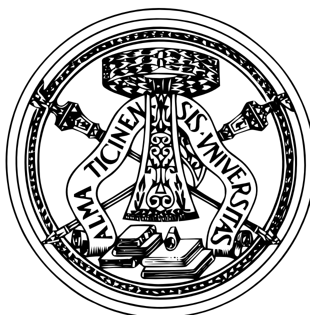


UNIVERSITÀ DEGLI STUDI DI PAVIA
Department of Electrical, Computer and Biomedical Engineering

Ph.D. School in Microelectronics
XXXI Cycle

Design, Modeling and Characterization of Circuits and Devices for Emerging Memories

Ph.D. Thesis of
Yilkal Andualem Belay

Supervisors:
Prof. Guido Torelli
Prof. Alessandro Cabrini

Coordinator:
Prof. Guido Torelli

January 2019

Abstract

Novel data storage device concepts and high density architectures have been under exploration to meet the memory performance and storage capacity demand, which is growing exponentially and becoming challenging to be met by the mainstream memory technologies. Scaling down to advanced technology nodes is needed for increasing storage capacity and area efficiency. However, the mainstream memories namely, Static Random Access Memory (SRAM), Dynamic Random Access Memory (DRAM), and Flash memory are facing issues such as, reliability degradation and increasing leakage power.

As a result, emerging memory technologies such as Resistive Random Access Memory (RRAM), Phase Change Memory (PCM), and Spin-Transfer Torque Magnetic Random Access Memory (STT-MRAM) are under research and development due to their promising scalability, reduced standby power, access speed, and other interesting features. They are expected not only to replace the mainstream memories but also to create new memory markets for instance, a Storage Class Memory (SCM) that combines a storage capacity comparable to that of NAND Flash memory and speed comparable to that of DRAM. The predominant emerging memories are based on resistance switching principle: the resistance of the storage device is switched between a high-resistance state (HRS) and a low-resistance state (LRS), which are used to store binary data. In addition, high density storage solutions such as multi-level cell (multiple bits in a single cell), 3D integration, and crosspoint arrays are under exploration to meet the high storage capacity demand. For instance, in crosspoint arrays, memory cells are built at the junctions of a lower and an upper plane of parallel metal lines running at right angles to each other, and hence if both the width of the metal lines and the spacing between them is equal to the minimum lithographic feature size, F, the memory cell can be allocated within the smallest footprint of $2F \times 2F$.

This thesis presents a model-based study of STT-MRAM cell, array and sensing circuits, an experimental electrical characterization of two (types of) RRAM device stacks, and a detail analysis of the design considerations for write and read operations and device technology requirements in crosspoint memory arrays. With respect to STT-MRAM, the thesis mainly contributes a comprehensive behavioral model of STT-MRAM cell for circuit simulations and review and analysis of sensing circuit schemes for conventional and crosspoint STT-MRAM arrays. The performance (i.e., sense margin) of various sensing circuit schemes are analyzed by taking into account the impact of cell-to-cell variations and parasitics in bitlines, BLs, and wordlines, WLs.

As for RRAM, a detailed array-level experimental electrical characterization of OxRAM (Oxide RAM) with TiN/Hf/GdAlO/TiN device stack and and device-level reliability study of CBRAM (Conductive Bridging RAM) with Cu/TiW/SrTiOx/WOx/W stack are presented. The characterization of the OxRAM stack focuses on analyzing operating voltages and cyclic endurance at array level, with the prospect of tuning the performance for embedded and storage class memory applications. The impacts of the thickness of the GdAlO layer and device size on operating voltages and endurance are discussed. On CBRAM, the experimental tests, which the results are presented in this thesis, were aimed at studying the Cu/TiW/SrTiOx/WOx/W stack and to optimizing the operating voltages and currents to obtain optimum memory performance and reliability.

Finally, this thesis presents a comprehensive study on design considerations for write and read operations and device technology requirements of 1S1R (one-selector one-resistor) crosspoint memory arrays. Indeed, crosspoint memory arrays has gained much attention as an architecture to obtain high-density storage. However, the successful implementation of large-size crosspoint arrays is hindered by some issues that need to be addressed. The most critical issue is that when we activate certain memory cell(s) many other cells that are not intended to be written/read will be partially activated resulting in sneak current paths, which lead to excessive leakage power consumption and write and read performance degradation. To solve this issue, the memory cell should have a strongly nonlinear current-voltage (I-V) characteristic, i.e. it should turn off at low bias voltages and turn on at adequately larger bias. One of the approaches for introducing nonlinearity is by integrating a two-terminal nonlinear selector device in series with each memory element, thus giving rise to an 1S1R configuration. By using analytical and circuit models, the thesis analyzes the dependence of 1S1R crosspoint memory array performance on the characteristics of selector device (nonlinearity and operating voltage), memory element (ratio of resistance of HRS to that of LRS, switching current and voltage) and interconnection metal line (parasitic resistance).

Acknowledgements

First of all, I would like to express my special thanks to my supervisors Professor Guido Torelli and Professor Alessandro Cabrini for the guidance and encouragement throughout my research. Without your support and guidance, completing this work would have been unthinkable. I would also like to extend my gratitude to fellow members of my research team Dr. Riccardo Zurla and Flavio Giovanni Volpe, for their feedback, cooperation and friendship.

My next gratitude goes to Dr. Andrea Fantini, who was my mentor when I was in imec (Leuven, Belgium) for six months research internship. Thank you for your very generous supervision and experience sharing, and also for familiarizing me not only with the experimental procedures of the labs but also with everything in the work environment of imec. I also worked with Dr. Attilio Belmonte during my last two months at imec. I want to say thank you for giving me the opportunity to work with you and for making me learn a lot in this brief time. My gratitude also goes to Ludovic Goux, manager of memory device design group, and Gouri Sankar Kar, Program Director of emerging memories at imec, for giving me the opportunity to join the research team.

Last but not least, I would like to say thank you to my parents, sisters and brothers for their love support. My friends, Moses, Awet, Beza, Meseret and residents of Collegio Spallanzani, thank you for all the good moments.

Contents

Contents iv

List of Figures viii

List of Tables ix

1 Introduction 1
- 1.1 General Overview 1
- 1.2 The Memory Sub-System 3
  - 1.2.1 Memory Hierarchy 3
- 1.3 Mainstream Solid-State Memories 5
  - 1.3.1 Static Random Access Memory (SRAM) 5
  - 1.3.2 Dynamic Random Access Memory (DRAM) 6
  - 1.3.3 Flash Nonvolatile Memory 7
- 1.4 Challenges of Mainstream Memories and Current Research and Development Trends 9
- 1.5 Objectives of the Thesis 11
- 1.6 Organization of the Thesis 12

2 Spin-Transfer Torque Magnetic RAM (STT-MRAM) 14
- 2.1 Overview 14
- 2.2 Key Technology Advances and Challenges of STT-MRAM 15
- 2.3 Basic Physics of STT-MRAM 17
  - 2.3.1 Principle of Operation 17
  - 2.3.2 Switching Current 19
  - 2.3.3 Probability of Switching 21
  - 2.3.4 Static Behavior of STT-MRAM Cell 22
- 2.4 Model of STT-MRAM 22
- 2.5 Sensing Circuits for STT-MRAM 26
  - 2.5.1 Conventional Sensing Scheme 28
  - 2.5.2 Non-destructive Self-Reference Sensing Scheme 28

2.5.3 Destructive Self-reference Sensing Scheme 29
2.5.4 Slope Detection Self-Reference Sensing Scheme 30
2.5.5 Variation-Aware Analysis of Sensing Margin in Slope Detection Sensing Scheme 32
2.6 Conclusion 36

3 Electrical Characterization of Resistive Memories 38
3.1 Overview 38
3.2 Basics of Resistive RAM (RRAM) 39
3.3 Electrical Characterization of GdAlO-Based OxRAM 41
3.3.1 OxRAM Device Stacks and Experimental Setup 41
3.3.2 Forming 43
3.3.3 RESET Voltage 46
3.3.4 SET Voltage 47
3.3.5 Cyclic Endurance 48
3.4 Electrical Characterization of $SrTiO_3$-Based CBRAM 51
3.4.1 Device Stack and Experimental Setup 51
3.4.2 Memory Window, Endurance and Retention 52
3.5 Conclusion and Outlook 54

4 Crosspoint Memory Arrays for High-Density Storage 56
4.1 Overview of Crosspoint Memory Arrays 56
4.2 Challenges of Crosspoint Memory Arrays 58
4.3 Selector Devices for Crosspoint Memory Arrays 61
4.3.1 Selector Device Requirements 61
4.3.2 Selector Device Technologies 63
4.3.3 Model of Selector Device 64
4.4 Model for Resistance-Switching Memory Elements 66
4.5 Interconnection Metal Line Scaling 68
4.6 Model for Large 1S1R Crosspoint Arrays 69
4.7 Design considerations for Write Operation in Crosspoint Arrays 71
4.7.1 Simplified Analysis of Boundary Conditions 71
4.7.2 Write Requirements in Practical-Size Arrays 73
4.8 Design considerations for Read Operation in Crosspoint Arrays 81
4.9 A Variability-Aware Analysis of the Voltage Compatibility of Memory Element and Selector Device 83
4.10 Biasing Crosspoint Memory Arrays for Minimum Leakage Power 90
4.10.1 x Biasing Scheme 91
4.10.2 Estimation of Leakage Power 92
4.11 Conclusion 94

5 General Conclusions and Future Prospects **96**
5.1 Conclusions 96
5.2 Future Prospects 98

List of Figures

1.1 Memory array organization 4
1.2 Physical distribution and hierarchy of memory in a computer . 4
1.3 Static RAM operation 5
1.4 6-Transistor SRAM cell 6
1.5 1-Transistor DRAM cell 6
1.6 Floating gate flash memory cell 7
1.7 Flash memory: (a) NOR (b) NAND 8
1.8 32-layer 3D NAND Flash memory 10

2.1 STT-MRAM cell with a MOS transistor as a selector 17
2.2 STT switching (a) antiparallel to parallel switching (b) parallel to antiparallel switching 19
2.3 Block diagram of the proposed STT-MRAM Verilog-A model . 23
2.4 R-I characteristic: asymmetrical (dashed line) switching currents for the AP→P and P→AP transitions; symmetrical switching (solid line) 25
2.5 Switching probability for different values of the thermal stability factor Δ. The inset shows variation of switching current with temperature 25
2.6 Simulated switching characteristic of STT-MRAM cell: normalized write current (blue, left y-axis) and the state of the cell (right y-axis) 26
2.7 Illustration of STT-MRAM arrays (a) conventional array (b) crosspoint array 27
2.8 Reading in STT-MRAM: (a) simplified read path model (b) conventional scheme with the simplified read path model . . . 27
2.9 Nondestructive self-reference sensing (source: [1]) 29
2.10 Slope detection sensing: (a) circuit scheme (b) timing diagram 30
2.11 Sampling in slope detection sensing 31
2.12 Normal inverse distribution of cell resistance 34
2.13 Normal inverse distribution of sense margin for $\alpha = 0$ and $\alpha = 1$ ($R_{line} = 0$) 35

2.14 Normal inverse distribution of sense margin for $R_{line} = 0$ and $R_{line} = 10$ kΩ ($\alpha = 0$) 35
2.15 Comparison of best-case and worst-case sense margins 36

3.1 Embedded RRAM requirements 39
3.2 RRAM device structure and filamentary switching mechanism 40
3.3 OxRAM pillar device stack 41
3.4 (a) Isolated (ISO) and dense (DS) memory cells (b) Memory chip with different Mbit arrays 42
3.5 OxRAM Memory stack implemented between Metal-3 (M3) and Metal-4 (M4) 42
3.6 Schematic view of 1T1R configuration and biasing condition for RESET, FORMING, SET, and read operations 42
3.7 Programing pulses: Incremental Single Pulse (left) and single Pulse programing (right) 43
3.8 Forming in D02 (5 nm GdAlO), ISO (isolated), 150 nm (cell diameter): (a) forming R-V-t plot and (b) Distribution of post-forming resistance 44
3.9 Forming in D02 (5 nm GdAlO), DS (dense), 60 nm (cell diameter): (a) forming R-V-t plot and (b) Distribution of post-forming resistance 44
3.10 Forming in D03 (3 nm GdAlO), ISO (isolated), 150 nm (cell diameter): (a) forming R-V-t plot and (b) Distribution of post-forming resistance 45
3.11 Forming in D03 (3 nm GdAlO), DS (dense), 60 nm (cell diameter): (a) forming R-V-t plot and (b) Distribution of post-forming resistance 45
3.12 RESET R-V-t in D02 (5 nm GdAlO stack): (a) Isolated, 150 nm and (b) Dense, 60 nm 46
3.13 RESET R-V-t in D03 (3 nm GdAlO): (a) Isolated 150 nm and (b) Dense 60 nm 46
3.14 SET R-V-t in D02 (5 nm GdAlO stack) plot: (a) Isolated 150 nm cells and (b) Dense 60 nm cells 47
3.15 SET R-V-t in D03 (3 nm GdAlO stack): (a) Isolated, 150 nm and (b) Dense, 60 nm 47
3.16 Cyclic Endurance in D02 (5 nm GdAlO) stack (150 μA compliance current): (a) Isolated 150 nm and (b) Dense 60 nm . . 48
3.17 Cyclic Endurance in D03 (5 nm GdAlO) stack (150 μA compliance current): (a) Isolated 150 nm and (b) Dense 60 nm . . 48
3.18 CBRAM (a) 1T1R structure (b) STO-based CBRAM device stack 51

3.19 Reliability of STO-based CBRAM at 10 μA, 10 ns, V_{SET} = 3.5 V, V_{RESET} = -3.0 V: (a) cyclic endurance and (b) retention 52
3.20 Comparison of endurance (V_{SET} = 3.5 V, V_{RESET} = -3.0 V): (a) forming and SET at 50 μA and (b) forming at 10 μA and SET at 50 μA 53
3.21 Endurance comparison (V_{SET} = 3.5 V, V_{RESET} = -2.5 V, forming at 10 μA and SET at 50 μA): (a) rectangular pulse (b) triangular pulse (5 ns-1 ns-5 ns) 53
3.22 Comparison of endurance (V_{SET} = 3.5 V, V_{RESET} = -2.5 V, forming at 10 μA and SET at 50 μA): (a) rectangular pulse (b) triangular pulse (5ns-1ns-5ns) 54

4.1 Schematics of 3D crosspoint memory arrays 57
4.2 1S1R crosspoint array: (a) physical schematics of a 2 layer array (b) a circuit schematic of a planar array with one selected cell shown 59
4.3 Representative I-V characteristics of selector devices: exponential selectors (left) and threshold selectors (right) 63
4.4 I-V characteristics of exponential (left)) and threshold (right) selector devices 64
4.5 Typical I-V characteristics of bipolar RRAM, STTRAM, and PCM 67
4.6 Physical schematics of a 2×2 crosspoint array with parameters of interconnection metal lines 68
4.7 Simplified circuit model of a crosspoint array when programming a memory cell at the lower right corner 70
4.8 Dependence of feasible crosspoint array size on low-leakage voltage margin, V_m, of exponential selector device 77
4.9 Maximum feasible size of square crosspoint array built with threshold selector versus threshold voltage of selector 77
4.10 Off-state resistance versus leakage current in threshold selector devices 78
4.11 Maximum feasible crosspoint array size versus nonlinearity of selector device 79
4.12 Dependence of maximum array size of on I_{sw} and V_{sw} of memory element 79
4.13 Dependence of maximum array size of on I_{sw} and V_{sw} of the memory element (for higher values of I_{sw} and V_{sw}) 80
4.14 Impact of of interconnection metal line scaling on array size . 80

4.15 Total sensing margin (using 10 μA read current), as a function of crosspoint array size and memory element resistance ratio, $\frac{R_H}{R_L}$ . . . 82
4.16 Total sensing margin (using 10 μA read current), as a function of array size and cell state resistance ratio, $\frac{R_H}{R_L}$ . . . 83
4.17 Circuit simulation of a crosspoint array with 256×256 memory (V_{th} = 1.1 V and V_{sw} =0.5 V) . . . 86
4.18 Upper and lower bounds of write voltage, V_W, for V_{th} = 0.5 V, and V_{sw} = 0.5 V . . . 87
4.19 Upper and lower bounds of read voltage, V_R, for V_{th} = 0.5 V and V_{sw} = 0.5 V, and β =0.4 . . . 87
4.20 Lower bound of nominal selector threshold voltage, $\overline{V}_{th}$ as a function of different spread of parameters $\alpha_{V_{sw}}$ and $\alpha_{V_{th}}$ . . . 88
4.21 Lower and upper boundaries for the ratio V_{th}/V_{sw} as a function of the spread parameters $\alpha_{V_{sw}}$ and $\alpha_{V_{th}}$ . . . 89
4.22 Lower and upper boundaries for the ratio V_{th}/V_{sw} as a function of the spread parameters $\alpha_{V_{sw}}$ and $\alpha_{V_{th}}$ for different cases of read and write biasing . . . 90
4.23 Total leakage power as a function of array size and bias scheme (sub-threshold nonlinearity of selector = 0.2 V/decade) . . . 93
4.24 Contour plot of total leakage power as a function of biasing scheme and nonlinearity of selector (leakage power is constant over each indicated contour line) . . . 94

List of Tables

2.1 STT-MRAM Technology Parameters 24

3.1 Summary of results: experimental electrical characterization at 150 μA operating current (embedded memory target) . . . 49

3.2 Summary of results: experimental electrical characterization at 50 μA operating current (storage class memory target) . . . 50

4.1 Example of effective resistivity of metal lines in different technology nodes 69

Chapter 1

Introduction

1.1 General Overview

Inventing better mechanisms for storing information has always been an integral part of the advancement of human civilization. The mechanisms evolved from carvings on stones and clays to markings on papyrus and parchment then to writing on modern paper, and then to the more recent electronic/digital data storage technologies. Today, memory devices that store digital data temporarily or permanently are key components of all electronic systems. Indeed, the demand for electronic storage capacity has been steadily increasing since the invention of the electronic computer and in the future, due to the expected massive number of embedded systems and Internet of Things (IoT) devices, the demand will increase at a much higher rate than ever before. It is estimated that the amount of data that needs to be processed and stored annually will reach 160 billion Terabytes by 2025, which is a 10 fold increase from the 16 billion Terabytes data generated in 2016 [2].

Moreover, since the memory subsystem inside today's electronic systems serves diversified functions, the functionality and performance of electronic systems are increasingly becoming dependent on the performance of the memory subsystems. In today's microprocessors, the memory subsystem has a strongly hierarchical organization, which mostly integrates the mainstream memory technologies namely, Static Random Access Memory (SRAM), Dynamic Random Access Memory (DRAM), and NAND Flash memory that serve as cache memory, main memory and secondary storage, respectively. However, with the evolution of fast microprocessors, there is an increasing performance disparity between the microprocessors and the memory subsystems.

Driven by the growing demand for memory performance and storage capacity, industry is exploring solutions from continuing the conventional scaling to other high-density storage techniques such as multi-level cell (MLC) storage, three-dimensional (3D) integration and crosspoint array architecture. These solutions are required to increase storage capacity and efficiency, thus to achieve high capacity and efficiency at low cost. For instance, in crosspoint arrays, memory cells are built at the junctions of a lower and an upper plane of parallel metal lines running at right angles to each other, and hence if both the width of the metal lines and the spacing between them is equal to the minimum lithographic feature size, F, the memory cell can be allocated within the smallest footprint of $2F \times 2F$. The effective area per cell can be reduced even further with 3D integration. In addition, if the memory arrays are sufficiently large, much of the peripheral circuitry (including address decoders, sense amplifiers, and control circuitry) can be placed underneath the arrays, thereby increasing area efficiency (i.e., reducing the fraction of silicon area associated with the memory array) and this implies lower cost per bit.

As technology scales down, increasing leakage power dissipation and significant degradation of the reliability of SRAM and DRAM are of increasing concern. In the case of NAND Flash, in addition to reliability concerns, its performance is still limited. As a result, emerging memory technologies, such as Resistive Random Access Memory (RRAM), Phase Change Memory (PCM), and Spin-Transfer Torque Magnetic Random Access Memory (STT-MRAM), are actively under research and development due to not only their promising scalability but also their reduced standby power and access speed. Unlike SRAM, DRAM and NAND Flash, which are charge-based, the most attractive emerging memories, including the aforementioned ones, operate based on resistance switching principle: the basic storage device is switched between a high-resistance state (HRS) and a low-resistance state (LRS) to store bits. The transition between the two states is triggered by applying electrical voltage/current pulses. These emerging memories are expected to be used mainly as a Storage class memory (SCM), a new class of memory with a storage capacity and cost per bit similar to NAND Flash but with a speed that approaches DRAM. Due to this features, SCM helps to address the aforementioned performance disparity between microprocessors and memory.

In particular, STT-MRAM, the basic storage element of which is a Magnetic Tunnel Junction (MTJ), has demonstrated a high write speed and a practically infinite cyclic endurance, which makes it a potential candidate not only as a storage class memory but also as DRAM and SRAM replacement. However, STT-MRAM faces some challenges. One of the crucial challenges is that the sense margin (i.e., the voltage/current margin available for sensing by the sense amplifier to differentiate the LRS and the HRS states) is very small due to the intrinsically small ratio of the resistance value of the two states, which is usually worsened by process variations in the STT-MRAM cells and parasitics of interconnection metal lines. Resistive random access memory (RRAM), with a typical metal-oxide-metal device structure, has also gained great attention with the prospects of obtaining a replacement for both embedded memory and mass storage applications, which have their own requirements. For example, low operating voltage (for compatibility with the core CMOS technology) and high cyclic endurance are required embedded memory whereas acceptable power consumption and high cell density are required issue in mass storage applications.

1.2 The Memory Sub-System

Semiconductor memory chips consist of millions to billions of memory cells organized in array structure. In particular, in Random Access Memory (RAM), individual cells can be accessed by activating the appropriate wordlines and bitlines as shown in Figure 1.1a. The binary address inputs are decoded by row and column decoders to locate the corresponding target cells, so that read and write operations can be performed [3]. Furthermore, large memory arrays are organized as sub-array blocks of as shown in Figure 1.1b.

1.2.1 Memory Hierarchy

In today's microprocessors, a hierarchical memory organization, which integrates different memory technologies, is used to achieve optimal overall performance, area, and cost [5, 6]. Accordingly, SRAM, which is the fastest one, is used in cache memories [7, 8]. DRAM comes next in the hierarchy as a main memory due to its higher cell density [5, 6, 8]. Flash memory features low cost per bit and non-volatility, which have made it the technology of choice for secondary mass storage [6, 8]. Figure 1.2 shows the physical distribution of memory in computer, starting from the fastest registers and L1 cache to the slowest tertiary storage. From left to right, speed decreases while storage capacity increases.

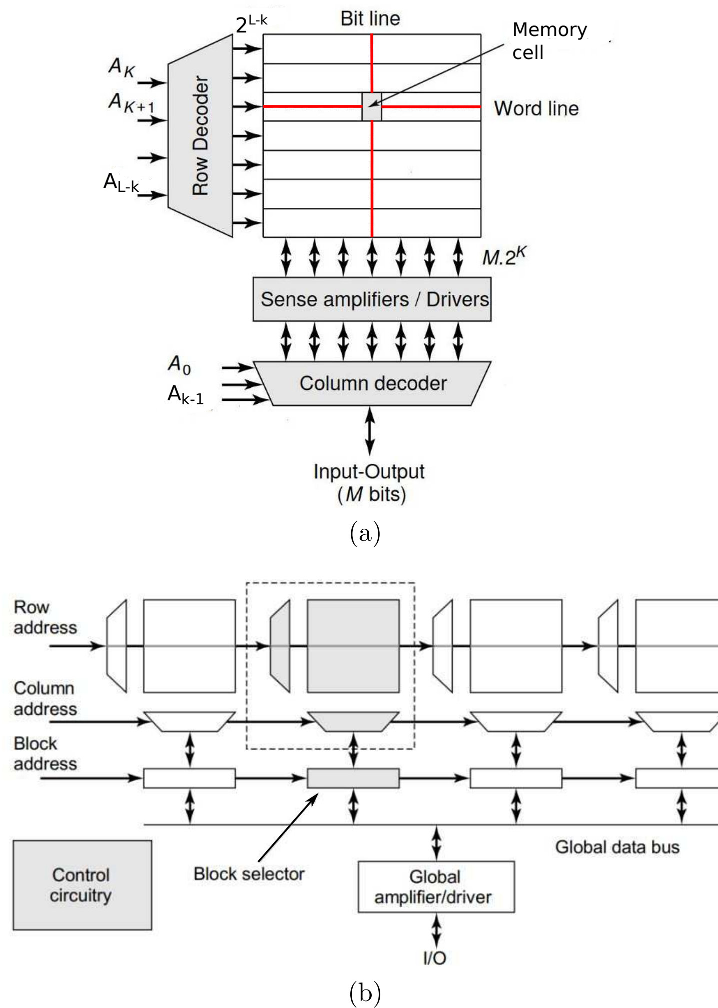


Figure 1.1: Memory array organization (source: [4])

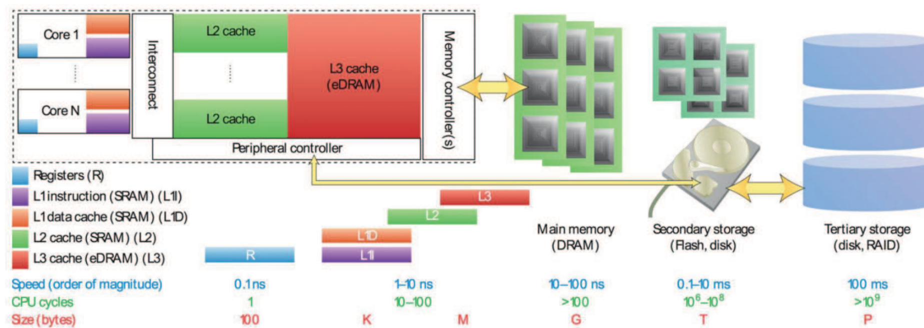


Figure 1.2: Physical distribution and hierarchy of memory in a computer (source: [9]). The lower panel shows speed (as order of magnitude), number of processor cycles needed for accessing the memory, and storage capacity in bytes (as order of magnitude).

1.3 Mainstream Solid-State Memories

1.3.1 Static Random Access Memory (SRAM)

Static Random Access Memory uses a bistable latch circuit to hold the stored bit, as shown in Figure 1.4. The latch circuit is formed by the two cross-coupled inverters and the two transistors (M 5 and M 6) connected with the wordline and two bitlines, are used as access transistors to select target cells. Here, Q represents the data stored in the SRAM, which could be either bit '0' or bit '1' and $\overline{Q}$ is the complementary data. The term 'static' is derived from the fact that the stored data can be retained with no need for refreshing as long as power is being supplied. In the most common CMOS SRAM cell, each of the two inverters are implemented with complementary NMOS and PMOS transistors, hence resulting in the 6-transistor SRAM cell structure shown in Figure 1.4 [3].

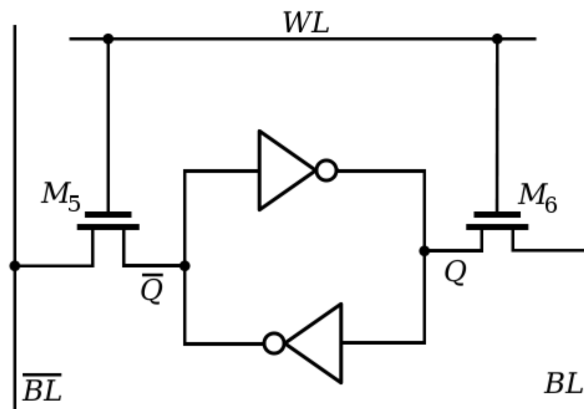


Figure 1.3: Static RAM operation

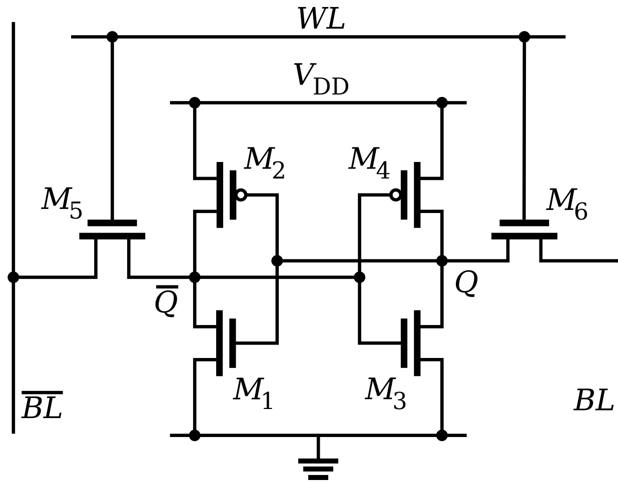


Figure 1.4: 6-Transistor SRAM cell

1.3.2 Dynamic Random Access Memory (DRAM)

The basic storage element of a Dynamic Random Access Memory (DRAM) is a capacitor. The data (bits '0' and '1') is represented by whether the capacitor is fully charged or discharged. A DRAM cell with a select transistor is shown in Figure 1.5.

However, the electrical charge on the capacitor will gradually leak away, and after a period of time, the voltage on the capacitor will be too low to differentiate between '0' and '1'. As a result, all DRAM cells need to be read out and written back periodically (known as a refresh) to ensure data integrity [3].

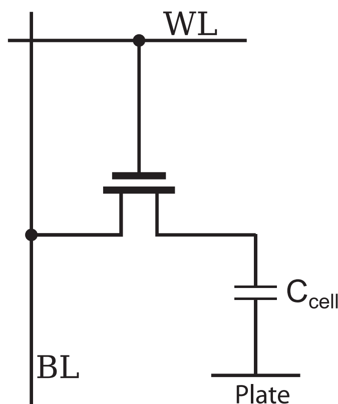


Figure 1.5: 1-Transistor DRAM cell

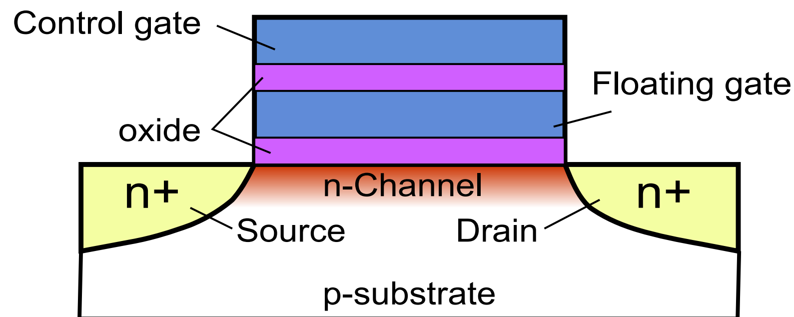


Figure 1.6: Floating gate flash memory cell

1.3.3 Flash Nonvolatile Memory

Flash memory is the most widely used nonvolatile memory technology today. The key device in this prevailing memory is a floating gate transistors whose cross section is shown in Figure 1.6. Unlike a MOSFET transistor, an additional floating gate (FG) is added between the control gate (CG) and channel. The data is encoded based on the presence or absence of electrons trapped in the FG and the data is retained without power. Since, it is isolated by oxide layers, the floating gate is able to trap charges and keep them for years, giving a non-volatility of Flash memory.

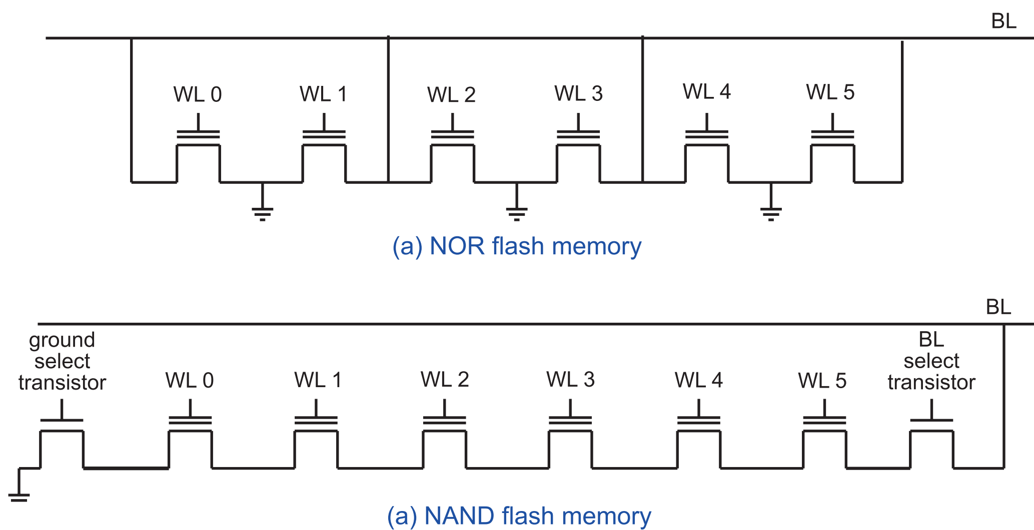


Figure 1.7: Flash memory: (a) NOR (b) NAND

There are two common layouts for flash memory namely, NAND flash memory with Floating gate transistors in series and NOR flash memory with floating gate transistors in parallel, as shown in Figure 1.7. The names NAND and NOR are derived from the fact that their connection fashion in series or parallel resembles a NAND gate or NOR gate. Since NAND Flash has one ground connection and two select transistors for each row, it provides with higher storage density than NOR Flash. Hence, it is widely used in external storage. Whereas, NOR Flash has lower latency and thus is widely used in embedded systems, where high performance is required.

1.4 Challenges of Mainstream Memories and Current Research and Development Trends

The direction of research and development on memory device technologies and architectures is shaped by some key technology trends. The first trend is the increasing difficulty of scaling down of the mainstream memory technologies [6, 7, 10, 11]. Scaling is needed for increasing storage capacity and efficiency; lack of it will make it difficult to achieve high capacity at low cost [11]. In Flash memory, scaling down the storage-cell tunneling area results in the reduction of the number of electrons that are injected to the floating gate during programming. These electrons determine the threshold voltage that is used to differentiate bit '1' and bit '0'. However, due to the reduction of the number of electrons due to scale down, a slight variation in the number of these electrons or the loss of some of the stored electrons over time may produce significant threshold voltage variations, thus lead to read errors and reliability concerns [10]. In DRAM, the unavoidable scaling of storage capacitance to scale down the overall memory cell size leads to poor data retention time and insufficient read margin [12]. SRAM also faces its own challenges when its constituent transistors are scaled down [13]. As technology scales down, it becomes more difficult to control variations when fabricating minimum-sized transistors. As a result, the quality and the reliability of SRAM's constituent transistors and, hence, SRAM cell reliability, are degraded [13]. This motivated the emergence of the aforementioned emerging memory technologies.

