

UCLA

UCLA Electronic Theses and Dissertations

Title

Integration of Voltage-Controlled Spintronic Devices in CMOS Circuits

Permalink

<https://escholarship.org/uc/item/0cx596qk>

Author

Lee, Hochul

Publication Date

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Integration of Voltage-Controlled Spintronic Devices in CMOS Circuits

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Electrical Engineering

by

Hochul Lee

2017

© Copyright by

Hochul Lee

2017

ABSTRACT OF THE DISSERTATION

Integration of Voltage-Controlled Spintronic Devices in CMOS Circuits

by

Hochul Lee

Doctor of Philosophy in Electrical Engineering

University of California, Los Angeles, 2017

Professor Kang Lung Wang, Chair

Spintronics is an emerging field that studies the properties of electron spin and discovers the methods to detect and manipulate its associated magnetic moment in a solid-state device, in addition to its fundamental electronic charge. Utilization of spintronic devices has been considered as a possible alternative for beyond CMOS technology. One of the most promising spintronic devices is a magnetic tunnel junction (MTJ) that has attracted the attention of academia and industry owing to its remarkable characteristics such as non-volatility, virtually unlimited endurance, and CMOS compatibility. Also, due to the discovery of the spin-transfer torque (STT) and spin Hall effect (SHE) as new switching mechanisms, a nanosecond switching speed has been demonstrated in MTJ devices. However, these current-driven switching methods inherently cause a significant ohmic loss since they require relatively a large amount current to generate sufficient spin torque. Recently, a voltage-controlled effect has been utilized to mitigate the energy issue by drastically reducing ohmic dissipation during

switching in a noble memory architecture called magnetoelectric RAM (MeRAM). In addition to achieving high-energy efficiency, voltage-induced switching leads to further improvement in terms of density and switching speed, opening the door to new possibilities of next generation low-power and high-speed system architectures.

In this dissertation, we explore the characteristics of voltage-controlled magnetic anisotropy (VCMA) effect driven precessional switching based on an MTJ macrospin compact model including the VCMA effect in its built-in Landau-Lifshitz-Gilbert (LLG) equation. In particular, this compact model allows predicting required bias conditions for switching, monitoring the three-dimensional magnetization dynamics, and extracting the write error rate (WER). Furthermore, we demonstrate a wide variety of spintronics-CMOS circuits utilizing unique features of voltage-controlled MTJ for many applications. Overall, the performances of the proposed circuits are improved by an order of magnitude, especially, in terms of energy and area. Also, we develop several practical design techniques to improve the reliability of the read and write operations in MeRAM. Lastly, a synchronous 4Kbit MeRAM macro is designed based on IBM 130 nm technology. After discussing the MeRAM macro specification and constraints, each circuit component of the macro and its verification results are presented.

The dissertation of Hochul Lee is approved.

Yong Chen

Puneet Gupta

Dejan Markovic

Kang Lung Wang, Committee Chair

University of California, Los Angeles

2017

To my family

This dissertation is specially dedicated to my beloved wife *Julie Yom*

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
1.1 Overview of Spintronics	4
1.2 Magnetic Tunnel Junction (MTJ)	7
1.2.1 MTJ Device Structure	7
1.2.1.1 In-plane MTJ	8
1.2.1.2 Perpendicular MTJ	9
1.2.2 Tunneling Magnetoresistance (TMR) Ratio	11
1.3 Magnetoresistive Random-Access Memory (MRAM)	15
1.3.1 Oersted-field-switched MRAM	15
1.3.2 Spin-Transfer Torque MRAM (STT-MRAM)	16
1.3.3 Magnetoelectric RAM (MeRAM)	19
1.3.4 Spin-Orbit Torque MRAM (SOT-MRAM)	20
1.3.5 Performance Comparison	22
1.3.5.1 Device Level	22
1.3.5.2 Array Level	26
CHAPTER 2: MACROSPIN COMPACT MODELING OF MTJ	33
2.1 Motivation	33
2.2 Physical and Dynamic Model of MTJ	34

2.2.1 Landau–Lifshitz–Gilbert (LLG) equation	34
2.2.2 Effective Magnetic Field	34
2.2.2.1 Perpendicular Magnetic Anisotropy (PMA)	35
2.2.2.2 Voltage-Controlled Magnetic Anisotropy (VCMA)	35
2.2.2.3 Shape Anisotropy and Demagnetization	36
2.2.2.4 External Field	36
2.2.3 Thermal Noise	37
2.2.4 Modeling of Spin-Transfer Torque	38
2.2.5 Modeling of Tunnel Magnetoresistance (TMR) Ratio	39
2.3 Two Terminal MTJ Switching Mechanisms and Compact Model Simulations	40
2.3.1 Voltage Dependence of Effective Magnetic Field	40
2.3.2 Timing of Voltage-Driven Precessional Switching	41
2.3.3 Thermal Noise Effect	43
2.3.4 External Magnetic Field Dependence of Switching Speed	44
2.3.5 Write Error Rate (WER)	45
2.3.6 Thermal Stability and Retention Time	48
2.4 Three Terminal MTJ Switching Mechanisms and Compact Model Simulations	49
2.4.1 Modeling of Spin Hall Effect	49
2.4.2 SHE-driven Switching and VCMA Assisted SHE-driven Switching	52

2.5 Scalability of Voltage-Controlled MTJ	54
<u>CHAPTER 3: VOLTAGE-CONTROLLED SPINTRONICS-CMOS CIRCUITS</u>	56
3.1 Introduction	56
3.2 Voltage-controlled MTJ based Ternary Content-Addressable Memory	56
3.2.1 Overview of TCAM	56
3.2.2 Design of MTJ based TCAM Cell	57
3.2.3 Configuration and Search Operations	59
3.2.4 Performance Evaluation and Comparison	61
3.3 A Spintronic Voltage-Controlled Stochastic Oscillator	62
3.3.1 Uniform Sampling versus Non-uniform Sampling	62
3.3.2 Advantages of using MTJ for the Non-uniform Clock Generator	64
3.3.3 Design and Performance Evaluation	66
3.4 Voltage-controlled MTJ based True Random Number Generator	70
3.4.1 Overview of Random Number Generator	70
3.4.2 Advantages of using MTJ for the True Random Number Generator	71
3.4.3 Design and Performance Evaluation	73
3.5 Spintronic Programmable Logic (SPL) using Voltage-gated Spin Hall Effect	77
3.5.1 Overview of Programmable Logic	77
3.5.2 Advantages of using three terminal MTJ for the SPL	78

3.5.3 Configuration and Logic Operations of the SPL	78
3.5.4 Performance Evaluation: Sensing Margin, Power Consumption, and Area	79
3.6 Analog to Stochastic Bit Stream Converter	83
3.6.1 Overview of Analog to Stochastic Bit Stream Converter	83
3.6.2 Switching Probability of Voltage-Assisted spin Hall effect	85
3.6.3 Design and Performance Evaluation	87
3.7 Spin and CMOS based Neural Network	90
3.7.1 Motivation	90
3.7.2 Design and Performance Evaluation	91
<u>CHAPTER 4: DESIGN TECHNIQUES ENHANCING MeRAM PERFORMANCE</u>	96
4.1 Motivation	96
4.2 Source Line Sensing (SLS) Scheme to Improve Read Performance	96
4.2.1 Use of VCMA with reverse voltage	96
4.2.2 Circuit Architecture of the SLS	98
4.2.3 Evaluation	102
4.3 Word Line Pulse (WLP) based Write Operation for Reduced Write Error Rate	104
4.3.1 Pulse Shape Dependence of Magnetization Behavior	104
4.3.2 Timing of the WLP	105
4.3.3 Performance Evaluation: Write Error Rate and Cell Area Efficiency	109

4.4 Write Pulse Termination (WPT) Circuit Technique	112
4.4.1 Motivation	112
4.4.2 Schematic of the WPT	113
4.4.3 Simulation and Analysis	116
4.5 Circuit for Controlling Non-deterministic Switching	119
4.5.1 Motivation	119
4.5.2 Schematic and Simulation of the pre-read and write sense amplifier (PWSA)	112
4.5.3 Performance Evaluation	124
<u>CHAPTER 5: 4KBIT MeRAM MACRO DESIGN</u>	<u>127</u>
5.1 Specification	127
5.2 Cell and Array Design	129
5.3 Core Circuit Design	131
5.4 Peripheral Circuit Design	135
5.5 Full-chip Analog and Digital Mixed Signal Verification	140
5.6 Full-chip Layout	144
5.7 Test and Evaluation	145
<u>CHAPTER 6: CONCLUSION</u>	<u>150</u>
<u>REFERENCE</u>	<u>153</u>

LIST OF FIGURES

Fig. 1.1 Trends of power consumption and dark silicon for SOC	2
Fig. 1.2 State of the art computing system architecture	3
Fig. 1.3 Emerging system architecture with the embedded non-volatile memory	4
Fig. 1.4 In-plane MTJ utilizing the shape anisotropy	8
Fig. 1.5 Perpendicular MTJ utilizing the interfacial perpendicular anisotropy	10
Fig. 1.6 Spin-split density of states (DOS) for non-magnetic and magnetic metals	12
Fig. 1.7 Two channel Jullière model for the tunneling between two ferromagnetic layers	13
Fig. 1.8 Majority and minority DOS for the Fe Fermi energy Block states	14
Fig. 1.9 Oersted-field-switched MRAM cell structure	16
Fig. 1.10 1T-1MTJ cell architecture in STT-MRAM	17
Fig. 1.11 Amplitude of write current as a function of write pulse width for STT-driven switching	18
Fig. 1.12 1T-1MTJ cell architecture in MeRAM	19
Fig. 1.13 1T-1MTJ cell architecture in SOT-MRAM	21
Fig. 1.14 Memory Hierarchy in a conventional computer architecture	23
Fig. 1.15 Performance comparison of different memory technologies	25
Fig. 1.16 Memory bank architecture and Schematic of a single column connection of MeRAM	27
Fig. 1.17 Array level performance comparisons based on the same capacity 256 Kbit	30
Fig. 1.18 Array level performance comparisons based on the same array area $200 \mu\text{m} \times 200 \mu\text{m}$	31
Fig. 2.1 Voltage dependence of the components of the effective magnetic field	41
Fig. 2.2 Illustration of the voltage-induced precessional switching mechanism	42
Fig. 2.3 Transient simulation of the non-deterministic voltage-driven precessional switching	43
Fig. 2.4 Transient simulation in the absence/presence of the thermal noise	44

Fig. 2.5 Transient simulations for the in-plane H_{ext} dependence of the switching speed	45
Fig. 2.6 WER of STT-induced switching and voltage-controlled precessional switching	46
Fig. 2.7 Compact model transient simulations for a square/triangular shape write pulse	47
Fig. 2.8 WER simulation as a function of slew rate and amplitude for precessional switching	48
Fig. 2.9 Voltage dependence of energy barrier and thermal stability via the VCMA effect	49
Fig. 2.10 Schematic of the 3-terminal MTJ on the heavy metal layer with spin-orbit interaction	50
Fig. 2.11 Simulation of pure SHE switching and gate-voltage-modulated SHE switching	52
Fig. 2.12 Switching probability as a function of current amplitude and applied voltage	53
Fig. 2.13 Required VCMA coefficient and interfacial anisotropy for device scaling	54
Fig. 3.1 Conventional SRAM based TCAM cell architecture	57
Fig. 3.2 Voltage-controlled MTJ based MeTCAM cell architecture	58
Fig. 3.3 Voltage dependence of the coercivity	59
Fig. 3.4 Search operations of the MeTCAM	60
Fig. 3.5 Uniform/Non-uniform sampling and Non-uniform clock based system	63
Fig. 3.6 Measured coercivity as a function of the voltage across the device	65
Fig. 3.7 Measured MTJ's resistance fluctuation and thermal stability	66
Fig. 3.8 MTJ based voltage-controlled stochastic oscillator (VCSO)	67
Fig. 3.9 Transient circuit simulation of the VCSO with the MTJ compact model	68
Fig. 3.10 Average sampling frequency of the VCSO with different thermal stabilities	69
Fig. 3.11 Description of VCMA-induced switching mechanisms in a perpendicular MTJ	71
Fig. 3.12 Simulated VCMA-induced switching probability as a function of write pulse width	72
Fig. 3.13 Schematic of the proposed MTJ based true random number generator (MRNG)	73
Fig. 3.14 Simulation result of the MRNG for a random bit operation	74

Fig. 3.15 Schematic of the $m \times n$ MTJs array based multi-bit MRNG	75
Fig. 3.16 Simulation results of consecutive random bit generations by using the MRNG	75
Fig. 3.17 Switching probability of MTJ in the presence of z-direction external field	76
Fig. 3.18 Schematic of the proposed 2-input spintronic programmable logic (SPL)	79
Fig. 3.19 Write and logic operation of the proposed 2-input SPL	80
Fig. 3.20 Simulation of the sensing margin during the logic operation	81
Fig. 3.21 Power consumption of the SPL at write, logic, and stand-by modes	82
Fig. 3.22 Area comparison of different types of LUT	82
Fig. 3.23 CMOS based and MTJ based analog to stochastic bit stream converter (ASC)	84
Fig. 3.24 Critical current I_{C_SHE} of the SHE effect as a function of voltage across the MTJ	85
Fig. 3.25 Switching probability as a function of voltage across the MTJ	86
Fig. 3.26 Trajectory of the free layer's magnetization during the voltage assisted SHE switching	87
Fig. 3.27 Schematic of the proposed spintronic ASC	88
Fig. 3.28 Transient circuit simulation of the spintronic ASC	89
Fig. 3.29 Spin-based artificial neuron	91
Fig. 3.30 Schematic of CMOS and MTJ based artificial dendrite and synapse	92
Fig. 3.31 Structure of the spin-based neural network	93
Fig. 3.32 Circuit components of a CMOS based Axon	94
Fig. 3.33 Simulation result of CMOS based Axon operation	95
Fig. 4.1 One transistor and one MTJ of MeRAM cell structure	97
Fig. 4.2 Conventional write and read operations of MeRAM	98
Fig. 4.3 Proposed core circuit architecture for implementing the SLS	100
Fig. 4.4 Simulation results of the BLS and the SLS based on the MTJ compact model	101

Fig. 4.5 Measured thermal stability and retention time of an MTJ as a function of applied voltage	102
Fig. 4.6 Read disturbance, sensing margin, and TMR as a function of sensing voltage	103
Fig. 4.7 Schematic of cell array architecture, including the BL driver and WL driver	106
Fig. 4.8 Timing of conventional BLP scheme and proposed WLP scheme	107
Fig. 4.9 Circuit simulation results of the BLP and the WLP	109
Fig. 4.10 Write error rates of the BLP and the WLP with respect to the capacitive loading	110
Fig. 4.11 Required pulse shape that achieves an acceptable BER and normalized driver size	112
Fig. 4.12 Experimentally observed an oscillatory behavior of the switching probability	113
Fig. 4.13 Block diagram of the write pulse termination (WPT) circuit	114
Fig. 4.14 Simulation result of voltage divider	115
Fig. 4.15 Transistor-level schematic of the WPT circuit	117
Fig. 4.16 Circuit simulations of the WPT scheme (P state)	118
Fig. 4.17 Circuit simulations of the WPT scheme (AP state)	119
Fig. 4.18 Concept diagram of the proposed pre-read and write sense amplifier (PWSA)	112
Fig. 4.19 Proposed data program flow of the PWSA	121
Fig. 4.20 Transistor-level schematic of the PWSA	122
Fig. 4.21 Simulation of the PWSA operation (AP to P)	123
Fig. 4.22 Simulation of the PWSA operation (P to P)	124
Fig. 4.23 Simulations for the sensing margin	126
Fig. 5.1 MeRAM macro architecture	127
Fig. 5.2 1T-1MTJ structure based unit cell design for MeRAM	130
Fig. 5.3 Layout design of the main and test cell array	131
Fig. 5.4 Schematic of the core circuit	132

Fig. 5.5 Schematic and simulation of the pulse generator	133
Fig. 5.6 Simulation of the sense amplifier for the sensing	134
Fig. 5.7 Schematic and layout design of the peripheral circuit	135
Fig. 5.8 Verilog simulation of the test read and test write operations	137
Fig. 5.9 Verilog simulation of the peripheral circuit read operation	138
Fig. 5.10 Verilog simulation of the peripheral circuit program operation	139
Fig. 5.11 Architecture of ADMS simulation system	140
Fig. 5.12 ADMS simulation for the read operation	141
Fig. 5.13 Pulse generator enable signals corresponding to the comparison result of Table 5.6	142
Fig. 5.14 MTJ resistance change and core circuit control signals during the sensing and writing	143
Fig. 5.15 Layout of the MeRAM macro	144
Fig. 5.16 Picture of the fabricated MeRAM macro via 130 nm IBM RF-DM technology	146
Fig. 5.17 Test set up for the MeRAM macro	147
Fig. 5.18 Measured macro signals for the test write/read operations	148
Fig. 5.19 Measured macro signals for the read operation	149
Fig. 5.20 Measured macro signals for the program operation	149

LIST OF TABLES

Table 1.1 Performance comparison of emerging memory technologies in 1T-1R cell structure	24
Table 1.2 28 nm node technology parameters and device characteristics	28
Table 1.3 Memory technologies unit cell area based on 28 nm node	28
Table 1.4 Formulas for array level parameters	29
Table 2.1 Parameters of the macrospin compact model	40
Table 3.1 Truth table of TCAM cell including Don't care condition	57
Table 3.2 Bias conditions for the MeTCAM configuration mode	59
Table 3.3 Performance comparison of TCAMs associated with different types of memories	62
Table 3.4 Performance comparison with previous works	69
Table 3.5 Area and performance of 64×64 MTJs Array based multi-bit-MRNG (45 nm node)	76
Table 3.6 Performance comparison with previous works	89
Table 3.7 Area and performance of the proposed spin and CMOS based hybrid neuron	95
Table 4.1 Resistive and capacitive loads on the BL and the WL (28 nm node)	108
Table 5.1 Specification of 4Kbit MeRAM Macro	128
Table 5.2 Pin list and descriptions of the MeRAM macro	129
Table 5.3 Port list and descriptions of the core circuit	132
Table 5.4 Input port list and descriptions of the peripheral circuit	136
Table 5.5 Output port list and fan-out of the peripheral circuit	136
Table 5.6 New data and sensed data from the predefined cell patterns	142

ACKNOWLEDGEMENT

I have always believed that passion, persistence, practice, and self-motivation are key words that lead to success. Throughout my life, I have been driven by my passion and continued to practice until I was satisfied regardless of failure. As a result, I have progressively achieved many goals and accomplishments that I had dreamed of. However, as I reflect on my life, I come to realize that my accomplishments were not the result of myself alone but have been possible because of the people who surround me. These are my family, advisors, colleagues, and friends who have played a pivotal role in shaping my life. I feel sincerely grateful and fortunate to have such exceptional figures who have helped me achieve my goals throughout my educational and professional development. It is an honor to have an opportunity to thank them individually through the acknowledgment section of this dissertation.

First and foremost, I would like to thank my beautiful wife Julie Yom, who has provided me unconditional love and support. Because of her dedication throughout my Ph.D. program, I was able to better focus on my research and be productive. Although her profession is in a different discipline, she has shown a great interest in my research. She puts in time to read my journals and publications to understand my research and engages in constructive conversation with me on a daily basis. I would also like to thank my parents who taught me to have a positive outlook in life and encouraged me to pursue academic achievement. Additionally, I would like to express my gratitude to every member of my family.

It has been an honor to work with my advisor, Prof. Kang L. Wang, who is a remarkable educator and researcher. I am incredibly grateful for the opportunity to conduct my research in his group, the Device Research Laboratory (DRL), and learn from him not only about scientific knowledge and skills but also research ethics. His tenacious efforts to open doors to new ideas have inspired many young researchers and provided invaluable guidance. Owing to his support, I was able to develop unique engineering skills from the device level to the circuit design level for realizing emerging spintronic

based systems. In addition, I thank the committee members, Prof. Yong Chen, Prof. Puneet Gupta, and Dejan Markovic, for providing me with their valuable advice, time, and patience.

I would also like to thank my co-advisor, Prof. Pedram Khalili, who has led the Spintronics team of DRL since 2009. I feel fortunate to have learned about fundamental physics of voltage-controlled magnetism as well as communication and writing skills. Moreover, he has provided specific research directions and taught me how to manage my schedule for projects.

I have two colleagues to thank specifically. Albert Lee is a smart, passionate, and collaborative young researcher. Ever since joining the group, Albert and I have worked together and successfully completed numerous projects. I have been inspired by his problem-solving skills, excellent circuit design skills, and sincerity. Farbod Ebrahimi is a brilliant colleague who has supported me to setup experiment and simulation environments. I have enjoyed having countless productive discussions with both Albert and Farbod from brainstorming new ideas to realizing them. As a result of our synergetic efforts, our work has been published in several prestigious journals.

Furthermore, I am grateful for the opportunity to collaborate with Shaodi Wang in Prof. Gupta's group, Richard Dorrance and Sina Basir-Kazeruni in Prof. Marcovic's group, and Yilei Li in Prof. Chang's group. Last but not the least, many thanks to the members of the Spintronics team, Ayran Navabi, Xiang Li, Armin Razavi, Yuxiang Liu, Kin Wong, Guoqiang Yu, and all the members and alumni of the DRL. I wish each and every one of them successful careers, prosperity, and happiness.

VITA

Education

- 2005-2007 M.S. in Electrical Engineering
Seoul National University
Seoul, South Korea
- 2001-2005 B.S. in Electrical Engineering
Korea University
Seoul, South Korea

Employment history

- 2012-2017 Graduate Student Researcher
Department of Electrical Engineering
University of California, Los Angeles
Los Angeles, California
- 2015-2017 Circuit Design Engineer (Part-time)
Inston Inc.
Los Angeles, California
- 2007-2012 Circuit Design Engineer
Samsung Electronics Co., LTD.
Flash Memory Circuit Design Team
Hwaseong, South Korea

Publication

H. Lee, A. Lee, S. Wang, F. Ebrahimi, P. Khalili, and K. L. Wang, “Anaylsis and compact modeling of magnetic tunnel junctions utilizing voltage-controlled magnetic aninsotropy,” under review, IEEE Trans. Magn., pp. 1–1, 2015.

A. Lee, **H. Lee**, F. Ebrahimi, B. Lam, W. Chen, M. Chang, P. Khalili, and K. L. Wang, “A Dual Data Line Read Scheme for High-Speed Low-Energy Resistive Nonvolatile Memories,” under review (minor revision), IEEE Trans. Very Large Scale Integr. Syst. 2017.

H. Lee, A. Lee, F. Ebrahimi, P. Khalili, and K. L. Wang, “Analog to Stochastic Bit Stream Converter Utilizing Voltage-Assisted Spin Hall Effect,” IEEE Electron Device Lett., pp.1-1, 2017.

H. Lee, S. Wang, A. Lee, F. Ebrahimi, P. Gupta, P. Khalili, and K. L. Wang, “A Word Line Pulse Circuit Technique for Reliable Magnetoelectric Random-Access Memory,” IEEE Trans. Very Large Scale Integr. Syst., pp. 1–8, 2017.

A. Lee, C.-P. Lo, C.-C. Lin, W.-H. Chen, K.-H. Hsu, Z. Wang, F. Su, Z. Yuan, Q. Wei, Y.-C. King, C.-J. Lin, **H. Lee**, P. Khalili Amiri, K.-L. Wang, Y. Wang, H. Yang, Y. Liu, and M.-F. Chang, “A ReRAM-Based Nonvolatile Flip-Flop With Self-Write-Termination Scheme for Frequent-OFF Fast-Wake-Up Nonvolatile Processors,” *IEEE J. Solid-State Circuits*, pp. 1–14, 2017.

H. Lee, A. Lee, F. Ebrahimi, P. Khalili, and K. L. Wang, “Array-Level Analysis of Magnetoelectric Random Access Memory (MeRAM) for High-Performance Embedded Applications,” *IEEE Magn. Lett.*, pp. 1–1, 2017.

H. Lee, F. Ebrahimi, P. Khalili, and K. L. Wang, “Design of High-Throughput and Low-Power True Random Number Generator Utilizing Perpendicularly Magnetized Voltage-Controlled Magnetic Tunnel Junction,” *Cit. AIP Adv.*, vol. 7, no. 55934, 2017.

H. Lee, C. Grezes, A. Lee, F. Ebrahimi, P. Khalili Amiri, and K. L. Wang, “A Spintronic Voltage-Controlled Stochastic Oscillator for Event-Driven Random Sampling,” *IEEE Electron Device Lett.*, vol. 38, no. 2, pp. 281–284, Feb. 2017.

C. Grezes, **H. Lee**, A. Lee, S. Wang, F. Ebrahimi, X. Li, K. Wong, J. A. Katine, B. Ocker, J. Langer, P. Gupta, P. Khalili, and K. L. Wang, “Write Error Rate and Read Disturbance in Electric-Field-Controlled MRAM,” *IEEE Magn. Lett.*, pp. 1–1, 2016.

H. Lee, C. Grezes, S. Wang, F. Ebrahimi, P. Gupta, P. K. Amiri, and K. L. Wang, “Source Line Sensing in Magneto-Electric Random-Access Memory to Reduce Read Disturbance and Improve Sensing Margin,” *IEEE Magn. Lett.*, vol. 7, pp. 1–5, 2016.

H. Lee, F. Ebrahimi, P. K. Amiri, and K. L. Wang, “Low-Power and High-Density Spintronic Programmable Logic (SPL) Using Voltage-Gated Spin Hall Effect in Magnetic Tunnel Junctions,” *IEEE Magn. Lett.*, vol. PP, no. 99, pp. 1–1, 2016.

S. Wang, **H. Lee**, F. Ebrahimi, P. K. Amiri, K. L. Wang, and P. Gupta, “Comparative Evaluation of Spin-Transfer-Torque and Magnetoelectric Random Access Memory,” *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 6, no. 2, pp. 134–145, Jun. 2016.

P. Khalili Amiri, J. Alzate, X. Cai, F. Ebrahimi, Q. Hu, K. Wong, C. Grezes, **H. Lee**, G. Yu, X. Li, M. Akyol, Q. Shao, J. Katine, J. Langer, and B. Ocker, “Electric-Field-Controlled Magnetoelectric Random Access Memory: Progress, Challenges, and Scaling,” *IEEE Trans. Magn.*, pp. 1–1, 2015.

K. L. Wang, **H. Lee**, and P. Khalili Amiri, “Magnetoelectric Random Access Memory-Based Circuit Design by Using Voltage-Controlled Magnetic Anisotropy in Magnetic Tunnel Junctions,” *IEEE Trans. Nanotechnol.*, vol. 14, no. 6, pp. 992–997, Nov. 2015.

H. Lee, J. Alzate, R. Dorrance, X. Cai, D. Markovic, P. Khalili Amiri, and K. Wang, “Design of a Fast and Low-Power Sense Amplifier and Writing Circuit for High-Speed MRAM,” *IEEE Trans. Magn.*, vol. PP, no. 99, pp. 1–1, 2014.

CHAPTER 1

INTRODUCTION

We stand on a transition point between the third and the fourth industrial revolution where emerging technologies will dramatically change the way we live. One of the key characteristics of the fourth is that independently developed technologies start to be integrated into a single mobile device, and billions of these devices are connected to each other, creating new functionalities that humankind have never experienced. Also, the advent of artificial intelligence (AI), automation (autonomous), virtual reality, and 3-D printing makes the functionalities more sophisticated and comprehensive. However, all these advanced technologies create a large amount of data via their distributed sensors, and such big data should be transferred, processed and stored in a fast and efficient way, allowing systems to function properly. Hence, new electronic computing systems are required to have unprecedented performances in terms of throughput, power, capacity, and area. Furthermore, these demands are growing at an exponential rather than a linear pace.

For the past few decades, the performances of computing systems have been enhanced by independently improving the performance of processors and separated memory layers within Von Neumann architectures via scaling of CMOS technologies. During the third industrial revolution or the digital revolution, the scaling strategies have successfully fulfilled the growing demand for computation performances under the Moore's Law: the number of transistors in a dense integrated circuit doubles approximately every two years [1], [2]. As shrink the size of transistors, the response time of transistors and the energy consumption for computing are reduced while the cost of fabrications effectively decreases by producing more chips on a wafer. However, improvement through scaling has saturated due to several reasons. First, the channel length of transistor approach a physical limit where a transistor cannot perform as an electrical switch due to its exponentially growing leakage current [3], as shown in Fig. 1.1(a). Second, such leakage current causes a significant static power consumption,

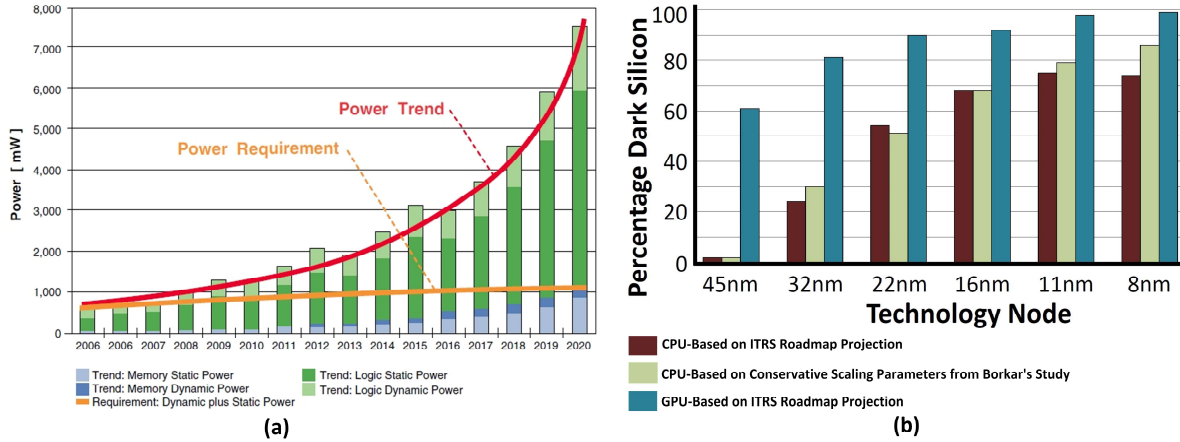


Fig. 1.1 (a) Power consumption trends for SOC. Memory and logic static power grow exponentially [3]. (b) Trend of dark silicon for different technology nodes. A significant portion of the total number of transistors in a chip needs to be turned off due to a thermal design power constraint, especially, in advanced nodes [4].

limiting low-power applications. More importantly, it becomes difficult to increase the clock frequency (throughput) due to dark silicon problem. In such a highly integrated chip, only part of the chip can be turned on at any given time, referred to dark silicon, due to a thermal design power (TDP) constraint. The percentage of dark silicon continues to increase as technology nodes are advanced [4], as shown in Fig. 1.1(b) Moreover, at the architecture level, the bandwidth of the system bus between the working memory (e.g., DRAM) and the processor cannot meet the demands of recent applications [5], which is known as Von Neumann bottleneck.

As a result, the semiconductor industry has begun to incorporate new system architectures for further performance improvements. For example, high bandwidth memory (HBM) and hybrid memory cube (HMC) techniques shorten the physical distance between working memories and processors and increase the number of channels by putting them together into a single package, as shown in Fig. 1.2. Although this approach has achieved hundreds of GB/s bandwidth in stacked 8GB DRAMs [6], the latency of signal transmissions via the interposer, ranging from 50 to 200 cycles depending on the network topology [7], is still longer than that of on-chip data transfer (4~60 cycles).

Another way of improving the throughput of a system is increasing the capacity of the processor's on-chip cache memory to reduce the cache miss rate. When the processor does not find needed data in

the cache, it allocates a memory space and fetches the required data from the main memory. This process typically takes hundreds of system clock cycles, thus acting as a bottleneck in terms of throughput. State-of-the-art processors have a few tens of MB L3 cache. However, the large area overhead of SRAM prohibits further increasing the cache capacity.

Regarding the power consumption issues, researchers have developed many standby power reduction schemes such as multiple power domains, reduced frequency, and body biasing. Other than a conventional electric power supply, additional energy sources have been implemented into a chip; for example, energy harvesting via vibration, temperature differentials, and light, to alleviate power related issues [8]. While these approaches are effective to some extent, they cannot completely eliminate leakage and often incur additional overheads.

All these approaches have pros and cons, motivating the need for a game-changing alternative that improves the performance of systems further without a trade-off. The proposed solutions are as follows. At the architecture level, integrating a high-density non-volatile memory device into a processor (non-volatile embedded system memory) has the potential to significantly improve throughput and energy efficiency, while reducing chip area and cost [9]. The improvement comes as a result of the following:

(i) On-chip data transfers are faster by one order of magnitude and more energy-efficient by a factor

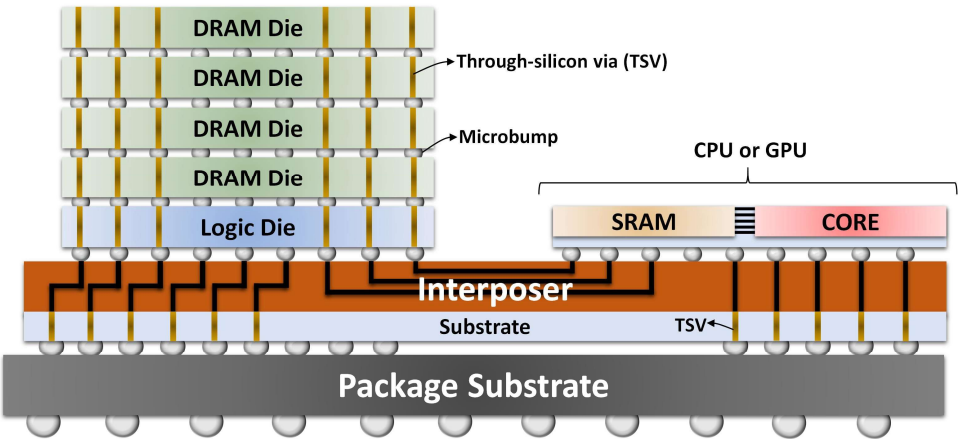


Fig. 1.2 State of the art computing system architecture where the main memory and the processor are connected via the interposer in a packaged chip.

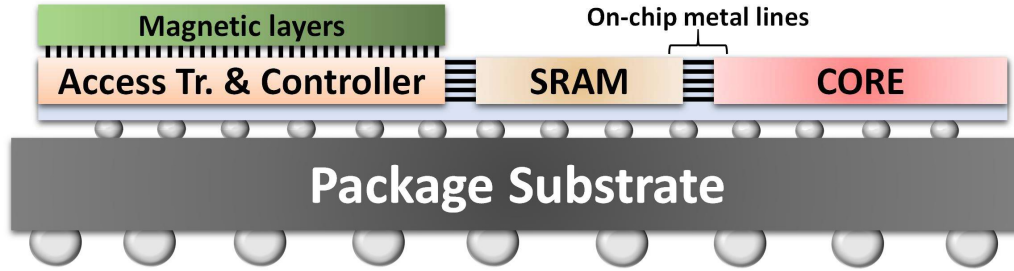


Fig. 1.3 Emerging system architecture with the embedded non-volatile memory, which has the potential to improve the bandwidth and reduce the data transfer energy further.

of 100, compared to off-chip data transfers with large capacitive loads of IO pads and off-chip wire connections. (ii) It is physically more feasible to expand the number of channels between different memory layers by using on-chip metal lines in the embedded memory compared to off-chip wires. (iii) Higher embedded memory capacity decreases the cache miss rate, effectively increasing the throughput of the system. (iv) Non-volatility can reduce the total system power consumption via zero standby power.

At the device level, the entire semiconductor community is looking for an emerging non-volatile device that can help CMOS continue to enhance system performances. Considering requirements of the embedded system memory in terms of speed, energy, area, and endurance, spintronic devices have been considered the strongest candidates among other emerging device technologies such as resistive RAM (ReRAM), phase-change RAM (PCRAM), and ferroelectric RAM (FeRAM). In the next section, the basic concept and history of spintronics are briefly discussed.

1.1 Overview of Spintronics

Spintronics (spin electronics) is an emerging interdisciplinary field that explores the methods to control the electrons' spin degrees of freedom and detect its associated magnetic moment in a solid-state device. The existence of a magnetic moment originating from electrons was discovered in the early 1920s, and later on, people found that it is related to the quantized electron spin. Specifically,

Wolfgang Pauli introduced Pauli matrices to explain the spin state and formalize the theory of spin based on quantum mechanics. In the 1970s, the conservation of the spin in electron tunneling was observed in ferromagnetic/insulator/superconducting aluminum junctions [10]. Although the spin of the electron has been theoretically and experimentally studied and contributed to areas of condensed matter physics, it had rarely used for practical applications due to its small magnetoresistance, a change in the electrical resistance of the device induced by an applied electrical or magnetic field, unlike the conventional electron charge-based electronics.

However, the discovery of the giant magnetoresistance (GMR) in ferromagnetic (FM) thin film multilayers separated by a non-magnetic metal layer attracted intense attention and became one of the significant milestones in spintronics in 1988 [11]. This is because the electrical resistance depending on the magnetization of adjacent ferromagnetic layers showed the potential that it can be used in many applications including memory devices, biosensors, and microelectromechanical systems (MEMS). Since the discovery of GMR, there have been two major directions in the field of spintronics, especially for memory applications. One has focused on further increasing the magnetoresistance (MR) to be utilized in practical memory applications, and another one has focused on developing new switching mechanisms to achieve energetically efficient and fast switching.

In 2001, it was theoretically proved that MR ratios of MgO-based tunnel barrier device could reach over 1000% based on first principle calculation [12]. A few years later, MR ratios of 200% were experimentally obtained in magnetic tunnel junctions (MTJs) with a crystalline MgO (001) barrier at room temperature in 2004 [13]. MR ratios of MTJ is called tunneling magnetoresistance (TMR), which is also defined the resistance difference between the high resistive state and low resistive state of the MTJ. The TMR has been continuously improved, achieving up to 604% at room temperature [14] via tremendous progress of MTJ fabrication, in particular, thin film deposition. From the memory circuit design perspective, high TMR not only causes reliable read operations by reducing sensing errors but also increases read speed by enhancing the sensing margin. Therefore, ever since demonstrating such

high TMR, the semiconductor industry have actively developed MTJs based magnetoresistive random-access memory (MRAM).

Regarding switching mechanism for MTJ devices, at the early development stage, a magnetic field was used to switch the state of spintronic devices (e.g. MTJ) by applying current to the adjacent metal line. However, in principle, directly applying a current or voltage to the MTJ for controlling its magnetization dynamic is more favorable in terms of energy and speed, compared to generating magnetic fields. This is because an enormous amount of energy is required to create a sufficient magnetic field to switch devices via a metal line, which in turn causes speed and scaling bottlenecks. Therefore, the magnetic field-driven switching method may be difficult to be used in practical applications.

The concept of the spin-transfer torque (STT) effect, controlling magnetization dynamics by using spin-polarized charge current directly flowing through the device, was proposed independently by Berger and Slonczewski in 1996. The STT effect has allowed achieving a fairly energetically efficient and fast switching of MTJ devices [15]–[18]. Also, the spin Hall effect (SHE), spin accumulation of the surface of a heavy metal (e.g. Ta, Pt) induced by flowing charge current, was predicted in 1971 [19] and experimentally observed in materials with a high spin-orbit coupling (SOC) in 2004 [20]–[23]. The SHE has the potential to reduce the switching energy further by one order of magnitude compared to a switching induced by the STT effect.

Although both current-driven switching mechanisms of MTJs have the promising advantages compared to that of conventional storage memory devices (e.g. flash memory) in terms of speed and energy, they need to be further improved to be comparable with embedded system memories (DRAM, SRAM). However, there are fundamental limits for the current-driven switching mechanisms to improve both speed and energy simultaneously since there exists a trade-off between them. Recently, a possible alternative has been proposed to achieve an ultra-low-energy and high-speed switching of the magnetization by using voltage-driven effect [24]–[26]. Specifically, the voltage-controlled

magnetic anisotropy (VCMA) effect, modulating the magnetic properties via interfacial effects, ideally does not involve any ohmic dissipation during the switching, resulting in extremely low switching energy (down to $\sim 1\text{fJ}$). Also, VCMA-driven switching of an MTJ offers the advantage of high switching speed (down to $\sim 100\text{ ps}$) due to the nature of the precessional magnetic dynamics. This dissertation focuses on the study of voltage-driven switching in an MTJ device including its modeling, circuit design, and memory applications.

1.2 Magnetic Tunnel Junction (MTJ)

Magnetic tunnel junction (MTJs) have been considered the most practical spintronic device due to their compatibility with CMOS fabrication, sufficient on/off ratio (i.e. TMR), and scalability. These characteristics allow MTJs interacting with the-state-of-the-art charge-based electronics on the same chip as a non-volatile storage block, leading to more compact, faster, and efficient electronics systems. An MTJ consists of two ferromagnetic layers (e.g. CoFeB) separated by a tunneling barrier (e.g. MgO). The magnetization of the pinned (or fixed) layer should not be changed under any bias conditions to achieve normal memory operations. Typically, two stable magnetic equilibrium exist along an easy magnetic axis in the free layer. The parallel state (P) occurs when the magnetic moments of the both layers are aligned in the same direction giving rise to a low resistance (R_P); on the other hand, the anti-parallel state (AP) occurs when the magnetic moment of the free layer is magnetized in the opposite direction to that of the pinned layer giving rise to a high resistance (R_{AP}). Note that the MTJ can also have an intermediate resistance value between R_P and R_{AP} if the magnetization of the free layer has transient positions.

1.2.1 MTJ Device Structure

The energy of the free layer is minimized if its magnetization is aligned with a certain axis called an easy axis resulting from the sum of different types of anisotropy in ferromagnetic systems. The

opposite directions along an easy axis are equivalent. Also, it is possible to have two or more thermally stable equilibrium states in an MTJ by engineering anisotropies. However, due to limited TMR, it is preferred to have two equilibrium states to fully exploit the resistance difference in the device for practical memory applications. Based on the direction of easy axis, MTJs can be categorized into two types: in-plane and out-of-plane (perpendicular) devices. In this dissertation, we mainly focus on the out-of-plane MTJ with the international system of units (SI) rather than the centimetre-gram-second system of units (CGS).

1.2.1.1 In-plane MTJ

An in-plane configuration has been used in early stage MTJ devices by engineering shape anisotropy. The demagnetization field is not equal for all directions in an elliptic cylinder shape device, creating easy axis. As shown in Fig. 1.4, the \hat{x} direction is the easy axis, and a rotation of the magnetization occurs via the \hat{y} direction (hard axis) where the smallest energy barrier is observed. Since the demagnetization field of the \hat{z} direction is very large ($H_d \sim M_s$), it is difficult to switching the magnetization through the \hat{z} direction. The magnitude of in-plane MTJ energy barrier is given by

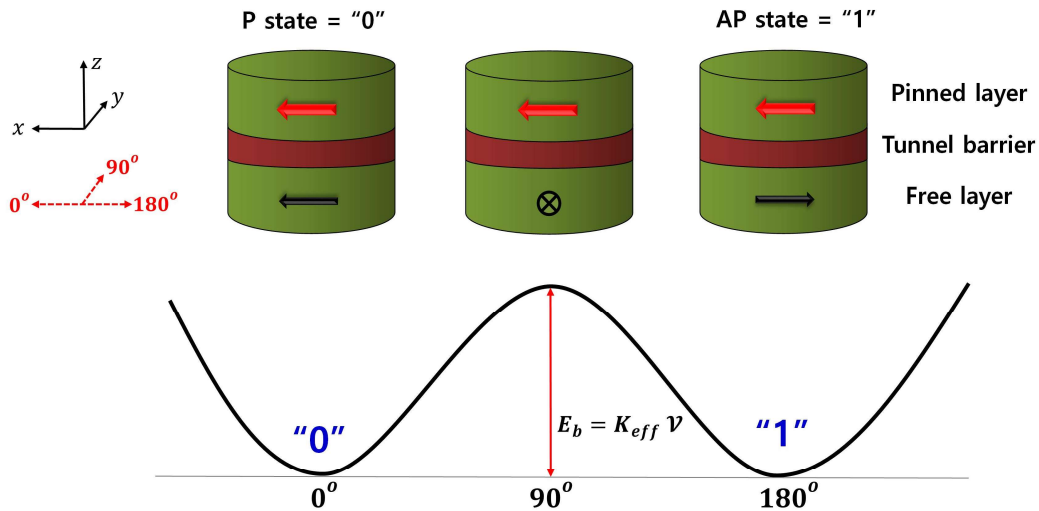


Fig. 1.4 In-plane MTJ utilizing the shape anisotropy for creating two equilibrium states. The ellipsoidal shape generates unequal demagnetizing fields for \hat{x} , \hat{y} , and \hat{z} directions, creating the energy barrier.

$$E_b = K_{eff}\mathcal{V} \quad (1.1)$$

where K_{eff} is the effective anisotropy energy [in units of energy per unit volume] and \mathcal{V} is the volume of the free layer. The effective anisotropy energy is equal to the shape anisotropy energy difference between the hard axis and easy axis, which is given by

$$K_{eff} = \mu_0 M_s H_{k,eff} / 2 = \mu_0 M_s^2 (N_y - N_x) / 2 \quad (1.2)$$

where μ_0 is the permeability [H/m], M_s is the saturation magnetization [A/m], N_x and N_y are the demagnetization factors in \hat{x} and \hat{y} direction, respectively, and $H_{k,eff}$ is the effective anisotropy field $H_{k,eff} = (N_y - N_x)M_s$ [A/m] where its magnitude is determined by the aspect ratio (AR) of the ellipse.

As the device size shrinks, the effective anisotropy energy K_{eff} needs to be enhanced to maintain the same magnitude of the energy barrier. Therefore, as the device scales down, either the AR should be increased, or the thickness of the free layer needs to be increased if the materials of the MTJ are identical. However, the former approach is feasible up to certain values of AR due to the limitation of patterning high aspect ratio memory bit. As for the latter approach, the free layer thickness is quadratically increased to compensate the volume reduction due to the scaled MTJ diameter. Hence, the required free layer's thickness would be comparable or larger than the lateral dimension of the device, causing the other direction (\hat{z}) to have the minimum energy barrier which is different to an in-plane rotation of the magnetization. Although in-plane MTJ devices typically have lower damping factor and larger TMR ratio [27]–[29], the semiconductor industry has moved toward developing perpendicular MTJ devices due to the scalability issue of the in-plane device.

1.2.1.2 Perpendicular MTJ

Another type of MTJ has the perpendicularly magnetized pinned and free layers as shown in Fig. 1.5. To make a perpendicularly magnetized MTJ, the both layers overcome the large demagnetization field H_{dem} , especially, in the \hat{z} direction by introducing a relatively stronger uniaxial anisotropy energy in

the perpendicular axis. In case of the perpendicular MTJ, the demagnetization energy is calculated by subtracting the in-plane demagnetization energy from the perpendicular demagnetization energy, which is given by

$$K_d = \mu_0 M_s H_{dem}/2 = \mu_0 M_s^2 (N_z - N_{x,y})/2 \approx \mu_0 M_s^2 /2 \quad (1.3)$$

where N_z is the demagnetization factor in the \hat{z} direction. We can approximate $N_z \approx 1$, $N_{x,y} = N_x = N_y \approx 0$, and $H_{dem} \approx M_s$ in a cylinder (circular) shape MTJ where the diameter of the device is much larger than the free layer thickness. Therefore, the magnitude of the uniaxial anisotropy energy K_u should meet the requirement $K_u > \mu_0 M_s^2 /2$ to realize a perpendicular MTJ.

The uniaxial anisotropy energy K_u for a perpendicular MTJ can be created by an interface effect between the tunneling barrier and the ferromagnetic layer or a magnetocrystalline effect [30]. Also, the uniaxial anisotropy can be expressed in terms of the uniaxial perpendicular anisotropy $H_{k\perp}$.

$$K_u = \mu_0 M_s H_{k\perp} /2 \quad (1.4)$$

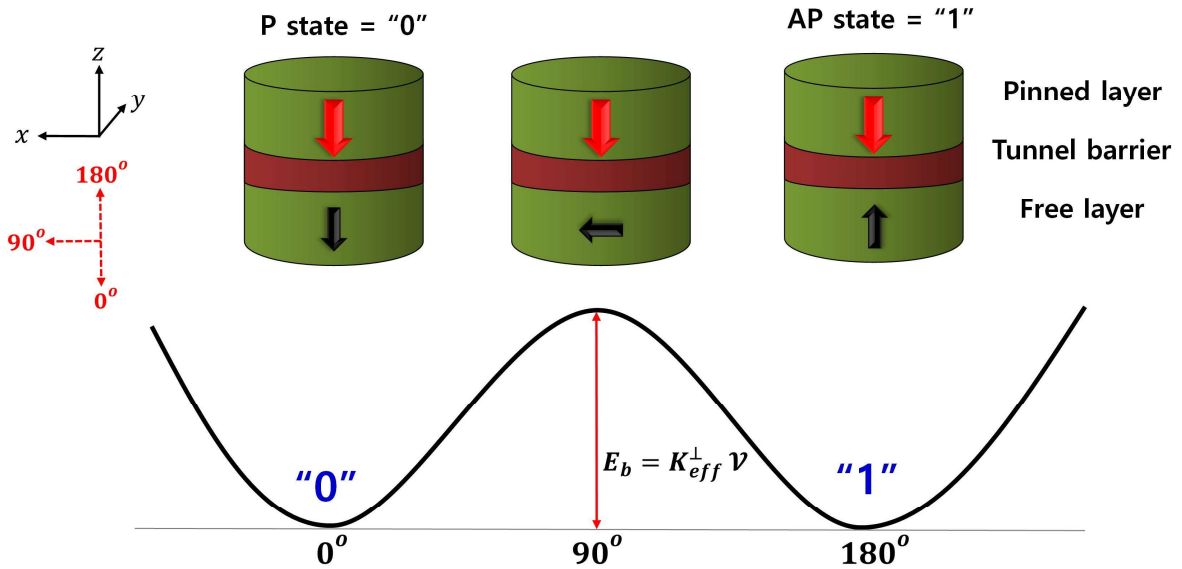


Fig. 1.5 Perpendicular MTJ utilizing the interfacial perpendicular anisotropy for creating two equilibrium states by overcoming the demagnetization field in the \hat{z} directions. Since the shape anisotropy is not necessary for forming the energy barrier, the geometry of the device can be circular, leading to a better scalability.

Therefore, the effective anisotropy energy K_{eff}^\perp is obtained by subtracting the demagnetization energy from the uniaxial anisotropy energy based on the single-domain approximation, which is given by

$$K_{eff}^\perp = K_u - K_d = \mu_0 M_s H_{eff}^\perp / 2 = \mu_0 M_s (H_{k\perp} - H_{dem}) / 2 = \mu_0 M_s (H_{k\perp} - M_s) / 2 \quad (1.5)$$

where the perpendicular effective anisotropy field H_{eff}^\perp is defined as $H_{eff}^\perp = H_{k\perp} - M_s$, but it will be revised by adding other mechanisms in the following chapter. The energy barrier between the two equilibrium states is simply given by $E_b = K_{eff}^\perp \mathcal{V}$ in a perpendicular MTJ.

In the perpendicular MTJ, the magnetization rotation between the \hat{z} and $-\hat{z}$ directions occurs via traveling the in-plane (hard axis) by overcoming the energy barrier as shown in Fig. 1.5. One of the most promising advantages of the perpendicular MTJ is that its energy barrier is not dominated by the geometry of the device but by the uniaxial perpendicular anisotropy. Since the uniaxial perpendicular anisotropy can be readily enhanced by engineering the interface effect, the perpendicular MTJ has a better scalability compared to that of the in-plane device governed by the AR. Recently, the large interfacial perpendicular anisotropy with a TMR ratio larger than 100% has been demonstrated in a Fe-rich CoFeB based MTJ [31], opening the door to practical applications. More importantly, it has been experimentally observed that such perpendicular anisotropy can be modulated by applying electric fields across the MTJ device, resulting in a low-energy and high-speed voltage-controlled switching scheme. This voltage-controlled switching mechanism in perpendicular MTJ will be discussed throughout this dissertation.

1.2.2 Tunneling Magnetoresistance (TMR) Ratio

Magnetic tunnel junctions (MTJs) have become the most promising spintronic devices as their tunneling magnetoresistance (TMR) ratio has been increased, resulting in a large readout sensing margin in memory applications. The TMR ratio is defined as the ratio between the high resistance (R_{AP}) and the low resistance (R_P), which is given by

$$\text{TMR} = \frac{R_{AP} - R_P}{R_P} \quad (1.6)$$

Although Jullière theoretically proposed the TMR effect and observed it only at low temperature in 1975 [32], it took 20 years to realize the effect in amorphous Al_2O_3 based MTJ devices at room temperature [33], [34]. For a better understanding of the Jullière model, it is necessary to introduce the spin-split density of states (DOS). Figure 1.6 shows the simplified DOS of non-magnetic and magnetic transition 3d metals. Especially, in magnetic metals, the DOS for spin-up and spin-down are unequal, causing the net magnetization and spin-polarization of the current.

It is important to distinguish two concepts between magnetization and spin-polarized current in the context of the DOS. The magnetization (M) originates from the difference between the total number of occupied electrons in the majority DOS and the minority DOS, which is given by

$$M \propto \int_{E=0}^{E=\infty} N_M(E) - \int_{E=0}^{E=\infty} N_m(E) \quad (1.7)$$

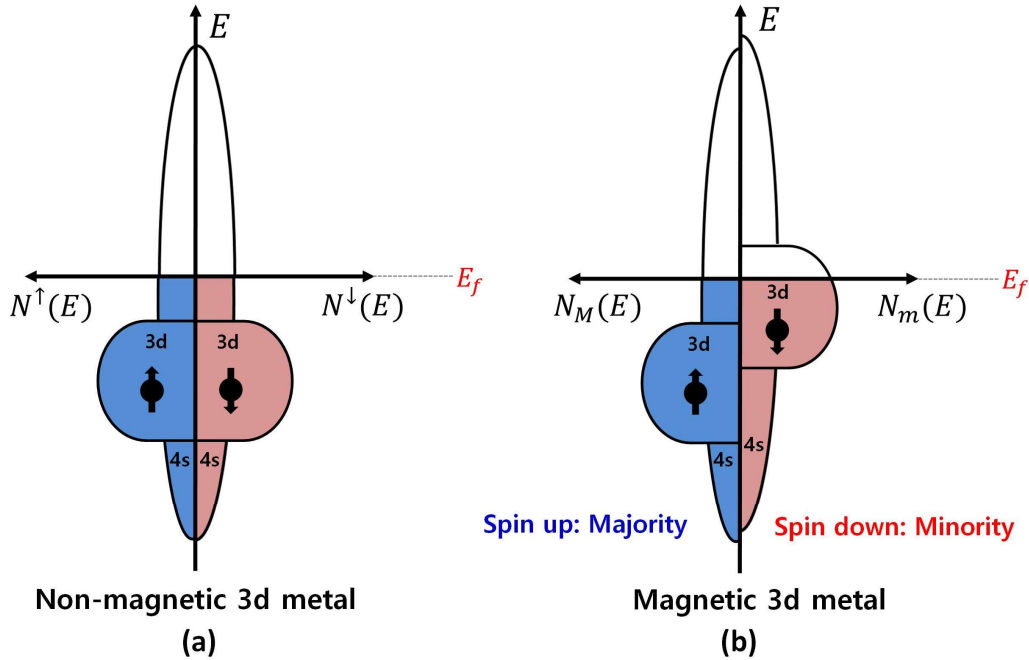


Fig. 1.6 Spin-split density of states (DOS) for (a) non-magnetic transition 3d metal which has the symmetry of the DOS for spin-up and spin-down (b) magnetic transition 3d metal that has the asymmetry of the DOS for spin-up (majority) and spin-down (minority).

However, electrical transport is caused by charge carriers close to the Fermi edge E_F . Then, we can define the spin polarization with the majority (N_M) and minority (N_m) DOSs at the Fermi edge E_F .

$$P = \frac{N_M - N_m}{N_M + N_m} \quad (1.8)$$

Jullière model assumes that not only the tunneling rate is proportional to the product of the DOS at the Fermi edge E_F for the two magnetic layers, but also there is no decay of the DOS within the tunnel barrier. Therefore, the tunneling currents of parallel and anti-parallel configuration in the MTJ would be expressed by

$$I_P \propto N_{M1}N_{M2} + N_{m1}N_{m2} \quad (1.9)$$

$$I_{AP} \propto N_{M1}N_{m2} + N_{m1}N_{M2} \quad (1.10)$$

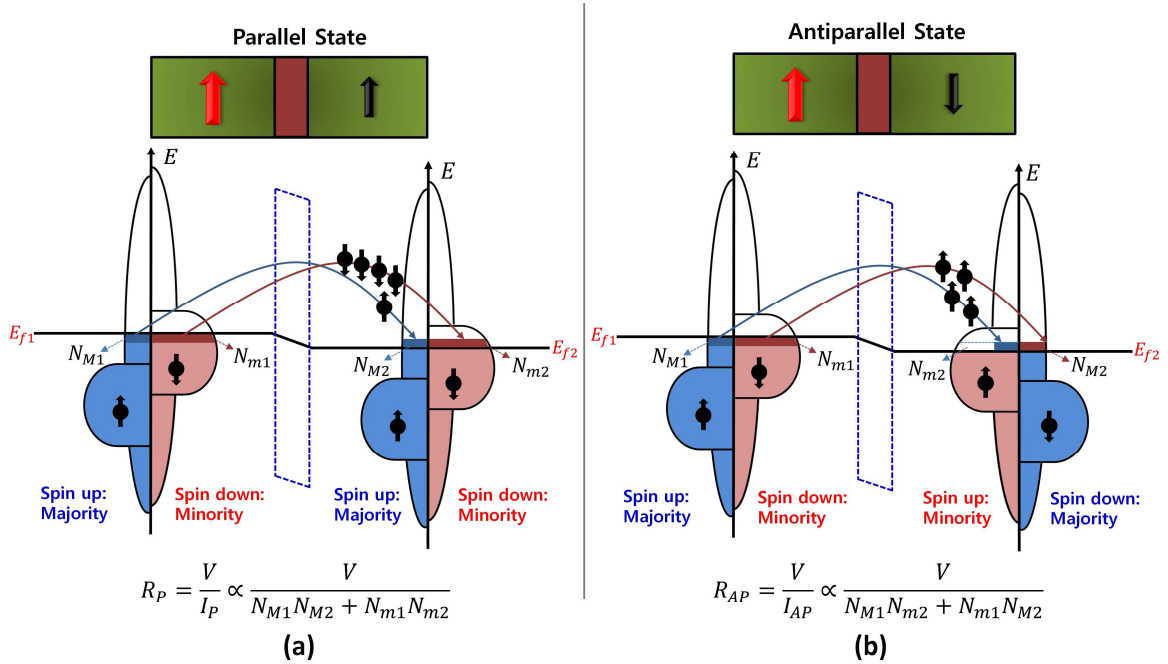


Fig. 1.7 Two channel Jullière model for the tunneling between two ferromagnetic layers with (a) parallel (b) anti-parallel magnetization. The model shows that the parallel state is dominated by the minority channel. However, in reality, the majority channel always dominates the tunneling process due to the severe decay of the minority DOS within the tunnel barrier despite the fact that the minority DOS is relatively larger than that of majority in the ferromagnetic layers.

