



**University of  
Nottingham**

UK | CHINA | MALAYSIA

**DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING**

**Non-Volatile FPGA Architecture based on Resistive Random-Access  
Memory**

**Ph.D. Candidate: Chee Hock Leong**  
**Supervisor: Prof. T. Nandha Kumar**  
**Co-Supervisor: Prof. Haider Abbas Mohammed**

**DATE: 11 August 2022**

**A PhD thesis submitted in part fulfilment of the requirements for the degree of Doctor of Philosophy (PhD)  
Electrical and Electronic Engineering, University of Nottingham.**

## Abstract

---

The advent of massive-scale data-intensive applications from Internet-of-Things (IoT), Artificial Intelligence (AI), neural networks, cloud computing and its services, machine learning, and 21st century modern technology adoption coupled with the increasing environmental awareness of modern society has led to the demand for more digital devices with higher computing power and increased energy consumption efficiency. One of the widely used computing device is the Field-Programmable-Gate-Array (FPGA).

This PhD research aims to develop a nonvolatile FPGA (nvFPGA) as an evolution of the conventional complementary-metal-oxide-semiconductor (CMOS)-based volatile FPGAs. To achieve this, nonvolatility (NV) is implemented through the use of next-generation emerging memory devices called Resistive Random-Access Memories (ReRAMs).

An analysis of the conduction mechanisms behind multi-filamentary ReRAMs is first performed and a model based on oxygen vacancy ( $V_O$ ) migration and trap-assisted tunnelling (TAT) is developed. The model successfully demonstrates the contribution of each individual filament in the ReRAM's metal-oxide layer to the resistive switching (RS) behaviour. In addition to that, the model also shows that the barrier height between the conductive filaments (CFs) in the metal-oxide and the electrode is a strong factor in the CF formation/rupture process. The multi-filamentary switching ReRAM also enables multi-bit (MB) switching; increasing the number of bits that can be stored in the ReRAM cell. The model matches well with experimental results with the largest difference seen in the first intermediate resistive state (IRS) at 30.35%. Compared to the single-filament ReRAM model, the current level of the multi-filamentary ReRAM is 190% higher due to the extra conduction paths through the additional filaments.

NV is then implemented in two fundamental FPGA elements; the configurable logic block (CLB) which is responsible for the FPGA combinatorial logic and the switch block (SB) which is responsible for the FPGA routing configuration.

Inside the CLB are two important components, the lookup table (LUT) and the D flip-flop (DFF). An analysis of a single-bit NV LUT (SB-nvLUT) array and its controller is first performed and the SB-nvLUT successfully eliminates the sneak path problem that is common in ReRAM arrays. A voltage-mode sense-amplifier is then developed to raise the output voltages of the ReRAMs in the array from subthreshold voltages to voltage levels that are

detectable to a differential comparator. The voltage-mode sense-amplifier utilizes 20% lower transmission gates compared to conventional ReRAM sense-amplifier designs and shows 32% and 54% improvements in READ time and energy dissipation compared to an existing voltage-mode design and 85% and 59% improvements in the same respective metrics compared to an existing current-mode design.

The MB NV LUT (MB-nvLUT) and its controller is then developed with 2-bit ReRAMs in the LUT array. The MB-nvLUT reduces the array size by 0.5x and the number of controller gates by 0.25x compared to the SB-nvLUT. Compared to the SB-nvLUT, the MB-nvLUT has an average of 2x lower delay, 1.22x lower energy consumption, and 2.46x lower energy delay product (EDP) for WRITE 0; 2x lower delay, 2x lower energy consumption, and 4.6x lower EDP for WRITE 1; 2x lower delay, 1x lower energy consumption, and 2x lower EDP for WRITE 01 $\rightarrow$ 10; 9.2x lower delay, 128x lower energy consumption, and 153x lower EDP.

Two NV electronic storage elements are developed alongside the NV DFF (nvDFF) which are the NV D latch (nvD latch) and the NV dynamic random-access memory (nvDRAM). Measurements of performance metrics such as Clk-to-Q delay, Q rise time/Q<sub>b</sub> fall time, Q fall time/Q<sub>b</sub> rise time, power dissipation, and WRITE0 and WRITE1 delay and power dissipations are performed. The NV storage elements demonstrate successful data retention during power disruption events, providing advantage over their CMOS-based counterparts; the nvDRAM does not require an energy consuming refresh operation compared to the CMOS-based DRAM and the nvD latch and nvDFF allows proper SLEEP modes and stores data in the event of power disruption. Although introduction of NV incurs a drawback in the aforementioned performance metrics, the NV designs nevertheless demonstrate comparable power consumptions and delay timings when compared to conventional electronic designs.

The nvFPGA made up of an nvCLB (comprising the MB-nvLUT and the nvDFF) and an nvSB (replacing the SRAM storage elements with the nvDRAM) is then presented. The NV properties of the nvCLB is demonstrated together with performance metrics such as the delay timing, energy dissipation, and EDP. The average WRITE delay and EDP of the nvLUT are respectively 73.463% and 99.79% higher than the SRAM LUT while the READ delay and EDP of the nvLUT are respectively 97.295% and 91.184% lower than the SRAM LUT. The improvements in READ performance metrics is advantageous for FPGA applications which typically have numerous READ operations after a single WRITE. The nvDFF implements NV with increase of 5.089% and 61.1045% respectively in the average delay and EDP. The average

WRITE delay and EDP of the nvSwB are 12253.17 ps and 7.588 ps nJ respectively while the average READ delay and EDP of the the nvSwB are 200 ps and  $3.814 \times 10^{-2}$  ps nJ respectively. Although the nvSwB performance metrics are higher than SRAM-based SwBs, the nvSwB values are in picoseconds and are capable of operating at conventional high frequencies.

## **Acknowledgement**

---

Thank you to God for all the blessings I have received in life.

I would like to convey my greatest gratitude to my PhD supervisor, Professor T. Nandha Kumar for providing me this unique opportunity to embark on my PhD journey and being a constant and the strongest guiding light throughout. Patient, meticulous, kind and knowledgeable, he is more than well-suited as a PhD supervisor and the best I could have asked for. The numerous meetings and discussions with him during this PhD have not only served to develop my technical skills, but also imparted a positive effect on my humanistic and professional skills; all of which I will forever remember in life. He is truly an inspiration to those around him.

I would also like to thank my co-supervisor, Professor Haider Abbas Almurib for his valuable support in my research. I'm also thankful to my internal accessor Dr. Belle Ooi for her invaluable advices given to me in my internal assessments. My gratitude also to Dr. Firas Odai Hatem who took his personal time to contribute his valuable knowledge and assistance to kickstart this research project.

Throughout this research, I have been accompanied with a great co-author and a dear friend, Dr. Arya Lekshmi Jagath. I deeply cherish our many discussions and her morale support during this research.

Words cannot express my gratitude to my beloved family, my parents, Chee Beng King and Sew Yoong Yoong and Mary Tang Swee Sing, my brother, Chee Hock Wu and my sisters, Chee Li Hui and Chee Sze Ching for their patience, love, support and motivation. To my family, truly, thank you and I love you.

My thanks to the friends I grew up with who offered their constant backing and encouragement, to the late Jack Tan Yee Jack, to Raymenjit Singh, to Rajkumar Sagunanathan; ours is a friendship I deeply cherish.

My special thanks also to Loh Yi who is more than a best friend and confidante and someone I hope to build many life experiences with.

## Table of Contents

---

Abstract.....	i
Acknowledgement.....	iv
Table of Contents .....	v
List of Figures.....	ix
List of Tables.....	xviii
List of Abbreviations.....	xx
List of Publications.....	xxii
1. Chapter One: Introduction.....	1
1.1 Next Generation Non-volatile FPGA.....	3
1.2 Next Generation Memory Device: Resistive Random-Access Memory .....	4
1.3 Research Gaps in the Field and Research Motivations.....	6
1.4 Aim and Objectives.....	8
1.4.1 Aim.....	8
1.4.2 Objectives .....	8
1.5 Research Outcomes .....	9
1.6 Thesis Outline .....	9
2. Chapter Two: Literature Review .....	12
2.1 Field Programmable Gate Arrays .....	13
2.1.1 Configurable Logic Blocks.....	16
2.1.2 Programmable Interconnects .....	19
2.1.3 Non-volatile FPGAs.....	20
2.1.4 The Volatile and Non-Volatile D Latch.....	22
2.1.5 The Volatile and Non-Volatile Flip-Flop .....	26
2.1.6 The Non-Volatile DRAM .....	27
2.2 Emerging Non-Volatile Memory (NVM) Devices .....	27

2.2.1	Suitability of NVM Device .....	28
2.2.2	Resistive Random-Access Memory .....	29
2.2.3	Multibit ReRAM.....	30
2.3	ReRAM Modelling.....	31
2.3.1	SET and RESET Switching Behaviour.....	32
2.3.2	Onset of the Origin of Gap Opening in the Conductive Filament .....	33
2.3.3	RS Mechanisms in Analytical Models.....	33
2.3.4	Required Parameters to Model RS in ReRAMs .....	39
2.3.5	Current Conduction Mechanisms Adopted in Physics-based Models.....	44
2.3.6	Comparison of Models with Experimental Results .....	48
2.4	ReRAM Fabrication and Layout .....	53
2.5	Research Gaps in the Literature Review .....	56
2.6	Summary of Literature Review .....	58
3.	Chapter Three: Methodology .....	59
3.1	Multifilamentary Model .....	60
3.2	Non-Volatile LUT .....	60
3.3	The Sequential Memories.....	61
3.4	The nvFPGA .....	62
4.	Chapter Four: Multi-Bit ReRAM Model .....	63
4.1	Multifilamentary ReRAM.....	64
4.2	Electrical Modelling of Multi-Filament Bi-Layered Ta <sub>2</sub> O <sub>5</sub> /TaO <sub>x</sub> ReRAM.....	65
4.2.1	Migration of V <sub>O</sub> .....	65
4.2.2	Trap-Assisted Tunelling.....	66
4.2.3	Current Equation.....	67
4.3	Simulation Results.....	67
4.4	Summary .....	74

5. Chapter Five: Single-Bit Non-Volatile LUT .....	76
5.1 Analysis of an SB-NVLUT .....	77
5.1.1 Results and Discussion .....	78
5.2 Sense Amplifier for SB-NVLUT .....	81
5.2.1 Proposed Sense Amplifier Design .....	83
5.2.2 Inverting-Buffer Circuit .....	84
5.2.3 Differential Comparator Circuit .....	86
5.2.4 Simulation and Results .....	89
5.2.5 Summary .....	90
6. Chapter Six: Multi-Bit Non-Volatile LUT .....	92
6.1 The MB-ReRAM .....	93
6.2 The MB-ReRAM LUT Array .....	96
6.3 The MB-LUT Controller .....	98
6.3.1 WRITE Operation .....	100
6.3.2 READ Operation .....	101
6.4 Simulation Results .....	106
6.4.1 1-bit WRITE Operation Results .....	106
6.4.2 2-bit WRITE Operation Results .....	107
6.4.3 READ Operation Results .....	109
6.5 Single Cell Performance Comparison .....	110
6.6 Evaluation on Benchmark Circuits .....	110
6.7 Summary .....	113
7. Chapter Seven: Non-Volatile Memories .....	114
7.1 Memories in Large-Scale Architecture .....	115
7.2 Non-Volatile D Latch .....	116
7.2.1 The Proposed nvD Latch .....	116

7.2.2	Results and Simulation .....	117
7.3	Non-Volatile D Flip-Flop.....	120
7.3.1	The Proposed nvDFF .....	120
7.3.2	Simulation and Results .....	125
7.4	Non-Volatile DRAM.....	136
7.4.1	The Proposed nvDRAM .....	137
7.4.2	2T1R nvDRAM Assessment.....	141
7.5	Summary .....	145
8.	Chapter Eight: The Non-Volatile FPGA Architecture.....	147
8.1	The nvFPGA .....	148
8.2	Analysis of the nvCLB.....	151
8.3	Analysis of the nvSwB.....	153
8.4	Analysis of the nvFPGA .....	153
8.5	Summary .....	156
9.	Chapter Nine: Conclusion and Future Works.....	157
10.	References .....	160

## List of Figures

---

Fig. 1-1. The four fundamental circuit elements and their mathematical relations [25].	4
Fig. 1-2. The current-voltage (IV) hysteresis curve showing change of resistance according to applied voltage [25].	5
Fig. 1-3. Cross-section of a MIM ReRAM cell.	6
Fig. 2-1. Generic FPGA structure and internal components [33].	13
Fig. 2-2. The 6T SRAM [34].	14
Fig. 2-3. Standard transistor (left) vs Flash transistor (right) [35].	15
Fig. 2-4. Flash transistor layout [35].	15
Fig. 2-5. MUX-based LB [35].	17
Fig. 2-6. SRAM-based LUT [35].	17
Fig. 2-7. Xilinx Virtex-5 6-input LUT architecture [36].	18
Fig. 2-8. Altera Stratix-II ALM LUT architecture [37].	18
Fig. 2-9. Switch blocks (C Box and S Box) form the configurable connections between CLBs [38].	19
Fig. 2-10. (a) and (b) FeFET-based nvLUT with different logic configurations, (c) the LUT array, and (d) the 6-input power comparison for LUTs [20].	20
Fig. 2-11. Magnetic tunnel junction (MTJ)/CMOS-based LUTs in (a) [44] and (b) [45].	21
Fig. 2-12. Phase change memory (PCM)-based LUT from [39].	21
Fig. 2-13. CRS ReRAM-based six-input LUT [46].	22
Fig. 2-14. A traditional D latch layout.	23
Fig. 2-15. nvD latch with CRS ReRAMs [52].	24
Fig. 2-16. nvD latch design with single ReRAM cell [53].	24
Fig. 2-17. nvD latch design with T-gates single ReRAM cell [54].	25
Fig. 2-18. Examples of STT-MRAM latches [55].	25
Fig. 2-19. Nonvolatile D flip-flop using control CTRL and SWL signals from [56].	26
Fig. 2-20. Memory taxonomy tree.	28

Fig. 2-21. Three-layer ReRAM. HfO <sub>2</sub> is the TMO layer, Ti and TiN are the electrodes [72]. .....	30
Fig. 2-22. Affect of n <sub>D</sub> on the (a) electrical conductance preexponential factor $\sigma_0$ and (b) activation energy for conduction EAC. The devices switch into LRS when n <sub>D</sub> =0.2x10 <sup>21</sup> cm <sup>-3</sup> , $\sigma_0$ =700Ω <sup>-1</sup> cm <sup>-1</sup> , and EAC=0.006eV for TiN/HfO <sub>2</sub> /TiN, when n <sub>D</sub> =5x10 <sup>21</sup> cm <sup>-3</sup> , $\sigma_0$ =23....	36
Fig. 2-23. Simulated I-V curves from (a) [92] showing different SET switching voltages as a result of different RESET stop voltages, (b) [93] showing the effect of different EA values, and (c) [75].	37
Fig. 2-24. (a) I-V curve of model from [95]. (b) Schematic of CF growth.	38
Fig. 2-25. (a) I-V curve of model from [97]. (b) Schematic of CF growth.	38
Fig. 2-26. Thermal conductivity k <sub>th</sub> as a function of local doping density, n <sub>D</sub> .	40
Fig. 2-27. Simulated microscopic filament evolution during RESET-SET with (A)--(H) corresponding to the I-V curve in Fig. 2-24.	42
Fig. 2-28. Simulation result from [94].	43
Fig. 2-29. Simulated result from [90]. No RS is observed when the temperature is fixed (red line).	43
Fig. 2-30. (a) Experimental [87] and (b) simulation results for Ta <sub>2</sub> O <sub>5</sub> /TaO <sub>x</sub> ReRAM with Schottky conduction [87]. (c) Simulation result from [27]. (d) Simulation result from [98].	49
Fig. 2-31. (a) Experimental results for Ta <sub>2</sub> O <sub>5</sub> /TaO <sub>x</sub> multilevel ReRAM device with Schottky conduction [99] b) Modeling multilevel for Ta <sub>2</sub> O <sub>5</sub> /TaO <sub>x</sub> device taken from [99].	49
Fig. 2-32. Experimental result and model taken from [88].	50
Fig. 2-33. (a) Experiment data and model data plotted for TiN/TiO <sub>x</sub> /HfO <sub>x</sub> /Pt device taken from [100]. (b) Experiment data and model data plotted for TiN/Hf/HfAlO <sub>x</sub> /TiN device taken from [100].	50
Fig. 2-34. Experimental I-V characteristics of the device, measured data, and calculated data plotted in linear scale taken from [91].	51
Fig. 2-35. (a) Typical linear I-V characteristics of HfO <sub>x</sub> -based device reported in [102], (b) and (c) show measured data and model simulation as reported in [92].	51

Fig. 2-36. Experimental I-V characteristics of Pd/Ta <sub>2</sub> O <sub>5</sub> /TaO <sub>x</sub> /Pd device and corresponding model as given in [93].....	52
Fig. 2-37. Experimental I-V characteristics of Pd/Ta <sub>2</sub> O <sub>5</sub> /TaO <sub>x</sub> /Pd device and corresponding model as given in [101].....	52
Fig. 2-38. ReRAM CBA array, ReRAM cells are located between wordlines, WL and bitlines, BL. Sneak path current is represented by dotted red line. Solid black line indicates desired READ current path [109].....	54
Fig. 2-39. (a) Pt/solid electrolyte/Cu ReRAM and its I-V characteristic (b). (c) Cu/solid electrolyte/Pt ReRAM and its I-V-characteristic (d). (e) CRS resulting from the combination of ReRAMs in (a) and (b) and its I-V characteristic (f) [110]. .....	54
Fig. 2-40. (a) Cross-sectional TEM image of a 4-layer 3D VReRAM array. (b) Magnified image of TiN/TiO <sub>x</sub> /HfO <sub>2</sub> /Ru 1S1R cell. [117] .....	55
Fig. 4-1. Schematic representation of the multi-filament ReRAM. Filaments f1, f2, and f3 have different diameters, $\phi$ .....	64
Fig. 4-2. Circuit blocks of (a) state variable migration and (b) current transport for three filaments. SPICE model for (a) single level cell and (b) multi-level cell (three filaments are used for our model). The state variable block produces the change in the ReRAM current. ..	66
Fig. 4-3. Simulation I-V characteristics of the single filament model (blue) and the multi-filament model (red) for 3V/-3V 100Hz transient sine-wave.....	67
Fig. 4-4. (a) Simulated resistance over time graph of single filament model (blue) and multi-filament model (red) for 3V/-3V 100Hz transient sine-wave. The circled area is magnified and plotted in (b) and the variation during RS can be clearly seen. (c).....	68
Fig. 4-5. (a) Semi-log I-V plot with experimental results from [75]. (b) Partial RS (IRS-1) to resistance=13k $\Omega$ , (c) partial RS (IRS-2) to Resistance=20k $\Omega$ , and (d) full RS (HRS) to resistance=50k $\Omega$ . Schematic depiction of the filaments are included in (e), (f), and (g) corresponding to the curves in (b), (c), and (d) respectively. ....	69
Fig. 4-6. Experimental I-V result during reset switching taken from (a) [75] and (b) [165]...	70
Fig. 4-7. Migration of $\omega$ over time. The filament with the lowest barrier height reaches maximum distance first.....	70

Fig. 4-8. (a) Temperature evolution over time and (b) Electric field strength on the individual filaments in the multi-filament model.....	71
Fig. 4-9. 3V triangular input voltage simulation result for (a) multi-filament model and (b) single filament model.....	71
Fig. 4-10. 3V 5ms square input voltage simulation result for (a) multi-filament model and (b) single filament model.....	72
Fig. 4-11. Simulated change of resistance for (a) multi-filament model and (b) single filament model corresponding to the input voltage in Fig. 4-11.....	72
Fig. 4-12. Output current simulation results of 1ms consecutive square input voltages with varying amplitudes of $V_1=1.56V$ , $V_2=1.7V$ , $V_3=1.9V$ for (a) multi-filament model and (b) single filament model.....	73
Fig. 4-13. Simulated change of resistance for (a) multi-filament model and (b) single filament model corresponding to the input voltage in Fig. 4-12.....	73
Fig. 4-14. Output current simulation results of 4ms consecutive square input voltages with varying amplitudes of $V_1=1.56V$ , $V_2=1.7V$ , $V_3=1.9V$ for (a) multi-filament model and (b) single filament model.....	74
Fig. 4-15. Simulated change of resistance for (a) multi-filament model and (b) single filament model corresponding to the input voltage in Fig. 4-14.....	74
Fig. 5-1. Layout of the controller scheme taken from [13]. M11, M12, M21, and M22 are RRAMs in a 2x2 array. Each column (bitline) is separated. G1 and G2 are the same as T1 and T2 signals in TABLE 5-1.....	77
Fig. 5-2. Circuit diagram of the controller taken from [29].....	77
Fig. 5-3. Time evolution plot of $\pm 2V$ , 0.5ms WRITE 1 and 0 pulse scheme from (a) D0 line and (b) D1 line to the ReRAM.....	79
Fig. 5-4. State variable evolution for all RRAMs in a 2x2 array during WRITE 1 and 0 for (a) M11, (b) M21, (c) M12, and (d) M22. The state variable of the selected cell changes from 0 (HRS) to 1 (LRS) during WRITE 1 and returns back to 0 during WRITE 0. ....	79
Fig. 5-5. READ output current from M11 for (a) state 1, LRS and (b) state 0, HRS. READ output current for the other RRAMs in the array demonstrate similar behaviour. ....	80

Fig. 5-6. (a) The schematic of the ReRAM circuit and (b) the input voltage fed to the circuit and the corresponding device current. ....	81
Fig. 5-7. The ReRAM crossbar array.....	82
Fig. 5-8. The (a) READ voltage and (b) state variable for the HRS ReRAM and the (c) READ voltage and (d) state variable for the LRS ReRAM.....	83
Fig. 5-9. The schematic of the proposed sense amplifier circuit. SE and SEN are input signals. A and B are the circuit output lines. ....	84
Fig. 5-10. The input and output voltage waveforms of the inverting-buffer segment (M1, M2). The output voltage is 1.1V when the input is 0V and the output falls to around 850mV when the input voltage rises to ~200mV, the maximum Bitline voltage from the ReRAM memory cell.....	85
Fig. 5-11. The input and output voltage waveforms of each inverter in the inverting-buffer circuit. (a) The linear input voltage from 0V to 200mV to simulate the maximum READ bitline voltage and the response to the input voltage of the (b) first inverter pair (M1, M2), (c) second inverter pair (M3,M4), and (d) the third inverter pair (M5,M6).....	86
Fig. 5-12. The differential comparator circuit. ....	87
Fig. 5-13. The (a) switch voltage, (b) gate voltage for transistor M11, and (c) the output voltage at nodes A and B from the differential comparator circuit. ....	88
Fig. 5-14. The operation waveforms for the sense amplifier circuit.....	89
Fig. 5-15. The memory crossbar array used in the simulation. One sense amplifier is connected to one bitline. ....	89
Fig. 5-16. READ (a) Boolean low (0) and (b) Boolean high (1) waveform for a 2ns 0.1V READ pulse. ....	90
Fig. 6-1. Switching behaviour comparison between SB-switching (left) and MB-switching (right). ....	95
Fig. 6-2. Comparison of device current for SB (a) and MB (b) and power consumption for SB (c) and MB (d). ....	95
Fig. 6-3. The MB-ReRAM array. The schematic of the Controller block is given in Fig. 6-5 while the external READ decoder is given in Fig. 6-4.....	97

Fig. 6-4. Block diagram of the READ controller circuit designed for two input MB-nvLUT. .....	98
Fig. 6-5. (a) Schematic of the controller block designed for MB-NVLUT. This block is able to receive 2 LUT inputs. (b) Additional controller blocks are used for higher-input-LUTs. ....	99
Fig. 6-6. Device current with $\pm 0.05$ VDD. Each WRITE operation is followed by a READ to check the device current at the RS. The closest margin is found between IRS-2, $0.05V_{DD}$ and IRS-1, $+0.05V_{DD}$ . ....	103
Fig. 6-7. WRITE variability test for LRS and HRS of the SB-RERAM and LRS, IRS-1, IRS-2, and HRS of the MB-ReRAM. ....	103
Fig. 6-8. READ endurance test for LRS and HRS of the SB-RERAM and LRS, IRS-1, IRS-2, and HRS of the MB-ReRAM. ....	104
Fig. 6-9. (a) The 4-input MB-LUT and (b) the equivalent RC circuit during WRITE to a single cell (M1). ....	104
Fig. 6-10. Parasitic RC effect on output voltage, $V_{out}$ rise time. Simulation performed with parasitic RC increments of 10%. ....	105
Fig. 6-11. Condition 1: WRITE 0 performance for SB vs MB for 8-bit, 32-bit, 72-bit, and 128-bit arrays; (a) delay, (b) energy, and (c) EDP. ....	106
Fig. 6-12. Condition 2: Write 1 performance for SB vs MB for 8-bit, 32-bit, 72-bit, and 128-bit arrays; (a) delay, (b) energy, and (c) EDP. ....	107
Fig. 6-13. Condition 3: 2-bit WRITE performance (01 to 10) for SB vs MB for 8-bit, 32-bit, 72-bit, and 128-bit arrays; (a) delay, (b) energy, and (c) EDP. ....	108
Fig. 6-14. Condition 4: 2-bit WRITE performance (10 to 01) for SB vs MB for 8-bit, 32-bit, 72-bit, and 128-bit arrays; (a) delay, (b) energy, and (c) EDP. ....	108
Fig. 6-15. Read 00, 01, 10, 11 for SB vs MB; (a) delay, (b) energy, and (c) EDP. ....	109
Fig. 6-16. MB and SB-LUT (a) KT and (b) KE for Virtex4 benchmarks and (c) KT and (d) KE for Virtex5 benchmarks. ....	113
Fig. 7-1. Flexible processor chip layout from [143]. ....	115
Fig. 7-2. DRAM-embedded FPGA layout from [144]. ....	115

Fig. 7-3. Schematic of the nvD latch. Two ReRAMs, M1 and M2 with a ground resistor, R1 form the NV segment of the nvD latch. ....	116
Fig. 7-4. nvD-latch process waveform. ....	117
Fig. 7-5. Successful restore of Q=1 and Q_b=0 after 100ns VDD cut. ....	118
Fig. 7-6. Successful restore of Q=0 and Q_b=1 after 100ns VDD cut. ....	119
Fig. 7-7. Schematic of the nvDFF. The NV component consists of two ReRAMs, M1 and M2 and the grounding resistor, R1. ....	120
Fig. 7-8. Electrical characteristics of the ReRAM model showing (a) the response of the state variable to the input voltage and (b) the device current. ....	121
Fig. 7-9. Process waveforms of the nvDFF. Two VDD cut scenarios are shown. ....	122
Fig. 7-10. Two nvDFFs, FF1 and FF2 combined with an XOR gate to form a NV 2-bit counter. ....	124
Fig. 7-11. Process waveforms of the nvDFF-based 2-bit counter showing recovery after VDD cut. ....	124
Fig. 7-12. Four nvDFFs (FF1, FF2, FF3, and FF4) combined to form a 4-bit NV shift-register. ....	125
Fig. 7-13. Process waveforms of the nvDFF-based 4-bit shift register showing recovery after VDD cut. ....	125
Fig. 7-14. nvDFF displaying normal DFF behaviour. ....	126
Fig. 7-15. Successful restore of Q=0 and Q_b=1 after 137ns VDD cut. Q switches from 1 to 0 at 25ns before VDD interrupt. ....	127
Fig. 7-16. Successful restore of Q=1 and Q_b=0 after 137ns VDD cut. ....	127
Fig. 7-17. Removal of NV segment results in failure to restore Q=0 and Q=1 after VDD cut. ....	128
Fig. 7-18. Switching delay timings for nvDFF and vDFF for different transistor technology nodes. ....	128
Fig. 7-19. -0.1VDD displaying normal DFF behaviour and successful restore after power supply cut. ....	129

Fig. 7-20. +0.1VDD displaying normal DFF behaviour and successful restore after power supply cut.....	130
Fig. 7-21. Power dissipation of active components in the nvDFF during active operation...	131
Fig. 7-22. Process variation effect on Clk-to-Q delay time. ....	133
Fig. 7-23. Process variation effect on Q rise/Q_b fall time. ....	133
Fig. 7-24. Process variation effect on Q fall/Q_b rise time. ....	133
Fig. 7-25. Process variation effect on maximum Q voltage. ....	134
Fig. 7-26. VDD, Clk, Q1, and Q2 waveforms for NV 2-bit counter. Regular 2-bit counter operation occurs between 0—26ns before VDD interrupt tests during Clk high (A1 and A2) and Clk low (B1 and B2). ....	135
Fig. 7-27. VDD, Clk, Input, D1, D2, D3, and D4 waveforms for NV 4-bit shift-register. Regular shifting operation occurs between 0—26ns before VDD interrupt tests during Clk high (A1 and A2) and Clk low (B1 and B2). ....	136
Fig. 7-28. 2T1R nvDRAM cell storing logical ‘1’ and ‘0’.....	137
Fig. 7-29. 2T1R nvDRAM READ1 and READ0 operation. ....	138
Fig. 7-30. WRITE1 and WRITE0 operations of the 2T1R nvDRAM cell.....	140
Fig. 7-31. READ1 and READ0 operations of the 2T1R nvDRAM cell.....	141
Fig. 7-32. Bit-flip 0-to-1 test of the 2T1R nvDRAM cell.....	141
Fig. 7-33. Bit-flip 1-to-0 test of the 2T1R nvDRAM cell.....	142
Fig. 7-34. Bit hammer test for stored DATA1 and DATA0.....	142
Fig. 7-35. Comparison of WRITE0 and WRITE1 EDPs for nvDRAM cells.....	144
Fig. 7-36. Comparison of READ0 and READ1 EDP for nvDRAM cells. ....	145
Fig. 8-1. Schematic of a BLE [161]......	148
Fig. 8-2. Schematic of a CLB comprised of <i>N</i> number of BLEs [161]. ....	149
Fig. 8-3. Block diagram of nvCLB. ....	149
Fig. 8-4. Traditional interconnect layout [163]......	150
Fig. 8-5. FPGA architecture showing CLBs and the interconnects [163]. ....	150

Fig. 8-6. Block diagram of nvSB combined with nvCLB. .... 151

Fig. 8-7. ReRAM states in the MB NV LUT during normal operation (left) with the insertion of data '100111' and after power restoration (right). .... 155

Fig. 8-8. The Q/Q\_b outputs and ReRAM states in the nvDFF during normal operation (left) with the insertion of data from the MB-nvLUT and after power restoration (right). .... 156

## List of Tables

---

TABLE 1-1. Comparison of FPGA Technologies.....	2
TABLE 2-1. Device characteristics of mainstream and emerging memory technologies.....	28
TABLE 2-2. Comparison between metal-oxide ReRAM and conductive bridge ReRAM [75] .....	30
TABLE 2-3. Activation energy for RS models .....	35
TABLE 2-4. Parameters used in Schottky Modelling .....	46
TABLE 2-5. Error Analysis of the Simulated Results against Experimental Measurements .	53
TABLE 5-1. Logic Table for Controller READ and WRITE Scheme.....	78
TABLE 5-2. Read Delay and Energy Comparison.....	90
TABLE 6-1. Model Parameters .....	93
TABLE 6-2. Array Logic Table for MB-nvLUT.....	97
TABLE 6-3. Array Logic for SB-nvLUT .....	98
TABLE 6-4. Controller READ and WRITE Logic Table for M1 .....	100
TABLE 6-5. Controller READ and WRITE Logic Table for M2.....	100
TABLE 6-6. READ Controller Logic.....	101
TABLE 6-7. Comparator Logic Block for READ Controller .....	101
TABLE 6-8. Comparison of Condition 1: WRITE 0 Delay, Energy and EDP .....	107
TABLE 6-9. Comparison of Condition 2: WRITE 1 Delay, Energy and EDP .....	107
TABLE 6-10. Comparison of Condition 3: 2-bit WRITE (01 to 10) Delay, Energy and EDP .....	108
TABLE 6-11. Comparison of Condition 3: 2-bit WRITE (01 to 10) Delay, Energy and EDP .....	109
TABLE 6-12. Comparison of READ 00, 01, 10, 11.....	110
TABLE 6-13. Virtex4 benchmark circuit comparison of average write delay and EDP for MB- LUT, SB-LUT, and SRAM LUT.....	112

TABLE 6-14. Virtex5 benchmark circuit comparison of average write delay and EDP for MB-LUT, SB-LUT, and SRAM LUT. ....	112
TABLE 6-15. TITAN23 benchmark circuit comparison of average write delay and EDP for MB-LUT, SB-LUT, and SRAM LUT. ....	112
TABLE 7-1. Clk-to-Q Timings for nvD latch and vD latch.....	119
TABLE 7-2. Clk-to-Q Timings for nvD latch and vD latch.....	120
TABLE 7-3. ReRAM Model Parameters.....	121
TABLE 7-4. Clk-to-Q Timings for nvDFF and vDFF .....	129
TABLE 7-5. nvDFF Clk-to-Q Timings With $\pm 0.1$ VDD Variability .....	130
TABLE 7-6. Average Power Dissipation .....	132
TABLE 7-7. Worst-case Power Dissipation.....	132
TABLE 7-8. nvDFF Design Comparison .....	134
TABLE 7-9. READ Voltages and Cycles to Failure without Restore.....	139
TABLE 7-10. WRITE and READ delay, Energy Dissipation, Retention Time, and EDP ...	143
TABLE 8-1. nvFPGA vs vFPGA performance comparison (A number higher than 1 is better, lower than 1 is worst).....	154

## List of Abbreviations

BE	Bottom Electrode
BLE	Basic Logic Elements
CF	Conductive Filament
CLBs	Combinatorial Logic Blocks
CMOS	Complementary-Metal-Oxide-Semiconductor
CPUs	Central Processing Units
DFFs	D Flip-flops
EDA	Electronic Design Automation
FPGAs	Field Programmable Gate Arrays
GPUs	Graphics Processing Units
HRS	High Resistance State
I/Os	Inputs/Outputs
IOT	Internet-of-Things
IRS	Intermediate Resistance States
LRS	Low Resistance State
LUTs	Lookup Tables
MB	Multi-Bit
MB-ReRAM	Multi-Bit ReRAM
MGISM	Metal-Graphene-Insulator-Semiconductor-Metal
MIGM	Metal-Insulator-Graphene-Metal
MIM	Metal-Insulator-Metal
MISM	Metal-Insulator-Semiconductor-Metal
MLCs	Multi-level cells
MOS	Metal-Oxide-Semiconductor
NV	Non-volatility
nvDFFs	Non-volatile DFFs
nvLUTs	Non-volatile LUTs
ReRAMs	Resistive Random-Access Memories
RS	Resistive Switching
SBs	Single-Bit

SRAMs	Static Random-Access Memories
SwBs	Switch Blocks
TE	Top Electrode
TMOs	Transition-Metal-Oxides
VLSI	Very-Large-Scale-Integration
V <sub>Os</sub>	Oxygen Vacancies

## List of Publications

---

### Peer Reviewed Journals:

1. Jagath, A.L., H.L. Chee, Kumar, T.N., and H.A. Almurib, "Insight into physics-based RRAM models – review," *IET The Journal of Engineering*, vol. 2019, no. 7, pp. 4644–4652, 2019.
2. H. L. Chee, T. N. Kumar, and H. A. F. Almurib, "Electrical model of multi-level bipolar Ta<sub>2</sub>O<sub>5</sub>/TaO<sub>x</sub> Bi-layered ReRAM," *Elsevier Microelectronics Journal*, vol. 93, no. March, p. 104616, 2019.
3. H. L. Chee, Y. Z. Kok, T. N. Kumar, and H. A. F. Almurib, "Sense amplifier for ReRAM-based crossbar memory systems," *Taylor and Francis International Journal of Electronics Letters*, pp. 1–13, 2022.
4. H. L. Chee, T. N. Kumar, and H. A. F. Almurib, "Low energy non-volatile look-up table using 2 bit ReRAM for field programmable gate array," *IOP Semiconductor Science and Technology*, vol. 37, no. 6, 2022.

### Conferences:

5. H. L. Chee, T. N. Kumar, and H. A. Almurib, "Multifilamentary Conduction Modelling of Bipolar Ta<sub>2</sub>O<sub>5</sub>/TaO<sub>x</sub> Bi-Layered RRAM," *Proceedings in 7th IEEE Non-Volatile Memory Systems and Applications Symposium (NVMSA)*, 2018, pp. 113–114. (Japan)
6. H. L. Chee, T. N. Kumar, H. A. F. Almurib, and D. W. H. Kang, "Analysis of a Novel Non-Volatile Look-Up Table (NV LUT) Controller Design with Resistive Random-Access Memories (RRAM) for Field-Programmable Gate Arrays (FPGA)," *Proceedings in 2019 IEEE Regional Symposium on Micro and Nanoelectronics (RSM 2019)*, pp. 87–90, 2019, doi: 10.1109/RSM46715.2019.8943560. (Malaysia)
7. W. J. X. Ng, H. L. Chee, T. N. Kumar, and H. A. F. Almurib, "A ReRAM-based Nonvolatile PIM," *Proceedings in 20th IEEE Student Conference on Research and Development (SCoReD 2022)*, 2022. (Malaysia)
8. H. L. Chee, T. N. Kumar, and H. A. F. Almurib, "A ReRAM-based Nonvolatile FPGA," *Proceedings in 20th IEEE Student Conference on Research and Development (SCoReD 2022)*, 2022. (Malaysia)

### Manuscripts Submitted and Under Review:

9. A Low Power Nonvolatile DRAM Cell based on ReRAMs, *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*.

10. Low Power ReRAM-based Nonvolatile D Flip-Flop Design for Nonvolatile Processor, *Circuits, Systems, and Signal Processing*.

**Manuscripts Under Preparation:**

11. A Nonvolatile D-Latch using ReRAMs
12. Radiation Hardened ReRAM-based Arrays

## 1. Chapter One

# Introduction

---

Modern 21<sup>st</sup> century society is highly driven and supported by electronics and technology. Ever since the discovery of semiconductors, electronic devices have impacted human culture and have become the solution or ‘go-to’ devices for a wide range of applications be it in communications, automotive, medical, entertainment, and many more fields. In the industrial sector, the past 10 years have seen the development of Internet-of-Things (IoT), industrial automation via the Fourth Industrial Revolution, cryptography, massive cloud-scale computing services, and neuromorphic computing amongst many others. These applications offer solutions to a wide variety of challenges but are common in their requirements of high computing power and low energy consumption or alternatively high energy efficiency. Correspondingly, these demands are expected to increase as these sectors continue to mature.

A promising solution to these challenges is to look beyond traditional central processing unit (CPU) and graphics processing unit (GPU)-based architectures; Field Programmable Gate Arrays (FPGAs) are highly flexible massive scale silicon circuits that are able to be configured down to their hardware configurations and routings. First, due to parallel processing ability FPGAs offer better efficiency in processing streaming data from inputs/outputs (I/Os) than CPUs and GPUs, leading to better throughput and latency [1], [2]. Second, FPGAs offer high reconfigurability at a fine granularity on a massive scale and can be adapted to best fit any algorithm; its high computational throughput allows acceleration of high-concurrency and high-dependency algorithms. Third, FPGAs consume an order of magnitude lower power than CPUs and GPUs and have two orders of magnitude in energy consumption efficiency for stream data processing and high-dependency task execution leading to savings in power consumption and costs [3], [4], [5]. The reconfigurability and high throughput of FPGAs are also highly desirable and have led to the development of embedded FPGAs [6], [7].

TABLE 1-1. Comparison of FPGA Technologies

Parameters	FPGA Technologies		
	SRAM	Flash	Anti-fuse
Volatile	Yes	No	No
Reprogrammable	Yes	Yes	No
Area Utilization	High	Med	Low
Cell Size	4-6 transistors	1 transistor	0 transistor

Current FPGAs in the market are mainly based on three different memory technologies to store the configuration bits; static random-access memories (SRAMs), Flash, and anti-fuse (TABLE 1-1) [8]. Amongst the three technologies SRAM-based FPGAs dominate the market share due to the maturity of its technology while Flash-based FPGAs are a close second. Anti-fuse FPGAs are a small percentage of the market as they cannot be reconfigured after the first programming but convey advantages in terms of NV, lower latency, smaller area utilization, and resistance to radiation.

SRAM-based FPGAs typically comprise of 4 to 6 complementary-metal-oxide-semiconductor (CMOS) transistors in a single SRAM cell. Metal-oxide-semiconductor (MOS) Transistors have been very successful for digital electronics but are facing limitations in terms of physical scaling. Moore's Law which predicts the doubling of transistors on a silicon wafer annually has slowed down in recent times due to the physical limits of CMOS technology as transistors reach ~10 nm nodes [9]. At the time of writing, Intel, the largest CPU producers have announced setbacks on 7 and 10 nm manufacturing process [10], [11] while the second largest CPU producers, AMD are using TSMC's 7 nm node technology which has been shown to be marketing terminology and is actually similar to Intel's 10 nm node processes [12], [13]. SRAMs also use large silicon estate as the technology requires 4-6T per cell. This is a disadvantage when it comes to large high-density arrays, which is a hallmark and an increasing requirement of 21st century computing devices. Another disadvantage of SRAM-based FPGAs is NV, as it can only retain data with power supply. As a result, they always consume power regardless if they are in active or idle states. They also need to be reprogrammed every time a computing device is started up and are therefore unable to be used as storage devices.

On the other hand, Flash memories are based on floating-gate transistors and are very stable devices (data retention of >10 years) [14]. They are therefore non-volatile, providing retention and energy efficiency advantages over SRAMs. Recent developments in the technology have seen the introduction of multi-level cells (MLC) where multiple bits can be stored in a Flash

cell to improve memory array density [15]. Flash technology-based FPGA however has multiple disadvantages. The stability of the stored charge in the cell results in high WRITE voltages and long WRITE times. READ times are also long due to the Flash architecture where a whole row in the memory array has to be read instead of just a single cell. Flash technology also has a drawback in terms of endurance as the cell can be written  $\sim 10^4$  times; much lower than the SRAM cell which has an endurance of  $>10^{16}$ .

### **1.1 Next Generation Non-volatile FPGA**

FPGAs host a large number of interconnected programmable configurable logic blocks (CLBs) which are made up of lookup tables (LUTs) and D Flip-flops (DFFs). Conventional FPGA LUTs consist of SRAM cells and together with DFFs are both non-volatile components. FPGAs therefore have to be reprogrammed at startup; either manually or by storing configuration bits in a separate storage device ie. Flash. Every CLB in the FPGA are connected to each other through programmable interconnects to account for design flexibility. The configurations of these interconnects are also stored in volatile SRAM cells and have the same startup requirements as the SRAM-based LUTs. As large-scale silicon circuits, SRAM-based FPGAs are energy inefficient due to the constant power supply requirements and volatility while Flash-based FPGAs are limited by slow WRITE/READ times and the technology's low cycling endurance, negating the reconfigurability advantage of the FPGA.

Due to these facts, there have been numerous attempts to introduce non-volatility (NV) into FPGAs through next-generation NV memory devices such as phase change memory (PCM) [16], [17], spin-torque-transfer magnetic random-access-memory (STT-MRAM) [18], [19], ferroelectric field-effect transistor (FeFET) [20], [21], and resistive random-access memories (ReRAMs) [22], [23]. A holistic implementation of NV into FPGAs through the use of ReRAMs is an aim of this research project. The components that make up the FPGA; LUTs, CLBs, switching blocks (SwBs), and the dynamic random-access memory (DRAM) are improved in this project with the introduction of ReRAMs.

The downscaling capabilities and electrical performances of ReRAMs coupled with their intrinsic NV makes them suitable candidates as the main memory technology for next generation non-volatile FPGAs. Since the FPGA is a very-large-scale-integration (VLSI) device made up of many subcircuits, there are multiple opportunities to introduce the ReRAM as the memory component to replace conventional memory devices and also implement NV.

Replacing the SRAM cells in the LUTs with multi-bit (MB) ReRAMs (MB-ReRAMs) would introduce NV to the LUT array and increase the array density due to the higher number of bits-per-cell of the MB-ReRAM cell compared to the SRAM cell. The MB-ReRAM is also a single device compared to the SRAM cell which typically consists of four or six transistors. Additionally, implementing ReRAMs into the DFF design would also introduce NV to the DFF thus producing a non-volatile CLB. The CLB interconnects can also be made to be non-volatile by replacing the SRAMs that store the interconnect configurations in the switch blocks (SBs) with ReRAMs. These changes would allow the FPGA to be non-volatile, removing the necessary programming at startup as well as negating the need for separate storage devices.

## 1.2 Next Generation Memory Device: Resistive Random-Access Memory

A solution to the challenges of the SRAM and Flash-based FPGAs exist in the ReRAM which are next generation memory devices. ReRAMs offer NV, low WRITE/READ voltages and latencies, low power consumption, and high retention and endurance. ReRAMs show high potential amongst the emerging next generation memory devices. They have fast switching ( $\sim 10$  ns), fast READ times ( $< 10$  ns), low WRITE voltages ( $< 3$  V), low WRITE energy ( $\sim 0.1$  pJ per bit), high retention ( $> 10$  years), and high endurance ( $\sim 10^{12}$ ).

ReRAMs belong to a group of devices called memristors which were first reasoned by Leon Chua as a fourth fundamental circuit element for symmetry arguments (Fig. 1-1) [24]. At the time, there were six different mathematical relations used to connect the four fundamental circuit variables; the electric current  $i$ , the electric voltage  $v$ , the electric charge  $q$ , and the magnetic flux  $\phi$ . The mathematical equation describing the relation between charge and flux

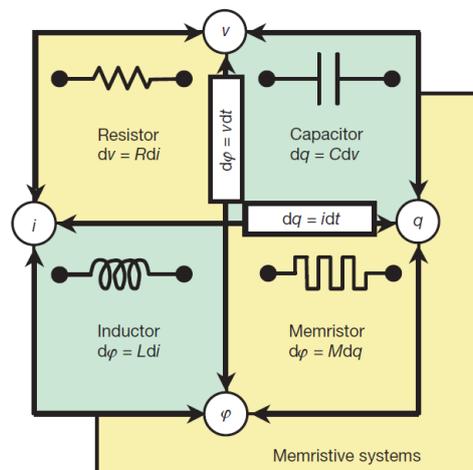


Fig. 1-1. The four fundamental circuit elements and their mathematical relations [25].

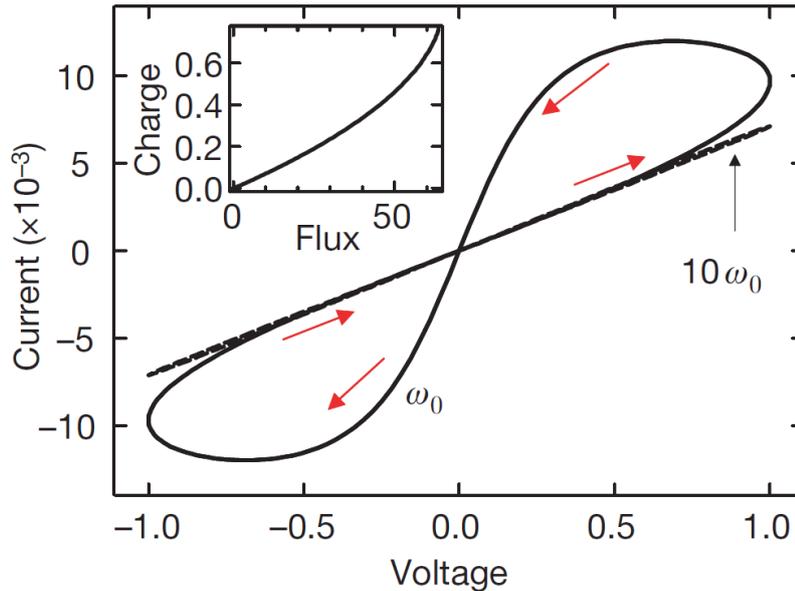


Fig. 1-2. The current-voltage (IV) hysteresis curve showing change of resistance according to applied voltage [25].

was missing and was presented as  $d\varphi = M dq$ ,  $M$  being the memristance. He theorized that memristive devices would be the equivalent of resistors that have memory; hence the name memristor is a portmanteau of memory-resistor.

This discovery of memristors as physical devices occurred in 2008 in [25]. It was demonstrated that when  $M$  is a function of  $\varphi$ , the device shows a hysteresis curve (Fig. 1-2), a unique property irreproducible by any combination of nonlinear resistive, capacitive, and inductive components. The hysteresis is dependent on the frequency of the applied voltage and collapses into a straight line at high frequencies. The device is thus identical to a linear resistor at high frequencies.

The physical memristors were called ReRAMs as they functioned on the RS mechanism. The resistance of the device is dependent on the applied voltage and the device switches between two distinct resistance levels; a low resistance state (LRS) and a high resistance state (HRS). The most basic ReRAM configuration is a multilayer two-terminal device with two metal electrodes sandwiching an insulator layer made up of transition-metal-oxides (TMOs) to form a metal-insulator-metal (MIM) structure; the cross-section of an MIM ReRAM is given in Fig. 1-3.

When voltage is applied across the two terminals (top electrode, TE and bottom electrode, BE), a conductive filament (CF) made up of mobile ions or oxygen vacancies ( $V_{Os}$ ) is formed

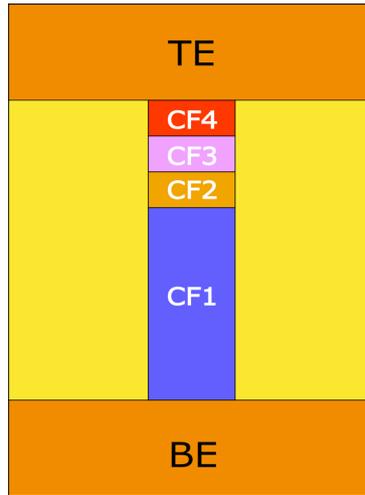


Fig. 1-3. Cross-section of a MIM ReRAM cell.

in the insulator layer that shunts the two electrodes, achieving LRS. This CF can then be broken to increase the resistance of the device and switch the ReRAM into HRS. Due to the mobility of the ions/ $V_{OS}$  in the CF, the CF can be partially broken by stopping the voltage supply to produce intermediate resistance states (IRS); IRS states exist in Fig. 1-3 when the fully formed CF ( $CF1+CF2+CF3+CF4$ ) is partially broken into  $CF1+CF2+CF3$  or  $CF1+CF2$ , LRS is achieved when the CF is fully broken (only CF1). The ReRAM therefore has an intrinsic ability to support MB switching and can store more than one-bit-per-cell. The MB-ReRAM increases the density of the memory cell and would provide an advantage in memory array sizes.

Continuous research and development into ReRAM technology has produced further enhancements on this basic device configuration; the addition of extra layers between the metal-insulator interface has led to metal-insulator-semiconductor-metal (MISM) ReRAMs and metal-insulator-graphene-metal (MIGM) or metal-graphene-insulator-semiconductor-metal (MGISM) ReRAMs. These auxiliary layers have been shown to provide improvements in device retention and control over switching characteristics although the introduction of the nanometer-thick layers presents quantum behavioural and variation effects [26], [27].

### 1.3 Research Gaps in the Field and Research Motivations

The FPGA is made up of multiple components that are suitable for the introduction of NV, namely the LUTs and DFFs in the CLBs and the SwBs in the routing architecture. As these components are conventionally MOSFET-based, an opportunity exists to replace them with next-generation ReRAM devices. Advances in ReRAM devices have shown the existence of intrinsic MB capabilities which convey advantages in memory densities and array sizes, providing further benefits to the ReRAM-based LUT array structure. The implementation of

the MB-ReRAM into circuit designs requires study into the modelling and the physical phenomenon behind the device.

Although plenty of MB-ReRAM models based on single CF rupture/formation exist, there is only one model for multi-CFs ReRAMs [28]. The multi-filament formation is modelled in [28] using the drift/diffusion equation for  $V_O$  migration, carrier continuity equation, and Fourier equation for Joule heating. There is however no model for multi-filament ReRAMs that consider the effects of the Schottky tunnel barrier that exists between the electrode and the TMO. The development of an electrical model based on this tunnelling barrier effects would contribute to further understanding of the ReRAM, especially the MB-ReRAM as well as a model that can be used in electronic design automation (EDA) circuit simulations. This is the contribution of the first stage of this research.

The next stage of the research looks at the implementation of ReRAMs in LUT arrays. ReRAM-based nonvolatile LUTs (nvLUTs) exist in literature in [29], [30], [31] but use single-bit (SB)-per-cell ReRAMs. The memory array density of the LUTs which make up the majority of layout area in FPGAs can be further improved by using MB-ReRAMs that can hold >1-bit-per-cell. Unlike the SB-ReRAM LUTs which can serve as a direct replacement for SRAM LUTs, MB-ReRAM LUT arrays require specially designed controllers to store input data into the MB cells.

The D Latch and DFF serves as the register latch components in the FPGA's CLB. They are CMOS-based and are therefore volatile. The data in these devices are affected during power disruption events such as power cuts or when the FPGA is shut down and they have to be reprogrammed every time it happens. The power consumption from these requirements can be reduced with the implementation of NV and this is the third contribution of this research. The NV D Latch (nvD Latch) and NV DFF (nvDFF) would be robust to these events and serve as its own data storage device, negating the need for external configuration storage.

Dynamic random-access-memories (DRAMs) in FPGAs are used as storage block memories and consist of conventional one transistor and one capacitor per cell (1T1C) architecture. They are volatile memory devices and face massive scaling limitations as a result of the capacitor's leakage and retention variability [32]. The capacitor's charge leakage also means that DRAMs have to be refreshed after a certain time period during continuous operation. The fourth

contribution of this research addresses these drawbacks through the design of an NV DRAM (nvDRAM) by substituting the capacitor with the ReRAM.

The final contribution of this research produces the holistic nvFPGA comprised of the discussed NV components. This holistic nvFPGA design based on ReRAMs is not available in literature to the author's knowledge. Inside the nvFPGA, the NV MB-RERAM LUT and nvDFF replaces their respective counterparts in the FPGA's CLBs. As for the FPGA's SwBs, which are configurable and control the routing paths between CLBs and I/Os, the nvDRAMs store the routing configuration bits instead of the volatile SRAMs.

## **1.4 Aim and Objectives**

### **1.4.1 Aim**

The aim of this PhD project is to develop an nvFPGA architecture using ReRAM cells with better features than Flash-based and SRAM-based FPGAs.

### **1.4.2 Objectives**

1. To design an electrical model of an MB-ReRAM with multi-filaments is first developed for use in EDA simulations. First, a literature review of the physics and mechanisms of ReRAM behaviour, the growth of multi-filaments, and the tunnelling barrier effects is conducted. The model is created for usage in EDA software LTSpice and is then verified on the same platform. To verify the model, the simulation result is analyzed for similarity with experimental results.
2. The study of existing nvLUTs is first carried out to look for gaps in the field. From that, a voltage-mode sense amplifier (SA) circuit design that is capable of sensing the low output voltages from reading a ReRAM cell in an array is developed.
3. To determine the potential for the introduction and study of MB-nvLUTs using MB-ReRAMs which provide improvements in array density, area utilization, and power consumption. The MB-nvLUT structure is then designed and verified through EDA. The established MB-nvLUT is found to require a novel controller design to account for the MB-ReRAMs in the array. A MB-controller is thus designed for the nvLUT.
4. To implement NV into the D latch and DFF, components that function as the register latch of the CLB with the introduction of ReRAMs. The nvD Latch and nvDFF are then tested for robustness during power disruption events.

5. To design a ReRAM-based nvDRAM that can be used in the FPGA's SwBs for FPGA routing configurations storage. The ReRAM-based nvDRAM is designed and tested as a next generation replacement for SRAM cells that have high area utilization (6T per cell) and are volatile. The nvDRAM is a necessary design as the SwB requires continuous power during operation.
6. To design a holistic nvFPGA architecture comprised of the previous NV components which is then tested and compared with SRAM-based FPGA with parameters such as power consumption, energy delay product (EDP), and WRITE/READ latency timings.

### 1.5 Research Outcomes

1. An electrical model of a MB-ReRAM based on multi-filamentary conduction is developed. The model is tested for accuracy and the multi-filaments achieve MB-RS due to the difference of activation energy ( $E_A$ ) and is shown to match experimental results.
2. A low-voltage SA is developed that is capable of sensing low-voltage (nV-mV) outputs of a ReRAM memory array.
3. A MB-nvLUT and its controller is developed. The operations for 4-bit-per-cell ReRAM arrays are demonstrated for 4, 6, and 8-input LUTs. The results are then compared with SB-nvLUT and SRAM-based LUT arrays.
4. A nvDFF is developed with the incorporation of two ReRAMs which successfully demonstrates NV.
5. A nv D latch is developed as a replacement for SRAM cells in the FPGA SwB.
6. The parameters for a holistic nvFPGA architecture are measured and compared with SRAM-based FPGA.

### 1.6 Thesis Outline

**Chapter 1: Introduction** presents the background and preliminaries of the FPGA and the ReRAM and thus introduces the basis of this research work. The gaps and opportunities in the field are discussed followed by the aims and objectives of tackling them. The idea of a holistic nvFPGA using combinations of SB and MB-ReRAMs is introduced.

**Chapter 2: Literature Review** is a thorough analysis and discussion of the components that make up this research work. Current FPGA architectures are studied together with the fundamental FGPA components; namely the LUT and the DFF in the CLB which make up the

majority of FPGA cell area and are responsible for combinatorial and sequential logic and the SwB which stores and handles the FPGA's routing configuration bits. The SRAM and DRAM cell is analyzed, showing the benefits and limitations of their transistor-based technology. This is followed by the discussion of next generation memory devices that have potential to replace transistor-based memory devices. Focus is placed on the ReRAM as one of the more promising next generation memory devices with a discussion of its physics, behaviour, and modelling (analytical and electrical).

**Chapter 3: Methodology** demonstrates the flow of this research work. A description of the methods and tools used is provided in this chapter. The workflow of this project is also included here.

**Chapter 4: Multi-bit ReRAM Model** presents the electrical model of the MB-ReRAM based on multiple filaments. The model is verified with experimental results and is tested under different voltage input conditions.

**Chapter 5: Non-volatile LUT** is subdivided into three sections. The first is an EDA analysis of a ReRAM-based SB-nvLUT. The WRITE/READ disturb on unselected cells are measured as well as the performance of the LUT controller. It is then followed by a proposed design of a ReRAM array specific sense amplifier circuit capable of READ operation for subthreshold output voltages from ReRAMs and its analysis.

**Chapter 6: MB-nvLUT** presents the design of the MB-nvLUT and its controller circuit. A specific WRITE scheme to incorporate MB array cells and a controller capable of this scheme are designed and simulated. WRITE/READ delay, energy dissipation, and EDP performance metrics for the MB-nvLUT are measured and compared with the SB-nvLUT. Comparison amongst the MB-nvLUT, SB-nvLUT, and SRAM-based FPGAs performance benchmark tests rounds out the chapter.

**Chapter 7: Non-volatile Memories** presents the nvD latch, nvDFF, and the nvDRAM circuits. Detailed discussions of the EDA simulation results are provided for each circuit. Clk-to-Q delay, Q rise time/Q<sub>b</sub> fall time, Q fall time/Q<sub>b</sub> rise time, and power dissipation performance metrics are measured for the nvD latch and nvDFF design. These metrics are then compared with their SRAM-based circuits and with NV designs presented in other literatures. As for the nvDRAM, performance metrics for the

WRITE0/READ0/WRITE1/READ1 delay, energy dissipation, and EDP performance metrics are measured and compared with other works in literature.

**Chapter 8: The Nonvolatile FPGA Architecture** discusses the holistic nvFPGA design incorporating NV components from the previous chapters. The layout of the nvFPGA which consists of an nvCLB and an nvSwB are presented. An EDA analysis is performed for circuit parameters such as power consumption, EDP, and WRITE/READ latency timings.

**Chapter 9: Conclusion** provides a roundup of the thesis and the work carried out in this research project. A brief section discusses the possibilities and future work possibilities that emerges from this research project.

## 2. Chapter Two

# Literature Review

---

### Related Publications

---

1. Jagath, A.L., H.L. Chee, Kumar, T.N., and H.A. Almurib, “Insight into physics-based RRAM models – review,” *IET The Journal of Engineering*, vol. 2019, no. 7, pp. 4644–4652, 2019. (This paper was co-authored by A. Lekshmi Jagath and this thesis’ author)

## 2.1 Field Programmable Gate Arrays

FPGAs are integrated silicon circuits that consist of programmable blocks of logic called CLBs which are interconnected through routings called configurable interconnects. The LBs are the main digital processing resources and can be programmed to perform combinatorial or sequential logic operations depending on the design. Inside the LB, LUTs are responsible for the combinatorial logic while DFFs are tasked with the sequential logic. New generation LUTs are capable of additional functions such as local storage (distributed RAM), Shift Register (SR), multiplexer, and adder/subtractor operations.

Modern FPGAs also consist of input/output (I/O) blocks which function as the FPGA's means of communication with external devices, memory blocks for block RAM memory, and arithmetic Digital Signal Processing (DSP) blocks for DSP specific applications. A generic structure of an FPGA with the features mentioned above is shown in Fig. 2-1 [33].

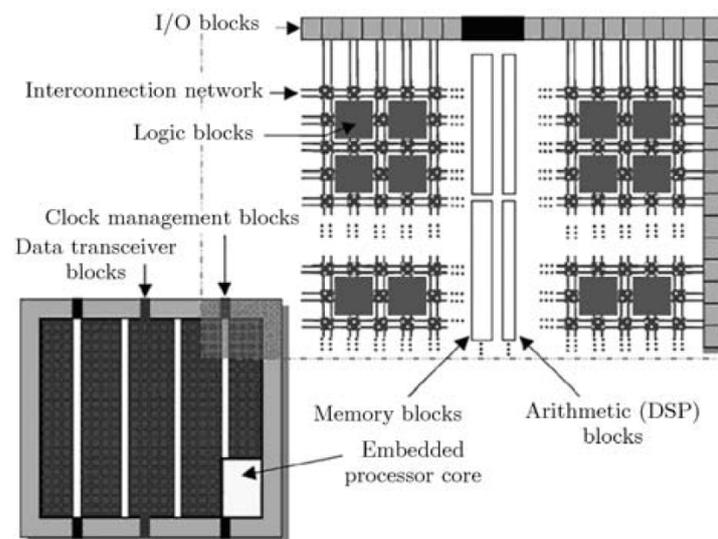


Fig. 2-1. Generic FPGA structure and internal components [33].

Created in the 1980s, FPGAs are mainly based on three types of programming technologies; static random-access-memories (SRAMs), FLASH, and fusible-link.

### 1. Static Random Access-Memory (SRAM) FPGAs

SRAM FPGAs are the most commonly used FPGAs owing to the fact that SRAM devices are the most mature technologies. SRAMs are volatile memory cells consisting of six transistors (Fig. 2-2) and are the fastest memory cells in consumer

electronics. They can be commonly found in consumer electronics such as the processor in personal computers (PCs), mobile phones, etc.

As a result of wide market adoption, SRAMs are at the forefront of fabrication processes and have the smallest process nodes. As of the time of writing, industry SRAMs are fabricated in 10 nm process nodes. To achieve fast speeds, SRAMs use a 6-transistor (6T) layout as shown in Fig. 2-2 [34] at a cost to the cell area footprint.

Transistors M5 and M6 form the access transistors and are controlled by the wordline, WL. The WL is set to logic '1' (high) during READ and WRITE to enable passthrough of the bitline, BL voltages which carry the data to be written into the memory component of the SRAM cell. The WL is set to logic '0' (low) outside of READ and WRITE operations to disconnect the BL from the SRAM's memory component. Transistors M1—M4 are cross-coupled inverters and latch the stored data provided there is power supply from  $V_{dd}$ . The SRAM cell is therefore volatile and the FPGA loses stored data when the power supply is removed. SRAM FPGAs thus require external storage components or reprogramming after every power disruption event.

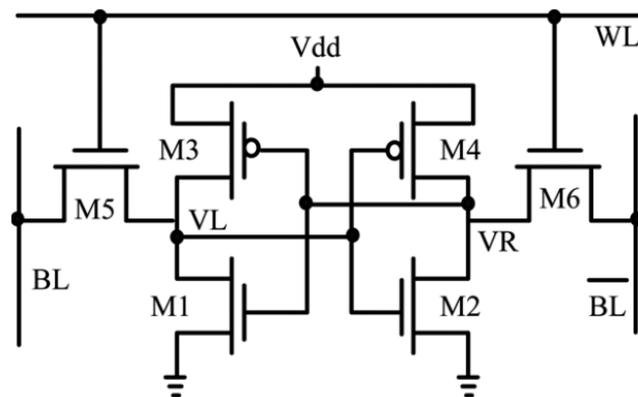


Fig. 2-2. The 6T SRAM [34].

## 2. Flash FPGAs

Flash memory technology is based on floating-gate transistors (Fig. 2-3). Compared to the standard metal-oxide-semiconductor (MOS) transistor, the floating-gate transistor has an isolated floating gate between the control gate and the drain-source channel. The floating gate is uncharged when it is unprogrammed and the control gate operates normally; control gate voltage that is larger than the

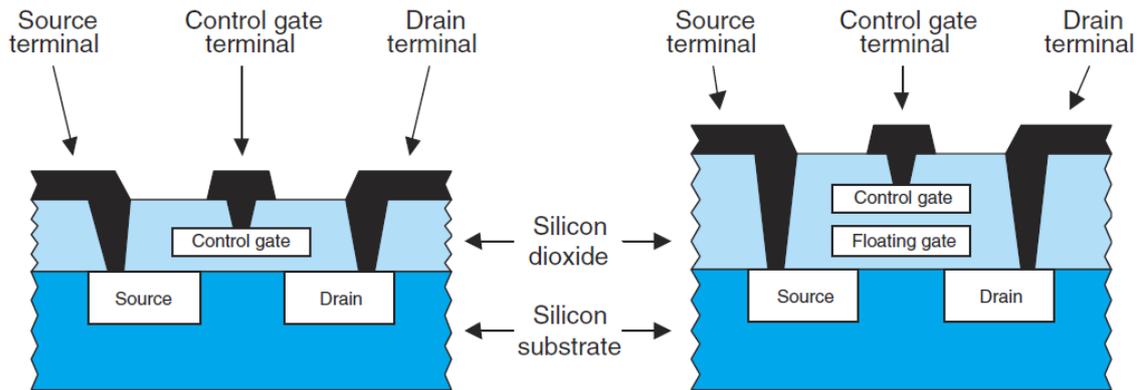


Fig. 2-3. Standard transistor (left) vs Flash transistor (right) [35].

threshold voltage of the transistor will switch the transistor on and allow current flow between the drain and source channels.

To program the transistor, high voltage (~12 V) is supplied at the control gate to force electrons through the oxide into the floating gate through a process called hot electron injection. This induces a negative charge in the floating gate and the strong insulation of the surrounding oxide makes the floating gate very stable; under normal conditions the gate will not discharge for years. The charged floating gate now affects the operation of the control gate.