The second technology trend is the increasing performance disparity between processor and memory. As discussed in Section 1.2.1, the conventional approach, a strongly hierarchical memory organization, which mostly integrates SRAM, DRAM, and Flash memory, is used to achieve optimal overall performance, area, and cost [5, 6]. However, with the evolution of fast microprocessors, fully exploiting their computational power with such memory hierarchy has become quite challenging. Hence, there are efforts to flatten the memory hierarchy [11, 12]. On technology side, the ideal solution would be using a single "universal" memory device that satisfies all the ideal characteristics: fast read/write speed, low cost per bit, low power consumption, nonvolatility, and so on. Although, it is almost impossible to get a "universal" memory device, some of the emerging memories have been pursued toward achieving part of the aforementioned ideal characteristics [14].

The third key trend is the steadily increasing demand for large data storage capacity [12], as also mentioned earlier. In fact, the amount of data to be processed is steadily increasing over time at a rate higher than that of Moore's law [12]. To satisfy this demand, industry is exploring, other options in addition to the conventional scaling down, such as multi-level cell (MLC) storage and three-dimensional (3D) integration to improve memory density [12], and hence, to reduce the cost per bit. For example, Figure 1.8 shows a 32-layer 3D NAND Flash memory for high-density storage. Crosspoint array has also become an attractive architecture to achieve high-density memory design [15, 16]; benefiting from the recent development of the partially-mentioned emerging memory technologies that are well-suited for the architecture due to their two-terminal device structure.

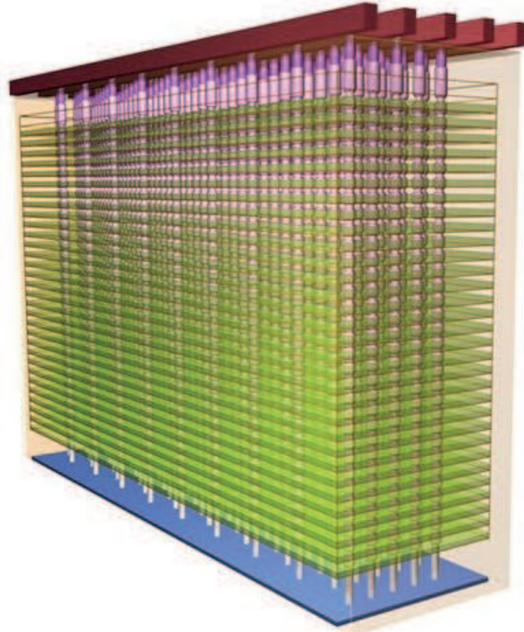


Figure 1.8: 32-layer 3D NAND Flash memory (source: Micron website, accessed on December 15, 2018)

1.5 Objectives of the Thesis

With the aim of contributing to the research and development on emerging memories, this PhD thesis presents a model-based study of Spin-Transfer Torque Magnetic Random Access Memory (STT-MRAM), an experimental electrical characterization of two Resistive Random Access Memory (RRAM) device technologies, and a comprehensive analysis of design and technology considerations and requirements for high-density crosspoint memory array architecture.

In the first part, the thesis presents a behavioral model of STT-MRAM cell that can be used for circuit simulations, review of different sensing circuit schemes applicable for STT-MRAM and a variability-aware analysis and design guideline of slope detection self-reference scheme, which is deemed to outperform other STT-MRAM sensing schemes reported in the literature. The performance of this sensing scheme is analyzed by taking into account the impact of cell-to-cell variations and parasitic resistance in bitlines, BLs, and wordlines, WLs.

The second study focuses on the electrical characterization of RRAM memory technologies, namely Oxide Resistive RAM (OxRAM) and Conductive Bridging Resistive RAM (CBRAM). For the case of OxRAM, the purpose of this study is to establish the performance of the memory array according to oxide thickness, density and size of devices. As for the CBRAM, the study focuses on reliability (endurance and retention) performance of the memory devices under different programming pulse amplitude and width, and at the same time, optimization of the pulses for better reliability.

The last study focuses on analysis of crosspoint memory array design based on 1S1R (one selector coupled to one resistive memory). Despite promise of crosspoint memory arrays for high-density storage, their implementation is challenged by some critical issues such as, sneak-path (leakage) currents and ohmic drop due to parasitic resistance in the interconnection metal lines, which will be elaborated in Chapter 4 (Section 4.2). Hence, the thesis presents a comprehensive analysis of crosspoint arrays by taking into account selector device and memory element operating characteristics, the impact of the BL and WL parasitic elements, the size of the memory array and the biasing scheme. The analysis is aimed at contributing to the understanding of crosspoint array implementation constraints. It also gives a guideline for circuit design and for memory/selector device technology optimization.

1.6 Organization of the Thesis

The thesis is structured as the following chapters. In Chapter 2, the study on STT-MRAM is discussed. After providing an overview on the application targets, key technology advances and challenges of STT-MRAM, the basic physical principle of operation of STT-MRAM is discussed with reasonable detail. Then, the developed behavioral model of STT-MRAM device is discussed and validated by circuit simulations. The next part of this chapter focuses on reading in STT-MRAM, which appears to be the fundamental challenge of this technology. In this regard, different circuit schemes are reviewed followed by analysis of self-referenced slope detection sensing architecture, which is deemed to give better read performance compared to the other sensing techniques proposed in the literature. A study of the impact of variability on the chosen self-reference slope detection sense amplifier architecture is then provided. A theoretical study, validated by simulation, proposes an original variation-aware optimization of the reading margin of this sensing technique. The chapter is concluded by providing an overall summary of the presented results.

Chapter 3 presents, consecutively, a detailed experimental electrical characterization of Oxide RAM with TiN/Hf/GdAlO/TiN device stack (or GdAlO-based OxRAM) and a reliability study of Conductive Bridging RAM with Cu/TiW/SrTiOx/WOx/W stack (or STO-based CBRAM). In the first part of the chapter, a brief discussion of the underlying operating principles of RRAM is presented followed by explanation of application targets for RRAM and the corresponding requirements for these applications. Then, the experimental setup for the electrical characterization of the GdAlO-based OxRAM is described and the obtained various results are analyzed in detail. The presented analysis focuses on the impacts of the thickness of the GdAlO layer, the size of the memory device, and memory cell density on array-level operating voltage/current performance and reliability. Similarly, for the CBRAM, the experimental setup used for the characterization is described and endurance and retention performance under different programming pulse amplitude and width conditions is presented. Analyzing the performances obtained on STO-based CBRAM devices, using optimized pulses completes this study. Finally, a conclusion is drawn from the experimental results and a recommendation for future work is made.

Chapter 4 is dedicated to the discussion of the study on crosspoint array design based on 1S1R (one selector coupled to one resistive memory). In the first part of this chapter, an introductory overview of crosspoint arrays and a review of research trends is provided, followed by a discussion of the challenges faced by crosspoint arrays. Then, a brief survey of selector together with their non-linearity feature is first reported, followed by the description of selector modeling. Resistive elements as well as interconnection metal line considerations are also discussed. A simplified array model for the worst-case scenario, which enables to reduce computation time to study large crosspoint arrays, is also presented. Then, a first analysis, which gives the boundary conditions of a proper write operation neglecting sneak-path as well as IR drop issues, is presented. The next part is dedicated to the evaluation of the array size considering all the design constraints (selector sensitivity, selector voltage margin, selector threshold, IR drop and voltage/current switching characteristic of the resistive element), which draws very interesting design guideline for crosspoint memory arrays. The analysis is extended for read operation evaluating the read margin versus the array size and the resistance ratio. The next part of this chapter is devoted to the analysis of the variability of the memory and of the selector element regarding voltage compatibility, which allows determining acceptable threshold voltage/switching voltage ratio for proper read and write operations considering variability. Then, a generic "x biasing" scheme is introduced in order to minimize leakage and is compared to conventional biasing schemes considering. Finally, some conclusions are drawn based on the results of the analyses presented in the chapter.

Finally, in Chapter 5, a general conclusion is drawn from the results presented in the preceding three chapters and an outlook for future work is provided.

Chapter 2

Spin-Transfer Torque Magnetic RAM (STT-MRAM)

2.1 Overview

Spin-Transfer Torque Magnetic Random Access Memory (STT-MRAM) has gained much attention due to its very desirable properties such as storage class nonvolatility, low standby power, area (and) current scalability, high write speed, and practically infinite cyclic endurance, which can be incorporated to all levels of the memory hierarchy [6, 17, 18]. In comparison with the competitor emerging memory technologies, STT-MRAM uniquely features a high write speed and a practically infinite endurance that make it a very good candidate not only as storage class memory but also as DRAM and SRAM replacement [6, 7, 19, 20]. The speed and the endurance of STT-MRAM are comparable to that of DRAM, while STT-MRAM also provides with some additional benefits. Most notably, STT-MRAM is a nonvolatile technology and, hence, it does not require refresh unlike DRAM, which gives energy and reliability advantages [18]. Indeed, Everspin's 256 Mb STT-MRAM DDR3 (Double Data Rate) is already in mass production replacing Synchronous DRAM (SDRAM). However, as of today, the cost of production of STT-MRAM is higher than that of DRAM, which is already a mature technology with huge production volumes and, thus, the adoption of STT-MRAM as DRAM replacement is only justified in cases where the need for the performance advantages outweigh the additional cost. As for SRAM replacement, STT-MRAM has advantage of nonvolatility and higher density, while maintaining write/read speed and endurance comparable to that of SRAM [7, 18].

In fact, it has been demonstrated in the literature that STT-MRAM can give a better performance (speed) than that of SRAM in relatively large memory array sizes like in the case of L2 and L3 caches [21, 22]. In such large-size memory arrays, the overall access (read/write) time is dominated by interconnection line delay. Due to the smaller size of STT-MRAM memory cell, the same array storage capacity can be implemented with a smaller silicon area, which results in a shorter global interconnect delay and hence, better cache performance (speed). However, in small array sizes like L1 cache at the core level, SRAM gives a better read/write speed [21] than that of STT-MRAM. Yet, the research and development efforts on STT-MRAM technology are continuing. In addition, it has been proposed an alternative magnetic memory concept, namely Spin-Orbit Torque MRM (SOT-MRM), that features sub-nanosecond switching speed making it a good candidate for L1 cache memory [23, 24].

2.2 Key Technology Advances and Challenges of STT-MRAM

The Magnetoresistance (MR) effect, which is the change of the electrical resistance of materials by an externally-applied magnetic field, has been known since 1856 [25]. However, it was the discovery of the Giant Magnetoresistance (GMR) effect, which is observed in structures made up of two ferromagnetic layers sandwiching a non-magnetic metallic spacer layer (or in multi-layers of such sandwich structure), that opened the way to a more efficient control of electron charge transport through magnetization [20, 26]. When the magnetization, which measures the local orientation of electron spins, of the two ferromagnetic layers are aligned to the same direction (parallel orientation), a relatively low resistance, R_P, is observed. On the contrary, when the magnetizations of the two layers are oriented in antiparallel, a relatively high resistance, R_{AP}, is observed. As a crucial advancement, a much higher Tunnel Magnetoresistance (TMR) ratio, a performance index defined as the relative difference of resistance between the antiparallel and parallel states, i.e. TMR ratio $= [(R_{AP} - R_P)/R_P] \cdot 100\%$, was obtained by replacing the metal spacer in the GMR structure with a thin non-magnetic (insulating) tunnel layer, hence obtaining a structure referred to as a Magnetic Tunnel Junction (MTJ). In this regard, the first room temperature MTJ with TMR ratio of 11.8% was demonstrated by Moodera et al. [27]. Further improvements of TMR ratio steadily continued. For instance, a TMR ratio of 200% was reached by using a single-crystal MgO tunnel barrier [28].

As shown in Figure 2.1, MTJ is the basic storage element in the STT-MRAM memory cell. One of the ferromagnetic layers is magnetically pinned during manufacturing while the remaining ferromagnetic layer is free to be switched into parallel or anti-parallel direction with respect to the fixed layer magnetization [29].

The other key technology advance was the discovery of spin-transfer torque effect as a way of switching the magnetization of nanomagnets. The spin-transfer torque (STT) effect was for the first time theoretically predicted by Slonczewski [30] and by Berger [31] in 1996 [29]. In the classic MRAM that preceded STT-MRAM, the memory cell lies between a wordline and a bitline arranged at right angles to one another, one below and the other above the cell and when a sufficiently high current passes through the wordline or the bitline, an induced magnetic field is produced at the intersection of the lines, which imparts torque that enables the switching of the free layer. However, since a magnetic field is difficult to localize, this technique is energy inefficient [29]. Using spin-transfer torque effect enabled a more efficient way of switching nanomagnets.

However, there are still efforts to reduce the switching current further as a high write current not only leads to energy inefficient write operations but also constrains the scalability of the driving transistors. In this respect, a perpendicular MTJ (as opposed to the in-plane MTJ shown in Figure 2.1) was introduced in 2010 [32]. Besides, more efficient write techniques such as voltage-controlled magnetic switching (VCMA), are being investigated to improve the write performance of the memory cells [33].

Insufficient sense margin is one of the critical challenges of STT-MRAM [1, 33]. This is mainly attributed to the intrinsically small TMR ratio [1]. Apart from the intrinsically small TMR ratio, process variations in the MTJ and the selector device of STT-MRAM cells within an array (cell-to-cell variations) put another key challenge on the design of the sensing circuit scheme [34]. The important sources of the cell-to-cell variations are the thickness of the tunnel oxide barrier (which is as thin as 1–2 nm) and the geometric size (surface area) of the MTJ device. Indeed, due to quantum mechanical tunneling, the resistance of an MTJ has an exponential relationship to the thickness of the oxide barrier between the two magnetic layers [34,35]. It is also worth to point out that the variation of the MTJ resistance will be aggravated by the further reduction of the oxide barrier thickness and the large MTJ geometry variation in scaled technologies [36].

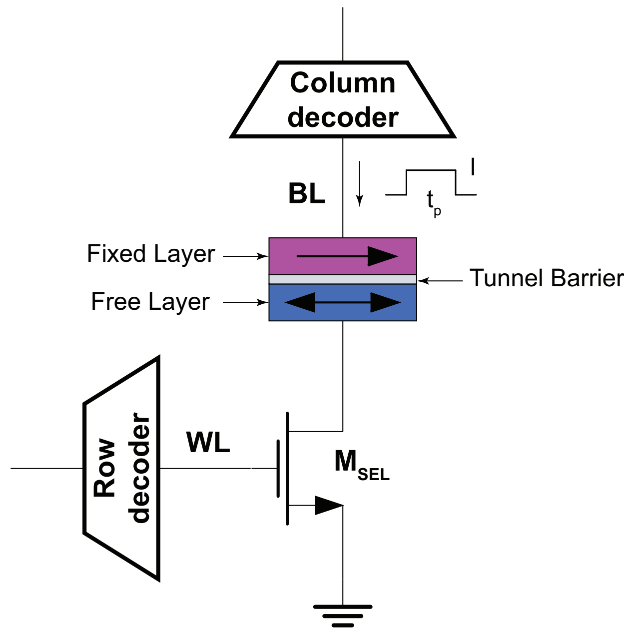


Figure 2.1: STT-MRAM cell with a MOS transistor as a selector

2.3 Basic Physics of STT-MRAM

2.3.1 Principle of Operation

Figure 2.1 shows STT-MRAM cell with a MOS transistor selector. The binary digits are stored as a low- or a high-resistance state of the MTJ, which is controlled by switching the magnetization of the free layer to parallel or anti-parallel direction, respectively, with respect to the fixed layer magnetization. The parallel configuration results in a low MTJ resistance state (assigned to bit '0'), whereas the anti-parallel configuration yields a high MTJ resistance state (assigned to bit '1'). The switching of the state of the STT-MRAM cell is achieved by using the spin-transfer torque effect, which is briefly discussed here referring to Figure 2.2.

Let us assume that electrons flow from the fixed (reference) layer to the free layer (equivalently, current flows from free to fixed layer) and assume that the two layers were initially magnetized along opposite directions as in Figure 2.2a. The electrons become spin polarized when they pass through the reference layer. As a result, the electrons coming out of the reference layer mainly hold a spin direction parallel to the magnetization of the reference layer. Then, the spin-polarized electrons tunnel through the thin insulating layer (maintaining their polarization) and reach the free layer, where their average spin is quickly re-aligned to the magnetization of the free layer. In the process, the electrons lose a spin angular momentum and, due to the conservation of total angular momentum, the lost angular momentum is transferred to the magnetization of the free layer. This results in a torque tending to align the magnetization of the free layer towards the spin momentum of the incoming electrons and, hence, towards the magnetization of the fixed layer [37, 38]. The magnetization of the free layer will be switched if the amount of the spin-polarized electrons exceeds a given threshold value.

In the opposite case, i.e., parallel-to-antiparallel switching (see Figure 2.2b), current should flow from the reference layer to the free layer (i.e., electrons flow from the free to fixed layer). The electrons become spin-polarized after going through the free layer, hence the majority of them will have a spin direction parallel to the magnetization of the free layer. These electrons will pass through the reference layer since their spin direction is also parallel to the magnetization of the reference layer.

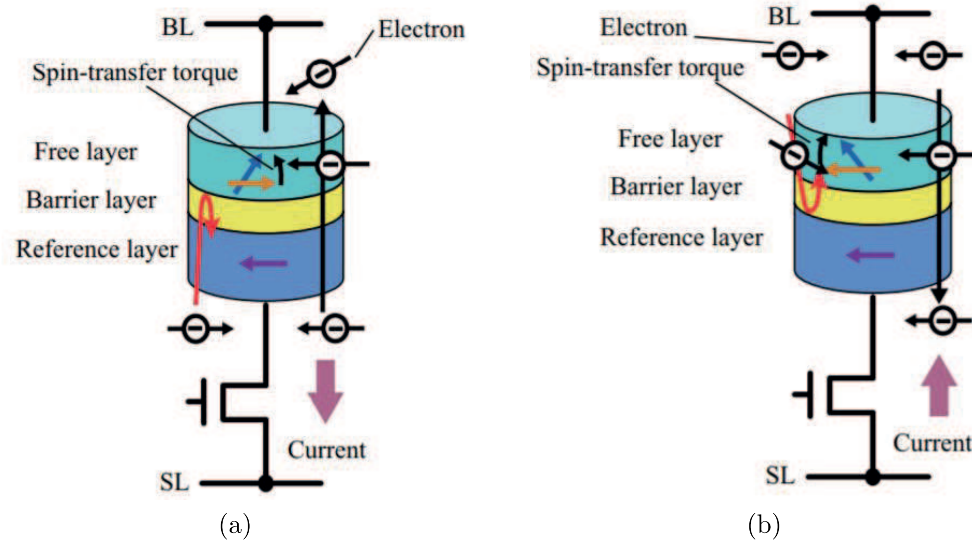


Figure 2.2: STT switching (a) antiparallel to parallel switching (b) parallel to antiparallel switching

However, those electrons with a spin direction antiparallel to the magnetization of the reference layer will be reflected at the interface between the junction and the reference layer. This results in a spin-polarized current injecting into the free layer and exerting spin-transfer torque on the magnetization of the free layer. Similar to the complementary case, the magnetization of the free layer will only switch when the amount of the spin-polarized electrons exceeds a given threshold value and an antiparallel state is obtained [37, 38].

For reading the stored data, one can force a small current (namely, a current smaller than the switching threshold current) through the cell and sensing the resultant voltage across the cell, which is then compared to a reference voltage to determine whether bit '1' or bit '0' was stored.

2.3.2 Switching Current

The intrinsic critical current (I_{c0}), proposed by Slonczewski [39], is used as a figure of merit in macro-spin models of current driven magnetization switching in nano-magnetic devices like STT-MTJ. It is defined as the minimum current which is able to cause a spin flip in the absence of any external magnetic field at absolute zero temperature.

For an in-plane MTJ, I_{c0} can be expressed in terms of the magnetic properties of the MTJ [40] as given by (2.1)

$$I_{c0} = \frac{2e\alpha M_S V \cdot (H_K + 2\pi M_S)}{\hbar\eta} 2x \tag{2.1}$$

where e is the absolute value of electron charge, α is the damping factor, M_S is saturation magnetization, V is the volume of the free layer, H_K is the effective anisotropy field, $\hbar$ is reduced Planck's constant, and η is the spin torque efficiency factor, which is a function of the current polarity, material polarization and the relative angle between the magnetization in the free and in the fixed layer [41].

Due to the inherent torque asymmetry in the MTJ cell, the critical currents required for switching from the parallel P to the anti-parallel AP state ($P \to AP$) and from the antiparallel to the parallel state ($AP \to P$) are different ; more specifically, the critical switching current for ($P \to AP$) switching being significantly larger. It is apparent from equation (2.1) that the switching current is directly proportional to the volume of the free layer, which means that I_{c0} decreases when scaling down the size of the MTJ cell. However, scaling down the size of the MTJ cell generally requires a trade-off between switching current and thermal stability factor, Δ, which is given by:

$$\Delta = \frac{H_k M_s V}{2K_b T} \tag{2.2}$$

where K_b is Boltzmann constant and T is (absolute) operating temperature. One wants to minimize I_{c0} while still preserving a reasonable thermal stability factor for non-volatility. In this respect, the intrinsic switching current to thermal stability ratio, $\frac{I_{c0}}{\Delta}$ given by (2.3) can be used as a more useful figure of merit:

$$\frac{I_{c0}}{\Delta} = \frac{4e\alpha K_b T \left(1 + \frac{2\pi M_s}{H_k}\right)}{\hbar\eta} \tag{2.3}$$

This equation was derived directly from (2.1) and (2.2).

As can be seen from equation (2.3) equation, by means of materials engineering, the critical current can be reduced without sacrificing thermal stability by using ferromagnetic materials with low magnetization M_s and/or high spin transfer efficiency η and/or small damping factor α [40]. Another promising method to decrease the switching current is to decrease the demagnetizing field by introducing perpendicular anisotropy in the free layer [40]. In a perpendicular MTJ, the magnetization direction is perpendicular to the plane of the ferromagnetic layers, which cancels out the effect of the demagnetization field, thus reducing the value of I_{c0} [40].

In long current pulses, switching can occur even with a current pulse amplitude lower than the critical current (i.e. $I < I_{c0}$) due to thermal fluctuations. In the presence of spin torque and long current pulses, when thermal activation plays a major role, the switching time (or, in other words the pulse width required for magnetization switching) can be calculated by applying [40, 42]

$$\tau_1 = \tau_0 \cdot e^{\left[\Delta\left(1-\frac{I}{I_{c0}}\right)\right]} \tag{2.4}$$

where, τ_0 (about 1 ns for storage class memory purposes) is the minimum time required to reverse the magnetization, referred to as thermal attempt time.The critical switching current (I_c) associated with a pulse width t_p can be obtained as [43]:

$$I_c = I_{c0} \cdot \left[1 - \frac{1}{\Delta} ln\left(\frac{t_p}{\tau_0}\right)\right] \tag{2.5}$$

When the current pulse has an amplitude I higher than the critical one, switching is dominated by the spin-transfer torque effect and the thermal random effect becomes less important. Consequently, the current pulse required for switching can be narrower. The switching time can be estimated by using [44]

$$\tau_2 = \frac{1}{\alpha\mu_0\gamma M_S} \cdot \frac{I_{c0}}{I - I_{c0}} \cdot ln\left(\frac{\pi}{2\theta_0}\right) \tag{2.6}$$

where γ is the gyromagnetic ratio, μ_0 is the permeability constant), and θ_0 is the root square average of the initial angle of the free layer magnetization (determined by thermal fluctuation).

2.3.3 Probability of Switching

Spin-transfer torque (STT)-based switching is intrinsically stochastic, and the probability of switching of state is generally a function of the amplitude and the width the current pulse passing through the MTJ device, I and t_p, respectively, and technology dependent parameters I_{c0} and Δ. For sufficiently long current pulses ($t_p \geq 10$ ns) with amplitude $I < I_{c0}$, the probability of switching can be expressed as: [40, 42, 43]:

$$P_{sw} = 1 - \exp\left\{\frac{-t_p}{\tau_0}\exp\left[-\Delta\left(1 - \frac{I}{I_{c0}}\right)\right]\right\} \tag{2.7}$$

Since intrinsic critical current is different for P→AP and AP→P switching, the probability of parallel to anti-parallel switching (denoted as $P_{sw,P\rightarrow AP}$) and the probability of anti-parallel to parallel switching (denoted as $P_{sw,AP\rightarrow P}$) are also different even with the same current pulse amplitude and width.

2.3.4 Static Behavior of STT-MRAM Cell

The static behavior of the MTJ in STT-MRAM cell can be represented by the resistance values of the anti-parallel and parallel states that are related to each other by the TMR ratio given by:

$$TMR\,ratio = \left(\frac{R_{AP} - R_P}{R_P}\right) \cdot 100\% \tag{2.8}$$

It is a measure of the distinguishability of the high- and the low- resistance states.

Besides, it has been demonstrated in the literature that the resistance values R_{AP} and R_P are dependent on the current/voltage bias of the MTJ [42, 44, 45]. In both the low- and the high- resistance state, the resistance value is maximum at zero current and starts to roll off when current flows through the MTJ. This characteristic can be modeled by using the following equations:

$$R_{AP} = R_{AP0} - S_{AP} \cdot I \tag{2.9a}$$

$$R_P = R_{P0} - S_P \cdot I \tag{2.9b}$$

where R_{AP0} and R_{P0} are the values of R_{AP} and R_P respectively, at zero current and S_{AP} and S_P are the corresponding slopes (or curve fitting parameters) for the resistance roll-off, which can be determined from experimental data. Generally, S_{AP} is larger than S_P.

2.4 Model of STT-MRAM

In this thesis work, a behavioral model of STT-MRAM cell was developed using Verilog-A language. This model is described by the block diagram shown in Figure 2.3 and was used to perform analyses using the device technology parameters given in TABLE 2.1. The storage element (MTJ) was modeled as a variable resistor whose resistance value is controlled by the state of the cell and the amplitude of the current pulse. The bias dependence of the resistance values is also considered according to (2.9a) and (2.9b).

The 'state decision' block controls the state of the cell. To take into account the intrinsically stochastic nature of STT-MTJ dynamic behavior, the block was implemented as a Bernoulli random binary number generator, in combination with the probability of switching function given by (2.7).

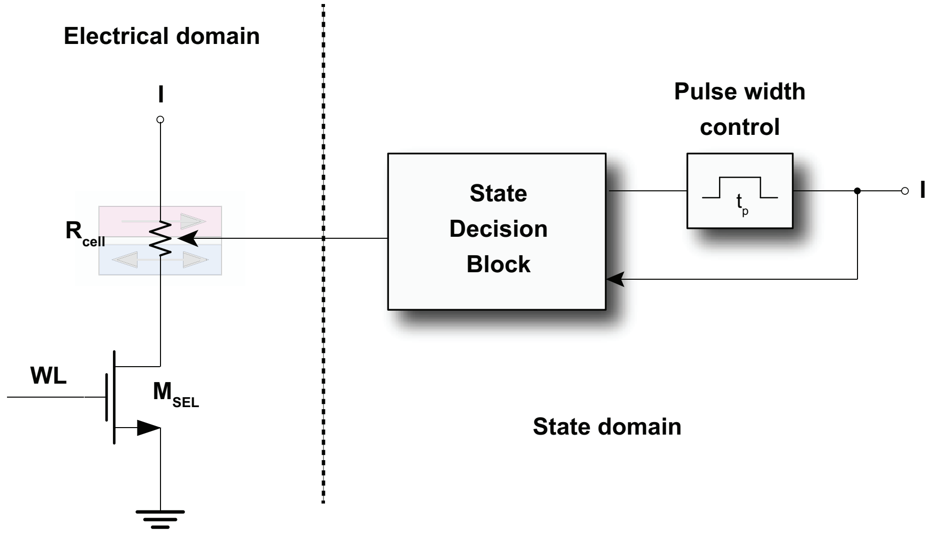


Figure 2.3: Block diagram of the proposed STT-MRAM Verilog-A model

If the previous state was bit '0' (parallel state), the state is retained if the write current is positive. In contrast, if the current is negative, state changes to bit '1' with a probability equal to $P_{sw,P \rightarrow AP}$ or remains unchanged with a probability equal to $1 - P_{sw,P \rightarrow AP}$. For the complementary case, if the previous state was bit '1', the bit '1' state is retained if the write current is negative, whereas if the current is positive, the state changes to bit '0' with a probability equal to $P_{sw,P \rightarrow AP}$ or remains bit '1' with a probability equal to 1-$P_{sw,P \rightarrow AP}$.

The STT-MRAM cell model is used in Cadence Virtuso platform to carry out simulations to demonstrate characteristics of the STT-MRAM cell. The simulated R-I characteristic is shown in Figure 2.4, with the resistance value of high-resistance AP state shown in blue color and the resistance value low-resistance P state shown in red color. In both states, the resistance value decreases when the amplitude of current pulse increases, with the high-resistance in the AP state dropping faster than the low-resistance in the P state, which results in decrease of TMR ratio. As expected, switching from one state to another state occurs when the programming current is sufficiently high (approximately equal to I_{c0}), as can be seen from the transitions. The dashed line corresponds to the case of a cell with an asymmetrical R-I characteristic i.e., AP→P and P→AP transitions occur at different currents. The switching currents in both transitions are normalized to $I_{C0,P \rightarrow AP}$. The results shown are consistent with the physical properties of the MTJ device and with results available in the literature [17, 20].

Table 2.1: STT-MRAM Technology Parameters

	Mean value	Variations (σ)
Resistance-area (RA) product corresponding to parallel state at low field (100 mV)	$6\ \Omega \cdot \mu m^2$	7.5%
Ratio of RA (corresponding to parallel state) at high field (400 mV) to RA at low field (100 mV)	0.8	
MTJ device diameter	20 nm	7.5%
Parallel state resistance (R_P) at zero bias	20 kΩ	7.5%
Tunnel Magnetoresistance Ratio (TMR) ratio at low field (100 mV)	100%	4.5%
Parallel state resistance roll off: R_P at 600 mV	16 kΩ	
Anti-parallel state resistance roll off: R_{AP} at 600 mV	22.4 kΩ	
Thermal stability factor Δ	60	7.5%
Critical current density for P to AP switching, $J_{c0,P\to AP}$	$4\cdot 10^6 A/cm^2$	
Critical current density for AP to P switching, $J_{c0,AP\to P}$	$2\cdot 10^6 A/cm^2$	
Write pulse width (t_p)	10 ns	

In Figure 2.5, the switching probability was simulated against normalized cell current for different values of thermal stability factor Δ. For current significantly lower than I_{c0}, the probability that the state of the cell changes is higher for lower values of Δ, whereas the probability is very low for higher values of Δ. According to the definition of Δ within equation (2.2), lower values of Δ correspond to high operating temperature or/and low volume of the free ferromagnetic layer. In all cases, the switching probability is close to unity when the current is approximately equal to I_{c0} and higher than this value. The inset of Figure 2.5 shows the variation of switching current with temperature for different values of pulse width t_p and Δ (i.e. assuming Δ varies due to factors other than temperature). Figure 2.6 provides a transient simulation which demonstrates that the state transitions of the cell from '1' state to '0' state (and vice versa) takes place when the normalized programing current is roughly equal to 1 (-1).

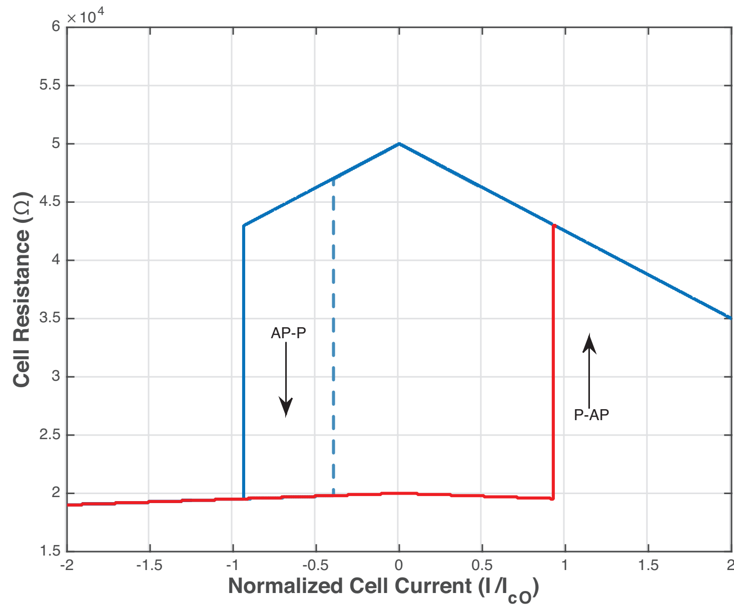


Figure 2.4: R-I characteristic: asymmetrical (dashed line) switching currents for the AP→P and P→AP transitions; symmetrical switching (solid line)

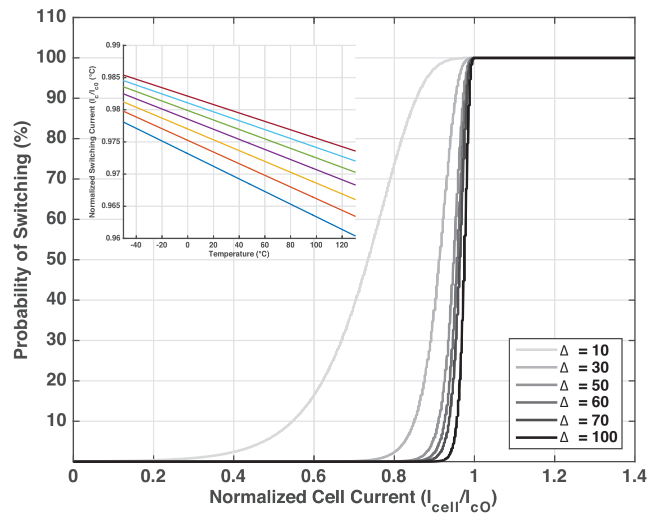


Figure 2.5: Switching probability for different values of the thermal stability factor Δ. The inset shows variation of switching current with temperature

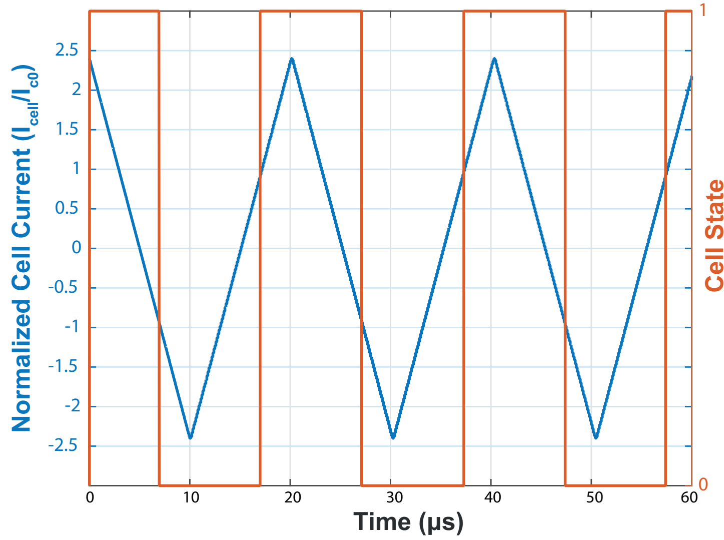


Figure 2.6: Simulated switching characteristic of STT-MRAM cell: normalized write current (blue, left y-axis) and the state of the cell (right y-axis)

2.5 Sensing Circuits for STT-MRAM

Improving the TMR ratio of STT-MRAM requires technology advancements. At the same time, circuit techniques that are robust to cell-to-cell variations in the array are also needed to be able to sense the state of the cell. In this chapter, the sensing margin of these circuit schemes were analyzed and compared. Subsection 2.5.5 is devoted to a detailed analysis and optimization of the sensing scheme with the best performance (namely, the slope detection self-reference sensing scheme) as applied to STT-MRAM arrays.