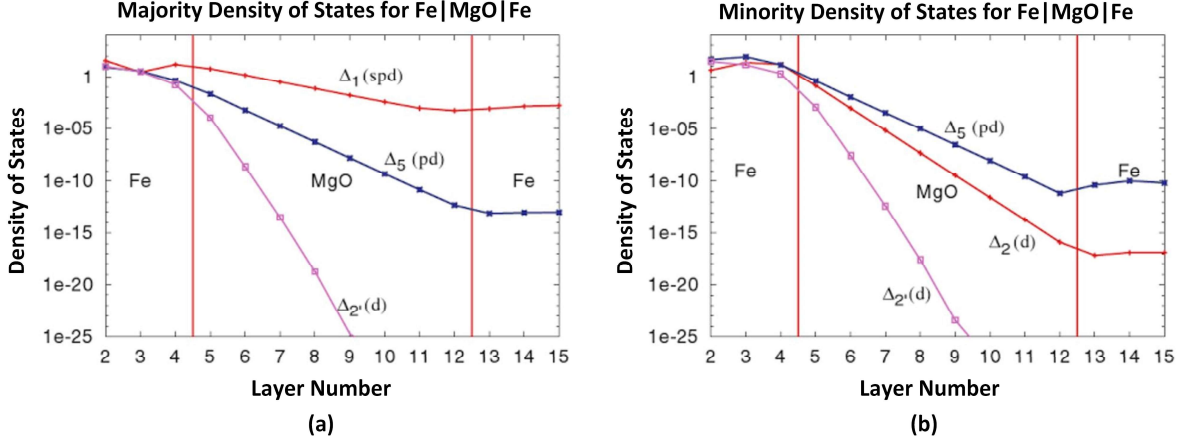


Fig. 1.8 (a) Majority and (b) minority DOS for the Fe Fermi energy Block states for parallel configuration [35]. Overall, the decay rate of the majority DOS is slower than that of the minority, allowing majority electron dominating the conductance.

where N_{M1} and N_{M2} are the majority Fermi energy DOS for the left and right ferromagnets, respectively, and N_{m1} and N_{m2} are the minority Fermi energy DOS for the left and right ferromagnets, respectively. We can define the polarization of the Fermi energy DOS for the two ferromagnets as,

$$P_1 = \frac{N_{M1} - N_{m1}}{N_{M1} + N_{m1}} \quad (1.11)$$

$$P_2 = \frac{N_{M2} - N_{m2}}{N_{M2} + N_{m2}} \quad (1.12)$$

Based on the equation from (1.9) to (1.12), we can rewrite the TMR ratio as,

$$\text{TMR} = \frac{R_{AP} - R_P}{R_P} = \frac{I_P - I_{AP}}{I_{AP}} = \frac{2P_1P_2}{1 - P_1P_2} \quad (1.13)$$

[$P_1P_2 > 0$: normal and $P_2 < 0$: inverse TMR effect]

Typically, the minority Fermi energy DOS in 3d transition metals such as cobalt and nickel is often several times larger than that of the majority, allowing the minority channel conductance to dominate the tunneling between two magnetic layers. However, this interpretation cannot be supported by experimental observations of tunneling in ferromagnets where the majority channel conductance

always dominates the tunneling regardless of the Fermi energy DOSs for both majority and minority. This is because the DOS decays within the tunnel barrier (MgO). The amount of the decay with the barrier can be calculated by including the incident Bloch states and all reflected Bloch states on the left and all possibly transmitted Bloch states on the right [35]. As shown in Fig. 1.8, the minority DOS has a faster decay compared to the majority DOS, making the majority DOS a dominant channel conductance in the actual tunneling process.

1.3 Magnetoresistive Random-Access Memory (MRAM)

Magnetoresistive random-access memory (MRAM), utilizing MTJ devices as memory elements, is categorized into several families based on the switching mechanisms: Oersted-field-switched (toggle) MRAM, STT-MRAM, MeRAM, and spin-orbit torque MRAM (SOT-MRAM). In this section, we will qualitatively describe the switching mechanisms and their bias condition for writing in one transistor and one MTJ (1T-1MTJ) cell structure. Also, advantages and disadvantages of each memory type will be discussed. Since a read (sensing) operation is quite common among the families of MRAM, detecting the current or voltage difference on the bit line (BL) or the source line (SL) depending on the state of an MTJ, we skip the detail description of the read operation in this section.

1.3.1 Oersted-field-switched MRAM

The Oersted-field-switched MRAM is considered as the first generations of MRAM and initially proposed in 1960 [36]. In order to switch the magnetization of the free layer, this type of MRAM generates the Oersted fields by applying a relatively large amount current via adjacent metal lines. The current flowing through the BL (I_{w1}) plays role in reducing the energy barrier between two states by generating the Oersted field in the direction of the hard axis as shown in Fig. 1.9. The other current flowing in the adjacent metal (I_{w2}) at the bottom side of the MTJ determines the final state depending on the sign of I_{w2} by changing the direction of the Oersted field which is parallel to the easy axis.

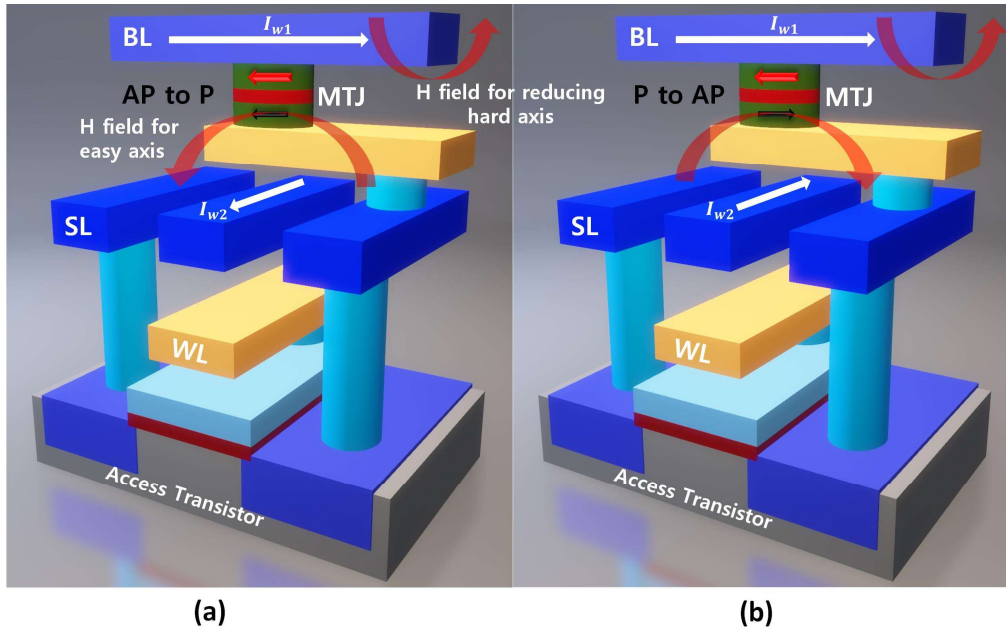


Fig. 1.9 Oersted-field-switched MRAM cell structure consisting of an access transistor and an in-plane MTJ. Two currents are applied to the adjacent metal lines to generate magnetic fields in the hard and easy axis of the MTJ. Bias conditions for switching from (a) AP to P and (b) P to AP.

Although the Oersted field based switching has been successfully demonstrated, this family of MRAM has not been widely used in practical electronic systems due to several critical issues. First of all, the magnitude of the write current for switching is substantial. For example, if an MTJ has 100 Oe (7957 A/m) coercivity, and the distance from the adjacent metal and the MTJ is 100 nm, the required current would be approximately 5 mA per bit. If the chip writes multiple MTJ devices simultaneously, it gives rise to large power consumption and heating problem. As the device size shrinks, these problems become severe because the switching current is inversely proportional the volume of the MTJ [37]. Also, as the distance between cells reduces, the stray fields from the selected cells more strongly affect the state of the adjacent cells, which may occur bit errors.

1.3.2 Spin-Transfer Torque MRAM (STT-MRAM)

Along with a giant TMR effect, the discovery of the spin-transfer torque (STT) manipulating the magnetization by spin-polarized currents in nanomagnets is considered as another milestone in the

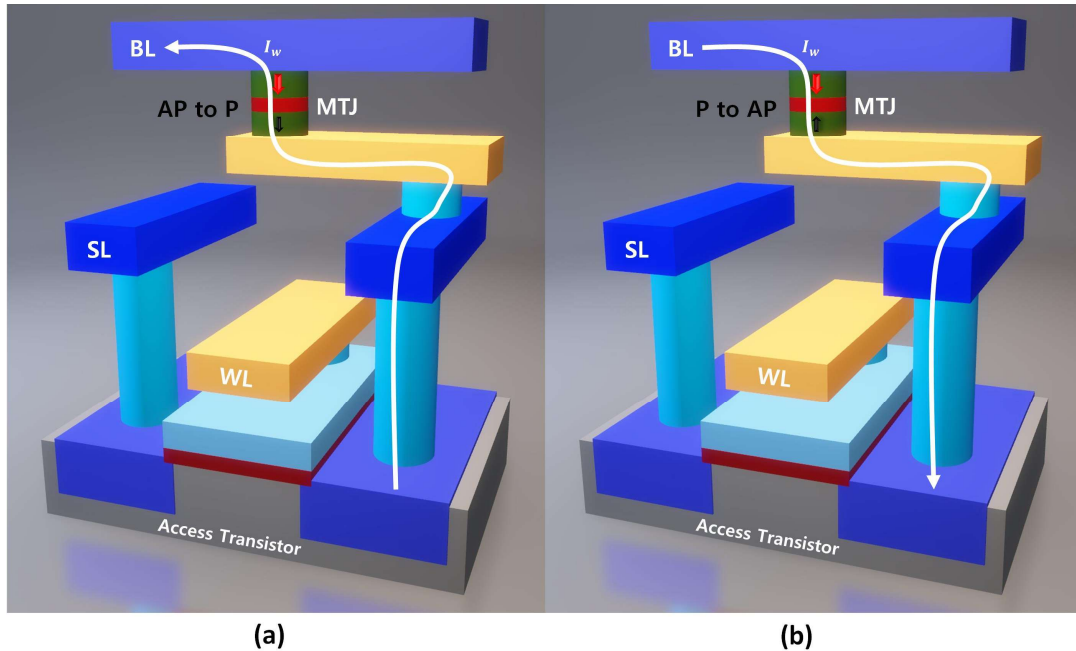


Fig. 1.10 1T-1MTJ cell architecture in STT-MRAM. Any magnetic fields are not involved in a write operation, allowing a relatively low energy switching compared to that of Oersted-field-switched MRAM. Switching can be achieved by passing a current through the MTJ, delivering the spin-transfer torque to the free layer. The final state of the MTJ is determined by the write current direction. (a) AP to P and (b) P to AP switching.

field of spintronics, accelerating the development of MRAM. STT-MRAM addresses the problems of Oersted-field-switched MRAM including scalability and switching energy issues while it maintains the general advantages of MRAM such as high endurance, high speed, and non-volatility. It has been demonstrated that the switching energy via STT effect (~ 100 fJ/bit) is two orders of magnitude smaller than that of Oersted-field-induced switching (~ 10 pJ/bit).

In a 1T-1MTJ cell structure of STT-MRAM as shown in Fig. 1.10, the writing operation is executed by applying a charge current via the MTJ device, and its final state is determined by the direction of the current. Specifically, for achieving AP to P switching, the current flows from the SL to the BL (from the free layer to the pinned layer), which allows electrons to be aligned with the magnetization of the pinned layer when they penetrate the pinned layer. This spin-polarized current transfers some of its momentum to the free layer, leading to switching of the MTJ from AP to P. On the other hand, to obtain switching from P to AP, the current flows from the BL to the SL (from the pinned layer to the

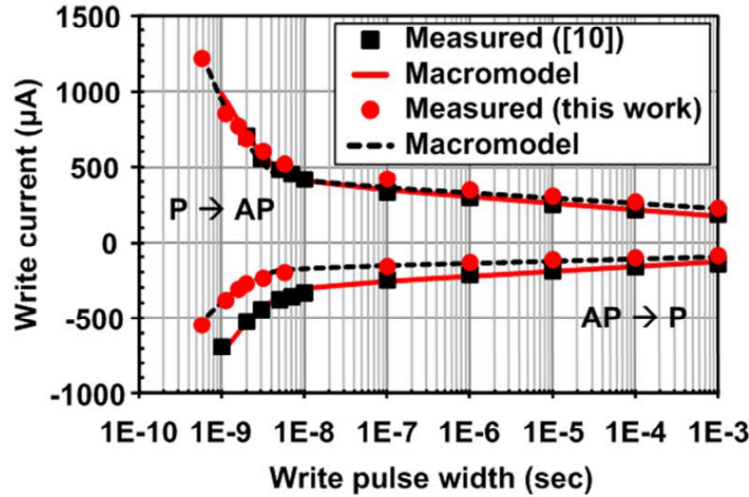


Fig. 1.11 Amplitude of write current as a function of write pulse width for switching. To achieve a high-speed switching, the required write current becomes very large, limiting the density of the array due to the large size of the access transistor [38].

free layer). In this case, electrons are transferred from the free layer to the pinned layer. However, some of the electrons, which are spin-polarized in the opposite direction of the magnetization of the pinned layer, bounce back to the free layer and transfer their momentums to the free layer, resulting in switching of the MTJ from P to AP. This current-driven STT switching is deterministic, which means that the current direction decides the final state of the MTJ. This deterministic switching allows having a simplified write circuitry, which is beneficial in terms of the circuit design effort.

Albeit STT-MRAM has many advantages, several problems need to be addressed. It is possible to change the magnetization of the free layer with relatively low current ($\sim 100 \mu\text{A}$) as long as the write operation time is long enough ($> 10 \text{ ns}$). However, to achieve a few nanosecond switching speed, the write current drastically increases as shown in Fig. 1.11 [38]. Also, the size of the access transistor should be sufficiently large to drive a required switching current, which limits the density of the cell array. Therefore, the high-performance STT-MRAM can be obtained at the expense of cell density and energy efficiency. Furthermore, the switching current from P to AP is typically larger than that of from AP to P due to different spin momentum transfer efficiencies as shown in Fig. 1.11. The asymmetry of the switching currents may lead to a uniformity issue in terms of write error rate (WER). This

problem can be more severe because of the effect of the source degeneration from the access transistor depending on the cell configuration (connection) of the access transistor and the MTJ.

1.3.3 Magnetoelectric RAM (MeRAM)

The tradeoff between speed and power of the STT-driven switching makes difficult for STT-MRAM to be used in high-performance and low-power applications. As an alternative, the VCMA effect can replace the STT and achieve both low-energy and high-speed switching by avoiding the flow of large current and exploiting the fast precessional magnetization dynamic during the switching process while maintaining other remarkable features of STT-MRAM such as high endurance, compatibility with CMOS process, and non-volatility. This type of MRAM is called magnetoelectric RAM (MeRAM), which is one of the main parts of this dissertation. We will scrutinize physics behind of the VCMA effect and its precessional switching mechanism in the next chapter. In this section, we

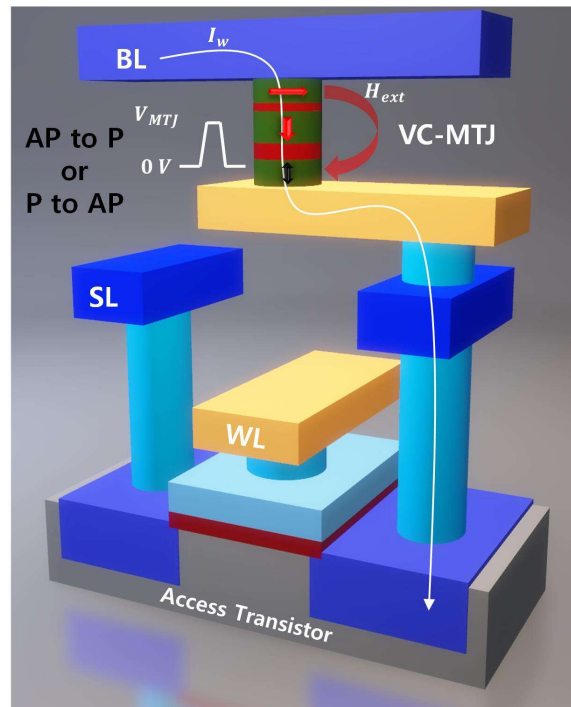


Fig. 1.12 1T-1MTJ cell architecture in MeRAM. Switching can be achieved by applying an electric pulse to the MTJ, which induces a magnetic precessional motion in the free layer. Since this type of switching mechanism is non-deterministic, a unipolar pulse can switch either from AP to P or from P to AP.

will discuss general features of MeRAM and its write bias conditions within the 1T-1MTJ cell structure.

There is no huge difference between MeRAM and STT-MRAM in terms of cell structure as shown in Fig. 1.12. However, the size of the access transistor in MeRAM can be reduced further in that the VCMA-driven switching ideally does not require the flow of current. Thus the bit cell array of MeRAM can achieve higher density compared to other families of MRAM. Also, the thickness of the tunnel barrier is relatively thicker than that of MTJ in STT-MRAM, practically reducing ohmic dissipation during the write operation.

For switching the MTJ device based on the VCMA effect, an electric pulse is applied to the device in the presence of an in-plane stray field which can be generated by the other ferromagnetic layer. The applied electric pulse modulates the magnetic property of the MTJ, resulting in precessional motion of the magnetization of the free layer. It is noted that the duration, slew rate, and amplitude of the pulse should be well defined depending on the MTJ characteristics to achieve high switching probability. Also, since the VCMA effect is unipolar, switching from AP to P or from P to AP is achieved by pulses with the same polarity. Thus, this type of switching is non-deterministic, which requires new considerations in terms of circuit design.

1.3.4 Spin-Orbit Torque MRAM (SOT-MRAM)

The spin Hall effect (SHE), which is one type of spin-orbit torques (SOTs), have been considered as another alternative for switching MTJs. Although the SHE was initially detected in thin films of the semiconductors GaAs and InGaAs by using Kerr rotation measurement [20], this mechanism has attracted a significant attention after demonstrating the SHE induced magnetic switching in a perpendicularly magnetized Ta/CoFeB/MgO/Ta film at room temperature [22]. This family of MRAM that utilizes spin-orbit torque is referred to SOT-MRAM.

Figure 1.13 shows a unit cell of SOT-MRAM where the access transistor is connected to the top electrode of the MTJ, which is fabricated on top of the non-magnetic metal with spin Hall angle, also

called the spin-orbit coupling (SOC) metal. For the write operation, the current flowing in the non-magnetic metal generates a spin current that delivers spin-orbit coupling (SOC) to the MTJ device for switching. The conversion efficiency between the charge current I_c and the spin current I_s is called spin Hall angle $\theta_{SH} \propto I_s/I_c$. The polarization of the spin current is the in-plane, which is the direction of the cross product of the charge current and the spin current. The desired MTJ state can be determined by the direction of the charge current.

The advantages of SOT-MRAM is as follows. First, it is possible to reduce the switching current by one order of magnitude by engineering the thickness of the metal layer with spin Hall angle compared to that of STT-MRAM. Second, the separation of the electrical paths between the write and the read operations in the three-terminal device decreases read disturbances, improving the reliability. In addition, the write circuit would be simple since the SHE-driven switching current from P to AP

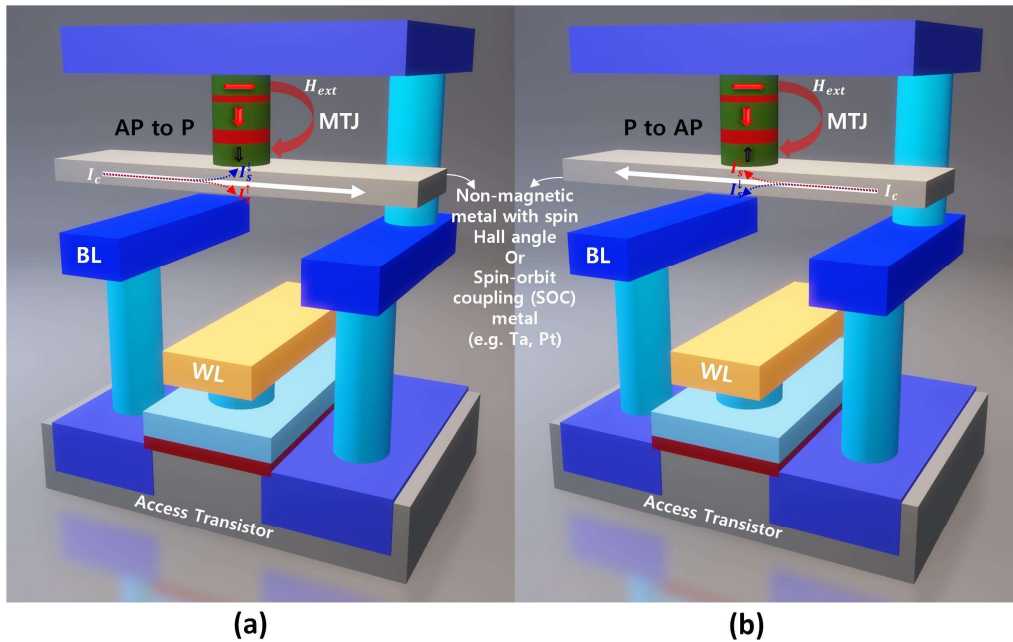


Fig. 1.13 1T-1MTJ cell architecture in SOT-MRAM. The three-terminal device utilizes the spin Hall effect of the SOC metal. For switching an MTJ, the charge current flowing in the SOC metal creates the spin-current via the SHE effect, which causes spin-orbit coupling in the free layer. A stray field is required to break the symmetry to switch perpendicular MTJ devices. The final state of the MTJ is determined by the write current direction. (a) AP to P and (b) P to AP switching.

and from AP to P is symmetric. Finally, a relatively low resistive metal layer (e.g. Ta, Pt) decreases ohmic loss further, resulting in the significant reduction of switching energy (~ 10 fJ/bit).

However, SOT-MRAM also faces challenges and issues that need to be solved. First, since the unit cell is based on the three-terminal magnetic device, it may require additional transistors at the edges of the metal with spin Hall angle for the write operation other than the access transistor, leading to penalties in terms of cell array density. Second, the spin current polarized into in-plane, in principle, cannot switch a perpendicularly magnetized MTJ device. Hence, additional methods are necessary to break the symmetry by using applying an external stray field, structural manipulation, and interfacial electromagnetic coupling [39], which increases the fabrication complexity.

1.3.5 Performance Comparison

1.3.5.1 Device Level

Conventional memory technologies for Von-Neumann structures have their physical limitations. Although SRAM has fast access time (< 1 ns) and unlimited endurance, this memory technology are power-hungry at the advanced CMOS technology nodes due to high leakage current [40]. Also, DRAM plays an important role in the main memory of the Von-Neumann architecture. However, it is volatile, requires refresh, and is not as fast as SRAM [41]. Although solid-state devices such as Flash memory are typically used in embedded systems as a secondary storage because of their high density, they suffer from low endurance ($< 10^5$), slow write time (> 10 μ s) and high write and erase energy [42]. The performance of existing memory technologies is summarized in Fig. 1.14.

Since each conventional memory has its pros and cons, sophisticated hierarchical memory structures consisting of several specialized layers have been used in current computing systems to achieve a high-performance with low cost. Even though potential performance breakthroughs are possible in specific layers based on emerging technologies, it remains difficult to improve the

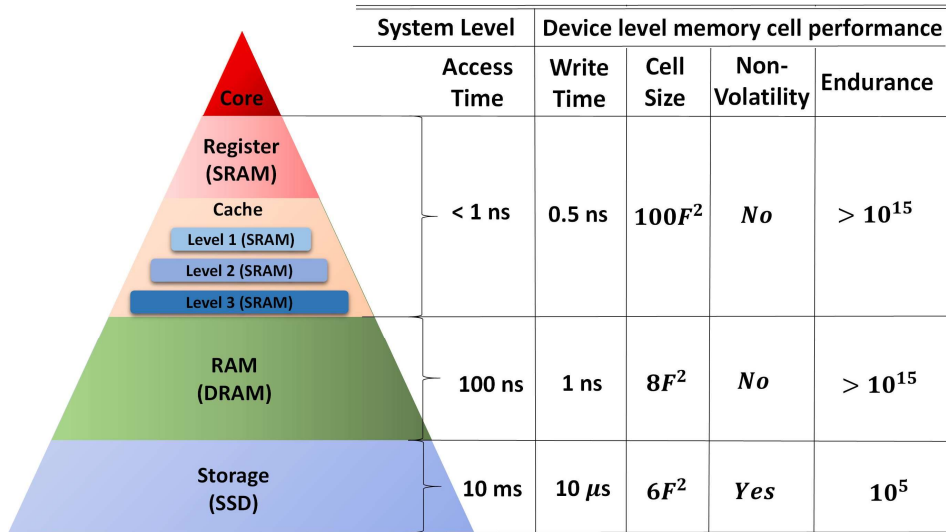


Fig. 1.14 Memory Hierarchy in a conventional computer architecture with system level and device level memory performance.

performance of an overall system because interactions between the layers often act in unexpected ways [43].

To solve these challenges, many researchers have worked on next generation memory technologies, with the ultimate goal often described as “Universal Memory” (UM), i.e. a solution satisfying all the key requirements for a combined embedded and storage memory device: cell size $< 4F^2$, read/write time < 1 ns, write energy < 1 fJ, endurance $> 10^{15}$, non-volatility with retention of more than 10 years, and low cost for fabrication [44]. Although currently none of the emerging memories in development meet all the requirements of the UM, some of the emerging memories, i.e. Resistive RAM (ReRAM), Phase-Change RAM (PCRAM), and Magnetoresistive (MRAM), may be able to merge more than one layer of the present memory hierarchy, hence bring the UM concept closer to realization.

The 1T-1R (i.e. one access transistor associated with one resistive element) structure is widely used in memory cells since a high on/off ratio of the metal–oxide–semiconductor field-effect transistor (MOSFET) allows to select an individual memory element effectively, preventing sneak paths. In addition, the access transistor can be used in not only unipolar operations but also bipolar operations, expanding its application to different cell characteristics. However, considering the layout design

limits of the MOSFET device, the feasible minimum size of an access transistor is $\sim 20F^2$ with standard logic processes, which limits the density of the memory cell array. Although the effective cell size can reach further down to $\sim 10F^2$ by sharing the source regions of the adjacent two cells with special via contacts [29], the access transistor is still the key factor limiting the density. Here, we briefly introduce several types of emerging memory technologies and compare their performance with that of MeRAM in the form 1T-1R structure.

First of all, PCRAM takes advantage of the resistance difference between the crystalline (low R) and amorphous (high R) states in phase change materials such as $Ge_2Sb_2Te_5$ and Sb_2Te . The on/off ratio is $\sim 10^3$ - 10^4 , allowing it to be used in multi-level cell (MLC) operations. In terms of retention, PCRAM can exceed 10 years at $85^\circ C$ [45], [46]. However, PCRAM is a power hungry device because the reset current for converting to the amorphous state is relatively large. For example, a PCRAM device with the 300 nm^2 aperture requires $300\text{ }\mu A$ (10^8 A/cm^2) reset current [47]. Therefore, the access transistor in a cell must be carefully chosen so that it can deliver a sufficient current. In order to provide such a high reset current, PCRAM based 1T-1R cell structures typically occupy $\sim 30F^2$ - $50F^2$ due to the large size of the access transistor, limiting its density.

	PCRAM	Re-RAM	STT-RAM	MeRAM
Non-volatility	Yes	Yes	Yes	Yes
Device size [nm^2]	7.5×22	10×10	11×11	30×30
Cell size with Tr. [F^2]	$30\text{--}50F^2$	$20\text{--}40F^2$	$20\text{--}50F^2$	$20F^2$
Cell size crossbar [F^2]	$4F^2$ (with diode)	$4F^2$	Cannot be applicable	$4F^2$ (with diode)
Write time [ns]	10–50 ns	10–100 ns	>3 ns	0.1–1 ns
Write current [μA]	80–200 μA	10–100 μA	>50 μA	1–10 μA
Write voltage [V]	1.5–3 V	1–2.5 V	~ 1 V	0.8–1.8 V
Write energy [J]	20 pJ/bit	0.1–1 pJ/bit	>90 fJ/bit	1–10 fJ/bit
ON/OFF ratio	10^{+1-4}	10^{+1-3}	TMR 100~200%	TMR 100~200%
Endurance [#]	10^{+5-9}	10^{+5-8}	10^{+15}	10^{+15}
Retention [year]	10 year	10 year	10 year	10 year

Table 1.1 Performance comparison of emerging memory technologies in 1T-1R cell structure.

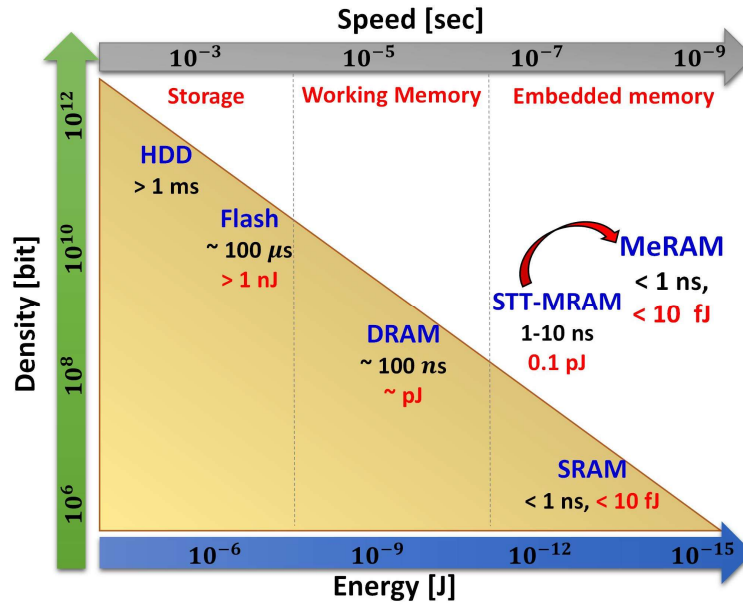


Fig. 1.15 Performance comparison of different memory technologies in terms of energy, speed, and density. MeRAM has the best combination of all these three categories.

ReRAM can be classified into two types: unipolar and bipolar devices. A unipolar ReRAM requires the different intensity of electric field to switch its states without changing its polarity. On the contrary, switching for a bipolar device needs different polarities of current [48]. The on/off ratio of ReRAM is as high as that of PCRAM. Due to its relatively smaller write current compared to PCRAM, i.e. 100 μA for $54 \text{ nm} \times 54 \text{ nm}$ cell [49], a ReRAM based 1T-1R cell can be reduced to $>20F^2$ in size. However, it has a low endurance 10^{5-8} and a relatively long switching time $>10 \text{ ns}$, limiting its applications [50].

In STT-MRAM, switching operation can be fast in the ballistic ($< 10 \text{ ns}$) and thermally activated regimes ($> 10 \text{ ns}$) based on the amplitude of the applied current. Hence, to increase switching speed, the write current must increase, requiring a large access transistor. Typically, the STT-driven 1T-1MTJ cell occupies $>30F^2$ to supply a sufficient write current $> 50 \mu\text{A}$ for the element with 50 nm-diameter [51] and shows unlimited endurance ($> 10^{15}$) [42]. Although these metrics makes STT-RAM a potential candidate to be used in SRAM replacement applications, the write energy (presently $\sim 100 \text{ fJ/bit}$) and switching speed still need to be improved further.

Table 1.1 and Fig. 1.15 summarize the performance comparison of MeRAM with other emerging memory technologies. Due to the VCMA-driven precessional switching, MeRAM dramatically improves the write energy (down to ~ 1 fJ/bit) and speed (< 1 ns). Furthermore, MeRAM cell only occupies $20F^2$ area with the minimum size transistor in a standard logic process due to its very low write current of less than $10 \mu\text{A}$. In addition, unlike STT-MRAM, MeRAM is based on unipolar devices, which allows utilizing a typical diode as the access device in a cell. This results in further reducing its cell size down to $4F^2$ and enabling 3D integration for very high bit densities [52].

1.3.5.2 Array Level

To date, there have been no studies evaluating the performance of MeRAM at the integrated array level, which may differ from the single-device characteristics. A systematic comparison to other existing or emerging embedded memories is also needed. This section compares the array level performance of MeRAM with that of SRAM, DRAM, eDRAM, and STT-MRAM based on the 28 nm node. The evaluation is conducted under two configurations: (i) same array capacity and (ii) same array area.

The structure of a memory bank consists of a crossbar array with columns (BLs) by rows (WLs) of memory cells and peripheral circuits such as drivers, decoder, multiplexer, and sense amplifiers as shown in Fig. 1.16(a). The storage capacity of a memory chip is often divided into several identical banks to reduce the critical paths at the cost of area efficiency [53].

It is important to note that there is a tradeoff between the total size of the memory array (capacity) and the performance (e.g. latency, energy). This is because increasing the number of cells in an array raises the capacitive and resistive loading on the shared signal lines, which in turn requires more energy and latency during operation. Therefore, the size of the memory array should be carefully designed based on a targeted application.

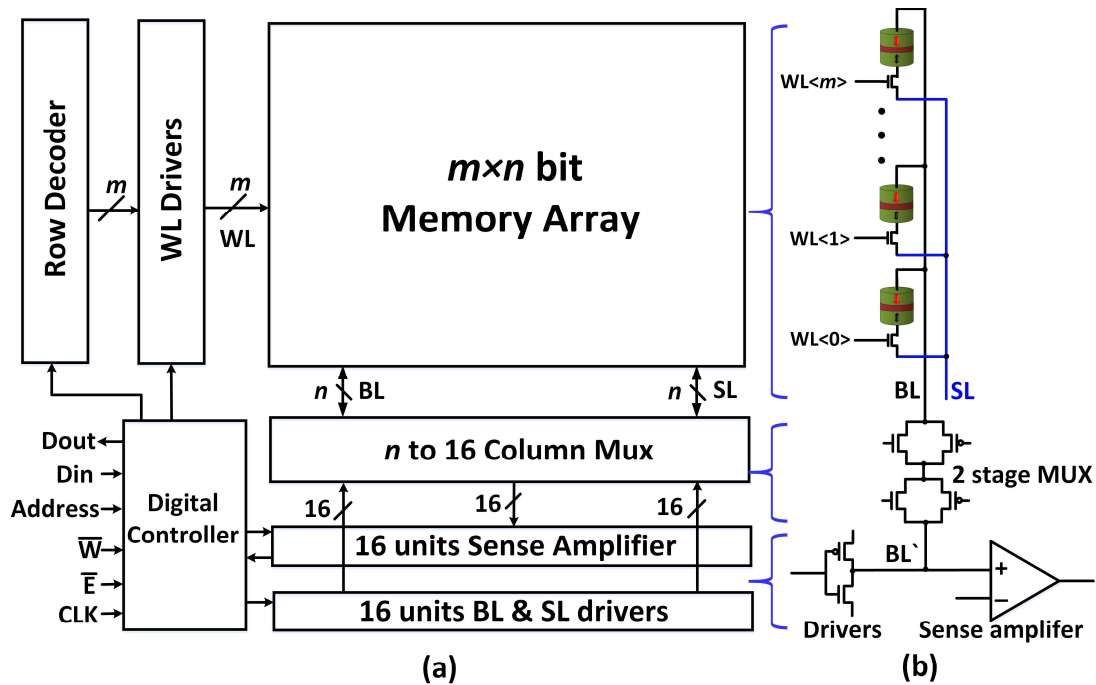


Fig. 1.16 (a) Memory bank architecture, where the row of $m \times n$ memory array is controlled by the WL drivers with the row decoder, and its column (BLs and SLs) are connected to the sense amplifier and drivers via the column mux. The control signals are generated by a digital controller based on the requested operation. (b) Schematic of a single column connection of MeRAM.

Although MeRAM also follows the general memory bank architecture, requirements for the drivers and sensing circuitry are a little bit different, as shown in Fig. 1.16(b). These requirements include sufficiently strong BL, SL, and WL drivers to maximize the slew rate ($> 1 \text{ V}/100 \text{ ps}$), and the write pulse width should be adjustable with a high resolution (100 ps) as these factors have a large impact on the write error rate (WER). Also, the sense amplifier should be able to distinguish a small sensing margin ($\sim 100 \text{ mV}$) due to the limited tunneling magnetoresistance ratio (TMR, the ratio between the two resistance states) of MTJs.

The array level performances of different memory technologies are compared under two conditions based on 28 nm node CMOS parameters: (i) fixed array capacity ($512 \times 512 \text{ bit}$), typical bank size of embedded memories, and (ii) fixed array area ($200 \times 200 \mu\text{m}^2$), which is equal to the area of a 256 Kbit SRAM array. Table 1.2 provides the values of the parameters used in this estimation. Table 1.3 shows the typical cell size for each memory technology in terms of the minimum feature size F .

Since an SRAM cell consists of six transistors and requires both NMOS and PMOS, its cell size reaches up to $190 F^2$ [54], [55]. An STT-MRAM typically has a $50 F^2$ cell size due to the large access transistors needed to supply the critical current ($> 10^6 A/cm^2$), required to achieve sub-10 nanosecond switching [56], [57]. The standalone trench-based DRAM technology (the only non-embedded memory listed on the table) allows its cell size to be as low as to $4\sim 8 F^2$ [58], [59]. With a logic process, however, the eDRAM cell occupies $> 40 F^2$ because a larger area is required to maintain sufficient capacitance [60], [61]. The cell area is approximately $20 F^2$ for a MeRAM cell with a standard logic process since an MTJ does not require a large current for switching. However, in principle, the cell area can be reduced to $8 F^2$ if MeRAM adopts a specialized process similar to DRAM process.

The write access time (t_{A_W}) of the bank is estimated via two methods. In the case of MeRAM and STT-MRAM, the write access time is extracted by combining the peripheral circuit delay (t_d) and the device write time (t_{w_cell}). The device write time of MeRAM and STT-MRAM are chosen to be 1 ns and 5 ns, respectively, in which a write error rate (WER) of 10^{-4} is guaranteed, and these numbers are

Parameters	Symbol	Value	Unit
Metal capacitance per unit area	C_M	1	fF/ μm^2
Junction capacitance per unit area	C_J	8	fF/ μm^2
Gate capacitance per unit area	C_G	22	fF/ μm^2
DRAM cell capacitance	C_{DRAM}	20	fF
eDRAM cell capacitance	C_{eDRAM}	10	fF
Metal line sheet resistance	R_M	0.3	Ω/\square
STT-MRAM cell resistance	R_{STT}	5	k Ω
MeRAM cell resistance	R_{MeRAM}	50	k Ω
SRAM cell resistance	R_{SRAM}	17	k Ω
Access transistor resistance	R_{Tr}	10	k Ω
device write time for MeRAM	t_{w_MeRAM}	1	ns
device write time for STT-MRAM	t_{w_STT}	5	ns

Table 1.2 28 nm node technology parameters and device characteristics.

	SRAM	MeRAM	STT-MRAM	DRAM	eDRAM
Cell Area	0.149	0.016	0.039	0.006	0.034
$A_C [\mu m^2]$	($190F^2$)	($20F^2$)	($50F^2$)	($8F^2$)	($40F^2$)

Table 1.3 Memory technologies unit cell area based on 28 nm node.

typically observed in many published works [48]. For SRAM, DRAM, and eDRAM, the write access time is the sum of the peripheral circuit delay (t_d), the charging time of the bit line (t_{A_Drive}) via a driver circuit, and the intrinsic RC delay of a single cell. Note that the delay of the chip IO interface is excluded in our estimation as we are assuming embedded applications.

The read access time (t_{A_R}) is obtained by adding the peripheral circuit delay (t_d) and the array delay required for generating a fixed margin (t_{A_RC}). In high-speed read operations, the bit line is pre-charged to a certain potential before read. The BL then discharges through the selected memory cell, until the voltage difference between the selected BL and the reference BL is sufficient for the sense amplifier to distinguish. Thus, the read access time is greatly dependent on the total RC load of the bit line and the resistance of the selected cell compared to the write access time.

The write energy is divided into two parts: (i) energy dissipation via capacitive charging (E_{A_C}) and (ii) ohmic loss (E_{A_O}). The former depends on the sum of bit line and word line metal capacitances, transistors' junction and gate capacitance, and is quadratically proportional to the amplitude of the write voltage. The latter is a function of the write voltage, the write time, and the total resistance through the current path. The formulas of these performance parameters are summarized in Table 1.4.

Parameters	Symbol	Formulas
Technology node [nm]	F	28
Array Length [μm]	L_A	$(N^2 * A_C)^{0.5}$
Array metal line capacitance [fF]	C_{A_M}	$2F * L_A * C_M$
Array Tr. junction capacitance [fF]	C_{A_J}	$N * F * 0.1 * C_J$
Array metal line resistance [Ω]	R_{A_M}	$(L_A * R_M)/2F$
Array RC delay [ps]	t_{A_RC}	$(R_{\text{cell}} + R_{A_M}) * (C_{A_M} + C_{A_J})$
Array area [μm^2]	A_A	$A_C * N^2$
Peripheral circuit delay [ps]	t_d	$\log_2(N) * 10 * \text{FO4}$
Write time [ns]	t_{A_W}	$(t_d + t_{w_cell})$ or $(t_d + t_{A_Drive})$
Read time [ns]	t_{A_R}	$(t_d + t_{A_RC})$
Write energy (capacitive) [fJ]	E_{A_C}	$(C_{A_M} + C_{A_J}) * (V_{\text{write}})^2/2$
Write energy (ohmic loss) [fJ]	E_{A_O}	$(V_{\text{write}})^2/(R_{\text{cell}} + R_{A_M}) * t_{w_cell}$

Table 1.4 Formulas for array level parameters. N is the number of cells that connect to a single bit line or a word line. V_{write} and t_{w_cell} are write voltage and intrinsic device write time, respectively. t_{A_Drive} is the charging or discharging time of the bit line via a driver circuit, and FO4 is the delay of an inverter.

Condition I: Array capacity (512 × 512 bit)

An array capacity of 512 × 512 bit is used for performance comparison. The physical length of the array (L_A) is extracted by using the array capacity and the unit cell dimension in each memory technology (A_C). The access time, write energy, and array area of each memory technology are shown in Fig. 1.17.

In terms of the write access time, logic-based SRAMs achieve the fastest write operation. While DRAM and eDRAM share similar structures, DRAM adopts a specialized process to allow larger cell capacitance at a relatively small area. This results in DRAM having slightly slower write time than its embedded counterpart but provides wider margins and greater retention times. MeRAM can achieve the same level of performance as volatile working memory (SRAM, DRAM, eDRAM), whereas there is still a gap (>4x) for STT-MRAM.

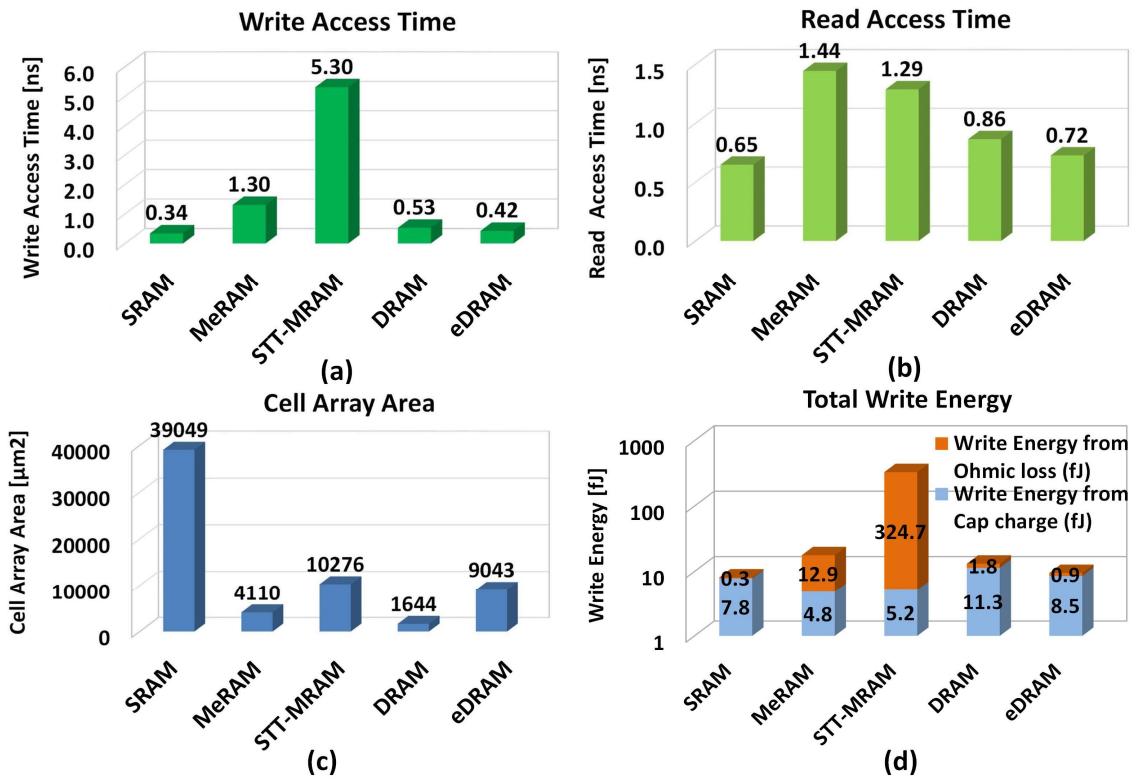


Fig. 1.17 Array level performance comparisons based on the condition that the array capacity is 256 Kbit. (a) write access time (b) read access time (c) array area (d) write energy.

While the read access times of both eDRAM and DRAM are sub-nano second, it should be noted that here we are excluding the effect of retention, which greatly decreases the read margin as the charge in a DRAM cell leaks. Magnetic memories suffer from smaller margins, which reflect in the read operation time, but is still in the acceptable regime ($\sim 2x$) compared to that of SRAM. The write energy follows a similar trend as the write time, with MeRAM achieving the same level of energy efficiency as working memories and a 20x improvement compared to STT-MRAM. However, the non-volatility of MeRAM can potentially save orders of magnitude in standby energy compared to that of volatile memories.

Condition II: Array area ($200 \mu\text{m} \times 200 \mu\text{m}$)

A 256 Kbit SRAM array occupies the area $40,000 \mu\text{m}^2$ in which the other memory technologies can have up to Mbit of capacity, as shown in Fig. 1.18(c). The performance of these memories thus

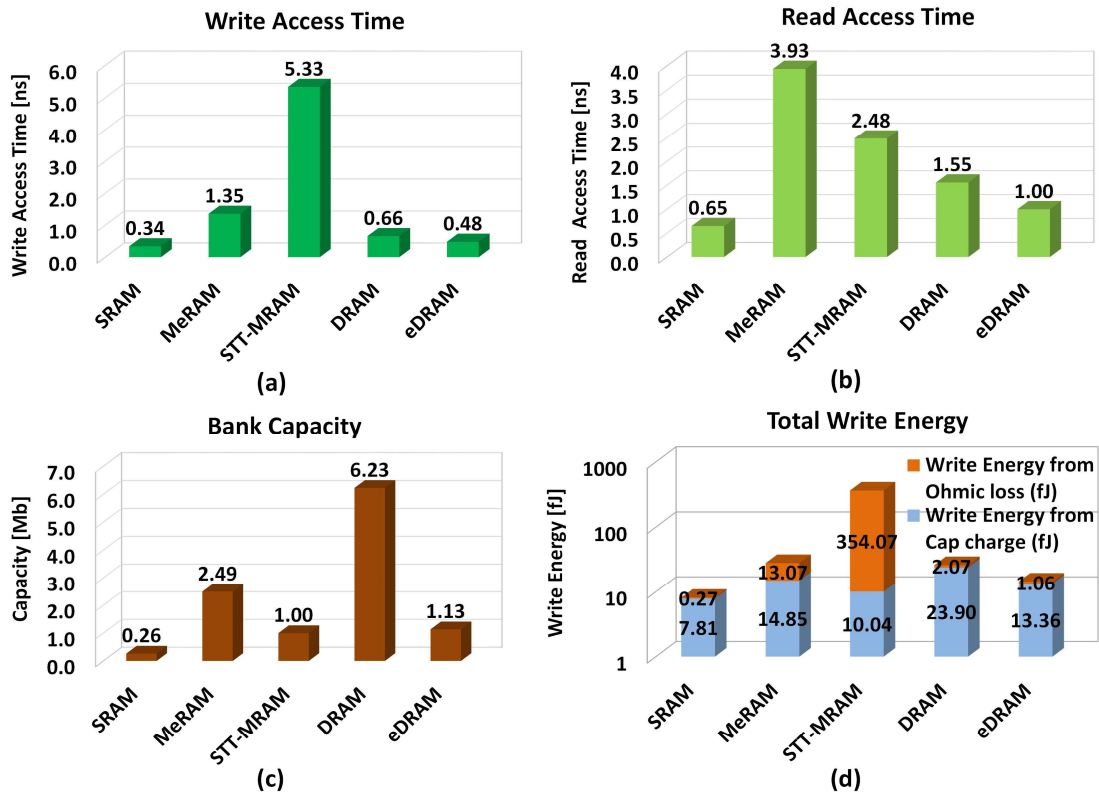


Fig. 1.18 Array level performance comparisons based on the condition that the array area is $200 \mu\text{m} \times 200 \mu\text{m}$. (a) write access time (b) read access time (c) array capacity (d) write energy.

degrades due to the increased RC loading resulting from the increased capacity. While MeRAM suffers from a slightly longer read access time (~ 4 ns) as shown in Fig. 1.18(b), it is possible that the high capacity ($\sim 10\times$) can compensate for the increased read access time and even achieve higher system throughput by decreasing the cache miss rate. Since the write access time is strongly dependent on the device write time in the cases of MeRAM and STT-MRAM, there is little change in the write time despite a change in the array capacity. However, the size of WL drivers and BL drivers should be adjusted to fulfill the write condition. For the write energy, MeRAM consumes 28 fJ per bit, which is comparable to that of DRAM and higher than that of SRAM and eDRAM by 3.5x and 2x, respectively. STT-MRAM requires at least an order of magnitude higher energy compared to other memory technologies due to its high Ohmic dissipation as shown in Fig. 1.18(d).

In conclusion, the write access time and dynamic write energy of MeRAM are comparable to those of conventional embedded memories. However, MeRAM provides a large improvement in terms of density over embedded SRAM and STT-MRAM. As the memory array size increases, the read access time may limit the entire system throughput. Here, the higher bit density (smaller area) can favor a faster read, while the relatively high cell resistance of MeRAM can increase read delay. To alleviate this issue, high-speed sensing schemes need to be developed at the circuit-level. At the device-level, TMR can be improved further to enhance the sensing margin, which in turn reduces the read access time.

CHAPTER 2

MACROSPIN COMPACT MODELING OF MTJ

2.1 Motivation

Magnetic tunnel junctions (MTJs) are being actively developed by the semiconductor industry as one of the most promising memory devices, opening the door to new possibilities of next-generation low-power and high-speed system architectures. As we illustrated in Chapter 1, MTJs have the potential to be implemented as an embedded system memory (e.g., L3 cache), which directly transmits information to arithmetic logic units (ALUs) or digital signal processors (DSPs) with low latency, and stores the data in a nonvolatile, fast, and energetically efficient way.

A compact model accurately capturing the VCMA-induced magnetization dynamics is essential for successful development of VCMA-based memory. Although several works related to single-domain Landau–Lifshitz–Gilbert (LLG)-based macrospin models have been reported [62]–[65], a few have included the VCMA effect [66], [67]. It has been experimentally observed that the VCMA effect impacts magnetization dynamics both in the presence of STT, as well as on its own, giving rise to an oscillatory behavior of the switching probability as a function of applied pulse width [68], [69], which differs from that of pure STT and thermally activated STT switchings. Hence, previous models need to be complemented by incorporating the voltage dependence of anisotropy at the interface of the free layer and the tunnel barrier.

In this chapter, we include the VCMA effect as a component of the effective magnetic field \vec{H}_{eff} in an LLG-based macrospin compact model, allowing implementation in a hardware description language such as Verilog-A. The main contributions of this work are as follows. (i) The model calculates quantitative values of the effective magnetic field and its components as a function of applied voltage across the device, predicting the required bias condition in which switching can occur.

(ii) The three-dimensional magnetization dynamics and resistance change can be monitored with respect to time. (iii) The model provides the in-plane external magnetic field dependence of the switching speed. (iv) The model accounts for the write error rate (WER) as a function of the pulse duration, amplitude, and slew rate by including the effect of thermal noise under repeated write trials. (v) The change of the retention time and thermal stability can be monitored under the different bias conditions.

2.2 Physical and Dynamic Model of MTJ

2.2.1 Landau–Lifshitz–Gilbert (LLG) equation

We assume a single-domain MTJ structure where the three-dimensional dynamics of the free layer's magnetic moment $\vec{m} = \{m_x, m_y, m_z\}$, with $|m^2| = 1$, can be described via a Landau-Lifshitz-Gilbert (LLG) equation in the presence of an effective field \vec{H}_{eff} [70].

$$\frac{d\vec{m}}{dt} = -\gamma'(\vec{m} \times \vec{H}_{eff}) - \alpha\gamma'\vec{m} \times (\vec{m} \times \vec{H}_{eff}) \quad (2.1)$$

where α is the material-dependent Gilbert damping factor, \vec{m} is a unit vector in the direction of magnetization, and γ' is the reduced gyromagnetic ratio, which is equal to $(\gamma_e\mu_0)/(1 + \alpha^2)$ in which γ_e is the gyromagnetic ratio and μ_0 is the relative permeability of the free layer. The first term in the equation (2.1) is responsible for precessional motion while the second term provides a damping torque that makes \vec{m} align with \vec{H}_{eff} .

2.2.2 Effective Magnetic Field

The compact model predicts the magnetization vector trajectory by solving the LLG equation in a given electromagnetic bias condition. The time derivative of the magnetic moment $d\vec{m}/dt$ is extracted

based on the effective magnetic field \vec{H}_{eff} in every simulation time step. \vec{H}_{eff} is a combination of the following magnetic anisotropies,

$$\vec{H}_{eff} = \vec{H}_{PMA} - \vec{H}_{VCMA} + \vec{H}_{dem} + \vec{H}_{ext} \quad (2.2)$$

The vector of the effective magnetic field points towards the direction of the energy minima, providing torques to the magnetic moment to be aligned with this direction via precession and damping motions. However, each component of the effective field can vary with respect to time depending on the bias condition, temperature, and its current magnetic moment direction. Hence, adopting a reliable model of each component determines the accuracy of the compact model.

2.2.2.1 Perpendicular Magnetic Anisotropy (PMA)

To make the magnetic layers perpendicular to the plane of thin-films, a sufficient interfacial anisotropy energy should exist in the interface between a MgO and a CoFeB layer to overcome the relatively large demagnetization field in the \hat{z} -axis. Such anisotropy, called perpendicular magnetic anisotropy (PMA), may stem from a preference of the spins to align in the perpendicular direction because of interfacial or magnetocrystalline effects [71] and can be described in the form of an anisotropy field [72].

$$\vec{H}_{PMA} = \frac{2K_i}{t_{fl}\mu_0 M_S} \vec{m}_z \quad (2.3)$$

where K_i is the PMA coefficient (interfacial anisotropy), t_{fl} is the thickness of the free layer, M_S is the saturation magnetization, and \vec{m}_z is the unit vector of the magnetic moment along the \hat{z} -axis.

2.2.2.2 Voltage-Controlled Magnetic Anisotropy (VCMA)

In ultrathin ferromagnetic films (< 2 nm), electric fields can change the magnetic properties via interfacial effects, providing a mechanism for coupling the electric field to the magnetic anisotropy.

The PMA can, therefore, be modulated via the voltage-controlled magnetic anisotropy (VCMA) [24], [73], [74]. The VCMA effect is theoretically explained by the electric field-induced change of occupancy of atomic orbitals [75] and the Rashba spin-orbit coupling at the interface [76]. In general, the VCMA effect may have a nonlinear dependence on the applied voltage. However, in the voltage range of memory applications, the linearized form of the dependence can be considered valid where

$$\vec{H}_{VCMA} = \frac{2\xi V_{MTJ}}{t_{f1}\mu_0 M_S d_{MgO}} \vec{m}_z \quad (2.4)$$

Here, ξ is the VCMA coefficient that is a material dependent parameter quantifying the change of interfacial anisotropy energy per unit electric field, V_{MTJ} is the voltage across the MTJ, and d_{MgO} is the thickness of the tunnel oxide. Depending on the polarity of the applied voltage across the device, the VCMA can either enhance or reduce the PMA.

2.2.2.3 Shape Anisotropy and Demagnetization

The demagnetization field \vec{H}_{dem} has a tendency to reduce the total magnetization and gives rise to shape anisotropy in ferromagnets. In cylindrical or ellipsoid based devices such as MTJs, the strength of the demagnetization field linearly depends on the magnetization with the geometrical demagnetizing tensors described by

$$\vec{H}_{dem,x} = -M_S(N_x \cdot \vec{m}_x) \quad (2.5)$$

$$\vec{H}_{dem,y} = -M_S(N_y \cdot \vec{m}_y) \quad (2.6)$$

$$\vec{H}_{dem,z} = -M_S(N_z \cdot \vec{m}_z) \quad (2.7)$$

where the sum of the demagnetizing tensors N_x , N_y , and N_z is equal to 1 in SI units and 4π in CGS unit.

2.2.2.4 External Field

In the compact model, any external magnetic bias field directly contributes to \vec{H}_{ext} . An in-plane bias field is typically required to realize a uniform voltage-induced precessional switching with defined switching speed. In a practical device for VCMA-based memory applications, the external magnetic field can be incorporated via an extra in-plane exchange-biased ferromagnetic layer within the MTJ stack.

2.2.3 Thermal Noise

The thermal noise \vec{H}_{th} can be included in the precessional term of the LLG equation, following the Langevin approach [77], which captures thermally-induced (e.g., stochastic) processes. Although omitting \vec{H}_{th} in the damping term may alter the particular realizations of magnetization trajectories, the average system property remains the same if the noise power is appropriately rescaled [78]. The stochastic LLG equation is as follows,

$$\frac{d\vec{m}}{dt} = -\gamma \vec{m} \times (\vec{H}_{eff} + \vec{H}_{th}) - \alpha \gamma \vec{m} \times (\vec{m} \times \vec{H}_{eff}) \quad (2.8)$$

This aspect of the model is required to simulate the write error rate (WER). The thermal noise is modeled as a Gaussian noise-like magnetic field [79], where \vec{H}_{th} is given by

$$\vec{H}_{th} = \vec{\sigma} \sqrt{\frac{2k_B T \alpha}{\mu_0 M_S \gamma \mathcal{V} \Delta t}} \quad (2.9)$$

Here, k_B is the Boltzmann constant, T is the temperature, \mathcal{V} is the volume of the free layer, Δt is the simulation time step, and $\vec{\sigma}$ is a unit vector whose \hat{x} , \hat{y} , and \hat{z} components are independent Gaussian random variables with a mean of 0 and a standard deviation of 1. These components are produced using the Verilog-A random number generator functions. Room temperature of $T=300$ K is used throughout the simulation.

