STT-MRAM Characterization and its Test Implications

by

Govindakrishnan Radhakrishnan

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

Waterloo, Ontario, Canada, 2020

© Govindakrishnan Radhakrishnan 2020

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

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

Abstract

Spin torque transfer (STT)-magnetoresistive random-access memory (MRAM) has come a long way in research to meet the speed and power consumption requirements for future memory applications. The state-of-the-art STT-MRAM bit-cells employ magnetic tunnel junction (MTJ) with perpendicular magnetic anisotropy (PMA). The process repeatability and yield stability for wafer fabrication are some of the critical issues encountered in STT-MRAM mass production. Some of the yield improvement techniques to combat the effect of process variations have been previously explored. However, little research has been done on defect oriented testing of STT-MRAM arrays. In this thesis, the author investigates the parameter deviation and non-idealities encountered during the development of a novel MTJ stack configuration. The characterization result provides motivation for the development of the design for testability (DFT) scheme that can help test and characterize STT-MRAM bit-cells and the CMOS peripheral circuitry efficiently.

The primary factors for wafer yield degradation are the device parameter variation and its non-uniformity across the wafer due to the fabrication process non-idealities. Therefore, effective in-process testing strategies for exploring and verifying the impact of the parameter variation on the wafer yield will be needed to achieve fabrication process optimization. While yield depends on the CMOS process variability, quality of the deposited MTJ film, and other process non-idealities, test platform can enable parametric optimization and verification using the CMOS-based DFT circuits. In this work, we develop a DFT algorithm and implement a DFT circuit for parametric testing and prequalification of the critical circuits in the CMOS wafer. The DFT circuit successfully replicates the electrical characteristics of MTJ devices and captures their spatial variation across the wafer with an error of less than 4%. We estimate the yield of the read sensing path by implementing the DFT circuit, which can replicate the resistance-area product variation up to 50% from its nominal value. The yield data from the read sensing path at different wafer locations are analyzed, and a usable wafer radius has been estimated. Our DFT scheme can provide quantitative feedback based on in-die measurement, enabling fabrication process optimization through iterative estimation and verification of the calibrated parameters.

Another concern that prevents mass production of STT-MRAM arrays is the defect formation in MTJ devices due to aging. Identifying manufacturing defects in the magnetic tunnel junction (MTJ) device is crucial for the yield and reliability of spin-torque-transfer (STT) magnetic random-access memory (MRAM) arrays. Several of the MTJ defects result in parametric deviations of the device that deteriorate over time. We extend our work on the DFT scheme by monitoring the electrical parameter deviations occurring due to the defect formation over time. A programmable DFT scheme was implemented for a sub-array

in 65 nm CMOS technology to evaluate the feasibility of the test scheme. The scheme utilizes the read sense path to compare the bit-cell electrical parameters against known DFT cells characteristics. Built-in-self-test (BIST) methodology is utilized to trigger the onset of the fault once the device parameter crosses a threshold value. We demonstrate the operation and evaluate the accuracy of detection with the proposed scheme. The DFT scheme can be exploited for monitoring aging defects, modeling their behavior and optimization of the fabrication process.

DFT scheme could potentially find numerous applications for parametric characterization and fault monitoring of STT-MRAM bit-cell arrays during mass production. Some of the applications include a) Fabrication process feedback to improve wafer turnaround time, b) STT-MRAM bit-cell health monitoring, c) Decoupled characterization of the CMOS peripheral circuitry such as read-sensing path and sense amplifier characterization within the STT-MRAM array. Additionally, the DFT scheme has potential applications for detection of fault formation that could be utilized for deploying redundancy schemes, providing a graceful degradation in MTJ-based bit-cell array due to aging of the device, and also providing feedback to improve the fabrication process and yield learning.

Acknowledgement

I would like to express my sincere gratitude and thanks to my supervisor Professor Manoj Sachdev, University of Waterloo, for believing in my thesis while I was lost. Without his guidance and patience, the work would not have been possible. I would also like to thank my co-supervisor, prof Youngki Yoon who has taught me how to research in this field. I would also like to thank my examination committee members, Dr. Guoxing Miao, Dr. David Nairn, Dr. Mustafa Yavuz, and Dr. Nicola Nicolici. I understand the effort put by all of you to be part of my Ph.D. examination committee and I am thankful for it.

I would also like to thank my colleagues for helping me and being part of many interesting conversations. It was great to have worked with you. My sincere thanks go to Adam, Sakib, Assem, Dhruv, Anthony, Kai, Qing, Sunil, Mahdi, Morteza, Hugo, Yugal, Ata, and Matthew.

Lastly, I would like to thank the memory devices group in Imec, Belgium for allowing me to work with real MTJ devices, CMC Microsystems for fabricating the test chips, and the University of Waterloo for funding and supporting my research.

Table of Contents

List of Tables x

List of Figures xi

List of abbreviations xvii

1 Introduction 1

1.1 Motivation 1

1.2 Evolution of MRAM and Future 3

1.3 MTJ Device Physics 4

1.4 MTJ Fabrication and Implementation 8

1.5 Test and Characterization of STT-MRAMs 12

1.6 Challenges in State-of-the-art STT-MRAM 13

1.7 Research Goal 14

1.8 Outline 14

2 STT-MRAM Design Considerations 15

2.1 Bit-Cell Operation 15

2.2 Write Operation 18

2.3 Read Operation 18

2.3.1 Reference Signal Generation Techniques 20

2.3.2 Sensing Topologies . . . 22
2.4 MTJ Physics & Modelling . . . 27
2.4.1 Compact Modelling . . . 28
2.4.2 Nanoscale Device Modelling based on NEGF . . . 29
2.4.3 Micro-magnetic Modelling . . . 30
2.4.4 Model Used for Bit-cell Analysis . . . 31
2.5 Preliminary Comparative Analysis of 1T-1MTJ STT-RAM Cells . . . 33
2.5.1 Simulation Methodology . . . 33
2.6 Results . . . 35
2.6.1 Read Performance . . . 36
2.6.2 Write Performance . . . 39
2.7 Summary . . . 41

3 MTJ Device Characterization **42**
3.1 Introduction . . . 42
3.2 Film-Level Characterization . . . 43
3.2.1 RA Product and TMR Evaluation . . . 43
3.2.2 Damping Factor . . . 43
3.2.3 Perpendicular Magnetic Anisotropy (PMA) . . . 44
3.3 Device-Level Characterization . . . 44
3.4 Objective . . . 48
3.5 Test Setup . . . 49
3.6 R_P Analysis . . . 50
3.7 TMR Analysis . . . 52
3.8 Summary . . . 52

4 A Parametric DFT Scheme **55**

4.1 Overview . . . 55

4.2 Approach for the DFT Scheme . . . 55

4.2.1 Parameter Generation Framework . . . 58

4.2.2 DFT-Cell Operation . . . 61

4.2.3 Replication of MTJ Characteristics Based on the DFT Circuit . . . 63

4.3 Compensating CMOS-Based Non-Idealities in the DFT Circuit . . . 64

4.3.1 Local Variation . . . 65

4.3.2 Global Process Corner and Temperature . . . 67

4.4 Test Chip Implementation and Results . . . 68

4.4.1 DC Resistance Voltage Behavior . . . 69

4.4.2 Switching Characteristics . . . 70

4.4.3 Retention Characteristics . . . 72

4.4.4 Transistor Area Overhead . . . 75

4.5 Application: Yield Characterization and process optimization . . . 76

4.6 Summary . . . 80

5 DFT for Long-Term Reliability **81**

5.1 Introduction . . . 81

5.2 Proposed BIST Scheme and Test Methodolgy . . . 83

5.2.1 DFT Circuit . . . 83

5.2.2 BIST Structure . . . 85

5.2.3 DFT Fault Classification and Identification . . . 87

5.2.4 Fault Analysis, Scheduling and Complexity . . . 89

5.2.5 Read-Sensing Path Characterization . . . 95

5.3 Case Study and Simulation . . . 99

5.3.1 Read-Sensing Circuitry Simulations . . . 99

5.4 Results . . . 104

5.5 Summary . . . 110

6 Conclusion **111**

6.1 Research Contribution 111

6.1.1 In-Die Parametric Characterization 111

6.1.2 Faster Wafer Screening and MTJ stack development 111

6.1.3 Bit-Cell Health Monitoring 112

6.1.4 65nm Test-Chip Design and Implementation 112

6.1.5 STT-MRAM Characterization 112

6.1.6 Future Work 113

References **114**

A Appendix **125**

A.1 Publications From This Work 125

List of Tables

1.1 Summary of the state of the art PMA-based STT-MRAM array implementations over the past years. 11

2.1 MTJ thermal stability needed for different memory capacity and FIT rates, Adapted from [1]. 18

2.2 MTJ device parameters. 32

2.3 MTJ structure dimension. 33

5.1 Parameter tuning range for the DFT cell. 85

5.2 Area estimate with respect to the total area of the STT-MRAM array . . . 106

List of Figures

1.1 STT-RAM performance comparison with other memory technologies [2]. . . 2
1.2 (a) Field induced MRAM, (b) Toggle MRAM, (c) TAS MRAM, and (d) STT-RAM. 3
1.3 Operating states of a magnetic tunnel junction (a) the MTJ symbol showing the free layer (FL) and pinning layer (PL) (b) AP state (c) P state. 5
1.4 (a) MTJ in parallel state, (b) in Anti parallel state. Adapted from [3]. . . . 6
1.5 Spin Torque Transfer (STT) mechanism,(a) AP → P,(b) P → AP. Adapted from [4]. 7
1.6 PMA-MTJ stack in a bit-cell. (a) MTJ stack layers, (b) MTJ states and switching, (c) STT-MRAM bit-cell. 8
1.7 Experimentally achieved TMR results reported till 2006. Adapted from [4]. 9
1.8 TEM image of the 50nm MTJ stack [5]. 10

2.1 Operating modes of 1T-1MTJ cell. 16
2.2 (a) Sense amplifer latch with bidirectional write driver, (b) illustration of bidirectional write operation. Adapted from [6]. 19
2.3 TMR variation vs. bias voltage applied across the MTJ for various MgO dielectric thickness, Adapted from [7]. 20
2.4 (a) Schematic of a conventional reference cell read scheme, (b) shows voltage created at the input of the sense amplifier , Adapted from [6]. 22
2.5 (a) Differential read scheme, (b) conventional read scheme with 2 reference cells, (c) Cross coupled current mirror amplifier based scheme, adapted from [8]. 23

2.6 (a) Dual array equalized reference scheme, (b) Simulation waveform. Adapted from [6]. 23
2.7 Current sense amplifier design using a single reference cell and a clamped reference, adapted from [9]. 24
2.8 Simulation waveforms for the sense amplifier (a) V_{WL}, (b) $SE1$, (c) $SE2$ (d) shows reference current (black), read '0' current (red) and read '1' current (e) Sense amplifier output for reading '0', (f) Sense amplifier output for a read '1'. 25
2.9 MTJ switching regimes vs. write pulse duration, adapted from [?] 27
2.10 (a) Shows the MTJ split into individual 2D layer unit cells (b) shows the NEGF representation of MTJ in form of 1D model. Adapted from [10]. . . 29
2.11 (a) 1T-1MTJ STT-RAM cell. (b) Four different configurations of bit cell. (c) Details of the extracted simulation parameters. The superscript of P_i or AP_i indicates the initial state of MTJ being P or AP mode. 34
2.12 MTJ hysteresis of 40-nm PMA-MTJ model. V_{MTJ} is defined on the PL with respect to the FL. 35
2.13 (a) R_{MTJ} and R_{TRAN} vs. V_{BL} for 1T-1MTJ cell with MTJ initialized in AP mode for case 1 and 4 (V_{WL} = 1.2 V; W_{NORM} = 1). (b) R vs. V_{BL} (V_{WL} = 1.2 V; W_{NORM} = 1) for all bit cell cases (the curves for cases 1 and 3 are overlapped). 36
2.14 (a) ΔI vs. V_{BL} (W_{NORM} = 1), (b) V_{ST}, (c) V_{BLM}, and (d) ΔI_{MAX} variation as a function of W_{NORM}. V_{WL} is 1.2 V in (a)(b)(c)(d). 37
2.15 ΔR vs. V_{WL} for (a) cases 1 and 3 (source follower) and (b) cases 2 and 4. V_{BL} = 0.2 V is used. ΔI vs. V_{WL} for (c) cases 1 and 3 (source follower) and (d) cases 2 and 4. V_{BL} = 0.2 V is used. 38
2.16 2D surface plots of (a) ΔI_{MAX}, (b) V_{BLM}, as a function of V_{WL} and W_{NORM} for case 2. 39
2.17 2D surface plots of $\Delta I_{MAX} - V_{BLM}$ product as a function of V_{WL} and W_{NORM} for (a) case 1, (b) case 2, (c) case 3 and (d) case 4. 40
2.18 (a) EDP vs. V_{BL} for case 4 (at V_{WL} = 1.6 V, W_{NORM} = 2). (b) The EDP_{MIN} vs. W_{NORM} (at V_{WL} = 1.6 V). 2D surface plots of the EDP_{MIN} as a function of V_{WL} and V_{BL} for a fixed width (W_{NORM} = 2) for (c) cases 2 and 3, and (d) cases 1 and 4. 41

3.1 Electrical characteristics of the MTJ device. (a) Shows the MTJ resistance vs. applied magnetic field for an R-H loop. (b) shows the R_P resistance vs. device dimension. (c) TMR of the MTJ device vs. device size. 44

3.2 Computed electrical diameter of the MTJ device vs. device size. 45

3.3 Magnetic characteristics of the MTJ device (a) shows the R-H loop showing the H_C and H_{off}. (b) H_C vs device size (c) H_{off} vs. device size. 46

3.4 Generalization of parameter spatial variation across the wafer. 47

3.5 R_P resistance plots for MTJ stack of 60, 100 and 150 nm sizes. (a) Spatial distribution of Normalized R_P resistance across wafer 1. (b) spatial distribution of R_P resistance on wafer 2 (CMOS substrate). 48

3.6 Parabolic fitting for the R_P spatial variation along the x and y axis of the wafer. 49

3.7 Mean to the standard deviation ratio for R_P resistance. (a) shows the spatial distribution for wafer 1 (b) spatial distribution for wafer 2 (CMOS substrate). 50

3.8 Normalized TMR median value for MTJ devices for 60, 100 and 150nm devices. (a) spatial variation on wafer 1. (b) spatial variation on wafer 2 (CMOS substrate). 51

3.9 Mean to standard deviation ratio for TMR (a) wafer 1. (b) wafer 2 (CMOS substrate deposition). 52

3.10 Application of the spatial variation model on to a generic device model. . . 53

4.1 (Left panel) STT-MRAM fabrication process flow, where the wafer testing and pre-qualification step is included. (Right) The process for testing and qualification adopted is shown in detail. The grey blocked region illustrates the proposed scheme. 56

4.2 (a) Generalized MTJ device parameter variations observed during single device level characterization. (b) MTJ device parameter mean and variance translated to control voltage. (c) Top level diagram for the DFT scheme. Here BL and SL are bit-line and source-line of the MTJ column. (d) Typical parameter values used for MTJ. 57

4.3 Parameter generation framework to generate bias voltages for the DFT circuit operation. 59

4.4 DFT cell consisting of R-V bias circuit, latch, and the NMOS access transistor. (b) Latch circuit. (c) Generalized R-V bias circuit. BL, SL and WL represent bit-line, source-line and word-line inputs, respectively. 62

4.5 Selective control of DFT characteristics using individual control voltages. (a) R_{DS} resistances contributed by each transistor group in the DFT circuit. (b) Changing the R_P resistance using V_{DD_P}. (c) Impact of changing V_G on RAP. Here $V_G = V_{G0} = V_{G1} = V_{G2} = V_{G3}$. (d) Changing the R_{AP} resistance using V_{SS_AP}. (e) DFT cell parameter tuning range. The results were based on 65–nm process pre–layout simulations. 64

4.6 R-V variation for different columns of the DFT cell arrays based on post layout simulations. (a) DFT_COL_V0 (L = 1X), (b) DFT_COL_V1 (L = 2X), (c) DFT_COL_V2 (L = 3X), and (d) DFT_COL_V3 (L= 4X). 65

4.7 (a) Process parameters for the target MTJ characteristics (b) Summary of the resistance variation due to transistor variability for each design. 66

4.8 Impact of global process corner (a)-(c) before correction and (d)-(f) after correction. R_{DFT} vs. V_{DFT} for (a,d) AP state and (b,e) P state. (c,f) V_{SW} vs. t_{pw} for P to AP. TT, FS, SF, SS, FF represent typical-typical, fast-slow, slow-fast, slow-slow, fast-fast, respectively. 67

4.9 Temperature dependency behavior of the DFT circuit resistance. The RV loop is shown for (a) 25° C, (b) 85° C, and (c) 125° C (b) RV bias compensation voltage based on post-layout simulation. 69

4.10 Die micrograph of the test chip with peripheral circuitry in 65-nm CMOS technology. 70

4.11 Schematic design for the R-V bias circuit used for DFT cell. (b) Sizing and voltage for the transistors used in the design. 71

4.12 (a) Bit-cell current (I_{CELL}) vs. voltage (V_{BL}–V_{SL}) loop from measurement. (b) DFT vs. V_{DFT} characteristics of the DFT circuit from measurement and post-layout simulation. 72

4.13 Block diagram of column with the shared tunable capacitor bank circuitry. 73

4.14 Post-layout simulation waveforms for the DFT cell switching operation. . . 74

4.15 (a) Switching voltage vs. write pulse width for AP→P and P→AP switching. (b) DFT cell switching probability measured for different thermal stability factors (Δ). The markers and the lines represent the chip measured data and model simulation results, respectively. 75

4.16 Read error rate, showing the read-disturb mechanism exhibited by the DFT cell for different thermal stability factors (Δ). The markers and the lines represent the chip measured data and model simulation results, respectively. 76

4.17 Read access path used for yield characterization simulation of the STT-MRAM column. 77

4.18 Yield characterization and optimization process flow chart. 78

4.19 RA product vs. radial distance from the center of wafer. The markers represent DFT cell output values. (b) Resistance distribution profile for AP and P states at 0 mm, 50 mm and 80 mm radial distance from the center of the wafer based on post layout simulation results. (c) 3-sigma read margin voltage based on sense amplifier outputs across the wafer. 79

5.1 DFT cell used for testing and replication of fault characteristics. 84

5.2 (a) System level diagram of the BIST scheme. (b) Device under test (DUT), is the conventional STT-MRAM sub-array consisting of bit-cells, read sensing column and multiplexers. 86

5.3 BIST engine diagram within the sub array. 87

5.4 Bias voltage generation and analog multiplexer array for selecting DFT control voltages. 88

5.5 Generalized parameter range of the MTJ device with defect. 89

5.6 Flow chart for classification and identification of faults using the DFT scheme. 90

5.7 (a) Switching test cycle. (b) Block diagram of column with the shared tunable capacitor bank circuitry. 91

5.8 (a)Test scheduling based on priority for different tests. (b) Test complexity for each test. Here, N and L correspond to the number of bit-cells in a bit-line column and number of read sense paths accessed simultaneously. M corresponds to the number of bits used to select the bias control voltage. . 93

5.9 Detection circuit within a sub-array. The table shows the different configuration modes for the circuit. 94

5.10 Schematic for the column read sense path and the current sense amplifier (CSA). CC is the compensation circuit. 96

5.11 (a) Simplified equivalent circuit model for offset current calculation. (b) Offset characterization of read sensing circuitry using the DFT cells. 97

5.12 (a) The read offset compensation circuit. (b) The offset compensation current from PMOS (I_{OFFP}) and NMOS (I_{OFFN}) chains. 98

5.13 Sense amplifier circuit design used for characterization. (a) Type I CLSA circuitry. (b) Type II VLSA sense amplifier circuit design. 100

5.14 Post layout simulation showing the read operation in a CLSA sensing scheme. 101

5.15 Statistical yield simulation results for the read sensing yield as a function of the bit–line. 102

5.16 Bias voltage generation and analog multiplexer array for selecting DFT control voltages. 103

5.17 R_P vs. pin-hole area. The resistance degradation based on MATLAB plot is shown in blue. The black and red corresponds to the measured resistance from the test structure, and the simulation respectively. The dot represents the mean value and the bounds represent the worst case max and min values. 104

5.18 Die micrograph of the DFT sub-array implemented in 65 nm bulk CMOS process. 105

5.19 Impact of temporal variation based on the test chip measured. (a) The SAOUT distribution vs V_{BIAS} for ideal (blue) and temporal variation case (red). (b) Standard deviation of the V_M for different resistance measured. 106

5.20 Impact of spatial variation, (a) Measured DFT variation along the column, the bias voltage for SA distribution with V_M is shown on the top right. (b) Distribution from the DFT cells obtained for a given SA location. (c) Resistance distribution before (top) and after (bottom) read path circuit offset compensation. The red line indicates the injected fault voltage and the blue line indicates the measured match voltage (V_M) bottom shows the corresponding resistance values computed from post layout circuit simulation. 107

5.21 Worst case LSB bin errors contributed by read sense amplifier, read voltage clamp circuit and the DFT circuit in the proposed scheme. 109

5.22 Resistance identification accuracy based with measurement and post-layout simulations for the DFT scheme pre-compensation (red solid line) and post-compensation (blue dashed line). The red stars and blue dots represent measured values before and after compensation. 110

List of abbreviations

1T-1MTJ 1 Transistor 1 MTJ

AP Anti-Parallel

BEOL Back-End-of-the-Line

BL Bit-line

bcc Body Centre Cubic

BIST Built in Self Test

CC Compensation Circuit

CIPT Current In-Plane Tunneling

CMOS Complementary Metal Oxide Semiconductor

CLSA Current Latch Sense Amplifier

CSA Current Sense Amplifier

DFT Design for Testability

DOS Density of States

DRAM Dynamic Random Access Memory

DUT Device Under Test

EDP Energy Delay Product

fcc Face Centre Cubic

FEOL Front-End-of-the-Line

FIMS Field Induced Magnetic Storage

FeRAM Ferroelectric Random Access Memory

FM Ferromagnetic Material

HKMG High-K Dielectric metal Gate technology

IMA In-Plane Magnetization Anisotropy

imec Interuniversity Microelectronics Centre

IPMA Interfacial Perpendicular Magnetic Anisotropy

IQR Inter Quartile Range

LSB Least Significant Bit

LLG Landeu Lifshitz Gilbert

MRAM Magnetic Random Access Memory

MTJ Magnetic Tunnel Junction

NEGF Non-Equilibrium Greens Function

PMA Perpendicular Magnetic Anisotropy

P Parallel

PRAM Phase Change Random Access Memory

PMA Perpendicular Magnetic Anisotropy

RAM Random Access Memory

RRAM Resistive Random Access Memory

RA Resistance-Area Product

RH Resistance - Magnetic Field

RSM Read Sense Margin

RV Resistance - Voltage bias

SA Sense Amplifier

STT Spin Torque Transfer

SRAM Static Random access memory

TAS Thermal Assisted Switching

NVM Non-Volatile Memory

SAF Synthetic Anti-Ferromagnet

SL Source-line

TAS Thermal Assisted Switching

TEM Transmission Electron Microscopy

TMR Tunnelling Magnetic Resistance

VCMA Voltage Controlled Magnetic Anisotropy

VSM Vibrating Spectral Magnetometry

WL Word-line

Chapter 1

Introduction

New technologies are changing the way we see embedded memory architectures. Motivated by the need for lower power and higher densities, new non-volatile memory technologies such as STT-MRAM, Resistive RAM, and PCRAM are likely to become a viable replacement for the DRAM technology in the near future. From the given class of memories, the Spin-Torque-Transfer (STT) Magnetoresistive Random-Access Memory (MRAM) has become a potential candidate for emerging non-volatile memory (NVM) devices [11]. In this chapter, we are introducing the non-volatile memories with emphasis on STT-MRAMs. We explore the motivation for STT-MRAM and the history of MRAM technology. The basic building block of the STT-MRAM is the Magnetic Tunnel Junction (MTJ) used in the bit-cells to store data. Firstly, we describe the functionality of the MTJ device and the physics of operation. Secondly, we look into the fabrication of MTJ devices and challenges faced during the mass production of large-scale STT-MRAMs. The yield and reliability limitations that prevent STT-MRAM from mass production are the main focus driving this research. The chapter concludes with the research goal and the outline.

1.1 Motivation

A good candidate for universal memories should have high read and write speeds, unlimited endurance, low power consumption. STT-RAMs have the potential to achieve the speed of SRAM, the density of DRAM with low power consumption, and it is a non-volatile solution, making it an ideal candidate for a universal memory. Figure 1.1 shows the performance comparison of STT-RAM with other memory technologies.

	SRAM	DRAM	Flash (NOR)	Flash (NAND)	FeRAM	MRAM	PRAM	STT-RAM
Non-volatile	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Cell size (F^2)	50–120	6–10	10	5	15–34	16–40	6–12	6–20
Read time (ns)	1–100	30	10	50	20–80	3–20	20–50	2–20
Write / Erase time (ns)	1–100	50 / 50	1 μs / 10 ms	1 ms / 0.1 ms	50 / 50	3–20	50 / 120	2–20
Endurance	10^{16}	10^{16}	10^{5}	10^{5}	10^{12}	$>10^{15}$	10^{10}	$>10^{15}$
Write power	Low	Low	Very high	Very high	Low	High	Low	Low
Other power consumption	Current leakage	Refresh current	None	None	None	None	None	None
High voltage required	No	2 V	6–8 V	16–20 V	2–3 V	3 V	1.5–3 V	<1.5 V
	Existing products						*Prototype*	

Figure 1.1: STT-RAM performance comparison with other memory technologies [2].

Comparing with SRAM cells, STT-RAM is approaching the read and write speed of the SRAM cell designs, and has better performance than other non-volatile memory technologies. Power consumption can be compared by looking at the static leakage and dynamic power consumption during memory access. STT-RAM's have low static power consumption due to its non-volatile nature with very low leakage current compared to SRAM cells. Considering dynamic power, STT-RAM's need higher write power consumption at high speeds compared to SRAM cells. This increase in switching current at high write speeds is due to shifting of the MTJ switching regimes. Several new switching strategies are being explored to reduce switching current. The STT-RAM has been considered actively as a replacement for on-chip SRAM cache memories because of its promising results [12] [13]. STT-RAMs are also compared with DRAM in terms of circuit density and performance. They may not provide the same density advantages as the DRAM cells [14]. Comparison of read and write performance shows that the STT-RAM has the same read performance as DRAM's, but its inferior in terms of write performance.

Research is also done on alternative non-volatile technologies for endurance, but most of them are not meeting the endurance capabilities (E = Number of cycles of successful

operation/life-time) of STT-RAM. Some of the memory technologies and their endurance values are as follows. a) Phase change memory(PRAM) (E $=10^{10}$), b) Resistive change memory (RRAM) (E $= 10^{10}$, c) FeRAM ($E = 10^{12}$), d) Flash Technologies (10^5) lack the endurance of existing SRAM, DRAM technologies. Currently used STT-RAM memories have an endurance of 10^{15}

1.2 Evolution of MRAM and Future

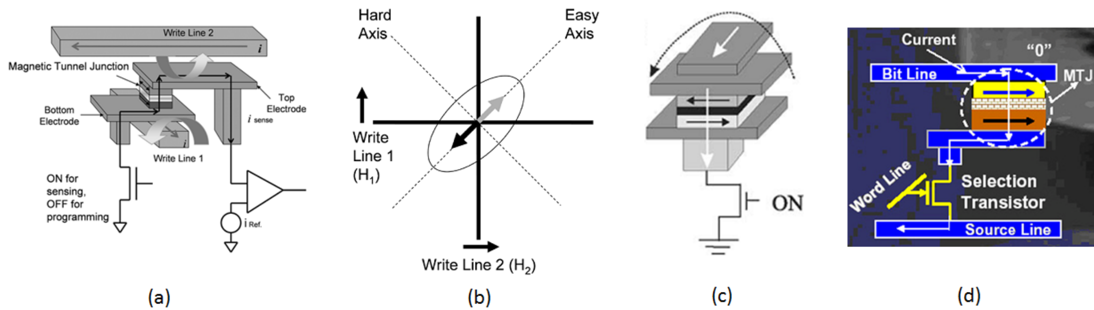


Figure 1.2: (a) Field induced MRAM, (b) Toggle MRAM, (c) TAS MRAM, and (d) STT-RAM.

MRAM development can be classified based on the development of write mechanisms employed for each generation of MTJ devices. Figure 1.2 shows the evolution of MRAM write technology. The write operation in the MRAM has changed drastically with the change in MTJ technology. The first generation of MRAM, called the field-induced magnetic storage (FIMS)-MRAM [15], relied on passing a current through the write line. Here the write line was placed close to the free layer so that they are magnetically coupled. A high current is then passed through the write line to magnetize the free layer. However, there were limitation to this writing scheme such as half-selection problem of bit cells and scalability issues. The half selection problem occurs in FIMS-MRAM, where all the cells in the column are partially selected when only a single cell in the column is accessed. The half-selection problem was mitigated by the introduction of toggle mode MRAM [16]. The Toggle-RAM shown in Fig. 1.2(b) uses Savtchenko switching to switch the Synthetic Anti-Ferromagnetic (SAF) MTJ structure. Here two perpendicular write lines 1 and 2 are used in a two-step write sequence to program the MTJ.

Another technique adopted was to program the MTJ in a thermally assisted manner. These class of MRAM's were called as TAS MRAM [17]. Initially, a current is passed through the

MTJ junction to heat it above the curie temperature (T_C) as shown in Fig. 1.2(c). Then a current pulse is passed through the write line to magnetize the free layer. Here the free layer is at a temperature greater than the curie temperature ($T \geq Tc$), hence the MTJ can be switched with a smaller write current. This memory writing technique also poses major drawbacks, such as a longer time to heat the MTJ element and scalability issues.

Figure. 1.2(d) shows the write process in a typical STT-MRAM cell. The theoretical prediction for writing mechanism based on spin torque transfer (STT) technique was proposed by Slonzewski [18]. STT process involves a spin polarised current being passed through the MTJ, where the polarised electrons passing through impart its angular momentum to the free layer, magnetizing the material. Currently different novel switching mechanisms are being investigated to improve the write performance of the MRAM devices. Current focus of research is on the MTJ device physics level to improve the write performance of the MTJ. The breakdown in the dielectric barrier is one of the most prominent failure mechanisms in PMA-MTJ devices that work based on STT write mechanism. Scaling with constant write energy demands decreasing of the RA product which requires decreasing in the MgO barrier thickness. This would lead to reliability issues due to extremely thin layer thickness of the dielectric material. Alternate device-level techniques such as voltage-controlled magnetic anisotropy (VCMA) [19] or spin-orbital torque (SOT) assisted STT [20] promises a reduction in the switching current without reducing MgO thickness, that could alleviate the stress on the dielectric material. Some of the alternate areas of research are a) Multiferroic [21] b) Spin-Hall effect switching c) Magnetostrictive switching [22].

1.3 MTJ Device Physics

Magnetic tunnel junction consists of 2 ferromagnetic (FM) materials separated by a dielectric material. The dielectric thickness is set at a few angstroms so that electrons can tunnel through the insulation barrier. Julliere proposed a model for the conduction of electrons through such a structure. Based on his proposal, the electron conduction consists of spin-up and spin-down polarized electrons, and the spins are conserved while tunneling through the dielectric material from one FM layer to another. The conductance across the junction depends on the relative magnetization between the fixed and the free layer of the ferromagnetic material. The state of the MTJ when $\theta = 0°$ is called as the Parallel (P) state. The MTJ has its maximum conductance (minimum resistance R_P). The state of the MTJ when $\theta = 180°$. The MTJ exhibits the least conductance here (maximum resistance R_{AP}) is the Anti-Parallel (AP) state. The symbolic representation of AP and P states are as shown in the Fig 1.3.

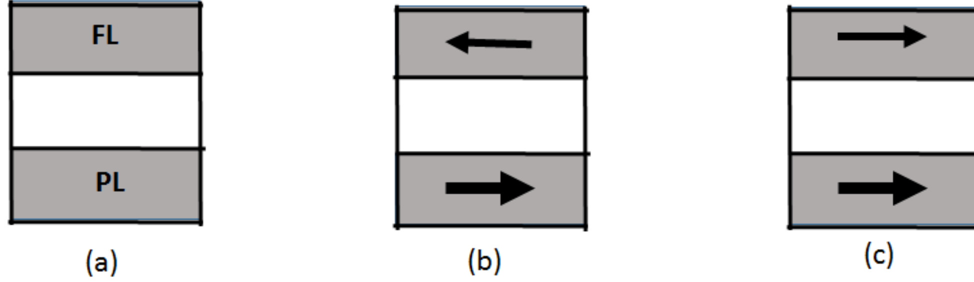


Figure 1.3: Operating states of a magnetic tunnel junction (a) the MTJ symbol showing the free layer (FL) and pinning layer (PL) (b) AP state (c) P state.

The equation for conductance is described as shown

$$G(\theta) = \frac{G_P - G_{AP}}{2} + \frac{G_P + G_{AP}cos(\theta)}{2} \tag{1.1}$$

From this the maximum Tunnelling Magnetic Resistance (TMR) ratio can be derived as ($\theta = 0°$)

$$G(\theta) = \frac{G_P - G_{AP}}{G_{AP}} \tag{1.2}$$

$$R(\theta) = \frac{R_{AP} - R_P}{R_P} \tag{1.3}$$

where G_{AP} is the conductance when $\theta = 180°$, and G_P is the conductance when $\theta = 0°$. R_{AP} and R_P are corresponding resistances respectively. As per classical band theory, the reason for the TMR effect in such hetero-structures is due to the difference in density of states (DOS) for spin up (N↑) and spin down (N↓) electrons at the Fermi level (E_F) for each of the sandwiched ferromagnetic material. When the dielectric thickness approach the tunneling barrier thickness, the spin polarised electrons can tunnel through the barrier into their corresponding sub-band without losing its spin momentum as shown in Fig 1.4. Here the MTJ conductance for a given spin is proportional to the product of the DOS(E_F) at each ferromagnetic material. During parallel state, the sub-bands for one of the spins (either spin up or spin down), have higher DOS product ($N_{FM1} \uparrow \times N_{FM2} \uparrow \quad or \quad N_{FM1} \downarrow \times N_{FM2} \downarrow$), resulting in higher conductance. However, for AP state, the product of the DOS values from the FM material are minimum, resulting in a lower conductance. In this perspective, the polarisation of electrons in a FM layer can be defined as

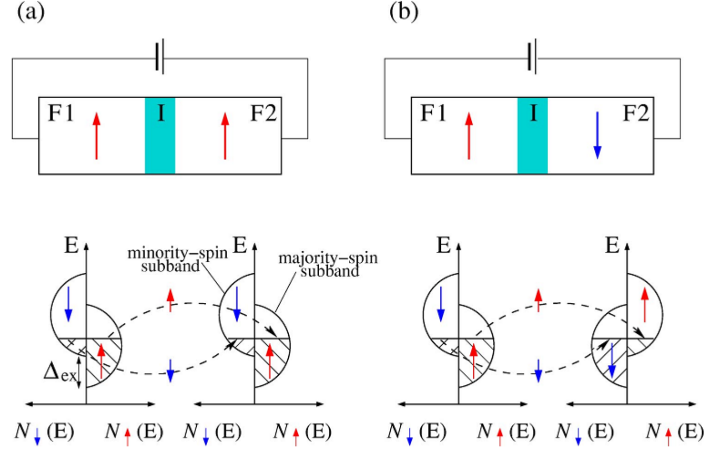


Figure 1.4: (a) MTJ in parallel state, (b) in Anti parallel state. Adapted from [3].

$$P = \frac{N\uparrow - N\downarrow}{N\uparrow + N\downarrow} \tag{1.4}$$

Fig. 1.5(a) shows the case for AP → P switching where electrons are passing through the MTJ material from the pinning layer to the free layer. The electrons get spin polarised after passing through the 1st layer of ferromagnetic material (pinning layer). This spin polarised current then tunnels through the thin dielectric material and interacts with the 2nd ferromagnetic layer (free layer), where the interaction leads to the electrons imparting its spin angular momentum to the free layer, resulting in the magnetization switching of the free layer. Figure 1.5(b) shows the case for P → AP switch, where the electrons move from free to pinning layer. Here the tunneling process is similar to the previous case, however, the spin polarised current is reflected into the free layer, toggling its existing magnetization state to AP mode. The P → AP switching is a lower efficiency process compared to the (a), hence the switching threshold current are asymmetrical $I_C(AP \rightarrow P) \neq I_C(P \rightarrow AP)$.

Initial STT based MTJ structures used In-Plane Magnetization Anisotropy (IMA), where the magnetization plane is parallel to the MTJ plane. Due to MTJ scaling, the read operation of MTJ became more error-prone and the data retention capability was affected. This could mean that the thermal stability for IMA based MTJ structures drop rapidly with device scaling. Additionally, during the write operation, the energy has to be spent to counteract the unnecessary demagnetization field of the free layer[23], which does not contribute to thermal stability. In order to avoid these effect, the MTJ materials with perpendicular magnetic anisotropy (PMA) was introduced. Here the MTJ is magnetized

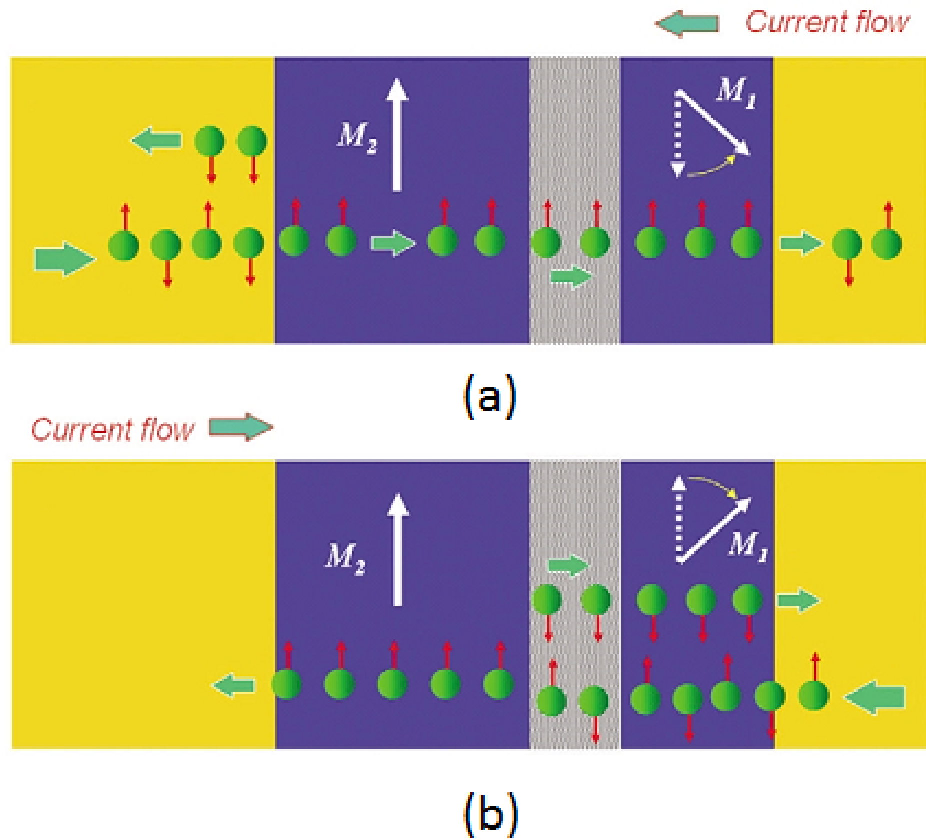


Figure 1.5: Spin Torque Transfer (STT) mechanism,(a) AP $\rightarrow$ P,(b) P $\rightarrow$ AP. Adapted from [4].

perpendicular to the plane by arranging the crystalline structure of the free layer material perpendicular to the MTJ plane. The technique severely reduced the write current needed without sacrificing on the thermal stability, which allowed for the scaling of the devices to diameters of 40 nm and below. The composition of the modern MTJ stack is shown in Fig. 1.6(a). The pinned layer provides a stable reference magnetic orientation for accessing the data stored in the MTJ stack. Depending on the magnetization of the free layer to its anisotropic easy axis provided by the pinned layer, the MTJ can be defined to be in Anti-Parallel (AP) state or Parallel (P) state as shown in Fig. 1.6(b). MTJ exhibits resistance in both parallel state (R_P) and anti-parallel state (R_{AP}). Fig. 1.6(c) shows the conventional STT-MRAM bit-cell consisting of the MTJ device and the NMOS access transistor.

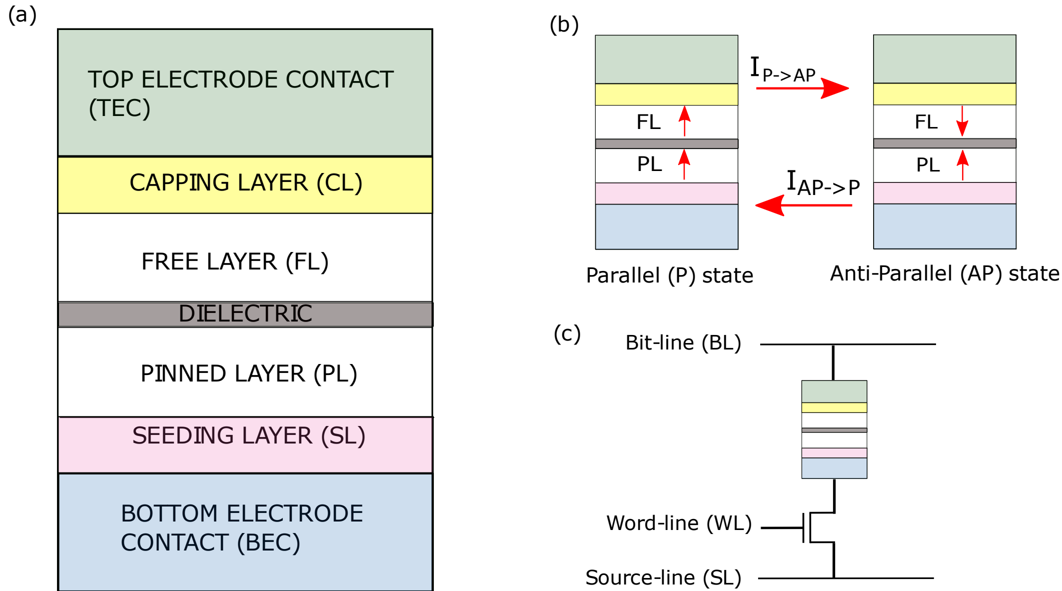


Figure 1.6: PMA-MTJ stack in a bit-cell. (a) MTJ stack layers, (b) MTJ states and switching, (c) STT-MRAM bit-cell.

1.4 MTJ Fabrication and Implementation

Materials and fabrication techniques can play a crucial role in deciding the quality of the MTJ created. Juliere demonstrated the TMR effect using a Fe/Ge/Co junction which shows a TMR of 14% at a temperature of 4 K [24]. This work was improved to get higher TMR values (15%) at room temperature using AlO_X as the dielectric material. The material processing techniques were improved further to yield a TMR of up to 70% using junctions such as CoFeB/AlOx/CoFeB. Several theoretical predictions were made for higher TMR using the crystalline structure of MgO dielectric which would make the tunneling process to be more dependant on the spin of electrons. Two major experimental data proved this concept, Yuasa et.al proposed an MTJ structure where the MgO layer was fabricated using molecular beam epitaxy yielding a TMR of 180% [7]. Another structure proposed by Parkin et.al uses a sputtering technique for MgO deposition that showed a TMR of 220% [25]. The Fig. 1.7 shows the rise of TMR over the past years [4].