Figures 2.7a and 2.7b show an example of read operation in conventional and crosspoint STT-MRAM memory arrays carried out by applying a read current to the bitline. Crosspoint memory arrays will not be discussed here as an entire chapter (Chapter 4) is devoted to such discussion. Without going to details, in a crosspoint STT-MRAM array (Figure 2.7b), each STT-MRAM cell along with a series-connected two-terminal, non-linear selector device is placed at the intersection of a bitline and a wordline. The selected STT-MRAM cell (lower right-corner cell in this case) can be read, for instance, by forcing a read current I_R into the selected bitline and connecting the selected wordline to ground, thus creating a current flow path indicated by the arrow line.

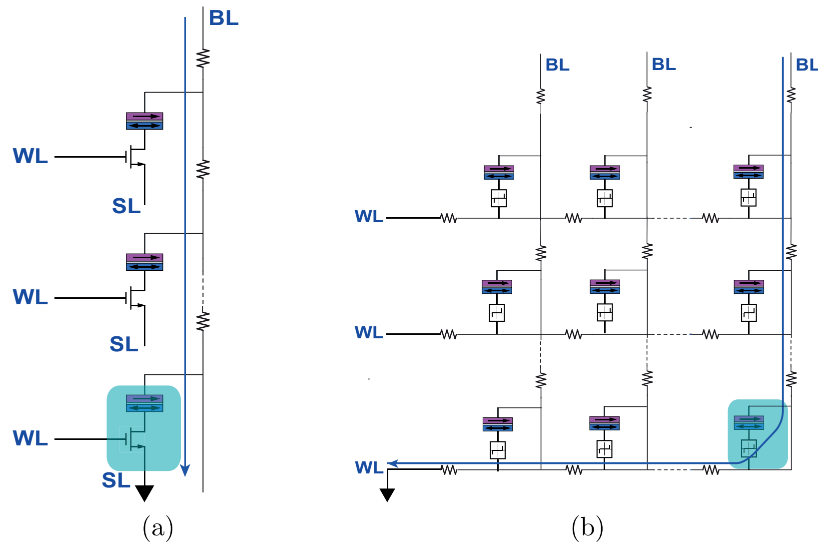


Figure 2.7: Illustration of STT-MRAM arrays (a) conventional array (b) crosspoint array

Without loosing generality, the read path (shown by an arrow line) in both array architectures can be modeled by Figure 2.8a, where R_{line} and C_{line} are, respectively, the total interconnection line parasitic resistance and capacitance of the read path.

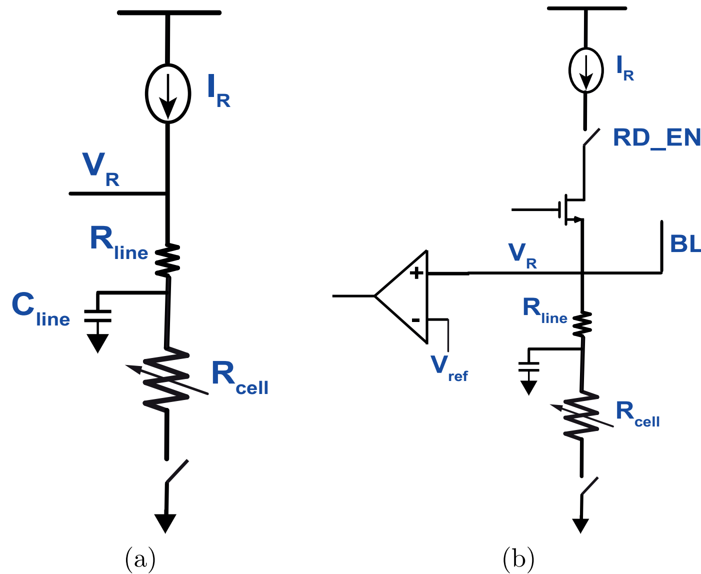


Figure 2.8: Reading in STT-MRAM: (a) simplified read path model (b) conventional scheme with the simplified read path model

2.5.1 Conventional Sensing Scheme

A conventional STT-MRAM sensing scheme (Figure 2.8b) involves applying a sufficiently small (compared to the switching current) read current I_R to the selected bitline and sensing the voltage on the bitline, which is compared to an external reference voltage V_{ref} to determine if the cell is in the high- or in the low-resistance state [1, 34]. The state of a memory cell located somewhere in the array can be correctly determined as long as the following condition is satisfied:

$$V_{R,L} = I_R\left(R_P + R_{line}\right) < V_{ref} < V_{R,H} = I_R\left(R_{AP} + R_{line}\right) \tag{2.10}$$

where $V_{R,L}$ and $V_{R,H}$ are the sensed bitline voltages when the cell is in the low-resistance P and the high-resistance AP state, respectively.However, with I_R sufficiently smaller than the switching current of the STT-MRAM cell and the small gap between R_P and R_{AP}, which is worsened by cell-to-cell variations on the values of R_P and R_{AP}, it is almost impossible to set a V_{ref} for some practically large-size arrays. Hence, the conventional sensing scheme is not suited for large-size STT-MRAM arrays.

2.5.2 Non-destructive Self-Reference Sensing Scheme

As a solution to the cell-to-cell variation problem, self-reference techniques have been proposed in the literature. In these techniques, sensing the bit stored in a cell relies on only the cell itself (rather than on using an external reference). For instance, a non-destructive self-reference sensing scheme is reported in the literature, which exploits the fact that the resistance of the high-resistance AP state decreases faster than that of the low-resistance P sate when an increasing read current is applied [34, 35, 46].

This sensing technique works as follows (Figure 2.9). A read current I_{R1} is applied to generate a BL voltage, V_{BL1}, which is stored in capacitor C_1: the value of V_{BL1} may be either low, $V_{BL1,L}$, or high, $V_{BL1,H}$ depending on the state of the cell. Then, a second read current, I_{R2}, which is larger than I_{R1} (say I_{R2}= βI_{R1}, with $\beta > 1$), is applied and generates a BL voltage $V_{BL,2}$ (i.e., $V_{BL2,L}$ or $V_{BL2,H}$). A fraction of $_{BL,2}$ (given as $V_{BL2,0}$= $\gamma V_{BL,2}$, where $\gamma = \frac{R_D}{R_D+R_U} < 1$) is then compared to $V_{BL,1}$.

Since the resistance of the high-resistance state decreases faster when the current through the cell increases from I_{R1} to I_{R2}, the current ratio β and the voltage dividing ratio γ can be properly designed so as to satisfy the condition: $V_{BL1,L} < V_{BL2,L0} < V_{BL2,H0} < V_{BL1,H}$ (where $V_{BL2,L0} = \gamma V_{BL2,L}$ and $V_{BL2,H0} = \gamma V_{BL2,H}$).

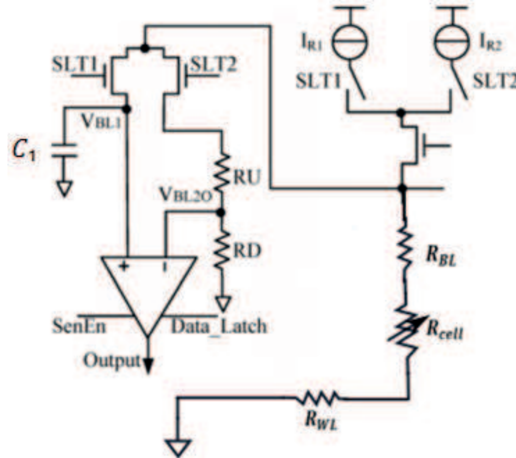


Figure 2.9: Nondestructive self-reference sensing (source: [1])

2.5.3 Destructive Self-reference Sensing Scheme

Another sensing technique proposed in the literature is destructive self-reference-sensing [1, 34, 35], which involves 4 steps: first read by applying current I_{R1} and store the resultant bitline voltage $V_{BL,1}$, then erase the data and write bit '0', next read by applying $I_{R2} > I_{R1}$ and store the new bitline voltage $V_{BL,2}$ (the original data is determined by comparing $V_{BL,1}$ and $V_{BL,2}$), finally write the original data back into the cell. I_{R1} should be smaller than the switching current so as not to modify the original data during the first readout. It is obvious from the reading process that this technique takes longer time and consumes more power when compared to the nondestructive sensing technique described earlier. However, since a read current as large as the switching current is used, the sense margin is much better than for the non-destructive and conventional sensing techniques discussed above.

2.5.4 Slope Detection Self-Reference Sensing Scheme

The principle of operation of slope detection sensing (Figure 2.10a) [1] is similar to the aforementioned destructive self-reference sensing. In slope detection sensing, a read current ramp is applied to the bitline instead of two currents pulses with amplitude I_{R1} and I_{R2} (in different time intervals). The slope detection sensing circuit scheme and the timing diagram are shown in Figures 2.10a and 2.10b.

Let us assume that the cell switches from the high- to the low-resistance state (or vice versa) when a current is positive (or negative) and higher than the critical current in magnitude. The red curve in Figure 2.11 shows a high-resistance state switching to low-resistance state when a positive current ramp is applied, whereas the blue curve, which corresponds to a cell is in the low-resistance state, shows that no switching occurs when the same current ramp is applied. At the switching instant, the bitline voltage drops (and, hence, shows a negative slope), whereas, if no switching occurs, the bitline voltage keeps always increasing and, hence, always shows a positive slope. The bitline voltage is sampled at time instants ϕ_1 and ϕ_{1d} (after some time delay), as indicated by the blue dots in Figure 2.11, and the sampled values are stored in capacitors C_1 and C_{1d}, respectively (Figure 2.10a). If the value stored in C_{1d} is less than the value stored at C_1 the memory cell state was '1' since a negative slope is detected. Otherwise, the memory cell state was '0'.

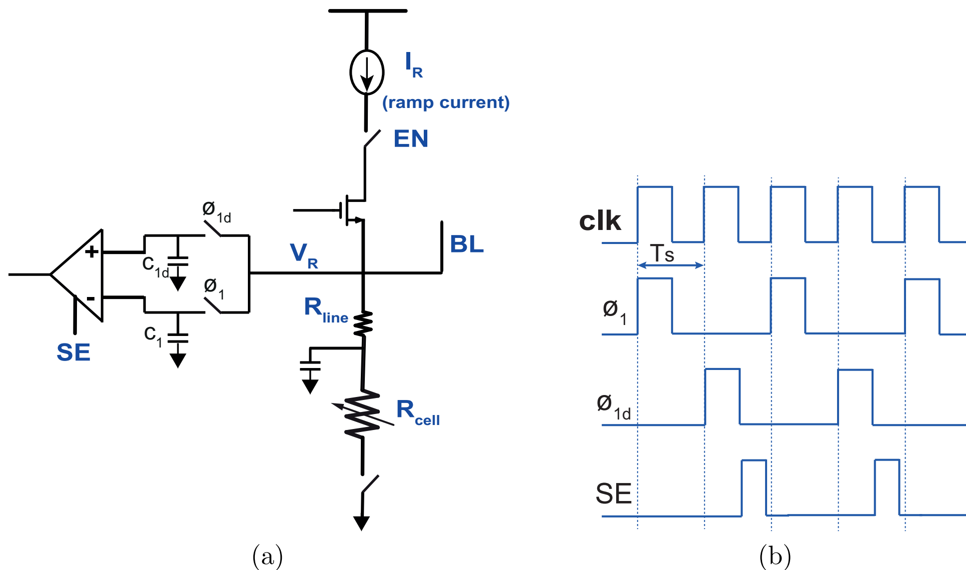


Figure 2.10: Slope detection sensing: (a) circuit scheme (b) timing diagram

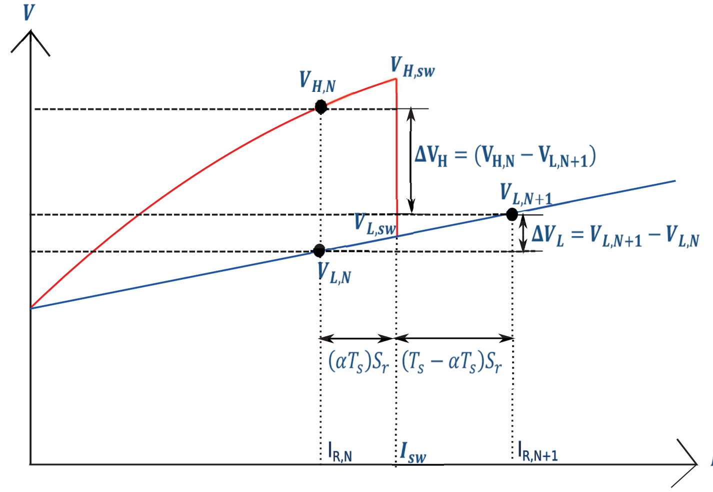


Figure 2.11: Sampling in slope detection sensing

To improve sensing robustness, the multiple sampling technique can be employed, in which a number of sampling-and-hold circuit pairs each one with its own sense amplifier are employed [1].

It is worth to point out that, in the destructive sensing technique (Subsection 2.5.3), the values of I_{R1} and I_{R2} are set in the design phase based on theoretical analysis and simulation, but their optimal values may be different in any real chip from the chosen values due to fabrication process induced variations. In contrast, slope detection sensing can be considered as a special case of destructive self-reference sensing in which, irrespective of process spreads and operating condition variations, we always can read with optimal I_{R1} and I_{R2} values among our multiple $I_{R1} - I_{R2}$ pairs of sampling currents. Hence, slope detection sensing has the best performance when compared with the other reviewed schemes. Subsection 2.5.5 is devoted to a detailed analysis and optimization of slope detection selfreference sensing scheme as applied to conventional and crosspoint STT-MRAM arrays.

2.5.5 Variation-Aware Analysis of Sensing Margin in Slope Detection Sensing Scheme

Let $S_r = \frac{\partial I}{\partial t}$ represent the slope of the current ramp signal, T_s represent the sampling period, αT_s $(0 \leq \alpha \leq 1)$ be the time interval between the instant when the first sample is taken and the instant when resistance switching (if any) occurs, and I_{sw} represent the switching current. Hence, the current difference between two consecutive samples (or current step) can be represented as $\Delta I = S_r \cdot T_s$. Using the above notations and the simplified model of the read path in the memory arrays shown in Figure 2.8a, the voltage changes indicated in Figure 2.11 can be calculated. The magnitude of the voltage change at the switching instant, if any, is given by:

$$\Delta V_{sw} = V_{H,sw} - V_{L,sw} = I_{sw}(R_{AP} - R_P) \tag{2.11}$$

The voltage at the selected bitline at the N-th sampling instant, if the cell was in high-resistance state, will be:

$$V_{H,N} = (I_{sw} - \alpha \Delta I) \cdot (R_{AP} + R_{line}) \,. \tag{2.12a}$$

similarly, if the memory cell was in low-resistance state

$$V_{L,N} = (I_{sw} - \alpha \Delta I) \cdot (R_P + R_{line}) \tag{2.12b}$$

Let us assume the resistance roll-off of R_{AP} and R_P is negligible so that their value does not change (apart from due to state switching) when the current changes from $I_{R,N}$ to I_{sw} and from I_{sw} to $I_{R,N+1}$. This approximation simplifies mathematical analysis while still having minimum impact over the obtained results. With this assumption, the voltage across the cell at the (N + 1)-th sampling instant can be expressed as:

$$V_{L,N+1} = (I_{sw} + (1-\alpha)\Delta I) \cdot (R_P + R_{line}) \tag{2.13}$$

The sense margins (i.e., the difference of the voltages sensed at the (N + 1)-th and the N-th instants) for the high- and the low-resistance state can now be calculated. For a cell in high resistance state (or bit ‘1’), the sense margin, denoted as $|\Delta V_H^c|$ (taking the absolute value), is given by:

$$|\Delta V_H^c| = |V_{L,N+1} - V_{H,N}| = \Delta V_{sw} - \alpha \Delta I \left(R_{AP} - R_P\right) - \Delta I \left(R_P + R_{line}\right) \tag{2.14}$$

Similarly, the sensing margin for a cell in low-resistance state, ΔV_L^c is easily obtained as

$$\Delta V_L^c = V_{L,N+1} - V_{L,N} = \Delta I \left(R_P + R_{line}\right) \tag{2.15}$$

From the above equation, the sum of the two sense margins for a single cell depends on technology (I_{sw}, R_{AP}, R_P), the choice of ΔI, and α; it decreases for increasing values of ΔI and α.

Let us now analyze the sensing margins considering a certain array size. From the above expressions of ΔV_L [equation (2.15)] and ΔV_H [equation 2.14)], it is apparent that the worst case for ΔV_L takes place when R_{line}= 0 (cell location closest to the wordline and bitline bias terminals), whereas the worst case for ΔV_H takes place when R_{line} is maximum (cell located at the most distant corner from the bitline and wordline bias terminals. The worst cases for ΔV_H in conventional and crosspoint STT-MRAM arrays occur when the shaded cells in Figures 2.7a and 2.7b) are selected. Hence, the nominal values of sensing margins in these worst-case cells are expressed as

$$|\Delta V_H| = I_{sw}(R_{AP} - R_P) - \alpha\Delta I\,(R_{AP} - R_P) - \Delta I\,(R_P + R_{line,t}) \quad (2.16a)$$

and

$$\Delta V_L = \Delta I\,(R_P) \quad (2.16b)$$

where $R_{line,t}$ is the total (parasitic) resistances of the selected path.

In the work that proposed the slope detection sensing technique, the authors argue that, from the design point of view, the key parameters to optimize sense margins are the slope of the current ramp signal and the sampling period [1] and aim at optimizing the sampling frequency and the slope of the current ramp signal separately. However, as is evident from equations (2.16a) and (2.16b), what is really important for sense margin optimization is the product of the two parameters (i.e., $\Delta I = S_r \cdot T_s$). Hence, a slow current ramp (which means longer sensing time) with a low sampling frequency or a fast current ramp with high sampling frequency (which requires a fast sample-and-hold circuit) can be used to obtain the same sensing margins.

Based on the above mathematical equations and the technology parameters given in Table 2.1, a variation-aware analysis was carried out using the given nominal and standard deviation values. Switching from the AP to the P state is assumed to occur at the critical switching threshold, calculated from Table 2.1: $I_{sw} = I_{c0,AP\rightarrow P} = 6.25\ \mu A$. For each case of the analyses, as a design choice, the current step ΔI is set so that the sense margins for bit '0' (2.16b) and bit '1' (2.16a) will be equal. A Gaussian distributed R_P and TMR spreads with the standard deviations (σ) given Table 2.1 and 5σ as a maximum variation is assumed and R_{AP} is calculated from R_P and TMR. Figure 2.12 shows the normal distribution of R_P (blue curve) and R_{AP} (red curve). The solid lines show the resistance values at zero current and the dashed lines show the actual resistance values after considering roll-off.

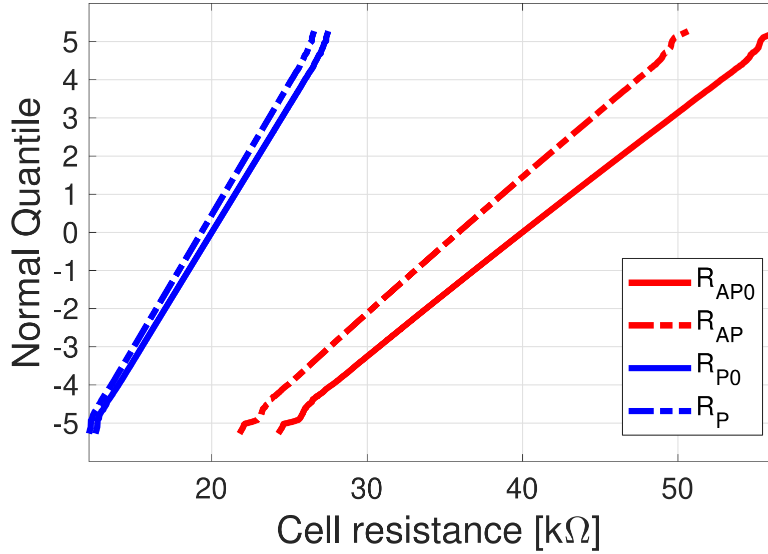


Figure 2.12: Normal inverse distribution of cell resistance

Figure 2.13 shows the impact of the parameter α on the sense margins. Refering to Figure 2.11, ideally we would like to take the N-th sample at the instant when the switching occurs ($\alpha = 0$), which gives the highest sense margin as indicated by the solid curves in Figure 2.13. However, due to the presence of variations, it is not practical to know the exact switching time (or current). The maximum deviation that may occur with respect to the ideal case is equal to one sampling time T_s, which corresponds to $\alpha = 1$. The sense margin when $\alpha = 1$ is shown by the dashed curves in Figure 2.13. In both cases, the impact of interconnection line resistance is neglected (i.e. $R_{line} = 0$). The impact of the parasitic resistance of interconnection lines on sense margin is illustrated in Figure 2.14, which shows the sense margins for $R_{line} = 0$ and $R_{line} = 10\ k\Omega$. If the parasitic resistance per memory cell is for instance, equal to 5 Ω, the 10 kΩ resistance can represent the total parasitic resistance along the read path during reading a memory cell located farthest from the BL bias edge in a conventional STT-MRAM array having 2000 memory cells along the BL (2.7a) or when reading the lower right corner cell in a 1000×1000 crosspoint STT-MRAM array (Figure 2.7b).

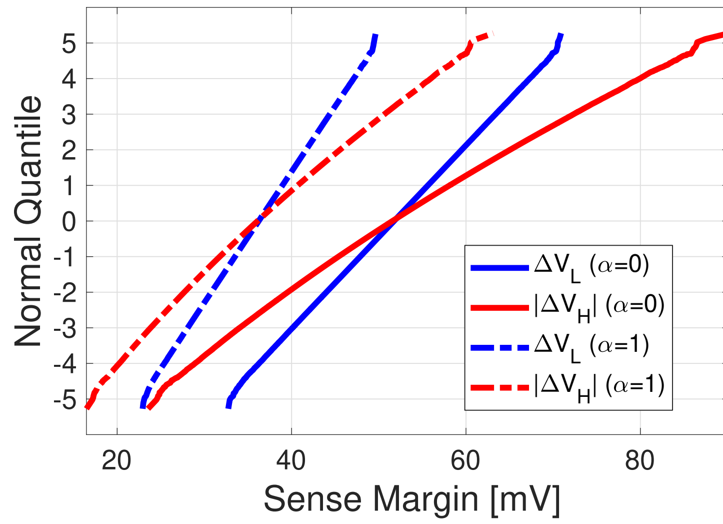


Figure 2.13: Normal inverse distribution of sense margin for $\alpha = 0$ and $\alpha = 1$ ($R_{line} = 0$)

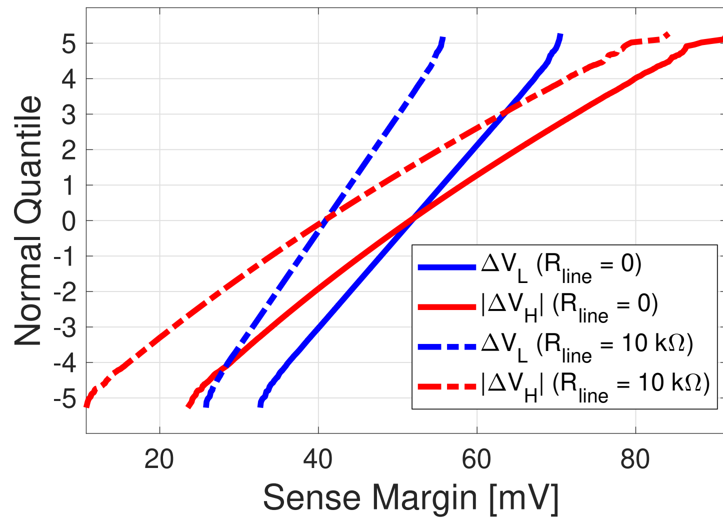


Figure 2.14: Normal inverse distribution of sense margin for $R_{line} = 0$ and $R_{line} = 10$ kΩ ($\alpha = 0$)

A comparision of the best- and worst-case sense margins is provided in Figure 2.15. Here, best refers to the case where $\alpha = 1$ and $R_{line} = 0$ and worst refers to case where $\alpha = 1$ and $R_{line} = 10\ k\Omega$. We can see that the median of the sense margin decreases from 50 mV in the best-case to to 30 mV in the worst-case. Another observation from all the plots of sense margin distributions (Figures 2.13, 2.14 and 2.15) is that $|\Delta V_H|$ is more affected by cell-to-cell variations in comparison to ΔV_L. This is because $|\Delta V_H|$ is affected by variations in both R_{AP} and R_P, whereas ΔV_L is affected only by the variations in R_P.

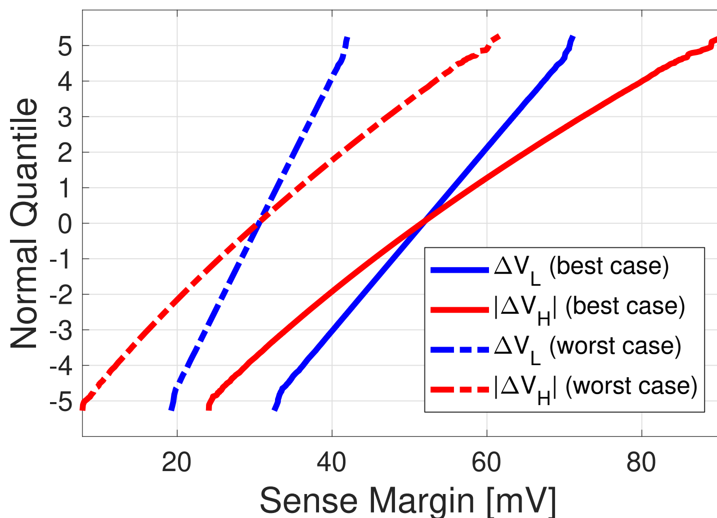


Figure 2.15: Comparison of best-case and worst-case sense margins

2.6 Conclusion

In this Chapter, a behavioral model of STT-MRAM cell has been presented. The STT-MRAM cell model is composed of a variable resistance whose value is controlled by a state control block and the current flowing through (the voltage across) the resistance. The state control block decides whether switching from one state to the other occurs or not using a probability of switching obtained from basic physics. The model mimics the dynamic (or switching) and static characteristics of the STT-MRAM cell and it is suited for circuit simulations, as demonstrated by the presented results.

In addition, it has been presented a review of STT-MRAM sensing circuit schemes and a variability-aware analysis and design guideline of slope detection self-reference scheme, which is deemed to outperform other STT-MRAM sensing schemes available in the literature. Using a simplified model for reading in conventional and crosspoint STT-MRAM arrays, the performance (i.e., sense margin) of the SD sensing scheme has been analyzed by taking into account the impact of cell-to-cell variations of resistance, variations in sampling time instant in the SD scheme and parasitic resistance in bitlines, BLs, and wordlines, WLs. A best-case and worst-case median sense margin of 50 mV and 30 mV, respectively, was obtained by assuming a HRS to LRS switching current of 6.25 μA, which are sufficient to determine the state of the STT-MRAM cell using a CMOS comparator with a good sensitivity.

Chapter 3

Electrical Characterization of Resistive Memories

3.1 Overview

Enormous research and development has been recently devoted to Resistive random access memories (RRAMs) as candidate emerging memory technology for embedded nonvolatile memory (eNVM) and storage class memory (SCM) application targets. As also briefly discussed in Chapter 1, a RRAM device is composed of a Metal/Insulator/Metal structure, i.e., an insulator material sandwiched between two metal electrodes [a top electrode (TE) and a bottom electrode (BE)]. There are two distinct types of resistive memory namely, Oxide Resistive RAM (OxRAM) and Conductive-bridge RAM (CBRAM), that differ (mainly) in the type of insulator material used: in OxRAM the insulator (also called switching layer) is metal oxide while CBRAM uses electrolytes as a switching material.

Each of the application targets have their own requirements. On the one hand, for embedded memory application targets, the operating voltages should be low enough for compatibility with the core CMOS transistor. Besides, compatibility with the thermal budget of CMOS Back-End-of the Line (BEOL) process is required. However, the array size generally ranges from few KB to few MB, and hence cell size is as critical as high-density storage applications. The main requirements for three categories of embedded RRAM application targets (automotive, general market and IoT sensor nodes) are summarized in Figure 3.1. On the other hand storage-class memory (SCM) requires low write current/power and small cell size (for storage density).

Application	Key attributes	Other requirements
Automotive	▪ Array size: 1MB-16MB ▪ Endurance: 1E6 ▪ Retention: 20yrs @150℃ ▪ Zero Defect capability	▪ Needs more mature technology ▪ Good understanding of failure modes, reliability mechanisms ▪ Reduced variability for zero defect
General Market (MCU eFlash replacement)	▪ Array size: 16kB-4MB ▪ Endurance: 1E4-1E5 ▪ Retention: 10yrs @ 85℃ (industrial 10yrs @105℃)	▪ Voltages compatibility with core transistor ▪ Compatibility with CMOS BEOL thermal budget ▪ Process cost should be minimized
IoT **Sensor node**	▪ Array size: 16KB-512KB ▪ Endurance: >1E6 ▪ Retention: 10yrs @ 70℃	▪ Cost is critical ▪ Minimal write/read energy ▪ Minimal variations to minimize write verify algorithms

Figure 3.1: Embedded RRAM requirements

As part of the effort to meet the aforementioned requirements for embedded and storage class memory application targets, this thesis presents an experimental electrical characterization OxRAM with TiN/Hf/GdAlO/TiN memory stack and Conductive-Bridging Resistive RAM (CBRAM) with Cu/TiW/SrTiOx/WOx/W stack layers, both fabricated on 300 mm wafers in imec's FAB (Leuven, Belgium). For the OxRAM, the results of an array-level study and optimization of FORMING, SET and RESET voltage pulses, and cyclic endurance are presented. In addition, the impacts of GdAlO thickness, cell size (i.e., diameter of the OxRAM device stack) and pitch size (i.e., distance between devices in the array) on performance and reliability are investigated. Similarly, for the case of CBRAM, optimization of memory performance and reliability (i.e., endurance and retention) by optimizing operating current and amplitude/shape of RESET/SET pulses, is discussed.

3.2 Basics of Resistive RAM (RRAM)

A RRAM storage device with its Metal/Insulator/Metal (MIM) structure and its operation principle is shown in Fig. 3.2. When stressed by an adequate external electrical voltage/current applied to the electrodes, the sandwich structure shows reversible and nonvolatile change of electrical resistance [47], meaning that the device can be repeatedly switched between, high-resistance and low-resistance state. Indeed, such resistive change property has been observed and studied in various metal oxide materials such as HfO_2, TaO_2, TiO_2, WO_2, Al_2O_3, MgO, $SrTiO_3$, NbO_2 [48]. Some oxides of rare-earth materials such Gd_2O_3 are also attractive for high-performance RRAM applications [49, 50].

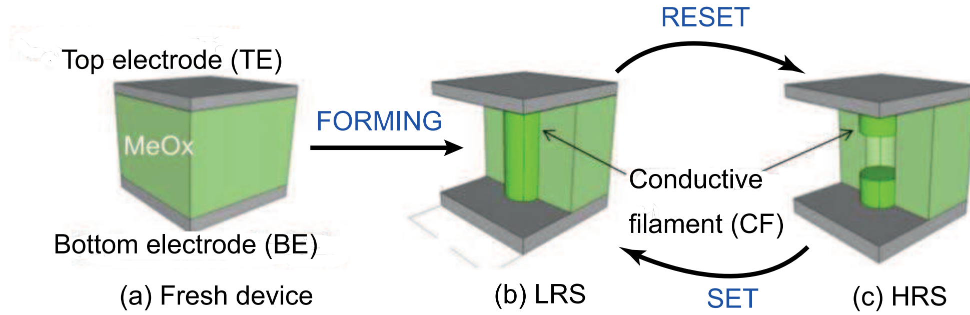


Figure 3.2: RRAM device structure and filamentary switching mechanism (source: [52])

Particularly, in filament-based RRAM, switching to the LRS and HRS relies, respectively, on the formation and the dissolution of a conductive filament (CF) within the insulating layer [7]. As a convention, switching to the HRS and to the LRS are called RESET and SET operation, respectively. The device is initially subjected to the operation of electro-forming (or simply FORMING), where a conductive filament is formed by dielectric breakdown (Figure 3.2(b)). The current should be limited to a compliance current I_c by a compliance system or a series resistor/transistor during forming, which allows the size of the CF to be controlled and avoids the destructive (hard) breakdown of the switching layer. After forming, the device manifests improved conductance as the CF connects the TE and BE by shunting the insulating layer, thus resulting in the low-resistance state (LRS) of the RRAM. The RESET operation can then be carried out to disconnect the CF, resulting in a high-resistance state (HRS), as shown in Figure 3.2(c) [51]. Just like the initial FORMING operation, a SET operation can be performed by applying a suitable positive voltage (typically lower than forming voltage) to the TE. Alternating the SET and RESET operations this way, the CF can be repeatedly connected/disconnected, thus allowing multiple transition cycles between HRS and LRS. If SET and RESET switching operations require opposite voltage polarities, the RRAM device is considered bipolar whereas a RRAM device is considered unipolar if the switching operation can be carried out without changing the voltage polarity [52]. In this case the magnitude, duration, and waveform of the write pulse distinguish the SET/RESET operations [53].