2.2.4 Modeling of Spin-Transfer Torque

The magnetization dynamics of the free layer is also affected by a current-dependent spin-transfer torque (STT), which can be included in the LLG equation as a third term [64]:

$$\frac{d\vec{m}}{dt} = -\gamma\vec{m} \times (\vec{H}_{eff} + \vec{H}_{th}) - \alpha\gamma\vec{m} \times (\vec{m} \times \vec{H}_{eff}) + \Gamma_{STT} \quad (2.10)$$

$$\Gamma_{STT} = -\gamma P \frac{\hbar J}{2q t_{fl} \mu_0 M_S} [\vec{m} \times (\vec{m} \times \vec{p})] \quad (2.11)$$

The STT term is derived based on two assumptions: (i) The whole transverse part of the spin current is absorbed next to the interface. (ii) The incident spin direction is aligned along the magnetization of the pinned layer (\vec{p}). Here, J/q represents the number of electrons per unit time and unit area, flowing through the MTJ device. Each electron carries an average angular momentum $P\hbar/2$, where P is the spin-polarization factor, the percentage of electrons whose intrinsic angular momentum (spin) aligns with the magnetization direction as current flows through the MTJ. Hence, $\gamma P(\hbar/2)(J/q)$ is the net flow of magnetization into a unit area, and its ratio of the specific magnetic moment provides the torque to the free layer.

In a perpendicularly magnetized MTJ, the PMA is large enough to cancel the demagnetization field, resulting in the out-of-plane easy axis. The out-of-plane component of the effective field $H_{k,eff}^\perp$ is given by

$$\vec{H}_{k,eff}^\perp = \vec{H}_{PMA} - \vec{H}_{VCMA} + \vec{H}_{dem} \quad (2.12)$$

This creates the energy barrier which needs to be overcome for achieving magnetization reversal. To switch the MTJ via the STT effect, the STT term should be larger than the damping term in the LLG equation, giving rise to the critical current density expression,

$$J_{c_STT} = \frac{2q \alpha t_{fl} \mu_0 M_S H_{k,eff}^\perp}{\hbar \eta} \quad (2.13)$$

where η is the spin transfer efficiency. However, the STT effect is not significant in the voltage-controlled MTJ since the relatively thick tunneling barrier (> 1.5 nm) limits the current density. Also, recent experiments have shown that an additional field-like torque $(\beta_1 + \beta_2 \cdot A \cdot J)(\vec{m} \times \vec{p})$ can be added in the STT term (2.11) where β_1 and β_2 are field-like torque constants, and A is the MTJ area. Although the compact model includes the current-driven field-like torque term, again this effect is also negligible in voltage-controlled MTJ devices.

2.2.5 Modeling of Tunnel Magnetoresistance (TMR) Ratio

The dynamic behavior of the magnetic moment \vec{m} itself does not provide useful information for circuit simulation. Therefore, the motion of \vec{m} needs to be expressed in terms of the MTJ conductance to interact with other electric components. According to the Jullière model [32], the conductance is described by the following equation:

$$G(\theta) = G_T(1 + P^2 \cos\theta) + G_{SI} \quad (2.14)$$

We assume that the spin polarization P in both the pinned and free layers has the same value ($P_1 = P_2$), and G_{SI} , additional conductance resulting from imperfections in the tunneling barrier MgO, is equal to zero. Here G_T is the pre-factor for direct elastic tunneling conductance $G_T^{-1} \approx R_P(1 + P^2)$. θ is the angle between \vec{m} and \vec{p} . In a perpendicular MTJ, $-\cos\theta$ is equal to the magnitude of \hat{z} -component of the magnetic moment \vec{m}_z , which leads to $G(\vec{m}_z) \approx G_T(1 - P^2 \vec{m}_z)$.

Tunnel magnetoresistance (TMR) ratio can be expressed by the spin-polarization factor P [32],

$$\text{TMR} = \frac{R_{AP} - R_P}{R_P} = \frac{2P^2}{1 - P^2} \quad (2.15)$$

A high TMR ($> 100\%$) is desirable for a reliable sensing operation since it causes larger sensing margins. Based on equations (13) and (14), we can derive the final MTJ resistance in our compact model as a function of TMR and \vec{m}_z ,

$$R(\vec{m}_z) = \frac{R_P(1 + P^2)}{1 - P^2\vec{m}_z} = \frac{R_P(1 + \frac{TMR}{TMR + 2})}{1 - \frac{TMR}{TMR + 2}\vec{m}_z} \quad (2.16)$$

If the magnetization of the pinned layer \vec{p} is aligned along the minus \hat{z} -axis ($\vec{p} = -1$), the parallel state corresponds to $R(\vec{m}_z = -1) = R_P$. On the other hand, the anti-parallel state is $R(\vec{m}_z = +1) = R_P(1 + TMR) = R_{AP}$.

2.3 Two Terminal MTJ Switching Mechanisms and Compact Model Simulations

2.3.1 Voltage Dependence of Effective Magnetic Field

The effective magnetic field \vec{H}_{eff} in the free layer of the MTJ is a function of the applied voltage across the device due to the VCMA effect. A quantitative analysis of voltage dependence of each component in \vec{H}_{eff} allows extracting the critical voltage V_c , the minimum voltage that induces the voltage-driven precessional switching.

Figure 2.1 shows each component of \vec{H}_{eff} as a function of the applied voltage in terms of an anisotropy field [$\text{kA/m} = 12.56 \text{ Oe}$] based on the parameters of Table 2.1, assuming that the data is captured immediately after applying voltage where the magnetic moment of the free layer is aligned with the out-of-plane direction ($\vec{m}_z = 1 \text{ or } -1$). Since $H_{k,eff}^\perp$ is the dominant component at zero bias

Parameters	Symbol	Value	Unit
MTJ diameter	l	60	nm
MgO thickness	d_{MgO}	1.62	nm
Free layer thickness	t_{fl}	1.1	nm
TMR	TMR	100	%
Temperature	T	300	K
Damping factor	α	0.02	
Saturation magnetization	M_s	1.2×10^6	A/m
PMA coefficient	K_i	1.06×10^{-3}	J/m ²
VCMA coefficient	ξ	61	fJ/V · m
Demagnetizing tensor (z)	N_z	0.96	
Demagnetizing tensor (x,y)	$N_{x,y}$	0.02	

Table 2.1 Parameters of the macrospin compact model.

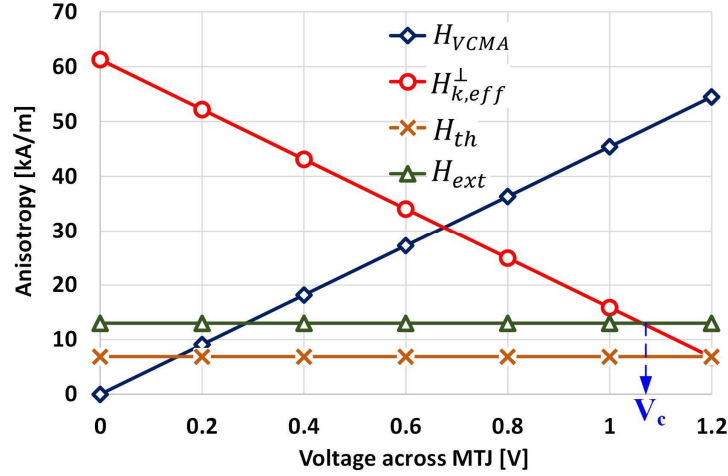


Fig. 2.1 Voltage dependence of the components of the effective magnetic field. $H_{k,eff}^{\perp}$ is reduced as the applied voltage increases due to H_{VCMA} . If $H_{k,eff}^{\perp}$ is smaller than H_{ext} , a voltage-driven precessional switching can occur where the applied voltage across the MTJ at that point is defined as the critical voltage V_c . We introduced the amplitude of the thermal noise for the comparison to other components.

due to the PMA, the easy axis of the device is in the out-of-plane direction. Depending on the amplitude of the applied voltage, switching can occur via thermal activation or a precessional reorientation. In particular, the latter can happen if the applied voltage reaches the critical voltage V_c , in which $H_{k,eff}^{\perp}$ is smaller than the in-plane external field H_{ext} . In this bias condition, the easy axis temporarily changes to the in-plane, causing the precessional dynamic motion along with the in-plane external field H_{ext} .

2.3.2 Timing of Voltage-Driven Precessional

In addition to the voltage amplitude across the device ($V_{MTJ} > V_c$), the pulse duration also plays a significant role in precessional reorientation. To achieve switching, the pulse needs to be removed when the magnetic moment achieves 180° reorientation. If the pulse duration is too long or too short, the magnetic moment may return to its initial state. Figure 2.2 illustrates the VCMA-induced precessional switching mechanism in the free layer of the MTJ.

Figure 2.3 shows the transient simulation of the compact model where the first two pulses with the half cycle of precession switch the MTJ state from P to AP and from AP to P, respectively. Because of such non-deterministic and unipolar switching, the initial state of the MTJ needs to be known to

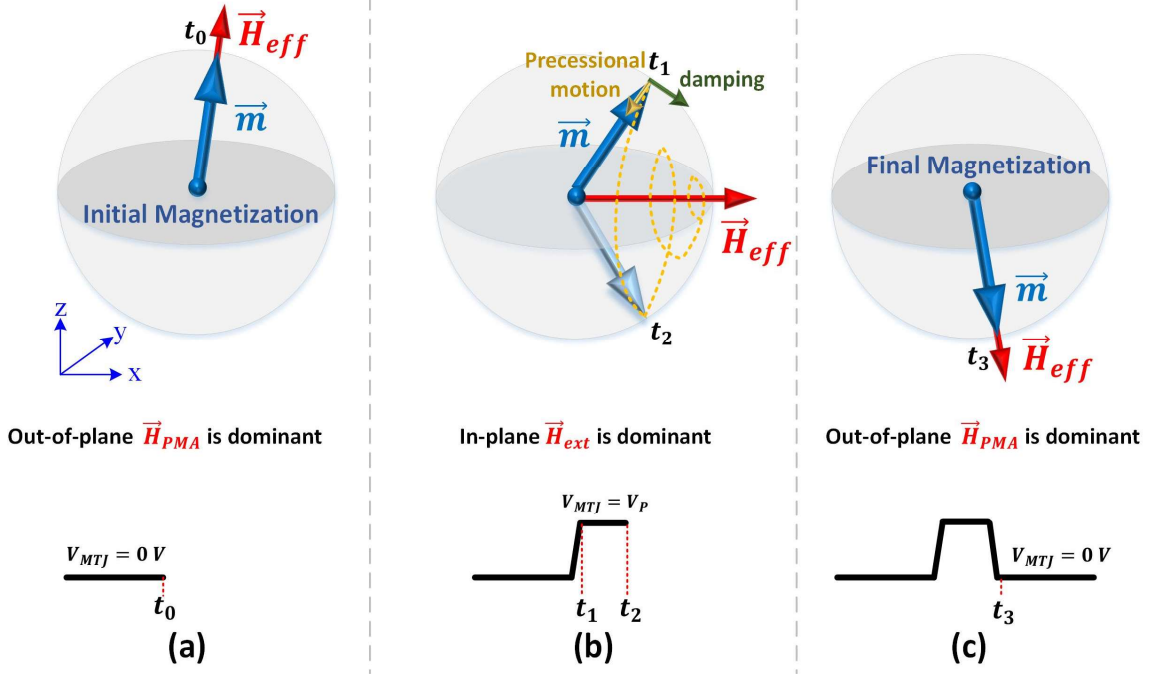


Fig. 2.2 Illustration of the voltage-induced precessional switching mechanism in the free layer of a perpendicularly magnetized MTJ. (a) Under zero electric bias condition ($V_{MTJ} = 0 V$ at $t < t_0$), the free layer is aligned with the out-of-plane direction because the perpendicular magnetic anisotropy \vec{H}_{PMA} is a dominant component in $\vec{H}_{k,eff}^\perp$. (b) When an applied voltage across the device reduces $\vec{H}_{k,eff}^\perp$ via the VCMA effect, the magnetic moment starts to precess around the in-plane direction. (c) If the width of the applied pulse is designed to coincide with half the precession period, a full 180° switching can be achieved. Note that voltage with opposite polarity cannot switch the device because it enhances $\vec{H}_{k,eff}^\perp$.

determine whether a reversal of the free layer is necessary to write the desired information. However, the unipolar writing process allows having a diode-controlled memory device for crossbar arrays, increasing the density and scalability of the overall memory [52]. Alternatively, in a 1T-1MTJ implementation, the reverse voltage can be used for high-speed and disturb-free sensing [80]. The third pulse fails to switch the MTJ since its duration is equal to the time of a full round trip, resulting in $P \rightarrow AP \rightarrow P$.

The bottom two simulations of Fig. 2.3 presents components of the effective field. When ($V_{MTJ} > V_c$), the VCMA cancels out the sum of the PMA and the demagnetization field. Hence, the out-plane component of effective field $H_{k,eff}^\perp$ becomes smaller than the external field H_{ext} , fulfilling the condition for the precessional switching.

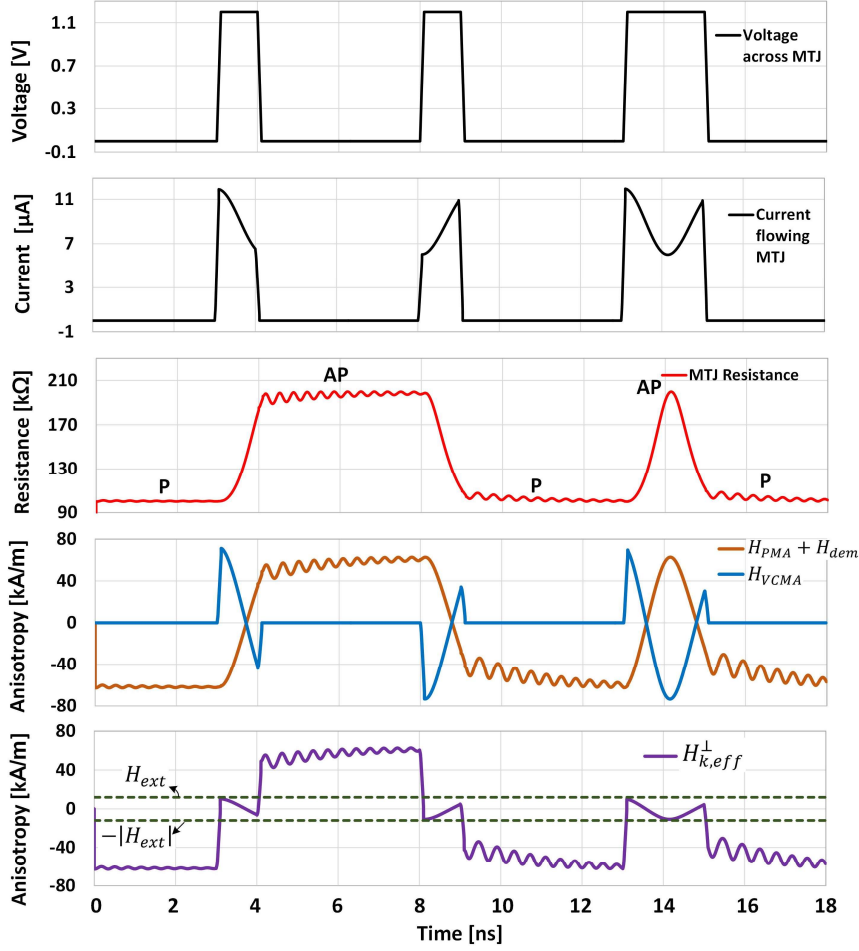
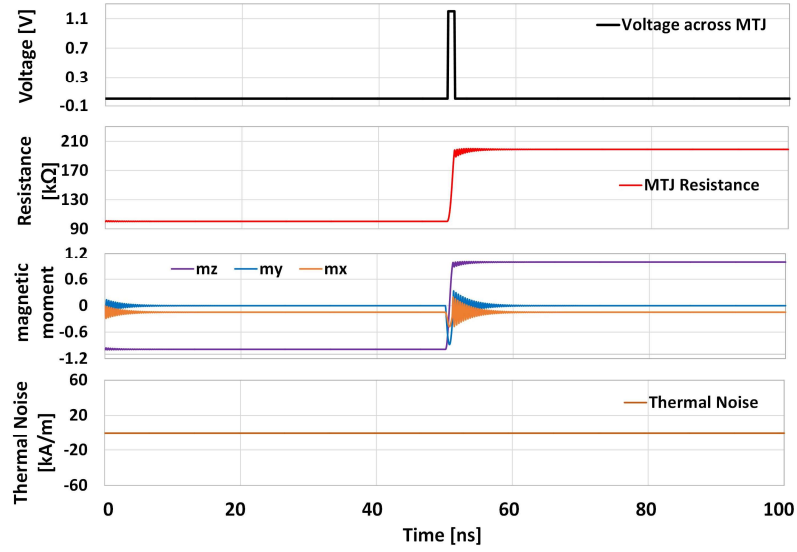


Fig. 2.3 Transient simulation of the non-deterministic voltage-driven precessional switching in which a unipolar pulse can switch either P to AP or AP to P. If the pulse duration is equal to the round trip precession time, the switching might not occur since the magnetic moment can return to the original state. While applying the write pulse, $\vec{H}_{k,eff}^\perp$ becomes smaller than H_{ext} , causing a precessional motion along with the in-plane H_{ext} .

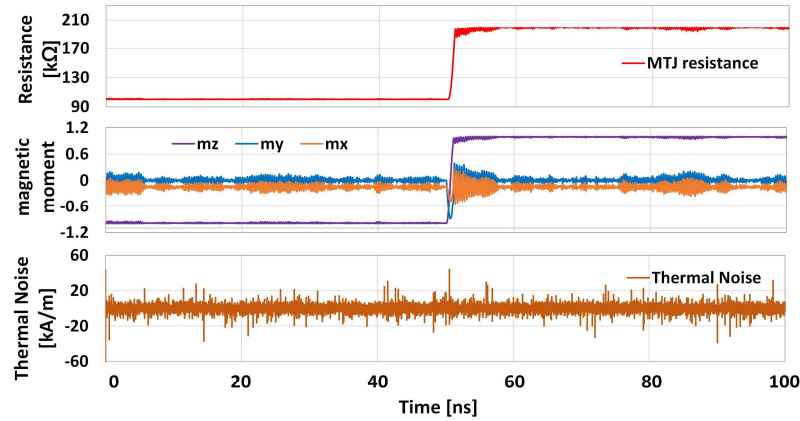
2.3.3 Thermal Noise Effect

The thermal noise influences the dynamic motion of the free layer magnetic moment by randomly modulating the precessional term of the LLG equation. In our model, the thermal noise is included in each \hat{x} , \hat{y} , and \hat{z} components of the LLG equation, resulting in a stochastic switching behavior. This allows us to obtain write error rates (WERs) in various electrical and magnetic bias conditions.

The static state of the MTJ is also affected by the thermal noise. Even under zero bias, the thermal noise has a possibility to overcome the energy barrier between the two states, switching the device to



(a)



(b)

Fig. 2.4 Transient simulation (a) in the absence of the thermal noise (b) in the presence of the thermal noise. The thermal noise influences both the dynamic and static behaviors of the magnetic moment, resulting in stochastic switching.

the other state. The time interval of this switching process is the retention time. Figure 2.4 shows the thermal noise effect on the dynamic and static state of the device in terms of the magnetic moment and resistance.

2.3.4 External Magnetic Field Dependence of Switching Speed

The switching speed of the voltage induced precessional switching depends not only on the applied pulse across the MTJ but also on the amplitude of the external field H_{ext} since the effective field in

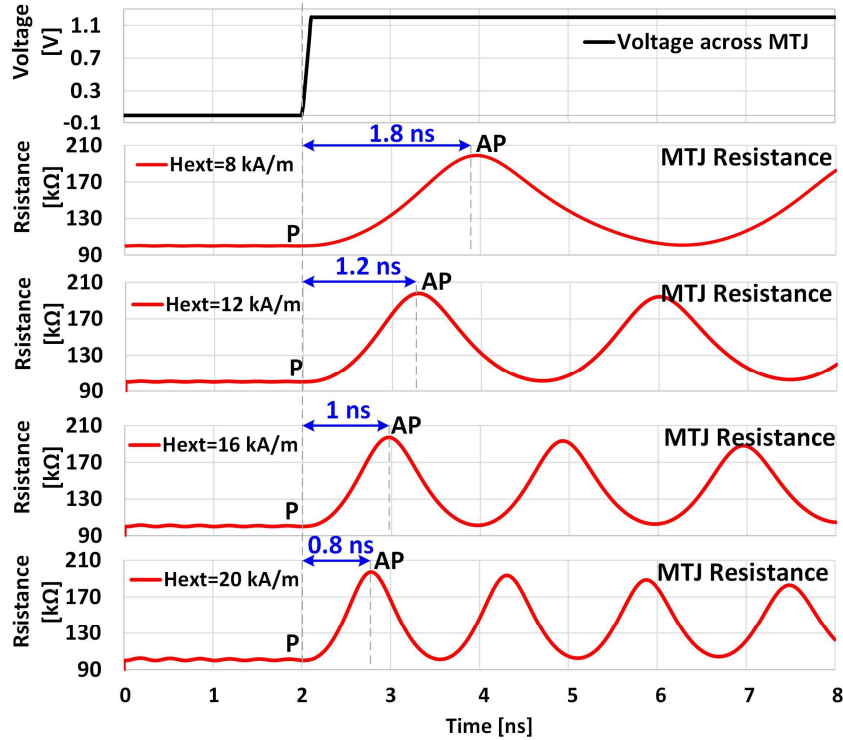


Fig. 2.5 Transient simulations for the in-plane external magnetic field H_{ext} dependence of the switching speed. As the H_{ext} increases, the switching speed increases.

the precessional term of the LLG equation provides a net torque to the magnetic moment. Assuming the condition that the applied voltage is larger than the critical voltage ($V_{MTJ} > V_C$), the switching speed is proportional to the magnitude of the in-plane component of the external field. The compact model provides the switching speed as a function of the in-plane external field (\hat{x} -axis) as shown in Fig. 2.5. Sub-nano second switching can be achieved if the H_{ext} is larger than 16 kA/m ($\sim 200 \text{ Oe}$). However, if the external field has a relatively strong in-plane component, it tilts the magnetization of the free layer at the equilibrium, reducing the thermal stability.

2.3.5 Write Error Rate (WER)

The write error rate ($\text{WER} = 1 - P_{sw}$, where P_{sw} is the switching probability) is defined as the number of switching failures divided by a total number of write trials in a single cell. The access time of a memory system is largely affected by the WER since multiple write operations are often necessary

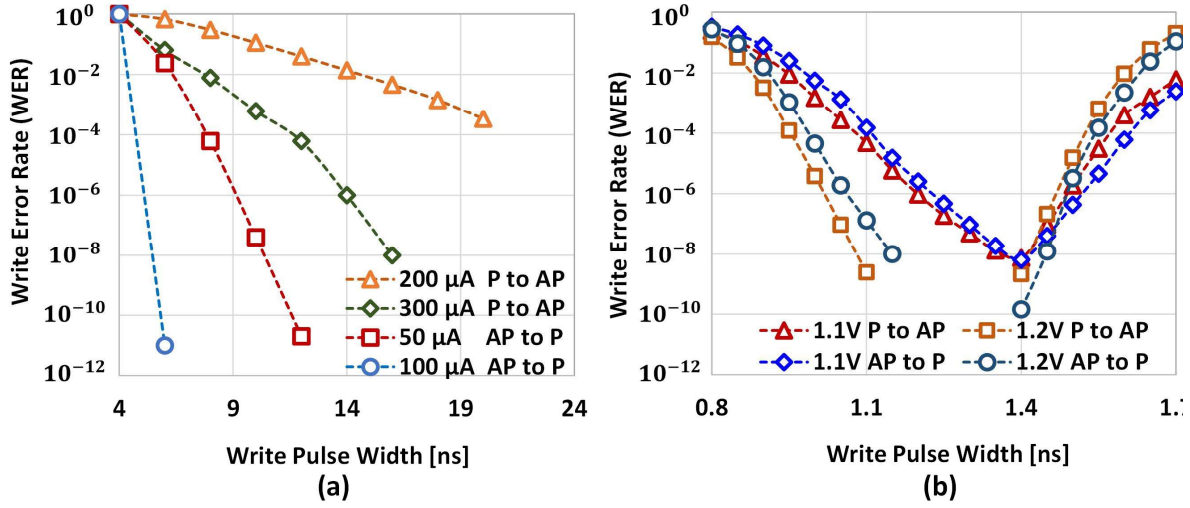


Fig. 2.6 (a) WER of STT-induced switching as a function of the duration and amplitude of the write current, and (b) Oscillatory WER of voltage-controlled precessional switching as a function of the width and amplitude of write voltage.

to achieve a desired bit error rate if the WER is high. Typically, a WER in the order of $\sim 10^{-9}$ is required for working memory applications. The magnitude of the WER also depends on the performance and acceptable overhead of the error correction code (ECC) algorithm built in the memory system [81], [82].

There are differences between the STT-induced switching and the VCMA-driven precessional switching in terms of WER, even though both are stochastic processes. Specifically, the WER of STT-induced switching in the thermally-activated regime is exponentially reduced by increasing the write time and/or current density as shown in Fig. 2.6(a). However, the VCMA-driven switching shows an oscillatory behavior of the WER as a function of the write pulse duration due to the precessional motion of the magnetization as shown in Fig. 2.6(b) [68], [69], [83]. Note that the WERs of the both cases also rely on device characteristics such as the damping factor, thermal stability, and device dimension [84].

In addition to the timing of the applied pulse for the VCMA-driven precessional switching, the dynamics of the magnetic moment is also largely affected by the pulse shape, in particular, the rising

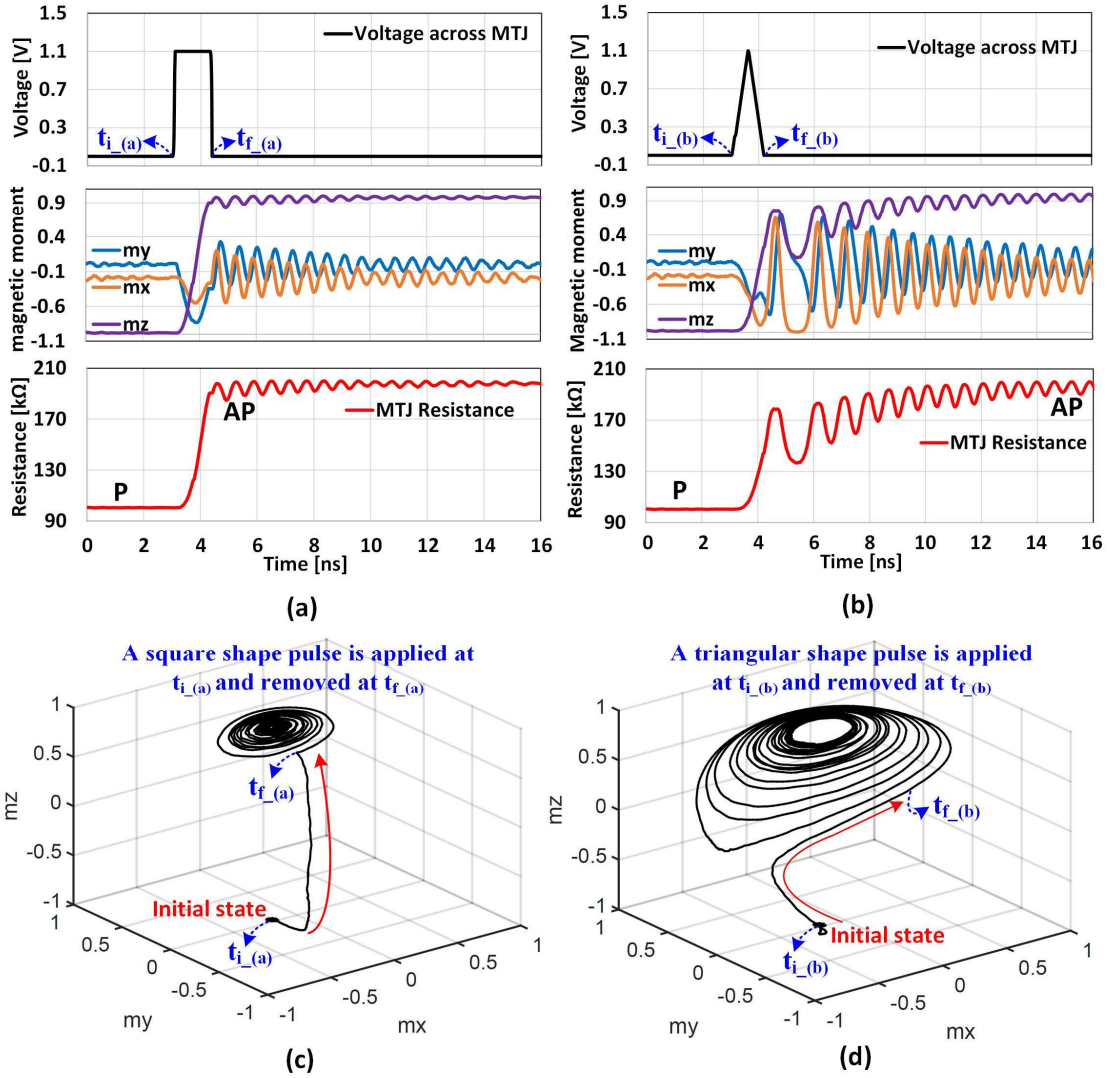


Fig. 2.7 Compact model transient simulations for (a) a square shape write pulse and (b) a triangular shape write pulse. (c) The magnetic moment based on the square pulse shows a more stable precessional trajectory compared to (d) the triangular pulse based one. This allows the square pulse driven switching to have a low write error rate by reducing the susceptibility to noise.

time, falling time, and amplitude of the pulse. To create a stable precessional motion, \vec{H}_{eff} needs to be a constant field, pointing in the in-plane direction, during the electric bias condition. Otherwise, the trajectory of the magnetic moment may be altered, resulting in higher error probability. Figure 2.7 shows simulation results of the magnetization dynamics based on the proposed model where a square shaped pulse (see Fig. 2.7(a)) and a triangular shaped pulse (see Fig. 2.7(b)) are applied. Their three-dimensional magnetic moment trajectories are presented in Figure 2.7(c) and (d), respectively.

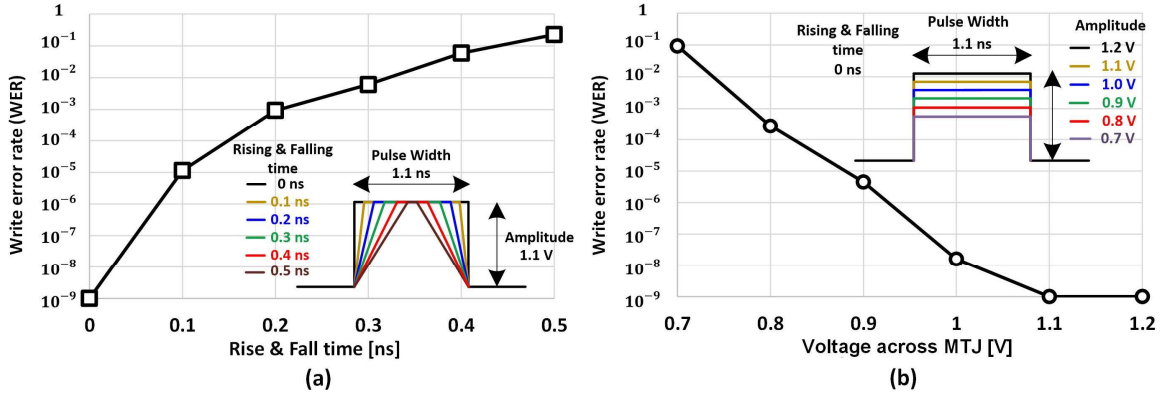


Fig. 2.8 WER simulation of voltage-controlled precessional switching with an ideal voltage source (a) as a function of rising and falling time (slew rate), and (b) as a function of amplitude.

To evaluate the effect of the slew rate and amplitude on the WER, we independently executed the macrospin simulation with 10^9 trials. Figure 2.8 shows that both factors influence the WER in a significant manner. This is because the energy barrier of the free layer linearly decreases as the applied voltage increases, and hence the slew rate fundamentally decides the effective pulse duration.

2.3.6 Thermal Stability and Retention Time

The thermal stability Δ is responsible for the ability of a memory element to maintain its current state at a certain temperature and thereby determines the retention time. The thermal stability of a perpendicularly magnetized MTJ is obtained via the following equation (assuming $H_{k,eff}^\perp \gg H_{ext}$),

$$\Delta = \frac{E_b(V_{MTJ})}{k_B T} = \frac{\mu_0 M_s H_{k,eff}^\perp(V_{MTJ}) \mathcal{V}}{2k_B T} = \frac{[K_i - \xi \frac{V_{MTJ}}{d_{MGO}} - \frac{\mu_0 M_s^2 t_{fl} (N_z - N_{x,y})}{2}] A}{k_B T} \quad (2.17)$$

where E_b is the energy barrier between the two stable states of the MTJ, k_B is the Boltzmann constant, T is the temperature, and \mathcal{V} is the volume of the free layer. We can also explain the VCMA effect in the context of the energy barrier change. Since $H_{k,eff}^\perp$ is a function of the applied voltage V_{MTJ} , the energy barrier can be modulated by V_{MTJ} as shown in Fig. 2.9(a). For a non-zero energy barrier, the switching can occur through thermal fluctuations. If the energy barrier is completely removed (the

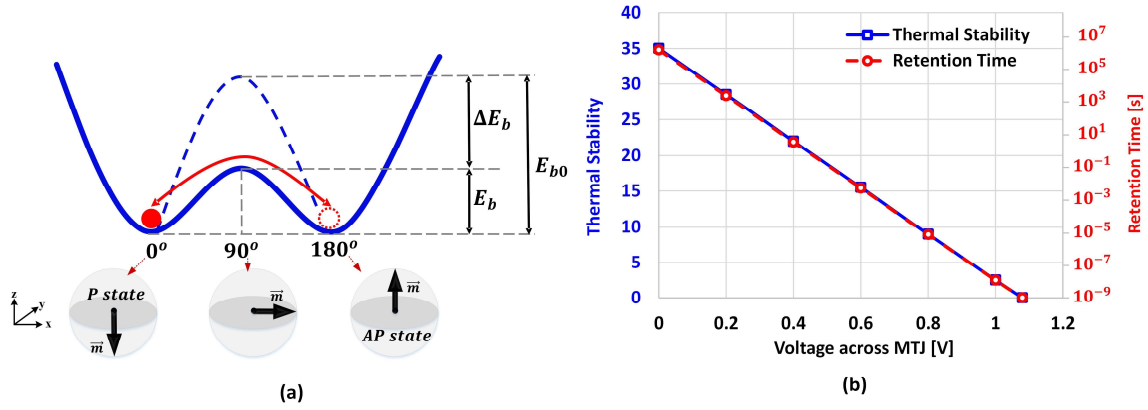


Fig. 2.9 (a) Schematic of the energy barrier E_b lowering via the VCMA effect. The reduced energy barrier provides more chance to switch via thermal activation, shortening the retention time. (b) Voltage dependence of the thermal stability and its corresponded retention time. The PMA, demagnetization field, and the volume of the free layer mainly determine the amplitude of the thermal stability. The slope of the line is changed by the VCMA coefficient

condition for precessional switching) in the presence of an in-plane external magnetic field \vec{H}_{ext} , the magnetization of the device is mainly governed by \vec{H}_{ext} .

At zero voltage bias condition with the given MTJ parameters in Table 2.1, the thermal stability is calculated as 35, which is suitable for working memory applications [85]. However, for storage class applications, the value of the thermal stability should be larger than 40~60 depending on the capacity of the memory chip. Figure 2.9(b) shows the voltage dependence of the thermal stability and retention time with a VCMA coefficient of 61 [fJ/V·m], which determines the slope of the line in the graph. In the applied voltage range ($0 < V_{MTJ} < 1.0$), the thermal stability is modulated from 35 to 0, and the retention time decreases by ten orders of magnitude via the following equation $\tau = \tau_0 \exp(\Delta)$, where τ_0 is equal to 1 ns.

2.4 Three Terminal MTJ Switching Mechanisms and Compact Model Simulations

2.4.1 Modeling of Spin Hall Effect

Three types of torques (VCMA, STT, SHE) are implemented into a macrospin Verilog-A based 3-terminal MTJ compact model. An electric field modulates the perpendicular magnetic anisotropy

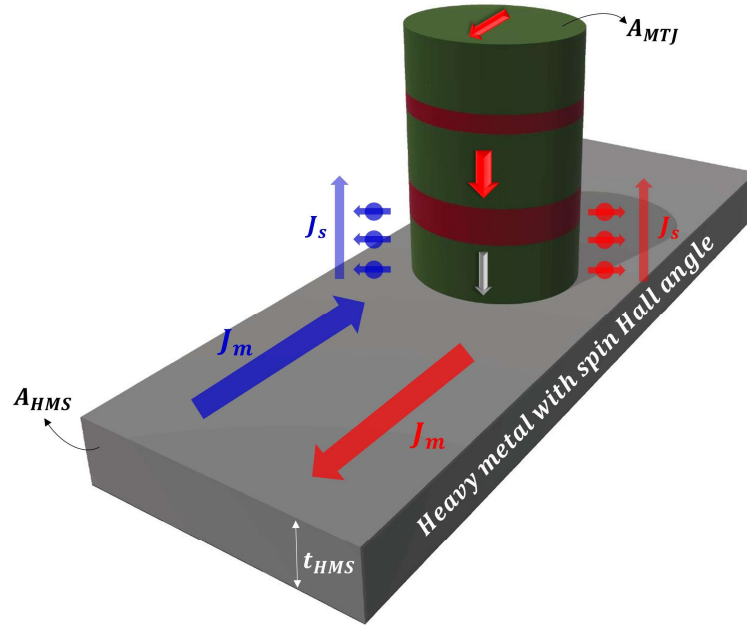


Fig. 2.10 Schematic of the 3-terminal MTJ on the heavy metal layer with spin-orbit interaction (i.e. Tungsten, W) for the SHE effect. A charge current I_m transfers the spin torque via a polarized (in-plane) spin current. The fixed symmetry-breaking in-plane magnetic field is needed to set the direction of the SHE based switching in the perpendicularly magnetized MTJ.

(PMA) of a ferromagnetic film and has been modeled by a voltage-dependent effective field. The STT and SHE are similar in that both generate spin currents resulting in (anti-)damping-like spin torques. STT relies on spin polarization created via interactions of electrons with a pinned layer and the subsequent transfer of said electrons into the free layer (with the charge current flowing through the pinned layer and then tunneling through the MgO and then into the free layer) where due to the polarization of the charge current, a spin current is also produced. The SHE, on the other hand, generates a spin current at the interface between the free layer and heavy metal in the presence of an in-plane charge current passing through the heavy metal thin film (no charge current passes through the MTJ) or topological insulator. Field-like spin torques were not considered in the present study.

Figure 2.10 shows the 3-terminal MTJ structure where the CoFeB/MgO/CoFeB based MTJ is placed on top of the heavy metal with spin-orbit interaction (e.g. Ta, Pt), giving rise to a sizeable spin Hall effect (hereafter, the HMS layer). Since the MgO thickness is 1.5 nm, which produces a relatively

large RA ($>10^3 \Omega \cdot \mu m^2$), the STT effect is negligible. To deterministically switch a perpendicularly magnetized MTJ by using the SHE effect, in-plane magnetic field is necessary to break the symmetry with respect to spin torque [86]. This in-plane magnetic field is generated by an additional pinned layer, which can be fabricated as part of the MTJ.

The circuit simulator with the 3-terminal compact model extracts the charge current (I_m) flowing through the HMS layer and calculates spin current by using the following equation [87],

$$I_s = \frac{A_{MTJ}}{A_{HMS}} \theta_{SHE} [1 - \operatorname{sech}\left(\frac{t_{HMS}}{\lambda_{sf}}\right)] I_m \quad (2.18)$$

where A_{MTJ} and A_{HMS} are the cross-sectional areas of the MTJ and the HMS layer, respectively, θ_{SHE} is the spin Hall angle, t_{HMS} is the thickness of the HMS layer, and λ_{sf} is the spin flip length. The compact model uses the obtained spin current (I_s) as one of the variables in the Landau-Lifshitz-Gilbert (LLG) equation to predict the dynamics of the free layer magnetic moment at a given magnetic field and electric bias condition, which is given by

$$\frac{d\vec{m}}{dt} = -\gamma \vec{m} \times (\vec{H}_{eff} + \vec{H}_{th}) - \alpha \gamma \vec{m} \times (\vec{m} \times \vec{H}_{eff}) + \Gamma_{STT} + \Gamma_{SHE} \quad (2.19)$$

$$\Gamma_{SHE} = -\gamma \frac{\hbar J_s}{2 q t_{fl} \mu_0 M_S} [\vec{m} \times (\vec{m} \times \vec{\sigma})] \quad (2.20)$$

where J_s is the spin current density (I_s/A_{MTJ}), and $\vec{\sigma}$ is the polarization orientation of the pure spin current induced by SHE [88], [89]. It is noteworthy that the direction of the torque induced by the SHE in the perpendicular MTJ geometry is along with $\vec{m} \times (\vec{m} \times \vec{\sigma})$ and thus does not compete with the damping torque directly [90]. In this case, the critical current density is independent of the damping factor but depends on the effective anisotropy field of the free layer, which is given by

$$J_{C_SHE} = \frac{2q t_{fl} \mu_0 M_S}{\hbar \theta_{SH}} \left[\frac{H_{k,eff}^\perp(V_{MTJ})}{2} - \frac{H_{ext}}{\sqrt{2}} \right] \quad (2.21)$$

If the external magnetic field H_{ext} points to the in-plane direction, the critical current density is linearly modulated as a function of the H_{ext} . Since the perpendicular component of the effective field $H_{k,eff}^\perp$ is a function of the applied voltage across the MTJ via the VCMA effect, the critical current density can also be modulated by V_{MTJ} . Hence, it is possible to combine both SHE and the VCMA effect to reconfigure the magnetization of the MTJ, which we refer to as gate-voltage-modulated SHE switching (V-SHE).

2.4.2 SHE-driven Switching and VCMA Assisted SHE-driven Switching

For the macrospin compact model simulation, we assume that the spin Hall angle is 0.3, the area of the HMS layer is 80 nm × 5 nm, and the spin flip length is 1.5 nm. Fig. 2.11 shows simulation results

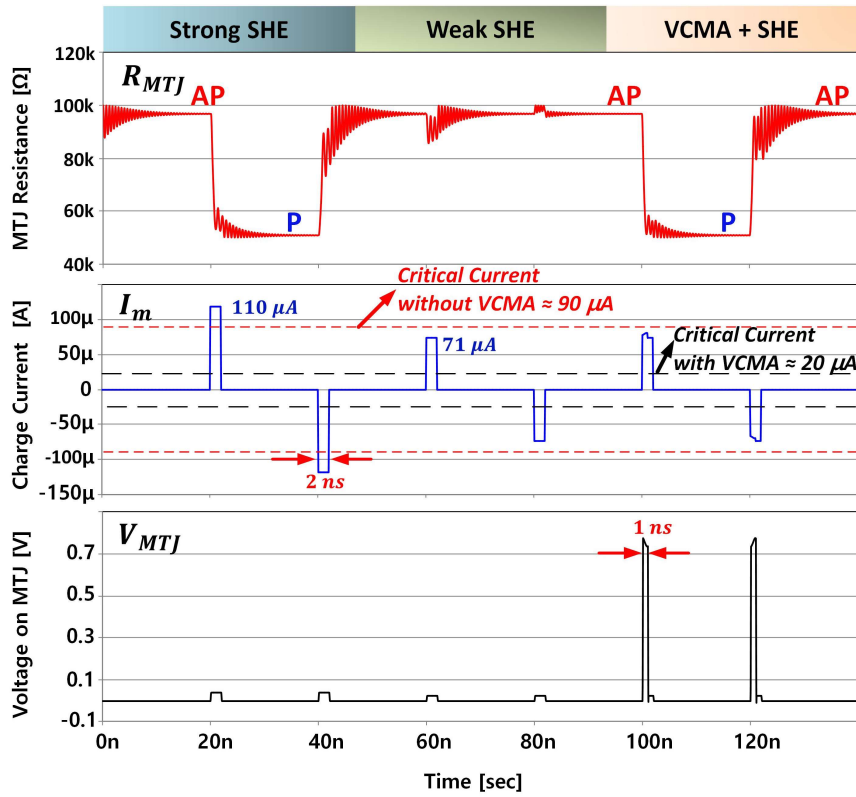


Fig. 2.11 Simulation results of pure SHE switching and gate-voltage-modulated SHE switching in the presence of the external in-plane magnetic field. If I_m is larger than the absolute value of the critical current, it switches the MTJ state from AP to P as well as from P to AP (strong SHE). Below the critical current, switching does not occur (weak SHE). However, applying a pulse to the MTJ decreases the critical current, which induces a switching by using a relatively smaller current.

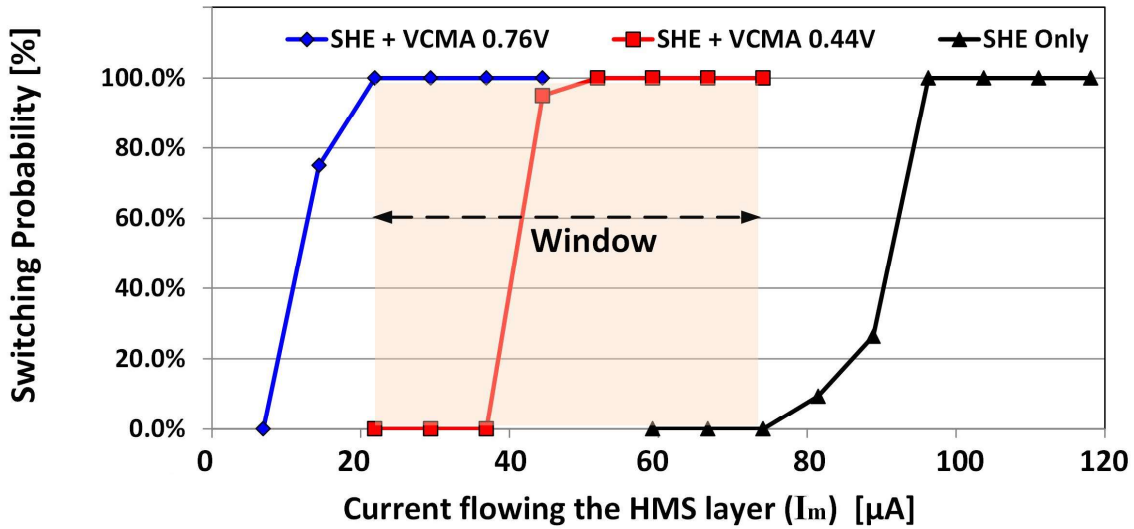


Fig. 2.12 Switching probability as a function of current (I_m) amplitude and applied voltage (V_{MTJ}). The switching probability for each condition was extracted from 1,000 attempts based on the 3-terminal MTJ compact model simulation. As the applied voltage (V_{MTJ}) increases, less current (I_m) is necessary to switch the MTJ state.

for a 3-terminal MTJ with a critical current for SHE switching of $90 \mu A$ in the absence of any voltage applied across the MTJ structure and in the presence of 100 Oe in-plane magnetic field. If the amplitude of the charge current flowing through the HMS layer with 2 ns duration is larger than the critical current, I_m can switch the MTJ both from high resistance state (denoted as AP) to low resistance state (denoted as P) and from P to AP depending on the current direction. However, it is possible to modulate the critical current for the SHE effect via the VCMA mechanism by applying a voltage across the MTJ. By varying the perpendicular magnetic anisotropy of the free layer, the critical current, therefore, becomes a function of the applied voltage across the device. This enables local gating of switching in structures where multiple MTJs are placed on top of a single HMS layer. In other words, we can selectively switch MTJs based on the presence of the applied voltage. Specifically, as shown in Fig. 2.12, applying 0.76 V across the MTJ reduces the critical current from $90 \mu A$ to $20 \mu A$, resulting in a wide bias window for controlling the switching current.

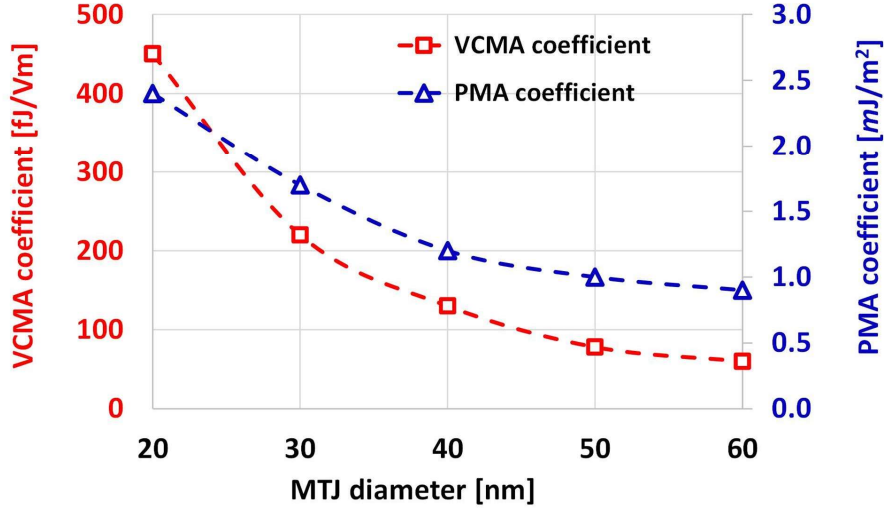


Fig. 2.13 Required VCMA coefficient and interfacial anisotropy in the scaled MTJ while maintaining the same value of the critical voltage ($V_C/d_{MgO} = 1$ V/nm) and thermal stability ($\Delta_0 = 40$).

2.5 Scalability of Voltage-Controlled MTJ

The thermal stability Δ is one of the most significant metrics that evaluate a memory cell characteristics, especially, for retention time, and can be calculated via the equation (2.17). Since the $H_{k,eff}^\perp$ is a function of the voltage across the MTJ (V_{MTJ}), the thermal stability Δ also changes with respect to V_{MTJ} . The ratio between critical switching current and the thermal stability (I_C/Δ_0) is an indicator of the scalability of STT-RAM where the Δ_0 is thermal stability at zero bias. Similarly, the scalability of MeRAM can be analyzed by the analogous critical voltage over thermal stability (V_C/Δ_0), meaning that any voltage larger than V_C can reconfigure the magnetic easy-axis to in-plane at a given thermal stability [91]. As shown in the equation (2.17), the thermal stability of the MTJ is a function of the energy barrier, which is proportional to $K_i A V_{MTJ}$ where the A is the device area. Hence, the interfacial anisotropy K_i needs to be increased as the MTJ size scales while maintaining the same level of thermal stability.

The ratio of critical voltage over thermal stability V_C/Δ_0 can be represented as $d_{MgO} k_B T / \xi A$. Therefore, as the device area scales down, the VCMA coefficient needs to be increased quadratically

to compensate for the reduction of the area of MTJ to keep the same rate of the energy barrier controllability by using the applied voltage. Figure 2.13 shows the required VCMA coefficient and interfacial anisotropy as the MTJ size is scaled while maintaining the same value of the critical voltage ($V_C/d_{MgO} = 1$ V/nm) and thermal stability ($\Delta_0 = 40$).

CHAPTER 3

VOLTAGE-CONTROLLED SPINTRONICS-CMOS CIRCUITS

3.1 Introduction

The macrospin compact model of MTJ including the VCMA effect has been successfully implemented into the circuit design platform, Cadence Virtuoso, based on Verilog-A programming language. In this chapter, we introduce several spintronics-CMOS circuits that show promising performance in terms of power, area, and throughput while achieving desired functions correctly. The proposed circuits exploit following physical phenomenon: (i) the applied voltage across the MTJ can either enhance or weaken the energy barrier (coercivity) of the device depending on the polarity of the voltage. (ii) The modulated energy barrier results in creating different switching probability and retention time. (iii) Combining the current-driven switching mechanisms (STT, SHE) with the voltage-driven switching causes a deterministic switching while enabling parallel configuration of multiple MTJ devices and consuming low switching energy. This chapter may provide a design methodology how a conventional CMOS circuitry benefits from the voltage-controlled spintronic devices.

3.2 Voltage-controlled MTJ based Ternary Content-Addressable Memory

3.2.1 Overview of TCAM

Ternary content addressable memory (TCAM) is an associative computing module, which has many practical applications such as in anti-virus scanners, IP filters, and network switches due to its ultra-fast, fully-parallel searching scheme [92]. The most important qualification for a TCAM cell is fast operation speed for data searching. Due to this reason, SRAM has been widely used in memory elements of the conventional TCAM cell, even though it has high bit-cell cost, typically requiring 12~16 transistors per cell as shown in Fig. 3.1 [93], [94]. However, as CMOS shrinks to nanometer-

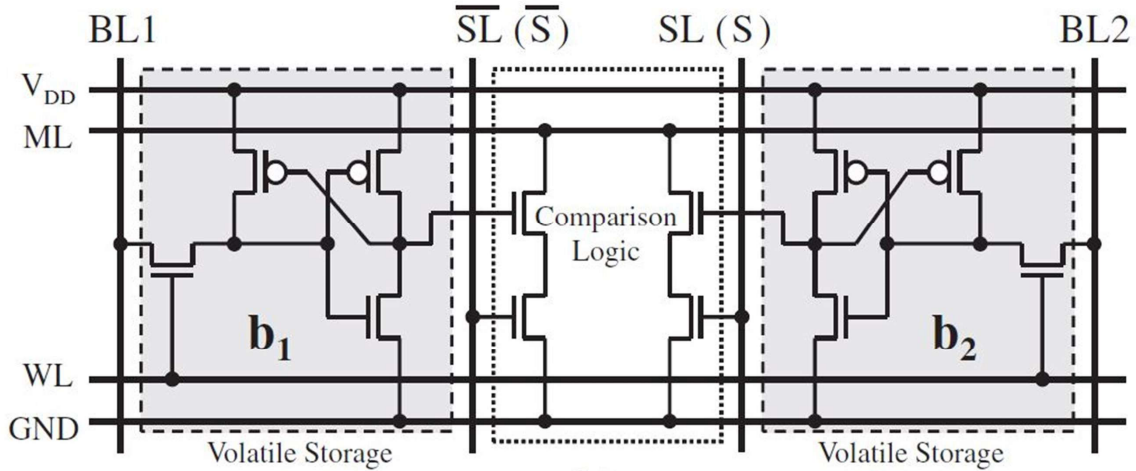


Fig. 3.1 Conventional SRAM based TCAM cell architecture consisting of two volatile storage elements and comparison logic [94].

Stored Data	(b1,b2)	(\overline{SL} , SL)	ML
0	(0,1)	(1,0)	High (Match)
	(0,1)	(0,1)	Low (Mismatch)
1	(1,0)	(1,0)	Low (Mismatch)
	(1,0)	(0,1)	High (Match)
Don't care (X)	(0,0)	(1,0)	High (Match)
	(0,0)	(0,1)	High (Match)

Table 3.1 Truth table of TCAM cell including Don't care condition.

scale, the other major issue has emerged: a high standby power due to leakage current. A scaled-down channel length increases the leakage current, and hence the use of SRAM in TCAM applications is not a sustainable pathway. Table 3.1 shows the truth table of TCAM cell for its search operation.

3.2.2 Design of MTJ based TCAM Cell

To design a low-standby power, high-speed accessibility, and low bit-cell cost TCAM cell, we propose a voltage-controlled MTJ based TCAM, referred to as magnetoelectric TCAM, or MeTCAM. A MeTCAM cell consists of four transistors and two MTJs, i.e. a 4T-2MTJ structure, as shown in Fig.

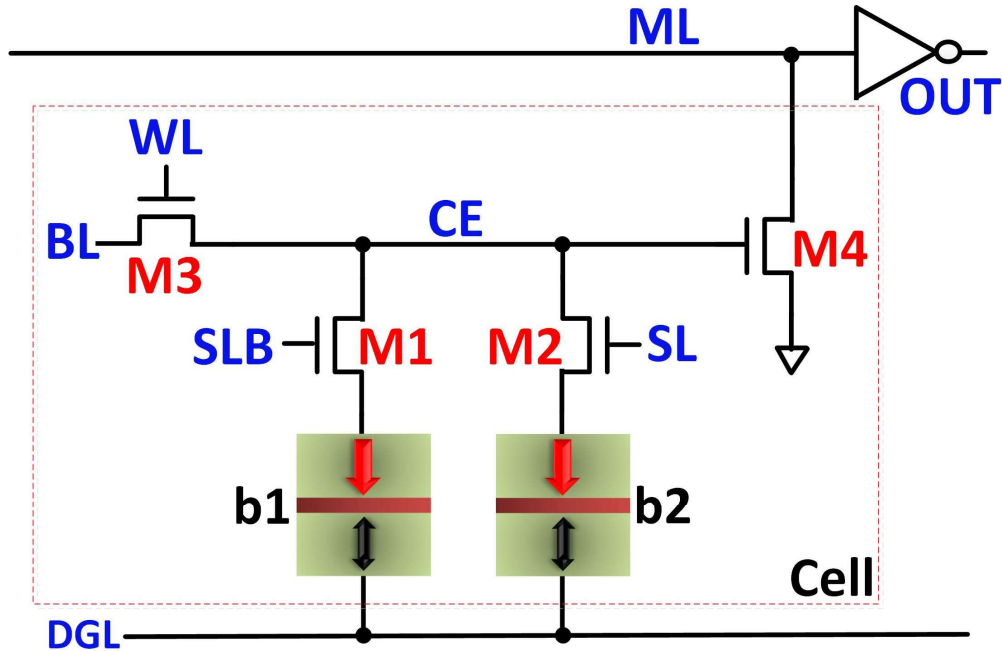


Fig. 3.2 Voltage-controlled MTJ based MeTCAM cell architecture. M1 and M2 are the comparison transistors which are connected to SLB and SL. Based on the combination of the stored data in b1 and b2 with search data, the circuit determines the ML state. M3 is used for the write operation, applying short pulses to induce the precessional switching. During the search operation, DGL pre-charges the CE node, reducing disturbance.

3.2. M1 and M2 are comparison transistors whose gates are connected to the search lines (SL and SLB), and the sources are connected to the pinned layers of the MTJs. M3 is an access transistor, which is required for the configuration (write) operation, and shared by two storage elements (MTJs), b1 and b2, to reduce the cell area. M4 is a match line (ML) driver transistor, which determines the state of the ML between ‘0 and ‘1 based on the potential of the center (CE) node during the search operation.

Figure 3.3 shows the voltage dependence of the coercivity (energy barrier) in MTJ devices. Applying a positive voltage on top of the MTJ decreases the coercivity of the free layer due to the accumulation of electrons between the interface of MgO/CoFeB [26]. On the other hand, applying a positive voltage to the bottom side makes the free layer energetically more stable by increasing its coercivity. The voltage dependence of the coercivity is an important mechanism for the operation of the proposed MeTCAM cell.