Fig. 2-4 shows an example of a Flash cell layout. The WL is connected to both the control and floating gate and is responsible for the WRITE operation. The BL which is the data line and its voltage levels are used to READ the stored data in the Flash transistor. Negative charge is stored in the floating gate during WRITE1 and the floating gate is discharged during WRITE0. The Wordline is brought high and a voltage is supplied at the the Columnline during READ operation. For stored data=0, the control gate's normal operation is unaffected so the transistor is switched

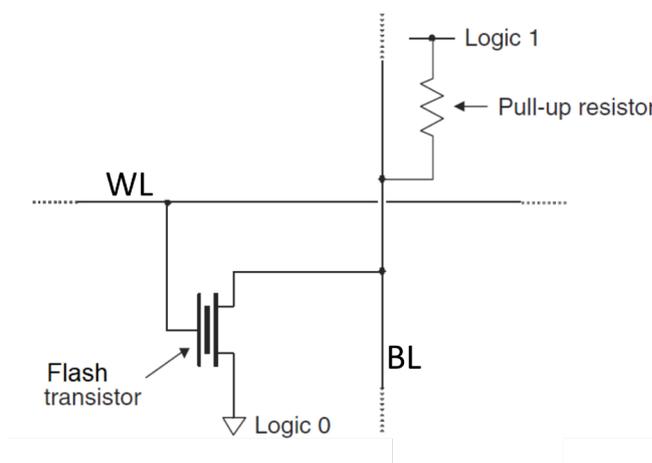


Fig. 2-4. Flash transistor layout [35].

on and a path to ground is created; the Columnline experiences a voltage drop.

Alternatively for stored data=1, the control gate's operation is now blocked by the charged floating gate. The Wordline voltage does not switch on the transistor and no path to ground is created; the Columnline voltage thus does not change.

The high stability of the charged floating gate makes the Flash cell nonvolatile and the FPGA does not require reprogramming after power disruption. Flash cells are also more area efficient compared to SRAM cells. The disadvantages of Flash technology are higher WRITE and READ delays compared to SRAM, erasing is done by clearing the whole device or large portions thereof, non-standard CMOS fabrication, and limited number of reprogramming operations. For example, the Flash-based Microsemi IGLOO2 FPGA is rated for 500-1000 programming cycles.

### 3. Anti-fuse one-time programming

Anti-fuse FPGAs utilize programmable fuses that exist in every routing path and are high resistance in their unprogrammed state. These fuses can be broken to reduce their resistance and 'create' routing paths according to the required design. In contrast to the previously mentioned technologies, the fuses cannot re-programmed once broken and the device is thus one-time programmable (OTP). Anti-fuse FPGAs are nonvolatile and are faster and have smaller footprint compared to the previous mentioned technologies. Its massive drawback is the inability to be reprogrammed and the programming requires special devices to break the fuses.

#### 2.1.1 Configurable Logic Blocks

CLBs are complex fundamental components in the FPGA where combinatorial and sequential logic operations are carried out. External inputs from the FPGA are stored in the CLBs which can implement any  $K$  Boolean logic for  $K$  inputs. The CLB output can be combinatorial or registered to the clock signal to make it sequential. There are two types of LBs; the multiplexer (MUX)-based LBs and LUT-based CLBs.

##### 1. MUX

MUX-based LBs take advantage of the many possible functions available for a two input MUX shown in Fig. 2-5. More complex functions can be obtained by combining MUX inputs and outputs or by connecting the inputs to a constant or a signal. MUX-based LBs offer high functionality for a small number of transistors

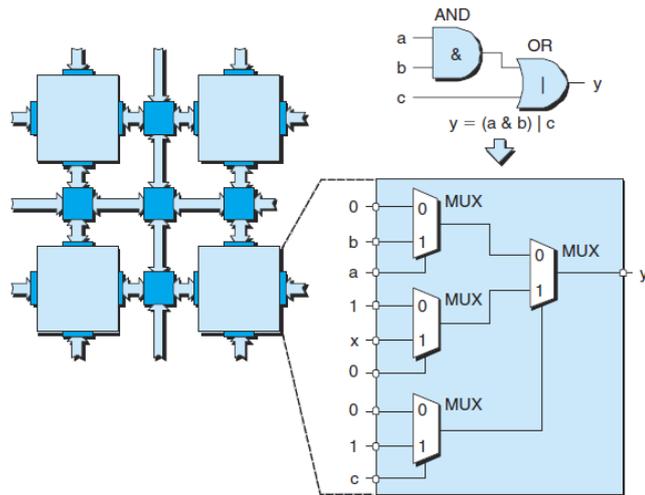


Fig. 2-5. MUX-based LB [35].

but have high demands on routing resources and are not as efficient as LUTs for arithmetic processing.

## 2. LUT

LUTs are arrays of one-bit memory cells. Each cell in the array is programmed so that for a certain group of inputs into the array produces a desired output. The group of input signals function as index that is looked up in the programmed table. LUTs can be formed by SRAMs, antifuses, or FLASH cells and are the most commonly used architectures in commercial FPGAs [35]. Most common LUTs have  $2^n$  SRAMs for  $n$  number of inputs. An example of a SRAM-based array LUT is given in Fig. 2-6. In this array, an active transmission gate passes on the input signal to its output and a disabled transmission gate prevents the input signal from passing through to

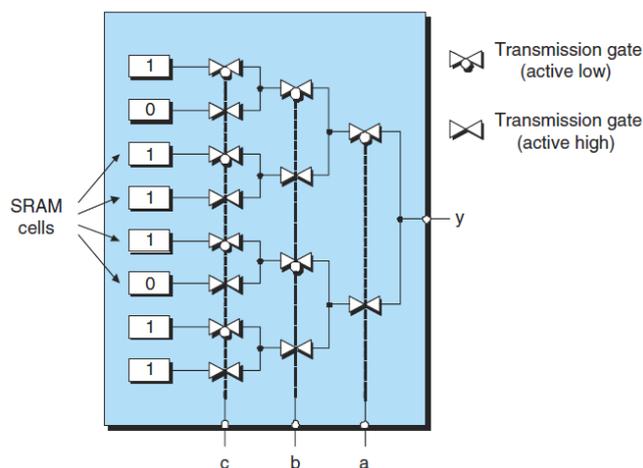


Fig. 2-6. SRAM-based LUT [35].

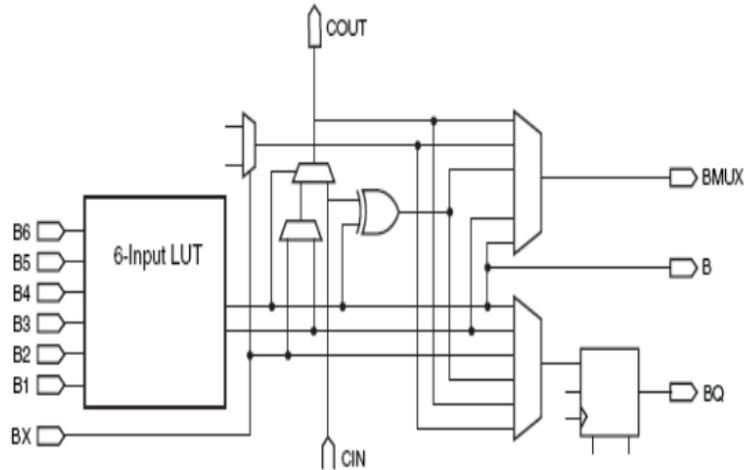


Fig. 2-7. Xilinx Virtex-5 6-input LUT architecture [36].

the output. The array is programmed to allow the function:

$$y = (a \& b)|c \quad (1.1)$$

A typical LB structure consists of an LUT for the combinatorial logic, a DFF for the sequential logic, and a MUX to interconnect the external input, LUT, and DFF. Fig. 2-7 and Fig. 2-8 are examples of a contemporary LB architecture from Xilinx [36] and Altera [37] respectively. Commercial industry FPGAs are mostly SRAM LUT-based due to the maturity and widespread use of SRAM technology. However, SRAMs are nonvolatile devices which require constant power supply and reprogramming every time the FPGA is turned on and contributes to high power usage over time. SRAMs cells also have a large area footprint as SRAMs are usually made up of 4T-6T. This is a drawback of the SRAM architecture and a newer next-

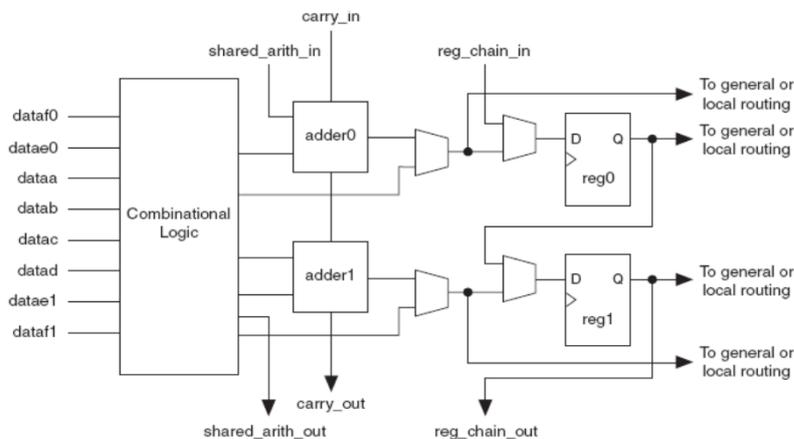


Fig. 2-8. Altera Stratix-II ALM LUT architecture [37].

generational architecture that uses fewer fundamental components has potential to bring down silicon area usage. In addition to that, transistor scaling according to Moore's Law has slowed down in recent decades due to physical limitations in the device. Charge leakage, etc. effects drastically increase when transistors are scaled down to single digit nanometer process nodes. A well-known industry case is the struggle Intel has had scaling down their i-series flagship processors below 14nm.

### 2.1.2 Programmable Interconnects

As the FPGA is highly configurable, the routing or interconnects responsible for signal communications between LBs and I/Os in the FPGA matrix are also required to be programmable. The connections are controlled by programmable interconnect points (PIPs) in SwBs as shown in Fig. 2-9 [38]. The pathing in CMOS-based PIPs is controlled by single pass transistors or transmission gates which are made up of complimentary NMOS and PMOS transistor pairs. The configurations of the pass transistors are stored in CMOS-based D-latch memory cells as shown in Fig. 2-9.

These D-latches memory cells are made up of CMOS inverters and face the same limitations as the SRAM-based LUTs, requiring constant power supply during operations and reprogramming every time the device is switched on. Operating power consumption is thus high and there is potential for improvement by introducing nonvolatility to the circuit.

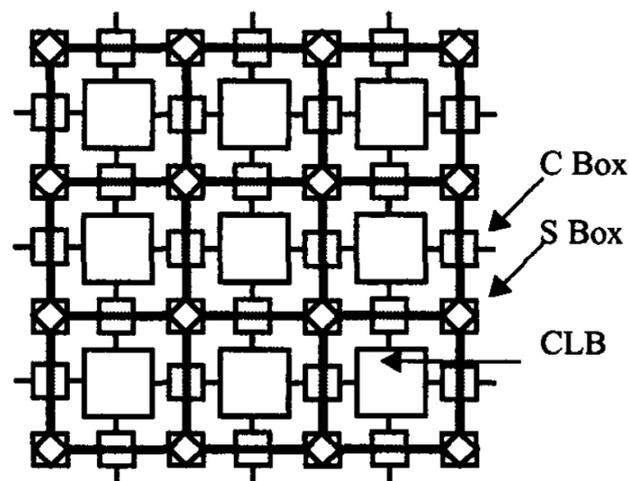


Fig. 2-9. Switch blocks (C Box and S Box) form the configurable connections between CLBs [38].

### 2.1.3 Non-volatile FPGAs

One of the core next-generation approaches to FPGAs focuses on the addition of nonvolatility to the volatile fundamental components in the architecture. As previously discussed, SRAMs and Flash are the dominant technologies in current market FPGAs but face increasing limitations from the physical scaling, energy consumption, leakages, and operational performances. NV is implemented inside these components through the use of emerging next-generation memory devices which typically comes with intrinsic NV. More detailed discussions on these devices are presented in the following subsections.

PCM-based nvFPGAs were proposed in [16], [39], and [40] achieving improvements in cell area (1.7x smaller than FLASH), WRITE time (33.3x smaller than Flash), and programming energy consumption (4.1x smaller than FLASH). However, the drawbacks of the Joule Heating mechanism in PCM leads to  $1.2 \times 10^4$  higher programming energy consumption compared to SRAMs.

A proposed solution using ReRAMs in [41] introduces a generic memristive structure (GMS) wherein two serially-connected ReRAMs form a single memory cell to replace SRAM and Flash cells in the FPGA LUT. The GMS structure shows x3 improvement in cell area, x16.6 improvement in WRITE time, x8.3 programming energy consumption compared to FLASH.

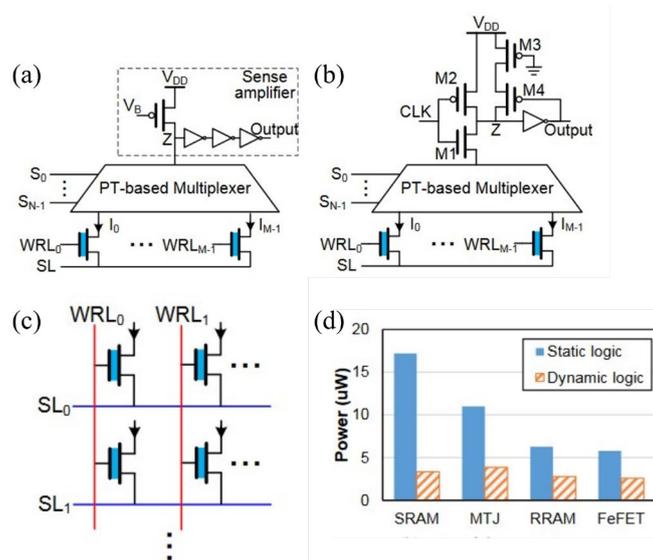


Fig. 2-10. (a) and (b) FeFET-based nvLUT with different logic configurations, (c) the LUT array, and (d) the 6-input power comparison for LUTs [20].

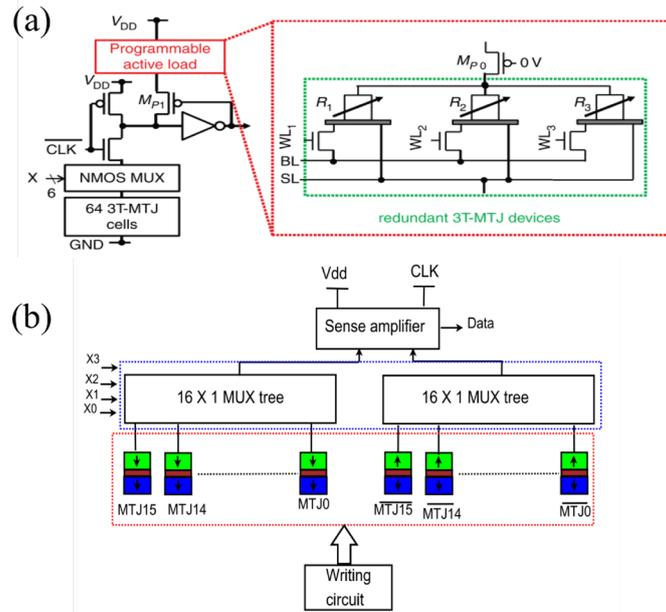


Fig. 2-11. Magnetic tunnel junction (MTJ)/CMOS-based LUTs in (a) [44] and (b) [45].

A study performed in [42] looked into the performance of 2 transmission gates 1 ReRAM (2TG1R)-based and 4 transistor 1 ReRAM (4T1R) as replacements for standard access transistors in SRAM-based FPGAs. Due to the additional components per cell (2TG and 4T) the cell areas were on average 1.4x and 1.03x larger than SRAM cells. WRITE delays were also higher by 1.6x and 0.99x. Further improvements on ReRAM-based architecture can be obtained by using faster switching and lower power consuming devices such as in [43].

Examples of NV implementation in the LUT array in literature are presented in Fig. 2-10— Fig. 2-13 [20], [44], [45], [39], [46]. In [20] (Fig. 2-10), the LUT is made up of FeFETs as the

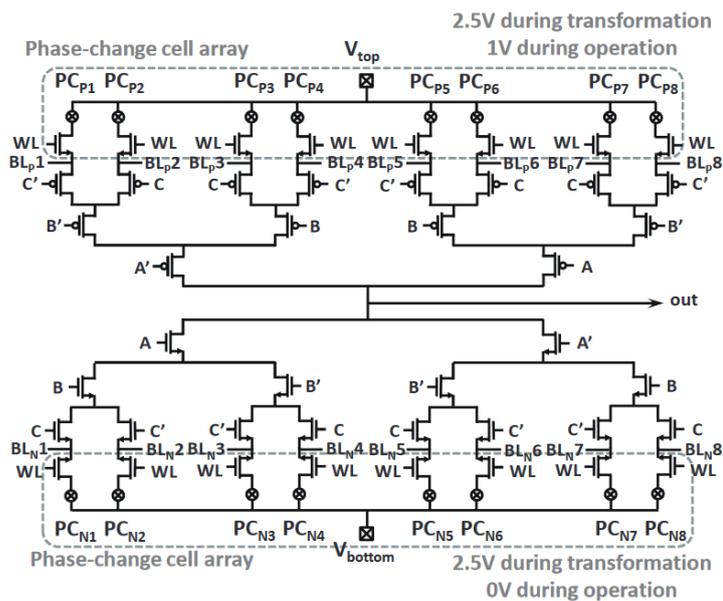


Fig. 2-12. Phase change memory (PCM)-based LUT from [39].

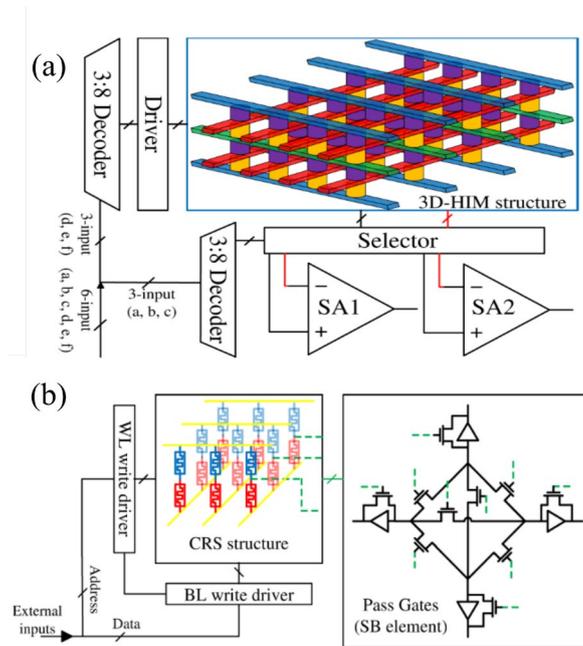


Fig. 2-13. CRS ReRAM-based six-input LUT [46].

memory cells. Hybrid STT-MRAMs or magnetic tunnel junction (MTJ) and CMOS designs are presented in [44] and [45] (Fig. 2-11). PCM memory cells are utilized in [39] (Fig. 2-12) while complementary resistive switching (CRS) ReRAM cells are used in [46] (Fig. 2-13). These designs demonstrate the feasibility of implementing NV in the LUT arrays through next-generational memory devices. However, these arrays are SB arrays and require combinations of CMOS circuits or peripheral circuits. In the case of the CRS ReRAM LUT, the CRS cell uses two ReRAMs per cell thereby increasing the array component count.

Improvements on the SB arrays have also been presented in literature. Since MB is an intrinsic property of ReRAMs, MB ReRAM LUTs were proposed in [47], [48], and [49]. However the presented LUT designs utilize a simplistic crossbar array which leaves the LUT array vulnerable to sneak path current effects [50]. These sneak path effects are amplified in MB cells due to the smaller resistance window between resistance levels.

#### 2.1.4 The Volatile and Non-Volatile D Latch

The D latch is a basic digital memory element that takes in single bit input data (D) when the control signal is high. This control signal can be supplied by clock pulses (CLK) to control the D latch in electronic circuits. An example of a basic CLK-controlled D latch design is given in Fig. 2-14. The drawback of the conventional D latch is that they are comprised of SRAMs and are volatile memories, storing data temporarily as long as power is supplied ( $V_{dd}$  is high).

As they form the basic memory elements of digital circuits, the requirement of power supply and their inability to be fully switched off during SLEEP leads to high leakage power [51]. To improve the D latch design and reduce power consumption, there have been numerous attempts to introduce NV into the D latch architecture in literature. For example, in [52] NV is achieved by using a CRS configuration consisting of 2 anti-serially connected ReRAMs ( $MEM_1$  and

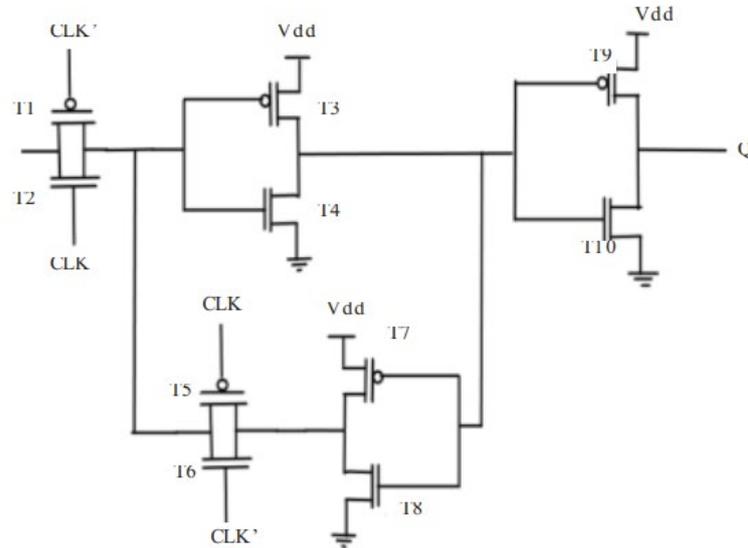


Fig. 2-14. A traditional D latch layout.

$MEM_2$ ) (Fig. 2-15).  $MEM_1$  and  $MEM_2$  are initially in HRS and LRS respectively. When  $In$  voltage increases above  $MEM_1$ 's switching voltage threshold,  $MEM_1$  becomes LRS leading to both  $MEM_1$  and  $MEM_2$  being in LRS states. Continuing voltage application then switches  $MEM_2$  into HRS. The nvLatch is achieved by this design as data is stored in the CRS cells however, when the input voltage changes, this design requires the ReRAM pairs to switch; this takes roughly double the time required for a single ReRAM switching scheme.

A single ReRAM nvD-latch was proposed in [53] using a ReRAM, ME and Resistor,  $R_L$  pair as a voltage divider (Fig. 2-16). CP is the CLK pulse and the D input is directly connected to the Q output through two inverters, INV2 and INV3. When CP is high, transistors M1 and M2 are on while P1 is off; when D is logic '1' (logic '0'), Q follows and is logic '1' (logic '0'). At the same time, the output of Inv1 is logic '0' (logic '1') and the input at Inv2 is logic '1' (logic '0'), leading to a voltage potential across ME and switching the ReRAM into  $R_{ON}$  ( $R_{OFF}$ ). During CP low,  $M_1$  and  $M_2$  are off and P1 is on, separating the D input from the Q output and enabling a current path for  $V_{SS}$ . Depending on the resistance state of ME set during CP high, the input of Inv2 is logic '1' ( $V > V_{threshold, Inv2}$ ) if  $R_{ME}$  is  $R_{ON}$  or logic '0' ( $V < V_{threshold, Inv2}$ ) if

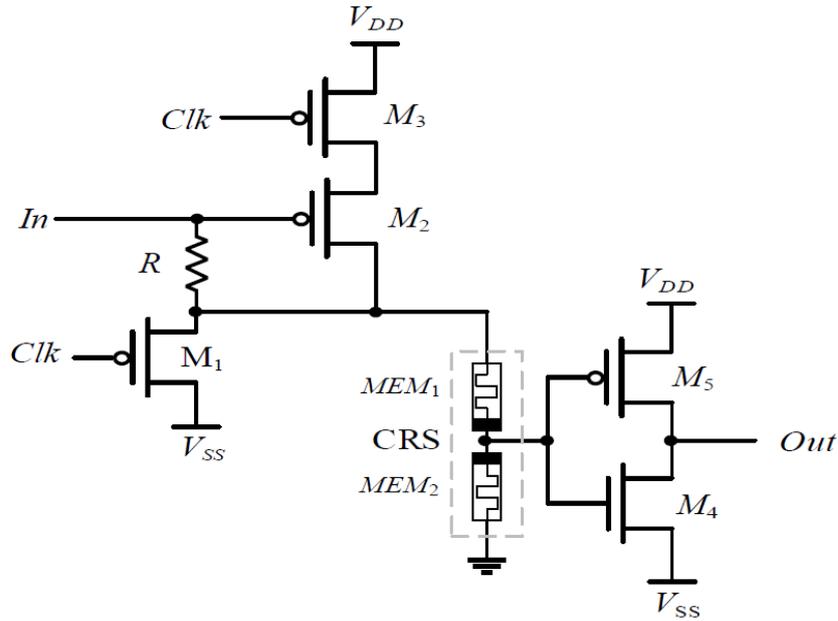


Fig. 2-15. nvD latch with CRS ReRAMs [52].

$R_{ME}$  is  $R_{OFF}$ . However, when CP is high the threshold loss voltages across M1 and M2 reduces the voltage across ME which increases the switching time. Additionally, during CP low the  $V_{SS}$  is fed across ME to provide the Q output and the design relies entirely on ME's resistance; continuous voltage application across the ReRAM causes a resistance state drift leading to variances at Inv2's input.

A solution to the switching time problem was presented in [54], by replacing the transistors

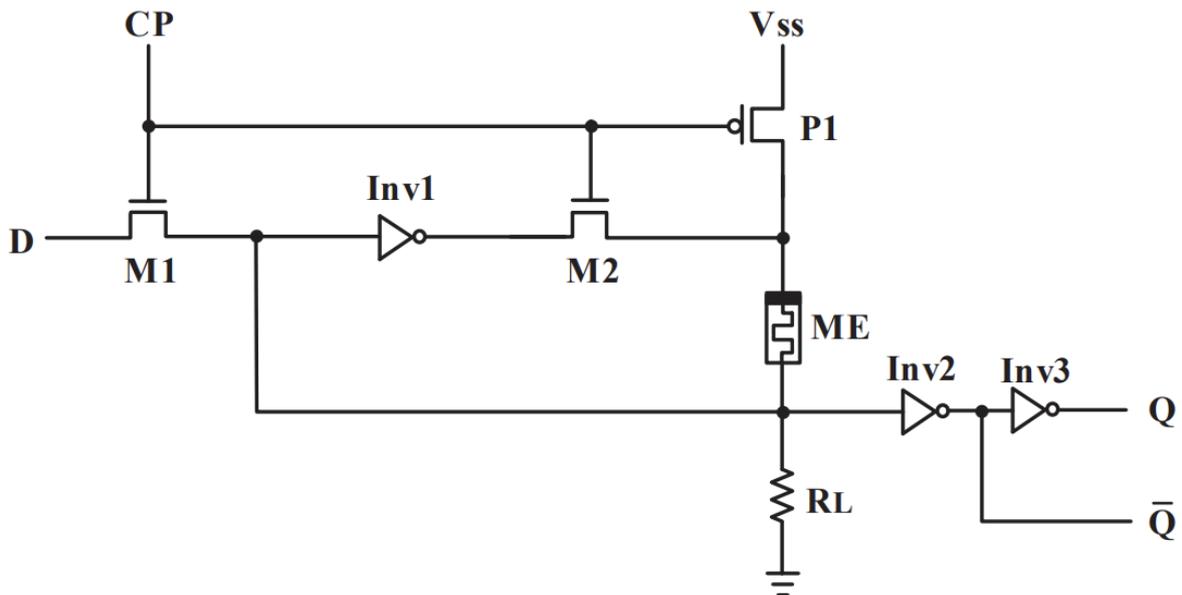


Fig. 2-16. nvD latch design with single ReRAM cell [53].

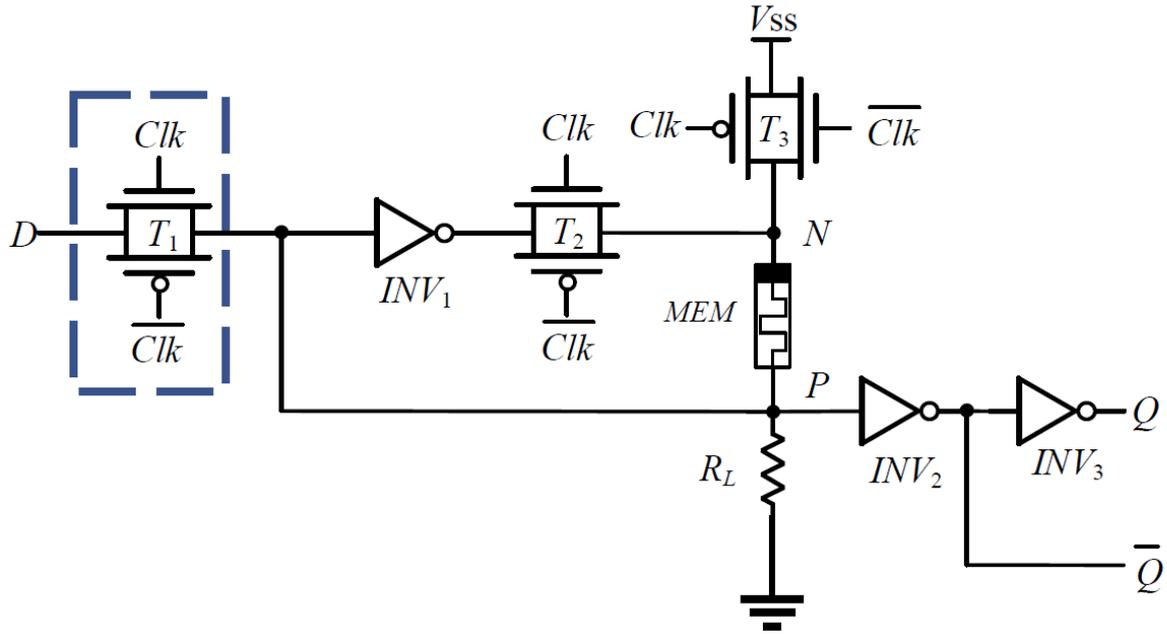


Fig. 2-17. nvD latch design with T-gates single ReRAM cell [54].

with transmission gates (T-gates) (Fig. 2-17). The implementation of T-gates ensures a proper logic '0' and logic '1' through the use of complementary NMOS and PMOS transistors and eliminates the threshold voltage drop problem. The design is however similar to Fig. 2-16 and still relies fully on the resistance of ME during CP low. Examples of STT-MRAM latches are given in Fig. 2-18 [55]. These layouts achieve NV with the addition of additional circuitry

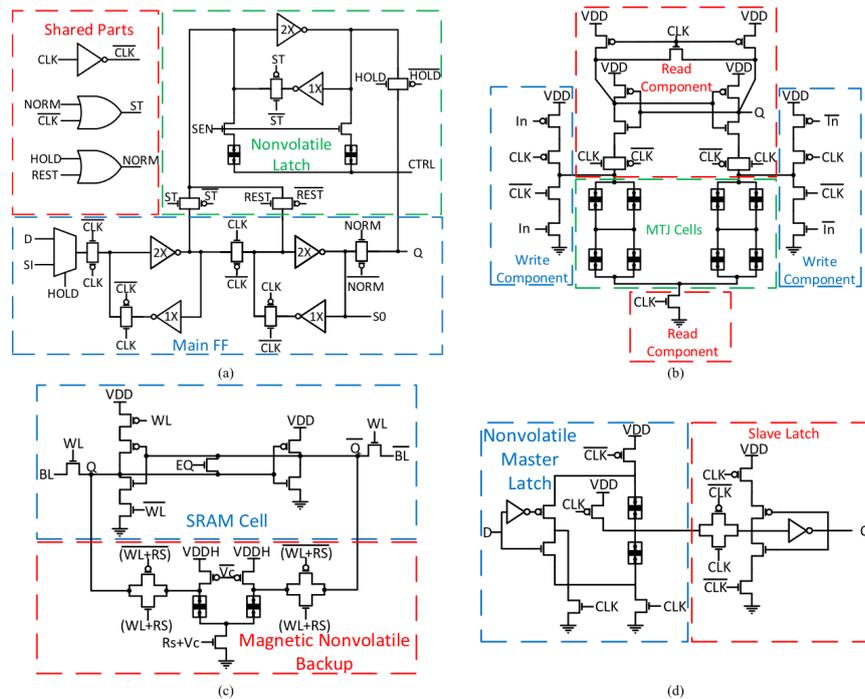


Fig. 2-18. Examples of STT-MRAM latches [55].

storage components, requiring additional STORE and RESTORE sequences. Combined with the increased cell footprint requirements of the technology, the STT-MRAM-based nvD latch incurs a significant cost increase in cell area.

### 2.1.5 The Volatile and Non-Volatile Flip-Flop

The D Flip-Flop or delay Flip-Flop is a Flip-Flop circuit that retains memory. They make up registers in the Processor where they receive logic calculation outputs from the Processor which are stored during a clock cycle and then released back to the Processor for further calculations in the next clock cycle. D Flip-Flops are however conventionally made up of SRAMs and are volatile so they do not retain the data if the supply power,  $V_{DD}$  is removed.

The benefits of zero leakage power consumption have led to the development of nvDFF architectures. Fig. 2-19 shows an example of a nvDFF from [56]. Two ReRAMs are gated by two pass transistors that are controlled by SWL signals. NV data storage occurs when the CLK signal is high, achieving passive NV. The RESTORE sequence is however non-passive as a READ signal is required. The auxiliary signals are controlled by an OR gate leading to the addition of 6 transistors together with an additional CTRL signal line.

A similar approach is adopted in [57] by using an external signal to control the pass transistors on the ReRAM branch. This design removes the OR gate but introduces non-passive STORE and RESTORE sequences that increase the READ and WRITE time to the NV components. The NVFF in [30] only utilizes one ReRAM cell in the NV segment but introduces complex circuitry to implement READ and WRITE onto the NV component. In [31], NV is introduced by adding a NV block that consists of an input tri-state inverter, an output multiplexer, and the ReRAM branch in between which increases the number of components in the circuit as well as the footprint of the NVFF. Other approaches using ferroelectric transistors (FEFETs) which

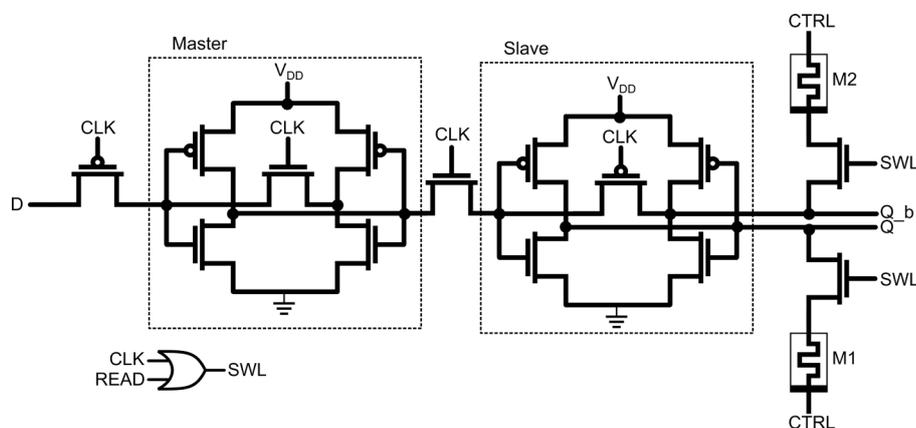


Fig. 2-19. Nonvolatile D flip-flop using control CTRL and SWL signals from [56].

have NV properties also require NV blocks with additional components such as in [32].

### **2.1.6 The Non-Volatile DRAM**

The nvDRAM cell has been previously proposed in [61], [62], [63], [64]. In [61], a novel metal-ferroelectric-metal (MFM) capacitor was proposed but ferroelectric materials experienced severe limitations in scaling. This was then improved on in [62] by using anti-ferroelectric materials. In [63] the capacitor is replaced by NV PCM. The performance metrics of the proposed DRAM storage components however lag behind ReRAMs [63], [65].

A ReRAM-based nvDRAM was proposed in [64]. The design utilizes a control access transistor for READ and WRITE operations into ReRAM. As a result, the cell structure contains four transistors, one gated diode and one ReRAM (4T1D1R) compared to the 1T1C volatile DRAM. The addition of the control access transistor also requires extra STORE/RESTORE operations in addition to the conventional READ and WRITE.

## **2.2 Emerging Non-Volatile Memory (NVM) Devices**

As MOS transistors reach their scaling limits and Moore's Law appears to be slowing down, research has been done into new memory technologies. Devices operating on the concept of RS which was first brought up in the 1970s and 1980s [66], [67], [68] have been widely explored.

Unlike electrical charge-based MOS transistors that are susceptible to charge leakage effects that become more pronounced with device downscaling, RS devices operate on the manipulatable resistances of thin-layer materials. Switching between two different resistance values in a device represents binary 0 and 1 and the ability to retain those resistance values lead to digital memory devices. Emerging memory devices will have to contend with strict digital memory criteria in order to compete successfully with very-matured conventional transistor technology in the industry. Amongst the requirements for emerging memories are low operating voltages ( $<1$  V), long cycling endurance ( $>10^{17}$  cycles), large data retention duration ( $>10$  years), low energy consumption ( $\sim$ fJ/bit), and scalability ( $<10$  nm) [69].

This section will delve into promising emerging technologies; exploring their fundamental operating mechanisms and properties.

### 2.2.1 Suitability of NVM Device

Fig. 2-20 shows a taxonomy tree of memory technologies. From the previous discussed emerging NVM devices, each of them has ‘pros’ and ‘cons’ when compared to conventional MOS devices. Through a comparison of the operational properties of the devices listed in TABLE 2-1 [70], the ReRAM was chosen for this PhD research as it has shown advantages compared to the other emerging NVM devices.

The ReRAM operates on localized CF in the TMO layer and is not affected by cell width. It

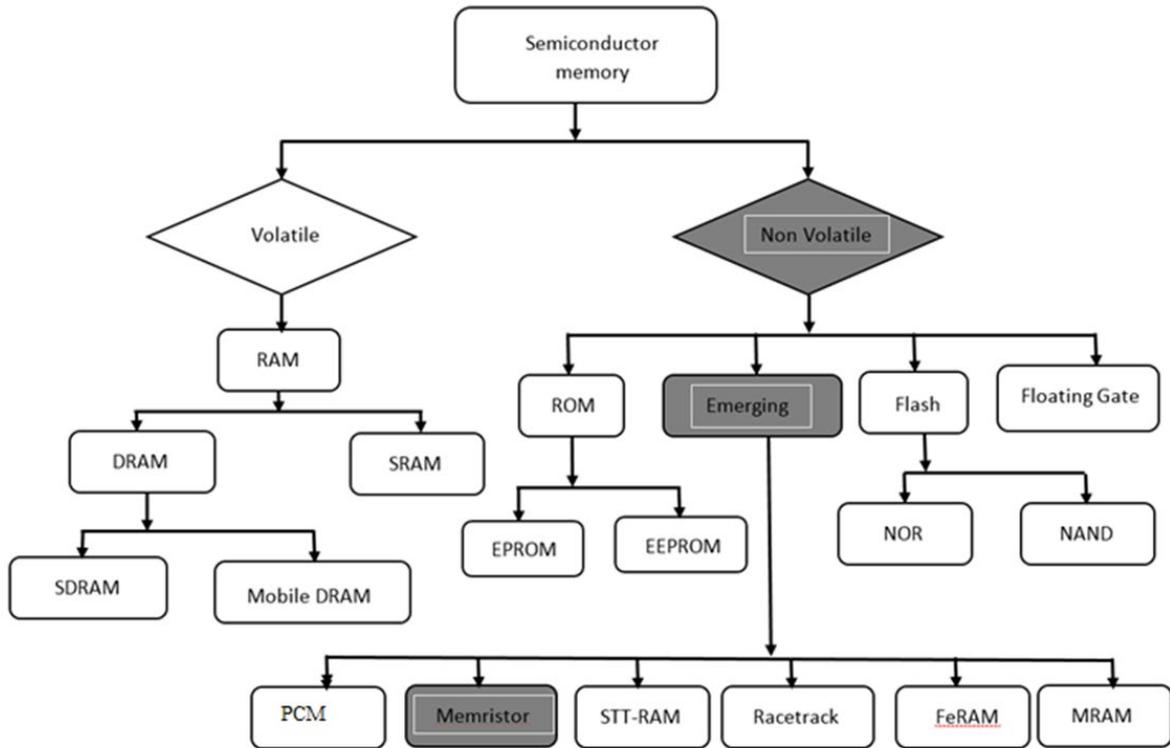


Fig. 2-20. Memory taxonomy tree.

TABLE 2-1. Device characteristics of mainstream and emerging memory technologies.

	MAINSTREAM MEMORIES				EMERGING MEMORIES		
	SRAM	DRAM	FLASH		STT-MRAM	PCRAM	ReRAM
Cell Area	>100 F <sup>2</sup>	6 F <sup>2</sup>	10 F <sup>2</sup>	<4 F <sup>2</sup> (3D)	6~50 F <sup>2</sup>	4~30 F <sup>2</sup>	4~12 F <sup>2</sup>
Multibit	1	1	2	3	1	2	2
Voltage	<1 V	<1 V	>10 V	>10 V	<1.5 V	<3 V	<3 V
READ time	~1 ns	~10 ns	~50 ns	~10 μs	<10 ns	<10 ns	<10 ns
WRITE time	~1 ns	~10 ns	10 μs-1 ms	100 μs-1 ms	<10 ns	~50 ns	<10 ns
Retention	N/A	~64 ms	>10 y	>10 y	>10 y	>10 y	>10 y
Endurance	>1E16	>1E16	>1E5	>1E4	>1E15	>1E9	>1E6~1E12
WRITE energy (J/bit)	~fJ	~10fJ	~100pJ	~10fJ	~0.1pJ	~10pJ	~0.1pJ

therefore, has much better potential for size downscaling ( $<4 F^2$ ) compared to PCM ( $\sim 4\text{--}20 F^2$ ) and spin-transfer torque magnetic random-access memory (STT-MRAM) ( $\sim 6\text{--}20 F^2$ ). PCM devices are limited by the requirement of the heating layer while STT-MRAM has three layers (free layer, tunnel barrier, fixed layer) between electrodes. All technologies however show improvement over conventional SRAM ( $>100 F^2$ ).

Compared to PCM, the ReRAM also demonstrates lower WRITE energy consumption (10 pJ/bit vs  $\sim 0.1$  pJ/bit) and higher cycling endurance ( $10^9$  vs  $10^{12}$ ). This is because the Joule heating-based RS of PCMs are energy intensive and degrade the endurance performance. ReRAMs also require less manufacturing processes to fabricate and are highly compatible with conventional CMOS fabrication [71].

### 2.2.2 Resistive Random-Access Memory

ReRAMs are nonvolatile memory devices that operate behind the RS mechanisms of transition metal-oxides (TMOs). ReRAMs are typically three-layered devices consisting of a TMO layer sandwiched between two metal electrodes. An example of a ReRAM cross section is given in Fig. 2-21 [72].

RS is achieved in the device by application of voltage across the electrode terminals. In this regard, there are two kinds of ReRAMs; bipolar ReRAMs switch between high and low resistances through application of opposite polarity supply voltages and unipolar ReRAMs switch between resistances through application of different amplitude but similar polarity voltages. Unipolar ReRAMs typically require higher voltage amplitudes and therefore have higher power consumption [73].

Depending on the TMO used, two differing physical mechanisms influence the RS properties of ReRAMs which are:

- 1) Forming/Destruction of conductive filaments (CFs)
- 2) Tunneling barrier modification

It has been shown that filamentary switching occurs with the existence of free moving metal ions or oxygen vacancies ( $V_{Os}$ ) in TMOs [74]. Application of voltage to the very-thin TMO layer creates a huge electrical field across the layer which causes a dielectrical breakdown of the TMO crystalline structure. In, the dielectric breakdown leads to drift-diffusion of copper ions that aggregate to form a CF in the TMO layer. A similar process occurs in Tantalum-oxide ReRAMs [75] but the CF is made up of  $V_{Os}$  instead of metal ions. TABLE 2-2 [76] shows a



Fig. 2-21. Three-layer ReRAM. HfO<sub>2</sub> is the TMO layer, Ti and TiN are the electrodes [72].

TABLE 2-2. Comparison between metal-oxide ReRAM and conductive bridge ReRAM [76]

Parameter	Speed (ns)	Operation voltage (V)	Operation current ( $\mu$ A)	Endurance (cycles)	On/Off Ratio	Retention@85°C (s)	Multilevel Capacity	CMOS compatible	Fabrication	Scalability
Metal-Oxide ReRAM	5	~3	5	$10^{12}$	$10^7$	$10^6$	Yes	Yes	Easy	Good
Conductive Bridge ReRAM	1	~7	10	$10^6$	$10^7$	$10^6$	Yes	Yes	Easy	Good

comparison of operational properties between metal ion-based ReRAMs and  $V_O$ -based ReRAMs. CBRAMs show good operation speeds due to the easier drift-diffusion of metal ions compared to  $V_O$ s but OxRAMs show better endurance and have lower operating voltages and currents.

The TMO layer is initially high resistance in its natural state and the localized formation of the CF that shunts the two metal electrodes leads to a drop in the ReRAM resistance. This formation process is called SET and switches the ReRAM into a low resistance state (LRS). Alternatively, the rupture of the CF removes this conduction path between the electrodes and switches the ReRAM into a HRS in a process called RESET.

### 2.2.3 Multibit ReRAM

As RS in ReRAMs operate behind the formation/rupture of CFs, it is possible to halt the process before the CF is fully formed/ruptured. This incomplete process leaves the ReRAM in an intermediate resistive state (IRS) and opens up the possibility for multiple-bits-per-cell

ReRAMs, also known as MB-ReRAMs. This increase in the number of bits per cell would lead to higher density ReRAM arrays and lower costs.

Another benefit is that no additional components are required to achieve MB storage in ReRAMs since the only requirement is to halt the RS process before the CF reaches the fully formed/ruptured thresholds. However, MB-ReRAMs require precise voltage control to ensure a wide margin between every resistance level in the cell to avoid data WRITE errors and a modified READ scheme or sense amplifying circuit which is a general requirement for all MB devices.

### 2.3 ReRAM Modelling

Scaling limitations of conventional non-volatile memory leads to the invention of emerging technologies like PCMs [77], ReRAMs, and STTRAMs. Amongst these devices, RRAMs have demonstrated potential for adoption as next generation solid state non-volatile memory devices because of its simple structure [24], compatibility with CMOS [78], high endurance, retention property and nanoscale dimension [25].

ReRAMs are two-terminal devices with a top and bottom electrode sandwiching a TMO layer in a metal-insulator-metal (MIM) structure. The TMO layer is capable of switching between stable oxidation states with different resistances, allowing data storage in the form of a high '1' and low '0'. Typically, ReRAMs require a forming step to form a conduction path in the TMO layer which shunts the two electrodes. The existence of this conductive filament (CF) lowers the resistance of the ReRAM and can be ruptured to raise the resistance. The rupture and formation of the CF is repeatable and the ReRAM is now able to switch between a LRS and HRS.

Further studies on engineering the material and structure of ReRAMs to finetune its performance have been performed [79], [80] and ReRAMs can now be further categorized by its structure and the RS characteristics. ReRAMs can be divided into two categories, oxide-based ReRAMs (OxRAM) [81] and conductive bridge ReRAMs (CBRAM) [82], [83] depending on the mechanism behind their RS behaviour. RS in OxRAMs stems from the drift-diffusion of oxygen vacancies ( $V_O$ ) while the migration of metal ions contributes to RS in CBRAMs [84], [85]. As the general RS mechanisms and the models for both are different, this section will focus on the modelling of OxRAMs. Any mention of ReRAMs will refer to OxRAMs in this section from here onwards.

To facilitate studies on the ReRAM behaviour in electronic circuits, different types of models have been developed namely electrical, analytical, mathematical, physics-based models and more general compact models [86]. Among these models, physics-based models explain device behaviour with physical concepts, which is more acceptable to define the working of the device [27], [87], [88], [89].

### 2.3.1 SET and RESET Switching Behaviour

The switching behaviour of the devices during state transitions from LRS to HRS (RESET) and from HRS to LRS (SET) can be either abrupt or gradual depending on the electric field strength, internal temperature, thermal conductivity, hopping distance, and various other properties of the ReRAM material. A high electric field across the filament gap length ( $g$ ) defines the abrupt SET and a low field across  $g$  causes gradual RESET in [79] and [27] whereas temperature influences the rate of  $V_O$  formation/recombination to control the changes in CF radii and gap length in [90]. Similarly, thermal feedback mechanisms affect SET and RESET switching in [91] and [92]. Positive feedback loops are generated from the self-acceleration of ions to cause a sharp SET process by enhancing the electric field and temperature of the filament tip while negative feedback is generated from the accelerated ionic migration and subsequent creation of the filamentary gap to form the basis of gradual RESET. The physical mechanism reported in [93] is the drift-diffusion of  $V_O$ s in the filament with changes in field and temperature. In this case, gradual SET switching is a result of the electric field and temperature driven drift-diffusion processes acting in same direction whereas the opposing drift and diffusion fluxes produce a cancelling effect to account for gradual RESET. The model presented in [88] reports the abrupt SET and RESET switching governed by the change in the volume of  $V_O$  doped region and its corresponding variation with Schottky barrier.

In addition to the electric field and temperature, the material properties such as hopping distance and thermal conductance are vital in predicting the abrupt and gradual changes in the switching behaviour. A smaller hopping distance yields analog switching characteristics while a larger hopping distance causes abrupt transition during switching. ReRAMs with high thermal conductivity electrodes tend to exhibit gradual switching behaviour by drawing more heat and thereby decreasing the temperature to impact the drift and diffusion.

### 2.3.2 Onset of the Origin of Gap Opening in the Conductive Filament

The modelling of ReRAM devices by adopting a physical approach provides an insight on the rupture of CF during RS. The prominent factors which affect the site of CF rupture are; the location of oxygen/metallic ion reservoir [94], temperature profile along the CF [91] and thermal properties of oxide and electrode.

The region near the oxygen/metallic ion reservoir layer provides favourable conditions for the growth and destruction of filaments as there are better chance of  $V_O$  recombination and release or migration of ions. For instance, filament rupture starts near top electrode in [95] and at interface between the switching and bulk layer in bi-layered structures [94], [96].

In temperature driven models, the location of the depletion gap starts at the point of highest temperature in the CF. The CF initially ruptures at middle of the CF is reported in [91] and at top electrode interface is given in [27] and [93]. The onset of gap can be determined by selecting the suitable materials for electrode and oxides. Since the metal electrodes are great heat conductors, they act as heat sinks. The thermal conductivities of both top and bottom electrodes play a role in the CF rupture site formation in single layered devices whereas the thermal properties of the electrode next to switching layer determines the location of gap formation in bi-layered devices. Therefore, to determine the change in temperature during RS, the heat flow in CF is calculated by taking the ratio of thermal conductivity of electrode and oxide [93], [97]. Moreover, in a few models, the Schottky contact formed at metal-semiconductor junction facilitates as an interface for the onset of CF rupture [79], [98], [99].

From the analyses above, it can be understood that CF formation or rupture is modelled based on two major physical mechanisms namely: drift-diffusion and reduction-oxidation (redox). Both mechanisms are driven by electric field or temperature or a combination of both. The following sections analyse the physics behind the two approaches in detail as well as the physical concepts adopted to model the current conduction.

### 2.3.3 RS Mechanisms in Analytical Models

#### 1. Drift-diffusion Model:

In this approach, RS is modelled through defect migration induced by an electric field and an increase in the localized temperature during SET and RESET. The application of an electric field drives the migration of ions with migration direction dependent on the ionic charge and field strength direction. The ionic migration is a

result of the interaction between two fluxes; Fick's diffusion and ionic drift. Fick's diffusion is dependent on the concentration gradient while electric field induced barrier lowering influences ionic drift. These fluxes contribute to the change in  $V_O$  concentration in the CF which is given by [92]:

$$\frac{\partial n_D}{\partial t} = \nabla \cdot (D\nabla n_D - \mu F n_D) \quad (2.1)$$

where  $D\nabla n_D$  and  $\mu F n_D$  are the Fick's diffusion and ionic drift terms, respectively.  $n_D$  is the doping concentration,  $D$  is ion diffusivity,  $\mu$  is the ion mobility, and  $F = -\nabla\psi$  where  $\psi$  is the local electrostatic potential demonstrating the relationship between drift and electric field.  $D$  and  $\mu$  are based on Mott-Gurney ion hopping law. The equation above can be solved with the carrier continuity equation (2.2) for electronic conduction to incorporate the temperature effect [92]:

$$\nabla \cdot \sigma \nabla \psi = 0 \quad (2.2)$$

where  $\sigma$  is the electrical conductivity, and with the steady-state Fourier equation [92]:

$$-\nabla \cdot k_{th} \nabla T = \sigma |\nabla \psi|^2 \quad (2.3)$$

where  $k_{th}$  is the thermal conductivity and  $T$  is the local temperature. In [93] a fitting parameter  $\gamma$  is added to the rightmost term to become  $\gamma \cdot \sigma |\nabla \psi|^2$  where  $\gamma = 1$  and  $\gamma = 2$  are used for DC and AC programming conditions respectively, for the simulation. Solving these expressions provide the change of  $n_D$ ,  $\psi$ , and  $T$  during RS.

Additional physical terms included in (1) are incorporated in [93] and [75] to fine tune the model. In [75], the migration or generation of  $V_O$  through thermally activated hopping during SET is considered termed  $G$ :

$$\frac{\partial n_D}{\partial t} = \nabla \cdot (D\nabla n_D - \mu F n_D) + G \quad (2.4)$$

As a result, the growth of the CF is considered proportional to the rate of ion migration and the generation term,  $G$  is given by [75]:

$$G = A \exp\left(-\frac{E_b - q\beta E}{KT}\right) \quad (2.5)$$

where  $A$  is the pre-exponential constant,  $E_b$  is the energy barrier for ion hopping and  $K$  is the Boltzmann constant. The  $q\beta E$  term is the barrier lowering effect of the applied electric field,  $E$  where  $\beta$  is the simulation's mesh size. The addition of the

TABLE 2-3. Activation energy for RS models

Parameter	Equation	Activation Energy $E_A$ in eV
E1	$\nabla \cdot (D\nabla n_D - \mu F n_D)$	$1^*/0^\#$ [92]
E2	$\nabla \cdot (D\nabla n_D - \mu F n_D) + G$	$0.052^*/0.016^\#$ [75]
E3	$\nabla \cdot (D\nabla n_D - \mu F n_D) + DS n_D \nabla T$	$-0.006^*/0.05^\#$ [93]
Ea1	$\exp\left(\frac{-(E_{a1} - Zq\alpha E)}{KT}\right)$	1.2 [91], [95] 1.12 <sup>a</sup> /1.35 <sup>b</sup> [90] 1.4 <sup>c</sup> /0.30 <sup>d</sup> [94]
Ea2	$\exp\left(\frac{-E_{a2}}{KT}\right)$	1.2 [91]
Ea3	$\exp\left(\frac{E_{a3} + q(1 - \alpha)V_{cell}}{KT}\right)$	0.70 [101]
Ea4	$\exp\left(\frac{-E_{a4}}{KT}\right) \sinh\left(\frac{\gamma q \alpha E}{KT}\right)$	1 [79] [95]

\*Denotes  $E_A$  at minimum  $n_D$ ; # denotes  $E_A$  at maximum  $n_D$ ; a, b denotes values for SET and RESET respectively; c, d denotes  $E_A$  for Ta<sub>2</sub>O<sub>5</sub> and TaO<sub>2</sub> respectively.

$G$  term is shown in (E2) of TABLE 2-3. The  $G$  term in (2.5) describes  $V_O$  generation during SET and is not applied during the RESET phase of the model.

A temperature influenced Soret's effect or thermophoresis term is added to (2.1) to account for the contribution from Joule heating in [93]. Thermophoresis which describes migration of vacancies along the temperature gradient is given as [93]:

$$\frac{\partial n_D}{\partial t}_{Soret} = DS n_D \nabla T \quad (2.6)$$

where  $S$  is the Soret coefficient. This term describes the tendency for  $V_O$  to move toward a hotter region [93]. The revised form of (2.1) is given as (E3) in TABLE 2-3.

The change in  $n_D$  during RS affects the electrical and thermal conductivity,  $\sigma$  and  $k_{th}$  as well as the activation energy for conduction,  $E_{AC}$  through the thermally activated electrical conductance Arrhenius equation:

$$\sigma = \sigma_0 e^{-\frac{E_{AC}}{KT}} \quad (2.7)$$

where  $\sigma_0$  is the preexponential factor,  $E_{AC}$  is the activation energy for conduction, and  $k$  is the thermal conductivity. Both  $\sigma_0$  and  $E_{AC}$  are dependent on  $n_D$  as the increase in  $V_O$  increases the device conductance and reduces the activation energy. The change in  $\sigma_0$  and  $E_{AC}$  during RS are plotted in Fig. 2-22.

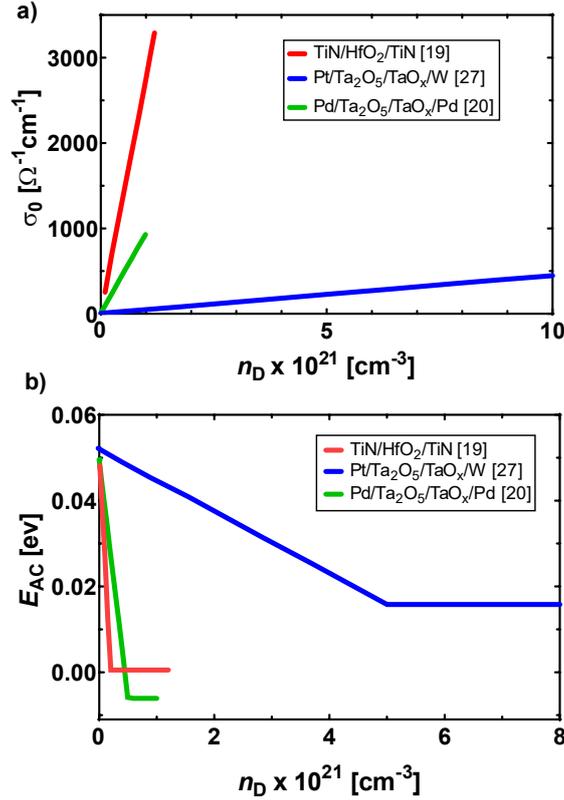


Fig. 2-22. Affect of  $n_D$  on the (a) electrical conductance preexponential factor  $\sigma_0$  and (b) activation energy for conduction  $E_{AC}$ . The devices switch into LRS when  $n_D=0.2 \times 10^{21} \text{cm}^{-3}$ ,  $\sigma_0=700 \Omega^{-1} \text{cm}^{-1}$ , and  $E_{AC}=0.006 \text{eV}$  for TiN/HfO<sub>2</sub>/TiN, when  $n_D=5 \times 10^{21} \text{cm}^{-3}$ ,  $\sigma_0=23$ .

The PDEs from (2.1)-(2.3) should be solved for all layers including the electrode layers to account for the contribution of the entire device to RS. In [92], the electrode layers act as ideal heat sinks with boundary condition  $T = T_0 = 300 \text{K}$ . This is a reasonable assumption as the electrode area is generally large compared to the CF.

Since [92], [93], and [75] modelled a ReRAM with inert electrodes, they do not contribute to ion migration and therefore have constant  $k_{th}$  and no  $V_O$  flux at the Top electrode (TE)/oxide and oxide/ bottom electrode (BE) interfaces. In [92] and [93] bi-layered RRAMs are modelled and the PDE is likewise solved for the bulk layer.

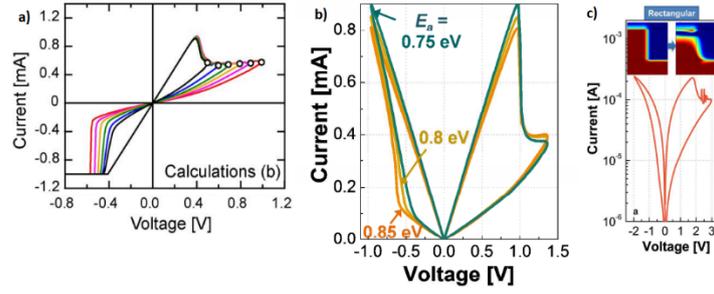


Fig. 2-23. Simulated I-V curves from (a) [92] showing different SET switching voltages as a result of different RESET stop voltages, (b) [93] showing the effect of different EA values, and (c) [75].

As the bulk layer acts as an  $V_O$  reservoir, only Fick's diffusion was considered and the drift term,  $\mu F n_D$  in (2.1) is zero.

The I-V curves generated from this approach in [92], [93], and [75] are shown in Fig. 2-23.

## 2. Reduction-Oxidation (Redox)-based Modelling Approach:

In this approach, redox mechanisms are applied to model defect evolution during CF formation and rupture. The redox mechanism is a result of the generation and recombination (G-R) of  $V_O$  with oxygen ions, oxide phase transition (P-T), as well as the hopping and migration of generated defects. The SET process comprises the P-T from semiconductor to metal phase and the generation of oxygen vacancies [94], [95] while the RESET process involves the reverse P-T from metal to semiconductor phase and the release and hopping of oxygen ions into the switching layer to recombine with  $V_O$  [97], [100]. The variation of the filament gap is given as:

$$\frac{dg}{dt} = a \cdot f \cdot \left( \exp \frac{-E_A}{KT} \right) \quad (2.8)$$

where  $g$  is the gap length,  $\alpha$  is the field enhancement factor, and  $f$  is the attempt-to-escape frequency. When a gap in the filament is present, the activation energy for conduction,  $E_A$  is lowered due to the Poole-Frenkel effect by the electric field in the gap given as:

$$E_A = E_{A0} - \alpha_0 q V_{gap} \quad (2.9)$$

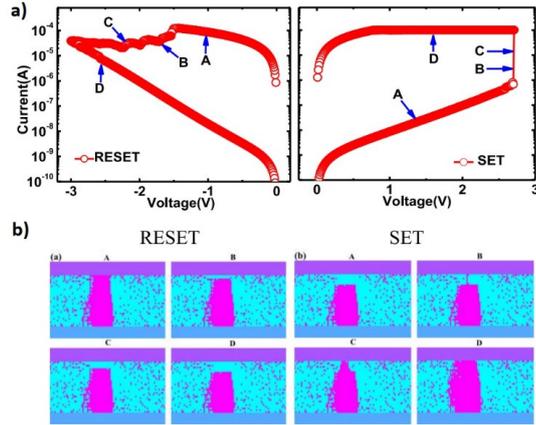


Fig. 2-24. (a) I-V curve of model from [95]. (b) Schematic of CF growth.

where  $\alpha_0$  is the activation energy barrier lowering factor,  $E_{A0}$  is the barrier at zero field, and  $V_{\text{gap}}$  is the voltage across the gap.

In addition to the field induced evolution of the filamentary gap, the redox models in [91], [95], [97] consider the change in filamentary radius during RS. The I-V curves and process schematic from [95] and [97] are shown in Fig. 2-24(a) and Fig. 2-25(a), and Fig. 2-24 (b), and Fig. 2-25(b) respectively.

The radii evolution is calculated as:

$$\frac{dr}{dt} = \left( \Delta r + \frac{\Delta r^2}{2r} \right) f \left( \exp \frac{-(E_A - Zq\alpha_0 E)}{KT} \right) \quad (2.10)$$

where  $r$  is the CF radius and  $Z$  is the charge of ion/vacancy. The I-V curve generated is shown in Fig. 2-25(a).

Another physical mechanism governing RS is the Butler-Volmer redox reaction which considers the role of oxidation and reduction on material property. This reaction is given as [101]:

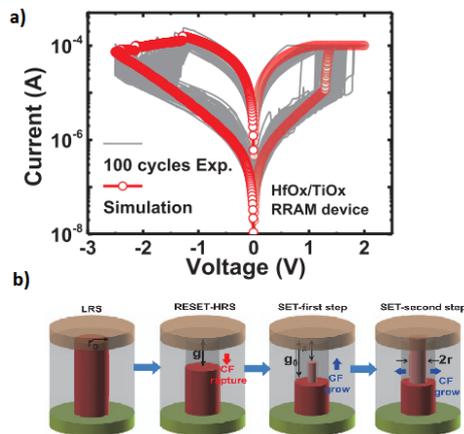


Fig. 2-25. (a) I-V curve of model from [97]. (b) Schematic of CF growth.

$$\frac{dr}{dt} = \frac{r_{max}-r}{\tau_{redox} \left( \exp \frac{E\alpha - q\alpha V_{cell}}{KT} \right)} - \frac{r}{\tau_{redox} \left( \exp \frac{E\alpha + q(1-\alpha)V_{cell}}{KT} \right)} \quad (2.11)$$

where  $r$  and  $r_{max}$  are the changing radius and maximum radius of the CF respectively,  $\tau_{redox}$  is the nominal redox rate, and  $V_{cell}$  is the voltage across the device.

The RS process can be modelled with the same defect kinetics for both SET and RESET by considering the rate of migration of ionized defects from one part of filament to another with the Arrhenius rate expression (2.8) and (2.12) [91], [100]. Gap length variations due to migration can be modelled with (7) and the radial rate of growth can be implemented with (2.10) or (2.11). The movement of oxygen ions and G-R of vacancies can be included to form a single expression:

$$\frac{dg}{dt} = a \cdot f \cdot \left( \exp \frac{-E_a}{KT} \right) \sinh \left( \frac{\gamma q a E}{KT} \right) \quad (2.12)$$

to model SET and RESET [100]. Similarly, the thin film growth rate is described with the concept of oxidation in the Mott-Gurney model in [79] which is utilized to model the CF gap length modulations during RS process.

From the expression,  $\gamma$  can be considered as a constant or variable [79], [100] depending on the models. Since CF geometry determines the state during RS, the filament growth can be modelled three dimensionally by considering the volume of the CF [90] which takes the form (2.13) for SET process while RESET (2.14) is given as:

$$\frac{dV_{TC}}{dt} = -v_0 \left( \exp \frac{-(E_A - Zq\alpha_a V)}{KT} \right) \quad (2.13)$$

$$\frac{dV_{TC}}{dt} = -v_0 \left( \exp \frac{-(E_A - Zq\alpha(g)V)}{KT} \right) \quad (2.14)$$

where,  $\alpha_a$  and  $\alpha(g)$  are the electric field enhancement factors.

### 2.3.4 Required Parameters to Model RS in ReRAMs

1. Activation Energy:

A list of activation energies used in different models that differ according to the switching material are given in TABLE 2-3. It is found that the activation energy required for redox mechanism models are much higher than drift-diffusion models.