3.3 Electrical Characterization of GdAlO-Based OxRAM

3.3.1 OxRAM Device Stacks and Experimental Setup

For the experimental characterization, 4 memory stack implementations were prepared, as shown in Figure 3.3. In the stack labeled as D02, the thickness of the GdAlO layer is 5 nm. The memory stack structure is similar to the TiN/HfO_2 /Hf/TiN stack whose electrical characterization was reported by [54] apart from the fact that in this thesis work the HfO_2 switching layer is replaced by GdAlO. In D03 stack, the thickness of the GdAlO layer was reduced to 3 nm to reduce forming voltage. The stack labeled as D04 is similar to D03 except that the former has undergone an extended thermal treatment at 400^0C for 80 minutes. This was done to mimic the BEOL deposition of metal layers. In D05 stack, the bottom electrode TiN was replaced by Ru. However, the D04 and D05 stacks were found to be too leaky during the experimental test, and hence this work focuses only on the D02 and D03 stacks. Moreover, within each stack type, different cell size (diameter of device): 150 nm, 75 nm, 60 nm (Figure 3.4b) dense and isolated cells (Figure 3.4a) were integrated each as 1 Mb array. In particular, the switching performance and reliability of the 150 nm, isolated cells and the 60 nm/dense cells for the cases of D02 and D03 stacks, are investigated.

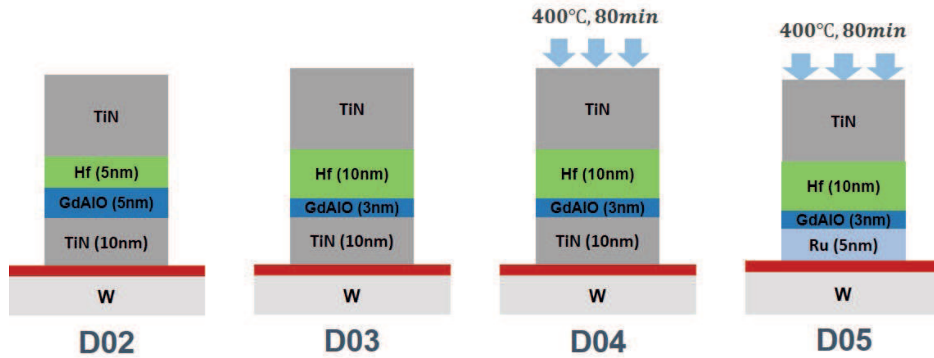
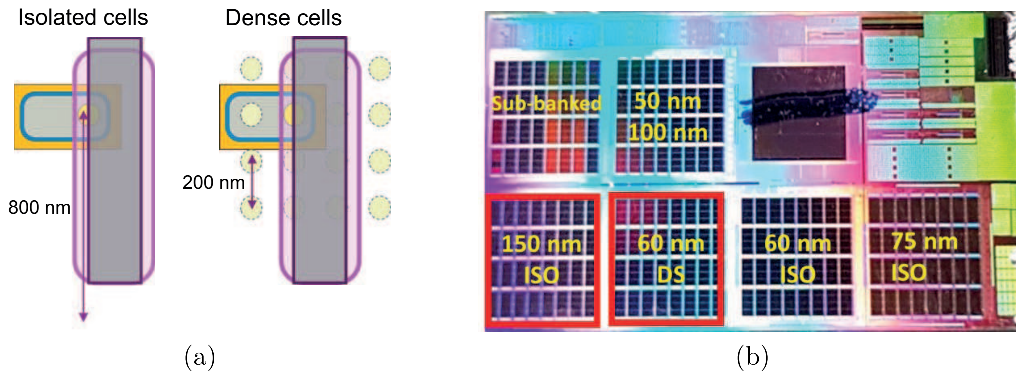


Figure 3.3: OxRAM pillar device stack



(a) (b)

Figure 3.4: (a) Isolated (ISO) and dense (DS) memory cells (b) Memory chip with different Mbit arrays

The memory device (or element) was implemented between Metal-3 and Metal-4 on top of a select and current-limiting transistor (65nm, 3.3V CMOS process) (Figure 3.5) in the so called 1T1R configuration. The schematic of the 1T1R configuration and the biasing conditions for forming, SET, read, and RESET operations are illustrated in Figure 3.6. Forming, SET and read pulses are applied to the source line, SL, while the bitline, BL, is connected to ground. The compliance current, I_c, for the forming and SET operations is controlled by setting biasing the gate bias of the transistor, i.e., by setting the WL voltage. On the contrary, to perform RESET that requires reversing polarity, the RESET pulse is applied to the BL while the SL is grounded.

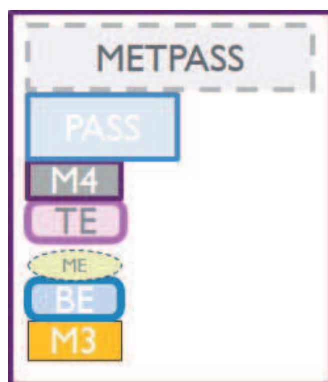


Figure 3.5: OxRAM Memory stack implemented between Metal-3 (M3) and Metal-4 (M4)

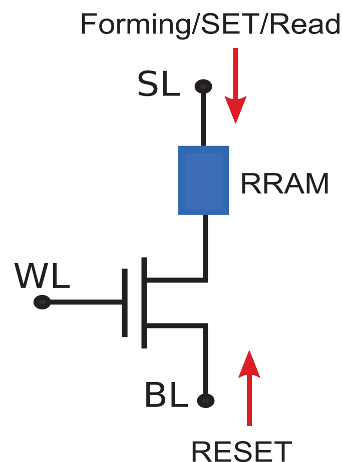


Figure 3.6: Schematic view of 1T1R configuration and biasing condition for RESET, FORMING, SET, and read operations

Reading is performed by applying a 0.1 V pulse and by sensing the change in current (using an amplifier connected at the source line) depending on the state of the memory cell is sensed. For the forming, SET and RESET operations, Single Pulse (SP) and Incremental Step Programming (ISP) were used depending on the type of test to be performed. In incremental step programming, a certain applied forming/SET/RESET voltage pulse of width PW is applied then read is performed and then the magnitude of the pulse is increased by a certain step size. These steps are repeated until the predetermined maximum pulse magnitude is reached [Figure 3.7(left)]. On the other hand, in single pulse (SP) programming, only a single pulse is applied following an initial read an then another read is performed [Figure 3.7(right)].

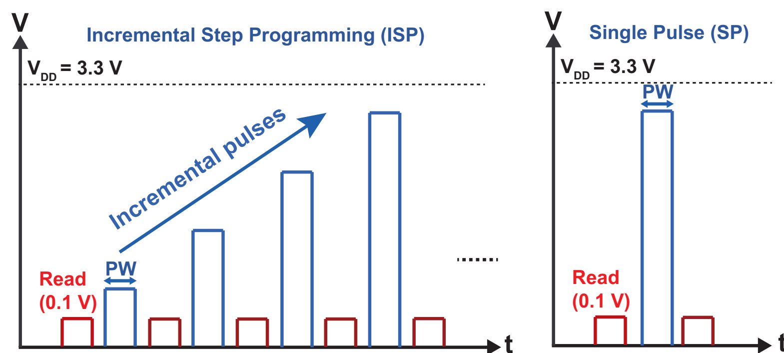


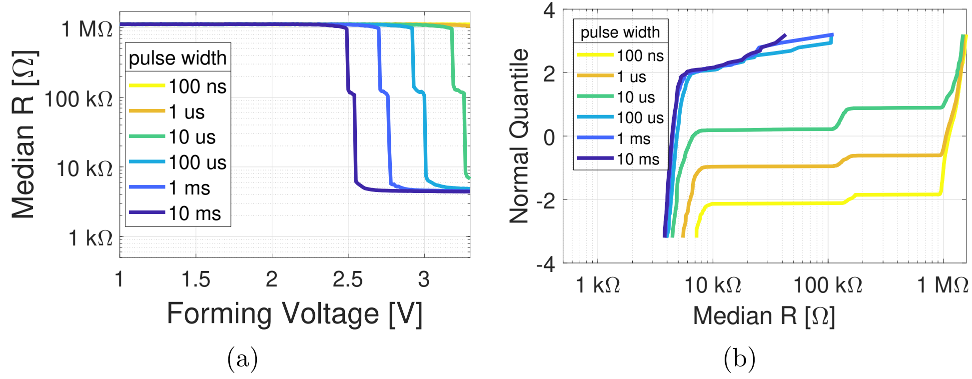
Figure 3.7: Programing pulses: Incremental Single Pulse (left) and single Pulse programing (right)

In general the objectives of this experimental characterization are evaluating switching performance as function of cell size (150 nm and 60 nm diameter) and pitch size between cells (isolated and dense), evaluating the statistical behavior of the performance at array level and studying endurance properties and optimizing SET/RESET voltages for better performance.

3.3.2 Forming

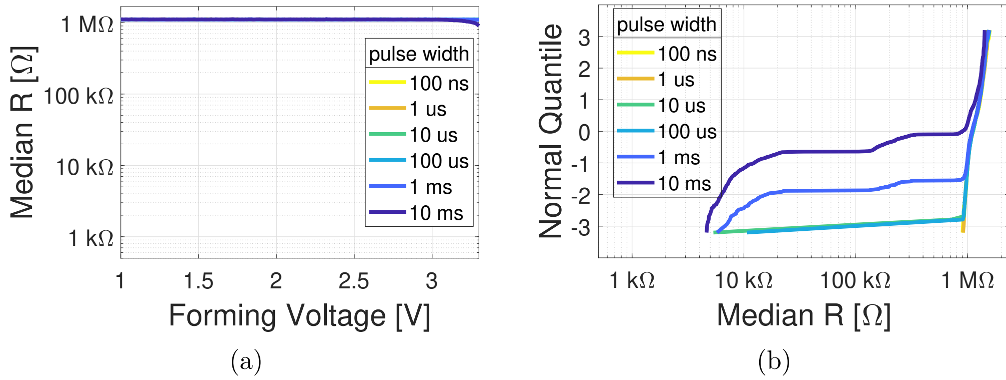
The Resistance-Voltage-time (R-V-t) plot of the 150 nm Ssize isolated cells (i.e., the 1 Mb array in the lower left corner of the memory chip shown in Figure 3.4b) having a D02 stack, is shown in Figure 3.8a. The experimental test was done by an incremental step programming (ISP) forming voltage at 150 μA compliance current. The vertical axis shows the median resistance of 1024 memory cells. Here, we can see the expected relationship between forming pulse width and magnitude: the longer the pulse the lower the required pulse magnitude.

Figure 3.8b shows the statistical distribution of the post-forming resistance; 100 kΩ is used as a threshold to differentiate the LRS and the HRS. For instance, we can see that for 10 ms and 1ms pulses almost all the cells were successfully formed whereas for 100 ns only up to -2σ cells were formed. On the other hand, the forming yield (i.e., the percentage of memory cells that are successfully formed) is much lower for the case of 60 nm size, dense cells with the same D02 stack (Figure 3.9a and Figure 3.9b).



(a) (b)

Figure 3.8: Forming in D02 (5 nm GdAlO), ISO (isolated), 150 nm (cell diameter): (a) forming R-V-t plot and (b) Distribution of post-forming resistance



(a) (b)

Figure 3.9: Forming in D02 (5 nm GdAlO), DS (dense), 60 nm (cell diameter): (a) forming R-V-t plot and (b) Distribution of post-forming resistance

A similar R-V-t and post-forming resistance distribution plots for 150 nm cell size, isolated cells with a thinned (3 nm GdAlO) D03 stack, are shown in Figure 3.10a and Figure 3.10b, respectively. We can see from these plots that the median forming voltage for all pulse widths are significantly lower than that of the 150 nm cell size, isolated cells with D02 stack (Figure 3.8a). Besides, a 100% forming yield is achieved as shown in Figure 3.10b. A similar trend is observed for the 60 nm dense cells (Figures 3.11a) as compared with 60 nm dense cells on with D03 stack (Figures 3.9a and 3.9b).

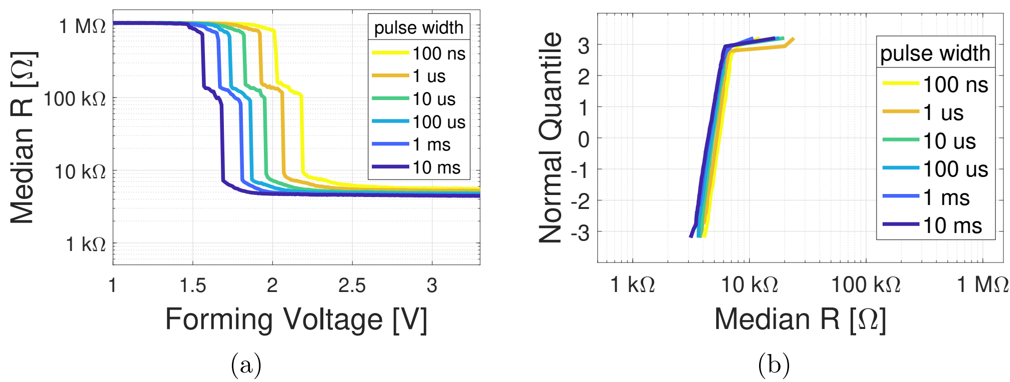


Figure 3.10: Forming in D03 (3 nm GdAlO), ISO (isolated), 150 nm (cell diameter): (a) forming R-V-t plot and (b) Distribution of post-forming resistance

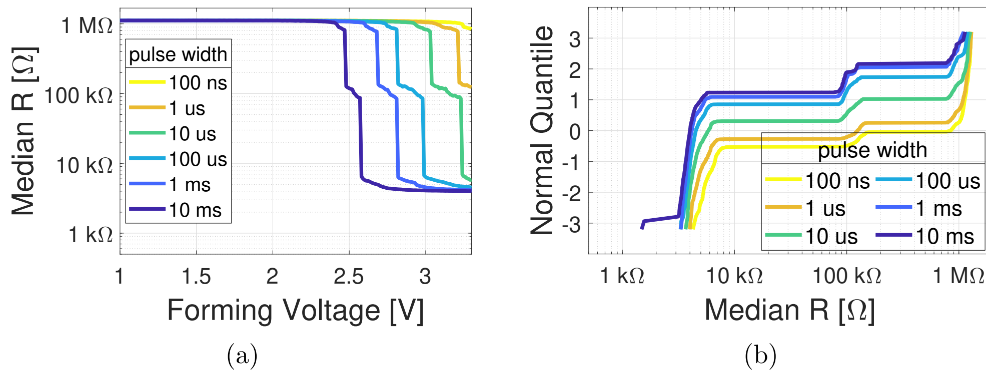


Figure 3.11: Forming in D03 (3 nm GdAlO), DS (dense), 60 nm (cell diameter): (a) forming R-V-t plot and (b) Distribution of post-forming resistance

On the one hand, we can conclude that reducing the GdAlO thickness from 5 nm to 3 nm enables to decrease the forming voltage both in 150nm isolated and 60 nm dense cells. On the other hand, the required forming voltage in 60 nm dense cells is higher than that of the 150 nm isolated cells.

3.3.3 RESET Voltage

First a 3.3 V Single Pulse (SP) forming voltage was applied to 1024 fresh cells. To increase the forming yield, a pulse width of 100 ms was used. Then after identifying the successfully formed cells, a RESET Incremental Step Programming (ISP) was applied to them. Figure 3.12a and Figure 3.12b shows the Median cell resistance versus reset pulse magnitude (for different pulse width) in isolated, 150 nm and dense, 60 nm cells with D02 stack. We can see that the peak RESET voltages in dense, 60 nm cells are higher than that of the isolated, 150 nm cells. We can also see that in both D02 and D03 stacks, the peak RESET voltage does not significantly change with RESET pulse width. Besides, peak RESET voltage does not show much dependence on the type of stack, as can be seen by comparing Figure 3.12a with 3.13a (isolated, 150 nm cells) and also Figure 3.12b with 3.12b (dense, 60 nm cells).

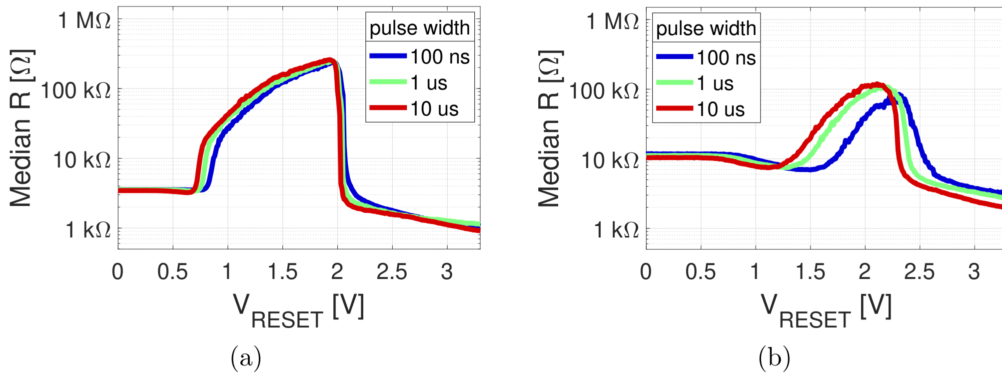


Figure 3.12: RESET R-V-t in D02 (5 nm GdAlO stack): (a) Isolated, 150 nm and (b) Dense, 60 nm

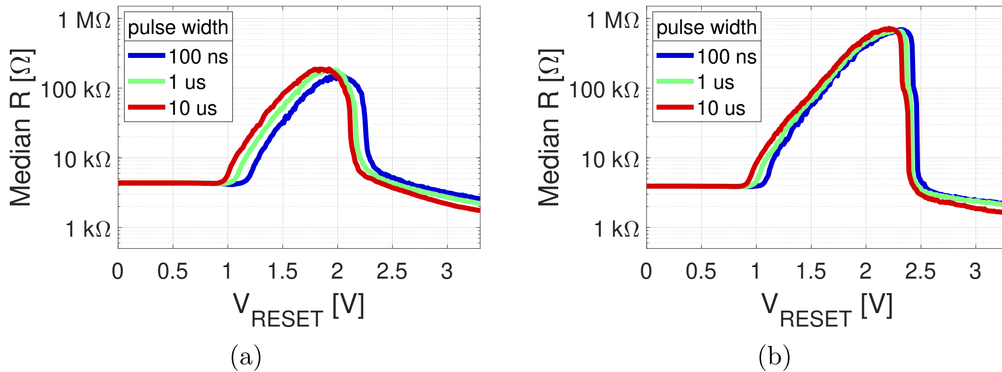


Figure 3.13: RESET R-V-t in D03 (3 nm GdAlO): (a) Isolated 150 nm and (b) Dense 60 nm

3.3.4 SET Voltage

Similar to the RESET, the SET characteristic was studied by applying a 3.3 V/100 ms forming voltage on 1024 fresh memory cells, then applying single RESET pulse followed by filtering the successfully RESET cells, and then applying SET ISP to them. Figures 3.14a and 3.14b show the SET switching characteristics for cells with D02 (5 nm GdAlO) stack and Figures 3.15a and 3.15b illustrate the SET switching characteristics for cells with D03 (3 nm GdAlO) stack. In all cases, unlike in the case of RESET, we can see a significant dependence of the SET voltage on pulse width. We can also observe that, similar to the case of RESET, the SET voltages of the dense, 60 nm cells are generally higher (though by only a small amount) than the for the isolated, 150 nm cells.

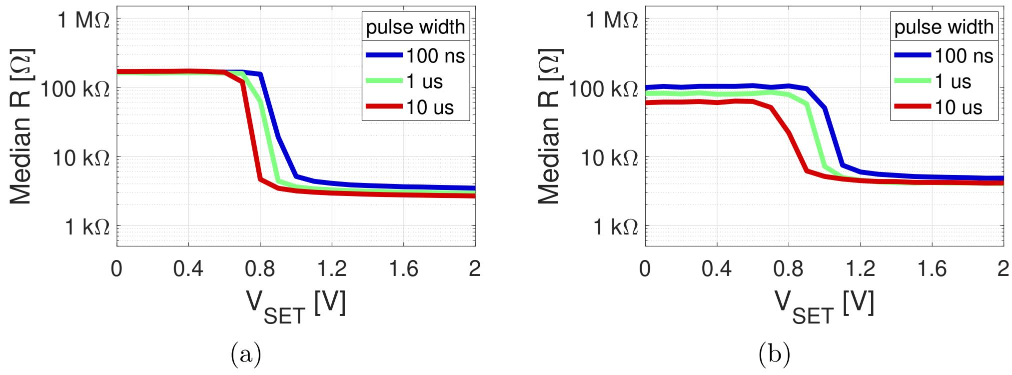


Figure 3.14: SET R-V-t in D02 (5 nm GdAlO stack) plot: (a) Isolated 150 nm cells and (b) Dense 60 nm cells

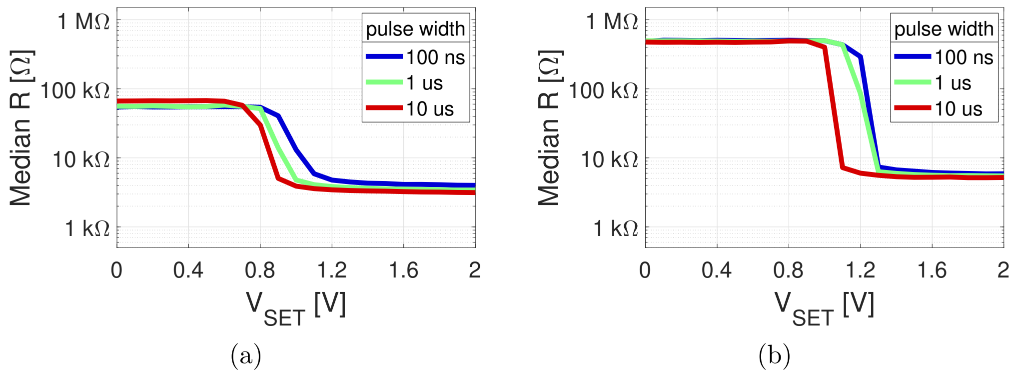


Figure 3.15: SET R-V-t in D03 (3 nm GdAlO stack): (a) Isolated, 150 nm and (b) Dense, 60 nm

3.3.5 Cyclic Endurance

The cyclic endurance of successfully formed cells was tested by applying 10^6 consecutive RESET/SET pulses. Resistance Window (RW), defined as the ratio of the HRS resistance (R_H) to the LRS resistance (R_L) is plotted against the number of SET/RESET cycles. Since the main source of endurance failure was found to be the HRS state operation in an early experimental test, only the RESET pulse was varied: -1.5 V, -1.75 V and -2 V while the SET pulse was kept to 3.3 V. From Figures 3.16a, 3.16b, 3.17a, and 3.17b, we can observe that the dense 60 nm cells have a better cyclic endurance and, especially for -1.5 V/3.3 V and -1.75 V/3.3 V RESET/SET pulses, the cyclic endurance are above 10^5, and hence are acceptable for instance for embedded Flash memory replacement.

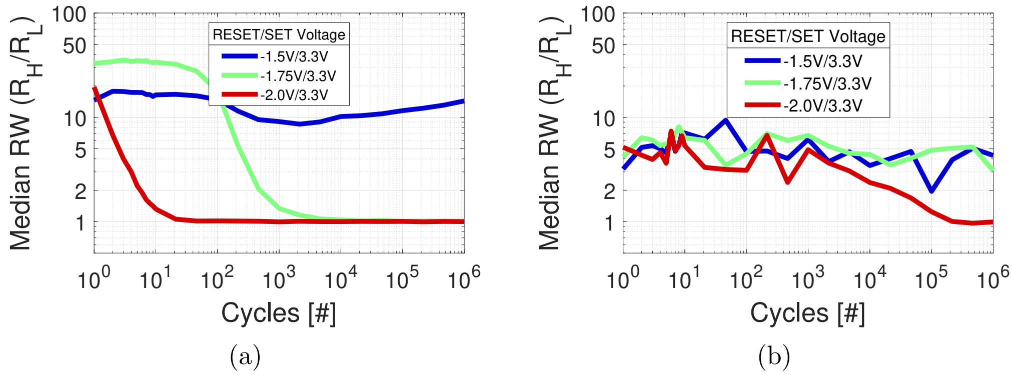


Figure 3.16: Cyclic Endurance in D02 (5 nm GdAlO) stack (150 μA compliance current): (a) Isolated 150 nm and (b) Dense 60 nm

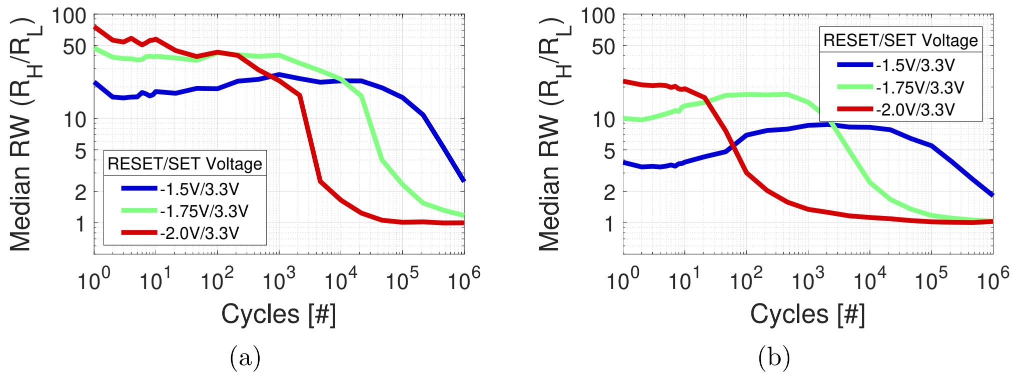


Figure 3.17: Cyclic Endurance in D03 (5 nm GdAlO) stack (150 μA compliance current): (a) Isolated 150 nm and (b) Dense 60 nm

The summary of the presented results at 150 μA operating current (for embedded memory target) are summarized in Table 3.1. A similar full array-level characterization was done at 50 μA compliance current targeting storage class memory (SCM) application. The results are provided in Table 3.2.

Table 3.1: Summary of results: experimental electrical characterization at 150 μA operating current (embedded memory target)

Array/device description	D02-ISO-150 (5 nm GdAlO, isolated cells, 150 nm cell diameter)	D02-DS-60 (5 nm GdAlO, dense cells, 60 nm cell diameter)	D03-ISO-150 (3 nm GdAlO, isolated cells, 150 nm cell diameter)	D03-DS-60 (3 nm GdAlO, dense cells, 60 nm cell diameter)
Forming voltage (Median)	2.5 V	>3.3 V	1.7 V	2.6 V
SET voltage (100ns pulse)	0.95 V	1.1 V	1 V	1.25 V
RESET voltage (100ns pulse)	-0.9 V	-1.75 V	-1.3 V	-1.2 V
Peak RESET voltage (100ns pulse)	-2.0 V	-2.2 V	-2.0 V	-2.3 V
Peak RW (or R_H/R_L)	70	10	35	180
Endurance (# cycles)	$5 \cdot 10^3$	$> 10^6$	10^5	10^5
R_H/R_L after 10 cycles	35	<2	40	20

Table 3.2: Summary of results: experimental electrical characterization at 50 μA operating current (storage class memory target)

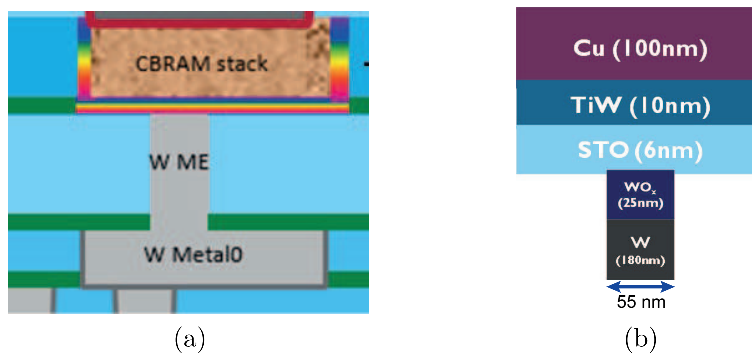
Array/device description	D02-ISO-150 (5 nm GdAlO, isolated cells, 150 nm cell diameter)	D02-DS-60 (5 nm GdAlO, dense cells, 60 nm cell diameter)	D03-ISO-150 (3 nm GdAlO, isolated cells, 150 nm cell diameter)	D03-DS-60 (3 nm GdAlO, dense cells, 60 nm cell diameter)
Forming Voltage (Median)	2.5 V	>3.3 V	1.7 V	2.6 V
SET Voltage (100ns pulse)	1 V	1.3 V	1.1 V	1.4 V
RESET Voltage (100ns pulse)	-0.8 V	-1.1 V	-1.2 V	-1.1 V
Peak RESET Voltage (100ns pulse)	-2.0 V	-2.35 V	-1.6 V	-2.3 V
Peak RW (or R_H/R_L)	35	60	10	100
Endurance (# cycles)	10^3	$> 10^6$	10^6	$5 \cdot 10^3$
RW after 10 cycles	15	<8	15	12

3.4 Electrical Characterization of $SrTiO_3$-Based CBRAM

In this section, an experimental electrical characterization of a dual-layer CBRAM with Cu/TiW/SrTiOx/WOx/W stack is presented. In [55] an interesting performance enhancement of CBRAM device is demonstrated by using dual switching layers: Al_2O_3 and WOx. By incorporating Al_2O_3 into a CRAM stack Cu/TiW/Al_2O_3/WOx/W, the authors were able to form an hourglass-shaped conductive filament [56], which enabled them to achieve a large memory window with high write speed and high cyclic endurance for an 10 μA operating current. However, it has been demonstrated that a better HRS retention is obtained when the $Al_2O_3$is replaced by $SrTiO_3$ (simply called STO) [57]. In this section, the results of an experimental study carried out to optimize the memory window, endurance and retention of Cu/TiW/SrTiOx/WOx/W CBRAM stack are presented.

3.4.1 Device Stack and Experimental Setup

The CBRAM device implemented in 1T1R configuration is shown in Figure 3.18a and 3.18b. Forming/SET/Read and RESET operations were carried out by applying a positive and negative voltage, respectively, to the top electrode of the device while the source terminal of the select transistor is always connected to ground.



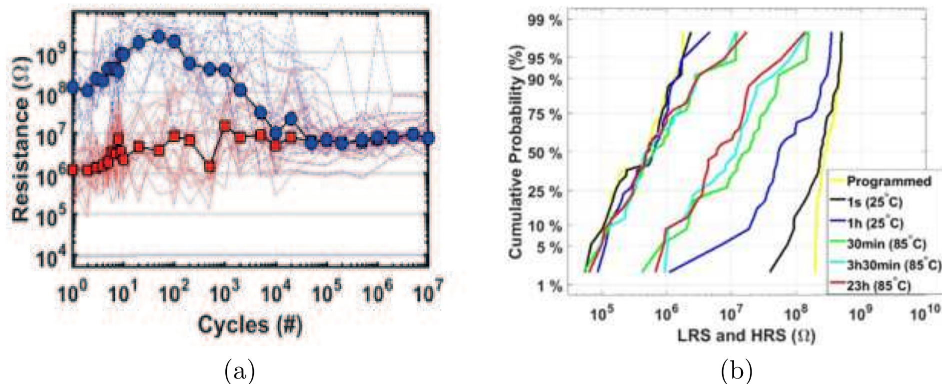
(a) (b)

Figure 3.18: CBRAM (a) 1T1R structure (b) STO-based CBRAM device stack

The endurance tests are performed by applying successive SET/RESET pulses. For the retention test, first 2 DC programming cycles are performed and then half of the devices will be programmed to HRS while the other half are programmed to LRS. Then the memory cells are read after 1 hr at room temperature and then they are baked for a predetermined temperature(s) and time(s).

3.4.2 Memory Window, Endurance and Retention

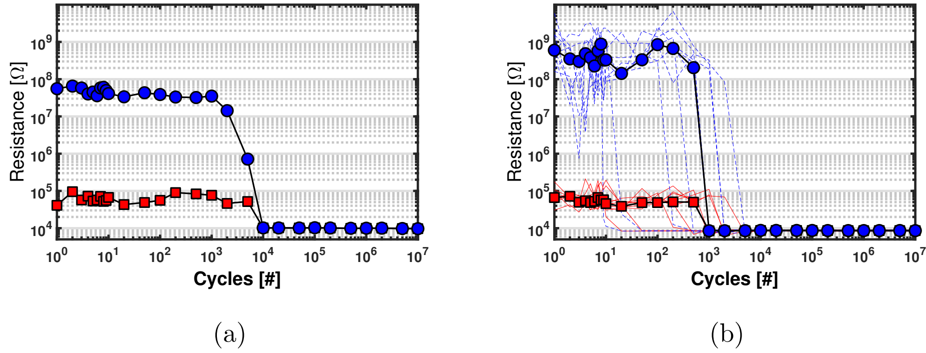
The cyclic endurance and the retention of the STO-based CBRAM stack at 10 μA forming and SET compliance current, 3.5 V SET voltage and -3.0 V RESET voltage are shown in Figure 3.19a and 3.19b, respectively. Figure 3.19a shows the cell median resistance of the tested devices versus the number of SET/RESET cycles. The blue curve shows the high resistance and the red curve shows the low resistance. As we can see the resistance window closes at 10^4 cycles. Figure 3.19b shows the cumulative distribution of the HRS (right-side curves) and the LRS (left-side curves) at room temperature and at different baking time and 85°C temperature.