Current STT-MRAM bit-cells utilize magnetic tunnel junction (MTJ) with perpendicular magnetic anisotropy (PMA) [23]. These MTJ stacks are deposited in a post-CMOS back-

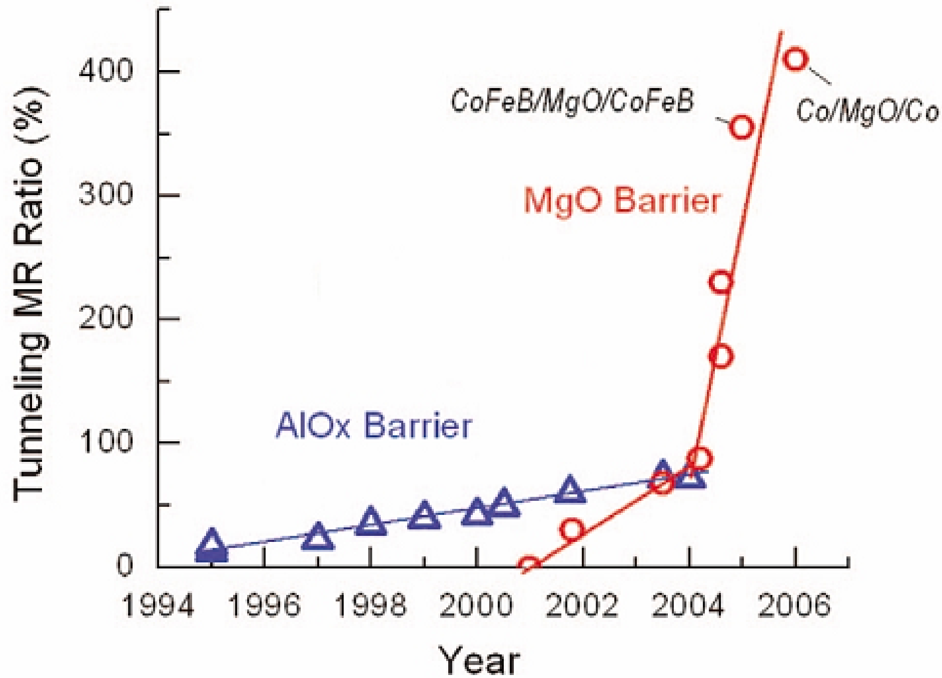


Figure 1.7: Experimentally achieved TMR results reported till 2006. Adapted from [4].

end-of-the-line (BEOL) deposition process [26][27][28] in the regular CMOS fabrication flow. The PMA-MTJ stacks is composed CoFeB ferromagnetic material for the pinned (PL) and the free layer (FL). The MgO dielectric material is used for seperating the ferromagentic materials. The transmission electron microscopy image of the current PMA MTJ stack for a 50 nm MTJ device is as shown in Fig. 1.8. The MTJ integration schemes and the level of metallization schemes are critical in forming uniform deposition over the wafer. Table 1.1 summarizes the features and the challenges addressed during the implementation of PMA-based STT-MRAM array over the years. Asymmetrical write switching current has been an issue during the advent of STT-MRAM bit-cells. This was equalized by reverse connecting the access transistor to the MTJ device and using its source degeneration effect [29]. During the initial years of PMA-based STT-MRAM, the MTJ stack integration has been highly dependent on the CMOS node, and the stack was optimized uniquely for each process. However, this changed when the STT-MRAM integration scheme started following existing fabrication processes such as the high-density DRAM process [30]. Also, the need for standardization of the STT-MRAM integration process motivated the MTJ

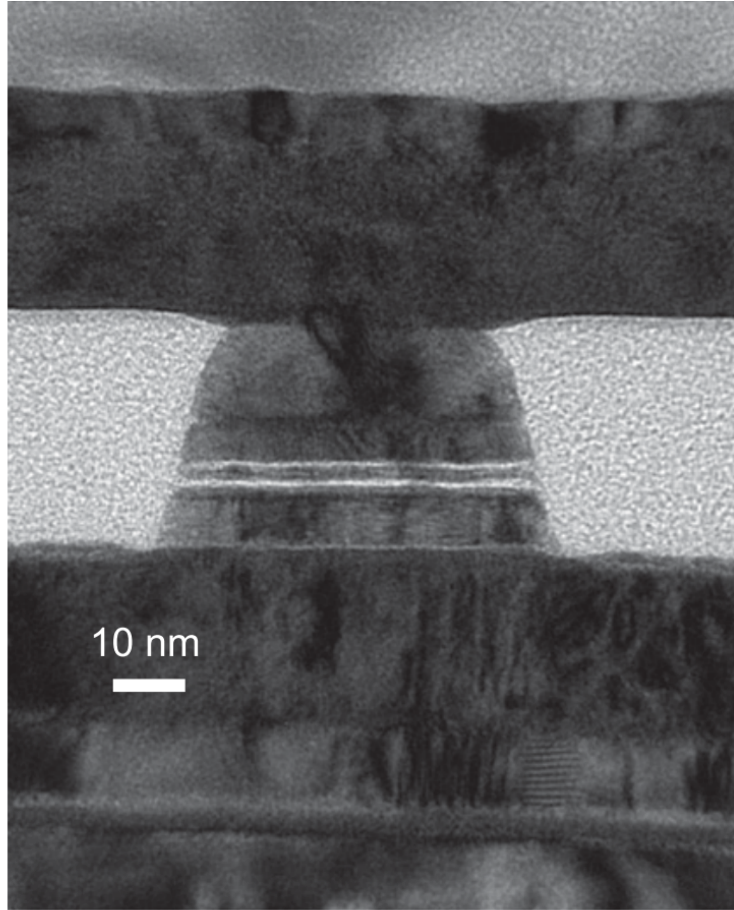


Figure 1.8: TEM image of the 50nm MTJ stack [5].

fabrication and deposition to be compatible with bulk, FINFET, and FDSOI CMOS processes. Thus current MTJ stack is implemented at higher metal layers to decouple the stack from underlying front-end-of-line (FEOL) layers. The work by Lee et al. [11] proposed an integration scheme that is compatible with bulk, FDSOI or FinFET processes in the 28-nm node. This implementation methodology provides opportunities for incorporating test structures that connect to lower metal levels for initial characterization and testing of peripheral circuitry in the CMOS process prior to and post MTJ deposition. Additionally with scaling, the interconnect resistance is rapidly increasing, forcing the implementations to adopt 2 metal schemes for bit-line implementation [31]. The CMOS process scaling also results in increased BL and SL parasitic resistance variation impacting the read and write performance of the bit-cells. Modern designs are incorporating low-power read and write

Publication	Lin et al.	Lu et al.	Song et al.	Rho et al.	Shum et al.	Shih et al.	Wei et al
Technology	TSMC 45 nm	40 nm	Samsung 28 nm	-	Global foundries 2x CMOS	TSMC 40 nm	Intel 22 nm
Core Voltage	1.1 V	1.2 V	1.0 V	1.8 / 1.2 V	-	1.1 V	1.1 V
Metalization	M3	M4	-	-	M5	M3	M2
Features	(a) Reverse connected for write current equalization	(a) 20 ns read access (b) Lowest write power (3.2 μW / MBps)	(a) Novel integration, patterning and stacking. (b) 10 year retention at 85 deg C	(a) Hierarchical bit-line (b) self-aligned DRAM process	(a) High density and process compatibility with HKMG, Finfet, FDSOI (b) Flash memory replacement	(a) Read current resolution $<$ +/- μA	(a) Write-verfiy-write scheme (b) Low voltage read (4 ns at 0.9 V)

Table 1.1: Summary of the state of the art PMA-based STT-MRAM array implementations over the past years.

schemes. The read sense margin is directly impacted by the bit-line parasitic resistance and the accuracy of the reference current generated. Thus, the accurate characterization of parasitic resistance variation across bit-lines is needed to adopt appropriate resistance matching schemes. To mitigate the impact of bit-line resistance and capacitance, hierarchical bit-line schemes [30] are adopted for larger memory sizes.

1.5 Test and Characterization of STT-MRAMs

Parametric testing has been widely used for MRAM technology development [32], which complements the simulation process by offering better coverage of testing conditions on the processed wafer. The previous works used analog multiplexer modules for monitoring cell resistance characterization and digital testing [32]. Alternatively, other strategies for memory built-in self-testing are employed to detect stochastic retention failures in STT-MRAM arrays [33]. Although these techniques characterizes specific MTJ cell resistance distribution issues in the lab environment, it provides limited avenues for exploration of parameter sensitivity. The MTJs have limitations in terms of its parameter tunability once deposited. Ideally, fabrication optimization requires the viability of different parameter tuning options. For instance, understanding the contribution of each MTJ device parameter, such as resistance-area (RA) variation or tunnel magnetoresistance (TMR), to the yield could be achieved by implementing MTJ stacks with different parameter combinations on multiple wafers. However, this strategy would be cumbersome due to the numerous process parameter options to be manipulated and would require significant wafer turnaround time and effort.

Furthermore, the imperfections in the manufacturing process raise several reliability concerns during the long term operation of the MRAM. The issues have grown in importance in light of recent scaling trends in sub-100 nm device dimensions [34][35]. The imperfections during the manufacturing process result in the formation and growth of defects in the MTJ stack layers. The perpendicular magnetic anisotropy (PMA) property of the MTJ device is obtained from the hybridization of atoms near the dielectric/free layer/capping layer interface. This interfacial PMA in MTJ devices suffers significantly from defects [34], non-uniformity of the stack layers and the device surface roughness [36]. For instance, a defect formed at the interface could alter the electrical and magnetic properties of the device from the intended design specification, creating a fault in the device [37]. Thus, defect formation during the MTJ stack deposition and integration should be monitored to improve the yield and reliability of STT-MRAM arrays.

Many recent studies have focused on the functional testing and characterization [38] to identify the reliability issues and accelerate the STT-MRAM wafer production [32]. Design-for-testability (DFT) structures have been proposed to identify various recoverable failure mechanism, such as read-disturb condition [39], in STT-MRAM bit-cells. However, there have only been few studies on monitoring defect-oriented failures in MTJ. Periodically monitoring the quality of MTJ stack may provide the information of the defect and its progression, which could be utilized to identify the onset of a fault during a read/write operation. A way to monitor the defect symptom is to evaluate the deviations in electrical

and magnetic characteristics of STT-MRAM bit-cells. However, observing large samples of data obtained from bit-cells in an array can be quite cumbersome. One of the solutions to identify the deviation is to compare the MTJ electrical characteristics with a known reference device that is not affected by the same defect-causing mechanisms. The impact of variations on MOS transistors has been extensively studied [40] and thereby can be designed to replicate the intended reference behavior of the MTJ device with minimum variations [41].

1.6 Challenges in State-of-the-art STT-MRAM

Spin torque transfer (STT)-magnetoresistive random-access memory (MRAM) has come a long way in research to meet the speed and low power consumption requirements for future memory applications [11]. The STT-MRAM bit-cells are ready for mass production; however, the imperfections in the MTJ manufacturing process still raise reliability and yield concerns. The MTJ based bit-cells consumes most of the array space and is the most crucial part of the STT-MRAM array. The process repeatability and yield stability during wafer fabrication for advanced novel MTJ stack is one of the critical issues limiting STT-MRAM mass production [26].

The primary factors for wafer yield degradation are the MTJ device parameter variation and its non-uniformity across the wafer due to the fabrication process non-idealities. An extensive study of fault models [42] and statistical modeling of devices before fabrication is done to represent parametric variations and achieve the optimum design parameter values for STT-MRAM array fabrication. In spite of these design efforts, the fabricated wafer could result in yield loss due to non-idealities in fabrication process steps. Thus in-die characterization is essential for maintaining high yield and reliability specifications demanded by current memory applications.

The PMA property of the MTJ device is obtained from the hybridization of atoms near the dielectric / free-layer / capping-layer interface. This interfacial PMA in MTJ devices suffers significantly from defects [34], non-uniformity of the stack layers, and device surface roughness [36], which could alter the electrical and magnetic properties of the device from the intended design specification [37]. After fabrication, the non-idealities embedded within the device could result in the formation and growth of defects within MTJ layers over time. The issue has grown in importance in light of recent scaling trends in sub-100 nm devices [34][35]. One such mechanism is the pin-hole defect formation in MTJ barrier layers [43][44]. Pin-hole defect grows over time with use and result in the failure of the MTJ device completely. In order to ensure good yield and reliability for STT-MRAM

arrays in large density arrays, we need to explore

1. A way to analyze the sensitivity of each device parameter to the system level yield on the wafer immediately after fabrication. Effective in-process testing strategies for exploring and verifying the impact of the parameter variation on the wafer yield will be needed to achieve fabrication process optimization.
2. We need a bit-cell health monitoring scheme for observing and controlling defect formation online after fabrication.

1.7 Research Goal

The main focus of the research is to explore the issues in yield and reliability encountered during advanced MTJ stack development that limit the STT-MRAM production. We develop an on-chip embedded DFT scheme for the test and characterization of bit-cell arrays and peripheral CMOS circuitry. The DFT cells allow to replicate the MTJ device electrical behavior accurately and, with a BIST scheme, is utilized for monitoring aging degradation effects in MTJ devices. The DFT scheme is implemented in 65 nm CMOS technology and its performance and accuracy are verified both through simulation and test chip measurement results.

1.8 Outline

The rest of the chapters are organized as follows. Chapter 2 gives the background for the current research. We discuss the major building blocks of the MRAM array such as bit-cell design, read and write circuitry. In addition, the modeling approach that was used for the analysis is discussed. Chapter 3 explores the MTJ device-level characterization performed to analyze the variations observed during advanced MTJ stack development. The results and possible implications of the work are discussed. Chapter 4 we introduce a parametric DFT scheme for peripheral circuitry characterization. Chapter 5 discusses the BIST scheme that utilizes the DFT cells to monitor aging effects observed in STT-MRAM bit-cells. Chapter 6 concludes the report and summarizes the contributions of this research.

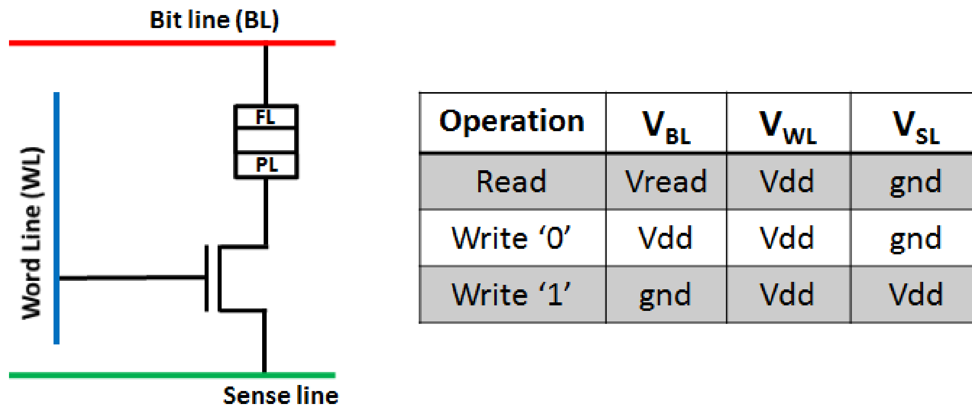
Chapter 2

STT-MRAM Design Considerations

2.1 Bit-Cell Operation

A full-scale bit-cell MRAM array that utilizes magnetic tunnel junction was implemented and reported in 2005 by Sony, who has demonstrated a 180 nm 1-MB STT-RAM array [45]. A basic STT-RAM bit cell consists of MTJ and an access NMOS transistor. The schematic and layout of a 1 transistor - 1 MTJ (1T-1MTJ) STT-RAM bit-cell design is as shown in Fig. 2.1. The STT-MRAM bit-cell works on the principle of Spin-Torque-Transfer, where the bit-cell is programmed by injecting a spin-polarised current through the MTJ. The bit-cell has two modes of operation, read and write mode. The general read and write operation of the bit-cell is illustrated in the schematic in Fig. 2.1. During operation, the MTJ exhibits a resistance across its junction, depending on the magnetic state of the MTJ, called as tunneling magneto-resistance effect. The difference in resistance at 2 different magnetic states is exploited for the read operation. During read mode, a small read-voltage is applied across the bit-line (BL) and source-line (SL), and the current flowing through the bit cell is measured to detect the magnetic state of the MTJ. A similar procedure is used for the write mechanism, except that the current flow here is bidirectional, i.e. the current can flow in both directions through the MTJ during programming depending on whether a 0 or 1 is written. For successful write operation, the currents injected should be greater than the minimum MTJ critical switching current (I_C), and since the writing process is bidirectional, the MTJ encounters 2 critical switching currents for each direction of the current $I_C(AP \rightarrow P)$ and $I_C(P \rightarrow AP)$.

A PMA-MTJ stack must be capable of providing the read and write margins, endurance, power consumption, retention, and switching performance compatible with the target mem-



Operation	V_{BL}	V_{WL}	V_{SL}
Read	Vread	Vdd	gnd
Write '0'	Vdd	Vdd	gnd
Write '1'	gnd	Vdd	Vdd

Figure 2.1: Operating modes of 1T-1MTJ cell.

ory application. For this the MTJ device parameters need to meet the design criteria in terms of MTJ fundamental device parameters such as TMR, device resistance, STT efficiency, offset field, dielectric breakdown, defect rate [5]. A way to make sure the device meets these specifications is to characterize the parameters of the MTJ device in an isolated manner before deposition over the CMOS sub-structure. Some of the fundamental parameters considered are,

1. Tunneling magnetic resistance: The ratio of the R_{AP} to R_P is given by the tunneling magnetoresistance (TMR) ratio as,

$$TMR = (R_{AP} - R_P)/R_P \tag{2.1}$$

For high read margin, the ratio of R_{AP} to R_P should be high. Current MTJ devices demand high TMR values with low RA product. This requires good crystallization of the MgO layer and the magnetic layer with minimal oxidation. The magnetic layers require a body-centrencubic (bcc) lattice (001) and MgO requires face-centre cubic (fcc) lattice structure (001) with 45° rotation of the in-plane lattice to the magnetic material interface [38]. Surface defects can form in these regions during the deposition and hence minimizing it using modern fabrication steps is crucial.

2. Switching current: The switching of the MTJ device is stochastic and is based on spin-torque-transfer (STT) described by the Landeu Lifshitz Gilbert (LLG) equation [18]. Depending on the direction and the magnitude of the injected current, the MTJ can switch to an AP state or P state as shown in Fig. 1.6(c), where the critical current of $I_{P->AP}$ or $I_{AP->P}$ is needed for each switching operation. The time taken to switch states is given

by the switching time (t_{SW}). Some other parameters that impact the switching current directly are switching current density, STT-efficiency, damping factor.

Additionally, the critical currents needed for MTJ switching are not equal i.e. $I_C(AP \rightarrow P) \neq I_C(P \rightarrow AP)$. This assymetry can pose problems such as unequal write time for programming '1' and '0' on to the MTJ, which has to be considered during the design of the write circuitry. Some circuit level and device level strategies have been proposed to reduce this effect. Lin et.al has proposed a reverse bit cell design [29] that leverages the source degeneration property of transistors to equalize the switching threshold currents. On the device side, MTJ structures have been designed that reduce this assymmetry by introducing nano-current channel structure into the MTJ [46] or by providing external magnetic field [11].

3. Thermal stability: Stability is an important factor that determines the endurance of the data stored in the MTJ. MTJ stability is defined as the ability of the MTJ to retain the magnetization in its free layer without any errors during thermal disturbances. It is the amount of the energy barrier (E_B) the free layer has to overcome in order to switch between P and AP states. It is given by,

$$\Delta = E_B/(k_B T) \tag{2.2}$$

Where k_B is the Boltzmann constant and T is the temperature. The data retention property of the MTJ device is decided by the thermal stability factor. Scaling of switching current while maintaining the thermal stability is one of the crucial merits needed for the MTJ stack. The FIT rate is common metric used to quantify the endurance of the cell, where 1 FIT corresponds to 1 failure per 10^9 s of operation. During operation, the MTJ's are usually operated in the thermal switching regime Fig. 2.9, where the switching is a stochastic process. The switching probability ($P_{SW}(t)$) can be defined in terms of FIT rate as [1]

$$P_{SW}(t) = \frac{FIT\ rate}{Bits} = 1 - e^{\frac{t}{\tau}} \tag{2.3}$$

$$\tau = \tau_0\ e^{\frac{K_U V}{k_B T}} \tag{2.4}$$

Here K_U is the magnetic anisotropy, V is the volume of the free layer of the MTJ, T is the temperature and τ_0 is the attempt frequency which is 1 ns. Here the term τ is called as the thermal stability parameter, which can be used to define the storage stability and read/write stability. The table 2.1 shows the thermal stability needed for MTJ's at different memory capacity and FIT rates

Capacity — 1000 FIT at 80°C — 0.1 FIT at 80°C — 0.1 FIT at 160°			
1MB	66.6	77.4	95
16MB	69.9	80.7	99
256MB	73.1	84	103
512MB	73.9	84.8	104
1GB	74.7	85.6	105

Table 2.1: MTJ thermal stability needed for different memory capacity and FIT rates, Adapted from [1].

Thermal stability decreases with scaling of the MTJ structure, and hence it is proportional to the volume of the MTJ structure. It is seen that the critical switching current (I_C) is decreased with device scaling, which will lead to drastic reduction of write current. However the MTJ thermal stability also deteriorates with this dimension shrinking. Hence the $\frac{I_C}{\Delta}$ as the figure of merit for MTJ writability rather than using I_C alone.

2.2 Write Operation

The key functionality of the MTJ write process is to store the '1' or '0' data based on the direction of flow of current. The write circuitry used should have a bidirectional driving capability which should be able to sink and source current with equal strength. A scheme for bidirectional write is as shown in figure 2.2 [6]. The circuit consist of both read and write circuitry which are connected to a common output line called SALT and SALB. During write operation, the sense amplifier is operated as a latch with SALT and SALB outputs determining the direction of current flow. The write operation takes place in 2 phases. First is the data loading phase, where the IO line loads the data into the latch changing the SALB, SALT output. The next phase is the write phase (WE = 1), where the NAND gate based write driver turn on one of the driver transistors. The current is then injected into the BL / SL lines using parallel connected PMOS (controlled by WEFB / WEMB) and NMOS transistors (WEFT / WEMT).

2.3 Read Operation

The read operation in an STT-RAM is done by sensing the resistance of the MTJ structure. Based on the magnetization of the MTJ free layer with respect to the pinned layer the MTJ

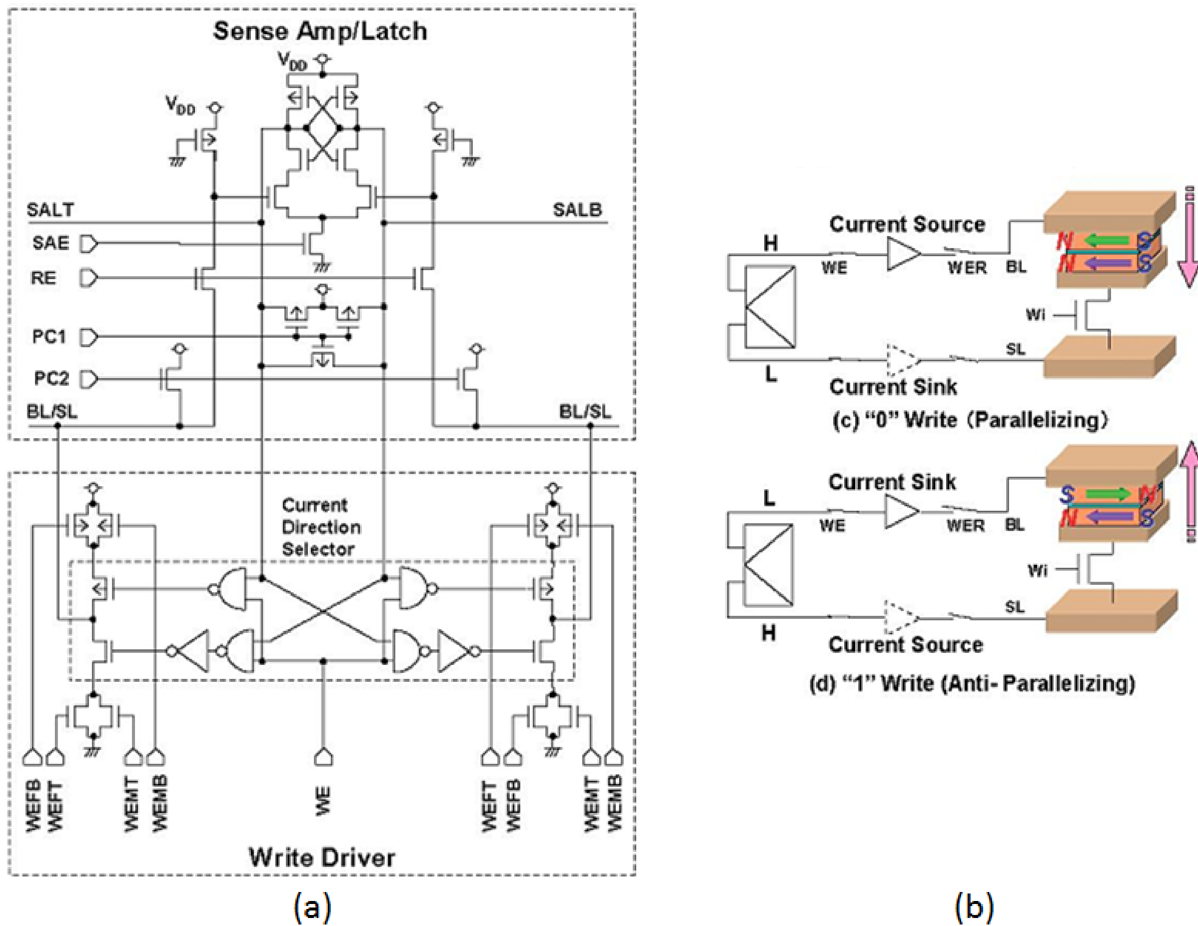


Figure 2.2: (a) Sense amplifer latch with bidirectional write driver, (b) illustration of bidirectional write operation. Adapted from [6].

can have a high or low resistance respectively. Read operation can be done using a current sensing or voltage sensing technique, where the sensed data is compared with a reference data to determine the state of the MTJ cell.

The read sensing scheme has not drastically changed compared to the modifications and innovations happening in the writing mechanism of the MTJ cell. One of the key criterion for effective reading is the reference signal generation method. The reference signal is chosen appropriately so that the raw read signal margin generated must be large enough to compensate for MTJ parameter mismatch as well as the sense amplifier offset. Hence the

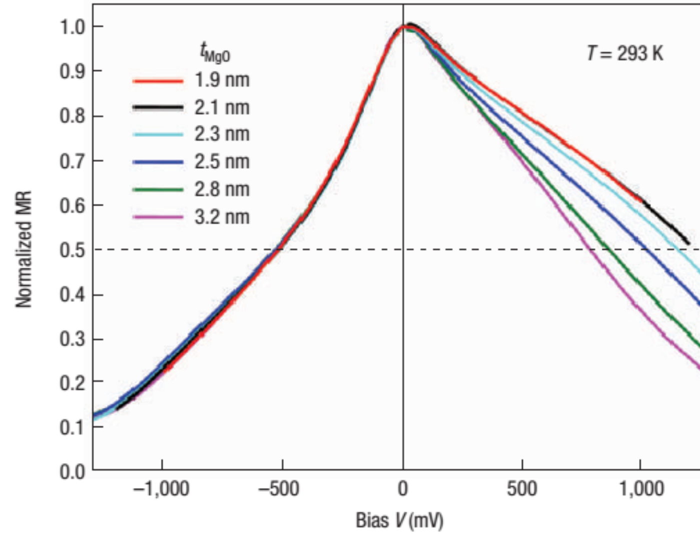


Figure 2.3: TMR variation vs. bias voltage applied across the MTJ for various MgO dielectric thickness, Adapted from [7].

reference signal generation scheme imposes design requirements on TMR, MTJ parameter variation and sense amplifier offset. The MTJ resistance and the tunneling magnetic resistance (TMR) can vary with the voltage bias applied across the MTJ structure. The MTJ resistance and TMR decreases with increase in bias voltage across MTJ as shown in Fig. 2.3 [7], which limits the use of higher read voltage to achieve a higher read sense current. Additionally, higher read voltage can also result in read disturb condition, where the MTJ is written accidentally during a read operation. However, this doesn't mean that choosing a low read voltage is good since the absolute difference between the read sense current and the reference current decreases drastically when the voltage bias approaches zero. Hence it is a general practise to set the read bias voltage around 200 to 400mV [8] [9]. This voltage is established using transistors that clamp the bit-line voltage to get the required read voltage.

2.3.1 Reference Signal Generation Techniques

Different read sensing methods are used to generate the reference signals for distinguishing the state of the MTJ cell being read. Some general techniques based on the usage of reference cells are a) twin cell, b) self reference schemes and c) single ended reference

cell. The twin cell method utilizes differential mode of sensing, where 2 MTJ's (true and complimentary) are used to store data of opposite states [47]. While reading, the current in the true MTJ is sensed and compared with the complimentary MTJ current to determine the data stored. The twin cell technique gives twice the raw read current margin compared to the conventional reference cell technique, but at the expense of lower circuit density. Alternate technique is the self reference scheme where temporal sensing of the current flowing through the bit cell is measured at various states. This method of sensing scheme operates in 3 read phases [48]. First phase, the current passing through the MTJ bit cell is sensed and the value is stored temporarily. On the second phase, the current flowing through the same MTJ will be sensed again with the MTJ written to a known state. On the third phase, the current will measured with the MTJ written to the opposite state. The original current is then compared with the average current, which is determined earlier by programming MTJ's to known states, to find out the initial state of the MTJ. Advantage of this technique includes no chip density sacrifice and insensitivity to MTJ process mismatches. The disadvantage is that the increased number of read access cycle will deteriorate the read speed and increase read power. However, leveraging on this concept chen et.al has proposed a non destructive self reference scheme for STT-RAM designs [49]. Here the MTJ need not have to be written with the known states to determine reference value, hence the read speed can be improved.

The reference cell method is the most popular sensing technique considering the circuit density, low power consumption and read speed. Here the sensed signal is compared with the reference signal generated from one or more reference cells. For a single ended reference cell approach, the reference signal generated has to be scaled and positioned midway between the AP and P state bit-cell read currents. This can be done by using 2 reference cell programmed to opposite states. The simplified 2 reference cell architecture for a 1T-1MTJ sensing is illustrated as shown in Fig. 2.4 [6]. Here the current flowing through the bit cell is compared with the reference current generated by averaging the currents flowing through the low and high programmed reference cell. A V_{CLAMP} voltage is applied at the reference cell node to provide the appropriate read voltage using transistors in source follower configuration. The bit and reference cell currents are passed through a load to create a voltage signal which is sensed sense amplifier. Compared to the twin cell approach, the reference cell method offers only half the raw read signal margin, but the circuit density is much compact. In the simulation waveform shown in fig 2.4(b), it can be seen that the reference voltage settles in between the read cell '1' and read cell '0' voltages only after a delay (15 ns) from the RE signal activation. This is due to the unequal currents at input nodes of the SA. Here, the SAE can be enabled only after the reference voltage has settled between the read cell values as shown in 2.6, which slows down the read speed.

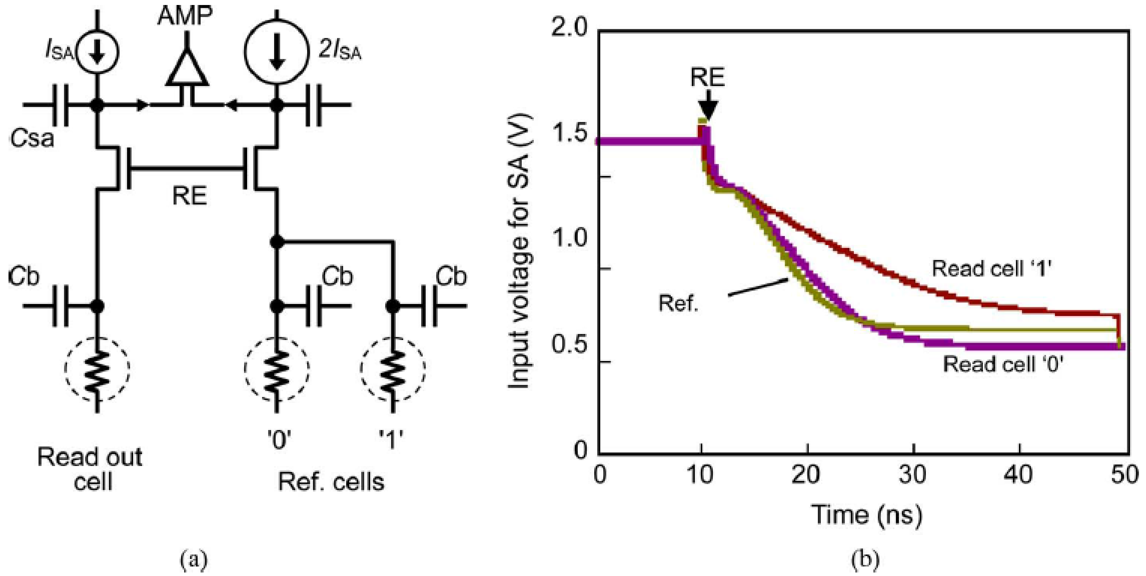


Figure 2.4: (a) Schematic of a conventional reference cell read scheme, (b) shows voltage created at the input of the sense amplifier , Adapted from [6].

2.3.2 Sensing Topologies

Current and voltage sense amplifiers can be used for MRAM sensing. For voltage sensing, a current source is used to inject current into the bit cell and the voltage is sensed. The bit-line voltage has to be developed and the equivalent read delay depend on the current into the bit cell, bit line capacitance and the MTJ resistance. In current sensing, a voltage source is used, where the equivalent circuit delay depend on the series bit line resistance and bit line capacitance rather than on the MTJ resistance. STT-MRAM applications demand high density of bit-cells that results in high bit-cell capacitance. Current sensing method offers high speed read operation under large bit-line capacitance conditions and is generally preferred for higher performance [8]. The 3 common current sense amplifier schemes are as follows.

Fig. 2.5(a) shows a differential scheme with 2 SA and 2 reference bit lines. I_{ref0} and I_{ref1} are averaged to create the reference current, which is passed through PMOS load P1 and P2 to create the reference voltage. Differential amplifiers are used to read data on both the bit lines. Fig. 2.5(b), shows the 2 reference cell scheme with a single SA. Fig. 2.5(c), shows one of the 2 SA that shares the reference line (note the short circuit at Iref), P0 to

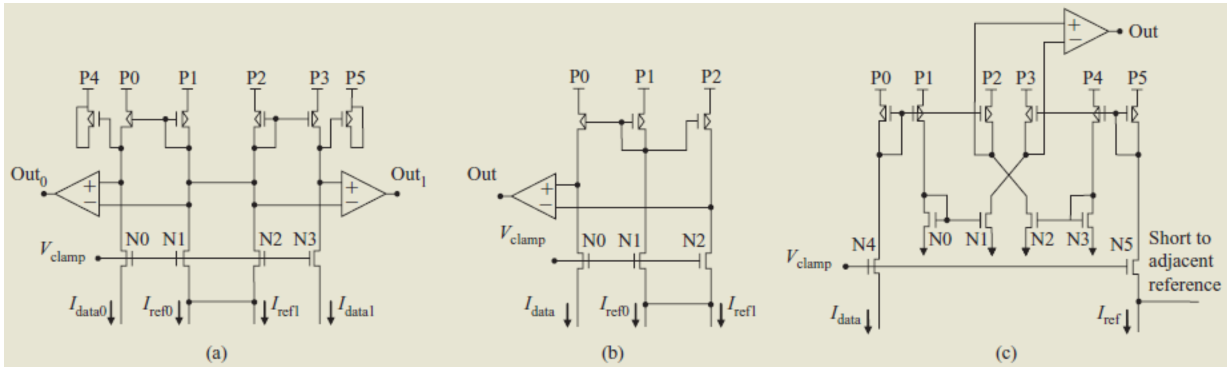


Figure 2.5: (a) Differential read scheme, (b) conventional read scheme with 2 reference cells, (c) Cross coupled current mirror amplifier based scheme, adapted from [8].

P5 and N0-N3 forms a cross coupled current mirror amplifier that creates the differential voltage for the SA. This scheme provides twice the amplification compared to other designs, however the additional matched pairs degrade the overall SA offset.

Takemura et.al has proposed a dual array equalized reference read scheme [6] that is suitable for differential sensing. The schematic and waveform are as shown in Fig. 2.6. Here a

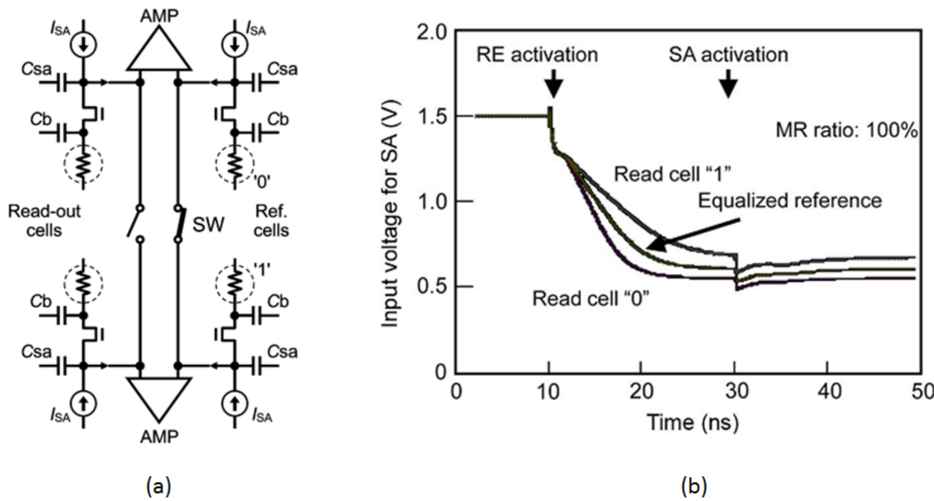


Figure 2.6: (a) Dual array equalized reference scheme, (b) Simulation waveform. Adapted from [6].

switch is used at the input nodes of the sense amplifier. The node is short circuited

when the nodes are used as the reference inputs for the SA. Thus, the read circuit of the memory cells are connected to SA and a common reference input is shared. This keeps the reference voltage always within the read '0' and read '1' values, allowing us to enable the sense amplifier earlier and thereby improving read speed.

An alternate approach is utilizing the current sense amplifier proposed by Tsuchida et.al [9]. The read scheme design used is different from the previous scheme since it uses only a "low" cell reference as shown in Fig. 2.7. Here the read current direction flowing through the reference cell only reinforces its existing state. Hence it eliminates the problem of accidental writing of reference cell during read operation. Additionally a clamped reference voltage is provided to bias the bit line using a biasing circuit. The schematic of the current sense amplifier and the bias generator circuit is as shown in 2.7. The sense amplifier circuit

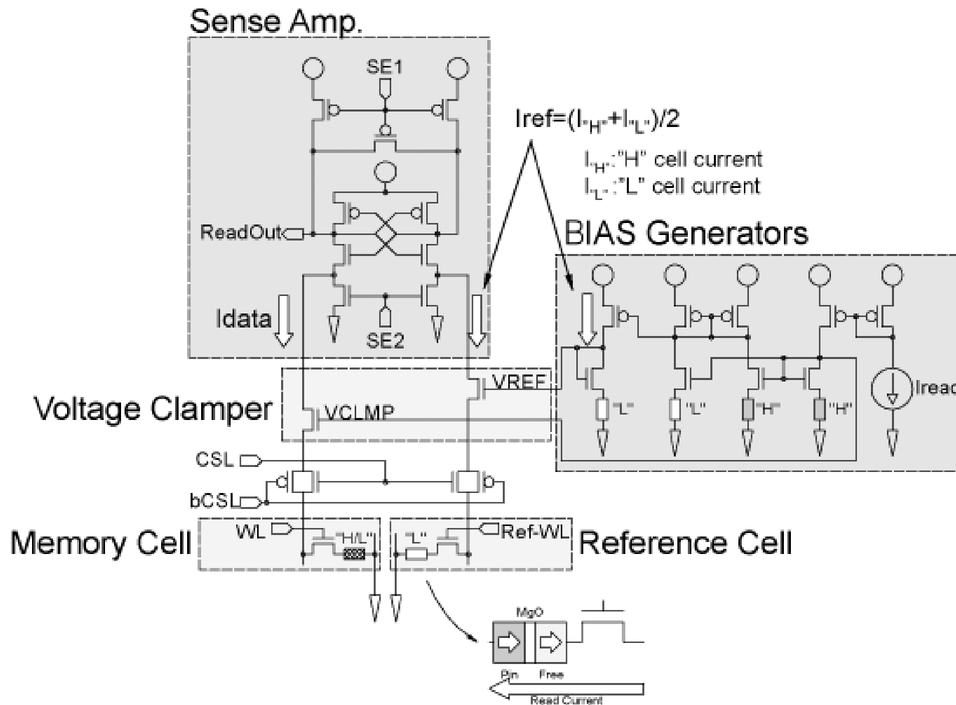


Figure 2.7: Current sense amplifier design using a single reference cell and a clamped reference, adapted from [9].

is simulated in cadence using 65 nm CMOS technology. Here the bias circuit was designed to provide a reference current of 15 μA. The timing waveform for the circuit operation at various input and output nodes of the circuit is as shown in Fig. 2.8.

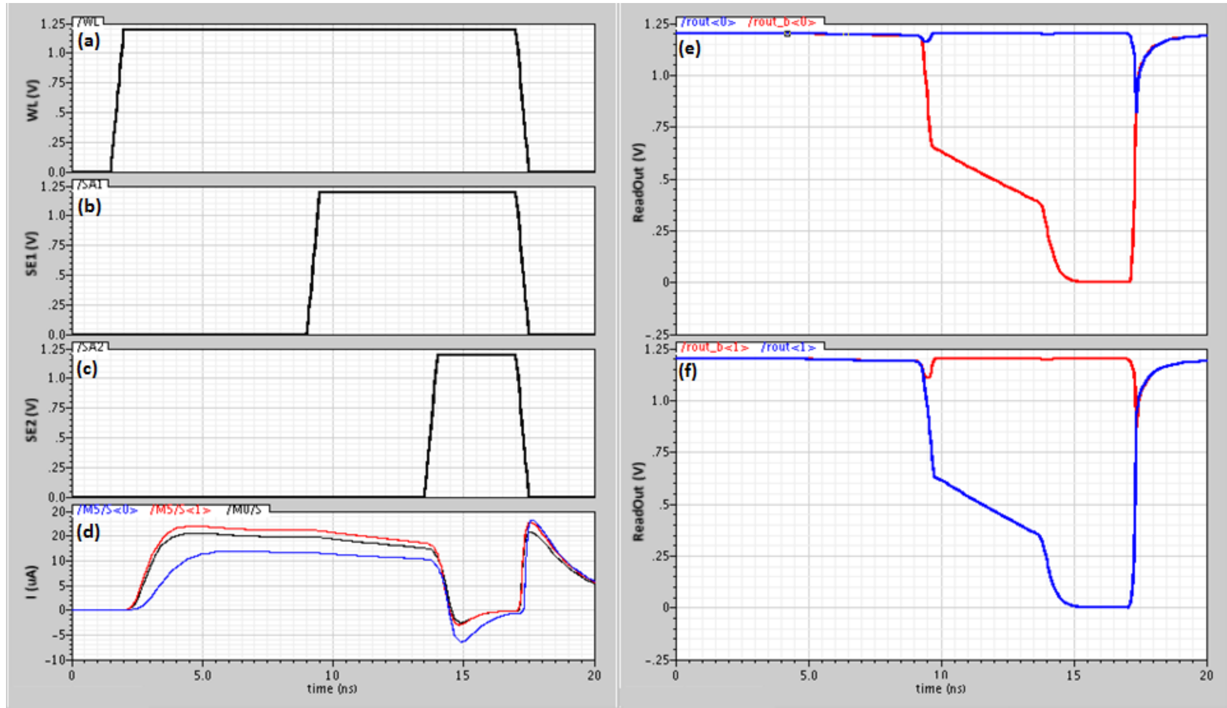


Figure 2.8: Simulation waveforms for the sense amplifier (a) V_{WL}, (b) $SE1$, (c) $SE2$ (d) shows reference current (black), read '0' current (red) and read '1' current (e) Sense amplifier output for reading '0', (f) Sense amplifier output for a read '1'.

The operation of the read scheme is described as follows. Initially, the word line voltage is applied, turning ON the bit cell. This charges the bit lines and establishing the I_{DATA} and I_{REF} currents in the bit line and the reference line respectively. At this point the precharge transistors are ON, keeping the ReadOut node in Fig. 2.7 to V_{DD}. Based on Fig. 2.8, when the sense enable 1 (SE1) is '1', the precharge transistors are turned off and the sensing operation begins. The difference in the I_{DATA} with respect to I_{REF} forces the latch to fall into one of the states. Finally, the sense enable 2 (SE2) is enabled to latch the data, which reduces the read power consumption and the sense amplifier output is obtained in the ReadOut node.