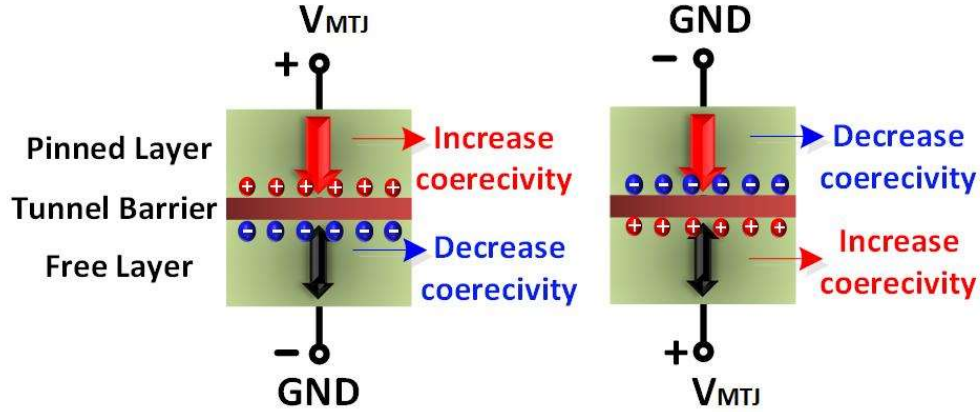


Fig. 3.3 Voltage dependence of the coercivity. Based on the polarity of the applied voltage, the coercivities of the free layers are changed, which can be used in both the configuration and search operations for the proposed MeTCAM application.

3.2.3 Configuration and Search Operations

In the configuration operation (write operation for MTJs), a two-step write method is used where writing of the memory elements b1 and b2 is performed in a serial manner. The pre-read step is necessary to deal with the non-deterministic behavior of the VCMA-driven precessional switching before the MTJ is written. Typically, a pulse with ~ 1 V amplitude and 1 ns duration should be applied to the MTJ to achieve precessional switching. To generate the write pulse in the proposed cell, the dynamic ground line (DGL) node is discharged to ground level, and M1 or M2 must be turned on to electrically connect between the CE node and DGL node. Then, the write pulse applied to the BL node is propagated to the CE node through M3 as shown in Fig. 3.2.

		BL	WL	SLB	SL	DGL
1 st step for b1 (3 ns)	Pre-read (2 ns)	V_{read}	V_{dd}	V_{dd}	0 V	0 V
	Write (1 ns)	V_{write}	V_{dd}	V_{dd}	0 V	0 V
2nd step for b2 (3 ns)	Pre-read (2 ns)	V_{read}	V_{dd}	0 V	V_{dd}	0 V
	Write (1 ns)	V_{write}	V_{dd}	0 V	V_{dd}	0 V

$V_{write} = 1.2$ V with 1 ns duration.

$V_{read} = 0.6$ V with 2 ns duration.

Table 3.2 Bias conditions for the MeTCAM configuration mode.

The reduction of the cell area is achieved by using the shared access transistor M3 at the expense of configuration time. The increased configuration time is not a critical issue since the configuration operation of TCAM applications is not frequently performed compared to the search operation. Furthermore, the VCMA-driven precessional switching speed is fast enough to compensate for the increase in the number of the write operation. The corresponding bias conditions of the configuration operation are summarized in Table 3.2.

In the search operation, the ML node is initially pre-charged up to V_{dd} during the ML pre-charge phase as shown in Fig. 3.4. For the evaluation phase, a 1 V pulse with T_{SE} (250 ps) duration is applied to the DGL node. Giving a positive pulse at the bottom of the MTJ enhances the coercivity of the free layer, hence reducing the search (read) disturbance of the MTJs. Under the pulsed condition, the potential of the CE node is determined by RC delay originating from the resistance of the MTJs and the intrinsic capacitance of the CE node. For example, consider the case where the MeTCAM cell

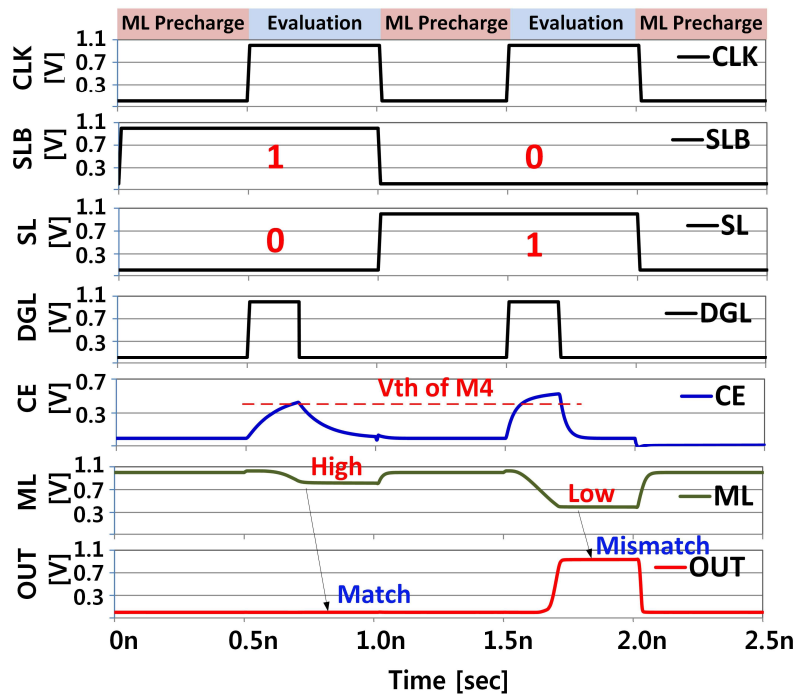


Fig. 3.4 Search operations of the MeTCAM, consisting of two phases: ML Precharge and Evaluation. During the ML Precharge phase, the ML node rises up to V_{dd} . The evaluation phase takes advantage of the time difference between the RC delay of the R_P path and that of the R_{AP} path, determining the mode (on/off) of transistor M4 without search disturbance.

stores a logic value ‘0’ ($b1-R_{AP}$, $b2-R_P$). In the case of a search ‘0’ operation ($SLB=1$, $SL=0$), the CE node cannot reach the threshold voltage of the ML-driver transistor M4 within the period T_{SE} due to the relatively long RC delay through the high resistance of $b1-R_{AP}$, resulting in ML remaining high. On the other hand, for a search ‘1’ operation ($SLB=0$, $SL=1$), the potential of the CE node is able to reach the threshold of M4 due to the relatively short RC delay through the low resistance of $b2-R_P$, enabling M4 to generate high enough pull down current I_{ML} from the ML node to the ground, which causes the ML node to become low as shown in Fig. 3.4. If the MeTCAM has a “don’t care” condition ($b1-R_{AP}$, $b2-R_{AP}$), it prohibits the CE node from exceeding the threshold of M4 for either search ‘0’ or search ‘1’ due to the long RC delay.

3.2.4 Performance Evaluation and Comparison

Since the write current of voltage-controlled MTJs is typically below $10\ \mu\text{A}$, M1~M3 can be designed based on the minimum size transistors in Fig. 3.2. The size of M4 largely influences the performance in terms of search delay and the cell area overhead. The proposed MeTCAM cell achieves 210 ps delay with 32 bit (ML length) by using the minimum size of M4. The total size of the cell is $44F^2$. Table 3.3 shows the performance comparison of MeTCAM with different types of TCAMs which are based on other memory technologies [92], [94]–[96]. It is noteworthy that the cell size of MeTCAM is 12% smaller than that of PCRAM based TCAM, which uses a 2T-2R structure. This is because the PCRAM-based TCAM cell needs to supply a relatively high reset current ($>100\ \mu\text{A}$), requiring larger transistors for providing a sufficient drivability. To increase the search speed of TCAMs, the critical path (from the ML node to GND) must be shortened and have a low resistance. The ML node of the proposed structure is directly connected to GND through M4, achieving even faster search speed than that of the 2T-2R structure, since the 2T-2R has a storage element on the critical path between the ML node and GND, causing longer RC delay.

Author/ Year	Conventional	Jing Li / 2014	Le Zheng / 2014	Shoun Matsunaga /2012	Proposed work/2014
Storage type	SRAM	PCRAM	ReRAM	STT-RAM	MeRAM
Structure	16T 	2T-2R 2bit-encoding	5T-2R 	9T-2MTJ 	4T-2MTJ
Area	250F²	50F²	> 100F²	> 108F²	44F²
Search Speed	0.1 ns for 32 bits	1.9 ns for 64 bits	4 ns for unknown C_{ML}	0.1 ns for 32 bits	0.2 ns for 32 bits
Voltage for search/write	1.2 V	1.2 V / 2.5 V	1 V / 2.5 V	1 V / 1 V	1 V / 1 V
Storage on I _{ML} path	No	Yes	No	No	No
Disturbance	No	Yes	Yes	Yes	No
Leakage from ML	No	Yes	Small (depend on on/off ratio)	No	Small (depend on TMR)

Table 3.3 Performance comparison of TCAMs associated with different types of storage memories [92], [94]–[96].

Utilizing voltage-controlled MTJs in MeTCAM also reduces errors originating from search-induced false writing events. In principle, currents flowing through storage elements may cause a disturbance (i.e. unwanted switching) during the data search operation. This is especially severe in the case where a storage element is located on the critical path like a 2T-2R structure. The more search operations execute, the higher possibility of the disturbance generates. However, the MeTCAM is less susceptible to the disturbance because applying a positive pulse to the DGL node enhances the coercivity of the free layers, leading to more stable MTJ free layers during the search operation [97].

3.3 A Spintronic Voltage-Controlled Stochastic Oscillator

3.3.1 Uniform Sampling versus Non-uniform Sampling

In the era of the Internet of Things (IoT), a tremendous amount of information in the analog domain is sampled and converted to digital data in a large palette of applications such as mobile communications, wearable devices, medical imaging, and radar detection. This digital information is

typically processed by a digital signal processor (DSP) or a central/graphics processing unit (CPU/GPU) and stored in memory devices. However, the data deluge significantly increases energy consumption for computations/transmission and requires higher memory capacity to record the data.

One of the promising ways to alleviate these issues is reducing the amount of data by adopting a non-uniform sampling scheme, which is an essential part of compressive sampling (CS) techniques, instead of using a conventional uniform sampling scheme [98]–[100]. Both sampling schemes are briefly described in Fig. 3.5(a) and (b).

The uniform sampling has been developed and optimized in modern hardware and software since the efficient fast Fourier transform (FFT) is executed based on uniformly sampled data. However, using uniform sampling is inefficient in certain types of application where much of the generated data does not significantly contribute to the overall information. This redundant data increases the computational load, causing energy waste for processing, transmitting and recording of the data [101]. Moreover, uniform sampling cannot efficiently avoid aliasing, leading to distortion in the signal reconstruction [102].

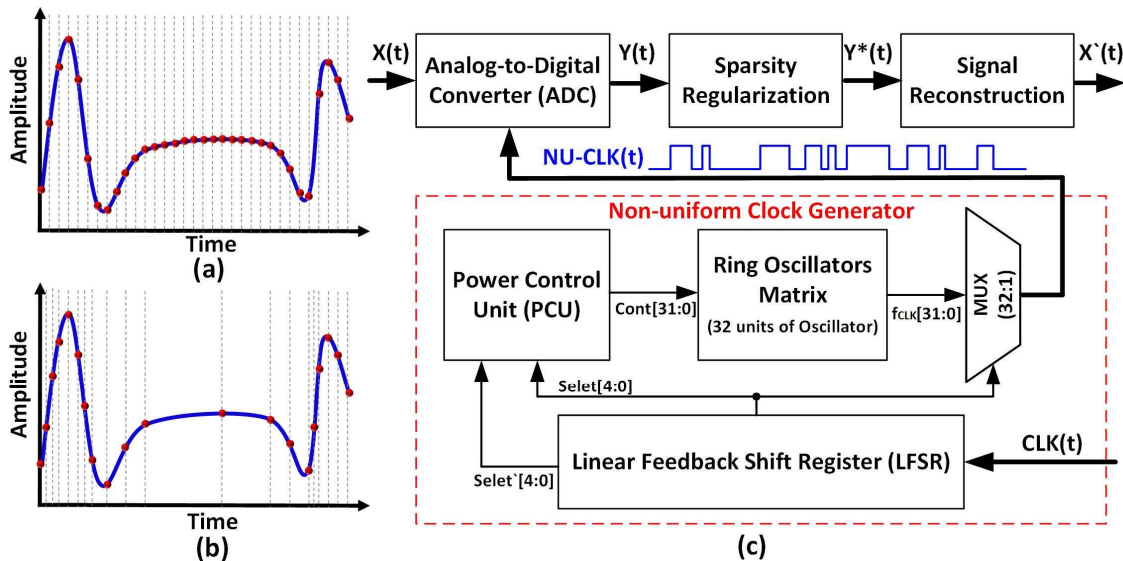


Fig. 3.5 (a) Uniform sampling that has a constant time interval between samples. (b) Non-uniform sampling which has a variable time interval between samples. (c) Non-uniform clock based analog to digital information conversion and reconstruction system using a conventional periodic non-uniform clock generator.

Non-uniform sampling is categorized into two groups: periodic and non-periodic [103]. In the periodic non-uniform sampling, the sampling noise is added to each periodic sampling time; on the other hand, the sampling time of the non-periodic methods is constructed by adding the noise to the previous sampling time, typically called additive random sampling. Another type of non-periodic non-uniform sampling is the level-crossing sampling scheme (LCSS) which takes samples when the input signal crosses predefined threshold levels, referred to event-driven sampling [100]. Non-uniform sampling can have advantages especially with low activity signals such as electrocardiograms and other biological signals, temperature, pressure, voice, and patterns, which remain constant most of the time and change sporadically. Since the total system energy consumption is a function of the sampling rate, event-driven random sampling drastically reduces the computation and data transmission energy by capturing the relevant samples based on the signal characteristic [100]. Also, the randomness of the time interval between samplings improves the dynamic range of the system and addresses aliasing issues [104].

Figure 3.5(c) shows a conventional CMOS-based analog-to-digital information conversion and reconstruction system where the conventional periodic non-uniform clock generator plays a role in determining sampling bandwidth, computational complexity, and overall power consumption. Many periodic non-uniform clock generators have been proposed based on a linear feedback shift register (LFSR) that randomly selects one clock signal among a matrix of ring oscillators, which have different frequencies and phases [105]–[108]. However, these circuits require a large number of transistors and additional controllers, resulting in area and energy consumption overhead, and only provide a fixed average sampling frequency.

3.3.2 Advantages of using MTJ for the Non-uniform Clock Generator

We propose an alternative approach to address the area and power issues and to provide flexibility in terms of sampling frequency, based on voltage-controlled magnetic tunnel junctions (MTJs). In

principle, a CMOS compatible MTJ is a memory device which has two discrete resistance states switched by an electrical or magnetic bias condition. However, if an MTJ is engineered to have sufficiently low thermal stability, the state of the device can be stochastically switched via thermal fluctuations. The average time interval between thermal switching events is called the retention time, which can be modulated by an applied voltage across the MTJ via the VCMA effect. These characteristics allow the MTJ to be used as a voltage-controlled stochastic oscillator (VCSO) that can generate an event-driven stochastic signal (ESS). In perpendicularly magnetized MTJs for typical storage memory applications, the interfacial perpendicular magnetic anisotropy (PMA) is enhanced by choosing suitable materials and adjusting the thickness of the ferromagnetic layers (< 2 nm) to achieve a high thermal stability ($\Delta > 60$) [109]. In this work, we deliberately engineered the PMA to obtain a relatively low thermal stability (20~35) for random sampling applications.

In an ultrathin magnetic film structure (e.g. MTJs), an applied voltage across the device can modulate the PMA of the free layer, an effect broadly known as voltage-controlled magnetic anisotropy (VCMA) [25], [72], [110], [111]. Figure 3.6(b) shows the measured corresponding coercivity change of the MTJ (diameter 60 nm, 1.1 nm thick free layer, $\Delta = 22$ at zero bias) as a function of voltages across the device. The modulation of the coercivity means that the energy barrier

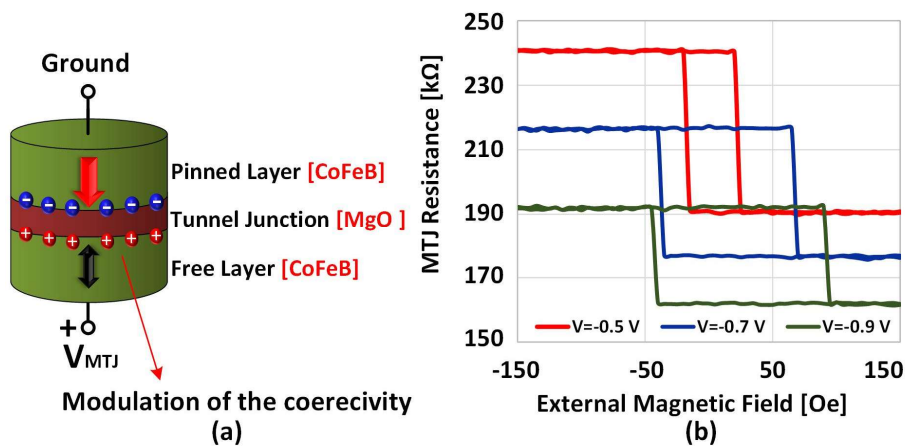


Fig. 3.6 (a) Voltage dependence of the coercivity. Based on the polarity of the applied voltage, the coercivity of the free layer changes due to the VCMA effect. (b) Measured coercivity with respect to the applied voltage across the device. As the amplitude of the bias increases, the coercivity is enhanced.

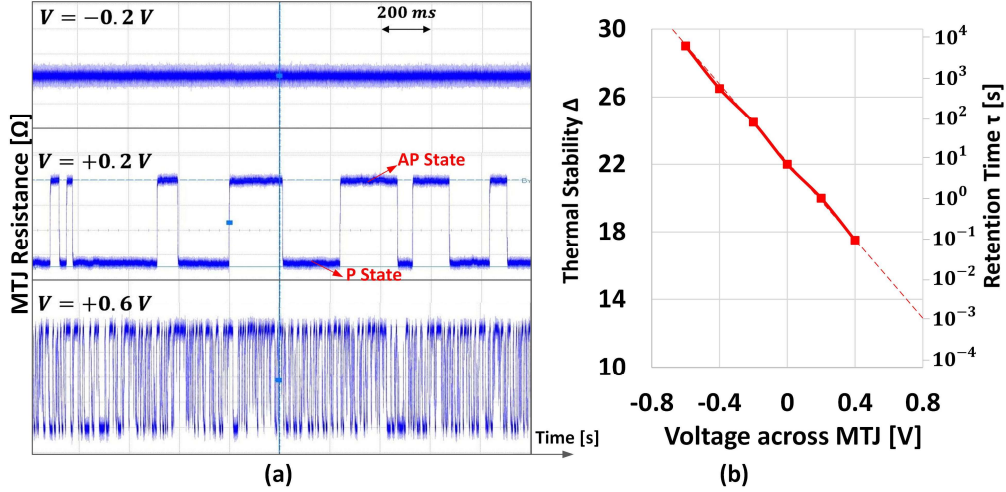


Fig. 3.7 (a) Measured time domain data of the MTJ's resistance fluctuation under the different electric bias conditions (b) thermal stability of the measured MTJ with respect to voltage across the device. Retention time is calculated based on the amplitude of the thermal stability at room temperature.

$E_b = \mu_0 M_s H_{k,eff}^\perp(V_{MTJ}) \mathcal{V} / 2$ between the two stable states can be changed by a voltage across the MTJ where $H_{k,eff}^\perp$ is the out-of-plane component of the effective magnetic anisotropy, \mathcal{V} is the volume of the free layer, and V_{MTJ} is the applied voltage across the MTJ. The equation (2.17) provides more detail information. Since the PMA is a dominant component in the out-of-plane component effective magnetic anisotropy $H_{k,eff}^\perp$ of the MTJ, the retention time of the MTJ is modulated by the applied voltage across the device as given by $\tau = \tau_0 \exp(\Delta)$, where τ_0 is equal to 1 ns, thermal stability Δ is equal to $E_b / (k_B T)$, and k_B is the Boltzmann constant. Figure 3.7(a) shows measured resistances of the MTJ as a function of time under different bias conditions where a positive voltage decreases the retention time by reducing the thermal stability, while a negative voltage increases the retention time via enhancing the thermal stability, demonstrating the fundamental concept behind this work. The voltage dependence of thermal stability is shown in Fig. 3.7(b).

3.3.3 Design and Performance Evaluation

We intentionally adjusted the MTJ model parameters (e.g. $\Delta = 35$ at zero bias, $\xi = 61$ fV/mV) to allow the MTJ to reliably operate with CMOS supply voltages (1.2 V for 65 nm node) and cover a

wider range of frequency. The rationale of using a thermally unstable MTJ in designing a VCSO is as follows: (i) Switching driven by thermal noise is a Poisson process in which the occurrences of certain events happen at a certain rate, but completely random, and guarantees non-uniform intervals between samplings; (ii) The voltage dependence of the retention time can be used for realizing event-based sampling; (iii) The two discrete resistance states of the MTJ can be easily converted to a digital signal.

The function of the VCSO in digital information conversion system is as shown in Fig. 3.8. Based on the frequency of the analog input signal, the VCSO generates an event-driven stochastic signal (ESS) to trigger the asynchronous analog to digital converter (A-ADC) so that the system can efficiently adjust its sampling frequency. The A-ADC samples the analog input signal at each edge of the ESS and converts the input signal into a digitized code based on the signal's amplitude. The average frequency of the ESS is determined by the potential of the BIAS node. The frequency to voltage converter (FVC) converts the maximum frequency of analog input signal into a certain level of voltage on the BIAS node in real time [112], [113].

In the VCSO, the MTJ is connected to the voltage clamp circuit (M1~M3) and the amplifier (M6~M8). The voltage clamp maintains the potential on the N1 node regardless of the MTJ resistance

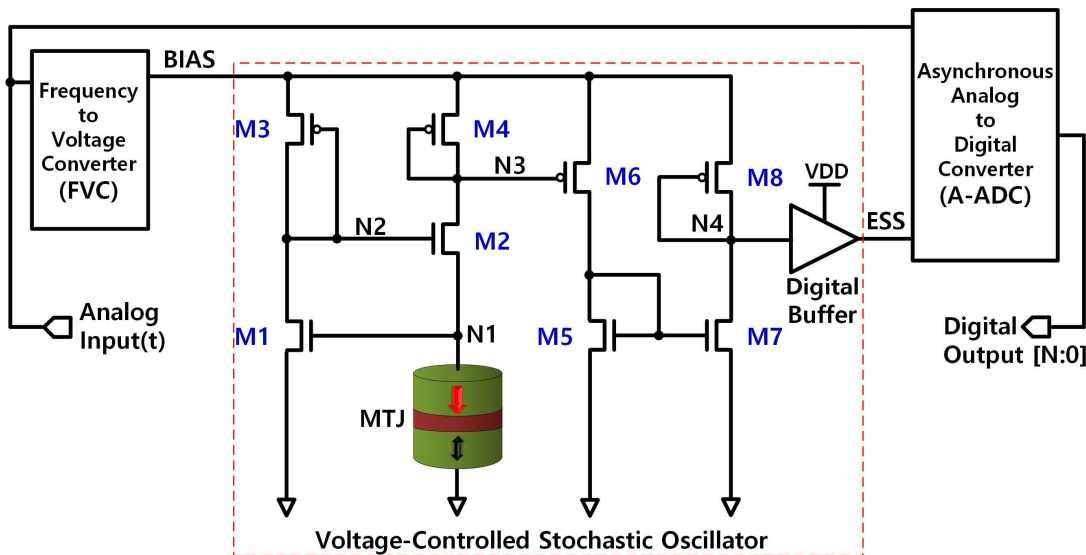


Fig. 3.8 MTJ based voltage-controlled stochastic oscillator (VCSO). The switching rate of the MTJ depends on the potential on the BIAS node. The two resistance states of the MTJ are sensed by the amplifier (M6-M7) whose output is converted to an event-driven stochastic signal (ESS) via the buffer.

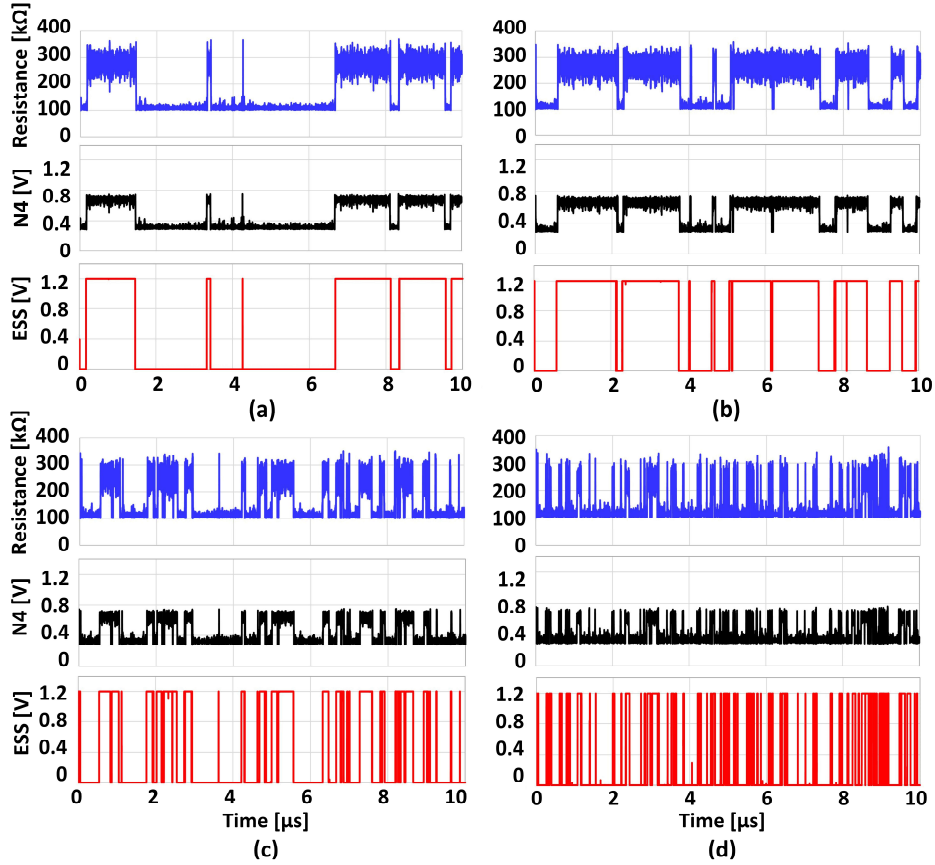


Fig. 3.9 Transient circuit simulation of the VCSO with the MTJ compact model. The potential on the BIAS node changes the switching rate of the MTJ. The potential of the N1 node is equal to (a) 0.72 V (b) 0.74 V (c) 0.76 V (d) 0.78 V.

fluctuations, which allows the voltage across the MTJ to be purely dependent on the potential of the BIAS node. The circuit operation for generating an ESS is as follows. If the frequency of the input signal is high, the potential of the BIAS node increases, reducing the energy barrier of the MTJ, leading to a higher rate of switching. If the frequency of the input signal is low, the potential of the BIAS node is low such that the energy barrier of the MTJ remains high, and switching occurs at a slower rate. The MTJ's resistance fluctuation is converted to a voltage variation on the N3 node, which is amplified by the amplifier whose output (N4) is digitized via the buffer, generating the ESS as shown in Fig. 3.9.

The average sampling frequency of the VCSO exponentially varies as a function of the potential of the BIAS node since the energy barrier decreases linearly as a function of voltage as shown in Fig. 3.10. The change of switching rate depending on the voltage range across the MTJ can be modulated

NUCG	Tech node	Power (μW)	Area (μm^2)
This work	65 nm	< 26.7	10.6
[104]	65 nm	89.7	222.6
[105]	90 nm	115.7	1053.7

Table 3.4 Performance comparison with previous works.

by engineering the VCMA coefficient. In this simulation, the average sampling frequency can be modulated from 1 kHz to 100 MHz under voltages ranging from 0.6 V to 0.9 V, which indicates that the VCSO can perform wide-dynamic-range random sampling. However, the FVC needs to convert an exponentially varying analog input signal in terms of frequency into a linearized bias voltage via a calibration to reduce errors in the signal reconstruction. Also, to guarantee reliable CMOS operations in the VCSO, the potential on the BIAS node should be larger than 0.6 V so that the amplifier can drive the digital buffer.

The total power consumption of the VCSO is a sum of the power consumption of the analog circuit (M1~M8) and the digital buffer. The former is mainly proportional to the amplitude of potential on the BIAS node. The latter depends on the switching rate of the MTJ since the dynamic power of the digital buffer is proportional to the number of switching during a certain period. The number of transistors in

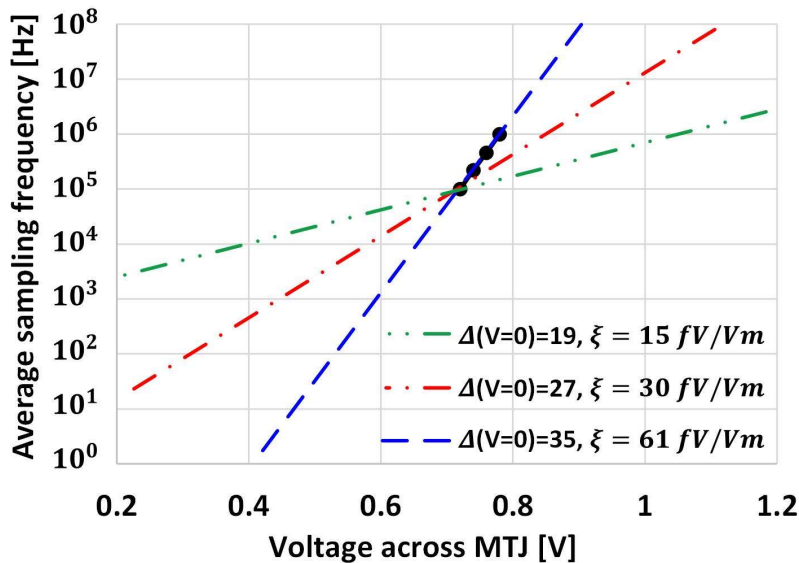


Fig. 3.10 Average sampling frequency of the VCSO with different thermal stabilities and VCMA coefficients as a function of the voltage across the MTJ.

designing the VCSO is drastically reduced by taking advantage of the MTJ characteristics. Table 3.4 summarizes the performance of the VCSO based on 65 nm technology node, assuming that the average frequency of the ESS is 100MHz.

3.4 Voltage-controlled MTJ based True Random Number Generator

3.4.1 Overview of Random Number Generator

The use of electronic financial transactions, on-line communications, and digital signature applications have exponentially increased over the last few decades. Demand for the secure transfer of confidential information raises the importance of cryptography. Researchers have worked a great deal on designing a random number generator (RNG) as it is one of the most indispensable components for cryptography.

A wide variety of integrated circuit (IC) based RNGs have been developed and implemented into secure digital chips, exploiting thermally induced jitter from ring oscillators, block RAM write collisions, and optical effects [114]–[117]. However, pure semiconductor-based IC RNGs have encountered several issues in terms of speed, power, and their quality of randomness. In the case of analog RNGs, since the amplitude of noise sources such as flicker noise is too small for the semiconductor circuit, amplification of the signal(s) is necessary, making the generation frequency of the random numbers less than 10 MHz.

These amplified noise based analog RNGs cannot meet the throughput requirements of cutting-edge high-speed information security applications due to the limited bandwidth of the entropy source [118]. Although the digital RNGs meet the demand for the throughput of the emerging security applications, they still have a few problems: area overhead, high power consumption, and injection locking problems [119].

To solve the issues, a spin-transfer torque (STT) MTJ based random number generators have been proposed by many research groups [120], [121]. However, the STT-MTJ based random number

generators require a precise write pulse width at given amplitude to achieve high-quality randomness (e.g. 50% switching probability), which in turn necessitates a dedicated control circuit and a calibration process [122]. Furthermore, a current-driven STT-MTJ intrinsically requires a significant amount of charge current to switch, resulting in high ohmic dissipations.

3.4.2 Advantages of using Voltage-Controlled MTJ for the True Random Number Generator

We propose a voltage-controlled MTJ based true random number generator (MRNG) where the electric field is used to induce switching instead of substantial current flow in the MTJ device, drastically reducing ohmic loss. Furthermore, unlike STT-driven switching, generating a random bit is not sensitive to the write pulse width because the magnetic moment converges to an in-plane (metastable) direction under the long enough electric bias condition.

The MTJ is used as a noise source device which is connected to a comparator. An MTJ has two discrete states, high (denoted as AP) and low (denoted as P) resistance states. However, it is possible that it can be in a metastable state under certain electrical bias conditions which sufficiently eliminate the energy barrier between the MTJ states as shown in Fig. 3.11. Once the voltage bias is removed, the

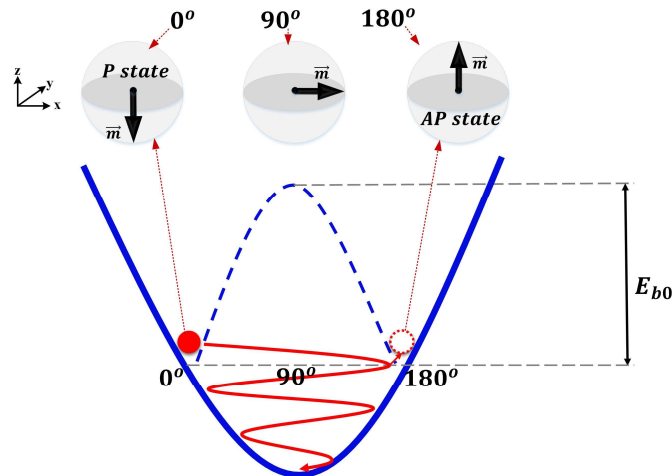


Fig. 3.11 Description of VCMA-induced switching mechanisms in a perpendicularly magnetized MTJ. At equilibrium with zero bias field ($V = 0$), the energy barrier E_b separates the two stable states of the free layer magnetization. If the energy barrier E_b is sufficiently removed due to the VCMA effect, damping and precessional motion of the magnetization can occur.

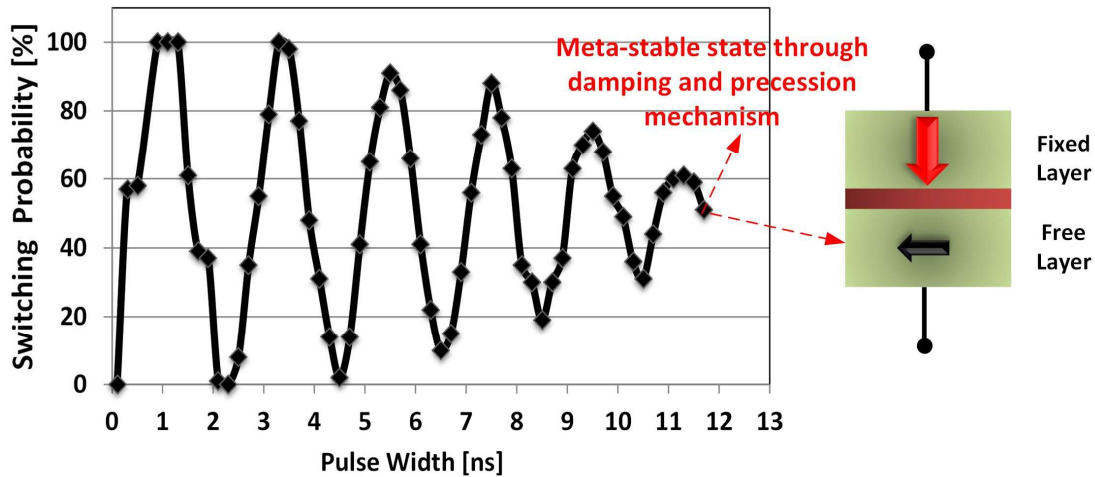


Fig. 3.12 Simulated VCMA-induced switching probability as a function of write pulse width. If the pulse duration is increased further, the magnetic moment of the free layer can be aligned with the in-plane direction, entering the metastable state via damping and precessional motion, achieving 50% switching probability.

state of the MTJ is randomly chosen between the high resistance state and the low resistance state due to thermal fluctuations. In terms of area, since they can be fabricated on top of CMOS circuitry via the back-end-of-line (BEOL) process, MTJs do not require additional area. Furthermore, a single comparator, which consists of 11 transistors, can control an array of MTJs, achieving a more compact design. For the power consumption, the non-volatile characteristic of MTJs allows the circuit to be in sleep mode, realizing zero leakage.

Due to the VCMA effect, the magnetic moment of the free layer eventually aligns with in-plane direction via damping and precessional motion after applying relatively long electric pulse (~ 10 ns) with a sufficient amplitude that can remove the energy barrier between two states. The macrospin MTJ compact model simulation shows that the switching probability of MTJ is a function of the applied pulse width as shown in Fig. 3.12. This oscillatory switching behavior of the voltage-controlled MTJ as a function of the pulse width has been experimentally observed in the presence of in-plane external field [68], [69]. Since the in-plane direction is not the easy axis of the device, the final state of the MTJ is chosen via thermal fluctuations after removing the bias. In this metastable state, the switching

probability converges toward 50% in the absence of the external magnetic field, which can be exploited for generating random bits.

3.4.3 Design and Performance Evaluation

The proposed MRNG consists of an MTJ device and a comparator (M1~M11) with a reference resistor (R_{ref}) as shown in Fig. 3.13. The MTJ device and the reference resistor are connected to the SEN node and the REF node, respectively. The circuit generates a random bit at the DOUT node by using damping and precessional motion of the MTJ, making the quality of randomness less sensitive to an applied pulse width and achieving high-speed. Specifically, the MRNG consecutively executes two different operations: write and sense for generating a random bit as shown in Fig. 3.14. During the write operation, M2 and M3 are turned on by a ‘low’ potential of \overline{CLK} , raising the SEN and REF up to VDD level. The voltage across the MTJ causes the precession and damping of the magnetic moment in the free layer. As a result, the magnetic moment of z-direction component is converged to around zero value after a certain period, implying that the magnetic moment is aligned with in-plane direction.

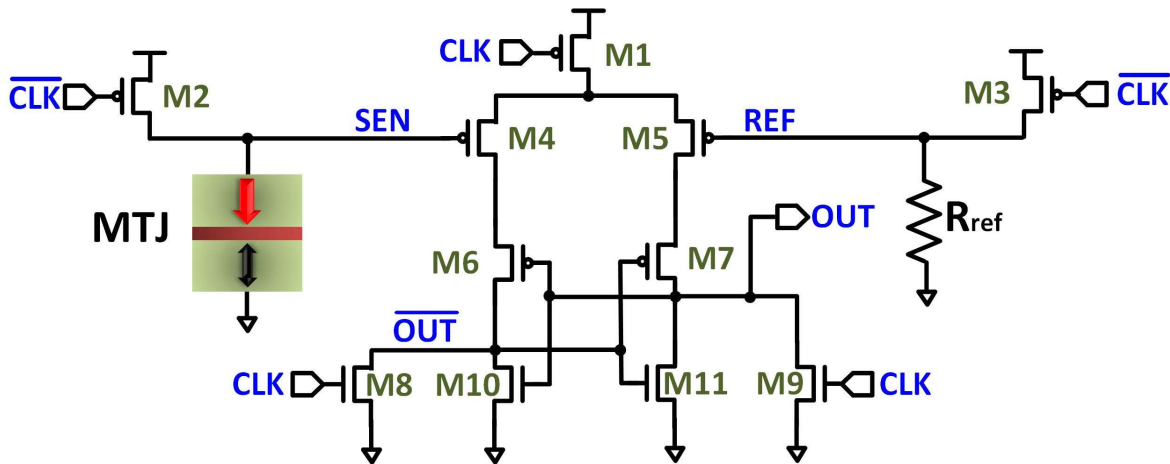


Fig. 3.13 Schematic of the proposed MRNG which is made of 11 transistors, a reference resistor, and an MTJ device. To achieve simple controllability, two control signals CLK and CLKb (\overline{CLK}), which is a delayed complementary signal of CLK, are used. If CLK is low, M2 raises the potential of the SEN node up to VDD, which causes the precession of the MTJ. When CLK becomes high after a certain period, the MTJ state is randomly selected from the metastable state, and the SEN node starts to discharge. Since the potential of the node SEN is determined by the MTJ state, the random data from the DOUT node is generated by comparing the potentials between the SEN node and the REF node.

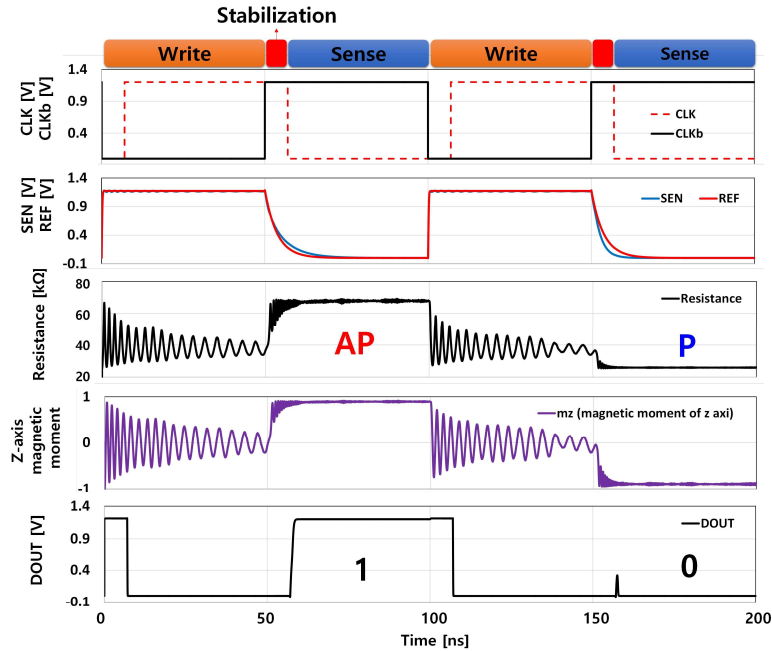


Fig. 3.14 Simulation result of the MRNG for a random bit operation which consists of two phases, the write and sense modes. During the write mode, the VDD potential at the node SEN causes the oscillatory behavior of the magnetic moment in z direction (precession), which eventually converges to zero value (being metastable) via damping factor. Then, the randomly selected MTJ state is sensed by the comparator (M1~M11) and is presented at the OUT node during the sense mode.

At this moment, the magnitude of the MTJ resistance is observed between the P and the AP states. Once M2 and M3 are turned off, the MTJ state is randomly selected via thermal fluctuations. Before the circuit goes into the sense operation, a few nanoseconds are necessary to stabilize its state.

For the sense operation, the previously applied write voltage VDD at the SEN node is reused as a pre-charge voltage to save operation time and energy since the discharge of the SEN node depends on the RC delay caused by the resistance of MTJs and the capacitance of the SEN node. As CLK turns M1 on, the random bit is generated depending on the potential difference between the SEN node and the REF node. As a result, the AP state and P state are converted to digital signal '1' and '0' at the DOUT node, respectively. If the resistance and capacitance of the SEN node are 50 kΩ and 10 fF, respectively, the RC time constant is around 0.5 ns, theoretically allowing a GHz level sense operation.

To realize a few Gbps throughput, we also proposed the multi-bit MRNG, which can generate random bits in a parallel manner, where $m \times n$ is the array size with access transistors as shown in Fig.

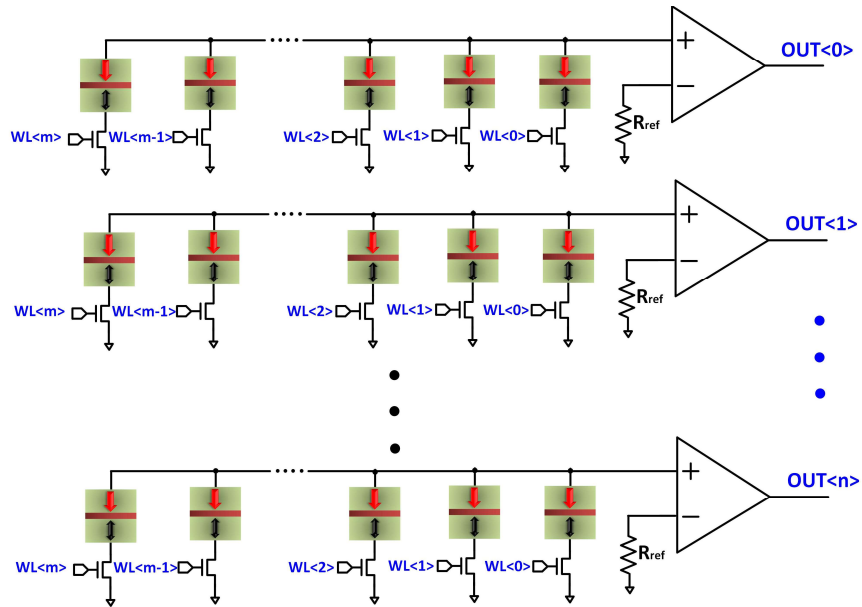


Fig. 3.15 Schematic of the $m \times n$ MTJs array based multi-bit MRNG. All $m \times n$ MTJs devices can have new random states during the single write operation, and n bit random data are sensed for each sense operation (~ 1 ns), achieving $\sim n$ Gbps random bit generation.

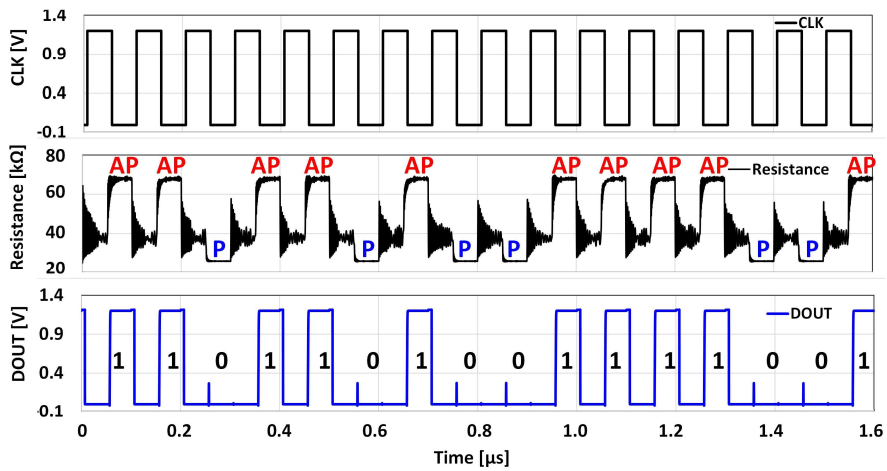


Fig. 3.16 Simulation results of consecutive random bit generations by using the MRNG. The random bit is generated at the falling edge of CLK, which is associated with the MTJ state. The random states of AP and P are converted to digital signal '1' and '0' at the DOUT node.

3.15. Figure 3.16 shows the generation of a random bit sequence by using the proposed MRNG. The random bit, which corresponds to the MTJ state, is consecutively generated at the falling edge of CLK. The energy requirement of the MRNG for a random bit generation is as follows: the $24 \mu\text{A}$ write current with ~ 10 ns duration is necessary to place a $50 \text{ k}\Omega$ based MTJ device into a metastable state when VDD is 1.2V. The energy consumption of the MTJ for a single bit is 288 fJ. However, this energy

Performance	Value
Area	139.96 μm^2
Throughput	29.6 Gbps
Energy	311 fJ/bit

Table 3.5 Area and performance of 64×64 MTJs array based multi-bit-MRNG (45 nm technology node).

can be reduced further by increasing the resistance of MTJs. Also, the comparator consumes an average of 13 μA of dynamic current for 1.5 ns to sense an MTJ state. Therefore, the total energy consumption per random bit is 311.4 fJ/bit. Considering the non-volatile nature of MTJs, the MRNG achieves zero leakage current (zero static power) during the standby mode.

Since an external magnetic field can manipulate the randomness of the MRNG, it is necessary to evaluate the switching probability as a function of external magnetic field magnitude. Figure 3.17 shows the switching probability of an MTJ after applying a write pulse in the presence of an out of plane direction magnetic external field H_z . As H_z increases, the MTJ is more likely to switch to the AP state, deteriorating the quality of randomness. In the region A ($H_z < 6$ Oe), the switching probability distributes between 48% and 52%, which is a suitable number for practical applications. The performance is summarized in Table 3.5.

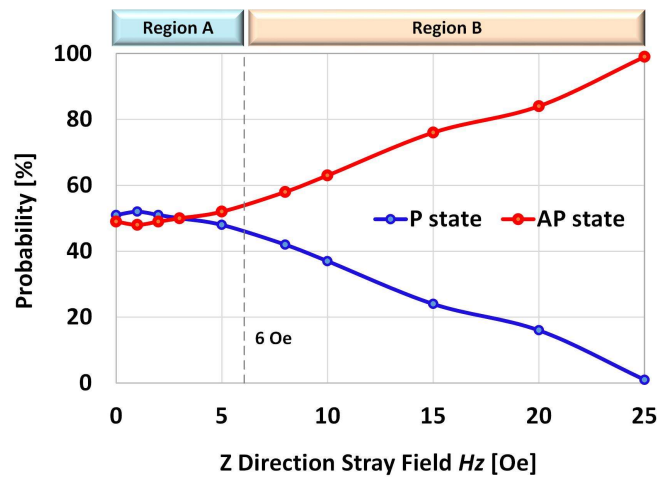


Fig. 3.17 Switching probability of MTJ after applying a write pulse in the presence of z-direction external field. The existence of external field H_z changes the switching behavior. In the region A where it has a weaker field < 6 Oe, the distribution of the switching probability is around 48% ~ 52 %, satisfying one of the conditions for the randomness. However, in the region B where the external field is larger than 6 Oe, the switching probability start to deviate from 50%.

3.5 Spintronic Programmable Logic (SPL) using Voltage-gated Spin Hall Effect

3.5.1 Overview of Programmable Logic

Conventional static random-access memory (SRAM) technology has been widely used as cache memory in modern microprocessors, and as the memory element in look-up tables (LUT) in programmable logic circuits. SRAM has a number of advantages such as fast access time (< 1 ns) and unlimited endurance ($> 10^{15}$) [123]. However, at present nanometer-scaled CMOS technology based SRAM cells have become a power hungry component in embedded systems, especially in terms of static power dissipation, due to the fact that the leakage current has exponentially increased by a continued shrinking of transistors. To alleviate this issue, the implementation of non-volatile memories into systems has been proposed by many researchers, completely eliminating standby power consumption [124]–[127].

An STT-MTJ based 6-input non-volatile lookup table (NV-LUT) has been proposed by [125]. The LUT circuit is implemented compactly by replacing SRAM cells with STT-MTJs. In this LUT, the circuit utilizes a shared write transistor especially for switching the STT-MTJ from P to AP to provide a sufficient write current since the switching from P to AP has a higher critical current than that of the opposite switching. Although the shared transistor can reduce the area overhead to some extent, the current-driven STT-MTJ requires high current (> 100 μ A) compared to a voltage-controlled MTJ, which in turn limits the scaling down of the access transistor.

A VCMA based MTJ, on the other hand, exploits magnetoelectric effects, significantly reducing the need for currents to switch the device. In addition to reduced switching energy, the use of an electric field for writing provides an advantage in terms of enhanced bit density and fast switching (< 1 ns) via precessional (i.e. resonant) switching. This type of switching, however, has a non-deterministic feature, where the state of the bit is always reversed regardless of its initial state for a given pulse duration. Therefore, the state of the MTJ needs to be read before writing [128].

3.5.2 Advantages of using three terminal MTJ for the SPL

We propose a spintronic programmable logic (SPL) concept, based on a 3-terminal MTJ device that combines both SHE (from spin polarized electrons due to a current) and the VCMA effect (i.e. voltage control), which we refer to as gate-voltage-modulated SHE switching (V-SHE). The SPL can be configured to perform not only any arbitrary combinational logic function but also any sequential logic function. Due to the V-SHE switching, which is deterministic, the SPL is able to configure its MTJ devices in a parallel manner, achieving 1 ns configuration time and low switching energy (28 fJ/bit). Compared to the conventional 6-input STT-MTJ based LUT, which writes in a serial manner, the configuration time of the SPL is significantly reduced up to 100x. In terms of area, the proposed SPL achieves 61% and 32% area reductions compared to SRAM and an STT-MTJ based LUTs, respectively, when used in the form of a 6-input LUT structure.

3.5.3 Configuration and Logic Operations of the SPL

We employed 65 nm CMOS technology with a 3-terminal MTJ compact model to design and evaluate the SPL in the Cadence circuit design environment. Figure 3.18 shows the proposed 2-input SPL, which consists of three major parts: a write circuit, a selection tree, and a current conveyor. The write circuit is used to configure the data MTJs (MTJ1~MTJ4). For instance, to configure the XOR logic function a 58 μA charge current (I_m), provided by the write circuit (M1, M2), flows through the heavy metal with spin Hall angle (HMS) to generate the SHE, and ~ 0.5 V pulses are simultaneously applied to the BL2 and the BL3. This results in changing MTJ2 and MTJ3 from AP to P via the gate-voltage-modulated SHE switching, hence realizing a parallel write configuration. The write (configuration) operation is followed by two logic operations separated by the standby mode (power off) as shown in Fig. 3.19.

The selection tree is used for the logic operation to choose one of the current paths, selecting an MTJ associated with the digital input of A and B. The current conveyor, composed of two transistors

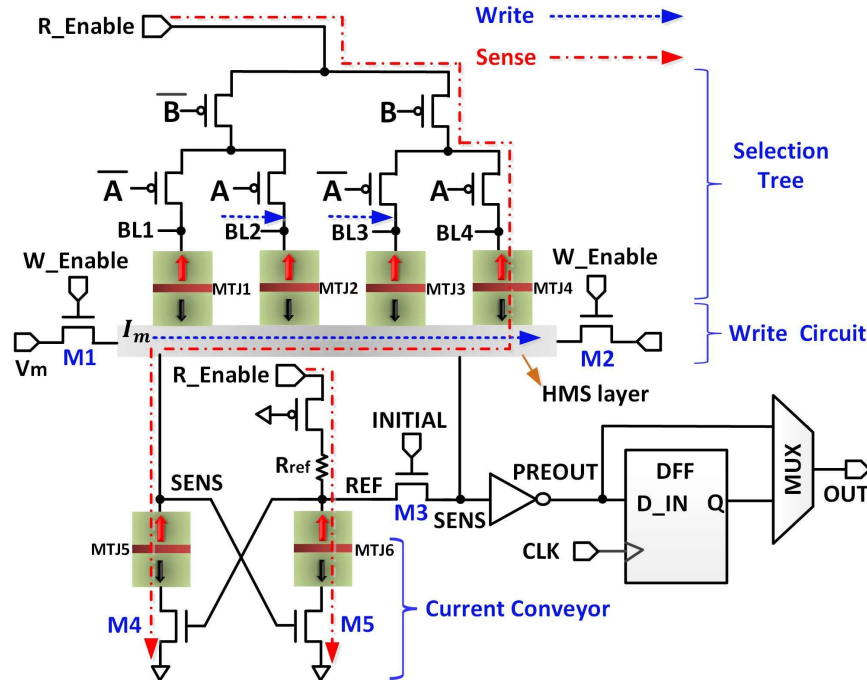


Fig. 3.18 Schematic of the proposed 2-input spintronic programmable logic (SPL). The truth table of a 2-input function is stored into 4 Data MTJs (MTJ1~MTJ4). In the write operation, M1 and M2 turn on to generate current I_m for SHE and V_{MTJ} is applied for the VCMA effect through the BLs. In the logic operation, one of the Data MTJs is selected based on the input signal A and B. The state of the selected MTJ is detected by a current conveyor, generating a stable logic value ‘high’ (AP) or ‘low’ (P).

(M4, M5) and two AP state MTJs (MTJ5, MTJ6), increase the sensing margin by using the feedback loop. The detail operation is described in the next section.

Based on the configuration of MTJ1~MTJ4, the SPL can perform any types of combinational and sequential logic functions by combining a flip-flop. One can increase the number of inputs of the SPL by adding more MTJs, realizing more sophisticated logic functions. It is noteworthy that the same logic operation can be resumed after returning from the standby mode (power off) due to non-volatile MTJs, thus achieving zero standby power and instant-on recovery without necessitating a data fetch from external memory.

3.5.4 Performance Evaluation: Sensing Margin, Power Consumption, and Area

The sensing margin is limited by intrinsic characteristics of the MTJs (in particular the tunnel magnetoresistance, TMR), as well as circuit design parameters and the sensing scheme used. Previous

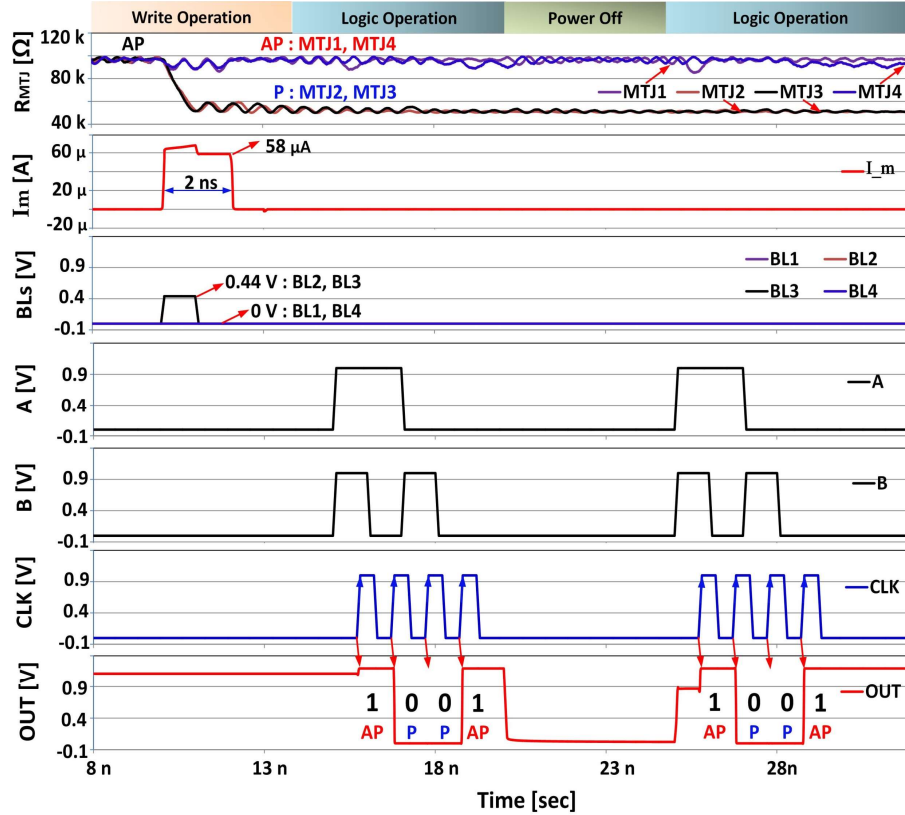


Fig. 3.19 Write and logic operation of the proposed 2-input SPL. The MTJ2 and MTJ3 are switched from AP to P by using the 2 ns duration of the charge current $I_m \approx 58\mu\text{A}$ with the 0.44 V pulse on the BL2 and BL3. During the logic operation, MTJs are consecutively selected based on two inputs A and B, and an OUT, which corresponds with the stored data in MTJs, is available at the rising edge of the CLK. After power off and on, the same logic function can be realized due to the non-volatile nature of MTJs.

works have shown that 0.18 V sensing margin can be achieved based on 1T-1MTJ topology with 200% TMR and a 65 nm technology [129]. In the proposed circuit as shown in Fig. 3.18, the sensing margin is defined by the voltage difference between the REF node and the SENS node. To maximize the sensing margin, a modified version of the current conveyor circuit is implemented into the SPL [130]. Based on a TMR of 100%, the SPL achieves a 0.8 V sensing margin in this work.