(a) (b)

Figure 3.19: Reliability of STO-based CBRAM at 10 μA, 10 ns, V_{SET} = 3.5 V, V_{RESET} = -3.0 V: (a) cyclic endurance and (b) retention

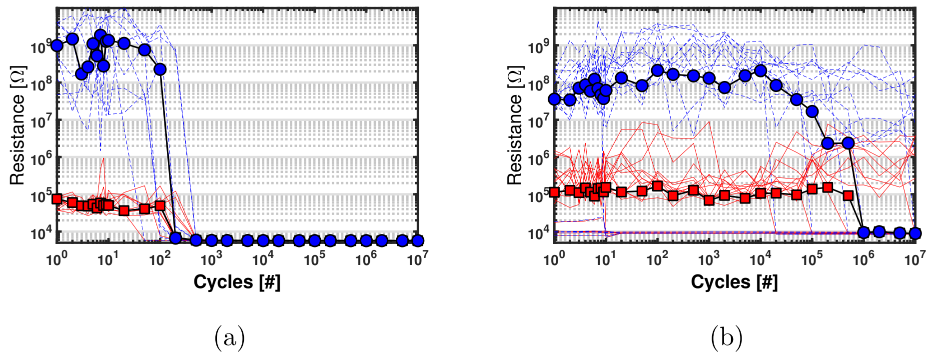
To increase the memory window, one can increase the operating current. For instance, Figure 3.20a shows the endurance test result at 50μA operating current. We can see that the low resistance decreases. By fixing the SET current to 50 μA, we can optimize the memory window by using a forming current different from the SET current Indeed, Figure 3.21a shows an improved resistance window for the case of 10 μA forming current and 50 μA SET current. However, the cyclic endurance is not good.



(a) (b)

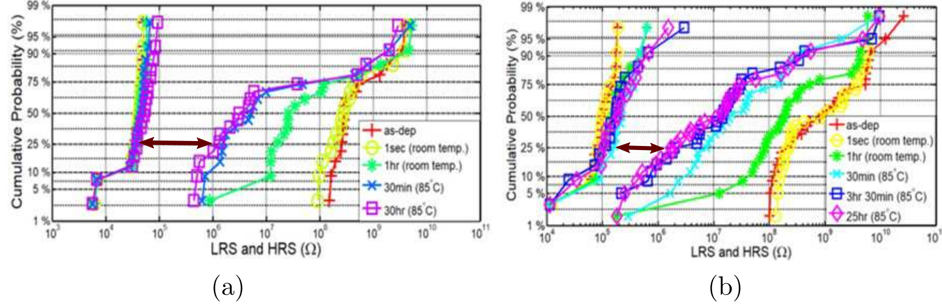
Figure 3.20: Comparison of endurance (V_{SET} = 3.5 V, V_{RESET} = -3.0 V): (a) forming and SET at 50 μA and (b) forming at 10 μA and SET at 50 μA

To improve the endurance, different combinations of SET and RESET pulse shapes and amplitudes optimizations were tried and a high endurance was obtained by using an optimized 5ns-1ns-5ns (i.e., rise time and fall time of 5ns and width of 1ns) triangular pulse. The achieved cyclic performance is shown in Figure 3.21a. The retention after programming the cell with such triangular pulses is shown in 3.21b, which is slightly worse when compared to the case of rectangular programming pulses. Hence, there is an endurance-retention trade off here.



(a) (b)

Figure 3.21: Endurance comparison (V_{SET} = 3.5 V, V_{RESET} = -2.5 V, forming at 10 μA and SET at 50 μA): (a) rectangular pulse (b) triangular pulse (5 ns-1 ns-5 ns)



(a) (b)

Figure 3.22: Comparison of endurance (V_{SET} = 3.5 V, V_{RESET} = -2.5 V, forming at 10 μA and SET at 50 μA): (a) rectangular pulse (b) triangular pulse (5ns-1ns-5ns)

3.5 Conclusion and Outlook

In this chapter it has been presented a detailed array-level electrical characterization of TiN/Hf/GdAlO/TiN OxRAM and device-level reliability study of Cu/TiW/SrTiOx/WOx/W CBRAM.

As for the OxRAM, the characterization was focused on analyzing the forming, SET and RESET voltages and Endurance at array level, with the prospect of tuning the performances for embedded memory (at 150 μA) and storage class memory (at 50 μA) applications. The impacts of the thickness of the GdAlO layer and the size of the memory cell on forming, SET and RESET voltages and on endurance was investigated. Accordingly, compared with the stack with 5 nm GdAlO (labeled as D02), the thinned stacks with 3 nm GdAlO thickness (labeled as D03) has shown a lower forming voltage. The median forming voltage was reduced from 2.5 V to 1.7 V for isolated, 150 nm cells and from 3.3 V to 2.6 V for dense, 60nm cells. However, no significant dependence of SET and RESET voltages on thickness of the GdAlO layer was observed. As for the impact of the size of cells, dense 60 nm cells have shown a higher forming voltage, reset voltage, set voltage in comparison with the isolated, 150 nm memory cells. Particularly, the peak RESET voltages in dense 60 nm cells are higher. Dense 60 nm cells have demonstrated a better cyclic endurance($> 10^5$) while isolated 150 nm cells have shown endurance of $\approx 10^3$. This is attributed to the higher peak RESET voltage of the dense, 60 nm cells.

It has been also attempted to study the impact of BEOL thermal treatment and using Ru as an alternative bottom electrode. However, the devices were found too leaky and the collected data is not good enough to draw conclusions. The issue of Ru as a bottom electrode could be due to a predisposition and perhaps the devices need to be studied with Transmission Electron Microscopy to study and solve the problem. Concerning the thermal treatment, it seems the devices can not withstand 400°C for 80 minutes. Hence, it would be a good idea to reduce the treatment time. Devices fabricated with different treatment time will help to understand more the impact of the thermal treatment.

As for the CBRAM, the characterization was aimed at studying the Endurance and retention of the Cu/TiW/SrTiOx/WOx/W stack and to optimize the SET and RESET pulses to obtain optimum memory performance and reliability. By optimizing the amplitude and shape of SET/RESET pulses, a resistance ratio of close to 10^3 and a cyclic endurance of 10^6 was achieved.

Chapter 4

Crosspoint Memory Arrays for High-Density Storage

4.1 Overview of Crosspoint Memory Arrays

In crosspoint arrays, each memory cell (i.e., each memory element plus selector device, if any) is built at the junctions (or 'crossing point') of a lower and an upper plane of parallel interconnection metal lines (bitlines, BLs, and wordlines, WLs) running at right angles to each other. If both the width of the metal lines and the spacing between them is equal to the minimum lithographic feature size, F, the memory cell is allocated within the smallest single-layer cell footprint of $2F \times 2F$ (or $4F^2$), thus providing high cell density [7, 15, 58, 59]. The effective area per cell can be reduced even further to $4F^2/N$ with N-layer 3D integration [59–61]. Typically, the overall memory chip will be composed of crosspoint sub-arrays, which might each be on the order of 1000 × 1000 interconnection metal lines in size. We want each of these sub-arrays to be as large as possible, such that much of the peripheral circuitry can be placed underneath the arrays, thereby reducing their silicon footprint. The larger the crosspoint sub-array, the higher the area efficiency and this usually implies lower cost per bit [62].

For the 3D integration of crosspoint arrays, two alternative schemes have been proposed. The first alternative is stacking multiple 2D arrays into a 3D configuration (or horizontal 3D crosspoint) as shown in Figure 4.1a and the second alternative is a vertical 3D integration (or vertical 3D crosspoint) as shown in Figure 4.1b [60, 61]. On the one hand, the horizontal 3D crosspoint scheme has some advantages such as higher peripheral circuit efficiency and lateral scalability [60].

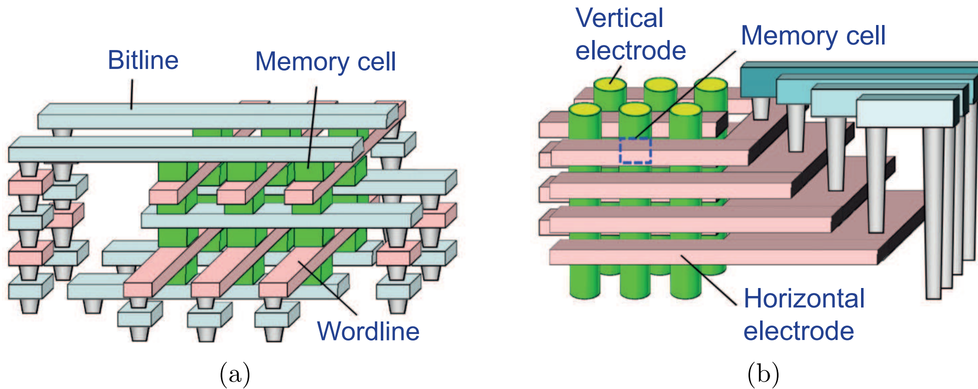


Figure 4.1: Schematics of 3D crosspoint memory arrays (adapted from: [63]]: (a) horizontal and (b) vertical configuration

However, the fabrication cost in horizontal 3D crosspoint arrays increases linearly with the number of layers, and hence the cost per bit does not always scale with the increasing number of layers [60, 61]. On the other hand, the vertical 3D crosspoint scheme features bit-cost scalability with increasing number of layers. However, vertical 3D crosspoint faces some implementation challenges. It is not allowed to add an intermediate electrode to separate the memory element and the selector device as the intermediate electrode would short circuit the memory cells in the same column [60, 61]. As a result, the only promising option for this scheme is to use a self-selective memory element with rectifying or built-in nonlinear current-voltage (I-V) characteristic [60, 61]. However, implementation of a self-selective memory cell with the required memory performance and nonlinearity is quite challenging [60].

This work focuses on planar crosspoint memory arrays. Although there are some additional issues in 3D crosspoint arrays that require consideration, a study on the fundamental design considerations and technology requirements of planar crosspoint arrays, which are addressed in this chapter, are equally applicable to 3D crosspoint arrays. In particular, from the circuit-designers' point of view, the main design challenges of crosspoint memory arrays come from the crosspoint architecture, not from the 3D integration.

4.2 Challenges of Crosspoint Memory Arrays

Figure 4.2a schematically illustrates a two-layer horizontal 3D crosspoint array, with each memory cell comprising of a memory element implemented on top of a two-terminal selector device, thus giving rise to an 1S1R (one-selector one-resistor) crosspoint memory. The function and the desired features of the selector device will be clarified below. Figure 4.2b shows a circuit schematic of a planar 1S1R crosspoint array with one selected memory cell located at the lower right corner. It also indicates the unselected memory cells (i.e. cells connected to unselected WLs and BLs) and the half-selected cells along the selected WL and BL (i.e. non-addressed cells sharing either the WL or the BL with the selected cell), for simplicity referred to as half-selected WL cells and half-selected BL cells, respectively. Generally, in an $N_{WL} \times N_{BL}$ crosspoint array (where N_{WL} is the number of WLs and N_{BL} is the number of BLs) with one cell selected at a time for write/read, there will be $N_{BL} - 1$ half-selected WL cells, $N_{WL} - 1$ half-selected BL cells, and $(N_{WL} - 1) \cdot (N_{BL} - 1)$ unselected cells. Let us assume that to operate the array, the selected WL is connected to the write voltage (or current) source V_W (or I_W) and the selected BL is connected to ground, whereas, the unselected WLs and BLs are connected to bias voltages V_{uWL} and V_{uBL}, respectively. The voltages across half-selected WL cells, half-selected BL cells and unselected cells are denoted as ΔV_{hWL}, ΔV_{hBL}, and ΔV_u, respectively. The selected path is shown with the blue arrow curve extending from the selected WL edge to the selected BL edge.

The first critical challenge of crosspoint memory arrays is that when we activate certain memory cell(s) by applying the required bias voltage(s)/current(s) to the wordline(s) and bitline(s) for performing write/read, other cells that are not intended to be written/read will be partially activated, which results in parasitic sneak (or leakage) current paths [7, 59, 64]. For instance, in Figure 4.2b connecting the selected WL and BL to V_W and ground, respectively, to write the selected cell will result in leakage currents through half-selected and unselected cells due to the voltages ΔV_{hBL}, ΔV_{hWL}, and ΔV_u. Sneak path currents uselessly raise the power consumption, and for the write operation, may produce write disturbance conditions. For the read operation, sneak path currents add noise to the signal being sensed, thus diminishing the read margin [5, 65, 66].

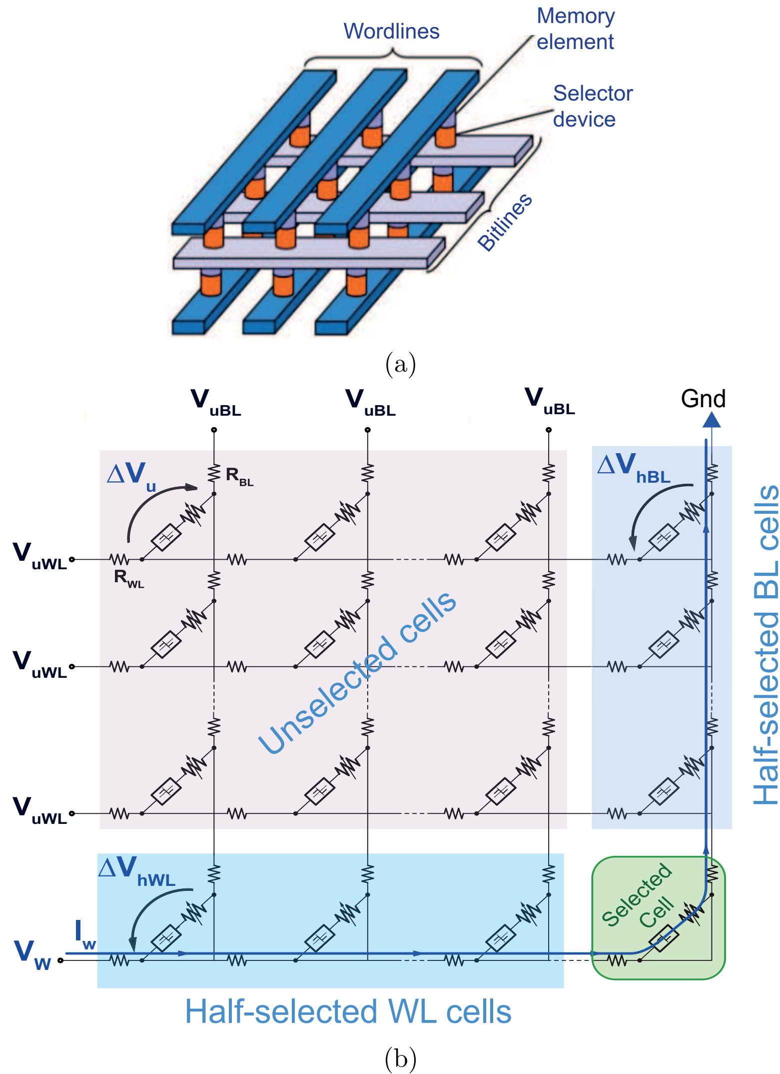


Figure 4.2: 1S1R crosspoint array: (a) physical schematics of a 2 layer array (b) a circuit schematic of a planar array with one selected cell shown

To solve the sneak path currents issue, the memory cell should as to provide it with a strongly nonlinear current-voltage (I-V) characteristic, i.e. it should turn off at low bias voltages (i.e. ΔV_{hBL}, ΔV_{hWL}, and ΔV_u) and turn on at adequately larger bias voltage (i.e. V_W if ohmic voltage drop along the selected path is neglected). This way, by properly biasing the array, the parasitic sneak (leakage) currents paths through half-selected and unselected memory cells can be suppressed while still delivering sufficient voltage and current to the selected cell(s). In this respect, there are two approaches for introducing nonlinearity into the memory cell.

The first approach is engineering the memory device to have built-in nonlinearity, referred to as self-rectifying cell (SRC). The second approach is integrating a separate nonlinear selector device in series with each memory element, thus giving rise to the 1S1R configuration, as shown in Figure 4.2b. The second approach, which is discussed in this work, has the advantage that the memory element and the selector device can be optimized separately and then integrated together in the final semiconductor processing scheme. It has been reported in the literature that the higher the nonlinearity in the selector device the larger the size of the implementable array, thanks to the better sneak current suppression [58, 67, 68].

The second critical challenge of crosspoint memory arrays is the ohmic (IR) drop due to parasitic resistance in the interconnection metal lines, which degrades the accessibility of the target cell, especially during memory write operation on cells located far from the array bias voltage/current sources, hence resulting in write failure. Sense margin also diminishes due to parasitic resistance and capacitance [5, 15, 16, 68]. Clearly, the larger the crosspoint array size the longer the interconnection metal line, and, hence, the higher the impact on write and read performance [16]. The peculiarity of the IR drop problem in crosspoint memory arrays as opposed to other circuits is that, even when they are not constrained by the supply voltage, meaning there is a room to increase the write/read bias voltage to compensate for the IR drop, it may not be allowed to raise the bias voltage above a certain value since doing so will increase the aforementioned sneak path current and hence, the disturbance and leakage power in half-selected and unselected memory cells. Hence, at some point, the drawbacks to using large arrays can partially counteract or even overwhelm the area efficiency benefits.

In general terms, memory write performance in 1S1R crosspoint memory arrays depends on the amount of the switching current and voltage of the memory element [69], the nonlinearity and the operating voltage of the selector [70, 71], the interconnection line resistance [70, 72], and the biasing scheme employed to operate the memory array [16]. On the other hand, read performance (or read margin) depends on the ratio of the HRS resistance value, R_H, to the LRS resistance value, R_L, [64], as well as on interconnection line resistance [71], biasing scheme, and selector device characteristic [71].

This chapter presents a comprehensive analysis of a generic 1S1R crosspoint memory array by considering circuit design and device technology parameters. In particular, it discusses the constraints on the allowed combinations of memory element and selector device to realize crosspoint arrays. For this purpose, circuital and mathematical array models have been developed to analyze the design constraints in terms of parameters of memory element, selector device, interconnection line, array biasing scheme, and array size.

4.3 Selector Devices for Crosspoint Memory Arrays

There are some desirable characteristics for a selector device to be used for building 1S1R crosspoint arrays. From the circuit designer viewpoint, the most important characteristics are the nonlinearity and the operating voltage/current. To a first order, the nonlinearity of the selector device used to implement the crosspoint memory array determines the maximum feasible array size and the operating voltage/current determines the compatibility of the selector device with the resistive memory element technology. Hence, we will focus on this two characteristics in this work. However, for the sake of providing the complete picture, the general selector requirements that come from circuit performance, device and process aspect, are briefly discussed in this section. Then, a brief survey of different selector technologies reported in the literature is presented, followed a discussion of the modeling of selector devices.

4.3.1 Selector Device Requirements

4.3.1.1 High Current Density

As a prerequisite, the selector device should be able to provide sufficiently high current to the memory element during all memory operations [5, 58]. It is necessary to consider that this high current must be delivered in spite of the very low cross-section area of the selector. For instance, programming a 20 nm diameter STT-MRAM cell into its high-resistance state by applying 25 μA current, translates into a current density of $^{8MA}/_{cm^2}$. Hence, the selector device, assuming its area to be the same as the area of the memory cell, should be able to provide at least $8MA/cm^2$ current density.

4.3.1.2 Two-Terminal Structure

To exploit the high cell density benefit inherent in the crosspoint array architecture, in other words to achieve the the minimum cell footprint, the selector has to be a two-terminal device. The use of a a three-terminal transistor as a crosspoint array selection device is hindered by this cell density requirement although it acts as a perfect switch for blocking leakage current [5].

4.3.1.3 Nonlinearity (Selectivity)

Nonlinearity, which is a measure selectivity, is one of the crucial requirements of selector devices for 1S1R crosspoint arrays [5,58]. The nonlinearity can be defined in different ways. The most commonly used one is half-bias [66,68] nonlinearity, $NL_{1/2}$, defined as the ratio of the current at the operating voltage of the selector, V_{op}, to the the current when the selector is biased at $V_{op}/2$, i.e., $NL_{1/2} = I(V = V_{op})/I(V = V_{op}/2)$. This definition is derived from the commonly used 1/2 array biasing scheme, in which for instance, V_{uBL} and V_{uWL} in Figure 4.2b are set to $V_W/2$ and, hence, the voltage across the half-selected memory cells will be equal to $V_W/2$, while the voltage across the selected memory cell will be equal to V_W. Hence, V_{op} and $V_{op}/2$ are the voltages across the selector device when the voltage across the 1S1R memory cell are V_W and $V_W/2$, respectively. According to this definition of nonlinearity, the higher the value of $NL_{1/2}$, the better the selectivity.

Another way to define selector nonlinearity is by using the inverse slope, denoted as δ, of the selector's current versus voltage (I_{slc}-V_{slc}) characteristic in a Log-lin plot [i.e., $\delta = {}^{dV_{slc}}/{dLog_{10}(I_{slc})}$] [70], where V_{slc} and I_{slc} are the voltage and the current of the selector device, respectively. According to this definition of nonlinearity, the lower the value of δ, the higher the selectivity. Since the operating voltage of the selector device may change depending on where it is used, and biasing schemes other than the 1/2 scheme can be used, the former definition of nonlinearity is not well-suited for the study in this thesis work. Whereas, in the latter definition, a single value of δ, otherwise two values namely, turn-off and turn-on slope (depending on the type of the selector device as discussed in Subsection 4.3.2) can be used to define the nonlinearity over the entire operating voltage range of the selector device. Hence, this definition of nonlinearity is used throughout this chapter.

4.3.1.4 Voltage Compatibility with Memory Element

The polarity (i.e., bipolar or unipolar) and the operating voltage range of the selector device should be compatible with that of the memory element [5,73]. For example, STT-MRAM and bipolar RRAM require a bipolar selector device, which should feature high current drive capability and high non-linearity at both polarities. Resistive memory elements have various SET and RESET voltages depending on the material system and the underlying working mechanisms. It is important that the selector element is compatible with the memory cell, in order to transfer selector non-linearity to the 1S1R full cell, so as to ensure limited leakage current from the unselected memory elements during both read and write operations.

4.3.1.5 Process Compatibility

Process compatibility is another key requirement of a selector device to be used in crosspoint memory arrays. The materials utilized to fabricate the selector device should be CMOS process compatible and, to enable 3D stacked memory arrays, the thermal budget of selector device fabrication should be compatible with the back-end-of-line (BEOL) process steps of CMOS technology [5, 58]. It is also desired that a selector has a simple structure and low aspect ratio, to reduce process complexity [5].

4.3.2 Selector Device Technologies

Even though there is a large number of selector device implementations proposed to date, we can group them into two main categories based on their I-V behavior [5, 66, 74]. In the first category, we have threshold selectors, which show an I-V behavior characterized by an abrupt increase of the current at a given threshold voltage, V_{th} [5, 7, 69], as shown in Figure 4.3 (right side). For example, Field Assisted Superlinear Threshold (FAST), Threshold Vacuum Switching (TVS), and Chalcogenide based Ovonic Threshold Switching (OTS) [5, 69] fall into this category. The selectors in the second category are referred to as exponential (or diode-like) selectors. They show nonlinear I-V characteristics, where the current changes gradually with the applied voltage without any abrupt I-V transition [5, 67, 69, 74], as shown in Figure 4.3(left side). Some of the selectors in this category include: Mixed-Ionic Electron Conduction (MIEC) device, varistor, and Metal-Silicon-Metal (MSM) device [5, 69, 74].

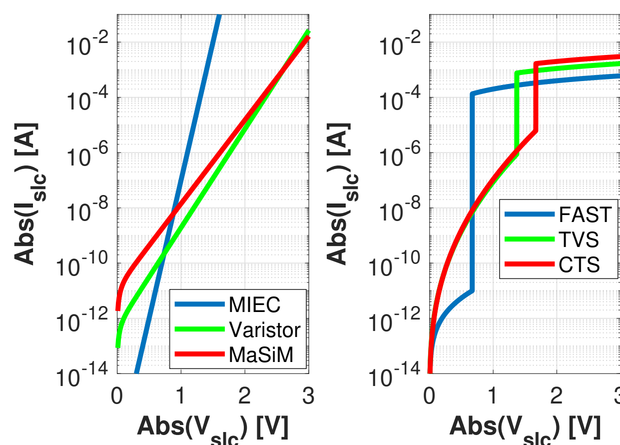


Figure 4.3: Representative I-V characteristics of selector devices: exponential selectors (left) and threshold selectors (right)

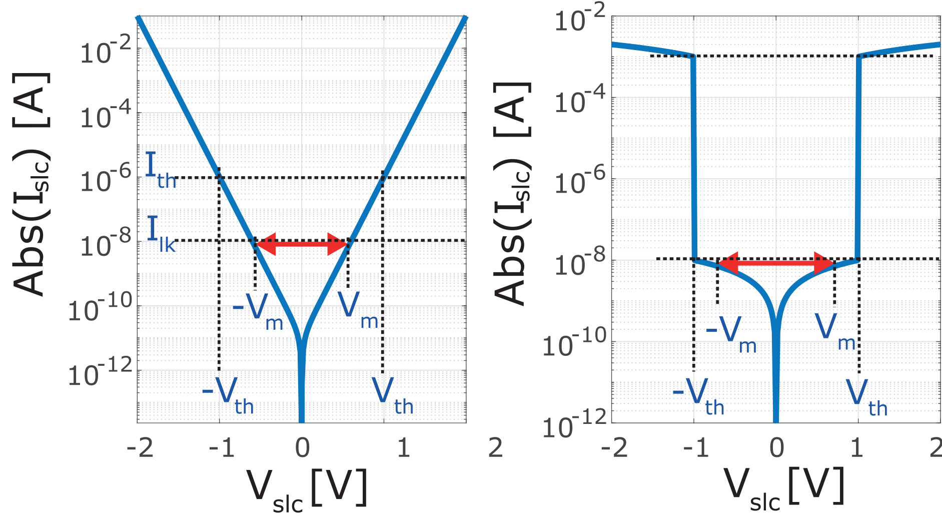


Figure 4.4: I-V characteristics of exponential (left)) and threshold (right) selector devices

4.3.3 Model of Selector Device

In this chapter, the bipolar exponential selector devices can be are modeled by two diodes connected in anti-parallel configuration to mimic bipolar I-V characteristic, that is required for memories such as STT-MRAM and bipolar RRAM). However, since the mathematical analyses in this work are done by considering only the magnitudes of voltages and currents with no assumption of their polarity, the results presented are equally applicable to PCM crosspoint arrays, in which a unipolar selector can be used. Mathematically, the I-V characteristic of the exponential selector device is modeled by a hyperbolic sine function given by (4.1), which is obtained by summing up the current contributed by the two anti-parallel diodes (the I-V characteristic in each diode is modeled by Shockley ideal diode equation). The model is well suited for fitting the I-V characteristics of various exponential selectors to published experimental data simply by varying the two parameters in the equation, i.e., I_{ss} and δ. The parameter I_{ss} corresponds to the reverse bias saturation current in the diode equation and the parameter δ, as defined in Subsection 4.3.1, is the inverse slope of the current versus voltage characteristic in a log-lin plot [i.e., $\delta = \frac{dV_{slc}}{dLog_{10}(I_{slc})}$].

$$I_{slc} = 2I_{ss} \cdot \sinh\left(\frac{V_{slc}}{\delta} \cdot \ln(10)\right) \tag{4.1}$$

Similarly, the I-V characteristic of threshold selectors can be modeled by separating two regions of operation. While the on-state ($V_{slc} > V_{th}$) shows an Ohmic relationship with a turn-on resistance, R_{on}, the off-state ($V_{slc} \leq V_{th}$) is reported to obey Poole-Frenkel (PF) conduction mechanism that can be stated as: $I = \kappa V \exp(\lambda\sqrt{V})$, where κ and λ are fitting parameters [5]. However, the off-state can also be approximated by a linear off-state resistance, R_{off}, and hence the I-V characteristic can be described as:

$$I_{slc} \approx \frac{V_{slc}}{R_{off}}, \; V_{slc} \leq Vth \tag{4.2a}$$

and in the on state:

$$I_{slc} = \frac{V_{slc}}{R_{on}}, \; V_{slc} > Vth \tag{4.2b}$$

For the sake of convenience when dealing with the design and the study of crosspoint arrays, some additional parameters can be introduced. The first parameter is selector threshold voltage, V_{th}. As already mentioned in Subsection 4.3.2, in threshold selectors, V_{th} is the voltage at which the selector device is turned on. An equivalent parameter can also be defined for exponential selectors. Accordingly, V_{th} in these selector devices is defined as the voltage across the selector that is required to draw a specific current (typically a reasonable fraction of the switching current of the memory element), referred to as threshold current, I_{th} (e.g. 1 μA), as shown in Fig. 4.4 . Another important design parameter is the selector voltage margin, V_m (see Fig. 4.4), defined as the selector bias voltage interval inside which the current is lower than a predetermined value, I_{lk}, that can be considered as the maximum acceptable leakage current in half-selected and unselected cells.

Since, for $|V_{slc}| \gg 0$ the magnitude of current in exponential selectors that was given by (4.1) can be approximated by

$$|I_{slc}| \approx I_{ss} \cdot 10^{\frac{|V_{slc}|}{\delta}} \tag{4.3a}$$

a relationship between V_{slc} and V_{th} can be easily derived:

$$V_{slc} = V_{th} + K_{th}^{sw} \cdot \delta \tag{4.3b}$$

where $K_{th}^{sw} = \log_{10}(\frac{I_{sw}}{I_{th}})$. For threshold selectors the voltage across the selector at higher cell current, i.e., $I_{slc} > I_{th}$, can be approximated by V_{th} , i.e., $V_{slc} \approx V_{th}$.

4.4 Model for Resistance-Switching Memory Elements

The memory elements are modeled as a variable resistor that is set to the LRS or the HRS by applying an electrical switching voltage and current with the required magnitude and polarity. The SET and the RESET operation generally require different magnitudes of switching voltage and current. Hence, we consider the maximum (or peak) switching voltages/currents corresponding to the worst-case write operation that imposes a stronger constraint on the feasible crosspoint array size. The worst-case write operation may be the SET or the RESET depending on which one of them requires a larger switching voltage and/or current. For example, in PCM the RESET operation can be considered as the worst-case scenario since both the RESET current and voltage are generally higher. In STT-MRAM, while SET and RESET voltages are roughly equal, the switching current required for RESET is generally larger than the corresponding SET current due to the inherent asymmetry of the MTJ device. Hence, the RESET operation defines the worst-case scenario for write operation. In RRAM, while the SET and RESET voltages are approximately constant, usually the SET voltage being higher, the SET and RESET current levels may change by orders of magnitude depending on the compliance current set to limit the size of the conductive filament during forming and SET operations [52]. In this chapter, the the worst-case write operation in RRAM is assumed to be the SET operation.

According to the literature, peak switching currents for RRAM may range from less than 10 μA to 100 μA and peak switching voltages range from 1 V to 3 V [8, 75, 76]. For STT-MRAM , peak switching currents from less than 10 μA up to 50 μA and voltages ranging from 0.4 V to 0.8 V were reported [8,77,78]. In PCM, the RESET current rangs from 100 μA to few mA, and the peak switching voltage ranges from 0.8 V to 1.8 V [8, 75, 79]. Fig. 4.5 shows typical I-V characteristics in RRAM, STT-MRAM, and PCM cells obtained by fitting published data. As we can see from the plots, in STT-MRAM and bipolar RRAM, SET and RESET occur at opposite polarities. In order to SET the cells, it is necessary to apply a voltage higher than a given positive switching voltage, V_{sw}^{+}, whereas to RESET the cells, it is necessary to apply a negative voltage whose magnitude is higher than the magnitude of $-V_{sw}^{-}$. In contrast, in PCM cells, SET and RESET are controlled by the magnitude, the shape, and the duration of the programming pulse, since the phase transition is controlled by the temperature inside the active material, and SET and RESET pulses can have the same polarity (but a different shape).

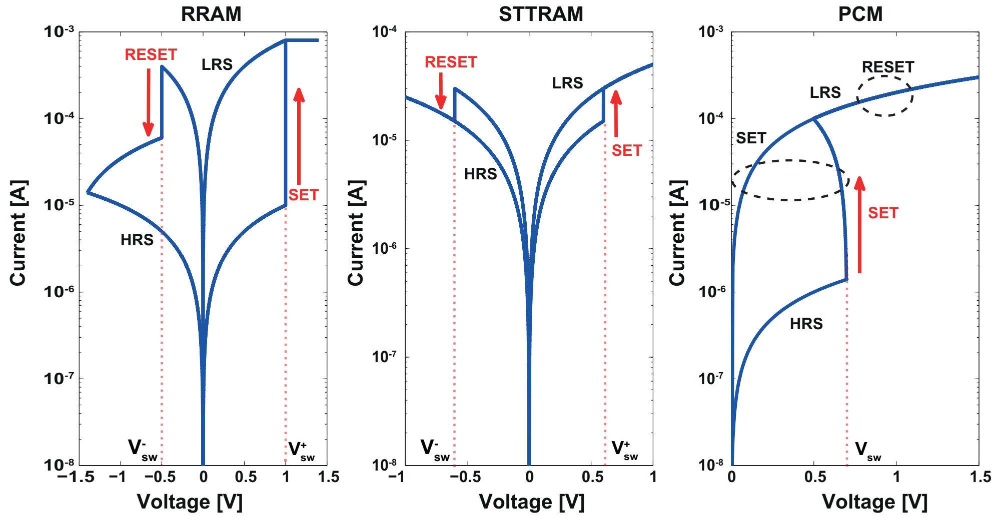


Figure 4.5: Typical I-V characteristics of bipolar RRAM, STTRAM, and PCM

It is worth mentioning that, in crosspoint RRAM arrays, the RRAM devices need to be subjected to an initial forming operation where a conductive filament is formed, and in fact, the required voltage for forming is often far higher than SET and RESET voltages [51, 52]. However, this chapter will consider only the SET and RESET operating voltages/currents with the assumption that the RRAM devices are already formed.