Read Disturb / Write Error Condition

Based on the modes of operation of the bit cell, we can observe that a high bidirectional current is used for writing the cell and a small unidirectional current is used to read the

data stored in the MTJ cell. Thus the same circuitry is used for both the read operation and the write operation. However, CMOS process and MTJ variability pose serious failures during read or write operation. The 2 main kinds of failure are a) write failure b) read disturb failure. A write failure event can be defined as a condition where the write current is too low that it has not exceeded the switching threshold current of the MTJ ($I_{write} \leq I_C$), thereby failing to write the cell. However, write current cannot be increased drastically to avoid write failure, since high write current can results in MTJ operating beyond its breakdown voltage. Hence the write operating region depends on the margin available between a write failure case and the MTJ breakdown case.

Similar kind of failures can also occur during read operation of MTJ bit cells. A read disturb condition occurs when the read current is too high that the MTJ cell gets accidentally written during a read operation ($I_{read} \geq I_C$). On the other side, read current cannot be decreased below the minimum current needed for the read sensing scheme and the sense amplifier to detect the signal. Hence the read margin depends on the margin available between the read disturb case and the minimum detectable current for the sense amplifier.

Dependence of Critical Switching Current to Write Pulse Duration

Switching threshold can be defined as the minimum current that has to be passed through the MTJ junction so as to program it to AP or P state. Its also seen that the switching threshold current is not a constant value, but it also depends on the duration of the write pulse being applied. The figure 2.9 shows the variation of critical switching threshold current as a function of the pulse duration.

Based on the STT writing regimes, the MTJ can be written in 3 different modes. They are precession, dynamic and thermal [50]. The precessional switching is the fastest mode of switching, where the switching depends on the initial thermal distribution and the switching current. However the critical current needed is larger. For precession switching, the relationship between the switching probability and pulse duration can be understood from the Sun's model [51].The thermal switching regime is a thermal activated process that occurs for long duration write pulses, Here the switching mechanism is mainly due to the thermal distribution. Neel's Brown model [52] that can be used to model the relationship between the switching probability and the write pulse duration for thermal switching regimes. The region in between is called as the dynamic reversal region, where the switching mechanism is the combination of precession and thermal switching regime. This is the region where most of the MTJ are operated. Here the switching process depends on both the intial thermal state and thermal distribution. However due to the complicated

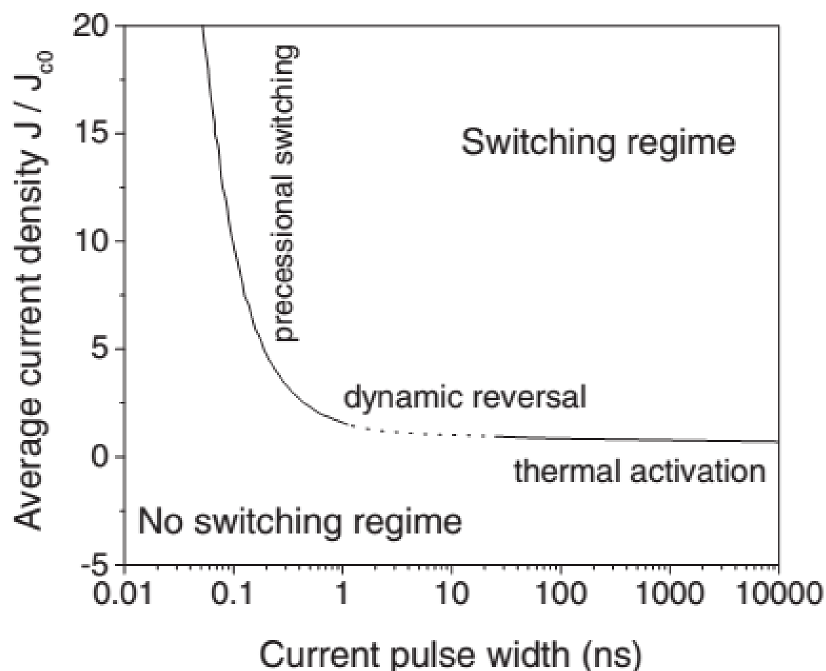


Figure 2.9: MTJ switching regimes vs. write pulse duration, adapted from [?]

nature of this operating region, there is no single formula to relate the switching probability to pulse duration.

2.4 MTJ Physics & Modelling

An understanding of the MTJ modeling is essential for the design of STT-RAM circuitry. The MTJ model should be able to capture the static resistance behavior based on the magnetic orientation of the free layer with respect to the pinning layer. The model should also be able to assimilate the transient switching behavior of free layer magnetization with passing of the current. There are different approaches for modeling nanoscale devices such as magnetic tunnel junctions, but the level of complexity of the model is usually a compromise between accuracy and computational cost. MTJ modelling can be broadly classified into a) Compact modeling b) Nano scale modeling based on Non-Equilibrium Greens Function (NEGF) c) Micro magnetic modeling.

2.4.1 Compact Modelling

Compact modelling approach is an efficient way of capturing the functionality of MTJ by capturing the most essential physics based equations and integrating them with CMOS circuit models for simulation. A basic MTJ model consists of two parts. a) Resistance characteristics b) Switching behavior. The first part of MTJ models the magneto resistance characteristics of the MTJ junction at both AP and P states. The relationship between these two resistances is called as the tunneling magnetic resistance which is found experimentally by

$$TMR = \frac{R_{AP} - R_P}{R_P} \tag{2.5}$$

The second parts is the dynamic part of the MTJ, which describes the switching behavior of the free layer as a function of time. Several compact models have been published for MTJs during the last decades. These models can be categorized broadly into 2 types [53] a) Static models and b) Dynamic models. In static models, the magnetisation switching of the free layer are approximated based on the current flow direction and its magnitude with respect to the critical switching current. The MTJ resistance values are then set based on the magnetization condition of the free layer. Advantage of static models is that they are suitable for DC I-V simulations with very low computational cost. The disadvantage is that static models do not capture the switching behavior accurately, hence they may not be suitable for high speed transient circuit simulations.

The dynamics models consider the magnetic switching behavior as a function of time to accurately model the transient behavior of the free layer magnetization. The instantaneous magnetisation (M) is estimated using the Landau Liftshitz Gilbert Equation (LLG), as shown below [54].

$$\left(\frac{1}{\gamma}\right)\frac{d\boldsymbol{M}}{dt} = (\boldsymbol{M} \times \boldsymbol{H}_{\text{eff}}) - \alpha\frac{\boldsymbol{M}}{M} \times (\boldsymbol{M} \times \boldsymbol{H}_{\text{eff}}) - \alpha_J\frac{\boldsymbol{M}}{M} \times (\boldsymbol{M} \times \boldsymbol{s}) \tag{2.6}$$

where M is the magnetization at any given time, γ is the gyromagnetic ratio, α is the damping constant, α_J is the flux of polarised electrons and H_{eff} is the effective magnetic field, which accounts for the anisotropy field and the external magnetic field.

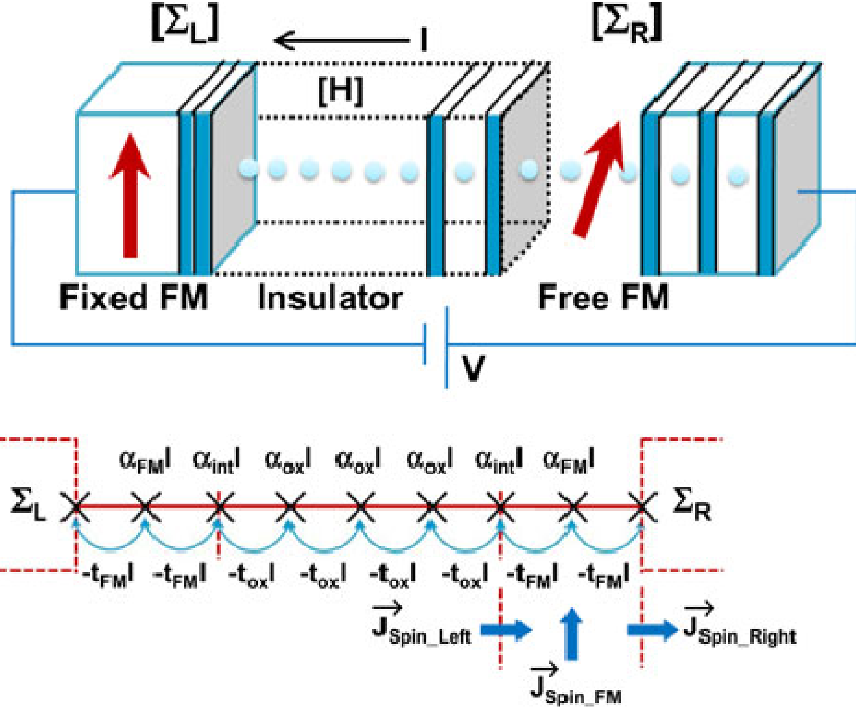


Figure 2.10: (a) Shows the MTJ split into individual 2D layer unit cells (b) shows the NEGF representation of MTJ in form of 1D model. Adapted from [10].

2.4.2 Nanoscale Device Modelling based on NEGF

An Non-equilubrium Greens Function (NEGF) model for MTJ based on single band effective mass Hamiltonian is described by Datta.et.al [10]. The MTJ model is described by parameters such as equilibrium fermi level, spin splitting, barrier height of the insulator and effective mass of electrons in the insulator and the ferromagnetic layers. The NEGF process starts with the finding of the single band effective mass Hamiltonian matrix of the MTJ layer and the self energy matrices at the contacts of the MTJ. Then the parameters such as the Greens function and electron correlation function are found. These parameters are then used to find the charge and spin densities in between each lattice points.

In Fig. 2.10(a), The MTJ is split into individual 2d unit cells and it is repeated along the z axis to create the MTJ structure.Here [H] represent the Hamiltonian matrix, the Σ_L and Σ_R represent the self energies at MTJ contacts. 1D representation of the NEGF model is as shown in Fig. 2.10(b). Here t_{ox} represents the coupling energy from one unit

lattice to another and α represents on site energies at each individual unit lattice. The model also assumes periodic boundary condition along the direction of flow of current, which allows for decoupling of the transmission mode waves along the MTJ's transverse path, enabling it to be decoupled as individual 1D wires. The charge and spin densities are then found individually for 1D wires. Finally all the transmission wave modes are summed up together to find the final current for the realistic cross section area. The main aim of this NEGF model is to determine the amount of torque transferred into the free layer of the MTJ, which is responsible for the switching mechanism. For this, we find the spin current ($\overrightarrow{J}_{SpinFM}$) responsible for torque transfer, which is found by subtracting the out flux current from the influx current at the Free layer region, given by

$$\overrightarrow{J}_{SpinFM} = \overrightarrow{J}_{SpinLeft} - \overrightarrow{J}_{SpinFM} \tag{2.7}$$

The spin torque transferred is equal to the divergence of the spin current between the lattice region being considered, which in this case is the free layer given by

$$\overrightarrow{\tau} = \int \delta S \ J_{Spin\ FM} \tag{2.8}$$

2.4.3 Micro-magnetic Modelling

The compact models considered in previous section are macro models, where the free layer magnetization of the MTJ is considered as one single domain. Here the MTJ behaviour is captured using basic MTJ physics models that are obtained from experimental data. These models are then integrated with circuit models for simulations in a particular CMOS technology. However, in order to see the nanoscale effects of material structure, device scaling and MTJ structural defects, it is essential to perform a micro-magnetic simulation. The micro-magnetic simulators takes in the magnetic parameters and the material structure as input parameters, and solves the LLG equation to find the instantaneous direction of magnetization of individual elemental magnetic domains. For instance Zhao et.al has studied the switching current characteristics for an elliptical MTJ Free layer with dimensions of 160 nm × 90 nm × 2.5 nm [50]. The results were consistent with the macroscopic models and the experimental data obtained.

2.4.4 Model Used for Bit-cell Analysis

A PMA-MTJ model proposed by Zhang et.al is used for circuit simulations in this research [55]. The model is based on a 40 nm circular CoFeB/MgO which exhibit good tunnel magneto-resistance ratio (120%) and switching performance. The model integrates the static, dynamic and stochastic behaviors in the model and its simulation results are in close agreement with the experimental data provided [56] [57]. The MTJ model is written in verilog-A, and is compatible with 65 nm Cadence circuit simulation tools. The model used is a static, where the switching behavior is modelled based on approximating the LLG behavior. This reduces the transient switching behavior during high frequency simulations, however for the scope of our DC simulations, the model accuracy is adequate. Additionally, since the model is not designed for transient behavior, the time duration of the voltage signal that is applied across the MTJ has to be specified (assuming rectangular input waveform) explicitly in the Verilog-A model. Different physics models / equations that are used to describe the functionality are,

Tunneling Resistance Model

Here the Brinkman's tunneling model [58] is used to calculate the parallel resistance of the MTJ. The simplified form of the equation is as shown below

$$R_P = \frac{t_{ox}}{F \ \times \phi^{-1/2} \ \times Area} \times e^{1.025 \times t_{ox} \times \phi^{-1/2}} \tag{2.9}$$

where t_{ox} is the oxide thickness, F is a fitting parameter calculated from the resistance area product, Area of the MTJ, ϕ is the barrier potential.

Bias Voltage Dependant Model

It was seen that the TMR decreases with increase in the bias voltage [7]. This model is important to be considered in the analysis since it directly dictates the difference between R_{AP} and R_P resistances. The model is described as a parabolic function that is fitted to the experimental data provided in [7] shown below.

$$TMR = \frac{TMR(0)}{1 + (\frac{V_{bias}}{V_H})^2} \tag{2.10}$$

where TMR(0) is the peak value of TMR, V_H is the bias voltage where the TMR is half the peak value, V_{bias} is the bias voltage appearing across the MTJ.

Static STT Switching Model

The model calculates the critical switching current from the magnetic and material parameters provided for the MTJ. The equation is as shown below

$$I_{c0} = \alpha \frac{\gamma e}{\mu_B g} (\mu_0 M_s) H_k V \tag{2.11}$$

where α is the damping parameter, γ is the gyromagnetic ratio, M_S is the saturation magnetization, H_k is the anisotropy field, V is the volume of the MTJ, μ_B is the Bohr magnetron constant.

Stochastic Switching Model

The static model is used to calculate the switching duration (τ) based on the critical threshold current (I_C) and the current flowing through the MTJ (I_{write}). The switching model provided in [59], given by

$$\frac{1}{\tau} = [\frac{2}{C + ln(\frac{\pi^2 \xi}{4})}] \frac{\mu_B P_{ref}}{em(1 + P_{ref} P_{free})} (I_{write} - I_{c0}) \tag{2.12}$$

where τ is the switching pulse duration, P_{ref} and P_{free} are the spin polarizations of the reference and the free layers, ξ is the activation energy in $k_B T$, C is the Euler constant = 0.577, I_{c0} is the critical switching current and I_{write} is the applied write current. The parameters used for the simulation is as shown in Table 2.2

Device Parameters	
Resistance-Area Product	$5\Omega - \mu m^2$
TMR Ratio	120%
Electron Polarisation	0.52
Damping Parameter (α)	0.025
Gyromagnetic Ratio (γ)	1.76×10^7
Out of Plane Anisotropy (H_k)	0.1734×10^4 Oe
Saturation Field in Free Layer (M_S)	15800 Oe
Barrier Hieght (ϕ_e)	0.4 eV
Pulse Width	5 ns

Table 2.2: MTJ device parameters.

The details of the MTJ dimensions is as shown in Table 2.3

MTJ Dimensions	
Free Layer Height	1.3 nm
Diameter	40 nm
Oxide Thickness (t_{ox})	0.85 nm

Table 2.3: MTJ structure dimension.

2.5 Preliminary Comparative Analysis of 1T-1MTJ STT-RAM Cells

In this section, the exploration of design boundaries for read and write operation is done by evaluating different configurations of the 1T-1MTJ cell. Currently, a comprehensive analysis of various cell configurations is absent, which is essential for designing 1T-1MTJ bit cells with optimum read performance and improving read sense margin (RSM) for sensing. Analysis of the write performance is also done based on the Energy delay Product measured during the write operation of the 1T-1MTJ bit cell.

2.5.1 Simulation Methodology

During the read operation, the current flowing through the bit cell is determined by its total resistance (R_{TOTAL}), which consists of MTJ resistance (R_{MTJ}) and NMOS source-drain resistance (R_{TRAN}). A 40-nm CoFeB/MgO circular (PMA) MTJ model with a resistance-area (RA) product of 5-μm^2 and TMR of 120% are used [55]. The model was verified with the actual experimental data provided in papers [57] [56]. The basic characteristic of MTJ is determined by

$$TMR = \frac{R_{MTJ}^{AP} - R_{MTJ}^{P}}{R_{MTJ}^{P}} \tag{2.13}$$

where R_{MTJ}^{AP} and R_{MTJ}^{P} are resistance of MTJ at anti-parallel (AP) and parallel (P) state, respectively. For the analysis, the MTJ model uses a symmetrical critical switching currents, $I_C(AP \rightarrow P) = I_C(P \rightarrow AP)$. Additionally, the NMOS resistance is a non-linear function of the voltage applied across drain-source of transistor. Hence the access transistor and the MTJ is coupled together during simulations to accurately estimate total resistance characteristics of the 1T-1MTJ cell.

Figure 2.11(a) shows a schematic of a 1T-1MTJ bit cell, where BL, WL, and SL denote bit-line, word-line, and source-line, respectively, and V_{SL} is grounded. PL and FL denote

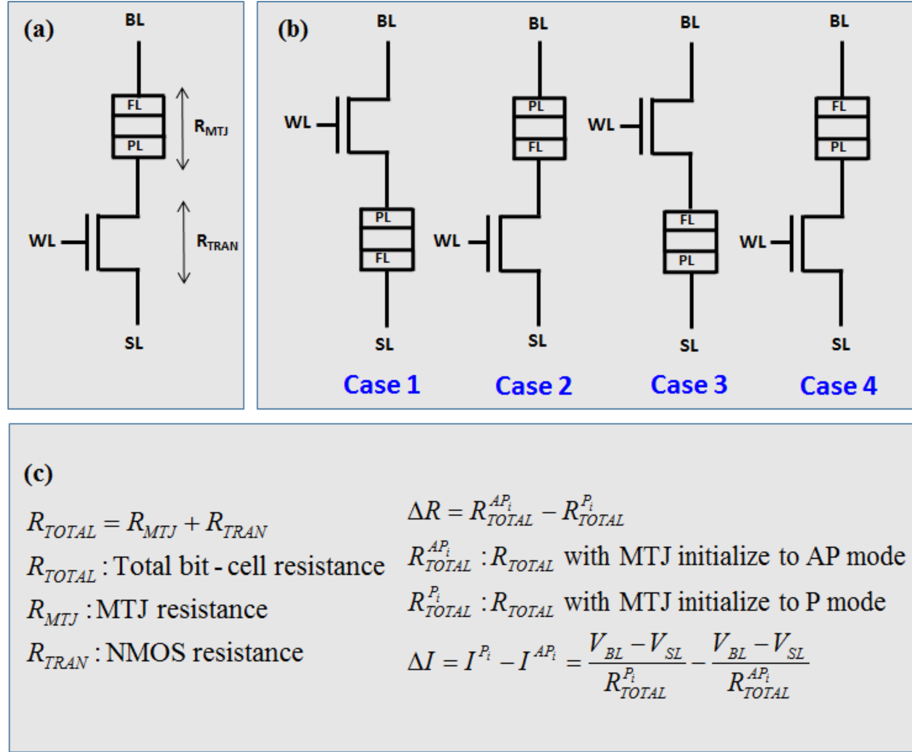


Figure 2.11: (a) 1T-1MTJ STT-RAM cell. (b) Four different configurations of bit cell. (c) Details of the extracted simulation parameters. The superscript of P_i or AP_i indicates the initial state of MTJ being P or AP mode.

pinning layer and free layer of the MTJ, respectively. The 1T-1MTJ bit-cell design exhibits different resistance characteristics based on the position and the orientation of MTJ with respect to the NMOS access transistor. Four different configurations of the transistor and the MTJ are considered depending on the relative orientation of the MTJ to the access transistor, as shown in Fig 2.11(b). Case 4 is the conventional architecture that is widely used in most STT-RAM applications. The case 2 design, proposed by [29], is also used when there is asymmetry in the MTJ critical switching currents.

The simulation methodology involves exploring the RSM of different cell configurations at various bit-line voltages (V_{BL}), word-line voltages (V_{WL}) and transistor widths (W_{NORM}). The transistor widths are normalized to the minimum size of W = 250 nm. First, the MTJ of each bit cell configuration case is initialized to either AP or P state (shown explicitly on the superscript with AP_i or P_i in the equations in Fig 2.11(c)), and the V_{BL} is varied

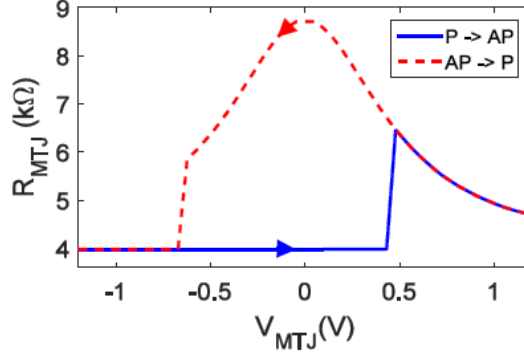


Figure 2.12: MTJ hysteresis of 40-nm PMA-MTJ model. V_{MTJ} is defined on the PL with respect to the FL.

for different V_{WL} and W_{NORM} to monitor the variation in total cell resistance (ΔR), and switching threshold. The current margin (ΔI) is then derived from these resistance values based on the equations shown in Fig 2.11(c), where ΔI is defined as the difference in current though the 1T-1MTJ cell in the P_i and AP_i mode.

The maximum current margin (ΔI_{MAX}) can be used to compare RSM among different cases. However, it alone does not provide the complete picture. For bit cells, ΔI_{MAX} occurs when V_{BL} is at switching threshold voltage (V_{ST}). Choosing this V_{BL} can have a large chance of read disturb failure during read operation. It is required to keep the read V_{BL} voltage lower than the V_{ST} without sacrificing ΔI significantly. Hence, a new metric called the bit-line voltage margin (V_{BLM}) is introduced to compare the voltage margin available for each case where ΔI is not significantly degraded. The V_{BLM} is defined as the V_{BL} range over which the drop in ΔI is less than 20% from the ΔI_{MAX}.

2.6 Results

Figure 2.12 shows the MTJ resistance characteristics of the MTJ model at AP and P states as a function of V_{MTJ}. Here the AP resistance is determined by the TMR and the resistance at the P state ($R^{P}_{MTJ} = 4$ kΩ). Figure 2.13(a) shows R_{MTJ} and R_{TRAN} as a function of V_{BL} ($V_{WL} = 1.2$ V; $W_{NORM} = 1$) for cases 1 and 4 with the MTJ initialized to the AP state. The MTJ resistance is maximum when V_{BL} appraches 0 V (8.8 kΩ) and decreases with increase in V_{BL} continuously for case 1 without switching. However, for case 4, the MTJ switches from AP to P state at $V_{BL} = 0.8$ V. Considering the access transistor

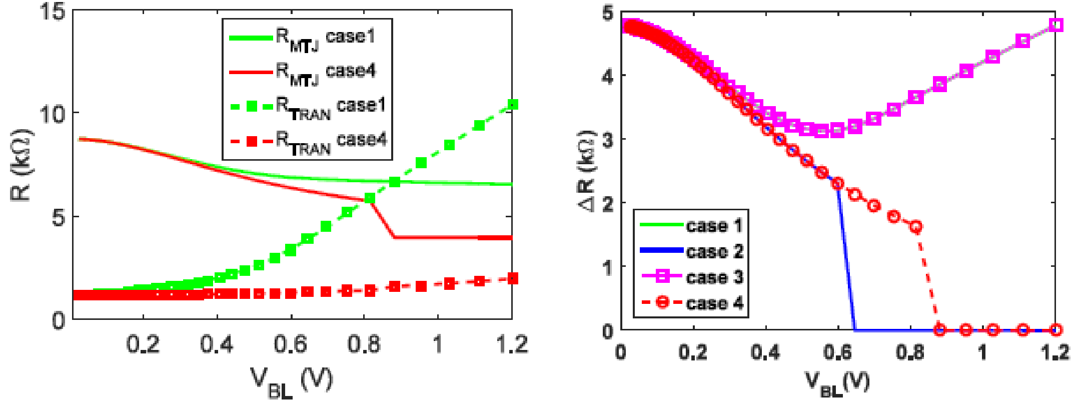


Figure 2.13: (a) R_{MTJ} and R_{TRAN} vs. V_{BL} for 1T-1MTJ cell with MTJ initialized in AP mode for case 1 and 4 (V_{WL} = 1.2 V; W_{NORM} = 1). (b) R vs. V_{BL} (V_{WL} = 1.2 V; W_{NORM} = 1) for all bit cell cases (the curves for cases 1 and 3 are overlapped).

operation, the NMOS operates in the linear mode at low V_{BL} resulting in low R_{TRAN}, and it operates in saturation mode at high V_{BL}, resulting in the increase of R_{TRAN}. It can be seen that for case 4, R_{MTJ} is significantly larger than R_{TRAN} for the entire V_{BL}, whereas case 1 shows significantly larger R_{TRAN} than R_{MTJ} at higher V_{BL}. This increase in R_{TRAN} is due to the threshold voltage increase caused by negative body bias ($V_{BS} < 0$) which is inherent in the source-follower read scheme (cases 1, 3) differentiating it from others (cases 2, 4) in Fig. 2.13(a).

Figure 2.13(b) shows ΔR for different cases at a given circuit condition (V_{WL} =1.2 V and W_{NORM} = 1). It is seen that for all the cases, ΔR is similar upto V_{BL}=0.3 V. After that the ΔR for case 1 and 3 (source degenerated bit cell) increases, where as cases 2 and 4 switches to 0 because of the read disturb failure. Here ΔR alone cannot be considered as a good metric for comparison, as the data does not provide information about the absolute current flowing through the bit cell, which is practically used for the sensing. Hence the absolute ΔI is derived from the measured resistance data for comparison.

2.6.1 Read Performance

In general, read sense amplifier designs depend on the TMR and RA product of the MTJ [60]. It is crucial to improve the TMR since it increases the difference between R_{MTJ}^{AP} and R_{MTJ}^{P}. The MTJ with large TMR and hence large difference in R_{MTJ} improves the RSM of the bit cell if voltage sensing is used. However, currently used MTJs have relatively low

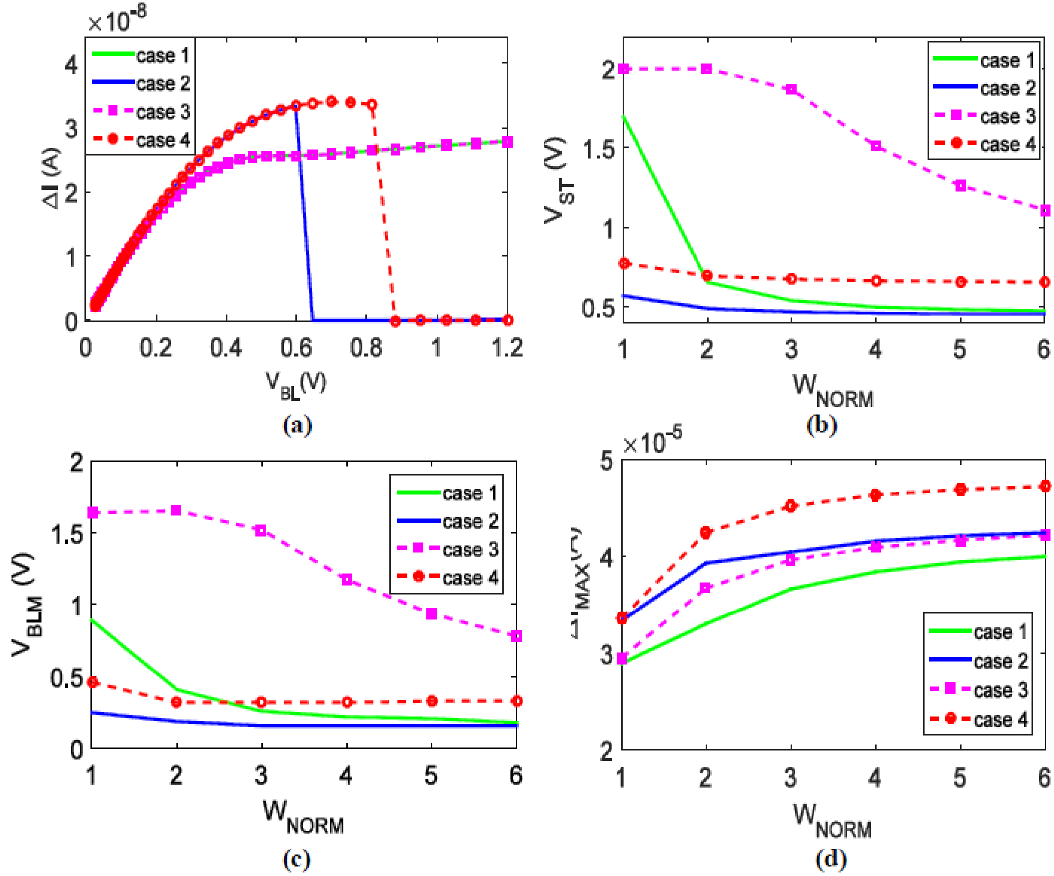


Figure 2.14: (a) ΔI vs. V_{BL} ($W_{NORM} = 1$), (b) V_{ST}, (c) V_{BLM}, and (d) ΔI_{MAX} variation as a function of W_{NORM}. V_{WL} is 1.2 V in (a)(b)(c)(d).

TMRs and also low RA product, for which measuring of current at bit lines becomes more efficient. Additionally, the current sensing is generally considered for high performance sensing since the equivalent charging time delay constant for the bit-line depends on the bit line capacitance and the resistance of the voltage source, which can be smaller than the R_{MTJ} [8]. The conventional CSA compares the current at the bit cell node against a reference node $((I^{P}_{TOTAL} + I^{AP}_{TOTAL})/2)$ to detect bit cell data [8]. Thus, a large difference in current between AP state and P state will improve the RSM of the bit cell.

Fig. 2.14(a) shows ΔI as a function of V_{BL} for different bit cell configurations. Here cases 2 and 4 give the highest ΔI with its peak value near the switching threshold. Operating the bit cells near the ΔI_{MAX} can result in accidental writing of the MTJ due to the variation

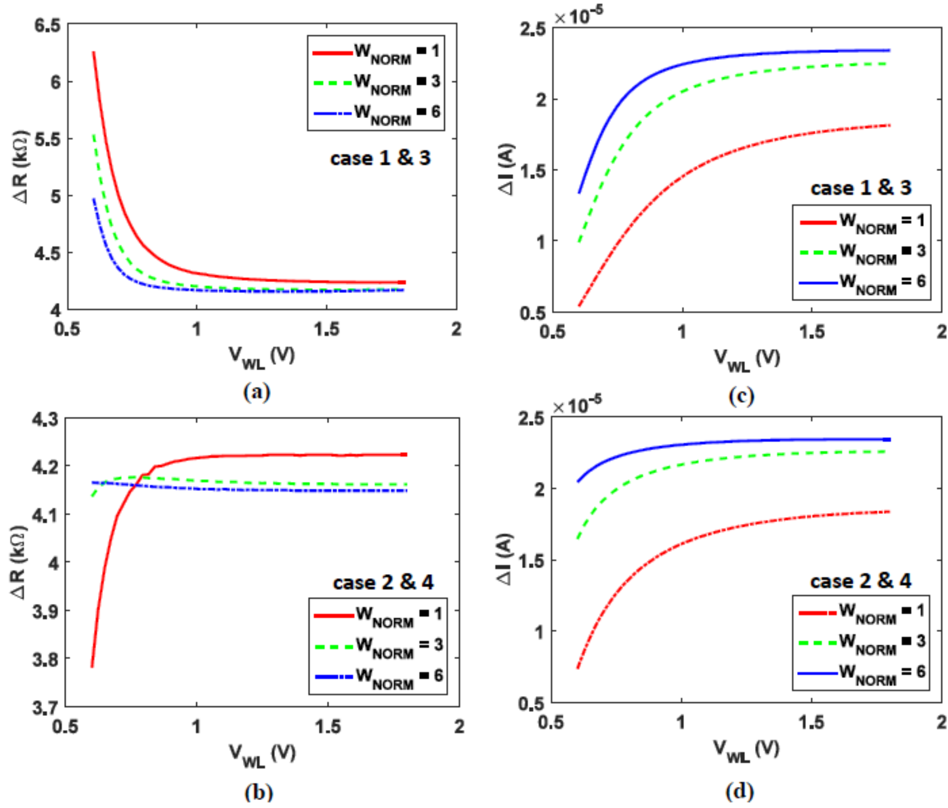


Figure 2.15: ΔR vs. V_{WL} for (a) cases 1 and 3 (source follower) and (b) cases 2 and 4. $V_{BL} = 0.2$ V is used. ΔI vs. V_{WL} for (c) cases 1 and 3 (source follower) and (d) cases 2 and 4. $V_{BL} = 0.2$ V is used.

of V_{BL}. Figure 2.14(b), 2.14(c) and 2.14(d) show the variation of V_{ST}, V_{BLM} and ΔI_{MAX} for different W_{NORM}. Simulation results show that case 3 provides the highest V_{ST} and V_{BLM}, while case 4 provides the highest ΔI_{MAX} for thee same area.

Next the impact of V_{WL} and W_{NORM} on the RSM is studied. Here the variation of ΔR with V_{WL} and W_{NORM} is investigated for cases 1 and 3, and cases 2 and 4 as shown in Figs. 2.15(a) and 2.15(b), respectively. Here V_{BL} is set to 0.2 V (well below write switching threshold of MTJ) to avoid any read disturb condition. The source follower configuration (case 1 and 3) shows an increase in ΔR for an increase in V_{WL}, which is different from case 2 and 4 that shows a decrease in ΔR value. This result may signify that the source follower configuration has better RSM at low V_{WL} voltages, but the absolute sense current decreases for the cell when V_{WL} decreases. Hence the absolute current margin (ΔI) data

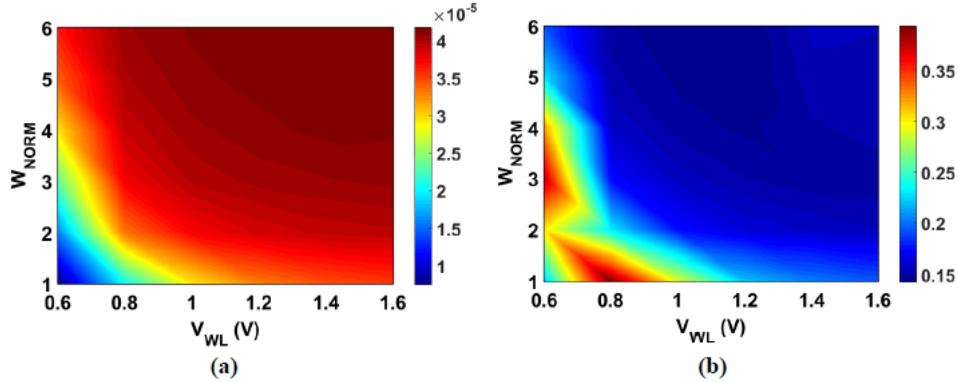


Figure 2.16: 2D surface plots of (a) ΔI_{MAX}, (b) V_{BLM}, as a function of V_{WL} and W_{NORM} for case 2.

provides a more meaningful comparison among cases as shown in Fig. 2.15(c) and 2.15(d). Considering the current margin for each bit cell case, As the V_{WL} is decreased, ΔI is maintained constant till 1.2 V. After that the ΔI decreases for all cases. However, the magnitude of decrease for cases 2 and 4 (to 87% for $V_{WL} = 0.6$ V) is much smaller than that for cases 1 and 3 (to 53% for $V_{WL} = 0.6$ V), particularly when the transistor width is large ($W_{NORM} = 6$), indicating that cases 2 and 4 can maintain its read current margin even at low V_{WL} voltages, which is an advantage for the conventional read scheme.

Next, the effects of V_{WL} and W_{NORM} on ΔI_{MAX} and V_{BLM} for case 2 are shown in Figs. 2.16(a) and 2.16(b), respectively. Here ΔI_{MAX} increases and V_{BL}M decreases with the increase of V_{WL} and W_{NORM}. In order to find an optimum point, a new figure of merit (FOM), ΔI_{MAX}-V_{BLM} product is introduced for comparison across cases. Figures 2.17(a) - 2.17(d) show the ΔI_{MAX}-V_{BLM} product for case 1, 2, 3 and 4, respectively. From the figure, case 3 gives the highest FOM (3.61×10^{-5}) at $W_{NORM} = 5$ and $V_{WL} = 1.2$ V, which is up to 380% larger compared to other configurations.

2.6.2 Write Performance

Finally, the write performance is also investigated in order to have a complete picture of the bit cell performance. For write performance, the switching delay and average write power are considered in the simulation for a clock rate of 100 MHz. As the write current increases, the switching delay decreases and the write power increases. These metrics were found as a function of V_{BL} to calculate the energy-delay product (EDP) for various V_{WL}

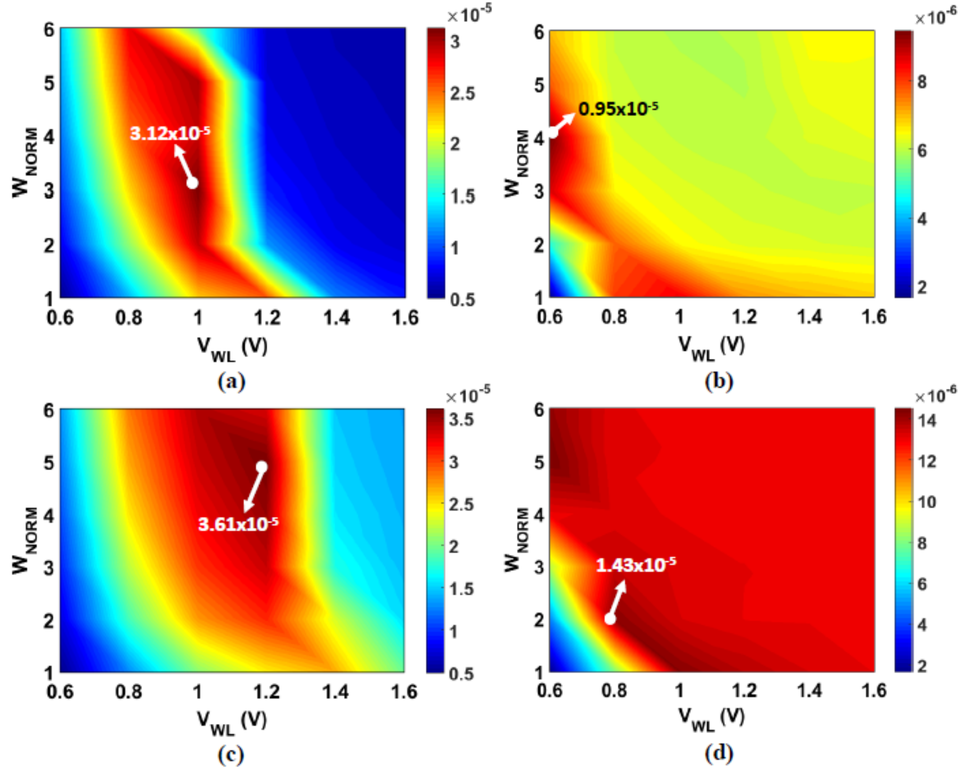


Figure 2.17: 2D surface plots of $\Delta I_{MAX} - V_{BLM}$ product as a function of V_{WL} and W_{NORM} for (a) case 1, (b) case 2, (c) case 3 and (d) case 4.

and W_{NORM}. An EDP vs. V_{BL} for case 4 (V_{WL} = 1.6 V; W_{NORM} = 2) is shown in Fig. 2.18(a). The minimum EDP (EDP_{MIN}) was found to be at V_{BL} = 1.0 V.

Fig. 2.18(b) shows the EDP_{MIN} vs. W_{NORM} at V_{WL} = 1.6 V for various cases. The EDP_{MIN} decreases with increase in W_{NORM} but an area penalty will be involved here due to the increase of NMOS width. Additionally, cases 1 and 4 have a lower EDP by 25% compared to cases 2 and 3, favoring the bit cell configuration with MTJ pinning layer connected to the NMOS. Figures 2.18(c) and 2.18(d) show the EDP_{MIN} of cases 2 and 3 and cases 1 and 4, respectively, as a function of both V_{BL} and V_{WL} at W_{NORM} = 2. The grey area in Figs. 2.18(c) and 2.18(d) shows the write failure region where the MTJ fails to be switched within the write cycle. For a bias condition of V_{WL} = 1.6 V and V_{BL} = 1 V (shown with white dots in Figs. 2.18(c) and 2.18(d)), the EDP_{MIN} for cases 1 and 4 (5.38×10^{-22}) is lower than that of cases 2 and 3 (7.67×10^{-22}) by 30%. Additionally, it

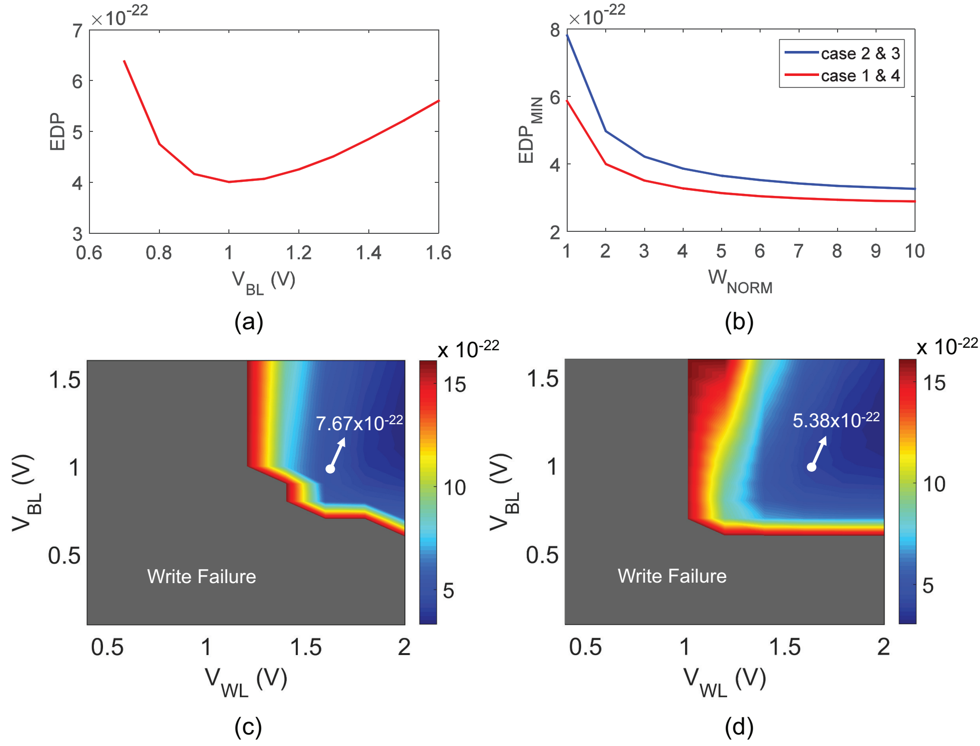


Figure 2.18: (a) EDP vs. V_{BL} for case 4 (at $V_{WL} = 1.6$ V, $W_{NORM} = 2$). (b) The EDP_{MIN} vs. W_{NORM} (at $V_{WL} = 1.6$ V). 2D surface plots of the EDP_{MIN} as a function of V_{WL} and V_{BL} for a fixed width ($W_{NORM} = 2$) for (c) cases 2 and 3, and (d) cases 1 and 4.

is observed that bit cell cases 1 and 4 can be written at a relatively lower V_{BL} (0.7 V) and V_{WL} (1.1 V) compared to cases 2 and 3.