## 2. Contribution of Thermal Conductance of Electrode and Oxide Layers in ReRAM:

Thermal conductivity is a material property which captures the change of defect concentration in the gap length [91]. The change in vacancy concentration in CF affects the its thermal conductivity whereby the thermal conductivity at minimum vacancy concentration takes the value of the insulating oxide (e.g. HfO<sub>2</sub>, Ta<sub>2</sub>O<sub>5</sub>) which linearly increases to the conductivity of the metallic CF (Hf, Ta) at maximum vacancy concentration. The enhancement of the thermal conductivity from the increase in vacancy concentration is a result of the free-carrier contribution to heat conduction. To solve (2.3), the change in thermal conductivity is modelled using Wiedemann-Franz law as [92]:

$$k_{HfO/Ta2O5} = k_{HfO_0/Ta2O5_0} (1 + \lambda(T - T_0)) \quad (2.15)$$

where  $k_{HfO/Ta2O5}$  is the thermal conductivity of the oxide at  $T = T_0 = 300K$  and  $\lambda$  is the linear thermal coefficient. The values for  $k_{th}$  differs according to the metal. The linear evolution of  $k_{th}$  is shown in Fig. 2-26.

In addition to thermal conductivity, resistivity of oxide material and CF forming material have an influence on the device characteristics. Resistivity of the oxide layer in the gap region directly affects Joule heating and influences  $V_{SET}$  and thus  $R_{on}/R_{off}$  ratio [102]. With an increase in resistivity of oxide layer, the device can achieve

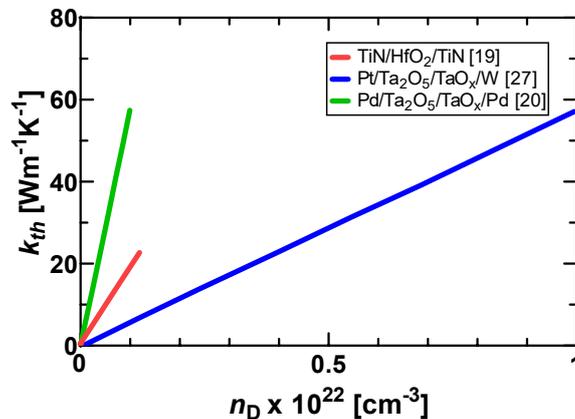


Fig. 2-26. Thermal conductivity  $k_{th}$  as a function of local doping density,  $n_D$ .

increase in  $V_{\text{SET}}$  and decrease in the leakage current. These two electrical parameters significantly enhance device functionality and efficiency. Likewise, resistivity of base layer and electrodes also contribute to RS. Resistivity of base layer has direct dependency on providing self-compliance in bilayer devices [79], [94] during SET process by facilitating a large resistance series to low resistance of CF. In [90] the resistance of electrodes were considered in addition to the resistance of switching and bulk layers (Fig. 2-26).

### 3. Barrier Height Reduction Factor:

The lowering of potential barrier due to the application of electric field is shown to be an important parameter in RS approaches. The term  $\alpha_a ZeE$  denotes general form of barrier reduction term. [91], [95], [97], [100] where  $\alpha_{a(h)}$ , is the field enhancement factor which can be a constant or variable quantity based on the models [91], [95]. A few models formulated variable enhancement factor for barrier lowering [90], [100].  $\alpha_a$  is function of  $g$  in [90], whereas in [100], enhancement factor ( $\gamma$ ) which depends on polarizability of the material [100].  $Z$  is the charge number of oxygen ions/vacancies [95].

### 4. CF Geometry Dependent Physical Parameters:

State of ReRAM devices are derived from RS process which is derived with variations in CF geometry as well as shape. Important physical parameters which facilitate RS are, namely, the gap length ( $g$ ) between top electrode (TE) and CF tip, the width ( $w$ ) or radius ( $r$ ) of CF and the rate of change of  $g$  or  $w$ , ( $dg/dt$  or  $dw/dt$ ). The temporal evolution of  $g$  [75] or the is the common state parameter in 1D models while the change in filamentary  $r$  and/or  $w$  is the state parameter adopted in 2D and 3D models [95], [97], [101]. The volume of the CF is also considered in the 3D model in [90]. The SET process in many models depends only on the gap length [100], while in a few models both radius (2.10) and conduction filament length are taken into account [95], [97]. The RESET process in most models are strongly dependent on  $g$  [91], [95], [97], [100].

The shape of conduction filament is another factor to be considered while RS modelling. RS in the filamentary switching models in [91] is obtained by adopting a cylindrical shaped CF while a conical shaped CF with variable upper and lower radius is considered in [90].

## 5. Temperature:

RS in many models apply Fourier heat flow equations and its developed forms [90], [94], [95] to calculate the effect of temperature during RS. Major parameters required to implement these processes are; thermal conductivity and specific heat per unit volume of the oxide material. In [94], the local temperature for the Pt/Ta<sub>2</sub>O<sub>5</sub>/TaO<sub>x</sub>/Ta ReRAM is calculated from

$$C \frac{\partial T}{\partial t} = \nabla(k \cdot \nabla T) + Q \quad (2.16)$$

where  $C$  represents the specific heat per unit volume of the Ta<sub>2</sub>O<sub>5</sub> layer,  $k$  is the thermal conductivity of Ta<sub>2</sub>O<sub>5</sub> (9.6 W K<sup>-1</sup> m<sup>-1</sup>), and  $Q$  is the Joule heat power density. Fig. 2-27 shows the simulated HRS and LRS transitions where A—H corresponds to the points in I-V curve. The structure layout is from left to right: top electrode, Ta<sub>2</sub>O<sub>5</sub>, TaO<sub>x</sub> (not shown), bottom electrode (not shown). The TaO<sub>2</sub> layer exists as an intermediary layer between the Ta<sub>2</sub>O<sub>5</sub> and TaO<sub>x</sub> interfaces due to the migration of oxygen ions (Fig. 2-28).

The ReRAM is initially in a LRS state at A. Reset begins at B when the supplied voltage is increased and the temperature can be seen to rise along the entire CF. This induces V<sub>O</sub> migration toward the bottom electrode and the filament ruptures at C and the gap region increasing in D. The high temperature is now focused on the gap region as the electric field is concentrated there. The CF gap prevents high temperatures during SET (F) but enough thermal effect is present to induce V<sub>O</sub>

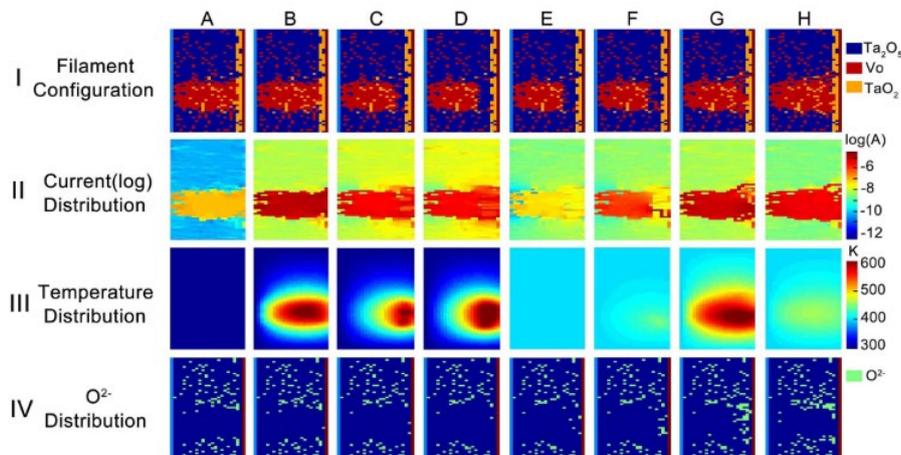


Fig. 2-27. Simulated microscopic filament evolution during RESET-SET with (A)--(H) corresponding to the I-V curve in Fig. 2-24.

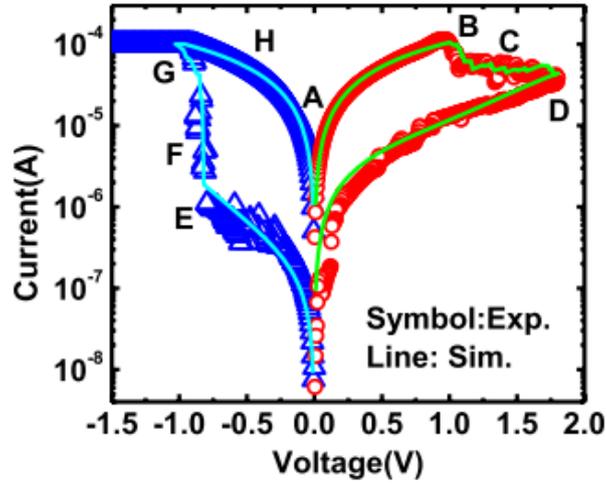


Fig. 2-28. Simulation result from [94].

migration and the reformation of the CF. The temperature profile in G is similar to B from the current flowing through the now connected CF.

In [97], the localized temperature effect induces both vertical and lateral migration of  $V_{OS}$  where the Joule heating effect on the local CF temperature is calculated as

$$T = T_0 + IVR_{th} \quad (2.17)$$

where  $T_0$  is the ambient temperature and  $R_{th}$  is the thermal resistance of the CF. During SET switching, the CF in this model grows vertically to connect the TE with the BE and laterally once the connection is formed (Fig. 2-24(b) and Fig. 2-25(b)).

The effect of temperature on RS is studied in [90] where a model with fixed

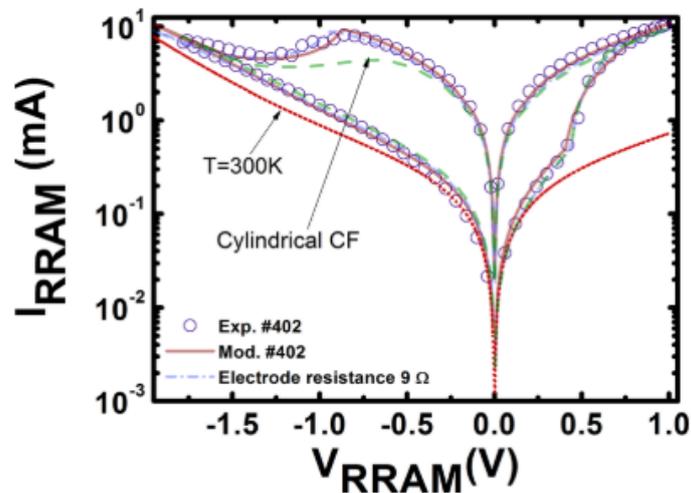


Fig. 2-29. Simulated result from [90]. No RS is observed when the temperature is fixed (red line).

temperature ( $T=300\text{K}$ ) is plotted against a model dependent on the heat equation [103] (Fig. 2-29). No RS occurs with fixed CF temperatures.

A simplified heat equation (2.20) is used to model the filament temperature variation in [79], [97], [100] while a 1D steady state Fourier equation given as

$$k_{th} \frac{d^2 T}{dz^2} + J^2 \rho = 0 \quad (2.18)$$

where  $z$  is the space coordinate along the CF ( $z=0$  at the injecting electrode and  $t_{ox}=20\text{nm}$  is the oxide thickness and total CF length),  $k_{th}$  is the thermal conductivity,  $\rho$  is the resistivity, and  $J$  is the current density is utilized to obtain the temperature variation in [91].

To obtain the temperature along CF, solution of 1D steady state Fourier equation is utilized in [91].

### 2.3.5 Current Conduction Mechanisms Adopted in Physics-based Models

Several electronic conduction mechanisms coexist in analytic models to facilitate current conduction characteristics, namely ohmic/metallic conduction [90], [91], [94], [95], [97], [101], tunnelling [79], [101], Schottky [79], hopping [95], [97], [100]. Among those, two major mechanisms have been analysed are; Schottky and generalized hopping model. The parameters used in these current conduction modelling are discussed in this section.

#### 1. Schottky-based Current Conduction:

Several models followed Schottky current conduction for modelling the interface between metal-semiconductor layers [79], [98], [99]. Schottky current expression takes general form as [87]:

$$I = AA^* T^2 \exp\left(\frac{-\phi_{Bi}(t)}{kT}\right) \left(\exp\left(\frac{V_{sch}(t)}{\eta kT}\right) - 1\right) \exp(-\sqrt{\xi} \delta) \quad (2.19)$$

In the given expression,  $A$  is area of the electrode,  $\eta$  is the ideality factor,  $k$  is the Boltzmann's constant,  $T$  is the temperature,  $V_{sch}$  is the voltage across the Schottky junction,  $A^*$  is the Richardson constant.  $\exp(-\sqrt{\xi} \delta)$  is the tunneling probability

parameter which includes the effect of tunnelling [104],  $\xi$  is the effective barrier in eV and  $\delta$  is the interface layer thickness,  $\phi_{Bi}$  is the effective barrier height which depends on barrier height modulation due to  $V_O$  doping effect ( $\phi_{Bn0}$ ) and barrier lowering due to image force and the electric field ( $\phi_{del}$ ) given as:

$$\phi_{Bi} = \phi_{Bn0} - \phi_{del} \quad (2.20)$$

Different approaches adopted in modelling the current conduction are Richardson-Schottky and Simmon's methods [105], [104], [106]. Richardson-Schottky conduction is state dependent and is well suited to models with the mean free path of electron comparable to the thickness of the oxide film [104].

## 2. Generalized Hopping Conduction:

To correlate the gap length with applied voltage, a generalized hopping current conduction is used in conduction models [90], [95], [97], [100]:

$$I = I_0 \exp\left(\frac{-g}{g_T}\right) \sinh\left(\frac{V}{V_T}\right) \quad (2.21)$$

where  $g_T$  and  $V_T$  are the characteristic length and voltage,  $g$  is the hopping/tunneling length,  $V$  is the applied voltage across the cell.

## 3. Parameters Required to Model Current Conduction Mechanisms:

### 3.1 Barrier Height:

The barrier height is the most critical electrical parameter in the thermionic emission process [107] which is determined by the physical properties of the oxide layer, structure of ReRAM devices, image force lowering effect and electric field. Effective barrier height can be calculated using (2.20) by taking the effects of image force lowering and doping (TABLE 2-4).

TABLE 2-4. Parameters used in Schottky Modelling

Parameter	Model from [88]	Model from [99]	Model from [87]	Model from [27]
Image force lowering effect $\phi_{del}$	$\left[ \frac{q^3 N \psi}{8\pi^2 \epsilon_S \phi_b^3} \right]^{\frac{1}{4}}$ $\psi = \phi_{Bn0} - \phi_n - V_{Sch}$ $\epsilon_S \phi_b = \left[ \sqrt[3]{\epsilon_{ins} * \epsilon_\infty^2} \right]$	$\sqrt{\frac{qE_m}{4\pi\epsilon}}$	$\left[ \frac{q^3 N \phi_{B0}}{8\pi^2 \epsilon_S^3} \right]^{\frac{1}{4}} \left( 1 - \frac{V}{4\phi_{B0}} \right)$	$\left[ \frac{q^3 N \psi}{8\pi^2 \epsilon_S^3} \right]^{\frac{1}{4}}$ $\psi = \phi_{Bn0} - \phi_n \pm V_{Sch}$
Dependence on channel doping effect ( $\phi_{Bn0}$ )	Not considered	$\phi_{Bi} \left( \sqrt{\frac{l(t)}{L_0}} \right)$	$\phi_{Bi} \left( 1 - \sqrt{\frac{w(t)}{D}} \right)$	$\phi_{Bi} \frac{g(t)}{D}$
Ideality factor	Not considered	$\eta=3$	$\eta = 1 + \frac{D_{its}}{D_{itm}}$	$\eta$ $= m \left[ \eta_{LRS} \left( 1 - \frac{g(t)}{D} \right) + \eta_{HRS} \left( \frac{g(t)}{D} \right) \right]$
Electric field	$E = \frac{V_s}{w_{eff}}$	$E = E_{max}$	$E = \frac{1}{\epsilon_s} [n_s W_d + n_d w(t)]$	$E = \frac{V_{undoped}}{g(t)}$

### 3.2 Image Force Lowering Effect with Uniform Electric Field within the Depletion Layer:

In the presence of an electric field, the opposite charges accumulated at the interface of the metal-dielectric have an effect of reducing the energy required for the carriers to overcome the barrier and is termed as image-force-lowering. Different methods are adopted to implement this concept as shown in TABLE 2-4, under the category of image force lowering [27], [88], [98], [99]. These models derive parameters based on different concepts. For example, the permittivity of the semiconductor layer ( $\epsilon_S$ ) is obtained from the product of optical dielectric constant  $\epsilon_\infty$  and insulator dielectric  $\epsilon_{ins}$  in [88].

### 3.3 Temperature:

As a result of the strong current flow and electric field present in ReRAMs, Joule heating is responsible for the temperature variations in the CF. For example, the models in [88], [97] use a simple Joule heating temperature model controlled by electronic current in the filament region given in as [95]:

$$T = V_{el} I_{el} R_{th} + T_0 \quad (2.22)$$

$$T = RI^2R_{th} + T_0 \quad (2.23)$$

However, many Schottky-based models approximate room temperature for current conduction (300K) [79], [98], [99]. The model outputs for [98] and [99] are shown in Fig. 2-28 and Fig. 2-29 respectively. Models based on generalized hopping do not consider temperature variations for current conduction. It can thus be concluded that CF temperature variation does not play a significant role in ReRAM current conduction.

### 3.4 Physical Parameters of CF and Structure of Device:

Physical parameters like channel length, filament thickness, location of Schottky junction and  $V_O$  doping concentration have significant impact on the variation in Schottky barrier height whereby influence electronic current. The length of the doped or un-doped section of the filament is spotted as a variable to control the Schottky barrier height in Schottky conduction models. The application of input voltage enables the oxygen vacancies to move towards the TE to form a CF and this channel doping effect ( $\phi_{Bn0}$ ) is the major factor for reducing the barrier height. To incorporate  $\phi_{Bn0}$ , several approaches are adopted in several models such as simply a constant value has been assigned on the basis of material property in [88] whereas  $\phi_{Bn0}$  has been proportionally related to insulating volume,  $l(t)$  in [99], to doped region length in [98] and to undoped tunnelling length ( $g$ ) in [27], where,  $L_0$  is the initial filament length and  $\phi_{Bi}$  is the initial barrier height,  $w$  is the length of doped region,  $D$  is the switching layer thickness.

Ideality factor ( $\eta$ ) is used in ReRAM current conduction models to assimilate the deviation of junction characteristics from its ideal behaviour. Different approaches are used to assign the ideality factor in models. A constant value is assigned in [99] while two distinct values are assigned to this parameter to show the trap effect in ON state and OFF state of device operation in [98] (TABLE 2-4), where  $D_{its}$  and  $D_{itm}$  are interfacial trap densities in the semiconductor and insulator [98]. Value '1' for  $\eta$  indicates an ideal Schottky junction and '2' represents non-ideal behavior due to interface traps.  $\eta$  is used in [79] to represent the variation in trap charging and discharging throughout the cycles. In the same model, the tunneling probability factor (TPF) is introduced as a critical parameter which incorporates the effect of

tunneling in the model [79]. TPF depends on the thickness of the oxide layer while, other models with Schottky approach ignored this parameter.

Alternatively, current through the device is implemented with hopping conduction in [97], [100] models, with gap length  $g$  and applied input voltage  $V$  as the control variables. However, in filamentary models, the total current in the device is associated with the shape of CF. In [90], a conical shape is approximated to the CF where the current is not only contributed by the top face of CF, but the lateral sides as well.

The position of Schottky contact in the device structure also has contribution in modeling approaches. Schottky junction concept can either applied to the interface between the top electrode and insulator layer [88], [98], [99], [101], [108]. However, Schottky barrier junction is assigned at the interface between the switching and bulk layers in Pt/Ta<sub>2</sub>O<sub>5-x</sub>/TaO<sub>2-x</sub>/Ta device at HRS [96].

The position of Schottky junction in the device structure also has a contribution in modeling approaches. Schottky junction concept can either applied to the interface between the top electrode and insulator layer [88], [98], [99].

### 2.3.6 Comparison of Models with Experimental Results

This section compares the models discussed in the previous sections with experimental data to understand the efficiency of the previously discussed physical mechanisms to obtain the model.

The experimental results reported for Pt/Ta<sub>2</sub>O<sub>5</sub>/TaO<sub>x</sub>/Pt ReRAM with a 4nm switching layer is given in Fig. 2-30(a) and the corresponding model curves are shown in Fig. 2-30 (b), Fig. 2-30(c), and Fig. 2-30(d). In Fig. 2-30(b) and Fig. 2-30(c), the current conduction mechanisms adopted for the models are the Schottky current conduction at HRS and ohmic conduction at LRS. The model in Fig. 2-30(d) further considers a state independent trap assisted tunnelling through the region around the CF along with Schottky at HRS. The Schottky barrier at the Pt and Ta<sub>2</sub>O<sub>5</sub> interface produces the asymmetric current during forward and reverse biases at HRS. The HRS current is a result of both the doping effect of  $V_{OS}$  and tunnelling in [27] while it is

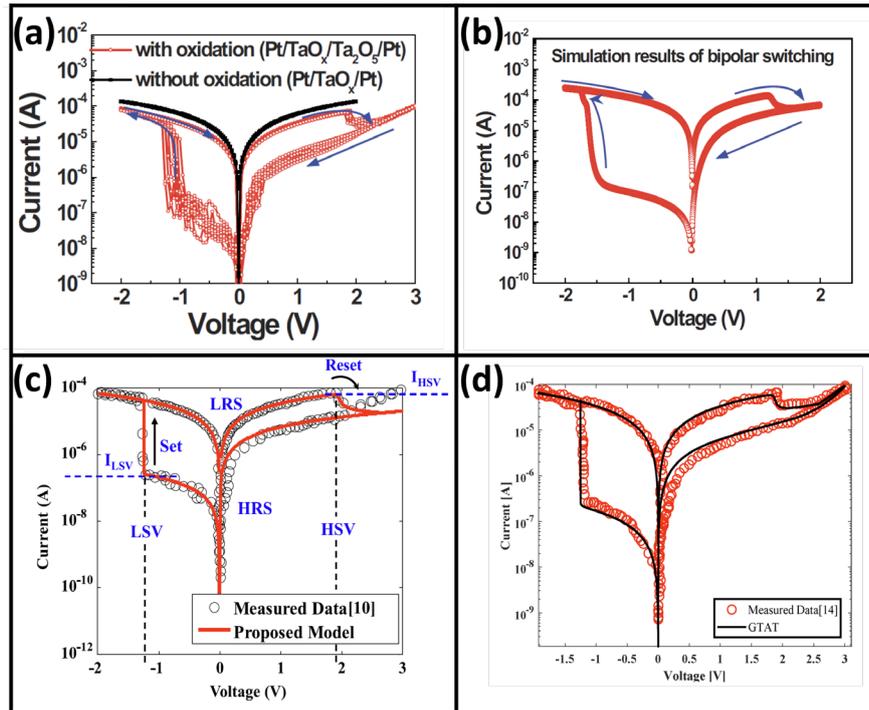


Fig. 2-30. (a) Experimental [87] and (b) simulation results for  $\text{Ta}_2\text{O}_5/\text{TaO}_x$  ReRAM with Schottky conduction [87]. (c) Simulation result from [27]. (d) Simulation result from [98].

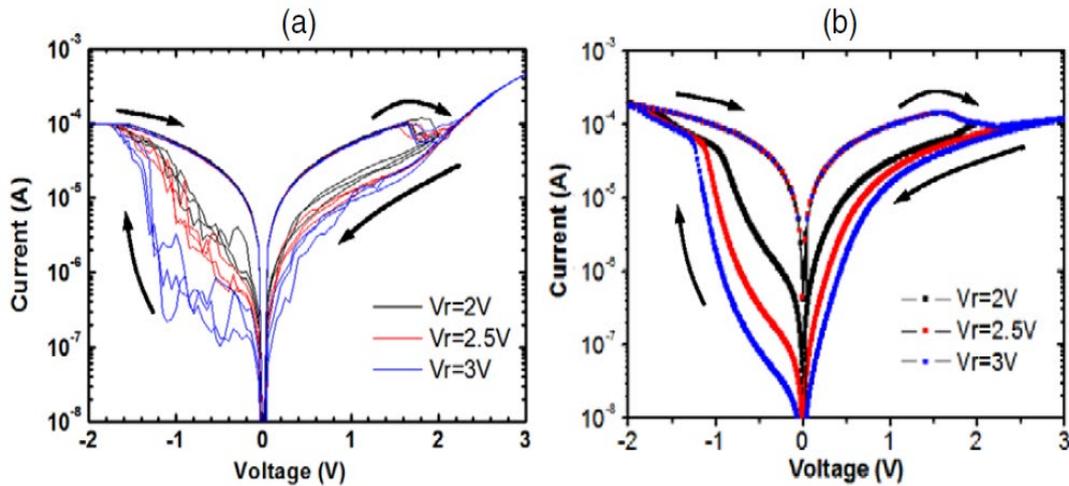


Fig. 2-31. (a) Experimental results for  $\text{Ta}_2\text{O}_5/\text{TaO}_x$  multilevel ReRAM device with Schottky conduction [99] b) Modeling multilevel for  $\text{Ta}_2\text{O}_5/\text{TaO}_x$  device taken from [99].

primarily dependent on the Schottky conduction in [98].

The multilevel capability of the  $\text{Ta}_2\text{O}_5/\text{TaO}_x$  has been proven experimentally as in Fig. 2-31(a) and theoretically modelled shown in Fig. 2-31(b) as given in [99]. The multiple resistance states are the reflections of variations in the insulating volume of conduction filament with Schottky barrier height. Hence, Schottky conduction is derived as the major current mechanism in the model, while the RS is resulting from the formation and rupture of CF originated from the

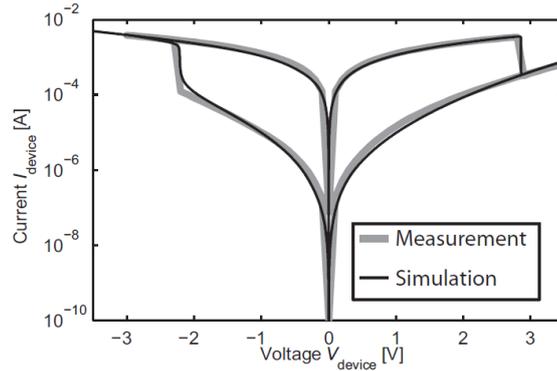


Fig. 2-32. Experimental result and model taken from [88].

physical mechanism modelled with a differential equation derived from the equation of motion of ions.

In [88], HRS and LRS are modelled by a Schottky contact with tunneling current and Poole Frenkel conduction areal current and the generated I-V are symmetrical as a result (Fig. 2-32). The simulation result is shown to match well with the experimental measurement as shown in Fig. 2-32. Meanwhile, the model presented in [88] defined the total current as the sum of state dependent Schottky current contributed by ionic and electronic components in addition to the state independent areal current that activated only at HRS. The state variable is considered as a function of average  $V_O$  concentration which determines the RS process in the model.

Alternatively, the experimental results and models based on  $HfO_2$  [100] are plotted in Fig. 2-33, while instead of adopting Schottky mechanism, different physical approaches are implemented such as hopping current conduction and redox based bipolar RS, where RS is demonstrated with the gap dynamics due to the generation recombination and migration based redox modelling approach. However, the gap evolution exponentially related to hopping conduction along with the hyperbolic sinusoidal function of applied voltage. Similarly, the characteristics

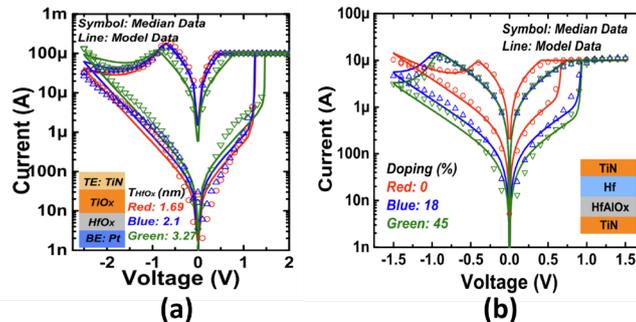


Fig. 2-33. (a) Experiment data and model data plotted for TiN/TiOx/HfOx/Pt device taken from [100]. (b) Experiment data and model data plotted for TiN/Hf/HfAlOx/TiN device taken from [100].

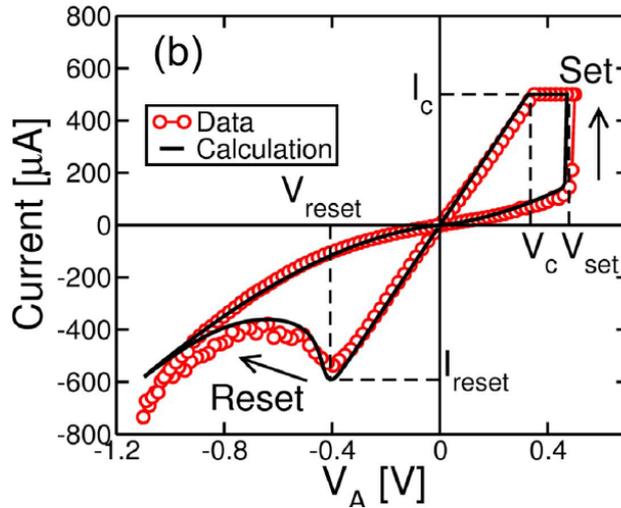


Fig. 2-34. Experimental I-V characteristics of the device, measured data, and calculated data plotted in linear scale taken from [91].

of  $\text{AlO}_x/\text{HfO}_2$  is also modelled with the same physical concepts of generalized hopping conduction [100].

To validate the model presented in [91], the plot of measured and calculated data given in Fig. 2-36 is analyzed and found that the model is fitting well with experimental results where the current profile is derived by the temperature and material dependent current conduction model where the variable thermal conductivity of the gap region is achieved with exponential law and variable resistivity is derived from the Poole conduction law while constant values are assigned for resistivity and thermal conductivity of the metal. Further, the RS is modelled with thermally activated ion kinetics falling under the redox reaction mechanisms.

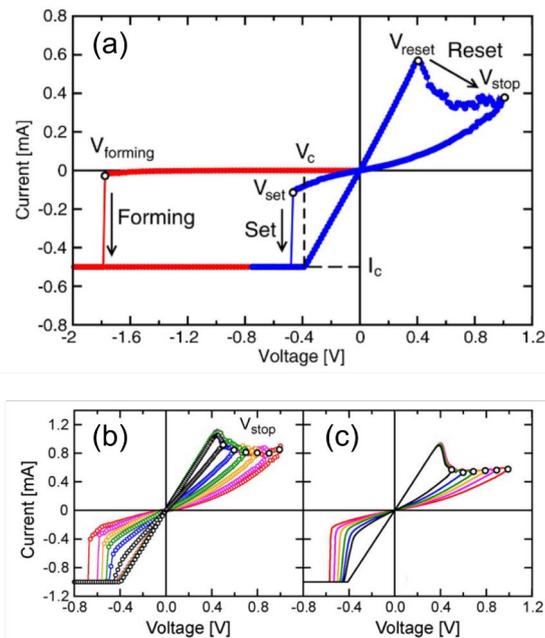


Fig. 2-35. (a) Typical linear I-V characteristics of  $\text{HfO}_x$ -based device reported in [102], (b) and (c) show measured data and model simulation as reported in [92].

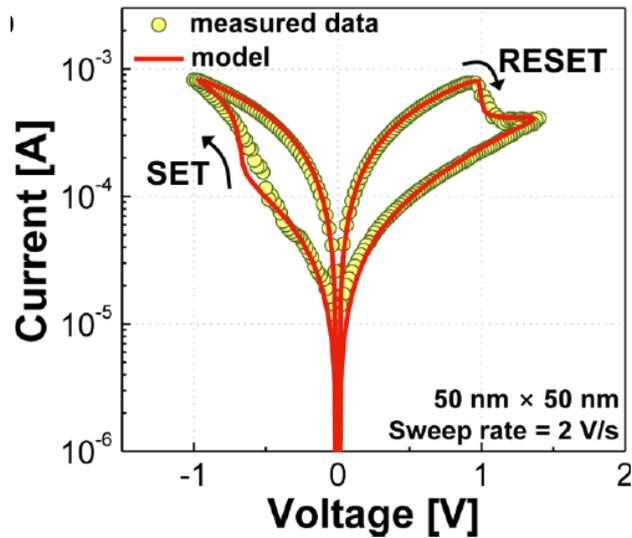


Fig. 2-36. Experimental I-V characteristics of Pd/Ta<sub>2</sub>O<sub>5</sub>/TaO<sub>x</sub>/Pd device and corresponding model as given in [93].

Similarly, the model shown in Fig. 2-35, is based on the drift-diffusion RS mechanism and electronic current is represented with carrier continuity expression and Fourier equation is used for Joule heating. However, in contrast to above mentioned model, instead of resistivity, variable thermal conductivity as well as thermally activated electrical conductance has been utilized to model the electronic current where the magnitude of electronic current changes with transition of oxide material from insulator to metallic phase [102].

The simulation results of model based on Pd/Ta<sub>2</sub>O<sub>5</sub>/TaO<sub>x</sub>/Pd structure is given in Fig. 2-36, where the current conduction is modelled with the basic current continuity approximation

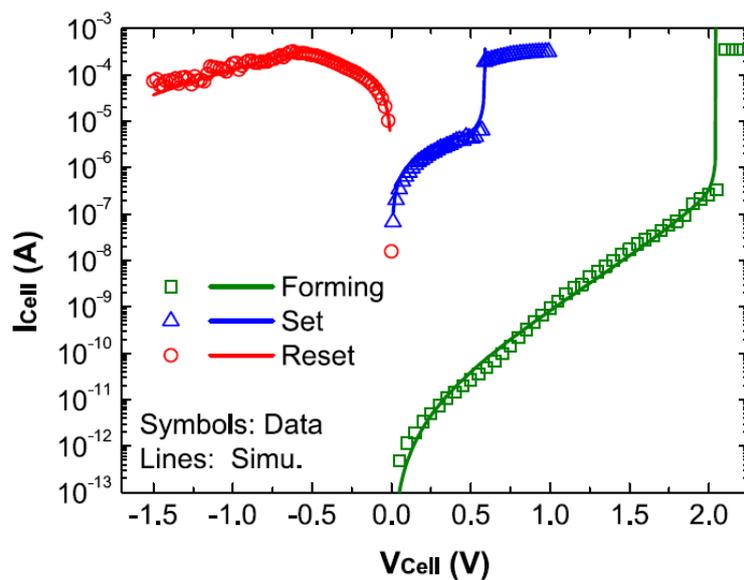


Fig. 2-37. Experimental I-V characteristics of Pd/Ta<sub>2</sub>O<sub>5</sub>/TaO<sub>x</sub>/Pd device and corresponding model as given in [101].

TABLE 2-5. Error Analysis of the Simulated Results against Experimental Measurements

Model	[98]	[100]	[100]	[93]	[88]	[27]	[91]	[90]
Error Percentage (%)	19	12	24	9	23	20	8	3

where the variation of current is linear with the applied voltage and the RS is obtained by the drift diffusion-based approach. Even though the model describes the fine tuning of the hopping distance parameter ( $a$ ) to select analog or digital switching behaviour, the measured data shown in Fig. 2-36 depicts analog switching characteristics while the model displays a digital switching behavior [93].

To validate the physics based compact model of TiN/HfO<sub>2</sub>/TiN stack with real device characteristics, an analysis has been done and found that the model is in well agreement with the experimental data showing the suitability of applied physical mechanisms such as the redox-based RS and the modelling of co-existing current mechanisms such as tunnelling through the pristine oxide structure and the ohmic conduction through the conduction filament as well as sub oxide phase. Additionally, the model reports the importance of temperature on RS process and modelled with heat equation. In addition to the modelling of SET/RESET, the model given in [101] included the characterization of electroforming stage in which highly resistive pristine oxide is converted to conductive channel on the application of high bias and the concept has been modelled with the rate of growth of sub oxide defined with the redox mechanism (Fig. 2-37).

The models presented in this review have been compared with their respective experimental results. Due to the different materials and physical mechanisms adopted, a direct comparison between models is unfeasible. An error analysis is however carried out to determine each model's accuracy compared to the experiments. The results are listed in TABLE 2-5.

## 2.4 ReRAM Fabrication and Layout

As previously discussed, one of the main advantages of ReRAM technologies are their low cell footprint;  $4 F^2$  where  $F$  is minimum feature size of the fabrication lithography and their high BeOL compatibility with existing CMOS fabrication processes. This low footprint area can be truly exploited by packing ReRAMs into dense crossbar arrays (CBAs).

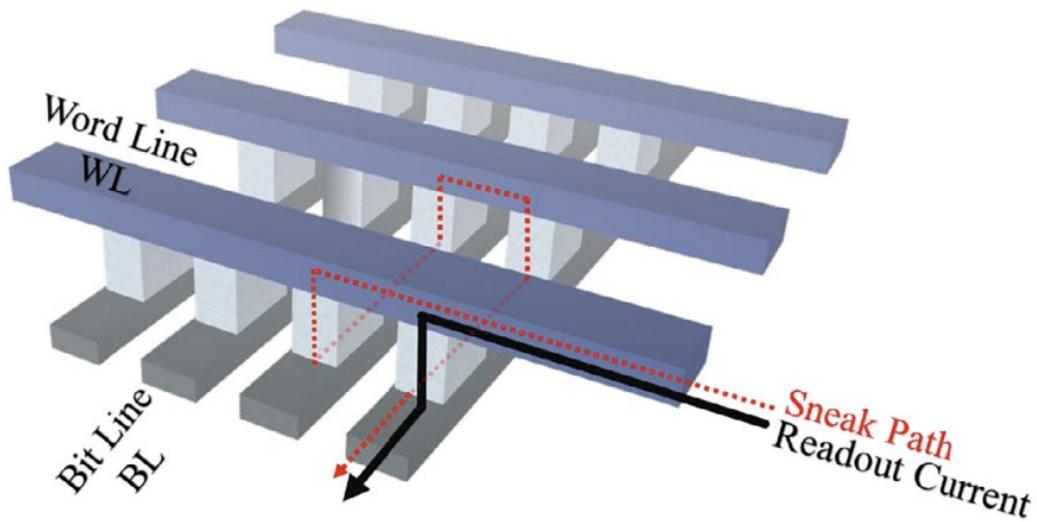


Fig. 2-38. ReRAM CBA array, ReRAM cells are located between wordlines, WL and bitlines, BL. Sneak path current is represented by dotted red line. Solid black line indicates desired READ current path [109].

An early example of a basic ReRAM CBA consists of ReRAM cells stacked between two orthogonal metal lines as shown in Fig. 2-38 [109]. The CBA density can be further increased to  $4 F^2/n$  by stacking  $n$  crossbars on top of each other [110]. An inherent problem with these layouts is the existence of sneak path currents, exhibited in Fig. 2-38. Unselected cells that are

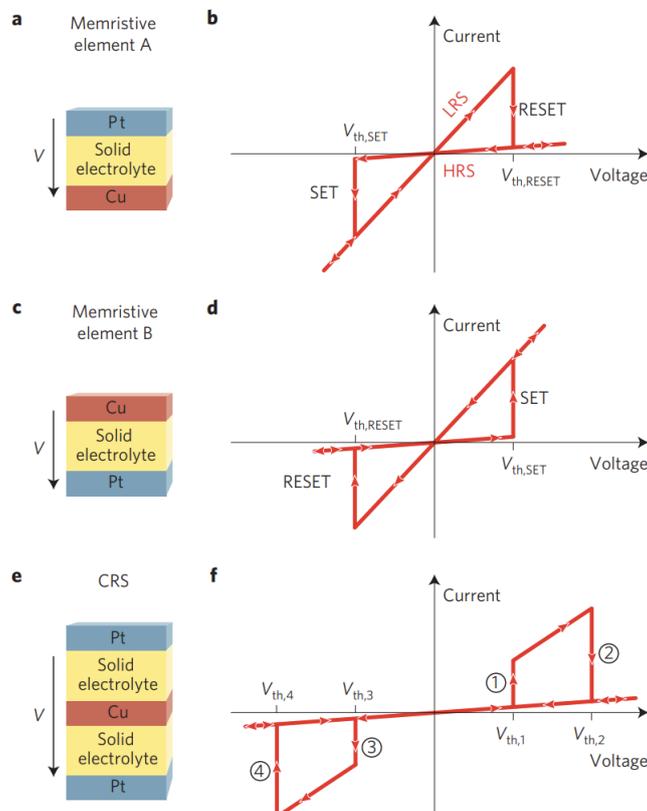


Fig. 2-39. (a) Pt/solid electrolyte/Cu ReRAM and its I-V characteristic (b). (c) Cu/solid electrolyte/Pt ReRAM and its I-V characteristic (d). (e) CRS resulting from the combination of ReRAMs in (a) and (b) and its I-V characteristic (f) [110].

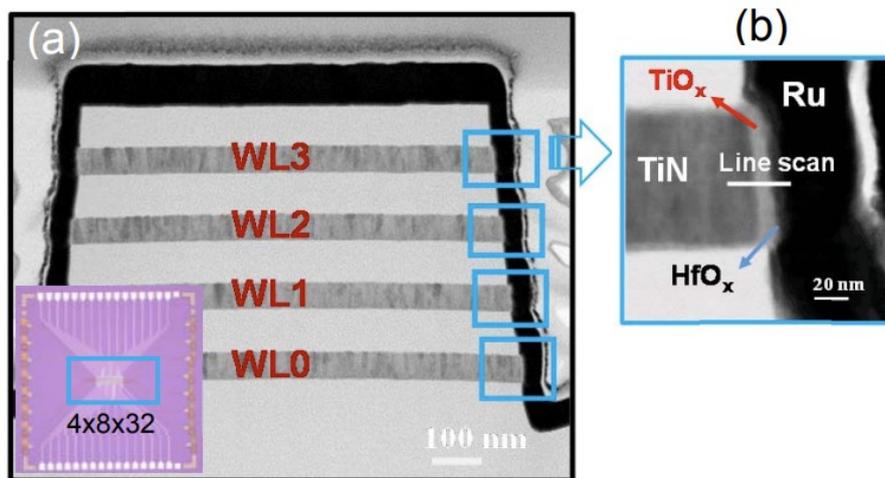


Fig. 2-40. (a) Cross-sectional TEM image of a 4-layer 3D VReRAM array. (b) Magnified image of TiN/TiO<sub>x</sub>/HfO<sub>2</sub>/Ru 1S1R cell. [117]

connected to the metal BLs and WLs provide a current path which leads to resistance drift in the unselected devices as well as deteriorating READ and WRITE margins.

Approaches to mitigate this problem involve floating the unselected ReRAM cells or introducing nonlinearity to the ReRAM cell itself. Floating unselected ReRAM is a relatively simple approach; adding a transistor or a diode to each ReRAM cell to produce a 1T1R or 1D1R cell respectively [111]. This approach however severely reduces CBA density and negates the advantage of ReRAM's low cell footprint.

Altering the ReRAM cell's linearity is a preferable approach although it requires more complex refinements. A CRS where two ReRAMs are arranged antiseriably was proposed in [110]. This antiseriably configuration constitutes a voltage divider between the two ReRAMs and the cell is only 'turned on' when both ReRAMs are in LRS as shown in Fig. 2-39.

Another approach is the introduction of a selector layer to produce a 1S1R cell; ie. 2D graphene in [112] or a RS selector layer [113]. The 1S1R configuration in effect increases the cell's threshold voltage and the cell is not turned on when low current is applied to it. Selector properties have to be optimally finetuned as the cell's READ margin is affected by HRS and LRS resistances [114].

In addition to individual cell improvements, new 3D CBA layouts using vertical ReRAMs (VReRAMs) have been proposed. A vertically stacked VRRAM\_1 was proposed in [115] with high pillar density in the metal plane direction.

A ‘VRRAM\_2’ CBA design from [115] was proposed where the memory cell is located between horizontal Wordlines and vertical Bitline pillars. This design was shown in [116] to achieve better READ margin than horizontal CBAs for large arrays (>400) due to lower leakage and power consumption.

The benefits from 1S1R cells and 3D vertical ReRAM (3DVReRAM) layouts led to the combination of both in [117] (Fig. 2-40). Inter-layer leakage was shown to be eliminated with off-state leakage current of 0.1 pA,  $\mu$ A operating current, and high endurance (>10<sup>7</sup> cycles) and retention.

Research into high density ReRAM arrays with good performance are ongoing as the ReRAM’s nonvolatility and tunable linearity/nonlinearity shows potential for in-computing memories (ICM) in upcoming non-Von Neuman architectures as well as vector-matrix cells for neuromorphic computing.

## **2.5 Research Gaps in the Literature Review**

The literature review of ReRAM modelling and nvFPGA leads to several gaps that currently exist in the field and an opportunity for this PhD project. They are:

1. Existing ReRAM models in the literature have addressed RS behaviour during HRS and LRS and account for gradual and abrupt RS characteristics. However, experimental results of gradual HRS (RESET) switching demonstrates current peaks and spikes. This can be attributed to the properties of the metal-oxide layer. Previous ReRAM models achieve RS through the equations that describe the ionic/ $V_O$  migration through a single CF or at the electrode/metal-oxide tunnel barrier. Studies on the grain boundaries in the metal-oxide’s crystalline structure indicates that there is a preferable condition for the growth of more than a single filament in the metal-oxide in a strong electric field. There is therefore an opportunity to produce a ReRAM model that incorporates multi-filamentary switching.
2. The FPGA is a widely used very large-scale integration (VLSI) circuit that was introduced in the 1980s. Since then, the design has advanced with transistor technology as SRAMs, DRAMs, and Flash cells provide a significant portion of FPGAs. The drawbacks of these transistor-based technologies have nonetheless become more serious as they scale down to single digit nanometer gate sizes. A next-generation memory technology that has shown potential to operate at or below these sizes would allow further advancement in

- FPGA technologies. The ReRAM is a nominal candidate as it satisfies this criterion and has desirable performance advantages over competing next-generation memory devices.
3. LUTs perform combinational operations in the FPGAs and make up a significant percentage of FPGA circuitry. Conventional LUTs are made up of volatile SRAM cells that are each made up of six transistors (called a 6T configuration). The SRAM cells therefore have large cell areas. Replacing the SRAM cells in the LUT with NV devices like ReRAMs with smaller cell footprint would introduce NV and reduce the physical sizes of LUTs in the FPGA.
  4. Additionally, using MB-ReRAMs enable storing more than one bit per cell and would further reduce LUT cell area.
  5. The output voltage from ReRAM cells in the array are in the subthreshold region due to the high intrinsic ReRAM resistances. Voltage-mode SAs are widely used but are unable to function at these subthreshold voltages. The design of a voltage-mode SA that is capable of sensing subthreshold voltages from ReRAM outputs would contribute to adoption of ReRAMs in the array.
  6. The D Latch and D Flip-Flop are traditional memory circuits that are utilized in FPGAs. Literature review of these circuits have shown numerous successful implementations of NV. However, existing nvD Latches and nvD Flip-Flops require STORE/RESTORE sequences which come at the cost of additional timings and energy consumption. The external circuitry required for these active STORE/RESTORE sequences also increase the circuit's footprint. Designs for nvD Latches and nvD Flip-Flops with passive STORE/RESTORE would provide a solution for these problems.
  7. Another memory component used in FPGAs is the DRAM. DRAMs are volatile 1T1C cells with large sizes due to the physical limitations of the capacitor component. Capacitors also experience charge leakages during usage and DRAMs therefore have to be refreshed after a period of time. Replacing the capacitor with a NV device like the ReRAM would drastically reduce the cell size and eliminate the refresh requirement of the 1T1C DRAM, thus reducing device power consumption.
  8. The nvFPGA which comprises of an nvCLB and an nvSB can be designed with the NV designs discussed above. The nvCLB will consist of the nvLUT and the nvDFF while the nvSB will store its configuration bits in the nvDRAMs.

## 2.6 Summary of Literature Review

Different emerging NVM devices are compared for selection for this PhD project. The analysed devices were the PCM, STT-MRAM, and ReRAM. Amongst the three, the ReRAM is a suitable candidate as it utilizes lower cell area, has multibit capabilities, and long endurance. Operational voltages and energy as well as READ/WRITE and retention times for the ReRAM are similar to the other NVM devices.

A comprehensive analysis of physics-based models has been carried out by identifying various approaches and parameters adopted in RS and current conduction mechanisms. The investigated approaches to modelling RS are drift-diffusion and redox-based mechanisms. As for current conduction, Schottky conduction and a general hopping mechanism are analysed along with their required parameters. This review has found that the RS process is dependent on the electric field and temperature regardless of the mechanism. Meanwhile for current conduction mechanisms, the variation in temperature does not contribute significantly in the models. The CF geometry is found to have an important role on RS behaviour as well as current conduction, with 2D and 3D models which calculate both radius and gap length variations generally exhibiting more accurate results.

A literature review of the memory circuitries that make up the nvFPGA, namely the nvD Latch and nvD Flip-Flop. Existing memory circuits successfully exhibit NV but require an active STORE phase to WRITE data into the NV components in the circuit as well as an active RESTORE phase to restore data from the NV component to the memory circuit after power restoration. The active STORE/RESTORE phases also require additional circuit components in addition to the NV components, leading to larger cell areas.

### 3. Chapter Three

## **Methodology**

---

This research aims to present an nvFPGA comprised of nvCLBs that is made up of MB-nvLUTs and nvDFFs and nvSwBs made up of nvDRAMs. This section delves into how the designs of these components were accomplished. Unless mentioned, the measurements in this PhD research were carried out in electronic design automation (EDA) simulator LTSpice.XVII from Analog Devices. The transistors models used in this research are from Predictive Technology Model (PTM) by Nanoscale Integration and Modeling (NIMO) Group [118]. PTM models are available in different technology nodes (ie. 32nm, 45nm, 65nm) and were chosen due to the models' realistic results and for their compatibility with LTSpice.

### **3.1 Multifilamentary Model**

The implementation of MB into an LUT array requires a study of the physical mechanisms of ReRAMs. This then led to opportunity to present a model of a multifilamentary ReRAM that captures the tunnelling barrier effect in the CFs. The following are the steps undertaken to create the multifilamentary model presented in this work.

1. The existence of multifilamentary phenomena is confirmed through literature review.
2. The average ionic drift velocity of the CF is modelled using Mott and Gurney rigid point-ion model and Joule heating effect. This equation describes the state variable motion during RS.
3. The current conduction is modelled using the Trap-Assisted Tunnelling equation, taking into account the tunnelling barrier height effect on the CFs in the ReRAM.
4. Model parameters for the equations are confirmed through testing in MATLAB.
5. The physical differences between the CFs are represented by their respective activation energies. Activation energy values are tested and confirmed in MATLAB.
6. Equivalent circuit models for the physical equations are created to produce an electronic design automation (EDA) compatible version of the model.
7. The circuit model behaviour is tested in LTSpice.XVII.

### **3.2 Non-Volatile LUT**

Next, this research focuses on the creation of an nvLUT based on MB-ReRAMs. First, a study on SB-ReRAM nvLUT was performed to analyze the sneak-path current effects. This led to an opening to develop a sense amplifier (SA) circuit sensitive to the low subthreshold output voltages from a ReRAM array. Designs of an nvLUT comprised of MB-ReRAMs and the nvLUT controller are finally presented.

1. The characterization of an 2-input SB-nvLUT (4 ReRAM cells) array is performed in LTSpice.XVII.
2. READ/WRITE testing was performed with analysis of the sneak-path effects on unselected cells in the array.
3. Design of an SA buffer circuit using transistor models from Predictive Technology Model (PTM) by Nanoscale Integration and Modeling (NIMO) Group [118] that correctly amplifies the low output voltage from ReRAM arrays is conducted using LTSpice.XVII.
4. Output from the SA buffer circuit is fed into a designed differential comparator circuit using PTM transistors. The behaviour of the comparator circuit is analyzed in LTSpice.XVII.
5. Simulation results of the SA circuit are obtained in LTSpice.XVII and compared with results from literature.
6. The truth table logic for the MB-ReRAM LUT is conceptualized using resistance windows for four resistance states from an MB-ReRAM model in LTSpice.XVII.
7. MB-ReRAM nvLUT controller circuit capable of providing the required WRITE/READ voltages is designed and tested in LTSpice.XVII.
8. The MB-ReRAM LUT array and the controller circuit are combined and circuit characterization is performed in LTSpice.XVII.
9. WRITE/READ performance tests of the MB-ReRAM nvLUT are performed in LTSpice.XVII and compared with SB-ReRAM nvLUT and SRAM LUT.
10. Benchmark circuit tests are performed for the MB-nvLUT and the results are compared with the the SB-ReRAM nvLUT and SRAM LUT. The average delay and EDP benchmark results are obtained by calculating the respective average delays and EDP of the MB-ReRAM nvLUT, SB-ReRAM nvLUT, and SRAM LUT for different  $n$ -input LUTs and summing the values according to the number of  $n$ -input LUTs in the benchmark circuits.

### **3.3 The Sequential Memories**

NV is then implemented into the sequential memory components of the FPGA namely, the D Latch, DFF, and the DRAM.

1. The designs of the NV versions of the D Latch and the DFF were carried out in LTSpice.XVII.

2. Both nvD Latch and nvDFF circuits are designed with the ability to passively STORE/RESTORE data into/from the ReRAMs, negating the need for peripheral control circuitry.
3. Designs of the nvD Latch and nvDFF circuits using 32, 45, and 65 nm technology node PTM transistors are produced and tested in LTSpice.XVII.
4. Performance criteria measurements are carried out in LTSpice.XVII and compared with their respective equivalent circuits in literature.
5. Design of nvDRAM using PTM transistors is created in LTSpice.XVII.
6. Performance tests for the nvDRAM circuit against conventional DRAM such as refresh time and bitflip tests performed in LTSpice.XVII.
7. Analysis of nvDRAM circuit with 22, 32, and 45 nm technology nodes PTM transistor carried out and compared with nvDRAM in literature.

### **3.4 The nvFPGA**

The NV components produced in this research are combined to present a holistic nvFPGA architecture that is based on ReRAMs. The combined architecture which consists of an nvCLB made up of 6-input MB-ReRAM nvLUTs and nvDFFs and SwBs comprised of nvDRAMs is compiled in LTSpice.XVII. The nvFPGA is then tested with an input application. Performance metrics such as the device area utilization, path delays, and power dissipation of the nvFPGA is obtained through analysis of its subcomponents.

## 4. Chapter Four

# Multi-Bit ReRAM Model

---

This chapter delves into the physical and electrical modelling of a multi-filamentary ReRAM. The phenomena of multi-filamentary RS are first discussed, then the equations and parameters used to model the multiple filaments in the ReRAM are presented. Simulations of the presented model are carried out and compared with a single filament ReRAM model and experimental results. The voltage input to the multi-filament ReRAM model is then varied to analyse the characteristics of each individual filament and how they contribute to RS and MB behaviour.

## Related Publications

---

1. H. L. Chee, T. N. Kumar, and H. A. F. Almurib, "Electrical model of multi-level bipolar Ta<sub>2</sub>O<sub>5</sub>/TaO<sub>x</sub> Bi-layered ReRAM," *Elsevier Microelectronics Journal*, vol. 93, no. March, p. 104616, 2019.
2. H. L. Chee, T. N. Kumar, and H. A. Almurib, "Multifilamentary Conduction Modelling of Bipolar Ta<sub>2</sub>O<sub>5</sub>/TaO<sub>x</sub> Bi-Layered RRAM," *Proceedings in 7th IEEE Non-Volatile Memory Systems and Applications Symposium (NVMSA)*, 2018, pp. 113–114. (Japan)

#### 4.1 Multifilamentary ReRAM

**R**S in ReRAMs occurs from the forming/rupture of a CF or the increase/decrease of the tunnelling barrier height at the interface between layers. This behaviour is possible as transition metal-oxides have multiple oxidation states with different resistances. RS is induced by applying a voltage across the ReRAM, inducing an internal electric field that generates migrations of ions which aggregate to form CFs which can comprise of metal ions, oxygen vacancies or a mixture of both [119].

The formed CFs are preferentially located at Grain Boundaries (GB) and their distribution is unaffected by RESET voltage. Depending on the polycrystalline structure of the device, defects can aggregate to form a single large conduction channel or multiple conduction channels. In multi-channel configurations, the channels or filaments with larger cross-sections have higher number of traps. The existence of multiple CFs is widely known and have been shown in transmission C-AFM and TEM analysis [120].

The proposed model captures the behaviour of three CFs in the  $Ta_2O_5$  layer with different physical properties ie. diameter and barrier height, during HRS RESET switching. The barrier height is affected by filaments with differing diameters and thus varying filamentary concentration of traps and defects [121], [28], [89]. The different barrier heights (33% variation between smallest and largest filament in range with existing literature [120], [28]; the lowest barrier height of 0.9 eV and the highest barrier height of 1.2 eV are selected as respective lower and higher values from the nominal barrier height of 1 eV taken from [75]) cause the filaments

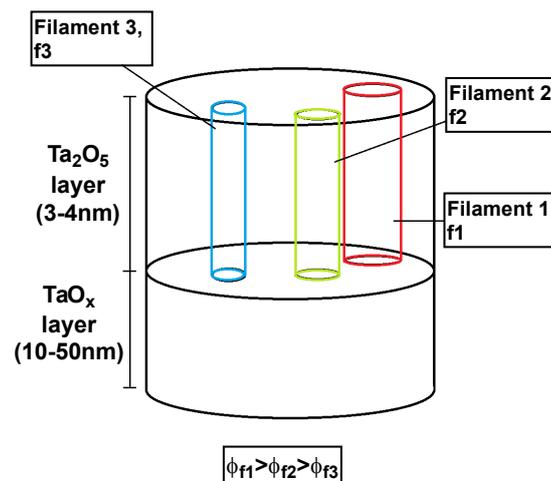


Fig. 4-1. Schematic representation of the multi-filament ReRAM. Filaments  $f_1$ ,  $f_2$ , and  $f_3$  have different diameters,  $\phi$ .

to rupture at different voltages producing resistance spikes and two IRS states during HRS switching. A schematic of the model is shown in Fig. 4-1.

## 4.2 Electrical Modelling of Multi-Filament Bi-Layered Ta<sub>2</sub>O<sub>5</sub>/TaO<sub>x</sub> ReRAM

### 4.2.1 Migration of V<sub>o</sub>

The movement of V<sub>o</sub> is taken as the state variable, in this model and given by [27]:

$$\frac{d\omega}{dt} = v(t) \approx \begin{cases} v_{1,2} \cdot a \cdot c \cdot \exp^{-\frac{E_A}{kT}} \cdot \sinh \frac{x_{1,2}E}{E_0}, & 0 < \omega < D \\ 0, & 0 \geq \omega \geq D \end{cases} \quad (4.1)$$

where  $\omega$  is the state variable,  $v(t)$  is the velocity over time,  $v_{1,2}$  is the velocity prefactor,  $a$  is the hopping distance,  $c$  is the attempt-to-escape frequency,  $E_A$  is the intrinsic barrier for ion hopping,  $kT$  is the thermal energy,  $x_{1,2}$  is the field enhancement factor,  $E$  is the applied electric field and  $E_0=2kT/qa$  where  $q$  is the electron charge.  $\omega$  is limited within the boundaries of 0 and  $D$ , the length of the Ta<sub>2</sub>O<sub>5</sub> film.

A simplification is adopted for this multi-filament model where each filament is assigned different barrier values,  $E_A$  to account for the differing junction properties at the filament-electrode interface. In this model the ReRAM has three filaments with  $E_{A1} = 0.9$ ,  $E_{A2} = 1.0$ , and  $E_{A3} = 1.2$  respectively. These values are chosen as they are in range of the activation energies in literature [122].

The electric field,  $E$  is given as [27]:

$$E = \frac{V_U}{\omega} = \frac{V_D}{\omega + \frac{R_{ON}}{R_{OFF}}(D-\omega)} \quad (4.2)$$

where  $V_D$  is the voltage across the Ta<sub>2</sub>O<sub>5</sub> layer, and  $R_{ON}$  and  $R_{OFF}$  are the lowest and highest resistance values of the Ta<sub>2</sub>O<sub>5</sub> layer respectively.  $V_D$  is calculated from [27]:

$$V_D = \left( R_{ON} \left( 1 - \frac{\omega}{D} \right) + R_{OFF} \cdot \frac{\omega}{D} \right) \cdot I \quad (4.3)$$

## 4.2.2 Trap-Assisted Tunelling

The intrinsic variability of ReRAMs stems from the mechanisms behind  $V_O$  migration during RS –ionic drift-diffusion, thermally activated hopping, generation/recombination of  $V_O$ s, and trap-assisted-tunnelling (TAT) [123], [124]. Amongst them, there is evidence that  $V_O$  defect-assisted TAT is a major contributor to the intrinsic variability [124], [125]. The TAT current in this model is affected by the modulation of the interfacial barrier height based on [80] and correlated to the electrical model as [126]:

$$I = AA^*T_0^2 \exp^{-(n_{1,2}\phi_b)/V_T} \left( \exp^{\frac{V_S}{\eta V_T}} - 1 \right) \left( \exp^{-\sqrt{\xi}\omega \times 10^{10}} \right) \quad (4.4)$$

where  $A$  is the surface area of the electrode in contact with the filament,  $A^*$  is the Richardson constant,  $T_0$  is the ambient temperature,  $n_1$  and  $n_2$  are barrier height factors that act on the ideal Schottky barrier height,  $\phi_b$ ,  $\eta$  is the ideality factor and  $V_T$  is the thermal voltage.  $V_S$  is the contact voltage which is the voltage drop across the Schottky barrier at the oxide-metal interface. The

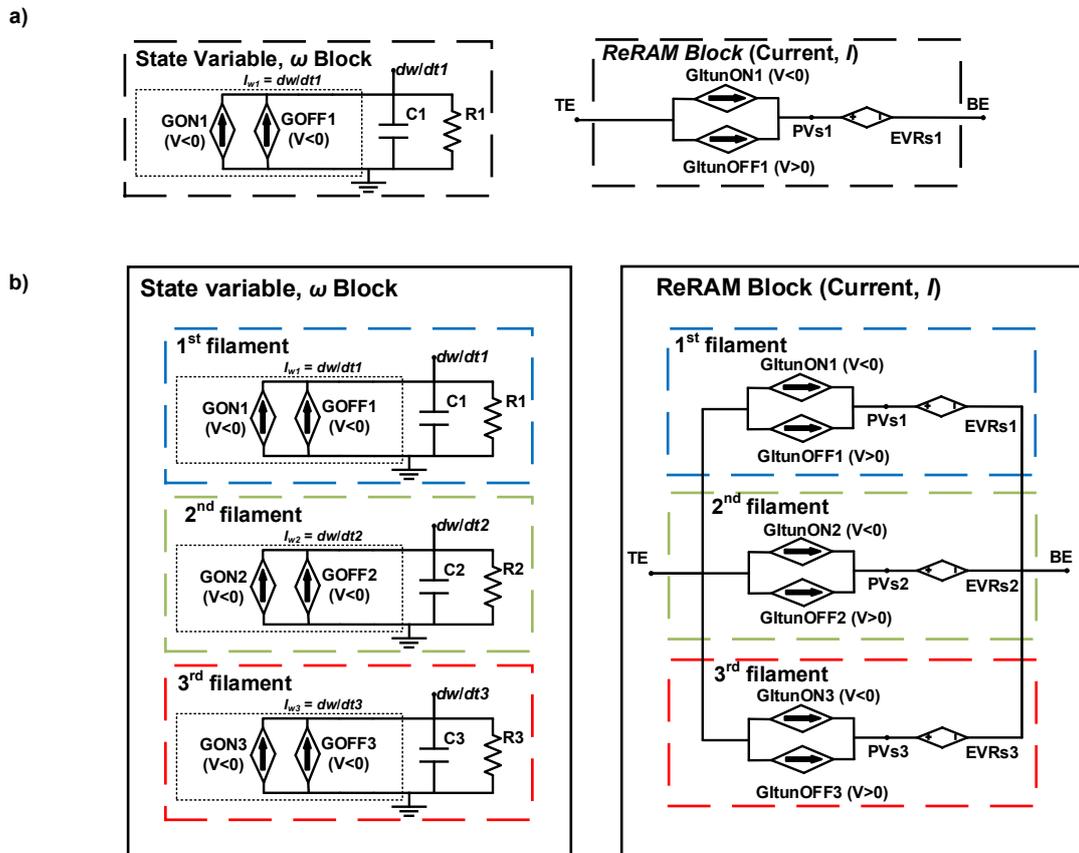


Fig. 4-2. Circuit blocks of (a) state variable migration and (b) current transport for three filaments. SPICE model for (a) single level cell and (b) multi-level cell (three filaments are used for our model). The state variable block produces the change in the ReRAM current.

rightmost term in (4.1) represents the tunnelling probability factor where  $\zeta$  is the effective tunnelling barrier height which is a fitting parameter,  $\omega \times 10^{10}$  is the thickness of the insulator volume in Å. The electrical circuit block for (4.1) is shown in (Fig. 4-2(a)) for each  $n$  number of filaments.

### 4.2.3 Current Equation

The current equation is given as (4.1). For  $n$  number of filaments,  $n$  blocks are calculated and the total output summed as shown in Fig. 4-2(b).

### 4.3 Simulation Results

The model is implemented as a SPICE macro-model for electrical simulations. Using a 3 V/-3 V 100 Hz sine-wave transient simulation, the model undergoes a gradual reset starting at  $\sim 1.5$  V and reaches its HRS at  $\sim 2.4$  V while set switching occurs at negative polarity at  $-1.15$  V. The current peaks during reset switching corresponds to the rupture of each filament; reset switching starts when the first filament ruptures (IRS-1 at  $1.5 \text{ V} \leq V < 1.6 \text{ V}$ ), the second peak appears when the 2nd filament ruptures (IRS-2 at  $1.6 \text{ V} \leq V < 1.8 \text{ V}$ ), and the final peak before HRS is from the rupture of the 3rd filament (HRS at  $V \geq 1.8 \text{ V}$ ). The small switching window of  $\sim 0.9 \text{ V}$  ( $0.1 \text{ V}$  for IRS-1,  $0.2 \text{ V}$  for IRS-2) results in small margin windows for each resistance level, magnifying the drastic effect of the device's intrinsic variation properties. Switching voltage margins should be used to ensure stability of the IRS states. A 5% switching margin between  $1.505 \text{ V}$  and  $1.595 \text{ V}$  can be applied for IRS-1 while a 25% switching margin between  $1.65 \text{ V}$  to  $1.75 \text{ V}$  can be applied for IRS-2. A precise input supply is thus essential for

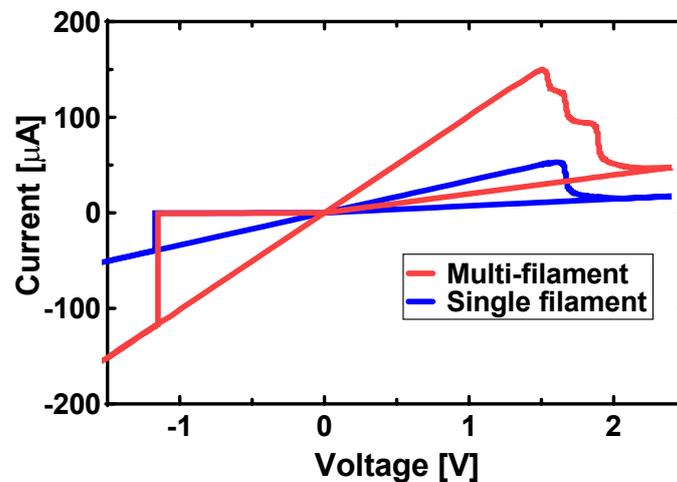


Fig. 4-3. Simulation I-V characteristics of the single filament model (blue) and the multi-filament model (red) for 3V/-3V 100Hz transient sine-wave.