The logic operation based on the current conveyor circuit is as follows: the R_Enable rises up to 1.0 V and pre-charges the REF node and the SENS node at a certain voltage level, which slightly turns M4 and M5 on. Simultaneously, both REF and SENS nodes become virtually-shorted by turning M3 on, as shown in Fig. 6. Once M3 is turned off, the potential of the REF node and SENS node are determined by the strength of the pull-up and pull-down paths. If the selected MTJ has AP, which is

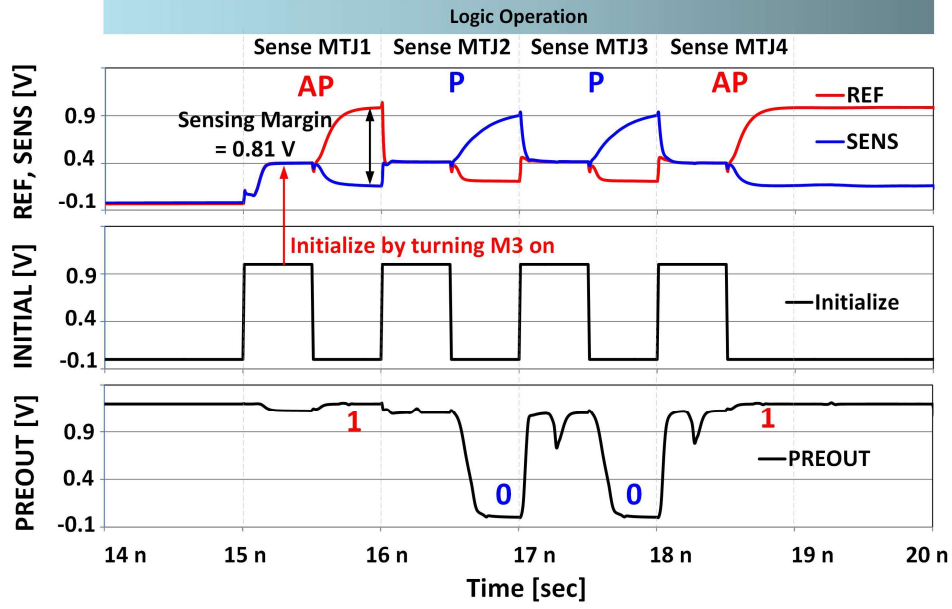


Fig. 3.20 Simulation of the sensing margin during the logic operation. Because of the current conveyor (M4, M5), a 0.81V sensing margin is achieved. To execute a logic operation, M3 is turned on by INITIAL for 0.5 ns, causing the equipotential between REF and SENS node. After turning M3 off, the potential of the SENS node is determined by the state of the selected MTJ; AP makes SENS node ‘low’, and P makes it ‘high’.

higher than R_{ref} , the SENS node is discharged much faster than the REF node. The reduced potential of the SENS node leads M5 to the subthreshold region, discharging the REF node slowly. These events are continuously repeated through the feedback loop until the circuit clearly distinguishes the state of MTJs; AP and P cause ‘high’ and ‘low’ on the PREOUT, respectively, as shown in Fig. 3.20.

The configuration energy depends on various factors: the PMA ($K_i = 1.005 \times 10^{-3} \text{ J/m}^2$), VCMA coefficient ($\xi = 37 \text{ fJ/Vm}$), saturation magnetization ($M_S = 1.2 \times 10^6 \text{ A/m}$), spin Hall angle ($\theta_{SHE} = 0.3$), and parasitic loading of the circuit. Based on the above assumptions with the compact model simulation, the switching energy of an MTJ via gate-voltage-modulated SHE switching was extracted as 28.7 fJ/bit as shown in Fig. 3.21.

For dynamic logic operation, the circuit consumes below $20 \mu\text{W}$ with 1GHz speed, which is similar to that of the conventional SRAM based LUT. However, it can achieve zero power consumption during sleep mode due to the non-volatility of MTJs. The total power dissipation is thus determined by the duty cycle, i.e. ratio of sleep mode to active mode.

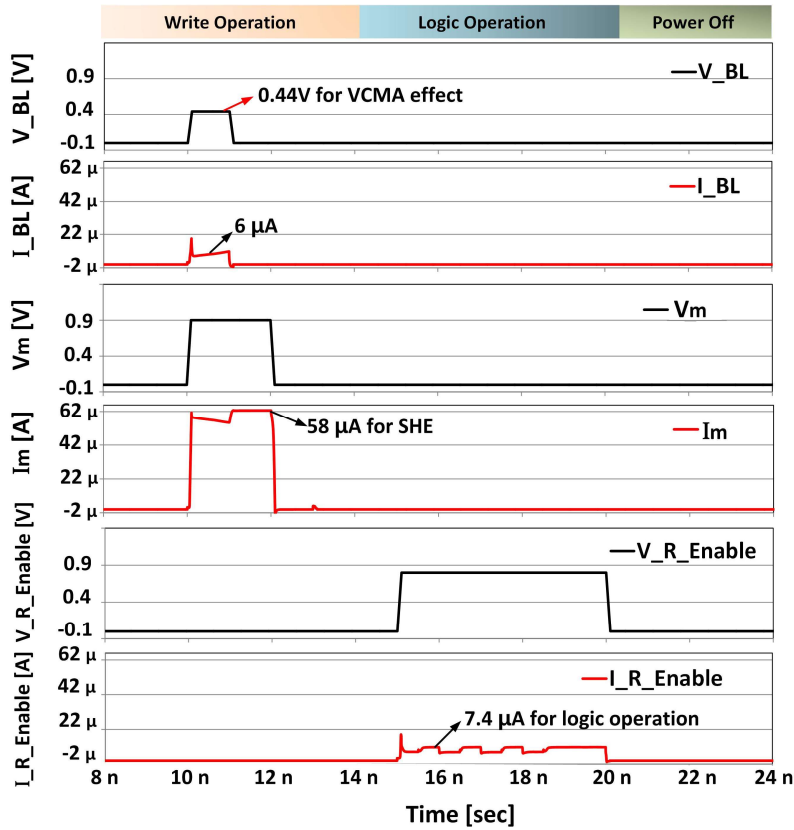


Fig. 3.21 An applied voltage V_{MTJ} induces the VCMA effect, causing a modulation of the critical current for spin Hall effect switching. In this structure, write energy and time require 28.7 fJ/bit and 2ns, respectively. The amount of dynamic power for the logic operation is below 20 μW. The SPL consumes zero power during sleep mode (power off mode).

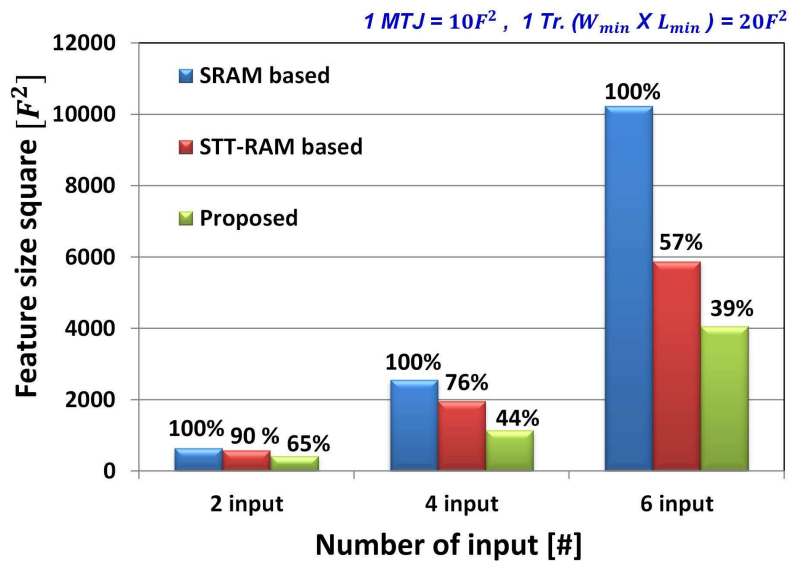


Fig. 3.22 Area comparison of different types of LUT. The proposed SPL is the most area-efficient structure compared to both SRAM and STT-RAM based LUTs. As the number of inputs increases, the area-efficiency of the proposed SPL also improves.

The 2-input SPL can achieve 35% and 28% area reduction compared to SRAM and STT-RAM based 2-input programmable logic, respectively. This is obtained by replacing a 6 transistor SRAM cell with an MTJ and using the minimum size of transistors for the write circuit. For these reasons, as the number of inputs increase, the proposed SPL can have more of an advantage in terms of area as shown in Fig. 3.22.

3.6 Analog to Stochastic Bit Stream Converter

3.6.1 Overview of Analog to Stochastic Bit Stream Converter

Stochastic computing (SC) processes information in the form of digitized probabilities such that the hardware does not produce the same outputs even if it is given the same input, unlike deterministic computing. The digitized probabilities are typically represented by randomly distributed binary numbers in the temporal domain, known as stochastic bit streams (SBS). Although the basic concept of SC was proposed in the 1960s as an area-efficient and low-power alternative to traditional binary computing [131]–[133], SC has been considered impractical as most previous applications have required fast and accurate computation.

Recently, however, SC has attracted increasing attention since the error resilience originating from its probabilistic feature may enhance the performance of applications such as artificial intelligence (learning and recognition) and informatics (sensor and social networks) [134]–[139]. In conventional computing, in terms of error-tolerance, a single error in the most-significant-bit (MSB) would result in an enormous impact on the entire computational error.

On the contrary, an error bit in stochastic computation induces only a small amount of error because each bit in an SBS equally contributes to the accuracy of the information. Moreover, since SC processes a stochastic bit stream in a serial manner, theoretically, the size of hardware and its power

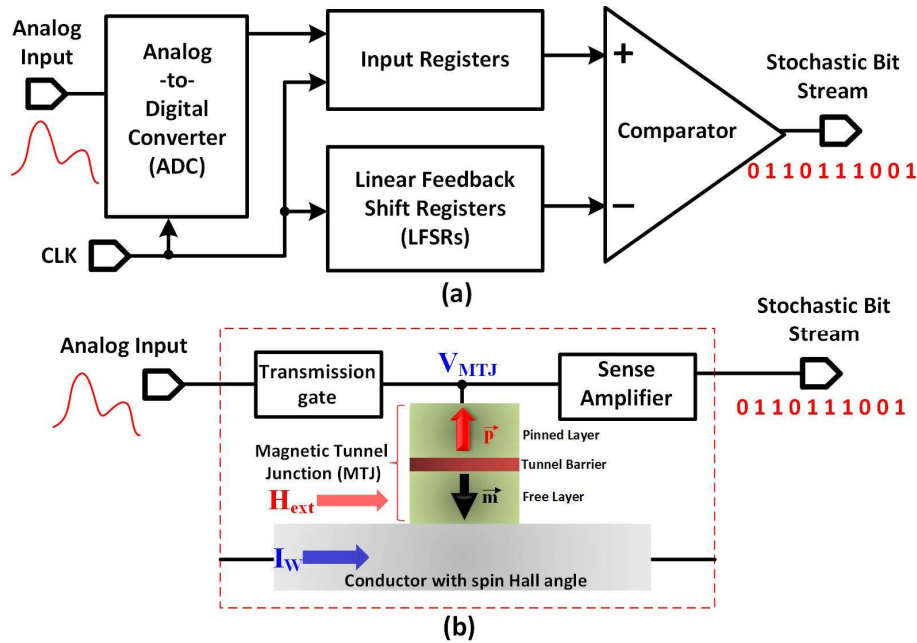


Fig. 3.23 (a) Conventional CMOS based analog to stochastic bit stream converter (b) Proposed three terminal MTJ based ASC utilizing the voltage assisted spin Hall effect for alleviating area overhead and reducing power.

consumption can be significantly reduced. This meets the requirements of a high area efficiency and ultra-low power electronic system for IoT, wearable devices, and implantable medical devices.

To utilize the promising features of stochastic computing in these applications, electrical analog signals coming from sensors need to be converted to SBS via an ASC [138]. However, in conventional circuit implementation, a pure CMOS based ASC requires a relatively large number of transistors due to the absence of a quantized noise source in CMOS. Thus, the linear feedback shift registers (LFSR) require random number generators whose designs are typically complicated and intricate and require multiple stages [137]. Also, the analog input needs to be quantized by an analog to digital converter (ADC), then stored in input registers. The bits of data stored in input registers are compared with random bits from the LFSR in a serial manner to generate an SBS [140]. The structure of a CMOS-based ASC is shown in Fig. 3.23 (a). This signal conversion process is energetically inefficient due to the large dynamic power consumed by the ADC, comparator, and LFSR as well as leakage current from the many registers during the process [120], [141].

To reduce the area overhead for signal conversion, the probabilistic switching behavior of a magnetic tunnel junction (MTJ) has been exploited in designing ASCs [142]–[144]. Although an MTJ based ASC improves the area efficiency by an order of magnitude, spin-transfer torque (STT)-driven MTJ switching results in a significant ohmic dissipation (>100 fJ/bit) and limited bandwidth due to its relatively long switching time (> 5 ns) [48], [145]. We introduce a new spintronic ASC that utilizes an emerging MTJ switching mechanism, voltage-assisted spin Hall effect, to achieve high area efficiency, ultra-low power (20 fJ/bit), and fast switching (< 2 ns) in an SC system.

3.6.2 Switching Probability of Voltage-Assisted spin Hall effect

A three terminal MTJ consists of two ferromagnetic layers separated by a tunnel barrier, and those layers are fabricated on top of a heavy metal layer which has a spin Hall angle (θ_{SH}) as shown in Fig. 3.23(b). In the three terminal MTJ, a charge current flowing through the heavy metal layer delivers a spin torque by injecting a spin current to the free layer due to the SHE effect [86], [89], [146], [147]. If the charge current is larger than the critical current, reorientation of the free layer’s magnetization can occur in the presence of the external magnetic field.

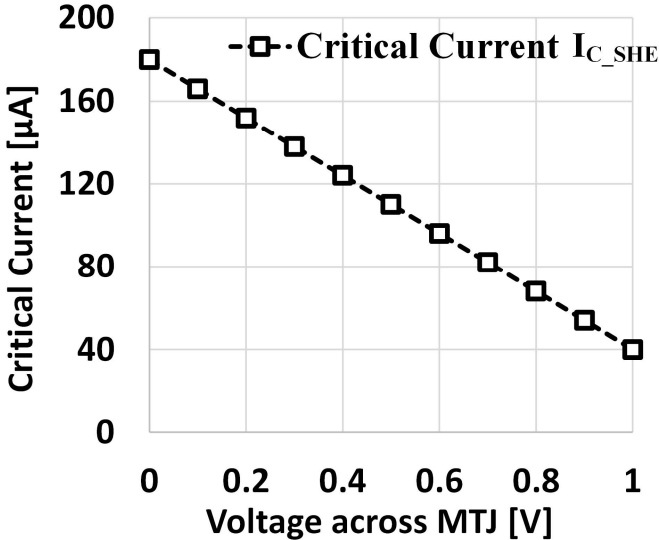


Fig. 3.24 Critical current I_{C_SHE} of the SHE effect as a function of the voltage across the MTJ. Reduced energy barrier leads to lowering the critical current.

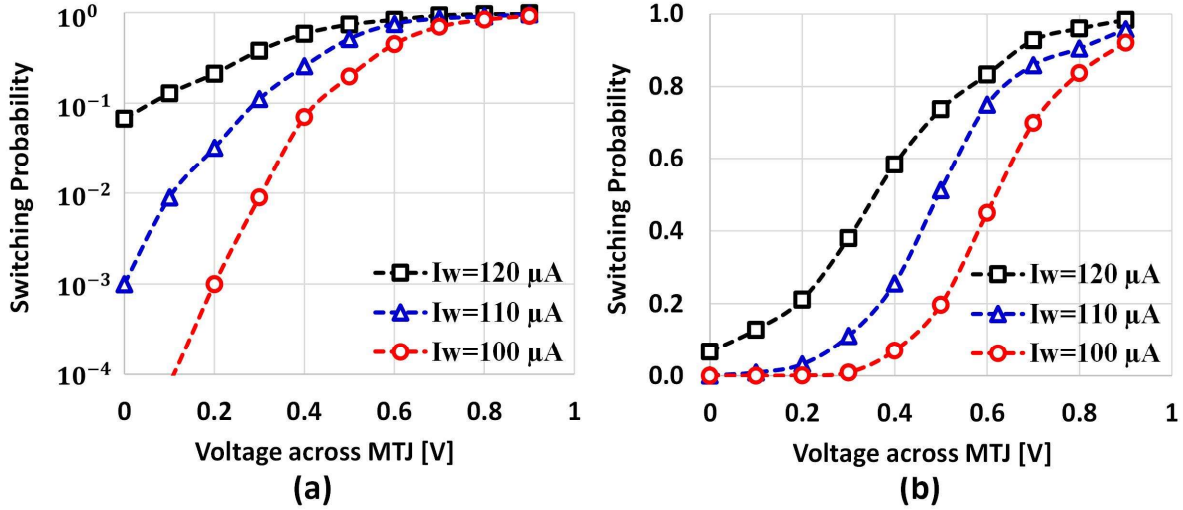


Fig. 3.25 Switching probability as a function of the applied voltage V_{MTJ} across the MTJ with different write current I_W flowing in the heavy metal. The switching probability approximately linearly increases as a function of V_{MTJ} within a certain range if $I_W \geq I_{C_SHE}$. (a) Log plot (b) Linear plot.

The VCMA effect assists energetically efficient SHE-driven switching and provides controllability of switching probability by modulating the critical current as shown in Fig. 3.24, which is expressed by the equation (2.21). Given a charge current (or write current) I_W flowing in the heavy metal layer, we can divide into two switching regimes depending on the magnitude of the applied voltage across the MTJ (V_{MTJ}). If the charge current I_W is smaller than the critical current I_{C_SHE} for the SHE effect, the switching probability increases exponentially as a function of V_{MTJ} . This is because the applied voltage effectively reduces the energy barrier between two states, facilitating the spin torque to overcome the barrier, as shown in Fig. 3.25(a). If I_W is equal or larger than I_{C_SHE} , the switching probability increases linearly as a function of V_{MTJ} , as shown in Fig. 3.25(b). This is because I_W provides sufficient spin torque to the free layer to overcome the energy barrier. However, finalization of the magnetization is completed by the MTJ's own effective anisotropy. Figure 3.26 shows the magnetization trajectory of the voltage assisted SHE-driven switching with different bias conditions where Fig. 3. 26(a) and (b) includes the thermal noise, and Fig. 3.26(c) and (d) do not. A higher V_{MTJ} causes a more stable magnetization trajectory during the finalization, making it less susceptible to

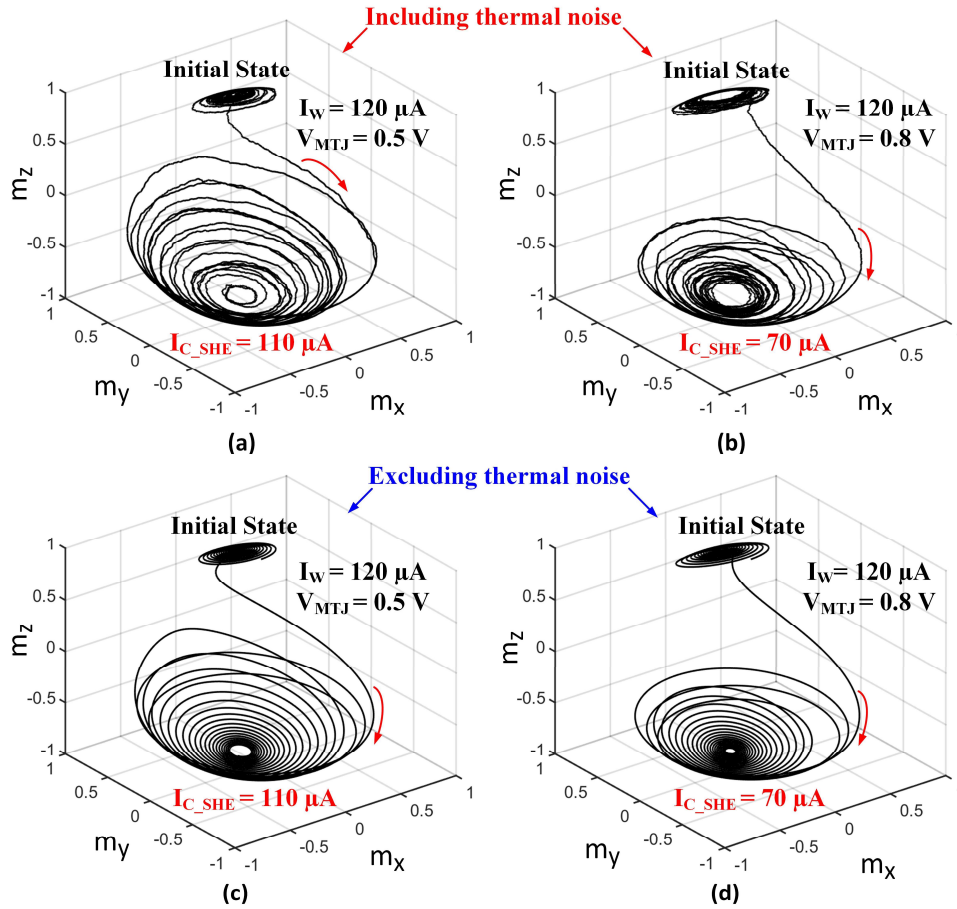


Fig. 3.26 Trajectory of the free layer's magnetization during the voltage assisted SHE switching with $I_W = 120 \mu\text{A}$. (a) $V_{MTJ} = 0.5 \text{ V}$ and (b) $V_{MTJ} = 0.8 \text{ V}$ include the thermal noise. (c) $V_{MTJ} = 0.5 \text{ V}$ and (d) $V_{MTJ} = 0.8 \text{ V}$ exclude the thermal noise.

thermal noise, which in turn increases switching probability. By providing a suitable write current, we can take advantage of the linear switching probability regime.

3.6.3 Design and Performance Evaluation

In the proposed spintronic ASC, the top electrode of the three terminal MTJ device is connected to the sensing circuit and the input transmission gate as shown in Fig. 3.27. Two write drivers (M2~M5) are connected to the edges of the heavy metal layer. Considering the resistance of the heavy metal ($\sim 330 \Omega/\text{cell}$) and the supply voltage ($\sim 1.2 \text{ V}$), the write drivers on a single heavy metal layer can be shared by sixteen MTJs and can provide a sufficient write current. Sharing the write driver can improve

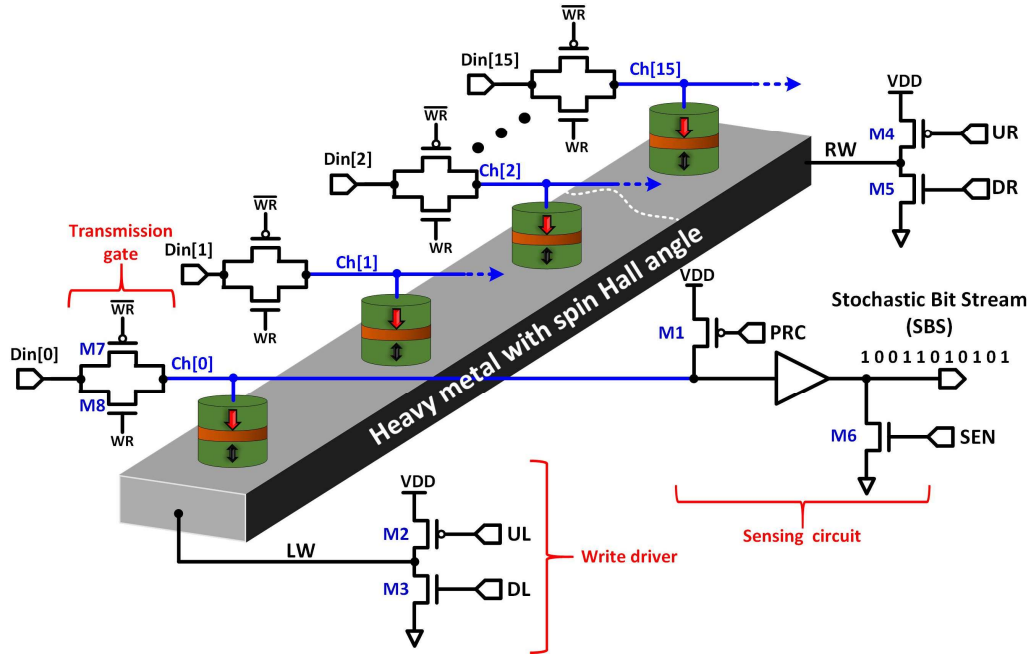


Fig. 3.27 Schematic of the proposed spintronic ASC consisting of sixteen MTJs on the heavy metal, shared write drivers, transmission gates, and sensing circuits. Multiple stochastic bit streams can be simultaneously generated.

the area, energy efficiency, and bandwidth via parallel conversion.

To generate each bit of an SBS, three operation modes are required: reset, write, and sensing. During the reset mode, all of the MTJs are reset to R_p by applying a relatively large charge current ($220 \mu\text{A} \gg I_{C_SHE}$) with a sufficient duration 3~4 ns to achieve high switching probability. For the write mode, the transmission gates are turned on, allowing the analog input signal to propagate from the $Din[n]$ node to the $Ch[n]$ node, which applies voltages across the MTJs (V_{MTJ}). A moderate write current ($I_W = 120 \mu\text{A}$) then flows in the heavy metal layer. Depending on the amplitude of an analog input signal, the MTJs switch to R_{AP} with a certain switching probability. During the sensing mode, the M1 transistor provides a proper sensing current flowing through the MTJ. Then, the MTJs states are converted to digitized signals via the sensing circuit.

The resolution of an SBS is determined by the number of iterations of these three modes described above. Figure 3.28 shows transient simulations of the proposed ASC with ten resolutions based on a macrospin three terminal MTJ compact model and 45 nm CMOS technology. As the amplitude of the

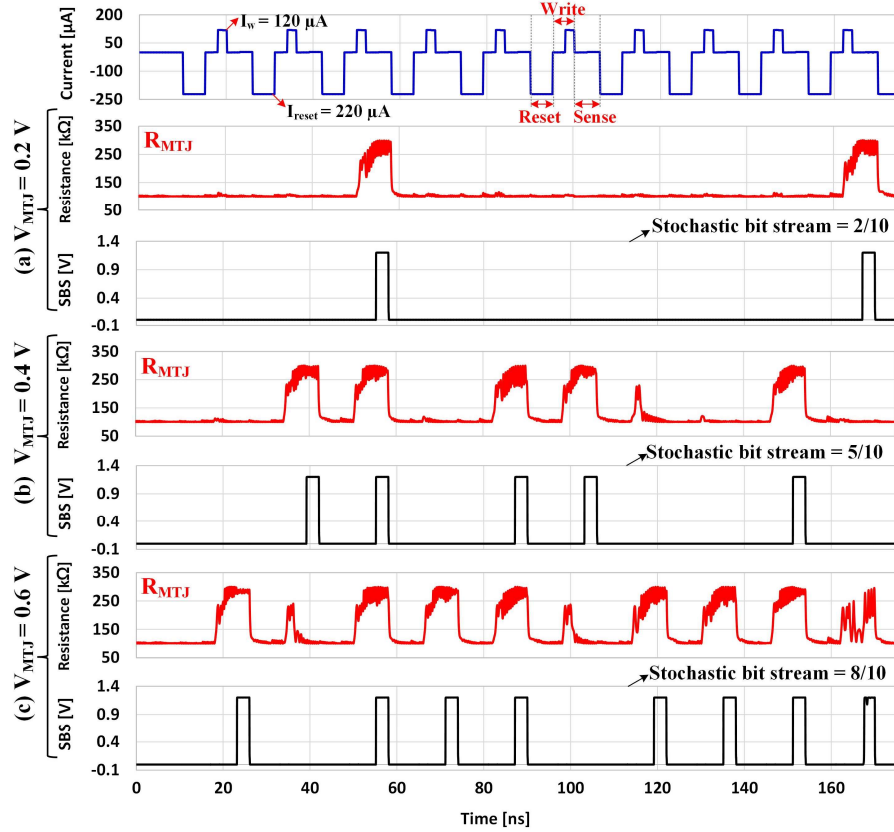


Fig. 3.28 Transient circuit simulation of the spintronic ASC with the three terminal MTJ compact model, generating a ten-bit resolution SBS. As the amplitude of analog input V_{MTJ} increases, the circuit generates more ‘1’ on its SBS. (a) $V_{MTJ} = 0.2$ V, SBS=2/10 (b) $V_{MTJ} = 0.4$ V, SBS=5/10 (c) $V_{MTJ} = 0.6$ V, SBS=8/10.

ASC	Node	Device	Power (μ W)	Area (μ m ²)
This work	45 nm	MTJ	< 13.8	0.41
[142]	90 nm	MTJ	95.2	1.62
[158]	65 nm	Memristor	370	1.80
[137]	45 nm	Pure CMOS	1383.5	79.3

Table 3.6 Performance comparison with previous works. Operation clock speed is 100 MHz.

analog input signal rises, the number of ‘1’s on the SBS also increases, corresponding to the data of Fig. 3.25 ($I_w = 120$ μ A).

The energy consumption of the CMOS component is only 10% of the entire conversion operation, which is 14 fJ per bit, including the ohmic loss in the write driver and the energy for sensing. The consumed energy of the SHE components for the reset and write operation are 88 fJ and 36 fJ per bit, respectively. Therefore, the total energy consumption for generating a stochastic bit is 138 fJ. We can

estimate that if the circuit operates at 100 MHz clock cycle, the total power consumption is 13.8 μ W. Table 3.6 summarizes the performance of the proposed work and provides a comparison with previous works utilizing other memory technologies.

3.7 Spin and CMOS based Neural Network

3.7.1 Motivation

Neuromorphic computing systems, which can mimic natural neuro-biological methods of information processing, are among the possible candidates for highly reconfigurable, intelligent system architectures. Artificial neural systems have been shown to have outstanding computational performance in many real-world applications, including pattern recognition, artificial vision, and robotics, compared to the Von Neumann system [148]. However, a purely CMOS-based neuromorphic processor is very inefficient in terms of both power and area since each neuron may require over 400 transistors [149]. Also, the volatility of CMOS memories is systematically unfavorable in a neuromorphic system, resulting in large static power dissipation.

To overcome the above-mentioned challenges, researchers have begun to develop artificial neurons using emerging, nonvolatile devices like the MTJ: a device whose magnetization can be manipulated by spin-polarized currents via STT [18], [111], [150]. This has led to a large reduction in the size of the artificial neuron, and a significant reduction in leakage current. Furthermore, the giant SHE has been shown to reduce the switching current of MTJs by at least 10x [21], [87]. In SHE, the charge current flowing in a metallic material with a large spin-orbit coupling (e.g. Ta, W, etc.) is converted into a spin current, which can exert spin torque and induce switching in an adjacent ferromagnetic material [22]. The SHE provides several potential advantages in the design of a practical neuromorphic system: (i) the symmetric generation of spin torques through SHE, unlike the asymmetric nature of STT, is analogous to the excitatory and inhibitory dendrites of neurons; (ii) the stochastic firing system is well emulated by the SHE-induced switching behavior in an MTJ, which can also be designed to be

stochastic by utilizing thermal activation in the switching process; and (iii) the SHE-based MTJ is a multi-terminal device with separated read and write current paths, allowing each to be independently optimized.

3.7.2 Design and Performance Evaluation

We propose a non-volatile, ultra-low-power, high speed, spin-based neural network (SNN), which outperforms the power, area, and speed of a pure CMOS neuromorphic system. The basic building block of the proposed neural network, the spiking core with weight systems, is shown in Fig. 3.29 where an MTJ (core MTJ) is located on top of a heavy metal with spin Hall angle (HMS). An STT-MTJ (synapse MTJ) based weight system is connected to each excitatory and inhibitory dendrite. The electromagnetic dynamics of the spiking core and weight system are captured by Verilog-A compact models based on the macrospin behavior for HMS and MTJ. The core MTJ with HMS mimics the

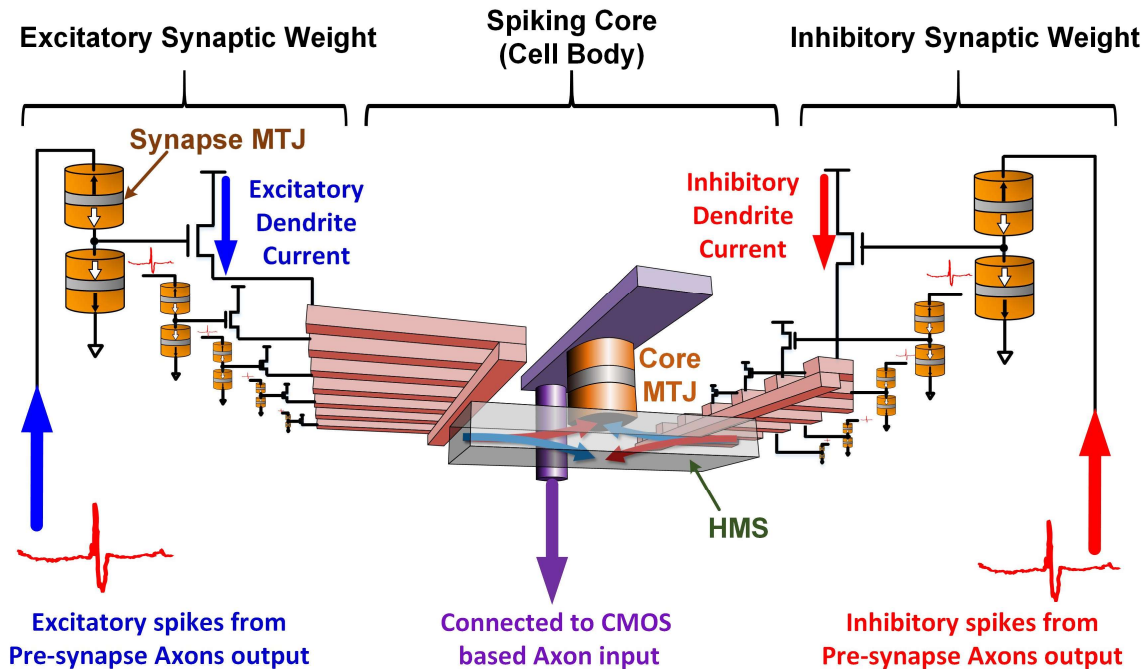


Fig. 3.29 Spin-based artificial neuron consisting of excitatory and inhibitory dendrites with MTJ based synaptic weight, an HMS with the core MTJ (spiking core), and an axon. An HMS can extract a net spin current by integrating all spiking current, switching the core MTJ. Once the free layer of the core MTJ is switched to the parallel (P) state, the CMOS based axon recognizes the state as the excitatory state.

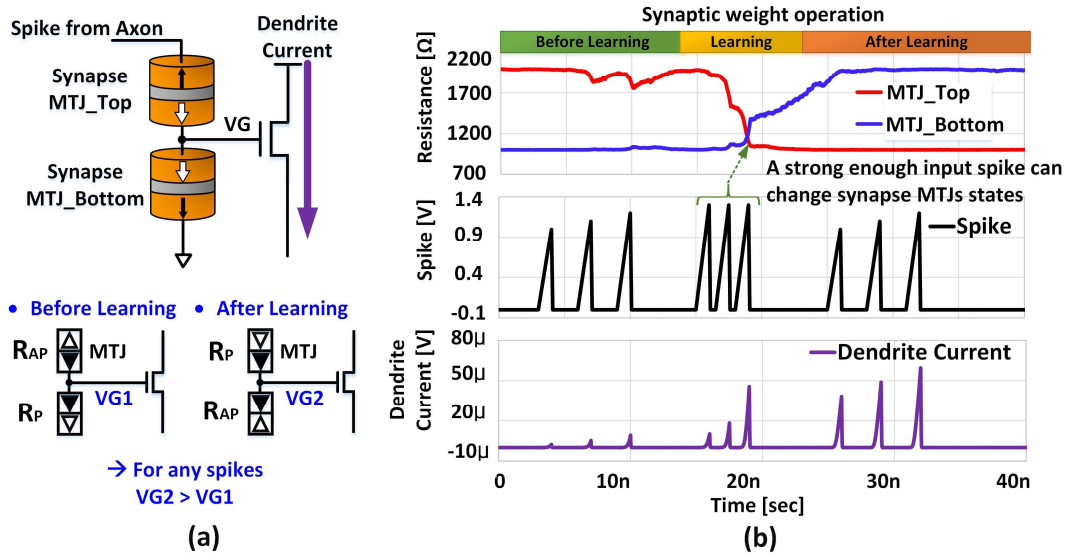


Fig. 3.30 (a) Schematic of CMOS and MTJ based artificial dendrite and synapse. Since the resistance of the synapse MTJs can be modulated depending on the strength of the spike, its state dramatically changes the conductance of the dendrite NMOS (b) Synaptic weight operation. Spikes above the threshold with sufficient frequency enable switching the state of the synapse MTJs, causing a critical learning. Then, a dendrite conductance exponentially increases.

functionality of the cell body in a biological neuron. In a network of biological neurons, information is passed between neurons via electrical impulses called “spikes”. Specifically, the spikes from excitatory neurons raise the membrane potential, while the spikes from inhibitory neurons reduce it. Spin-polarized current pulses function in the same manner in our proposed spin-based artificial neuron: current from an excitatory dendrite delivers a spin current to the free layer of the core MTJ via the SHE, contributing to a parallel (P) state (excitatory state), while current from an inhibitory dendrite transfers a spin current with the opposite polarization to the free layer, contributing to an anti-parallel (AP) state (inhibitory state). The HMS effectively integrates all these excitatory and inhibitory signals to generate a net spin-polarized current, removing the need for dedicated summation circuitry. Additionally, since the switching behavior of the MTJ is probabilistic, the proposed artificial neuron can emulate the stochastic resonance of a biological spiking system. Another significant component of the proposed artificial neuron is the STT-MTJ based synaptic weight, consisting of two MTJs and one NMOS where the states of the synapse MTJs depend on the strength of the spike. A strong enough

spike or high-frequency spike causes a critical learning, changing the conductance of synapse MTJs via the STT effect. After the critical learning, the dendrite current associated with the spike exponentially increases due to the raised gate voltage of the NMOS as shown in Fig. 3.30. The resolution of the weight can be increased further by connecting additional STT-MTJs in a serial manner.

A connection between the hybrid spin and CMOS based pre-synapse and the post-synapse is shown in Fig. 3.31. Each neuron can increase the number of connections with other neurons through synapses, creating a more complicated SNN. The output of the spiking core in the pre-synapse is connected to the CMOS based Axon input (*axon_in*). Its output (*axon_out*) derives the post-synapse dendrites as

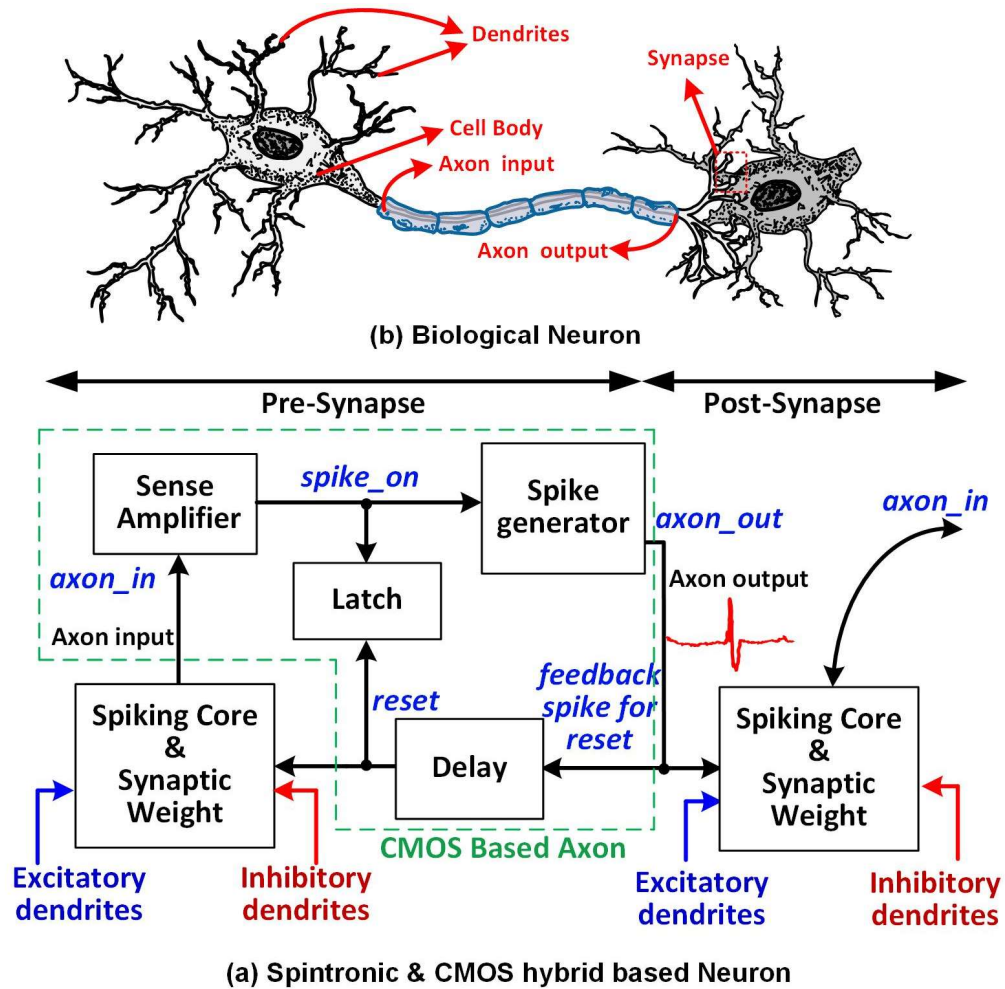


Fig. 3.31 Structure of the spin-based neural network in (a) analogous to that of biological neurons shown in (b). Each spin-neuron produces a binary spike, which is modulated by the synaptic weight (and the excitatory/inhibitory dendrites) to produce complex spiking behavior.

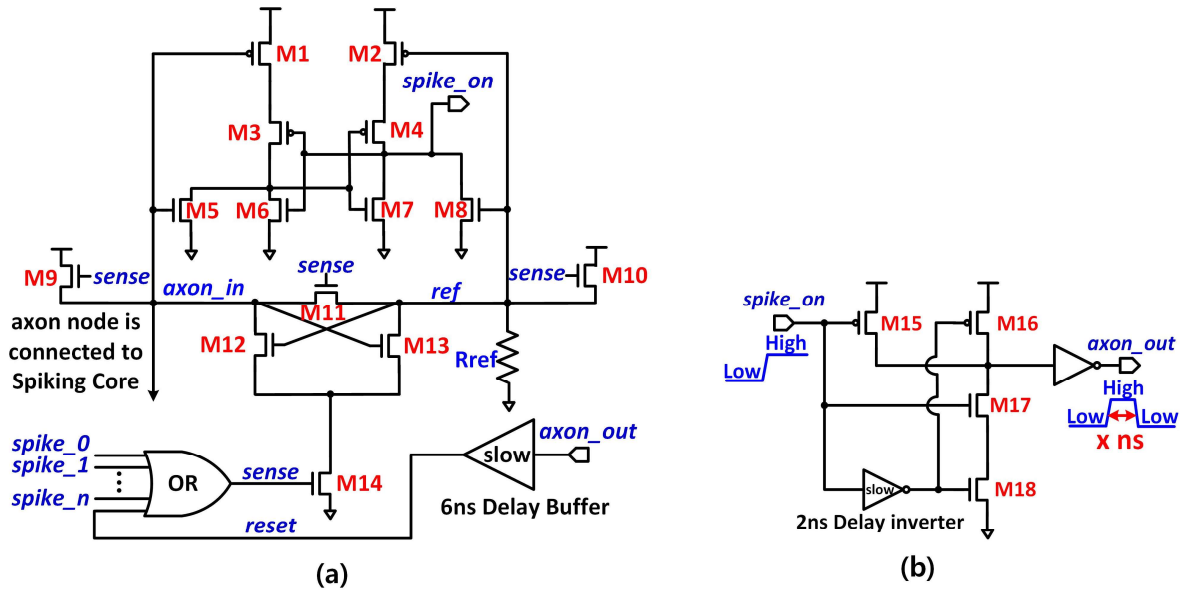


Fig. 3.32 Circuit components of a CMOS based Axon: (a) sense amplifier, latch, delay buffer. Whenever a spike or a reset signal comes in, the circuit evaluates the resistance of the core MTJ by comparing the potential of the *axon_in* node to that of the *ref* node. The comparison result is transferred to the *spike_on* as a digital signal. (b) Spike generator circuit in a CMOS based Axon. If the potential of *spike_on* changes from low to high, a spike (action potential) with 2 ns duration is generated at the *axon_out* node.

well as the pre-synapse inhibitory dendrite to provide a feedback spike for reset, which reverses the core MTJ state to the inhibitory state (initial state) after the MTJ changes to the excitatory state.

Figure 3.32 shows the Axon circuit, consisting of the sense amplifier, latch, delay buffer, and spike generator. An event-driven asynchronous operation of the proposed SNN further reduces power consumption. Specifically, the sense amplifier and the spike generator are triggered by spikes and the change of the core MTJ state, respectively. For example, when a spike comes in, the sense node changes its potential from '0' to '1', which in turn charges up both the *axon_in* node and the *ref* node. If the core MTJ state switches to the excitatory state due to the spike from excitatory dendrite, the sense amplifier detects the resistance difference by comparing the potential between the *axon_in* node and the *ref* node and updates its state as a digital signal '1' in the latch. The state '1' in the latch triggers the spike generator to fire a spike. Figure 3.33 shows the simulation result of event-driven CMOS based Axon operations with the spiking core. In addition to removing the dedicated summation

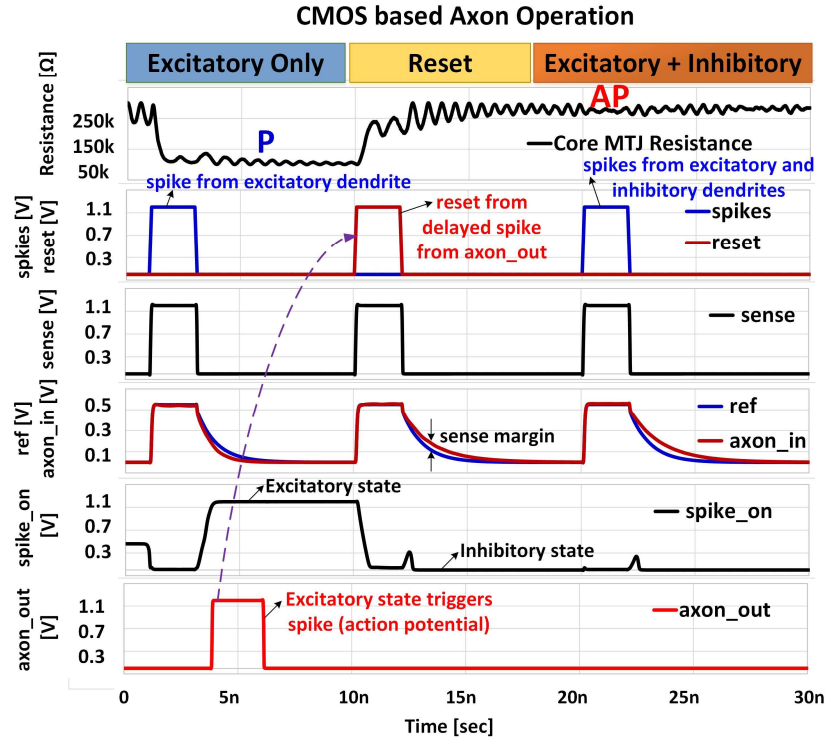


Fig. 3.33 Simulation result of CMOS based Axon operation (i) Excitatory only: a spike into an excitatory dendrite switches the core MTJ to P state, which triggers the action potential at the axon_out node. (ii) Reset: the action potential is propagated with a 6 ns delay and reset the state to AP state (iii) Excitatory + Inhibitory: spikes from both excitatory and inhibitory dendrites at the same time cannot switch the core MTJ state.

Symbol	Proposed SNN	SNN from Reference [148]
Energy	<100 fJ	1.95 pJ
Speed	< 6.5 ns	1,000 ns
Area	504 F^2	12,300 F^2

Table 3.7 Area and performance of the proposed spin and CMOS based hybrid neuron, including sixteen dendrites for synaptic weight. F is a minimum feature size.

circuitry, the sense amplifier and latch can be combined into a single circuit, allowing us to reduce the size of an artificial neuron by more than 25x. The area and performance comparisons are summarized in Table 3.7.

CHAPTER 4

DESIGN TECHNIQUES ENHANCING MeRAM PERFORMANCE

4.1 Motivation

An individual memory element has physical limits (i.e. speed, energy, endurance, and size) due to its intrinsic characteristics at the device level, however, the performance of a memory macro at the system level can be improved by using circuit design techniques in terms of access time, write error rate, and read disturbance. Likewise, the performance enhancement of MeRAM can be achieved by either fully utilizing unique features of a voltage-controlled MTJ via circuit design schemes or modifying a memory macro circuit. In this chapter, several circuit design techniques are introduced to enhance the performance of MeRAM, especially, for reducing read disturbance and write error rate, and enhancing sensing margin and cell area efficiency.

4.2 Source Line Sensing (SLS) Scheme to Improve Read Performance

4.2.1 Use of VCMA with reverse voltage

There are some challenges that currently prevent MeRAM from being implemented in embedded system memory applications. One problem is read failure, which occurs when a sensing circuit cannot distinguish between two states of the memory cell due to the small sensing margin. This is caused by the low tunneling magnetoresistance (TMR) ratio in typically used material systems in STT-RAM and MeRAM. As the sensing margin decreases, the memory becomes more susceptible to noise, increasing the read failure and requiring a dedicated circuit to amplify signals. The other issue is the read disturbance, a chance of flipping the MTJ state after applying an electric read pulse (i.e. the probability of a destructive read), which is not affected by TMR but by the thermal stability. The read disturbance

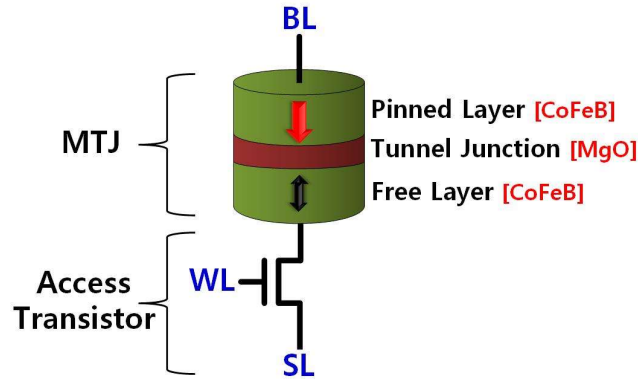


Fig. 4.1 One transistor and one MTJ of MeRAM cell structure with a transistor as the access device. The bit line (BL) and the source line (SL) are connected to the pinned layer of the MTJ and the source of the access transistor, respectively. Note that the orientation of the MTJ layers may be reversed depending on the sign of the VCMA coefficient. A word line (WL) controls the gate of the access transistor.

happens when a read operation occurs by charging up the bit lines to a certain voltage level (sensing voltage). To be specific, this bit line voltage reduces the energy barrier E_b between the two stable states of an MTJ via the voltage-controlled magnetic anisotropy (VCMA) effect. The possibility of the destructive read, therefore, increases exponentially as the sensing voltage of the bit lines increases since the applied voltage lowers the thermal stability ($\Delta = E_b/k_B T$) of the MTJ where k_B is the Boltzmann constant and T is temperature. To increase the sensing margin and reduce the read disturbance, we propose a source line sensing (SLS) scheme for MeRAM, which reversely exploits the VCMA effect to stabilize the bit during sensing. The basic concept of the reverse use of VCMA effect was introduced in our previous work, MeRAM based ternary content addressable memory (MeTCAM) application. However, we did not provide quantitative assessments based on experimental data and simulation data from a large number of trials.

In this section, besides providing a corresponding memory core circuit architecture for the SLS, we measured the retention time from nanoscale (60 nm) MTJs by changing the sensing voltage across the MTJ and extracted read disturbance via executing 10^{10} attempts based on the MTJ compact model simulation. The experiment data shows that the SLS (applying -0.6 V) lengthens the retention time by up to a million times ($10^6 \times$) compared to the bit line sensing (applying 0.6 V). Also, the simulation

results show that the SLS can significantly reduce read disturbance up to 10^9 x and improves its sensing margin by 3x.

4.2.2 Circuit Architecture of the SLS

There are two basic modes which need to be considered in memory operations: write and read (sense) operations. For the write operation, in the case of MeRAM, a pulse generator circuit applies a write pulse of suitable duration to the bit line (see Fig. 4.1) to switch the MTJ state. The basic modes are illustrated in Fig. 4.2, showing simulation results from a macrospin MTJ compact model incorporating the VCMA effect. In this simulation, a write pulse (1 ns, 1.2 V) can switch the MTJ state from P to AP or from AP to P. This demonstrates the resonant but non-deterministic feature of precessional switching, where the state of the bit is always reversed regardless of its initial state for a given pulse duration. If the control circuit uses the conventional BL sensing (BLS) scheme, a moderate

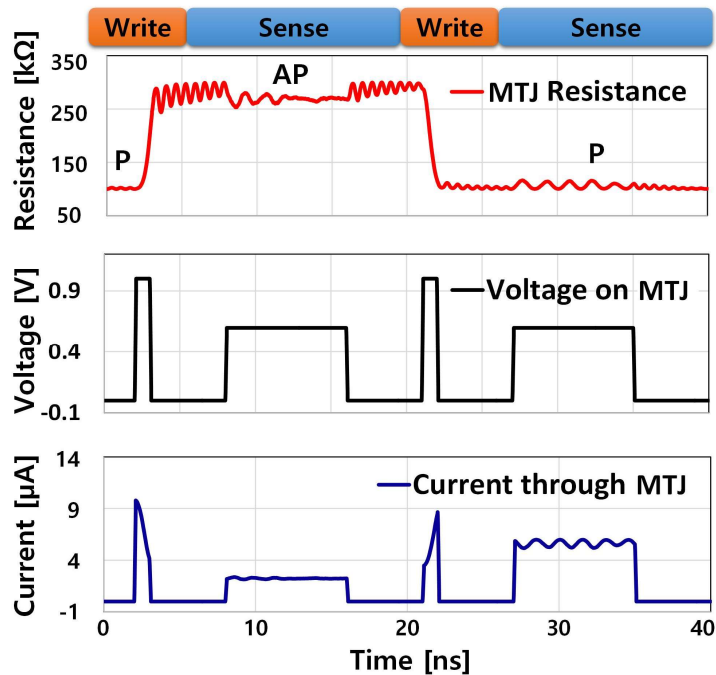


Fig. 4.2 Conventional write and read operations of MeRAM. For the write operation, a 1V write pulse with 1 ns duration is applied to the bit line to switch the MTJ state. A sensing voltage (0.6 V) is applied to the device during the read operation, which might cause unwanted switching due to the reduced energy barrier of the free layer.

sensing voltage (e.g. ~ 0.6 V) is applied to the bit lines for the read operation. Then, the sense amplifiers detect the voltage or current level of the bit line to sense the MTJ state.

Reading of a voltage-controlled MTJ, unlike typical STT devices, is strongly affected by the choice of voltage polarity during the read operation since the VCMA effect results in a change of PMA in the free layer under voltage application, which in turn leads to a coercivity change. This is illustrated in Fig. 3.6(a) for the case of MgO|CoFeB|Ta based MTJs, where a negative (positive) voltage across the perpendicularly magnetized MTJ increases (decreases) the coercivity of the bottom free layer. Figure 3.6(b) shows the measured corresponding coercivity change of an MTJ (RA product $650 \text{ } \Omega \cdot \mu\text{m}^2$, diameter 60 nm, 1.1 nm thick CoFeB free layer, 1.4 nm thick MgO barrier layer) as a function of voltages across the device. In this case, the MTJ has its free layer at the bottom, and the coercivity is enhanced as the amplitude of negative voltage increases. The change of coercivity, in turn, varies the thermal stability of the free layer. Although the non-vanishing STT effect shifts the offset field due to the electric DC bias (a few second), the effect is negligible in an actual read operation (5 ns \sim 50 ns).

As a result of the coercivity modulation, the BLS scheme has a possibility of causing read disturbances in MeRAM cell arrays. This is especially severe for embedded system memory applications that may require a relatively short retention time (< 1 ms) since they have a relatively low thermal stability ($\Delta \sim 20\text{-}30$) compared to storage applications (typically $\Delta > 40$). To reduce the read disturbance during the BLS scheme, the sensing voltage (pre-charge voltage) on the bit line should be limited. However, too low sensing voltage reduces the sensing margin.

The VCMA effect can be used to enable the SLS scheme, resulting in the reduction of the read disturbance and the improvement of sensing margin. A sensing margin can be defined by the potential difference between V_{sen} (the SL node) and V_{ref} (the REF node) as shown in Fig. 4.3. The key idea of the SLS is to apply a sensing voltage to the source line, hence increasing the coercivity of the MTJs during the read operation, taking advantage of the odd dependence of PMA on voltage in voltage-controlled material systems. Although applying a negative read voltage to the BL has the same effect

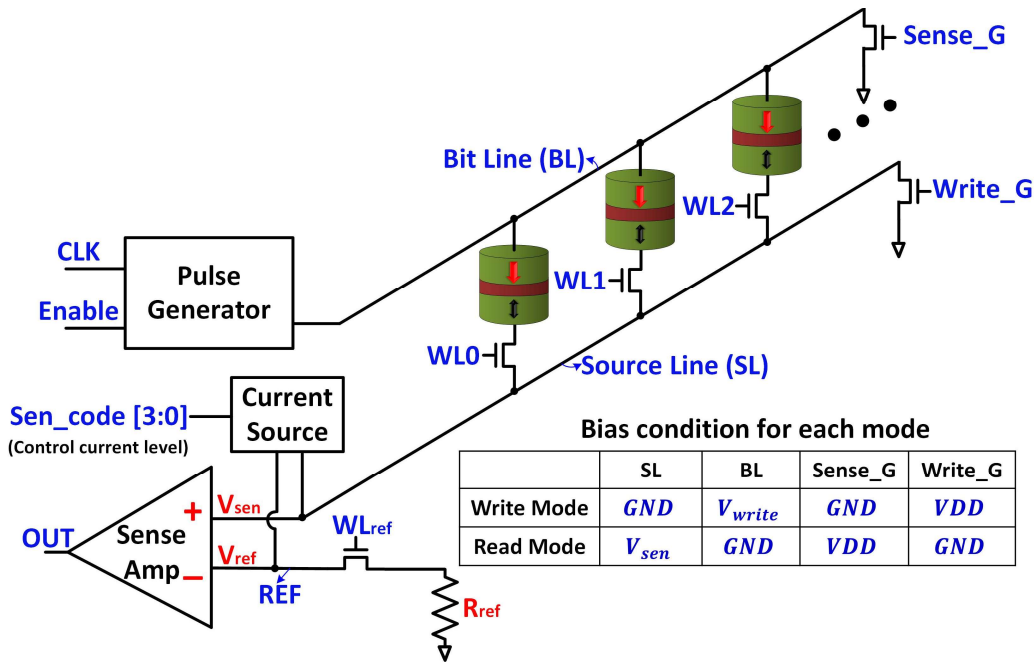


Fig. 4.3 Proposed core circuit architecture for implementing the SLS. The pulse generator is connected to the BL and provides a write pulse to switch a selected MTJ. The sense amplifier and the current source circuit are connected to the SL so that they generate a sensing voltage in an opposite polarity of the write pulse, reducing the possibility of the read disturbance.

as the SLS does, generating a negative bias requires more resources (e.g. charge pump circuit) in the chip where it has a positive power supply and a common ground. Figure 4.3 shows the proposed MTJ cell array and core architecture for realizing the SLS. In this architecture, both sense amplifier (sense amp) and current source are connected to the SL, and a number of MTJ cells are attached to between the BL and the SL. The pulse generator is connected to the BL. To select an MTJ, VDD should be applied to one of word lines (WLs) during the each operation.