2.7 Summary

Bit-cell forms the fundamental memory blocks in the STT-MRAM array. Here we explored the design space for the STT-MRAM bit-cell design in different configurations to see the read and write performance. Additionally, an overview of the critical peripheral read and write circuitry used for STT-MRAM implementation was discussed. The analysis may be used to find the optimum bias conditions and peripheral circuit design needed to obtain high read-sense margin without sacrificing on the write performance. The nature of the MTJ parametric variations will be discussed in the subsequent chapters.

Chapter 3

MTJ Device Characterization

3.1 Introduction

Development of the novel MTJ stacks for the bit-cell needs innovation in meteorology and characterization; Specifically, characterizing magnetic films, interfaces, and patterned MTJ stacks. Depending on the stage of fabrication, the MTJ device characterization is performed at the film-level, device-level or bit-cell level. Film-level characterization is performed prior to patterning the MTJ device to characterize the film and interface properties. Some of the techniques used for film measurements are Vibrating Spectral Magnetometry (VSM), Magneto-Optic Kerr effect magnetometry and Current in-plane tunneling (CIPT) [61]. Device-level characterization is performed after patterning of the deposited MTJ device and allows observing the MTJ device properties after the etching process. Here a large sample array of patterned devices are considered for characterization and the data is statistically analyzed to extract device parameters. Finally, the bit-cell level characterization is performed to evaluate the functionality and performance after integration with the CMOS circuitry. In this chapter, the main focus is on the device level characterization to extract the MTJ device parameter variations observed during a novel MTJ stack development. The work was done on novel MTJ stacks under development to observe the process dependant issues faced during MTJ stack fabrication. The characterization work was done on the single device MTJ stacks developed at IMEC, Belgium.

3.2 Film-Level Characterization

The first stage of the development process is to develop suitable magnetic film layers that meet the target application. Film-level characterization helps us understand the properties of the material and their interfaces. These un-patterned films are used only for characterization purposes, and the MTJ film layers have to be patterned before using the CMOS structure. Nevertheless, characterization of the blanket film is of paramount importance, since this can reduce the wafer turnaround time significantly. The magnetic materials must possess high PMA and have to be compatible with the MgO barrier to provide high STT efficiency and TMR. Additionally, the free layer material should possess low magnetic damping ($\alpha < 0.01$) [38], which is a challenge for CoFeB based PMA materials. During film-level characterization, many properties of the MTJ junction are evaluated such as the Resistance area (RA) product, TMR, damping factor, PMA parameters and coupling of the pinning layer to the free layer. Several techniques are used for blanket film characterization and some of the methods and the parameters measured are as below.

3.2.1 RA Product and TMR Evaluation

For RA evaluation, the CIPT technique [61] is the most efficient way to characterize the electrical properties from the MTJ interface. Here a 4-point probe is used to inject the current into the top electrode of the MTJ stack. During characterization, the magnetic field can be used to initialize the MTJ state to AP or P state. Based on the distance between the probes placed, the current can flow through the top layer for smaller probe distances, and can flow through the top and bottom layer interfaces for large seperation distance. The film-level TMR, in this case, can be defined as,

$$TMR_{FILM} = (RA_{AP} - RA_P)/RA_P * 100 \tag{3.1}$$

3.2.2 Damping Factor

Gilbert damping constant determines the rate at which the magnetization relaxes to the equilibrium position. Damping constant is contributed from both extrinsic and intrinsic effects in the material. Intrinsic contribution is related to energy transfer between spin and lattice subsystems and is dependant on the materials high atomic number. Thus bulk PMA materials with Pt have high damping constant (0.05 - 1). Materials with low atomic number such as CoFeB have a low damping constant (0.001 - 0.01). Extrinsic contribution

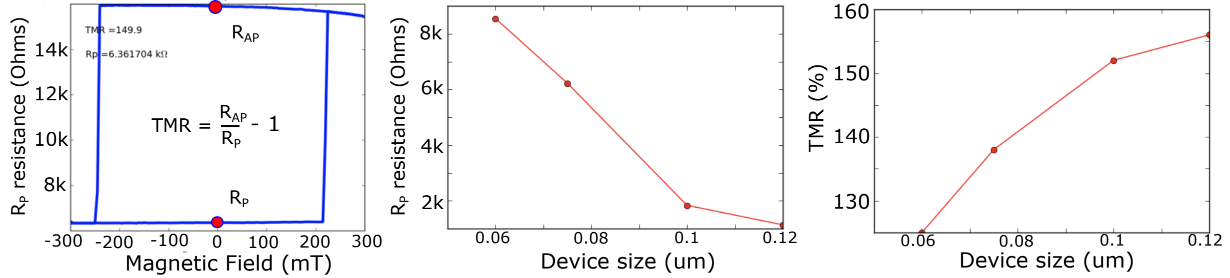


Figure 3.1: Electrical characteristics of the MTJ device. (a) Shows the MTJ resistance vs. applied magnetic field for an R-H loop. (b) shows the R_P resistance vs. device dimension. (c) TMR of the MTJ device vs. device size.

may be due to magnonic scattering, scattering at magnetic defects [62], and dissipation via interaction with adjacent materials. Experimentally, the ferromagnetic resonance can be used to measure the damping constant of the material.

3.2.3 Perpendicular Magnetic Anisotropy (PMA)

Perpendicular magnetic anisotropy is contributed by both the bulk of the PMA material and the interface formed by the magnetic material with the barrier layers. Bulk PMA in the magnetic material contributes to the high anisotropy of the ferromagnetic material, which is improved by doping,or using a buffer layer to provide lattice matching to the magnetic material. Interfacial anisotropy (σ) depends on the change in the crystal symmetry at the interface with respect to the bulk of the material. IPMA materials can be used to provide a good interface and improve the STT-efficiency and TMR of the MTJ. Interface roughness and formation of microstructures (columnar growth or patches) can result in the decrease of the inter-facial anisotropy degrading the TMR of the MTJ junction formed. Thus deposition and fabrication optimization is crucial for minimizing roughness and improving the anisotropy.

3.3 Device-Level Characterization

Patterning is the process of etching the excess material off the MTJ stack to create the desired pattern. A mask is created for this etching process and a suitable etching technique or a combination is utilized. Some of the main etching techniques used are Reactive-ion

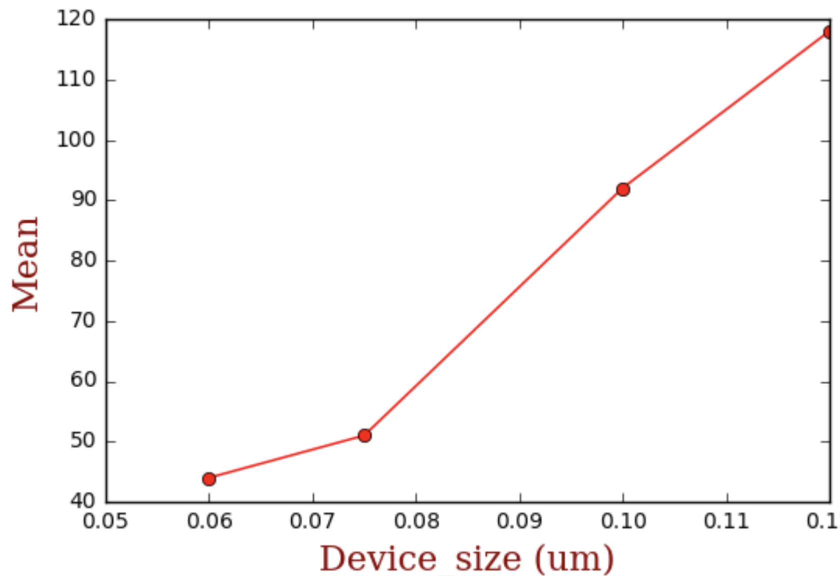


Figure 3.2: Computed electrical diameter of the MTJ device vs. device size.

etching and Ion-beam etching. Gajek et. al [56] utilizes a combination of both the techniques for etching, which results in good patterning results for sub-nm scaled MTJ stacks. Additionally, during the development stage for the advanced nodes, the stack characteristics and fabrication process are not fully optimized. Several devices can incur failure due to defect causing mechanisms such as shape variation, side-wall damage, shorting and opening of the top contact electrode of the MTJ stack [38]. The patterning process can also introduce defects and damage into the MTJ stack, thus early device-level characterization is essential for faster wafer turnaround time. Additionally, the device characterization can provide isolated testing of the MTJ stack alone without the CMOS substructure.

A fast screen process for RH loops can be performed on devices across wafer using an automatic prober, with multiple devices measured in parallel. We can utilize CMOS integration, which is utilizing more transistors in the pathway to the device that is accessed. The transistors, in this case, need to be large and exhibit linear characteristics over a large range of sweep during accessing of the cell. The device-level parameters can then be compared with the measurements obtained from film-level characterization to observe different discrepancies. For instance, we could compare the TMR from the film-level vs. device-level data to see any TMR degradation occurring due to the patterning of the device.

For this work, the characterization was performed on MTJ single device wafers in Imec,

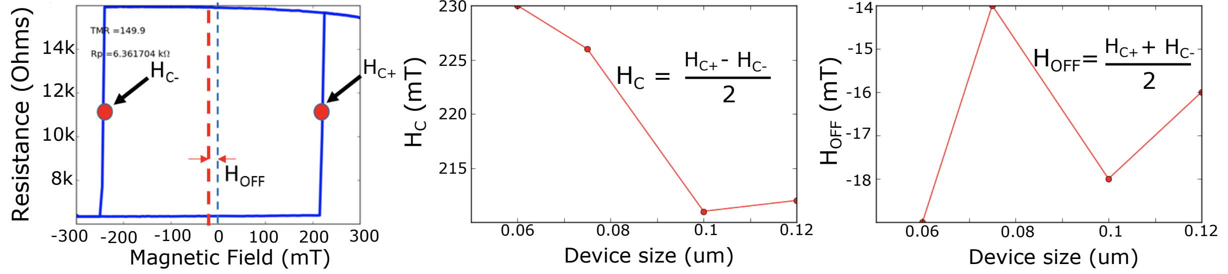


Figure 3.3: Magnetic characteristics of the MTJ device (a) shows the R-H loop showing the H_C and H_{off}. (b) H_C vs device size (c) H_{off} vs. device size.

Belgium. The MTJs are probed directly at the metal pads using the wafer prober stations to measure the electrical characteristics such as R_P, R_{AP}, TMR, R-V Slope, and magnetic parameters such as Coercivity (H_C) and Offset-field (H_{off}). Either an R-H sweep with an electromagnet or an R-V sweep was performed to extract the electrical characteristics. The impact of device scaling on the MTJ device parameters was observed as the first step in characterization. The resistance versus field hysteresis loop is used to quickly obtain the low and high-resistance states, TMR, H_C and H_{off} (characterizing the average stray field at the FL) of the device. Here the MTJ stack wafer is placed within an electromagnet and a magnetic field is applied to observe its resistance. Fig. 3.1 shows the electrical parameters captured from the R–H Loop measurements. Fig 3.1(a) shows the R_{AP} and R_P resistance at near-zero magnetic fields. The R-H loop also shows abrupt switching which signifies a single domain response of the MTJ device. Fig. 3.1(b) shows the median R_P resistance vs device size. This is expected, since as the device size reduces the resistance should increase for a given RA ($12\Omega.\mu m^2$). However, we need to confirm if it is indeed decreasing proportionally. Assuming the un-patterned RA remains constant across the wafer, The electrical diameter of the MTJ device can be computed by,

$$E_D = 2\sqrt{(RA/(\pi R_P))} \tag{3.2}$$

Fig. 3.2 shows the electrical diameter computed. It can be observed that the electrical diameter is slightly lower than the expected device diameter. This is common in advanced MTJ stack processes [56], the difference is resistance may be due to the etch induced damage at the periphery of the barrier that results in increasing of the RA at the edge of the junction [56]. Based on Fig. 3.1(c), the TMR is shown to decrease with a decrease in the size of the MTJ stack. Assuming, the MgO thickness remains constant, this could

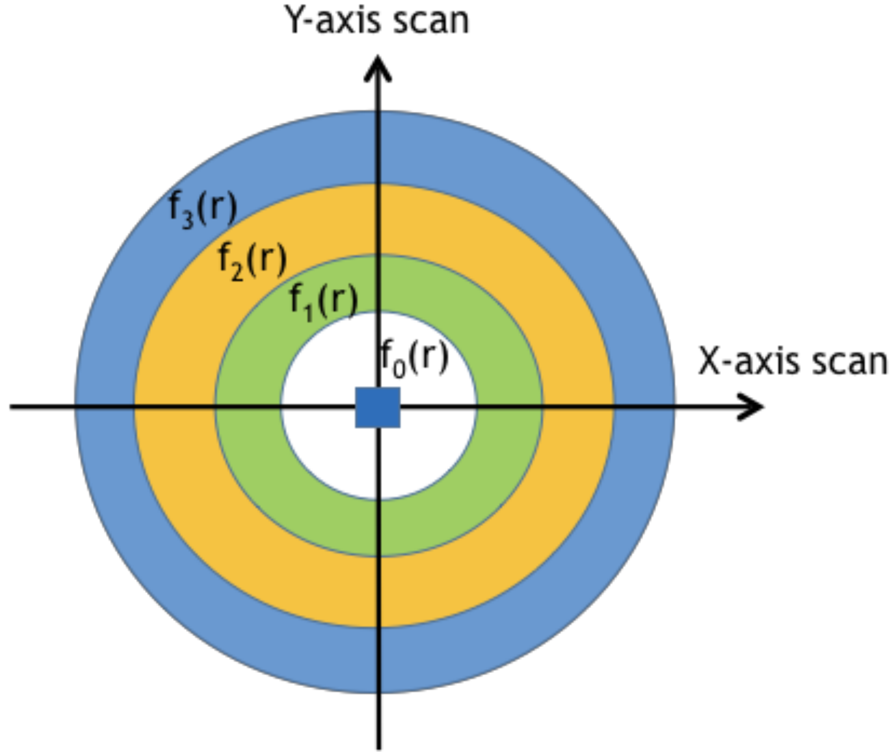


Figure 3.4: Generalization of parameter spatial variation across the wafer.

be due to factors such as pin-hole formation or sidewall deposition at the periphery, which for a given defect area size would have a higher degradation effect on the smaller device dimension due to higher resistance value and smaller device size.

Additionally, the magnetic parameters are observed for the device sizes. Fig. 3.3(a) shows the magnetic parameters such as coercivity (H_C) and offset field (H_{off}) for the MTJ stack. The offset field for the MTJ stack remains constant across device sizes as shown in Fig. 3.3(c), which is expected as per experimental results from work by Gajek et.al [56]. However, H_C was observed to be increasing with a decrease in device size as shown in Fig. 3.3(b), which was contrary to the experimental results in the work [56]. More analysis is required to understand why this is happening. For this thesis, the characterization of the device-level patterned MTJ stacks was performed to understand the variation of electrical characteristics commonly seen during the development of novel advanced stack architectures. The MTJ devices considered here are stand-alone devices with metals pads

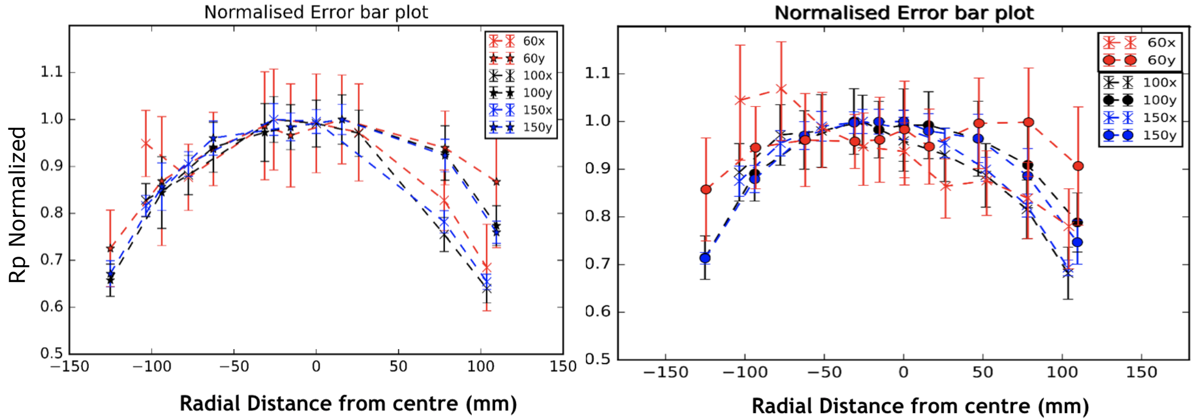


Figure 3.5: R_P resistance plots for MTJ stack of 60, 100 and 150 nm sizes. (a) Spatial distribution of Normalized R_P resistance across wafer 1. (b) spatial distribution of R_P resistance on wafer 2 (CMOS substrate).

to probe the device. Two types of the wafer are considered for characterization. Wafer 1 has MTJ devices deposited on metal surface which is smooth. Wafer 2 has the MTJ devices deposited on the CMOS substrate that also accounts for the surface roughness of the CMOS substrate. Similar measurement conditions are followed for both the wafers during the characterization.

3.4 Objective

The primary motivation for the characterization is to understand the variation behavior of the fabricated devices for, a) Accurate defect modeling b) Feedback for improving the fabrication process steps. Measurement and subgrouping analysis were performed for evaluating the process variation in patterned MTJ single devices for 60, 100, 150 nm devices across the wafer. Fig. 3.4 illustrates the generalization made to ease the analysis. We are assuming and verifying the spatial parameter variation as a radial function of the distance from the center of the wafer. The MTJ parameters such as R_P, and TMR characteristics are sub-grouped to find an analytical expression for spatial variation of device parameters as a function of radial distance. We are observing the TMR and R_P parameters across the wafer. The mean and standard deviation of the parameter is evaluated to extract a suitable fitting function that can model the custom variation characteristics observed in the wafer.

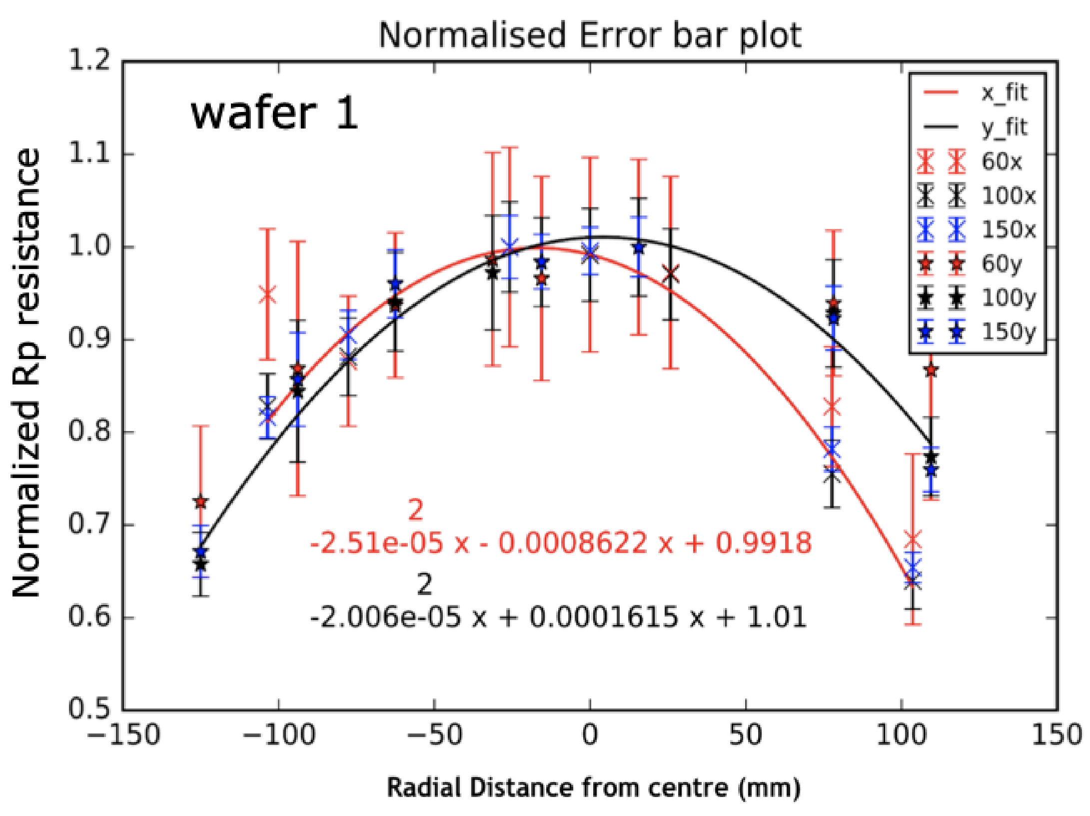


Figure 3.6: Parabolic fitting for the R_P spatial variation along the x and y axis of the wafer.

3.5 Test Setup

Each die consists of patterned single-device MTJ of different dimensions organized in a regular manner. The aim of the measurement is to identify the MTJ parameter variation along the wafer as a function of radial distance from the center of the 300 mm wafer. The dies in the wafer are scanned along the x and y-axis directions to estimate the MTJ parameter variation. Each data point used for statistical analysis is based on the measurement of 3 adjacent dies per X / Y axis position. This is done to get a sufficient number of samples for statistical analysis per X, Y location. An RV scan was performed across the devices in the wafer at 25°C and post-processing data analysis is done to remove shorts and poor TMR devices from the analysis. R_P and TMR are then extracted from the peak

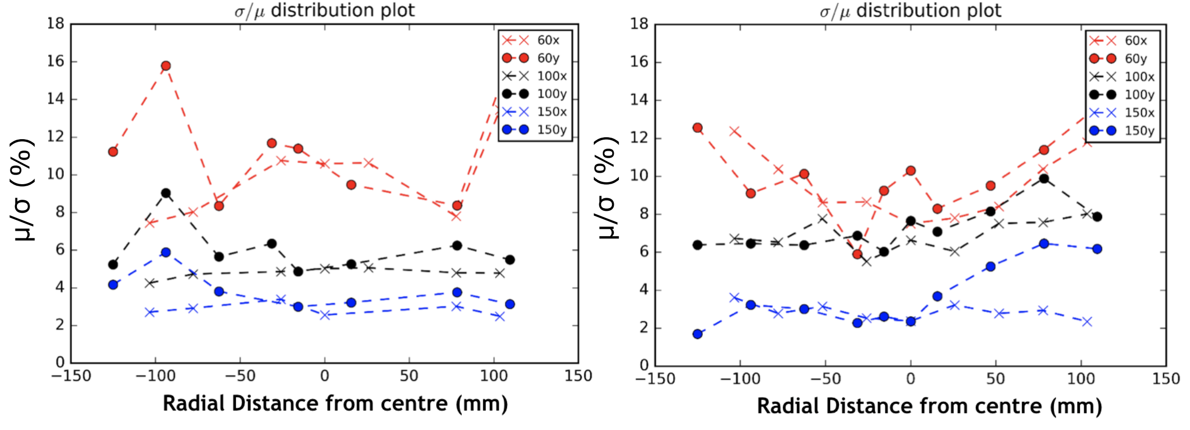


Figure 3.7: Mean to the standard deviation ratio for R_P resistance. (a) shows the spatial distribution for wafer 1 (b) spatial distribution for wafer 2 (CMOS substrate).

of the RV curve. A 1.5 times the inter-quartile range (IQR) rule was followed for outlier detection and removal of parameters, the data points outside the lower and upper limit was considered as an outlier and removed from the analysis,

$$Lowerlimit = P25 - 1.5 * IQR \tag{3.3}$$

$$Upperlimit = P75 + 1.5 * IQR \tag{3.4}$$

Here P25 and P75 and 25th and 75th percentile data, and IQR is the interquartile range given by P75 - P25. Outliers are removed based on the MTJ parameter values (R_P and TMR), and the sample size for each data point. Datapoint that has an insufficient sample size per location (e.g less than 10 samples) is also discarded from the analysis. The figures show the 1σ error bar plot showing the R_P and TMR variation as a function of radius from the center of the wafer. A course and fine scan were done on the wafer to extract the parameters and establish the variation trend.

3.6 R_P Analysis

Resistance voltage together with R-H measurement was done to measure the R_P and TMR parameters for the MTJ devices. Fig. 3.5 shows the normalized R_P resistance across the

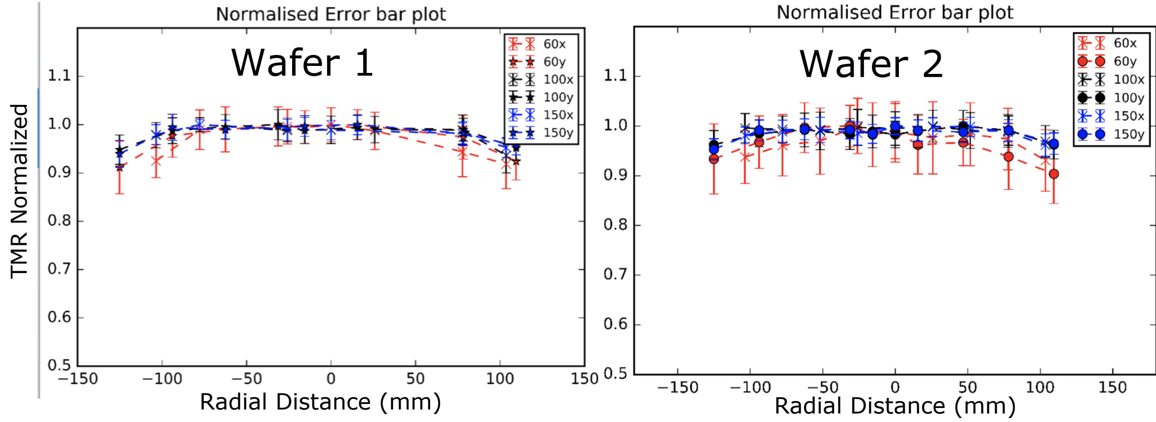


Figure 3.8: Normalized TMR median value for MTJ devices for 60, 100 and 150nm devices. (a) spatial variation on wafer 1. (b) spatial variation on wafer 2 (CMOS substrate).

wafer. Here the resistance is normalized to the peak resistance obtained for a given stack distribution. An error bar plot is done with the median value of the resistances represented by the marker. The dash limit lines of the bar plot indicate the 1-sigma margin of the parameter data point measured. The dashed lines are used to indicate the trend of the parameter deviation across the wafer. It was observed that the R_P parameters deviate based on its spatial location on the wafer and maximum was observed at the center, and its decreases to the edges of the wafer as shown in the normalized R_P distribution shown in Fig. 3.5. It can be seen that the resistance distribution follows a parabolic trend with maximum resistance at the center of the wafer. Provided the RA of the blanket MTJ device is constant across the wafer. The resistance variation could be due to multiple factors such as barrier thickness variation, or etching induced variation in the electrical diameter of the MTJ device. This could be possible provided the etching process induced periphery damage decreases with radial distance from the center of the wafer. An alternate explanation for the resistance decrease farther from the center of the wafer could be due to the increase in sidewall deposition across the MTJ device. The resistance spatial distribution is captured and fitted using a parabolic approximation as shown in Fig. 3.6 the lines show the best fit parabolic curve to model the spatial distribution of the parameter. Fig. 3.7 shows the variation characteristics for R_P resistance, which is the ratio of the standard deviation to the mean value. The 100 and 150 nm have very stable and minimal variation, however, it was observed that the 60 nm variation was slightly higher than the variation of other device sizes. Additionally, the MTJ devices in the center die have minimal variation and increasing outwards. Here an inverted parabolic fitting is done to model the variation

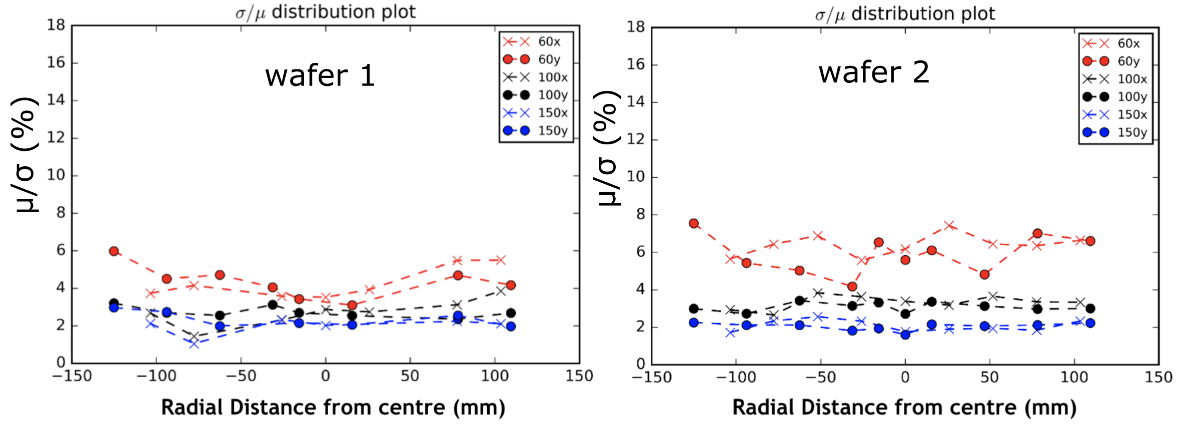


Figure 3.9: Mean to standard deviation ratio for TMR (a) wafer 1. (b) wafer 2 (CMOS substrate deposition).

behavior. The measurements were almost similar in the 2nd wafer.

3.7 TMR Analysis

Based on the measurement of TMR for the MTJ devices, only a smaller deviation in value was observed in wafer 1 due to spatial location as shown in Fig. 3.8. A significant deviation was observed only towards the edge of the wafer as shown in Fig. 3.8(a). However, comparing the TMR values seen in wafer 2 and wafer 1, there is more variation in the TMR value in wafer 2 for 60 nm devices. This could be due to the surface roughness of the CMOS substrate over which the MTJ stack is deposited. This is again confirmed with the TMR spatial variation characteristics seen in Fig. 3.9(b). Here it can be seen that the σ/μ ratio for 60 nm in wafer 2 is slightly higher than the 60nm devices in wafer 1.

3.8 Summary

Equal and uniform parameter distribution is important for ensuring minimal deviation of device parameters and ensure the reliable operation of the system. However, during stack development, it is common to encounter the fabrication process-induced variations that are customized to the foundry used. Characterization provides a means to analyze the

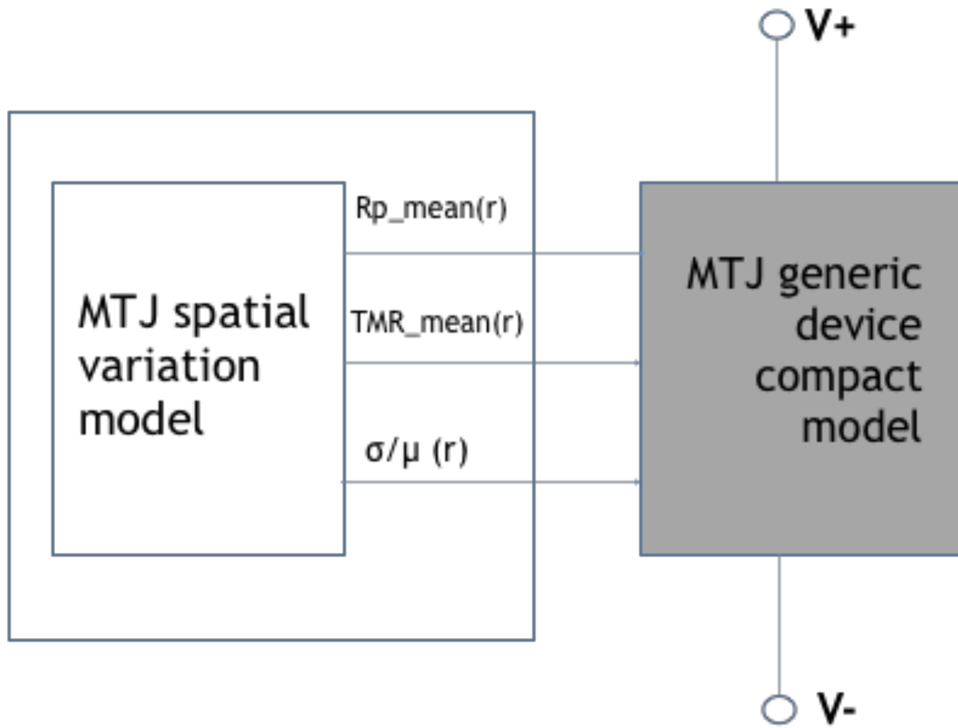


Figure 3.10: Application of the spatial variation model on to a generic device model.

parameter behavior of the MTJ device across the wafer. However, an empirical approach might be useful to capture the variation characteristics to accurately model the spatial variation behavior. From the measurements observed it can be seen that the R_P variation is sensitive to spatial variation, as it decreases with an increase in distance from the center of the wafer and follows a parabolic trend. However, the TMR shows a flat response, with a negligible decrease at the edge of the wafer. The parabolic fitting function can be easily stored and utilized for recreating the mean and variation characteristics of the parameter for accurate modeling of the device during simulation. Fig. 3.10 shows the proposed scheme where the MTJ spatial variation model provides the median and σ/μ to the MTJ generic model for the given parameter based on the location of the MTJ device being simulated. This could be beneficial for the yield estimation of STT-MRAM bit-cell arrays while considering the CMOS circuitry variation across the wafer.

Due to the complexity of fabrication of MTJ and lack of infrastructure, we are attempting to explore other alternatives to creating the functionality of an MTJ used in the STT-RAM bit cell array. Currently, MTJ are difficult to be fabricated in the university environment and

only selected universities with collaboration of industry partners are capable of fabricating MTJ in large quantity with good manufacturing repeatability. Hence, the fabrication of MTJ is difficult in this university, we are considering other options where the MTJ element can be replaced with device that would ideally mimic its resistance characteristics, which should be a voltage controlled current source element. The closest element that can be used to emulate the resistance characteristics of an MTJ is a transistor. This motivated us to pursue transistor based DFT schemes for STT-MRAM characterization.

A DFT based test chip is proposed that could capture the variation behavior observed in MTJ devices and replicate the behavior for testing the peripheral circuitry and system level functionality of STT-RAM sub-array. The test chip is designed to be the first step towards building the STT-RAM array.

Chapter 4

A Parametric DFT Scheme

4.1 Overview

A test platform can enable parametric optimization and verification process using CMOS-based design-for-testability (DFT) circuits. In this chapter, a DFT algorithm and a DFT circuit is implemented for parametric testing and pre-qualification of the critical circuits in the CMOS wafer. The DFT circuit successfully replicates the electrical characteristics of MTJ devices and captures their spatial variation across the wafer with an error less than 4%. We implemented the integrated DFT circuit and the read access sensing path in 65-nm CMOS technology to evaluate the read margin characteristics of the column. The yield estimation results provide insight into the response of the MRAM to MTJ parameter variations across the wafer as well as an estimate on the wafer usable area. The read sensing path implementing the DFT circuit can replicate resistance-area product variation up to 50% from its nominal value. The yield data from the read sensing path at different wafer locations are analyzed, and a usable wafer radius up to 75 mm has been estimated.

4.2 Approach for the DFT Scheme

The spatial non-uniformity in device parameters is one of the factors that impact the MTJ wafer yield. To understand this, extensive characterization was carried out by measuring the single device MTJ film level parameter variations across the wafer [38]. Spatial non-uniformity in MTJ device parameters can be attributed to the variations in RA product, barrier oxide thickness (t_{ox}), barrier quality, as well as geometric dimension variations

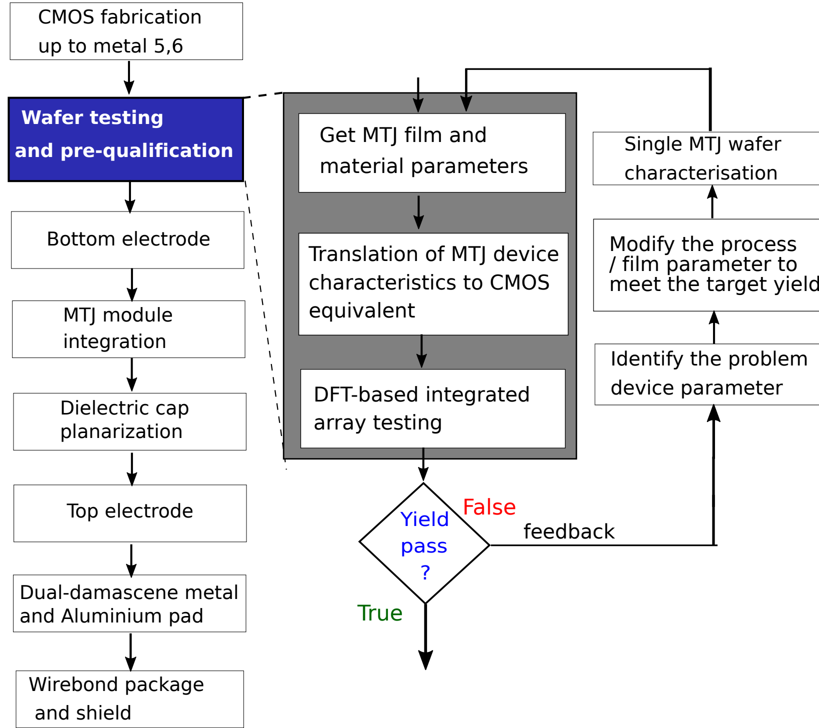
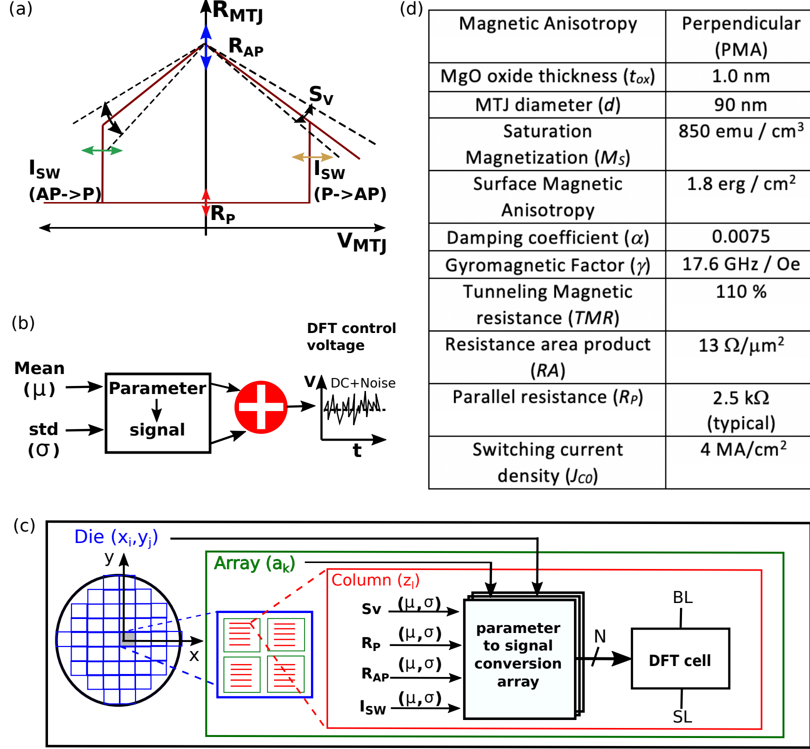


Figure 4.1: (Left panel) STT-MRAM fabrication process flow, where the wafer testing and pre-qualification step is included. (Right) The process for testing and qualification adopted is shown in detail. The grey blocked region illustrates the proposed scheme.

along the wafer. Some of the magnetic parameters such as magnetic anisotropy (H_K) and saturation magnetization (M_S) also contribute to MTJ device variations [42], resulting in fluctuations in thermal stability and switching characteristics of the MTJ device [38]. MTJ parameter variations can be classified as within-die variations and die-to-die (D2D) variations based on the spatial information, where D2D process variations are arbitrary and relatively larger compared to within-die variations [63]. For instance, die-to-die variation of more than 10% in RA product was reported for MTJ in a 200-mm wafer [64]. Other non-MTJ related factors that affect the yield margin are the variations in contact and interconnect resistances [65], which depend on the quality of the connection formed while integrating the MTJ.

STT-MRAMs employ a post-CMOS back-end-of-the line (BEOL) deposition process, where the MTJ module is integrated after CMOS fabrication steps [26][28]. Fig. 4.1 shows



Magnetic Anisotropy	Perpendicular (PMA)
MgO oxide thickness (t_{ox})	1.0 nm
MTJ diameter (d)	90 nm
Saturation Magnetization (M_S)	850 emu / cm^3
Surface Magnetic Anisotropy	1.8 erg / cm^2
Damping coefficient (α)	0.0075
Gyromagnetic Factor (γ)	17.6 GHz / Oe
Tunneling Magnetic resistance (TMR)	110 %
Resistance area product (RA)	13 Ω/μm^2
Parallel resistance (R_P)	2.5 kΩ (typical)
Switching current density (J_{C0})	4 MA/cm^2

Figure 4.2: (a) Generalized MTJ device parameter variations observed during single device level characterization. (b) MTJ device parameter mean and variance translated to control voltage. (c) Top level diagram for the DFT scheme. Here BL and SL are bit-line and source-line of the MTJ column. (d) Typical parameter values used for MTJ.

the embedded perpendicular STT-MRAM process flow used for the CMOS 2X low-power platform where the MTJ is integrated after either the metal 4 (M4) or metal 5 (M5) copper layers [26][28]. At this process step, the CMOS circuits are already integrated (except for the MTJ module) using the lower-level metal layers. The DFT scheme allows us to test and prequalify the wafer at the earlier stages of the fabrication process. We place the proposed DFT circuit in each column and electrically connect them between bit-cell access transistor and the bit-line present at the last-level metal (M3) below the MTJ module bottom electrode. Each DFT cell in the column is then biased, and accessed in a time-multiplexed manner to replicate the position-dependent MTJ device behavior at that particular location. The yield output is then measured using the read sensing path

at each column of the array as shown in 4.1. If the yield requirements are met, then the wafer is ready for the MTJ integration. However, if the expected yield is not satisfactory, the feedback from the pre-qualification step can be used to modify the parameters and optimize the fabrication process. Thus, this technique can provide a platform for testing and pre-qualification of the wafer.

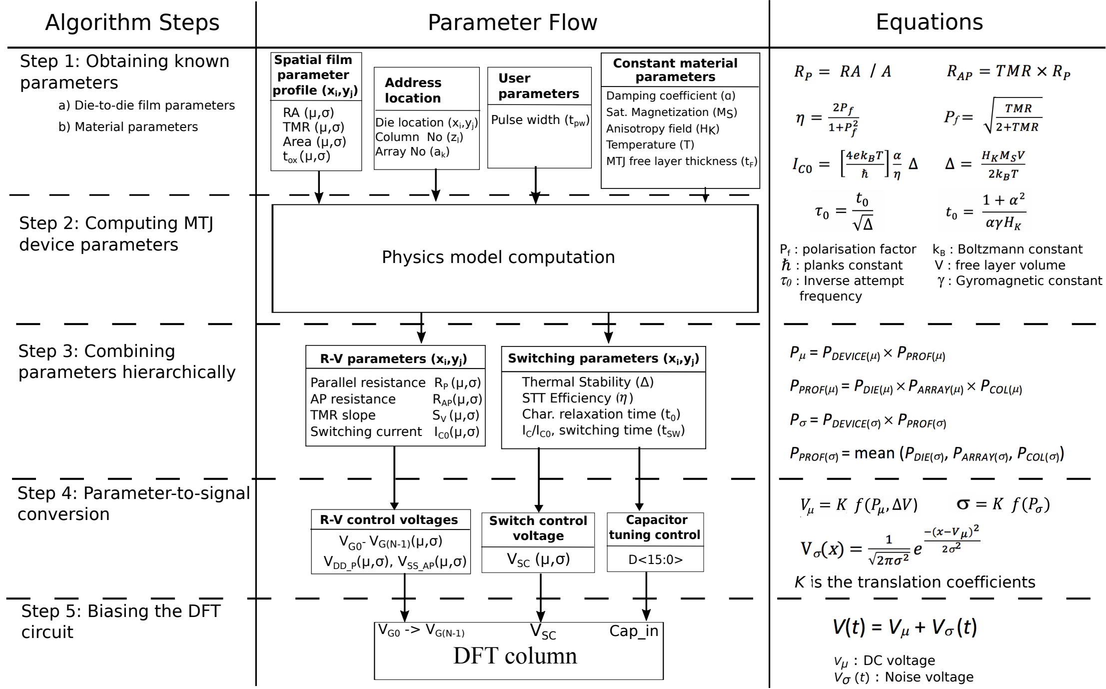
In this work, the DFT scheme is composed of the DFT cells at each column of the memory array with the parameter generation framework to bias them. We compute the MTJ device parameters from film-level parameters and material parameters. Some of the main MTJ device parameters that are captured in the testing process are shown in Fig. 4.2(a). These parameters are translated to voltage signals, which are applied to the DFT circuit during testing as shown in Fig. 4.2(b). Fig. 4.2(c) illustrates the overall DFT scheme, where the MTJ device parameter generation is performed hierarchically at the column, array and die location within the wafer.