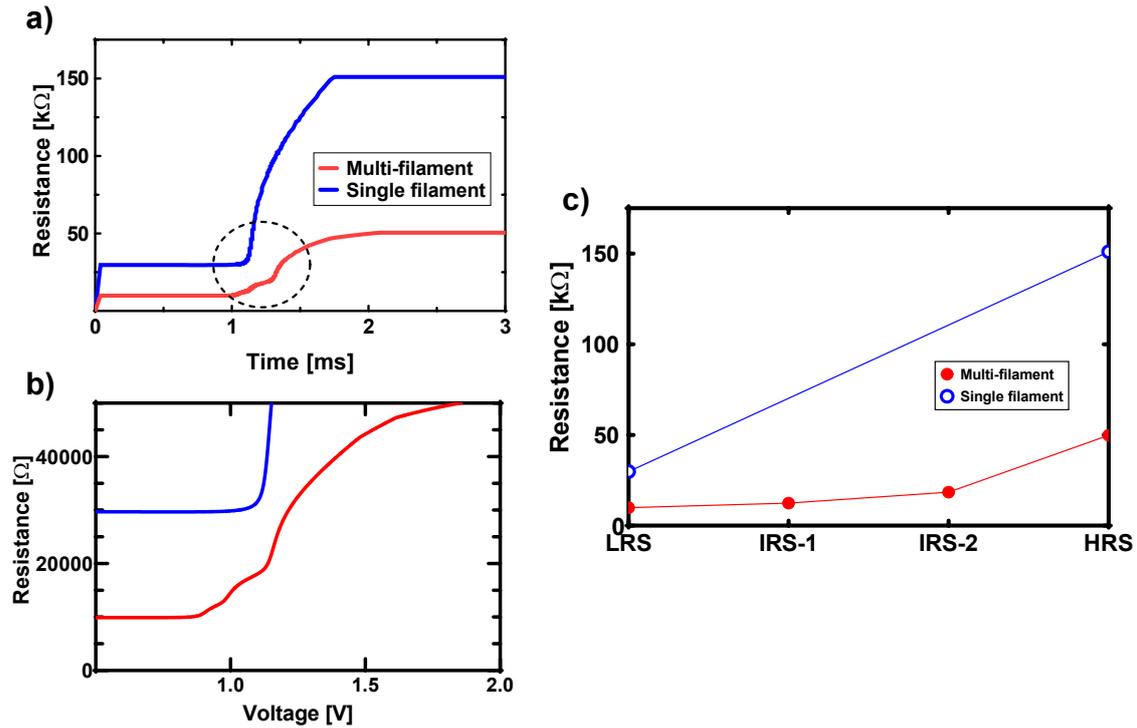


Fig. 4-4. (a) Simulated resistance over time graph of single filament model (blue) and multi-filament model (red) for 3V/-3V 100Hz transient sine-wave. The circled area is magnified and plotted in (b) and the variation during RS can be clearly seen. (c)

MB switching to prevent WRITE errors [127], [128].

To understand the difference between the multi-filament and single filament models, the I-V simulation result and the change of resistance over time for both are plotted. As can be seen in Fig. 4-3, the current level of the multi-filament model is 190% higher due to the extra conduction paths provided by the additional filaments. Consequently, the HRS resistance of the single filament model at  $\sim 150$  kΩ is much larger than the  $\sim 50$  kΩ of the multi-filament model (Fig. 4-4). The on/off ratio is also reduced by half from the presence of multiple filaments.

Reset switching is initiated earlier in the multi-filament model at around  $\sim 1.5$  V compared to  $\sim 1.6$  V in the single filament model. This is because the barrier height of the first filament,  $E_{A1}$  is lower than the barrier height of the single filament model. This induces the earlier rupture of the filament. The 2nd filament is however assigned the  $E_A$  value as the single filament model and they rupture at the same voltage,  $V \approx 1.6$  V.

Overall, the simulated result (Fig. 4-5(a) is in good agreement with the experiment result from [75]. Fig. 4-5 shows different resistance states achieved by the model from partial reset switching. IRS-1 has a resistance of  $\sim 13$  kΩ (Fig. 4-5(b)), IRS-2 has a resistance of  $\sim 19$  kΩ

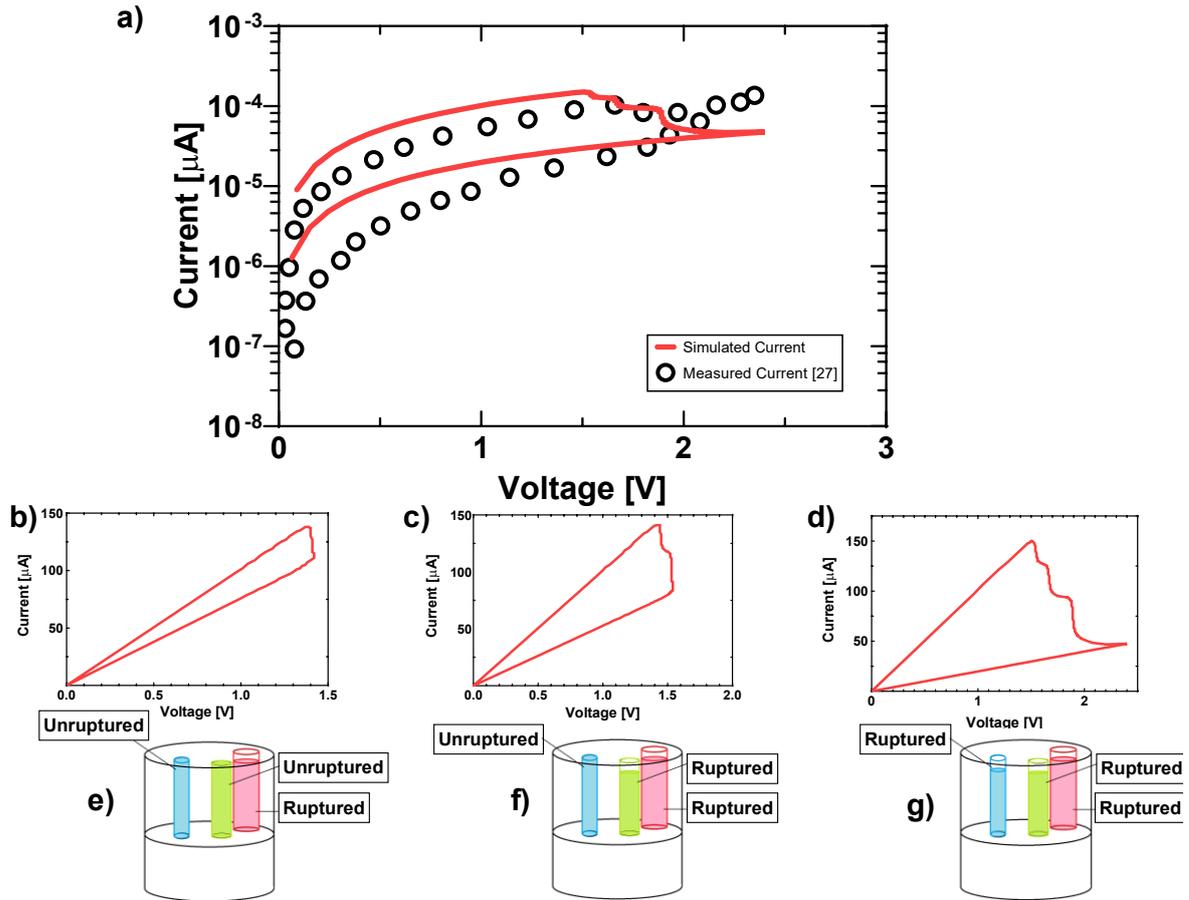


Fig. 4-5. (a) Semi-log I-V plot with experimental results from [75]. (b) Partial RS (IRS-1) to resistance=13kΩ, (c) partial RS (IRS-2) to Resistance=20kΩ, and (d) full RS (HRS) to resistance=50kΩ. Schematic depiction of the filaments are included in (e), (f), and (g) corresponding to the curves in (b), (c), and (d) respectively.

(Fig. 4-5(c)), and HRS has a resistance of  $\sim 56$  kΩ (Fig. 4-5(d)). IRS-2 matches well with experimental results while IRS-1 has a 30.35% percentage difference. The I-V curve after HRS ( $>2$  V) differ greatly due to over-reset phenomenon not accounted for in this model [129]. The RS peaks that are produced by the introduction of multi-filaments to the RS process are in agreement with the observable experimental RS switching curves such as in Fig. 4-6. The migration of the state variable,  $\omega$  is plotted in Fig. 4-7 and shows that each filament completely ruptures during reset.

The temperature and electric field on each filament are plotted in Fig. 4-8. The temperature and electric field strength on the filament increase proportionately with barrier height. This can be understood from (4.1) and (4.2). The higher barrier height impedes the migration of  $V_O$  and the filament requires higher electric field strength to rupture. This in turn increases the localized temperature on filaments with higher barrier height values. The filament with the highest barrier height, similarly experiences the highest temperature and  $E_{field}$  in Fig. 4-8. To further

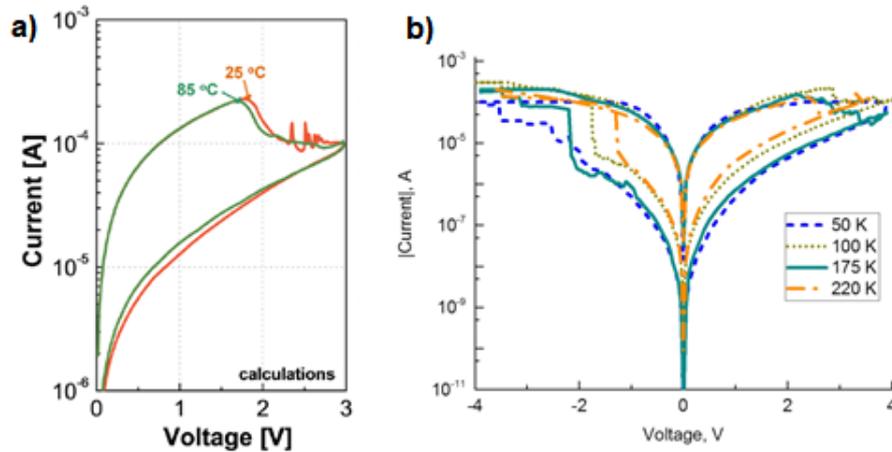


Fig. 4-6. Experimental I-V result during reset switching taken from (a) [75] and (b) [165].

understand the model’s versatility and the difference between the multi-filament and single filament models, a triangle and pulse voltage of 3 V reset is applied to both. The triangular voltage input results for the single filament and multi-filament model are shown in Fig. 4-9(a) and Fig. 4-9(b) and the pulse input voltage in Fig. 4-10(a) and Fig. 4-10(b) respectively.

Both models are initially in LRS for the triangular input voltage and the reset happens at  $\sim 1.78$  V for the multi-filament model and  $\sim 2.53$  V for the single filament model (Fig. 4-9(a) and Fig. 4-9(b)). Both reset voltages ( $V_{\text{reset}}$ ) are at similar levels to the sinewave  $V_{\text{reset}}$ . Repeated increasing positive triangular input voltages is supplied to simulate a multi-level switching

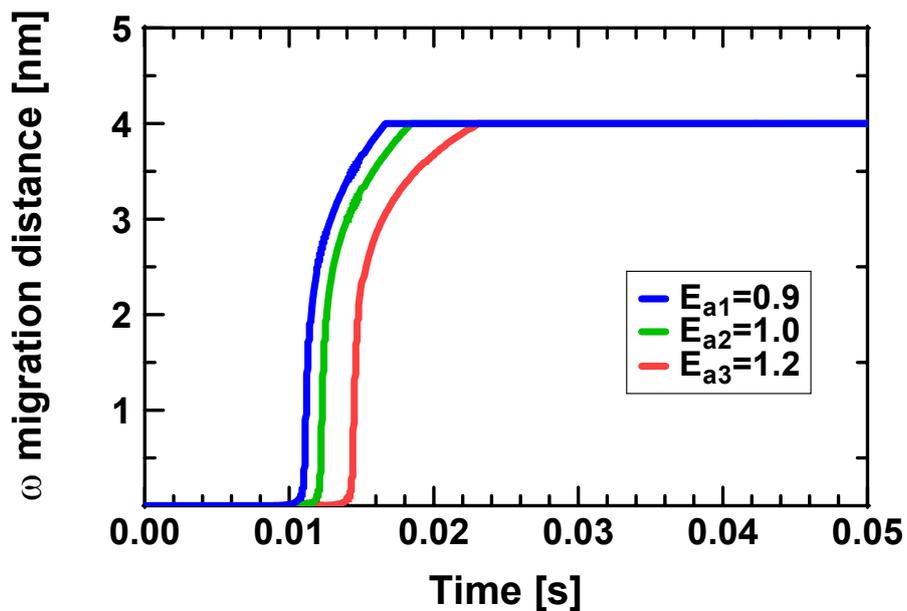


Fig. 4-7. Migration of  $\omega$  over time. The filament with the lowest barrier height reaches maximum distance first.

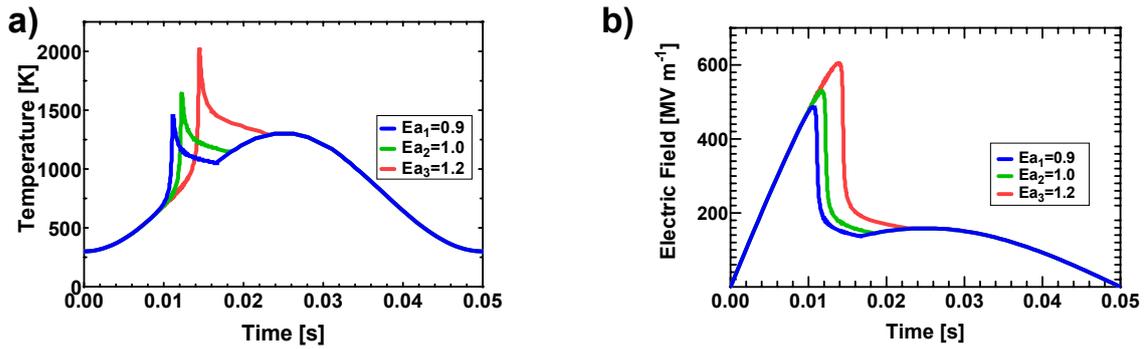


Fig. 4-8. (a) Temperature evolution over time and (b) Electric field strength on the individual filaments in the multi-filament model.

environment scheme. The values of the triangle inputs are  $V_1=1.56$  V,  $V_2=1.7$  V, and  $V_3=1.9$  V with a period of 10ms to reset the model into IRS-1, IRS-2, and HRS respectively. This input is simulated for both models and the change of resistance is plotted in Fig. 4-11(a) and Fig. 4-11(b) for the single filament and multi-filament models respectively.

The increase in the peak triangular voltage leads to a gradual resistance change in both models. In Fig. 4-9(a) the first triangle pulse is too low to induce any change in the single filament model. Reset switching begins at the second pulse leads to a resistance of  $\sim 64$  k $\Omega$  while the third pulse resets the device to  $\sim 108$  k $\Omega$ . The model resistance clearly increases in step according to the supplied input with IRS-1, IRS-2, and HRS values of 11 k $\Omega$ , 17 k $\Omega$ , and 20 k $\Omega$  respectively.

The multi-filament model performs accurately when supplied with a pulse voltage of 3V for 5ms. Fig. 4-10(a) and Fig. 4-10(b) plots the pulse current for both models. The multi-filament model resets immediately and maintains a stable resistance of  $\sim 50$  k $\Omega$  while the single filament model resets to a resistance of  $\sim 141$  k $\Omega$ . The 5ms pulse of 3 V is sufficiently resets both models

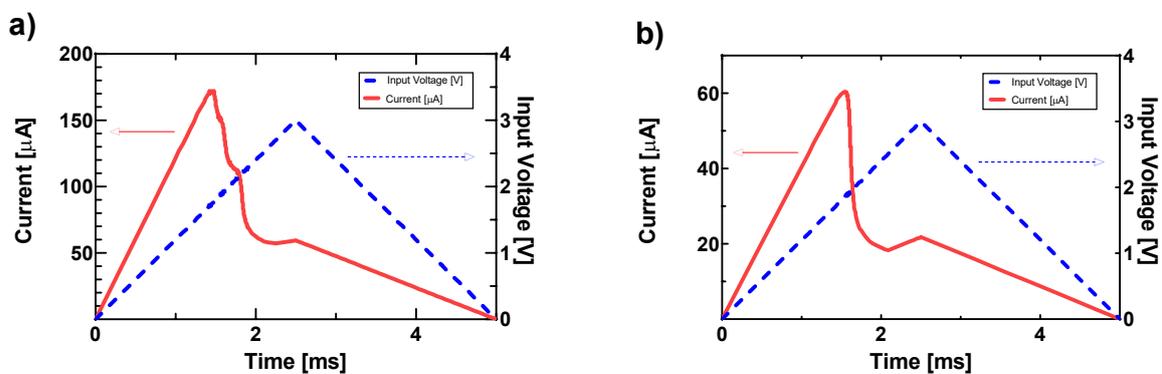


Fig. 4-9. 3V triangular input voltage simulation result for (a) multi-filament model and (b) single filament model.

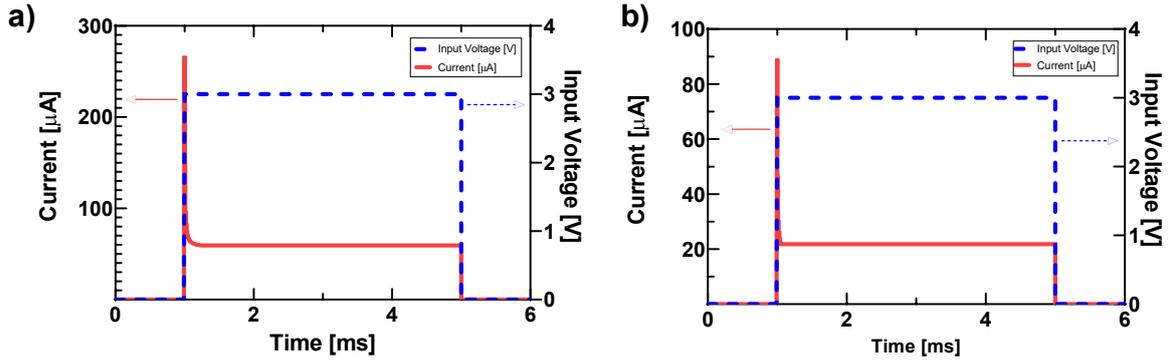


Fig. 4-10. 3V 5ms square input voltage simulation result for (a) multi-filament model and (b) single filament model.

to their highest resistance.

To test MB behaviour, two pulse voltage schemes with 1ms and 4ms pulse widths are utilized. Three sequential pulses of increasing voltages ( $V_1=1.56$  V,  $V_2=1.7$  V,  $V_3=1.9$  V) are provided and the results are plotted in Fig. 4-12 and Fig. 4-14 respectively.

The 1ms pulse scheme in Fig. 4-12 shows the expected trend of decreasing current and increasing resistance with each pulse for both models. Unlike the triangular voltage scheme, the maximum current for each  $n$ th pulse is initially higher than the lowest current from the  $n$ th-1 pulse due to the voltage spike from the pulse voltage input. This does not affect the multi-level switching as the pulse is long enough for the current to decrease to expected levels. The first pulse does not provide a strong enough electric field to instigate reset switching in the single filament model and it remains in LRS at  $\sim 30$  k $\Omega$  (Fig. 4-13(b)). The second pulse increases the model's resistance to  $\sim 80$  k $\Omega$  while a further increase in resistance occurs in the third pulse to reset the model into  $\sim 113$  k $\Omega$ . The multi-filament model achieves multi-state resistances of IRS-1 $\approx 12$  k $\Omega$ , IRS-2 $\approx 18$  k $\Omega$ , and HRS $\approx 30$  k $\Omega$  (Fig. 4-13(a)) from this scheme.

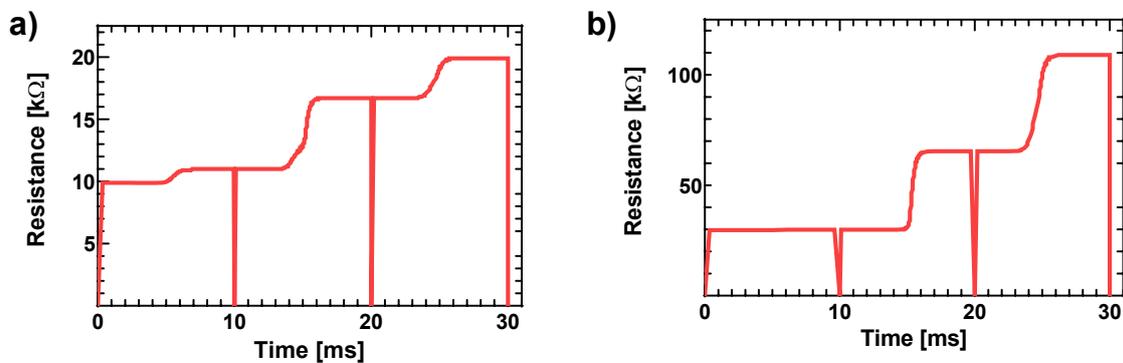


Fig. 4-11. Simulated change of resistance for (a) multi-filament model and (b) single filament model corresponding to the input voltage in Fig. 4-11.

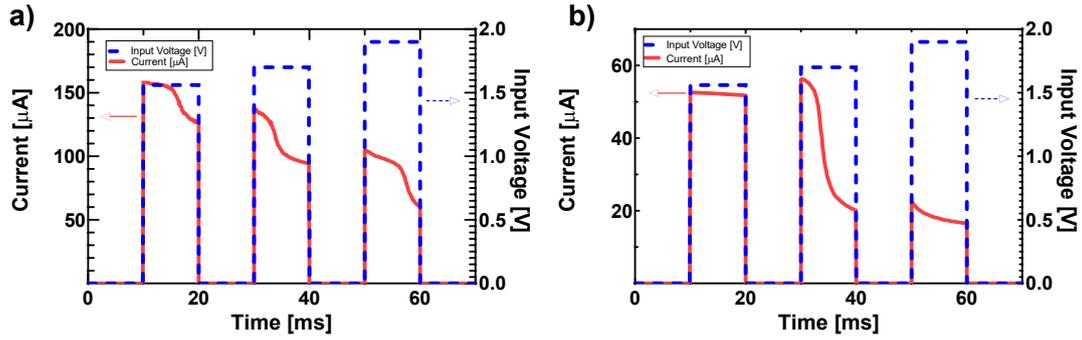


Fig. 4-12. Output current simulation results of 1ms consecutive square input voltages with varying amplitudes of  $V_1=1.56V$ ,  $V_2=1.7V$ ,  $V_3=1.9V$  for (a) multi-filament model and (b) single filament model.

The maximum resistances of the models are considerably lower than the maximum HRS resistance of achieved in the sinewave and triangular voltage inputs. This is because the 1 ms pulse width is too short for the filaments to undergo complete rupture from three pulses. This indicates that the pulse scheme allows an extra resistance state which was successfully achieved in simulation (results not shown).

Fig. 4-14 and Fig. 4-15 shows the current and resistance plot of the 4 ms pulse scheme. This pulse scheme provides a longer reset voltage exposure allowing higher resistances for each pulse than the 1ms scheme. The initial pulse is now strong enough to induce switching for the single filament model. The model resets from  $\sim 30 \text{ k}\Omega$  at 2 ms into the pulse to  $\sim 80 \text{ k}\Omega$  at the end of the pulse with a high rate of change of resistance of  $26 \text{ M}\Omega \text{ s}^{-1}$  (Fig. 4-14(b)). The model experiences a more gradual reset during the second and third pulses switching to resistances of  $\sim 112 \text{ k}\Omega$  and  $\sim 140 \text{ k}\Omega$  respectively. The multi-filament model's multi-state resistances for the 4ms scheme are  $\text{IRS-1} \approx 18 \text{ k}\Omega$ ,  $\text{IRS-2} \approx 19 \text{ k}\Omega$ , and  $\text{HRS} \approx 40 \text{ k}\Omega$  (Fig. 4-14(a)). The resistance of IRS-1 and IRS-2 are similar so this scheme is unpractical for the multi-filament model. The longer pulse width is strong enough to activate the migration energy for two filaments during the first pulse. As a result, two noticeable current and resistance peaks are present in Fig. 4-15(a).

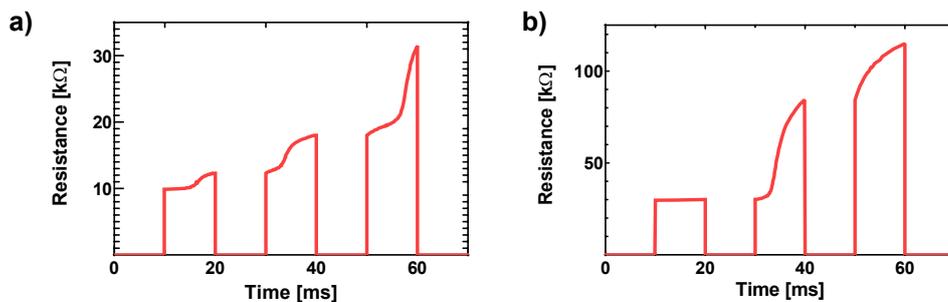


Fig. 4-13. Simulated change of resistance for (a) multi-filament model and (b) single filament model corresponding to the input voltage in Fig. 4-12.

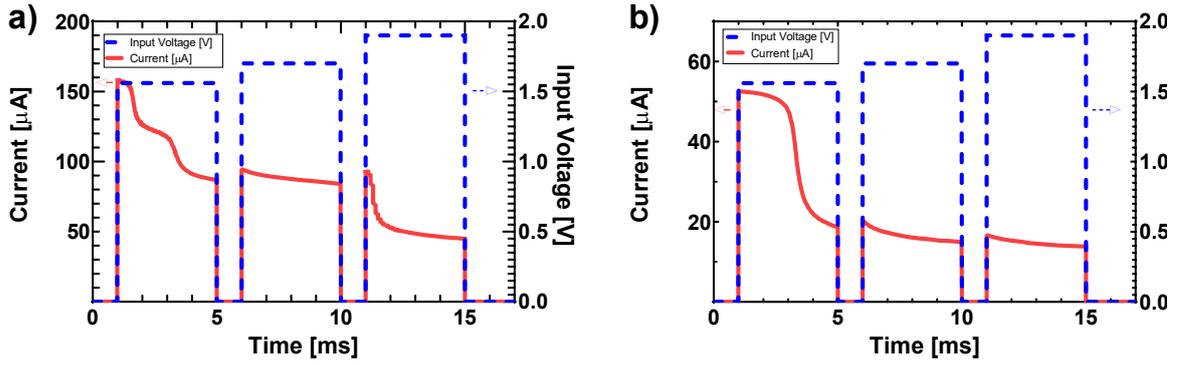


Fig. 4-14. Output current simulation results of 4ms consecutive square input voltages with varying amplitudes of  $V_1=1.56V$ ,  $V_2=1.7V$ ,  $V_3=1.9V$  for (a) multi-filament model and (b) single filament model.

Fig. 4-15(a) also shows there a gradual change in resistance during the second pulse. The third filament has not fully ruptured to provide a significant tunnel barrier. The rupture happens almost immediately at the start of the third pulse with a high resistance rate of change of  $22 \text{ M}\Omega\text{s}^{-1}$  for 0.543 ms. The 4 ms pulse scheme provides too long reset voltage exposure and is therefore unfeasible for the multi-filament model. Like in the 1ms pulse scheme, the  $n$ th pulse current is initially higher than the lowest current from the  $n$ th-1 pulse but settles into correct current levels during the pulse.

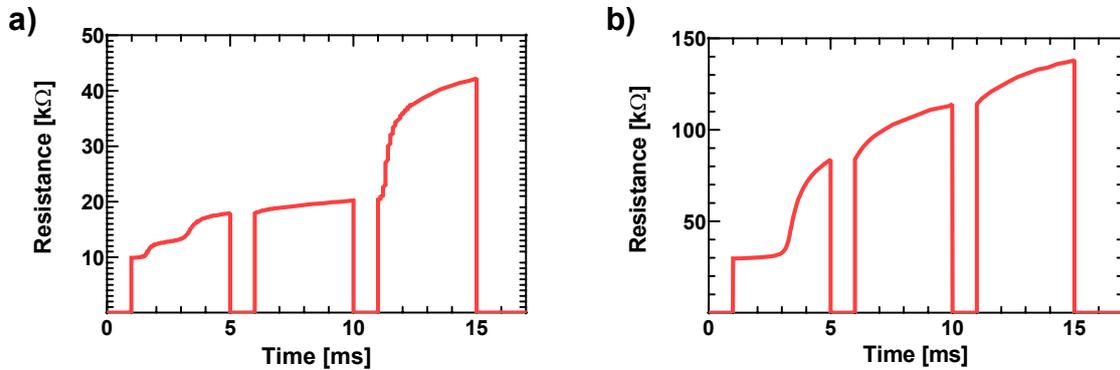


Fig. 4-15. Simulated change of resistance for (a) multi-filament model and (b) single filament model corresponding to the input voltage in Fig. 4-14.

#### 4.4 Summary

An electrical model of a multi-filament bi-layered  $\text{Ta}_2\text{O}_5/\text{TaO}_x$  ReRAM that is capable of MB switching with four resistance states ( $\text{LRS} \approx 10 \text{ k}\Omega$ ,  $\text{IRS-1} \approx 13 \text{ k}\Omega$ ,  $\text{IRS-2} \approx 20 \text{ k}\Omega$ , and  $\text{HRS} \approx 50 \text{ k}\Omega$ ) has been presented. The simulated results of the multi-filament  $\text{Ta}_2\text{O}_5/\text{TaO}_x$  ReRAM has higher current than the single filament  $\text{Ta}_2\text{O}_5/\text{TaO}_x$  ReRAM from the extra conduction paths provided by the additional filaments. Multi-bit switching is accomplished for different input voltages and it is found that filaments with lower barrier height allows easier migration

of  $V_{OS}$  and undergo the lowest  $E_{field}$  and temperature during reset switching. A simulation of two different pulse schemes demonstrated the model's versatility and an extra state is successfully achieved in one pulse scheme while the other pulse scheme was deemed unsuitable for multi-bit applications.

## 5. Chapter Five

# Single-Bit Non-Volatile LUT

---

This chapter looks into the study of nvLUTs done in this project. An analysis of a single-level nvLUT (SB-nvLUT) is first performed. The READ, WRITE, and sneak path characteristics of the SB-nvLUT is analysed. Next, a design of a voltage-mode sense amplifier (SA) circuit that is capable of sensing the low subthreshold output voltages from ReRAMs in an array is proposed. Performance characteristics of the presented SA circuit are performed and the circuit is compared to existing subthreshold current-mode and voltage-mode SAs.

## Related Publications

---

1. H. L. Chee, T. N. Kumar, H. A. F. Almurib, and D. W. H. Kang, "Analysis of a Novel Non-Volatile Look-Up Table (NV LUT) Controller Design with Resistive Random-Access Memories (RRAM) for Field-Programmable Gate Arrays (FPGA)," *Proceedings in 2019 IEEE Regional Symposium on Micro and Nanoelectronics (RSM 2019)*, pp. 87–90, 2019. (Malaysia)
2. H. L. Chee, Y. Z. Kok, T. N. Kumar, and H. A. F. Almurib, "Sense amplifier for ReRAM-based crossbar memory systems," *Taylor and Francis International Journal of Electronics Letters*, pp. 1–13, 2022.

## 5.1 Analysis of an SB-NVLUT

The controller scheme that is tested in this work was proposed in [29]. This scheme presents a novel LUT circuit design that separates the rows of a memristor array which allows better control over the ReRAM state variable during RS and addresses the sneak path current problem.

The controller scheme with ReRAMs in a 2x2 array is shown in Fig. 5-1 and the circuit diagram is shown in Fig. 5-2. There are six input pins to the controller. When the controller is

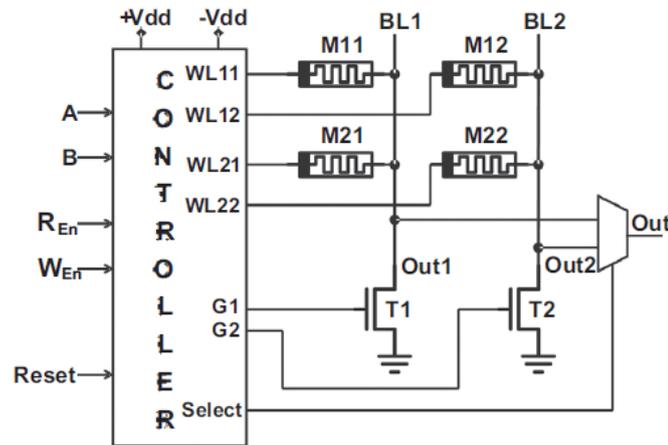


Fig. 5-1. Layout of the controller scheme taken from [13]. M11, M12, M21, and M22 are RRAMs in a 2x2 array. Each column (bitline) is separated. G1 and G2 are the same as T1 and T2 signals in TABLE 5-1.

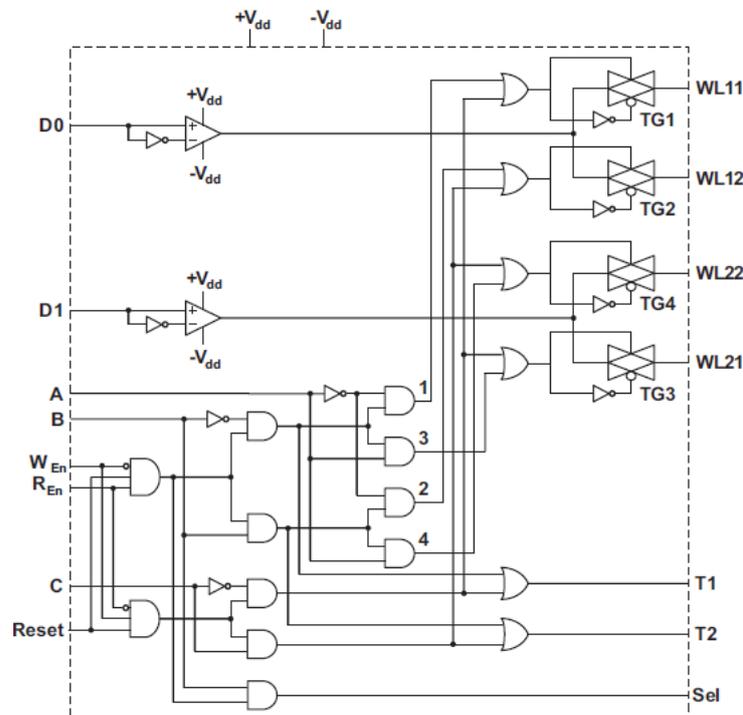


Fig. 5-2. Circuit diagram of the controller taken from [29].

TABLE 5-1. Logic Table for Controller READ and WRITE Scheme

Operation	Signals							Input at ReRAM			
	$R_{EN}$	$W_{EN}$	$C$	$A$	$B$	$T1$	$T2$	$M11$	$M12$	$M21$	$M22$
WRITE	Low	High	0	$x^a$	$x^a$	High	Low	D0	$z^a$	D1	$z^a$
	Low	High	1	$x^a$	$x^a$	Low	High	$z^a$	D0	$z^a$	D1
READ	High	Low	$x^a$	0	0	High	Low	D0	$z^a$	$z^a$	$z^a$
	High	Low	$x^a$	0	1	Low	High	$z^a$	D0	$z^a$	$z^a$
	High	Low	$x^a$	1	0	High	Low	$z^a$	$z^a$	D1	$z^a$
	High	Low	$x^a$	1	1	Low	High	$z^a$	$z^a$	$z^a$	D1

enabled, the RESET input is always high.  $R_{EN}$  and  $W_{EN}$  input pins are used to select the read and WRITE mode of the controller respectively. In READ mode, A and B input pins are used to select the memory address of the ReRAM and the output fed to the OUT pin. The input pin C is used to select the WRITE destination column during WRITE mode. WRITE operation is done parallelly for all ReRAMs in a column in this scheme. The logic table of the scheme is given in TABLE 5-1.

### 5.1.1 Results and Discussion

WRITE and READ operations of each ReRAM (M11, M12, M21, and M22) in a 2x2 array are performed and the state of the unselected RRAMs are checked for any unwanted changes (during M11 WRITE /READ, the state variables of M12, M21, and M22 are checked and so on).

The writing scheme check consists of writing ‘1’ and ‘0’ to each ReRAM. The state variable of the ReRAM model is initially at 0.2 and a RESET pulse of -2V is supplied to the device for 0.2 ms to bring the state variable to the HRS value of 0. The state variable is then capable of switching to boundary 0 (HRS) and boundary 1 (LRS). A +2V pulsewidth of 0.5ms is used for WRITE. A +2V pulsewidth of 0.1 ms is applied for READ. Reading can be done with a lower amplitude voltage to conserve power consumption and a higher value is chosen in this work is to produce a read current that is clearly visible for presentation.

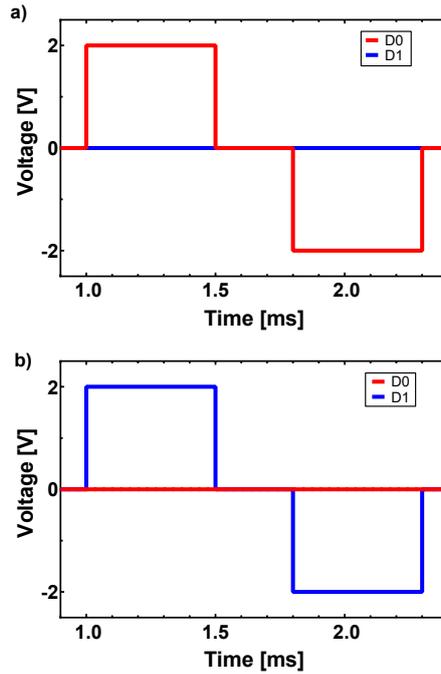


Fig. 5-3. Time evolution plot of +2V, 0.5ms WRITE 1 and 0 pulse scheme from (a) D0 line and (b) D1 line to the ReRAM.

The WRITE pulse for M11 and M21 is shown in Fig. 5-3(a) while the WRITE pulse for M12 and M22 is shown in Fig. 5-3(b). The change in the state variables of every ReRAM during are recorded and shown in Fig. 5-4. The state variable evolution during consecutive WRITE 1 and 0 for M11/M21/M12/M22 are shown in Fig. 5-4(a)/ Fig. 5-4(b)/ Fig. 5-4(c)/ Fig. 5-4(d). The

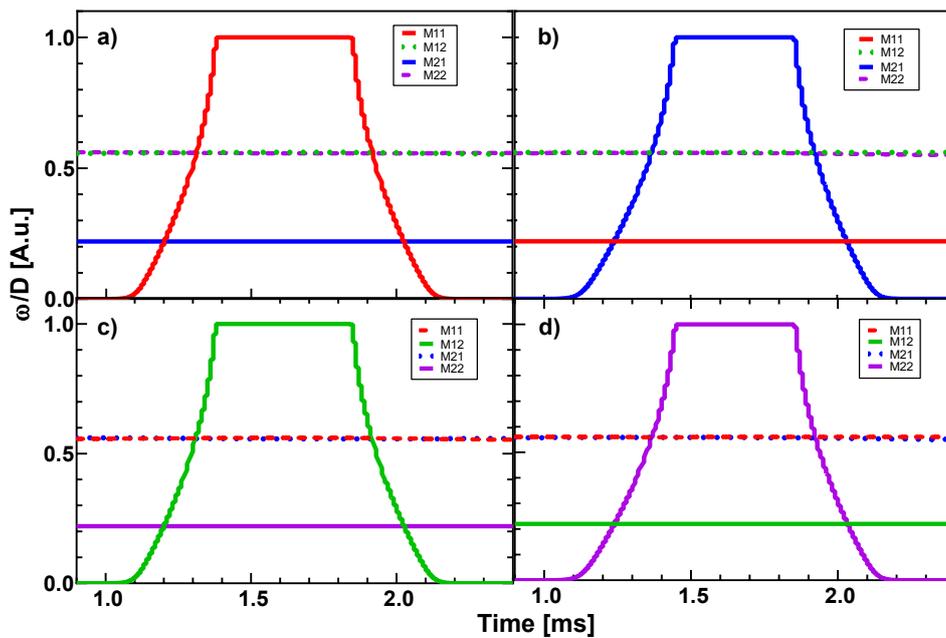


Fig. 5-4. State variable evolution for all RRAMs in a 2x2 array during WRITE 1 and 0 for (a) M11, (b) M21, (c) M12, and (d) M22. The state variable of the selected cell changes from 0 (HRS) to 1 (LRS) during WRITE 1 and returns back to 0 during WRITE 0.

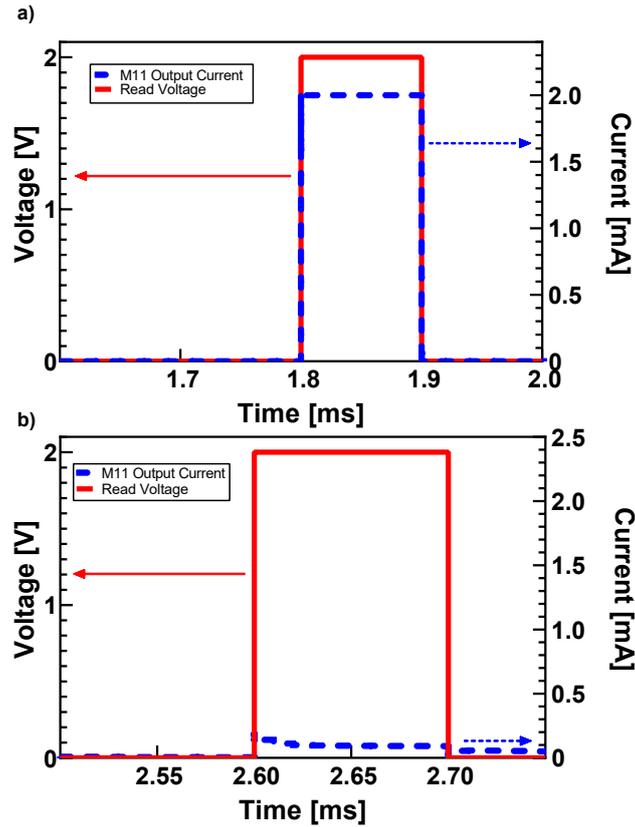


Fig. 5-5. READ output current from M11 for (a) state 1, LRS and (b) state 0, HRS. READ output current for the other RRAMs in the array demonstrate similar behaviour.

state variable of the selected ReRAM moves successfully reaches 1 during the SET pulse and 0 during the RESET pulse. The state variables of the unselected ReRAMs show almost no variation (the largest recorded variation in an unselected ReRAM was 0.003) which demonstrates the successful elimination of the sneak path current and write half-select effect.

READ testing was done for high ‘1’ and low ‘0’ states of the selected ReRAM. For read ‘1’/‘0’ the selected ReRAM is first switched to LRS/HRS before a read pulse is applied. To determine the state of the selected ReRAM, the output current from the device is read. The results for READ 1 and READ 2 are shown in Fig. 5-5(a) and Fig. 5-5(b) respectively. The output currents show successful writing and reading of the devices by the controller circuit.

This circuit provides a solution to the sneak path current problem and is viable for use in NV-based LUTs for FPGAs. Further studies on larger array sizes to increase memory density as well as circuit simulation with ReRAMs made of different materials and with different characteristics are required.

## 5.2 Sense Amplifier for SB-NVLUT

The READ voltage that is supplied to a ReRAM in an array should ideally be small enough to not induce changes to the ReRAM's resistance. Coupled with high values of  $R_{HRS}$ , a small voltage output is expected at the bitline and the sensing amplifiers. It is therefore important to design a sense amplifier (SA) that is capable of detecting the small output Bitline voltages during READ.

A voltage-mode sense amplifier capable of amplifying the output Bitline voltage and is therefore able to read a highly resistive ReRAM cell for memory crossbar systems such as LUTs is proposed and designed in this research as a solution. The sense amplifier utilizes an inverting-buffer circuit to amplify and inverse the bitline voltage before performing differential logic comparison with a reduction in transistor count that improves its READ time compared to the latch voltage-mode sense amplifier provided in [130].

This section discusses the background of the proposed work by first analyzing the ReRAM cell used in the work along with highlighting the necessity of a SA circuit for ReRAM-based memory arrays. Voltage application to the ReRAM drives the reversible and repeatable creation and dissolution of a CF in the metal oxide layer to achieve LRS or HRS with typically high resistance ratios of  $>1000$  [114]. The ReRAM model used in this work is based on MIM layered ReRAMs and demonstrate the characteristics mentioned above.

Fig. 5-6(a) shows a bipolar ReRAM cell circuit and Fig. 5-6(b) shows its switching behaviour.

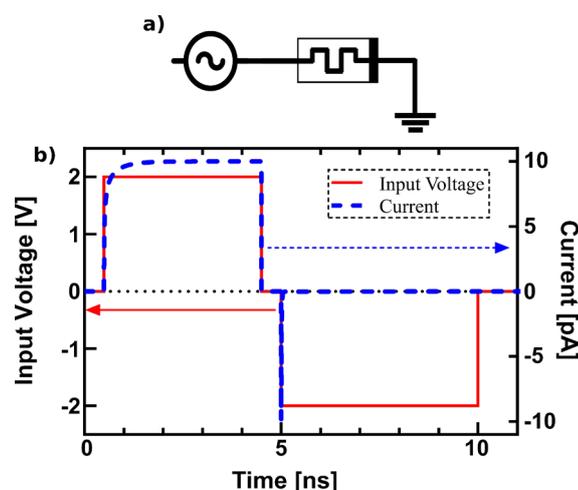


Fig. 5-6. (a) The schematic of the ReRAM circuit and (b) the input voltage fed to the circuit and the corresponding device current.

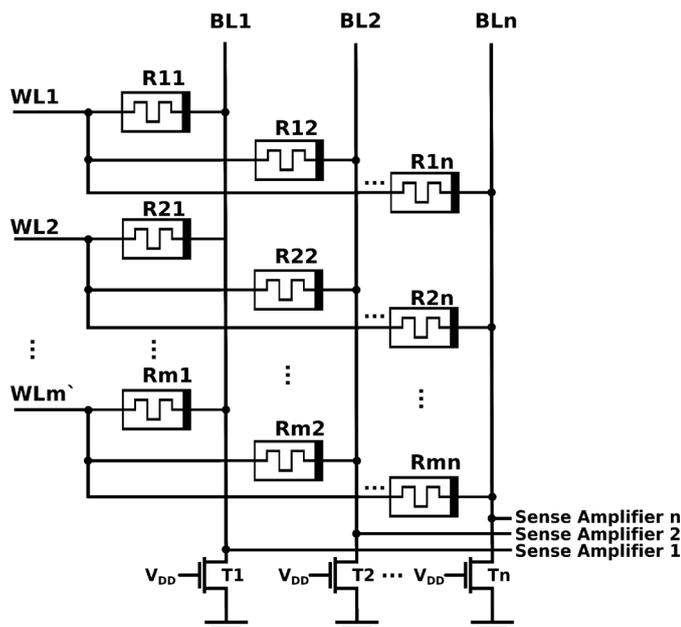


Fig. 5-7. The ReRAM crossbar array.

It switches to LRS (bit: '1') with a positive 2 V input pulse (0.5—4.5 ns) and the device current (blue dotted line) rises to  $\sim 5 \mu\text{A}$ . The device switches to HRS (bit:0) with a negative 2 V input pulse (5—10 ns) and the device current falls to  $\sim 0 \text{ A}$ .

A schematic of the typical  $m \times n$  ReRAM crossbar arrays which have been proposed as a solution to non von Neumann architectures [76] is shown in Fig. 5-7. The ReRAM cells ( $R_{mn}$ ) are connected to the wordlines ( $WL_m$ ) through their top electrodes (TEs) and are connected to the bitlines ( $BL_n$ ) through their bottom electrodes (BEs). A WL is shared by ReRAMs across a row while the BLs are shared by ReRAMs across a column.

Any individual ReRAM in the array can be selected by activating specific  $WL_m$  and  $BL_n$  address combinations, ie. memory cell R22 is selected by activating  $WL_2$  and  $BL_2$ . For the WRITE operation, the WRITE voltage is supplied through the WL while the required Bitline is activated by switching the appropriate access transistor ( $T_1$ ,  $T_2$ , and  $T_n$  in Fig. 5-7). For READ, READ voltages with significantly lower amplitude than the WRITE voltage is supplied through the WL to prevent high READ voltages to the ReRAM that will drive ionic migration in the oxide layer, risking data corruption. This leads to READ output voltages and currents in ranges that are too low to pass through the Bitlines into existing conventional SAs.

A solution proposed in [131] replaces the access transistors with transmission gates to prevent disturbance of the bitline subthreshold voltages with a drawback of microseconds READ delays. The SA proposed in [132] incorporates two capacitors to offset the subthreshold bitline voltages with an increase in SA footprint. The bitline subthreshold problem is overcome in

[133] with the use of amplifiers which lead to increase component counts. Sensitive CMOS-based level shifter circuits are an option but they have to be specially designed with increased transistor counts [134].

The following section thus proposes a voltage-mode SA that is capable of amplifying the output subthreshold bitline voltage to read a highly resistive ReRAM cell for memory crossbar systems. The SA utilizes an inverting-buffer circuit consisting of three inverters to amplify and inverse the bitline voltage before performing logic comparison with a reference voltage in the differential comparator circuit with a reduction in transistor count that improves its READ time compared to the latch voltage-mode SA provided in [130].

### 5.2.1 Proposed Sense Amplifier Design

The conditions considered for the proposed SA design are such that the READ scheme must first be designed in order to analyse the SA circuit. The READ voltage scheme should be selected such that its pulse amplitude and duration do not change the resistive state of the ReRAM. The READ scheme used in this work is a READ pulse-width voltage of 0.5V for 0.2ns (Fig. 5-6).

The state variable which represents the resistive state of the ReRAM are plotted for the HRS and LRS ReRAMs in Fig. 5-8(a) and Fig. 5-8(b) respectively. It moves between the boundaries ‘0’ and ‘1’ and the ReRAM is in LRS(HRS) when the state variable is ‘1’(‘0’).

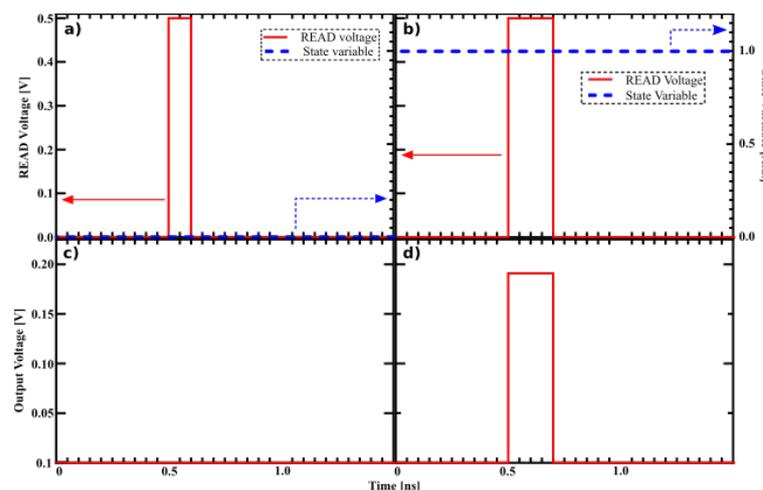


Fig. 5-8. The (a) READ voltage and (b) state variable for the HRS ReRAM and the (c) READ voltage and (d) state variable for the LRS ReRAM.

The LRS and HRS output voltages from the ReRAM during READ that goes through the Bitline and into the SA are plotted in Fig. 5-8(c) and Fig. 5-8(d) respectively. It should be noted that the voltages are very low—0V when the ReRAM is in HRS (Fig. 5-8(b)) and ~200mV

when the ReRAM is in LRS (Fig. 5-8(d)). A functioning SA has to be therefore designed around these constraints.

The complete SA circuit proposed in this work is shown in Fig. 5-9. The SA circuit is divided into two segments, an inverting-buffer and a differential logic. The main role of the inverting-buffer is to increase the low bitline READ voltage to levels required by the differential logic part of the circuit and inverting the bitline voltage. When the ReRAM is in HRS, its high resistance provides 0 V to the bitline during READ so an inverse amplified high voltage must be provided to the gate M11 of the differential circuit to switch on M11 and pull the voltage at the output node A to ground. Conversely the LRS state of the ReRAM should provide a nominal voltage to the Bitline during READ and 0 V must be provided to the gate of M11 to turn it off and keep the voltage at output node A high. The design is tested using EDA software, LTSpice with 45nm technology node PTM transistors from [118]. The electrical characteristics of these two subcircuits are discussed further as follows.

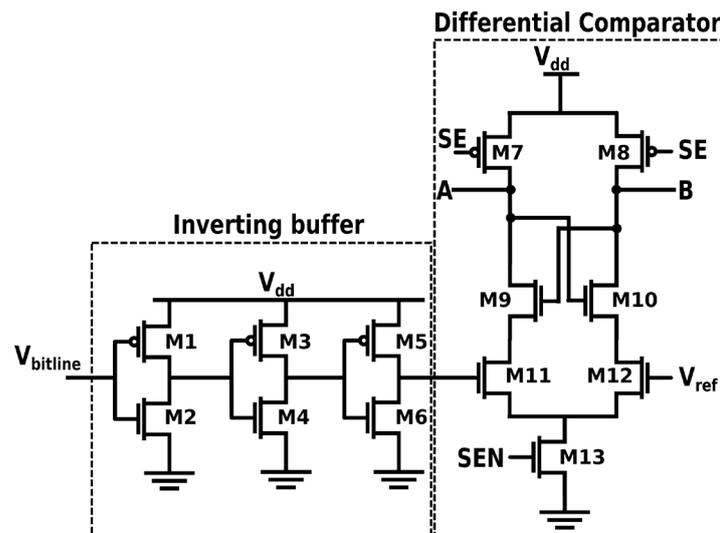


Fig. 5-9. The schematic of the proposed sense amplifier circuit. SE and SEN are input signals. A and B are the circuit output lines.

### 5.2.2 Inverting-Buffer Circuit

The inverting-buffer circuit provides a buffer that strengthens the subthreshold voltage from the ReRAM (150 mV) to switch to typical voltage of 1 V.

The buffer consists of 3 serially-connected inverters. It is crucial that the M1 and M2 transistor pair (Inverter 1) are sensitive to the subthreshold bitline READ voltage at the input. The switching point voltage,  $V_{sw}$  can be obtained from the inverter voltage transfer equation:

$$V_{SW} = \frac{V_{in} - V_{TP} + V_{TN} \sqrt{\frac{\beta_N}{\beta_P}}}{1 + \sqrt{\frac{\beta_N}{\beta_P}}} \quad (5.1)$$

where  $V_{in}$  is the input voltage,  $\beta_N$  and  $\beta_P$  are the respective NMOS and PMOS transconductance, and  $V_{TN}$  and  $V_{TP}$  are the NMOS and PMOS threshold voltages. Since the technology node sizes of the NMOS and PMOS transistors are kept the same, the aspect ratio of M1 and M2 in this work are 1 and 10 so that Inverter 1 is responsive to the subthreshold input voltage of  $\sim 200$  mV.

The electrical characteristic of Inverter 1 is plotted in Fig. 5-10. The input voltage of  $\sim 200$  mV is inverted to  $\sim 0.7$  V while the output voltage maintains a steady 1.1 V when there is no input voltage. The output of Inverter 1 has to be further rectified as it only drops to  $\sim 0.7$  V. The output is thus fed to Inverter 2 (M3//M4 pair in Fig. 5-9) with an aspect ratio of 20 for M3 and 1 for M4. The higher PMOS aspect ratio allows Inverter 2 to fully swing its output to 1.1 V with a 0.7 V input. Inverter 3 (M5//M6), with an aspect ratio of 2 for M5 and 22 for M6 then inverts the Inverter 2 output voltage for the differential circuit.

The outputs from all three inverters with a linear input voltage are plotted separately in Fig. 5-11. The linear input voltage increases from 0 V to 200 mV to match the expected bitline voltage from the ReRAM, Inverter 1, Inverter 2, and Inverter 3. The lower aspect ratio of Inverter 2 reduces the inverter capacitances and provide a sharper switching as compared to Inverter 1.

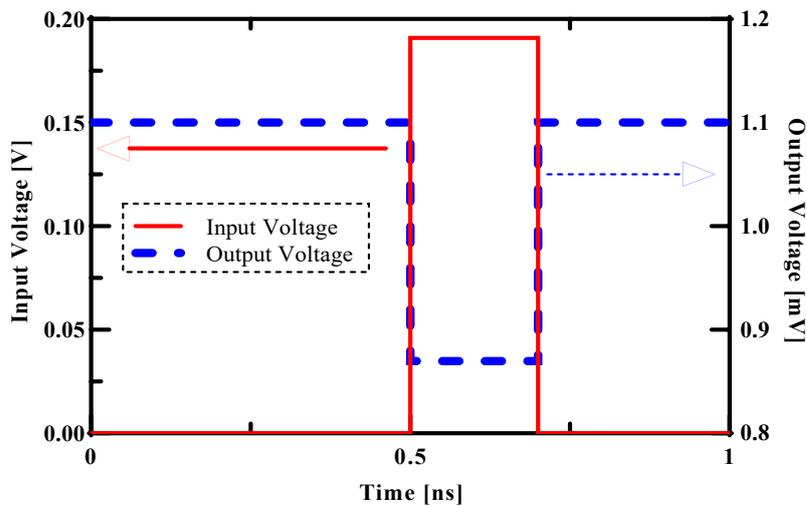


Fig. 5-10. The input and output voltage waveforms of the inverting-buffer segment (M1, M2). The output voltage is 1.1V when the input is 0V and the output falls to around 850mV when the input voltage rises to  $\sim 200$ mV, the maximum Bitline voltage from the ReRAM memory cell.

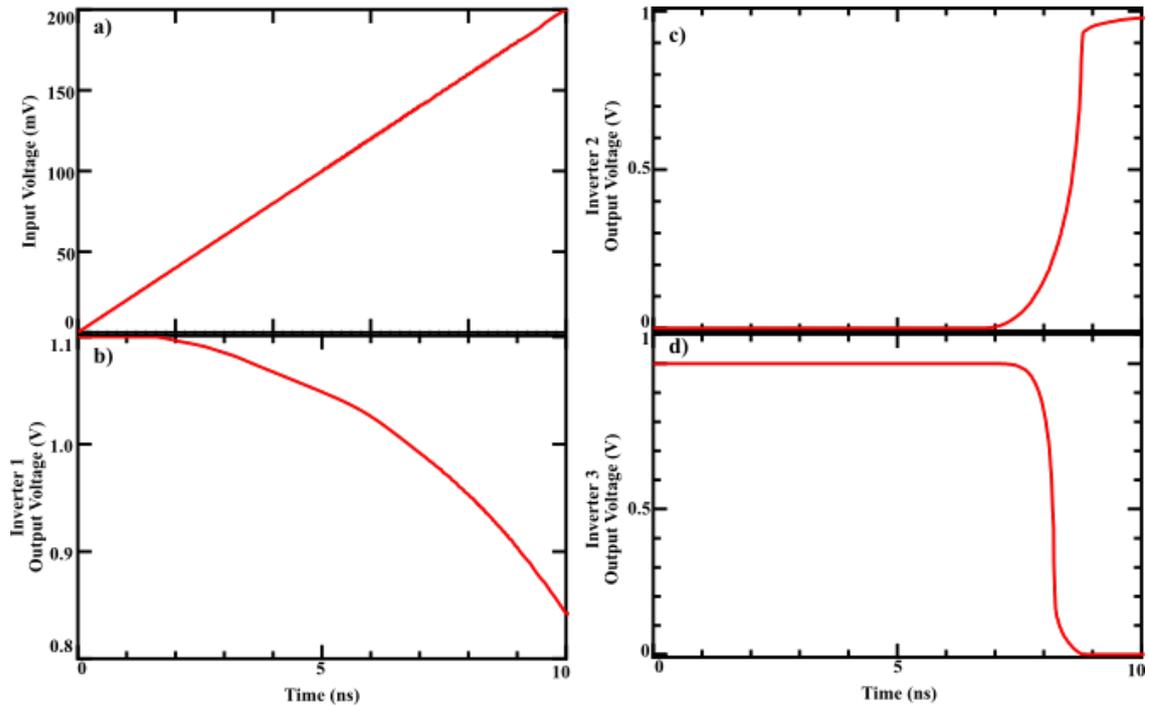


Fig. 5-11. The input and output voltage waveforms of each inverter in the inverting-buffer circuit. (a) The linear input voltage from 0V to 200mV to simulate the maximum READ bitline voltage and the response to the input voltage of the (b) first inverter pair (M1, M2), (c) second inverter pair (M3,M4), and (d) the third inverter pair (M5,M6).

Additionally, when there is no READ command and the SA is turned off, the bitline voltage does not reach the differential comparator circuit which prevents unwanted outputs from the SA.

### 5.2.3 Differential Comparator Circuit

The differential comparator segment comprises of the circuit to the right of the inverting buffer segment in Fig. 5-9 and shown here in Fig. 5-12. Transistors M7 and M8 form the precharge sections of the differential comparator circuit. These transistors are precharged before the READ operation begins. The drains of transistors M7 and M8 are connected to the gates of transistor M10 and M9 as well as nodes A and B respectively. As the precharge voltage is above the threshold voltages of M9 and M10, these transistors are switched on and nodes A and B are charged to  $V_{dd}$ .

The source of transistors M9 and M10 are connected to the drains of input transistors, M11 and M12. Depending on the input fed to the gate of M11 and M12, different voltages,  $V_A$  and  $V_B$  are achieved at node A and node B. In this configuration, the input for M12 (gate voltage of M12) is connected to a reference voltage,  $V_{ref}$  which is always off while the input for M11 (gate voltage of M11) is connected to the output of the inverting-buffer circuit.

The working principle is that when M12 is off and M11 is fed with an input from a memory,

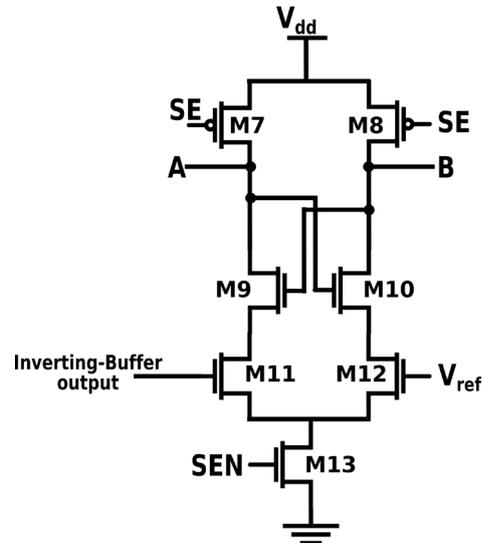


Fig. 5-12. The differential comparator circuit.

then, M12 stays off and M11 conducts. This means that the pre-charged voltage at node A goes through transistor M9 and M11 and is pulled to the ground. At the same time, the voltage at node A now falls to 0 V. Since, node A is also connected to the gate of transistor M10, when the voltage=0 V then, the transistor M10 is off and the pre- charged voltage,  $V_B$  at node B is not pulled to ground. Therefore, when the voltage is read at node B, the voltage is shown to be at pre-charged voltage.

Node B therefore acts as a reference while node A is the output of the SA. When, the voltage at node A is at 0 while the voltage at node B is at pre-charged voltage, the SA is reading a Boolean '0' and the reading is Boolean '1' if both nodes are at pre-charged voltages.

The electrical characteristics of the differential comparator circuit are plotted in Fig. 5-13. In Fig. 5-13(a) the switching voltage is the voltage that switches the SA circuit on. The SA circuit is switched on twice, first from 0.5 ns to 1 ns and from 1.5 ns to 2 ns. Two conditions are simulated during these two periods, the first one where the input at the M11 transistor's gate is 0 V and the next one when the input is raised to 1 V (Fig. 5-13(b)), values chosen to match the expected output voltage from the Inverter 3 in Fig. 5-9. The output of the SA is plotted in Fig. 5-13(c). As shown the voltages at both nodes remain high when there is no input at the M11 gate, indicating the ReRAM is in LRS and giving the Boolean logic 1. When the M11 gate voltage is increased to 0.2 V the voltage at node A,  $V_A$  falls to 0 V giving a reading of Boolean logic '0'.

Fig. 5-14 shows the waveform for the operation of the complete SA with the sequence given below:

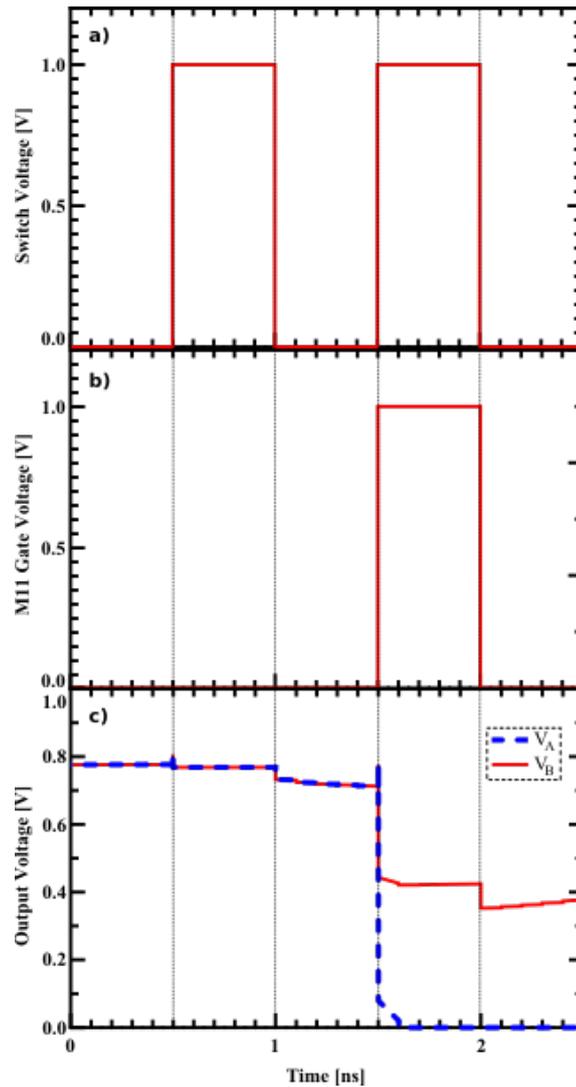


Fig. 5-13. The (a) switch voltage, (b) gate voltage for transistor M11, and (c) the output voltage at nodes A and B from the differential comparator circuit.

- (1) Transistors M5 and M6 are turned on by signals SE and precharge nodes A and B.
- (2) SEN signal is fed to M7 and turns on the circuit.
- (3) Output voltage from the ReRAM goes through the Bitline into the inverting buffer and is fed to M1. A reference voltage is always fed to M2.
- (4) The voltage at nodes A and B are detected to provide a reading of the memory cell.