For the write operation, the pulse generator provides a write pulse to the BL while the potential of the SL discharges to the ground level by applying VDD to the Write_G node. On the other hand, during the read operation, the BL is grounded by applying VDD to the Sense_G node. Then, the current source supplies a certain amount of current to the SL and the REF node, which generates electric potentials V_{sen} and V_{ref} at each node, respectively. The potential difference between V_{sen} and V_{ref} is amplified by the sense amp, generating a digital output '0' (P) or '1' (AP) at the OUT node.

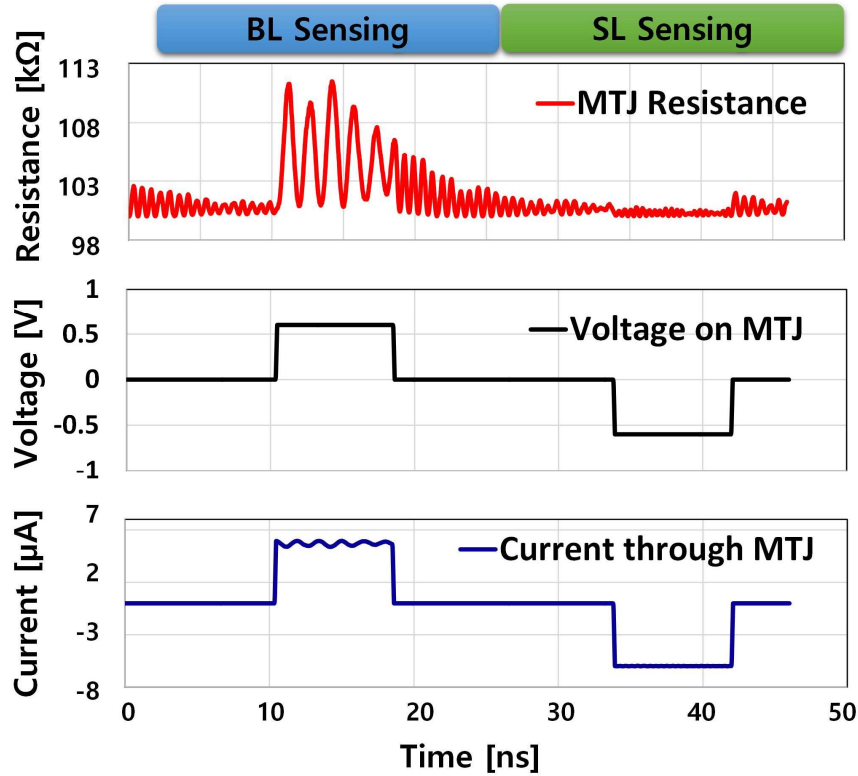


Fig. 4.4 Simulation results of the BLS and the SLS based on the MTJ compact model. A positive voltage across the MTJ causes resistance fluctuations up to 13% due to the reduced PMA. On the contrary, the resistance of MTJ is stable when a negative voltage is applied, achieving reliable sensing operation and reducing the read disturbance.

Figure 4.4 shows a comparison between the BLS and the SLS schemes based on the compact model transient simulation. In the case of the BLS, the MTJ resistance fluctuates up to 13%, resulting in an unstable sensing current which may cause the failure of the sensing operation. In addition, as it is being read, the MTJ state can be switched to the opposite state since the VCMA effect lowers the energy barrier, causing read disturbance. Therefore, the sensing voltage of the BLS needs to be carefully determined within a range where it avoids read disturbance while taking into account the energy barrier lowering. However, if the sensing voltage is too low, it limits sensing margins. During the SLS, in contrast, the resistance of the MTJ becomes more stable compared to the BLS case, which allows the sense amp to have more reliable sensing results. Furthermore, as the amplitude of the sensing voltage on the SL increases, the MTJ enhances its thermal stability. This implies that the sensing margin is not

limited by the sensing voltage, but rather by the voltage dependence of the TMR, hence obtaining a larger sensing margin compared to the BLS.

4.2.3 Evaluation

For quantitative assessment of the SLS, we measured the thermal stability and extracted the retention time as a function of the amplitude of an applied voltage to the MTJ where the device has a 60 nm diameter, and its thermal stability (Δ) is equal to 18 (retention time ~ 10 ms) at zero bias condition. The thermal stability is modulated at a rate of -16 V^{-1} as shown in Fig. 4.5(a). In the case of the BLS (0.6 V), the thermal stability is below 12 when the sensing voltage is applied. On the other hand, it reaches 25 in the case of the SLS (-0.6 V). These values, in turn, can be converted to the retention time as shown in Fig. 4.5(b) [151]. The retention times of the BLS and the SLS are 100 μs and 100 s, respectively.

The increase in the retention time of the SLS significantly improves the reliability of MeRAM for system memory applications. Although the MTJ compact model has different parameters ($K_i = 1 \times$

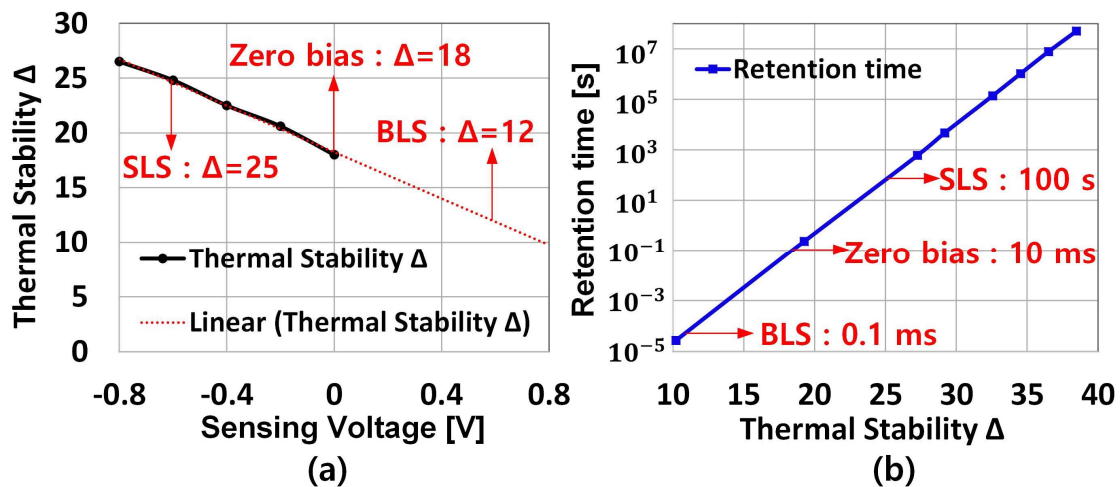


Fig. 4.5 (a) Measured thermal stability of an MTJ device with respect to voltage across the device. Due to the VCMA effect, the thermal stability is a function of the applied voltage, which has -16 V^{-1} slope. (b) Retention time is calculated based on the amplitude of the thermal stability at room temperature. At zero bias condition, retention time is 10 ms. During the BLS, the retention time is reduced down to 0.1 ms, increasing the possibility of the read disturbance while the retention time increases up to 100 s by using the SLS.

10^{-3} j/m^2 , $\Delta = 29.2$ at zero bias, $\xi = 100 \sim 200 \text{ fJ/V} \cdot \text{m}$, *diameter* = 60 nm) compared to the measured data, it is possible to quantitatively evaluate the read disturbances in terms of switching probability. In this simulation, we assume the MTJ has a relatively large VCMA coefficient. Figure 4.6(a) shows the read disturbance as a function of the sensing voltage whose duration is 50 ns. There is a significant read disturbance under the positive sensing voltage (BLS), resulting in the reliability issue. However, the SLS achieve the read disturbance below 10^{-12} , which is a fairly acceptable value in practical applications, and it decreases further by increasing the sensing voltage on the SL.

The sensing margin relies on the applied sensing voltage and TMR [152]. If TMR is a constant value, the sensing margin can be improved as the applied sensing voltage increases. However, actual TMR is reduced as the sensing voltage rises. Thus, there is an optimal sensing voltage that gives rise to the maximum sensing margin with an acceptable read disturbance. In this simulation (relatively high VCMA coefficients), we compare V_{sen} (the SL node) and V_{ref} (the REF node) and assume that TMR (at zero bias 52%) is a function of the applied sensing voltage. As shown in Fig. 4.6(b), in the BLS approach, the sensing margin is limited by the sensing voltage at 0.4V in which a significant read disturbance ($> 10^{-6}$) occurs, obtaining the maximum sensing margin of 50 mV. However, in the case

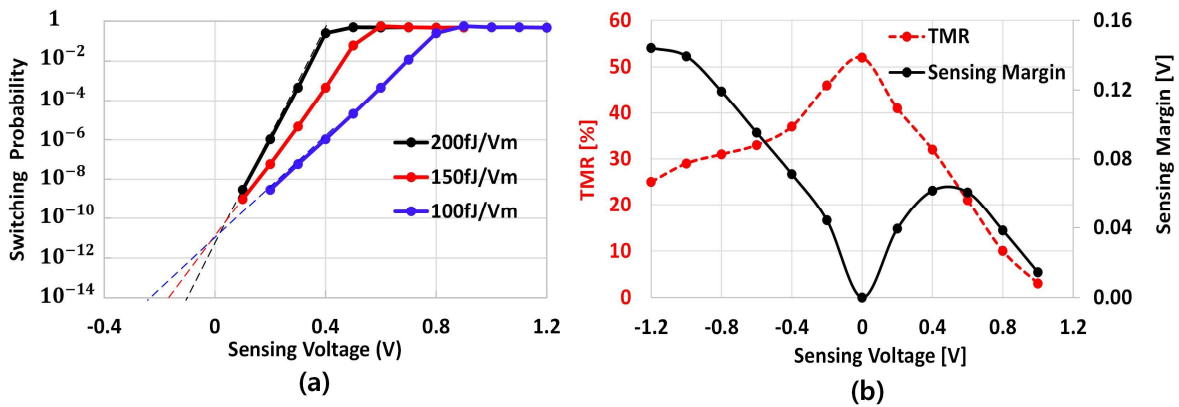


Fig. 4.6 (a) Read disturbance (in terms of switching probability) as a function of sensing voltage via the MTJ compact model simulations. If the sensing voltage exceeds 0.6 V (BLS), the read disturbance rapidly rises and converges to 50%. However, in the negative bias region (SLS), it achieves the read disturbance below 10^{-12} . (b) Maximum sensing margin as a function of sensing voltage. In the case of SLS, the sensing margin increases as the absolute value of sensing voltage increases, until it is governed by the voltage dependence of TMR.

of the SLS, the sensing margin can be achieved 150 mV at the sensing voltage -1.2V without causing any read disturbance. Above this sensing voltage ($>|-1.2\text{ V}|$), the sensing margin starts to diminish since the reduced TMR cancels them out. In short, the sensing margin of the SLS is limited not by the voltage which causes read disturbance but by the voltage that limits TMR, achieving 3x higher sensing margin compared to that of the BLS.

4.3 Word Line Pulse (WLP) based Write Operation for Reduced Write Error Rate

4.3.1 Pulse Shape Dependence of Magnetization Dynamics

To enable MeRAM to be realized in practical embedded system memory applications, low write error rate (WER) needs to be achieved. Write error in MeRAM is mainly caused by a degraded write pulse (e.g. slew rate and duration) and can limit its applications in high-speed memories. As an example, if WER is relatively high (e.g. $\sim 10^{-3}$), multiple write operations are required to achieve an acceptable bit error rate (BER) (i.e. $< 10^{-9}$) [128], hence, the total write access time could become too long to meet the speed requirement of embedded system memory. Although the WERs based on the VCMA effect assisting STT or spin Hall effect (SHE) switching methods are less sensitive to the write pulse shape and duration, they require additional time and energy compared to pure VCMA-driven precessional switching [153]–[155]. The main contribution of this work is to address the WER related challenge as described above by using a new scheme to improve the write pulse-shape for a significant reduction of WER based on 4x smaller word line (WL) and bit line (BL) drivers compared to a conventional method.

We already described that the duration and amplitude of the write pulse are important to control the WER. In addition to these factors, the dynamics of the magnetic moment is largely affected by the slew rate, rising and falling time of the pulse. During the write operation, \vec{H}_{eff} needs to be a constant field aligned with in-plane direction to maintain a stable precessional motion. Otherwise, the trajectory of the magnetic moment deviates from the precessional route. In chapter 2, Fig 2.7 compares the

magnetization dynamics between two extreme cases: applying a square pulse and a triangular pulse. In the case of applying a square pulse as shown in Fig 2.7(a) and (c), the initial state of the free layer is the P state ($m_z = -1$) and it starts to precess around the \hat{x} axis at $t=t_{i(a)}$. Since the PMA is abruptly reduced by the VCMA effect, which gives rise to a relatively constant in-plane component of \vec{H}_{eff} , the magnetic moment of the free layer can have a stable precessional trajectory and switch to the AP state ($m_z \approx 1$) at $t=t_{f(a)}$. However, in the case of applying a triangular pulse as shown in Fig 2.7(b) and (d), the direction of \vec{H}_{eff} is no longer in-plane. Instead, \vec{H}_{eff} gradually changes its direction from out-of-plane to in-plane as a function of time, which in turn causes an unstable precessional motion. At the end point of the triangular pulse $t=t_{f(b)}$, the magnetic moment cannot reach 180° reorientation ($m_z \approx 0.72$). After removing the pulse, hence, the magnetic moment converges to the AP state via the damping and precessional motion driven by its intrinsic anisotropies. During this process, the device becomes susceptible to thermal noise, which can produce a switching fail and increase the WER.

4.3.2 Timing of the WLP

Instead of applying the write pulse to the bit line (BL), called BLP, we propose a method of applying the write pulse to the word line (WL), which is referred to as WLP. The WLP can create a better square shaped write pulse across the MTJ, which in turn improves switching probability, and minimize the area overhead (e.g. driver size). There are three reasons why the WLP can have a better pulse shape compared to the conventional BLP scheme [128]: (i) eliminating discharge path during applying a pulse on the WL; (ii) using the gain of the access transistor in a selected cell; (iii) reducing the capacitive loading which needs to be driven. Further explanation of the reasons will be discussed below. Figure 4.7 shows a schematic design where MeRAM cells are connected to the BL and WL drivers. To achieve a fair performance comparison between the WLP and the BLP, we intentionally designed both drivers based on the same size transistors. We assume that the bit line capacitance C_{BL} is equal to

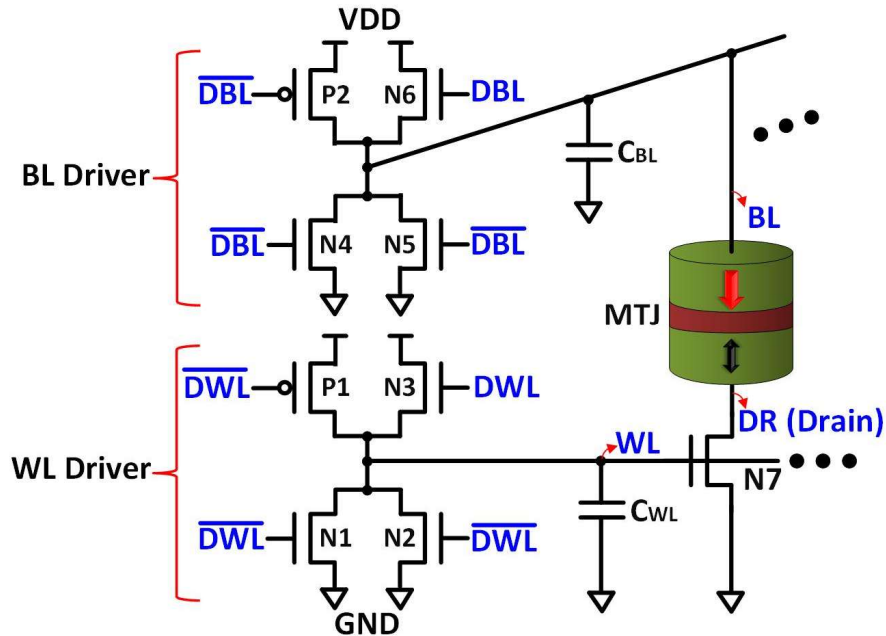


Fig. 4.7 Schematic of cell array architecture, including the BL driver and WL driver. The number of access transistors connecting the WL and the WL length determines its capacitive loading C_{WL} . The number of MTJs connecting the BL and the BL length decide its capacitive loading C_{BL} . The size of the drivers is carefully chosen based on the magnitude of the capacitance of each line to generate a suitable pulse shape.

the word line capacitance C_{WL} . Note that there are n-channel transistors (N3 and N6) in the pull-up path of each driver. These n-channel transistors can supply a large amount of charge at the beginning of the charge up for writing, compared to that of the same sized p-channel transistors, due to higher mobility, resulting in a better square pulse shape. But they gradually turn off as the potential of the target node (WL or BL) increases.

The control signals of the conventional BLP are shown in Fig. 4.8(a) where the DWL and the DBL enable the WL driver and the BL driver, respectively. The \overline{DWL} and \overline{DBL} are their complementary signals. We assume that the rising and falling times of the control signals are 100 ps. For the BLP, the WL driver is enabled first at $t=t_{WL_a}$, which charges up the selected WL to VDD, turning on the access transistor (N7). Then, the DBL triggers the BL driver that starts to charge up the BL at $t=t_{BL_a}$. However, this scheme deforms the write pulse shape because the BL driver directly drives the entire BL capacitive loading C_{BL} , and some portion of the electric charge leaks through the unselected

MeRAM cells, which prohibits the BL from reaching VDD within a 1 ns period and increases the rising time of the write pulse.

By contrast, in the case of the WLP, the waveform of the control signals (*DWL* and *DBL*) are shown in Fig. 4.8(b). The BL is charged first up to VDD, and the drain (DR) of the access transistor (N7) is also charged up to VDD since the N7 turns off at $t=t_{BL_b}$. Then, the WL driver is enabled and starts to increase the WL potential at $t=t_{WL_b}$. The slew rate of the WL is improved by 20% compared to that of the BLP since the gate of the access transistor (N7) provides a high input resistance, eliminating a discharge path. Furthermore, the WLP can efficiently utilize the current gain of the access transistor N7 through a common-source stage. Even below the threshold of the N7, the current flowing through N7 exponentially increases as a function the WL voltage. Above the threshold, the provided current increases quadratically as the WL voltage increases further.

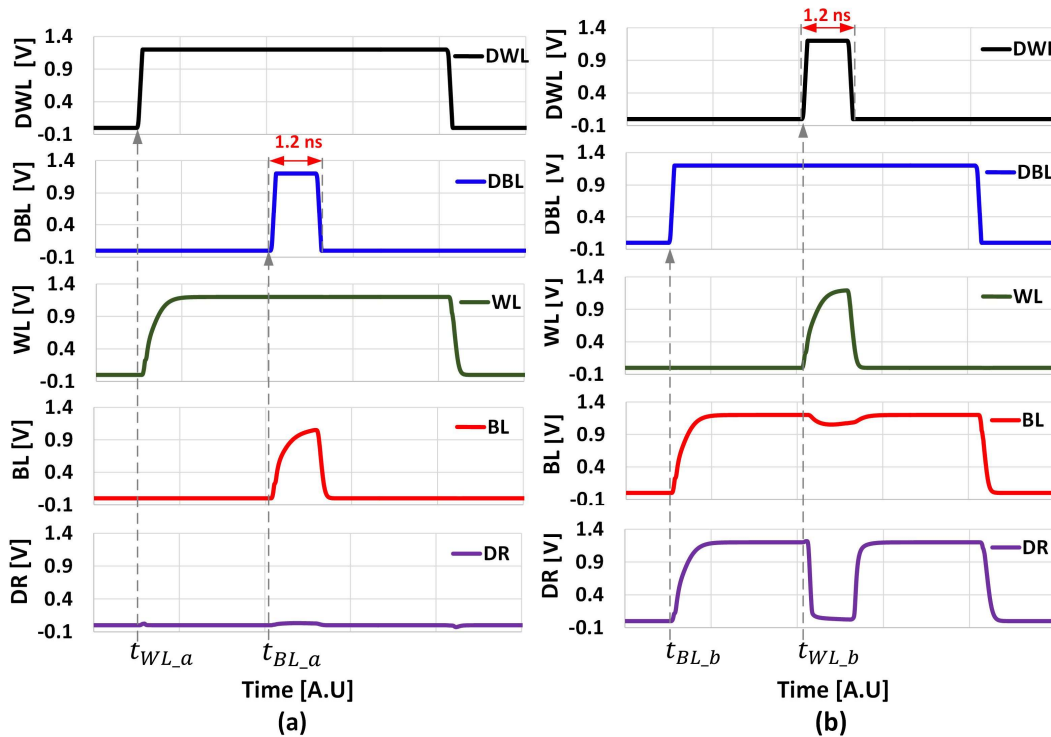


Fig. 4.8 (a) Conventional BLP scheme. After the WL driver completes charging the WL up to VDD, the BL driver applies a write pulse to the BL. (b) Proposed WLP scheme. The BL and DR are pre-charged to VDD, then a write pulse is applied to the WL via the WL driver. The WLP can make a better square shape write pulse based on the same size driver compared to the write pulse from the BLP.

# of cells	BL			WL		
	C [fF]	R [Ω]	RC [ps]	C [fF]	R [Ω]	RC [ps]
128	8	45	0.4	9	45	0.4
256	17	90	1.5	19	90	1.7
512	33	179	5.9	38	179	6.8
1024	66	358	23.6	75	358	26.9

Table 4.1 Resistive and capacitive loads on the BL and the WL. Assuming 28 nm node, $25F^2$ unit cell size.

Last but foremost, the WL voltage rapidly discharges the DR node to ground via the N7 transistor, since the capacitance on the DR node consists only of the MTJ and the access transistor (N7) itself, which is significantly smaller than the C_{BL} . Thus, these effects result in a better square shape pulse across the MTJ, allowing the circuit to achieve more reliable write operation.

The resistance and capacitance of the BL and the WL in an array level can be calculated by using the cell dimension, the sheet resistance, and the capacitance per unit length. We assume that the sheet resistances of the metal layer for the BL and WL are $0.14 \Omega/\square$, and the capacitance per unit length is $0.2 \text{ fF}/\mu\text{m}$ when the metal width is equal to $0.1 \mu\text{m}$. Based on CMOS 28 nm technology node with $25F^2$ cell size (F is the minimum feature size), the dimension of the unit cell is $0.14 \mu\text{m} \times 0.14 \mu\text{m}$. If the width and length of the access transistor (a standard logic transistor) are $100 \text{ nm} / 30 \text{ nm}$, its gate and junction capacitances are 57 aF and 48 aF , respectively. Table 4.1 shows the estimated values of RC loading on the WL and the BL in the array.

The voltage across the MTJ is the potential difference between the BL and the DR nodes ($V_{BL} - V_{DR}$). Figure 4.9 shows V_{MTJ} with a corresponding MTJ resistance change based on the BLP (black) and the WLP (red) as a function of the capacitive loading on the BL and the WL. As the capacitive loadings increase, the write pulse is severely degraded especially in the BLP case. Eventually, it fails to switch the MTJ with $C_{BL}=30 \text{ fF}$, which is approximately equivalent to the number of 512 memory cells on the BL since the pulse becomes a triangular shape and its amplitude also diminishes. In contrast, the WLP generates a square shape pulse regardless of the amount of the capacitive loading (within the capacitance range for the simulation), successfully switching the MTJ state, although the slew rate is

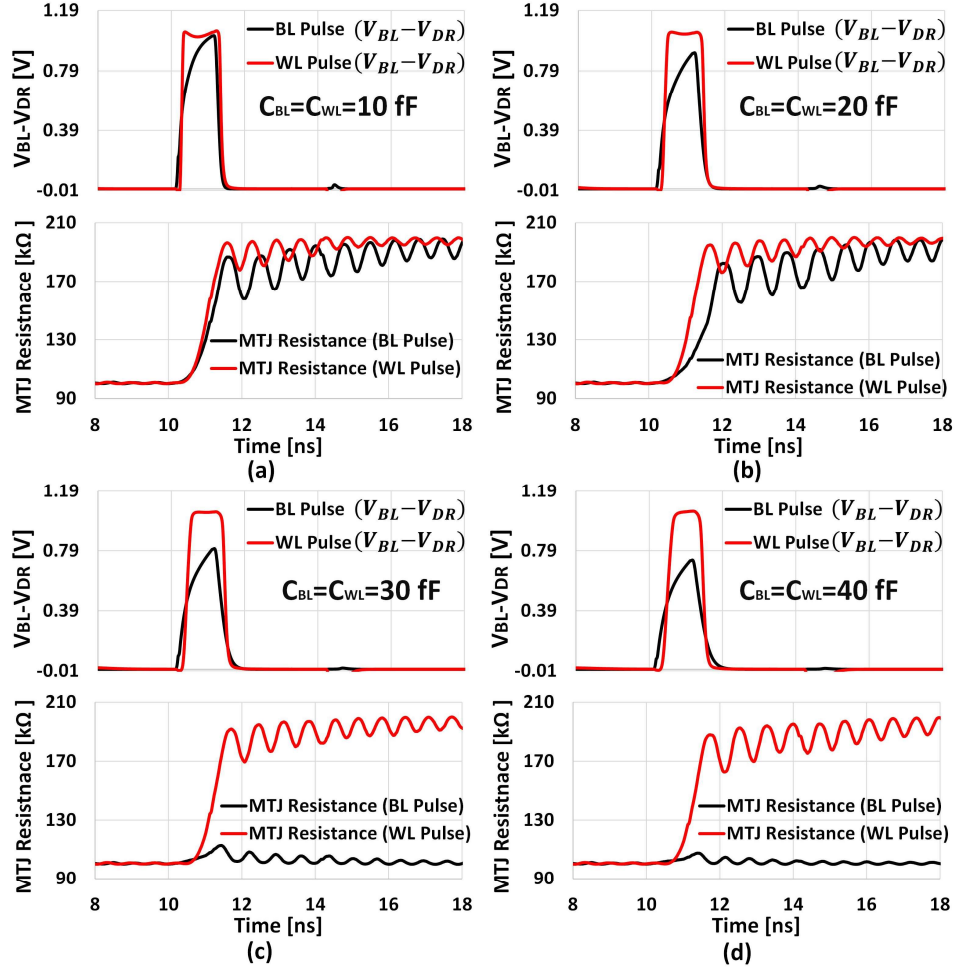


Fig. 4.9 Circuit simulation results of the BLP (black line) and the WLP (red line) based on the same size driver (the minimum size driver $W/L=160\text{ nm}/120\text{ nm}$). The write pulse from the BLP become degraded as the capacitance increases, and fail to switch the device beyond 30 fF. The WLP generates a square shape write pulse even under the largest loading of 40 fF and succeeds in switching the device.

slightly increased. A quantitative evaluation of switching probability (or WER) will be discussed in the next section.

4.3.3 Performance Evaluation: Write Error Rate and Cell Area Efficiency

The WER, defined as the number of switching failures divided by a total number of write trials, is an important indicator to evaluate the performance of a write operation. Specifically, the WER influence the total access time of a memory system [128]. Because if a memory cell has a high WER at given write pulse, multiple write operations are necessary to achieve an acceptable BER, which is

the maximum WER that can be successfully managed by an error correction code (ECC) algorithm built in the memory system [156].

In order to understand which component is a dominant factor on the WER between the slew rate and the amplitude of the write pulse, we independently executed the WER via the macrospin compact model simulations with an ideal voltage source based on two conditions as shown in Fig. 2.8 (a) and (b), respectively. Figure 2.8 shows that the both components influence on the WER in an exponential manner since the energy barrier between the two states linearly decreases as a function of the amplitude of the applied voltage, and the slew rate decides the trajectory of the magnetic moment and the effective pulse width.

To quantitatively evaluate the performance of the BLP and the WLP, the WER of both cases are extracted via 10^{10} trials under the condition where both BL and WL drivers use the minimum size transistors for the fair comparison [48]. However, in the case of actual memory design, the size of drivers should be adjusted with respect to the capacitive loading to achieve an acceptable BER. Figure 4.10 shows the WER comparison result between the BLP and the WLP. Since the BLP fails to generate

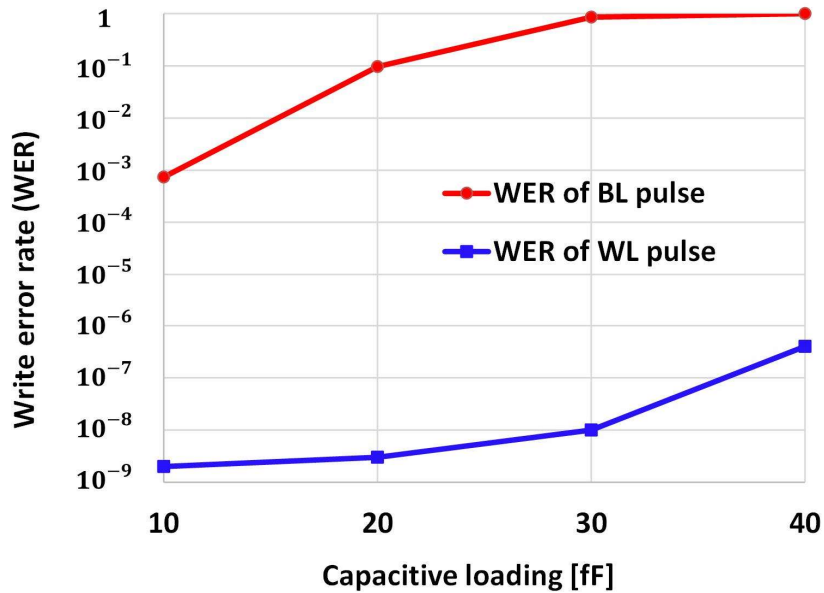


Fig. 4.10 Write error rates of the BLP and the WLP with respect to the capacitive loading via 10^{10} transient simulation trials with the minimum size driver. The WLP achieves on average seven orders of magnitude lower WER as compared to that of the BLP under the same condition (e.g. driver size, loading).

a proper write pulse in terms of slew rate and amplitude, the WER of the WLP is on average seven orders of magnitude lower than that of the BLP through the given capacitive loading (10 fF ~ 40 fF) based on the minimum size driver.

Note that the WER of the BLP with low capacitive loading (10 fF) is mainly due to the slew rate (rising time > 0.3 ns) compared to that of the WLP because both schemes reach the same amplitude (~1.1 V) as shown in Fig. 4.9(a). However, as the capacitive loading increases, the amplitude of the write pulse becomes the main reason for such high WER of the BLP. Because the amplitude decreases with a faster rate compared to that of the slew rate (see Fig. 4.9), which exponentially increases the WER.

A simple way to improve the write pulse shape is to increase the size of the transistor in the drivers associated with the loadings on the BL or the WL. However, the increase in the driver size might limit the memory capacity in a given die area, resulting in low cell area efficiency. The cell area efficiency is typically used as a target parameter to compare the compactness of memory designs, which is defined as follows:

$$Cell\ Area\ Efficiency = \frac{Area_{(cell\ array)}}{Total\ Die\ Size} = \frac{Area_{(cell\ array)}}{Area_{(cell\ array)} + Area_{(logic\ and\ analog\ circuit)}} \quad (4.1)$$

To achieve a high cell area efficiency the logic and analog circuits should be minimized while fulfilling the required performances such as access time and power. In a typical MeRAM macro design, the BL drivers and WL drivers occupy 14% and 4% in the total die area, respectively. Because the BL drivers must drive a significant amount of capacitive loading in the range of 10 - 500 fF within a nanosecond, depending on memory array size and technology node [157]. If we put the areas of the logic and analog circuits and the cell array into the equation (4.1), the cell area efficiency is 67.8% for a conventional BLP based MeRAM design.

A reduction in the driver size may improve the cell area efficiency. The proposed WLP allows reducing the driver size while generating a good square shape pulse. Figure 4.11(a) shows a required

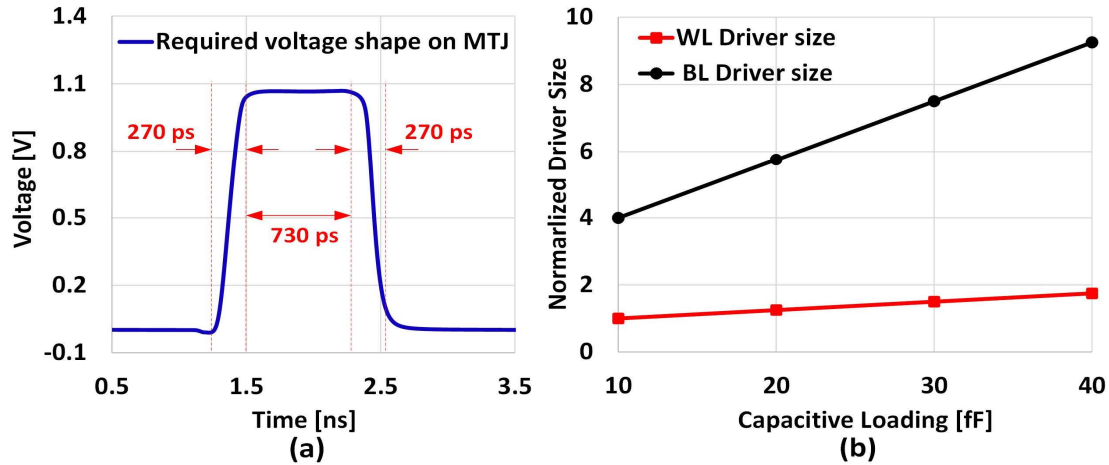


Fig. 4.11 (a) Required pulse shape that achieves an acceptable BER ($< 10^{-9}$). (b) Normalized driver size associated with a given capacitive loading to generate the required pulse shape. Note that the WLP allows a chip to have 4x the area-efficiency of drivers as compared to the BLP.

write pulse shape that provides an acceptable BER ($< 10^{-9}$). Figure 4.11(b) presents a normalized driver size, which can generate the required write pulse shape, in a given capacitive loading. Compared to the driver size of the BLP, the WLP can produce the required write pulse with a suitable slew rate by using a 4x smaller driver. Therefore, the WLP has the potential to reduce the occupancy rate of the drivers in the total die size, resulting in 76.8% cell area efficiency.

4.4 Write Pulse Termination (WPT) Circuit Technique

4.4.1 Motivation

It has been experimentally observed that the switching probability of voltage-controlled MTJ has an oscillatory behavior as a function of the applied write pulse width due to the nature of precessional switching as shown in Fig. 4.12 [68], [69]. Therefore, generating an optimum write pulse width to achieve the lowest WER is one of the most significant challenges that we need to solve for realizing MeRAM implementation. However, in an array level, it would be difficult to choose the optimum write pulse since an individual cell has a different switching behavior due to the process variations. There are several approaches that can deal with this challenge.

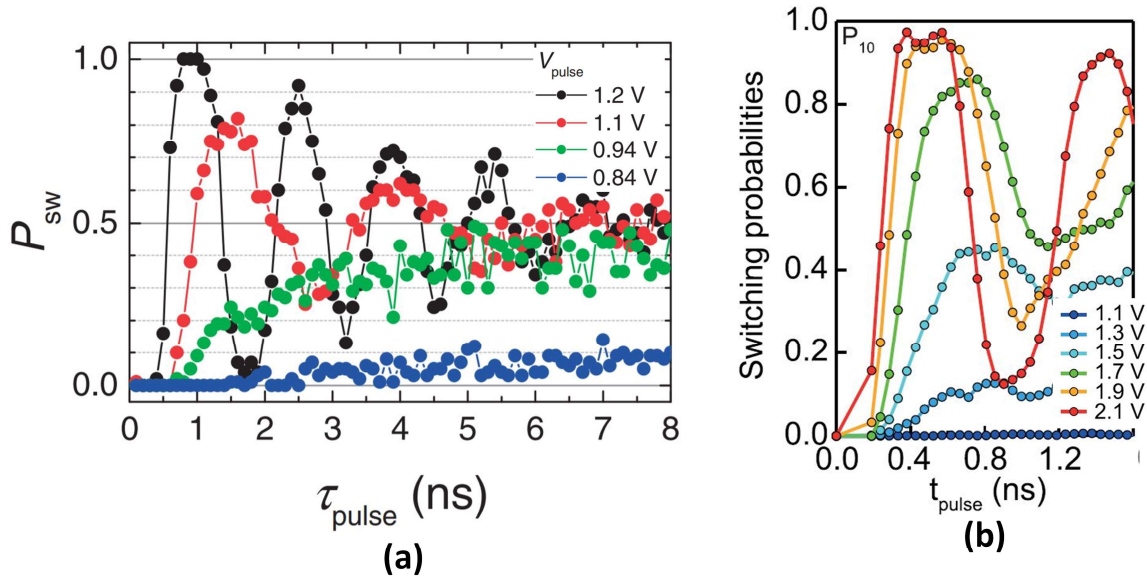


Fig. 4.12 Experimentally observed an oscillatory behavior of the switching probability of the voltage-controlled MTJ as a function of the applied pulse width and amplitude. Presented data from (a) reference [68] (b) reference [69].

The first method is an incremental pulse width scheme via multiple write-verification operations. If a memory cell fails to switch its state, the peripheral circuit increases the pulse width and applies only to the failed cells at the next write operation. Based on a CMOS technology (< 130 nm node), it is possible to control the pulse width with a 50 ps time resolution. However, this method causes a relatively long latency due to the multiple operations.

The other approach is to develop a write pulse termination (WPT) scheme that reduces the WER by applying write pulses to selected cells together via the word line (WL) and individually removing the write pulse via each bit line (BL) once the individual cell is switched to the desired state. As shown in Fig. 4.12, there is a time window (300 ps ~ 500 ps) where it has the relatively high switching probability. To realize the WPT scheme in MeRAM application, the feedback circuit needs to have a short response time to enable the pull-down transistor of the BL driver for removing the write pulse right after detecting the state change. Although it would be challenging to design a circuit for the WPT, especially in a large size cell array due to the inherent long RC delay of the array, in this section, we introduce a basic concept of the WPT technique.

4.4.2 Schematic of the WPT

A write pulse termination (WPT) scheme is proposed, along with a corresponding memory core circuit architecture, for the read and write operation of MeRAM. The WPT is aiming for reducing the WER and the latency by individually adjusting its write pulse width on each bit line (BL) and removing the pre-reading, respectively. The voltage divider allows the BL to vary its potential associated with the MTJ resistance changes when the write pulse is applied. The WPT circuit consists of an amplifier, a comparator, a multiplexer, and a termination digital controller. The WPT circuit amplifies a fluctuated small signals on the BL and enables the pull-down transistor of the BL driver, terminating the write pulse when the state of the MTJ device is converted to the desired state. If the MTJ initially has the desired bit, the WPT circuit cuts off the write pulse within a short period (e.g. 300 ps), which has an extremely low possibility to switch the MTJ. This quick feedback may remove the pre-read sequence in MeRAM, reducing the latency.

Figure 4.13 presents a block diagram of the WPT circuit where the BL driver, voltage divider, and amplifier are connected to the BL. A number of cells are attached to between the BL and the SL. The WL driver, which is controlled by the WLC, is connected to the WL. Multiple cells are connected to

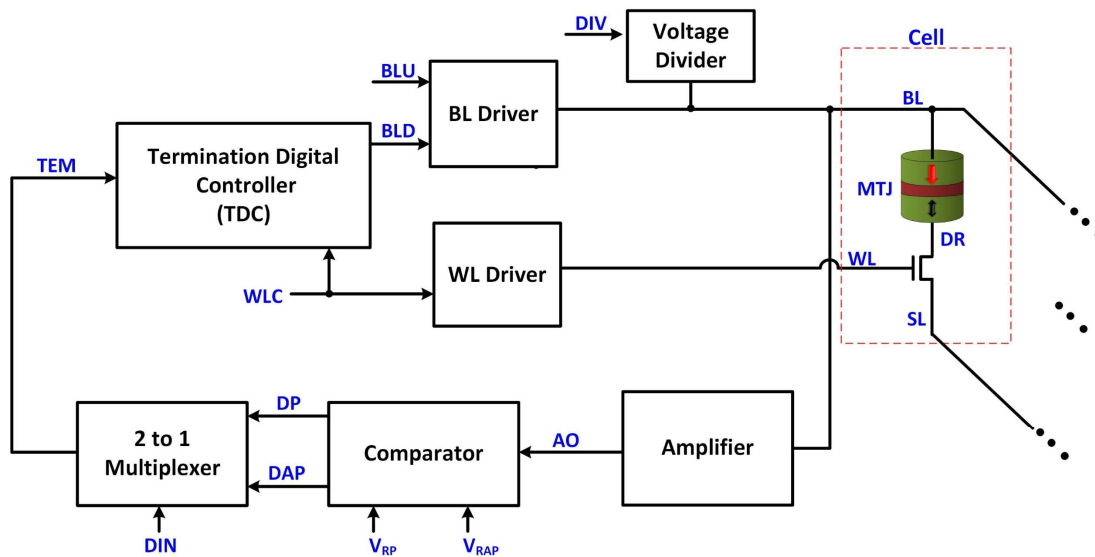


Fig. 4.13 Block diagram of the WPT circuit. The circuit modulates the write pulse width depending on individual MTJ's switching behavior by monitoring the resistance change.

the WL. The amplifier output, the AO, is connected to the input of the comparator. The multiplexer selects between two detection signals, the detection of parallel state (DP) and the detection of anti-parallel state (DAP), depending on the desired data on the DIN. The termination (TEM) is connected to the input of the termination digital controller that enables or disables the pull-down transistors of the BL driver.

The first idea of the proposed WPT scheme is related to the sequence of generating a write pulse. i) The WL initiates write pulses to the all selected MTJ devices. To be specific, the BLU enables the pull-up transistors of the BL driver, which makes the potential of the BL and the DR to be a write voltage simultaneously. Then, the WLC enables the pull-up transistors of the WL Driver, turning the

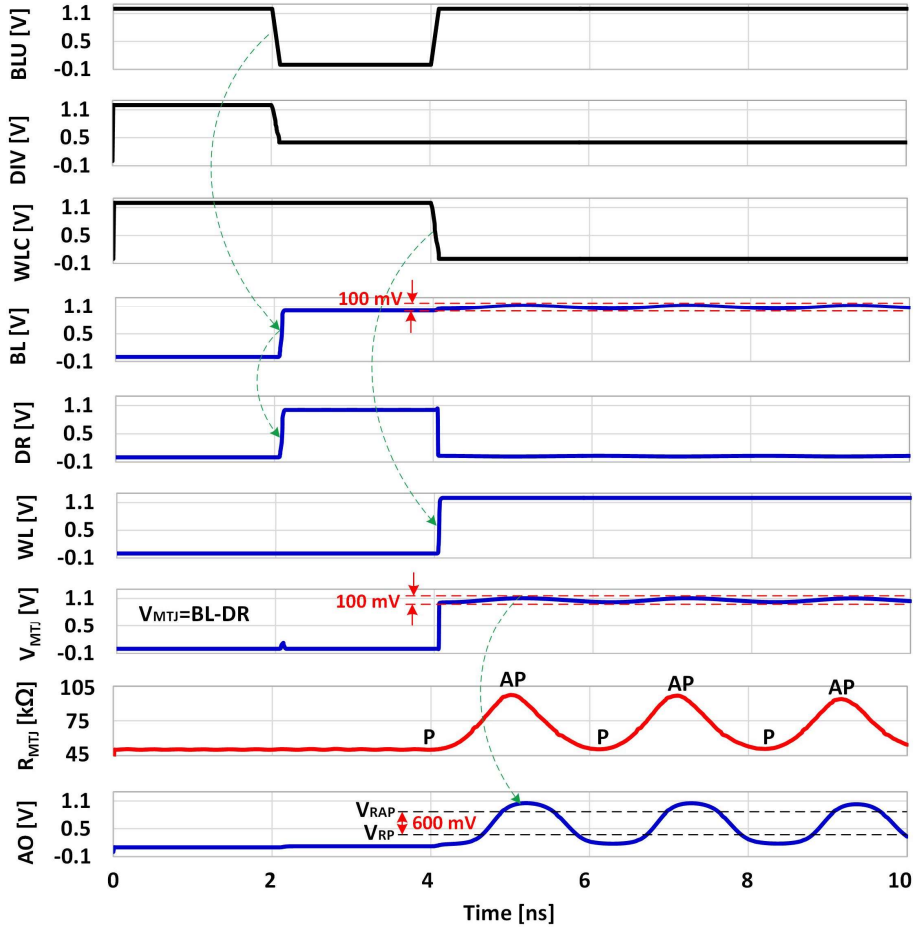


Fig. 4.14 Voltage divider converts the MTJ resistance fluctuations to the potential changes on the BL. These variations are enhanced by using the amplifier, generating a signal with the sufficient magnitude on the AO.

access transistor on, which create the potential difference across the MTJ. ii) The write pulse on each BL is individually removed once the MTJ switch to the desired state. If the TEM becomes '1', the BLD enables the pull-down transistors of the BL driver, discharging the BL potential to ground level.

Another idea is related to a function of the voltage divider, which allows the potential of the BL fluctuating correspondingly with the MTJ resistance changes under the write bias condition. If we continuously enable the BL driver, it maintains the stable BL potential regardless of the MTJ resistance fluctuation. This is because the BL driver typically has strong pull-up transistors. By contrast, the voltage divider has weak pull-up transistors, which causes the BL to respond to the MTJ resistance fluctuations. The operation sequence is as follows. Once the BL potential reaches the write voltage level, the BL driver is turned off while the voltage divider is turned on. Figure 4.14 shows the BL potential fluctuations with 100 mV amplitude corresponding to the MTJ resistance changes. The BL potential fluctuations are amplified to 800 mV peak-to-peak value via the amplifier. Since we disable the termination digital controller (TDC), the magnetization of the MTJ keeps precession, causing an oscillation. However, if the capacitive loading of the BL is larger than 50 fF, the signal delay between the BL and the AO might fail the WPT scheme.

4.4.3 Simulation and Analysis

The transistor-level schematic of the WPT circuit is shown in Fig. 4.15. The following steps are the detailed description of the write pulse termination (WPT) circuit operation for terminating a write pulse based on the condition that the initial state of the MTJ is P, and the desired state is AP as shown in Fig. 4.16(a). i) Pre-charge the BL to a write voltage level: the BLU and DIV enable the pull-up transistors of the BL driver and the voltage divider, respectively, increasing the potentials of both the BL and the DR up to the write voltage level. The BL driver is disabled before the WL becomes VDD. ii) Applying a write pulse across the MTJ: the WLC enables the pull-up transistors of the WL driver, allowing the WL potential to be VDD. When the access transistor is turned on, the DR becomes grounded (the SL

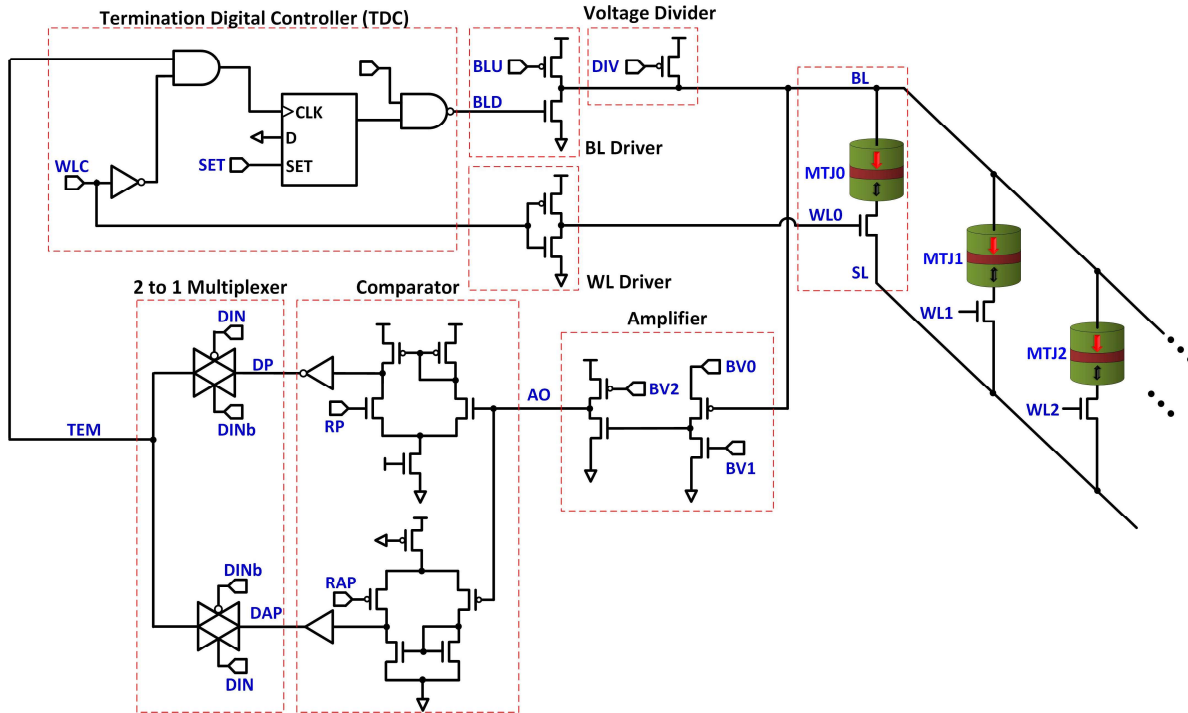


Fig. 4.15 Transistor-level schematic of the write pulse termination (WPT) circuit which generates a self-adjusted write pulse by monitoring the resistance change of the MTJ.

is initially grounded), generating a write pulse across the MTJ. iii) Amplification of the BL potential: the amplifier generates the AO by amplifying the small signal on the BL associated with MTJ resistance changes. iv) Detection of the MTJ states: the comparator simultaneously compares the potential of the AO with the two reference voltages, V_{RP} and V_{RAP} . If the potential of the AO is smaller than that of the V_{RP} , the DP maintains '1', indicating that the MTJ is P. On the other hand, if the potential of the AO is larger than that of the V_{RAP} , the DAP becomes '1' in which the MTJ completes to switch from P to AP. v) Selection of the write termination signal (TEM): Depending on the desired bit which is applied to the DIN, the multiplexer selects a signal between the DP and the DAP. In the case of the DIN is equal to '1' (desired bit is AP) as shown in Fig. 4.16(a), the multiplexer chooses the DAP as the TEM. vi) Termination of the write pulse: If the TEM become '1', and the WLC is equal to '0', the termination digital controller (TDC) makes the BLD to '1' for enabling the pull-down transistor of the BL driver, removing the write pulse across the MTJ.

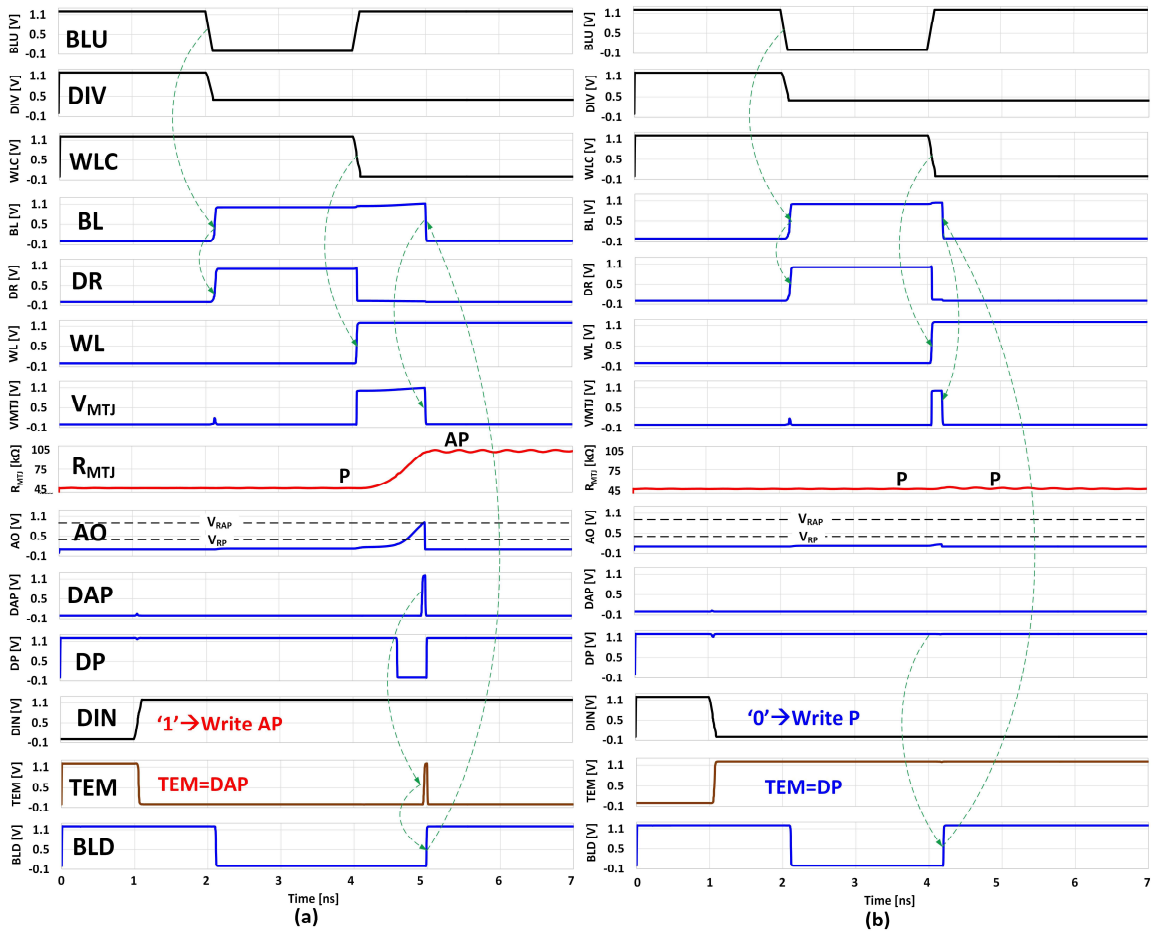


Fig. 4.16 Circuit simulations of the write pulse termination (WPT) scheme where the initial MTJ state is P, and desired state is (a) AP and (b) P.

Fig. 4.16(b) assumes that the initial state of the MTJ is P, and the desired state is also P. In this case, the switching should be prohibited. The circuit operation from step i) to iii) of Fig. 4.16(b) are the same as those of Fig. 4.16(a). However, at step iv) the comparator maintains the DP to '1' (the initial state is P) right after the write pulse is given to the MTJ because the AO is smaller than the V_{RP} . Since the desired bit is P, '0' is applied to the DIN, allowing the multiplexer selecting the DP as the TEM. Therefore, the write pulse can be terminated within 300 ps without destructing the stored data in the MTJ, allowing the MeRAM to eliminate the pre-reading sequence.

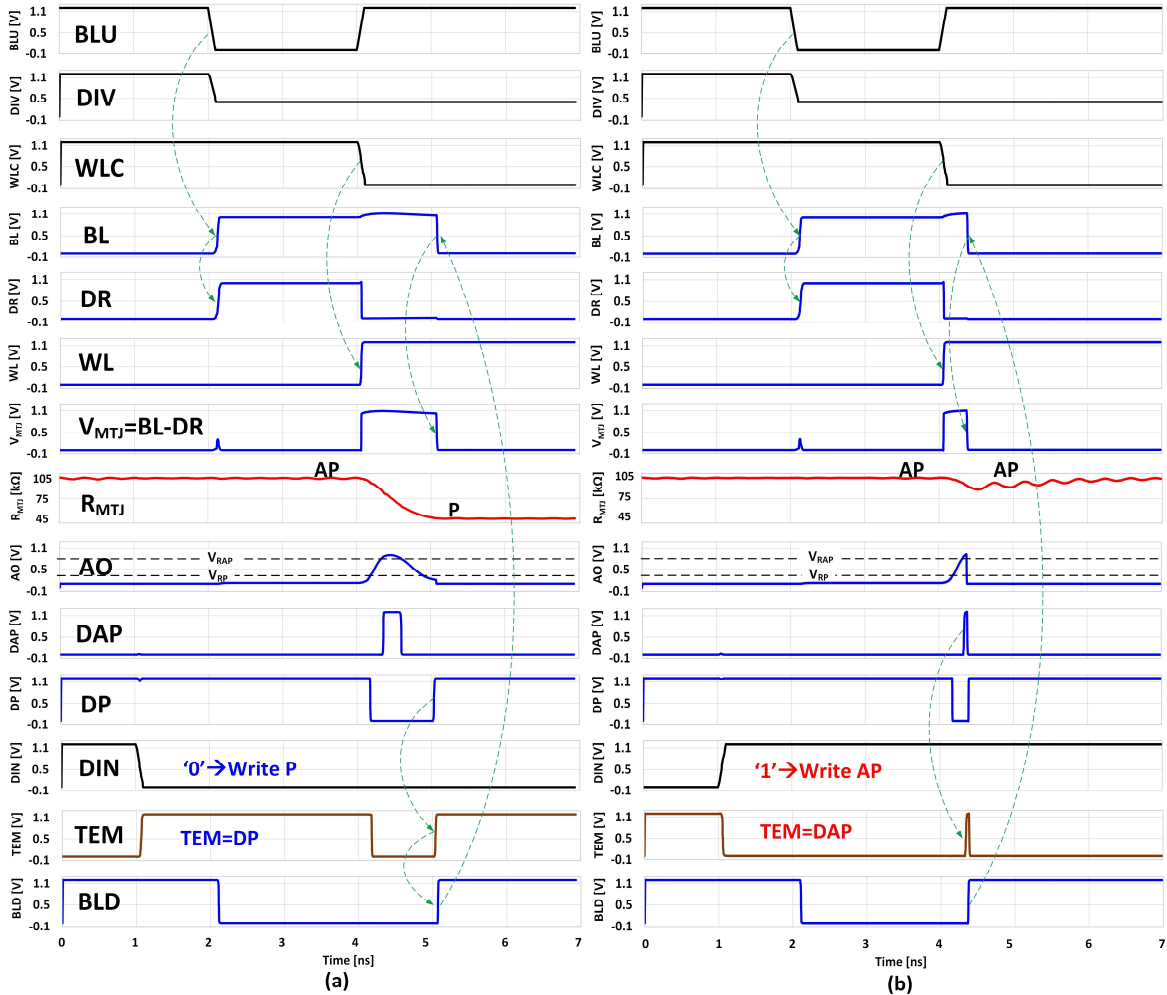


Fig. 4.17 Circuit simulations of the write pulse termination (WPT) scheme where the initial MTJ state is AP, and desired state is (a) P and (b) AP.