4.2.1 Parameter Generation Framework

First, we present a framework to generate the MTJ device parameters. The parameter computation, hierarchical signal combining, and translation to control signals are implemented in the MATLAB –LabVIEW interface based on the algorithm summarized in Fig 4.3(a). The algorithm steps for parameter generation (the left column of Fig. 4.3) are explained below.

1. *Obtaining known parameters:* Known parameters refer to the material and film parameters already available from the film-level characterization of MTJ single device and model-based simulation results [38]. Spatially varying parameters, such as RA profile, TMR and area of MTJ, thickness of oxide layer (t_{ox}) are introduced, which vary across the wafer. The pulse width duration (t_{pw}) is also provided extrinsically to compute the switching behavior. A summary of the typical MTJ parameter values is shown in Fig. 4.2 (d).

2. *Computing MTJ device parameters:* The R–V and switching parameters are calculated based on the MTJ physics equations from a previous work [38]. MTJ models based on the uniform switching approximation is used [38]. The main device parameters considered are (i) parallel resistance (R_P), (ii) tunnel magnetoresistance (TMR), (iii) TMR slope (S_V), and (iv) critical switching current (I_{C0}). Examples of the computed parameters are shown in Fig. 4.2(d). Each parameter in the subgroup is defined in terms of its mean (μ) and standard deviation (σ) at a spatial location



$$R_P = RA\ /\ A \qquad R_{AP} = TMR \times R_P$$
$$\eta = \frac{2P_f}{1+P_f^2} \qquad P_f = \sqrt{\frac{TMR}{2+TMR}}$$
$$I_{C0} = \left[\frac{4ek_BT}{\hbar}\right]\frac{\alpha}{\eta}\Delta \qquad \Delta = \frac{H_K M_S V}{2k_BT}$$
$$\tau_0 = \frac{t_0}{\sqrt{\Delta}} \qquad t_0 = \frac{1+\alpha^2}{\alpha\gamma H_K}$$
$$P_\mu = P_{DEVICE(\mu)} \times P_{PROF(\mu)}$$
$$P_{PROF(\mu)} = P_{DIE(\mu)} \times P_{ARRAY(\mu)} \times P_{COL(\mu)}$$
$$P_\sigma = P_{DEVICE(\sigma)} \times P_{PROF(\sigma)}$$
$$P_{PROF(\sigma)} = \text{mean}\,(P_{DIE(\sigma)}, P_{ARRAY(\sigma)}, P_{COL(\sigma)})$$
$$V_\mu = K\ f(P_\mu, \Delta V) \qquad \sigma = K\ f(P_\sigma)$$
$$V_\sigma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-V_\mu)^2}{2\sigma^2}}$$
$$V(t) = V_\mu + V_\sigma(t)$$

Figure 4.3: Parameter generation framework to generate bias voltages for the DFT circuit operation.

with respect to the center of the wafer. The equations used to compute mean and standard deviation of basic film parameters and material parameters are shown in the right column of Fig 4.3.

3. *Combining parameters hierarchically:* During single device MTJ characterization and statistical analysis, it is common practice to analyze and represent the measured parameter data in terms of the nominal value (P_{NOM}) and normalized spatial variation profile (P_{PROF}). Here we utilize the same convention for MTJ parameter generation. The normalized variation profiles are obtained by combining the die (P_{DIE}), array (P_{ARRAY}) and column-level (P_{COL}) variations. The device parameters are accessed during testing based on the address location of the DFT cell, such as column address (z_l), array address (a_k) and die location (x_i, y_j).

4. *Parameter-to-signal conversion*: The basic principle is to find the circuit equivalent parameters for transistors to replicate the MTJ device characteristics. Here the MTJ parameters, mean (P_μ) and standard deviation (P_σ) are translated to a constant

bias voltage (V_μ) and standard deviation (σ), respectively. Based on the type of translation fitting coefficient (K), we are performing two main types of conversions.

a. Resistance to MOS gate bias voltage (K_G): To find the V_{GS} needed to achieve the required R_{DS} for R-V biasing ($V_{G0} - V_{G3}$, V_{SS_AP}, V_{DD_P}). The ON resistance of transistor in triode mode can be expressed as

$$R_{\text{DS}} \propto \frac{K_G}{V_{\text{GS}}} \tag{4.1}$$

where

$$K_G = \frac{1}{\mu C_{\text{ox}} \frac{W}{L}}. \tag{4.2}$$

Here K_G is the R-V translation coefficient and μ, C_{ox} and W/L are the transistor process information and sizing. The coefficients for each transistor in the DFT are found empirically through SPICE simulations.

b. Voltage to latch bias voltage (K_V): To set the latch switching voltage (V_{SW}) given by

$$V_{\text{SW}} = \frac{rV_{\text{SC}}}{r+1}, \tag{4.3}$$

where

$$K_V = \frac{r}{r+1}. \tag{4.4}$$

Here K_V is the V-V translation coefficient, and r is the ratio of the relative drive strengths for the PMOS to NMOS transistors in the latch [66]. Since the DFT circuit is based on CMOS, both temperature and global process corner compensation voltages (ΔV) have to be applied to match it to the MTJ behavior. The strategy for compensation is described in detail later in Section III, design considerations for the DFT circuit.

The parameter mean values (P_μ) maintain the correlation based on the MTJ physics-based equations. However, we are assuming Gaussian distribution for the standard deviation of parameters (P_σ). Thus, the stochastic samples generated may have a lesser degree of correlation between parameters. The standard deviation, σ of the noise voltage depends on the parameter standard deviation (P_σ) and the translation coefficient (K_σ). A clocked Gaussian distributed noise source is used to create a stochastic random voltage value (V_σ) for a given test clock cycle. This scheme allows for the single DFT cell to mimic the spatially variant bit-cell behavior in a time-multiplexed manner.

5. *Biasing the DFT Circuit:* The DC (V_{μ}) and the noise voltage (V_{σ}) are combined to get the control voltages $V(t)$ that is applied to the DFT circuit. The test bench creates the following control voltages.

   a. *R–V* control voltages (V_{G0}–$V_{G(N\text{-}1)}$): represents the TMR slope with respect to the bias voltage. The gate and body bias of the PMOS transistors in the R–V circuit is varied for coarse and fine adjustment, respectively. N represents the number of PMOS gate inputs used.

   b. *Multiplexer control voltages (*$\mathrm{V_{DD_P}}$ *and* $\mathrm{V_{SS_AP}}$*)*: represents the V_{STATE} applied to create the MTJ parallel (P) and anti-parallel-state (AP) resistance. Offset voltages (V_{DD_P} and V_{SS_AP}) are added to this to compensate for the global process corner deviation and temperature. These voltages are applied to V_{DD} and GND terminals of the circuit as shown in Fig. 4.4(c).

   c. *Latch control (*V_{SC}*)*: sets the switching voltage of DFT cell (V_{SW}), which replicates the bit-cell switching operation. V_{SC} is used for fine and quick stochastic variation in switching voltages of the DFT cell.

   d. *Capacitor tuning control (D<0:15>)*: used for course tuning of the time-dependent switching characteristics of the DFT cell. Here the time constant for the DFT switching time is varied by selecting among different capacitor banks. We are manually adjusting it to obtain the switching characteristics. The testing configuration allows 16-bit control values for the capacitance selection and flexibility. The number of bits is decided based on the precision of capacitance tuning required. We are using a serial port interface and shift register to provide the 16-bit digital input to the die.

4.2.2 DFT-Cell Operation

Fig. 4.4(a) shows the DFT cell consisting of the *R–V* bias circuit, latch circuit and the NMOS access transistor (N1), operating together in unison based on the control signal (labeled red). The *R–V* bias and latch circuit is referred to as the DFT circuit whose function is to replicate the MTJ behavior. The latch circuit stores the state (AP or P) of the DFT circuit and controls the *R–V* bias circuit to exhibit the resistance behavior for the given state. Fig. 4.4(b) shows the schematic diagram of the latch circuit. Prior to operation, the Q or $\mathrm{Q_B}$ node of the latch is pre-charged to V_{DD} using a SET or RESET signal. Here the pre-charged node (Q or $\mathrm{Q_B}$) depends on the MTJ bit-cell access condition

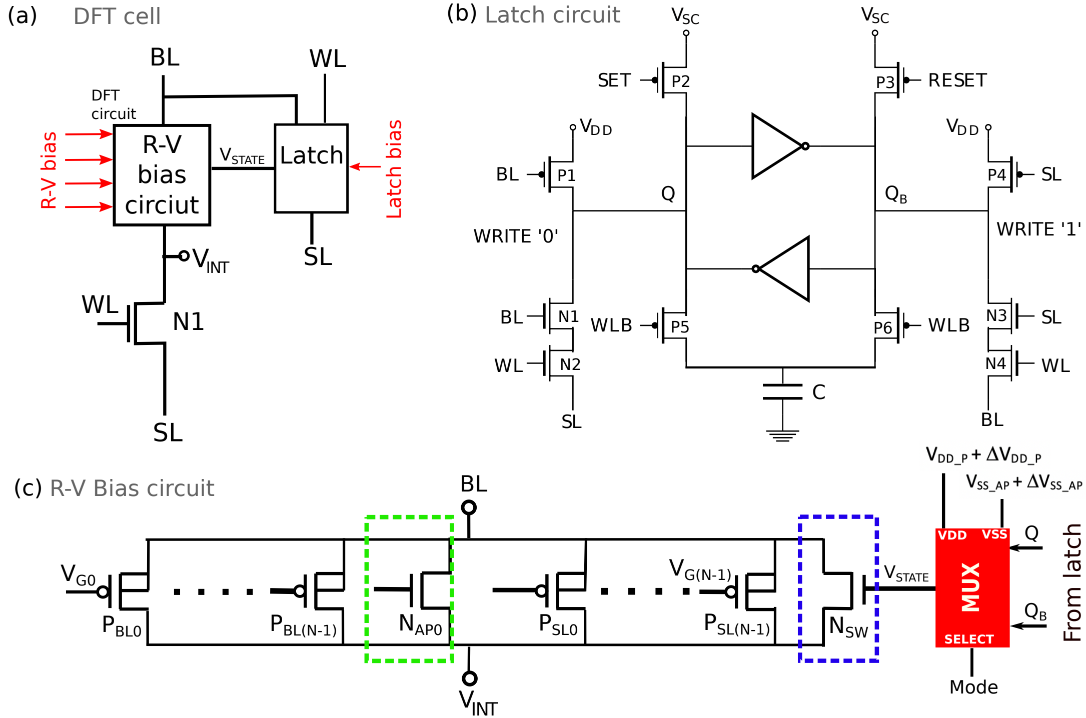


Figure 4.4: DFT cell consisting of R-V bias circuit, latch, and the NMOS access transistor. (b) Latch circuit. (c) Generalized R-V bias circuit. BL, SL and WL represent bit-line, source-line and word-line inputs, respectively.

(regular (V_{SL} = GND) or source-degenerated (V_{BL} = GND)). When the bit-cell is accessed, the Q or $\mathrm{Q_B}$ node is pulled down to GND by the pair of NMOS transistors connected in series (N1, N2) or (N3, N4) as shown in Fig 4.4(b). The pull-down strength depends on the applied bit-line (BL), word-line (WL) and source-line (SL) voltages. The NMOS pull-down transistors lower the Q node voltage until the latch reaches its switching threshold (V_{SW}), resulting in the flipping of the latch state.

The multiplexer (MUX) circuitry selects the latch output (Q or $\mathrm{Q_B}$) and feeds back to the R–V circuit to signal the state. The mode select for the multiplexer allows replication of regular (fixed layer connected to NMOS) or reverse-connected configuration (free layer connected to NMOS). The multiplexer output (V_{STATE}) indicates the state of the DFT circuit (AP or P) as shown in Fig. 4.4(c). When $\mathrm{N_{SW}}$ is OFF (AP state), and the resistance is the parallel equivalent of the ON transistor resistances. When $\mathrm{N_{SW}}$ is turned ON (P state), and the NMOS $\mathrm{N_{SW}}$ source-drain resistance comes in parallel with other

transistors resistances, reducing the equivalent resistance to the designed R_P value.

4.2.3 Replication of MTJ Characteristics Based on the DFT Circuit

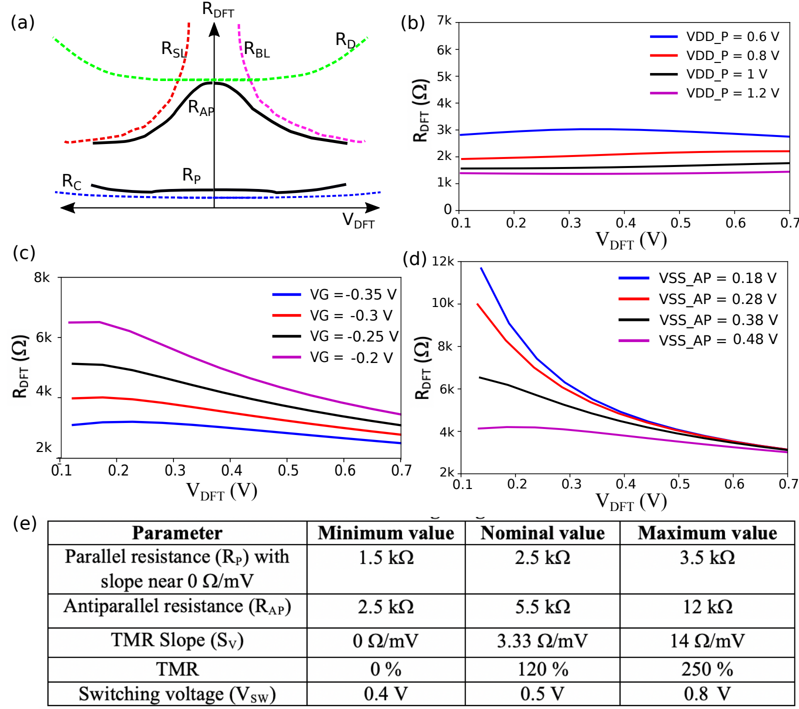
All the transistors in the R–V circuit are biased in the linear mode of operation, and individually contribute to different resistance curves as shown in Fig. 4.5(a). Thus, the architecture of the DFT circuit allows for selective activation of specific parameter variations by appropriately biasing the transistor sub-groups The PMOS transistors are sub-grouped into $\mathrm{P_{BL\mathit{x}}}$ and $\mathrm{P_{SL\mathit{x}}}$, to emulate the MTJ R–V curve in the positive and negative V_{DFT} voltages. In principle, N transistors enables flexibility to tune N resistances via the gate biasing, however it results in more number of control voltage generation. Here we are introducing variation in three key R-V parameters (R_P, R_{AP}, S_V) using 4 PMOS and 1 NMOS. The transistors group, functionality and their equivalent resistances as shown in Fig. 4.5(a) are,

1. $\mathrm{P_{BL}}$ transistors are responsible for the AP state R–V curve for positive bias (R_{BL}). Here $\mathrm{P_{BL0}}$ (with $\mathrm{N_{SW}}$ at $V_C = V_{SS_AP}$) controls the R_{AP} resistance by generating the R_D curve in the positive V_{DFT} voltages. $\mathrm{P_{BL1}}$ controls the S_V in the positive voltages.

2. $\mathrm{P_{SL}}$ transistors are responsible for the AP state R–V curve for negative bias (R_{SL}). Here $\mathrm{P_{SL0}}$ (with $\mathrm{N_{SW}}$ at $V_C = V_{SS_AP}$) controls the R_{AP} resistance by generating the R_D curve in the negative V_{DFT} voltages. $\mathrm{P_{BL1}}$ controls the S_V in the positive voltages.

3. $\mathrm{N_{AP0}}$ (optional) is sized so that equivalent parallel resistance of $\mathrm{N_{AP0}}$ and $\mathrm{P_{BL0}}$ can match the R_D for better R_{AP} match near $V_{DFT} = 0$ V. $\mathrm{N_{SW}}$ determines the R_P when it is ON ($V_C = V_{DD_P}$).

For the P state, the DFT resistance characteristic is obtained by using the parallel combination of both NMOS and PMOS for a constant resistance across varying V_{DFT}. Here the R_P and R_{AP} resistance can be independently varied by tuning V_{DD_P} and V_{SS_AP} of the MUX circuit as shown in Fig. 4.5(b) and 4.5(d). Variation in V_G results in shifting of R_{AP} at 0 V and R_{AP} at 0.7 V as shown in Fig 4.5(c). Here R_{AP} can be varied with constant slope (S_V) by changing V_G and V_{SS_AP}. Although the parameters can be varied independently, MTJ physics models, shown in Fig. 4.3, are used to closely capture the interdependency of the parameters as observed in real MTJ operation. The tuning range of DFT cell parameters are shown in Fig. 4.5(e)



Parameter	Minimum value	Nominal value	Maximum value
Parallel resistance (R_P) with slope near 0 Ω/mV	1.5 kΩ	2.5 kΩ	3.5 kΩ
Antiparallel resistance (R_{AP})	2.5 kΩ	5.5 kΩ	12 kΩ
TMR Slope (S_V)	0 Ω/mV	3.33 Ω/mV	14 Ω/mV
TMR	0 %	120 %	250 %
Switching voltage (V_{SW})	0.4 V	0.5 V	0.8 V

Figure 4.5: Selective control of DFT characteristics using individual control voltages. (a) R_{DS} resistances contributed by each transistor group in the DFT circuit. (b) Changing the R_P resistance using V_{DD_P}. (c) Impact of changing V_G on RAP. Here $V_G = V_{G0} = V_{G1} = V_{G2} = V_{G3}$. (d) Changing the R_{AP} resistance using V_{SS_AP}. (e) DFT cell parameter tuning range. The results were based on 65–nm process pre–layout simulations.

4.3 Compensating CMOS-Based Non-Idealities in the DFT Circuit

The DFT circuit is designed to match the target MTJ characteristics. Not only is an ideal DFT circuit able to replicate the MTJ device behavior but also should be immune to CMOS process variations. We analyze the effect of local, global process, and temperature variation and provide countermeasures to minimize the interfering of CMOS circuit non-idealities into the testing process.

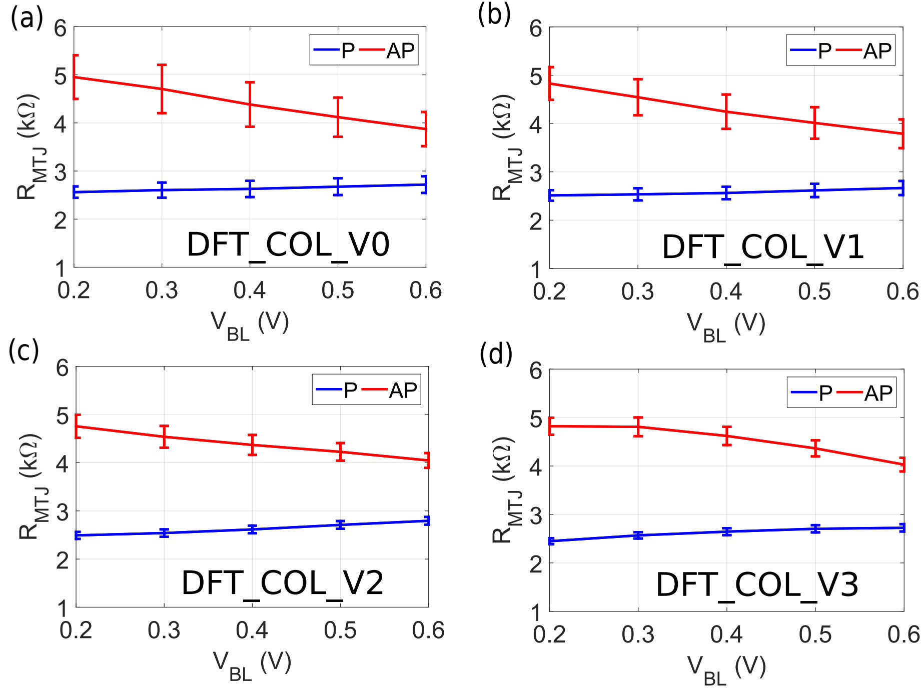


Figure 4.6: R-V variation for different columns of the DFT cell arrays based on post layout simulations. (a) DFT_COL_V0 (L = 1X), (b) DFT_COL_V1 (L = 2X), (c) DFT_COL_V2 (L = 3X), and (d) DFT_COL_V3 (L= 4X).

4.3.1 Local Variation

Transistor sizing plays a key role in determining the parameter tunable space and its variation. The width (W) and length (L) of transistors are sized such that its equivalent resistance matches the target nominal R_{AP} and R_P value. The ON resistance of PMOS in the linear mode can be given as,

$$R_{\mathrm{ON}} \propto \frac{1}{\frac{W}{L} C_{\mathrm{OX}} (V_{\mathrm{GS}} - V_{\mathrm{TP}})} \tag{4.5}$$

where C_{OX} is the oxide capacitance and V_{TP} is the threshold voltage for the PMOS. The gate resistance tuning sensitivity ($dR_{ON}/dVgs$) can be increased by decreasing the W/L of the PMOS transistors, thereby obtaining wider parameter operating range. The

(a)

Target STT-MRAM bit cell (1T-MTJ)	
V_{DD}	1.1 V / 1.8V
W/L for N1	270 nm / 40nm
TMR	110 %
R_P	2.5 kΩ
R_{AP}	5.1 kΩ
S_V	6.5 Ω/mV

(b) 65 nm CMOS process

DFT cell name	(W/L) sizing for P_{BL0}, P_{SL0}, N_{SW}	1-Sigma Transistor variability			
		R_{P_MIN}	R_{P_MAX}	R_{AP_MIN}	R_{AP_MAX}
DFT_COL_V0	140 nm / 60 nm (1X)	3.75 %	5 %	4.38 %	7.8 %
DFT_COL_V1	240 nm / 120 nm (2X)	2.23 %	2.65 %	2.4 %	4.7 %
DFT_COL_V2	360 nm / 180 nm (3X)	1.6 %	1.9 %	1.85 %	4 %
DFT_COL_V3	500 nm / 240 nm (4X)	1.2 %	1.5 %	1.42 %	3.2 %

Figure 4.7: (a) Process parameters for the target MTJ characteristics (b) Summary of the resistance variation due to transistor variability for each design.

parameter tuning range is bound by the maximum control voltage that can be applied without breaking down the transistor. The previous approach could imply increasing parameter tuning range while decreasing the length of the channel. However, there is a trade-off involved here. As the length decreases, the random dopant fluctuation and line edge roughness in the CMOS fabrication process begin to impact the transistor operation by inducing unwanted variations. Pelgrom's model gives the threshold voltage variance experienced by the transistor as $\sigma_{\mathrm{VT}} = \frac{\sqrt[4]{N}}{\sqrt{WL}}$, where N is the channel doping concentration [67]. In the design, the transistor WL should be large enough but the variability exhibited by the DFT circuit should be less than the variation of MTJ parameter being replicated.

Monte-Carlo simulations are performed for 1,000 DFT cells to extract the resistance deviation behavior at AP and P mode. Fig. 4.6 illustrates the CMOS process-induced variability in each column (V0 to V3) having different transistor size (L). Here L is increased from minimum size (60 nm) to 2X, 3X and 4X. The resistance deviation in the read/write operating region of 0.2 V to 0.6 V is considered for the analysis. W/L ratio between transistors is maintained so that the target MTJ resistance characteristics, as shown in Fig. 4.7(a), is achieved. From the table in Fig. 4.7(b), it can be seen that the resistance variation decreases with increase in L, resulting in more accurate replication of the test resistance as

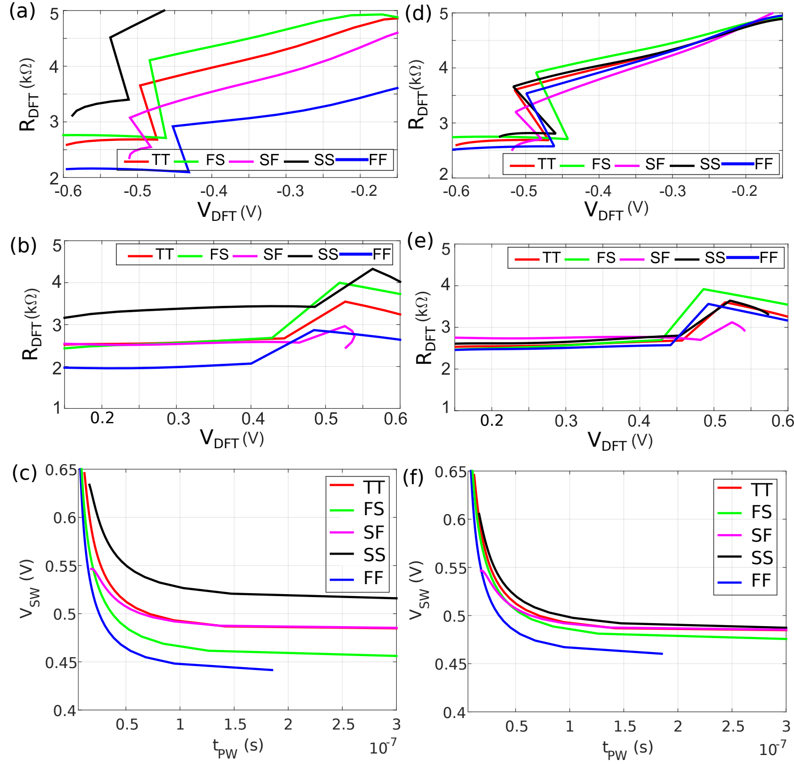


Figure 4.8: Impact of global process corner (a)-(c) before correction and (d)-(f) after correction. R_{DFT} vs. V_{DFT} for (a,d) AP state and (b,e) P state. (c,f) V_{SW} vs. t_{pw} for P to AP. TT, FS, SF, SS, FF represent typical-typical, fast-slow, slow-fast, slow-slow, fast-fast, respectively.

intended. The gate tuning sensitivity remains the same, since W/L remains constant. It was seen to be $dR_{AP}/d_{VSS_AP} = 24\ \Omega/mV$ and $dR_P/d_{VDD_P} = 4\ \Omega/mV$. For further yield testing, DFT_COL_V2 is chosen to limit the resistance variation below ±4%.

4.3.2 Global Process Corner and Temperature

Global process corner effects are lumped together as a threshold voltage (V_T) deviation that affects the R_{DS} of a transistor. The test chip simulation results in Fig. 4.8(a)-(c) show the effect of global process corners on the R_{DFT} - V_{DFT} for AP→P and P→AP, and switching characteristics of the DFT circuit. It is seen that the change in V_T spreads the circuit

characteristics away from the intended behavior. However, the circuit resistance deviation can be corrected by applying compensation bias voltages (V_{G_GLOB} and V_{SC_GLOB}) during lab testing. The compensation voltage is implemented as an offset to the V_{DD_P} and V_{SS_AP} of the MUX to correct for the P and AP resistance deviation, respectively. Fig. 4.8(d)-(f) show the DFT circuit characteristics after correction, where the resistance and switching curves closely match the typical-typical (TT) corner. Similar strategy is adopted to compensate for the impact of temperature. In MTJ's, the resistance and TMR decrease with increase in temperature [29]. On the contrary, the transistor R_{DS} increases with temperature, leading to the increase of the DFT circuit resistance. Here the DC bias offset voltages (V_{G_TEMP} and V_{SC_TEMP}) are applied to compensate for the temperature-induced deviation in a similar manner as the correction for global process corner. Figs. 4.9(a)-(c) show the R-V characteristics of the DFT circuit along with the actual MTJ data at temperatures of 25º C, 85º C and 125º C . Fig. 4.9(d) shows the voltage-to-temperature fitting curves. This DC bias compensation technique shifts the DFT circuit resistance and the switching curve to fit the MTJ experimental data at different temperatures within an error of ±3%. The combined effects of global process corner and temperature can be corrected by applying the bias voltage to the DFT cell as shown below

$$\Delta V_G (= \Delta V_{G_GLOB} + \Delta V_{G_TEMP}) \tag{4.6}$$

$$\Delta V_{SC} (= \Delta V_{SC_GLOB} + \Delta V_{SC_TEMP}) \tag{4.7}$$

4.4 Test Chip Implementation and Results

Fig. 4.10 shows the die micrograph of the DFT circuit array layout in 65 nm CMOS technology. The test chip consists of 4 pairs (A and B) of columns (DFT_COL_V0, DFT_COL_V1, DFT_COL_V2, DFT_COL_V3), where each column consists of 64 DFT cells and a reference cell. Each column pair has a read sense amplifier and tunable capacitor banks placed at the end of the column, allowing for the integrated testing of the DFT cell with the read access path circuitry. All the timing signals are generated from an on-chip common timing and control circuitry. Internal programmable clock signals with tunable pulse width is used for generating the read word-line pulse signal (PULSE).

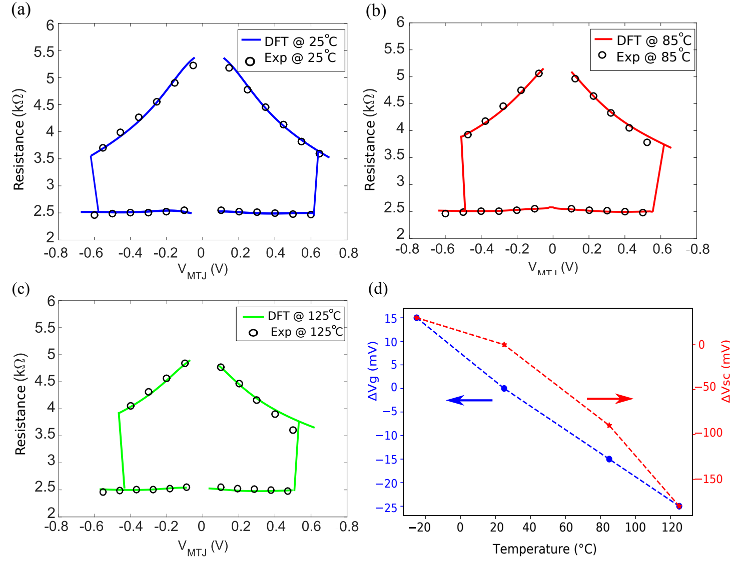


Figure 4.9: Temperature dependency behavior of the DFT circuit resistance. The RV loop is shown for (a) 25° C, (b) 85° C, and (c) 125° C (b) RV bias compensation voltage based on post-layout simulation.

4.4.1 DC Resistance Voltage Behavior

Here the DFT circuit characteristics is fitted to the experimental data published by Lin *et al.* [29] The schematic for the R-V bias circuit is provided on Fig. 4.11(a), along with the sizing and voltage information in Fig. 4.11(b). We have used four PMOS (P_{BL0}, P_{BL1}, P_{SL0}, P_{SL1}) and one NMOS (N_{SW}) transistor for the R–V circuit. Here W/L of P_{BL1} is sized twice as W/L of P_{BL0} to obtain the R_{AP} ($V_{DFT} = 0.7$ V) at half the resistance as R_{AP} peak for the same applied gate voltage. The access NMOS transistor (N1) is included with the DFT cell during matching process to account for its source degeneration effect of the NMOS. Fig. 4.12(c) shows the measurement results of bit-cell current vs. V_{BL}-V_{SL}, where the arrow indicates the direction of I-V loop. Here the voltages V_{DD_P} and V_{SS_AP} are obtained based on previous simulations, and then additional offset voltages (V_{DD_P} and V_{SS_AP}) were globally applied to match the resistance behavior to the target MTJ characteristics shown in the work [29]. The post-layout simulations are done with the internal node signals (V_{INT}) of the bit-cell being captured and post-processed to extract the voltage (V_{DFT}) and resistance (R_{DFT}) across the DFT R-V circuit, as shown in Fig. 4.11(d).

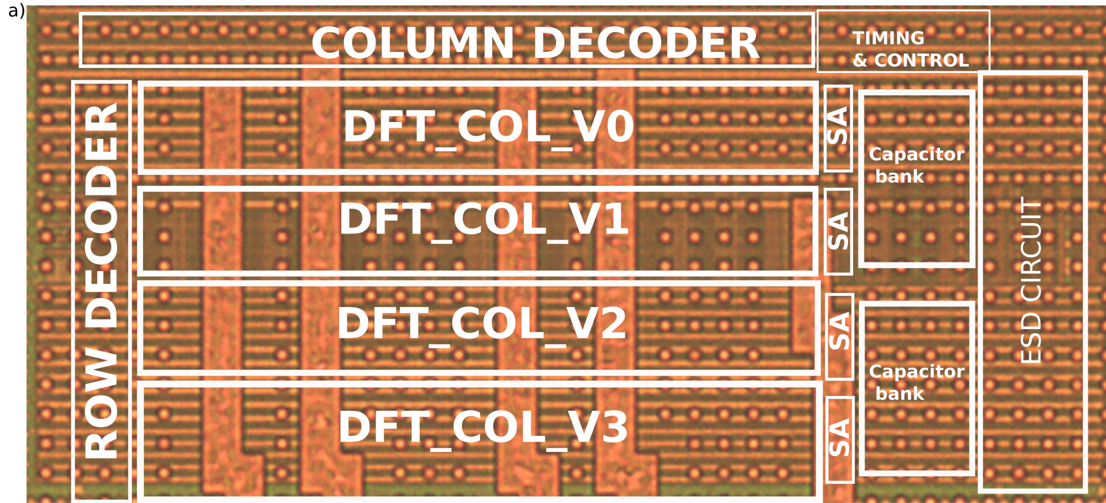
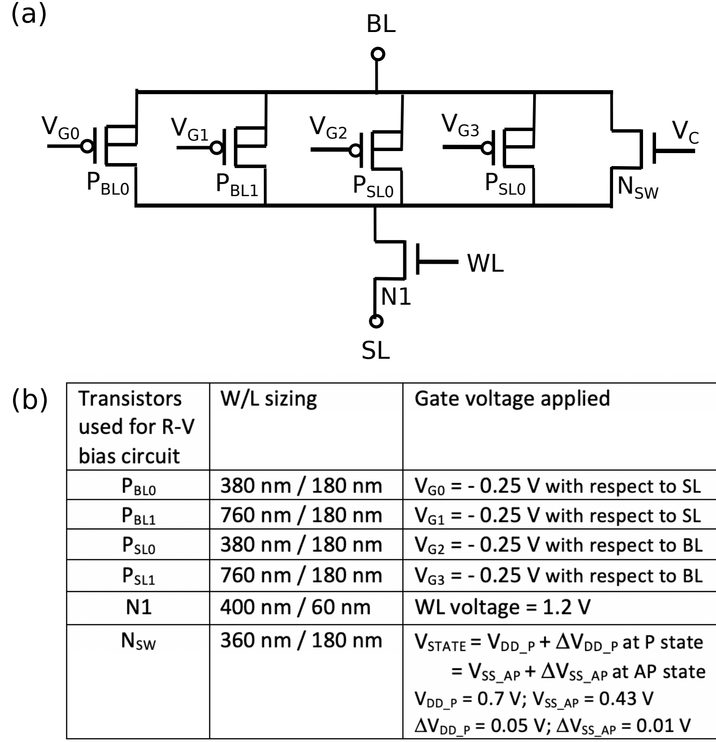


Figure 4.10: Die micrograph of the test chip with peripheral circuitry in 65-nm CMOS technology.

4.4.2 Switching Characteristics

The array architecture consists of the DFT cells along with a capacitor bank shared among the column as shown in Fig. 4.13. The write operation on the column can be illustrated using the waveforms as shown in Fig. 4.14. Initially the capacitors are precharged to V_{DD}. Then, PULSE signal activates the DFT cell, connecting the latch Q node to the capacitor bank via the PMOS. The write current (I_{CELL}) and the voltage at the latch Q node (V_Q) is shown in the waveform. Here the intrinsic switching delay (t_{SW}) is measured from the rising edge of the PULSE to the time that V_Q falls abruptly to GND. The switching time is inversely proportional to BL and WL voltages applied to the NMOS pair (N1 and N2). However, applying a positive SL voltage degrades the pull-down action of Q node. The capacitor output (SOUT) is buffered and taken off-chip for monitoring the switching operation.

MTJ device operates in two major switching regimes depending on the applied pulse width duration, which are the thermal activated (for long pulse widths) and precessional switching (shorter pulse width). In the thermal activated switching, the magnetization is independent of the initial conditions and is determined by the thermal agitation during the switching process. However, for precessional switching, the switching depends on the initial thermal distribution [50]. Switching current increases linearly in the thermal regime and increases exponentially in the precessional switching regime [50]. The switching interface between



Transistors used for R-V bias circuit	W/L sizing	Gate voltage applied
P_{BL0}	380 nm / 180 nm	V_{G0} = - 0.25 V with respect to SL
P_{BL1}	760 nm / 180 nm	V_{G1} = - 0.25 V with respect to SL
P_{SL0}	380 nm / 180 nm	V_{G2} = - 0.25 V with respect to BL
P_{SL1}	760 nm / 180 nm	V_{G3} = - 0.25 V with respect to BL
N1	400 nm / 60 nm	WL voltage = 1.2 V
N_{SW}	360 nm / 180 nm	$V_{STATE} = V_{DD_P} + \Delta V_{DD_P}$ at P state $= V_{SS_AP} + \Delta V_{SS_AP}$ at AP state V_{DD_P} = 0.7 V; V_{SS_AP} = 0.43 V ΔV_{DD_P} = 0.05 V; ΔV_{SS_AP} = 0.01 V

Figure 4.11: Schematic design for the R-V bias circuit used for DFT cell. (b) Sizing and voltage for the transistors used in the design.

the regimes depends on the anisotropy field (H_K). Here we capture the MTJ switching behavior and the interface between the precessional and thermal regime using latches. The dynamic noise margin characteristics present in the latch circuit can assimilate the MTJ switching mechanism. The latch noise margin increases exponentially with decrease in cell access times [68] [69]. The switching voltage depends on the latch transistor sizing and the capacitance at the input nodes Q and Q_B. The switching voltage at the thermal regime based on the pulse width (t_{PW}) can be computed from the thermal switching model [50].

Tuning the capacitance at the Q and Q_B node allows for modifying the switching characteristics due to variation in the anisotropy field (H_K) [38]. Although large external tunable capacitors can be used to achieve larger (> 1 μs) switching time constants, it is avoided by using a thermal switching model to compute the switching voltage. We chose to use internal on-chip capacitor banks to achieve high-speed operation and flexibility to tune the

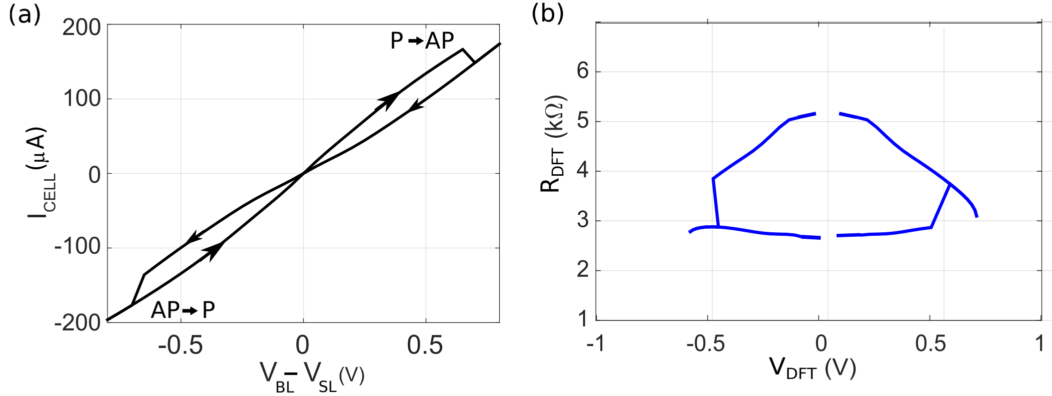


Figure 4.12: (a) Bit-cell current (I_{CELL}) vs. voltage (V_{BL}–V_{SL}) loop from measurement. (b) DFT vs. V_{DFT} characteristics of the DFT circuit from measurement and post-layout simulation.

precession/thermal interface region of the DFT cell for the column. The computed H_K is used to generate the 16–bit digital value that sets the capacitance. The DFT circuit switching characteristics along with the experimental data for P→AP and AP→P switching are shown in Fig. 4.15(a). The DFT circuit switching characteristics closely fit the MTJ experimental data [29] for write pulse widths from 10 ns to 100 ns with an error of only ±4% for $V_{SW.}$

4.4.3 Retention Characteristics

Here we are incorporating the stochastic switching behavior of MTJ devices in the testing environment. For an applied DC V_{SC}, the conventional latch inherently exhibits a sharp switching behavior depending on the applied V_{SC} noise voltage. The switching characteristics of the latch is made stochastic by increasing the noise (σ) in V_{SC} with a Gaussian normal distribution. The noise is correlated to the thermal stability factor (Δ) of the MTJ device,

$$\Delta = \frac{E_B}{k_B T} \tag{4.8}$$

where E_B is the energy barrier for magnetic switching in $k_B T$. The Δ is computed from device dimensions and magnetic material properties as shown in Fig. 4.3. It is observed

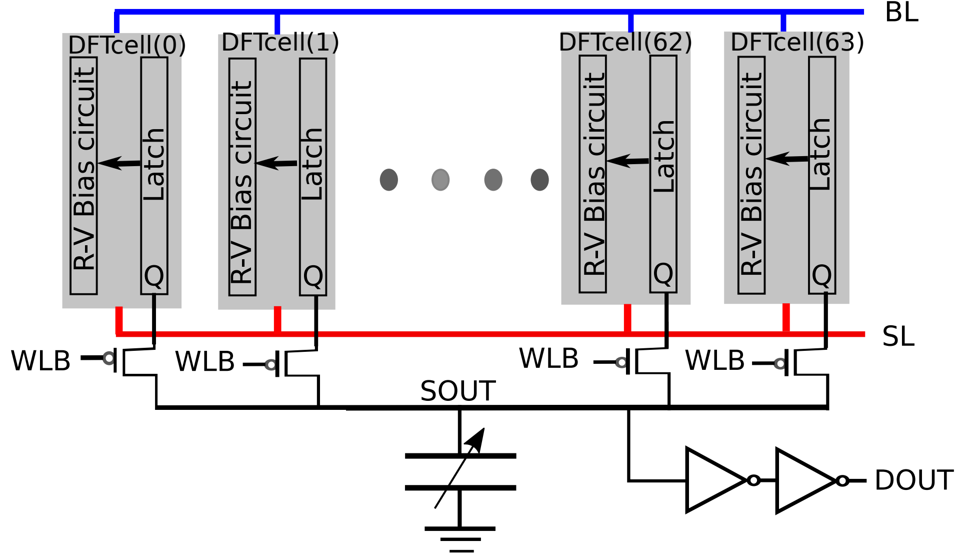


Figure 4.13: Block diagram of column with the shared tunable capacitor bank circuitry.

that the latch switching becomes more probabilistic for higher magnitudes of $V_{SC,\sigma}$ as shown in Fig. 4.15(b). We are considering a thermal-assisted STT switching operation with a uniform MTJ switching model approximation. The parametric generation algorithm can compute I_C/I_{C0} and switching probability (P_{SW}) using the equations below,

$$I_c = I_{c0}\left(1 - \frac{1}{\Delta} ln \frac{t_{\text{pw}}}{\tau_0}\right) \tag{4.9}$$

$$P_{\text{SW}}\left(I_c\right) = 1 - exp\left[-\frac{t_{\text{pw}}}{\tau_0} \exp\left\{-\Delta(1 - I_c/I_{c0})\right\}\right] \tag{4.10}$$

where τ_0 and t_{pw} denote the inverse of the MTJ attempt frequency and the applied pulse width, respectively. The mean DC component of V_{SC} is proportional to I_C/I_{C0} and R_{DFT}. Alternatively, an empirical approach can also be utilized to generate the sub-100 nm (< 40 nm) MTJ device parameters based on a non-uniform switching mechanism. Here a look-up table approach may be used to generate the critical switching current, and switching probability data for testing based on the MTJ wafer measurements [5].

