrack memory [99, 100], and Skyrmion-based memory [101, 102], have the potential to overcome the above-mentioned constraints of conventional memory.

Spin configurations with a stable spatial variation, such as helimagnets and skyrmions, have recently attracted a great deal of interest in the research community as they are promising candidates for magnetic memory, spin-logic devices, and future spintronics applications [103, 104, 105]. One type of spatial variation, where magnetic moments are spatially rotating in a plane that is perpendicular to the propagation plane, is known as a bloch-like chiral helimagnet [106]. In a standard micromagnetic system, this chiral helimagnet can be realized by the competition between exchange energy and Dzyaloshinskii–Moriya interaction (DMI) that arises due to a lack of inversion symmetry in non-centrosymmetric materials [107, 108]. Upon the application of an external magnetic field, the helical period of the helimagnet can be tuned because the Zeeman energy, which arises upon the external magnetic field, tends to align spins along the applied field direction [109]. In this letter, we propose a memory device structure named as helimagnet-based random-access memory (HMRAM), consisting of a helimagnetic layer sandwiched by two ferromagnetic layers, with the capability of multi-bit storage and non-volatile memory characteristics. The exchange-coupled free layer on one side of the helimagnet (the other side is strictly pinned) can follow the external fields, thus applying controllable torques onto the helimagnet. Once the external magnetic field is removed, the free layer tends to align to its anisotropy axis and lock the helimagnet into predefined configurations. Thus, a non-volatile behavior can be achieved in the helimagnet in the absence of an external magnetic field. This spring-like helimagnet can be tuned in 180° increments until it can no longer hold more torque. The number of stored bits is, therefore, determined by the strength of DMI in the helimagnet

and the anisotropy energy in the free layer.

We first show how the helimagnetic layer can be formed as a result of the competition between exchange energy and DMI energy. After that, we examine how a helimagnet period can evolve as a function of DMI energy while maintaining constant exchange stiffness. We then proceed to illustrate the writing mechanism in our proposed memory system. A complete 360-degree cycle rotation of the top free layer as a function of an applied magnetic field is depicted. It is found that the helimagnetic layer follows the free layer and starts twisting in the presence of a magnetic field, while at zero fields, the free layer's anisotropy energy locks the system in non-volatile spin configurations. The non-volatile states' helicity can be increased or decreased by simply rotating the external field in the clockwise or counter-clockwise directions. This way we can write multi-bit memory states and they are stable without an external magnetic field. Since there is a change in the relative spin angle within the helimagnet, reading of the memory states can be done simply by sensing the resistance change as a result of spin scattering [110].

5.2 Simulation Details

In this research, Object Oriented MicroMagnetic Framework (OOMMF) was used to simulate the dynamics of the magnetization due to its accuracy and acceptance as a standard micromagnetic problem solver [111]. The total magnetic free energy, E, of our system is comprised of various energy densities, including exchange energy density (w_E), DMI energy density (w_D), Anisotropy energy density (w_A), and Zeeman energy density (w_Z). The energy terms selected are in agreement with the experimental evidence of helimagnet [112, 113, 114]. Thus, Energy, E, can be written as:

$$E = \int [w_E + w_D + w_A + w_Z] d^3r \tag{5.1}$$

Exchange energy density, w_E, term can be expressed as:

$$w_E = A[(\nabla m_x)^2 + (\nabla m_y)^2 + (\nabla m_z)^2] \tag{5.2}$$

where A is the exchange stiffness of the material and m_x, m_y, and m_z are the x,y, and z components of the magnetization vector $\mathbf{m}$, respectively. Magnetization vector $\mathbf{m}$ is defined as the ratio of normal magnetization, M, to the saturation magnetization, M_s. To replicate the layered quantum material that has a strong in-plane exchange while a

Table 5.1: List of parameters values to realize the HMRAM

Name of the Parameters	Values
Dimension	X: 500nm, Y: 500nm, Z: 120nm
Cell Size	X: 5nm, Y: 5nm, Z: 2nm
DMI, D (J/m^2)	Helimagnet Layer: 2.5×10^{-3}
Anisotropy, K (J/m^3)	Free Layer: 10^6 Pinned Layer: 10^{10}
Exchange Stiffness, A_z (J/m)	Free Layer: 5×10^{-11} Helimagnet Layer: 1×10^{-11} Pinned Layer: 1×10^{-11}
Saturation Magnetization, M_s (A/m)	10^6

small out-of-plane exchange component, we have modified the exchange energy extension provided by OOMMF(exchange6ngbr). In the extension, the x,y, and z components are assumed to be equal, whereas in our case, we have considered the x and y component is 100 times higher than the z components in order to simulate the behavior in a 2D layered material. This extension could be provided to the reader upon request.

The next term, DMI energy density, w_D, can be determined by combining all the different possible Lifshitz invariants for the T crystallographic class and defined by the following equation:

$$w_D = D[\mathbf{m}.(\nabla \times \mathbf{m})] \tag{5.3}$$

where D depends on the structure of the material and the unit is in J/m^2. The theoretical and experimental value of D ranges from -1.5 mJ/m^2 to 10 mJ/m^2 [115, 116, 117]. In our calculation, we used the OOMMF extension developed by David Cortés-Ortuño et al. in ref. [118, 119].

The third term, Anisotropy energy density term, w_A, can be understood as a prefer-

ence of materials' magnetization towards a certain lattice direction of the crystal, and the preferred axis is known as Easy Axis. This term is expressed as follows:

$$w_A = K[(\mathbf{m}.\mathbf{u})^2] \tag{5.4}$$

where K is the material-dependent parameter and is usually expressed in J/m^3, and $\mathbf{u}$ is the direction of crystalline anisotropy.

The final term, w_Z, is known as Zeeman energy density, which results from the interaction between the external field and internal magnetic moment. The following equation is used to describe the term:

$$w_Z = \mu_0 M_s[(\mathbf{m}.\mathbf{H})] \tag{5.5}$$

where H is the amplitude of the external magnetic field direction with a unit of A/m.

Once we define all the system's energy we let our system relax for minimum energy and extract the feature to understand the memory effects. The simulation parameter details of different layers are given in Table. 5.1.

5.3 Results and Discussions

5.3.1 Competition between Exchange and DMI

Exchange energy (J) is associated with the interaction between the magnetic moments of neighboring atoms. It is a result of electrons' quantum mechanical exchange interaction, which arises from the Fermi nature of electrons. In a ferromagnet, the exchange energy is responsible for aligning the magnetic moments of neighboring atoms in the same direction, leading to a strong net magnetic moment. This energy can be realized using a continuous model with constant A as given in the Equation. 5.2. Figure 5.1(a) represents the relaxed spin configurations of a typical magnetic system that has only Exchange energy. As expected, after relaxation, all the spins are aligned in the same direction. DMI arises from the breaking of inversion symmetry at the atomic scale and can lead to the formation of a non-collinear magnetic structure. In a material with a strong DMI, the magnetic moments of neighboring atoms are arranged in a perpendicular fashion from each other as can be seen in Figure 5.1(b).

When both exchange energy and DMI energy are present in a material, they can compete with each other and stabilizes neighboring spins at a fixed angle. Once the energy

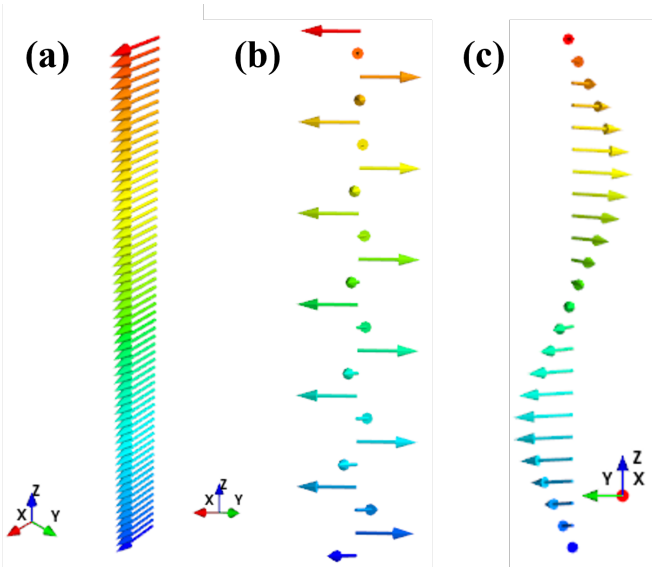


Figure 5.1: Relaxed spin configurations with (a) Exchange energy term only, (b) DMI energy only, and (c) both Exchange and DMI energy [This figure is constructed by using Mayavi [21]].

of the system is minimized, the competition between these two energy terms leads to the formation of a helimagnet, where the magnetic moments are arranged in a spiral pattern. A portion of a relaxed spin structure is shown in Figure 5.1(c). Thus, depending on the relative of DMI with respect to the exchange energy, a helical magnetic structure with different periodicity can be formed.