The working SA is then connected with a 2x2 ReRAM crossbar array in the following section.

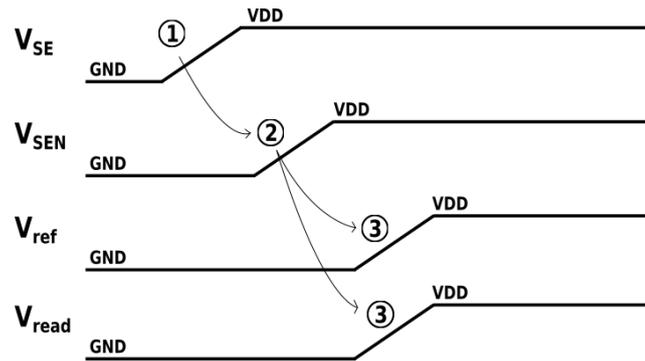


Fig. 5-14. The operation waveforms for the sense amplifier circuit.

### 5.2.4 Simulation and Results

The SA circuit is then connected with a ReRAM memory array and simulations are conducted for analysis. READ high and low operations are carried out and the delay and energy consumption are recorded and compared with [130].

Fig. 5-15 shows the schematic of the ReRAM crossbar array used for the simulation. One SA is connected to each bitline. To select the cell for writing or reading a WRITE or READ voltage is supplied to the Wordline and the NMOS bitline selector (T1 or T2) is activated. The output voltage from the ReRAM is then fed to the SA circuit.

The ReRAM cell is supplied with a pulswidth WRITE voltage of 2 V, 4ns for LRS and 2 V, 5 ns for HRS. For READ, the scheme mentioned in Section 5.3.4 of 0.5 V pulswidth of 0.2 ns is used.

The ReRAMs are first written into either HRS (Boolean: ‘0’) or LRS (Boolean: ‘1’) before the READ voltage is supplied. The output is then read at the nodes A and B of the SA. The results of high and low READ operations are plotted in Fig. 5-16(a) and Fig. 5-16(b) respectively. The SA correctly detects the states of the ReRAM memory with the voltage at

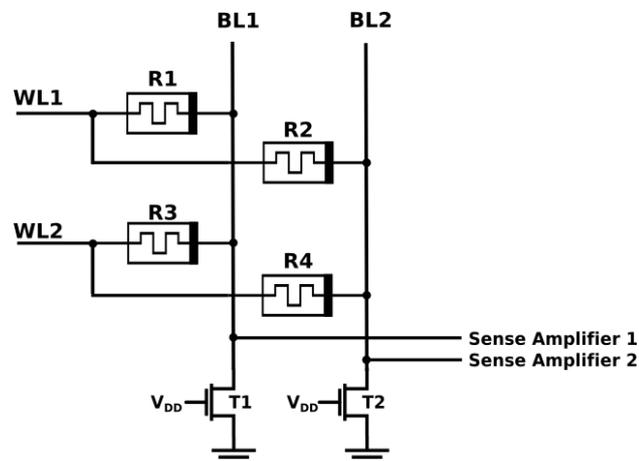


Fig. 5-15. The memory crossbar array used in the simulation. One sense amplifier is connected to one bitline.

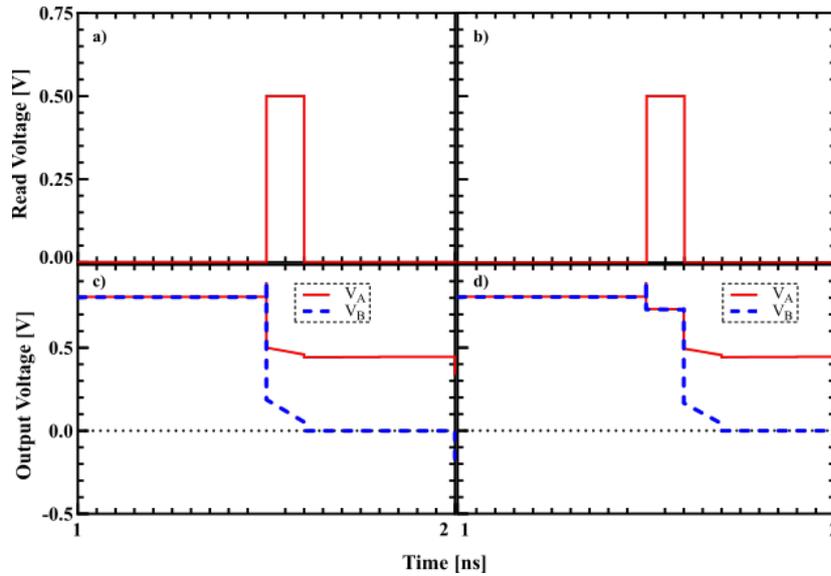


Fig. 5-16. READ (a) Boolean low (0) and (b) Boolean high (1) waveform for a 2ns 0.1V READ pulse.

TABLE 5-2. Read Delay and Energy Comparison

	<b>Voltage-mode circuit</b> [130]	<b>Current-mode circuit</b> [130]	<b>Proposed Circuit</b>
READ delay [ps]	112	516	76
READ energy [ $\mu$ W]	1620	1840	749.9

node A dropping when the ReRAM HRS (Boolean: ‘0’) is read (Fig. 5-16(c)) and remaining high when reading LRS (Boolean: ‘1’) (Fig. 5-16(d)).

The READ delay and energy consumption are listed in TABLE 5-2 and are compared with the delay and energy of voltage-mode and current-mode SAs from [15]. The proposed SA shows better performance in both categories. This is due to the usage of only two pull-up transistors (M7 and M8 in Fig. 5-9) in the precharge segment of the circuit compared to four pull-up transistors in the referred circuit.

### 5.2.5 Summary

A simulation of the controller circuit was carried out and the circuit is capable of switching a selected ReRAM without disturbing the other RRAMs in a 2x2 array. This circuit provides a solution to the sneak path current problem and is viable for use in NV-based LUTs for FPGAs.

A design of an SA circuit that is capable of reading the low Bitline voltages in a typical ReRAM memory crossbar, showing 36% improvement in READ delay and 57.3% in READ energy consumption. It is of great interest to increase the density of memory arrays and MB-ReRAM arrays are currently being investigated [76], [135]. MB RS has smaller resistance

windows which further reduces the Bitline voltage margins during READ. Since the inverter-buffer SA is already capable of subthreshold switching, an opportunity exists to tailor this work for MB-ReRAM arrays.

## 6. Chapter Six

# **Multi-Bit Non-Volatile LUT**

---

This chapter will look into the study of nvLUTs done in this project. A design of a novel MB-nvLUT array is presented together with a novel controller for the MB-nvLUT capable of accommodating the requirements of the MB per cell MB-nvLUT. Further characterization of the MB-nvLUT is performed with analysis of the WRITE and READ delay timings, energy dissipation, and EDP which are then compared with similar metrics of the SB-nvLUT. FPGA benchmark tests are then performed for the MB-nvLUT and SB-nvLUT.

## **Related Publications**

---

1. H. L. Chee, T. N. Kumar, and H. A. F. Almurib, "Low energy non-volatile look-up table using 2 bit ReRAM for field programmable gate array," *IOP Semiconductor Science and Technology*, vol. 37, no. 6, 2022.

## 6.1 The MB-ReRAM

The ReRAM model used in this section is a generic model from [136]. The model relates the ionic motion to a state variable that moves between two boundaries to obtain the HRS and LRS. The boundary characteristics are controlled by limiting window functions and the RS can be tuned to be gradual or abrupt. The model also allows for adjustment of parameters to match experimental results.

In this work, changes are made to the fitting parameters for the IV relationship:  $a_1$ ,  $a_2$ , and  $b$ , for the positive-negative voltage thresholds:  $V_p$  and  $V_n$ , for the state variable motion multiplier:  $A_p$  and  $A_n$ , and the SV decay motion rates:  $\alpha_p$  and  $\alpha_n$ . The parameter values are listed in TABLE 6-1.

TABLE 6-1. Model Parameters

<i>Parameter</i>	<i>Value</i>	<i>Symbols</i>
$a_1$	0.1	-
$a_2$	0.1	-
$b$	5.00E-12	-
$V_p$	0.15	V
$V_n$	0.16	V
$A_p$	1.00E+11	-
$A_n$	8.50E+09	-
$\alpha_p$	1.00E+00	-
$\alpha_n$	5.00E+00	-

Parameters  $a_1$ ,  $a_2$ , and  $b$  affect the device conductivity through

$$I(t) = \begin{cases} a_1 x(t) \sinh(bV(t)), & V(t) \geq 0 \\ a_2 x(t) \sinh(bV(t)), & V(t) < 0 \end{cases} \quad (6.1)$$

where  $x(t)$  is the state variable modelled by the functions

$$f(x) = \begin{cases} e^{-\alpha_p(x-x_p)} \omega_p(x, x_p), & x \geq x_p \\ 1, & x < x_p \end{cases} \quad (6.2)$$

$$f(x) = \begin{cases} e^{\alpha_n(x+x_n-1)} \omega_n(x, x_n), & x \leq 1 - x_n \\ 1, & x > 1 - x_n \end{cases} \quad (6.3)$$

The function  $f(x(t))$  describes the motion of the state variable until it reaches the threshold points  $\omega_p$  and  $\omega_n$ .  $\omega_p$  and  $\omega_n$  consist of windowing functions that provides a decay to the exponential when the thresholds are reached. They are described by the equations

$$\omega_p(x, x_p) = \frac{x_p^{-x}}{1-x_p} + 1 \quad (6.4)$$

$$\omega_n(x, x_{pn}) = \frac{x}{1-x_n} \quad (6.5)$$

The state change is also governed by the function  $g(V(t))$  which provides a hyperbolic sinusoidal motion once the voltage threshold is exceeded. This function is given as

$$g(V(t)) = \begin{cases} A_p(e^{V(t)} - e^{V_p}), & V(t) > V_p \\ -A_n(e^{-V(t)} - e^{V_n}), & V(t) < V_n \\ 0, & V_n \leq V(t) \leq V_p \end{cases} \quad (6.6)$$

The rate of the change of the state variable,  $x$  is then given as

$$\frac{dx}{dt} = g(V(t)) f(x(t)) \quad (6.7)$$

These changes allow the MB-ReRAM to have four resistance levels and two-bits-per-level; LRS='11', IRS-1='10', IRS-2='01', and HRS='00'. In contrary the 1-bit SB-ReRAM only consists of 2 levels with one-bit-per-level; LRS='1' and HRS='0'. This ReRAM model is bipolar and switches to LRS (HRS) with positive (negative) supply voltage.

The input voltage scheme in this work is pulse-width with parameters as follows:

- For SB switching: 1V for 0.4 ns for LRS switching, -0.5 V for 0.46 ns for HRS switching
- For MB switching: 1V for 0.4 ns for LRS switching, -0.5 V for 0.06 ns for IRS-1 switching, -0.5 V for 0.1 ns for IRS-2 switching, -0.5 V for 0.3 ns for HRS switching

Lower voltage is used for HRS switching to allow precise control of switching into the MB states. The input voltage and the state variable motion are plotted in Fig. 6-1(a) and Fig. 6-1(b).

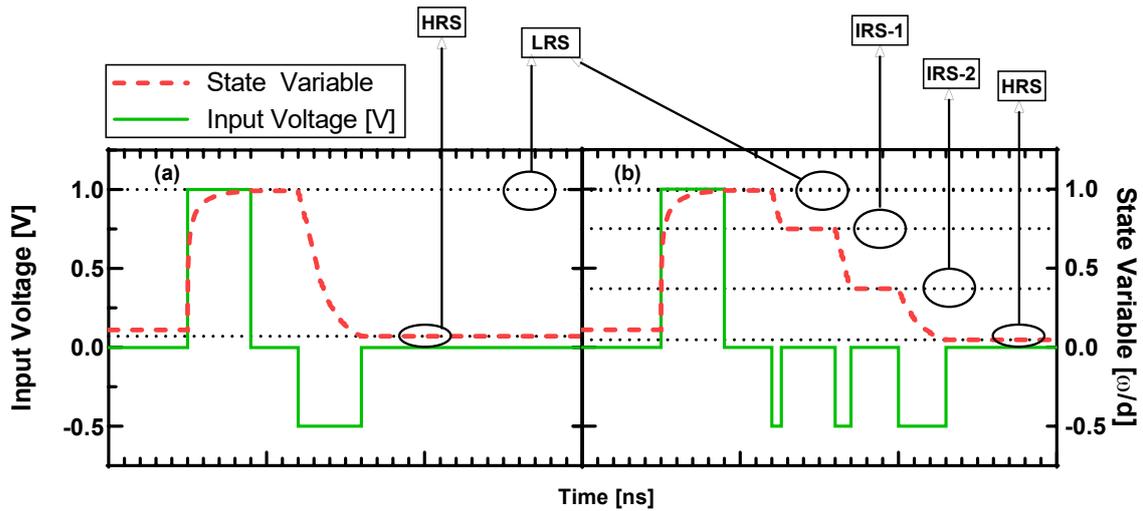


Fig. 6-1. Switching behaviour comparison between SB-switching (left) and MB-switching (right).

It should be noted that the state variable of the model does not begin at 0 as a requirement for the electrical model is that the initial state does not begin at the boundary.

The current and power consumption for each SB and MB resistive state in this switching scheme are plotted in Fig. 6-2. The maximum current and power consumption during the positive voltage cycle for both SB and MB devices is  $\sim 440$  fA (Fig. 6-2(a) and Fig. 6-2(b)) and  $\sim 500$  fJ (Fig. 6-2(c) and Fig. 6-2(d)). They are the same for both as there is no IRS in the

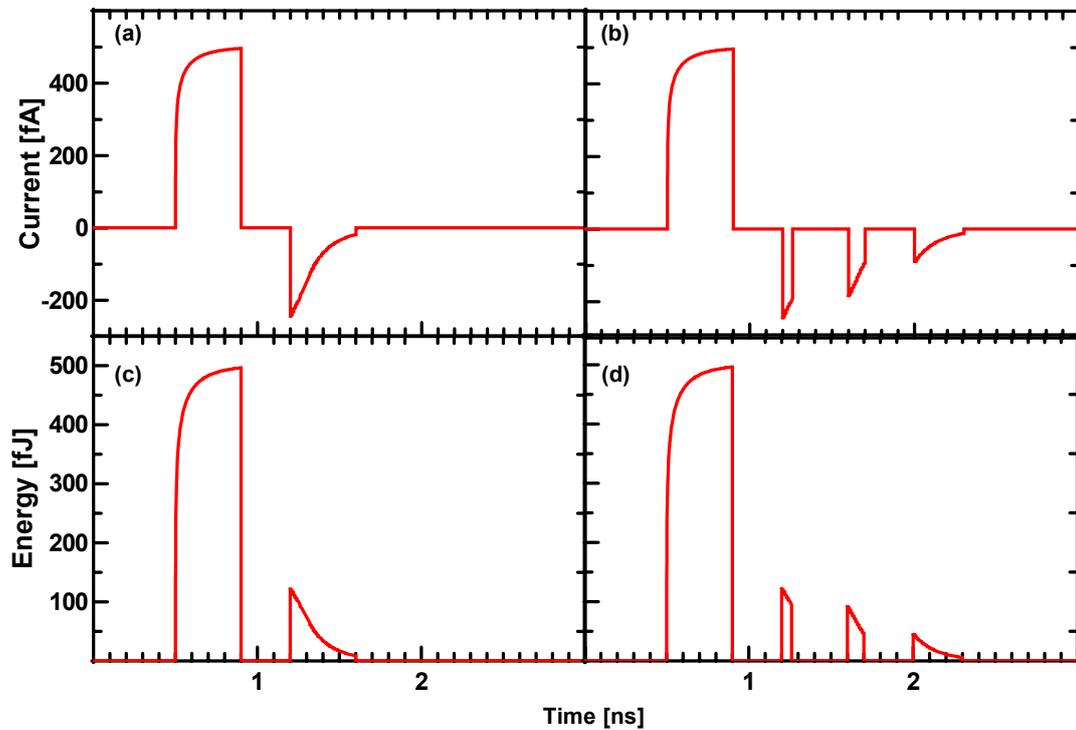


Fig. 6-2. Comparison of device current for SB (a) and MB (b) and power consumption for SB (c) and MB (d).

positive cycle.

As for HRS switching, the maximum device current for both the SB and MB-ReRAM (Fig. 6-2(a) and Fig. 6-2(b)) is  $\sim 200$  fA as they are in similar LRS states before switching. The total power consumption for HRS switching is also similar for both SB and MB-ReRAMs at 4.7 pJ. This is because the voltage supply time is the same for both; 0.46 ns HRS switching for SB-ReRAM and 0.06+0.1+0.3 ns HRS switching for the MB-ReRAM. It should be noted that although they both show the same latency and energy consumption, the MB-ReRAM is switching two bits while the SB-ReRAM is only switching one bit ('1'→'0' for SB and '11'→'00' for MB). The MB-ReRAM's advantage over the SB-ReRAM in circuit applications will be discussed further in Section 6.4.

## 6.2 The MB-ReRAM LUT Array

The MB-nvLUT array schematic is given in Fig. 6-3. Each ReRAM cell ( $M_n$  where  $n \geq 1$ ) is connected to the controller through a wordline and to the output transistors (TG $_n$ ) through a bitline. The WRITE and READ voltages are supplied through the wordline while the bitline voltages are outputs that are fed to an external READ decoder. To select a ReRAM cell for READ/WRITE operations, the controller activates the wordline internally and the bitline through the TG $_n$  transistors.

In this setup, the MB-ReRAM cell contains two memory address locations while the four resistive states represent the four possible output-bit combinations for the two addresses. For example, the addresses AB='00' and AB='01' are now located in ReRAM cell M1 (Fig. 6-3). Now a single address in a conventional LUT can store an output bit of 0 or 1 so that the output bit for address AB='00' can be either '0' or '1'. This is the same for address AB='01'. From this we obtain four possible output-bit combinations for the two addresses in a cell and list them in TABLE 6-2. TABLE 6-2 is the logic table for a 2-input MB-nvLUT with four memory address divided into two ReRAM cells (M1 and M2). The Memory Address A and B columns correspond to the A and B inputs of a conventional LUT. M1 now contains the addresses AB='00' and AB='01' as described above while M2 contains the addresses AB='10' and AB='11', accounting for all the memory locations for a 2-input LUT.

The output bits of the two addresses located in a single cell depends on the four resistive states of the ReRAM cell namely HRS, IRS2, IRS1 and LRS. As mentioned above we have four possible output-bits combinations for two addresses. Now if we take the output-bit of

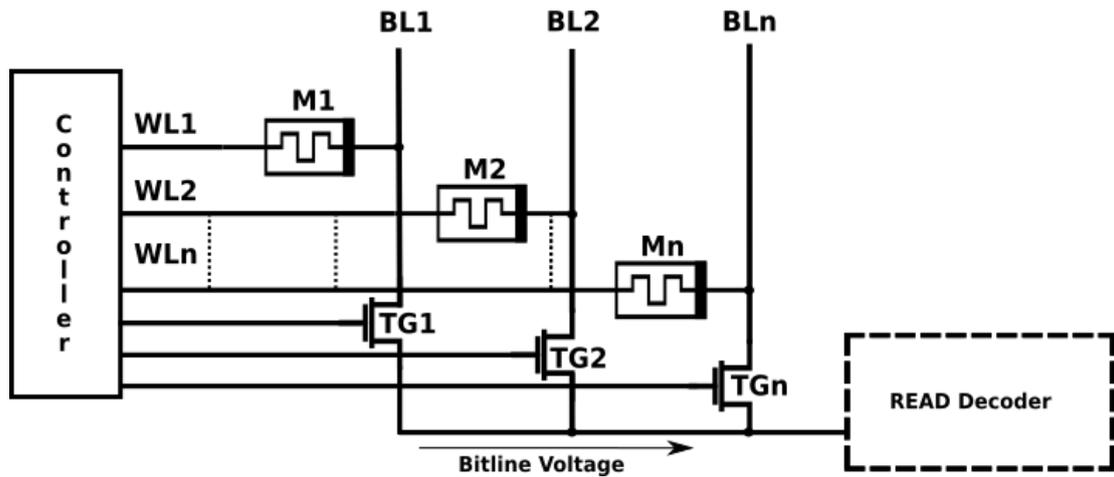


Fig. 6-3. The MB-ReRAM array. The schematic of the Controller block is given in Fig. 6-5 while the external READ decoder is given in Fig. 6-4.

TABLE 6-2. Array Logic Table for MB-nvLUT

Cell	Memory Address		Output				
	A	B	HRS	IRS-2	IRS-1	LRS	
M1	0	0	0	0	1	1	<i>x</i>
	0	1	0	1	0	1	<i>y</i>
M2	1	0	0	0	1	1	<i>x</i>
	1	1	0	1	0	1	<i>y</i>

AB='00' as *x* and the output-bit of AB='01' as *y* then the possible combinations of *xy* are '00', '01', '10', '11'. A single *xy* combination is then represented by a single resistive state. Referring to TABLE 6-2, when the ReRAM is in HRS the *xy* output bit of the cell is 00 which indicates that the output bit for address AB='00' is '0' and the output-bit for address AB='01' is 0. In another example, when the ReRAM is in IRS-1 the *xy* output-bit of the cell is 10 which indicates that the output-bit for address AB='00' is 1 and the output-bit address AB='01' is '0'.

The logic table for the SB-nvLUT is given in TABLE 6-3 where a single ReRAM cell contains only one memory address location and one output-bit (output-bit can be either '0'=HRS or '1'=LRS for each *M<sub>n</sub>* cell) [29]. By storing the outputs of 2 memory addresses per cell, the number of cells in the MB-nvLUT array is now  $2^{n-1}$  compared to the  $2^n$  of the SB-nvLUT.

To READ the stored output-bits, the analog bitline voltage from a selected cell is fed to an external READ decoder where it is compared with four reference voltages to determine the resistive state of the ReRAM. The decoder then feeds out the appropriate digital output bit. Due to the MB design where each ReRAM cell now holds two addresses the decoder can

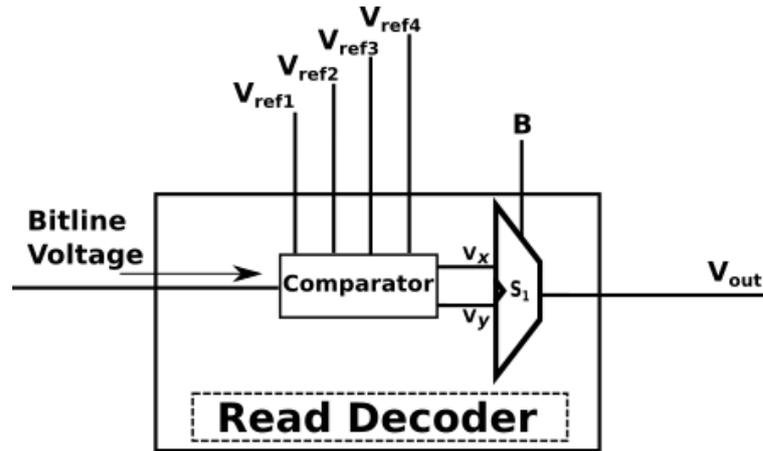


Fig. 6-4. Block diagram of the READ controller circuit designed for two input MB-nvLUT.

TABLE 6-3. Array Logic for SB-nvLUT

Cell	Memory Address	Output	
		HRS	LRS
M1	00	0	1
M2	01	0	1
M3	10	0	1
M4	11	0	1

potentially read two addresses at a time. The block diagram for the designed READ decoder is shown in Fig. 6-4 and the READ process will be discussed in further detail in Section 6.3.2.

### 6.3 The MB-LUT Controller

To implement an MB-NVLUT, a controller circuit has been designed (Fig. 6-5(a)) to accommodate the different writing and reading requirements of the MB-nvLUT cells. The controller is made up of multiple controller blocks. In Fig. 6-5(a), the Controller Block 1 is the layout for the first block in the controller with input from an external input decoder and signals DATA1, DATA2, DATA3, DATA4,  $W_{EN}$ , and  $R_{EN}$ . This block controls two ReRAM cells through the wordlines, WL1 and WL2 and output signals, G1 and G2.

The input decoder decodes the conventional LUT-input-addresses so the controller block can select the correct memory cell during READ or WRITE as the array now consists of two-addresses-per-cell. The method used to assign the conventional LUT-input-addresses in this design is explained below in Sections 6.3.1 and 6.3.2.

Controller block 1 alone works as a 2-input LUT. WL1 (WL2) and TG1 (TG2) are selected depending the signal from the input decoder following the logic in TABLE 6-2 where addresses beginning with A='0' are assigned to M1 and addresses with A='1' are assigned to M2. This

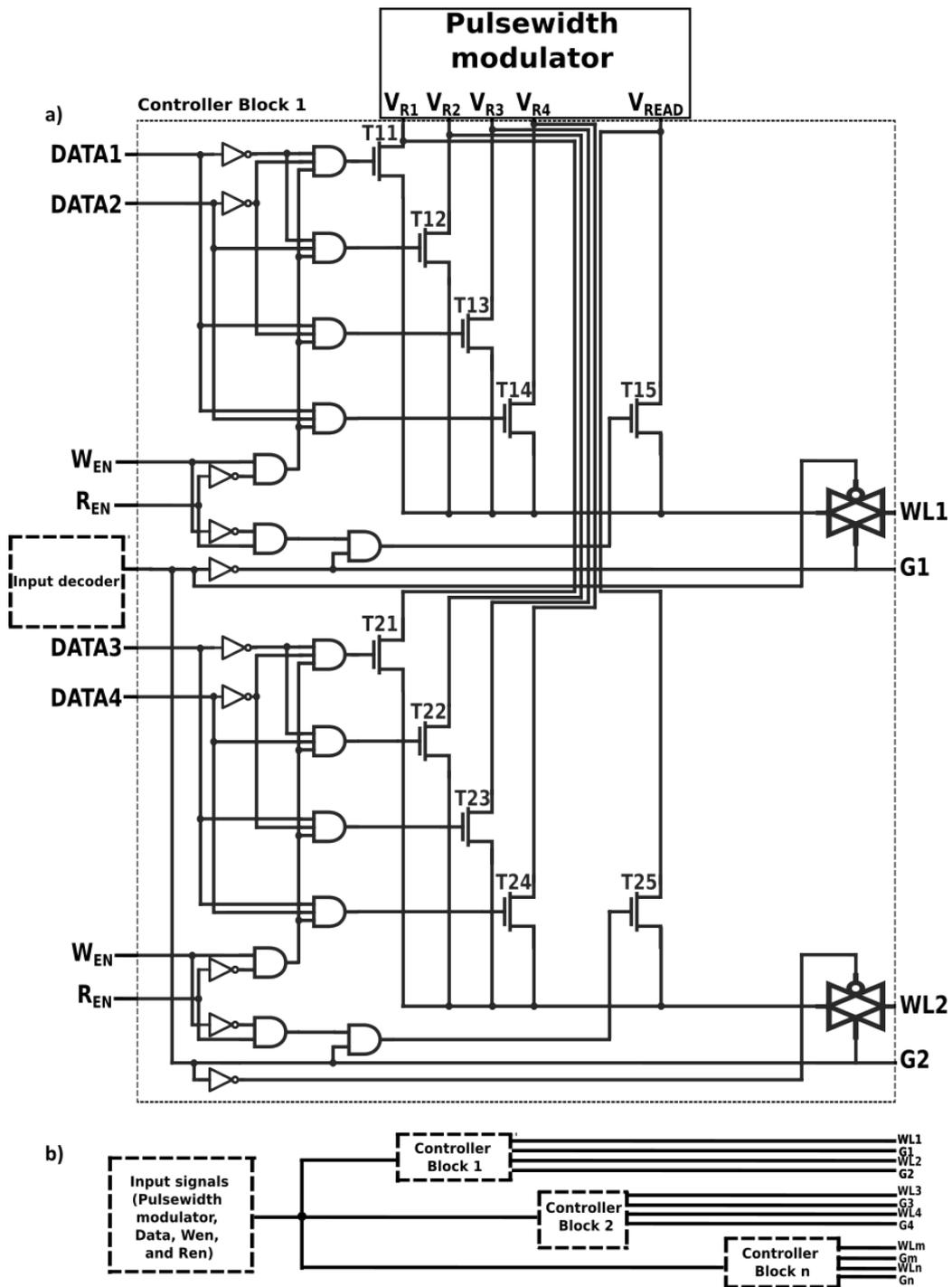


Fig. 6-5. (a) Schematic of the controller block designed for MB-NVLUT. This block is able to receive 2 LUT inputs. (b) Additional controller blocks are used for higher-input-LUTs.

ensures a holistic cell select by controlling the wordline and bitline while the unselected wordline and bitline remains in a high impedance state.

The lines DATA1 and DATA2 are the  $x$  and  $y$  for M1 while DATA3 and DATA4 are the  $x$  and  $y$  for M2 and receive the output bits that are to be written before activating the appropriate WRITE transistor, T11-T14 and T21-T24.  $W_{EN}$  and  $R_{EN}$  are the WRITE and READ enable

pins respectively.  $R_{EN}$  activates the READ transistor T15 or T25 depending on the selected address to be read. The WRITE and READ logic for the controller circuit is given in TABLE 6-4 and TABLE 6-5 for M1 and M2 respectively. Additional controller blocks are used for higher-input-LUTs (Fig. 6-5(b)). For example, eight ReRAM cells are required for a 4-input MB-nvLUT. Since there are two cells per controller block, the 4-input LUT controller will consist of four controller blocks.

### 6.3.1 WRITE Operation

The WRITE operation is started by asserting the  $W_{EN}$  signal and the input at A selects either WL1 or WL2. As shown in TABLE 6-4 and TABLE 6-5, the DATA1 and DATA2 (DATA3 and DATA4) signals then select which resistance state the cell M1(M2) is written into. It should be noted their values correspond to the outputs of the chosen AB address; for instance, in TABLE 6-2, DATA1(x) and DATA2(y) are ‘0’ and ‘1’ respectively if the output bits for address AB=‘00’ and AB=‘01’ are ‘0’ and ‘1’.

Depending on the DATA1(x) and DATA2(y) (DATA3(x) and DATA4(y)) values, one of the transistor gates T11, T12, T13, or T14 (T21, T22, T23, or T24) are then activated to feed one of the WRITE voltages  $V_{R1}$ ,  $V_{R2}$ ,  $V_{R3}$ , or  $V_{R4}$ . These WRITE voltages follow the WRITE scheme in Fig. 6-1 where  $V_{R1}$  is the HRS voltage,  $V_{R2}$  is the IRS-2 voltage,  $V_{R3}$  is the IRS-1

TABLE 6-4. Controller READ and WRITE Logic Table for M1

Operation	Signals											Input at ReRAM	
	$R_{EN}$	$W_{EN}$	DATA1	DATA2	IM1	IM2	T11	T12	T13	T14	T15	M1	M2
WRITE	L	H	0	0	H	L	H	-	-	-	-	VR1	-
	L	H	0	1	H	L	-	H	x	-	-	VR2	-
	L	H	1	0	H	L	-	-	H	-	-	VR3	-
	L	H	1	1	H	L	-	-	-	H	-	VR4	-
READ	H	L	-	-	H	L	-	-	-	-	H	Vread	-

H is High, L is Low, - is don't care

TABLE 6-5. Controller READ and WRITE Logic Table for M2

Operation	Signals											Input at ReRAM	
	$R_{EN}$	$W_{EN}$	DATA1	DATA2	IM1	IM2	T21	T22	T23	T24	T25	M1	M2
WRITE	L	H	0	0	H	L	H	-	-	-	-	-	VR1
	L	H	0	1	H	L	-	H	x	-	-	-	VR2
	L	H	1	0	H	L	-	-	H	-	-	-	VR3
	L	H	1	1	H	L	-	-	-	H	-	-	VR4
READ	H	L	-	-	H	L	-	-	-	-	H	-	Vread

H is High, L is Low, - is don't care

voltage, and  $V_{R4}$  is the LRS voltage.

### 6.3.2 READ Operation

The READ operation begins when the  $R_{EN}$  signal is asserted. The A input activates the either one of the Wordline WL1 or WL2 and pass transistors G1 or G2 (Fig. 6-3 and Fig. 6-5). The bitline READ voltage is then fed to the comparator in the READ Decoder and compared with four reference voltages that correspond to the four resistive states and four combination output bits— $V_{ref1}$  is HRS ( $xy='00'$ ),  $V_{ref2}$  is IRS-2 ( $xy='01'$ ),  $V_{ref3}$  is IRS-1 ( $xy='10'$ ), and  $V_{ref4}$  is LRS ( $xy='11'$ ). The output of the comparator then goes through an analog-digital converter, ADC to convert the analog voltage to digital. From there, the digital voltage contains the  $x$  and  $y$  bits for the selected cell. To extract one bit from the two-bit output, the two bits are passed through selector  $S_1$  to determine if output bit  $V_x$  or  $V_y$  is fed to the  $V_{out}$  line. The selector signal is input B following the logic in TABLE 6-2— when input B is '0'('1') it selects the  $x(y)$  for a chosen cell. The READ decoder logic is given in TABLE 6-6 while the comparator block logic is given in TABLE 6-7.

TABLE 6-6. READ Controller Logic

Memory Cell	Memory Address		Pass Transistor		Bitline Voltage		Selector Output
	A	B	G1	G2	BL1	BL2	
M1	0	0	1	-	1	-	$x$
	0	1	1	-	1	-	$y$
M2	1	0	-	1	-	1	$x$
	1	1	-	1	-	1	$y$

- is don't care

TABLE 6-7. Comparator Logic Block for READ Controller

Comparator Block			
Reference Voltage value	ReRAM Resistance State	Output Bits	
		$x$	$y$
$V_{ref1}$	HRS	0	0
$V_{ref2}$	IRS-2	0	1
$V_{ref3}$	IRS-1	1	0
$V_{ref4}$	LRS	1	1

The following sequence demonstrates the READ operation for the address  $AB='00'$  for M1 that contains the output bits  $xy='01'$  (the cell is in IRS-2) (TABLE 6-2):

- $R_{EN}$  is asserted and the READ transistor T15 is turned onto supply READ voltage to M1. The values  $A='0'$  and  $B='0'$  are fed to the A and B input lines to indicate which address is to be read (Fig. 6-5). The A input turns on the Wordline WL1 and transistor G1 to select cell M1. (Fig. 6-3 and Fig. 6-4)

- The Bitline READ voltage is then fed to the comparator in the Read Decoder. This block compares the incoming voltage with four reference voltages,  $V_{ref1}$ ,  $V_{ref2}$ ,  $V_{ref3}$ , and  $V_{ref4}$ . Since M1 is in IRS-2, the comparator matches the incoming voltage with  $V_{ref2}$  and sends out the appropriate digital output,  $xy='01'$  (Fig. 6-4, TABLE 6-2)
- At the same time the B input is fed to selector  $S_1$ . The selector selects only the output for the address  $AB='00'$  ( $x$ ) and the value of  $x$  which is '0' is then fed to the  $V_{out}$  line.

Similarly, the READ operation sequence for address  $AB='11'$  for M2 that is in IRS-1 ( $xy='10'$ ) is given below:

- $R_{EN}$  is asserted and the READ transistor T25 is turned on to supply READ voltage to M2. The values  $A='1'$  and  $B='1'$  are fed to the A and B input lines to indicate which address is to be read (Fig. 6-5). The A input turns on the wordline, WL2 and transistor G2 to select cell M2. (Fig. 6-3 and Fig. 6-4)
- The Bitline READ voltage is then fed to the comparator in the Read Decoder. This block compares the incoming voltage with four reference voltages,  $V_{ref1}$ ,  $V_{ref2}$ ,  $V_{ref3}$ , and  $V_{ref4}$ . Since M2 is in IRS-1, the comparator matches the incoming voltage with  $V_{ref3}$  and sends out the appropriate digital output,  $xy='10'$  (Fig. 6-4, TABLE 6-2)
- At the same time the B input is fed to selector  $S_1$ . The selector selects only the output for the address  $AB='11'$  ( $y$ ) and the value of  $y$  which is '0' is then fed to the  $V_{out}$  line.

This READ operation explains the READ for a single address in a MB-ReRAM cell. It is possible to READ two addresses from a MB-ReRAM cell simultaneously. The B input into selector  $S_1$  is responsible for removing the output bit from the unwanted address and only feeding the output from the selected address (in the READ  $AB='00'$  example, the output for  $AB='00'$ ( $x$ ) is selected while the output from  $AB='01'$ ( $y$ ) is ignored. Two-address-simultaneous- READ is thus obtained by removing  $S_1$ . One-address-read is used in this work to allow a clearer one-output-per-cell comparison with the SB-nvLUT.

With tighter margins from MB-switching, the effect of  $V_{DD}$  and ReRAM intrinsic device variabilities have to be considered. To test the effect of  $V_{DD}$  variability on the functionality of the MB-ReRAM and the MB-nvLUT  $V_{DD} \pm 0.05$  V was supplied to the ReRAM cell and the device current for each state is measured (Fig. 6-6). It is important that the margin between each of the four states, LRS, IRS-1, IRS-2, and HRS are maintained and the states do not overlap. The current for each RS achieves clear distinctive levels with the closest measured margin of 42.29% between the  $-0.05 V_{DD}$  IRS-2 and the  $+0.05 V_{DD}$  IRS-1. The ReRAM model

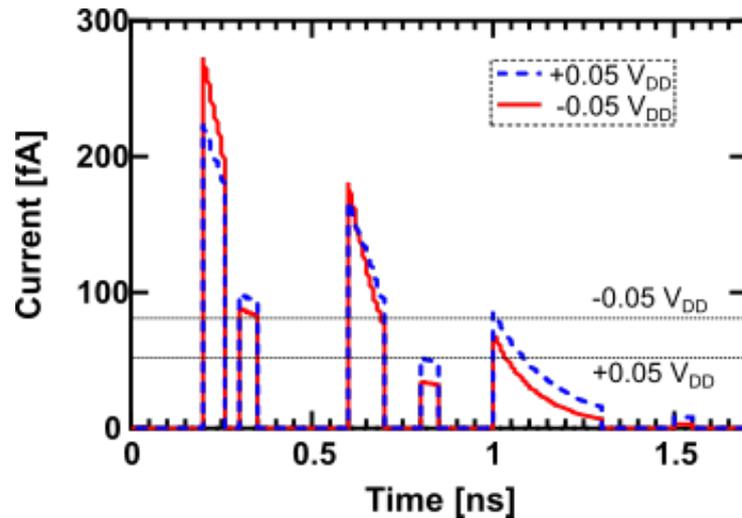


Fig. 6-6. Device current with  $\pm 0.05 V_{DD}$ . Each WRITE operation is followed by a READ to check the device current at the RS. The closest margin is found between IRS-2,  $0.05V_{DD}$  and IRS-1,  $+0.05V_{DD}$ .

used in this work retains the four distinct states even with variability. To translate this to a real device, the device must firstly be able to display similar characteristics to this model in terms of MB-ability and gradual RS. The  $V_{DD}$  is not necessarily fixed to the 1 V HRS and -0.5 V LRS used in this work; they should be adapted accordingly to the real device. The READ decoder circuit should then also be adjusted for the different  $V_{DD}$ . These changes are however not crucial to the function of the controller and the MB-nvLUT in this work with the main criteria being a MB-capable ReRAM.

For the ReRAM's intrinsic variability, a stochastic model obtained from [137] is used to generate a Poisson distribution for the model's RS thresholds, introducing stochasticity to the threshold parameters,  $V_n$  and  $V_p$  (TABLE 6-1). WRITE variability and READ endurance tests are then conducted for the LRS and HRS of the SB-ReRAM and the LRS, IRS-1, IRS-2, and

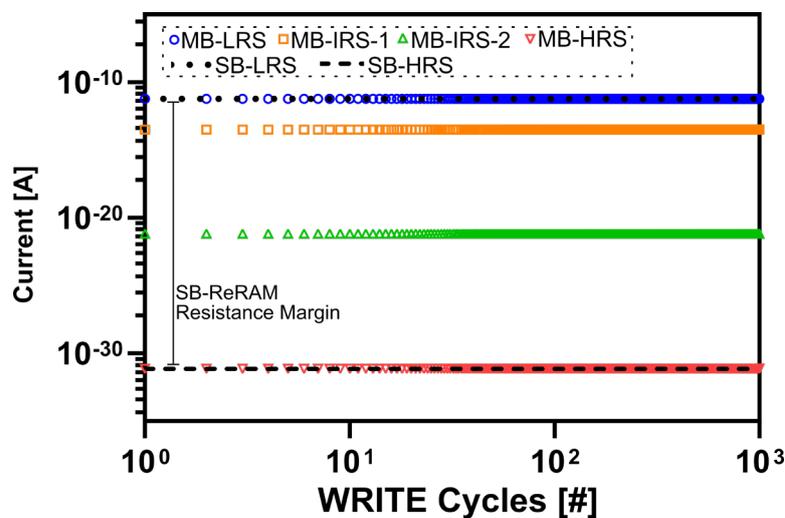


Fig. 6-7. WRITE variability test for LRS and HRS of the SB-ReRAM and LRS, IRS-1, IRS-2, and HRS of the MB-ReRAM.

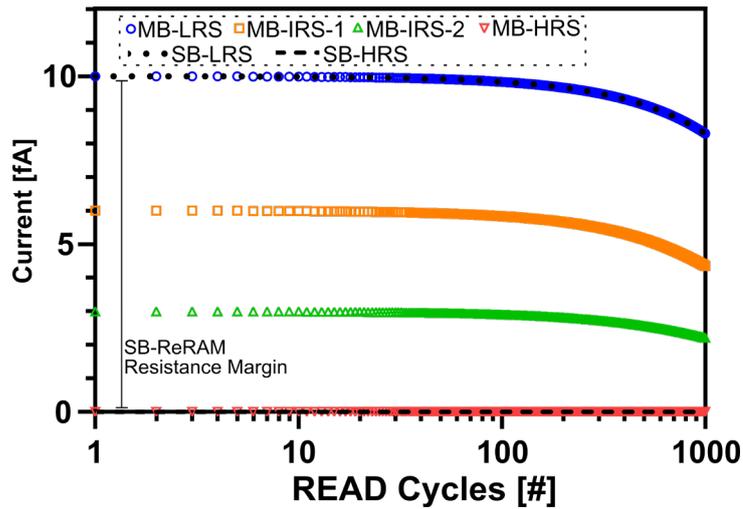


Fig. 6-8. READ endurance test for LRS and HRS of the SB-ReRAM and LRS, IRS-1, IRS-2, and HRS of the MB-ReRAM.

HRS states of the MB-ReRAM for comparison.

To test the WRITE variability, the SB-ReRAM is first set to LRS and then written to HRS, this is repeated 1000 times. The test is similar for the MB-ReRAM which is first set to LRS before a WRITE to another state (IRS-1, IRS-2 or HRS) is performed, repeated for 1000 times. As for the READ endurance test, the SB-ReRAM is set to either LRS or HRS followed by 1000 READ pulses. Likewise, the test is repeated for the MB-ReRAM for each of its resistance states (LRS, IRS-1, IRS-2 or HRS). The existence of IRS levels result in lower margins for the MB-ReRAM and this test was performed to analyze the MB-ReRAM's performance after multiple WRITES. In both cases, the device's output current is measured.

The WRITE variability results are plotted in Fig. 6-7 and shows that the MB-ReRAM WRITE current for each level retains good margins. The READ endurance test is performed up to 1000

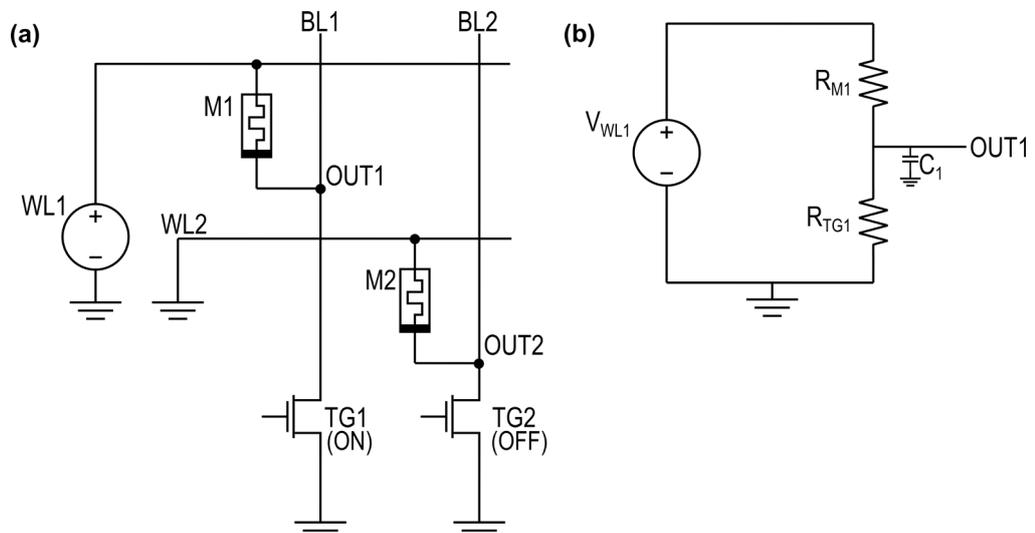


Fig. 6-9. (a) The 4-input MB-LUT and (b) the equivalent RC circuit during WRITE to a single cell (M1).

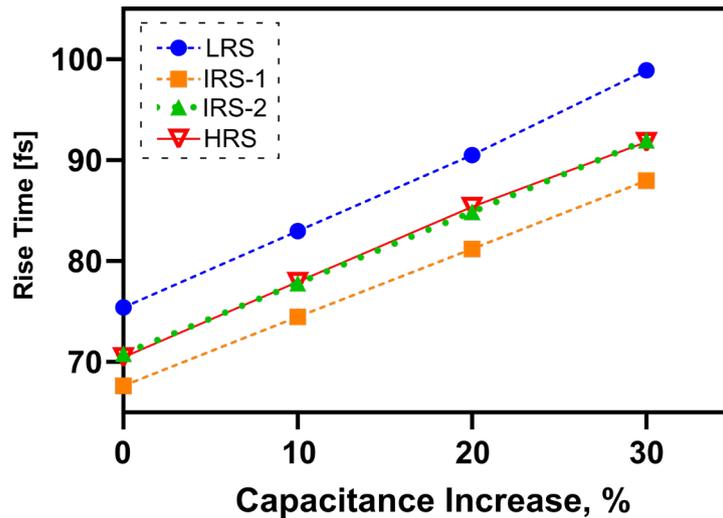


Fig. 6-10. Parasitic RC effect on output voltage,  $V_{out}$  rise time. Simulation performed with parasitic RC increments of 10%.

cycles with notable drift upon reaching the 1000th READ (Fig. 6-8). Since the same ReRAM is used for both SB and MB simulations, the SB-ReRAM's LRS and HRS WRITE variability and READ endurance results follow the MB-ReRAM's LRS and HRS variability and endurance. There is however a larger resistance margin in the SB-ReRAM as a result of the lower number of resistance states and the resistance drift in the READ endurance test is not as drastic as the MB-ReRAM's drift.

Although a refresh WRITE cycle is required upon reaching 1000 READ cycles, the nvLUT's WRITE performance metrics are advantageous when compared to the SB-ReRAM. These results from these tests are consistent with the experimental results, [138] and [139]. In addition to the variability tests, the LUT array is tested with parasitic RC characterizations. Fig. 6-9 shows the 4-input MB-nvLUT and the equivalent RC circuit when writing to an array cell and the capacitance,  $C_1$  is the total capacitance of the transistor, TG1 as well as the wires.

The output voltage,  $V_{out}$  rise times corresponding to 10% increments in parasitic capacitance are plotted in Fig. 6-10. The  $V_{out}$  rise time delays closely follow the increase in parasitic capacitance with the rise times of all states showing an average increase of 30%. This shows that the parasitic capacitance carries an important effect on the circuit and should be considered during fabrication processes.

This MB-nvLUT design reduces the cell count in the LUT array by 0.5x, provides and around 0.25x reduction of gates in the controller circuit while also minimizing the connection lines in the layout. In addition to that the MB-nvLUT provides faster writing, reading with better energy efficiency compared to the SB-nvLUT which will be demonstrated in Section 6.4.

## 6.4 Simulation Results

Simulation results for WRITE and READ delay, energy, and EDP are obtained for the common 2, 4, and 6-input LUTs and a larger 8 input LUT to demonstrate the capabilities of the NV- LUTs. Taking the number of inputs to the LUT as  $n$ , the SB-NVLUT has  $2^n$  number of cells in the array while the MB-NVLUT has  $2^{n-1}$  number of cells. The SB-NVLUT readings are obtained by using the layout in [140]. The LTSpice EDA tool is used for circuit simulations with 100 $\mu$ m transistor technology node while the benchmark tests were carried out using Xilinx ISE and Titan23 [141].

The SB and MB-nvLUTs were simulated with the following selected extreme conditions to highlight the differences between array performances: Condition 1: WRITE ‘0’ from ‘1’ to all cells. Condition 2: WRITE ‘1’ from ‘0’ to all cells. Condition 3: WRITE 2 bits at a time (flip ‘01’ to ‘10’). Condition 4: WRITE 2 bits at a time (flip ‘10’ to ‘01’). Condition 5: READ states ‘00’, ‘01’, ‘10’, ‘11’ from the array. For Condition 3 and 4, half of the cells in the SB-array are flipped from ‘0’ to ‘1’ and the other half are flipped from ‘1’ to ‘0’.

### 6.4.1 1-bit WRITE Operation Results

For Condition 1, all cells in the LUT array are written from ‘1’ to ‘0’—a single MB-ReRAM is switched from ‘11’ to ‘00’ and a single SB-ReRAM is switched from ‘1’ to ‘0’. The reverse is performed for Condition 2. The results for both conditions are listed respectively in TABLE 6-8 and TABLE 6-9 and plotted in Fig. 6-11 and Fig. 6-12. In both cases the WRITE delay of the MB-nvLUT is half that of the SB-nvLUT, showing the advantage of storing 2-bits per ReRAM.

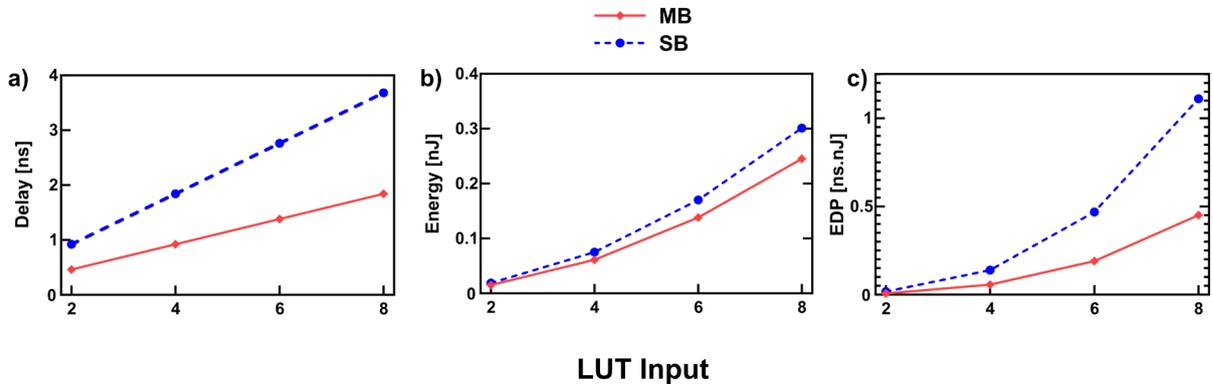


Fig. 6-11. Condition 1: WRITE 0 performance for SB vs MB for 8-bit, 32-bit, 72-bit, and 128-bit arrays; (a) delay, (b) energy, and (c) EDP.

TABLE 6-8. Comparison of Condition 1: WRITE 0 Delay, Energy and EDP

LUT Size	Delay (ns)		Energy (nJ)		EDP (nJ ns)	
	MB-NVLUT	SB-NVLUT	MB-NVLUT	SB-NVLUT	MB-NVLUT	SB-NVLUT
2	0.46	0.92	0.0153	0.0188	0.0071	0.0173
4	0.92	1.84	0.0613	0.0753	0.0564	0.1390
6	1.38	2.76	0.1380	0.1700	0.1900	0.4680
8	1.84	3.68	0.2450	0.3010	0.4510	1.110

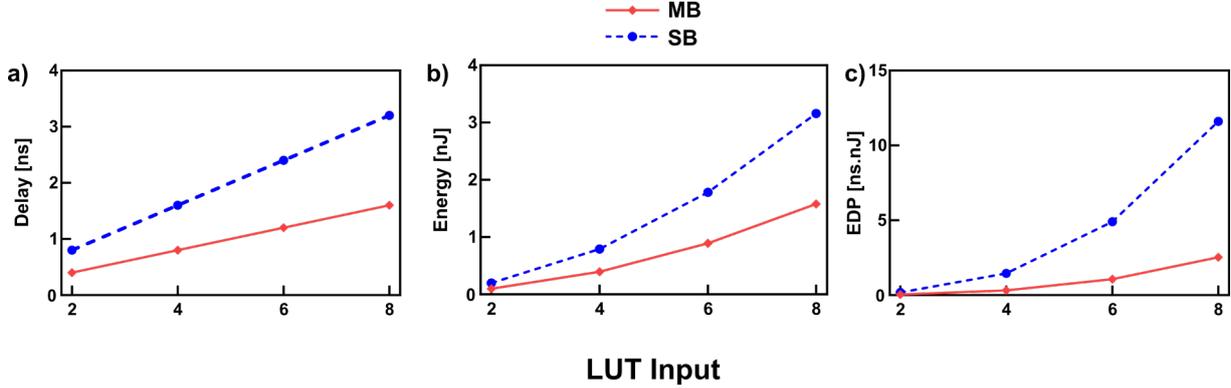


Fig. 6-12. Condition 2: Write 1 performance for SB vs MB for 8-bit, 32-bit, 72-bit, and 128-bit arrays; (a) delay, (b) energy, and

TABLE 6-9. Comparison of Condition 2: WRITE 1 Delay, Energy and EDP

LUT Size	Delay (ns)		Energy (nJ)		EDP (nJ ns)	
	MB-NVLUT	SB-NVLUT	MB-NVLUT	SB-NVLUT	MB-NVLUT	SB-NVLUT
2	0.4	0.8	0.0989	0.198	0.0396	0.1820
4	0.8	1.6	0.3960	0.791	0.3160	1.4600
6	1.2	2.4	0.8900	1.780	1.0700	4.9700
8	1.6	3.2	1.5800	3.160	2.5300	11.600

The energy consumption of the MB-nvLUT is  $1.22\times$  lower on average for Condition 1 and  $2\times$  lower on average for Condition 2 than the SB-nvLUT. This is due to the combination of the MB-ReRAM and having lesser ReRAM cells in the array. The difference in energy consumption can be seen to exponentially increase with the array size.

The EDP of the MB-nvLUT follows the same trend being  $2.46\times$  lower for Condition 1 and  $4.6\times$  lower for Condition 2. The energy consumption is lower for Condition 1 because of the higher resistivity of the cell and thus lower current through the cell.

#### 6.4.2 2-bit WRITE Operation Results

Conditions 3 and 4 produce the results for simultaneous 2-bit WRITE performances. To test both conditions in the SB-nvLUT, 2 SB-ReRAMs are switched at the same time (1 ReRAM switched from '0'  $\rightarrow$  '1' and another ReRAM switched from '1'  $\rightarrow$  '0'). Since state '10' is

IRS-1 and state '01' is IRS-2, the MB-ReRAM must be switched back to LRS before switching to state '10' (transition from: '01' → '11' → '10') for Condition 3.

There is however no increase in delay time and the energy and EDP still remains lower than the SB-nvLUTs as shown in Fig. 6-13 and TABLE 6-10. In this condition, the extra switching step required increases the energy and EDP for the MB-nvLUT, raising the energy consumption to within the range of SB-nvLUT values. The MB-nvLUT still retains the advantage in WRITE delay, averaging about half the time of the SB-NvLUT. The Condition 3 energy for the MB-nvLUT is lower by 1.03× on average compared to the SB-NvLUT. The EDP of the MB-nvLUT is 2× lower on average than the SB-nvLUT due to the shorter WRITE

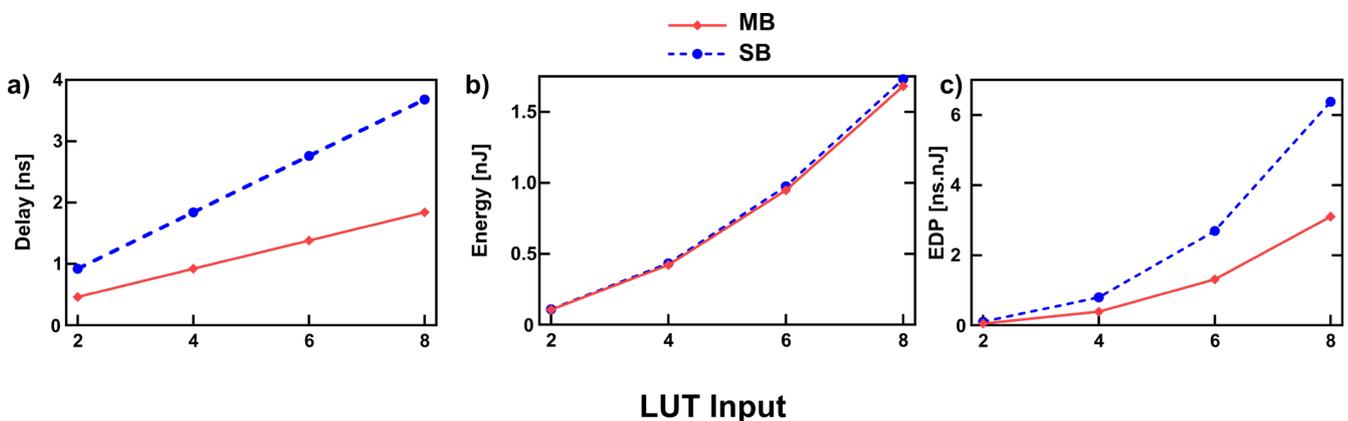


Fig. 6-13. Condition 3: 2-bit WRITE performance (01 to 10) for SB vs MB for 8-bit, 32-bit, 72-bit, and 128-bit arrays; (a) delay, (b) energy, and (c) EDP

TABLE 6-10. Comparison of Condition 3: 2-bit WRITE (01 to 10) Delay, Energy and EDP

LUT Size	Delay (ns)		Energy (nJ)		EDP (nJ ns)	
	MB-NvLUT	SB-NvLUT	MB-NvLUT	SB-NvLUT	MB-NvLUT	SB-NvLUT
2	0.46	0.92	0.105	0.108	0.0484	0.0997
4	0.92	1.84	0.421	0.433	0.3870	0.7970
6	1.38	2.76	0.947	0.975	1.3100	2.6900
8	1.84	3.68	1.680	1.730	3.1000	6.3800

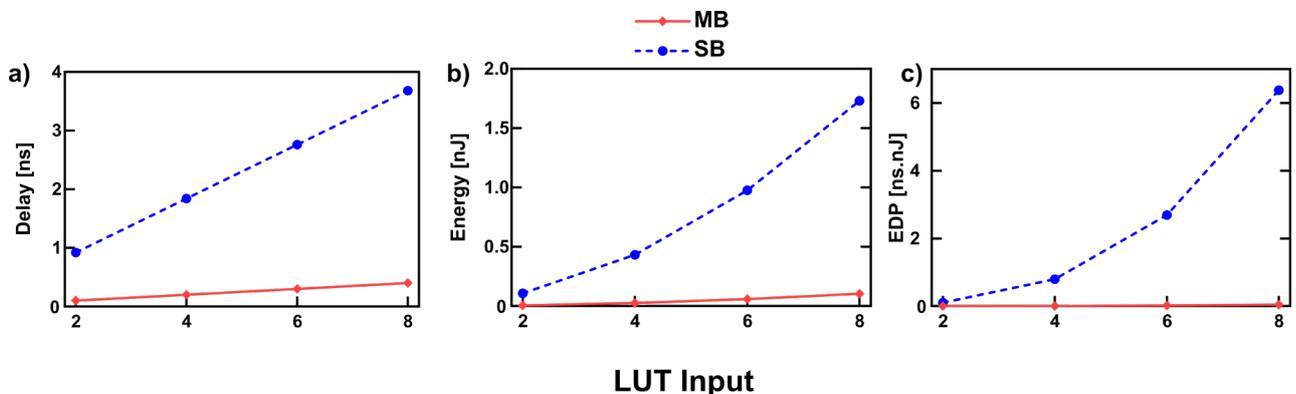


Fig. 6-14. Condition 4: 2-bit WRITE performance (10 to 01) for SB vs MB for 8-bit, 32-bit, 72-bit, and 128-bit arrays; (a) delay, (b) energy, and (c) EDP.

TABLE 6-11. Comparison of Condition 3: 2-bit WRITE (01 to 10) Delay, Energy and EDP

LUT Size	Delay (ns)		Energy (nJ)		EDP (nJ ns)	
	MB-nvLUT	SB-nvLUT	MB-nvLUT	SB-nvLUT	MB-nvLUT	SB-nvLUT
2	0.46	0.92	0.105	0.108	0.0484	0.0997
4	0.92	1.84	0.421	0.433	0.3870	0.7970
6	1.38	2.76	0.947	0.975	1.3100	2.6900
8	1.84	3.68	1.680	1.730	3.1000	6.3800

delay.

The MB-nvLUT displays a massive advantage over the SB-nvLUT in Condition 4 tests shown in Fig. 6-14 and TABLE 6-11. Switching from ‘10’ → ‘01’ in a 2-input LUT only takes 0.1 ns for the MB-nvLUT compared to the 0.92 ns for the SB-nvLUT.

The average difference in WRITE delay for Condition 4 is 9.2×. The energy required to switch from IRS-1 to IRS-2 in the MB-ReRAM is significantly lower and the Condition 4 results for the MB-nvLUT are lower by a significant 128× on average. This benefit is also seen in the EDP where the MB-nvLUT is lower by 153× on average compared to the SB-nvLUT.

Throughout the WRITE tests, the MB-nvLUT consistently demonstrates a significant improve in performance in terms of WRITE latency and energy consumption over the SB-nvLUT.

### 6.4.3 READ Operation Results

The READ operation analysis is carried out by reading a single MB cell (2-bits) and compared with the result of 2 SB cells (2-bits). To perform SB-nvLUT read of 2-bits, 2 cells of different columns are selected. This prevents simultaneous 2-bit read and provides for an increment in delay time. The average READ delay for SB-nvLUT is 5× higher than the MB-nvLUT. The reading of 2 cells for SB-nvLUT compared to just 1 cell for MB-nvLUT also increases the energy consumption of the operation. The SB-nvLUT average READ energy and average

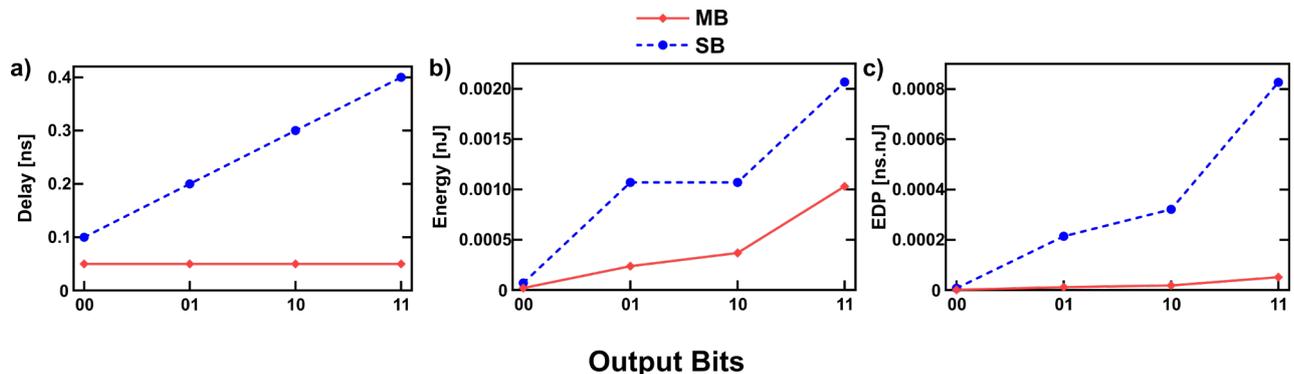


Fig. 6-15. Read 00, 01, 10, 11 for SB vs MB; (a) delay, (b) energy, and (c) EDP.

TABLE 6-12. Comparison of READ 00, 01, 10, 11

LUT Input	Delay (ns)		Energy (nJ)		EDP (nJ ns)	
	MB-NVLUT	SB-NVLUT	MB-NVLUT	SB-NVLUT	MB-NVLUT	SB-NVLUT
00	0.05	0.1	1.99e-5	6.95e-5	9.95e-7	6.95e-6
01	0.05	0.2	2.37e-4	1.07e-3	1.19e-5	2.14e-4
10	0.05	0.3	3.68e-4	1.07e-3	1.84e-5	3.21e-4
11	0.05	0.4	1.03e-3	2.07e-3	5.17e-5	8.27e-4

READ EDP is  $3.2\times$  and  $14.6\times$  higher than the MB-nvLUT.

The MB-nvLUT is therefore much more efficient for READ operation by virtue of having twice the bits per cell compared to the SB-nvLUT. This efficiency can also be further increased with ReRAMs with  $>2$  bits per cell. The READ delay, energy, and EDP values are plotted in Fig. 6-15 and given in TABLE 6-12.

### 6.5 Single Cell Performance Comparison

Additionally, the single cell performance matrices of the MB-nvLUT presented in this work is compared with the 2TG1M single cell performance in [49]. The WRITE time for the 2TG1M cell is a consistent 2.70 ns for all WRITE conditions whereas the WRITE times for the presented MB-nvLUT cell is 0.4 ns for WRITE ‘11’ and 0.46ns for WRITE ‘00’. The WRITE time is therefore lower by an average of 84% compared to the 2TG1M cell. The READ time of the 2TG1M is 15 ns compared to the READ time of the 0.05 ns for the presented MB-nvLUT cell, which is 96.7% lower. These performance improvements are a result of using just ReRAM cells in the array and the omission of TGs.

### 6.6 Evaluation on Benchmark Circuits

The performances of MB-ReRAM LUTs are then evaluated in FPGA benchmark circuits; ISCAS’89 benchmark circuits in Xilinx Virtex4 FPGA (XC4VLX100) and Virtex5 FPGA (XC5VLX220) as well as four Titan23 benchmarks: denoise, directrf, bitcoin\_miner, and gaussianblur [141]. The SRAM-based LUTs are replaced with the SB and MB-LUTs from this work while the interconnect routing is kept the same. The comparison results with Virtex4 and Virtex5 benchmark circuits are listed in TABLE 6-13 and TABLE 6-14 respectively while the comparison result for Titan23 is given in TABLE 6-15.

The MB-nvLUT performs better than the SB-nvLUT in both average WRITE and READ delays as well as average EDP in the benchmark tests. The average READ delay and EDP is significantly lower for MB-nvLUTs at higher LUT counts as evident in Titan23 benchmarks.