The noise standard deviation (σ) is determined based on the $1/\Delta$ slope needed and the I_{C0} to generate the probabilistic switching behavior. Gaussian noise source with memory

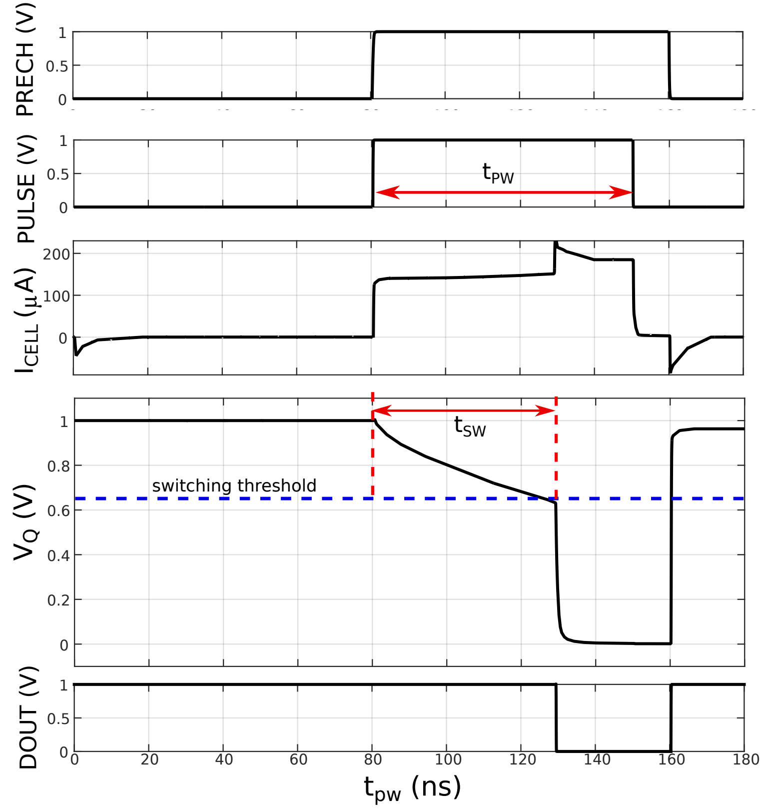


Figure 4.14: Post-layout simulation waveforms for the DFT cell switching operation.

depth of 32.4 million samples are used to generate the control signal samples $V_{SC,}$, and each sample is 14-bit wide allowing for 2^{14} unique bias voltage values. This enables to generate a larger set of unique bias voltage samples that replicate bit error rate as low as 10^{-6}. We applied a pulse width of $t_{pw} = 20$ ns and the ratio of the read time to total cycle time $t_{read}/t_{cycle} = 0.0625$. The sense amplifier output (SAOUT) and the output from the sensing amplifier latch (DOUT) are compared to detect a read disturb condition. Fig. 4.16 shows the read disturb bit error rate of the DFT cell, measured for different MTJ thermal stability factors at $t_{pw} = 20$ ns, where the switching voltage is normalized. The distribution curves exhibit a $-1/\Delta$ dependance as a function of switching voltage [5]. The markers represent the measurement from the chip and the dashed lines correspond to simulation results for comparison. The measured data fits closely to the simulation results for the switching equation.

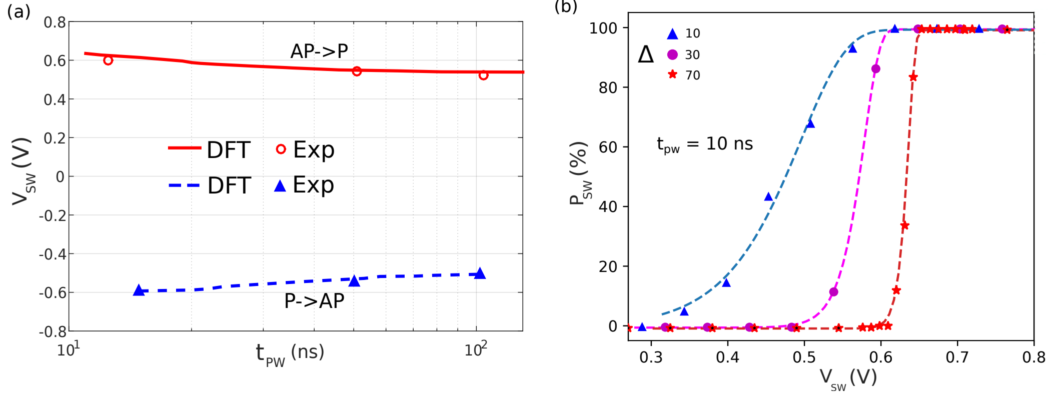


Figure 4.15: (a) Switching voltage vs. write pulse width for AP→P and P→AP switching. (b) DFT cell switching probability measured for different thermal stability factors (Δ). The markers and the lines represent the chip measured data and model simulation results, respectively.

4.4.4 Transistor Area Overhead

For the 256×256 bit-cell subarray, DFT structure is placed at the edge of each column with a tunable capacitor bank shared among all DFT cells in the entire subarray. The DFT transistors are oversized (2× times the MTJ access NMOS) to lower the variability. We estimated the area overhead to be 5-10 % of the entire design. For area optimization, analog multiplexers can be used to share the DFT cell among 4–8 columns to reduce the area overhead to 2–4 %. MIM capacitors can be used to place the capacitor bank on top of the CMOS circuitry to further reduce area overhead.

For the current DFT cell operation, we have chosen to replicate a commercial implementation of the MTJ device based on the work by Lin et al, that provided complete experimental data for the resistance and switching characteristics [29]. However, the DFT cell can be redesigned for lower R_P values that will suitable for replicating other MTJ stack devices. In such cases, we would anticipate an increase in the transistor area due to the higher W/L sizing requirement for meeting the lower resistance requirement.

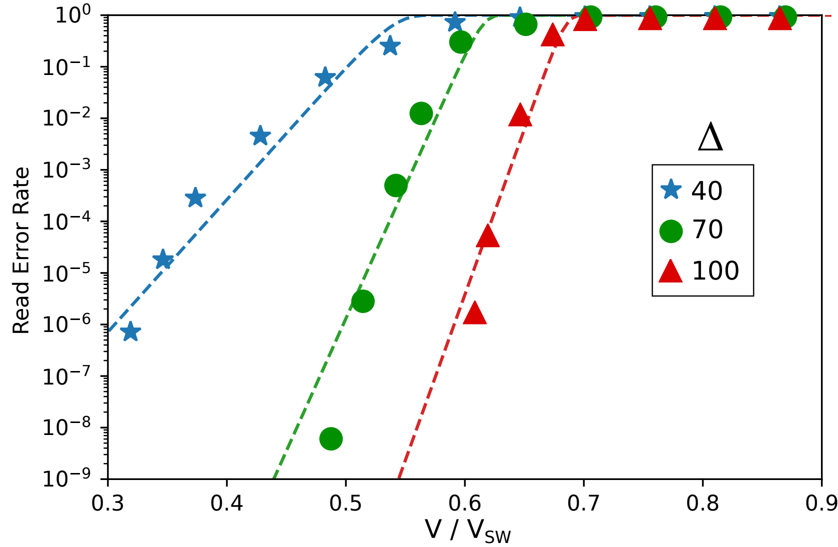


Figure 4.16: Read error rate, showing the read-disturb mechanism exhibited by the DFT cell for different thermal stability factors (Δ). The markers and the lines represent the chip measured data and model simulation results, respectively.

4.5 Application: Yield Characterization and process optimization

MTJ parameter variations can have a detrimental effect on the yield of the wafer. The key to achieving better STT-MRAM wafer yield is to ensure spatial uniformity of MTJ device parameters across the wafer. Thus, quantifying the maximum spatial non-uniformity of the parameter that can be tolerated in the wafer, is crucial. Process and device parameters of the MTJ device can be tuned to optimize the wafer yield. The choice of the parameter needed for optimization depends on the sensitivity of the yield to the parameter and the flexibility with which the parameter can be modified in the fabrication process. Thus, the testing process would require an exploration of the contribution of the individual parameter to the wafer yield. This section introduces the yield characterization and optimization flow to quantify the impact of parameter variation on the wafer yield and verify it within the wafer.

The optimization process is implemented in three steps. Firstly, the CMOS-based DFT circuit is matched to the target MTJ behaviors on the wafer. This allows for testing and

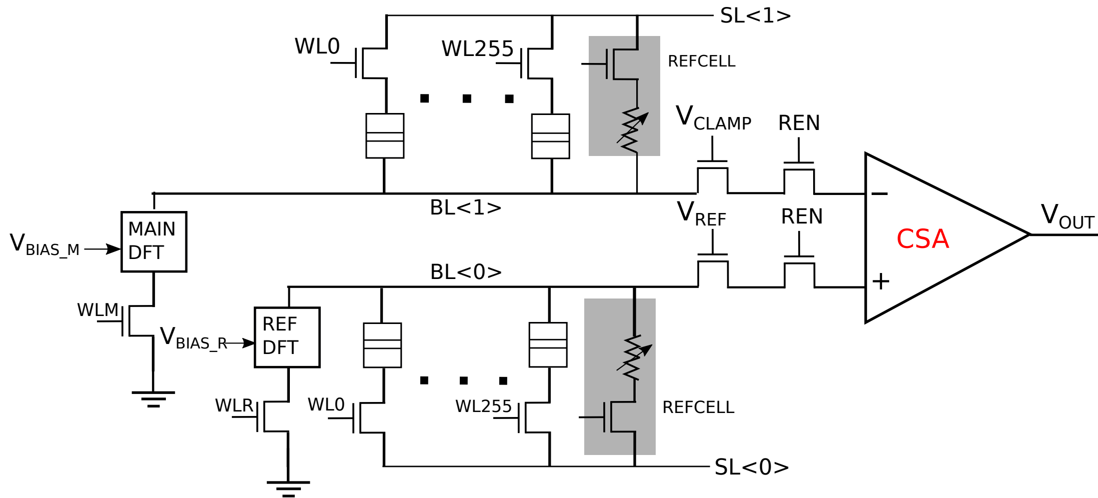


Figure 4.17: Read access path used for yield characterization simulation of the STT-MRAM column.

pre-qualification of the CMOS circuitry in the wafer before and after MTJ deposition, helping us identify and estimate the known good die statistics in the wafer. Secondly, we exploit the ability to tune the parameters of the DFT circuit. The DFT circuit can be tuned to reproduce the MTJ characteristics in the die as a function of its spatial location on the wafer, which allows us to replicate the effect of the individual parameter variations at different locations of the wafer. This helps us identify the sensitivity of the parameter to the yield. Finally, we estimate corrective measures that can be carried out to meet the yield requirement. The yield characteristics are analyzed directly on the wafer using the read access path present in each column of the array. This pre-verification step has the potential to reduce the overall testing and packaging costs. Additionally, the outcome of the proposed testing scheme can provide feedback for customizing the MTJ deposition quality and film parameters needed to optimize wafer yield and reduce its turn-around time. The main contributions of this chapter are as follows.

Figure 4.17 shows the regular STT-MRAM column pair with the DFT circuits inserted at the edge of the column. During testing, the bit-cells are disabled and the DFT circuit is connected directly to the bit-lines. Separate word-line control signals (WLR, WLM) are generated to activate these DFT cells, where one DFT cell acts as the reference and the other as the main cell replicating the target MTJ characteristics. A digital sense amplifier-based resistance (DSR) measurement technique is used to map distribution at P and AP states as shown in Fig. 4.19(b).

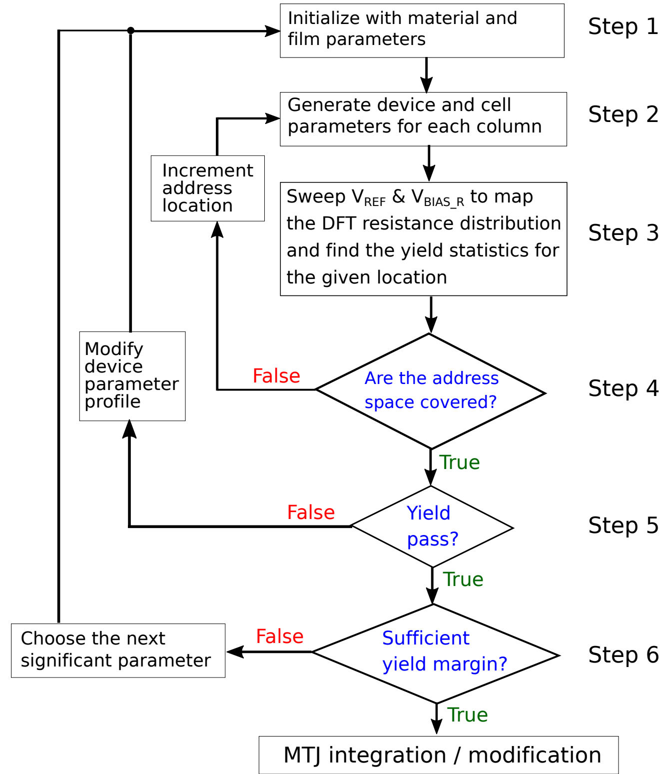


Figure 4.18: Yield characterization and optimization process flow chart.

The main DFT cell is biased to replicate MTJ variation in each test cycle by biasing and the reference bit-line is swept to read the sense amplifier output. This can be done by changing the V_{REF} or V_{BIAS_R} to sweep the current in the reference bit-line (BL<0>). Here we have chosen to use V_{REF} for reference current sweeping. The cell resistances variations are recorded as sense amplifier reference voltage distribution, which are then analyzed to obtain the read margin. The final reference voltage is generally placed at the center of the margin to obtain the equal read yield for both P and AP conditions. The flow for the yield optimization process is illustrated in Fig. 4.18 and the details are as follows.

Step 1: Obtain the film parameters with spatial non-uniformity considered. After matching, the parameter with the most probable impact on the yield can be chosen first for optimization.

Step 2: Generate the device parameters for each DFT main cell at different spatial locations

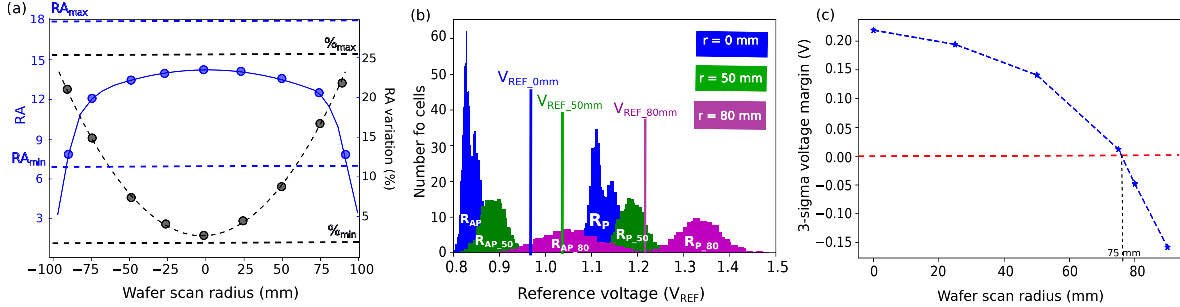


Figure 4.19: RA product vs. radial distance from the center of wafer. The markers represent DFT cell output values. (b) Resistance distribution profile for AP and P states at 0 mm, 50 mm and 80 mm radial distance from the center of the wafer based on post layout simulation results. (c) 3-sigma read margin voltage based on sense amplifier outputs across the wafer.

and match it to the MTJ characteristics observed. This is done based on the parameter generation algorithm explained previously.

Step 3: Sweep the reference current flowing though the bit-line, and record the sense amplifier output for the given state. Proceed to next step after obtaining sufficient samples for the cell resistance distribution for the given column.

Step 4: Check if all the address space in the wafer is covered; otherwise increment to the next address location and repeat steps 2 and 3 to retrieve the parameter set for the new location.

Step 5: Verify if the DSR-based yield measurements pass the requirements for the given parameter variation profile. Otherwise, modify the parameter profile and verify in wafer by repeating steps 1–4.

Step 6: Proceed to MTJ integration if the CMOS wafer meets the yield margin requirements. Otherwise, choose the next most significant parameter for the yield margin optimization by repeating steps 1–5.

In general, the spatial parameter variations follow a radial profile across the wafer. Thus, we consider statistical samples of bit-cells for dies along the vertical and horizontal axis in the center of the wafer to make sure we have chosen test cases with significant spatial variations. Although, we can consider several MTJ parameters (RA, TMR, MTJ device area, I_{C0}, H_K, Δ) and their combination, here a case study for the spatial variation in resistance-area product of the MTJ device is considered. During the analysis, the spatial variation of all

the other process parameters is held constant for simplicity. However, in reality the impact of the RA on other film parameters needs to be considered [70]. Fig. 4.19(a) shows the RA profile variation for the testing scenario. The spatial distribution profile considers both the parameter mean and the standard deviation similar to the profile shown in the typical line scan measurements across a 200-mm wafer [64]. For illustration, we assume a symmetric standard deviation profile that is parabolic with a minimum variation at the center of the wafer as shown in Fig. 4.19(a). The dashed lines indicate the tuning range upper and lower boundaries within which the RA value can be replicated using the DFT cell.

In order to understand the impact of RA variation on the wafer, we observe the sense amplifier reference voltage distribution, which is the number of cells that is in AP and P state in a die location vs. sense amplifier reference voltage for different radial locations on the wafer. Fig 4.19(b) shows the AP and P voltage distribution at the center ($r = 0$), middle ($r = 50$ mm) and near the edge ($r = 80$ mm) and it can be seen that the AP and P distributions are crossing the reference voltage at $r = 80$ mm, resulting in loss of read margin. This could be due to the film parameter variation that result in spatial variation in the RA product. Fig. 4.19(c) shows the read margin based on 1,000 Monte Carlo samples for each radial location (r) on the wafer. The estimate for the 3σ read margins also clearly shows the margin available between the AP and P states. Instead of a single common V_{REF} bias applied to all the dies, varying reference voltages ($V_{REF_0\ \mathrm{mm}}$ to $V_{REF_90\ \mathrm{mm}}$) can be applied to improve the read margin for individual die locations. However, it can be seen that near to the edge of wafer ($r > 75$mm) the yield begins to deteriorate, even though the reference voltage is placed at the middle of the distribution, due to the overlapping of the R_P and R_{AP} distributions. Thus, the RA profile here needs to be modified by changing the parameter profile. Some of the steps that can be taken are to maintain the RA for the R_P state above 9 $\Omega/\ \mu\mathrm{m}^2$ or decrease the variation in RA value. This feeds back to modify the process steps and improve the quality of the MTJ film deposited.

4.6 Summary

The post-layout simulation and measurement results show that the CMOS-based DFT circuit behavior matches the target MTJ experimental data within an error of 4%. The CMOS-based DFT circuit could introduce non-idealities due to the fabrication process variations; however, we provided methods to minimize the impact of these non-idealities in the testing process. The DFT scheme can provide quantitative feedback based on in-die measurement, enabling fabrication process optimization through iterative estimation and verification of the calibrated parameters.

Chapter 5

DFT for Long-Term Reliability

5.1 Introduction

Several of the MTJ defects are parametric in nature to begin with, and deteriorate over time. In this chapter, we integrate the design-for-testability (DFT) scheme within the array to monitor the electrical parameter deviations occurring due to the defect formation over time. The scheme utilizes the read sense path to compare the bit-cell electrical parameters against known DFT characteristics to identify faulty bit-cell behavior. Built-in-self-test (BIST) methodology is utilized to trigger the onset of the fault once the device parameter crosses a threshold value. We perform a case study to demonstrate the operation, and evaluate the accuracy of detection with the proposed scheme.

Many recent studies have focussed on functional testing and characterization [38] for STT-MRAM arrays [32]. Design-for-testability (DFT) structures have been proposed to identify various recoverable failure mechanisms, such as read-disturb condition [39]. Memory BIST structures have been proposed to efficiently evaluate the thermal stability and retention behavior of STT-MRAM bit-cells [33]. Most of these works done previously were used to detect failures that are recoverable or failures that occur due to the complete damage of the stack, which can be identified by regular bit-cell read/write tests. However, often subtle failures grow over time and a substantial fraction of their population can escape the initial test and screening process. These defects need to be identified before they result in permanent faults. One such defect forming mechanism is the pin-hole defect formation. There have only been a few studies on monitoring defect-oriented failures in MTJ that occur over time, such as pin-hole defect formation [43][44].

Periodic monitoring of the MTJ stack quality by observing its device parameter characteristics can provide information on the aging defect and its progression. A way to identify these electrical symptoms is to monitor the deviations in electrical and magnetic characteristics of STT-MRAM bit-cells. However, acquiring and observing large samples of data from bit-cells in a die can be quite cumbersome. One solution for identifying the deviation within die is to compare the MTJ electrical characteristics with a known reference device that is not affected by the same defect-causing mechanisms. Given that metal-oxide-semiconductor field-effect transistor (MOSFET) variation and aging are well understood [40], appropriately designed MOSFET-based DFT circuits can replicate the intended reference behavior of the MTJ device with minimum variations.

A fault is a representation of the defect in the device behavior, may result in parameter deviations or, in the worst case, failure of the bit-cell. Several types of defects could be formed during the MTJ deposition, patterning and bit-cell integration process. Research has been done that surveys different defects formed during the fabrication and integration processes. The work by Zhao et al. [34][37] summarizes the defects and issues faced during different stages of the bit-cell integration process. A comprehensive analysis of bit-cell level defects and faults are done by Chintaluri et al. [42]. Here the work explores resistive and capacitive faults between bit-cells. Additionally, the bit-cell integration can result in defect formation at each junction creating an open circuit. Finally, the work by Vatajelu et al. [71] provides a bit-cell defect model that illustrates the defects formed during deposition and integration and result in an open circuit in the bit-cell.

The faults in the bit-cell can be classified into defects related to CMOS and MTJ device. CMOS defects have been extensively studied for mainstream SRAM and DRAM memory applications. As the process node shrinks, some of the challenges include threshold voltage variation due to random dopant fluctuation and W/L variation [40]. Additionally, the aging of the CMOS device is also becoming a concern. Modern write schemes for STT-MRAM [72] employing boosted write voltage can put stress on both the access transistor and the MTJ device resulting in dielectric breakdown. For transistors, the gate dielectric breakdown can lead to bit-cell stuck-at faulty behavior and parametric failures. The defects in the MTJ are often irrecoverable and result in permanent faults. They occur due to extreme defect formation, resulting in the functional failure of the MTJ device. Some of the permanent faults [42] [73] are as follows:

1. Open faults: Occur rarely. Open faults occurring due to contact fails between the MTJ stack and the top electrode have been observed. This is due to the polymer remains after an unstable etching process. However, they can be eliminated by optimizing the etching process [27].

2. Short faults: Occurs frequently. The most common short faults are due to dielectric breakdown, pin-hole formation, and side-wall re-deposition. [43][44]. In this case, the short circuit path is formed by the CoFeB free layer material through the dielectric barrier.

3. Stuck at P or AP faults: Occurs moderately. The MTJ does not switch states after applying the required write current through the MTJ device. These faults occur due to the failure of the dielectric barrier or synthetic anti-ferromagnetic (SAF) layer. [42]

Defect injection methodology and fault models have been constructed previously based on layout characteristics to inject and test different ideal open and short circuits [74], but they focus on functional test coverage based on ideal defects. The work doesnt account for the pin-hole degradation mechanisms [43] which is an aging-related defect that needs parametric defect oriented testing. These ideal short and open permanent faults appear as linear resistor values. However, most of the defect formation exhibit deviation in the electrical characteristics intermediate between ideal short and open resistance values before the permanent fault. The resistance characteristics exhibited during this phase are non-linear, which is similar to a degraded MTJ device. One such example is the resistance and the TMR behavior exhibited during a pin-hole growth [44]. These deviations in their electrical parameters can be observed through the read sense path within a column of the bit-cell array. Some of the parameter deviation that can be observed are TMR degradation, R_P variation, slope (S_V), switching voltage (V_{SW}), and switching time (t_{SW}). In order to understand the defect progression over time, we need to monitor the electrical parameters of the bit-cell over time.

5.2 Proposed BIST Scheme and Test Methodolgy

5.2.1 DFT Circuit

We previously explored the design and implementation of the DFT circuit [41]. We designed it with tunable electrical characteristics (R_P, R_{AP}, TMR and switching voltage and switching time). However, the circuit could only replicate the MTJ behavior across a narrow parameter range. We extend the parameter tuning range of MTJ stack to replicate cases for weak open and short circuit failures by using 2 additional R-V blocks. Each DFT cell within the circuit (open, nominal, and short cell) represents three circuit fault scenarios

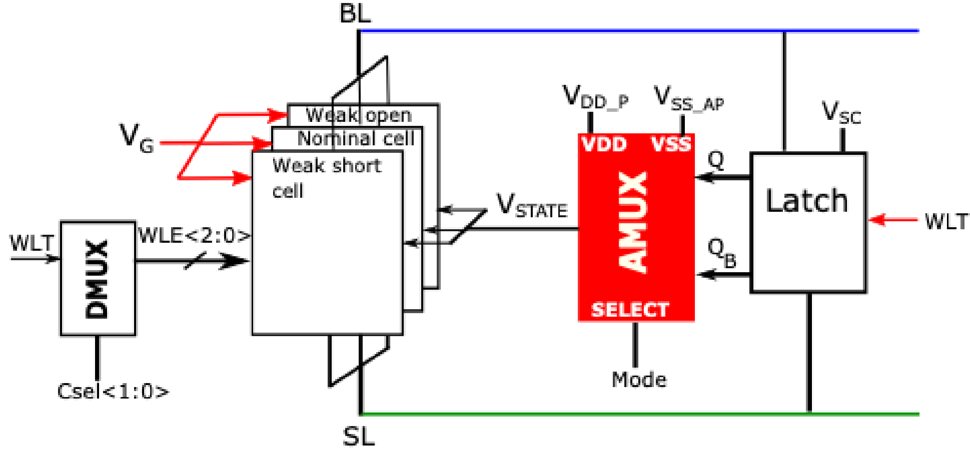


Figure 5.1: DFT cell used for testing and replication of fault characteristics.

of the MTJ device. Fig. 5.1 shows the DFT circuit composed of cells that replicate the resistance behavior in three different fault scenarios. The parameter tuning range for each R-V circuit is as shown in Table 5.1. The latch circuit is used to replicate the switching operation. The circuit consists of Analog (AMUX) and digital (DMUX) multiplexer circuits. The mode bit is used to define the direction of the applied BL, SL voltages. Based on the fault scenario, each cell is selectively activated using DMUX based on Csel input. The latch controls the R-V bias circuit to exhibit the resistance for a given state. Based on the latch output (Q and QB) and the bias voltage (VDD_P and VSS_{AP}) the AMUX output biases the N_{SW} NMOS in the DFT circuit to achieve the R_P and R_{AP} resistances. The switching operation is controlled by the V_{SC} voltage applied, which allows the tuning of the switching voltage from 0.4 V to VDD = 1.2 V. The DFT circuits are sized larger and activated only during the memory test phase for minimal electrical stress and aging effects. The switching time is varied by tuning the capacitance shared across the latches across the array. Capacitance banks of different capacitance values are implemented and each element is selectively activated to get the desired switching voltage for the applied write pulse width and voltage. A 5-bit tuning control is used to select from 32 different capacitance combinations. The DFT circuits are sized larger and activated only during the memory test phase for minimal electrical stress and aging effects.

Parameters	Minimum	Maximum	Comments
$R_{WEAKOPEN}$	7 kΩ	18 kΩ	R > 20 kΩ is beyond range; considered strong open.
$R_{NOMINAL}$	1.5 kΩ	8 kΩ	Typical design value R_P = 2.5 kΩ; R_{AP} = 5.5 kΩ
$R_{WEAKSHORT}$	400 Ω	1.7 kΩ	R < 400 Ω considered strong short for the MTJ
V_{SW}	0.3 V	0.9 V	Tunable with V_{SC} input of latch
t_{SW}	5 ns	150 ns	Tunable with 5 ns steps

Table 5.1: Parameter tuning range for the DFT cell.

5.2.2 BIST Structure

Modern STT-MRAM array consists of a large number of bit-cells (typically 256 kB) per sub-array. Individual manual testing is nearly impossible from the user end, thus monitoring the health of bit-cells and identifying the deviation requires the chip to run diagnostic tests in an automated manner. The BIST scheme is suitable for running automated tests during the test sequence and detecting faults in STT-MRAM bit-cell array. The system diagram for the BIST test structure used is shown in Fig. 5.2(a). The BIST is implemented at the die-level, where it uses the DFT array to provide input, and the detection circuit to get data from the device-under-test (DUT). Fig. 5.2(b) shows the DUT comprising of the MTJ bit-cell array, ref-cell, and the read sensing path (clamping circuitry and current sensing amplifier). The DFT cells are placed far and near to the read sense amplifier. Specifically, the near DFT cells are used for read-sensing path characterization with minimal effect from the interconnect resistance. The far DFT cells are used for open bit-line and switching tests, which considers the bit-line resistance. The resistive reference cell is used for creating reference current during read operation [27]. The DFT cell is enabled using the word-line test (WLT) input to the access transistor. The read enable transistors (REN) are turned on to connect the bit-line pair to the sensing path. A current sense amplifier (CSA) is then used to compare the bit-cell behavior with a known DFT cell to observe the parameter deviation. The architecture allows a means to calibrate and test the CMOS circuitry separate from the bit-cells.

Fig. 5.3 shows the functional diagram of the BIST engine. The test process and role of each module is described as follows. The test process is initiated by enabling the BIST start. The controller produces the internal MTJ bit-cell address as well as the DFT bias vector needed for the DFT array. It utilizes a finite state machine to generate the MTJ bit-cell

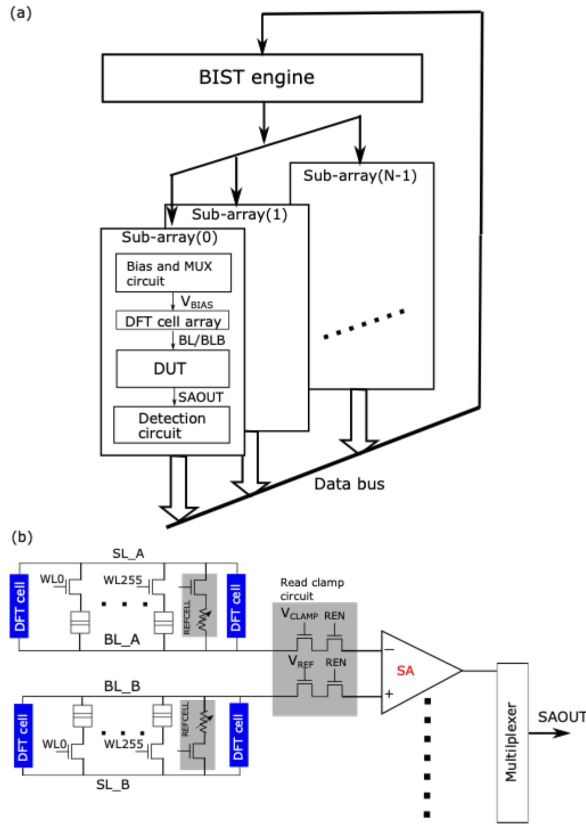


Figure 5.2: (a) System level diagram of the BIST scheme. (b) Device under test (DUT), is the conventional STT-MRAM sub-array consisting of bit-cells, read sensing column and multiplexers.

address and DFT bias vectors based on the testing schedule. The test controller coordinates all the operations of the sub-blocks within the array. The multiplexer (MUX) is used to select between the external address or internal address generation during testing. During the detection phase, the results compactor (RC) is used to compact the sense amplifier output (SAOUT) distribution and extract parameters. Once a fault is detected, RC stores the fault results, which includes the test address and the SA address in the off-chip external memory. External memory is utilized off-chip to store the fault and calibration data and keep a log of the test vector information. Finally, the RC signals the test controller to perform additional tests to identify the faulty bit-cell characteristics in the next boot-up sequence. Fig. 5.4 shows the bias network and the analog multiplexer (Analog MUX)

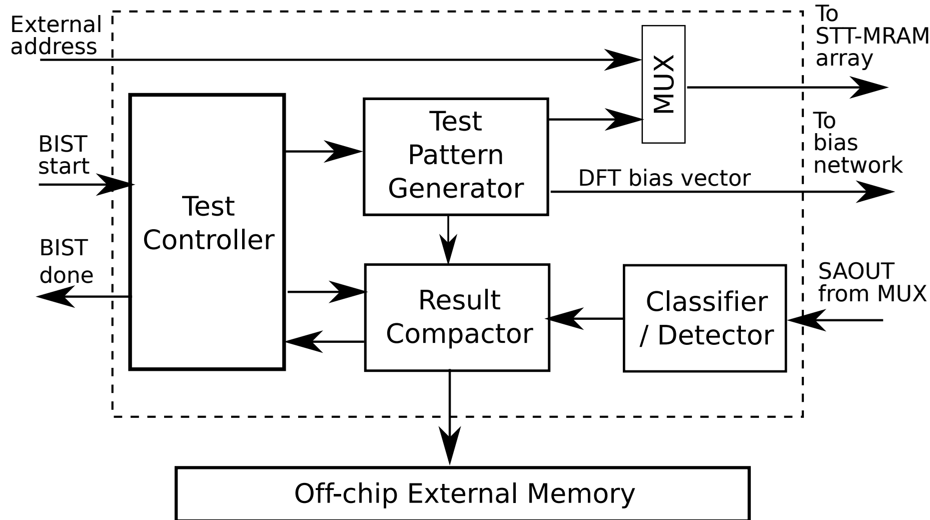


Figure 5.3: BIST engine diagram within the sub array.

circuitry. The MUX array selects the bias voltage and applies it as a gate bias to the DFT cell. Both bias network and detector circuits are implemented in each sub-array level, enabling simultaneous testing of sub-arrays within the die.

5.2.3 DFT Fault Classification and Identification

The MTJ electrical parameter can deviate to a range of values due to fabrication process variation and defect formation. Classification is done to prioritize the testing process and identify faulty cells that exhibit parameter deviation beyond the acceptable nominal range. Based on the DFT cell resistance being compared, the faults detected can be classified into regions based on the parameter impact on the operation of the STT-MRAM bit-cells. Fig. 5.5 shows the generalized classification of MTJ parameter ranges. The nominal range (green) corresponds to the safe parameter value within which the MTJ operates reliably. Beyond the nominal range, the weak short or weak open region (orange) is considered as the onset of the fault. Finally, the strong short or strong open region (red) is the extreme case where the MTJ incurs a functional failure. For the resistance comparison tests, the bit-cell is compared with the DFT circuit activated in the reference-line. Alternatively, for switching current and switching time comparisons, the DFT circuit in the same column as the DUT bit-cell is utilized.

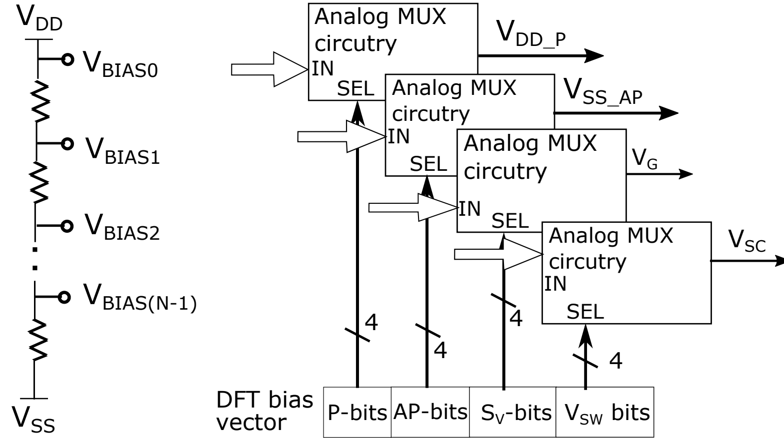


Figure 5.4: Bias voltage generation and analog multiplexer array for selecting DFT control voltages.

A self-triggered approach is followed to detect the onset of the faulty bit-cell behavior as shown in Fig. 5.6. We start the process by initializing the DUT MTJ to the required state (AP or P). The DFT cells are biased to exhibit a fixed upper (P_{NOMH}) and lower bound (P_{NOML}) based on its nominal parameter range of operation. The system triggers a fault and starts monitoring only when the MTJ bit-cell characteristics cross the specified bound. Once the nominal range bound is crossed, the BIST engine records the DUT bit-cell information for more tests to identify the deviation in the parameter. The faulty bit-cell resistance is compared multiple times with different known characteristics replicated by the DFT cell to see the best match. A binary-sort approach is utilized to compare and efficiently arrive at the closest matching known parameter deviation. When the DFT resistance matches the MTJ resistance, the sense amplifier input currents are equalized and the output exhibits a 50% probability of generating a logic '1' or '0' output. We can identify the probability distribution by taking multiple samples of the read sensing output data (SAOUT). The V_{BIAS} setting for the matching condition is then retrieved to estimate the resistance or switching voltage. The identification process is more computationally complex than the previous threshold detection process explained. Thus it is performed only on faulty bit-cells that have crossed the nominal range.

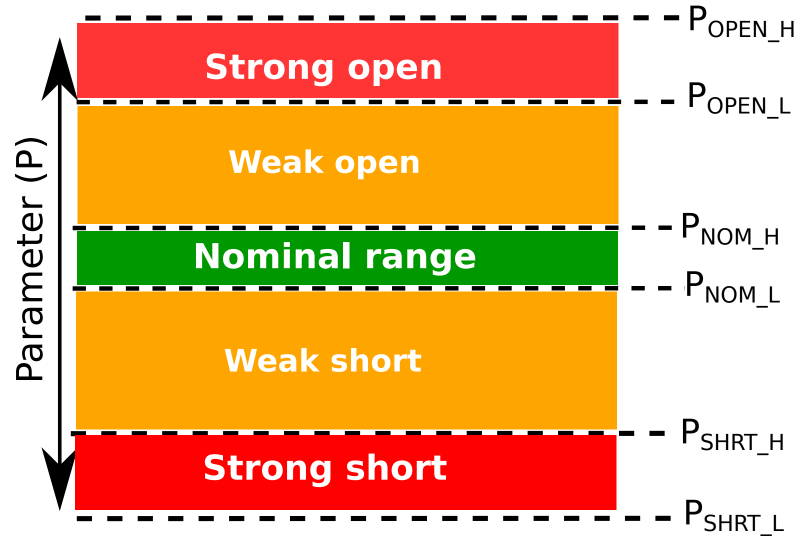


Figure 5.5: Generalized parameter range of the MTJ device with defect.

5.2.4 Fault Analysis, Scheduling and Complexity

Yield learning is done in post-processing, where the sense amplifier data collected over time from memory are analyzed to observe the MTJ fault progress behavior. Fault tests used for different parameter faults in their order of severity and complexity are described as follows.

Open bit-line test Degradation of the access transistor in the bit-cell results in a short circuit path between the gate and the drain/source regions. This fault will result in a decrease of the bit-cell resistance in the OFF-state condition, compromising the operation of all the bit-cells connected in the bit-line. Open bit-line tests are performed with the bit-cells not accessed, and the read voltage is applied at the bit-line. In the alternate column, the DFT cell is enabled to the weak-open resistance state. The bit-line to ground resistance is then evaluated to make sure it is above the given threshold high resistance. This test is performed once per bit-line, prior to diagnosing the bit-cells individually.

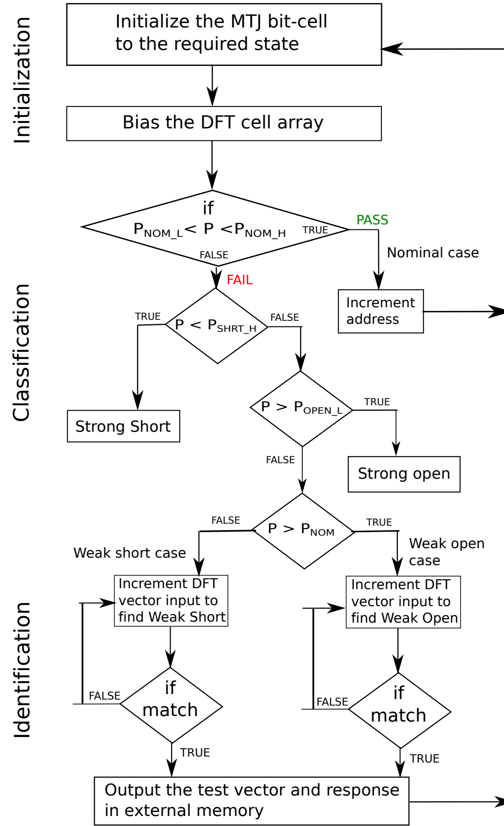


Figure 5.6: Flow chart for classification and identification of faults using the DFT scheme.

Resistance Characteristics We initialize the STT-MRAM bit-cells to the given state and enable the DFT cells. The DFT cells are biased to replicate the fault conditions for the given state. We perform 2 iterations during resistive fault testing for both AP and P state at near-zero bit-line voltage bias applied. This test is used to identify the deviation in peak value of R_{AP} and R_P for the MTJ. Furthermore, the initialization procedure tests help us see the stuck at P or AP fault based on the resistance of the MTJ cell. The TMR can then be computed from the ratio of R_{AP} and R_P values.

The slope of the AP curve is an indication of the barrier quality of the MTJ stack. For slope (S_V) evaluation, we perform the test in 3 steps.
1. Initialize the bit cell to AP state and near-zero bit-line voltage to evaluate the peak

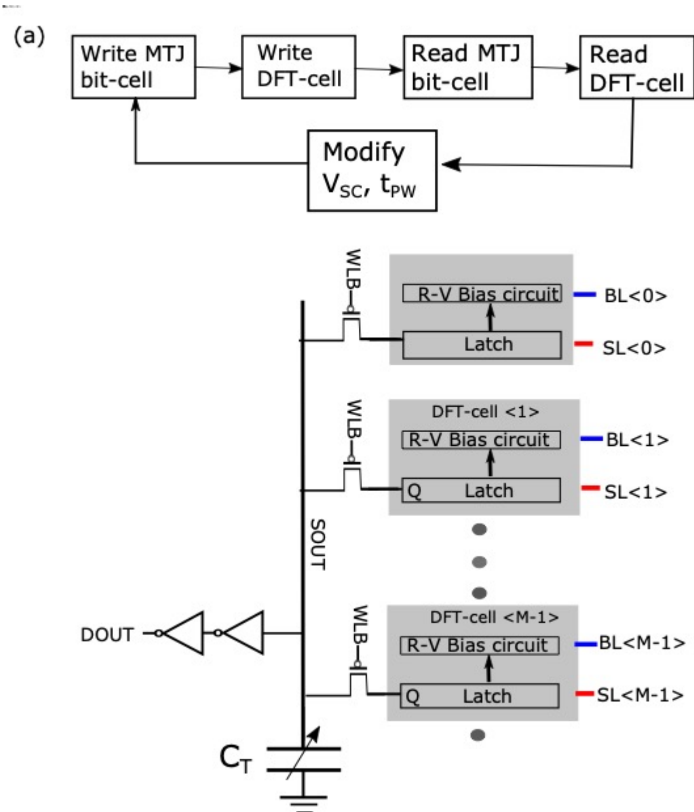


Figure 5.7: (a) Switching test cycle. (b) Block diagram of column with the shared tunable capacitor bank circuitry.

$R_{AP}(V_{BL} = 0)$ resistance.
2. Set the gate bias for the R-V circuit (V_{GSH} in this case) and sweep the V_{CLAMP} voltage to evaluate the SAOUT distribution.
3. Vary the gate bias until the slope of the DFT AP resistance curve matches the bit-cell AP curve. Step 1 and 2 have to be repeated to ensure that the MTJ has not flipped states during the V_{CLAMP} scans.