To get insights into how the helical period of the helimagnet transforms with respect to the DMI energy, a simulation of the helimagnet with different DMI energy and a fixed exchange energy was performed and the result is shown in Figure 5.2. The bottom layer is pinned to the x-axis with high anisotropy. As can be seen in the figure, with the increase in DMI energy, the helical period becomes smaller, meaning the helical period is inversely proportional to the DMI energy value. The results obey the equation of zero fields chiral length of a magnetic system and can be written as [120],

$$L(0) \cong \frac{2\pi a J}{D} \tag{5.6}$$

Where L is the helical period, a is the crystal lattice constant, J is the exchange energy, and D is the DMI energy.

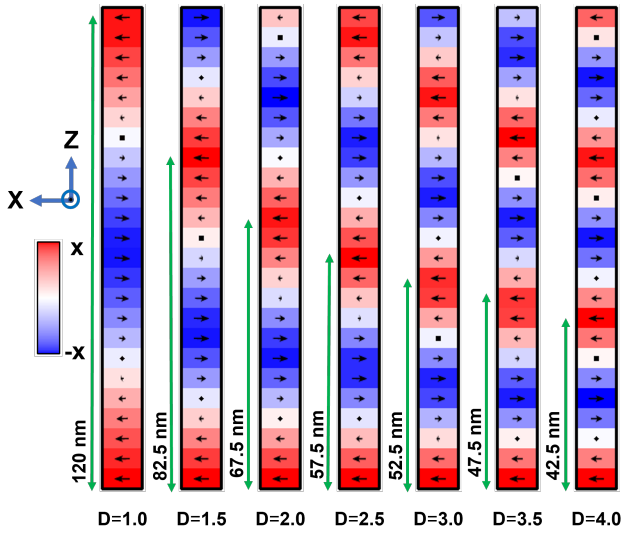


Figure 5.2: Simulated spin configuration of a magnetic system with a fixed exchange energy and varying DMI energy (units are in mJm^{-2}).

5.3.2 Writing Mechanism

The proposed memory device, helimagnet-based random-access memory (HMRAM), shown in the inset of Figure 5.3(a), consists of a helimagnet layer that is sandwiched between two ferromagnetic layers. The bottom ferromagnetic layer is pinned while the top layer can be rotated with the application of external fields. The thickness of the bottom pinned layer is set to 5 nm, and it is strictly locked in the $+x$ direction with very high anisotropy. As the thickness was very small, to achieve the locking the anisotropy of that layer was set to a large value.

Figure 5.3(a) illustrates the proposed writing schemes, which consist of a pass transistor, our proposed single-unit memory cell, electrodes on top and bottom, and two programming word lines that are perpendicular to one another. An electrical connection is formed between the bottom electrode of the memory cell and the ground through a stack of vias that are connected to a pass transistor below.

To write a memory state inside the memory cell, the pass transistor should be in the off state. After that desired current should be sent to the programming word lines (white arrows), and magnetic fields (yellow arrows) perpendicular to the current directions will switch the memory unit to the desired states. As the programming conductors are physically separated from the proposed memory cell, the parasitic delay would be minimized

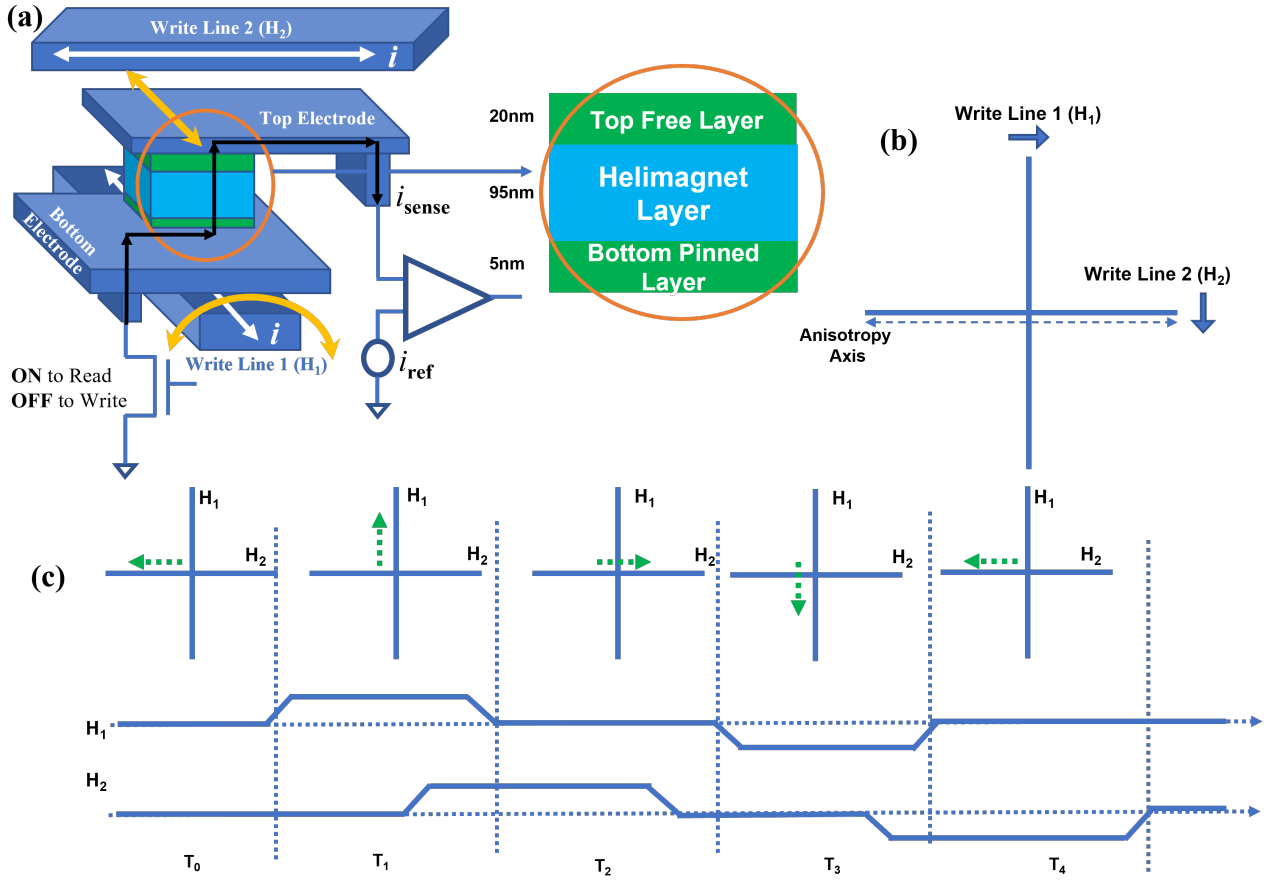


Figure 5.3: Writing scheme on our proposed memory with (a) sense path for reading and programming path for writing (following the scheme of Ref. [22]), (b) Anisotropy Axis along with Write Line axis, and (c) Current pulse diagram for writing a complete cycle. [inset shows the proposed memory structure]

in the proposed writing scheme. Reading the memory state of the proposed cell could be done by turning on the pass transistor and sending a small amount of current through the memory cell and sensing the change in resistance. The easy axis of the ferromagnetic layers and two different programming lines (Word Line1, WL_1, and Word Line2, WL_2) along with their field lines are depicted in Figure 5.3(b).