Both nvLUTs demonstrate higher WRITE delays than their volatile LUT counterparts but have much lower average READ delays and EDP. This is desirable for FPGA applications, where the READ operation is used more often after initial programming. The nvLUTs are also able to retain data in the event of a power loss so they don't require writing after reboots.

A method of comparison for performance metrics in [29] and [142] is carried out for the ISCAS'89 benchmarks in this work. The average performance of the MB and SB-nvLUTs are found by estimating the number of consecutive READ operations required immediately following a WRITE so that the total delay time incurred by the MB and SB-nvLUTs is equal to the delay incurred by the SRAM-based LUT.

Let  $T_{\text{ReRAMWrite},B}$  and  $T_{\text{SRAMWrite},B}$  be the average ReRAM and SRAM write delay for benchmark, B respectively. In the same way  $E_{\text{ReRAMWrite},B}$  and  $E_{\text{SRAMWrite},B}$  are the average ReRAM and SRAM EDPs under the same case. Both are given as:

$$T_{\text{ReRAMWrite},B} = T_{\text{SRAMWrite},B} = t_{\text{SRAMWrite},B} + K_T t_{\text{SRAMRead},B} \quad (6.8)$$

$$E_{\text{ReRAMWrite},B} = E_{\text{SRAMWrite},B} = e_{\text{SRAMWrite},B} + K_E e_{\text{SRAMRead},B} \quad (6.9)$$

where  $t_{\text{SRAMWrite},B}$  and  $e_{\text{SRAMWrite},B}$  are the average WRITE time and EDP for SRAM and  $t_{\text{SRAMRead},B}$  and  $e_{\text{SRAMRead},B}$  are the average READ time and EDP for SRAM under the same benchmark B.  $K_T$  and  $K_E$  are the least consecutive READ operations in integers required so that  $T_{\text{ReRAMWrite},B}$  is equal to  $T_{\text{SRAMWrite},B}$  and  $E_{\text{ReRAMWrite},B}$  is equal to  $E_{\text{SRAMWrite},B}$ . The average  $K_T$  for Virtex4 benchmark circuits (Fig. 6-16(a)) are 1.3 for MB-nvLUTs and 1.7 for SB-nvLUTs while the average  $K_E$  (Fig. 6-16(b)) is 3331 and 6950 for MB and SB-nvLUTs respectively with the lower value for MB-nvLUTs demonstrating the advantage in power consumption.

The average  $K_T$  for Virtex5 benchmark circuits (Fig. 6-16(c)) is 0.74 for MB-nvLUTs and 0.97 for SB-nvLUTs while the average  $K_E$  (Fig. 6-16(d)) is 57 and 120 for MB and SB-nvLUTs respectively. Average  $K_E$  values for SB-nvLUTs are consistently more than twice that of the MB-nvLUT.

The drop in both  $K_T$  and  $K_E$  in the nvLUT in the Virtex5 benchmark circuit compared to the Virtex4 benchmark circuit shows the large potential of nvLUTs in scaling. The performance of nvLUT catches up when more LUTs are used in an FPGA and are capable of surpassing

TABLE 6-13. Virtex4 benchmark circuit comparison of average write delay and EDP for MB-LUT, SB-LUT, and SRAM LUT.

Virtex 4 FPGA (XC4VLX160)															
Benchmark circuit	2-Input LUTs	3-Input LUTs	4-Input LUTs	WRITE						READ					
				Average delay, ns			Average EDP, pJ ns			Average delay, ns			Average EDP, pJ ns		
				MB	SB	SRAM	MB	SB	SRAM	MB	SB	SRAM	MB	SB	SRAM
298	4	6	11	13.49	33.82	3.92	33.47E2	131.14E2	38.90E-2	1.05	5.25	17.17	43.57E-2	7.19	0.452
400	11	14	19	27.34	68.53	7.94	65.76E2	257.60E2	1.51	2.20	11.00	33.29	91.29E-2	15.06	1.691
510	13	17	55	55.74	139.73	16.19	140.85E2	551.74E2	7.18	4.25	21.30	75.27	1.76	29.09	8.761
820	10	26	72	73.13	183.34	21.24	189.88E2	743.77E2	12.35	5.40	27.00	98.69	2.24	36.96	15.05
953	25	33	123	119.64	299.93	34.75	304.43E2	119.25E3	33.69	9.05	45.30	163.82	3.76	61.95	41.56
1238	27	41	155	148.75	372.91	43.20	381.43E2	149.41E3	52.40	11.20	55.80	204.63	4.63	76.32	64.88
1488	26	48	183	173.24	434.32	50.31	448.16E2	175.54E3	71.51	12.90	64.30	239.45	5.33	87.96	88.89
5378	65	96	206	237.50	595.41	68.97	593.31E2	232.40E3	123.56	18.40	91.80	307.87	7.61	125.60	145.82
15850	140	270	376	508.36	1274.48	147.64	126.94E3	497.22E3	532.56	39.30	197.00	629.47	16.31	269.00	606.08
35932	1192	489	1307	16.98E2	4257.76	493.23	372.07E3	145.75E4	603.17E1	149.00	747.00	21.26E2	62.00	10.22E2	6923.8

TABLE 6-14. Virtex5 benchmark circuit comparison of average write delay and EDP for MB-LUT, SB-LUT, and SRAM LUT.

Virtex 5 FPGA (XC5VLX220)																	
Benchmark circuit	2-Input LUTs	3-Input LUTs	4-Input LUTs	5-Input LUTs	6-Input LUTs	WRITE						READ					
						Average delay, ns			Average EDP, pJ ns			Average delay, ns			Average EDP, pJ ns		
						MB	SB	SRAM	MB	SB	SRAM	MB	SB	SRAM	MB	SB	SRAM
298	1	3	3	3	4	12.07	30.26	4.22	56.90E2	223.06E2	75.53E-2	0.70	3.50	29.61	29.05E-2	4.79	1.37
400	2	5	6	9	4	22.37	56.07	7.83	105.44E2	413.31E2	2.25	1.30	6.5	47.64	53.95E-2	8.90	3.54
510	0	2	5	11	17	34.79	87.22	12.99	194.13E2	761.08E2	8.35	1.75	8.75	105.30	72.62E-2	11.98	17.38
820	4	3	7	25	47	85.20	213.60	32.17	484.84E2	190.09E3	53.31	4.30	21.50	271.46	1.78	29.43	115.53
953	13	9	10	31	71	126.74	317.73	47.32	694.75E2	273.60E3	115.22	6.70	33.50	399.73	2.78	45.86	250.25
1238	9	9	19	39	60	128.51	322.18	47.52	697.85E2	272.37E3	109.36	6.80	34.00	377.64	2.82	46.54	223.40
1488	11	10	24	24	76	134.55	337.31	49.38	713.16E2	279.59E3	126.85	7.25	36.25	421.63	3.01	49.62	278.45
5378	32	41	49	59	80	223.30	559.81	79.18	107.71E3	422.22E3	272.80	13.05	65.25	569.75	5.42	89.32	506.96
15850	89	94	96	197	163	549.90	13.79E2	196.81	270.85E3	106.18E4	15.93E2	31.95	159.75	13.39E2	13.26	218.69	28.00E2
35932	1152	326	380	129	542	16.24E2	40.74E2	541.06	595.70E3	233.51E4	117.34E2	126.45	632.25	36.50E2	52.47	865.52	206.65E2

TABLE 6-15. TITAN23 benchmark circuit comparison of average write delay and EDP for MB-LUT, SB-LUT, and SRAM LUT.

Titan23 Benchmarks													
Benchmark Circuit	6-Input LUTs	WRITE						READ					
		Average delay, ms			Average EDP, pJ ms			Average delay, ms			Average EDP, pJ ms		
		MB	SB	SRAM	MB	SB	SRAM	MB	SB	SRAM	MB	SB	SRAM
denoise	322385	686.68	860.77	99.62	416.30E3	867.05E3	5.31E-2	16.12	80.60	108.39E4	6.69	1.10E2	1.02E4
directrf	471194	10.04E2	12.58	145.60	608.45E3	126.73E4	7.77E-2	23.56	117.80	158.42E4	9.78	1.61E2	1.49E4
bitcoin miner	455263	969.71	12.16E2	140.68	587.88E3	122.44E4	7.50E-2	22.76	113.82	153.06E4	9.45	1.56E2	1.44E4
gaussianblur	805079	17.15E2	21.50E2	248.77	103.96E4	216.53E4	13.27E-2	40.25	201.27	270.67E4	1.67E1	2.76E2	2.54E4

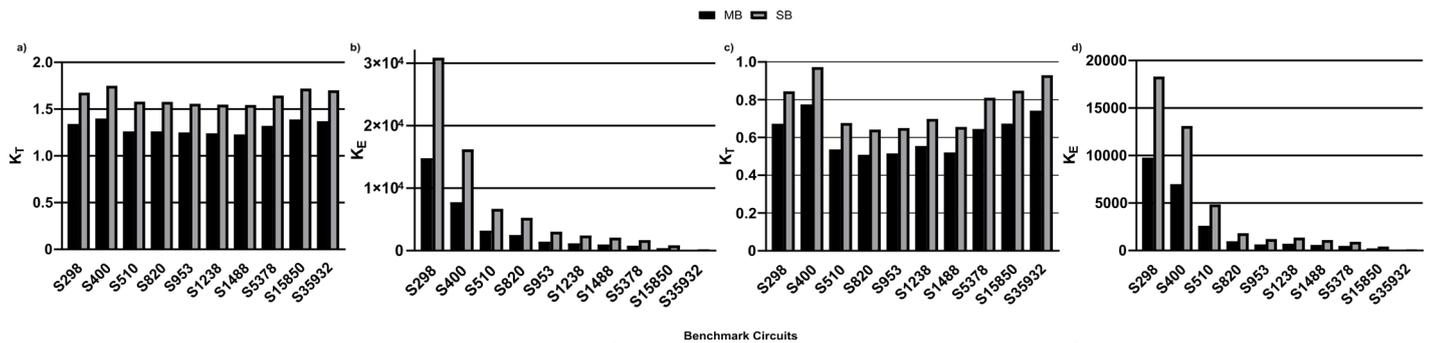


Fig. 6-16. MB and SB-LUT (a) KT and (b) KE for Virtex4 benchmarks and (c) KT and (d) KE for Virtex5 benchmarks.

conventional LUT performance in massive LUT designs. Traditional LUTs are also limited by physical constraints of the SRAM while the ReRAM is able to scale down in size.

## 6.7 Summary

The MB-nvLUT takes advantage of the intrinsic MB property of ReRAMs and each ReRAM cell in the MB-nvLUT is capable of storing 2 bits compared to the 1-bit-per-ReRAM architecture of SB-nvLUTs. The 2-bits in each cell stores the output values for two specific LUT inputs (ie. AB='00' and AB='01' in ReRAM M1). This method of storage requires two output bits to be written simultaneously different than conventional LUT writing schemes. An MB-nvLUT controller is also designed and proposed specifically for this MB-nvLUT writing scheme. The controller takes in two output values and selects which level of resistance to switch the selected ReRAM into.

WRITE and READ tests were performed on the MB-nvLUT and SB-nvLUT for a comparison. The MB-nvLUT is highly promising, demonstrating significant performance advantage in WRITE and READ delay in all but one test in which is still had a slight advantage over the SB-nvLUT.

Virtex4 and Virtex5 benchmark circuit tests were also performed and the MB-nvLUT once again displays significant lower WRITE and READ delays, energy consumption, and EDP. The MB-nvLUT performance catches up with traditional SRAM-based LUTs in larger FPGA designed as can be seen when looking at Virtex4 and Virtex5 results.

Overall, the MB-nvLUT is very promising as a next-generation nvLUT. The reduction in latency and energy consumption as well as the reduction in components used in the design gives the MB-nvLUT an edge over traditional LUTs as well SB-nvLUTs.

## 7. Chapter Seven

# Non-Volatile Memories

---

This chapter will look at additional opportunities for introducing nonvolatility to the nvFPGA in the sequential circuits which are commonly made up of D Flip-flops in the CLBs and the SwBs that are responsible for programmed routing and are commonly made up of SRAMs. These components are conventionally volatile and have to be reprogrammed when the FPGA is switched off or if there is a power disruption. These components also require constant power supply during operation which leads to high idle power consumption. Three NV designs are presented in this chapter; the nvD Latch, the nvD Flip-Flop, and the nvDRAM. Analysis of performance metrics that are typical for each design such as the Clk-to-Q delay timing for the nvD Latch and the nvD Flip-Flop and the refresh timing for the nvDRAM are carried out. These are then compared with the conventional volatile versions of each design.

### Related Publications

---

1. H. L. Chee, T. N. Kumar, and H. A. F. Almurib, "A ReRAM-based Nonvolatile FPGA," *Proceedings in 20th IEEE Student Conference on Research and Development (SCOReD 2022)*, 2022. (Malaysia)
2. A Low Power Nonvolatile DRAM Cell based on ReRAMs, *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* (under review).

## 7.1 Memories in Large-Scale Architecture

The memory designs presented form the memory components of VLSI designs such as processors or FPGAs. The D Latch and D Flip-Flops (DFFs) are memory circuits that store data bits with the difference between them being that the data in DFFs only changes at the Clk rising edge (edge triggered) whereas the data in the D Latch changes during Clk high. For example the DFF is used as the memory module for a flexible processor presented in Fig. 7-1 [143].

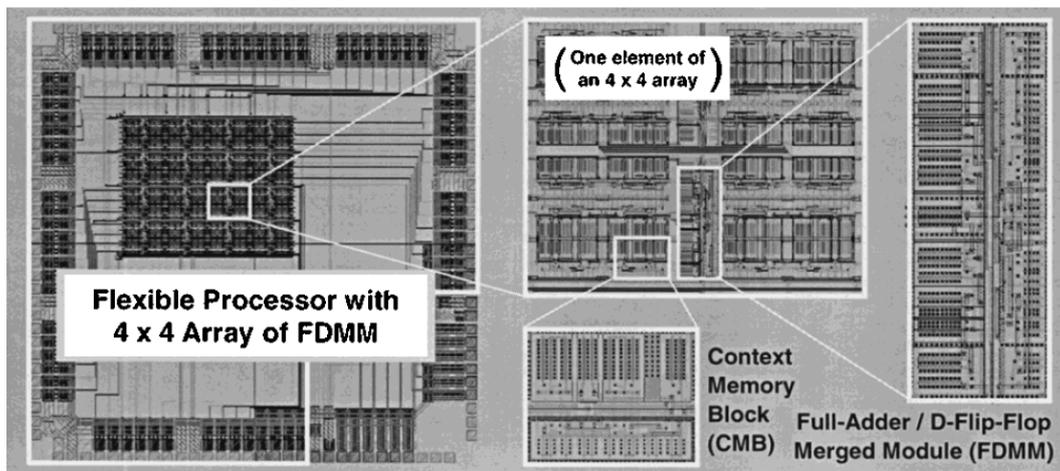


Fig. 7-1. Flexible processor chip layout from [143].

The DRAM is a common memory device based on 1T1C cells. Due to the small component counts, they have a smaller cell area compared to the D latch and D Flip-Flops but are slower in WRITE and READ operations. An example of an embedded DRAM in an FPGA is given in Fig. 7-2 [144].

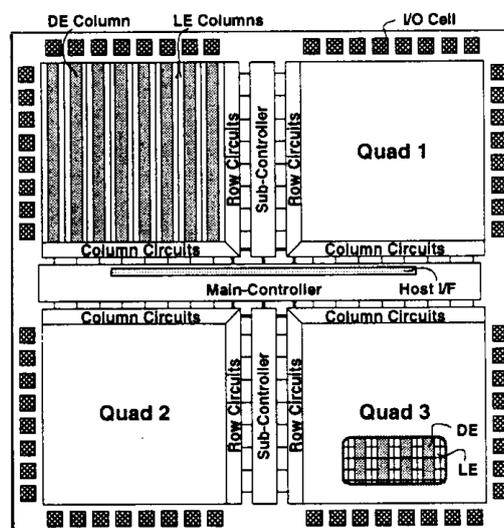


Fig. 7-2. DRAM-embedded FPGA layout from [144].

## 7.2 Non-Volatile D Latch

This section presents a nvD latch that uses two ReRAMs at the Q and Qbar (Q<sub>b</sub>) outputs to perform passive data storage, incorporating NV into the conventional CMOS D-latch design. The ability for passive storage ensures the data in the nvD latch is always stored in the case of power disruptions and the ‘passive’ component negates the need for time-consuming ‘active’ storage sequences.

### 7.2.1 The Proposed nvD Latch

The schematic of the proposed nvD-latch using two ReRAMs, M<sub>1</sub> and M<sub>2</sub> is presented in Fig. 7-3. The two ReRAMs are connected to the respective Q and Q<sub>b</sub> outputs of the D-latch and to a common node, G.

During Clk high, the D input from V<sub>D</sub> is passed through transmission gate, TG1 into the gates of transistors, T1 and T2 which form an inverter pair. The D input is then inverted and passed to the input gates of inverter pair transistors, T3 and T4. The Q output which follows the D input is tied to the output of T3/T4 while the Q bar, Q<sub>b</sub> output is tied to the output of T1/T2 which produces the inversion of D. Transmission gate TG2 is active during Clk low, therefore tying the Q line to T1/T2 input gates when Clk is low and storing the previous data bit.

As for the NV segment, when Clk is high and D input is logic ‘1’ (logic ‘0’), the Q output follows and is logic ‘1’ (logic ‘0’) while Q<sub>b</sub> is inverted and is logic ‘0’ (logic ‘1’). This forms a voltage potential across M<sub>1</sub> and M<sub>2</sub> with G through resistor R1, and switches M<sub>1</sub> to HRS (LRS) and M<sub>2</sub> to LRS (HRS). Thus, both ReRAMs are passively written during normal operation and do not require additional STORE sequences. It should be noted that both

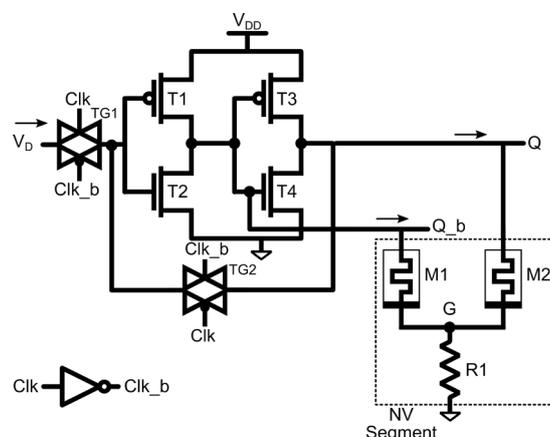


Fig. 7-3. Schematic of the nvD latch. Two ReRAMs, M<sub>1</sub> and M<sub>2</sub> with a ground resistor, R1 form the NV segment of the nvD latch.

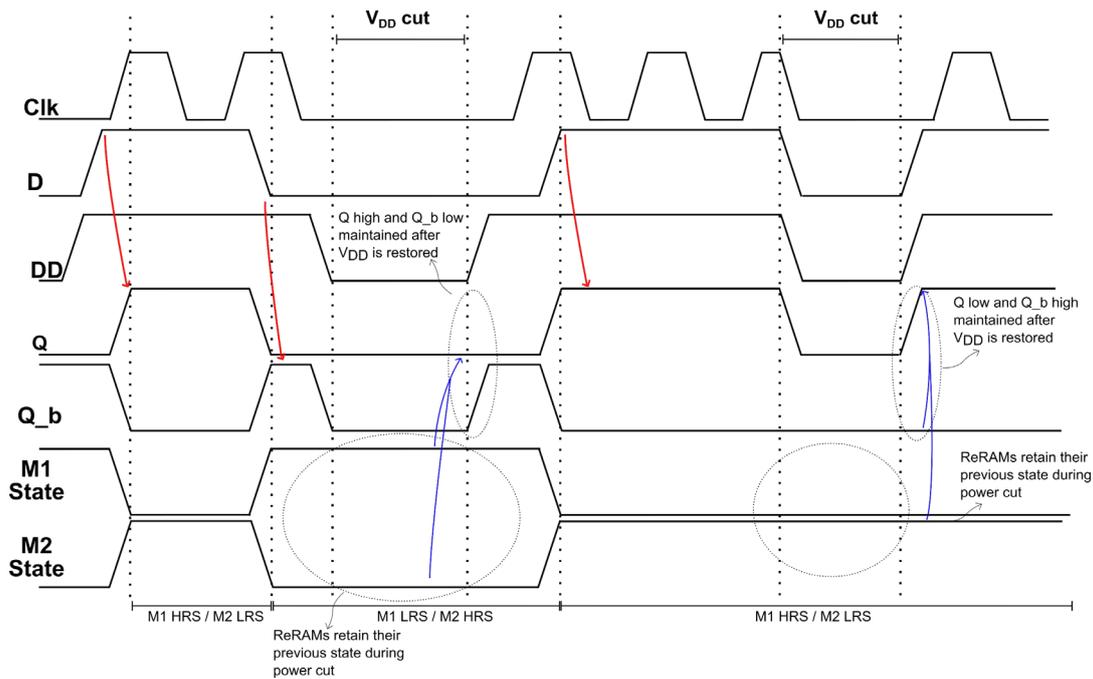


Fig. 7-4. nvD-latch process waveform.

ReRAMs switch at the same time during this period and is contrary to the CRS layout which requires one ReRAM to switch before the other.

The ReRAMs retain their resistance states in the event of power disruption and when  $V_{DD}$  is restored, the ReRAM in LRS pulls down its respective output line to ground while the ReRAM in HRS maintains its output line close to logic '1'. The nvD-latch's process waveform is given in Fig. 7-4.

## 7.2.2 Results and Simulation

In this section, the design is simulated in LTSpice for transistor technology nodes of 32 nm, 45 nm, and 65 nm.  $V_{DD}$  interrupts are performed to test the Restore function under two conditions;

1. Restore  $Q='1'$  and  $Q\_b='0'$
2. Restore  $Q='0'$  and  $Q\_b='1'$

The results are shown in Fig. 7-5 and Fig. 7-6 respectively. The nvD Latch successfully restores the respective  $Q$  and  $Q\_b$  conditions when power is restored after a disruption. The  $Q/Q\_b$  rising/falling times and clock-to- $Q$  delay ( $t_{CQ}$ ) are measured for the nvD Latch and the volatile D Latch (vD latch) and the results are given in TABLE 7-1.

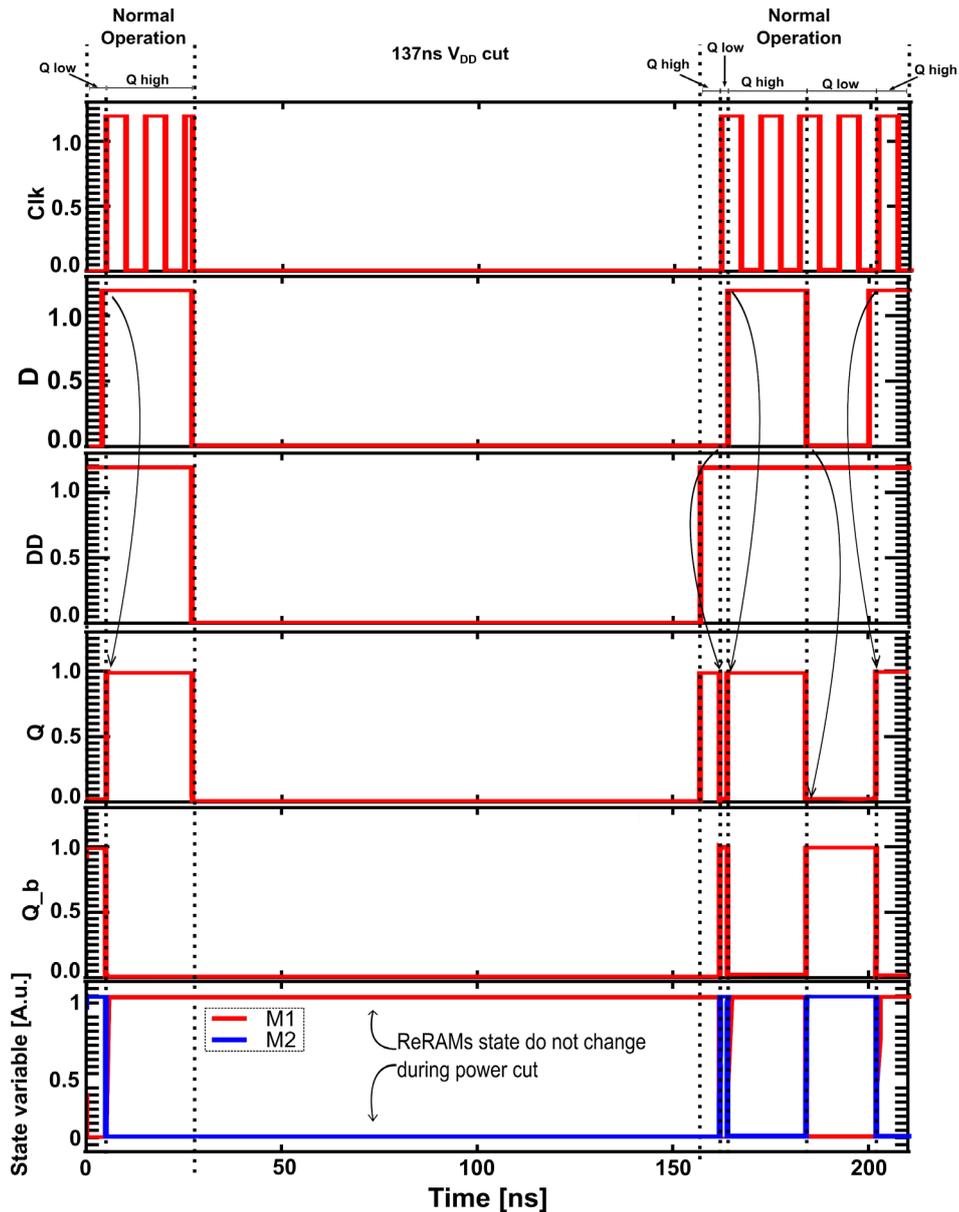


Fig. 7-5. Successful restore of  $Q=1$  and  $Q_b=0$  after 100ns VDD cut.

The nvD Latch has an increase of 72% in  $t_{CQ}$ , 56% increase in  $Q$  rising/ $Q_b$  falling time, and 76% increase in  $Q$  falling/ $Q_b$  rising time over the vDlatch due to the added switching time of the ReRAMs. The timings are however all measured in picoseconds and the nvD latch is thus capable of high frequency switching. The average RESTORE times and the worst-case restore times for  $Q/Q_b$  after power restoration after a disruption are also given in TABLE 7-1 with an average of 17.5 ps restore and 18.8 ps worst-case times across the transistor technology nodes.

The nvD latch's switching time is then compared with the D latch designs [53] and [54] and given in TABLE 7-2. It should be noted that this comparison was performed as is and the designs in [53] and [54] are using different technology nodes from the design in this work. This

TABLE 7-1. Clk-to-Q Timings for nvD latch and vD latch

Transistor node (nm)	Clk-to-Q delay (ps)		Switching delay (ps)				Average Restore Time (ps)	Worst-case Restore Time (ps)
			Q rising time / Q_b falling time		Q falling time / Q_b rising time			
	nvDlatch	vDlatch	nvDlatch	vDlatch	nvDlatch	vDlatch	nvDlatch	vDlatch
32	39.9	10.2	31.7	13.9	30.9	10.1	16.0	18.07
45	60.3	19.6	39.4	17.4	60.6	12.3	17.9	18.7
65	77.7	20.3	43.0	19.1	71.2	13.4	18.6	19.7

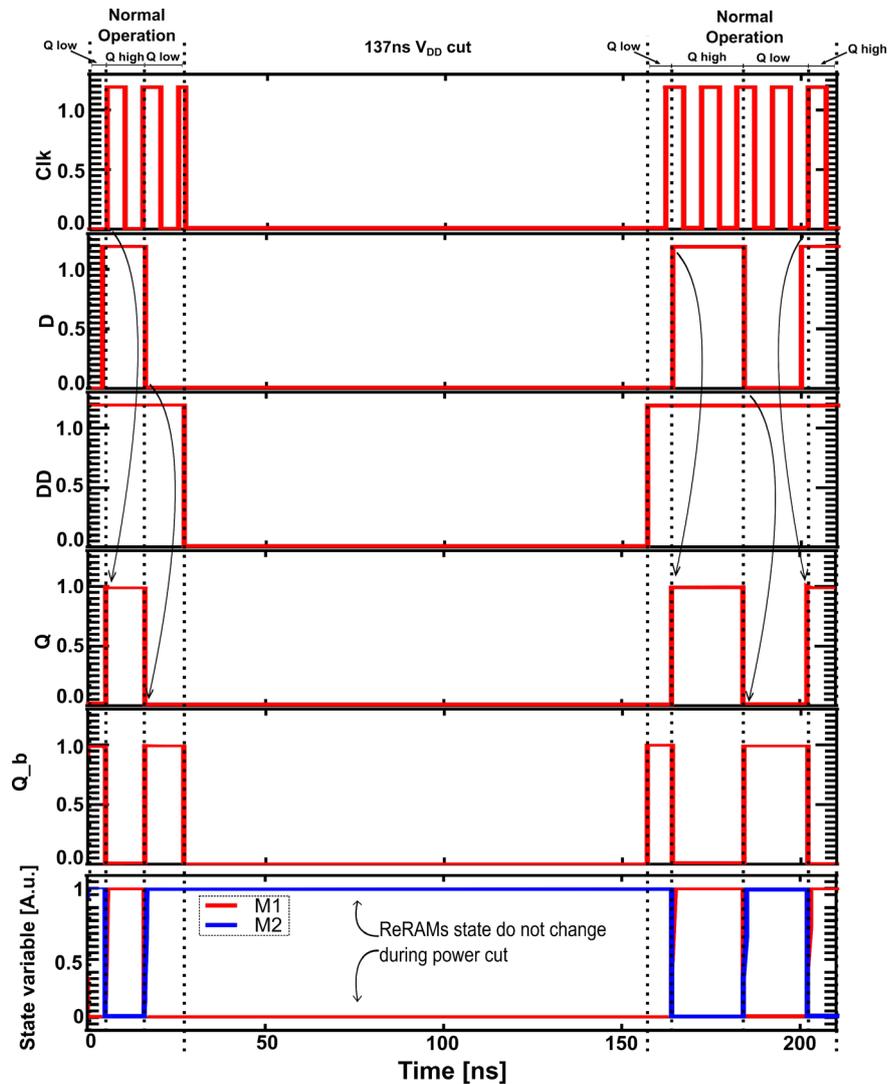


Fig. 7-6. Successful restore of Q=0 and Q\_b=1 after 100ns VDD cut.

TABLE 7-2. Clk-to-Q Timings for nvD latch and vD latch

Design	Threshold Voltage, $V_T$	Switching time (ns)
[53]	2.4 V	9.3 ns
[54]	2.4 V	4.0 ns
This work	1.2 V	65.9 ps

design is capable of faster switching as the Q/Q\_b lines are able to produce outputs without the need to wait for the ReRAMs to complete its switching.

### 7.3 Non-Volatile D Flip-Flop

This section presents a novel nvDFF design by using two ReRAMs to retain data in the event of power interruption. The NV STORE/RESTORE in this design is a passive operation and therefore does not require additional sequences or external control circuitry.

#### 7.3.1 The Proposed nvDFF

The design of the proposed nvDFF and the modified circuit model of the ReRAM used in the proposed nvDFF are presented in this section. The schematic of the proposed nvDFF is shown in Fig. 7-7 consisting of the Master latch, Slave latch and modified ReRAM NV segment highlighted in dotted boxes.

TABLE 7-3 lists the parameters' respective values and the model's electrical characteristics are shown in Fig. 7-8. The parameters of the ReRAM model [145] are selected to match the characteristics of experimental results [33], [147].

To switch the ReRAM to the LRS, a positive pulse input of 1.2 V amplitude with 3 ns pulse

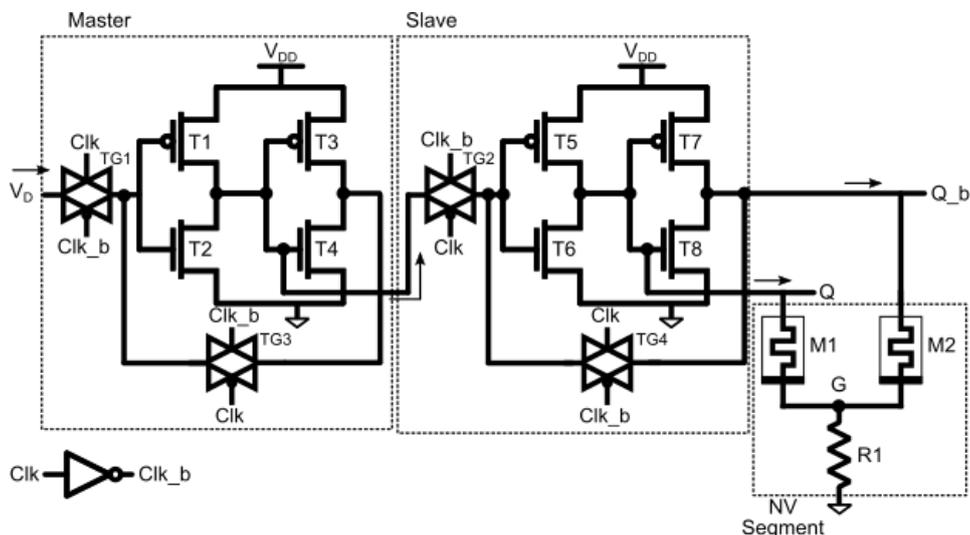


Fig. 7-7. Schematic of the nvDFF. The NV component consists of two ReRAMs, M1 and M2 and the grounding resistor, R1.

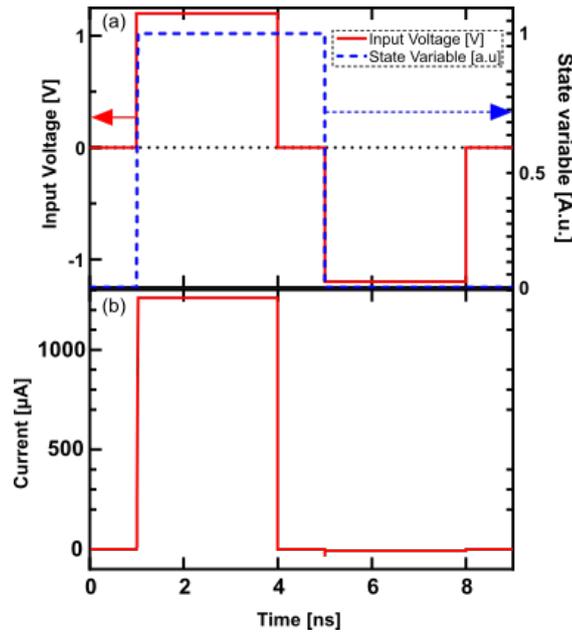


Fig. 7-8. Electrical characteristics of the ReRAM model showing (a) the response of the state variable to the input voltage and (b) the device current.

TABLE 7-3. ReRAM Model Parameters

Parameters	Functions	Values
a1	IV relationship modifier	8.5
a2		0.3
b		1.2e-3
Vp	Switching voltage thresholds	0.1
Vn		0.1
Ap	State variable motion modifier	5e10
An		1e10
xp	State variable motion limiter	0.9
xn		0.5
αp	State variable motion decay modifier	5
αn		0.5

width is applied while to switch the ReRAM to the HRS a negative pulse (-1.2 V) is applied with 3 ns pulse width duration.

The state variable of the ReRAM attains its maximum value when positive polarity voltage is applied from 1 ns to 4 ns (Fig. 7-8(a)). The LRS resistance of the ReRAM reaches  $\sim 100 \Omega$  with the maximum current of 10 mA flowing through the device in Fig. 7-8(b). This state of the ReRAM is used to store the logic low level (data,  $V_D=0$ ) in the nvDFF.

The ReRAM switches to HRS state when negative polarity voltage is applied from 5ns to 8ns and the state variable falls to its minimum value ((Fig. 7-8(a)). The HRS resistance is around 16 k $\Omega$  and the current through the device is a minimal 62  $\mu$ A. This state of the ReRAM is used

to store the logic high value in the nvDFF (data,  $V_D = 1$ ). It should be noted that the state variable remains constant until the SET/RESET voltage is applied and the device is able to store data when supplied power is removed.

Referring to Fig. 7-7, the Master latch consists of transmission gates, TG1 and TG3 and two serially connected inverters (T1/T2 and T3/T4). The nvDFF input is the TG1 input and TG1 switches on when the clock signal, Clk is high. The output of TG1 is connected to the input of the first inverter (T1/T2). The output of T1/T2 is connected to the T3/T4 input and the input of the Slave latch. The output of T3/T4 is connected to TG3 which switches on during Clk low. TG3 output is connected to the input of T1/T2 to form a feedback loop between T3/T4 output and T1/T2 input. This loop forms the latch part of the segment. The Slave latch has a similar layout to the Master latch but the transmission gates TG2 and TG4 are functionally reversed to their Master latch counterparts. TG2 switches on when Clk is low to connect the output from the T1/T2 inverter to the T5/T6 input.

Next, the function of the nvDFF is explained based on the WRITE/READ operation waveform shown in Fig. 7-9. When Clk is high the data ( $V_D = 1$  or 0) to be written passes through TG1 into inverter T1/T2 and from T1/T2 outputs into inverter T3/T4. The output at T3/T4 ( $V_D$ ) is driven through TG3 during the falling edge of Clk and thus latches the data in the Master.

The inverted data (now  $\overline{V_D}$ ) from T1/T2 is only passed through TG2 when Clk is low.  $\overline{V_D}$  is

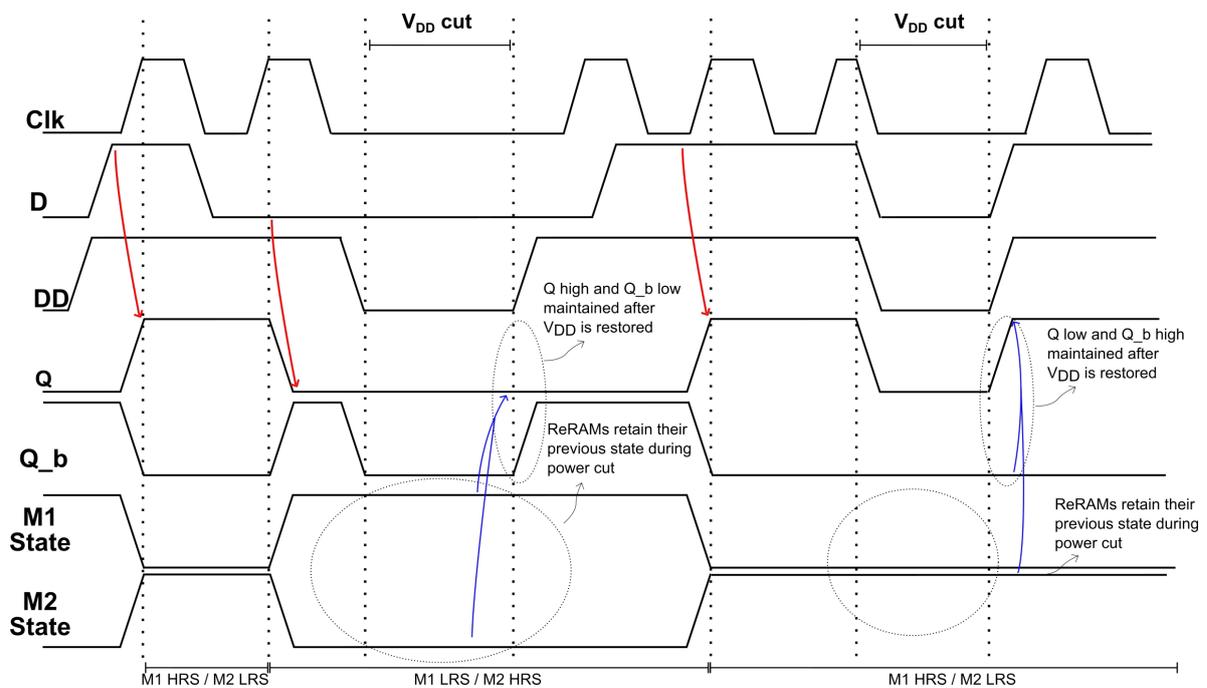


Fig. 7-9. Process waveforms of the nvDFF. Two VDD cut scenarios are shown.

inverted back to  $V_D$  after passing through T5/T6 and becomes the Q output. T7/T8 inverter outputs  $\overline{V_D}$  to the Q\_b line as well as to TG4 which switches on during Clk high and connects Q\_b to the T5/T6 input to form a latch. Thus, Q is always  $V_D$  and Q\_b is always  $\overline{V_D}$ .

When Q is high ('1') and Q\_b is low ('0'), the current flows from the BE to the TE of M1 and reversely from the TE to the BE of M2. As shown in Fig. 7-9, current flow from BE to TE switches M1 to HRS while the current flow from TE to BE switches M2 to LRS. The high resistance of M1 ensures the Q line is pulled to  $V_D$  while the low resistance of M2 pulls down the Q\_b line to ground. Conversely, the opposite happens when Q is low ('0') and Q\_b is high ('1'). M1 is now SET into LRS, pulling the Q line to ground while M2 is in HRS, feeding high ('1') into Q\_b.

As for the NV segment, when the power supply  $V_{DD}$  is disrupted, the ReRAMs M1 and M2 retain their respective resistance states (Fig. 7-9). When  $V_{DD}$  is restored after the first disruption, M1 is in LRS and M2 is in HRS and Q and Q\_b immediately assume their pre-power-disruption values (Q returns to low and Q\_b to high). The nvDFF then functions as normal until the next power cut. M1 and M2 are in HRS and LRS respectively and maintain their states until  $V_{DD}$  is restored.

Similarly, Q and Q\_b immediately assume their pre-power-disruption values of low and high respectively. Both M1 and M2 are always in opposing resistive states as can be seen in the waveform in Fig. 7-9.

The NVDFD transistors operate similar to conventional CMOS transistors where the switching threshold,  $V_M$  is given as:

$$V_M = \frac{rV_{DD}}{1+r} \quad (7.1)$$

where  $r$  is the transistor width ratios.

#### 1. 2-bit Counter:

The 2-bit counter is created out of 2 nvDFDs (FF1 and FF2) and an XOR gate (Fig. 7-10) and the process waveform is given in Fig. 7-9. Q1 and Q2 corresponds respectively to bit '0' and bit '1' of the counter and the counter counts up by one bit at each clock signal rise.

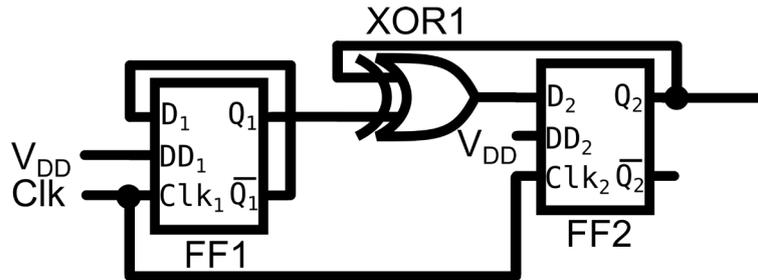


Fig. 7-10. Two nvDFFs, FF1 and FF2 combined with an XOR gate to form a NV 2-bit counter.

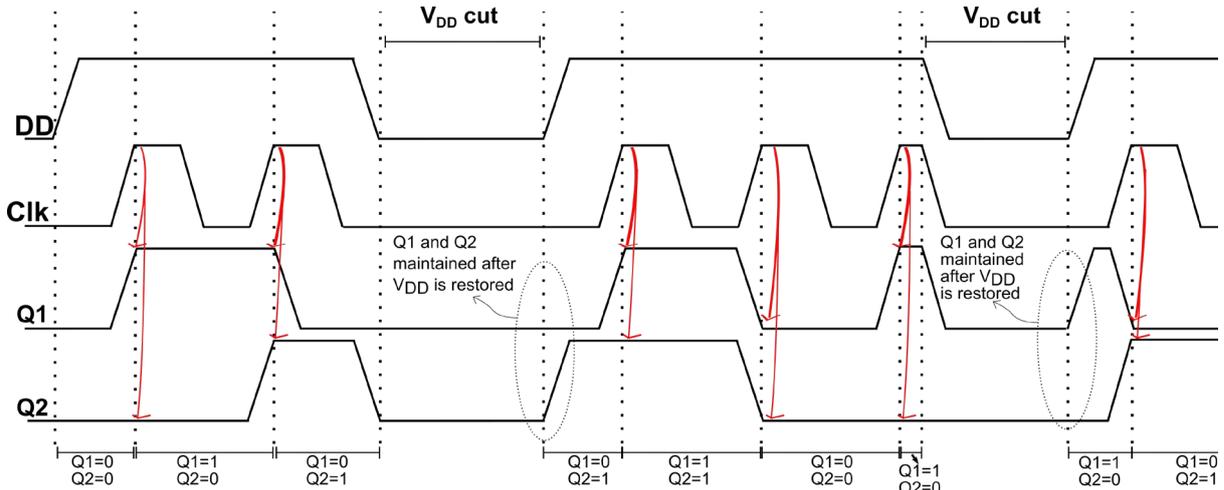


Fig. 7-11. Process waveforms of the nvDFF-based 2-bit counter showing recovery after  $V_{DD}$  cut.

In Fig. 7-11,  $V_{DD}$  is first disrupted after the second Clk pulse when Q1 and Q2 are ‘01’. When power is restored Q1 and Q2 immediately restore their 01 values and the counter continues from before the power cut. The second  $V_{DD}$  cut happens when Q1 and Q2 are ‘10’ and they successfully restore after power restoration.

## 2. 4-bit Shift-Register

Next, four nvDFFs (FF1, FF2, FF3, and FF4) are serially connected to form a 4-bit shift-register (Fig. 7-12). At the first Clk rising edge D1 follows the Input and becomes high (‘1’). At the next Clk rising edge D1 transmits its high (‘1’) value to D2 and D1 falls to low (‘0’). The same happens for D2 and D3 at the next Clk rising edge and the for D3 and D4 for the Clk rising edge after that.

The  $V_{DD}$  disruption happens after the second Clk pulse in Fig. 7-13. At this point, D2 is high (‘1’) after having received the transmitted value from D1 and D1, D3, and D4 are low (‘0’).

When  $V_{DD}$  is restored, D1, D2, D3, and D4 successfully restore their pre-power-disruption values and the register is not affected by the power disruption. The

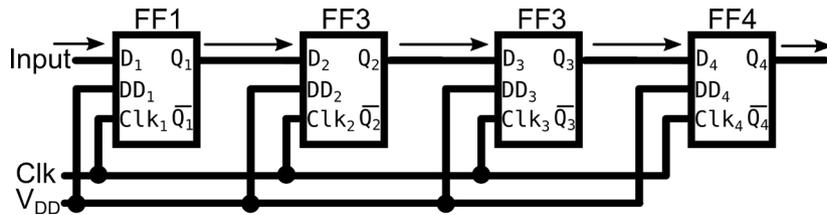


Fig. 7-12. Four nvDFFs (FF1, FF2, FF3, and FF4) combined to form a 4-bit NV shift-register.

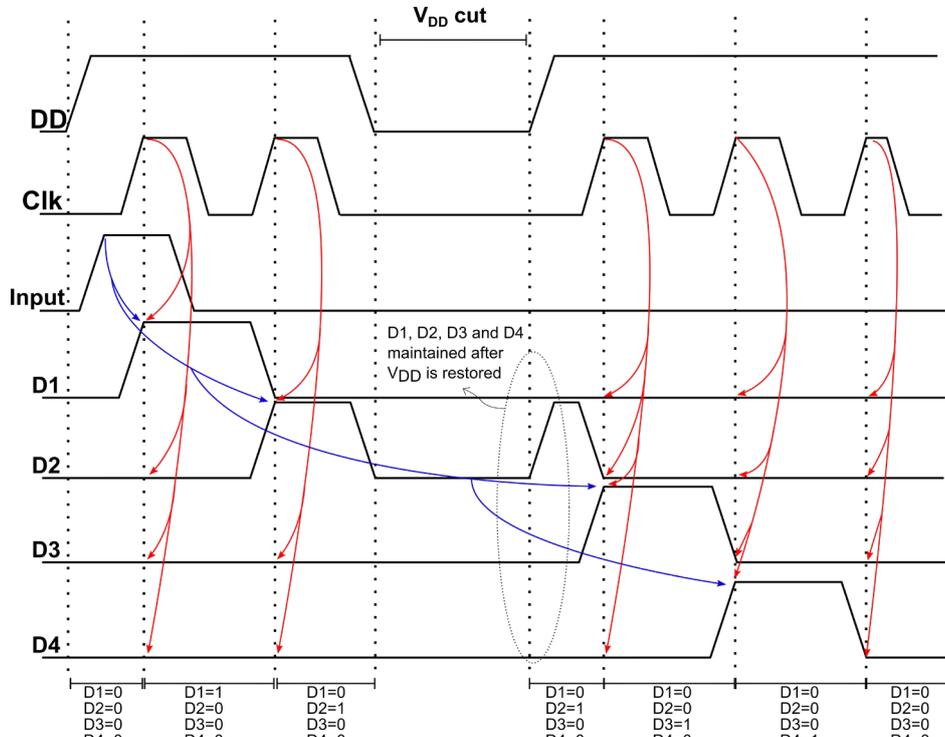


Fig. 7-13. Process waveforms of the nvDFF-based 4-bit shift register showing recovery after VDD cut.

nvDFFs introduce NV into these circuits and they demonstrate ability to retain their previous data even with disruption to the power supply,  $V_{DD}$ .

The following section will look into the simulation results of these proposed circuits and demonstrate the effect of the NV property.

### 7.3.2 Simulation and Results

In this section, the electrical model of the ReRAM used in this work is first presented before the proposed nvDFF design is evaluated discussed. A simulation of the design is carried out with 3 transistor technology node sizes; 32 nm, 45 nm, and 65 nm. The EDA tool used is LTSpice and the transistor models are obtained from [35].

The nvDFF performs like a conventional DFF during normal operations but the ReRAMs retain their resistive states during power interruption and the nvDFF is able to recover its previous configuration. As this design is a modification of the conventional DFF and ReRAMs

are CMOS back-end-of-line (BEOL) compatible, the nvDFF can be easily adapted to current manufacturing processes.

1. The nvDFF Electrical Behaviour:

The waveform for the nvDFF operating under normal DFF behaviour for 45nm node is plotted in Fig. 7-14. The design functions as expected and the NV segment does not have an effect on normal DFF behaviour.  $V_{DD}$  interrupts are then performed to test Restore function under two conditions;

- Restore Q='0' and Q\_b='1'
- Restore Q='1' and Q\_b='0'

The results are shown in Fig. 7-15 and Fig. 7-16 respectively. In both figures, the NV segment passively captures the data at the time of power interruption and recovers immediately when  $V_{DD}$  normalizes. The setup thus provides a quick RESTORE mechanism with no additional RESTORE-delays and sequences as the NV components are directly incorporated into the DFF layout and no external control circuitry is utilized. For comparison, the failed Restore Q='0' and Q\_b='1' sequence for the vFF is plotted in Fig. 7-17.

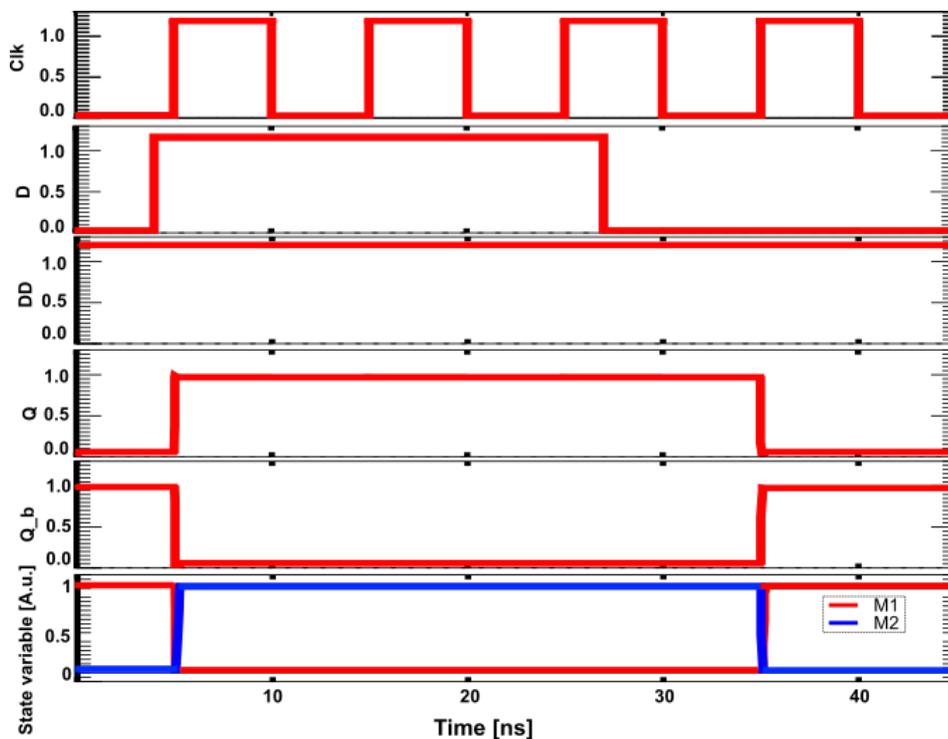


Fig. 7-14. nvDFF displaying normal DFF behaviour.

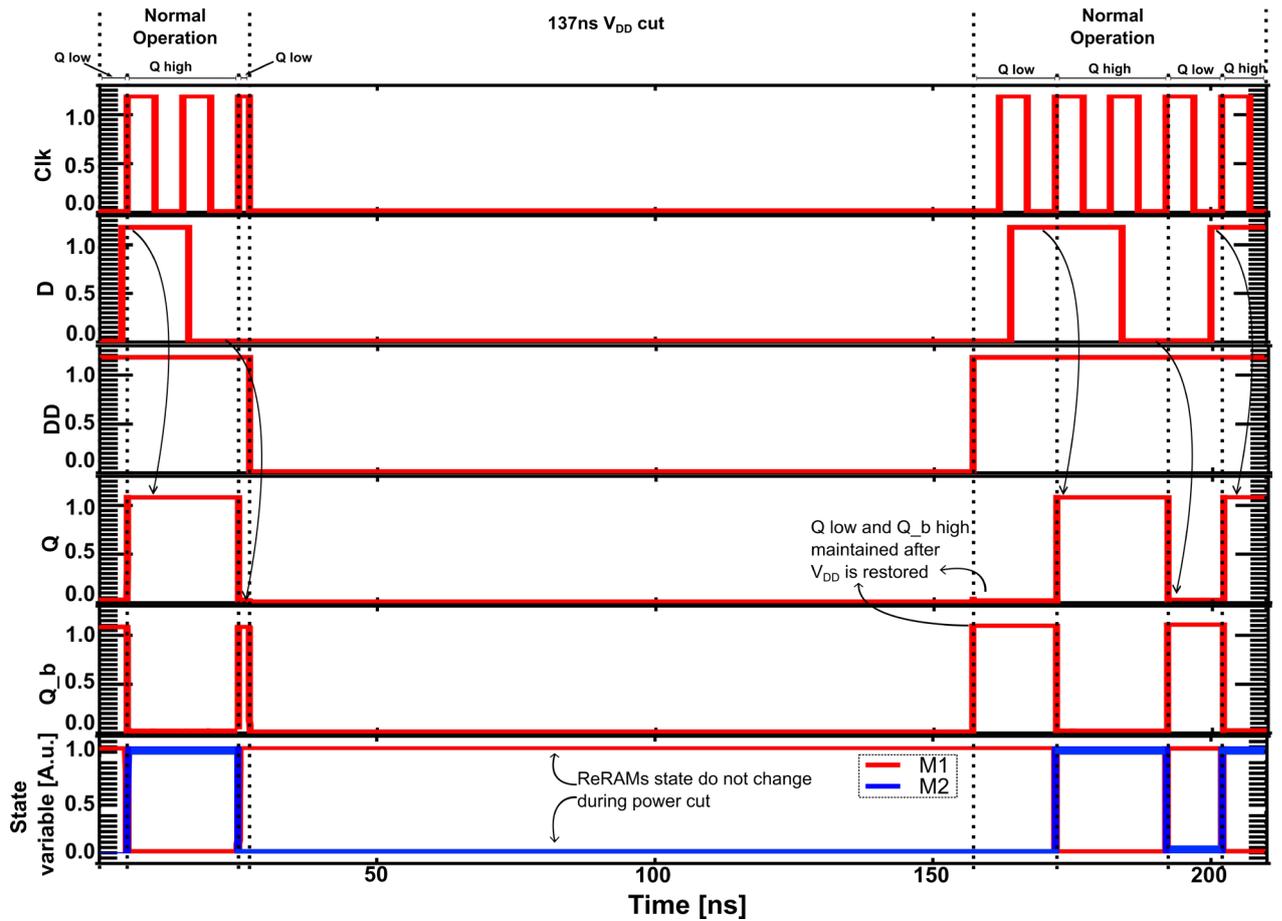


Fig. 7-15. Successful restore of  $Q=0$  and  $Q_b=1$  after 137ns VDD cut. Q switches from 1 to 0 at 25ns before VDD interrupt.

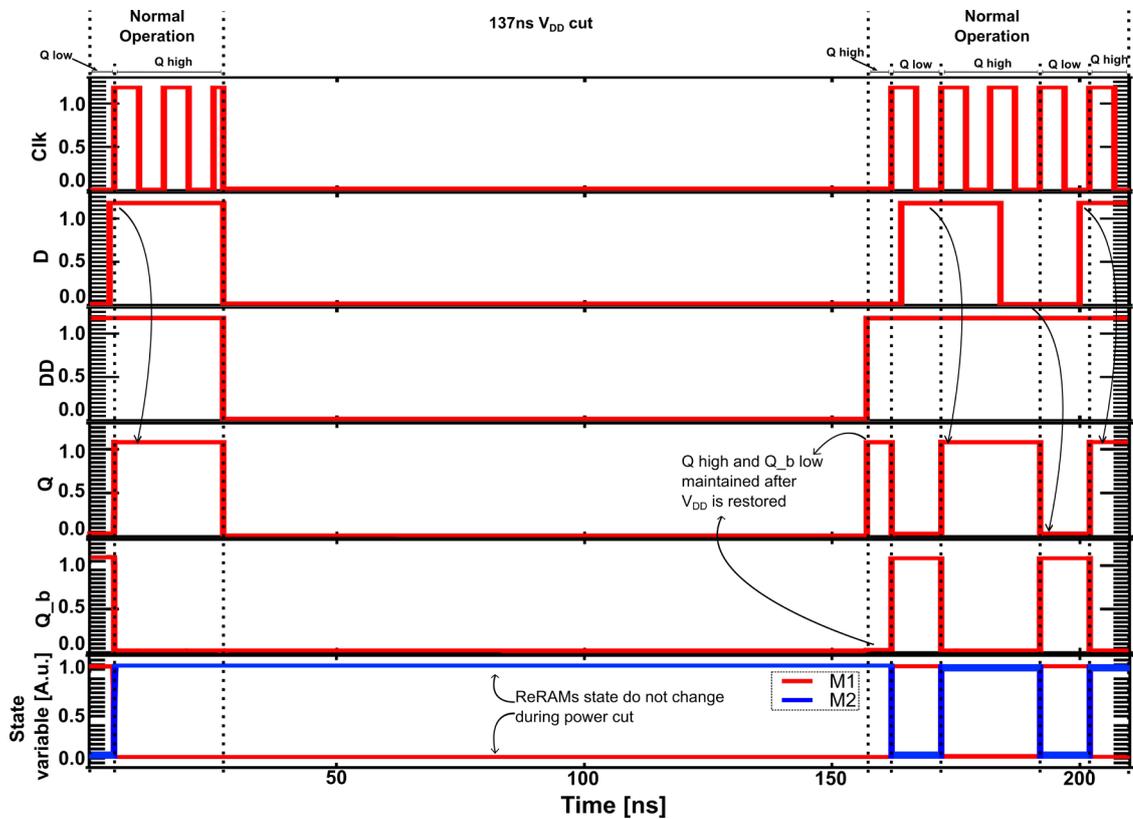


Fig. 7-16. Successful restore of  $Q=1$  and  $Q_b=0$  after 137ns VDD cut.

The  $Q/Q_b$  rising/falling times and clock-to-q delay ( $t_{CQ}$ ) are important parameters for a Flip-Flop to capture the propagation time for the Q input to match the D input and the time required for the Q output to stabilize. Both parameters for the nvDFF are measured for the three transistor technology nodes and compared with

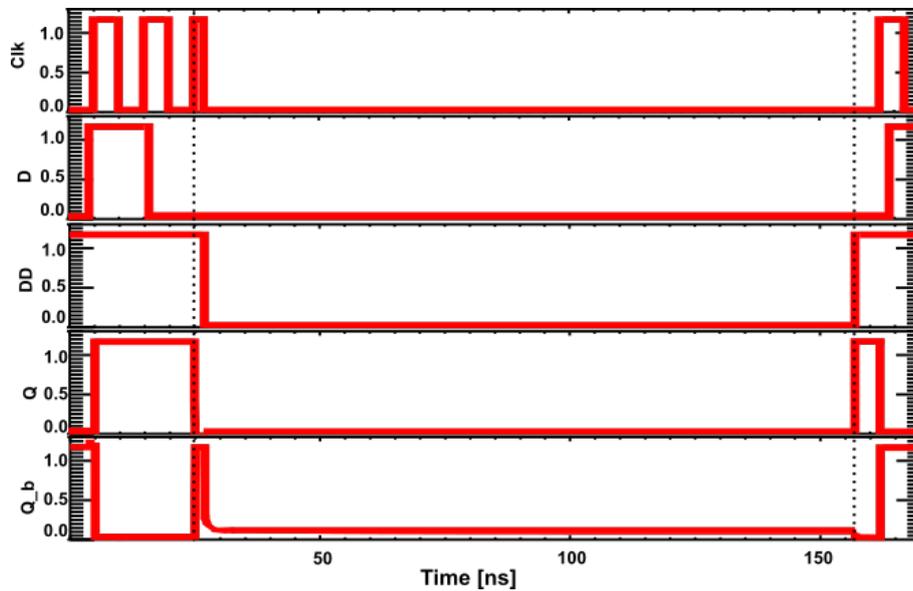


Fig. 7-17. Removal of NV segment results in failure to restore  $Q=0$  and  $Q=1$  after VDD cut.

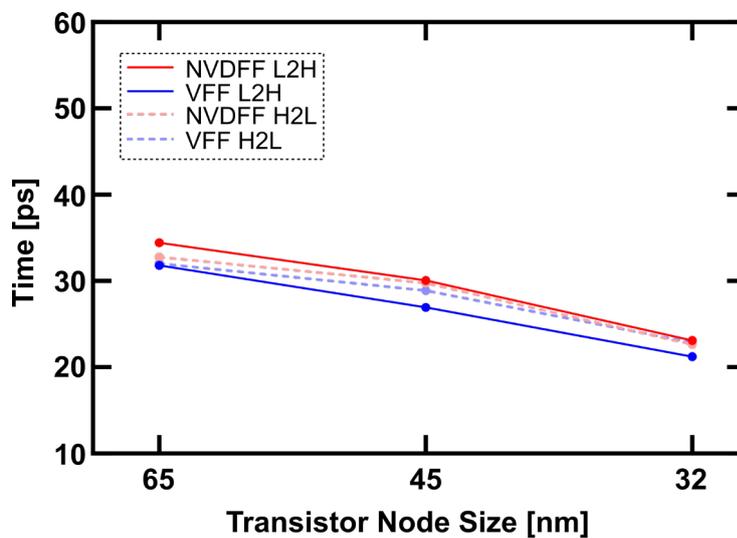


Fig. 7-18. Switching delay timings for nvDFF and vDFF for different transistor technology nodes.

TABLE 7-4. Clk-to-Q Timings for nvDFF and vDFF

Transistor node (nm)	Clk-to-Q delay (ps)		Switching delay (ps)				Fmax (GHz)		Average Restore Time (ps)	Worst-case Restore Time (ps)
			Q rising time / Q_b falling time		Q falling time / Q_b rising time					
	nvDFF	vDFF	nvDFF	vDFF	nvDFF	vDFF	nvDFF	vDFF	nvDFF	vDFF
32	11.54	11.55	23.09	21.22	22.65	22.85	21.87	22.69	19.56	20.34
45	15.9	15.65	30.07	26.94	29.76	28.89	16.71	17.91	34.99	38.43
65	19.07	19.63	34.43	31.8	32.79	32.02	14.88	15.67	38.78	40.01

their vDFF counterparts. The Q/Q\_b rising/falling times are measured during switching from Q='0', Q\_b='1' to Q='1', Q\_b='0' and vice versa. These results are plotted in Fig. 7-18 and listed in TABLE 7-4. TABLE 7-4 also lists the maximum

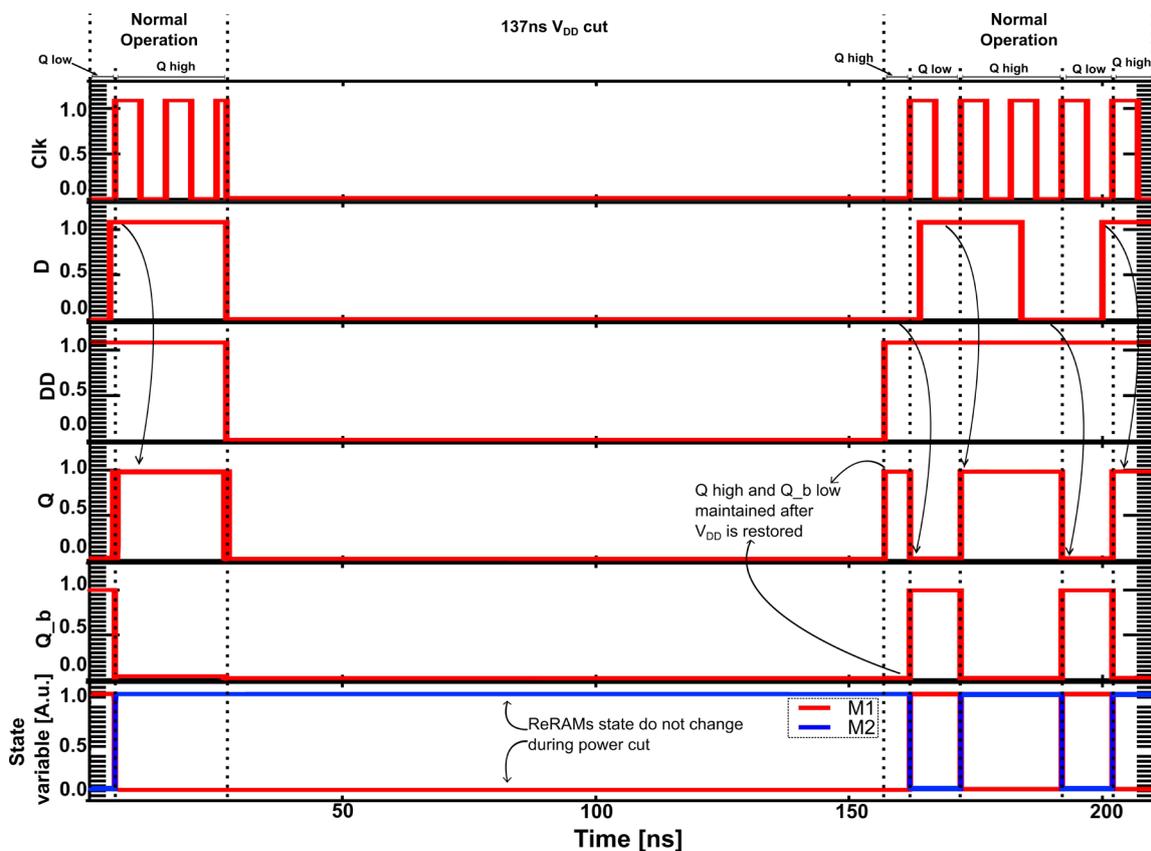


Fig. 7-19. -0.1VDD displaying normal DFF behaviour and successful restore after power supply cut.