Fig. 4.17(a) assumes that the initial state of the MTJ is AP, and the desired state is P. On the contrary, Fig. 4.17(b) assumes that the initial state of the MTJ is AP, and the desired state is AP. The description is similar to that of Fig. 4.16.

4.5 Circuit for Controlling Non-deterministic Switching

4.5.1 Motivation

The VCMA-induced precessional switching is non-deterministic in which the state of the bit is always reversed regardless of its initial state for a given pulse duration. In the absence of the optimized

write termination circuit (WTC) that we described in the previous section, the core circuit needs to recognize the state of the selected cell before executing write operation in order to deal with the non-deterministic switching behavior. In this section, we introduce one of the core circuits, pre-read and write sense amplifier (PWSA).

4.5.2 Schematic and Simulation of the pre-read and write sense amplifier (PWSA)

Figure 4.18 shows a block diagram of the proposed PWSA. It is composed of the sense latch (S Latch), data latch (D Latch), XNOR logic gate, differential amplifier (Diff-amp), and write and pre-charge circuit. The circuit is designed to perform a read operation and to compare the current MTJ state to the incoming data, leading to a decision on whether a write pulse should be applied.

Figure 4.19 shows the data program flow, including the pre-read, comparison, write, read, and pass/fail check steps. The PWSA reads out the initial MTJ state and stores it in the S Latch during the

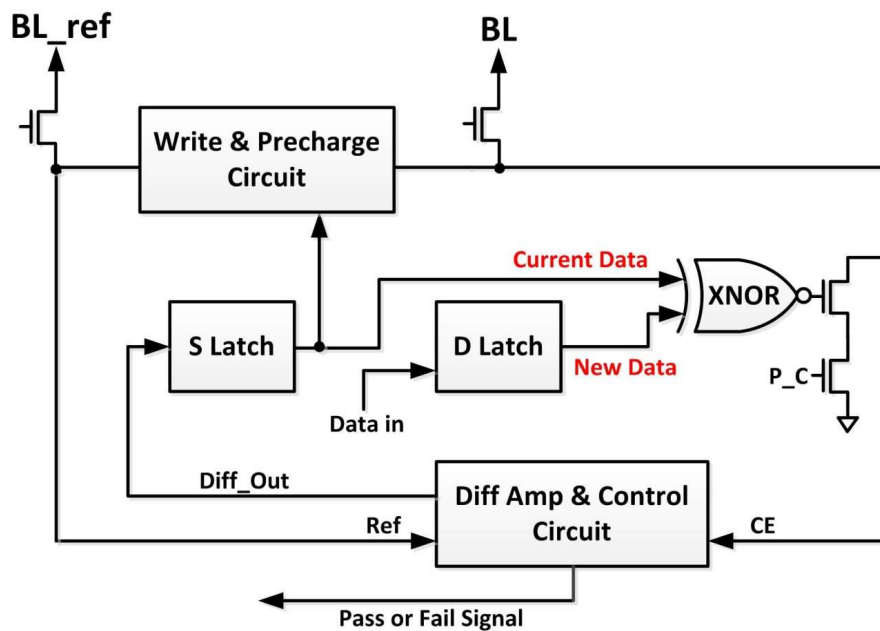


Fig. 4.18 Concept diagram of the proposed sense amplifier. The S Latch stores the MTJ state based on the voltage difference between the Ref node and the CE node, which is amplified by the Diff Amp. The D Latch stores external data which is transferred to the MTJ during the write operation if S Latch and D Latch data mismatch (XNOR=0). The write and pre-charge circuit provides write and pre-charge pulses to the BL during the write and read operations, respectively. When the circuit successfully completes a full sequence (data program), the control circuit generates a pass signal '1' to the external circuit.

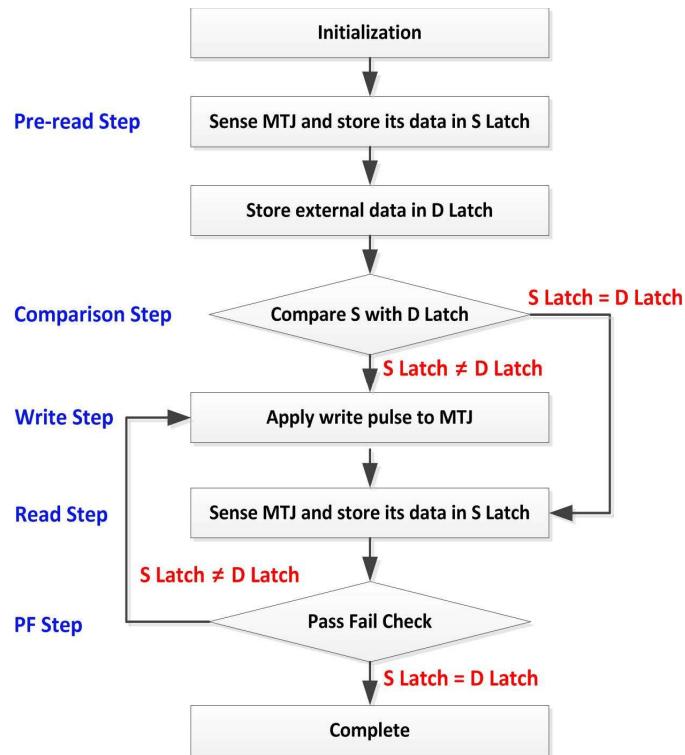


Fig. 4.19 Proposed data program (full sequence) flow of the PWSA. During the pre-read and comparison step, the circuit compares the initial MTJ state and external input data. A write pulse is generated based on the comparison result. At the Pass/Fail (PF) step, the PWSA creates a Pass signal ‘1’ when the newly programmed MTJ data is the same as the external data.

pre-read step and determines whether applying a write pulse to the MTJ based on the comparison result between the initial MTJ data and the external data. Therefore, redundant writes can be eliminated, leading to significant reduction of the total power consumption. Then, the MTJ state is compared again with the external data on the D Latch after applying the write pulse. If the MTJ is switched to the desired bit, the operation finishes, and a pass signal becomes ‘high’ which is transferred to the external circuit. Otherwise, the circuit iterates until the MTJ is switching to the desired state, or the number of iterations n reaches to the maximum value.

The schematic of the PWSA is shown in Fig. 4.20. The data program operation consists of five consecutive steps: pre-read, comparison, write, read, and pass/fail check steps. We assume that the AP represents logic value ‘1’, and P represents logic value ‘0’. We employed a 65 nm technology for circuit simulations with the MTJ compact model. The MTJ is assumed to have 100 k Ω resistance in

the parallel state and $200\text{ k}\Omega$ in the anti-parallel state, corresponding to the Tunneling Magnetoresistance (TMR) of 100%. The write pulse with a 1.2 V amplitude and a 1.2 ns width is applied to allow precessional switching. In this section, two simulation cases are discussed.

Figure 4.21 shows a simulation result in which the initial MTJ state is AP, and the external data is P. During the pre-read step, the S Latch stores '1' due to the AP state of the initial MTJ. For the comparison step, the D Latch becomes '0' as it receives the external data P. Mismatching between the S Latch data and the D Latch data generates a '0' on the XNOR node, turning the M6 off, which maintains the CE node in high potential. Since the potential of the CE node is higher than the Ref node, the output of the differential amplifier, Diff_Out, drops to '0', turning the M7 off. Thus, the S Latch remains in the '1' state due to the absence of a ground path, even though the M8 is turned on. Under

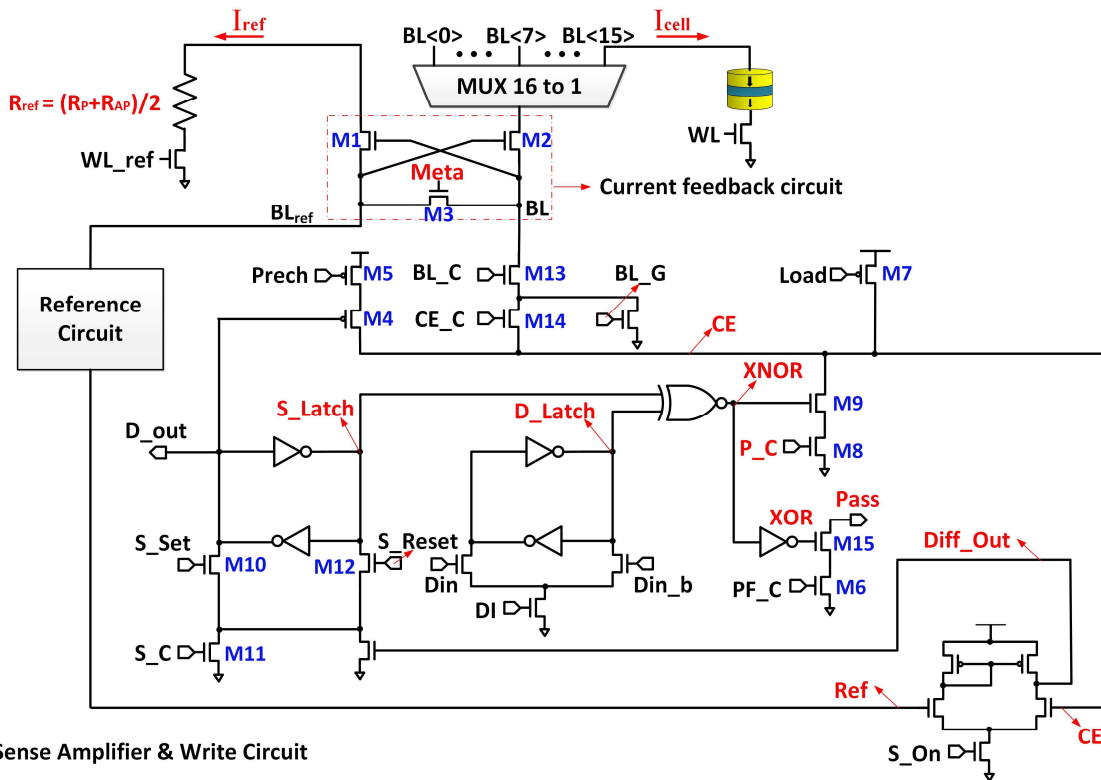


Fig. 4.20 Schematic of the proposed sense amplifier and write circuit (PWSA). The S Latch and D Latch store the initial MTJ state and the external data, respectively. During the read step, the differential amplifier amplifies the voltage difference between the Ref node and the CE node and creates a reliable logic value to the Diff_out node. The XNOR node holds the comparison value after evaluating the initial MTJ state versus the external data, determining whether a write pulse is generated during the write step. A current feedback circuit is used to increase the sensing margin.

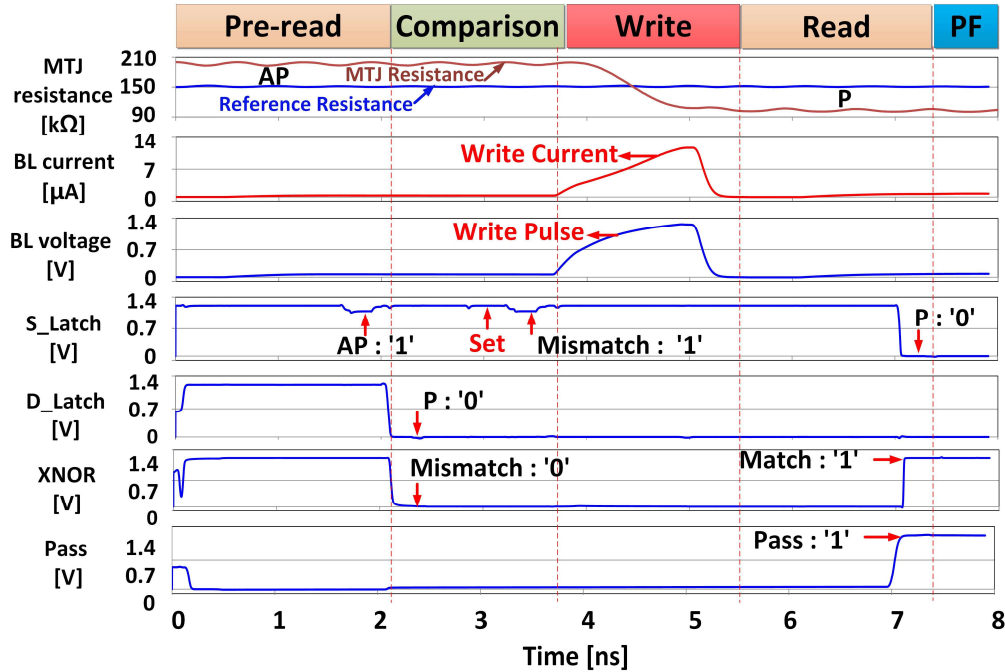


Fig. 4.21 Simulation of the PWSA operation in the case of an initial AP state for the MTJ, with P being the new external (input) data. Since the initial MTJ state (AP) is different from the external data (P), the output of the XNOR gate goes to ‘0’, which keeps the S Latch in a ‘1’ state. Hence, the PWSA provides the MTJ with a write pulse, then the MTJ state switches from AP to P during the write step. After the read step, the circuit compares the S Latch and the D Latch again for verification. In this case, the pass signal becomes ‘1’ since the MTJ is switched to the desired state.

this condition, the circuit can provide a 1.2V write pulse to the MTJ by using the transistors M4 and M5, and the MTJ state switches from AP to P during the Write step, which can be monitored by the resistance change. Next, the PWSA reads the MTJ state, and the S Latch changes to ‘0’, because the MTJ has switched from AP to P. At the final pass/fail check step, the ‘1’ pass signal is transferred to the external circuit because the MTJ is correctly programmed.

As the second scenario, we consider the case where switching does not require. Specifically, the initial MTJ state is P, and the external data is also P, causing the S Latch to be ‘0’ at the end of the comparison step, as shown in Fig. 4.22. If S Latch holds data ‘0’ (turning off M4), the circuit does not apply a write pulse to the MTJ, hence it remains in its initial P state. Next, the PWSA reads out the MTJ resistance again, and the pass signal becomes ‘1’ due to the matched data between the S Latch and the D Latch during the final pass/fail check step.

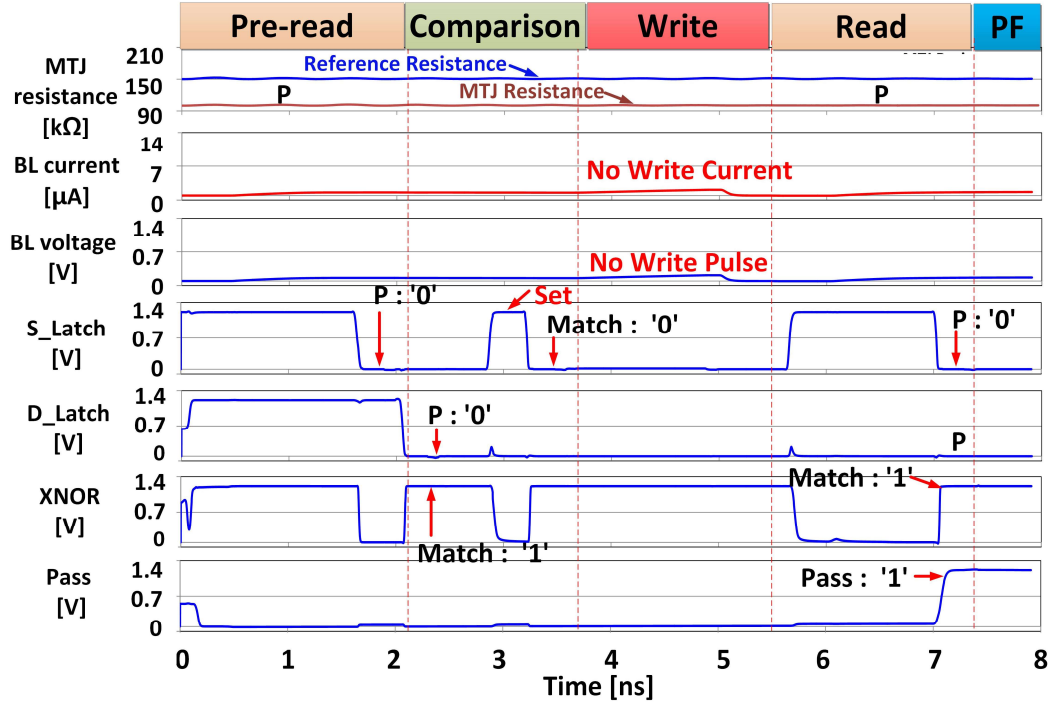


Fig. 4.22 Simulation of the PWSA operation in the case of an initial P state for the MTJ, with P being the external (input) data. Since the initial MTJ state (P) is the same as the external data (P), the output of the XNOR goes to '1', which switches the S Latch to '0'. Because of having a '0' state in the S Latch, the PWSA does not apply a write pulse to the MTJ.

4.5.3 Performance Evaluation

The sensing margin is affected not only by MTJ characteristics such as the TMR ratio but also by circuit design parameters such as the size and the bias conditions of the sense amplifier circuit. Previous works have reported sensing circuits with a 1T-1MTJ topology using MTJs with 200% TMR and 65nm technology, achieving 0.18V sensing margin [129]. Here we target to increase the sensing margin to ensure that the differential amplifier can generate a reliable output signal to control logics.

In the memory architecture, sensing margin and read disturbance are sensitive to the bias voltage applied to the bit line (BL). Applying higher voltage to the BL during the read operation generates larger read disturbance, causing reliability issues. On the other hand, applying a low voltage to the BL results in a decreased sensing margin. The sensing margin in the circuit of Fig. 4.20 is determined by the voltage difference between the Ref node and the CE node. To maximize the sensing margin and

minimize the read disturbance, we propose a current feedback circuit which consists of transistors M1, M2 and M3 as shown in Fig. 4.20, which is based on a current conveyor circuit [130].

The reference resistor R_{ref} is connected to M1, and its resistance is $(R_P + R_{AP})/2$ for a centered sensing margin of AP and P. Such reference resistor could be implemented, for example, via a serial and parallel combination of MTJs [$R_P + (R_P \parallel R_P)$] or a digitally tunable CMOS based resistor circuit.

The operation of the current feedback circuit is as follows: a read operation is made up of the pre-charge stage, BL discharge stage through an MTJ, and latch stage. During the pre-charge stage, the sense amplifier charges up both BL_{ref} and BL_{cell} to the same potential level because M3 is fully turned on. Once M3 is turned off, the BL discharge stage begins. If the MTJ has a P state, I_{cell} would be larger than I_{ref} , causing BL_{cell} to have a lower potential than that of BL_{ref} . The decreased potential of BL_{cell} slightly turns off M1, which reduces I_{ref} further and leads BL_{ref} to discharge slowly. Thus, the circuit is able to have a much larger potential difference between the Ref node and the CE node. Through circuit simulations, the average sensing margin 360 mV with 100% TMR was achieved, which is 2x larger compared to that of conventional sense amplifiers using 200% TMR. As a result, the improved sensing margin through the proposed current feedback circuit guarantees a stable logic swing as observed in Fig. 4.23.

To evaluate the speed of the proposed circuit, we constructed the RC model on the BL based on the value of sheet resistance and metal capacitance of the considered 65nm technology. The circuit achieved 2 ns read operation time as shown in Fig. 4.23. However, the write time is determined by the switching characteristics of the MTJ. If the intrinsic precessional switching time of MTJs is around 1 ns, it is possible to achieve 1.8 ns write time, accounting for both the peripheral circuit delay as well as the RC delay of the BL. Since the pass/fail check step is based on digital circuit operation, it takes only 0.5 ns to generate a Pass/Fail signal and does not cause a major penalty in terms of speed.

The large resistance of the MTJ devices assures small write and read current, reducing dissipated power without impact on data programming speed, as shown previously. The average power

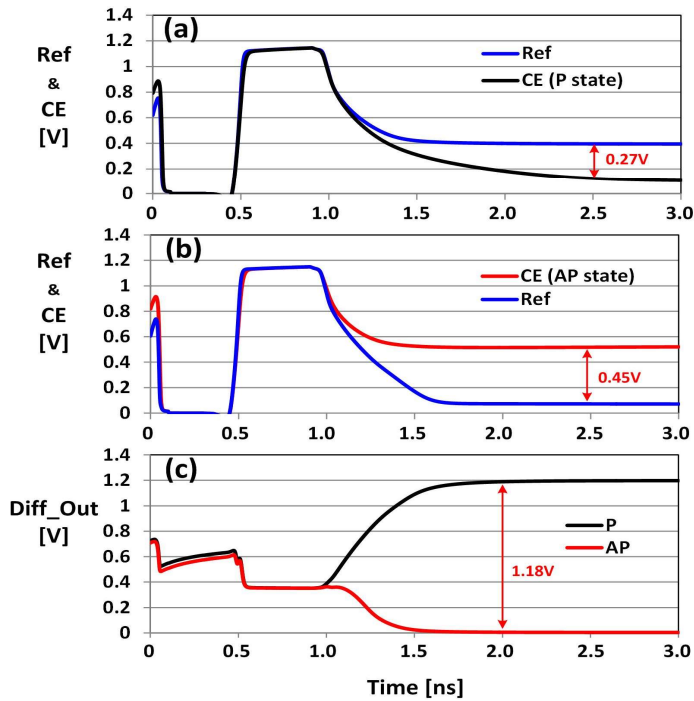


Fig. 4.23 (a) Simulations for the sensing margin of the P and reference resistance. Sensing margin is defined as the voltage difference between the CE node and the Ref node in which the sensing margin is 0.28 V. (b) Sensing margin of the AP and reference resistance where the sensing margin is 0.44 V. (c) The average sensing margin of 0.36 V is sufficient to generate ‘0’ or ‘1’ logic states at the output of the differential amplifier, while the overall read time is estimated to be ~ 2 ns at a given capacitive loading.

consumption of the PWSA is $18 \mu\text{W}$ (excluding a write power to BL), which is 13% higher than the power required to write a single VCMA-driven MTJ. However, if the PWSA controls a cell array, the write power is dramatically increased due to RC loading of the BL and a large number of unselected MTJs. Therefore, the total power consumption of the PWSA is significantly reduced by eliminating the redundant write pulses (and its required sequence), if there is a match between old and new data in the MTJs. The most frequently occurring matching probability is 50% as expected, translating into an additional 50% saving in write power consumption under random data pattern conditions.

CHAPTER 5

4KBIT MeRAM MACRO DESIGN

5.1 Specification

A synchronous 4 Kbit magnetoelectric random-access memory (MeRAM) macro has been designed based on IBM 130 nm RF-DM technology. The macro is divided into three major parts: core circuit, peripheral circuit, and IO circuit as shown in Fig. 5.1. The core circuit consists of the 1T-1MTJ and

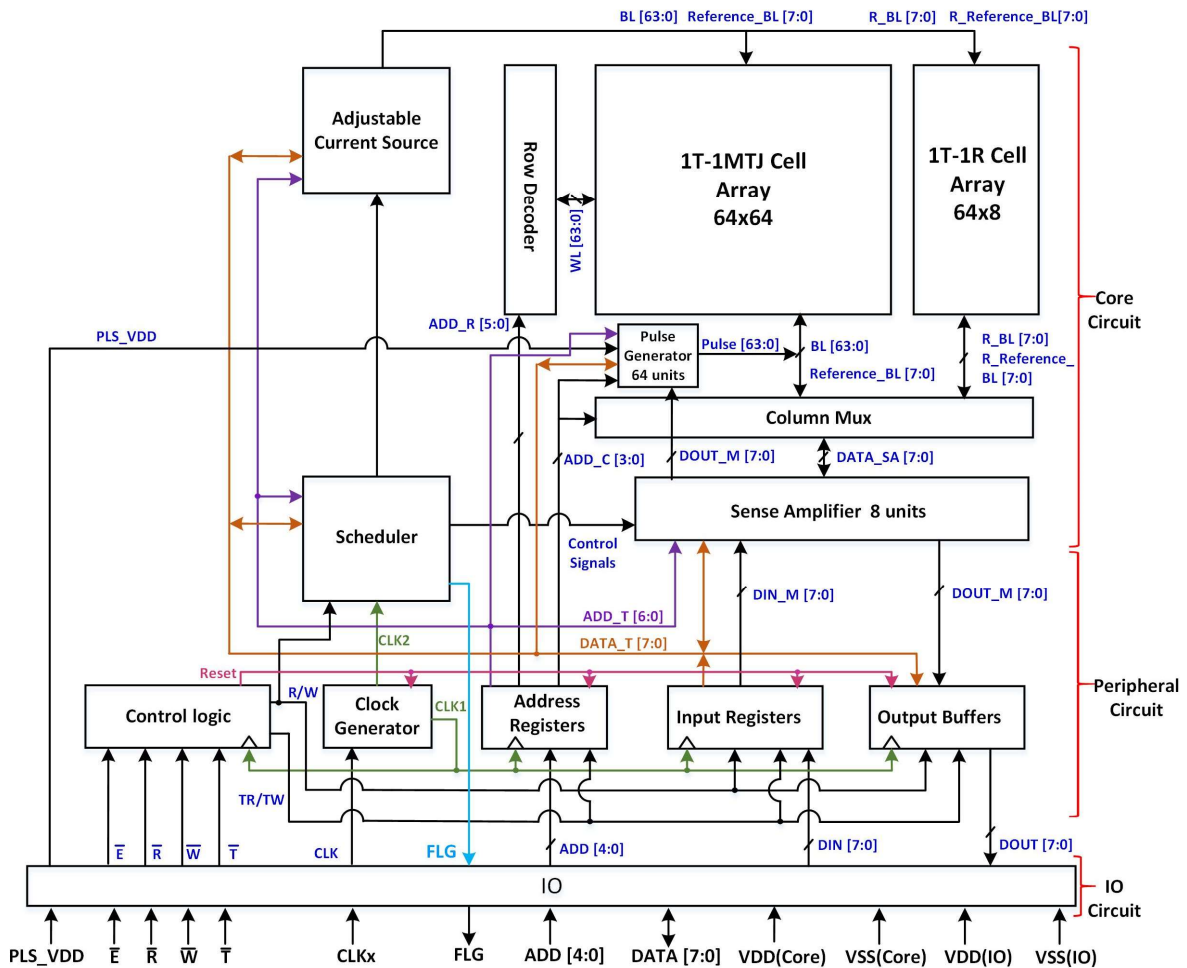


Fig. 5.1 MeRAM macro architecture consisting of three major parts: core, peripheral, and IO circuits. The core circuit selects eight memory cells at a given address by using the row decoder and the column mux. Selected memory cells are sensed by the sense amplifiers. The peripheral circuit generates signals to control the core circuit and temporarily stores the commands, input and output data, address in internal registers. The IO circuits is an interface between the external chips and the internal MeRAM circuitry.

Parameters	Value
Process Design Kit (PDK)	130 nm IBM RF-DM
Macro Size	1.25 mm × 1.25 mm
1T-1MTJ Cell Array Capacity	4 Kbit = 64 WLs × 64 BLs
1T-1R Cell Array Capacity	512 bit = 64 WLs × 8 BLs
Clock Speed	< 100 MHz
Supply Power	1.2 V for Core, 2.5 V for IOs
Supply Current	Typical = 14 mA, Max = 20 mA

Table 5.1 Specification of 4Kbit MeRAM Macro.

1T-1R cell arrays, row decoder, column mux, pulse generators, and sense amplifiers. The 1T-1MTJ cell array is composed of 64 word lines (WLs) and 64 bit lines (BLs), and there is a cell at each cross point of the BLs and WLs. The WLs and BLs are selected by the row decoder and the column mux, respectively. To be specific, if a 10-bit address (6 bit for the row, 4 bit for the column) is given during the read or program operation, the row decoder selects one of the WLs, and the column mux selects eight of the BLs. The sense amplifiers electrically sense these selected memory cells, and their states are converted to digital logic signals ‘0’ or ‘1’ depending on the MTJs state (R_P, R_{AP}). The pulse generators apply write pulses to the BLs if the stored data mismatch the desired data. The amplitude and duration of the write pulse can be externally modulated via the test mode operations, which allows the macro to find an optimum write pulse for achieving the low WER. The peripheral circuit is a digital circuit that generates signals to control the core circuit associated with the applied clock and temporarily holds external signals (i.e. command, address, data, and analog setting code) in the dedicated internal registers. The IO circuit is the interface between an external controller and the MeRAM macro. To protect internal circuits of the macro, the IO circuit involves an electrostatic discharge (ESD) protection circuitry.

The macro is equipped with the chip enable (\bar{E}), read enable (\bar{R}), write enable (\bar{W}), and test mode enable (\bar{T}), allowing the macro to perform in two normal operations: read and program; and two test mode operations: register configuration and register readout. Especially, the program operation includes the pre-read stage followed by the writing stage to manage the non-deterministic switching

IO Pins	Description
\bar{E}	Chip enable
\bar{T}	Test mode enable
\bar{R}	Read operation enable
\bar{W}	Program operation enable
CLK	Clock
ADD [4:0]	Address (two cycles are required for 10-bit address)
DATA [7:0]	Data in and out
PLS_VDD	Analog input for the amplitude of write pulse
FLG	Operation completion flag signal
VDD (Core)	Power supply for the core and peripheral circuits
VSS (Core)	Power ground for the core and peripheral circuits
VDD (IO)	Power supply for the IO
VSS (IO)	Power ground for the IO

Table 5.2 IO pin list and descriptions of the MeRAM macro.

behavior of the voltage-controlled MTJ. If the test mode is enabled, a user can access the internal registers and modulate analog settings (i.e. reference current level, pulse width, etc.) by changing the state of the registers. Since all these operations are based on 8-bit data, the macro has the bidirectional 8-bit data IO pins (DATA [7:0]). In terms of speed, the MeRAM macro can be operated up to 100 MHz clock, and each operation requires the different number of cycles. The specification summary and the IO pin list of the macro are shown in Table 5.1 and Table 5.2, respectively.

5.2 Cell and Array Design

A memory cell consists of one transistor and one MTJ (1T-1MTJ) structure where the free layer (bottom layer) of the MTJ is connected to the drain of the access transistor, and the pinned layer (top layer) is connected to the BL as shown in Fig. 5.2(a). The resistance of the MTJ should be at least 10x larger than that of the access transistor (on state) so that the state of MTJs mainly governs the sensing current during the read operation. Also, a relatively higher resistance of MTJ causes a significant portion of the voltage drop across the MTJ compared to that of the access transistor, effectively enhancing the amplitude of the applied voltage across the devices and reducing the WER further.

The IBM 130 nm RF-DM process has the eight metal layers. Hence the top metal layer (i.e. MA layer) was used for the bottom electrode of MTJs as shown in Fig. 5.2(b). As shown in Fig. 5.2(c), the

size of the unit cell is relatively large ($12.48 \mu\text{m} \times 9.24 \mu\text{m}$) due to the design rule of the top three metal layers which are originally aiming for RF circuit designs.

The macro was designed to control the four different diameters of MTJs on the same cell array: 100 nm, 400 nm, $1 \mu\text{m}$, and $2 \mu\text{m}$. To achieve this, the main cell array is divided into four zones: Zone0 (WL0~15), Zone1 (WL16~31), Zone2 (WL32~47), and Zone3 (WL48~63). Although this type of cell array provides the ability to analyze the characteristics of MTJ as a function of its dimension, it requires more circuit design effort.

In addition to the implementation of the 1T-1MTJ cell array, the macro has another cell array whose unit cell is made of one transistor and one polysilicon based resistor (1T-1R). The resistor is fabricated through front-end-of-line (FEOL) so that the implemented circuits are tested without MTJs fabrication. Especially, we predefined the value of resistors based on the row address and implemented them as the hard coded 1T-1R cell array in advance of FEOL fabrication, allowing us to verify the MeRAM macro by simply monitoring the readout data from the hard coded 1T-1R cell array.

Fig. 5.3 shows the layout of the $64 \text{ WLs} \times 64 \text{ BLs}$ 1T-1MTJ cell array and the $64 \text{ WLs} \times 8 \text{ BLs}$ 1T-1R cell array. In the 1T-1MTJ cell array, there is a reference BL at an interval of eight BLs to generate

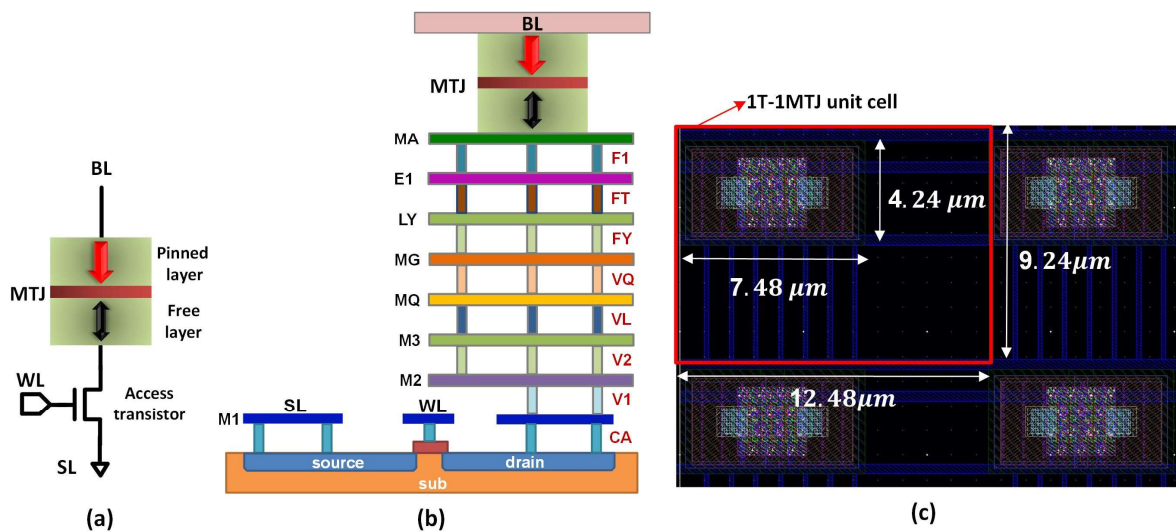


Fig. 5.2 (a) One transistor and one MTJ (1T-1MTJ) structure based unit cell for MeRAM. (b) Cross section view of the unit cell. (c) Layout design of the unit cell. Due to the design rule of the top metal (minimum space between two the metal is $5 \mu\text{m}$), the unit cell dimension is $12.48 \mu\text{m} \times 9.24 \mu\text{m}$.

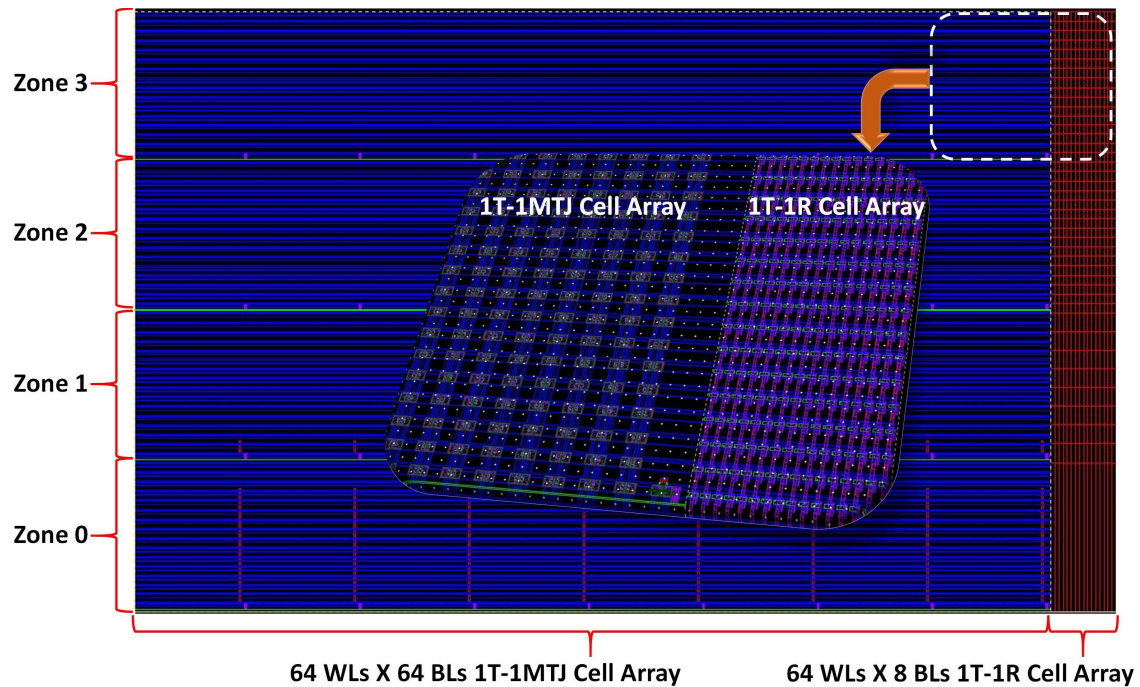


Fig. 5.3 Layout design of the 64 WLs \times 64 BLs 1T-MTJ cell array and the 64 WLs \times 8 BLs 1T-1R cell array. The unit cell of the 1T-1R array has a poly-silicon based resistor, and its BL is designed by using the metal layer 2 (M2). The cells in the 1T-1R array can be completely fabricated via the front-end-of-line process (FEOL). Then, the macro can be verified without MTJs fabrication.

a reference voltage for the sense amplifier. For the same purpose, there are also eight reference BLs in the 1T-1R cell array. The reference resistors are also designed based on the polysilicon by taking into account of the resistance of the MTJs in each zone.

5.3 Core Circuit Design

The core circuit plays a significant role in selecting memory cells and providing proper bias conditions to the WLs, BLs, and SLs to achieve sensing and writing. Figures 5.4 shows a schematic of the core circuit where the pulse generator is directly connected to the BL, and the sense amplifier is connected the BL via the column mux. The input and output ports of the core circuit are listed in Table 5.3.

Since the WER is sensitive to the applied write pulse shape (duration, amplitude, and slew rate), we designed the dedicated pulse generators which can modulate the pulse duration and amplitude by using

Type	Category	Port Name	Description
Input	Address	ADD_C [3:0]	Column Address
		ADD_R [5:0]	Row Address
Pulse Generator Control Signals	Pulse Generator Control Signals	PLS_VDD	Write pulse amplitude (0.8 V ~ 1.5V)
		PLS_CLK	Clock for generating a write pulse
		PLS_Code [10:0]	Pulse width control code
		PLS_Enable [63:0]	Pulse generator enable signals
		SA_Precharge	Precharge bit lines (BLs) up to reference voltage
Sense Amplifier Control Signals	Sense Amplifier Control Signals	SA_Current	Initiate the BL development
		SA_Enable	Comparator Enable (Data latch)
		SA_Zone [3:0]	Change the bias condition of the current source
		SA_Code [3:0]	Modulate the current level for the reference BL
		DOUT_M [7:0]	Output of the sense amplifier
Output	Data Out	DOUT_B_M [7:0]	Complementary output of the sense amplifier

Table 5.3 Core circuit input and output ports and descriptions

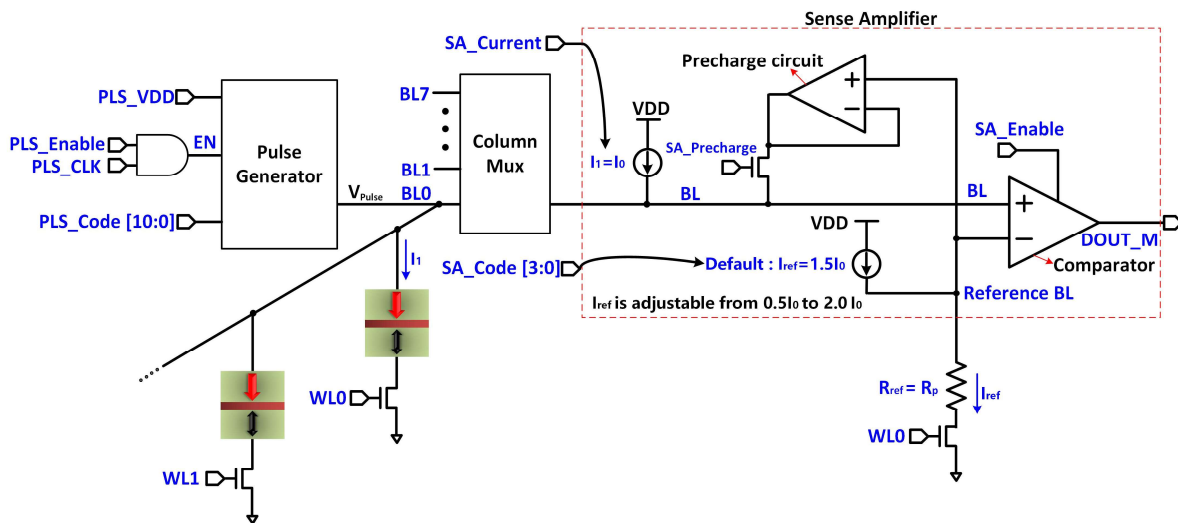


Fig. 5.4 Schematic of the core circuit consisting of the pulse generator, sense amplifier, column mux, and row decoder. To minimize the deterioration of write pulse shape, a pulse generator is directly connected to each BL. For the fast precharge, we designed a dedicated precharge circuit which is able to charge up 763 fF of a BL capacitance to the reference level within 3 ns. The reference current is adjustable for better sensing reliability.

a digital code and external analog input, respectively. Specifically, the pulse duration is adjusted ranging from 0.5 ns to 1.5 ns with 100 ps resolution depending on the configuration of the 11-bit digital code, and the pulse amplitude is adjusted ranging from 0.8 V to 1.5 V depending on the magnitude of the external voltage source.

Figure 5.5 shows a schematic of the pulse generator including the code-controlled delay circuit and its simulation result. The key idea of this circuit is that the EN decides the rising timing of a write pulse

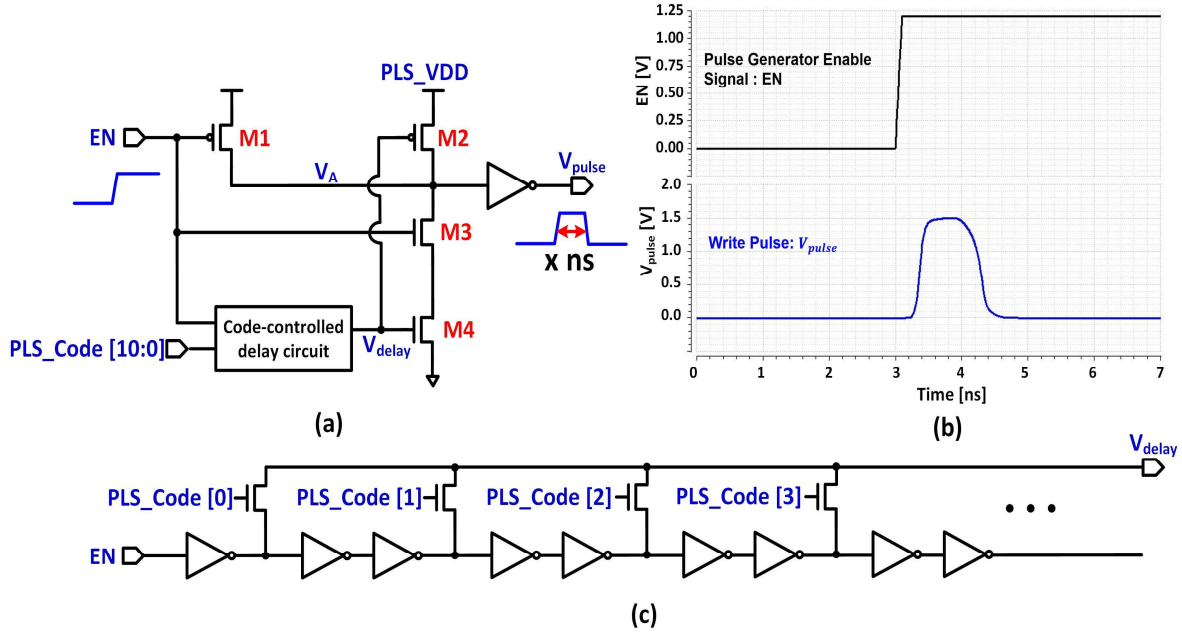


Fig. 5.5 Schematic and simulation of the pulse generator including the code-controlled delay circuit. The pulse width is modulated by the configuration of CODE [10:0] (a) pulse generator (b) circuit simulation (c) code-controlled delay circuit.

(V_{pulse}), and the code-controlled delay circuit decides the falling timing. The detail description is as follows. Initially, the EN is grounded, which turns M1 on and M3 off, making V_{pulse} to be grounded. If the EN becomes '1', M1 and M3 are turned off and on, respectively. Then, V_A is discharged via M3 and M4, becoming V_{pulse} '1'. After a certain amount of time depending on the configuration of PLS_Code [10:0], V_{delay} disconnects the ground path and turns on M2, becoming V_{pulse} '0'.

The sense amplifier plays a role in converting the resistive state of MTJ into a digitized signal during the sensing which consists of three stages: precharge, develop, and latch. Initially, the circuit generates a reference voltage via the adjustable current source and the reference resistance (R_{ref}). We implemented the polysilicon based reference resistor whose magnitude is equal to the low resistance state of MTJ ($R_{ref} = R_P$). Since the reference current (I_{ref}) can be controllable, ranging from $0.5I_0$ to $2.0I_0$ with $0.1I_0$ resolution, we can determine the voltage of the reference BL (V_{ref}) without changing the magnitude of the reference resistor. As a default value, I_{ref} and R_{ref} are equal to $1.5I_0$ and R_P ,

respectively, resulting in $V_{ref} = 1.5I_0R_P$, which is lower than the sensed voltage of AP state ($V_{BL} = I_0R_{AP}$) and higher than that of P state ($V_{BL} = I_0R_P$) as long as the MTJ has a sufficient TMR.

During the precharge stage, the precharge circuit pulls up the BL until it reaches the reference voltages. The reason for implementing the dedicated precharge circuit is that the capacitive load of the BL is large (~ 763 fF) due to the big cell size originating from the design rule of the technology. For the develop stage, the fixed current (I_0) flows into the selected MTJ, generating a certain voltage level (I_0R_P or I_0R_{AP}) depending on the state of the MTJ. During the latch stage, the comparator generates a digital signal by comparing the voltage difference between the BL and the reference BL. As a result, the eight sense amplifiers perform the read operation and transfer the 8-bit sensed data to DOUT_M[7:0] simultaneously. Figure 5.6 shows the simulation result of the 8-bit based operation where we obtained the expected sensing results: $R_P \rightarrow '0'$, $R_{AP} \rightarrow '1'$.

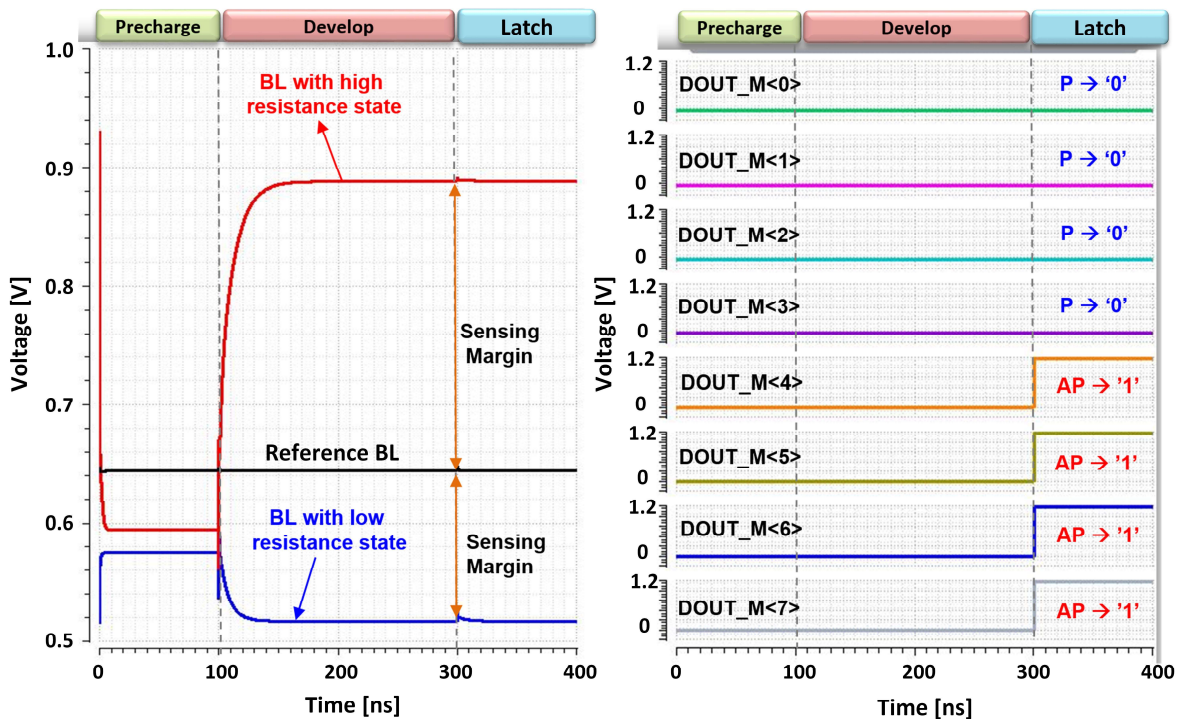


Fig. 5.6 Simulation of the sense amplifier for the sensing. The voltage of the reference BL can be adjusted by the digitally controlled current source. The states of the selected MTJs are converted to the digital signals at the DOUT_M nodes by comparing the voltages difference between the BL and the reference BL via the comparator.

5.4 Peripheral Circuit Design

The peripheral circuit is designed by using digital components from the standard cell libraries offered by ARM. We used the register transfer level (RTL) synthesis and physical synthesis via Cadence Encounter to reduce design time and errors. The synthesized schematic and physical layout of the peripheral circuit is shown in Fig. 5.7 where it consists of 315 standard cells, and the total area is 112^2 um^2 . The input and output ports are listed in Table 5.4 and Table 5.5, respectively. However, we skip the detail design methodologies and descriptions for the sub-blocks of the circuit.

The major functions of the peripheral circuit are as follows. First, the scheduler (state-machine) in the peripheral circuit generates different timing chains depending on the configuration of the command signals (i.e. \bar{E} , \bar{T} , \bar{R} , \bar{W}). Second, the circuit provides core control signals corresponding to each stage associated with a clock so that the circuit components of the core circuit successfully perform pre-defined functions. For example of the sense amplifier, it is necessary to apply control signals such as SA_Precharge, SA_Current, and SA_Enable at the right timing with a correct digital logic value to

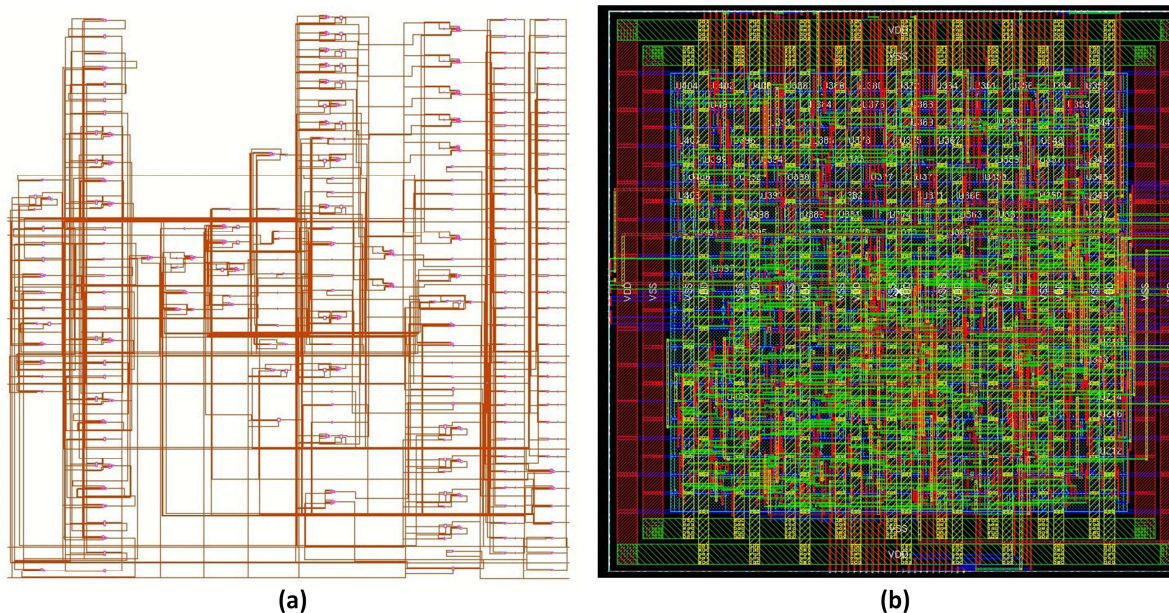


Fig. 5.7 (a) Schematic design of the peripheral circuit via RTL synthesis based on ARM's digital standard cell libraries. The circuit was initially described by a high-level Verilog code, which is converted to a gate level Verilog code by using the Design Compiler. (b) A layout design of the peripheral circuit. The gate level Verilog code is transferred to the physical dimension through physical synthesis.

Type	Category	Port Name	Description
Input	From IO Pins	\bar{E}	Chip enable
		\bar{T}	Test mode enable
		\bar{R}	Read operation enable
		\bar{W}	Write operation enable
		CLKx	Clock
		ADD [4:0]	Address (two cycles are required for 10 bit)
	DATA [7:0]	Data in and out	
From core circuit	DOUT_M [7:0]	Output of the sense amplifier (sensed data)	

Table 5.4 Peripheral circuit input ports and descriptions.

Type	Category	Port Name	Estimated fan-out				
			Width [μm]	Length [μm]	# of Tr.	# of unit	Total Capacitance
Output	To core circuit	ADD_C [3:0]	0.32	0.12	1	8	213 fF
			0.16	0.12	1	8	
			1.00	0.12	1	16	
			0.50	0.12	1	16	
			5.00	0.12	1	24	
		ADD_R [5:0]	0.32	0.12	1	32	289 fF
			0.16	0.12	1	32	
			1.73	0.12	1	32	
		PLS_CLK	0.4	0.75	1	64	2.25 pF
			0.2	0.75	1	64	
			0.4	0.12	25	64	
			2	0.12	1	64	
		PLS_Code [10:0]	0.4	0.12	1	64	48.2 fF
		PLS_Enable [63:0]	2	0.112	30	64	7.22 pF
	SA_Code [3:0]	2	0.12	5	8	151 fF	
	SA_Current	4	0.12	5	8	301 fF	
	SA_Enable	2	0.12	10	8	301 fF	
	SA_Precharge	2	0.12	10	8	435 fF	
	SA_Zone [3:0]	2	0.12	5	8	1.15 pF	
To IO pins	IE (input enable)	Standard IO cell				30 fF	
	OE (output enable)	Standard IO cell				30 fF	
	DOUT [7:0]	Standard IO cell				30 fF	
	FLG	Standard IO cell				30 fF	

Table 5.5 Peripheral circuit output ports and fan-out on each port. Each unit involves the number of transistors described on the table.

complete the read operation properly. Third, the internal registers of the peripheral circuit temporarily store new input data, output data, address, and several digital codes for adjusting the pulse width and the reference current level. The internal registers related to the analog setting can be accessed via the test mode operations such as register configuration and register readout.

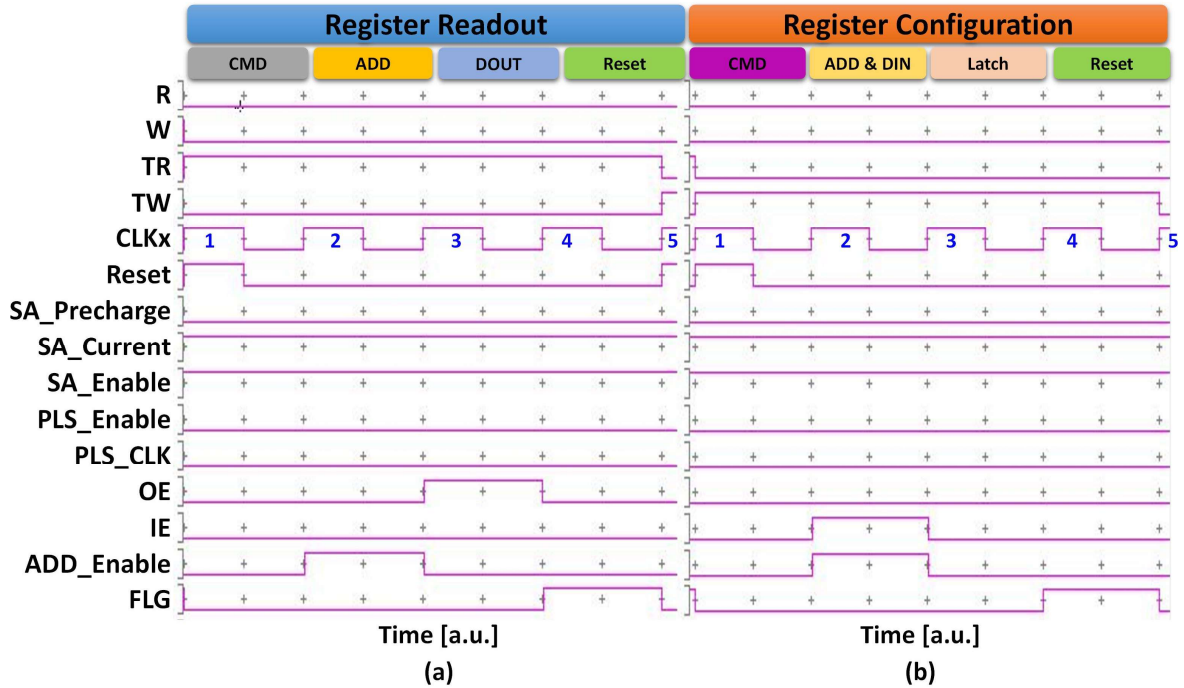


Fig. 5.8 Verilog simulation of the test mode operations (register readout and register configuration) of the peripheral circuit via ModelSim. A user can access the internal registers and change the logic configurations of them during these operations.

For the first test mode, register readout operation, if the macro receives the command ($\bar{E} = 0$, $\bar{T} = 0$, $\bar{R} = 0$, $\bar{W} = 1$) during the CMD stage, the TR becomes enabled as shown in Fig. 5.8(a). Once the address comes in during the ADD stage, the targeted registers are electrically connected to the output buffers. For the DOUT stage, as the output enable (OE) becomes ‘1’, the data stored in the registers are transmitted to the DATA [7:0] of the IO pins. Finally, the macro returns the FLG signal ‘1’, informing that the operation is successfully completed. During this test mode operation, the core control signals are not activated.

To execute the second test mode operation, register configuration, the combination of the command ($\bar{E} = 0$, $\bar{T} = 0$, $\bar{R} = 1$, $\bar{W} = 0$) needs to be received during the CMD stage, enabling the TW as shown in Fig. 5.8. At the next stage, the macro receives an address of the targeted registers and an input data by enabling the ADD_Enable and the IE, respectively. Then, the input data is stored in the targeted registers during the Latch stage, and the macro returns the FLG signal ‘1’.