For effective writing of a memory state, the spin configurations of the top layer must be changed by a specific arrangement of the applied current pulses with associated perpendicular magnetic fields. The pulse polarity and amplitude of the two-word lines along with the resultant top-layer spin configurations are illustrated in Figure 5.3(c).

Initially, i.e., at time T_0, we assume that the top layer is aligned with the negative x-axis, which is the easy axis of the bit. Next, at T_1, current flows only through the WL_1, producing a perpendicular magnetic field in the y-direction. The produced magnetic field would rotate the top layer configuration due to the Zeeman energy, while the bottom pinned layer spin configuration will not change due to high anisotropy. At the last quarter of the time T_1, WL_2 will be turned on in order to avoid the spin relaxation back to negative x-direction.

At T_2, we need to turn off the current flow through WL_1 and turn on the current for WL_2, resulting in a magnetic field along the x-axis and top layer spin configurations will lock in the same direction due to its anisotropy. The resultant spin configuration is stable even if we turned off both currents.

At T_3, a negative current flows through the WL_1. This results in a magnetic field that aligns the spin configuration of the top layer to the negative y direction. Thus another 90-degree rotation can be achieved. Similar to the previous steps, a negative current is sent through WL_2 before switching off WL_1 current to prevent bit flipping back.

Lastly, at T_4, a negative current will be conducted through WL_2, and a magnetic field in the negative x direction will be generated and flip the top layer towards that direction. Thus, a 360 rotation is achieved. It can be seen from the figure that these 360 rotations of the top layer are in a clockwise (CW) direction. It is worthwhile mentioning that by simply changing the current polarity, we could achieve a 360-degree counter-clockwise (CCW) rotation of the spin.

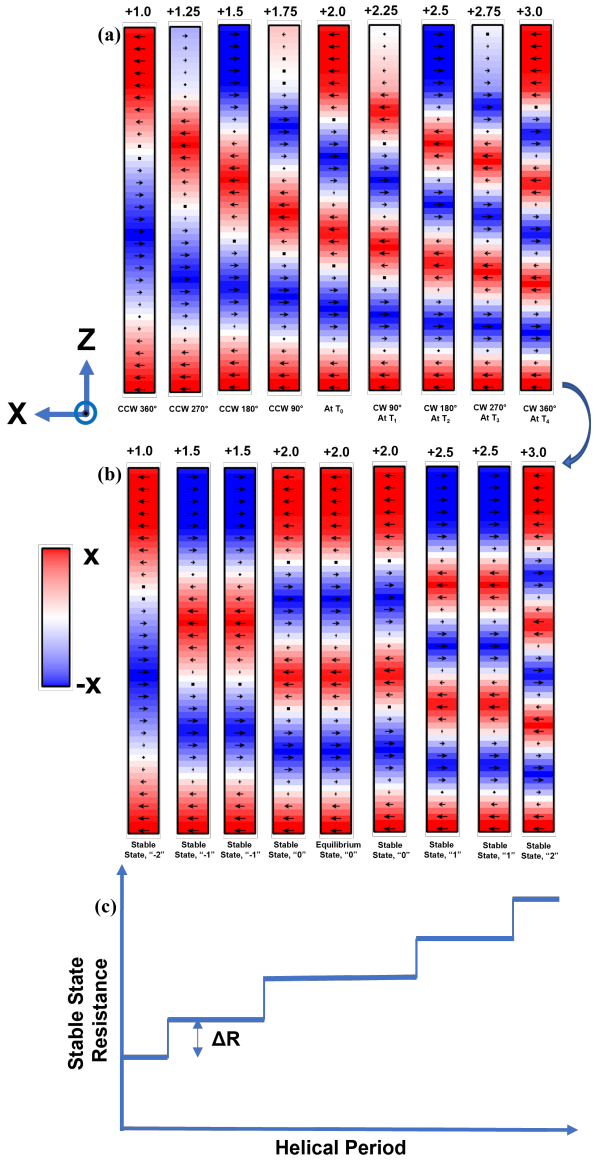


Figure 5.4: Spin configurations of our proposed memory device (a) when sweeping the magnetic field in clockwise (CW) and counterclockwise (CCW) direction and (b) when each corresponding configuration is relaxed without the magnetic field. (c) Schematic illustrations of the resistance states that correspond to the spin configurations without external magnetic field.

5.3.3 Application of Proposed Writing Method in HMRAM

Now, the writing scheme is modeled in a typical memory cell, where the total thickness was assumed 120nm, with a helimagnet thickness of 95nm, the bottom pinned layer thickness was 5nm, and the Top layer thickness was 20nm. The parameter details were given in Table. 5.1. Whichever direction we apply the magnetic field, a gradual change of the spin configurations was observed and captured in Figure 5.4(a). As can be seen from the figure, there is no change in the bottom layer magnetization since the anisotropy energy is dominant over Zeeman energy. The top layer configurations, however, were changed according to the specific writing scheme described in Figure 5.3(c). To make things more comparable, the corresponding writing time for the clockwise rotation has been given as a footnote of Figure 5.4(a). It is worthwhile to note that illustrated spin configurations are minimized by considering all the energy terms, including Zeeman energy. If we start from the equilibrium state, the clockwise rotation increases the helicity period, while the counterclockwise rotation results in a decreased helicity.

Upon finishing the analysis with the magnetic field, the next study was to find out the non-volatile stable memory state without the magnetic field. The spin configurations of the relaxed stable states are shown in Figure 5.4(b). As expected, when the top layer spin configurations are perpendicular to the anisotropy axis, upon removing the external fields, it immediately relaxes to the adjacent state with lower anisotropy energy. We define the stable state at T_0 as equilibrium state '0', states in the CW direction with positive state symbols, e.g., +1, +2, and so on, and the states in the CCW direction are denoted with a negative sign, e.g. -1, -2, and so on. As the spin configuration of the free layer is locked by the anisotropy energy of the top layer, the corresponding spin configurations will be stable after the CW or CCW rotation of the free layer. Also, spin configuration at T_1, which is a quarter turn rotation from the equilibrium state, T_0, reverts back to its previous state, as shown by Stable State "0". Similarly, for smaller rotations not exceeding the anisotropy barrier, they would relax back to the previous state as well.

In this particular setup, we have found multiple stable states with +6 states in the positive direction and -6 in the negative direction sweep. Thus, we could fit multi-bit of information in each memory units.

5.3.4 Reading of the Memory states

Reliable reading of a memory state is also crucial as real application depends on how easily and reliably we can read and write. In order to read the state of the memory cell, the pass

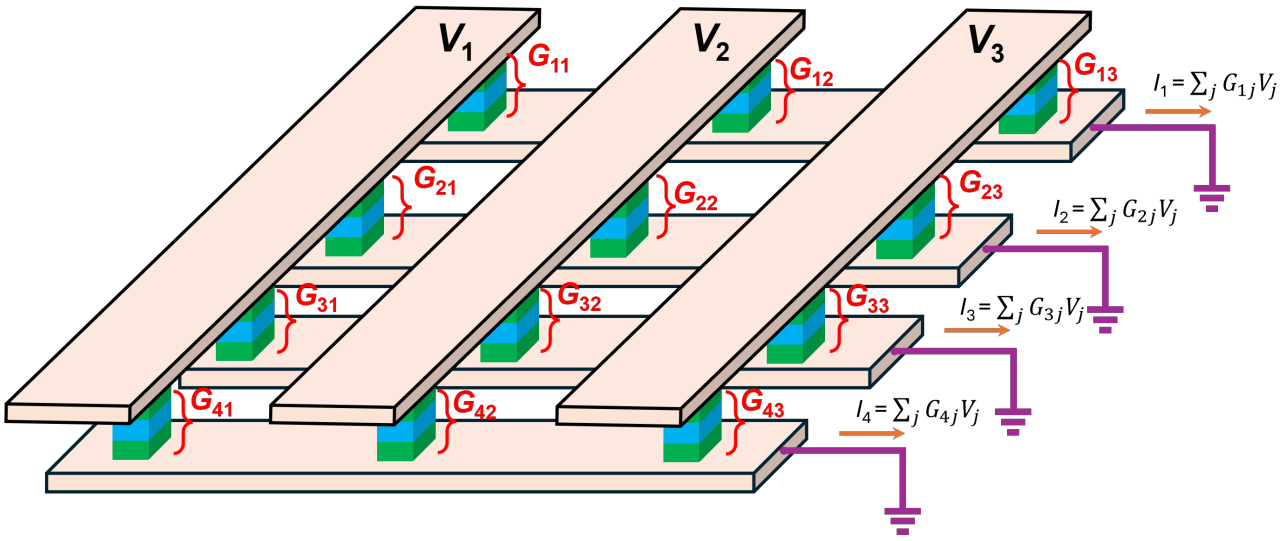


Figure 5.5: HMRAM crossbar array to realize im-memory computing architecture.

transistor must be turned "ON". The helimagnet period changes upon the writing scheme, and so does its resistance. This is due to the increased spin scattering when the relative angles between adjacent spins change [121]. Thus, to read the memory state inside the device structure, a small voltage is applied to the device and measures the current from the device, which provides the discrete resistance state of the device.