TABLE 7-5. nvDFF Clk-to-Q Timings With  $\pm 0.1$  VDD Variability

Transistor node (nm)	Clk-to-Q delay (ps)	Switching delay (ps)		Fmax (GHz)
		Q rising time / Q_b falling time	Q falling time / Q_b rising time	
-0.1	17.81	33.46	32.10	15.2544131
+0.1	14.42	27.64	27.25	18.21991437

frequency (Fmax) for the flip-flops using the measured switching delay timings and the average and worst-case RESTORE times for the nvDFF.

As the nvDFF has a passive STORE mechanism, the nvDFF's minimum required STORE time for a successful RESTORE after power failure should be regarded as similar to the switching times in TABLE 7-4. The nvDFF's passive RESTORE ensures an almost immediate RESTORE after start-up with average RESTORE timings of 31.11 ps for RESTORE0 and RESTORE1. The worst-case RESTORE

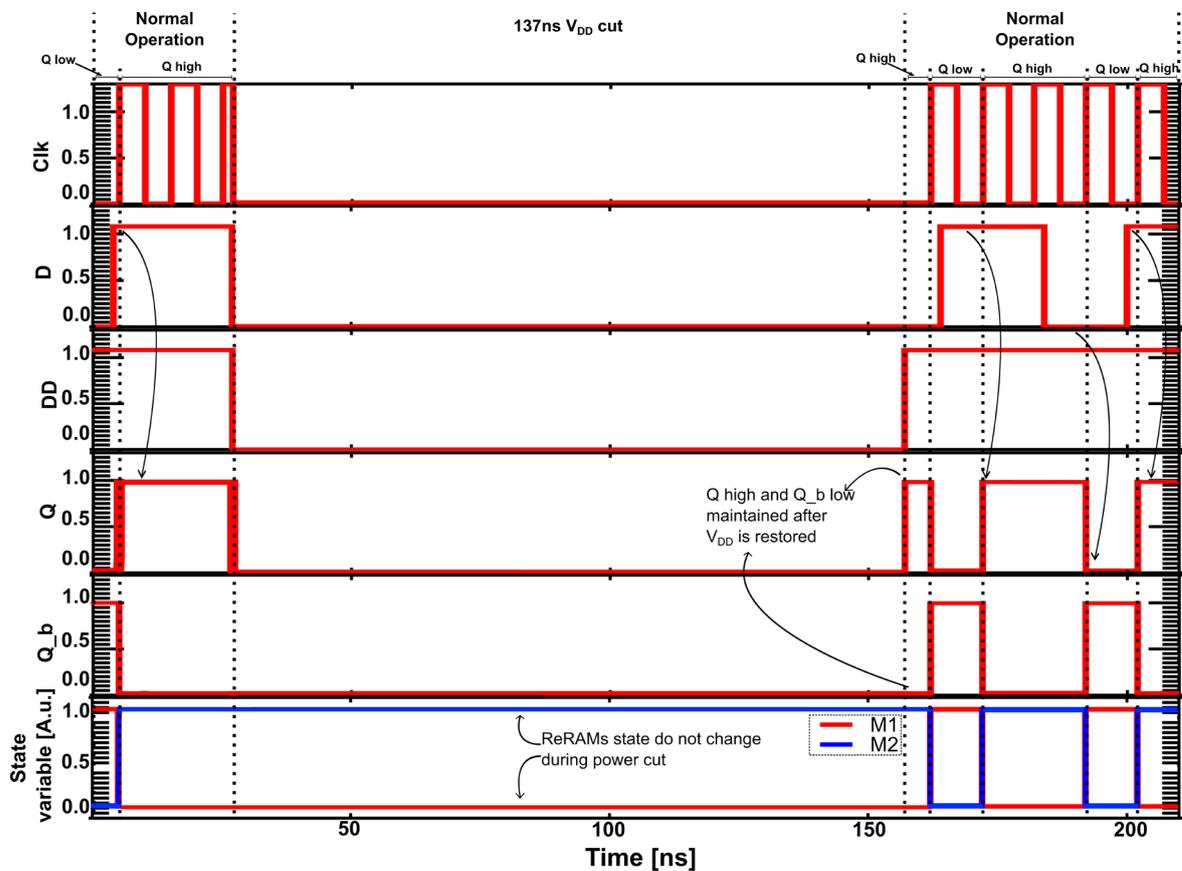


Fig. 7-20. +0.1VDD displaying normal DFF behaviour and successful restore after power supply cut.

measurements are obtained from the RESTORE simulations with the longest delay and show an average of 38.79 ps.

The Clk-to-Q delay timings of both nvDFFs and vDFFs are within range of each other because Clk-to-Q propagation is controlled by the transistors and the transistor technology node has a dominant effect on this timing. On the other hand, the rise and fall times of the nvDFFs are higher than the vDFFs for both low to high and high to low conditions. This is because additional time is taken to switch the ReRAM resistances. The nvDFFs are slower by an average of 1.1x and 1.01x for low to high and high to low switching respectively. This are not large values and indicate that nvDFFs comprised of fast-switching-ReRAMs can be viable. This delay is sensitive to the NV technology and using slower switching ReRAMs will increase delays. The nvDFF is then tested under  $V_{DD}$  variability conditions.

$V_{DD}$  of  $\pm 0.1$  V is supplied to the nvDFF which is tested for normal behaviour and  $V_{DD}$  interrupt during CLK high and low. The waveforms for  $V_{DD}-0.1$  V and  $V_{DD}+0.1$  V using 45 nm technology nodes are plotted in Fig. 7-19 and Fig. 7-20 respectively.

The circuit shows robustness and is able to retain data even with  $V_{DD}$  variation. The  $t_{CQ}$  of the nvDFF under  $V_{DD}$  variability is listed in TABLE 7-5.

## 2. Power Dissipation

Measurements for average and worst-case power dissipation of the nvDFF and vDFF are carried out next. Two conditions are tested for both measurements: Q change from '1'-to-'0' or '0'-to-'1' and vice versa for Q\_b.

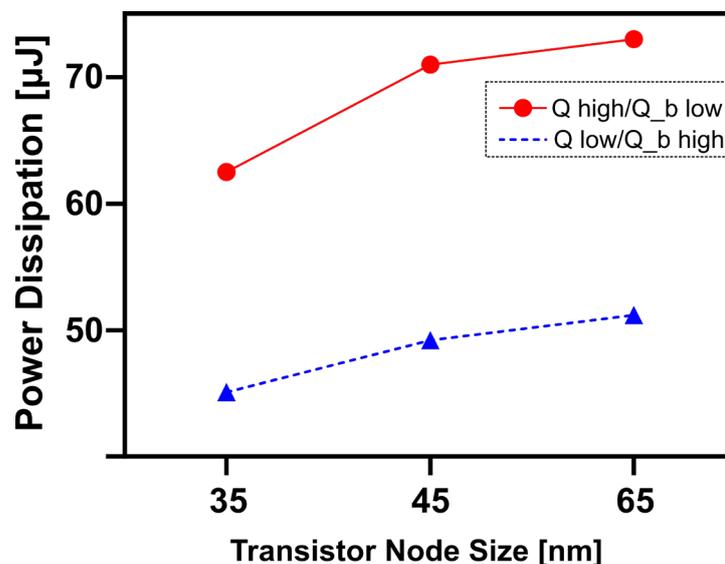


Fig. 7-21. Power dissipation of active components in the nvDFF during active operation.

TABLE 7-6. Average Power Dissipation

	Switching Power Dissipation ( $\mu\text{J}$ )	
	Q=0-to-1, Q_b=1-to-0	Q=1-to-0, Q_b=0-to-1
vDFF	24.156	32.614
nvDFF	70.236	67.394

TABLE 7-7. Worst-case Power Dissipation

	Switching Power Dissipation ( $\mu\text{J}$ )	
	Q=0-to-1, Q_b=1-to-0	Q=1-to-0, Q_b=0-to-1
vDFF	66.453	69.409
nvDFF	85.324	81.968

For the average power dissipation measurement, the average power is measured during Q switching. The peak power is taken for the worst-case measurement. To ensure similarity, the same input voltages are used throughout the tests. The average power dissipation is listed in TABLE 7-6 while the worst-case power dissipation is listed in TABLE 7-7.

Additionally, the average power dissipation of the active components in the nvDFF which are namely, the conventional DFF and NV segments during active operation (T5, M1, and R1 during Q high (Q\_b low) / T7, M2, and R1 during Q low (Q\_b high) in Fig. 7-7) is plotted in Fig. 7-21 for different transistor nodes.

### 3. ReRAM Process Variation Effects

The effects of the ReRAM process variations on the Clk-to-Q delay, Q rise/Q\_b fall time, Q fall/Q\_b fall time, and maximum Q voltage are then investigated and plotted in Fig. 7-22—Fig. 7-25. To simulate the process variation effects, the parameters are measured for each 5% increase/decrease of the LRS and HRS resistances of the ReRAM to a maximum of 10%.

The measured timings are increased at lower resistances as the lower ReRAM resistance increases the time required for Q and Q\_b lines to be driven to their correct value. The worst-case increase is seen in -10% ReRAM resistance with an increase of 8%, 7%, and 15% for the Clk-to-Q delay, Q rise/Q\_b fall time, and Q fall/Q\_b fall time respectively. Conversely, the increase in ReRAM resistance improves the timings by 9%, 7%, and 10% respectively.

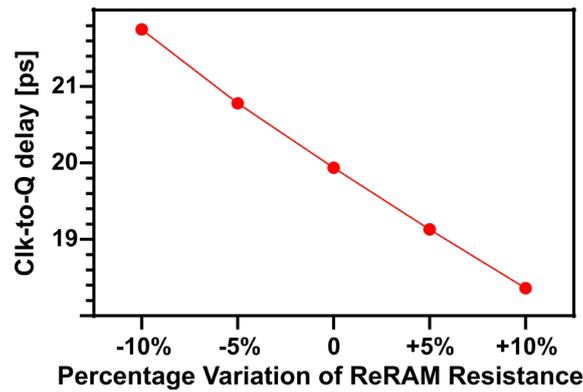


Fig. 7-22. Process variation effect on Clk-to-Q delay time.

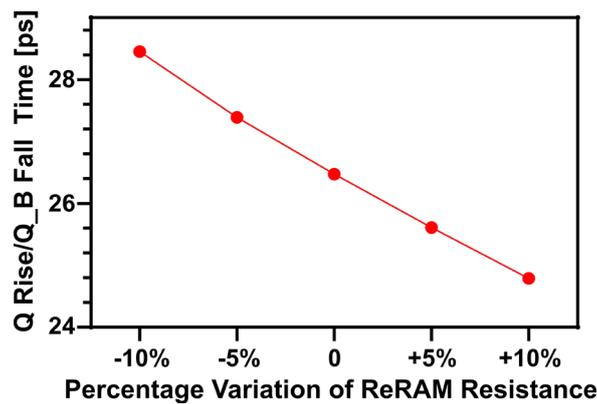


Fig. 7-23. Process variation effect on Q rise/Q\_b fall time.

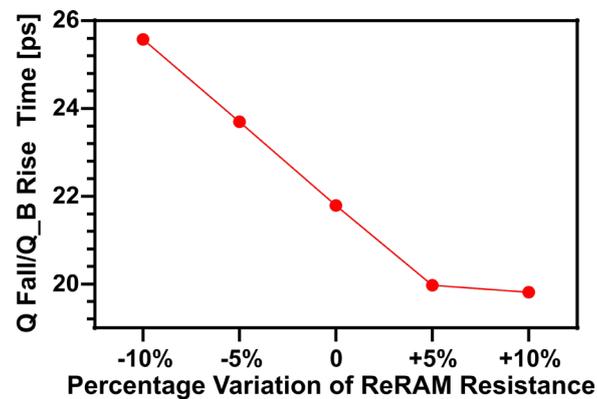


Fig. 7-24. Process variation effect on Q fall/Q\_b rise time.

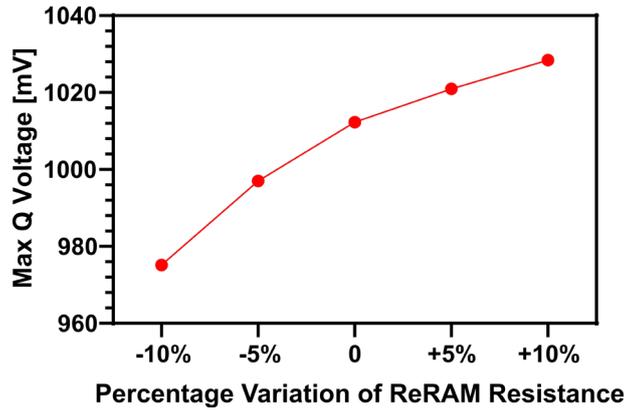


Fig. 7-25. Process variation effect on maximum Q voltage.

The variation in ReRAM resistance also affects the maximum Q voltage; a decrease in ReRAM resistance provides an easier path-to-ground for the current thereby lowering the Q voltage and vice versa when the ReRAM’s resistance is increased. The average change in maximum Q voltage however is low at 2%.

#### 4. Comparison of nvDFF Criteria

In this section an evaluation of nvDFF criteria such as cell area, STORE time, voltages, and power, and endurance is carried out with the benchmarks of other nvDFF circuits from [148], [149], [150], [151], [152], and [153]. TABLE 7-8 provides the summary of this comparison. The ReRAM endurance values of  $10^8$  are obtained from literature [154].

TABLE 7-8. nvDFF Design Comparison

Type	This work	[148]	[149]	[150]	[151]	[152]	[153]
Structure	16T2R	8T2R	12T	19T2R	6T2C	6T2R	7T2PCM
NV device	ReRAM	ReRAM	SONOS	MRAM	Fe. C	ReRAM	PCM
T <sub>STORE</sub>	—*	10ns	4ms	6ns	200ns	100ns	200na
V <sub>STORE</sub>	1.2/-1.2	1.8/-1.6	-10/11	N/A	3V	N/A	3V
P <sub>STORE</sub>	70μJ	64–75μJ	N/A	N/A	N/A	N/A	150μJ
Endurance	$10^8$	$2 \times 10^8$	$10^6$	Inf.	$10^8$	N/A	$10^6$

\*nvDFF passive store occurs during normal vDFF operation

#### 5. Incorporating the nvDFF in Complex Circuits

In this section the nvDFF is used to build the complex circuits discussed in the previous section and the NV character is tested by interrupting power supply.

The same inputs as the ones used in the VDD variability tests are supplied to the 2-bit counter circuit (Fig. 7-10) and the VDD, Clk, Q1 and Q2 outputs are plotted in Fig. 7-26. The counter counts up normally between 0—26 ns before the VDD is interrupted. A1 and A2 interrupts are performed during Clk high and B1 and B2 are

interrupts during Clk low. Both nvDFFs retain their state during the interrupt and the counter operation is not disturbed.

For the next design, four nvDFFs (FF1, FF2, FF3, and FF4) are incorporated to form a 4-bit shift-register (Fig. 7-12). The circuit undergoes the same testing conditions as the 2-bit counter and the VDD, Clk, Input, D1, D2, D3, and D4 waveforms are plotted in Fig. 7-27. Similarly, the normal shift-register operation occurs from 0-26 ns where one bit is shifted from FF1→FF2→FF3→FF4 (D1→D2→D3→D4). The input is raised at 22 ns to initialize another one-bit sequence where D1 goes high for the second time at the third Clk rising edge. VDD is again interrupted during CLK high at A1 and A2 and during Clk low at B1 and B2. The bit shifting is not affected and the D1→D2→D3→D4 sequence continues without disturbance.

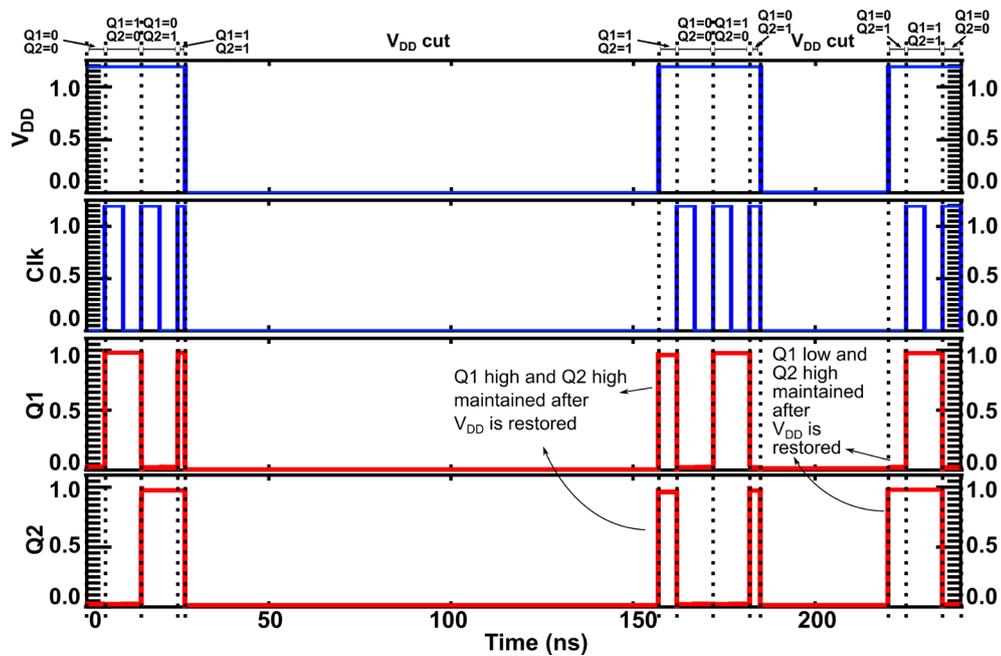


Fig. 7-26. VDD, Clk, Q1, and Q2 waveforms for NV 2-bit counter. Regular 2-bit counter operation occurs between 0—26ns before VDD interrupt tests during Clk high (A1 and A2) and Clk low (B1 and B2).

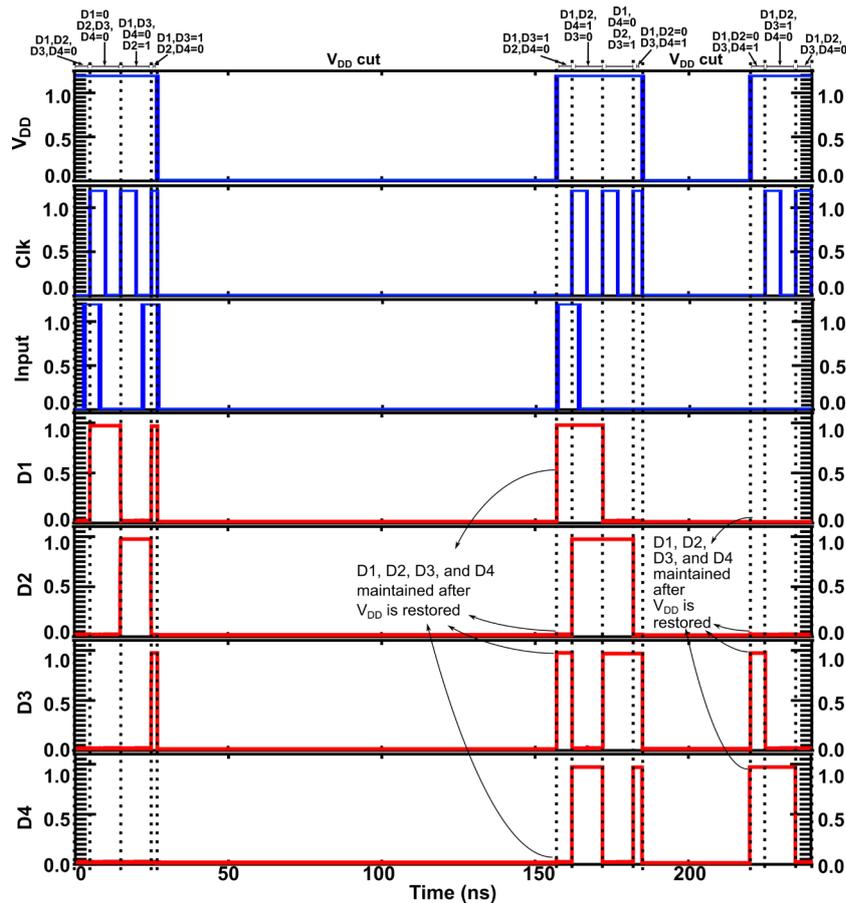


Fig. 7-27. VDD, Clk, Input, D1, D2, D3, and D4 waveforms for NV 4-bit shift-register. Regular shifting operation occurs between 0—26ns before VDD interrupt tests during Clk high (A1 and A2) and Clk low (B1 and B2).

## 7.4 Non-Volatile DRAM

The DRAM which consists of one access transistor and one charge-storing capacitor (1T1C) in its most basic form is facing massive scaling challenges with current feature sizes at >10nm, lagging behind processor fabrication processes [155]. A major limiting factor is in the scaling of the capacitor; the capacitor's aspect ratio is delicate and must be considered at every reduction in size. Downsizing the capacitor also reduces the cell capacitance, increasing charge leakage and variable retention time (VRT) [156]. The capacitor's charge leakage also makes the DRAM cell volatile and requires energy-consuming refresh cycles during operation.

Replacing the capacitor with a ReRAM would provide a solution to the challenges above. This section introduces a one transistor, one resistor, one ReRAM (1T2R) nvDRAM cell architecture that offers a solution to conventional 1T1C DRAM scaling whilst eliminating the need for energy-consuming refresh cycles by making use of the ReRAM's NV.

### 7.4.1 The Proposed nvDRAM

Fig. 7-28 shows the schematics of the proposed 2T1R cell consisting of two access transistors and one ReRAM which is the storage component.

The transistor,  $T_W$  is the WRITE access transistor and is connected to the bitline, wordline, and node N.  $T_W$  is switched on during the WRITE sequence by the wordline voltage, WL at its gate to form a path between the WRITE bitline voltage, BL and node N. At the same time, the READ segment is left floating and the voltage at N is the voltage flowing through the ReRAM:

$$V_N = V_{ReRAM} = V_{Bitline} \quad (7.2)$$

The READ sequence is handled by the READ transistor,  $T_R$  which provides a dividing resistance and is connected to the READ voltage line, RL and node N. This forms a simple voltage divider at N and the value of  $T_R$  has to be adjusted depending on the ReRAM's highest resistance to provide accurate Boolean outputs during READ.

Taking the voltage at N and the output,  $V_{OUT}$  during READ is given as:

$$V_N = V_{OUT} = \left( \frac{R_{ReRAM}}{R_T + R_{ReRAM}} \right) V_{READ} \quad (7.3)$$

When the ReRAM is in HRS,  $R_{ReRAM}$  should be much larger than  $T_R$  ( $R_{ReRAM} \gg T_R$ ) so that  $V_{OUT1} \approx V_{READ}$  and stored DATA1 is fed to the output. Alternatively, when the ReRAM is in LRS, ( $T_R > R_{ReRAM}$ ) so that  $V_{OUT2}$  is a fraction of  $V_{READ}$  and much lower than  $V_{OUT1}$  which is then read as stored DATA0 is at the output.

The WRITE and READ operations of the nvDRAM are shown in Fig. 7-28 and Fig. 7-29 and are described as follows:

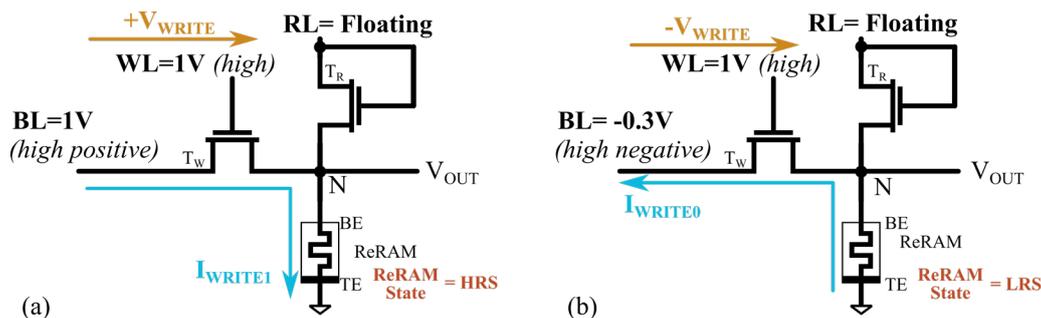


Fig. 7-28. 2T1R nvDRAM cell storing logical '1' and '0'.

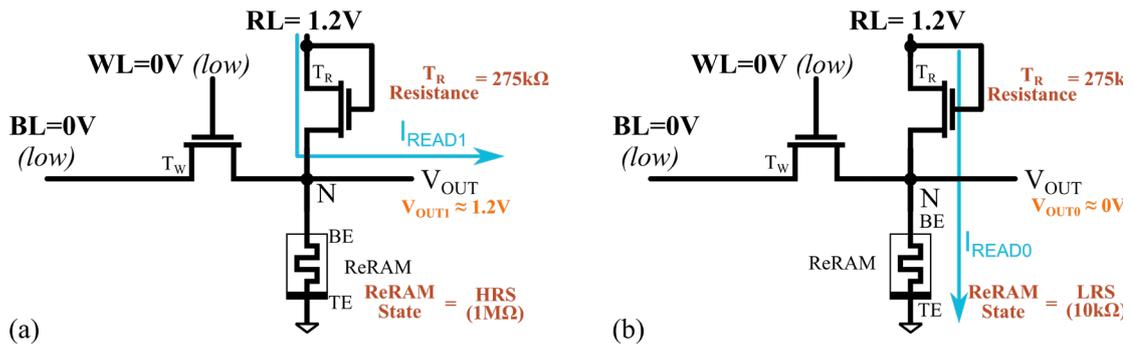


Fig. 7-29. 2T1R nvDRAM READ1 and READ0 operation.

## 1. WRITE Operation

As shown in Fig. 7-28, the  $V_{\text{READ}}$  line is left floating during WRITE to remove the path from N through  $T_R$ . The wordline is set to high to turn on  $T_W$ . This enables the bitline voltage to be driven to node N which will correspondingly determine the resistance of the ReRAM.

As the ReRAM is bipolar, the bitline voltage is positive polarity for WRITE 1 and switches the ReRAM from LRS to HRS (Fig. 7-28(a)). Alternatively WRITE 0 requires a negative polarity Bitline voltage which switches the ReRAM from HRS to LRS (Fig. 7-28(b)). The ReRAM's resistance does not change if it is already in HRS during WRITE 1 and LRS during WRITE 0.

## 2. READ Operation

For the READ operation, the wordline voltage is kept at 0 to turn off  $T_W$  and remove the connection between N and the bitline.  $V_{\text{READ}}$  is supplied and subsequently divided by both  $T_R$  and the ReRAM's resistance at N. If the ReRAM is in HRS (stored DATA1), the ReRAM resistance is much higher than the resistance of  $T_R$  (HRS resistance= $1\text{M}\Omega$ ,  $T_R = 275\text{k}\Omega$ ) and blocks the current path to ground (Fig. 7-29(a)). The output voltage,  $V_{\text{OUT}}$  is around  $V_{\text{READ}}$  levels and is therefore high (1).

For READ0, the ReRAM is in LRS ( $R_{\text{LRS}} = 10\text{k}\Omega$ ) and has a much lower resistance than  $T_R$ , therefore establishing a current path from N to ground and  $V_{\text{OUT}}$  levels drop to low (Fig. 7-29(b)).

In this work, a positive polarity  $V_{\text{READ}}$  is used to minimize external circuitry and wiring. This however, produces a positive current flow through the LRS-configured ReRAM (DATA0) similar to the current flow during WRITE 1 (Fig. 7-28(a) and

Fig. 7-28(b)), causing the ReRAM's resistance to begin drifting to HRS. This eventually leads to a catastrophic resistance flip after multiple READ0 sequences.

This is shown in TABLE 7-9 where simulation is carried out using repeated READ pulses without the Restore segment. The ReRAM is switched to LRS (DATA0) then fed with 400 READ pulses of differing READ voltage amplitudes. The higher amplitudes lead to lower READ cycles-after-WRITE before the ReRAM's state variable drifts from its LRS state to HRS. The nvDRAM cell's output voltage at the point of failure,  $V_{OUT0F}$  is included and shows that it does not significantly differ from the output voltage of DATA1,  $V_{OUT1}$ . Additionally, the  $V_{OUT1}$  in TABLE 7-9 shows that the 1.2 V READ voltage is desirable to maintain a high output voltage during READ1.

It is thus necessary to include a negative polarity Restore pulse after the Read pulse to recover the ReRAM's resistance drift during READ0. This is incorporated into the design with the READ operation pulse consisting of a 100 ps, +1.2 V Read pulse followed immediately by a 100 ps, -1.2 V Restore pulse.

The reverse scenario for the HRS-configured ReRAM (DATA1) during the negative Restore pulse is not a concern as the HRS resistance is large (in  $\sim M\Omega$  range) and the short duration of the Restore pulse is not sufficient to affect the ReRAM.

The WRITE and READ operations of the proposed design happen passively and the high retention time of ReRAMs ( $\sim 10^7$  seconds [43]) provides a passive NV solution that eliminates the periodic refresh requirements endemic in conventional DRAMs.

TABLE 7-9. READ Voltages and Cycles to Failure without Restore

READ Voltage (V)	READ cycles to failure without Restore	$V_{OUT0F}$ (mV)	$V_{OUT1}$ (mV)
0.4	66	267.619	313.747
0.6	56	401.102	471.162
0.8	35	531.150	628.056
1.0	30	669.370	783.594
1.2	26	805.339	937.015

### 3. Electrical Characteristics

Fig. 7-30 shows the simulated WRITE1 (dotted blue line) and WRITE0 (solid red line) operation waveforms for the 2T1R cell.

For WRITE1, a pulse-width of 10 ns length and 1V amplitude is supplied to the bitline. The wordline is raised to 1 V for the same duration to activate access transistor, TW. It can be seen in Fig. 7-30 that the ReRAM state variable, which represents the ReRAM resistance switches accordingly from 1 (LRS) to 0 (HRS).

A negative polarity pulse-width of 25 ns length and -0.3 V amplitude is supplied for WRITE0 operation (Fig. 7-30). In this case the state variable of the ReRAM switches from '0' to '1' and the ReRAM is in LRS.

The READ operation is demonstrated in Fig. 7-31. The output of the cell to the sense amplifiers, VOUT is provided during the positive-polarity READ pulse. VOUT high is represented by the blue-dotted line while VOUT low is represented by the solid red line. VOUT high (VOUT1) is around 1 V and VOUT low (VOUT2) is around 0 V during the READ pulse.

The following negative polarity pulse-width is the Restore pulse to recover any resistance drift in the ReRAM during READ. All values at VOUT at this time should be disregarded.

This way, the READ operation pulse is the same for both READ0 and READ1 and both READs can be controlled by the same singular external source circuit. The ReRAM state variable shows that the ReRAM resistances are not affected by this

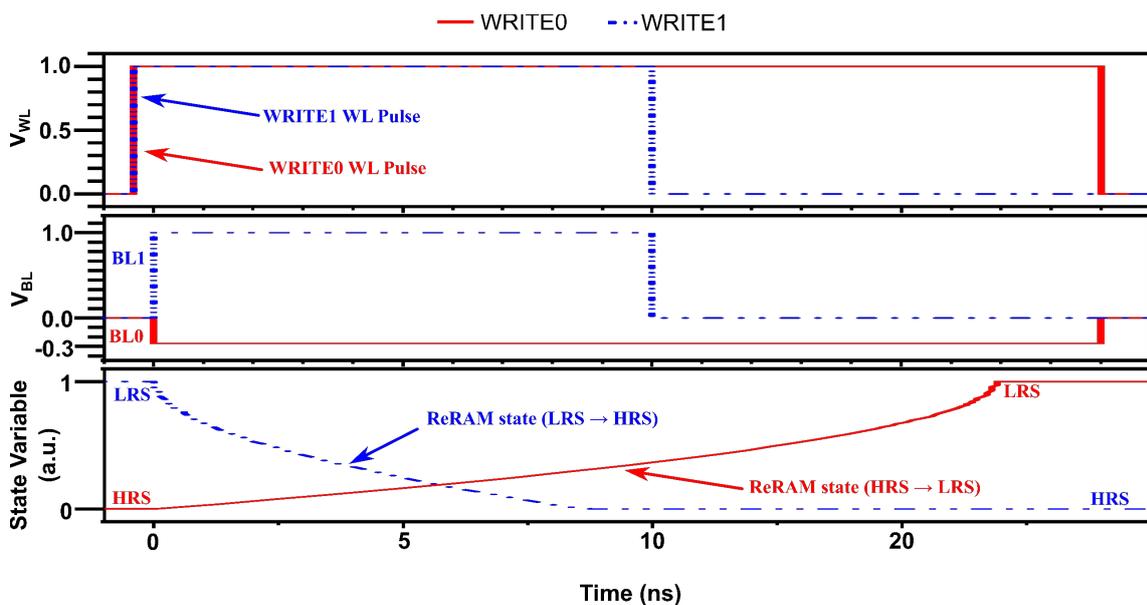


Fig. 7-30. WRITE1 and WRITE0 operations of the 2T1R nvDRAM cell.

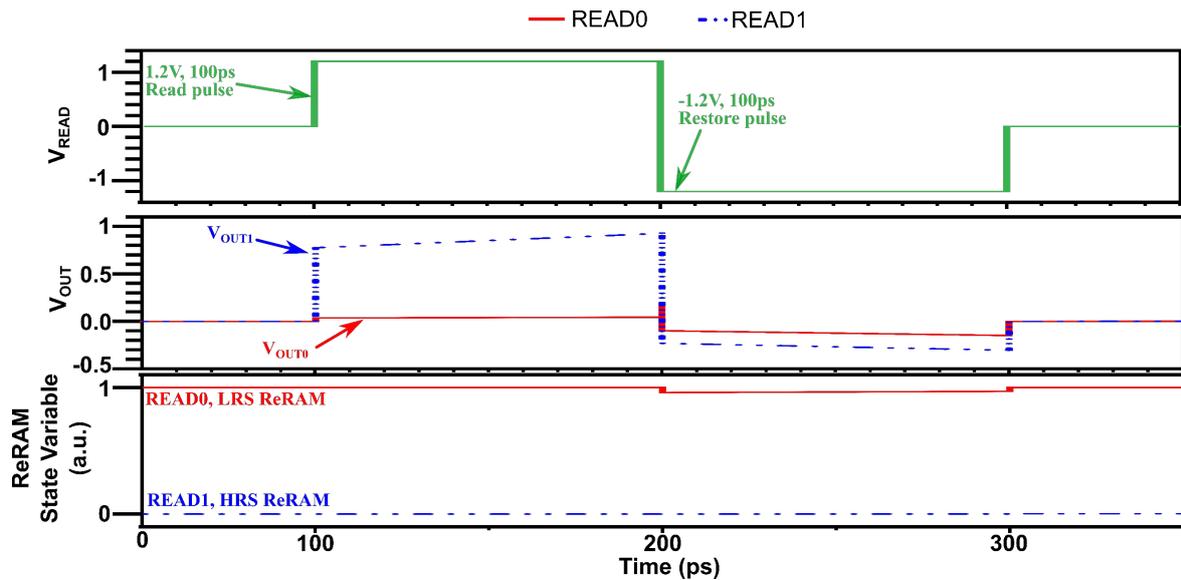


Fig. 7-31. READ1 and READ0 operations of the 2T1R nvDRAM cell.

READ scheme.

#### 7.4.2 2T1R nvDRAM Assessment

The proposed 2T1R cell is then tested for typical DRAM performances as well as its NV characteristics with simulations using LTSpice. The tests carried out are as follows:

- 1) Bit-flipping by using sequentially alternative WRITE operations to confirm complete ReRAM resistance switching.
- 2) Bit Hammer test. Repeated READ pulses are supplied to the cell after WRITE0 and WRITE1 to ensure no bit-flip error occurrences.

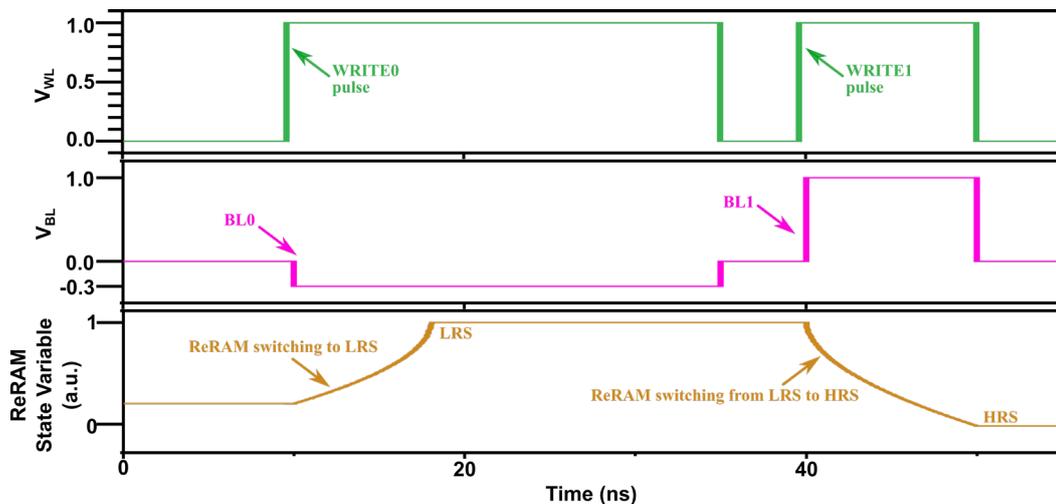


Fig. 7-32. Bit-flip 0-to-1 test of the 2T1R nvDRAM cell.

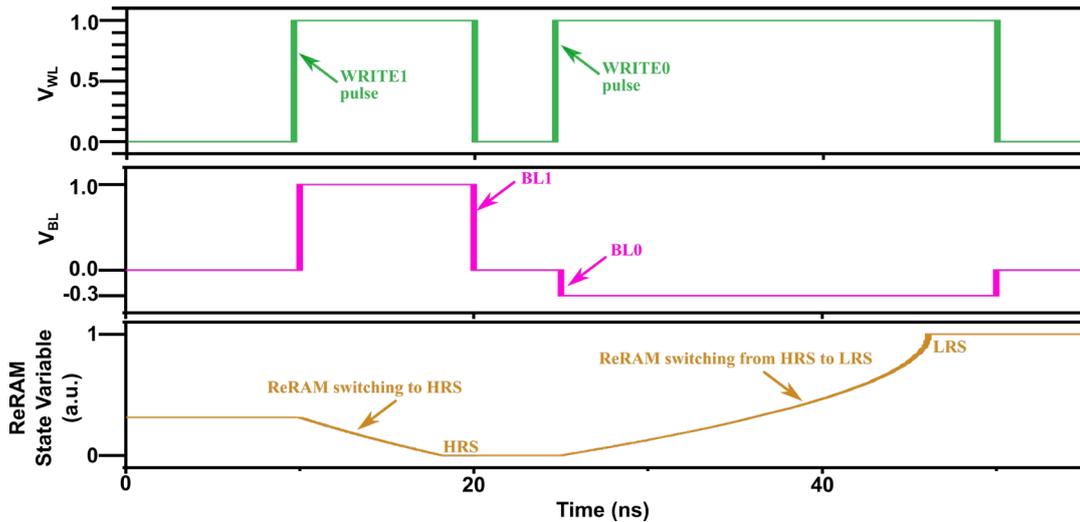


Fig. 7-33. Bit-flip 1-to-0 test of the 2T1R nvDRAM cell.

3) Measurements for WRITE and READ delays, energy dissipation, and EDP parameters of the nvDRAM cell. The results are then compared with existing DRAM cell architectures.

The transistors used in the simulations are predictive technology models (PTM) from [118].

Fig. 7-32 shows the simulation results for ‘0’-to-‘1’ and Fig. 7-33 shows the simulation for ‘1’-to-‘0’ tests. The ReRAM state variable in both simulations begin at an initial state and are completely switched to either HRS or LRS before being switched again to opposing states. For the ‘0’-to-‘1’ test, the ReRAM completes a full switch from LRS to HRS and the ReRAM completes a full switch from HRS to LRS

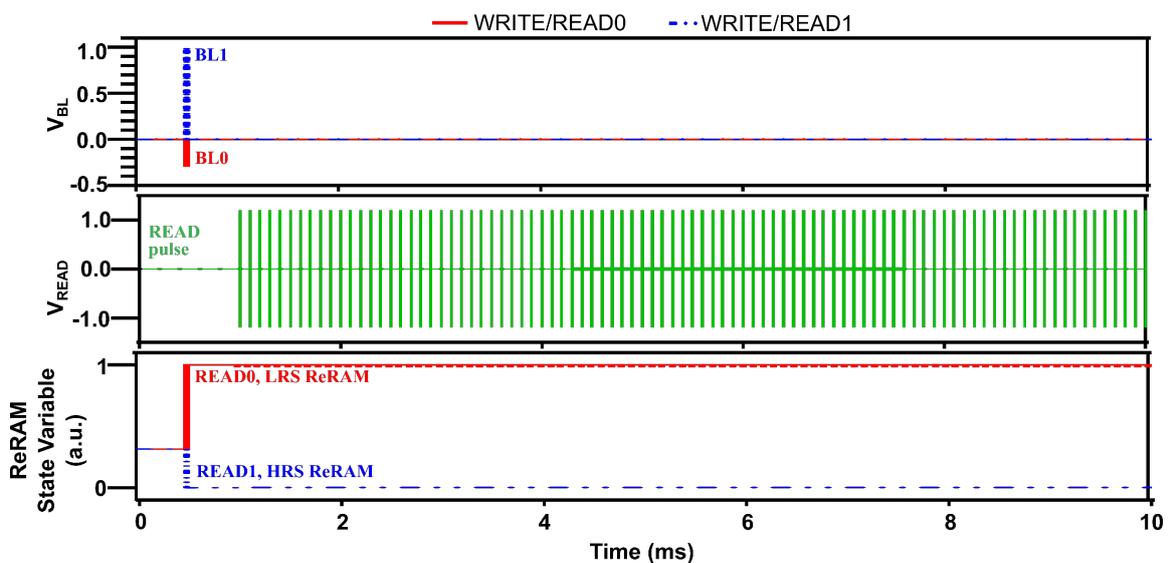


Fig. 7-34. Bit hammer test for stored DATA1 and DATA0.

TABLE 7-10. WRITE and READ delay, Energy Dissipation, Retention Time, and EDP

Transistor Node (nm)	Operation	Cell																			
		3T1D [157]			B3T [158]				4T1D1R [64]				4T1RP [64]				2T1R [this work]				
		Delay (ps)	Energy dissipation (nJ)	EDP (ps·nJ)	Retention Time (ns)	Delay (ps)	Energy dissipation (nJ)	EDP (ps·nJ)	Retention Time (ns)	Delay (ps)	Energy dissipation (nJ)	EDP (ps·nJ)	Retention Time (ns)	Delay (ps)	Energy dissipation (nJ)	EDP (ps·nJ)	Retention Time (ns)	Delay (ps)	Energy dissipation (nJ)	EDP (ps·nJ)	Retention Time (ns)
22	WRITE0	89.43	-	7.25e4	319.4	136.4	-	1.39e5	362.7	123.8	-	1.05e5	235.9	188.9	-	2.00e5	267.9	16637	6.53e-4	10.86	inf
	WRITE1	158.0	810.9	1.28e5		235.6	1016	2.39e5		218.7	845.9	1.85e5		326.2	1059	3.45e5		7633	3.54 e-4	2.704	
	READ0	172.1	-	1.51e5	165.3	-	1.57e5	175.7	-	1.61e5	169.3	-	1.67e5	200	1.66 e-4	3.3e-2					
	READ1	213.9	879.2	1.88e5	162.4	946.9	1.54e5	224.7	916.4	2.06e5	164.0	987.1	1.62e5	200	2.10 e-4	4.2e-2					
	RESTORE0	-	-	-	-	-	-	-	99.25	691.3	6.86e4	117.1	845.5	9.90e4	-	-	-				
	RESTORE1	-	-	-	-	-	-	-	125.3	8.66e4	147.8	1.25e5	-	-	-	-					
32	WRITE0	103.1	-	1.03e5	613.7	157.3	-	1.99e5	697.3	141.5	-	1.49e5	428.7	215.9	-	2.87e5	486.9	16640	7.17 e-4	11.92	inf
	WRITE1	171.6	1002	1.72e5		255.9	1265	3.24e5		235.5	1054	2.48e5		351.2	1331	4.67e5		7797	4.43 e-4	3.452	
	READ0	187.3	-	2.05e5	180.4	-	2.13e5	191.2	-	2.18e5	184.2	-	2.26e5	200	1.66 e-4	3.3e-2					
	READ1	229.1	1095	2.51e5	175.6	1179	2.07e5	235.9	1140	2.69e5	175.6	1227	2.16e5	200	2.17 e-4	4.3e-2					
	RESTORE0	-	-	-	-	-	-	-	99.25	870.1	8.64e4	136.3	1063	1.45e5	-	-	-				
	RESTORE1	-	-	-	-	-	-	-	125.3	1.09e5	171.4	1.82e5	-	-	-	-					
45	WRITE0	120.5	-	1.50e5	1112	183.9	-	2.89e5	1263	159.9	-	2.21e5	749.5	224.0	-	3.92e5	851.2	16651	7.78 e-4	12.96	inf
	WRITE1	189.1	1245	2.35e5		282.0	1573	4.44e5		250.9	1384	3.47e5		374.2	1748	6.54e5		8161	4.45 e-4	3.632	
	READ0	211.2	-	2.91e5	203.5	-	3.02e5	215.6	-	3.07e5	207.5	-	3.18e5	200	1.66 e-4	3.3e-2					
	READ1	253.0	1376	3.48e5	188.3	1482	2.79e5	264.6	1422	3.76e5	193.7	1531	2.97e5	200	2.19 e-4	4.4e-2					
	RESTORE0	-	-	-	-	-	-	-	117.6	1022	1.20e5	156.0	1249	1.95e5	-	-	-				
	RESTORE1	-	-	-	-	-	-	-	148.4	1.52e5	196.9	2.46e5	-	-	-	-					

'-' indicates unused operations.

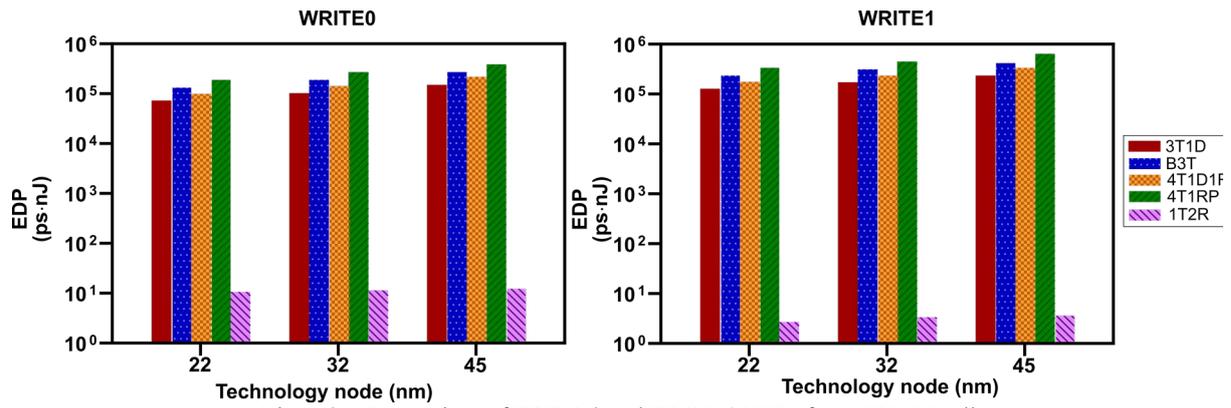


Fig. 7-35. Comparison of WRITE0 and WRITE1 EDPs for nvDRAM cells.

for the ‘1’-to-‘0’ test. The length and amplitude of the WRITE pulse are important criteria and are highly dependent on the ReRAM’s properties. ReRAM switching speeds differ according to the material it is fabricated from and the nvDRAM operation should be tuned appropriately.

Next, repeated READ pulses are supplied after WRITE1 and WRITE0 operations to test for bit-flip errors and the ReRAM’s resistance stability (represented by the ReRAM state variable in Fig. 7-34).

The READ pulse shows no significant effect on the ReRAM state variable. Similar to the WRITE pulse, the Read and Restore segment of the READ pulse in this design are tuned to match the characteristics of the ReRAM.

The nvDRAM WRITE and READ delays, energy dissipation, and EDP for transistor technology nodes of 22 nm, 32 nm, and 45 nm are then measured for the nvDRAM and listed in TABLE 7-10. Also included in the table are the parameters for existing DRAM architectures; namely, a three transistor, one gated diode (3T1D) cell [157], a boosted 3T1D (B3T) cell [158], a four transistor, one gated diode, one ReRAM (4T1D1R) cell [64], and a four PMOS transistor, one gated diode, one ReRAM (4T1RP) cell [64]. The 2T1R cell has larger WRITE delays compared to the other designs due to the switching timings of the ReRAM. This is a limitation of the ReRAM used in this work and can be improved by using faster switching ReRAMs (ie. picoseconds switching ReRAM [159]).

However, the average WRITE energy dissipation is calculated by taking each design’s average WRITE0 and WRITE1 energy dissipation for every technology node and shows that the 2T1R cell has an average of  $1.19 \times 10^3$  nJ lower energy consumption than the compared DRAM cells. This reduction in energy dissipation is a result of the use of a single WRITE transistor in the 2T1R. The 2T1R cell is also

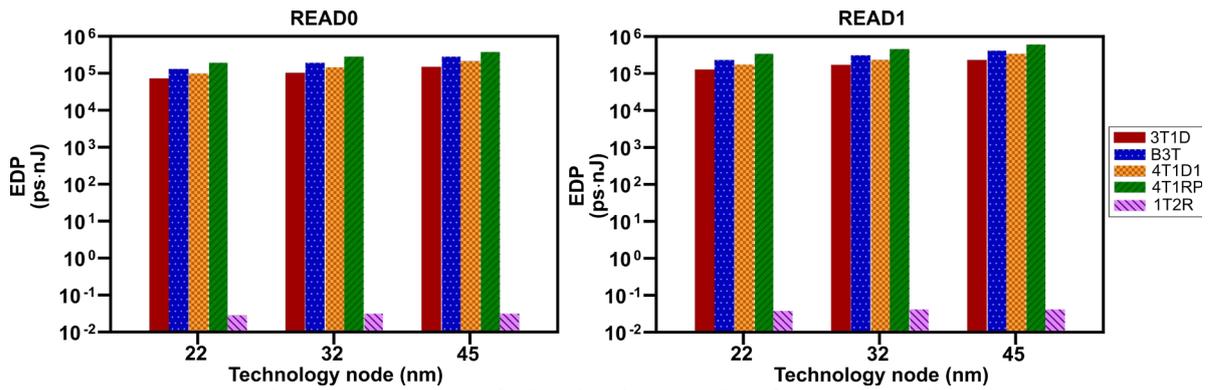


Fig. 7-36. Comparison of READ0 and READ1 EDP for nvDRAM cells.

NV and forgoes the retention time criteria of the compared designs. The 3T1D cell for example has a retention time of 319.4 ns for the 22 nm technology node and has to be refreshed at every retention time interval. Therefore, the energy consumption for a typical usage would be far larger than the values in TABLE 7-8.

The measured EDP of the designs in TABLE 7-10 are then plotted in Fig. 7-35 (WRITE0, WRITE1) and Fig. 7-36 (READ0, READ1) according to technology node sizes. The 2T1R is shown to provide an improvement in EDP and is closer to conventional 3T1C DRAM EDP of  $3.89 \times 10^{-13}$  ps·nJ in [160].

Each design's average READ energy dissipation and EDP for are calculated by taking the average energy dissipation for READ0 and READ1 across transistor sizes. The 2T1R shows lower average READ energy dissipations and EDPs by 1181.8 nJ and  $2.36 \times 10^5$  ps·nJ respectively than the compared cells. This is again due to the reduction in transistor count in the nvDRAM cell.

## 7.5 Summary

The nvD latch successfully introduces NV into the D latch design with the use of two ReRAMs and is able to achieve faster switching than compared designs. The design's NV property allows it to be used in devices where frequent power interruptions are expected or for devices that require intermittent usage, allowing for proper SLEEP mode where the power supply to the memory can be switched off and reducing power consumption of large-scale electronic circuits.

The nvDFF's is advantageous for NVPs where protection from power disruption is crucial or where intermittent power is expected. A proper SLEEP mode with close-to-zero power consumption during Processor SLEEP is also achievable with nvDFF-based NVPs. In this regard, the nvDFF is a suitable solution for IoT devices that expect to experience power

disruptions or devices that have long SLEEP times with intermittent operation. The nvDFF shows an average of 1.1x low-to-high switching delay and an average of 1.02x high-to-low switching delay compared to the vDFF while the Clk-to-Q delay between both shows an average difference of 1.01x. The nvDFF electrical performance matches well with the SRAM-based DFF and can be used as a substitute where NV is essential.

The nvDRAM design demonstrates NV and resilience to bit-flipping in simulation tests. Although the 2T1R has higher single WRITE delays, it has an average  $1.19 \times 10^3$  nJ lower WRITE energy consumption than the compared DRAM cells. The proposed cell's NV eliminates the need for refresh operations unlike the referenced designs. Its low energy WRITE coupled with the removal of refresh offers great promise as a next-generation DRAM cell candidate. The 2T1R's READ performance metrics are also lower than the reference DRAM cells by an average of 1181.8 nJ. The nvDRAM provides a solution for the DRAM scaling problem as well as eliminating the need for periodic refresh cycles that is the hallmark of conventional DRAM technology.

## 8. Chapter Eight

# **The Non-Volatile FPGA Architecture**

---

This chapter presents the holistic nvFPGA, composed entirely of NV components presented in the previous chapters. NV is introduced to the FPGA throughout its basic components; the logic elements and the routing interconnects. The nvFPGA is then analyzed through three performance metrics which are:

1. The FPGA device area
2. The critical path delay
3. The average EDP

These are then compared with traditional SRAM-based FPGA metrics.

### **Related Publications**

---

1. H. L. Chee, T. N. Kumar, and H. A. F. Almurib, "A ReRAM-based Nonvolatile FPGA," *Proceedings in 20th IEEE Student Conference on Research and Development (SCoReD 2022)*, 2022. (Malaysia)

## 8.1 The nvFPGA

As discussed in Chapter 2, the CLB forms the basic logic components of the FPGA architecture. Fig. 8-1 is an example of a generic basic logic element (BLE) [161]. It is comprised of a  $K$ -input LUT, a DFF register, and two multiplexers responsible for selecting the combinatorial LUT output or the registered output that is fed to the output of the BLE. Fig. 8-2 shows the CLB or logic block (LB) which is composed of  $N$  number of BLEs.

The aim of this research is to replace the LUTs in the BLE with the MB-nvLUT presented in Chapter 6 and the register DFF with the nvDFF from Chapter 7. The block diagram for the nvCLB proposed in this thesis is presented in Fig. 8-3. This nvCLB is composed of eight six-input MB-nvLUTs which is the conventional number of LUT inputs [162] together with dedicated nvDFFs for each MB-nvLUT. Each LUT has six independent inputs (MB-nvLUT1 inputs – A1 to A6, MB-nvLUT2 inputs – B1 to B6, MB-nvLUT3 inputs – C1 to C6, MB-nvLUT4 inputs – D1 to D6, MB-nvLUT5 inputs – E1 to E6, MB-nvLUT6 inputs – F1 to F6, MB-nvLUT7 inputs – G1 to G6, MB-nvLUT8 inputs – H1 to H6) and two independent outputs for each MB-nvLUT (Output 1 is the LUT output and Output 2 is the registered output). This way, each LUT can implement any arbitrary six-input Boolean function if the functions share the same inputs.

As for the FPGA's SwBs, the configurable interconnects are intersections of the wire routing that connects the input/output (I/O) pins of the logic elements (CLBs) to the I/O pins of the FPGA and the other components that exist in the FPGA. A traditional interconnect layout is given in Fig. 8-4 [163]. Each line has a pass transistor (1-6) that can be switched on or off to establish or sever a connection between the lines. For example, pass transistor (1) connects line 90-3 to line 90-4 when turned on.

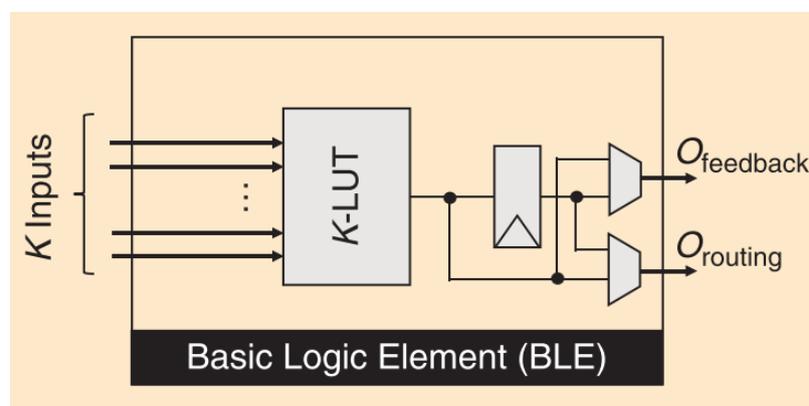


Fig. 8-1. Schematic of a BLE [161].

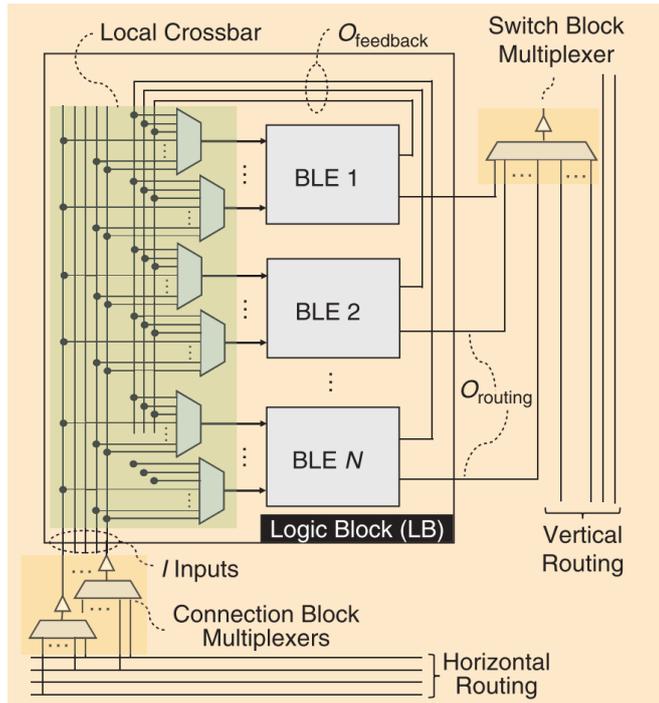


Fig. 8-2. Schematic of a CLB comprised of  $N$  number of BLEs [161].

This interconnect is present on every wire cross-section in the FPGA as shown in Fig. 8-5.

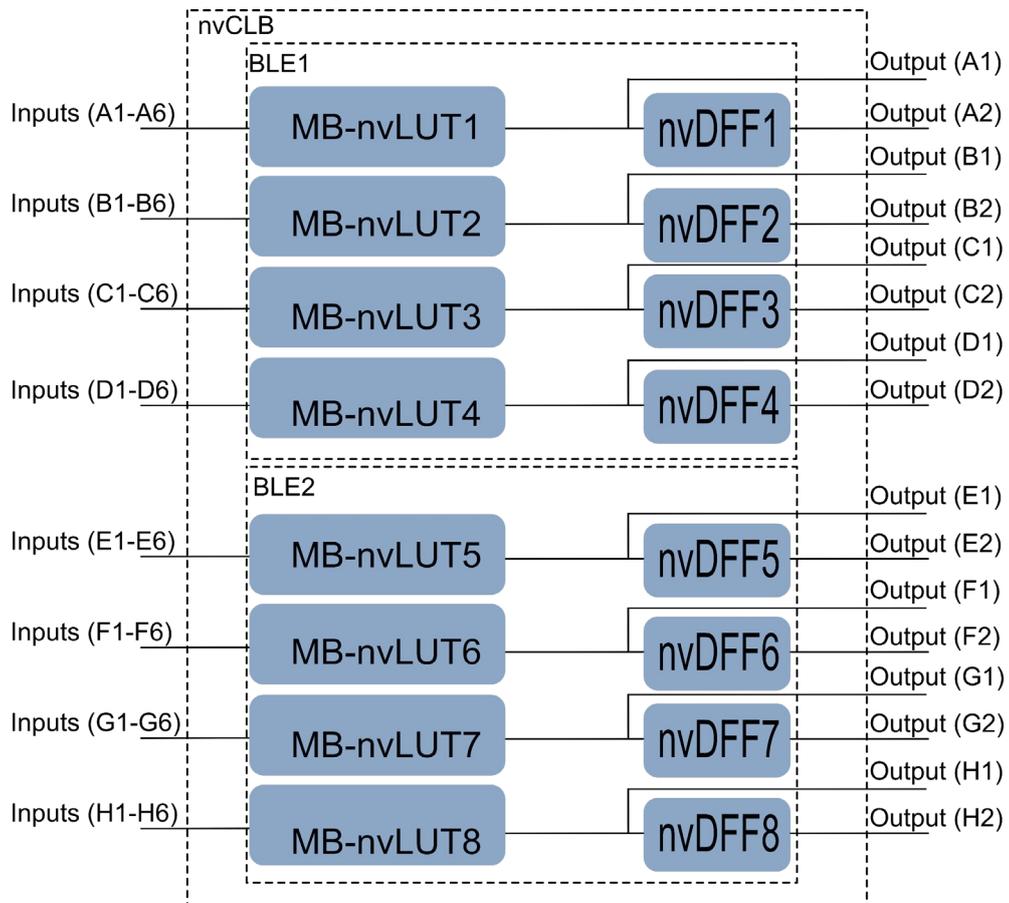


Fig. 8-3. Block diagram of nvCLB.

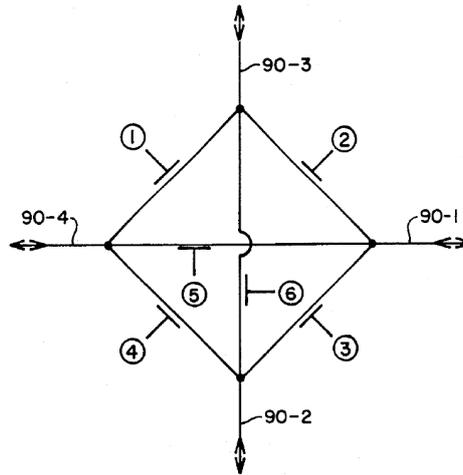


Fig. 8-4. Traditional interconnect layout [163].

The configuration of every pass transistor in the interconnects are stored in a SwB which consists of storage components, typically SRAMs that retain the configuration bits for a design.

The SRAMs which store the FPGA's routing configuration bits are replaced with the nvDRAMs from Chapter 7 to implement NV for the SwB architecture. The block diagram of the nvCLB and the nvSwB is shown in Fig. 8-6 where the routing configurations for the inputs and outputs of the nvCLB are controlled by the nvSwB.

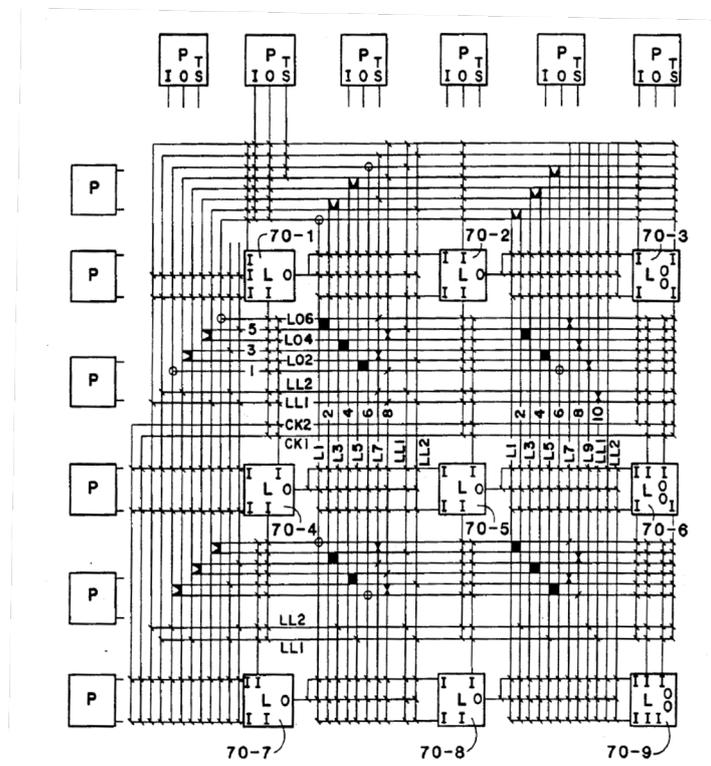


Fig. 8-5. FPGA architecture showing CLBs and the interconnects [163].

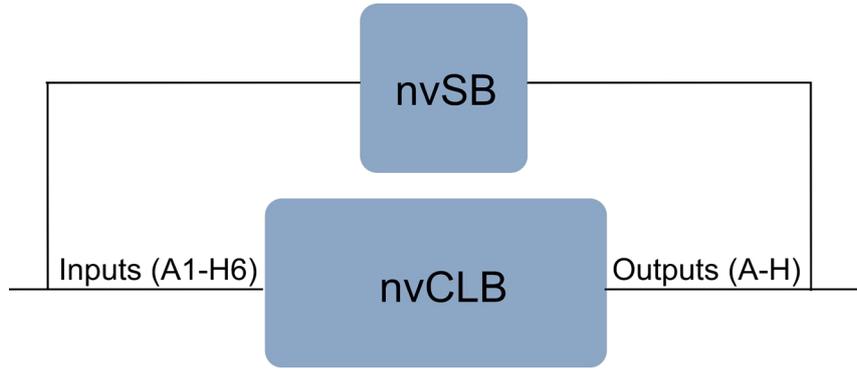


Fig. 8-6. Block diagram of nvSB combined with nvCLB.

## 8.2 Analysis of the nvCLB

To analyse the performance metrics, the resource usage is first obtained through synthesizing test designs in Xilinx ISE 14.1. The volatile CMOS-based CLBs (written as vCLBs in text from hereon) are then replaced with the nvCLBs to evaluate the performance metrics.