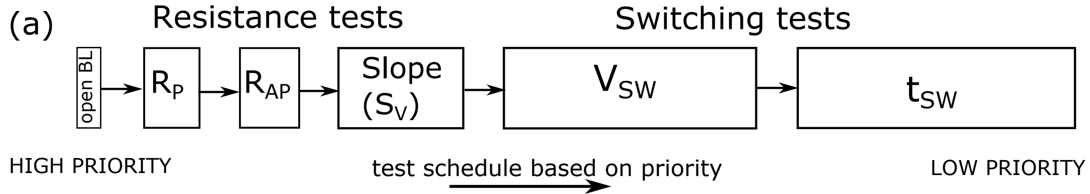
Switching Characteristics

During testing, the bit-cell and DFT-cell are written to AP and P states using the auxiliary write circuitry. This is similar to the regular write circuitry except that the control voltage

(V_{SC}) is tunable using the bias network. The switching test cycle is shown in Fig. 5.7(a). During the test scan, the write pulse width (t_{PW}) is varied for a given applied write voltage to see if the bit-cells are switching at the intended voltages and switching times. The same write voltage and pulse width are then applied to the DFT cell within the same column. The write tests are done sequentially so that the DFT cell can see the same source resistances as the bit-cell. Data stored in both cells are read back at the end of the cycle. The resistance states of both cells are sensed to see the success of the switching operation. We perform two different switching tests.
1. Switching voltage test: The applied switching voltage (V_{SC}) is varied for a fixed pulse-width word-line voltage. Auxiliary write circuitry with tunable write voltage is utilized to apply the $V_{BL} - V_{SL}$ voltage. The source line is grounded for the write operation.
2. Switching time test: Here, the applied switching time (t_{PW}) is varied for a given write voltage. Digital multiplexer-controlled delay lines are used to create tunable pulse width word-line voltage signals during testing.

Switching test requires multiple samples to be considered due to the pulse width dependent switching characteristics of the MTJ device [50]; The detection circuit captures 128 samples for a given switching voltage and pulse width setting. For threshold detection measurements, we look at the number of test samples that are within the bounds specified. A fault is triggered if the bound condition is not met. Once the samples are captured, the switching test parameters (t_{PW}, V_{SC}) are modified to perform the new tests. The write circuitry of the DFT scheme is shown in Fig. 5.7(b). It consists of the DFT circuits with the capacitor bank shared among the columns. The capacitor output (SOUT), which indicates the DFT cell state is buffered and taken off-chip for monitoring. The scheme allows for observing the read sensing path output (SAOUT) and the DFT cell state (DOUT) independently, which will be useful for distinguishing the sense amplifier error due to insufficient input margin, from a DFT cell switching states.

Testing complexity is a crucial factor in deciding the time and cost of the DFT-based BIST implementation. We prioritize the test type employed depending on the complexity of the test and the impact of the detected fault on the systems reliability. A scheduling scheme is utilized to evaluate the most critical functionality tests first. For instance, the open bit-line test is employed as the first pass/fail test to evaluate the resistance of the bit-line to the ground. Then, the resistance deviation tests are done, followed by the switching test. DFT test schedule based on priority is shown in Fig. 5.8(a). The width of the block for each test indicates the number of test cycles needed to perform it. The complexity of each test type, based on the number of test access cycles, is shown in Fig. 5.8(b). The detection scan process for a given test type is performed across all the bit-cells before moving to the next test type. Column 2 represents the number of test clock cycles needed



(b)

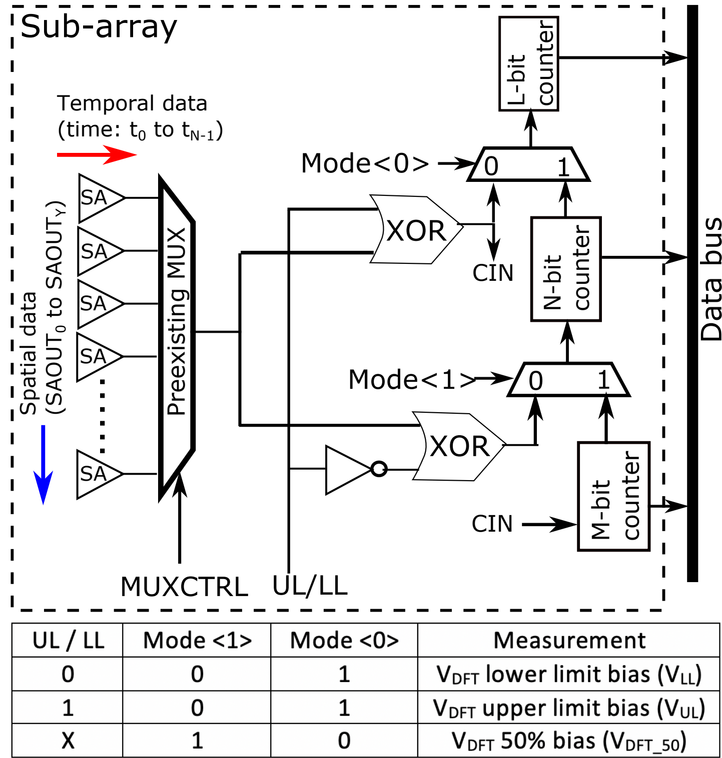
	Test type	*Detect* (*Cycles / bit-line*)	*Detection throughput* (*Cycles/die*)	*Identify* (*Cycles / bit-line*)
1	*Open bit-line*	1	1/L	log(M)
2	R_P / R_{AP}	2N	2N/L	2Nlog(M)
3	S_V *(AP)*	6N	6N/L	6Nlog(M)
4	*Switching test*	128N	128N/L	128Nlog(M)

Figure 5.8: (a)Test scheduling based on priority for different tests. (b) Test complexity for each test. Here, N and L correspond to the number of bit-cells in a bit-line column and number of read sense paths accessed simultaneously. M corresponds to the number of bits used to select the bias control voltage.

to scan across the cells in the bit-line, where N is the number of bit-cells per bit-line. The presence of a detection circuit in each sub-array allows simultaneous evaluation of bit-cells across multiple sub-arrays. Column 3 represents the test output data throughput, which depends on the number of sub-arrays (L) assuming 1 SA is accessed per sub-array. Column 4 represents the worst-case test cycles needed to identify the fault once it falls in the weak open or weak short case. The faulty bit-cell address is stored in the external memory so that it can be measured during each boot-up phase. This enables visibility on the parameter degradation trend and yield learning, which is done by post-processing the fault comparison results that are collected over time.

Fault Detection Circuit

A detector circuit is introduced to the interface between the sub-array and the BIST controller. The main function of the detector circuit is to provide storage of SAOUT data



UL / LL	Mode <1>	Mode <0>	Measurement
0	0	1	V_{DFT} lower limit bias (V_{LL})
1	0	1	V_{DFT} upper limit bias (V_{UL})
X	1	0	V_{DFT} 50% bias (V_{DFT_50})

Figure 5.9: Detection circuit within a sub-array. The table shows the different configuration modes for the circuit.

and average the samples obtained from single or several SA locations within the sub-array. The averaging is done temporally or spatially with respect to the time and location of the SAs being accessed respectively. The sample depth for each averaging can be adjusted by changing the MUXCTRL input during the testing process. Fig. 5.9 shows the detector logic used to evaluate the SAOUT samples. The circuit uses multiple counters to store the number of 0s and 1s generated by SAOUT. Some of the parameters extracted from the SAOUT distribution are,
1. DFT Upper bias voltage (V_{UL}): The minimum bias for 99.7% of the SAOUT samples to converge to 1.
2. DFT lower bias voltage (V_{LL}): the maximum bias for 0.03% of the SAOUT samples to converge to 0.
3. DFT 50% match voltage (V_M): the DFT bias voltage where the SAOUT generate an equal number of samples for 1 and 0.

The detection logic is configured in three different modes depending on the parameters extracted. We have used the 10-bit counter-based scheme, segmented as M = 5, N = 4 and L = 1. For high accuracy V_{UL} and V_{LL} measurement, we set mode = 01, which utilizes 9 bits for SAOUT majority counting and 1 bit for the fault counting. For V_M measurement, we use an equal number of bits (M=5, N+L=5) for both the samples. Mode 01 gives the overall distribution of SAOUT with a reasonable sample size (1024 samples). After evaluation, the BIST polls through each detector circuit output and retrieves the count data for analysis. The RC extracts V_{UL}, V_M, V_{LL} parameters from the SAOUT distribution based on the detection circuit mode. For instance, if the detection logic is set to identify 50% match bias, the RC will observe the counter value for each bias voltage setting and gives a match output 1 when the count values are equal. This indicates that the DFT cell resistance matches the MTJ bit-cell resistance for the given bias voltage. Here, the least significant bits (LSB) from the counter output may be truncated to obtain a faster convergence of the MATCH output signal. However, the bit truncation error should be maintained less than the match detection error imposed by the quantized bias voltage applied.

5.2.5 Read-Sensing Path Characterization

Detecting faults accurately in the production dies are challenging due to process-induced variations. The detection process is based on comparing the faulty characteristics of the bit-cell with the known DFT behavior using the existing read sense path. In an ideal scenario, all the read sense paths should measure the same resistance for a given injected fault resistance, irrespective of the location or time of SAOUT evaluation. However, the read sensing path can exhibit temporal and spatial variation in output that would influence the accuracy of the detection process. During measurement, we have observed both temporal and spatial variation in SAOUTs for the same injected fault resistance. The time-dependent variations can be due to the fluctuation in bias voltages applied to different nodes of the read path. This could occur due to coupling between the signal nodes to the bias lines or due to noise in the bias voltage source itself. Temporal variation can be minimized by averaging multiple data samples from the same SA. On the other hand, spatial variation can be due to the fabrication process and the IR drop incurred during bias voltage distribution. This can lead to a location-based deviation in the measured fault resistance. One of the solutions adopted is to characterize and compensate for the process variation in the read sense path. We will illustrate both these cases in the results section. The read sensing path used consists of the DFT circuit, the bit-lines, the clamping circuitry, and the current sensing amplifier. Some of the concerns that need to be addressed are as follows.

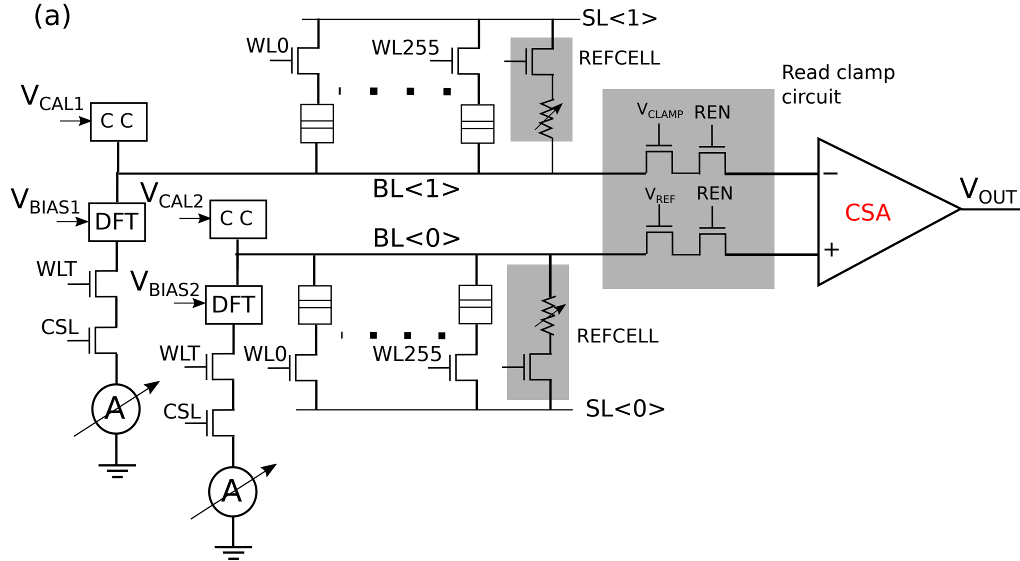


Figure 5.10: Schematic for the column read sense path and the current sense amplifier (CSA). CC is the compensation circuit.

1. Read-sensing path offsets due to sense amplifier and read peripheral circuitry can result in deviations in measured behavior.

2. Interconnect resistance and capacitance complicate the assessment of faulty bit-cell electrical behavior.

3. Column-to-column variation of the DFT R-V circuit parameters can result in measurement inaccuracy.

All of these factors contribute to the column-to-column offset current variations across the STT-MRAM sub-array, which limit the accuracy of the fault identification. Thus read sense path characterization is crucial prior to fault detection. One of the previous works in calibrating the sense amplifiers is provided by Cosemans et. al [75]. We are modifying the concept to suit the current sensing scheme used here. Here we show how the DFT scheme can be utilized for read-sense path characterization and compensation. Offset current characterization and compensation can be done in both global sub-array and local bit-line

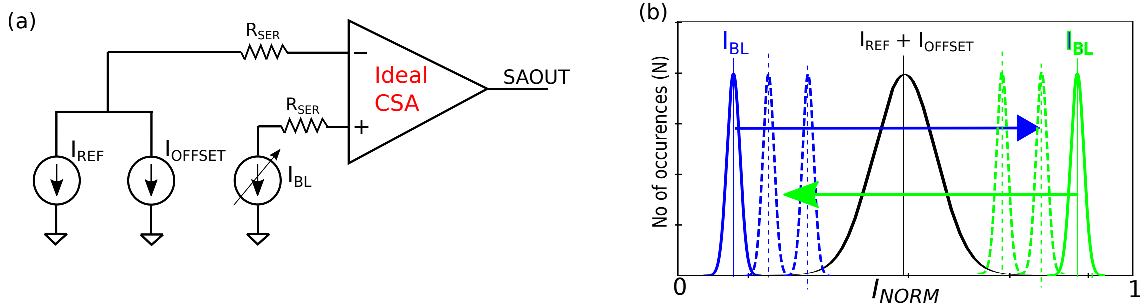


Figure 5.11: (a) Simplified equivalent circuit model for offset current calculation. (b) Offset characterization of read sensing circuitry using the DFT cells.

level within the die, and we are attempting compensation schemes in both these levels. The global-level characterization is done prior to deployment of the dies and the local-level characterization can be performed online during the boot-up phase of the die.

Global Offset Characterization and Compensation

The column select transistors are implemented on one read path per sub-array for global offset characterization and compensation immediately after fabrication. This allows initial accurate measurement of global offset currents for read paths across the die. Fig. 5.10 shows the schematic setup for the sequential mode column characterization. The column consists of 256 MTJ bit-cells with the DFT cell at the end of the column. Each DFT cell in the column pairs (e.g., BL<0> and BL<1>) is selected by using the column select transistors and the current is measured off-chip per column pair at a time. Initially, the DFT cell is enabled in the bit-line with the reference cell in the other bit-line. V_{BIAS} input is swept and the SAOUT data is collected. The test setup shown in the figure can be simplified into an equivalent model as shown in Fig. 5.11(a). Here the read sense path offset is referred to the DFT bit-cell as the system offset input current (I_{OFFSET}). Thus the total bit-line current I_{BLM} is composed of the reference current (I_{REF}) set by the DFT cell and the system input offset current. In order to find the offset current, we first find the I_{BL} current at which the $I_{BL} = I_{BL\text{-}M}$, this is also the point where the SA outputs exhibit a 50% probability of a logic '0' or '1' output, signifying the matching of the both input currents. The offset current is determined by

$$I_{OFFSET} = I_{BL_M} - I_{REF.} \tag{5.1}$$

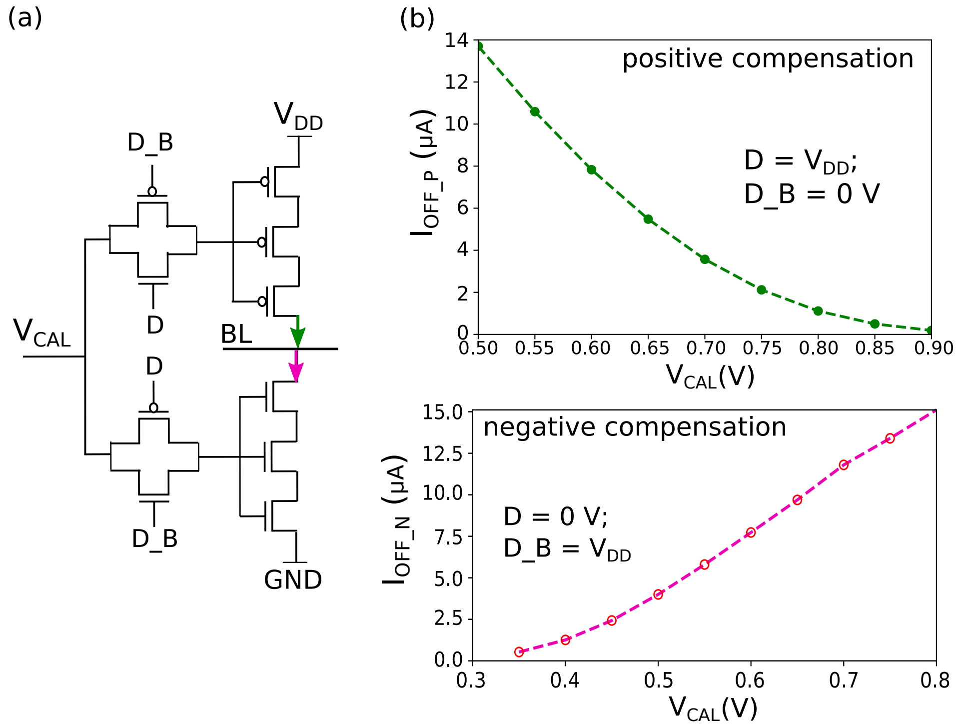


Figure 5.12: (a) The read offset compensation circuit. (b) The offset compensation current from PMOS (I_{OFFP}) and NMOS (I_{OFFN}) chains.

This offset current found is compensated across the bit-lines within the sub-array. The calibration circuit is enabled to provide a fixed offset currents and the corresponding V_{CAL} value is set to compensate for the offset current in the column pairs within the sub-array. The schematic for the compensation circuit is shown later in Fig. 5.12(a). The compensation current vs. the applied V_{CAL} bias is shown in figure. The circuit is able to compensate a current upto +/- 14 μA with a 5-bit control.

Local Offset Compensation

The previous global compensation technique provides a provision to adjust the global offset for a given sub-array, Thus allows the shifting of all the offsets by a predetermined amount for all read-sense paths. However, the process-induced variation can also impact each read sense paths differently, resulting in varying offset in each read sense path. Here a local

compensation scheme is introduced that focus on using the DFT cells to minimize offset within each read sense path of the sub-array. The DFT based current scanning is illustrated in Fig. 5.11. The DFT cells are also impacted by process induced variation and appropriate sizing of transistor width and length is required to limit the induced current variation to 1-2 order less than the read-path offset current characterized. For local compensation, the read-sense path voltage offset is defined as the difference in voltage applied to the DFT gate bias with respect to the known reference voltage on the alternate DFT cell(V_{BIAS_REF}), to obtain a 50% distribution of 0 or 1 in SAOUT.

The offset voltage is also dependent on the resistance being measured within the parameter range setting (Nominal mode in this case). To accommodate for the varying offset with resistance, we averaged the offset voltage obtained from 3 resistance values (lower, higher and middle) within the tuning range to get the final offset voltage. These offset values can be precomputed and loaded into the BIST memory based on the read sense path address evaluated. The average offset voltage corresponding to each read sense path and the DFT mode is stored in the external memory and applied to the V_{BIAS} during the fault measurement. The offset voltage stored is added to the V_{BIAS} measured from the read sense path to obtain the compensated V_{BIAS},

$$V_{\text{OFFSET}}(n) = V_{BIAS_REF} - \frac{\sum_{n=0}^{N-1} V_{\text{BIAS}}(n)}{N} \tag{5.2}$$

where N represents the total number of samples taken for different resistances using the read sense amplifiers. V_{BIAS_REF} is the V_{BIAS} applied to get the reference current (I_{REF}). We illustrate the local offset compensation and its effectiveness in the results section of this chapter.

5.3 Case Study and Simulation

5.3.1 Read-Sensing Circuitry Simulations

We have simulated the DFT circuit in CMOS general-purpose 65 nm technology to show the proposed DFT-based detection and monitoring scheme. All the simulations were done on CADENCE SPICE with a temperature of 25°C. For our study, we considered two commonly used read sensing circuits for DFT-based characterization. Type-I sensing in Fig. 5.13(a) utilizes a current latched sense amplifier (CLSA) scheme with a PMOS load [6] and Type-II current sensing [9], [30],[72] is similar to a voltage latched sense amplifier

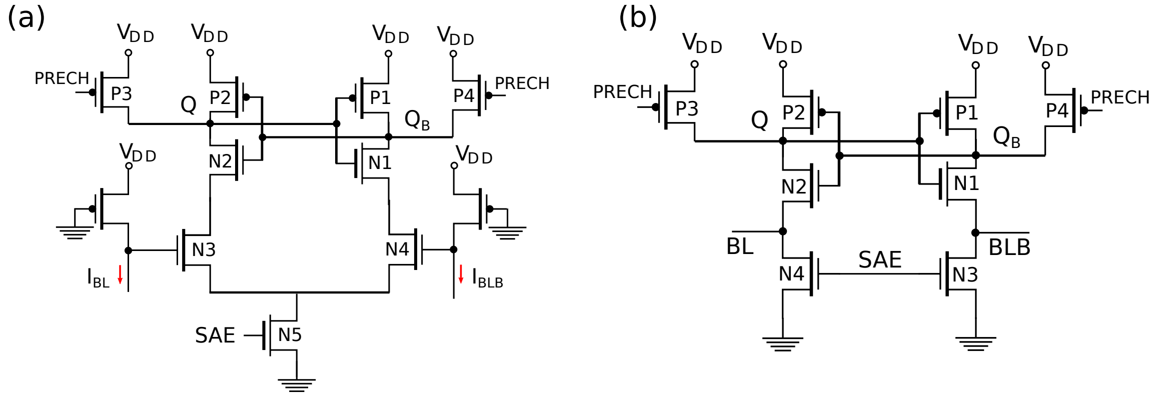


Figure 5.13: Sense amplifier circuit design used for characterization. (a) Type I CLSA circuitry. (b) Type II VLSA sense amplifier circuit design.

(VLSA) design, where the sensing currents for the bit-cell is applied to the source nodes of the sensing transistor as shown in Fig. 5.13(b).

Fig. 5.14 shows the post layout simulations for the read operation using a CLSA scheme. First, the bit-cells are activated and the pre-charge is released, charging and setting up the bit-line currents in the BL and REF line. Once a sufficient differential ΔI is established, the SAE is enabled to activate the latch circuitry in the CLSA circuit. The SAE starts the positive feedback that is aided by the differential current to make the latch converge to a given state. SAOUT provides the output of the latch which is used to read a 1 or 0 stored in the bit-cell. Prior to fault detection, we need to characterize the sensing scheme to process variation to evaluate the reliability of the SAOUT data with varying bit-line current. The SAOUT is sensitive to the differential current developed across the bit-lines and the mismatch of the transistors within the SA design. We performed a bit-line current sweep my tuning the DFT resistance characteristics for a given reference current. Fig. 5.15 shows the normalized SAOUT distribution for Type-I and Type-II current sensing schemes. Each bit-line current data point is based on 1000 Monte-Carlo transient runs considering process variation in both sense amplifiers and the DFT test structure. For the simulation, a DFT cell with the current variation of 0.135 μA, which is 2 orders of magnitude less than the offset current distribution was measured (1%). The mean bit-line flipping current ($I_{BL-FLIP}$) is the I_{BL} current at which the SA output generates an equal number of 1 and 0 (50% yield) under a given read-access time. Here it can be seen that the VLSA sensing scheme is closer to the ideal step response characteristics. The standard deviation of the

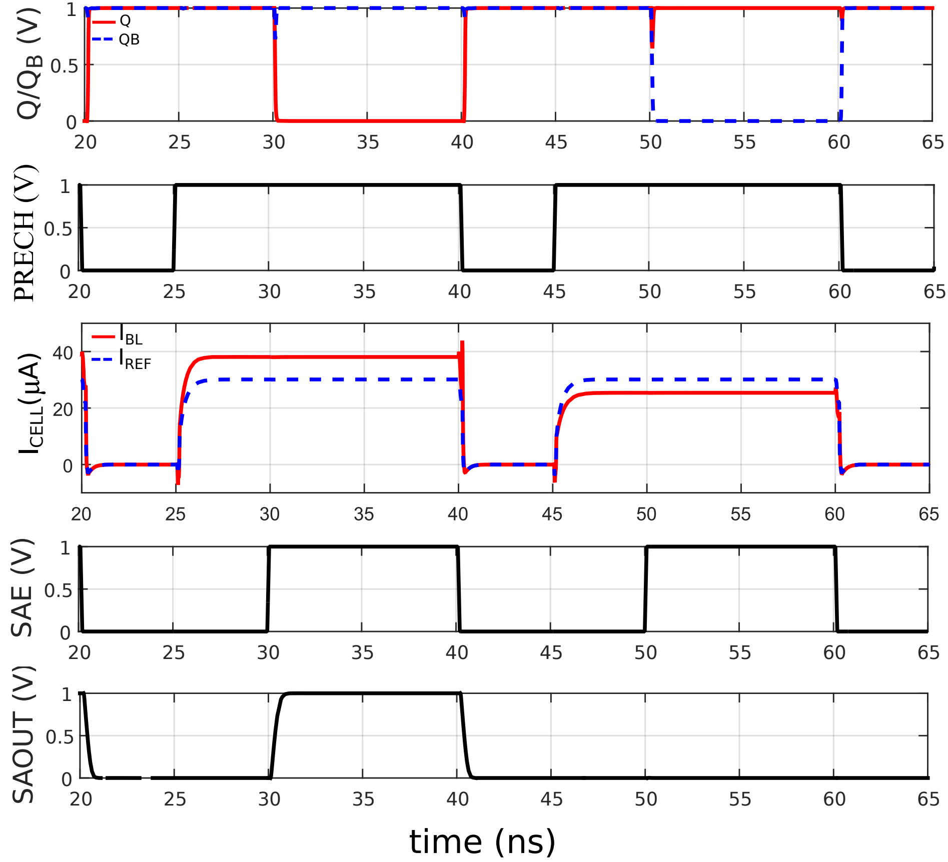


Figure 5.14: Post layout simulation showing the read operation in a CLSA sensing scheme.

distribution is computed for each scheme, and VLSA was more suitable for read-sensing compared to CLSA for the same transistor area.

Consider a pin-hole growth scenario on the MTJ stacks. The pin-hole growth occurs in 3 phases that result in the eventual barrier breakdown of the MTJ stack. In the initial phase, the junction doesn't show any degradation with applied bias voltage and current. In the second phase around the breakdown voltage, the pin-hole area keeps increasing and the MTJ junction area decreasing with the increase in the applied current. The R-V behavior of the MTJ device is non-linear in this region and works like a degraded MTJ. In the final phase, the pin-hole occupies most of the junction area, making the MTJ device behave like

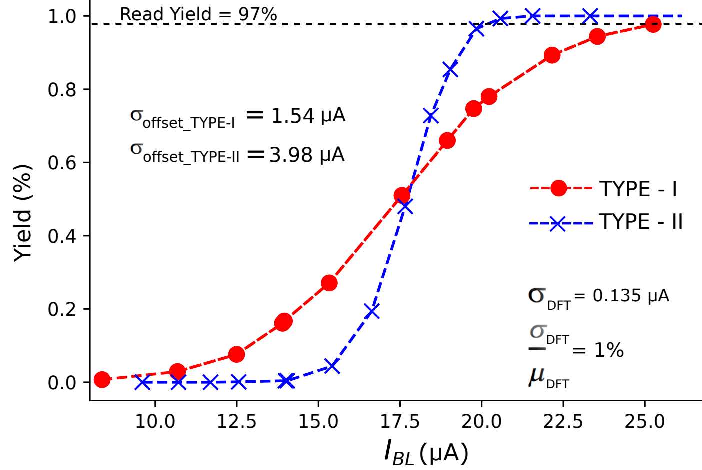


Figure 5.15: Statistical yield simulation results for the read sensing yield as a function of the bit–line.

a resistor. Symptoms of pin-hole area expansion are shown in the electrical characteristics of the MTJ cell [76], where the degradation in *RA* and *TMR* due to pin-hole formation can be described by [73][28]

$$\mathrm{RA_{eff}} = \frac{A_0}{\frac{A - A_{\mathrm{pin}}}{\mathrm{RA_{jun}}} + \frac{A_{\mathrm{pin}}}{\mathrm{RA_{pin}}}} \tag{5.3}$$

$$\mathrm{TMR_{eff}} = \frac{\mathrm{TMR_0}(\mathrm{RA_{eff}} - \mathrm{RA_{pin}})}{\mathrm{RA_{jun}} + RA_{\mathrm{pin}}}. \tag{5.4}$$

Here RA_{eff} and TMR_{eff} are the effective Resistance-Area (RA) product and TMR of the MTJ stack, respectively, with pin-holes considered. RA_{pin} and A_{pin} correspond to the RA and area of the pin-hole formed. TMR_0 and A_0 correspond to the initial TMR and area of the junction. Assuming RA, let us consider a pin-hole area of 5% of the total MTJ stack cross-sectional area. The R-V electrical characteristics of the faulty MTJ is as shown in Fig. 5.16(a). The DFT cell resistance ($\mathrm{R_{DFT}}$) is scanned until a match with the faulty R_P resistance is obtained. This is done by setting the DFT cell latch in R_P mode and sweeping the V_{BIAS} applied. Fig. 5.16(b) shows the change in R_{DFT} and SAOUT for varying V_{BIAS} for both R_P and R_{AP} of the faulty MTJ bit-cell. Fig. 5.16(c) shows the read sense amplifier

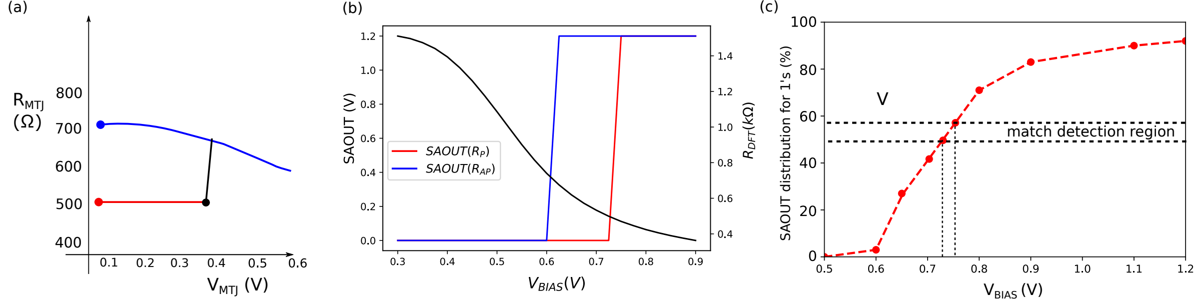


Figure 5.16: Bias voltage generation and analog multiplexer array for selecting DFT control voltages.

outputs for multiple DFT scanning of R_P resistance across several read sensing paths. The sense amplifier outputs (number of '1') show a cumulative distribution curve due to the CMOS process variation incurred in different read paths. Here the match detection region represents the bias voltage range within which the DFT resistance matches the faulty MTJ resistance. Similar resistance scan process is repeated to find R_{AP} by setting the DFT cell to AP state and sweeping the $V_{SS_AP.}$ We can then compute the TMR of the faulty MTJ based on the ratio of peak value of R_{AP} and R_P resistances. The switching voltage identification is similar to a write and read test as explained previously. The data in MTJ bit-cell and DFT cell are sensed separately with respect to the reference cell.

Symptoms of pin-hole formed in the MTJ are more prominent in the resistance characteristics of the MTJ device [73], where the degraded RA_{eff} due to pin-hole formed can be described by [43]. Based on the equation, we computed the R_P resistance and evaluated the accuracy of our scheme over the resistance range. For the setup, we considered a 80 nm diameter MTJ with $RA_{jun} = 12\ \Omega.\mu\text{m}^2$, $RA_{pin} = 0.2\ \Omega.\mu\text{m}^2$; Fig. 5.17 shows the R_P vs. normalized pin-hole area with respect to the total MTJ junction area. The resistance degrades rapidly with an increase in the pin-hole area as shown in Fig. 5.17 (blue curve). Based on the resistance ranges, we have split it into nominal, weak short and strong short. The fault resistance corresponding to each pin-hole area was injected into the test structure and evaluated using DFT and read sensing path. It can be seen that in the nominal range, the DFT measures and tracks the resistance with worst-case error of +/- 3.3 %. The weak short resistance tracking was done using Monte Carlo simulations in 65 nm CMOS technology since the test chip implementation included only the nominal cell in the DFT circuit. The relative worst-case error was observed to be +/- 8 % for the weak short case.

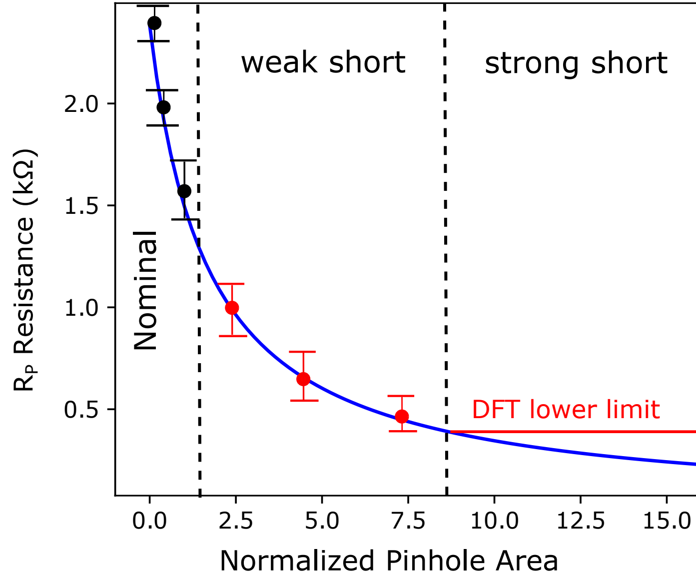


Figure 5.17: R_P vs. pin-hole area. The resistance degradation based on MATLAB plot is shown in blue. The black and red corresponds to the measured resistance from the test structure, and the simulation respectively. The dot represents the mean value and the bounds represent the worst case max and min values.

5.4 Results

We implemented a 512 DFT-cell sub-array in 65 nm CMOS technology to evaluate the feasibility of the DFT based BIST scheme. The chip micrograph is shown in Fig. 5.18. The sub-array consists of 8 bit-lines with 64 DFT-cells in each bit-line and a read sensing circuitry per bit-line pair. The design scheme allows replicating the faulty bit-cell behavior simultaneously in different locations within the sub-array. The BIST logic was implemented externally to control and gather data from the sense amplifier outputs within the sub-array. It was done externally in FPGA for ease of modification and adaptability; however, they can be easily adaptable for implementation within a die. Here we are showing several test strategies that demonstrate the capability of the DFT scheme. A proper selection of DFT circuitry and their synergistic usage are important for conserving the silicon real-estate below an acceptable level. Targeted testing conditions can be assigned and the DFT circuitry can be customized for catering to the targeted test scenarios.

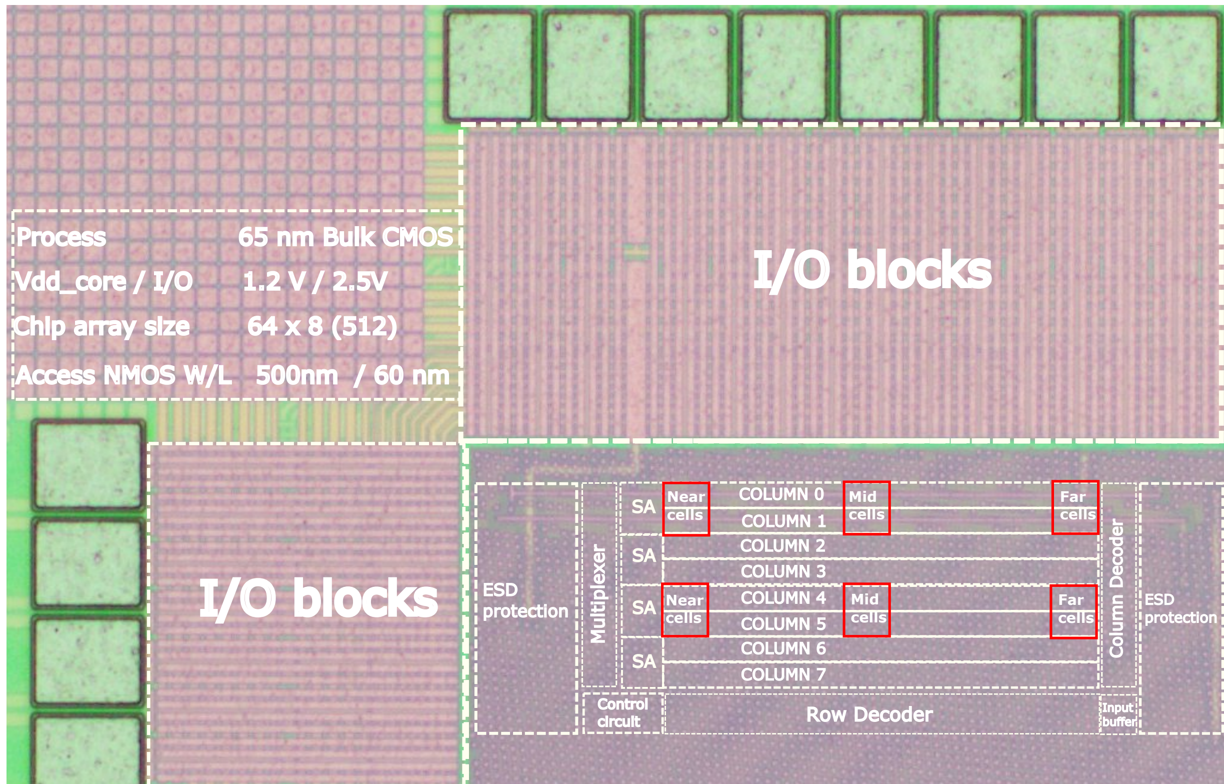


Figure 5.18: Die micrograph of the DFT sub-array implemented in 65 nm bulk CMOS process.

The DFT circuit transistors here are sized 2X-3X compared to the regular minimum size transistors to control variability. Here the DFT cell is used per column of 1024 bit-cells. The latch circuitry is shared between the DFT pairs in the bit-line. Based on layout estimation of 1024 row x 256 column [9][77] within the sub-array, the area overhead due to the DFT circuitry, the bias circuitry, and the classifier detection circuit account to 8-9% of the total design area. Based on the NVSIM work by Dong et. al. [77], the actual area of the STT-MRAM memory implemented in 65 nm CMOS technology is 39 mm^2 for 256 sub-arrays, each with 256 kB size. The DFT cell area is based on the layout we implemented in the 65 nm test chip. The table below illustrates the area computation for a sub-array,

Here the BIST controller is implemented in the die level, so the effective area contribution is distributed equally among the sub-arrays. The area of each cell type is represented in percentage (%) at the last column, which is computed with respect to the total STT-MRAM + periphery area shown in the last row of the table. The DFT cells in both the

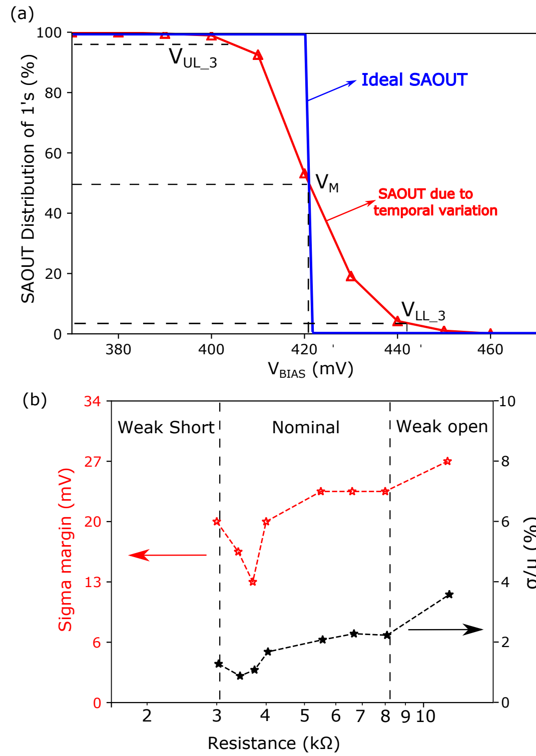
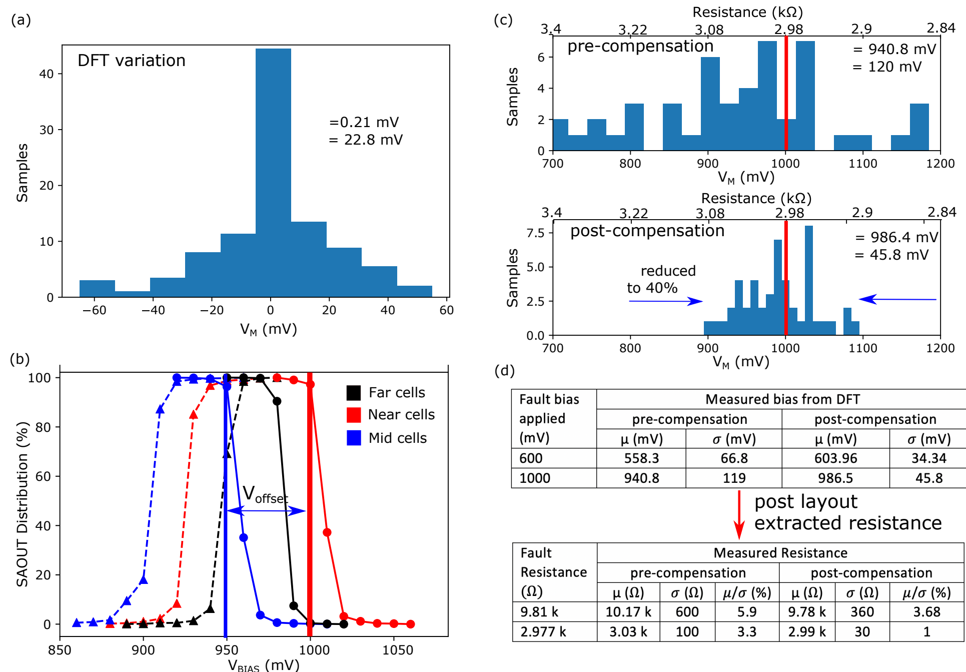


Figure 5.19: Impact of temporal variation based on the test chip measured. (a) The SAOUT distribution vs V_{BIAS} for ideal (blue) and temporal variation case (red). (b) Standard deviation of the V_M for different resistance measured.

Cell type	Area per cell (μm^2)	Number per sub-array	Area contribution (μm^2)	percentage contribution
DFT cell (RV + latch)	45	256	11,520	7.56%
Bias + detection circuit	1500	1	1500	0.98%
BIST controller per die		1/256	500	0.34%
STT-MRAM cells alone	0.3584	256 K	93,952	
STT-MRAM + periphery		256 K	152,343.75	

Table 5.2: Area estimate with respect to the total area of the STT-MRAM array

bit-line pair are enabled for the measurement. The fault is injected in one DFT cell and



Fault bias applied (mV)	Measured bias from DFT			
	pre-compensation		post-compensation	
	μ (mV)	σ (mV)	μ (mV)	σ (mV)
600	558.3	66.8	603.96	34.34
1000	940.8	119	986.5	45.8

Fault Resistance (Ω)	Measured Resistance					
	pre-compensation			post-compensation		
	μ (Ω)	σ (Ω)	μ/σ (%)	μ (Ω)	σ (Ω)	μ/σ (%)
9.81 k	10.17 k	600	5.9	9.78 k	360	3.68
2.977 k	3.03 k	100	3.3	2.99 k	30	1

Figure 5.20: Impact of spatial variation, (a) Measured DFT variation along the column, the bias voltage for SA distribution with V_M is shown on the top right. (b) Distribution from the DFT cells obtained for a given SA location. (c) Resistance distribution before (top) and after (bottom) read path circuit offset compensation. The red line indicates the injected fault voltage and the blue line indicates the measured match voltage (V_M) bottom shows the corresponding resistance values computed from post layout circuit simulation.

the cell in the alternate bit-line is swept using V_{BIAS}. The DFT bias voltages are swept at 10 mV steps and the read sensing outputs are recorded. Fig. 5.19 shows the effect of temporal variation on the read sensing. Fig. 5.19(a) shows the SAOUT distribution for the sweep in V_{BIAS} voltage. The SAOUT samples have a gradual transition with V_{BIAS} voltage, compared to the ideal distribution response. This would impact the accuracy of the resistance detected. Fig. 5.19(b) shows the standard deviation of the V_M captured for different resistance values measured in the nominal range. It can be seen that the relative resistance variation measured (black) is less than 4% over the nominal range after averaging scheme was utilized. Faults of the same resistance are replicated at different locations in the array to understand the impact of spatial variation on the read sensing

path. The data is measured from 5 different dies, and in each die, we tested on 4 bit-lines with DFT samples taken from 3 different locations (near, middle and far) as shown in Fig 5.18. The locations are chosen this way to extract as much spatial induced variations as possible. For each V_{BIAS} setting, we captured and averaged 32 SAOUT samples to reduce the V_M fluctuation due to temporal variation. The DFT cell variation has to be one order less than the total read sense path variation to accurately measure resistances. Based on the SAOUT data across multiple bit-lines, we analyzed the variation due to DFT along the bit-line. Fig 5.20(a) shows the V_M samples vs. the V_{BIAS} voltage. The contribution of the DFT cell variation on the V_M was observed to be having a standard deviation of 22.8 mV which is much smaller than the total read path variation shown in Fig. 5.20(c). In a practical implementation scheme, the V_{BIAS} can be swept with 50 mV step size considering the variation induced spreading of the V_M bias.