5.3.5 HMRAM crossbar array for IMC

As discussed earlier in Chapter 2.6.2, crossbar memory architecture is a promising candidate for highly energy-efficient and area-efficient matrix-vector multiplication (MVM) implementation due to non-volatility and reduced footprints ($4F^2$). Figure 5.5 illustrates an HMRAM crossbar array that consists of twelve intersections between a perpendicular row and column electrodes. At each intersection, an HMRAM cell is connected between the top (column) and bottom (row) electrodes. The voltage, V_j, is applied to the corresponding column j, and conductance, G_{ij}, is encoded into the memory cell. The multiplication, described by Ohm's law, results in circuit current through the device, and the sum of the current from rows results from Kirchhof's current law. Thus, in general, a total current flowing through any row (i), can be written as,

$$I_i = \sum_j G_{ij} * V_j \tag{5.7}$$

where i is the row number, G_{ij} is the conductance of the HMRAM at row number i and column number j.

5.4 Conclusion

In conclusion, a memory device structure, consisting of a helimagnetic element sandwiched by two ferromagnetic layers(a pinned layer and a free layer), with the capability of multi-bit and non-volatile data storage has been proposed. In order to tune the helical period of the memory cell, a rotating in-plane external magnetic field, resulting from two-word lines, was used in CW and CCW direction to increase and decrease the helical period from the equilibrium state. The spin configurations were stabilized by the anisotropy of the free layer when the external field was removed, and the states were non-volatile. The reading scheme of the memory is done through the different resistance states for different memory states due to spin scattering. In addition, HMRAM-based crossbar array architecture can be applied for IMC applications. Overall, the proposal of our memory device paves the path for future spin-based emerging memory.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

With the increased demand for artificial intelligence, computer vision, and natural language processing, existing computing schemes are constrained with respect to energy efficiency and cost. In addition, further scaling of CMOS chips is approaching a limit that is also coupled with power density issues. Most of the energy is consumed to access the external DRAM memory. Therefore, alternative computing schemes, such as neuromorphic computing (NC) and in-memory computing (IMC), have gained much attention due to the fact that computation and memory are co-localized. Among different candidates, memristors-based NC and IMC schemes have attracted a great deal of attention due to their simple two-terminal structure and capability to switch at low power. The main contribution of this thesis is the demonstration of a novel lithium-imbued TiO_x device that initially, before forming, shows analog volatile resistive switching, and after forming, the device can store multi-bit information in a non-volatile fashion. These characteristics are crucial for NC and IMC architectures. This thesis also discussed a fabrication method of BEOL integration of the memory device on top of a CMOS tape-out chip that could be utilized in NC architecture. Finally, a helimagnet-based multi-bit non-volatile memory device was presented that can be used in IMC architectures. A more detailed contribution from this thesis is presented below.

1. A **Li**-ion imbued **TiO_x** iontronic device, with both volatile and non-volatile memory capabilities, has been demonstrated, which is a first-of-its-kind demonstration (Chapter 4). Before forming, the devices show inherently uniform gradual conductance change upon the application of external bias due to the controlled lithium ionic

motion from a solid-state ionic layer. **Li**-ion intercalation in cathodic $\mathbf{TiO_x}$ increases the conductivity, and **Li**-ion deintercalation reduces the conductivity of the volatile memory device. A self-relaxation phenomenon confirms the volatile nature of the memory. The **Li**-ion-induced vacancy creation inside $\mathbf{TiO_x}$ resembles Calcium-ion-induced synaptic modulation inside biological synapses. Therefore, volatile memory devices mimic the short-term plasticity (STP) of the biological brain. Pulse-dependent results strongly resemble the different brain functionalities such as PPF, PTP, and SRDP characteristics. More interestingly, the time scale of the volatile memory devices can be tuned by changing the applied voltage pulse durations. By performing the temperature-dependent analysis, a transition temperature was revealed, above which the conductivity is thermally activated and below which the conductivity is variable-range hopping dependent.

In addition, the same volatile device can be transformed into non-volatile memory by an electro-forming process with unique multi-bit capabilities. To illustrate the devices' capabilities to perform neuromorphic computing, a unique reservoir computing (RC) scheme was proposed, in which tunable volatile memory is used as a non-linear dynamic reservoir and multi-bit non-volatile memory as a weight storage layer. In addition, to explore the possibilities of scaling down the representing bits of weights, post-training quantization (PTQ) is used to map the trained weights from float32 to int8. The quantized model is validated using the same validation data set, and the results show that there is no accuracy drop after quantization. Based on that, a 2T-2R IMC array is proposed, which is used inside the fully connected layer, with our devices working in non-volatile mode. Using the proposed RC computing architecture, voice recognition tasks were performed, and an accuracy of over 94% has been achieved. Therefore, a general iontronic platform that is customarily configurable to mimic brain functions on a hardware level has been achieved.

2. A BEOL integration method of **Li**-ion imbued $\mathbf{TiO_x}$ iontronic device on foundry CMOS chips has been demonstrated (Chapter 3). Several materials, including lithium, are strictly forbidden in the semiconductor foundry, rendering the integration of memory devices with CMOS almost impossible. Therefore, developing a fabrication method that could allow one to realize CMOS-integrated memory devices is of great interest. In this thesis, a BEOL integration of CMOS process-incompatible devices on CMOS chips has been implemented. The integrated $\mathbf{TiO_x}$ devices with tape-out CMOS show similar performance to standalone $\mathbf{TiO_x}$ devices. The transistor of the 1T-1R structure provides the benefit of a localized compliance current source to protect the integrated devices from dielectric breakdown. The BEOL integration of Li-based memristors on CMOS chips sets steps closer to an all-hardware imple-

mentation of reservoir computing to actually demonstrate its advantages in energy efficiency. The developed method can also be generalized to post-integrate other process-incompatible devices involving quantum materials such as two-dimensional (2D) materials, superconducting materials, and many more.

3. A novel memory device structure, consisting of a helimagnetic element sandwiched by two ferromagnetic layers (a pinned layer and a free layer), with the capability of multi-bit and non-volatile data storage, has been demonstrated (Chapter 5). The period of the helimagnet, i.e., helical spin configurations, is determined by the material's internal exchange and DMI energy value. This period is also an equilibrium state of the resistive memory device. The writing of the memory device is similar to toggle magnetic random-access memory (MRAM), in which a rotating in-plane external magnetic field, resulting from two-word lines, was used in CW and CCW directions to increase and decrease the helical period from the equilibrium state. The spin configurations are stabilized by the anisotropy of the free layer when the external field was removed, and the states are non-volatile. The reading scheme of the memory is done through the different resistance states for different memory states due to spin scattering. In addition, HMRAM-based crossbar array architecture can be applied for IMC applications. Overall, the proposal of our memory device paves the path for future spin-based emerging memory.

6.2 Future Work

1. The first scope lies in the understanding of the resistive switching mechanism by using equation-based modelling in COMSOL Multiphysics simulation tools. This analytical simulation will allow us to realize the underlying mechanism. In-situ TEM could also reveal changes in the phase of the material, which could also indicate an indirect mechanism of the device. Although we tried to look at the TEM images of the as-deposited and hundreds of cycled devices to investigate any sign of the crystal phase changes, we failed to find any change in the crystallinity of TiOx due to the amorphous nature of the films.