### 1. Device Area

The MB-nvLUT replaces the CMOS transistors in the LUT-array with ReRAMs. As ReRAMs are fabricated in the via between metal layers, their device area can be neglected compared to the transistor gate length. Due to the multibit storage per memory cell, the MB-nvLUT also halves the number of memory cells in the LUT array, thereby reducing the LUT array size. There are a total of sixty-four SRAM cells in the SRAM LUT array while the MB-nvLUT array only has thirty-two NMOS pass transistors in the array bitlines. In terms of array size where  $f$  is the half-pitched length of the transistors, the SRAM LUT array is  $64 f^2$  while the MB-nvLUT is  $32 f^2$ .

The controller for the MB-nvLUT however requires seven AND gates, five inverters, and one T-gate per LUT which contributes to the nvCLB cell area and increases the array size to  $518 f^2$ . The SRAM LUT has an additional thirty-one NMOS and thirty-one PMOS transistors that bring up the array size to  $605 f^2$ . For a transistor technology node of 45 nm ( $L=45$  nm,  $NMOS_{width,gate}=132.5$  nm,  $PMOS_{width,gate}=256$  nm), the total MB-nvLUT cell area becomes  $3.02291 \mu m^2$  compared to the SRAM cell area of  $3.59476 \mu m^2$ .

The nvDFF presented in Section 7.2 has the same cell area as the CMOS-based DFF which is  $24 f^2$ . The resistor, R1 is fabricated directly into the metal wires and do not affect the area of the cell. Thus, the size of one DFF/nvDFF for 45nm transistor node is  $0.1399 \mu m^2$ .

A single CLB which comprises of eight 6-input LUTs and eight DFFs therefore has a cell area of 25.3026  $\mu\text{m}^2$  for the nvCLB and 29.8769  $\mu\text{m}^2$  for the vCLB.

## 2. Path Delay

The path delay comparisons between the MB-nvLUT and the SRAM-LUT are listed in TABLE 6-13, TABLE 6-14, and TABLE 6-15. The MB-nvLUT has an average 73.463% higher WRITE delay compared to the SRAM-LUT because of the switching time of the ReRAMs. This is a drawback for the MB-nvLUT but FPGAs are generally expected to have much lesser WRITE operations compared to READ operations [29].

In the case of the READ operation, the MB-nvLUT has a 97.295% lower delay over the SRAM-LUT as the ReRAM READ operation is much faster than the SRAM READ [164]. This is desirable for FPGA applications that expect to see much more READ cycles than WRITE operations throughout its lifetime.

As for the path delays for the DFFs, the nvDFF has an average higher delay of 5.089% compared to the vDFF (TABLE 7-4). This a very low difference as the timings were measured in picoseconds and their performances are therefore similar.

The DFF segment in the CLB is however only used for the CLB output during READ and not used during WRITE operations which leads to the nvCLB having a 73.463% higher path delay during WRITE and a 92.206% lower path delay during READ over the vCLB.

## 3. Power Dissipation

The power dissipation of the MB-nvLUT for FPGA benchmarks have been given in TABLE 6-13, TABLE 6-14, and TABLE 6-15. The MB-nvLUT has an average 99.79% higher WRITE EDP compared to the SRAM-LUT because of the increase in WRITE delay as well as the increase in ReRAM WRITE energy.

The same observation for the path delay occurs for the EDP where the MB-nvLUT has a 91.184% lower EDP compared to the SRAM-LUT. Once again, this is also attributed to the faster speeds for ReRAM READ operations as well as lower ReRAM READ energy dissipation.

The nvDFF and vDFF values are given in TABLE 7-6. Due to the WRITE operations on the ReRAMs, the nvDFF has an average 61.1045% EDP compared to the vDFF.

As the DFF does not contribute during the WRITE phase, the nvCLB has a 99.79% higher WRITE EDP and a 30.0795% lower READ EDP over the vCLB.

### **8.3 Analysis of the nvSwB**

The design of the nvDRAM uses only two NMOS transistors and one ReRAM compared to the six/eight NMOS transistors of the SRAM cell. This leads to an nvDRAM cell area of  $2 f^2$  compared to the  $6 f^2$  of the SRAM cell. Using 45nm transistor nodes, the area of the nvDRAM cell is  $0.0119 \mu\text{m}^2$  while the area of the SRAM cell is  $0.03578 \mu\text{m}^2$  leading to a reduction in cell area of 67%.

The average delay and EDP values of the nvDRAM are listed in TABLE 7-4 and TABLE 7-6. The average WRITE delay of the nvDRAM cell is 12253.17 ps; 98% higher than the 213.0103 ps of the SRAM cell while the average READ delay of the nvDRAM cell is 200 ps; 99.5% higher than the SRAM cell's READ delay. Additionally, the average EDP values of the nvDRAM for WRITE and READ operations are 7.588 ps nJ and  $3.814 \times 10^{-2}$  ps nJ respectively. These are higher than the average SRAM WRITE and READ EDPs of  $1.658 \times 10^{-9}$  ps nJ and  $6.001 \times 10^{-13}$  ps nJ respectively.

Although the delay and EDP of the nvDRAM cell are higher than the SRAM cell, the SwB's function is to only store the configuration bits data of the design and the typical usage of an FPGA does not require high WRITE operations to the SwB once a design has been programmed. The nvSwB however manages to achieve NV therefore allowing the SB to be fully shut down in during deep SLEEP mode while also making the FPGA resilient to power interruptions.

Moreover, it should be noted that the delay and EDP values that are compared are in the ranges of picoseconds and nanojoules, well within the required timings of conventional electronic circuits.

### **8.4 Analysis of the nvFPGA**

The comparisons above are compiled in this section to provide a summation of the comparison between the nvFPGA's and vFPGA's performances and are listed in TABLE 8-1.

TABLE 8-1. nvFPGA vs vFPGA performance comparison (A number higher than 1 is better, lower than 1 is worst)

	LUT	DFF	SB
WRITE delay	0.260548858	0.947509	1.738E-02
WRITE EDP	0.006884248	0.386867	2.186E-10
READ delay	22431.00959	-	4.875E-03
READ EDP	632.3883984	-	1.573E-11

Next, the NV CLB's passive storage ability is tested out with an application with the following Boolean expression:

$$\text{Output} = A[\overline{BC}[\overline{D}(\overline{EF} + E\overline{F}) + D\overline{EF}] + \overline{DEF}(\overline{BC} + B\overline{C})] \quad (8.1)$$

In this expression, the A input is checked first where the LUT output will be '0' if A is '0' and the remaining five bits determine the LUT output if A is '1'. If the sum of the remaining five bits is '1' and only '1' then the output will be '1' otherwise the LUT's output is '0'. The A1 input comes from an external signal for the first LUT in the nvCLB (MB-nvLUT1 in Fig. 8-3) and the output of LUT1 will be fed into the first nvDFF (nvDFF1 in Fig. 8-3). The output of LUT1\_NV DFF is fed into the second LUT (MB-nvLUT2 in Fig. 8-3) as the A2 input. Expression (8.1) is stored into both LUTs.

In this simulation, the input '100010' is fed into MB-nvLUT1 which gives an output of '1' which is then fed into nvDFF1 and into the A2 input of MB-nvLUT2. The selected address for MB-nvLUT2 is '111111' which gives the output, '0' which is fed into nvDFF2.

Fig. 8-7 shows the written states for three cells, LUT1\_M17, LUT1\_M18, and LUT2\_M32; LUT1\_M17 and LUT2\_M32 are the locations of the selected addresses and LUT1\_M18 is plotted to show the state condition of an unselected cell in the array. As a 6-input LUT, the MB NV LUT has 32 cells in the array, half the number of array cells in a SB LUT array.

Cell LUT1\_M17 stores the data bits for addresses '100001' and '100010' which are '0' and '1' respectively based on (1) while cell LUT1\_M18 stores '1' and '0' for addresses '100010' and '100011', and cell LUT2\_M32 stores '0' and '0' for addresses '111110' and '111111'.

Therefore, cells LUT1\_M17, LUT1\_M18, and LUT2\_M32 are written into IRS\_2, IRS\_1, and HRS respectively following the WRITE logic in TABLE 2.

The state changes are plotted in Fig. 8-7 where LUT1\_M17 and LUT1\_M18 are first written into LRS and then subsequently written into the desired states. No WRITE operations were performed on LUT2\_M32 as it is already in HRS.

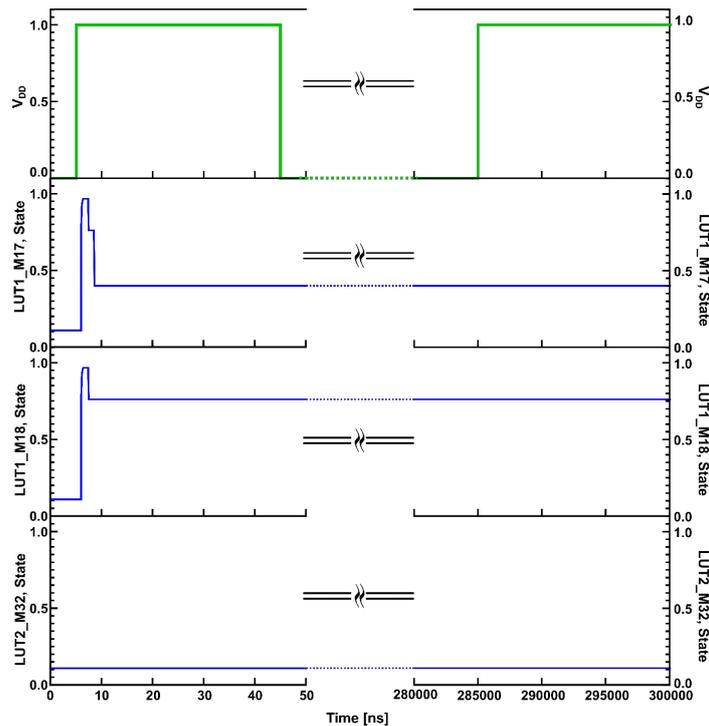


Fig. 8-7. ReRAM states in the MB NV LUT during normal operation (left) with the insertion of data ‘100111’ and after power restoration (right).

The supply voltage,  $V_{DD}$  is then cut at 45ns to simulate a power disruption scenario. Restoration of power supply occurs at 285µs and it can be seen that the ReRAMs retain their states during power cut-off and restoration.

The simulation was similarly carried out for the NV DFFs in the NV CLB, LUT1\_NVDFF and LUT2\_NVDFF. In this case, the data outputs from the LUT1 and LUT2 are fed into the LUT1\_NVDFF and LUT2\_NVDFF respectively. Similar  $V_{DD}$  conditions to Fig. 8-7 are used and the results are plotted in Fig. 8-8.

Data “1” is fed to NV DFF1 from the MB NV LUT and  $Q/Q_{\bar{b}}$  rises/falls while data “0” is fed to NV DFF2 and  $Q/Q_{\bar{b}}$  falls/rises. The ReRAMs also retain their respective states during power cut-off and restoration and successfully restore  $Q/Q_{\bar{b}}$  values of their respective NV DFFs.

An interesting outcome from NV implementation in the BLE is the opportunity for the reduction of multiplexers; the multiplexer in Fig. 8-1 can provide a direct output from the LUT or feed the LUT output into the DFF for sequential operations. This multiplexer is removed in the NV CLB so that the NV LUT output is split into the direct output and the input into the NV DFF. If the NV DFF is not in use, power supply to the NV DFF can be halted and the NV DFF retains the previous  $Q/Q_{\bar{b}}$  values.

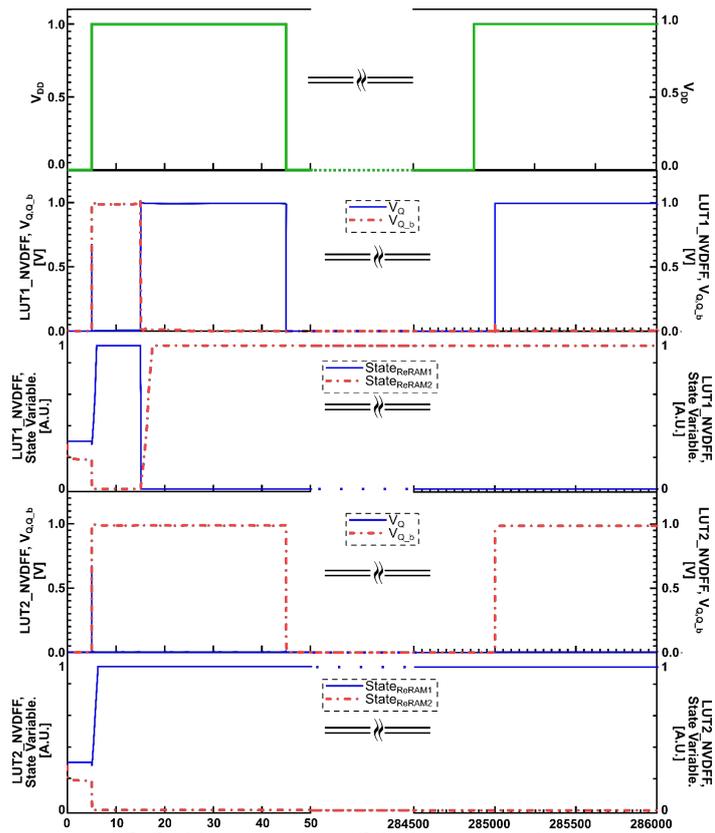


Fig. 8-8. The Q/Q<sub>b</sub> outputs and ReRAM states in the nvDFF during normal operation (left) with the insertion of data from the MB-nvLUT and after power restoration (right).

## 8.5 Summary

An nvFPGA consisting of an nvCLB made up of MB-nvLUTs and nvDFFs and an nvSB made up of nvDRAMs was analyzed in this work. Through usage of these NV memory components, the nvFPGA successfully retains data during power disruption events. All NV memory components incorporate ReRAM devices for storage and do not require constant power supply during operation, allowing novel SLEEP modes to reduce power consumption. This will lead to significant energy savings over the nvFPGA usage's lifetime and will surpass the energy cost incurred by the nvFPGA during WRITE and READ operations.

# Conclusion and Future Works

---

This PhD research has produced a nvFPGA architecture based on ReRAMs, an emerging nonvolatile memory device. The research produced NV versions of components that form the backbone of the FPGA structure namely, the LUTs and the DFFs that make up the CLBs, and the SwB which are responsible for routing the FPGA interconnects.

This work presents a nvLUT that is made up of MB cells in the LUT array. This was achieved due to the analog switching behaviour of ReRAMs where the ReRAM resistances undergo a gradual switching between resistance states. Removal of voltage supply before the ReRAM reaches its LRS or HRS halts the RS process and leaves the ReRAM in an IRS.

A study was first conducted on the physical phenomenon of ReRAMs and the multiple methods of modelling the RS mechanism in the ReRAM. A model of a MB-ReRAM was then developed based on the multi-filamentary phenomena. It was found in the literature that the CF that form in the metal-oxide layer of the ReRAM are preferentially located at GB in the metal-oxide polycrystalline structure. As the existence of multiple CFs have been noted in experiments, this research undertook the task of developing a MB-ReRAM model that simulates the effect of the creation and destruction of multi-filaments in the metal-oxide to achieve RS.

The created model was based on modelling the migration of  $V_{OS}$  in the metal-oxide layer under an electric field through a supplied voltage. The current through the ReRAM, which indicates its resistance is modelled through TAT equations. TAT describes the modulation of the interfacial barrier between the CFs and the top electrode of the ReRAM and includes the probability factor for electron tunnelling. Of interest to note is that the developed MB-ReRAM model demonstrates that multi-filamentary conduction is strongly controlled by the barrier height activation energy parameter. This indicates that the formation/rupture processes of

individual filaments in a multi-filamentary ReRAMs are controlled by the in-situ barrier height strength located where the CF meets the top electrode.

The research is then focused on developing a MB-ReRAM-based nvLUT. An SBn-vLUT is first analysed and a sense-amplifier is developed for an SB-ReRAM crossbar. The SB-nvLUT design from literature provides a controller design specially made for ReRAM-based LUT arrays. Of particular importance is the ability of the ReRAM array to shield unselected ReRAM cells from sneak path current during operation. The controller was shown to function well in this regard and the concept was adopted in the design for a MB-nvLUT.

On the other hand, it was noted that the ReRAM cells in the array would require specially designed sense-amplifiers to detect sub-threshold voltages at the array output. A sense amplifier design was produced that utilizes the concept of inverting buffers to first raise the sub-threshold voltages from the ReRAM cell's output before feeding the voltage to a differential comparator. The inverting-buffer circuit successfully raised sub-threshold voltages of 200mV to 1.1V in nanosecond ranges.

The MB-nvLUT is then developed using MB-ReRAMs that are capable of holding 2-bits-per-cell, giving 4 different resistance levels. As a result of using MB-ReRAMs, the number of cells in the array is halved compared to an SB-nvLUT array or an SRAM-based LUT array. A unique WRITE scheme was required to store 2-bits-per-cell and a controller was thus designed. Compared to the SB-nvLUT, the MB-nvLUT manages to reduce the cell count in the LUT array by 0.5x and reduces the gates in the controller by 0.25x. The MB-nvLUT also has an average of 2x lower delay, 1.22x lower energy consumption, and 2.46x lower EDP for WRITE 0; 2x lower delay, 2x lower energy consumption, and 4.6x lower EDP for WRITE 1; 2x lower delay, 1x lower energy consumption, and 2x lower EDP for WRITE 01 → 10; 9.2x lower delay, 128x lower energy consumption, and 153x lower EDP over the SB-nvLUT. Benchmark tests also demonstrated that the MB-nvLUT performance catches up to SRAM-based LUTs in arrays of larger sizes, which indicates that the MB-nvLUT is a suitable candidate for future FPGAs as the demand for larger arrays continue.

Three NV electronic storage elements were developed in the course of this research namely, the nvD Latch, the nvDFF, and the nvDRAM. The implementation of ReRAMs in the designs result in an increase of performance metrics like WRITE delay and switching times because of the ReRAM switching times. However, it should be noted that the performances of the ReRAM-based storage circuits are within nanosecond-picosecond ranges which are faster than

conventional circuits that exist today. The introduction of NV into these designs also convey an advantage over the volatile complementary-metal-oxide-semiconductor (CMOS)-based designs. In the case of the nvDRAM, the CMOS-based DRAM requires an energy consuming refresh operation every few hundred milliseconds whereas the nvDRAM eliminates this requirement. The nvD Latch and nvDFF are storage elements that are commonly found in registers located in processors. The CMOS-based designs are volatile and require constant power supply even in situations where the latches and FFs aren't in use ie. during SLEEP or intermittent usage. This is circumvented by the nvD Latch and nvDFF and presents a solution for energy-saving and energy-efficient electronics that can contribute to sustainable production and consumption.

The nvFPGA is then completed in this work by the design of nvSwBs in the FPGA. The nvSBs hold the configuration bits for a programmed design which are then used to control the routing architecture of the connecting wires in the FPGA. Conventional FPGAs use SRAMs for the storage elements which are NV. An nvSwB design is presented in this research utilizing the previously design nvDRAM as storage elements. The nvSBs eliminate the requirement for the FPGA to be reprogrammed after every start up.

The future directions of the work done in this PhD research should be the physical fabrications of the presented designs which would produce a viable nvFPGA for applications that require NV or low energy consumption. Interesting developments in CMOS technology such as negative-capacitance field-effect transistors would further reduce the energy consumption of the NV designs in this work. In addition to that, ReRAM development is still ongoing in the field with layer engineering producing 1S1R or 1G1R configurations that are able to lower the switching energy consumption of the ReRAM cell.

## References

---

- [1] R. Tessier and W. Burleson, “Reconfigurable computing for digital signal processing: A survey,” *J. VLSI Signal Process. Syst. Signal Image. Video Technol.*, vol. 28, no. 1–2, pp. 7–27, 2001, doi: 10.1023/A:1008155020711.
- [2] R. Woods, J. McAllister, G. Lightbody, and Y. Yi, *FPGA-based Implementation of Signal Processing Systems*. Hoboken, NJ, USA: Wiley, 2008.
- [3] M. Vestias and H. Neto, “Trends of CPU, GPU and FPGA for high-performance computing,” *Conf. Dig. - 24th Int. Conf. F. Program. Log. Appl. FPL 2014*, pp. 11–16, 2014, doi: 10.1109/FPL.2014.6927483.
- [4] E. Bank-Tavakoli, S. A. Ghasemzadeh, M. Kamal, A. Afzali-Kusha, and M. Pedram, “POLAR: A Pipelined/Overlapped FPGA-Based LSTM Accelerator,” *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 28, no. 3, pp. 838–842, 2020, doi: 10.1109/TVLSI.2019.2947639.
- [5] R. Dorrance, F. Ren, and D. Marković, “A scalable sparse matrix-vector multiplication kernel for energy-efficient sparse-blas on FPGAs,” *ACM/SIGDA Int. Symp. F. Program. Gate Arrays - FPGA*, pp. 161–169, 2014, doi: 10.1145/2554688.2554785.
- [6] D. Koch, N. Dao, B. Healy, J. Yu, and A. Attwood, “FABulous: An embedded fpga framework,” *FPGA 2021 - 2021 ACM/SIGDA Int. Symp. Field-Programmable Gate Arrays*, pp. 45–56, 2021, doi: 10.1145/3431920.3439302.
- [7] J. Qiu, J. Wang, S. Yao, K. Guo, and B. Li, “Going Deeper with Embedded FPGA Platform for Convolutional Neural Network • Deep Learning and Convolutional Neural Network – V2 : Brief introduction,” *F. Program. Gate Arrays*, pp. 26–35, 2016.
- [8] P. Babu and E. Parthasarathy, “Reconfigurable FPGA Architectures: A Survey and Applications,” *J. Inst. Eng. Ser. B*, vol. 102, no. 1, pp. 143–156, 2021, doi: 10.1007/s40031-020-00508-y.

- [9] J. Shalf, "The future of computing beyond Moore's Law," *Philos. Trans. R. Soc.*, vol. 378, no. 20190061, pp. 1–14, 2020.
- [10] C. Fox, "Intel's next-generation 7nm chips delayed until 2022," *BBC News*, 2020. <https://www.bbc.com/news/technology-53525710>.
- [11] J. Hruska, "Intel's Rocket Lake Roars to Life," *Extremetech*, 2021. <https://www.extremetech.com/computing/321350-intels-rocket-lake-roars-to-life>.
- [12] K. Morris, "No More Nanometers," *EEJournal*, 2020. <https://www.eejournal.com/article/no-more-nanometers/>.
- [13] H. S. P. Wong *et al.*, "A Density Metric for Semiconductor Technology [Point of View]," *Proc. IEEE*, vol. 108, no. 4, pp. 478–482, 2020, doi: 10.1109/JPROC.2020.2981715.
- [14] F. Pan, S. Gao, C. Chen, C. Song, and F. Zeng, "Recent progress in resistive random access memories: Materials, switching mechanisms, and performance," *Mater. Sci. Eng. R Reports*, vol. 83, no. 1, pp. 1–59, 2014, doi: 10.1016/j.mser.2014.06.002.
- [15] K. Ishimaru, "Challenges of Flash Memory for Next Decade," *IEEE Int. Reliab. Phys. Symp. Proc.*, vol. 2021-March, pp. 1–5, 2021, doi: 10.1109/IRPS46558.2021.9405182.
- [16] K. Huang, Y. Ha, R. Zhao, A. Kumar, and Y. Lian, "A Low Active Leakage and High Reliability Phase Change Memory (PCM) Based Non-Volatile FPGA Storage Element," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 61, no. 9, pp. 2605–2613, 2014, doi: 10.1109/TCSI.2014.2312499.
- [17] L. Gezi, C. Xiaogang, L. Shunfen, M. Bin, and S. Zhitang, "FPGA-enhanced data processing system using PCM technology," *Chinese J. Electron.*, vol. 29, no. 4, pp. 766–771, 2020, doi: 10.1049/cje.2020.06.004.
- [18] R. Rajaei, "Radiation-Hardened Design of Nonvolatile MRAM-Based FPGA," *IEEE Trans. Magn.*, vol. 52, no. 10, pp. 1–10, 2016, doi: 10.1109/TMAG.2016.2578278.
- [19] M. Natsui *et al.*, "A 47.14- $\mu$ W 200-MHz MOS/MTJ-Hybrid Nonvolatile Microcontroller Unit Embedding STT-MRAM and FPGA for IoT Applications," *IEEE J. Solid-State Circuits*, vol. 54, no. 11, pp. 2991–3004, 2019, doi: 10.1109/JSSC.2019.2930910.
- [20] X. Chen, M. Niemier, and X. S. Hu, "Nonvolatile Lookup Table Design Based on

- Ferroelectric Field-Effect Transistors,” *Proc. - IEEE Int. Symp. Circuits Syst.*, vol. 2018-May, pp. 1–5, 2018, doi: 10.1109/ISCAS.2018.8351375.
- [21] H. L. Chee, T. N. Kumar, and H. A. F. Almurib, “Low energy non-volatile look-up table using 2 bit ReRAM for field programmable gate array,” *Semicond. Sci. Technol.*, vol. 37, no. 6, 2022, doi: 10.1088/1361-6641/ac6903.
- [22] J. Cong and B. Xiao, “FPGA-RPI: A novel fpga architecture with rram-based programmable interconnects,” *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 22, no. 4, pp. 864–877, 2014, doi: 10.1109/TVLSI.2013.2259512.
- [23] Y. Y. Liauw, Z. Zhang, W. Kim, A. El Gamal, and S. S. Wong, “Nonvolatile 3D-FPGA with monolithically stacked RRAM-based configuration memory,” *Dig. Tech. Pap. - IEEE Int. Solid-State Circuits Conf.*, vol. 55, pp. 406–407, 2012, doi: 10.1109/ISSCC.2012.6177067.
- [24] L. O. Chua, “Memristor—The Missing Circuit Element,” *IEEE Trans. Circuit Theory*, vol. 18, no. 5, pp. 507–519, 1971, doi: 10.1109/TCT.1971.1083337.
- [25] 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, doi: 10.1038/nature06932.
- [26] S. Claramunt, A. Ruiz, Q. Wu, M. Porti, M. Nafria, and X. Aymerich, “MIS structures with interfacial graphene for ReRAM applications: a nanoscale and device level characterization,” in *2020 Joint International EUROSOI Workshop and International Conference on Ultimate Integration on Silicon (EUROSOI-ULIS)*, Sep. 2020, pp. 1–4, doi: 10.1109/EUROSOI-ULIS49407.2020.9365299.
- [27] F. O. Hatem, T. N. Kumar, and H. A. F. Almurib, “A SPICE Model of the Ta<sub>2</sub>O<sub>5</sub>/TaO<sub>x</sub> Bi-Layered RRAM,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 63, no. 9, pp. 1487–1498, 2016, doi: 10.1109/TCSI.2016.2579503.
- [28] S. Asapu and T. Maiti, “Multifilamentary conduction modeling in transition metal oxide-based rram,” *IEEE Trans. Electron Devices*, vol. 64, no. 8, pp. 3145–3150, 2017, doi: 10.1109/TED.2017.2709249.
- [29] T. N. Kumar, H. A. F. Almurib, and F. Lombardi, “Design of a memristor-based look-up table (LUT) for low-energy operation of FPGAs,” *Integr. VLSI J.*, vol. 55, pp. 1–11, 2016, doi: 10.1016/j.vlsi.2016.02.005.

- [30] W. P. Lin *et al.*, “A nonvolatile look-up table using ReRAM for reconfigurable logic,” *2014 IEEE Asian Solid-State Circuits Conf. A-SSCC - Proc. Tech. Pap.*, no. c, pp. 133–136, 2015, doi: 10.1109/ASSCC.2014.7008878.
- [31] Y. Guo, X. Wang, and Z. Zeng, “A Compact Memristor-CMOS Hybrid Look-Up-Table Design and Potential Application in FPGA,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 36, no. 12, pp. 2144–2148, 2017, doi: 10.1109/TCAD.2017.2681079.
- [32] M. Pešić *et al.*, “Conduction barrier offset engineering for DRAM capacitor scaling,” *Solid. State. Electron.*, vol. 115, pp. 133–139, 2016, doi: 10.1016/j.sse.2015.08.012.
- [33] E. Monmasson, L. Idkhajine, M. N. Cirstea, I. Bahri, A. Tisan, and M. W. Naouar, “FPGAs in industrial control applications,” *IEEE Trans. Ind. Informatics*, vol. 7, no. 2, pp. 224–243, 2011, doi: 10.1109/TII.2011.2123908.
- [34] V. Gupta and M. Anis, “Statistical Design of the 6T SRAM Bit Cell,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 57, no. 1, pp. 93–104, 2009, doi: 10.1109/tcsi.2009.2016633.
- [35] C. Maxfield, *FPGAs: Instant Access*. 2008.
- [36] I. Xilinx, “Xilinx UG190 Virtex-5 FPGA User Guide,” vol. 190. pp. 1–385, 2012, [Online]. Available: [http://www.xilinx.com/support/documentation/user\\_guides/ug190.pdf](http://www.xilinx.com/support/documentation/user_guides/ug190.pdf).
- [37] A. Corporation, “Introduction, Stratix II Device Family Data Sheet,” no. May. pp. 1–6, 2007.
- [38] G. Varghese, Z. Hui, and R. Jan, “The design of a low energy FPGA,” in *Proceedings of the 1999 International Symposium on Low Power Electronics and Design*, 1999, pp. 188–193, doi: 10.1145/313817.313920.
- [39] C. Y. Wen *et al.*, “A non-volatile look-up table design using PCM (phase-change memory) cells,” *IEEE Symp. VLSI Circuits, Dig. Tech. Pap.*, pp. 302–303, 2011.
- [40] P. E. Gaillardon *et al.*, “Design and architectural assessment of 3-D resistive memory technologies in FPGAs,” *IEEE Trans. Nanotechnol.*, vol. 12, no. 1, pp. 40–50, 2013, doi: 10.1109/TNANO.2012.2226747.
- [41] P. E. Gaillardon, D. Sacchetto, S. Bobba, Y. Leblebici, and G. De Micheli, “GMS:

- Generic memristive structure for non-volatile FPGAs,” *IEEE/IFIP Int. Conf. VLSI Syst. VLSI-SoC*, vol. 07-10-Octo, pp. 94–98, 2015, doi: 10.1109/VLSI-SoC.2012.7332083.
- [42] X. Tang, G. Kim, P.-E. Gaillardon, and G. De Micheli, “A Study on the Programming Structures for RRAM-Based FPGA Architectures,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, pp. 1–14, 2016, doi: 10.1109/TCSI.2016.2528079.
- [43] B. Govoreanu *et al.*, “10×10nm<sup>2</sup> Hf/HfO<sub>x</sub> crossbar resistive RAM with excellent performance, reliability and low-energy operation,” *Tech. Dig. - Int. Electron Devices Meet. IEDM*, pp. 729–732, 2011, doi: 10.1109/IEDM.2011.6131652.
- [44] D. Suzuki, M. Natsui, S. Ikeda, T. Endoh, H. Ohno, and T. Hanyu, “Design of a variation-resilient single-ended non-volatile six-input lookup table circuit with a redundant-magnetic tunnel junction-based active load for smart Internet-of-things applications,” *Electron. Lett.*, vol. 53, no. 7, pp. 456–458, 2017, doi: 10.1049/el.2016.4233.
- [45] S. D. Kumar and H. Thapliyal, “Exploration of Non-Volatile MTJ/CMOS Circuits for DPA-Resistant Embedded Hardware,” *IEEE Trans. Magn.*, vol. 55, no. 12, 2019, doi: 10.1109/TMAG.2019.2943053.
- [46] Y.-C. Chen, W. Wang, H. (Helen) Li, and W. Zhang, “Non-volatile 3D stacking RRAM-based FPGA,” in *22nd International Conference on Field Programmable Logic and Applications (FPL)*, 2012, pp. 367–372, doi: 10.1109/FPL.2012.6339206.
- [47] F. Zhang *et al.*, “The application of non-volatile look-up-table operations based on multilevel-cell of resistance switching random access memory,” *2018 Int. Symp. VLSI Des. Autom. Test, VLSI-DAT 2018*, pp. 1–4, 2018, doi: 10.1109/VLSI-DAT.2018.8373268.
- [48] C. E. Merkel, N. Nagpal, S. Mandalapu, and D. Kudithipudi, “Reconfigurable N-level memristor memory design,” *Proc. Int. Jt. Conf. Neural Networks*, pp. 3042–3048, 2011, doi: 10.1109/IJCNN.2011.6033622.
- [49] C. W. Ian Wong and P. W. C. Ho, “Multilevel memristive non-volatile look-up table using two transmission gates one memristor memory cells,” *Semicond. Sci. Technol.*, vol. 35, no. 10, 2020, doi: 10.1088/1361-6641/abaa59.
- [50] S. Kannan, J. Rajendran, R. Karri, and O. Sinanoglu, “Sneak-path testing of memristor-

- based memories,” *Proc. IEEE Int. Conf. VLSI Des.*, pp. 386–391, 2013, doi: 10.1109/VLSID.2013.219.
- [51] S. Borkar, “Design challenges of technology scaling,” *IEEE Micro*, vol. 19, no. 4, pp. 23–29, 1999, doi: 10.1109/40.782564.
- [52] H. Sun, X. Xu, and H. Shao, “Memristor-based non-volatile D trigger,” 2012.
- [53] J. Zheng, Z. Zeng, and Y. Zhu, “Memristor-based nonvolatile synchronous flip-flop circuits,” *7th Int. Conf. Inf. Sci. Technol. ICIST 2017 - Proc.*, pp. 504–508, 2017, doi: 10.1109/ICIST.2017.7926812.
- [54] Z. Chang, A. Cui, Z. Wang, G. Qu, and C. Park, “Novel Memristor-based Nonvolatile D Latch and Flip-flop Designs,” in *2021 22nd International Symposium on Quality Electronic Design (ISQED)*, 2021, pp. 244–250, doi: 10.1109/ISQED51717.2021.9424269.
- [55] A. Amirany, F. Marvi, K. Jafari, and R. Rajaei, “Nonvolatile Spin-Based Radiation Hardened Retention Latch and Flip-Flop,” *IEEE Trans. Nanotechnol.*, vol. 18, pp. 1089–1096, 2019, doi: 10.1109/TNANO.2019.2946108.
- [56] P. W. C. Ho, H. A. F. Almurib, and T. N. Kumar, “Configurable memristive logic block for memristive-based FPGA architectures,” *Integr. VLSI J.*, vol. 56, no. September 2016, pp. 61–69, 2017, doi: 10.1016/j.vlsi.2016.09.003.
- [57] M. Li, P. Huang, L. Shen, Z. Zhou, J.-F. Kang, and X.-Y. Liu, “Simulation of the RRAM-based flip-flops with data retention,” pp. 1–2, 2016, doi: 10.1109/inec.2016.7589321.
- [58] I. Kazi, P. Meinerzhagen, P. E. Gaillardon, D. Sacchetto, A. Burg, and G. De Micheli, “A ReRAM-based non-volatile flip-flop with sub-VT read and CMOS voltage-compatible write,” *2013 IEEE 11th Int. New Circuits Syst. Conf. NEWCAS 2013*, pp. 1–4, 2013, doi: 10.1109/NEWCAS.2013.6573586.
- [59] J. M. Portal, M. Bocquet, D. Deleruyelle, and C. Muller, “Non-volatile Flip-Flop based on unipolar ReRAM for power-down applications,” *J. Low Power Electron.*, vol. 8, no. 1, pp. 1–10, 2012, doi: 10.1166/jolpe.2012.1172.
- [60] D. Wang, S. George, A. Aziz, S. Datta, V. Narayanan, and S. K. Gupta, “Ferroelectric Transistor based Non-Volatile Flip-Flop,” *Proc. Int. Symp. Low Power Electron. Des.*,

- pp. 10–15, 2016, doi: 10.1145/2934583.2934603.
- [61] R. Moazzami, C. Hu, and W. H. Shepherd, “A ferroelectric DRAM cell for high density NVRAMs,” *Dig. Tech. Pap. - Symp. VLSI Technol.*, vol. 1, no. 10, pp. 15–16, 1990, doi: 10.1109/VLSIT.1990.110985.
- [62] M. Pešić, M. Hoffmann, C. Richter, T. Mikolajick, and U. Schroeder, “Nonvolatile Random Access Memory and Energy Storage Based on Antiferroelectric Like Hysteresis in ZrO<sub>2</sub>,” *Adv. Funct. Mater.*, vol. 26, no. 41, pp. 7486–7494, 2016, doi: 10.1002/adfm.201603182.
- [63] B. C. Lee, E. Ipek, O. Mutlu, and D. Burger, “Architecting phase change memory as a scalable DRAM alternative,” *Proc. - Int. Symp. Comput. Archit.*, pp. 2–13, 2009, doi: 10.1145/1555754.1555758.
- [64] W. Wei, K. Namba, and F. Lombardi, “Extending non-volatile operation to DRAM cells,” *IEEE Access*, vol. 1, pp. 758–769, 2013, doi: 10.1109/ACCESS.2013.2288312.
- [65] S. Mittal, J. S. Vetter, and D. Li, “A Survey of architectural approaches for managing embedded DRAM and non-volatile on-chip caches,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 6, pp. 1524–1537, 2015, doi: 10.1109/TPDS.2014.2324563.
- [66] G. Dearnaley, A. M. Stoneham, and D. V. Morgan, “Electrical phenomena in amorphous oxide films,” *Reports Prog. Phys.*, vol. 33, no. 3, pp. 1129–1191, 1970, doi: 10.1088/0034-4885/33/3/306.
- [67] H. Biederman, “Metal-insulator-metal sandwich structures with anomalous properties,” *Vacuum*, vol. 26, no. 12, pp. 513–523, 1976, doi: 10.1016/S0042-207X(76)81130-X.
- [68] H. Pagnia and N. Sotnik, “Bistable switching in electroformed metal–insulator–metal devices,” *Phys. Status Solidi*, vol. 108, no. 1, pp. 11–65, 1988, doi: 10.1002/pssa.2211080102.
- [69] S. Yu and K. Iniewski, *Resistive Random Access Memory (RRAM) From Devices to Array Architectures SyntheSiS LectureS on emerging engineering technoLogieS SyntheSiS LectureS on emerging engineering technoLo.* 2016.
- [70] Y. Xie and Y. Zhao, “Emerging memory technologies,” *IEEE Micro*, vol. 39, no. 1, pp. 6–7, 2019, doi: 10.1109/MM.2019.2892165.

- [71] I. Daniele, "Resistive switching memories based on metal oxides: mechanisms, reliability and scaling," *Semicond. Sci. Technol.*, vol. 31, no. 6, p. 63002, 2016, [Online]. Available: <http://stacks.iop.org/0268-1242/31/i=6/a=063002>.
- [72] M. Alayan, E. Vianello, B. De Salvo, L. Perniola, A. Padovani, and L. Larcher, "Correlated Effects on Forming and Retention of Al Doping in HfO<sub>2</sub>-Based RRAM," *IEEE Des. Test*, vol. 34, no. 3, pp. 23–30, 2017, doi: 10.1109/MDAT.2017.2682246.
- [73] P. W. C. Ho, F. O. Hatem, H. A. F. Almurib, and T. Nandha Kumar, "Comparison on TiO<sub>2</sub> and TaO<sub>2</sub> Based Bipolar Resistive Switching Devices," *IEEE Int. Conf. Electron. Des.*, pp. 249–254, 2014, doi: 10.1109/ICED.2014.7015808.
- [74] C. Nail *et al.*, "Understanding RRAM endurance, retention and window margin trade-off using experimental results and simulations," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, pp. 4.5.1-4.5.4, 2017, doi: 10.1109/IEDM.2016.7838346.
- [75] S. Kim *et al.*, "Physical electro-thermal model of resistive switching in bi-layered resistance-change memory," *Sci. Rep.*, vol. 3, no. 1, p. 1680, 2013, doi: 10.1038/srep01680.
- [76] F. Zahoor, T. Z. Azni Zulkifli, and F. A. Khanday, "Resistive Random Access Memory (RRAM): an Overview of Materials, Switching Mechanism, Performance, Multilevel Cell (mlc) Storage, Modeling, and Applications," *Nanoscale Res. Lett.*, vol. 15, no. 1, 2020, doi: 10.1186/s11671-020-03299-9.
- [77] N. H. El-Hassan, T. N. Kumar, and H. A. F. Almurib, "Phase change memory cell emulator circuit design," *Microelectronics J.*, vol. 62, no. February, pp. 65–71, 2017, doi: 10.1016/j.mejo.2017.02.006.
- [78] T. J. Cardinali *et al.*, "Memristor–CMOS Hybrid Integrated Circuits for Reconfigurable Logic," *Nano Lett.*, vol. 9, no. 10, pp. 3640–3645, 2009, doi: 10.1021/nl901874j.
- [79] Z. Fang, H. Y. Yu, X. Li, N. Singh, G. Q. Lo, and D. L. Kwong, "HfO<sub>x</sub>/TiO<sub>x</sub>/HfO<sub>x</sub>/TiO<sub>x</sub> multilayer-based forming-free RRAM devices with excellent uniformity," *IEEE Electron Device Lett.*, vol. 32, no. 4, pp. 566–568, 2011, doi: 10.1109/LED.2011.2109033.
- [80] F. O. Hatem, P. W. C. Ho, T. N. Kumar, and H. A. F. Almurib, "Modeling of bipolar resistive switching of a nonlinear MISM memristor," *Semicond. Sci. Technol.*, vol. 30,

- no. 11, p. 115009, 2015, doi: 10.1088/0268-1242/30/11/115009.
- [81] D. Kumar, R. Aluguri, U. Chand, and T. Y. Tseng, "Metal oxide resistive switching memory: Materials, properties and switching mechanisms," *Ceram. Int.*, vol. 43, no. xxxx, pp. S547–S556, 2017, doi: 10.1016/j.ceramint.2017.05.289.
- [82] M. Saremi, S. Rajabi, H. J. Barnaby, and M. N. Kozicki, "The Effects of Process Variation on the Parametric Model of the Static Impedance Behavior of Programmable Metallization Cell (PMC)," in *Materials Research Society Symposium Proceedings*, 2014, vol. 1692, doi: 10.1557/opl.2014.521.
- [83] M. Saremi, "A physical-based simulation for the dynamic behavior of photodoping mechanism in chalcogenide materials used in the lateral programmable metallization cells," *Solid State Ionics*, vol. 290, pp. 1–5, 2016, doi: 10.1016/j.ssi.2016.04.002.
- [84] M. Saremi, H. J. Barnaby, A. Edwards, and M. N. Kozicki, "Analytical Relationship between Anion Formation and Carrier-Trap Statistics in Chalcogenide Glass Films," *ECS Electrochem. Lett.*, vol. 4, no. 7, pp. H29–H31, 2015, doi: 10.1149/2.0061507eel.
- [85] M. Saremi, "Carrier mobility extraction method in ChGs in the UV light exposure," *Micro Nano Lett.*, vol. 11, no. 11, pp. 762–764, 2016, doi: 10.1049/mnl.2016.0132.
- [86] P. W. C. Ho, F. O. Hatem, H. A. F. Almurib, and T. N. Kumar, "Enhanced SPICE Memristor Model with Dynamic Ground," in *Circuits and Systems Symposium (ICSSyS), 2015 IEEE International*, 2016, pp. 130–132.
- [87] J. H. Hur, M. J. Lee, C. B. Lee, Y. B. Kim, and C. J. Kim, "Modeling for bipolar resistive memory switching in transition-metal oxides," *Phys. Rev. B*, vol. 82, no. 15, p. 155321, 2010, doi: 10.1103/PhysRevB.82.155321.
- [88] A. Siemon, S. Menzel, A. Marchewka, Y. Nishi, R. Waser, and E. Linn, "Simulation of TaO<sub>x</sub>-based complementary resistive switches by a physics-based memristive model," *Proc. - IEEE Int. Symp. Circuits Syst.*, no. Iwe Ii, pp. 1420–1423, 2014, doi: 10.1109/ISCAS.2014.6865411.
- [89] H. L. Chee, T. N. Kumar, and H. A. Almurib, "Multifilamentary Conduction Modelling of Bipolar Ta<sub>2</sub>O<sub>5</sub>/TaO<sub>x</sub> Bi-Layered RRAM," in *7th IEEE Non-Volatile Memory Systems and Applications Symposium (NVMSA)*, 2018, pp. 113–114, doi: 10.1109/NVMSA.2018.00029.

- [90] G. González-Cordero, F. Jiménez-Molinos, J. B. Roldán, M. B. González, and F. Campabadal, “In-depth study of the physics behind resistive switching in TiN/Ti/HfO<sub>2</sub>/W structures,” *J. Vac. Sci. Technol. B, Nanotechnol. Microelectron. Mater. Process. Meas. Phenom.*, vol. 35, no. 1, p. 01A110, 2017, doi: 10.1116/1.4973372.
- [91] S. Ambrogio, S. Balatti, D. C. Gilmer, and D. Ielmini, “Analytical modeling of oxide-based bipolar resistive memories and complementary resistive switches,” *IEEE Trans. Electron Devices*, vol. 61, no. 7, pp. 2378–2386, 2014, doi: 10.1109/TED.2014.2325531.
- [92] S. Larentis, F. Nardi, S. Balatti, D. C. Gilmer, and D. Ielmini, “Resistive switching by voltage-driven ion migration in bipolar RRAM Part II: Modeling,” *IEEE Trans. Electron Devices*, vol. 59, no. 9, pp. 2468–2475, 2012, doi: 10.1109/TED.2012.2202320.
- [93] S. Kim, S. Choi, and W. Lu, “Comprehensive physical model of dynamic resistive switching in an oxide memristor,” *ACS Nano*, vol. 8, no. 3, pp. 2369–2376, 2014, doi: 10.1021/nm405827t.
- [94] Y. Zhao *et al.*, “Modeling and optimization of bilayered TaO<sub>x</sub> RRAM based on defect evolution and phase transition effects,” *IEEE Trans. Electron Devices*, vol. 63, no. 4, pp. 1524–1532, 2016, doi: 10.1109/TED.2016.2532470.
- [95] P. Huang *et al.*, “A physics-based compact model of metal-oxide-based RRAM DC and AC operations,” *IEEE Trans. Electron Devices*, vol. 60, no. 12, pp. 4090–4097, 2013, doi: 10.1109/TED.2013.2287755.
- [96] Y. D. Zhao *et al.*, “Simulation of TaOX-RRAM with Ta<sub>2</sub>O<sub>5</sub>-X/TaO<sub>2</sub>-X stack engineering,” *Int. Conf. Simul. Semicond. Process. Devices, SISPAD*, vol. 2015-October, pp. 285–288, 2015, doi: 10.1109/SISPAD.2015.7292315.
- [97] H. Li *et al.*, “Variation-Aware, Reliability-Emphasized Design and Optimization of RRAM using SPICE Model,” *Des. Autom. Test Eur. Conf. Exhib. (DATE), 2015*, pp. 1425–1430, 2015, doi: 10.7873/DATE.2015.0362.
- [98] A. L. Jagath, T. N. Kumar, and H. A. F. Almurib, “Modeling of Current Conduction during RESET Phase of Pt/Ta<sub>2</sub>O<sub>5</sub>/TaO<sub>x</sub>/Pt Bipolar Resistive RAM Devices,” *Proc. - 7th IEEE Non-Volatile Mem. Syst. Appl. Symp. NVMSA 2018*, pp. 55–60, 2018, doi: 10.1109/NVMSA.2018.00014.
- [99] J. H. Hur *et al.*, “Modeling for multilevel switching in oxide-based bipolar resistive

- memory,” *Nanotechnology*, vol. 23, no. 22, 2012, doi: 10.1088/0957-4484/23/22/225702.
- [100] Z. Jiang *et al.*, “A Compact model for metal-oxide resistive random access memory with experiment verification,” *IEEE Trans. Electron Devices*, vol. 63, no. 5, pp. 1884–1892, 2016, doi: 10.1109/TED.2016.2545412.
- [101] M. Bocquet *et al.*, “Robust compact model for bipolar oxide-based resistive switching memories,” *IEEE Trans. Electron Devices*, vol. 61, no. 3, pp. 674–681, 2014, doi: 10.1109/TED.2013.2296793.
- [102] S. Larentis, F. Nardi, S. Balatti, D. C. Gilmer, and D. Ielmini, “Resistive Switching by Voltage-Driven Ion Migration in Bipolar RRAM—Part I: Experimental Study,” *IEEE Trans. Electron Devices*, vol. 59, no. 9, pp. 2468–2475, 2012, doi: 10.1109/TED.2012.2202320.
- [103] M. A. Villena *et al.*, “An in-depth simulation study of thermal reset transitions in resistive switching memories,” *J. Appl. Phys.*, vol. 114, no. 14, 2013, doi: 10.1063/1.4824292.
- [104] J. Simmons, “Richardson-Schottly Effects in Solids,” *Phys. Rev. Lett.*, vol. 15, no. 25, pp. 967–968, 1965, doi: 10.1103/PhysRevLett.15.967.
- [105] C. E. Graves, N. Dávila, E. J. Merced-Grafals, S. T. Lam, J. P. Strachan, and R. S. Williams, “Temperature and field-dependent transport measurements in continuously tunable tantalum oxide memristors expose the dominant state variable,” *Appl. Phys. Lett.*, vol. 110, no. 12, 2017, doi: 10.1063/1.4978757.
- [106] J. H. Yoon *et al.*, “Highly uniform, electroforming-free, and self-rectifying resistive memory in the Pt/Ta<sub>2</sub>O<sub>5</sub>/HfO<sub>2-x</sub>/TiN structure,” *Adv. Funct. Mater.*, vol. 24, no. 32, pp. 5086–5095, 2014, doi: 10.1002/adfm.201400064.
- [107] S. M. Sze and M. K. Lee, *Semiconductor Devices: Physics and Technology*. Wiley, 2012.
- [108] T. Chang, S. H. Jo, K. H. Kim, P. Sheridan, S. Gaba, and W. Lu, “Synaptic behaviors and modeling of a metal oxide memristive device,” *Appl. Phys. A*, vol. 102, no. 4, pp. 857–863, 2011, doi: 10.1007/s00339-011-6296-1.
- [109] F. Gül, “Addressing the sneak-path problem in crossbar RRAM devices using memristor-based one Schottky diode-one resistor array,” *Results Phys.*, vol. 12, pp.

- 1091–1096, 2019, doi: 10.1016/j.rinp.2018.12.092.
- [110] E. Linn, R. Rosezin, C. Kügeler, and R. Waser, “Complementary resistive switches for passive nanocrossbar memories,” *Nat. Mater.*, vol. 9, no. 5, pp. 403–406, 2010, doi: 10.1038/nmat2748.
- [111] J. Lee *et al.*, “Review of candidate devices for neuromorphic applications,” in *ESSDERC 2019 - 49th European Solid-State Device Research Conference (ESSDERC)*, 2019, pp. 22–27, doi: 10.1109/ESSDERC.2019.8901694.
- [112] M. M. Rehman, H. M. M. U. Rehman, J. Z. Gul, W. Y. Kim, K. S. Karimov, and N. Ahmed, “Decade of 2D-materials-based RRAM devices: a review,” *Sci. Technol. Adv. Mater.*, vol. 21, no. 1, pp. 147–186, 2020, doi: 10.1080/14686996.2020.1730236.
- [113] U. Dilna and S. N. Prasad, “Comparative Study of Selector Device Design for Sneak Current in 3D Crosspoint ReRAM,” *MPCIT 2020 - Proc. IEEE 3rd Int. Conf. "Multimedia Process. Commun. Inf. Technol.*, pp. 138–145, 2020, doi: 10.1109/MPCIT51588.2020.9350434.
- [114] L. Zhang, S. Cosemans, D. J. Wouters, G. Groeseneken, M. Jurczak, and B. Govoreanu, “On the optimal ON/OFF resistance ratio for resistive switching element in one-selector one-resistor crosspoint arrays,” *IEEE Electron Device Lett.*, vol. 36, no. 6, pp. 570–572, 2015, doi: 10.1109/LED.2015.2427313.
- [115] Y. Deng *et al.*, “Design and optimization methodology for 3D RRAM arrays,” *Tech. Dig. - Int. Electron Devices Meet. IEDM*, pp. 629–632, 2013, doi: 10.1109/IEDM.2013.6724693.
- [116] L. Zhang, S. Cosemans, D. J. Wouters, B. Govoreanu, G. Groeseneken, and M. Jurczak, “Analysis of vertical cross-point resistive memory (VRRAM) for 3D RRAM design,” *2013 5th IEEE Int. Mem. Work. IMW 2013*, pp. 155–158, 2013, doi: 10.1109/IMW.2013.6582122.
- [117] X. Xu *et al.*, “Fully CMOS compatible 3D vertical RRAM with self-Aligned self-selective cell enabling sub-5nm scaling,” *Dig. Tech. Pap. - Symp. VLSI Technol.*, vol. 2016-Sept, pp. 2015–2016, 2016, doi: 10.1109/VLSIT.2016.7573388.
- [118] ““Predictive Technology Model,”” *Nanoscale Integration and Modeling (NIMO) Group, ASU*, 2007. <http://ptm.asu.edu>.

- [119] A. Wedig *et al.*, “Nanoscale cation motion in TaO<sub>x</sub>, HfO<sub>x</sub> and TiO<sub>x</sub> memristive systems,” *Nat. Nanotechnol.*, vol. 11, pp. 67–74, 2015, doi: 10.1038/nnano.2015.221.
- [120] P. Bousoulas and D. Tsoukalas, “Understanding the Formation of Conducting Filaments in RRAM Through the Design of Experiments,” *Int. J. High Speed Electron. Syst.*, vol. 25, no. 01n02, p. 1640007, 2016, doi: 10.1142/S0129156416400073.
- [121] R. T. Tung, “The physics and chemistry of the Schottky barrier height,” *Appl. Phys. Rev.*, vol. 1, no. 1, p. 011304, 2014, doi: 10.1063/1.4858400.
- [122] M. J. Lee *et al.*, “A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta<sub>2</sub>O<sub>5-x</sub>/TaO<sub>2-x</sub> bilayer structures,” *Nat. Mater.*, vol. 10, pp. 625–630, 2011, doi: 10.1038/nmat3070.
- [123] J. Wu, L. F. Register, and E. Rosenbaumt, “Trap-assisted tunneling current through ultra-thin oxide,” in *1999 IEEE International Reliability Physics Symposium Proceedings. 37th Annual (Cat. No.99CH36296)*, 1999, pp. 389–395, doi: 10.1109/RELPHY.1999.761644.
- [124] F. M. Puglisi, L. Larcher, A. Padovani, and P. Pavan, “A Complete Statistical Investigation of RTN in HfO<sub>2</sub>-Based RRAM in High Resistive State,” *IEEE Trans. Electron Devices*, vol. 62, no. 8, pp. 2606–2613, 2015, doi: 10.1109/TED.2015.2439812.
- [125] D. Veksler *et al.*, “Random telegraph noise (RTN) in scaled RRAM devices,” in *2013 IEEE International Reliability Physics Symposium (IRPS)*, 2013, p. MY.10.1-MY.10.4, doi: 10.1109/IRPS.2013.6532101.
- [126] H. C. Card and E. H. Rhoderick, “Studies of tunnel MOS diodes I. Interface effects in silicon Schottky diodes,” *J. Phys. D. Appl. Phys.*, vol. 4, no. 10, pp. 1589--1601, 1971, doi: 10.1088/0022-3727/4/10/319.
- [127] A. Prakash, D. Deleruyelle, J. Song, M. Bocquet, and H. Hwang, “Resistance controllability and variability improvement in a TaO<sub>x</sub>-based resistive memory for multilevel storage application,” *Appl. Phys. Lett.*, vol. 106, no. 23, p. 233104, 2015, doi: 10.1063/1.4922446.
- [128] Z. Alamgir, K. Beckmann, J. Holt, and N. C. Cady, “Pulse width and height modulation for multi-level resistance in bi-layer TaO<sub>x</sub> based RRAM,” *Appl. Phys. Lett.*, vol. 111, no. 6, 2017, doi: 10.1063/1.4993058.

- [129] Z. Chai *et al.*, “The Over-Reset Phenomenon in Ta<sub>2</sub>O<sub>5</sub> RRAM Device Investigated by the RTN-Based Defect Probing Technique,” *IEEE Electron Device Lett.*, vol. 39, no. 7, pp. 955–958, 2018, doi: 10.1109/LED.2018.2833149.
- [130] M. R. Garg and A. Tonk, “A Study of Different Types of Voltage & Current Sense Amplifiers used in SRAM,” *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 4, no. 5, pp. 30–35, 2015, doi: 10.17148/IJARCCCE.2015.4507.
- [131] M. Uddin and G. S. Rose, “A Practical Sense Amplifier Design for Memristive Crossbar Circuits (PUF),” *Int. Syst. Chip Conf.*, vol. 2018-Septe, pp. 209–214, 2019, doi: 10.1109/SOCC.2018.8618502.
- [132] A. Lee, C. C. Lin, T. C. Yang, and M. F. Chang, “An embedded ReRAM using a small-offset sense amplifier for low-voltage operations,” *2015 Int. Symp. VLSI Des. Autom. Test, VLSI-DAT 2015*, vol. 1, pp. 8–11, 2015, doi: 10.1109/VLSI-DAT.2015.7114532.
- [133] J. Yin *et al.*, “A 0.75V reference clamping sense amplifier for low-power highdensity ReRAM with dynamic pre-charge technique,” *IEICE Electron. Express*, vol. 16, no. 12, pp. 1–6, 2019, doi: 10.1587/elex.16.20190201.
- [134] M. T. I. Badal, M. B. I. Reaz, A. Farayez, S. A. B. Ramli, and N. Kamal, “Design of a low-power CMOS Level Shifter for low-delay SoCs in silterra 0.13  $\mu\text{m}$  CMOS process,” *J. Eng. Sci. Technol. Rev.*, vol. 10, no. 4, pp. 10–15, 2017, doi: 10.25103/jestr.104.02.
- [135] H. L. Chee, T. N. Kumar, and H. A. F. Almurib, “Electrical model of multi-level bipolar Ta<sub>2</sub>O<sub>5</sub>/TaO<sub>x</sub> Bi-layered ReRAM,” *Microelectronics J.*, vol. 93, no. March, p. 104616, 2019, doi: 10.1016/j.mejo.2019.104616.
- [136] C. Yakopicic, T. M. Taha, G. Subramanyam, and R. E. Pino, “Memristor SPICE model and crossbar simulation based on devices with nanosecond switching time,” in *Proceedings of the International Joint Conference on Neural Networks*, 2013, pp. 1–7, doi: 10.1109/IJCNN.2013.6706773.
- [137] R. Naous, M. Al-Shedivat, and K. N. Salama, “Stochasticity modeling in memristors,” *IEEE Trans. Nanotechnol.*, vol. 15, no. 1, pp. 15–28, 2016, doi: 10.1109/TNANO.2015.2493960.
- [138] S. R. Lee *et al.*, “Multi-level switching of triple-layered TaO<sub>x</sub> RRAM with excellent reliability for storage class memory,” in *2012 Symposium on VLSI Technology (VLSIT)*,

- 2012, pp. 71–72, doi: 10.1109/VLSIT.2012.6242466.
- [139] X. Sheng *et al.*, “Low-Conductance and Multilevel CMOS-Integrated Nanoscale Oxide Memristors,” *Adv. Electron. Mater.*, vol. 5, no. 9, pp. 1–8, 2019, doi: 10.1002/aelm.201800876.
- [140] H. A. F. Almurib, F. Lombardi, and T. N. Kumar, “Design and evaluation of a memristor-based look-up table for non-volatile field programmable gate arrays,” *IET Circuits, Devices Syst.*, vol. 10, no. 4, pp. 292–300, 2016, doi: 10.1049/iet-cds.2015.0217.
- [141] K. E. Murray, S. Whitty, S. Liu, J. Luu, and V. Betz, “Timing-driven titan: Enabling large benchmarks and exploring the gap between academic and commercial CAD,” *ACM Trans. Reconfigurable Technol. Syst.*, vol. 8, no. 2, 2015, doi: 10.1145/2629579.
- [142] H. A. F. Almurib, T. N. Kumar, and F. Lombardi, “A memristor-based LUT for FPGAs,” in *The 9th IEEE International Conference on Nano/Micro Engineered and Molecular Systems (NEMS)*, 2014, pp. 448–453, doi: 10.1109/NEMS.2014.6908847.
- [143] S. Sakaidani, N. Miyamoto, and T. Ohmi, “Flexible Processor Based on Full-Adder/D-Flip-Flop Merged Module (FDMM),” *Jpn. J. Appl. Phys.*, vol. 40, no. Part 1, No. 4B, pp. 2581–2584, 2001, doi: 10.1145/370155.370254.
- [144] M. Motomura, Y. Aimoto, A. Shibayama, Y. Yabe, and M. Yamashina, “An embedded DRAM-FPGA chip with instantaneous logic reconfiguration,” *Proc. - IEEE Symp. FPGAs Cust. Comput. Mach. FCCM 1998*, vol. 1998-April, pp. 264–266, 1998, doi: 10.1109/FPGA.1998.707909.
- [145] C. Yakopcic, T. M. Taha, G. Subramanyam, R. E. Pino, and S. Rogers, “A Memristor Device Model,” *IEEE Electron Device Lett.*, vol. 32, no. 10, pp. 1436–1438, 2011, doi: 10.1109/LED.2011.2163292.
- [146] M. Ueki *et al.*, “Low-power embedded ReRAM technology for IoT applications,” in *IEEE Symposium on VLSI Circuits, Digest of Technical Papers*, 2015, vol. 2015-Augus, pp. T108–T109, doi: 10.1109/VLSIC.2015.7231367.
- [147] W. Feng, H. Shima, K. Ohmori, and H. Akinaga, “Investigation of switching mechanism in HfO<sub>x</sub>-ReRAM under low power and conventional operation modes,” *Sci. Rep.*, vol. 6, no. April, pp. 1–8, 2016, doi: 10.1038/srep39510.

- [148] P. F. Chiu *et al.*, “Low store energy, low VDDmin, 8T2R nonvolatile latch and SRAM with vertical-stacked resistive memory (memristor) devices for low power mobile applications,” *IEEE J. Solid-State Circuits*, vol. 47, no. 6, pp. 1483–1496, 2012, doi: 10.1109/JSSC.2012.2192661.
- [149] M. Fliesler, D. Still, and J. M. Hwang, “A 15ns 4Mb NVSRAM in 0.13u SONOS technology,” *2008 Jt. Non-Volatile Semicond. Mem. Work. Int. Conf. Mem. Technol. Des. Proceedings, NVSMW/ICMTD*, vol. 00, no. c, pp. 83–86, 2008, doi: 10.1109/NVSMW.2008.30.
- [150] N. Sakimura, T. Sugibayashi, R. Nebashi, and N. Kasai, “Nonvolatile magnetic flip-flop for standby-power-free SoCs,” *IEEE J. Solid-State Circuits*, vol. 44, no. 8, pp. 2244–2250, 2009, doi: 10.1109/JSSC.2009.2023192.
- [151] T. Miwa *et al.*, “NV-SRAM: A nonvolatile SRAM with backup ferroelectric capacitors,” *IEEE J. Solid-State Circuits*, vol. 36, no. 3, pp. 522–527, 2001, doi: 10.1109/4.910492.
- [152] W. Wang *et al.*, “Nonvolatile SRAM cell,” *Tech. Dig. - Int. Electron Devices Meet. IEDM*, pp. 2–5, 2006, doi: 10.1109/IEDM.2006.346730.
- [153] M. Takata, K. Nakayama, T. Izumi, T. Shinmura, J. Akita, and A. Kitagawa, “Nonvolatile SRAM based on phase change,” *21st IEEE Non-Volatile Semicond. Mem. Work. 2006, NVSMW 2006*, vol. 2006, no. c, pp. 95–96, 2006, doi: 10.1109/.2006.1629510.
- [154] Z. Swaidan, R. Kanj, J. El Hajj, E. Saad, and F. Kurdahi, “RRAM endurance and retention: Challenges, opportunities and implications on reliable design,” *2019 26th IEEE Int. Conf. Electron. Circuits Syst. ICECS 2019*, pp. 402–405, 2019, doi: 10.1109/ICECS46596.2019.8964707.
- [155] S. Shiratake, “Scaling and Performance Challenges of Future DRAM,” *2020 IEEE Int. Mem. Work. IMW 2020 - Proc.*, pp. 2020–2022, 2020, doi: 10.1109/IMW48823.2020.9108122.
- [156] U. Kang *et al.*, “Co-Architecting Controllers and DRAM to Enhance DRAM Process Scaling,” *Mem. Forum*, pp. 1–4, 2014.
- [157] S. Ganapathy, R. Canal, D. Alexandrescu, E. Costenaro, A. González, and A. Rubio, “A novel variation-tolerant 4T-DRAM cell with enhanced soft-error tolerance,” *Proc. -*

- IEEE Int. Conf. Comput. Des. VLSI Comput. Process.*, pp. 472–477, 2012, doi: 10.1109/ICCD.2012.6378681.
- [158] K. C. Chun, P. Jain, J. H. Lee, and C. H. Kim, “A sub-0.9V logic-compatible embedded DRAM with boosted 3T gain cell, regulated bit-line write scheme and PVT-tracking read reference bias,” *IEEE Symp. VLSI Circuits, Dig. Tech. Pap.*, vol. 38, no. 4, pp. 134–135, 2009.
- [159] B. J. Choi *et al.*, “High-Speed and Low-Energy Nitride Memristors,” *Adv. Funct. Mater.*, vol. 26, no. 29, pp. 5290–5296, 2016, doi: 10.1002/adfm.201600680.
- [160] S. Akashe, A. Mudgal, and S. B. Singh, “Analysis of power in 3T DRAM and 4T DRAM Cell design for different technology,” in *Proceedings of the 2012 World Congress on Information and Communication Technologies, WICT 2012*, 2012, pp. 18–21, doi: 10.1109/WICT.2012.6409043.
- [161] I. Kuon, R. Tessier, and J. Rose, “FPGA architecture: Survey and challenges,” *Found. Trends Electron. Des. Autom.*, vol. 2, no. 2, pp. 135–253, 2007, doi: 10.1561/1000000005.
- [162] T. Ahmed, P. D. Kundarewich, and J. H. Anderson, “Packing techniques for virtex-5 FPGAs,” *ACM Trans. Reconfigurable Technol. Syst.*, vol. 2, no. 3, 2009, doi: 10.1145/1575774.1575777.
- [163] R. H. Freeman, “Configurable electrical circuit having configurable logic elements and configurable interconnects,” 4 870 302, 1989.
- [164] H. Bazzi, A. Harb, H. Aziza, and M. Moreau, “Non-volatile SRAM memory cells based on ReRAM technology,” *SN Appl. Sci.*, vol. 2, no. 9, pp. 1–13, 2020, doi: 10.1007/s42452-020-03267-z.
- [165] N. Andreeva, A. Ivanov, and A. Petrov, “Multilevel resistive switching in TiO<sub>2</sub>/Al<sub>2</sub>O<sub>3</sub> bilayers at low temperature,” *AIP Adv.*, vol. 8, no. 2, p. 025208, 2018, doi: 10.1063/1.5019570.