One possible solution to avoid the additional constraint on the crosspoint RRAM array that comes from the high forming operation could be forming all the RRAM devices before the array is used. This way, the write disturbance [i.e., in this particular case, a fraction of the forming voltage applied, and hence partially activating the half-selected and unselected memory cells while forming a targeted selected memory cell(s)] will not be a problem. Let us assume that the RRAM devices need 3 V forming voltage. Depending on the the specific biasing scheme employed, the forming voltage may generate a write disturbance voltage in the range of 1 V to 1.5 V [80]. If the partially activated cells were already formed, this amount of write disturbance voltage, will be similar to doing or attempting to do a SET operation on formed cells, which is not usually a problem. Also, if the disturbance voltage is applied to memory cells that are not formed yet (or virgin cells), usually such a small voltage (compared to the required forming voltage) will not affect these virgin cells. Leakage power consumption during forming will also not be a critical issue since forming is done only once in the lifetime of the RRAM devices.

4.5 Interconnection Metal Line Scaling

As the dimension of copper interconnection metal lines scale down to a regime that is comparable to the mean free path of copper, surface scattering and grain boundary scattering become substantial, thus resulting increase in size-dependent resistivity [72]. More specifically, resistivity increases with scaling down [81]. As a result, the parasitic resistance, and, thus, the in IR drop along WLs and BLs is increases with scaling down of metal lines. Figure 4.5 shows a schematics of a 2×2 1S1R crosspoint memory array is with detailed parameters of interconnection metal lines. Resistance per memory cell, denoted as R_c, for an interconnection metal line with effective resistivity ρ_{eff}, length $L = 2F$, width $W = F$, and metal aspect ratio AR (hence, height $H = AR \cdot F$) can be calculated using:

$$R_c = \rho_{eff} \frac{L}{W \cdot H} = \rho_{eff} \frac{2}{AR \cdot F}. \tag{4.4a}$$

A metal aspect ratio, AR =2 is recommended by ITRS for high-density memory interconnects [81]. Hence, (4.4a) can be re-written as (4.4b). Table 4.1 shows the values of ρ_{eff} and the corresponding R_c for some technology nodes [74, 81, 83]. In this work, a 22 nm technology node is assumed, and, hence $R_c = 2.5\ \Omega$.

$$R_c = \frac{\rho_{eff}}{F} \tag{4.4b}$$

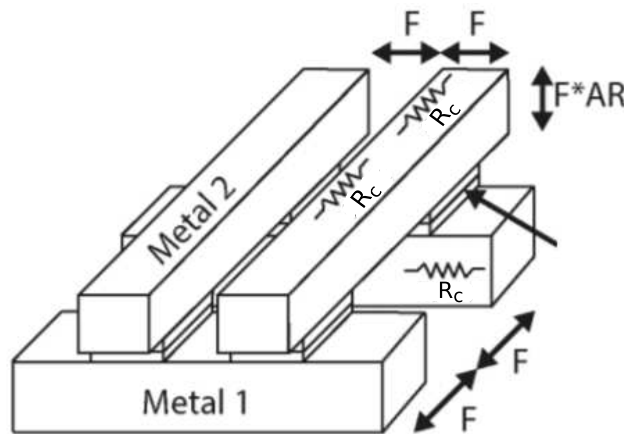


Figure 4.6: Physical schematics of a 2×2 crosspoint array with parameters of interconnection metal lines (source: [82])

Table 4.1: Example of effective resistivity of metal lines in different technology nodes

Half pitch, F [nm]	ρ_{eff} [$\mu\Omega.cm$]	R_c [Ω]
40	4.0	1
28	4.8	1.7
22	5.5	2.5
10	9.4	9.4

4.6 Model for Large 1S1R Crosspoint Arrays

The simulation or computation time of a full-array model of large crosspoint arrays with millions of memory cells can be prohibitively long. However, with some reasonable assumptions we can come up with a simplified computationally efficient model suited for making simulations of large size arrays. Firstly, since the current that flows through unselected WLs and BLs in a well-designed crosspoint array is generally small, the IR drop along these lines can be neglected. As a result, when a certain memory cell is selected for write/read, the unselected cells in the array can be replaced by a single aggregate 1S1R cell, which can draw a sneak (or leakage) current that is equal to the sum of all the individual leakage currents through the replaced unselected memory cells. Besides, we can assume that all the half-selected WL and BL cells are directly connected to the unselected BL and WL bias voltages, V_{uBL} and V_{uWL}, respectively, again because the IR drop along these lines is negligible. With these approximations, we obtain a simplified model for a certain write/read operation as indicated in Figure 4.7, which have been also used in some previous works [69]. The approach reduces the total number of cells included in the model, and hence, significantly speeds up simulations while still preserving accuracy.

The schematic shown in Figure 4.7 represents the worst-case selected cell location for writing operation: the selected cell is located at the array corner farthest from WL and BL bias edges, which translates into the longest selected path and hence, into the highest write current/voltage degradation. Indeed, the longest selected path results in the highest IR drop due to the write current flowing through the path, and the leakage current in WL half-selected cells diminishes the amount of write current arriving at the selected cell.

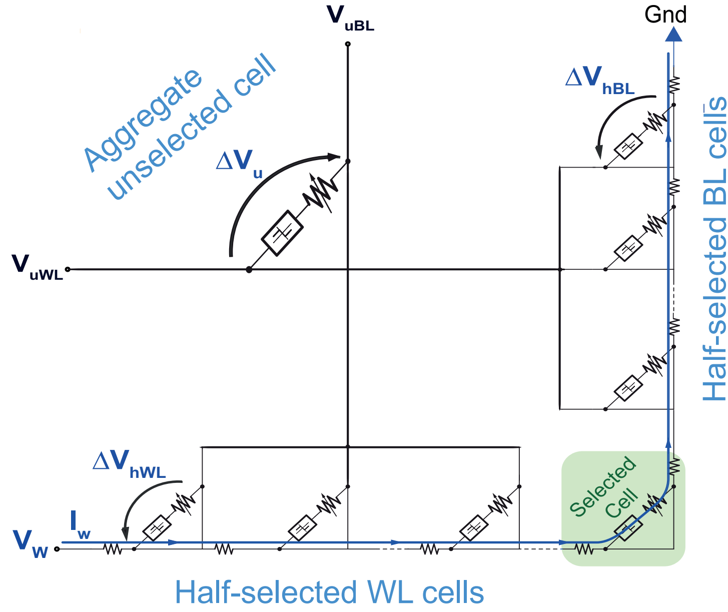


Figure 4.7: Simplified circuit model of a crosspoint array when programming a memory cell at the lower right corner

The overall worst-case scenario for write operation takes place when the worst-case selected cell location is combined with two additional operating worst-case conditions, i.e. when half-selected and unselected cells are in the the low resistance state (LRS), which corresponds to highest sneak current, and when the write operation requires high cell switching voltage and/or current. The higher the cell switching voltage, the higher the voltage that needs to be applied at the the selected WL edge and thus, the higher the write disturbance voltage across WL half-selected cells. Similarly, high cell switching current results in an increase in the IR drop along the selected path. The maximum feasible crosspoint array size, which is discussed in the next two sections, is determined by considering this worst-case scenario.

In particular, for the above worst-case scenario, we can make an additional reasonable assumption. Consider a half-selected or unselected 1S1R memory cell, in which the memory element is in LRS (worst-case) and the voltage across the 1S1R pair is low [a fraction of the write voltage (V_W)] to keep the selector in its off state. In such scenario, the resistance of a selector device (at low bias) is much higher than the resistance of the series-connected memory element (at LRS state). Hence, an unselected or a half-selected 1S1R cell can be effectively approximated by just the selector device (1S); this means, in such cells the voltage bias (or resistance) of the 1S1R pair can approximated by the voltage (or resistance) of just the selector device.

4.7 Design considerations for Write Operation in Crosspoint Arrays

Implementing a crosspoint array with given memory element, selector, and interconnection metal line technologies requires determining the feasible array size and choosing the suitable bias scheme. Write operation on the selected memory cell is performed by applying a write voltage V_W (or equivalently a current I_W) to the selected WL, as indicated in Fig. 4.7. The value of V_W (set to bias the array or the voltage at the WL edge as a result of I_W) depends on the voltage across the selected memory cell and the voltage drop along the selected path, which is a function of I_W, array size, and interconnection metal line resistance. The voltage drop across WL and BL half-selected cells and, thus, the leakage current through these cells depends on V_{uBL} and V_{uWL}, respectively. Therefore, these two bias voltages should be carefully chosen so as to avoid excessive leakage and write disturbance in half-selected cells. In this section, a simplified analysis of the requirements for write operation is presented followed by a discussion of the write considerations for designing practical size 1S1R crosspoint arrays.

4.7.1 Simplified Analysis of Boundary Conditions

In a proper write operation, the selected memory cell should be programmed successfully while the bits stored in all the half-selected and unselected memory cells are unaffected. Before going to practical crosspoint array design issues such as limited nonlinearity of selector device and IR drop along interconnection metal lines, the minimum requirements (or boundary conditions) for performing the write operation are analyzed. The analysis is aimed at studying the ultimate constraints that arise from the crosspoint architecture itself, which is often overlooked in the literature.

For the purpose of this analysis, let us assume that the selector device has a very high nonlinearity and current drive capability such that below its threshold voltage, V_{th}, the device is turned-off so as to completely suppress sneak current paths and above V_{th} the device is turned-on to provide any high current required for programming the memory element. Let us also assume that the IR drop along BLs and WLs is negligible and that the total leakage power (due to sneak current paths) is not an issue. This way, the WL and BL bias voltages are set under the only constraints of successfully programming the selected cell and avoiding undesired writing of half-selected and unselected cells. Such analysis will enable to determine the boundary conditions for write operation before going to practical design considerations.

On the one hand, for successful writing of the selected memory cell, the voltage delivered to cell must be large enough to turn-on the selector device and the memory element. Hence, the applied write voltage, V_W, must be at least equal to the sum of V_{th} and V_{sw}:

$$V_W > V_{th} + V_{sw} \tag{4.5}$$

On the other hand, to avoid undesired writing of half-selected WL cells, half-selected BL cells and unselected cells, the respective voltages across these cells, ΔV_{hWL}, ΔV_{hBL}, and ΔV_u, respectively, should be less than V_{th}. These voltages are set by the bias voltages of selected and unselected bitlines and wordlines and can be expressed as follows: $\Delta V_{hWL} = V_W - V_{uBL}$, $\Delta V_{hBL} = V_{uWL}$, and $\Delta V_u = V_{uBL} - V_{uWL}$. The required conditions are summarized as:

$$\Delta V_{hBL} = V_{uWL} < V_{th}. \tag{4.6a}$$

$$\Delta V_u = \ V_{uBL} - V_{uWL} < V_{th}. \tag{4.6b}$$

$$\Delta V_{hWL} = V_W - V_{uBL} < V_{th}. \tag{4.6c}$$

From the above equations, we can also easily derive a guideline for setting the biasing voltages of selected and unselected WLs and BLs:

$$V_{uWL} < V_{th} \tag{4.7a}$$

$$V_{uBL} = \Delta V_u + V_{uWL} < 2V_{th} \tag{4.7b}$$

$$V_W = V_{uBL} + \Delta V_{hWL} < 3V_{th} \tag{4.7c}$$

The above equations can be re-arranged to determine the minimum requirements on bias voltages, selector threshold voltage V_{th} and switching of memory element V_{sw}. For instance, the acceptable range for write voltage, V_W, can be obtained by combining (4.5) and (4.7c):

$$(V_{sw} + V_{th}) < V_W < 3V_{th} \tag{4.8}$$

A variability aware-analysis on the voltage compatibility of memory element and selector device is also presented in Section 4.9, based on the above equations and some additional considerations for read operation.

In practical large-size crosspoint arrays with millions of unselected cells and thousands of half-selected cells and implemented with practical selector devices, the leakage power will be excessively large if ΔV_{hWL} and ΔV_{hBL} and ΔV_u are not sufficiently lower than V_{th}. Hence, in the study of the design constraints in practical crosspoint arrays, voltage margin (V_m) that is lower than V_{th} and defined at an acceptable leakage current level, is used as maximum value of the voltage across the half-selected cells. Generally, crosspoint array design requires determining the maximum size of array that can be safely implemented and carefully biasing selected and unselected WLs and BLs under various write and read constraints. A detailed analysis of these constraints in practically large-size crosspoint arrays is provided in Subsection 4.7.2 and Section 4.8.

4.7.2 Write Requirements in Practical-Size Arrays

In practical 1S1R crosspoint memory arrays, the bias voltage applied at the edge of the selected WL during writing, V_W, should satisfy the following condition:

$$V_{sw} + V_{slc} + \Delta V_{line} \leq V_W \leq \widehat{V}_W \tag{4.9}$$

where ΔV_{line} is the total IR drop along the selected WL and BL, V_{sw} is the switching voltage of the memory element, V_{slc} is the voltage across the selector device during the write operation, and $\widehat{V}_W$ is the maximum acceptable value for V_W. In the ideal case of negligible leakage through half-selected and unselected cells, the amount of current is constant throughout the whole selected path and in this case, the IR drop during writing of a cell can be calculated as $I_{sw} \cdot R_c \cdot N_{tot}$, where R_c, is the interconnection metal line resistance between two adjacent cells as discussed in Section 4.5 and N_{tot} is the total number of cells in the selected path (e.g., assuming having selected the $i-$th BL and the $j-$th WL, $N_{tot} = i + j$). In the presence of non-negligible leakage currents, the current along the selected path varies along the path; in the schematics shown in Fig. 4.7, it decreases from I_W at the selected WL terminal to I_{sw} at the selected cell and also it increases from I_{sw} to some higher amount along the selected BL. This makes the accurate calculation of the IR drop challenging for mathematical analyses. However, the IR drop can be estimated by introducing a correction factor, $\gamma > 1$, to take into account the effect of leakage currents, and hence ΔV_{line} is calculated using the following equation:

$$\Delta V_{line} = (\gamma \cdot I_{sw}) \cdot R_c \cdot N_{tot} \tag{4.10}$$

where the value of γ is determined from circuit simulations.

Since V_W can be generally expressed as $\Delta V_{hBL} + \Delta V_u + \Delta V_{hWL}$, the maximum allowed write voltage, $\widehat{V}_W$, can also be expressed in terms of the maximum acceptable voltages across half-selected and unselected cells, denoted as $\Delta\widehat{V}_{hBL}$, $\Delta\widehat{V}_{hWL}$, and $\Delta\widehat{V}_u$:

$$\widehat{V}_W = \Delta\widehat{V}_{hBL} + \Delta\widehat{V}_u + \Delta\widehat{V}_{hWL} \tag{4.11}$$

By substituting (4.3b), (4.10), and (4.11) into (4.9), the condition to be satisfied in a crosspoint array for writing operation can be determined as:

$$\begin{aligned} V_{sw} + (V_{th} + K_{th}^{sw} \cdot \delta) + \gamma I_{sw} \cdot N_{tot} \cdot R_c \leq V_W \\ \leq \Delta\widehat{V}_{hBL} + \Delta\widehat{V}_{hWL} + \Delta\widehat{V}_u \end{aligned} \tag{4.12}$$

The values of ΔV_{hBL}, ΔV_{hWL}, and ΔV_u are set by properly biasing the selected and unselected WLs and BLs. In this respect, there are two commonly used biasing schemes, generally referred to as $\frac{1}{2}$ and $\frac{1}{3}$ schemes [16]. In the $\frac{1}{2}$ biasing scheme, both V_{uWL} and V_{uBL} are set to $\frac{V_W}{2}$, thus $\Delta V_{hBL} = \Delta V_{hWL} = \frac{V_W}{2}$ and $\Delta V_u = 0$. In the $\frac{1}{3}$ biasing scheme, V_{uWL} is set to $\frac{V_W}{3}$ and V_{uBL} is set to $\frac{2V_W}{3}$, thus $\Delta V_{hBL} = \Delta V_{hWL} = \Delta V_u = \frac{V_W}{3}$.

The allowed values of the biasing voltages are constrained by leakage current, which is determined by the characteristic of the selector device. In particular, similar to the conditions given by 4.6a, 4.6c, and 4.6c, for large-size crosspoint memory arrays, ΔV_{hBL}, ΔV_{hWL} and ΔV_u should be less than V_m of the selector device, which is corresponds to a predetermined maximum acceptable leakage current. In Section 4.10 of this chapter, it is presented a guideline for biasing 1S1R crosspoint memory arrays to minimize total worst-case leakage power consumption. More specifically, a generic x biasing scheme, in which V_{uWL} is set to $x \cdot V_W$ and V_{uBL} is set to $(1-x) \cdot V_W$, where $\frac{1}{3} \leq x \leq \frac{1}{2}$ is proposed. Therefore, instead of using the conventional $x = \frac{1}{2}$ or $x = \frac{1}{3}$ bias schemes, the value of x can be set to an optimal value that minimizes total leakage power consumption. It is demonstrated that for array size in the order of 1000×1000 and a selector device with nonlinearity $\delta = 0.1$ V/decade (in its sub-threshold region), the total leakage power consumption is minimized when $x \approx \frac{2}{5}$, hence can be referred to as $\frac{2}{5}$ biasing scheme. In terms of V_m of the selector device, ΔV_{hBL}, ΔV_{hWL} are set to V_m and ΔV_u is set to $\frac{V_m}{2}$ by biasing the selected WL to the write voltage $V_W = \frac{5V_m}{2}$, $V_{uWL} = V_m$ and $V_{uBL} = \frac{3V_m}{2}$. This biasing scheme is used in the next analysis of write and read requirements.

From equation (4.3a), the voltage margin V_m and threshold voltage V_{th} of a selector device can be related by the following equation:

$$V_m = V_{th} - K_{lk}^{th} \cdot \delta \tag{4.13}$$

where, $K_{lk}^{th} = \log_{10}(\frac{I_{th}}{I_{lk}})$. The leakage current (per cell), I_{lk}, is set to 10 nA. Hence, (4.12) can be re-written as

$$V_{sw} + (V_{th} + K_{th}^{sw} \cdot \delta) + \gamma I_{sw} \cdot N_{tot} \cdot R_c \leq 2.5\left(V_{th} - K_{lk}^{th}\delta\right) \tag{4.14a}$$

and, by re-arranging the terms, we obtain

$$V_{sw} + \gamma I_{sw} \cdot N_{tot} \cdot R_c \leq 1.5V_{th} - \left(K_{th}^{sw} + 2.5K_{lk}^{th}\right) \cdot \delta \tag{4.14b}$$

The maximum number of cells along the selected path can be determined by equating the right- and the left-hand sides of (4.14b). For square crosspoint arrays, there are equal number of cells along the selected WL and BL (i.e., $N_{BL} = N_{WL} = N_W = \frac{N_{tot}}{2}$), and, hence, the feasible square crosspoint array size that meets the minimum write requirements can be obtained as:

$$N_W \leq \frac{1}{2}\left[\frac{1.5V_{th} - \left(K_{th}^{sw} + 2.5K_{lk}^{th}\right) \cdot \delta - V_{sw}}{\gamma \cdot I_{sw} \cdot R_c}\right] \tag{4.15}$$

For threshold selectors, the maximum voltage across the selector device, V_{slc} during programming the memory cell can be approximated with its threshold voltage. In these selectors, the transition from the off to the on region of the selector device operation occurs abruptly at the threshold voltage V_{th}. Hence, the low-leakage voltage margin for biasing half-selected cells can also be assumed to be equal to its threshold voltage, i.e., $V_m = V_{th}$ and hence, ΔV_{hBL} and ΔV_{hWL} can be set to V_{th}. For the unselected cells, following the same approach followed for exponential selectors we could bias them to $\frac{V_{th}}{2}$. However, biasing to $\frac{V_{th}}{2}$ instead of V_{th} reduces the leakage current only by a factor of 2 unlike exponential selectors at the expense of loosing some margin. Hence, ΔV_u is also set to V_{th}. While in crosspoint arrays built with exponential selectors one can exponentially reduce the leakage current by decreasing the bias voltage, in the case of threshold selectors the leakage current reduces only linearly with the voltage. The only way to reduce leakage in threshold selectors is by using a selector device with a high off-resistance as given by

$$I_{th} \approx \frac{V_{th}}{R_{off}} \tag{4.16}$$

With the above considerations for the selected, half-selected and unselected cells, the required condition for write operation in crosspoint arrays built with threshold selectors can be obtained:

$$V_{sw} + V_{th} + \gamma I_{sw} \cdot N_{tot} \cdot R_c < 3V_{th} \tag{4.17}$$

Hence, the maximum feasible array size can be determined.

$$N_W \leq \frac{1}{2} \left[\frac{2V_{th} - V_{sw}}{\gamma \cdot I_{sw} \cdot R_c} \right] \tag{4.18}$$

Fig. 4.8 shows the maximum size of a square crosspoint memory array calculated using (4.15) for crosspoint arrays implemented with exponential selectors as a function of the voltage margin, V_m, of the selector device. The selector turn-on slope δ was set to 100 mV/dec, the leakage current I_{lk} was set to 10 nA and the voltage margin was varied V_m from 0 to 3 V. The result (Fig. 4.8) illustrates how the maximum feasible array size in the considered memory technologies is constrained by voltage margin (or threshold voltage) of the selector device. The figure also shows that the higher the voltage margin, the higher the feasible array size. This result (Fig. 4.8) also shows that there is a minimum voltage margin requirement for each memory technology to implement even very small-size crosspoint arrays. In comparison, PCM requires a selector device with higher voltage margin than STT-MRAM and RRAM while STT-MRAM has the lowest requirements. A similar trend is observed for crosspoint arrays implemented with threshold selectors as illustrated in Fig. 4.9, which shows the dependence of the maximum feasible array size on the threshold voltage of the selector device. For this plot, the off-state resistance of the selector device, R_{off}, was varied with V_{th} (see Fig. 4.10) to maintain a constant leakage current.

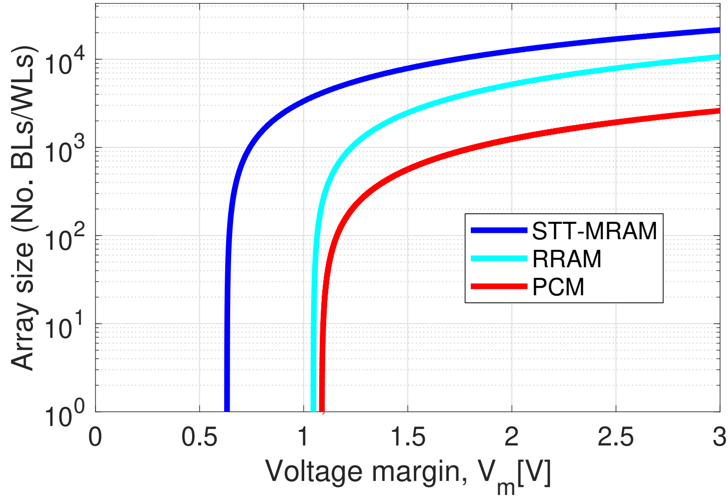


Figure 4.8: Dependence of feasible crosspoint array size on low-leakage voltage margin, V_m, of exponential selector device (δ = 100 mV/dec): the cases of STT-MRAM (I_{sw} = 30 μA, V_{sw} = 0.6 V), RRAM (I_{sw} = 50 μA, V_{sw} = 1.2 V), and PCM (I_{sw} = 200 μA, V_{sw} = 1.2 V) arrays as considered

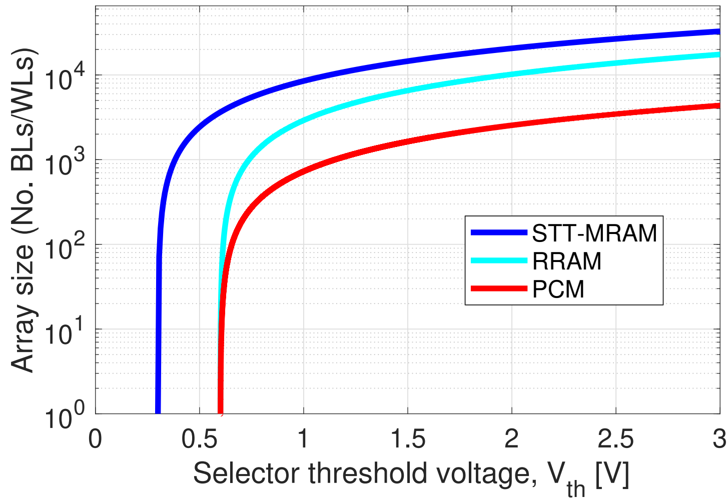


Figure 4.9: Maximum feasible size of square crosspoint array built with threshold selector versus threshold voltage of selector (for STT-MRAM: I_{sw} = 30 μA, V_{sw} = 0.6 V, RRAM: I_{sw} = 30 μA, V_{sw} = 1.2 V, and PCM: I_{sw} = 200 μA, V_{sw} = 1.2 V)

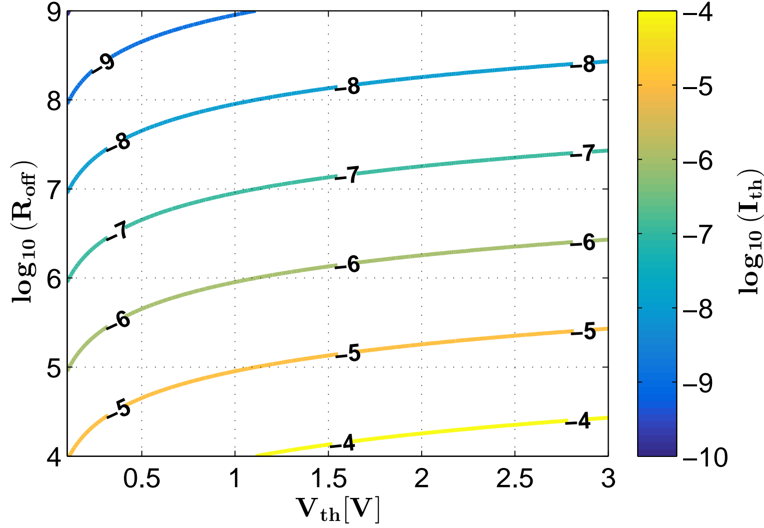


Figure 4.10: Off-state resistance versus leakage current in threshold selector devices

Fig. 4.11 shows the dependence of the maximum feasible array size on the nonlinearity of an exponential selector device. In this case, δ was varied from 5 mV/dec to 300 mV/dec and V_{th} was set to 1.5 V while V_m was calculated for each value of δ using (4.13). Here (Fig. 4.11), we can see the maximum feasible array sizes for the cases of STT-MRAM, RRAM and PCM. Clearly, the higher the nonlinearity (i.e., the smaller the slope δ), the higher the feasible array size and also there is a minimum nonlinearity (i.e., a maximum slope δ) requirement to implement even very small size arrays.

In addition to the characteristics of the selector device, maximum feasible array size heavily depends on the characteristics of the memory element. As demonstrated in Figs. 4.12 and 4.13, for a given selector, the higher the switching current and/or the switching voltage of memory element, the smaller the feasible array size.

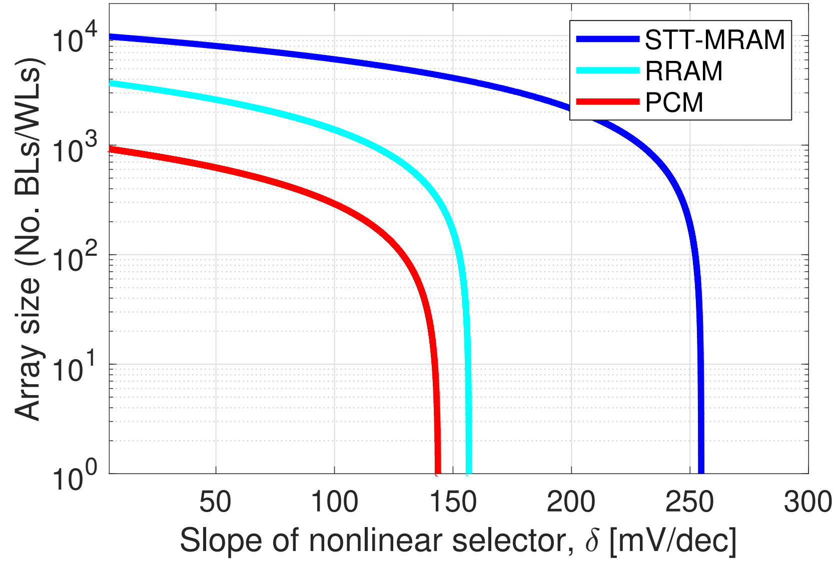


Figure 4.11: Maximum feasible square crosspoint array size versus nonlinearity of selector device (for STT-MRAM: $I_{sw} = 30\ \mu$A, $V_{sw} = 0.6$ V, RRAM: $I_{sw} = 50\ \mu$A, $V_{sw} = 1.2$ V, and PCM: $I_{sw} = 200\ \mu$A, $V_{sw} = 1.2$ V; threshold voltage, V_{th}, of selector device was set to 1.5 V in all cases.)

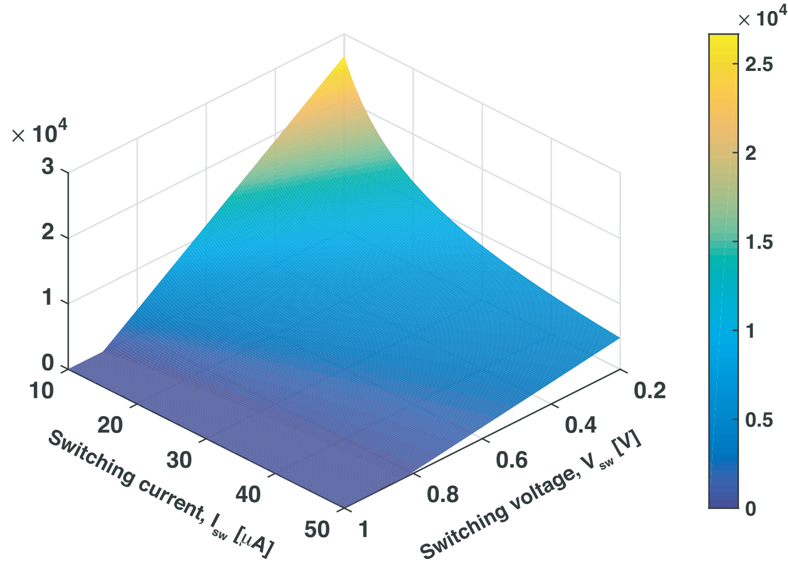


Figure 4.12: Dependence of maximum array size of on I_{sw} and V_{sw} of memory element. The selector slope and threshold voltage was set to $\delta = 100$ mV/dec and $V_{th} = 1$ V (and, hence, $V_m = 0.8$ V), respectively

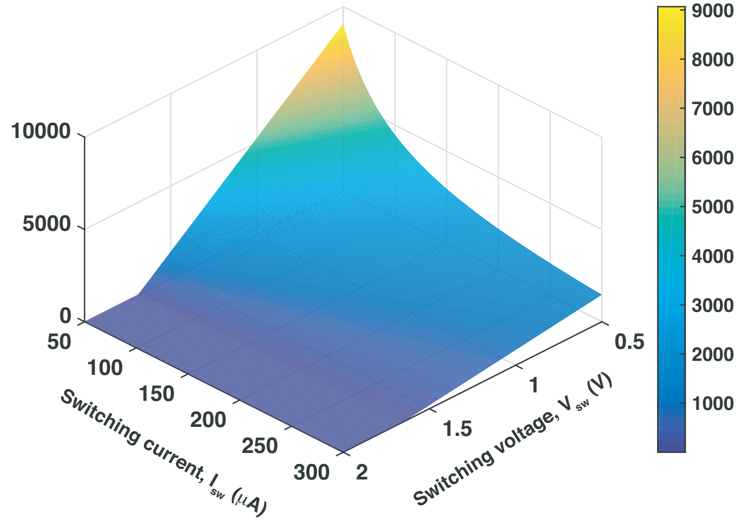


Figure 4.13: Dependence of maximum array size of on I_{sw} and V_{sw} of memory element (selector parameters: δ = 50 mV/dec, V_{th} = 1.35 V. The result is applicable for memory elements with high I_{sw} and V_{sw}

Parasitic resistance of metal lines is another issue which plays a key role and was investigated in this thesis work. In fact, the resistance of interconnections increases as technology scales down to more advanced nodes. The results of this analysis are shown in Fig. 4.14.