   There is room for performance improvement regarding the speed of the ion motion and control of the diffusion barrier to controllably change the decay time constant. The same battery-like structure can be imitated with other oxide layers to compare and contrast the performance. Regarding Non-volatile memory, the devices still suffer from limited endurance, which could be improved by a dedicated deposition

chamber. Although volatile memory devices readily endure thousands of cycles, the non-volatile devices' endurance needs improvement to meet industry requirements, which is about a million (10^6) cycles.

The voice recognition task was used to benchmark the performance of reservoir computing architecture. A more sophisticated model, such as weather forecasting and price prediction, can be used to take advantage of this architecture.

2. The timescale of the tunable volatile Lithium-ion imbued $\mathbf{TiO_x}$ devices depends on the pulse voltage excitation duration. The decay constant determines the duration of the memory. It is possible to extend the timescale of the memory by adding a barrier layer on top of the solid-state ionic layer. Adding this barrier layer can slow down the decay from the device, allowing one to realize a deterministic timescale fadeable memory. This fadeable memory can be used for secure information distribution.

3. The tape-out chip was designed with a decoder only, which will activate one transistor at a time, which limits the applications of the BEOL-integrated memory device. In the future tape-out chip design, more components are needed to realize the hardware-level artificial brain. The components include sense amplifiers to sense the state of the memory cell, analog-to-digital-converters to sum up the accumulated current, and Multiplexers to enable the row and column. Moreover, the tape-out chip size could also be increased. The whole assembly could be used to realize a hardware-level artificial neural network.

References

[1] A. Mehonic and A. J. Kenyon, "Brain-inspired computing needs a master plan," *Nature*, vol. 604, no. 7905, pp. 255–260, 2022. xiii, 1, 2

[2] M. A. Zidan, J. P. Strachan, and W. D. Lu, "The future of electronics based on memristive systems," *Nature electronics*, vol. 1, no. 1, pp. 22–29, 2018. xiii, 3

[3] C. Du, W. Ma, T. Chang, P. Sheridan, and W. D. Lu, "Biorealistic implementation of synaptic functions with oxide memristors through internal ionic dynamics," *Advanced Functional Materials*, vol. 25, no. 27, pp. 4290–4299, 2015. xiii, 11, 12

[4] Z. Wang, S. Joshi, S. E. Savel'ev, H. Jiang, R. Midya, P. Lin, M. Hu, N. Ge, J. P. Strachan, Z. Li *et al.*, "Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing," *Nature materials*, vol. 16, no. 1, pp. 101–108, 2017. xiii, xv, 11, 12, 26

[5] Y. Xi, B. Gao, J. Tang, A. Chen, M.-F. Chang, X. S. Hu, J. Van Der Spiegel, H. Qian, and H. Wu, "In-memory learning with analog resistive switching memory: A review and perspective," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 14–42, 2020. xiv, 13

[6] I. Valov, R. Waser, J. R. Jameson, and M. N. Kozicki, "Electrochemical metallization memories—fundamentals, applications, prospects," *Nanotechnology*, vol. 22, no. 25, p. 254003, 2011. xiv, 15

[7] S. Yu, B. Lee, and H.-S. P. Wong, *Metal oxide resistive switching memory*. Springer, 2012. xiv, 16

[8] Q. Wan and Y. Shi, *Neuromorphic Devices for Brain-inspired Computing: Artificial Intelligence, Perception, and Robotics*. John Wiley & Sons, 2022. xiv, 18

[9] H.-S. P. Wong, H.-Y. Lee, S. Yu, Y.-S. Chen, Y. Wu, P.-S. Chen, B. Lee, F. T. Chen, and M.-J. Tsai, "Metal–oxide rram," *Proceedings of the IEEE*, vol. 100, no. 6, pp. 1951–1970, 2012. xiv, 19

[10] B. Zhang, R. Tan, L. Yang, J. Zheng, K. Zhang, S. Mo, Z. Lin, and F. Pan, "Mechanisms and properties of ion-transport in inorganic solid electrolytes," *Energy Storage Materials*, vol. 10, pp. 139–159, 2018. xiv, 21

[11] Y. Wang, W. D. Richards, S. P. Ong, L. J. Miara, J. C. Kim, Y. Mo, and G. Ceder, "Design principles for solid-state lithium superionic conductors," *Nature materials*, vol. 14, no. 10, pp. 1026–1031, 2015. xiv, 21

[12] Y. Zhu, J. C. Gonzalez-Rosillo, M. Balaish, Z. D. Hood, K. J. Kim, and J. L. Rupp, "Lithium-film ceramics for solid-state lithionic devices," *Nature Reviews Materials*, vol. 6, no. 4, pp. 313–331, 2021. xiv, 22

[13] A. Lakshmi-Narayana, O. M. Hussain, A. Mauger, and C. M. Julien, "Transport properties of nanostructured li2tio3 anode material synthesized by hydrothermal method," *Sci*, vol. 1, no. 3, p. 56, 2019. xiv, 23

[14] R. A. John, F. Liu, N. A. Chien, M. R. Kulkarni, C. Zhu, Q. Fu, A. Basu, Z. Liu, and N. Mathews, "Synergistic gating of electro-iono-photoactive 2d chalcogenide neuristors: coexistence of hebbian and homeostatic synaptic metaplasticity," *Advanced Materials*, vol. 30, no. 25, p. 1800220, 2018. xiv, 25

[15] L. Q. Zhu, C. J. Wan, L. Q. Guo, Y. Shi, and Q. Wan, "Artificial synapse network on inorganic proton conductor for neuromorphic systems," *Nature communications*, vol. 5, no. 1, p. 3158, 2014. xv, 26

[16] G.-q. Bi and M.-m. Poo, "Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type," *Journal of neuroscience*, vol. 18, no. 24, pp. 10 464–10 472, 1998. xv, 27

[17] S. Yu, Y. Wu, R. Jeyasingh, D. Kuzum, and H.-S. P. Wong, "An electronic synapse device based on metal oxide resistive switching memory for neuromorphic computation," *IEEE Transactions on Electron Devices*, vol. 58, no. 8, pp. 2729–2737, 2011. xv, 27

[18] C. Du, F. Cai, M. A. Zidan, W. Ma, S. H. Lee, and W. D. Lu, "Reservoir computing using dynamic memristors for temporal information processing," *Nature communications*, vol. 8, no. 1, p. 2204, 2017. xv, 29

[19] Y. Zhong, J. Tang, X. Li, X. Liang, Z. Liu, Y. Li, Y. Xi, P. Yao, Z. Hao, B. Gao *et al.*, "A memristor-based analogue reservoir computing system for real-time and power-efficient signal processing," *Nature Electronics*, vol. 5, no. 10, pp. 672–681, 2022. xv, 30, 31, 49

[20] A. Amirsoleimani, F. Alibart, V. Yon, J. Xu, M. R. Pazhouhandeh, S. Ecoffey, Y. Beilliard, R. Genov, and D. Drouin, "In-memory vector-matrix multiplication in monolithic complementary metal–oxide–semiconductor-memristor integrated circuits: Design choices, challenges, and perspectives," *Advanced Intelligent Systems*, vol. 2, no. 11, p. 2000115, 2020. xvi, 32

[21] P. Ramachandran and G. Varoquaux, "Mayavi: 3d visualization of scientific data," *Computing in Science & Engineering*, vol. 13, no. 2, pp. 40–51, 2011. xviii, 75

[22] B. Engel, J. Akerman, B. Butcher, R. Dave, M. DeHerrera, M. Durlam, G. Grynkewich, J. Janesky, S. Pietambaram, N. Rizzo *et al.*, "A 4-mb toggle mram based on a novel bit and switching method," *IEEE Transactions on Magnetics*, vol. 41, no. 1, pp. 132–136, 2005. xviii, 77

[23] "IEEE International Roadmap for Devices and Systems, 'Beyond CMOS and Emerging Materials'," Institute of Electrical and Electronics Engineers, 2023, doi: https://doi.org/10.60627/0p45-zj55. 1

[24] E. Strubell, A. Ganesh, and A. McCallum, "Energy and policy considerations for modern deep learning research," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 09, 2020, pp. 13 693–13 696. 1