It can be seen in the pre-compensation plot in Fig. 5.20(c) that the process variation has resulted in the spreading of the V_M captured for a given resistance. This is due to the varying offset in different read sense paths. Fig 5.20(b) illustrates the SAOUT distributions for DFTs (near, middle and far) and from BL (solid line) and BLB (dashed line). The average offset for a given sense path is measured as the

$$V_{OFFSET} = V_{REF} - \sum_{n=0}^{5} V_M(n)/N \tag{5.5}$$

where n corresponds to the number of V_{BIAS} sweeps done for different resistance values. Fig. 5.20(c) shows the V_M distribution before and after the local compensation. From the figure, it can be seen that the V_M distribution has tightened to 40% of its previous standard deviation. In terms of resistance, we are able to achieve a variability of less than 4% of its mean value within the nominal resistance as shown in Fig 5.20(d). The mean of the distribution has also improved, and approach closer to the applied fault voltage after compensation.

The primary factor for spatial variation is due to the transistor V_T and W/L variations in the read sensing path (sense amplifier, read clamping circuitry) and the DFT circuit. However, the circuit part of the DUT that contributes more to the variation is unknown. We performed 1000 sample Monte-Carlo simulations considering variation in individual sections of the DUT. Process variation spreads the resistance distribution measured, resulting in overflowing across the V_{BIAS} resistance bin values. Fig. 5.21 shows the bin error contribution (bin size = 50 mV) due to different parts of the DUT before offset compensation. It can be observed that the DFT circuit maintains minimal variation across the resistance range and the sense amplifier contribution is prominent.

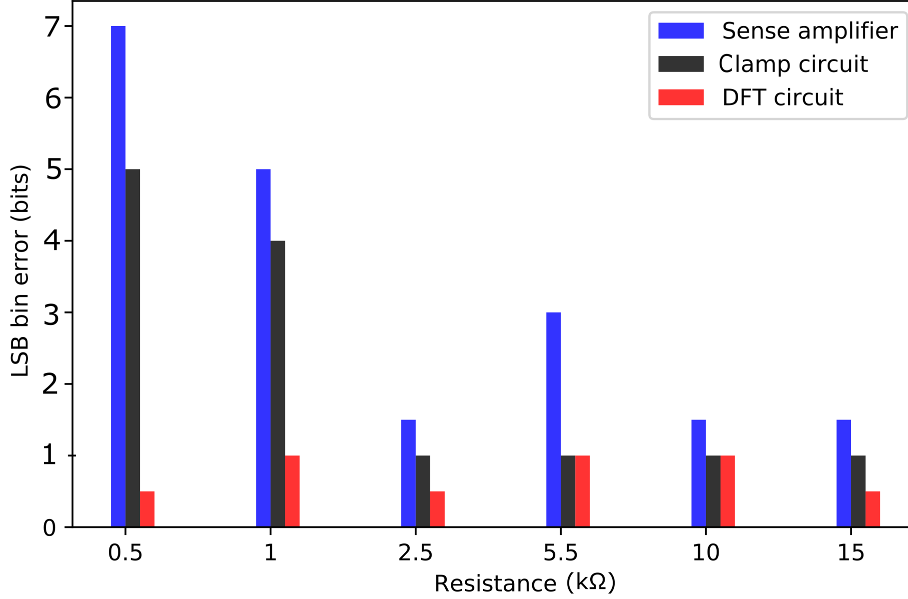


Figure 5.21: Worst case LSB bin errors contributed by read sense amplifier, read voltage clamp circuit and the DFT circuit in the proposed scheme.

The V_{BIAS} voltage bin values are translated to the equivalent resistance values to evaluate the accuracy. Fig. 5.22 shows the accuracy of the entire range of operating resistance values. Here, the post-compensation is applied to reduce the bin error to +/-1 LSB. The pre-compensation data (red solid line) shows the impact of process variation on the measured resistance value. The accuracy is improved by offset compensation as shown in the post-compensated plot (blue dashed line), which indicates the accuracy of the DFT scheme for the V_{BIAS} bin resolution of 50 mV. Based on the offset characterization data obtained from the sense amplifiers, we compute the new V_{BIAS} voltage that corresponds to almost zero offset in the sense amplifier. These are applied as offset compensation bits in the DFT bias vector. The compensation tightens the resistance distribution spread, confirming it within the bin range to avoid overflowing. During design, the bit-resolution for the compensation current can be improved to get an accuracy of 7% for 50 mV V_{BIAS} bin size. The accuracy could not be improved further for the current configuration since it is limited by the V_{BIAS} bin resolution. Improving the V_{BIAS} precision and averaging can provide better accuracy in exchange for higher test time.

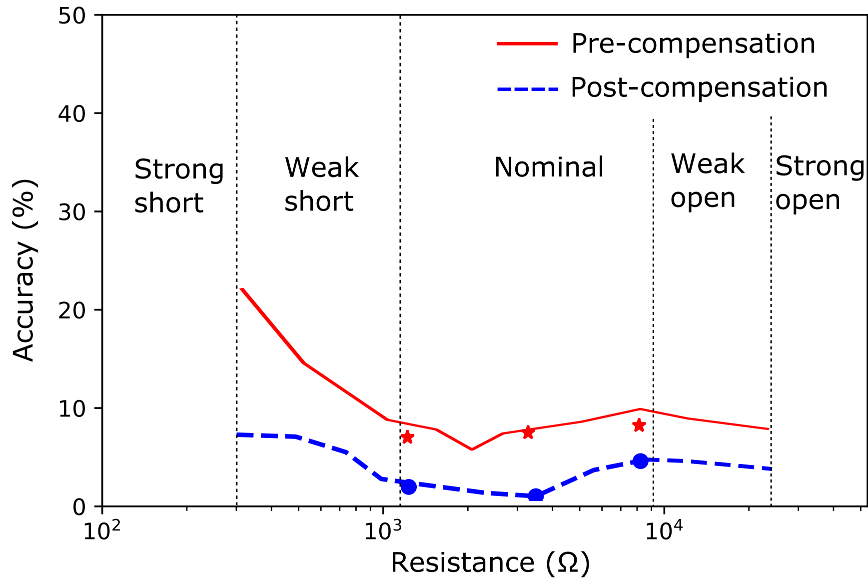


Figure 5.22: Resistance identification accuracy based with measurement and post-layout simulations for the DFT scheme pre-compensation (red solid line) and post-compensation (blue dashed line). The red stars and blue dots represent measured values before and after compensation.

5.5 Summary

Periodic monitoring of MTJ bit-cells is crucial for maintaining the reliability of the STT-MRAM arrays. In this chapter, we demonstrate the BIST methodology along with the offset compensation technique to provide a DFT-based BIST scheme for monitoring and detection of faulty MTJ behavior. An overview of the types of testing possible with the scheme and its complexity is presented. The simulation results verify the functionality of the BIST based scheme. The 65 nm silicon results from 5 different dies were used to verify the detection accuracy of the DFT based scheme.

Chapter 6

Conclusion

This chapter summarizes the potential applications and outlines the research direction moving forward.

6.1 Research Contribution

6.1.1 In-Die Parametric Characterization

We implemented the integrated DFT circuit and the read access sensing path in 65-nm CMOS technology to evaluate the read margin characteristics of the column. The yield estimation results provide insight into the response of the MRAM to MTJ parameter variations across the wafer as well as an estimate on the wafer usable area. The proposed in-circuit DFT scheme has the potential to provide a platform for parameter testing and optimization of the STT-MRAM fabrication process.

6.1.2 Faster Wafer Screening and MTJ stack development

DFT scheme offers more input or output ports and test options for wafer-level characterization per bit-line. The scheme allows us to test the bit-cell characteristics and the peripheral read and write circuitry performance in a decoupled manner. This provides more visibility in terms of isolating the manufacturing issues and finding fault contributions from each circuit segment in the array. The DFT scheme can improve the test visibility for novel MTJ stack development and reduce wafer turn-around time to deployment.

6.1.3 Bit-Cell Health Monitoring

Identifying defects in MTJ device is crucial for the long term reliability of STT- MRAM. The DFT technique is able to detect the trajectory of the parameter deviation by periodically monitoring the MTJ device parameter through nominal, weak short and weak open parameter deviation cases. This scheme can be used for alerting the system regarding the fault formation and providing warning to the user before the complete failure of the bit-cell, allowing time for introducing redundancy schemes to mask the error during operation or even relocate the memory data to another location. Furthermore, gathering large data set from bit-cell operations is useful for accurate defect modelling. The scheme can collect bit-cell data statistics and fault locations and provide feedback to improve the fabrication process and yield learning.

6.1.4 65nm Test-Chip Design and Implementation

Two test chips have been designed and implemented successfully in 65nm CMOS technology to provide silicon validation for the proposed DFT schemes. The chip was designed to provide a test platform for peripheral circuitry characterization and DFT circuit demonstration. Both the chips contain 512 DFT cell array with peripheral read and write circuitry. The chip was designed to operate in 1.2 V core voltage and 2.5 V peripheral IO voltage. Both the test chips were successfully tested for the given process. In the future, the DFT can be explored for 22 nm to evaluate the impact of interconnect resistance and capacitance.

6.1.5 STT-MRAM Characterization

Device level characterization was performed for single device MTJ stacks from 60 to 150 nm for novel MTJ stack compositions. The spatial parameter deviation and its variability was captured for modeling of the MTJ device. The technique provide an empirical model for spatial dependant MTJ device modeling and simulation based estimation of STT-MRAM array at different spatial locations of the wafer. The variation model based simulation can provide estimation of yield attainable for the given parameter spatial deviation profile. The characterization results also provide feedback for fabrication process optimization.

6.1.6 Future Work

The thesis work explores the DFT scheme to explore parameter sensitivity analysis and bit-cell health monitoring. In the future, the STT-MRAM technology is getting adopted to FDSOI and FINFET based sub-40 nm CMOS technologies. The rapid scaling has resulted in the increase in the parameter variability of the MTJ and transistors. However, the DFT cells based on transistors with appropriate sizing can still be implemented in these advanced nodes to provide a test platform for testing and characterization. The DFT cell can be sized to reduce the variability 1-2 order lower than the rest of the circuitry for testing. Apart from bit-cell monitoring, this work could result in characterization methodologies for advanced CMOS process nodes. Some of the future research directions that would be pursued based on the thesis work is,

a) As the semiconductor technology scales, it is anticipated to increase the interconnect resistance and capacitance variation between bit-lines of the STT-MRAM array. The modern STT-MRAM arrays utilize bit-lines routed across two metal layers to reduce the interconnect resistance [31]. Thus the characterization of interconnect resistance along with the bit-cell resistance is crucial for parametric characterization. The placement of two DFT scheme (far and near the SA) could be explored for interconnect resistance evaluation for 22 nm technology nodes. The first DFT cell in the bit-line is placed near to the current sense amplifier input, and the other DFT-cell is placed at the far end of the bit-line. The bit-line resistance can be evaluated by comparing the difference in DFT-cell resistances measured. We may explore the write operation of the bit-lines for capacitance characterization. Tunable capacitance can be provided at the bit-lines which can be tuned to match the bit-line capacitance.

b)For high-density MRAM implementations, the bit-lines are precharged to appropriate currents before the current sensing. The establishing of the read current may take significant time depending on the capacitance in the bit-line. A new current sensing based read scheme is investigated to improve the read signal margin and decrease the read time for high-density STT-MRAM architectures.

c) A DFT-cell based platform for comparative analysis of modern current sense amplifiers is proposed for STT-MRAM arrays. The array platform can be built using DFT modules to replicate bit-cell resistance characteristics. Selectable NMOS capacitor banks connected to the bit-line can replicate various capacitance in the bit-line. Here the objective is to perform side-by-side analysis of different sense amplifier schemes and observe their figure of merits for STT-MRAM applications.

References

[1] E. Chen, D. Apalkov, Z. Diao, A. Driskill-Smith, D. Druist, D. Lottis, V. Nikitin, X. Tang, S. Watts, S. Wang, S. Wolf, A. Ghosh, J. Lu, S. Poon, M. Stan, W. Butler, S. Gupta, C. Mewes, T. Mewes, and P. Visscher, "Advances and Future Prospects of Spin-Transfer Torque Random Access Memory," *IEEE Transactions on Magnetics*, vol. 46, no. 6, pp. 1873–1878, June 2010.

[2] S. A. Wolf, S. J. Lu, M. R. Stan, E. Chen, and D. M. Treger, "The Promise of Nanomagnetics and Spintronics for Future Logic and Universal Memory," *Proceedings of the IEEE*, vol. 98, no. 12, pp. 2155–2168, Dec 2010.

[3] I. Žutić, J. Fabian, and S. Das Sarma, "Spintronics: Fundamentals and applications," *Rev. Mod. Phys.*, vol. 76, pp. 323–410, Apr 2004.

[4] J. G. Zhu and C. Park, "Magnetic tunnel junctions," *Materials Today*, vol. 9, no. 11, pp. 36–45, 2006.

[5] L. Thomas, G. Jan, J. Zhu, H. Liu, Y.-J. Lee, S. Le, R.-Y. Tong, K. Pi, Y.-J. Wang, D. Shen, R. He, J. Haq, V. Lam, K. Huang, T. Zhong, T. Torng, and P.-K. Wang, "Perpendicular spin transfer torque magnetic random access memories with high spin torque efficiency and thermal stability for embedded applications (invited)," vol. 115, pp. 172 615.1–172 615.6, May 2014.

[6] R. Takemura, T. Kawahara, K. Miura, H. Yamamoto, J. Hayakawa, N. Matsuzaki, K. Ono, M. Yamanouchi, K. Ito, H. Takahashi, S. Ikeda, H. Hasegawa, H. Matsuoka, and H. Ohno, "A 32-Mb SPRAM With 2T1R Memory Cell, Localized Bi-Directional Write Driver and '1'/'0' Dual-Array Equalized Reference Scheme," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 4, pp. 869–879, Apr 2010.

[7] S. Yuasa, T. Nagahama, A. Fukushima1, Y. Suzuki, and K. Ando, "Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions," *Nature materials*, vol. 3, pp. 868–871, 2004.

[8] T. M. Maffitt, J. K. Debrosse, J. A. Gabric, E. T. Gow, M. C. Lamorey, J. S. Parenteau, D. R. Willmott, M. A. Wood, and W. J. Gallagher, "Design considerations for MRAM," *IBM Journal of Research and Development*, vol. 50, no. 1, pp. 25–39, Jan 2006.

[9] K. Tsuchida, T. Inaba, K. Fujita, Y. Ueda, T. Shimizu, Y. Asao, T. Kajiyama, M. Iwayama, K. Sugiura, S. Ikegawa, T. Kishi, T. Kai, M. Amano, N. Shimomura, H. Yoda, and Y. Watanabe, "A 64Mb MRAM with clamped-reference and adequate-reference schemes," *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 258–259, Feb 2010.

[10] D. Datta, B. Behin-Aein, S. Datta, and S. Salahuddin, "Voltage Asymmetry of Spin-Transfer Torques," *IEEE Transactions on Nanotechnology*, vol. 11, no. 2, pp. 261–272, Mar 2012.

[11] K. Lee and S. Kang, "Control of Switching Current Asymmetry by Magnetostatic Field in MgO-Based Magnetic Tunnel Junctions," *IEEE Electron Device Letters*, vol. 30, no. 12, pp. 1353–1355, Dec 2009.

[12] A. Jog, A. K. Mishra, C. Xu, Y. Xie, V. Narayanan, R. Iyer, and C. R. Das, "Cache Revive: Architecting Volatile STT-RAM Caches for Enhanced Performance in CMPs," *Proceedings of the 49th Annual Design Automation Conference*, pp. 243–252, 2012.

[13] M. Rasquinha, D. Choudhary, S. Chatterjee, S. Mukhopadhyay, and S. Yalamanchili, "An energy efficient cache design using Spin Torque Transfer (STT) RAM," *ACM/IEEE International Symposium on Low-Power Electronics and Design (ISLPED)*, pp. 389–394, Aug 2010.

[14] E. Kultursay, M. Kandemir, A. Sivasubramaniam, and O. Mutlu, "Evaluating STT-RAM as an energy-efficient main memory alternative," *IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)*, pp. 256–267, April 2013.

[15] P. Naji, M. Durlam, S. Tehrani, J. Calder, and M. DeHerrera, "A 256 kb 3.0 V 1T1MTJ nonvolatile magnetoresistive RAM," *IEEE International Solid-State Circuits Conference, 2001. Digest of Technical Papers. ISSCC.*, pp. 122–123, Feb 2001.

[16] B. Engel, J. Akerman, B. Butcher, R. Dave, M. DeHerrera, M. Durlam, G. Grynkewich, J. Janesky, S. Pietambaram, N. Rizzo, J. Slaughter, K. Smith, J. Sun, and S. Tehrani, "A 4-Mb toggle MRAM based on a novel bit and switching method," *IEEE Transactions on Magnetics*, vol. 41, no. 1, pp. 132–136, Jan 2005.

[17] I. Prejbeanu, W. Kula, K. Ounadjela, R. Sousa, O. Redon, B. Dieny, and J.-P. Nozieres, "Thermally assisted switching in exchange-biased storage layer magnetic tunnel junctions," *IEEE Transactions on Magnetics*, vol. 40, no. 4, pp. 2625–2627, July 2004.

[18] J. Slonczewski, "Current-driven excitation of magnetic multilayers," *Journal of Magnetism and Magnetic Materials*, vol. 159, no. 12, 1996.

[19] J. Alzate, P. Amiri, P. Upadhyaya, S. Cherepov, J. Zhu, M. Lewis, R. Dorrance, J. Katine, J. Langer, K. Galatsis, D. Markovic, I. Krivorotov, and K. Wang, "Voltage-induced switching of nanoscale magnetic tunnel junctions," in *IEEE International Electron Devices Meeting (IEDM)*, Dec 2012.

[20] Z. Wang, H. Zhou, M. Wang, W. Cai, D. Zhu, J.-O. Klein, and W. ZHAO, "Proposal of Toggle Spin Torques Magnetic RAM for Ultrafast Computing," *IEEE Electron Device Letters*, vol. 40, pp. 726–729, 03 2019.

[21] J. T. Heron, M. Trassin, K. Ashraf, M. Gajek, Q. He, S. Y. Yang, D. E. Nikonov, Y.-H. Chu, S. Salahuddin, and R. Ramesh, "Electric-Field-Induced Magnetization Reversal in a Ferromagnet-Multiferroic Heterostructure," *Phys. Rev. Lett.*, vol. 107, Nov 2011.

[22] A. Khan, D. E. Nikonov, S. Manipatruni, T. Ghani, and I. A. Young, "Voltage induced magnetostrictive switching of nanomagnets: Strain assisted strain transfer torque random access memory," *Applied Physics Letters*, vol. 104, no. 26, 2014.

[23] A. Driskill-Smith, D. Apalkov, V. Nikitin, X. Tang, S. Watts, D. Lottis, K. Moon, A. Khvalkovskiy, R. Kawakami, X. Luo, A. Ong, E. Chen, and M. Krounbi, "Latest Advances and Roadmap for In-Plane and Perpendicular STT-RAM," in *3rd IEEE International Memory Workshop (IMW)*, May 2011.

[24] M. Julliere, "Tunneling between ferromagnetic films," *Physics Letters A*, vol. 54, no. 3, pp. 225–226, 1975.

[25] S. S. P. 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*, no. 3, pp. 862–867, 2004.

[26] D. Shum, D. Houssameddine, S. Woo, Y. You, J. Wong, K. Wong, C. Wang, K. Lee, V. Naik, C. Seet, T. Tahmasebi, C. Hai, H. Yang, N. Thiyagarajah, R. Chao, J. Ting, N. Chung, T. Ling, and T. Andre, "CMOS-embedded STT-MRAM arrays in 2x nm nodes for GP-MCU applications," *2017 Symposium on VLSI Technology*, pp. T208–T209, Jun 2017.

[27] Y. Song, J. Lee, H. Shin, K. Lee, K. S. Suh, J. Kang, S. Pyo, H. Jung, S. Hwang, G. Koh, S. Oh, S. Park, J. Kim, J. Park, J. Kim, K. Hwang, G. Jeong, K. Lee, and E. Jung, "Highly functional and reliable 8Mb STT-MRAM embedded in 28nm logic," *2016 IEEE International Electron Devices Meeting (IEDM)*, pp. 27.2.1–27.2.4, Dec 2016.

[28] Y. Lu, T. Zhong, W. Hsu, S. Kim, X. Lu, J. Kan, C. Park, W.-C. Chen, X. Li, X. Zhu, P. Wang, M. Gottwald, J. Fatehi, L. Seward, J. Kim, N. Yu, G. Jan, J. Haq, S. Le, and S. Kang, "Fully functional perpendicular STT-MRAM macro embedded in 40 nm logic for energy-efficient IOT applications," *2015 IEEE International Electron Devices Meeting (IEDM)*, pp. 26.1.1–26.1.4, Dec 2015.

[29] C. Lin, S. Kang, Y. Wang, K. Lee, X. Zhu, W.-C. Chen, X. Li, W. Hsu, Y. Kao, M. Liu, Y. Lin, M. Nowak, N. Yu, and L. Tran, "45nm low power CMOS logic compatible embedded STT MRAM utilizing a reverse-connection 1T/1MTJ cell," *IEEE International Electron Devices Meeting (IEDM)*, pp. 1 – 4, Jan 2010.

[30] K. Rho, K. Tsuchida, D. Kim, Y. Shirai, J. Bae, T. Inaba, H. Noro, H. Moon, S. Chung, K. Sunouchi, J. Park, K. Park, A. Yamamoto, S. Chung, H. Kim, H. Oyamatsu, and J. Oh, "A 4Gb LPDDR2 STT-MRAM with compact 9F2 1T1MTJ cell and hierarchical bitline architecture," *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, pp. 396–397, Feb 2017.

[31] Y. Shih, C. Lee, Y. Chang, P. Lee, H. Lin, Y. Chen, K. Lin, T. Yeh, H. Yu, H. H. L. Chuang, Y. Chih, and J. Chang, "Logic Process Compatible 40-nm 16-Mb, Embedded Perpendicular-MRAM With Hybrid-Resistance Reference, Sub- μA Sensing Resolution, and 17.5-nS Read Access Time," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 4, pp. 1029–1038, April 2019.

[32] R. Robertazzi, J. Nowak, and J. Sun, "Analytical MRAM test," *2014 International Test Conference*, pp. 1–10, Oct 2014.

[33] I. Yoon, A. Chintaluri, and A. Raychowdhury, "EMACS: Efficient MBIST architecture for test and characterization of STT-MRAM arrays," *2016 IEEE International Test Conference (ITC)*, pp. 1–10, Nov 2016.

[34] W. Zhao, Y. Zhang, T. Devolder, J.-O. Klein, D. Ravelosona Ramasitera, C. Chappert, and P. Mazoyer, "Failure and reliability analysis of STT-MRAM," *Microelectronics Reliability*, vol. 52, p. 18481852, Sept 2012.

[35] G. Kar, W. Kim, T. Tahmasebi, J. Swerts, S. Mertens, N. Heylen, and T. Min, "Co/Ni based p-MTJ stack for sub-20nm high density stand alone and high performance embedded memory application," *Technical Digest - International Electron Devices Meeting, IEDM*, vol. 2015, pp. 19.1.1–19.1.4, Feb 2015.

[36] S. Peng, M. Wang, H. Yang, L. Zeng, J. Nan, J. Zhou, Y. Zhang, A. Hallal, M. Chshiev, K. Wang, Q. Zhang, and W. ZHAO, "Origin of interfacial perpendicular magnetic anisotropy in MgO/CoFe/metallic capping layer structures," *Scientific Reports*, vol. 5, p. 18173, Dec 2015.

[37] W. Zhao, X. Zhao, B. Zhang, K. Cao, L. Wang, W. Kang, Q. Shi, M. Wang, Y. Zhang, Y. Wang, S. Peng, J.-O. Klein, L. Naviner, and D. Ravelosona Ramasitera, "Failure Analysis in Magnetic Tunnel Junction Nanopillar with Interfacial Perpendicular Magnetic Anisotropy," *Materials*, vol. 9, p. 41, Jan 2016.

[38] A. Khvalkovskiy, D. Apalkov, S. Watts, R. Chepulskii, R. Beach, A. Ong, X. Tang, A. Driskill-Smith, W. Butler, P. Visscher, D. Lottis, E. Chen, V. Nikitin, and M. Krounbi, "Basic principles of STT-MRAM cell operation in memory arrays," *Journal of Physics D: Applied Physics*, vol. 46, p. 074001, Jan 2013.

[39] R. Bishnoi, M. Ebrahimi, F. Oboril, and M. Tahoori, "Read disturb fault detection in STT-MRAM," *2014 International Test Conference*, pp. 1–7, Oct 2014.

[40] Y. Ye, F. Liu, M. Chen, S. Nassif, and Y. Cao, "Statistical Modeling and Simulation of Threshold Variation Under Random Dopant Fluctuations and Line-Edge Roughness," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 19, pp. 987 – 996, Jul 2011.

[41] G. Radhakrishnan, Y. Yoon, and M. Sachdev, "A Parametric DFT Scheme for STT-MRAMs," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 27, pp. 1–12, Apr 2019.

[42] A. Chintaluri, H. Naeimi, S. Natarajan, and A. Raychowdhury, "Analysis of Defects and Variations in Embedded Spin Transfer Torque (STT) MRAM Arrays," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 6, pp. 1–11, Apr 2016.

[43] L. Wu, S. Rao, G. C. Medeiros, M. Taouil, E. J. Marinissen, F. Yasin, S. Couet, S. Hamdioui, and G. S. Kar, “Pinhole Defect Characterization and Fault Modeling for STT-MRAM Testing,” in *2019 IEEE European Test Symposium (ETS)*, 2019.

[44] L. Wu, M. Taouil, S. Rao, E. J. Marinissen, and S. Hamdioui, “Electrical Modeling of STT-MRAM Defects,” in *2018 IEEE International Test Conference (ITC)*, 2017.

[45] M. Hosomi, H. Yamagishi, T. Yamamoto, K. Bessho, Y. Higo, K. Yamane, H. Yamada, M. Shoji, H. Hachino, C. Fukumoto, H. Nagao, and H. Kano, “A novel nonvolatile memory with spin torque transfer magnetization switching: spin-ram,” *IEEE International Electron Devices Meeting*, pp. 459–462, Dec 2005.

[46] X. Yao, H. Meng, Y. Zhang, and J.-P. Wang, “Improved current switching symmetry of magnetic tunneling junction and giant magnetoresistance devices with nano-current-channel structure,” *Journal of Applied Physics*, vol. 103, no. 7, 2008.

[47] K. Yamada, N. Sakai, Y. Ishizuka, and K. Mameno, “A novel sensing scheme for an MRAM with a 5% MR ratio,” *Symposium on VLSI Circuits, Digest of Technical Papers*, pp. 123–124, Jun 2001.

[48] G. Jeong, C. Wooyoung, S. Ahn, J. Hongsik, G. Koh, H. Youngnam, and K. Kim, “A 0.24 μm 2.0-V 1T1MTJ 16-kb nonvolatile magnetoresistance RAM with self-reference sensing scheme,” *IEEE Journal of Solid-State Circuits,*, vol. 38, no. 11, pp. 1906–1910, Nov 2003.

[49] 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, Feb 2012.

[50] 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, pp. 165–209, 2007.

[51] J. Z. Sun, “Spin-current interaction with a monodomain magnetic body: A model study,” *Phys. Rev. B*, vol. 62, pp. 570–578, Jul 2000.

[52] R. H. Koch, J. A. Katine, and J. Z. Sun, “Time-Resolved Reversal of Spin-Transfer Switching in a Nanomagnet,” *Phys. Rev. Lett.*, vol. 92, Feb 2004.

[53] A. Vatankhahghadim, S. Huda, and A. Sheikholeslami, “A Survey on Circuit Modeling of Spin-Transfer-Torque Magnetic Tunnel Junctions,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 9, pp. 2634–2643, Sept 2014.

[54] D.D.Tang and Y. Lee, “Magnetic Memory: Fundamentals and Technology,” *Cambridge, U.K.: Cambridge University Press*, 2010.

[55] Y. Zhang, W. Zhao, Y. Lakys, J. O. Klein, J. V. Kim, D. Ravelosona, and C. Chappert, “Compact Modeling of Perpendicular-Anisotropy CoFeB/MgO Magnetic Tunnel Junctions,” *IEEE Transactions on Electron Devices*, vol. 59, no. 3, pp. 819–826, March 2012.

[56] M. Gajek, J. J. Nowak, J. Z. Sun, P. L. Trouilloud, E. J. OSullivan, D. W. Abraham, M. C. Gaidis, G. Hu, S. Brown, Y. Zhu, R. P. Robertazzi, W. J. Gallagher, and D. C. Worledge, “Spin torque switching of 20nm magnetic tunnel junctions with perpendicular anisotropy,” *Applied Physics Letters*, vol. 100, no. 13, 2012.

[57] S. Ikeda, K. Miura, H. Yamamoto, K. Mizunuma, H. D. Gan, M. Endo, S. Kanai, J. Hayakawa, F. Matsukura, and H. Ohno, “A perpendicular anisotropy CoFeB-MgO magnetic tunnel junction,” *Nature materials*, vol. 9, pp. 721–724, July 2010.

[58] W. F. Brinkman, R. C. Dynes, and J. M. Rowell, “Tunneling Conductance of Asymmetrical Barriers,” *Journal of Applied Physics*, vol. 41, no. 5, pp. 1915–1921, 1970.

[59] D. C. Worledge, G. Hu, D. W. Abraham, J. Z. Sun, P. L. Trouilloud, J. Nowak, S. Brown, M. C. Gaidis, E. J. OSullivan, and R. P. Robertazzi, “Spin torque switching of perpendicular Ta/CoFeB/MgO-based magnetic tunnel junctions,” *Applied Physics Letters*, vol. 98, no. 2, 2011.

[60] R. Dorrance, F. Ren, Y. Toriyama, A. Hafez, C. K. K. Yang, and D. Markovic, “Scalability and Design-Space Analysis of a 1T-1MTJ Memory Cell for STT-RAMs,” *IEEE Transactions on Electron Devices*, vol. 59, no. 4, pp. 878–887, 2012.

[61] D. Abraham, P. Trouilloud, and D. Worledge, “Rapid-turnaround characterization methods for MRAM development,” *IBM Journal of Research and Development*, vol. 50, pp. 55 – 67, Feb 2006.

[62] K. Lenz, H. Wende, W. Kuch, K. Baberschke, K. Nagy, and A. Jánossy, “Two-magnon scattering and viscous Gilbert damping in ultrathin ferromagnets,” *Phys. Rev. B*, vol. 73, p. 144424, Apr 2006.

[63] W. Kang, L. Zhang, J.-O. Klein, Y. Zhang, D. Ravelosona Ramasitera, and W. ZHAO, "Reconfigurable Codesign of STT-MRAM Under Process Variations in Deeply Scaled Technology," *Electron Devices, IEEE Transactions on*, vol. 62, Mar 2015.

[64] D. Kjaer, O. Hansen, H. Henrichsen, J. Chenchen, K. Noergaard, P. Nielsen, and D. Petersen, "Fast static field CIPT mapping of unpatterned MRAM film stacks," *Measurement Science and Technology*, vol. 26, Apr 2015.

[65] X. Zhu, Y. Lu, C. Park, and S. H. Kang, "Decoupling of source line layout from access transistor contact placement in a magnetic tunnel junction (MTJ) memory bit cell to facilitate reduced contact resistance," in *U.S. Patent 9 721 634 B2*, Jan 2017.

[66] N. Weste and D. Harris, "CMOS VLSI Design," *4th Edn*, 2011.

[67] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, Oct 1989.

[68] J. Lohstroh, "Static and dynamic noise margins of logic circuits," *IEEE Journal of Solid-State Circuits*, vol. 14, no. 3, pp. 591–598, June 1979.

[69] M. Sharifkhani and M. Sachdev, "SRAM cell stability: A dynamic perspective," *Solid-State Circuits, IEEE Journal of*, vol. 44, pp. 609 – 619, Mar 2009.

[70] Z. Zeng, P. Amiri, G. Rowlands, H. Zhao, I. Krivorotov, J.-P. Wang, J. Katine, J. Langer, K. Galatsis, K. Wang, and H. Jiang, "Effect of resistance-area product on spin-transfer switching in MgO-based magnetic tunnel junction memory cells," *Applied Physics Letters*, vol. 98, pp. 072 512–072 512, Feb 2011.

[71] E. I. Vatajelu, P. Prinetto, M. Taouil, and S. Hamdioui, "Challenges and Solutions in Emerging Memory Testing," *IEEE Transactions on Emerging Topics in Computing*, vol. 7, no. 3, pp. 493–506, 2019.

[72] L. Wei, J. Alzate, U. Arslan, J. Brockman, N. Das, K. Fischer, T. Ghani, O. Golonzka, P. Hentges, R. Jahan, P. Jain, B. Lin, M. Meterelliyoz, J. OrDonnell, C. Puls, P. Quintero, T. Sahu, M. Sekhar, A. Vangapaty, and F. Hamzaoglu, "A 7Mb STT-MRAM in 22FFL FinFET Technology with 4ns Read Sensing Time at 0.9V Using Write-Verify-Write Scheme and Offset-Cancellation Sensing Technique," *2019 IEEE International Solid- State Circuits Conference - (ISSCC)*, pp. 214–216, Feb 2019.

[73] H. Lv, D. C. Leitao, Z. Hou, P. P. Freitas, S. Cardoso, T. Kampfe, J. Muller, J. Langer, and J. Wrona, "Barrier breakdown mechanism in nano-scale perpendicular magnetic tunnel junctions with ultrathin MgO barrier," in *AIP Advances*, 2017.

[74] S. Nair, R. Bishnoi, M. Tahoori, G. Tshagharyan, H. Grigoryan, G. Harutyunyan, and Y. Zorian, "Defect injection, Fault Modeling and Test Algorithm Generation Methodology for STT-MRAM," *2018 IEEE International Test Conference (ITC)*, pp. 1–10, Oct 2018.

[75] S. Cosemans, W. Dehaene, and F. Catthoor, "A 3.6 pJ/access 480 MHz, 128 kb on-chip SRAM with 850 MHz boost mode in 90 nm CMOS with tunable sense amplifiers," *Solid-State Circuits, IEEE Journal of*, vol. 44, pp. 2065 – 2077, Aug 2009.

[76] B. Oliver, G. Tuttle, Q. He, X. Tang, and J. Nowak, "Two breakdown mechanisms in ultrathin alumina barrier magnetic tunnel junctions," *Journal of Applied Physics*, vol. 95, pp. 1315–1322, Feb 2004.

[77] 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*, pp. 994–1007, Jul 2012.

[78] C. Kim, K. Kwon, C. Park, S. Jang, and J. Choi, "A covalent-bonded cross-coupled current-mode sense amplifier for STT-MRAM with 1T1MTJ common source-line structure array," in *IEEE International Solid- State Circuits Conference - (ISSCC)*, Feb 2015.

[79] J. Kim, K. Ryu, S. H. Kang, and S.-O. Jung, "A Novel Sensing Circuit for Deep Submicron Spin Transfer Torque MRAM (STT-MRAM)," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 20, no. 1, pp. 181–186, Jan 2012.

[80] K. Ono, T. Kawahara, R. Takemura, K. Miura, H. Yamamoto, M. Yamanouchi, J. Hayakawa, K. Ito, H. Takahashi, S. Ikeda, H. Hasegawa, H. Matsuoka, and H. Ohno, "A disturbance-free read scheme and a compact stochastic-spin-dynamics-based MTJ circuit model for Gb-scale SPRAM," in *IEEE International Electron Devices Meeting (IEDM)*, Dec 2009.

[81] H. Koike and T. Endoh, "A new sensing scheme with high signal margin suitable for Spin-Transfer Torque RAM," in *International Symposium on VLSI Technology, Systems and Applications (VLSI-TSA)*, April 2011.

[82] H. Noguchi, K. Ikegami, K. Kushida, K. Abe, S. Itai, S. Takaya, N. Shimomura, J. Ito, A. Kawasumi, H. Hara, and S. Fujita, "A 3.3ns-access-time 71.2 μ W/MHz 1Mb embedded STT-MRAM using physically eliminated read-disturb scheme and normally-off memory architecture," in *IEEE International Solid-State Circuits Conference - (ISSCC)*, Feb 2015.

[83] H. Noguchi, K. Ikegami, N. Shimomura, T. Tetsufumi, J. Ito, and S. Fujita, "Highly reliable and low-power nonvolatile cache memory with advanced perpendicular STT-MRAM for high-performance CPU," in *Symposium on VLSI Circuits Digest of Technical Papers,*, June 2014.

[84] T. Na, J. Kim, J. P. Kim, S. Kang, and S.-O. Jung, "An Offset-Canceling Triple-Stage Sensing Circuit for Deep Submicrometer STT-RAM," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,*, vol. 22, no. 7, pp. 1620–1624, July 2014.

[85] S. Fujita, H. Noguchi, K. Ikegami, S. Takeda, K. Nomura, and K. Abe, "Technology Trends and Near-Future Applications of Embedded STT-MRAM," *IEEE International Memory Workshop (IMW),*, May 2015.

[86] S. Datta, "The non-equilibrium Green's function (NEGF) formalism: An elementary introduction," *International Electron Devices Meeting*, pp. 703–706, Dec 2002.

[87] T. Kawahara, R. Takemura, K. Miura, J. Hayakawa, S. Ikeda, Y. Lee, R. Sasaki, Y. Goto, K. Ito, T. Meguro, F. Matsukura, H. Takahashi, H. Matsuoka, and H. Ohno, "2Mb Spin-Transfer Torque RAM (SPRAM) with Bit-by-Bit Bidirectional Current Write and Parallelizing-Direction Current Read," *IEEE International Solid-State Circuits Conference, ISSCC 2007. Digest of Technical Papers.*, pp. 480–617, Feb 2007.

[88] Y. Huai, "Spin-transfer torque MRAM (STT-MRAM): Challenges and prospects," *AAPPS Bulletin*, vol. 18, 2008.

[89] A. V. Khvalkovskiy, D. Apalkov, S. Watts, R. Chepulskii, R. S. Beach, A. Ong, X. Tang, A. Driskill-Smith, W. H. Butler, P. B. Visscher, D. Lottis, E. Chen, V. Nikitin, and M. Krounbi, "Basics principles of the STT-MRAM cell operation in memory arrays," in *Journal of Physics D: Applied Physics*, 2013.

[90] W. Kang, L. Zhang, W. ZHAO, J.-O. Klein, Y. Zhang, D. Ravelosona Ramasitera, and C. Chappert, "Yield and Reliability Improvement Techniques for Emerging Non-volatile STT-MRAM," *on emerging and selected topics in circuits and systems*, vol. 1, pp. 1–13, Dec 2014.

[91] M. Wang, W. Cai, K. Cao, J. Zhou, J. Wrona, S. Peng, H. Yang, J. Wei, W. Kang, Y. Zhang, J. Langer, B. Ocker, A. Fert, and W. ZHAO, "Current-induced magnetization switching in atom-thick tungsten engineered perpendicular magnetic tunnel junctions with large tunnel magnetoresistance," *Nature Communications*, vol. 9, Aug 2017.

[92] Y. Lee, Y. Song, J. Kim, S. Oh, B.-J. Bae, S. Lee, J. Lee, U. Pi, B. Seo, H. Jung, K. Lee, H. Shin, H. Jung, M. Pyo, A. Antonyan, D. Lee, S. Hwang, D. Jang, Y. Ji, and E. Jung, "Embedded STT-MRAM in 28-nm FDSOI Logic Process for Industrial MCU/IoT Application," *2018 IEEE Symposium on VLSI Technology*, pp. 181–182, Jun 2018.

[93] A. Driskill-Smith, D. Apalkov, V. Nikitin, X. Tang, S. Watts, D. Lottis, K. Moon, A. Khvalkovskiy, R. Kawakami, X. Luo, A. Ong, and E. Chen, "Latest Advances and Roadmap for In-Plane and Perpendicular STT-RAM," *2011 3rd IEEE International Memory Workshop, IMW 2011*, May 2011.

[94] Y. Lu, T. Zhong, W. Hsu, S. Kim, X. Lu, J. Kan, C. Park, W.-C. Chen, X. Li, X. Zhu, P. Wang, M. Gottwald, J. Fatehi, L. Seward, J. Kim, N. Yu, G. Jan, J. Haq, S. Le, and S. Kang, "Fully functional perpendicular STT-MRAM macro embedded in 40 nm logic for energy-efficient IOT applications," *2015 IEEE International Electron Devices Meeting (IEDM)*, pp. 26.1.1–26.1.4, Dec 2015.

[95] S. Van Beek, K. Martens, P. Roussel, G. Donadio, J. Swerts, S. Mertens, G. Kar, T. Min, and G. Groeseneken, "Four point probe ramped voltage stress as an efficient method to understand breakdown of STT-MRAM MgO tunnel junctions," *IEEE International Reliability Physics Symposium Proceedings*, vol. 2015, pp. MY41–MY46, May 2015.

[96] K. C. Chun, H. Zhao, J. Harms, T.-H. Kim, J.-p. Wang, and C. Kim, "A Scaling Roadmap and Performance Evaluation of In-Plane and Perpendicular MTJ Based STT-MRAMs for High-Density Cache Memory," *Solid-State Circuits, IEEE Journal of*, vol. 48, pp. 598–610, Feb 2013.

[97] H. Noguchi, K. Ikegami, K. Kushida, K. Abe, S. Itai, S. Takaya, N. Shimomura, J. Ito, A. Kawasumi, H. Hara, and S. Fujita, *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, pp. 1–3, Feb 2015.

[98] Y. Wang, H. Cai, L. A. d. B. Naviner, Y. Zhang, X. Zhao, E. Deng, J. Klein, and W. Zhao, "Compact Model of Dielectric Breakdown in Spin-Transfer Torque Magnetic

Tunnel Junction," *IEEE Transactions on Electron Devices*, vol. 63, no. 4, pp. 1762–1767, April 2016.

[99] 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," *J. Phys.: Condens. Matter*, vol. 19, pp. 165 209–13, Apr 2007.

This chapter provides the summary of all the publications from this work accepted and pending.

[1] G. Radhakrishnan, Y. Yoon and M. Sachdev, "A Parametric DFT Scheme for STT-MRAMs," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 27, no. 7, pp. 1685-1696, July 2019. DOI: 10.1109/TVLSI.2019.2907549

[2] G. Radhakrishnan, Y. Yoon and M. Sachdev, "A DFT Scheme for Fault Monitoring in STT-MRAMs" poster session in International Test conference, Nov 2019.

http://www.itctestweek.org/itc-2019-posters/

[3] G. Radhakrishnan, Y. Yoon and M. Sachdev, "Monitoring Aging Defects in STT-MRAMs" IEEE transactions on CAD integrated circuits and systems, accepted pending press. DOI: 10.1109/TCAD.2020.2982145

[4] G. Radhakrishnan, Y. Yoon and M. Sachdev,"Accelerating STT-MRAM Ramp-up Characterization" IEEE NEWCAS 2020 conference, accepted pending press.