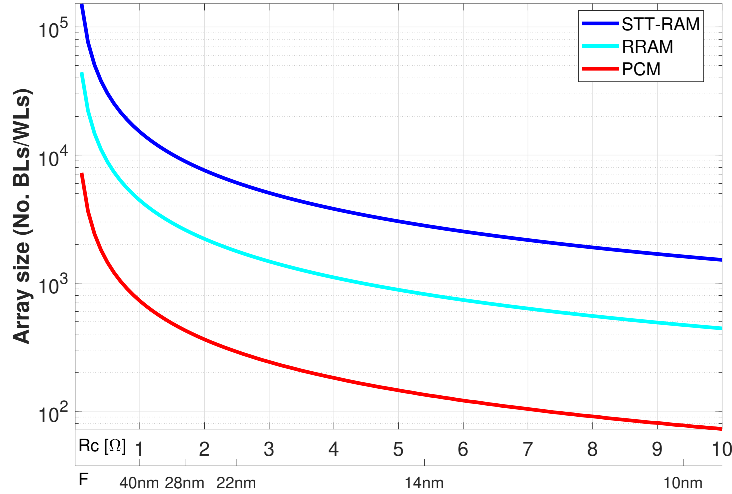


Figure 4.14: Impact of of interconnection metal line scaling on array size for the cases of STT-MRAM (I_{sw} = 30 μA, V_{sw} = 0.6 V), RRAM (I_{sw} = 30 μA, V_{sw} = 1.2 V), and PCM (I_{sw} = 200 μA, V_{sw} = 1.2 V) crosspoint arrays. Other parameters: δ = 100 mV/dec, V_{th} = 1.5 V (or V_m = 1.3 V)

4.8 Design considerations for Read Operation in Crosspoint Arrays

Let us assume a read operation carried out by injecting a read current, I_R, into the selected WL, connecting the selected BL to ground (or vice versa), and sensing the ensuing voltage, V_R, at the WL terminal. Depending on the state (i.e., HRS or LRS) stored in the cell being read, a high ($V_{R,H}$) or a low voltage ($V_{R,L}$) is sensed. The two resistance states can be differentiated as long as $V_{R,H}$ is sufficiently higher than $V_{R,L}$. For example, a comparator with a fixed reference voltage, V_{ref}, can be used as long as $V_{R,L} < V_{ref} < V_{R,H}$: if V_R is higher than V_{ref}, HRS is detected otherwise LRS is detected. In this case, the sense margins for the HRS and the LRS, respectively, can be defined as $S_{M,H} = V_{R,H}$ - V_{ref} and $S_{M,L} = V_{ref}$ - $V_{R,L}$, respectively. A variety of sensing schemes are available for different memory technologies, and the choice of V_{ref} and, hence, the sense margins may vary depending on the specific implementation of the sensing scheme. Nevertheless, irrespective of the particular scheme employed, a generic analysis can be done by considering the total available sense margin, S_M, defined as $S_M = V_{R,H} - V_{R,L}$. The read voltage, V_R, can be expressed analytically using (4.9) and (4.10), where γ is set to unity since the read current and, hence, the contribution of leakage currents to ΔV_{line} is relatively small. To be able to use a single reference voltage for the whole array block, we should consider the worst-case scenarios with respect cell locations, which take place when $V_{R,H}$ is minimum (cell at the upper left corner) and $V_{R,L}$ is maximum (cell at the lower right corner). With this consideration, the total worst-case sense margin, denoted as $\widehat{S}_M$, can be obtained as

$$\widehat{S}_M = I_R \cdot [(R_H - R_L) - N_{tot} \cdot R_c] \tag{4.19}$$

Hence, from above equation, the feasible size of a square crosspoint array, N_R, that meets the minimum sense margin (depending on the sensitivity of the sense amplifier), $\widehat{S}_{M,min}$, can be calculated as:

$$N_R \leq \frac{1}{2} \cdot \left[\frac{\frac{R_H}{R_L} - 1}{\frac{R_c}{R_L}} - \frac{\widehat{S}_{M,min}}{I_R \cdot R_c} \right] \tag{4.20}$$

Figures 4.15 and 4.16 indicate the dependence of the sensing (or read) margin upon the crosspoint array size and the resistance ratio of HRS to LRS, $\frac{R_H}{R_L}$, as calculated from (4.19). The color bar indicates the total available sense margin, $\widehat{S}_M$, in mV. The read current used for this simulation was 10 μA and R_L was set to 10 kΩ. If a different read current is used, the sense margin scales proportionally as per (4.19). In particular, Figure 4.15 shows the analysis results for memory elements with a small $\frac{R_H}{R_L}$. For example, in STT-MRAM cell, $\frac{R_H}{R_L}$ is as small as 2, which makes it very difficult to obtain sufficient sense margin even in the case of small array sizes. As we can see from the plot, for $\frac{R_H}{R_L} = 2$ and array size of $1Kb \times 1Kb$, the total available sensing margin is very small (less than 20 mV). This demonstrates that realization of a $1Kb \times 1Kb$ crosspoint STT-MRAM array is more challenged by read requirements more than write requirements. This illustrates that in memory cells with low $\frac{R_H}{R_L}$, like STT-MRAM, the read requirements, not the write requirements, set the most stringent constraints on crosspoint array size. In contrast, in memory technologies with high $\frac{R_H}{R_L}$ (e.g. PCM cells), large size arrays can be implemented while still obtaining sufficient sense margin (Figure. 4.16). Hence, the feasible array size of these memory technologies is mainly constrained by write requirements, which is consistent with the literature.

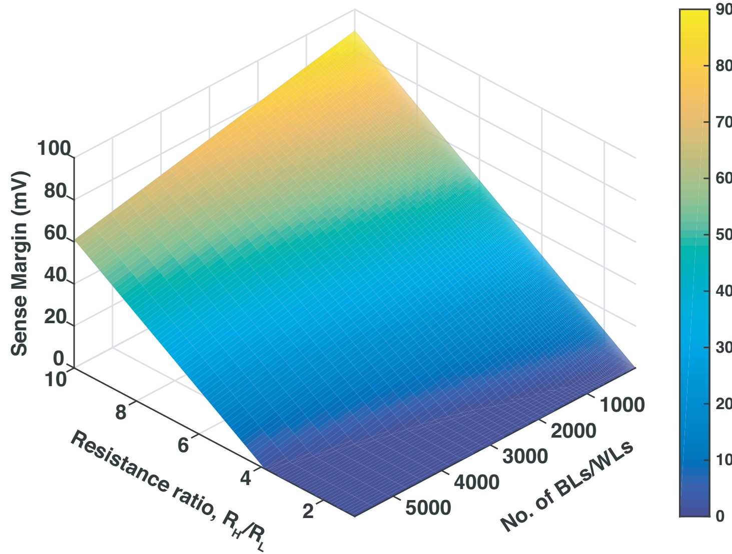


Figure 4.15: Total sensing margin (using 10 μA read current), as a function of crosspoint array size and memory element resistance ratio, $\frac{R_H}{R_L}$. The plot shows the sensing margin for the case of memory elements with a small $\frac{R_H}{R_L}$, for example STT-MRAM cells

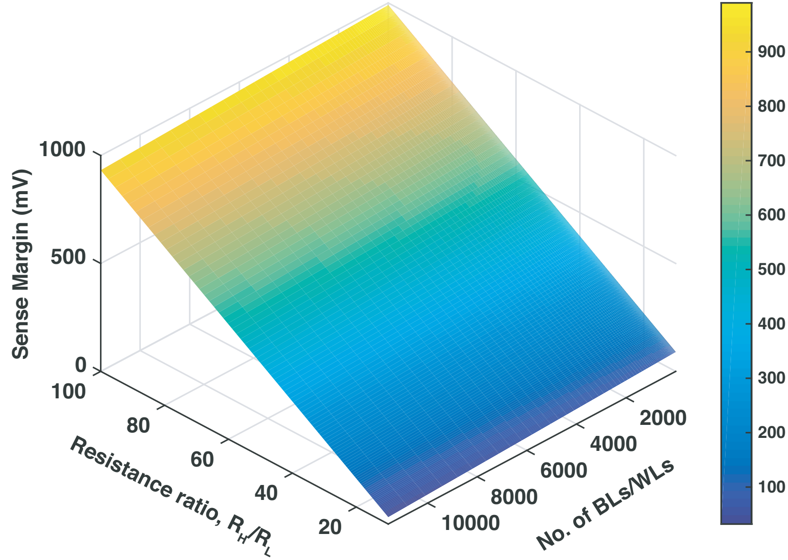


Figure 4.16: Total sensing margin (using 10 μA read current), as a function of array size and cell state resistance ratio, $\frac{R_H}{R_L}$. The plot shows the sensing margin for the case of memory elements with high $\frac{R_H}{R_L}$, for example in PCM and RRAM cells

4.9 A Variability-Aware Analysis of the Voltage Compatibility of Memory Element and Selector Device

Equation (4.8) is a useful guide that shows the allowed range of the write voltage provided from the driver, V_W, for a given selector device with threshold voltage, V_{th} and memory element with switching voltage, V_{sw}. In addition to defining the range for V_W, (4.8) shows the relationship between V_{th} and V_{sw}, hence, determining the voltage compatibility between selector device and memory element. By rearranging the terms of in the equation, we obtain

$$V_{th} \geq \frac{1}{2} V_{sw} \tag{4.21}$$

which defines the requirement on threshold voltage, V_{th}, of the selector device for a given memory element with a given switching voltage, V_{sw}, or vice versa.

Let us now consider device-to-device variations on the values of V_{th} and V_{sw}. In the worst-case scenario for the write operation, we should take into account the maximum values of V_{th} and V_{sw} for successful writing [i.e., for the condition stated in (4.5)], and the the minimum value of V_{th} for leakage through half-selected and unselected cells (i.e., conditions (4.6a) - (4.6b)). By denoting the nominal values of V_{sw} and V_{th} as $\overline{V}_{sw}$ and $\overline{V}_{th}$, respectively, we can express the maximum and minimum values in each case. The maximum and minimum values of the switching voltage are $V_{sw,max} = (1 + \alpha_{V_{sw}}) \cdot \overline{V}_{sw}$ and $V_{sw,min} = (1 - \alpha_{V_{sw}}) \cdot \overline{V}_{sw}$, respectively. Similarly, $V_{th,max} = (1 + \alpha_{V_{th}}) \cdot \overline{V}_{th}$ and $V_{th,min} = (1 - \alpha_{V_{th}}) \cdot \overline{V}_{th}$. Here, $\alpha_{V_{sw}}$ and $\alpha_{V_{th}}$ are equal to $n\sigma_{V_{sw}}/\overline{V}_{sw}$ and $n\sigma_{V_{th}}/\overline{V}_{th}$, respectively, where $\sigma_{V_{th}}$ and $\sigma_{V_{sw}}$ represent the relative standard deviations of V_{th} and V_{sw} distributions, respectively. As a result, (4.21), can be re-written for the worst-case variations as:

$$\overline{V}_{th} \geq \frac{1}{2} \cdot \left(\frac{1 + \alpha_{V_{sw}}}{1 - 2 \cdot \alpha_{V_{th}}} \right) \cdot \overline{V_{sw}} \tag{4.22}$$

which, clearly, sets a more stringent constraints on the pairing of a memory element and a selector device than (4.21) .

Memory read requirement also sets some constraints on the appropriate pairing of selector device and memory element. On the one hand, if we assume, for example, a current-mode read operation (where a read voltage, V_R, is applied to the WL/BL and the resulting current through the selected cell is compared to a reference current), V_R, should be high enough to turn on the addressed selector device but low enough not to switch the cell being read. Analytically, this condition can be stated as

$$V_R \leq V_{th} + V_{safe} \tag{4.23}$$

where V_{safe} is the voltage across the memory element during reading, which must be sufficiently smaller than V_{sw} in order to avoid undesired switching of the cell to be read. Without loss of generality we can express V_{safe} as a fraction of the minimum switching voltage as:

$$V_{safe} = \beta \cdot \overline{V}_{sw} \cdot (1 - \alpha_{V_{sw}}) \tag{4.24}$$

where β is a safety factor less than unity. In other words, V_{safe} must be set considering the minimum switching voltage and some additional safety margin. Besides, the possibility of read disturb caused by cumulative voltage stress due to repeated read events should be taken into consideration. As the selector device and the memory element are connected in series, for a given V_R, the voltage drop across the memory element increases when the voltage drop across the selector decreases.

Hence, the minimum value of selector threshold voltage, $V_{th,min}$, should be considered while determining the upper boundary value of V_R (4.23). On the other hand, V_R must be set larger than the highest value of the selector threshold voltage, $V_{th,max}$ to ensure that the device is turned on to provide a detectable read current. By combining the aforementioned considerations, the suitable range for V_R can be expressed as:

$$(1 + \alpha_{vth}) \cdot \overline{V}_{th} \leq V_R \leq (1 - \alpha_{V_{th}}) \cdot \overline{V}_{th} + V_{safe} \tag{4.25}$$

which, by using (4.24) and rearranging terms, gives the lower boundary value of $\overline{V}_{th}$ as a function of $\overline{V}_{sw}$:

$$\overline{V}_{th} \leq \frac{1}{2} \cdot \frac{\beta(1 - \alpha_{V_{sw}})}{\alpha_{Vth}} \cdot \overline{V}_{sw}. \tag{4.26}$$

By combining (4.22) and (4.26) the acceptable range for the ratio of the nominal threshold voltage of the selector, $\overline{V}_{th}$, to the switching voltage of the memory element, $\overline{V}_{sw}$, can be defined as:

$$\frac{1}{2} \cdot \left(\frac{1 + \alpha_{V_{sw}}}{1 - 2 \cdot \alpha_{V_{th}}} \right) \leq \frac{\overline{V}_{th}}{\overline{V}_{sw}} \leq \frac{1}{2} \cdot \left(\frac{\beta(1 - \alpha_{V_{sw}})}{\alpha_{Vth}} \right) \tag{4.27}$$

It is worth reminding that in practical large arrays, V_W has to be boosted from the stated ideal minimum value (i.e., $V_{sw} + V_{th}$) to compensate for the IR drop along the selected WL and BL. Besides, in arrays with millions of unselected and thousands of half-selected cells, the leakage power will be excessively large unless the voltage across those cells is kept much lower than V_{th}. Hence, the maximum allowed value of V_W is significantly less than the $3V_{th}$ ideal maximum value stated in (4.8). As a result, the actual design space gets narrower. As an illustration, Figure 4.17 shows result of a circuit simulation of a crosspoint array with 256×256 memory cells implemented with a selector device with V_{th} =1.1 V and memory element with $V_{sw} = 0.5$ V. The write voltage V_W was swept from 0.5 V to 5.5 V and the bias voltages for unselected WLs and BLs were set to $\frac{2}{5} \cdot V_W$ and $\frac{3}{5} \cdot V_W$, respectively. The current through selected, half-selected and unselected cells is plotted as in Figure 4.17. The upper sub-plot shows the current through the selected cells (blue curve) and a the current through a single half-selected cell (red curve) located at the lower left corner of the array (Figure 4.2b). The lower sub-plot shows the total leakage through all unselected and half-selected memory cells. We can see that in agreement with our analytical model (4.8), the acceptable value of V_W should be higher than $V_{sw} + V_{th}$ for successful writing but lower than the $3V_{th}$ to avoid excessive leakage power.

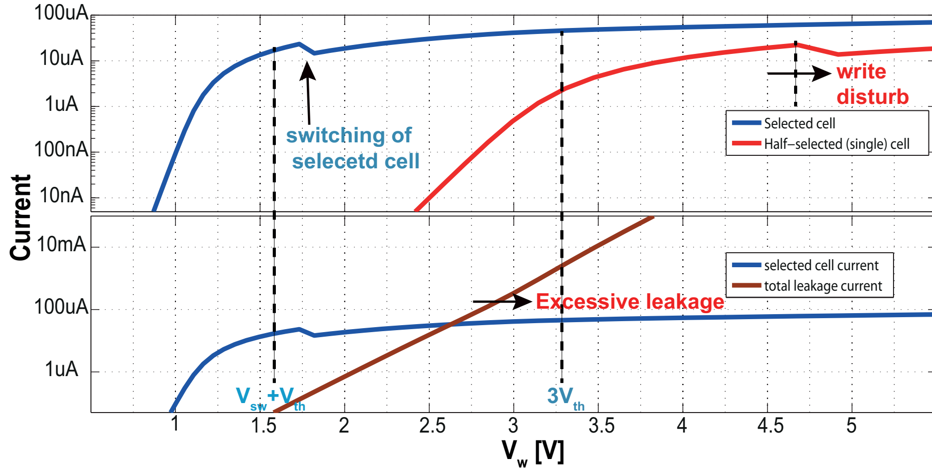


Figure 4.17: Circuit simulation of a crosspoint array with 256×256 memory (V_{th} = 1.1 V and V_{sw} =0.5 V)

The equations discussed above can be used as a design guideline for suitable values of biasing voltages for operating the crosspoint array during write and read operations by considering the impacts of process spread. The minimum (lower bound) and maximum (upper bound) allowed values for the write voltage, V_W, are stated in (4.8). Fore example the boundaries of V_W are shown in Figure 4.18 for a selector device with V_{th} = 0.5 V and memory element with V_{sw} = 0.5 V, considering a variability $\alpha_{V_{sw}}$ and $\alpha_{V_{th}}$ that were varied from 0 to 0.4. For a better visibility, contour lines of the difference between the upper bound and the lower bounds (denoted as ΔV_W) is shown on the right side of Figure 4.18. The $\Delta V_W = 0$ line marks the intersection of the two surfaces and positive values represent availability of a design space for V_W. As it is evident from (4.8), the higher the selector threshold voltage the larger the design space for the write voltage.

The design space for the read voltage V_R is also illustrated in Figure 4.19. As it is evident from (4.25), assuming similar percentage variation, the higher the V_{th} the narrower the design space for V_R. Hence, the threshold voltage of the selector device should be tuned to a suitable value that compromising write and read performance. In this regard, (4.27) gives a generic guideline for selecting a suitable selector device for a given switching voltage of memory element, or vice versa.

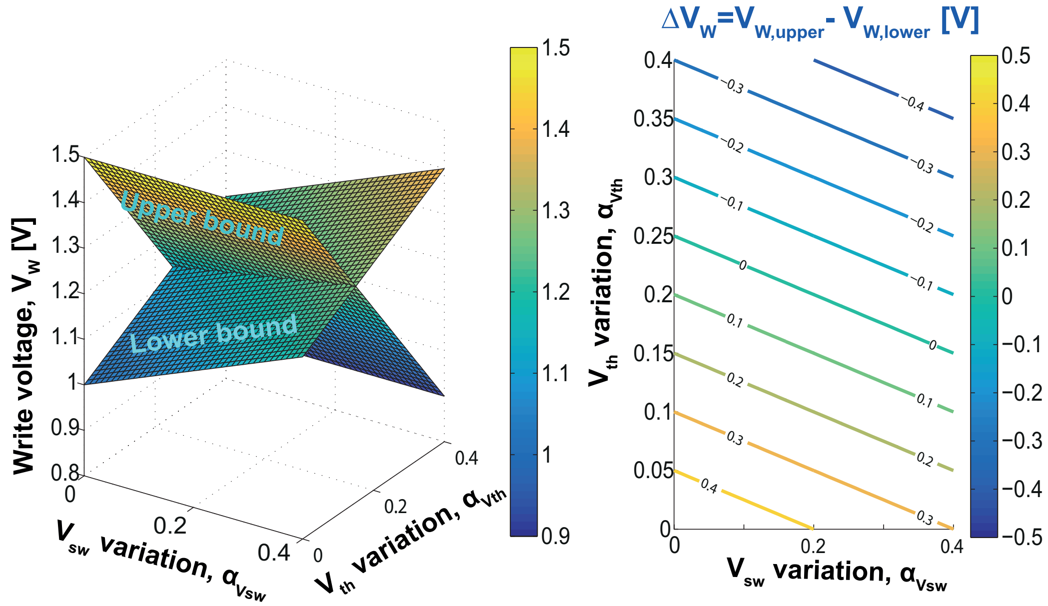


Figure 4.18: Upper and lower bounds of write voltage, V_W, for $V_{th} = 0.5$ V, and $V_{sw} = 0.5$ V

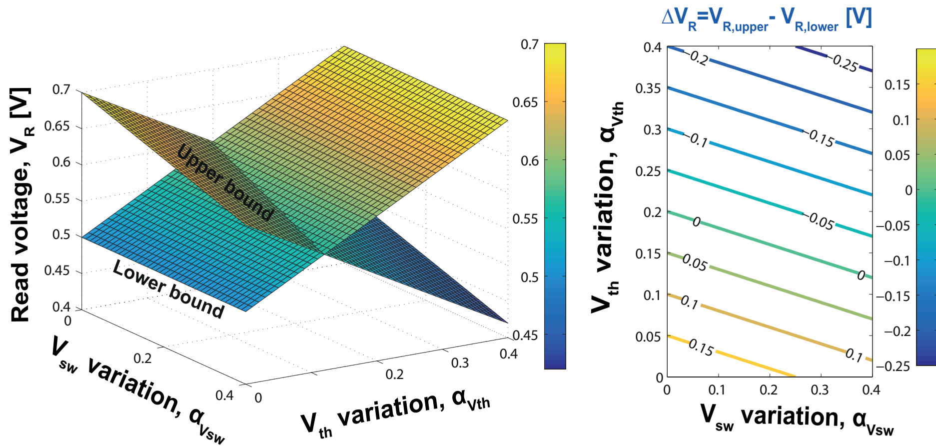


Figure 4.19: Upper and lower bounds of read voltage, V_R, for $V_{th} = 0.5$ V and $V_{sw} = 0.5$ V, and β =0.4

The obtained equations also provide a useful relationships between threshold voltage of selector device and the switching voltage of memory element in the presence of device-to-device variations. These expressions useful for determining the voltage compatibility of selector device and memory element.

In this regard, equations, (4.22) and (4.26) show the relationship between the nominal values of selector threshold voltage, $\overline{V}_{th}$, and memory element switching voltage, $\overline{V}_{sw}$, for meeting memory write and read requirements. In fact, these equations can be used to determine the allowed $\overline{V}_{th}$ for a given $\overline{V}_{sw}$, (or vice versa simply be rearranging the terms). As an illustration, the lower boundary of $\overline{V}_{th}$, (imposed by write operation requirements) is shown in Figure 4.20 as a function of different spread of parameters $\alpha_{V_{sw}}$ and $\alpha_{V_{th}}$, considering a nominal switching voltage $\overline{V}_{sw} = 1.5$ V (wich is roughly equal to the switching voltage of RRAM) for $\alpha_{V_{sw}}$ and $\alpha_{V_{th}}$ varied from 0 to 0.3 .

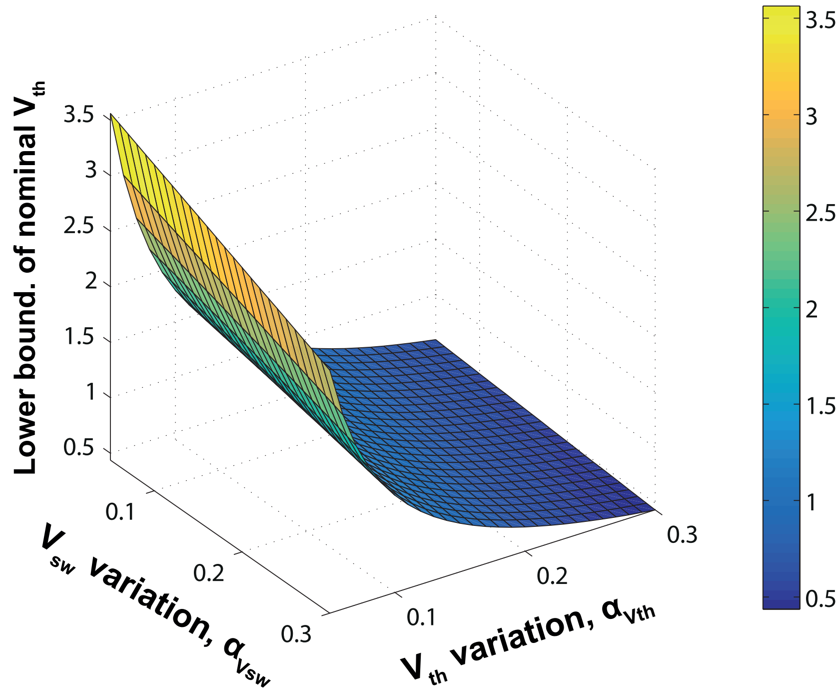


Figure 4.20: Lower boundary of the nominal selector threshold voltage, $\overline{V}_{th}$, (imposed by write operation requirements) as a function of different spread of parameters $\alpha_{V_{sw}}$ and $\alpha_{V_{th}}$, considering $\overline{V}_{sw} = 1.5$ V

Figure 4.21 shows the lower and the upper boundaries of the ratio between $\overline{V}_{sw}$ and $\overline{V}_{th}$, as a function of the amount of variations $\alpha_{V_{sw}}$ and $\alpha_{V_{th}}$ for $\beta = 0.3$. As evident from this plot, the design space (region between the two boundaries) decreases very significantly for increasing process spreads. It is also possible to calculate the maximum tolerable variations.

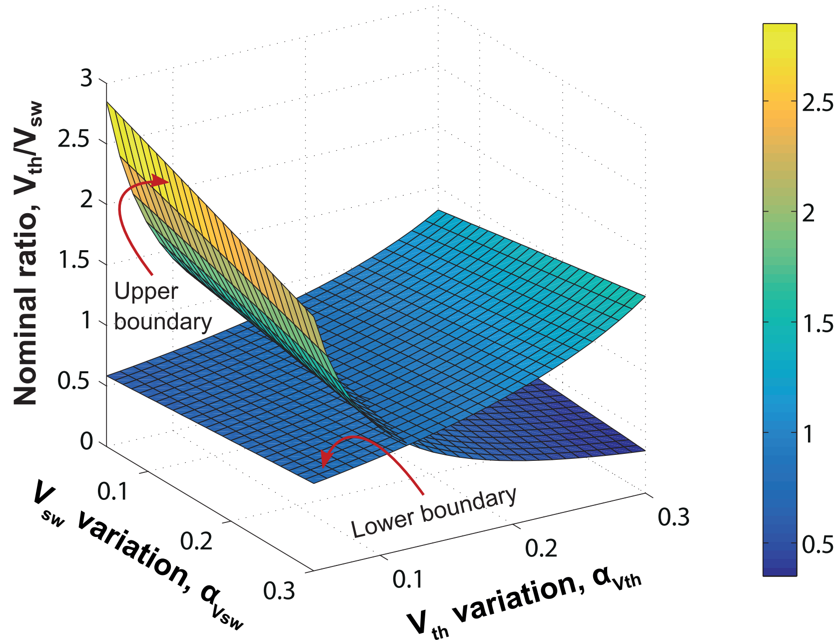


Figure 4.21: Lower and upper boundaries for the ratio $^{V_{th}}/_{V_{sw}}$ as a function of the spread parameters $\alpha_{V_{sw}}$ and $\alpha_{V_{th}}$

The results (Figs 4.21 and 4.22) also demonstrate that memory write requirements set the minimum value of $^{\overline{V}_{th}}/_{\overline{V}_{sw}}$ (due to the requirement to have enough sub-threshold operating region to accommodate half-selected and unselected cells) and memory read requirements set the maximum value of $^{\overline{V}_{th}}/_{\overline{V}_{sw}}$.

In particular, Figure 4.22 shows the upper boundaries for two memory read cases (β =0.3 and β = 0.5) and lower boundaries for two memory write cases (when the half-selected and unselected cells are biased to V_{th} and to some fraction (0.8 in this case) of V_{th}). Indeed, in practically large arrays, where the unselected and half-selected cells should be biased to some fraction of V_{th} to avoid excessive leakage power consumption, the lower boundary increases hence narrowing down the available design space. Also for the read operation β = 0.3 and β = 0.5 are optimistic values when the read disturbance due to cumulative voltage stress over multiple readouts is taken into account. The safe value of β is typically between 0.1 and 0.2 [84,85]. In this case, the design space gets even narrower.

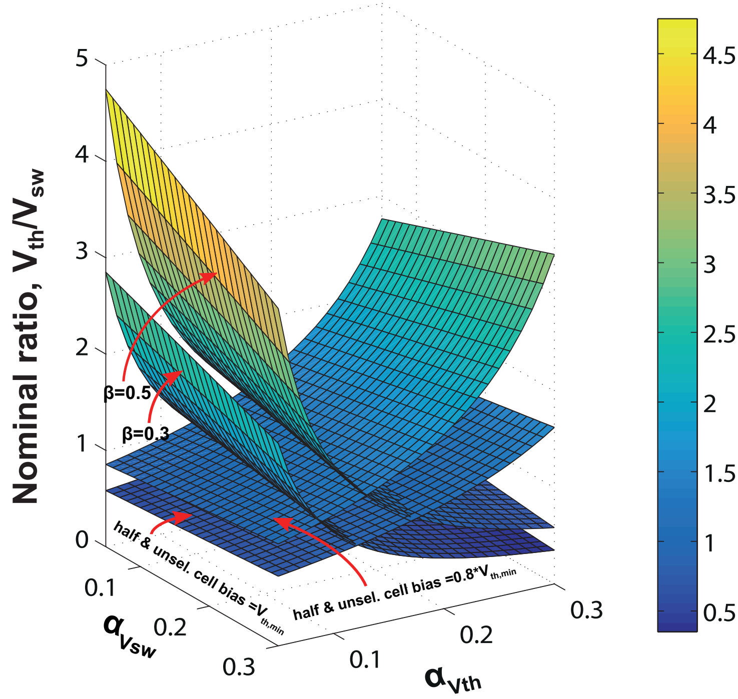


Figure 4.22: Lower and upper boundaries for the ratio $^{V_{th}}/_{V_{sw}}$ as a function of the spread parameters $\alpha_{V_{sw}}$ and $\alpha_{V_{th}}$ for different cases of read and write biasing

4.10 Biasing Crosspoint Memory Arrays for Minimum Leakage Power

As already discussed in Section 4.7 referring to the circuit schematics shown in Figure 4.7, selected and unselected WLs and unselected and selected BLs are biased to voltages V_W, $x \cdot V_W$, $(1-x) \cdot V_W$, and ground, respectively (this scheme is referred to as "x bias scheme"). In this regard, the most commonly used bias schemes are are the 1/2 and 1/3 bias schemes [16, 86, 87]. On the one side, for the same array sizes, the leakage in half-selected cells is less in the 1/3 bias scheme than in the 1/2 bias scheme due to the lower voltage across these cells [16, 16, 87]. On the other side, it is stated in the literature that the 1/2 bias scheme gives minimum leakage power. However, this work demonstrates that the bias scheme which gives the minimum leakage power consumption is largely dependent on array size and selector nonlinearity.

The voltages across half-selected cells (ΔV_{hWL} and ΔV_{hBL}) are equal to $1/2 \cdot V_W$ and $1/3 \cdot V_W$, in the 1/2 and the 1/3 bias schemes , respectively. Hence, for the same V_W, the leakage in half-selected cells is less in the 1/3 bias scheme than in the 1/2 biasing scheme. This means, less write disturbance, less write failure probability, and higher read margin [16, 87] as compared to the 1/2 bias scheme. On the other hand, the voltage across unselected cells, ΔV_u, is 0 V in the the 1/2 biasing scheme as opposed to $1/3 \cdot V_W$ in the 1/3 bias scheme and as in the number of unselected cells in a practically-size array is much higher than the number of half-selected cells, the total leakage power, which is the sum of leakage in half-selected and unselected cells, is in most cases, less in the 1/2 bias scheme than that of the 1/3 bias scheme. However, since total leakage power is contributed by both unselected and half-selected cells, the optimal bias scheme that gives the minimum leakage power consumption in large-size arrays largely highly depends the nonlinearity of selector device. In this section, a design guide for biasing arrays for minimum leakage power consumption is presented.

4.10.1 x Biasing Scheme

Let us assume a generic bias scheme where the unselected WLs are biased to some fraction of V_W, namely $V_{uWL} = x \cdot V_W$ and the unselected BL is biased to $V_{uBL} = (1 - x)$. Therefore, the maximum voltage across half-selected WL and BL cells will be: $\Delta V_{hWL} = \Delta V_{hBL} = x \cdot V_W$ and the voltage across unselected cells, ΔV_u will be equal to $(1 - 2 \cdot x) \cdot V_W$. We can take some considerations to determine the practical range for the value of x. Firstly, since unselected cells by far outnumber half-selected cells in a typical large crosspoint array, the bias voltage across unselected cells should be generally less than or at most equal to the bias voltage across half-selected cells to reduce leakage power. This sets the first constraint on the practical value of x: $(1 - 2 \cdot x) \cdot V_W \leq x \cdot V_W \rightarrow x \geq 1/3$. Secondly, for a given maximum ΔV_{hWL} wich is limited by the selected and since $\Delta V_{hWL} = V_W - V_{uBL}$, we would like to increase V_{uBL} so as to be able to accommodate a high V_W. Therefore, V_{uBL} should be generally greater than or at least equal to V_{uWL}: $V_{uBL} = (1 - x) \cdot V_W \geq V_{uWL} = x \cdot V_W$, which gives the second constraint on the choice of x: $x \leq 1/2$. Hence, the practical range for the choice of x is obtained by combining the two aforementioned constraints.

$$\frac{1}{3} \leq x \leq \frac{1}{2} \tag{4.28}$$

4.10.2 Estimation of Leakage Power

In Section 4.7, we have modeled the interconnection line voltage drop along the selected path, ΔV_{line}, by introducing a correction factor γ to take into account contribution of leakage currents for the drop and hence, to obtain bias voltage at the WL terminal, V_W. For more accurate calculation, the voltage drop, $\Delta V_{line,n}$ along the selected path between WL/BL bias terminal and a cell distant n positions from it (hereafter referred to as the nth cell) can be obtained as:

$$\Delta V_{line,n} = \left(\sum_{i=1}^{n} (I_{lk,i} \cdot i) + \sum_{i=n+1}^{N-1} (I_{lk,i} \cdot n) + n \cdot I_{sw} \right) \cdot R_c \tag{4.29}$$

where, n goes from 1 to N and i goes from 1 to $N-1$. Therefore, the voltage across the nth half-selected WL cell will be equal to $V_W - \Delta V_{line,n} - V_{uBL} = \Delta V_{hWL} - \Delta V_{line,n}$ and voltage across the nth half-selected BL cell will be equal to $V_{uWL} - \Delta V_{line,n} = \Delta V_{hBL} - \Delta V_{line,n}$. The leakage current through the nth half-selected WL or BL cell, $I_{lk,n}$, is obtained from the I-V equation of the selector device. Since $I_{lk,n}$ itself is a function of $\Delta V_{line,n}$, equation (4.29) is solved by applying iterative loops until all node voltages and leakage currents are determined with satisfactory accuracy. Then, leakage power values in all half-selected and unselected cells are calculated and summed up:

$$P_{hfWL} = \sum_{n=1}^{N-1} (V_W - \Delta V_{line,n} - (1-x) \cdot V_W) \cdot I_{lk,n} \tag{4.30a}$$

$$P_{hfBL} = \sum_{n=1}^{N-1} (x \cdot V_W - \Delta V_{line,n}) \cdot I_{lk,n} \tag{4.30b}$$

$$P_{un} = (N-1) \cdot (N-1) \cdot (1 - 2 \cdot x) \cdot V_W \cdot I_{lk,u} \tag{4.30c}$$

where, $I_{lk,u}$ is leakage current through unselected cells. Using the analytical equations discussed previously, simulations were done to determine the bias scheme for minimum power consumption. First, the bias scheme factor, x, and the array size, $N \times N$, were varied to investigate how the minimum-power bias scheme changes with array size. For every combination of N and x, the write current, I_W, required to provide enough switching current for the selected cell in the worst-case scenario, and also the WL bias voltage, V_W, were calculated. The voltages and the leakage currents in half-selected and unselected cells were then obtained, which enabled us to calculate the leakage power in each cell.