[25] M. Horowitz, "Computing's energy problem (and what we can do about it)," in *2014 IEEE international solid-state circuits conference digest of technical papers (ISSCC)*. IEEE, 2014, pp. 10–14. 3

[26] G. Indiveri, E. Chicca, and R. Douglas, "A vlsi array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity," *IEEE transactions on neural networks*, vol. 17, no. 1, pp. 211–221, 2006. 4

[27] D. Ielmini and H.-S. P. Wong, "In-memory computing with resistive switching devices," *Nature electronics*, vol. 1, no. 6, pp. 333–343, 2018. 5

[28] T. Soliman, S. Chatterjee, N. Laleni, F. Müller, T. Kirchner, N. Wehn, T. Kämpfe, Y. S. Chauhan, and H. Amrouch, "First demonstration of in-memory computing

crossbar using multi-level cell fefet," *Nature Communications*, vol. 14, no. 1, p. 6348, 2023. 5

[29] L. Chua, "Memristor-the missing circuit element," *IEEE Transactions on Circuit Theory*, vol. 18, no. 5, pp. 507–519, 1971. 8

[30] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," *nature*, vol. 453, no. 7191, pp. 80–83, 2008. 8

[31] Q. Xia and J. J. Yang, "Memristive crossbar arrays for brain-inspired computing," *Nature materials*, vol. 18, no. 4, pp. 309–323, 2019. 9

[32] A. Sawa, "Resistive switching in transition metal oxides," *Materials today*, vol. 11, no. 6, pp. 28–36, 2008. 17

[33] R. Yang, K. Terabe, G. Liu, T. Tsuruoka, T. Hasegawa, J. K. Gimzewski, and M. Aono, "On-demand nanodevice with electrical and neuromorphic multifunction realized by local ion migration," *ACS nano*, vol. 6, no. 11, pp. 9515–9521, 2012. 17

[34] Z. Q. Wang, H. Y. Xu, X. H. Li, H. Yu, Y. C. Liu, and X. J. Zhu, "Synaptic learning and memory functions achieved using oxygen ion migration/diffusion in an amorphous ingazno memristor," *Advanced Functional Materials*, vol. 22, no. 13, pp. 2759–2765, 2012. 17

[35] E. Mikheev, B. D. Hoskins, D. B. Strukov, and S. Stemmer, "Resistive switching and its suppression in pt/nb: Srtio3 junctions," *Nature communications*, vol. 5, no. 1, p. 3990, 2014. 17

[36] Y. Chen, "Reram: History, status, and future," *IEEE Transactions on Electron Devices*, vol. 67, no. 4, pp. 1420–1433, 2020. 18

[37] G. Long, Q. Xue, Q. Li, Y. Shi, L. Li, L. Cheng, P. Li, J. Zhang, X. Zhang, H. Guo *et al.*, "Interfacial control via reversible ionic motion in battery-like magnetic tunnel junctions," *Advanced Electronic Materials*, vol. 7, no. 9, p. 2100512, 2021. 19

[38] P. Lopez-Aranguren, M. Reynaud, P. Głuchowski, A. Bustinza, M. Galceran, J. M. Lopez del Amo, M. Armand, and M. Casas-Cabanas, "Crystalline lipon as a bulk-type solid electrolyte," *ACS Energy Letters*, vol. 6, no. 2, pp. 445–450, 2021. 22

[39] B. Put, P. M. Vereecken, J. Meersschaut, A. Sepúlveda, and A. Stesmans, "Electrical characterization of ultrathin rf-sputtered lipon layers for nanoscale batteries," *ACS applied materials & interfaces*, vol. 8, no. 11, pp. 7060–7069, 2016. 22

[40] X. Yu, J. Bates, G. Jellison, and F. Hart, "A stable thin-film lithium electrolyte: lithium phosphorus oxynitride," *Journal of the electrochemical society*, vol. 144, no. 2, p. 524, 1997. 22

[41] R. Midya, Z. Wang, S. Asapu, X. Zhang, M. Rao, W. Song, Y. Zhuo, N. Upadhyay, Q. Xia, and J. J. Yang, "Reservoir computing using diffusive memristors," *Advanced Intelligent Systems*, vol. 1, no. 7, p. 1900084, 2019. 30, 48, 68

[42] J. Moon, W. Ma, J. H. Shin, F. Cai, C. Du, S. H. Lee, and W. D. Lu, "Temporal data classification and forecasting using a memristor-based reservoir computing system," *Nature Electronics*, vol. 2, no. 10, pp. 480–487, 2019. 30, 48, 49, 68

[43] Y. Zhong, J. Tang, X. Li, B. Gao, H. Qian, and H. Wu, "Dynamic memristor-based reservoir computing for high-efficiency temporal signal processing," *Nature communications*, vol. 12, no. 1, p. 408, 2021. 31, 49, 68

[44] B. Cui, *Recent advances in nanofabrication techniques and applications*. BoD–Books on Demand, 2011. 37

[45] J. W. Coburn and H. F. Winters, "Ion-and electron-assisted gas-surface chemistry—an important effect in plasma etching," *Journal of Applied physics*, vol. 50, no. 5, pp. 3189–3196, 1979. 39

[46] V. M. Donnelly and A. Kornblit, "Plasma etching: Yesterday, today, and tomorrow," *Journal of Vacuum Science & Technology A*, vol. 31, no. 5, 2013. 40

[47] H. F. Winters, "The role of chemisorption in plasma etching," *Journal of Applied Physics*, vol. 49, no. 10, pp. 5165–5170, 1978. 40

[48] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal processing magazine*, vol. 29, no. 6, pp. 82–97, 2012. 48

[49] Z. Huang, S. Sun, J. Zhao, and L. Mao, "Multi-modal policy fusion for end-to-end autonomous driving," *Information Fusion*, vol. 98, p. 101834, 2023. 48

[50] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015. 48

[51] Y. Goldberg, *Neural network methods for natural language processing*. Springer Nature, 2022. 48

[52] P. Yao, H. Wu, B. Gao, J. Tang, Q. Zhang, W. Zhang, J. J. Yang, and H. Qian, "Fully hardware-implemented memristor convolutional neural network," *Nature*, vol. 577, no. 7792, pp. 641–646, 2020. 48

[53] P. Yao, H. Wu, B. Gao, S. B. Eryilmaz, X. Huang, W. Zhang, Q. Zhang, N. Deng, L. Shi, H.-S. P. Wong *et al.*, "Face classification using electronic synapses," *Nature communications*, vol. 8, no. 1, p. 15199, 2017. 48

[54] S. Yu, "Neuro-inspired computing with emerging nonvolatile memorys," *Proceedings of the IEEE*, vol. 106, no. 2, pp. 260–285, 2018. 48

[55] Z. Jiang, S. Yin, J.-S. Seo, and M. Seok, "C3sram: In-memory-computing sram macro based on capacitive-coupling computing," *IEEE Solid-State Circuits Letters*, vol. 2, no. 9, pp. 131–134, 2019. 48

[56] Q. Deng, L. Jiang, Y. Zhang, M. Zhang, and J. Yang, "Dracc: A dram based accelerator for accurate cnn inference," in *Proceedings of the 55th annual design automation conference*, 2018, pp. 1–6. 48

[57] S. Yu, W. Shim, X. Peng, and Y. Luo, "Rram for compute-in-memory: From inference to training," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 7, pp. 2753–2765, 2021. 48

[58] S.-O. Park, H. Jeong, J. Park, J. Bae, and S. Choi, "Experimental demonstration of highly reliable dynamic memristor for artificial neuron and neuromorphic computing," *Nature Communications*, vol. 13, no. 1, p. 2888, 2022. 48

[59] X. Duan, Z. Cao, K. Gao, W. Yan, S. Sun, G. Zhou, Z. Wu, F. Ren, and B. Sun, "Memristor-based neuromorphic chips," *Advanced materials*, p. 2310704, 2024. 48

[60] N. C. Dvornek and X. Li, "Chapter 13 - deep learning with connectomes," in *Connectome Analysis*, M. D. Schirmer, T. Arichi, and A. W. Chung, Eds. Academic Press, 2023, pp. 289–308. 48