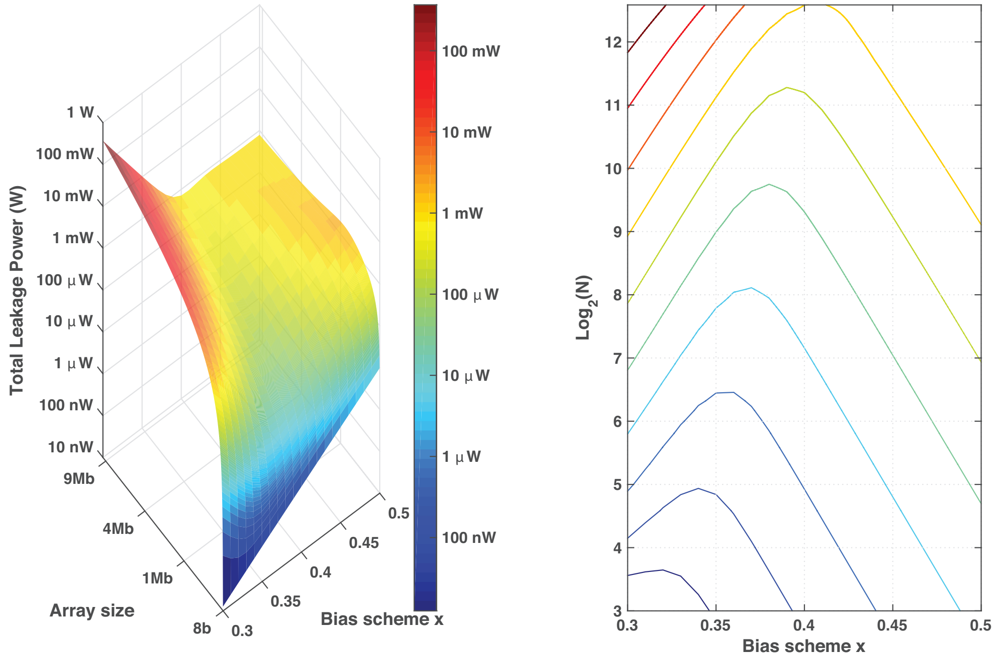


Figure 4.23: Total leakage power as a function of array size and bias scheme (sub-threshold nonlinearity of selector = 0.2 V/decade)

The total leakage power obtained by summing up (4.30a), (4.30b) and (4.30c) is shown in Figure 4.23 for different array sizes. The peaks of the contour show the optimum point for obtaining minimum leakage power consumption. As demonstrated by the 3D plot (left) and its contour plot (right) with leakage power in logarithmic scale (the array size is in $log_2()$scale), the leakage power is minimum for x somewhere between $1/3$ and $1/2$. Moreover, the optimum value of x depends on array size: it is close to $1/3$ for small array sizes and increases with increasing array size, approaching $1/2$ for very large array sizes.

The minimum-power bias scheme is also a function of selector nonlinearity in its sub-threshold region, as demonstrated by the contour plot of leakage power in $1Mb$ array size, as shown in Figure 4.24. For high nonlinearity (small δ_{sub}), the optimum value of x lies close to $1/3$ whereas, for low nonlinearity (large δ_{sub}) , the optimum value increases, approaching $1/2$.

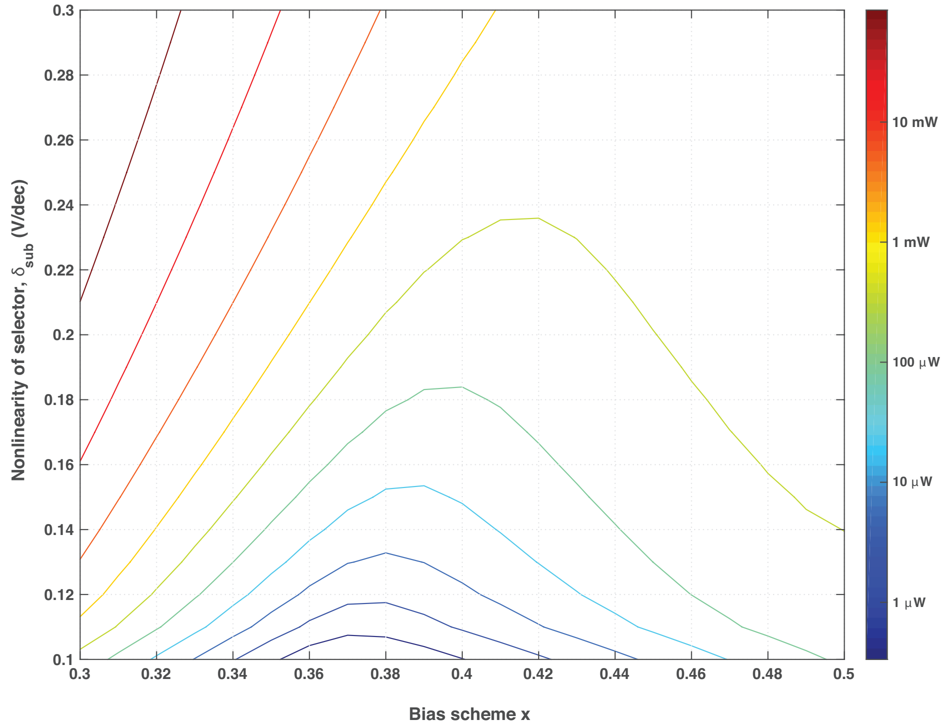


Figure 4.24: Contour plot of total leakage power as a function of biasing scheme and nonlinearity of selector (leakage power is constant over each indicated contour line)

4.11 Conclusion

In this Chapter, it has been presented a comprehensive analysis on the design considerations and technology requirements of 1S1R crosspoint memory arrays.

Firstly, the minimum requirements (boundary conditions) of array biasing and memory/selector operating voltages have been analyzed. As a guideline for array biasing, it has been presented an analysis on biasing arrays to obtain minimum leakage power consumption by employing a customized bias scheme instead of using the conventional schemes. In this respect, the dependence of the customized bias scheme on array size and nonlinearity of selector device has been analyzed. Besides, the voltage compatibility requirements of selector device and memory element have also been investigated by considering variabilities in the switching voltage of the memory element and threshold voltage of the selector device, which is useful for choosing a selector device for a particular memory element or vice versa.

Secondly, analysis of the design considerations in practically large-size arrays to meet a certain write/read performance and power consumption requirements has been analyzed. By using circuital model and analytical equations, and by setting the maximum feasible array size (i.e. the maximum array size that satisfies the minimum write and read performance requirements) as a design target, it have been demonstrated how this feasible array size is affected by characteristics of selector device, memory element, and interconnection metal line. A higher selector voltage margin (or threshold voltage) provides a broad low-leakage zone to accommodate half-selected and unselected cells and, hence, enables to implement larger arrays. Yet, the nonlinearity of the selector device is another critical parameter that limits the feasible array size. The higher the nonlinearity of the selector device (i.e. the smaller the turn-on slope, δ), the higher the feasible array size. Additionally, the achievable feasible array size highly depends on the characteristic of the memory element itself. For instance, a high switching current, such as the current required during RESET operation in PCM, results in high IR drop along interconnection metal lines, hence, limiting the feasible array size. Similarly, a high switching voltage across the selected cell raises the voltage level in the selected WL, which in turn increases the voltage across half-selected cells, thereby increasing leakage and, in the extreme case, disturbing the half-selected cells. It has also been shown that the feasible array size decreases significantly with an increase in parasitic resistance of WLs and BLs, which unavoidably occurs with advanced scaling down of interconnection metal lines.

According to the presented analysis, it seems challenging to implement large-size PCM crosspoint arrays using selector devices reported in the literature since the high switching current required for RESET operation leads to high ohmic voltage drop along the interconnection metal lines. In contrast, in the case of STT-MRAM, which exhibits lower switching voltage and current than PCM and RRAM, the write requirements do not forbid implementation of practically large-size arrays. Nevertheless, due to the very small R_H/R_L ratio, the size of STT-MRAM crosspoint arrays is constrained by read (or sensing) margin requirement, which will depend on the sensitivity of the comparator and the amount of read current. In the case of RRAM, the feasible array size lies in between that of STT-MRAM and that of PCM. Due to the high R_H/R_L ratio, sufficient read margin is achievable in large-size RRAM and PCM arrays, and thus, the most important constraint to PCM and RRAM crosspoint arrays comes from write requirements, whereas for STT-MRAM, it seems the more stringent constraint comes from the sensing requirements.

Chapter 5

General Conclusions and Future Prospects

5.1 Conclusions

With the objective of contributing to the research and development on emerging memory devices and high storage density architectures, this PhD thesis has presented a model-based study of device, array and sensing circuit schemes of Spin-Transfer Torque Magnetic Random Access Memory (STT-MRAM), an experimental electrical characterization of two Resistive Random Access Memory (RRAM) device stacks, and a detail analysis of the design considerations for write and read operations and device technology requirements in crosspoint memory arrays.

For the case of STT-MRAM, a behavioral model of STT-MRAM cell has been presented. The model has been developed in Veriolg-A language based on the physics of the basic storage Magnetic Tunnel Junction (MTJ) device and the spin-transfer torque (STT) effect. The model mimics the dynamic (or switching) and static characteristics of the STT-MRAM cell and it is suited for circuit simulations. In addition, it has been presented a review of STT-MRAM sensing circuit schemes and a variability-aware analysis and design guideline of slope detection self-reference scheme, which is deemed to outperform other STT-MRAM sensing schemes available in the literature. Using a simplified model for reading in conventional and crosspoint STT-MRAM arrays, the performance (i.e., sense margin) of the SD sensing scheme has been analyzed by taking into account the impact of cell-to-cell variations and parasitic resistance in bitlines, BLs, and wordlines, WLs.

On Resistive RAM, a detailed array-level experimental electrical characterization of Oxide RAM with TiN/Hf/GdAlO/TiN device stack and a reliability study of Conductive Bridging RAM with Cu/TiW/SrTiOx/WOx/W stack have been presented. The impacts of the thickness of the GdAlO layer and the size of the memory device on array forming, SET and RESET voltages and endurance have been discussed. When the thickness of GdAlO layer was reduced from 5 nm to 3 nm, the median forming voltage decreased by about 30% while the SET and RESET voltages were not affected significantly. Concerning the impact of cell size, on the one hand, when the diameter of the memory cell is reduced from 150 nm to 60 nm, an increase in forming, RESET and SET voltages have been observed. On the other hand, the 60nm cells has shown better endurance than that of the 150 nm cells. On the STO-based CBRAM, it has been presented a reliability study of the Cu/TiW/SrTiOx/WOx/W stack and an optimization of the SET and RESET voltage pulses and forming and SET currents for obtaining optimum memory performance and reliability

Finally, a detail analysis on the design considerations and technology requirements of 1S1R crosspoint memory arrays have been presented. In this respect, the impacts of selector and memory element characteristics, and interconnection metal line parasitic resistance on the maximum achievable crosspoint array size have been analyzed in detail. Besides, an analysis on biasing arrays to obtain minimum leakage power consumption, by employing a customized bias scheme instead of using the conventional schemes, has been presented. As a guideline for choosing a selector device for a particular memory element or vice versa, the voltage compatibility requirements of selector device and memory element have also been investigated by considering variabilities in the switching voltage of the memory element and threshold voltage of the selector device.

5.2 Future Prospects

As for the study on STT-MRAM, the developed model and the analyses should be validated by experimental tests. It would also be necessary to consider the latest device technology developments. A natural continuation of the experimental electrical characterization work on the OxRAM and CBRAM devices would be repeating the tests with new devices with different features (for example with different oxide thickness and duration of thermal treatment). As for the crosspoint arrays, the full characteristics of the 1S1R cell was predicted simply by combining the individual I-V characteristics of the selector device (1S) and memory element (1R). It would be interesting to develop an aggregate model of the the 1S1R cell and to compare it with the presented work. Besides, the aggregate approach will enable to easily model self-rectifying cells (SRC), which appears to be the solution for enabling high-density vertical 3D crosspoint memories. In connection with this, it would also be interesting to extend the presented analyses 3D crosspoint arrays.

Even though the idea of a 'universal' memory with ideal characteristics seems far from achieving, the emerging memories can be optimized to different targets from replacement of traditional semiconductor memories to opening new markets. To take advantage of the interesting features of these emerging memory technologies such as, non-volatility and fast access, it seems very necessary to re-think computation systems in general and the memory subsystem in particular. In addition to speed, cost and power targets, this re-thinking may add new functionality and features to computing systems. I also expect that the research and development efforts for adopting the emerging memory technologies for non-memory applications, such as in-memory computing, spin logic, neuromorphic computing, hardware security, will keep only increasing.

"After all, we can always reshape the future to our needs."

Bibliography

[1] S. Motaman, S. Ghosh, and J. P. Kulkarni, "A novel slope detection technique for robust STTRAM sensing," in *Low Power Electronics and Design (ISLPED), 2015 IEEE/ACM International Symposium on.* IEEE, 2015, pp. 7–12.

[2] D. Reinsel, J. Gantz, and J. Rydning, "Data age 2025: The evolution of data to life-critical don't focus on big data; focus on data that's big," *IDC, Seagate, April*, 2017.

[3] H. Yu and Y. Wang, *Design exploration of emerging nano-scale non-volatile memory.* Springer, 2014.

[4] J. M. Rabaey, A. P. Chandrakasan, and B. Nikolić, *Digital Integrated Circuits, 2/e.* Prentice hall, 2003.

[5] L. Zhang, "Study of the selector element for resistive memory," PhD dissertation, Ku Leuven Arenberg Doctoral School-Faculty of Engineering of Science, 2015.

[6] L. Torres, R. M. Brum, L. V. Cargnini, and G. Sassatelli, "Trends on the application of emerging nonvolatile memory to processors and programmable devices," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*, May 2013, pp. 101–104.

[7] S. Yu and P.-Y. Chen, "Emerging memory technologies: recent trends and prospects," *IEEE Solid-State Circuits Magazine*, vol. 8, no. 2, pp. 43–56, 2016.

[8] T. Endoh, H. Koike, S. Ikeda, T. Hanyu, and H. Ohno, "An overview of nonvolatile emerging memories—spintronics for working memories," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 6, no. 2, pp. 109–119, 2016.

[9] H.-S. P. Wong and S. Salahuddin, "Memory leads the way to better computing," *Nature nanotechnology*, vol. 10, no. 3, p. 191, 2015.

[10] S.-K. Park, "Technology scaling challenge and future prospects of DRAM and NAND flash memory," in *Memory Workshop (IMW), 2015 IEEE International*. IEEE, 2015, pp. 1–4.

[11] O. Mutlu and L. Subramanian, "Research problems and opportunities in memory systems," *Supercomputing frontiers and innovations*, vol. 1, no. 3, pp. 19–55, 2015.

[12] C. Xu, D. Niu, N. Muralimanohar, R. Balasubramonian, T. Zhang, S. Yu, and Y. Xie, "Overcoming the challenges of crossbar resistive memory architectures," in *2015 IEEE 21st International Symposium on High Performance Computer Architecture (HPCA)*, Feb 2015, pp. 476–488.

[13] S. Khan and S. Hamdioui, "Trends and challenges of SRAM reliability in the nano-scale era," in *5th International Conference on Design Technology of Integrated Systems in Nanoscale Era*, March 2010, pp. 1–6.

[14] S. Yu, "Overview of resistive switching memory (RRAM) switching mechanism and device modeling," in *Circuits and Systems (ISCAS), 2014 IEEE International Symposium on*. IEEE, 2014, pp. 2017–2020.

[15] A. Chen, Z. Krivokapic, and M.-R. Lin, "A comprehensive model for crossbar memory arrays," in *Device Research Conference (DRC), 2012 70th Annual*. IEEE, 2012, pp. 219–220.

[16] A. Chen, "Analysis of partial bias schemes for the writing of crossbar memory arrays," *IEEE Transactions on Electron Devices*, vol. 62, no. 9, pp. 2845–2849, 2015.

[17] K. Wang, J. Alzate, and P. K. Amiri, "Low-power non-volatile spintronic memory: Stt-ram and beyond," *Journal of Physics D: Applied Physics*, vol. 46, no. 7, p. 074003, 2013.

[18] K. Asifuzzaman, R. S. Verdejo, and P. Radojković, "Enabling a reliable STT-MRAM main memory simulation," in *Proceedings of the International Symposium on Memory Systems*. ACM, 2017, pp. 283–292.

[19] R. Carboni, S. Ambrogio, W. Chen, M. Siddik, J. Harms, A. Lyle, W. Kula, G. Sandhu, and D. Ielmini, "Understanding cycling endurance in perpendicular spin-transfer torque (p-STT) magnetic memory," in *Electron Devices Meeting (IEDM), 2016 IEEE International*. IEEE, 2016, pp. 21–6.

[20] Z. Xu, C. Yang, M. Mao, K. B. Sutaria, C. Chakrabarti, and Y. Cao, "Compact modeling of STT-MTJ devices," *Solid-State Electronics*, vol. 102, pp. 76–81, 2014.

[21] K. C. Chun, H. Zhao, J. D. Harms, T. Kim, J. Wang, and C. H. Kim, "A scaling roadmap and performance evaluation of in-plane and perpendicular MTJ-based STT-MRAMs for high-density cache memory," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 2, pp. 598–610, Feb 2013.

[22] P. Chi, S. Li, Y. Cheng, Y. Lu, S. H. Kang, and Y. Xie, "Architecture design with STT-RAM: Opportunities and challenges," in *2016 21st Asia and South Pacific Design Automation Conference (ASP-DAC)*, Jan 2016, pp. 109–114.

[23] K. Garello, F. Yasin, S. Couet, L. Souriau, J. Swerts, S. Rao, S. Van Beek, W. Kim, E. Liu, S. Kundu *et al.*, "SOT-MRAM 300mm integration for low power and ultrafast embedded memories," in *2018 IEEE Symposium on VLSI Circuits*. IEEE, 2018, pp. 81–82.

[24] G. Prenat, K. Jabeur, P. Vanhauwaert, G. Di Pendina, F. Oboril, R. Bishnoi, M. Ebrahimi, N. Lamard, O. Boulle, K. Garello *et al.*, "Ultra-fast and high-reliability SOT-MRAM: From cache replacement to normally-off computing." *IEEE Trans. Multi-Scale Computing Systems*, vol. 2, no. 1, pp. 49–60, 2016.

[25] W. Thomson, "XIX. on the electro-dynamic qualities of metals:—effects of magnetization on the electric conductivity of nickel and of iron," *Proceedings of the Royal Society of London*, vol. 8, pp. 546–550, 1857.

[26] M. N. Baibich, J. M. Broto, A. Fert, F. N. Van Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friederich, and J. Chazelas, "Giant magnetoresistance of (001) Fe/(001) Cr magnetic superlattices," *Physical review letters*, vol. 61, no. 21, p. 2472, 1988.

[27] J. S. Moodera, L. R. Kinder, T. M. Wong, and R. Meservey, "Large magnetoresistance at room temperature in ferromagnetic thin film tunnel junctions," *Physical review letters*, vol. 74, no. 16, p. 3273, 1995.

[28] S. S. Parkin, C. Kaiser, A. Panchula, P. M. Rice, B. Hughes, M. Samant, and S.-H. Yang, "Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers," *Nature materials*, vol. 3, no. 12, p. 862, 2004.

[29] X. Fong, Y. Kim, K. Yogendra, D. Fan, A. Sengupta, A. Raghunathan, and K. Roy, "Spin-transfer torque devices for logic and memory: Prospects and perspectives," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 35, no. 1, pp. 1–22, 2016.

[30] J. C. Slonczewski, "Current-driven excitation of magnetic multilayers," *Journal of Magnetism and Magnetic Materials*, vol. 159, no. 1-2, pp. L1–L7, 1996.

[31] L. Berger, "Emission of spin waves by a magnetic multilayer traversed by a current," *Physical Review B*, vol. 54, no. 13, p. 9353, 1996.

[32] S. Ikeda, K. Miura, H. Yamamoto, K. Mizunuma, H. Gan, M. Endo, S. Kanai, J. Hayakawa, F. Matsukura, and H. Ohno, "A perpendicular-anisotropy CoFeB–MgO magnetic tunnel junction," *Nature materials*, vol. 9, no. 9, p. 721, 2010.

[33] R. Andrawis, A. Jaiswal, and K. Roy, "Design and comparative analysis of spintronic memories based on current and voltage driven switching," *IEEE Transactions on Electron Devices*, 2018.

[34] Y. Chen, H. Li, X. Wang, W. Zhu, W. Xu, and T. Zhang, "A 130 nm 1.2 v/3.3 v 16 kb spin-transfer torque random access memory with nondestructive self-reference sensing scheme," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 2, pp. 560–573, 2012.

[35] Z. Sun, H. Li, Y. Chen, and X. Wang, "Voltage driven nondestructive self-reference sensing scheme of spin-transfer torque memory," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 20, no. 11, pp. 2020–2030, 2012.

[36] W. Kang, Y. Cheng, Y. Zhang, D. Ravelosona, and W. Zhao, "Readability challenges in deeply scaled STT-MRAM," in *Non-Volatile Memory Technology Symposium (NVMTS), 2014 14th Annual*. IEEE, 2014, pp. 1–4.

[37] C. Chappert, A. Fert, and F. N. Van Dau, "The emergence of spin electronics in data storage," in *Nanoscience And Technology: A Collection of Reviews from Nature Journals*. World Scientific, 2010, pp. 147–157.

[38] S. Peng, Y. Zhang, M. Wang, Y. Zhang, and W. Zhao, "Magnetic tunnel junctions for spintronics: principles and applications," *Wiley Encyclopedia of Electrical and Electronics Engineering*, pp. 1–16, 1999.

[39] J. C. Slonczewski, "Current-driven excitation of magnetic multilayers," *Journal of Magnetism and Magnetic Materials*, vol. 159, no. 1-2, pp. L1–L7, 1996.

[40] Z. Diao, Z. Li, S. Wang, Y. Ding, A. Panchula, E. Chen, L.-C. Wang, and Y. Huai, "Spin-transfer torque switching in magnetic tunnel junctions and spin-transfer torque random access memory," *Journal of Physics: Condensed Matter*, vol. 19, no. 16, p. 165209, 2007.

[41] Y. Huai, "Spin-transfer torque MRAM (STT-MRAM): Challenges and prospects," *AAPPS bulletin*, vol. 18, no. 6, pp. 33–40, 2008.

[42] J. D. Harms, F. Ebrahimi, X. Yao, and J.-P. Wang, "Spice macromodel of spin-torque-transfer-operated magnetic tunnel junctions," *IEEE transactions on electron devices*, vol. 57, no. 6, pp. 1425–1430, 2010.

[43] A. Vatankhahghadim, S. Huda, and A. Sheikholeslami, "A survey on circuit modeling of spin-transfer-torque magnetic tunnel junctions." *IEEE Trans. on Circuits and Systems*, vol. 61, no. 9, pp. 2634–2643, 2014.

[44] L.-B. Faber, W. Zhao, J.-O. Klein, T. Devolder, and C. Chappert, "Dynamic compact model of spin-transfer torque based magnetic tunnel junction (MTJ)," in *Design & Technology of Integrated Systems in Nanoscal Era, 2009. DTIS'09. 4th International Conference on.* IEEE, 2009, pp. 130–135.

[45] R. Garg and J. Kedia, "A novel verilog-a model of spin torque transfer magnetic tunnel junction," *Universal Journal of Physics and Application*, vol. 7, no. 3, pp. 290–294, 2013.

[46] Y. Chen and H. Li, "Emerging sensing techniques for emerging memories," in *Proceedings of the 16th Asia and South Pacific Design Automation Conference.* IEEE Press, 2011, pp. 204–210.

[47] I. H. Inoue and A. Sawa, "Resistive switchings in transition metal oxides," *Functional Metal Oxides: New Science and Novel Applications*, pp. 443–463, 2013.

[48] J. J. Yang, D. B. Strukov, and D. R. Stewart, "Memristive devices for computing," *Nature nanotechnology*, vol. 8, no. 1, p. 13, 2013.

[49] D. Jana, M. Dutta, S. Samanta, and S. Maikap, "RRAM characteristics using a new Cr/GdO x/TiN structure," *Nanoscale research letters*, vol. 9, no. 1, p. 680, 2014.

[50] C. Y. Chen, L. Goux, A. Fantini, A. Redolfi, G. Groeseneken, and M. Jurczak, "Low-current operation of novel Gd2O3-based RRAM cells with large memory window," *physica status solidi (a)*, vol. 213, no. 2, pp. 320–324, 2016.

[51] P. Feng, C. Chao, Z.-s. Wang, Y.-c. Yang, Y. Jing, and Z. Fei, "Nonvolatile resistive switching memories-characteristics, mechanisms and challenges," *Progress in Natural Science: Materials International*, vol. 20, pp. 1–15, 2010.

[52] D. Ielmini, "Resistive switching memories based on metal oxides: mechanisms, reliability and scaling," *Semiconductor Science and Technology*, vol. 31, no. 6, p. 063002, 2016.

[53] M. J. Marinella, "Emerging resistive switching memory technologies: Overview and current status," in *Circuits and Systems (ISCAS), 2014 IEEE International Symposium on.* IEEE, 2014, pp. 830–833.

[54] A. Fantini, D. Wouters, R. Degraeve, L. Goux, L. Pantisano, G. Kar, Y.-Y. Chen, B. Govoreanu, J. Kittl, L. Altimime *et al.*, "Intrinsic switching behavior in HfO_2 RRAM by fast electrical measurements on novel 2R test structures," in *Memory Workshop (IMW), 2012 4th IEEE International.* IEEE, 2012, pp. 1–4.

[55] A. Belmonte, L. Goux, J. Woo, U. Celano, A. Redolfi, S. Clima, and G. S. Kar, "Enhancement of CBRAM performance by controlled formation of a hourglass-shaped filament," in *Non-Volatile Memory Technology Symposium (NVMTS), 2017 17th.* IEEE, 2017, pp. 1–5.

[56] R. Degraeve, A. Fantini, N. Raghavan, L. Goux, S. Clima, Y.-Y. Chen, A. Belmonte, S. Cosemans, B. Govoreanu, D. Wouters *et al.*, "Hourglass concept for RRAM: a dynamic and statistical device model," in *Physical and Failure Analysis of Integrated Circuits (IPFA), 2014 IEEE 21st International Symposium on the.* IEEE, 2014, pp. 245–249.

[57] K. Ota, A. Belmonte, Z. Chen, A. Redolfi, L. Goux, and G. Kar, "Impact of the filament morphology on the retention characteristics of Cu/Al 2 O 3-based CBRAM devices," in *Electron Devices Meeting (IEDM), 2016 IEEE International.* IEEE, 2016, pp. 21–2.

[58] R. S. Shenoy, G. W. Burr, K. Virwani, B. Jackson, A. Padilla, P. Narayanan, C. T. Rettner, R. M. Shelby, D. S. Bethune, K. V. Raman *et al.*, "MIEC (mixed-ionic-electronic-conduction)-based access devices

for non-volatile crossbar memory arrays," *Semiconductor Science and Technology*, vol. 29, no. 10, p. 104005, 2014.

[59] R. Aluguri and T.-Y. Tseng, "Overview of selector devices for 3-D stackable cross point RRAM arrays," *IEEE Journal of the Electron Devices Society*, vol. 4, no. 5, pp. 294–306, 2016.

[60] B. Hudec, C.-W. Hsu, I.-T. Wang, W.-L. Lai, C.-C. Chang, T. Wang, K. Fröhlich, C.-H. Ho, C.-H. Lin, and T.-H. Hou, "3D resistive RAM cell design for high-density storage class memory—a review," *Science China Information Sciences*, vol. 59, no. 6, p. 061403, 2016.

[61] H.-Y. Chen, S. Brivio, C.-C. Chang, J. Frascaroli, T.-H. Hou, B. Hudec, M. Liu, H. Lv, G. Molas, J. Sohn *et al.*, "Resistive random access memory (RRAM) technology: From material, device, selector, 3d integration to bottom-up fabrication," *Journal of Electroceramics*, vol. 39, no. 1-4, pp. 21–38, 2017.

[62] G. W. Burr, R. S. Shenoy, K. Virwani, P. Narayanan, A. Padilla, B. Kurdi, and H. Hwang, "Access devices for 3d crosspoint memory," *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, vol. 32, no. 4, p. 040802, 2014.

[63] I. Baek, C. Park, H. Ju, D. Seong, H. Ahn, J. Kim, M. Yang, S. Song, E. Kim, S. Park *et al.*, "Realization of vertical resistive memory (vrram) using cost effective 3d process," in *Electron Devices Meeting (IEDM), 2011 IEEE International*. IEEE, 2011, pp. 31–8.

[64] 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 Letters*, vol. 36, no. 6, pp. 570–572, 2015.

[65] S. Choi, W. Sun, H. Lim, and H. Shin, "An analysis of the read margin and power consumption of crossbar ReRAM arrays," in *TENCON 2015-2015 IEEE Region 10 Conference*. IEEE, 2015, pp. 1–3.

[66] B. Govoreanu, L. Zhang, and M. Jurczak, "Selectors for high density crosspoint memory arrays: Design considerations, device implementations and some challenges ahead," in *IC Design & Technology (ICICDT), 2015 International Conference on*. IEEE, 2015, pp. 1–4.

[67] X. Dong, C. Xu, Y. Xie, and N. P. Jouppi, "Nvsim: A circuit-level performance, energy, and area model for emerging nonvolatile memory," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 31, no. 7, pp. 994–1007, 2012.

[68] A. Ciprut and E. G. Friedman, "Modeling size limitations of resistive crossbar array with cell selectors," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 25, no. 1, pp. 286–293, 2017.

[69] P. Narayanan, G. W. Burr, K. Virwani, and B. Kurdi, "Circuit-level benchmarking of access devices for resistive nonvolatile memory arrays," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 6, no. 3, pp. 330–338, 2016.

[70] P. Narayanan, G. W. Burr, R. S. Shenoy, S. Stephens, K. Virwani, A. Padilla, B. N. Kurdi, and K. Gopalakrishnan, "Exploring the design space for crossbar arrays built with mixed-ionic-electronic-conduction (MIEC) access devices," *IEEE Journal of the Electron Devices Society*, vol. 3, no. 5, pp. 423–434, 2015.

[71] L. Zhang, S. Cosemans, D. J. Wouters, G. Groeseneken, M. Jurczak, and B. Govoreanu, "Selector design considerations and requirements for 1S1R RRAM crossbar array," in *Memory Workshop (IMW), 2014 IEEE 6th International*. IEEE, 2014, pp. 1–4.

[72] J. Liang, S. Yeh, S. S. Wong, and H. S. P. Wong, "Scaling challenges for the cross-point resistive memory array to sub-10nm node - an interconnect perspective," in *2012 4th IEEE International Memory Workshop*, May 2012, pp. 1–4.

[73] G. W. Burr, R. S. Shenoy, K. Virwani, P. Narayanan, A. Padilla, B. Kurdi, and H. Hwang, "Access devices for 3D crosspoint memory," *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, vol. 32, no. 4, p. 040802, 2014.

[74] X. Peng, R. Madler, P.-Y. Chen, and S. Yu, "Cross-point memory design challenges and survey of selector device characteristics," *Journal of Computational Electronics*, vol. 16, no. 4, pp. 1167–1174, 2017.

[75] P. W. Ho, N. H. El-Hassan, T. N. Kumar, and H. A. F. Almurib, "PCM and memristor based nanocrossbars," in *Nanotechnology (IEEE-NANO), 2015 IEEE 15th International Conference on*. IEEE, 2015, pp. 456–459.

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

[77] S. Verma, A. A. Kulkarni, and B. K. Kaushik, "Spintronics-based devices to circuits: Perspectives and challenges." *IEEE Nanotechnology Magazine*, vol. 10, no. 4, pp. 13–28, 2016.

[78] L. Thomas, G. Jan, J. Zhu, H. Liu, Y.-J. Lee, S. Le, R.-Y. Tong, K. Pi, Y.-J. Wang, D. Shen *et al.*, "Perpendicular spin transfer torque magnetic random access memories with high spin torque efficiency and thermal stability for embedded applications," *Journal of Applied Physics*, vol. 115, no. 17, p. 172615, 2014.

[79] H.-S. P. Wong, S. Raoux, S. Kim, J. Liang, J. P. Reifenberg, B. Rajendran, M. Asheghi, and K. E. Goodson, "Phase change memory," *Proceedings of the IEEE*, vol. 98, no. 12, pp. 2201–2227, 2010.

[80] Y. A. Belay, A. Cabrini, and G. Torelli, "Analysis of array biasing in crosspoint memories for leakage power minimization," in *2017 13th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME)*, June 2017, pp. 17–20.

[81] International technology roadmap for semiconductors (itrs). [Online]. Available: http://www.itrs2.net/2013-itrs.html

[82] A. Levisse, P. Royer, B. Giraud, J. P. Noel, M. Moreau, and J. M. Portal, "Architecture, design and technology guidelines for crosspoint memories," in *2017 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*, July 2017, pp. 55–60.

[83] J. Liang and H.-S. P. Wong, "Size limitation of cross-point memory array and its dependence on data storage pattern and device parameters," in *Interconnect Technology Conference (IITC), 2010 International*. IEEE, 2010, pp. 1–3.

[84] A. Belmonte, U. Celano, A. Redolfi, A. Fantini, R. Muller, W. Vandervorst, M. Houssa, M. Jurczak, and L. Goux, "Analysis of the excellent memory disturb characteristics of a hourglass-shaped filament in Al 2 O 3/Cu-based CBRAM devices," *IEEE Transactions on Electron Devices*, vol. 62, no. 6, pp. 2007–2013, 2015.

[85] J. Kim, T. Ahmed, H. Nili, J. Yang, D. S. Jeong, P. Beckett, S. Sriram, D. C. Ranasinghe, and O. Kavehei, "A physical unclonable function

with redox-based nanoionic resistive memory," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 2, pp. 437–448, 2018.

[86] W. Sun, S. Choi, and H. Shin, "A new bias scheme for a low power consumption reram crossbar array," *Semiconductor Science and Technology*, vol. 31, no. 8, p. 085009, 2016.

[87] W. Sun, S. Choi, H. Lim, and H. Shin, "Guideline model for the bias-scheme-dependent power consumption of a resistive random access memory crossbar array," *Japanese Journal of Applied Physics*, vol. 55, no. 4S, p. 04EE10, 2016.