[61] W. Wang, M. Wang, E. Ambrosi, A. Bricalli, M. Laudato, Z. Sun, X. Chen, and D. Ielmini, "Surface diffusion-limited lifetime of silver and copper nanofilaments in resistive switching devices," *Nature communications*, vol. 10, no. 1, p. 81, 2019. 48

[62] Y. Wang, Z. Zhang, H. Li, and L. Shi, "Realizing bidirectional threshold switching in ag/ta 2 o 5/pt diffusive devices for selector applications," *Journal of Electronic Materials*, vol. 48, pp. 517–525, 2019. 48

[63] J. H. Yoon, Z. Wang, K. M. Kim, H. Wu, V. Ravichandran, Q. Xia, C. S. Hwang, and J. J. Yang, "An artificial nociceptor based on a diffusive memristor," *Nature communications*, vol. 9, no. 1, p. 417, 2018. 48

[64] R. Wang, J.-Q. Yang, J.-Y. Mao, Z.-P. Wang, S. Wu, M. Zhou, T. Chen, Y. Zhou, and S.-T. Han, "Recent advances of volatile memristors: Devices, mechanisms, and applications," *Advanced Intelligent Systems*, vol. 2, no. 9, p. 2000055, 2020. 48

[65] D. Kim, B. Jeon, Y. Lee, D. Kim, Y. Cho, and S. Kim, "Prospects and applications of volatile memristors," *Applied Physics Letters*, vol. 121, no. 1, 2022. 48

[66] G. Zhou, Z. Wang, B. Sun, F. Zhou, L. Sun, H. Zhao, X. Hu, X. Peng, J. Yan, H. Wang *et al.*, "Volatile and nonvolatile memristive devices for neuromorphic computing," *Advanced Electronic Materials*, vol. 8, no. 7, p. 2101127, 2022. 48

[67] S. A. Chekol, S. Menzel, R. W. Ahmad, R. Waser, and S. Hoffmann-Eifert, "Effect of the threshold kinetics on the filament relaxation behavior of ag-based diffusive memristors," *Advanced Functional Materials*, vol. 32, no. 15, p. 2111242, 2022. 48

[68] R. Wang, Q. Liang, S. Wang, Y. Cao, X. Ma, H. Wang, and Y. Hao, "Deep reservoir computing based on self-rectifying memristor synapse for time series prediction," *Applied Physics Letters*, vol. 123, no. 4, 2023. 48

[69] J. Guo, L. Liu, J. Wang, X. Zhao, Y. Zhang, and Y. Yan, "A diffusive artificial synapse based on charged metal nanoparticles," *Nano Letters*, 2024. 48

[70] H. Hu, A. Scholz, C. Dolle, A. Zintler, A. Quintilla, Y. Liu, Y. Tang, B. Breitung, G. C. Marques, Y. M. Eggeler *et al.*, "Inkjet-printed tungsten oxide memristor displaying non-volatile memory and neuromorphic properties," *Advanced Functional Materials*, p. 2302290, 2023. 48

[71] J. Hafner, "Ab-initio simulations of materials using vasp: Density-functional theory and beyond," *Journal of computational chemistry*, vol. 29, no. 13, pp. 2044–2078, 2008. 50

[72] M. Ernzerhof and G. E. Scuseria, "Assessment of the perdew–burke–ernzerhof exchange-correlation functional," *The Journal of chemical physics*, vol. 110, no. 11, pp. 5029–5036, 1999. 50

[73] G. Henkelman, B. P. Uberuaga, and H. Jónsson, "A climbing image nudged elastic band method for finding saddle points and minimum energy paths," *The Journal of chemical physics*, vol. 113, no. 22, pp. 9901–9904, 2000. 50

[74] M. Hacene, A. Anciaux-Sedrakian, X. Rozanska, D. Klahr, T. Guignon, and P. Fleurat-Lessard, "Accelerating vasp electronic structure calculations using graphic processing units," *Journal of computational chemistry*, vol. 33, no. 32, pp. 2581–2589, 2012. 50

[75] M. Hutchinson and M. Widom, "Vasp on a gpu: Application to exact-exchange calculations of the stability of elemental boron," *Computer Physics Communications*, vol. 183, no. 7, pp. 1422–1426, 2012. 50

[76] T. Gu, "Role of oxygen vacancies in tio2-based resistive switches," *Journal of Applied Physics*, vol. 113, no. 3, 2013. 56

[77] H. H. Pham and L.-W. Wang, "Oxygen vacancy and hole conduction in amorphous tio 2," *Physical Chemistry Chemical Physics*, vol. 17, no. 1, pp. 541–550, 2015. 56

[78] Z. Li, H. Liu, Z. Zhao, Q. Zhang, X. Fu, X. Li, F. Gu, H. Zhong, Y. Pan, G. Chen *et al.*, "Space-charge control of magnetism in ferromagnetic metals: Coupling giant magnitude and robust endurance," *Advanced Materials*, vol. 35, no. 8, p. 2207353, 2023. 56

[79] Z. Li, Y. Peng, X. Zhang, Y. Ren, W. Chen, F. Xu, N. Wang, C. Liu, L. Mo, Y. Ding *et al.*, "Oxygen vacancies enhance lithium-ion storage properties of tio2 hierarchical spheres," *Batteries & Supercaps*, vol. 4, no. 12, pp. 1874–1880, 2021. 56

[80] S. Paul, M. A. Rahman, S. B. Sharif, J.-H. Kim, S.-E.-T. Siddiqui, and M. A. M. Hossain, "Tio2 as an anode of high-performance lithium-ion batteries: A comprehensive review towards practical application," *Nanomaterials*, vol. 12, no. 12, p. 2034, 2022. 56

[81] Y. Zhang, X. Zhang, Q. Pang, and J. Yan, "Control of metal oxides' electronic conductivity through visual intercalation chemical reactions," *Nature Communications*, vol. 14, no. 1, p. 6130, 2023. 56

[82] A. Citri and R. C. Malenka, "Synaptic plasticity: multiple forms, functions, and mechanisms," *Neuropsychopharmacology*, vol. 33, no. 1, pp. 18–41, 2008. 57

[83] T. Ohno, T. Hasegawa, T. Tsuruoka, K. Terabe, J. K. Gimzewski, and M. Aono, "Short-term plasticity and long-term potentiation mimicked in single inorganic synapses," *Nature materials*, vol. 10, no. 8, pp. 591–595, 2011. 57

[84] R. S. Zucker and W. G. Regehr, "Short-term synaptic plasticity," *Annual review of physiology*, vol. 64, no. 1, pp. 355–405, 2002. 57

[85] Y. Abbas, A. S. Sokolov, Y.-R. Jeon, S. Kim, B. Ku, and C. Choi, "Structural engineering of tantalum oxide based memristor and its electrical switching responses using rapid thermal annealing," *Journal of Alloys and Compounds*, vol. 759, pp. 44–51, 2018. 57

[86] T. Chang, S.-H. Jo, and W. Lu, "Short-term memory to long-term memory transition in a nanoscale memristor," *ACS nano*, vol. 5, no. 9, pp. 7669–7676, 2011. 57

[87] B. J. Morgan and G. W. Watson, "Gga+ u description of lithium intercalation into anatase tio 2," *Physical Review B*, vol. 82, no. 14, p. 144119, 2010. 62

[88] J. Halim, E. J. Moon, P. Eklund, J. Rosen, M. W. Barsoum, and T. Ouisse, "Variable range hopping and thermally activated transport in molybdenum-based mxenes," *Physical Review B*, vol. 98, no. 10, p. 104202, 2018. 62

[89] J. Wang, H. Zeng, Y. Xie, Z. Zhao, X. Pan, W. Luo, Y. Shuai, L. Tang, D. Zhu, Q. Xie *et al.*, "Analog ion-slicing linbo3 memristor based on hopping transport for neuromorphic computing," *Advanced Intelligent Systems*, vol. 5, no. 10, p. 2300155, 2023. 62

[90] R. Li, H. Yang, Y. Zhang, N. Tang, R. Chen, Z. Zhou, L. Liu, J. Kang, and P. Huang, "Adjustable short-term memory of sio x: Ag-based memristor for reservoir computing," *Nanotechnology*, vol. 34, no. 50, p. 505207, 2023. 68

[91] S. Yu, *Semiconductor Memory Devices and Circuits*. CRC Press, 2022. 71

[92] B. Rawat and P. Mittal, "Single bit line accessed high-performance ultra-low voltage operating 7t static random access memory cell with improved read stability," *International Journal of circuit theory and applications*, vol. 49, no. 5, pp. 1435–1449, 2021. 71

[93] K. K.-W. Chang, D. Lee, Z. Chishti, A. R. Alameldeen, C. Wilkerson, Y. Kim, and O. Mutlu, "Improving DRAM performance by parallelizing refreshes with accesses," in *2014 IEEE 20th International Symposium on High Performance Computer Architecture (HPCA)*. IEEE, 2014, pp. 356–367. 71

[94] S. Boboila and P. Desnoyers, "Write endurance in flash drives: Measurements and analysis." in *FAST*, 2010, pp. 115–128. 71

[95] S. Gerardin, M. Bagatin, A. Paccagnella, K. Grürmann, F. Gliem, T. Oldham, F. Irom, and D. N. Nguyen, "Radiation effects in flash memories," *IEEE Transactions on Nuclear Science*, vol. 60, no. 3, pp. 1953–1969, 2013. 71

[96] A. D. Kent and D. C. Worledge, "A new spin on magnetic memories," *Nature nanotechnology*, vol. 10, no. 3, pp. 187–191, 2015. 71

[97] S. Bhatti, R. Sbiaa, A. Hirohata, H. Ohno, S. Fukami, and S. Piramanayagam, "Spintronics based random access memory: a review," *Materials Today*, vol. 20, no. 9, pp. 530–548, 2017. 71

[98] R. Islam, B. Cui, and G.-X. Miao, "Dry etching strategy of spin-transfer-torque magnetic random access memory: A review," *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, vol. 38, no. 5, p. 050801, 2020. 71

[99] S. S. Parkin, M. Hayashi, and L. Thomas, "Magnetic domain-wall racetrack memory," *Science*, vol. 320, no. 5873, pp. 190–194, 2008. 71

[100] S. Parkin and S.-H. Yang, "Memory on the racetrack," *Nature nanotechnology*, vol. 10, no. 3, pp. 195–198, 2015. 71

[101] A. Fert, V. Cros, and J. Sampaio, "Skyrmions on the track," *Nature nanotechnology*, vol. 8, no. 3, pp. 152–156, 2013. 71

[102] N. Nagaosa and Y. Tokura, "Topological properties and dynamics of magnetic skyrmions," *Nature nanotechnology*, vol. 8, no. 12, pp. 899–911, 2013. 71

[103] A. Hirohata, K. Yamada, Y. Nakatani, I.-L. Prejbeanu, B. Diény, P. Pirro, and B. Hillebrands, "Review on spintronics: Principles and device applications," *Journal of Magnetism and Magnetic Materials*, vol. 509, p. 166711, 2020. 71

[104] Y. Cao, Z. Huang, Y. Yin, H. Xie, B. Liu, W. Wang, C. Zhu, D. Mandrus, L. Wang, and W. Huang, "Overview and advances in a layered chiral helimagnet $cr_{1/3}nbs_2$," *Materials Today Advances*, vol. 7, p. 100080, 2020. 71

[105] M.-G. Han, J. Garlow, Y. Kharkov, L. Camacho, R. Rov, J. Sauceda, G. Vats, K. Kisslinger, T. Kato, O. Sushkov *et al.*, "Scaling, rotation, and channeling behavior of helical and skyrmion spin textures in thin films of te-doped cu2oseo3," *Science Advances*, vol. 6, no. 13, p. eaax2138, 2020. 71

[106] J.-i. Kishine and A. Ovchinnikov, "Theory of monoaxial chiral helimagnet," *Solid state physics*, vol. 66, pp. 1–130, 2015. 71

[107] I. E. Dzyaloshinskii, "On the magneto-electrical effects in antiferromagnets," *Soviet Physics JETP*, vol. 10, pp. 628–629, 1960. 71

[108] T. Moriya, "Anisotropic superexchange interaction and weak ferromagnetism," *Physical review*, vol. 120, no. 1, p. 91, 1960. 71

[109] J. Chovan and N. Papanicolaou, *Commensurate and Incommensurate Magnetism in Layered Antiferromagnets*. Springer Berlin Heidelberg, 2005, pp. 347–384. 71

[110] P. Li, R. Islam, G.-X. Miao, and et al., "Helimagnet behavior in Mn1/4TaS2," 2023, (In a helimagnet Mn-TaS2, we observed multi-resistance states with external fields, corresponding to the helicity change of the system, Manuscript in preparation). 72

[111] M. Donahue, "Oommf user's guide, version 1.0," 1999-09-01 1999. 72

[112] A. Tonomura, X. Yu, K. Yanagisawa, T. Matsuda, Y. Onose, N. Kanazawa, H. S. Park, and Y. Tokura, "Real-space observation of skyrmion lattice in helimagnet mnsi thin samples," *Nano letters*, vol. 12, no. 3, pp. 1673–1677, 2012. 72

[113] C. Zhang, J. Zhang, C. Liu, S. Zhang, Y. Yuan, P. Li, Y. Wen, Z. Jiang, B. Zhou, Y. Lei *et al.*, "Chiral helimagnetism and one-dimensional magnetic solitons in a cr-intercalated transition metal dichalcogenide," *Advanced Materials*, vol. 33, no. 35, p. 2101131, 2021. 72

[114] M. Beg, R. Carey, W. Wang, D. Cortés-Ortuño, M. Vousden, M.-A. Bisotti, M. Albert, D. Chernyshenko, O. Hovorka, R. L. Stamps *et al.*, *Ground state search, hysteretic behaviour and reversal mechanism of skyrmionic textures in confined helimagnetic nanostructures*. Nature Publishing Group UK London, 2015, vol. 5, p. 17137. 72

[115] M. Shen, X. Li, Y. Zhang, X. Yang, and S. Chen, "Effects of the interfacial dzyaloshinskii–moriya interaction on magnetic dynamics," *Journal of Physics D: Applied Physics*, vol. 55, no. 21, p. 213002, 2022. 73

[116] A. Hrabec, N. Porter, A. Wells, M. Benitez, G. Burnell, S. McVitie, D. McGrouther, T. Moore, and C. Marrows, "Measuring and tailoring the dzyaloshinskii-moriya interaction in perpendicularly magnetized thin films," *Physical Review B*, vol. 90, no. 2, p. 020402, 2014. 73

[117] M. Beg, M. Albert, M.-A. Bisotti, D. Cortés-Ortuno, W. Wang, R. Carey, M. Vousden, O. Hovorka, C. Ciccarelli, C. S. Spencer *et al.*, "Dynamics of skyrmionic states in confined helimagnetic nanostructures," *Physical Review B*, vol. 95, no. 1, p. 014433, 2017. 73

[118] D. Cortés-Ortuño, M. Beg, V. Nehruji, R. Pepper, and H. Fangohr, "Oommf extension: Dzyaloshinskii-moriya interaction (dmi) for crystallographic classes t and o," 2018. 73

[119] D. Cortés-Ortuño, M. Beg, V. Nehruji, L. Breth, R. Pepper, T. Kluyver, G. Downing, T. Hesjedal, P. Hatton, T. Lancaster *et al.*, "Proposal for a micromagnetic standard problem for materials with dzyaloshinskii–moriya interaction," *New Journal of Physics*, vol. 20, no. 11, p. 113015, 2018. 73

[120] Y. Togawa, T. Koyama, K. Takayanagi, S. Mori, Y. Kousaka, J. Akimitsu, S. Nishihara, K. Inoue, A. Ovchinnikov, and J.-i. Kishine, "Chiral magnetic soliton lattice on a chiral helimagnet," *Physical review letters*, vol. 108, no. 10, p. 107202, 2012. 75

[121] L. Wang, N. Chepiga, D.-K. Ki, L. Li, F. Li, W. Zhu, Y. Kato, O. Ovchinnikova, F. Mila, I. Martin *et al.*, "Controlling the topological sector of magnetic solitons in exfoliated cr 1/3 nbs 2 crystals," *Physical review letters*, vol. 118, no. 25, p. 257203, 2017. 81