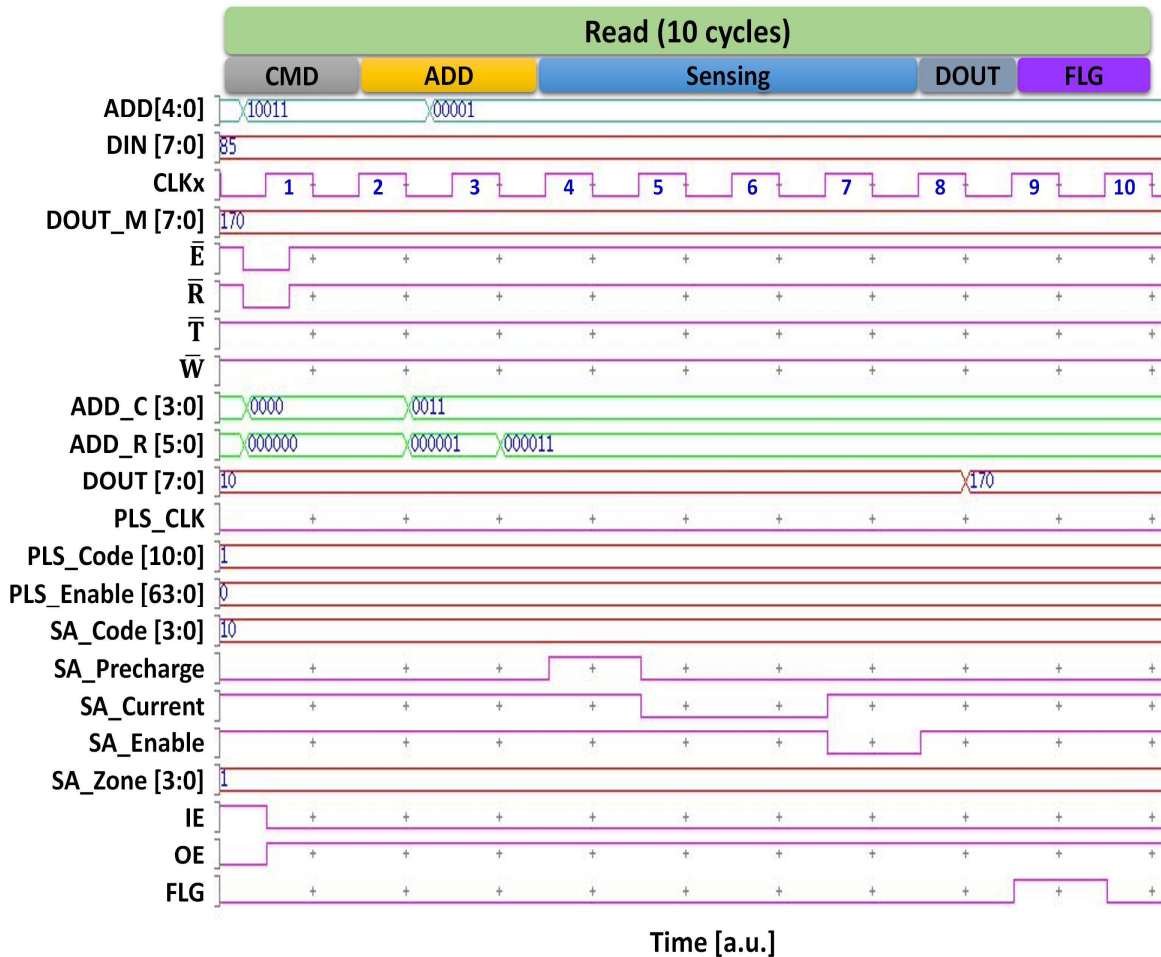


Fig. 5.9 Verilog simulation of the peripheral circuit read operation via ModelSim. The circuit decodes the command and generates the core control signals so that the sense amplifiers can sense the state of selected MTJs. After reading out the sensed data, the circuit returns the FLG signal ‘1’ to one of the IO pins, indicating that the operation is completed.

The read operation requires 10 clock cycles, which consists of five stages: CMD, ADD, Sensing, DOUT, and FLG as shown in Fig. 5.9. To perform the read operation, the configuration of the command should be $\bar{E} = 0$, $\bar{T} = 1$, $\bar{R} = 0$, $\bar{W} = 1$, which are received at the CMD stage. Since the number of IO pins for the address is five, the macro needs to receive 10-bit address via two clock cycles and store the address in the address registers at the ADD stage. Then, the peripheral circuit generates the core control signals, and the sense amplifiers store the state of the selected MTJs at the sensing stage. In the next DOUT stage, the sensed data is transmitted from the output of the sense amplifier (DOUT_M [7:0]) to the output buffer (DOUT [7:0]) which are electrically connected to the

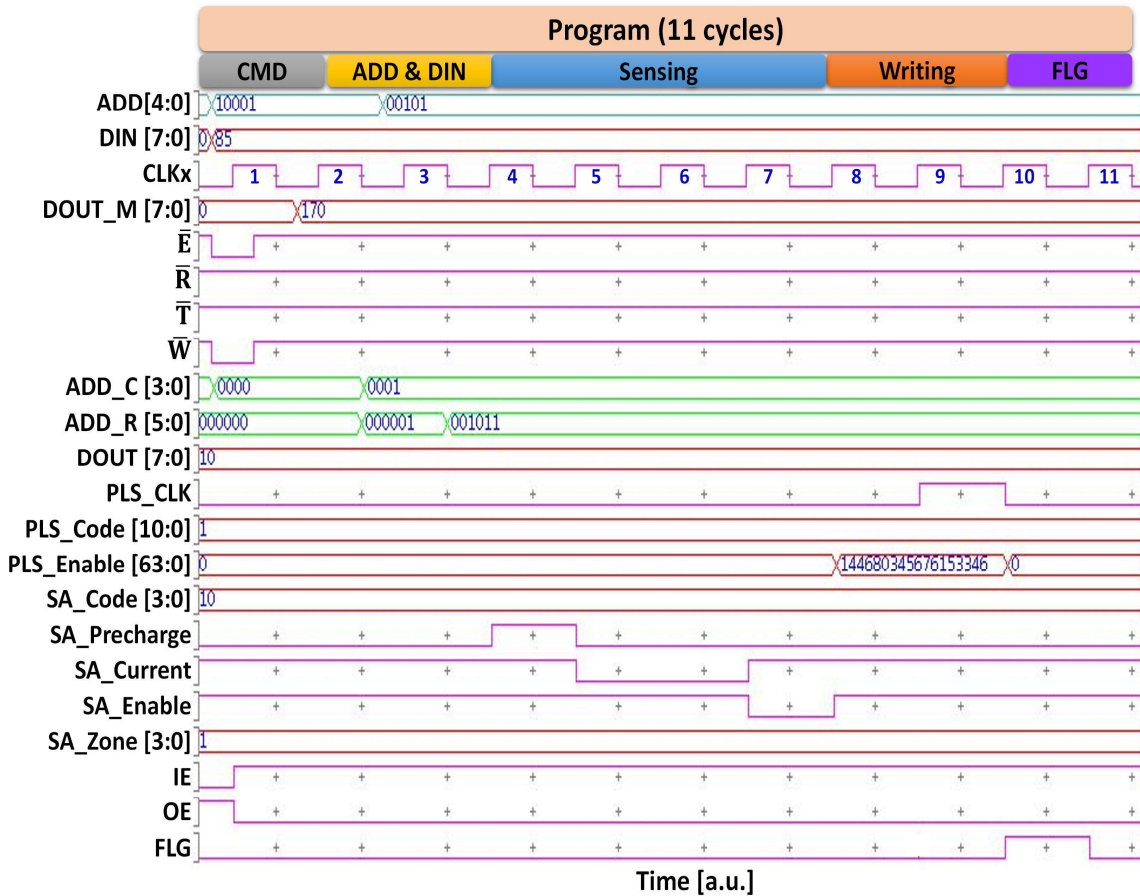


Fig. 5.10 Verilog simulation of the peripheral circuit program operation via ModelSim. The circuit decodes the command and generates the core control signals so that the sense amplifiers can sense the state of selected MTJs during the sensing stage, and the pulse generators can apply write pulses to the cells that need to be switched. After applying the write pulses, the circuit returns the FLG signal '1' to one of the IO pins, indicating that the operation is completed.

DATA [7:0] of the IO pins if the OE is enabled. Once the data is successfully read out, the macro sends the FLG signal '1' via the FLG of the IO pins.

The program operation requires 11 clock cycles, which consists of five stages: CMD, ADD & DIN, Sensing, Writing, and FLG as shown in Fig. 5.10. The configuration of the command for the program operation is $\bar{E} = 0$, $\bar{T} = 1$, $\bar{R} = 1$, $\bar{W} = 0$, which are received at the CMD stage. The targeted address and new data are received at the ADD & DIN stage. Due to the non-deterministic switching of the voltage-controlled MTJs, the macro executes a sensing (pre-read) before applying a write pulse. When the sensing is completed, the peripheral circuit compares the new data and the stored data. Based on

the comparison result, the macro selectively enables some of the pulse generators via the PLS_Enable [63:0]. Once the PLS_CLK becomes '1' at the Writing stage, these enabled pulse generators apply write pulses to the BLs simultaneously. Finally, the macro returns the FLG signal '1' and complete the operation.

5.5 Full-chip Analog and Digital Mixed Signal Verification

Analog-digital mixed signal (ADMS) simulator is a useful tool to verify a mixed-mode circuit. In our MeRAM macro, the digital based peripheral circuit and the analog based core circuit were designed and verified independently in different circuit design environments. Thus, they need to be simulated simultaneously so that we can verify that both circuits interact correctly. Figure 5.11 shows an architecture of ADMS simulation environment where the VHDL-based top module consists of the Verilog-based peripheral circuit and the SPICE netlist-based core circuit, and the both are received the digital stimuli and analog stimuli, respectively. At the interface between the peripheral and core circuits, all ports of the peripheral circuit are regarded as digital. However, all ports of the core circuit are

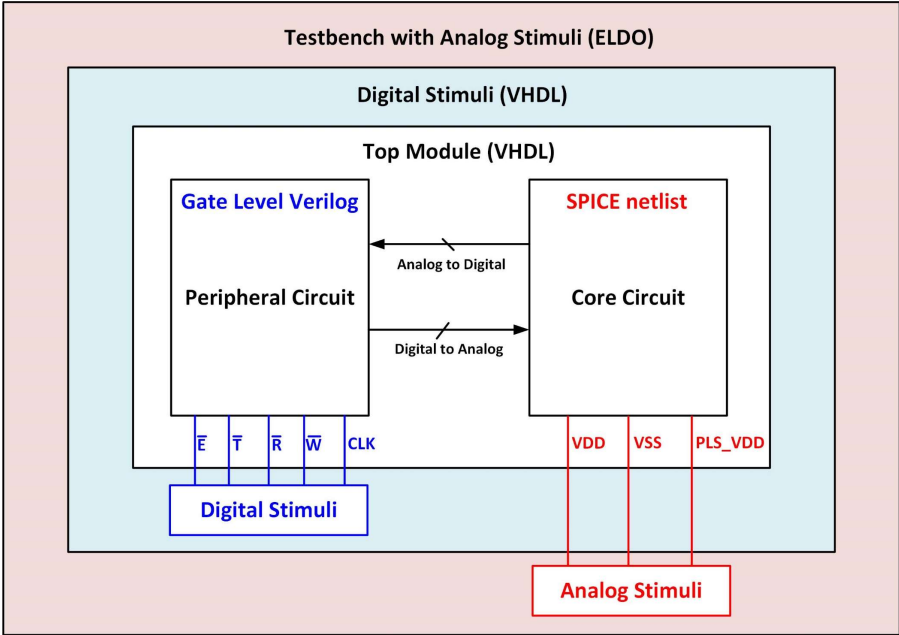


Fig. 5.11 Architecture of ADMS simulation system. The digital peripheral circuit and the analog core circuit can communicate to each other through the ADMS interface.

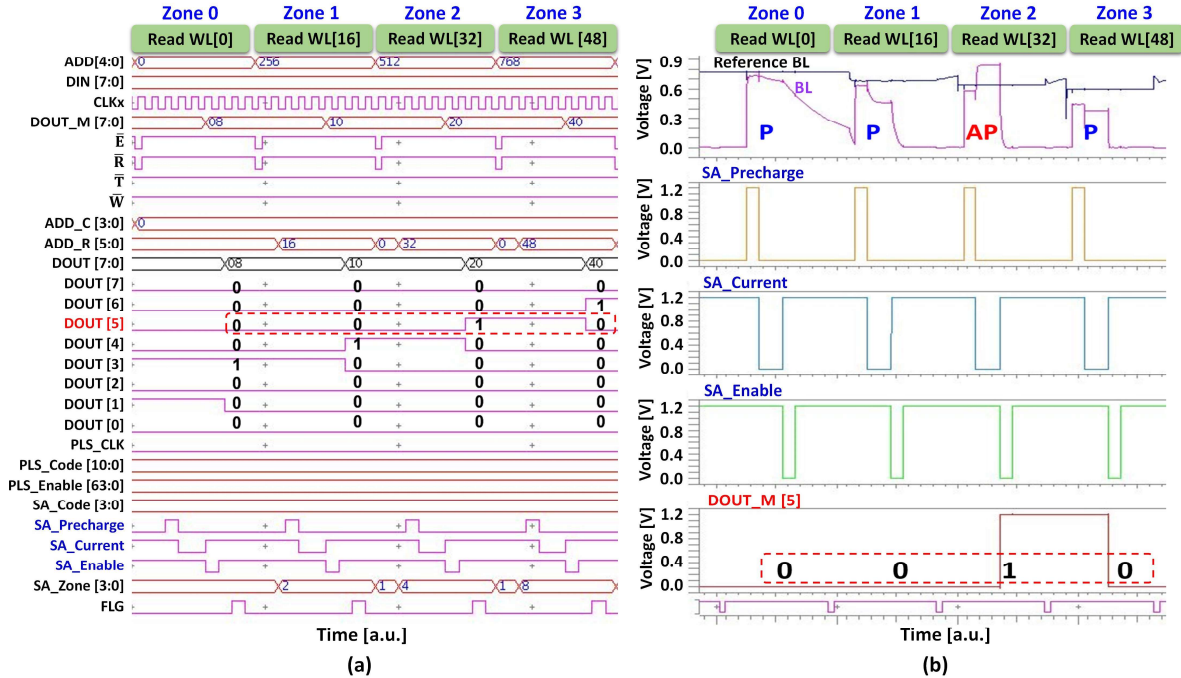


Fig. 5.12 ADMS simulation for the read operation. Input and output signals of (a) the digital based peripheral circuit (b) the analog based core circuit. The potential of the reference BL is automatically adjusted depending on which zone is selected.

regarded as analog. Hence, we must specify the low/high level and rise/fall time of the signals at the interface.

To verify the read operation via ADMS simulation, we implemented predefined data patterns at the first column address of WL[0], WL[16], WL[32], and WL[48] in the 1T-1MTJ cell array. If the readout data from the targeted address matches the predefined pattern, we can verify that the interaction between the peripheral circuit and the core circuit is correct. Four read operations are consecutively executed by increasing the row address, selecting a different zone as shown in Fig. 5.12. At each read operation, the peripheral circuit manipulates the core control signals (i.e. SA_Precharge, SA_Current, SA_Enable) corresponding to each stage as shown in Fig. 5.12(a). The ADMS converts these digital signals to the analog signals that can drive the sense amplifiers of the core circuit. As shown in Fig. 5.12(b), the core circuit performs the read operation and transmits the sensed data to the peripheral circuit via DOUT_M[7:0]. This data is temporarily stored in the output buffer of the peripheral circuit

Operation	Port	[7]	[6]	[5]	[4]	[3]	[2]	[1]	[0]
Program WL[0]	DIN [7:0]	0	1	0	1	0	1	0	1
	DOUT_M [7:0]	0	0	0	0	1	0	0	0
	Mismatch		Yes		Yes	Yes	Yes		Yes
Program WL[16]	DIN [7:0]	0	1	0	1	0	1	0	1
	DOUT_M [7:0]	0	0	0	1	0	0	0	0
	Mismatch		Yes				Yes		Yes
Program WL[32]	DIN [7:0]	0	1	0	1	0	1	0	1
	DOUT_M [7:0]	0	0	1	0	0	0	0	0
	Mismatch		Yes	Yes	Yes		Yes		Yes
Program WL[48]	DIN [7:0]	0	1	0	1	0	1	0	1
	DOUT_M [7:0]	0	1	0	0	0	0	0	0
	Mismatch				Yes		Yes		Yes

Table 5.6 New data and sensed data from the predefined cell patterns. The former and the latter are observed on DIN [7:0] and DOUT_M [7:0], respectively.

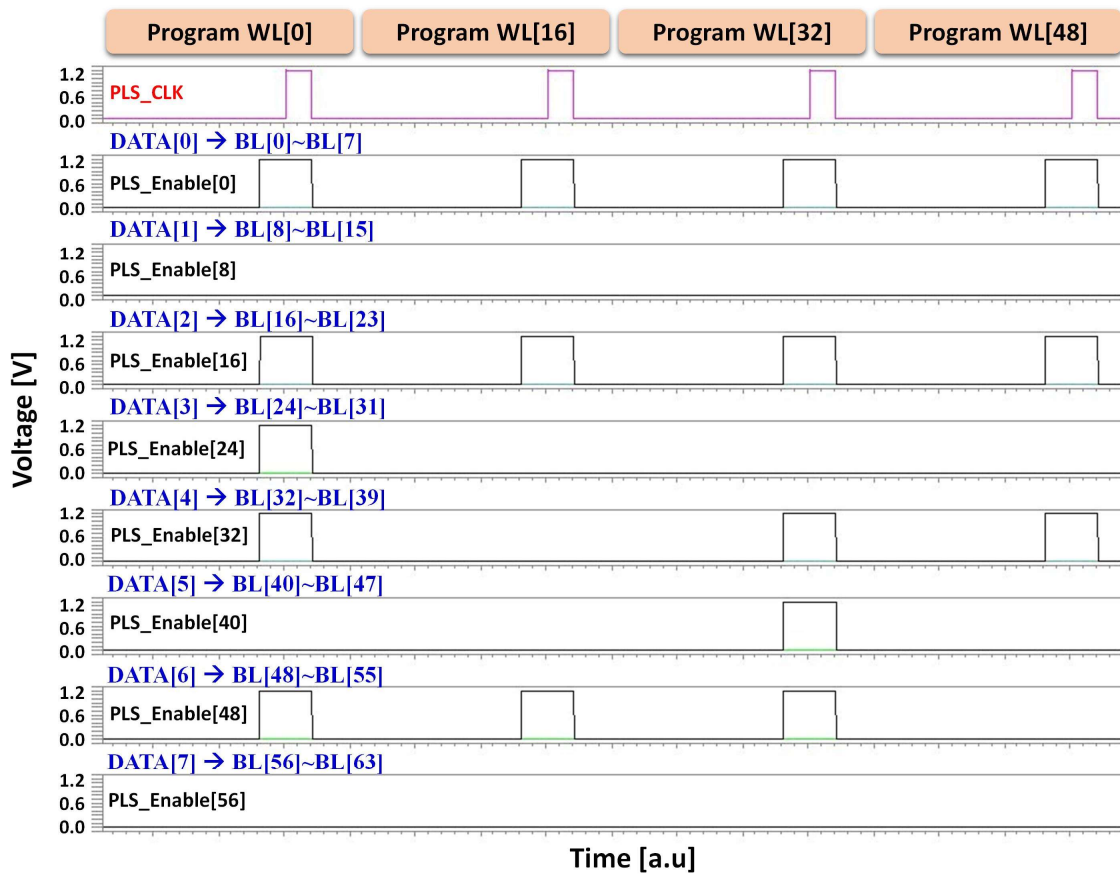


Fig. 5.13 Pulse generator enable signals corresponding to the comparison result of Table 5.6. The mismatched bit enables the pulse generator, applying a write pulse to the BL at the rising edge of PLS_CLK.

and sent to the IO pins via DOUT [7:0]. Noted that the potential of the reference BL is gradually adjusted depending on which zone is selected. We confirmed that the MeRAM macro precisely

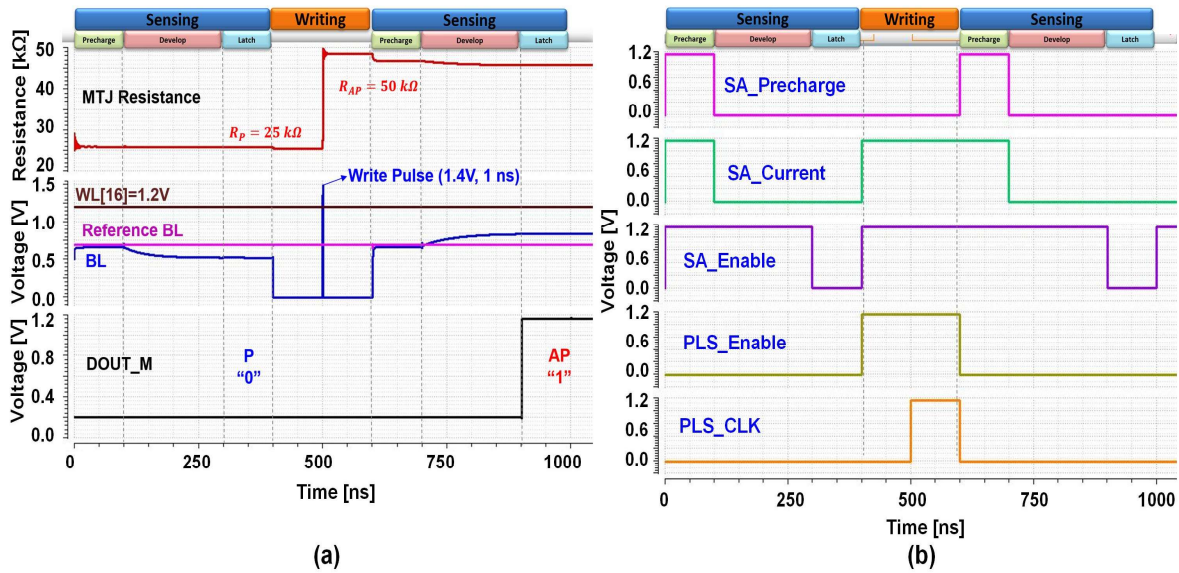


Fig. 5.14 (a) MTJ resistance change along with the connected BL and reference BL during sensing and writing stages. After applying the write pulse with 1.4 V amplitude and 1 ns duration, the MTJ state switches from P to AP. (b) Applied core control signals generated by the peripheral circuit.

performs read operations based on the fact that all of the readout data match the predefined cell patterns.

For the verification of the program operation, we also utilized the same cell data patterns which are used for the read operation. The new data are randomly generated and applied to the DIN [7:0] of the peripheral circuit. Due to the non-deterministic feature of the devices, write pulses are given based on the comparison results between a sensed data and a new data during the program operation. The new data and the sensed data are presented associated with the row address as shown in Table 5.6. Based on the simulation, we verified that the pulse generator enable-signals (PLS_Enable[63:0]) are correctly generated from the peripheral circuit and applied to the pulse generators in the core circuit as shown in Fig. 5.13, which allows the macro selectively applying write pulses to the MTJ devices whose states are different from the new data.

The macrospin MTJ compact model in the ADMS simulation allows to monitor the MTJ resistance change along with the connected BL and reference BL as shown in Fig 5.14(a). Before applying the write pulse, the P state of the device is sensed and converted to digital logic value '0' at DOUT_M.

The write pulse switches the state from P to AP, and the followed sensing readouts the state by converting it to logic value '1'. Figure 5.14(b) shows the applied core control signals for these operations.

5.6 Full-chip Layout

The full-chip layout of the MeRAM macro has been completed as shown in Fig. 5.15 where the layout of the BL drivers and WL drivers are drawn by considering the pitches of the BLs and WLs. The metal 1 layer (M1) is used to draw the WLs whose pitch is 9.24 μm and the SLs which are directly connected to the ground. Initially, we planned to draw the BLs by using the BEOL metal layer whose pitch is 12.50 μm for the 1T-1MTJ cell array. The metal layer 2 (M2) is used for drawing the BLs of the 1T-1R cell array. Although the WL driver fits in the WL pitch, the BL driver in the pulse generator does not fit in the BL pitch due to the large size of the BL driver (50 $\mu\text{m} \times 24 \mu\text{m}$). This is because

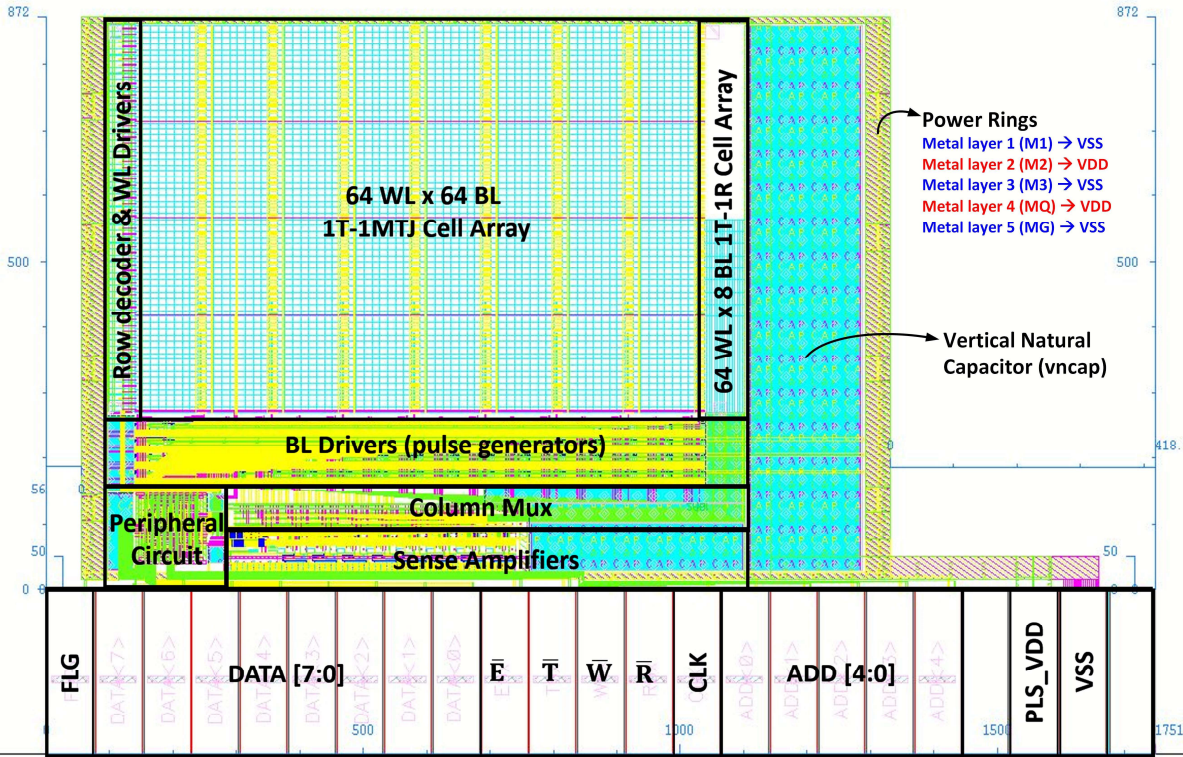


Fig. 5.15 Layout of the MeRAM macro where the WL drivers and BL drivers are arranged based on the pitches of the WLs and BLs, respectively.

the BL driver needs to have a sufficient drivability for generating a write pulse with high slew rate ($1\text{ V} / 100\text{ ps}$). Thus, we organized the 64 BL drivers in a way (16 horizontal \times 4 vertical) at the bottom of the cell arrays.

The column mux selects 8 BLs among 72 BLs of both cell arrays based on a column address and connects them to sense amplifiers. Also, eight reference BLs are connected to the sense amplifiers via the column mux. M2 and M3 are used to routing these two circuits. The size of an individual sense amplifier is $47\text{ }\mu\text{m} \times 49\text{ }\mu\text{m}$. As we described, the layout of the peripheral circuit has been done via the physical synthesis. The peripheral circuit is connected to the IO pins via M5 (MG) while connected to rest of the macro via M1, M2, and M3.

The power rings were drawn on the circumference of the macro by using the five metal layers (M1~M5) to deliver uniform power supply. Each layer width is $40\text{ }\mu\text{m}$ and connected to VSS or VDD. Also, the vertical natural capacitors (vncap) were added to the residue area so that the macro avoid from dropping VDD during the program operation. The capacitance density of the vncap is $0.6\text{ fF}/\mu\text{m}^2$, resulting in total 88 pF of the capacitance on the supply. Typically, if a chip consumes a large amount of current for a short period, a voltage level of power would be decreased, which might cause an operation failure. Similarly, the MeRAM macro requires a large current while applying write pulse to the BLs (each BL has 763 fF). In the worst case, the macro has to drive eight BLs simultaneously. Hence, the macro uses the vncap to stores enough charge to manage this case.

5.7 Test and Evaluation

The MeRAM macro was fabricated based on IBM 130 nm RF-DM technology via a multi-project wafer (MPW) program in which the macro shares the die ($2.5\text{ mm} \times 2.5\text{ mm}$) with other design projects. Figure 5.16 shows the fabricated 4Kbit MeRAM macro without MTJ devices. Since the layers of MTJs (i.e. Ta, CoFeB, MgO, etc.) are typically deposited and etched at the wafer level, we were unable to fabricate the MTJ array at the die level. However, it is possible to partially verify the functionality of

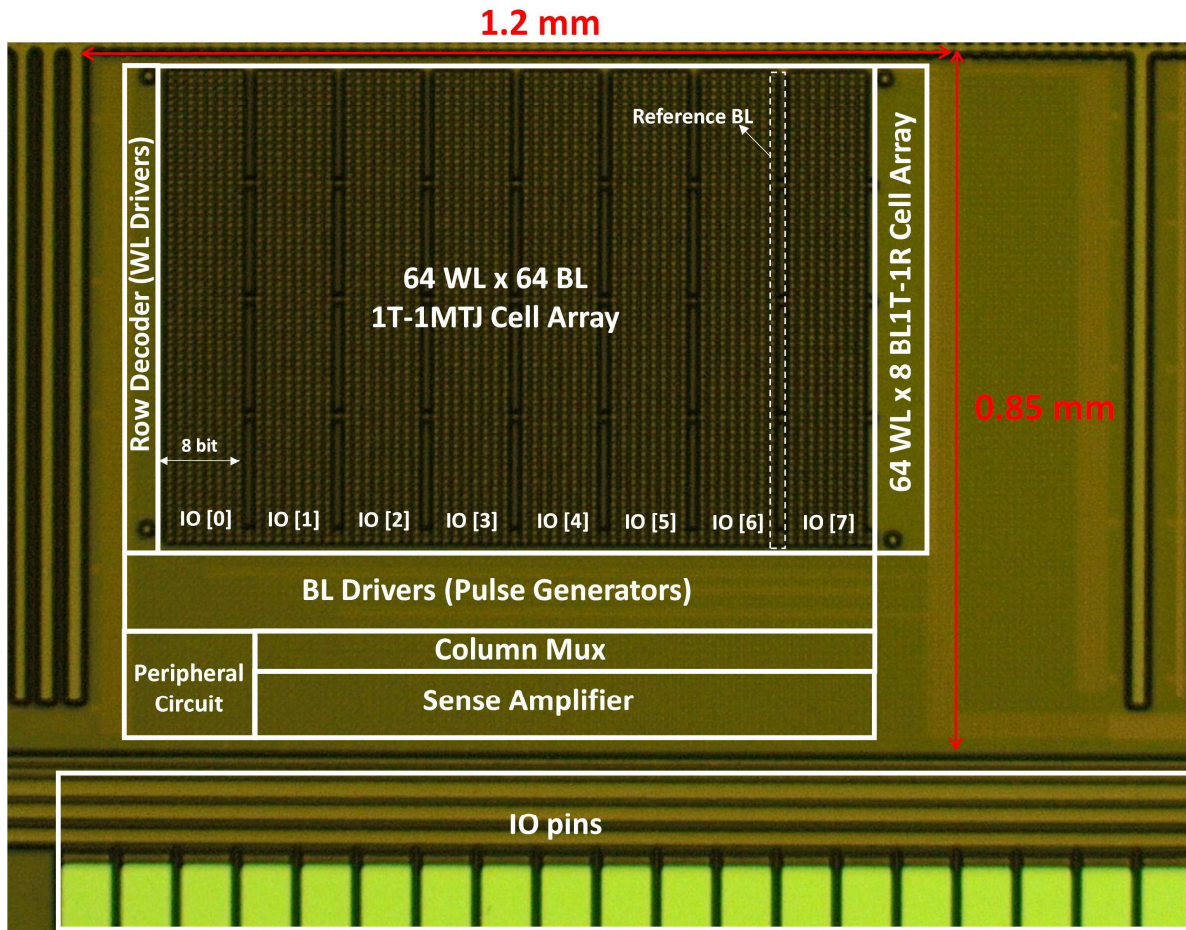


Fig. 5.16 Picture of the fabricated MeRAM macro via 130 nm IBM RF-DM technology. The magnetic layers were not deposited on top of the macro because the wafer was cut into dies after the CMOS fabrication.

the macro by executing the test mode operations accessing to the internal registers and the read operation accessing to the 1T-1R cell array.

The die was packaged, and a type of packaged chip is 100 lead OCP-QFN 12 mm ×12 mm. Then, the chip was soldered on the printed circuit board (PCB) in order to connect the macro to the breadboard as shown in Fig. 5.17. This allows the macro to be connected to the microcontroller (Arduino Due), power supply (Keithley 2400A), and oscilloscope (Agilent MSO7014B). The Arduino Due is a microcontroller board based on the ARM Cortex-M3 CPU, which generates control commands, clock, input data, and address by driving the input IO pins with 3.3 V and monitors the readout data from the macro. By using the built-in libraries of the Arduino programming platform, the

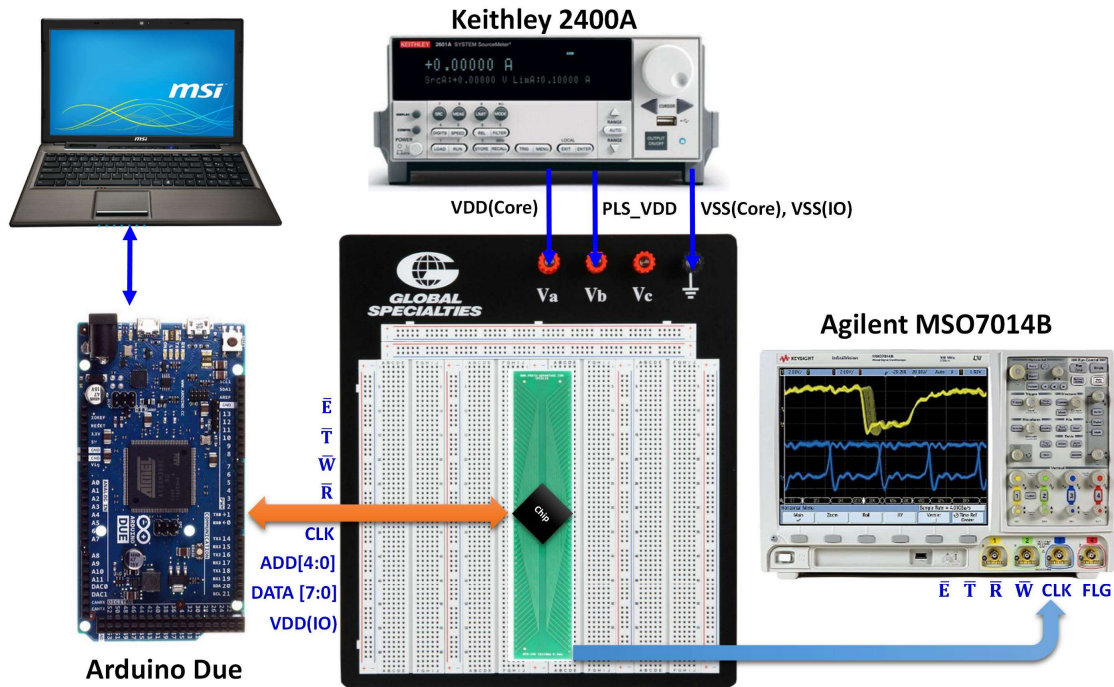


Fig. 5.17 Test set up for the MeRAM macro where the microcontroller (Arduino Due) generates the control signals and receives the readout data that can be displayed on the PC. The oscilloscope (Agilent MSO7014B) monitors the signals in real time.

microcontroller is easily programmed by using a PC to perform the desired functions. The oscilloscope probes some of the IO pins and displays them in real time.

First of all, the peripheral and IO circuits are tested by executing the register configuration and register readout operations, consecutively. Specifically, we store data in the targeted registers corresponding to the given address during the register configuration operation and read out the data by accessing to the same registers during the register readout operation. Figure 5.18(a) shows the measured signals by using the oscilloscope (top) and the part of the microcontroller program (bottom) for the register configuration operation where we store the data pattern 1010_1010 in the targeted registers. We verified that the register configuration operation is successfully completed with five clock cycles via monitoring the FLG becoming '1'. Then, we access to the same registers and read out the stored data via the register readout operation as shown in Fig. 5.18(b). The register readout operation is completed with five clock cycles and returns the readout data 1010_1010 that is identical

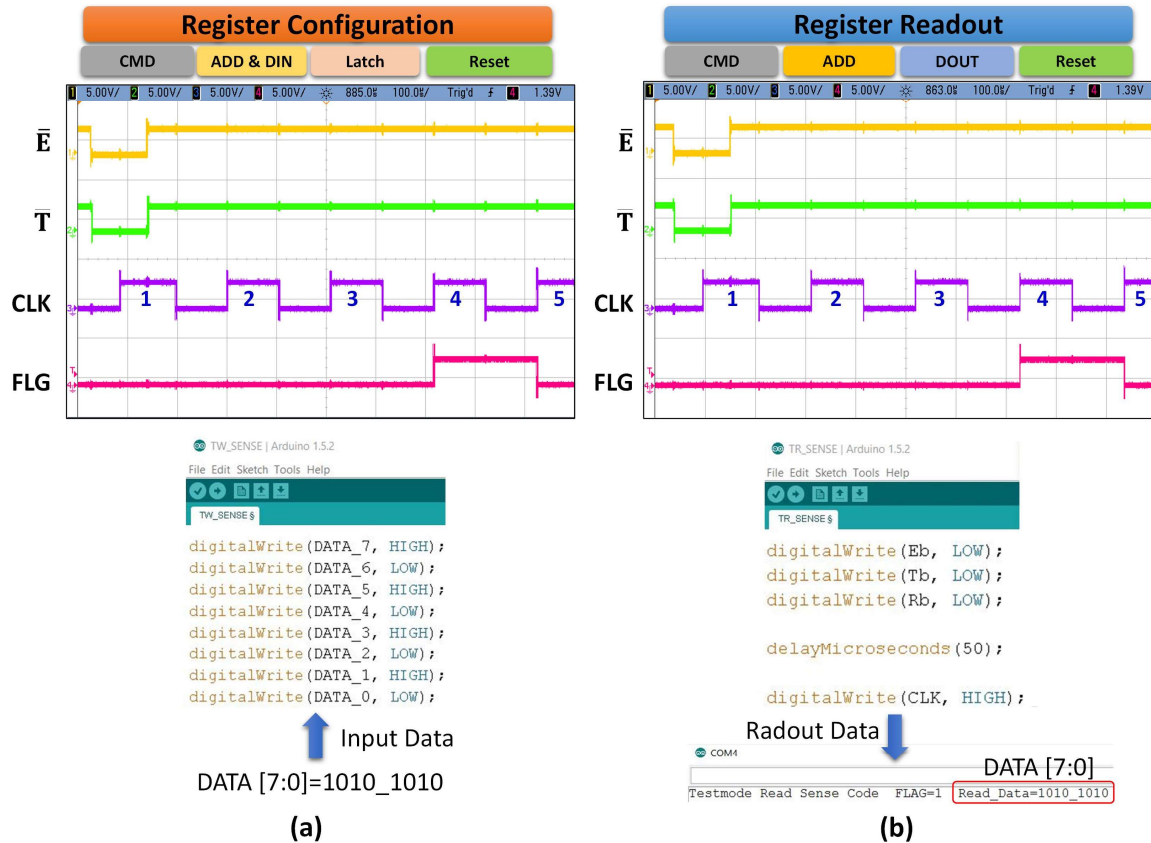


Fig. 5.18 (a) Measured signals via the oscilloscope (top) and part of the microcontroller program which shows the input data to be stored in the targeted registers (bottom) during the register configuration operation (b) Measured signals via the oscilloscope (top) and readout data from the same registers (bottom) during the test register readout operation. The input data and the read out data matches.

to the data that we used in the previous register configuration operation. Therefore, we verify that the both test mode operations are successful, indicating the peripheral and IO circuits are functioning properly.

Secondly, the entire circuit components (except the pulse generators) of the macro can be tested by reading the predefined cell patterns from the 1T-1R cell array. Figure. 5.19(a) shows the measured signals during the read operation where it requires 10 clock cycles to complete. Figure 5.19(b) presents the hard coded data associated with the WLs in the 1T-1R cell array, which were implemented at the circuit design stage. Figures 5.19(c) shows the readout data from the even WL of the Zone2

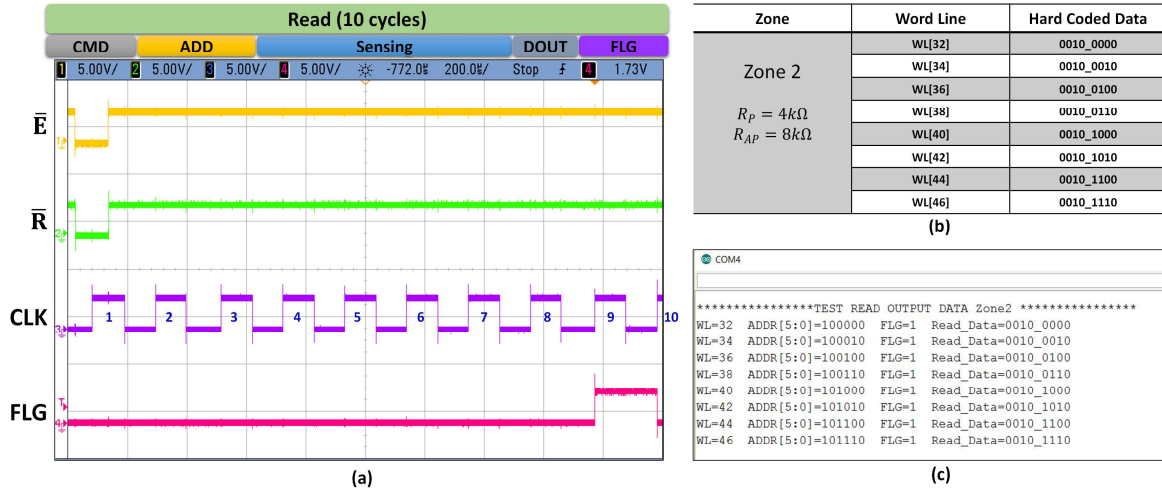


Fig. 5.19 (a) Measured signals for the read operation via the oscilloscope. The number of the clock cycle to complete the operation is equal to the simulation result. (b) Hard coded data in the 1T-1R test cell array associated with the WLs. (c) Received readout data from the macro via consecutive read operations. The hard coded (predefined) data and the read out data coincide with each other.

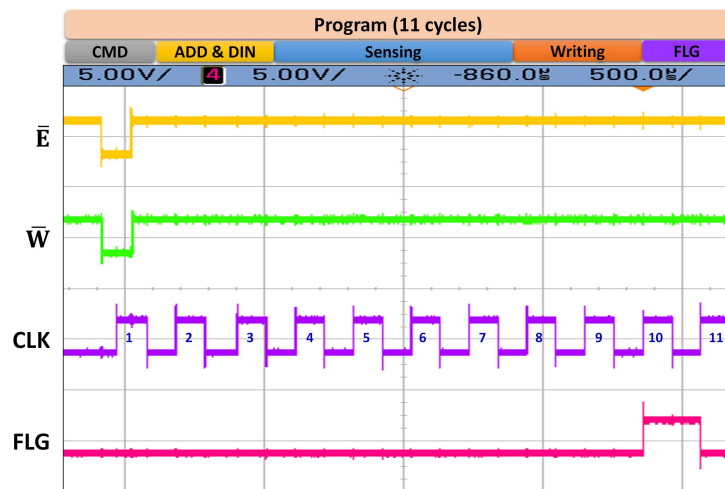


Fig. 5.20 Measured signals for the program operation via the oscilloscope. The number of the clock cycle to complete the operation is equal to the simulation result.

(WL[32]~WL[47]), which are captured by the microcontroller. We can verify that the read operations are successful based on the fact that the predefined data and the readout data are identical.

Although we cannot fully verify the macro functions during the program operation due to the absence of the MTJ cell array, it is possible to test the peripheral circuit. Figure. 5.20 shows the measured signals during the program operation where it requires 11 clock cycles to complete, which coincides with the simulation result.

CHAPTER 6

CONCLUSION

Today's buzzwords in the field of high-technology are machine learning, internet of things (IoT), and augmented/virtual reality, which have created unprecedented demands for the performance of electronic hardware in terms of computational throughput, power, and memory capacity. To meet these demands, semiconductor and electronics industries have not only renovated computer architectures (i.e. HBM, brain-inspired computer) at the system level but also continued to shrink transistors (<10 nm) at the device level. Seemingly, these efforts have achieved balance by fulfilling the demands. However, the demands are still growing, not linearly but rather exponentially, and have become difficult to keep pace with it. Moreover, the size of transistors has reached the physical limit of being a functional switch. Under these circumstances, engineers and scientists have researched beyond-CMOS devices and innovative computing architectures in order to maintain continuous improvement of electronic systems.

In Chapter 1, we introduced a magnetic tunnel junction (MTJ), a promising spintronic device, as one of the alternatives for beyond-CMOS devices and described how the magnetic properties of MTJs can enhance the performance of electronic systems. Specifically, high endurance and CMOS comparability of the MTJ allows it to be used as an embedded system memory device, alleviating static power based on its non-volatility. Also, fabrication of MTJs on top of a CMOS wafer makes possible for the devices to directly communicate with the CMOS circuits via on-chip interconnections, increasing throughput and reducing data transfer energy. We described several types of MRAM families distinguished by its switching mechanism, especially, magnetoelectric random-access memory (MeRAM), which uses voltage-controlled magnetic anisotropy (VCMA) driven precessional switching and outperforms over other MRAM families such as STT-MRAM and SOT-MRAM in

terms of speed and energy. At the end of Chapter 1, we quantitatively compared the performance of MeRAM with other conventional/emerging memory technologies at the device level as well as the array level.

Chapter 2 introduces an MTJ macrospin compact model where the VCMA effect is included in the effective anisotropy field of its built-in Landau-Lifshitz-Gilbert (LLG) equation. The compact model provides the ability to understand the voltage-controlled precessional switching mechanism by monitoring a trajectory of the magnetization. Also, it allows extracting proper bias conditions (amplitude and timing) for achieving the lowest write error rate (WER) by including the thermal noise term in the LLG equation. The two-terminal MTJ model was expanded to a three-terminal model by adding the spin Hall effect (SHE). Based on the three-terminal compact model, we analyzed the characteristics of the voltage-gated SHE switching. In the last part of this chapter, we dealt with the scalability of voltage-controlled MTJ in which both the VCMA and PMA coefficients should be quadratically increased to maintain the same thermal stability and switching voltage in the scaled device.

The compact models were implemented into a circuit design platform where we were able to simulate them with conventional CMOS circuits. In Chapter 3, we introduced several emerging MTJ-CMOS circuits where MTJs play a special role in the circuits besides performing as a memory. In the two-terminal MTJ device, an applied voltage across the device can modulate the energy barrier between the two states, which in turn changes the retention time. Based on this phenomenon, we proposed a voltage-controlled stochastic oscillator, which generates an event-driven stochastic signal for achieving energetically efficient non-uniform sampling. Also, we presented an MTJ based random number generator exploiting the VCMA effect where the MTJ can be in a meta-stable state under the sufficient bias, resulting in generation of a random bit via thermal noise. In the case of the three-terminal MTJ, we found that the critical current of the SHE is a function of the applied voltage, and multiple MTJs on the heavy metal layer can be switched simultaneously. The former was used to

design a spintronic analog to stochastic bit stream converter, and the latter was exploited to design a spintronic programmable logic.

In Chapter 4, we presented several circuit design techniques for a MeRAM macro to enhance the performance. The first design technique is called the source line sensing (SLS) scheme which reversely uses the VCMA effect by applying a sensing voltage with the opposite polarity of the conventional bit line sensing (BLS). Since the SLS linearly enhances the thermal stability as a function of the amplitude of the sensing voltage, it dramatically reduces the read disturbances and increases the sensing margin. Secondly, we introduced the word line pulse (WLP) scheme that applies a write pulse to the WL rather than the BLs. The WLP efficiently improves the slew rate of the write pulse without increasing the size of the drivers, which in turn reduces the WER. Besides these techniques, we proposed the write pulse termination (WPT) scheme and pre-read/write sense amplifier (PWSA) circuit. The former deals with the switching speed distribution of the MTJ array caused by the process variation, and the latter manages non-deterministic switching behavior of the voltage-controlled MTJ.

The CMOS part of the 4Kbit MeRAM macro has been successfully designed based on IBM 130 nm RF-DM technology, and Chapter 5 summarizes key features of the macro along with simulation and measurement results. In the core circuit, the pulse generators are designed to drive the BLs with sufficiently high slew rate and can modulate the duration of the write pulses by using an externally provided digital code. Also, the sense amplifiers are designed to sense MTJs with four different sizes by adjusting the potential of the reference BLs. The required control signals for the core circuit are generated by the digital peripheral circuit which is synchronized by the clock. The measured results of the four different operation modes are accorded with the simulation results, confirming the CMOS part of MeRAM macro functions properly. As an ongoing project, we have designed the CMOS part of a 1Mbit MeRAM based on the advanced technology (55 nm) and plan to fabricate MTJ arrays via BEOL process. We hope to demonstrate the world first voltage-controlled MTJ based high-speed and low-power memory in the near future.

REFERENCE

- [1] G. E. Moore, "Cramming more components onto integrated circuits," *Electron. Mag.*, vol. 38, no. 8, p. 144, Sep. 1965.
- [2] G. Moore, "Progress In Digital Integrated Electronics," *Int. Electron Devices Meet. IEEE*, pp. 11–13, Sep. 1975.
- [3] "International Technology Roadmap for Semiconductors (ITRS): Power Consumption Trends for SOC," in *ITRS*, 2005.
- [4] M. Shafique, S. Garg, J. Henkel, and D. Marculescu, "The EDA Challenges in the Dark Silicon Era," in *Proceedings of the The 51st Annual Design Automation Conference on Design Automation Conference - DAC '14*, 2014, pp. 1–6.
- [5] G. H. Loh, "3D-Stacked Memory Architectures for Multi-core Processors," in *2008 International Symposium on Computer Architecture*, 2008, pp. 453–464.
- [6] J. Jeddeloh and B. Keeth, "Hybrid memory cube new DRAM architecture increases density and performance," in *2012 Symposium on VLSI Technology (VLSIT)*, 2012, pp. 87–88.
- [7] I. Akgun, J. Zhan, Y. Wang, and Y. Xie, "Scalable memory fabric for silicon interposer-based multi-core systems," in *2016 IEEE 34th International Conference on Computer Design (ICCD)*, 2016, pp. 33–40.
- [8] Hulfang Qin, Yu Cao, D. Markovic, A. Vladimirescu, and J. Rabaey, "SRAM leakage suppression by minimizing standby supply voltage," in *SCS 2003. International Symposium on Signals, Circuits and Systems. Proceedings (Cat. No.03EX720)*, pp. 55–60.
- [9] C. W. Smullen, V. Mohan, A. Nigam, S. Gurusurthi, and M. R. Stan, "Relaxing non-volatility for fast and energy-efficient STT-RAM caches," in *2011 IEEE 17th International Symposium on High Performance Computer Architecture*, 2011, pp. 50–61.
- [10] P. M. Tedrow and R. Meservey, "Spin Polarization of Electrons Tunneling from Films of Fe, Co, Ni, and Gd," *Phys. Rev. B*, vol. 7, no. 1, pp. 318–326, Jan. 1973.
- [11] D. Samal and P. S. Anil Kumar, "Giant magnetoresistance," *Resonance*, vol. 13, no. 4, pp. 343–354, Apr. 2008.
- [12] J. Mathon and A. Umerski, "Theory of tunneling magnetoresistance of an epitaxial Fe/MgO/Fe(001) junction," *Phys. Rev. B*, vol. 63, no. 22, p. 220403, May 2001.
- [13] S. Yuasa and D. D. Djayaprawira, "Giant tunnel magnetoresistance in magnetic tunnel junctions with a crystalline MgO(001) barrier," *J. Phys. D: Appl. Phys.*, vol. 40, no. 21, pp. R337–R354, Nov. 2007.
- [14] S. Ikeda, J. Hayakawa, Y. Ashizawa, Y. M. Lee, K. Miura, H. Hasegawa, M. Tsunoda, F. Matsukura, and H. Ohno, "Tunnel magnetoresistance of 604% at 300 K by suppression of Ta diffusion in CoFeB/MgO/CoFeB pseudo-spin-valves annealed at high temperature," *Appl. Phys.*

Let., vol. 93, no. 8, p. 82508, Aug. 2008.

- [15] J. C. Slonczewski, “Current-driven excitation of magnetic multilayers,” *J. Magn. Magn. Mater.*, vol. 159, no. 1, pp. L1–L7, 1996.
- [16] D. C. Ralph and M. D. Stiles, “Spin Transfer Torques,” Nov. 2007.
- [17] J. A. Katine and E. E. Fullerton, “Device implications of spin-transfer torques,” *J. Magn. Magn. Mater.*, vol. 320, no. 7, pp. 1217–1226, Apr. 2008.
- [18] P. Khalili Amiri, Z. M. Zeng, P. Upadhyaya, G. Rowlands, H. Zhao, I. N. Krivorotov, J.-P. Wang, H. W. Jiang, J. A. Katine, J. Langer, K. Galatsis, and K. L. Wang, “Low Write-Energy Magnetic Tunnel Junctions for High-Speed Spin-Transfer-Torque MRAM,” *IEEE Electron Device Lett.*, vol. 32, no. 1, pp. 57–59, Jan. 2011.
- [19] D’Yakonov and V. Perel’, “Possibility of Orienting Electron Spins with Current,” *ZhETF Pisma Redaktsiiu*, vol. 13, 1971.
- [20] Y. K. Kato, R. C. Myers, A. C. Gossard, and D. D. Awschalom, “Observation of the Spin Hall Effect in Semiconductors,” *Science (80-.)*, vol. 306, no. 5703, 2004.
- [21] T. Seki, Y. Hasegawa, S. Mitani, S. Takahashi, H. Imamura, S. Maekawa, J. Nitta, and K. Takanashi, “Giant spin Hall effect in perpendicularly spin-polarized FePt/Au devices,” *Nat. Mater.*, vol. 7, no. 2, pp. 125–9, Feb. 2008.
- [22] L. Liu, C.-F. Pai, Y. Li, H. W. Tseng, D. C. Ralph, and R. A. Buhrman, “Spin-torque switching with the giant spin Hall effect of tantalum,” *Science*, vol. 336, no. 6081, pp. 555–8, May 2012.
- [23] Y. Fan, P. Upadhyaya, X. Kou, M. Lang, S. Takei, Z. Wang, J. Tang, L. He, L.-T. Chang, M. Montazeri, G. Yu, W. Jiang, T. Nie, R. N. Schwartz, Y. Tserkovnyak, and K. L. Wang, “Magnetization switching through giant spin–orbit torque in a magnetically doped topological insulator heterostructure,” *Nat. Mater.*, vol. 13, no. 7, pp. 699–704, Apr. 2014.
- [24] T. Maruyama, Y. Shiota, T. Nozaki, K. Ohta, N. Toda, M. Mizuguchi, A. A. Tulapurkar, T. Shinjo, M. Shiraishi, S. Mizukami, Y. Ando, and Y. Suzuki, “Large voltage-induced magnetic anisotropy change in a few atomic layers of iron,” *Nat. Nanotechnol.*, vol. 4, no. 3, pp. 158–61, Mar. 2009.
- [25] Y. Shiota, T. Nozaki, F. Bonell, S. Murakami, T. Shinjo, and Y. Suzuki, “Induction of coherent magnetization switching in a few atomic layers of FeCo using voltage pulses,” *Nat. Mater.*, vol. 11, no. 1, pp. 39–43, Jan. 2012.
- [26] W.-G. Wang, M. Li, S. Hageman, and C. L. Chien, “Electric-field-assisted switching in magnetic tunnel junctions,” *Nat. Mater.*, vol. 11, no. 1, pp. 64–8, Jan. 2012.
- [27] S. IKEDA, H. SATO, M. YAMANOUCHI, H. GAN, K. MIURA, K. MIZUNUMA, S. KANAI, S. FUKAMI, F. MATSUKURA, N. KASAI, and H. OHNO, “RECENT PROGRESS OF PERPENDICULAR ANISOTROPY MAGNETIC TUNNEL JUNCTIONS FOR NONVOLATILE VLSI,” *SPIN*, vol. 2, no. 3, p. 1240003, Sep. 2012.

- [28] E. Chen, D. Apalkov, Z. Diao, A. Driskill-Smith, D. Druist, D. Lottis, V. Nikitin, X. Tang, S. Watts, S. Wang, S. A. Wolf, A. W. Ghosh, J. W. Lu, S. J. Poon, M. Stan, W. H. Butler, S. Gupta, C. K. A. Mewes, T. Mewes, and P. B. Visscher, “Advances and Future Prospects of Spin-Transfer Torque Random Access Memory,” *IEEE Trans. Magn.*, vol. 46, no. 6, pp. 1873–1878, Jun. 2010.
- [29] 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 *2011 3rd IEEE International Memory Workshop (IMW)*, 2011, pp. 1–3.
- [30] J. Stohr and H. C. Siegmann, *Magnetism: From fundamentals to nanoscale dynamics*, vol. 152. 2006.
- [31] 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,” *Nat. Mater.*, vol. 9, no. 9, pp. 721–724, 2010.
- [32] M. Julliere, “Tunneling between ferromagnetic films,” *Phys. Lett. A*, vol. 54, no. 3, pp. 225–226, Sep. 1975.
- [33] J. S. Moodera, L. R. Kinder, T. M. Wong, and R. Meservey, “Large Magnetoresistance at Room Temperature in Ferromagnetic Thin Film Tunnel Junctions,” *Phys. Rev. Lett.*, vol. 74, no. 16, pp. 3273–3276, Apr. 1995.
- [34] T. Miyazaki and N. Tezuka, “Giant magnetic tunneling effect in Fe/Al₂O₃/Fe junction,” *J. Magn. Magn. Mater.*, vol. 139, no. 3, pp. L231–L234, Jan. 1995.
- [35] A. (Anthony) Bland and B. (Bretislav) Heinrich, *Ultrathin magnetic structures*. Springer, 1994.
- [36] L. Smith HorstCzichos, Tetsuya Saito, *Springer handbook of materials measurement methods*, vol. 9, no. 7–8. 2006.
- [37] C. Chappert, A. Fert, and F. N. Van Dau, “The emergence of spin electronics in data storage,” *Nat. Mater.*, vol. 6, no. 11, pp. 813–823, Nov. 2007.
- [38] K. C. Chun, H. Zhao, J. D. Harms, T.-H. Kim, J.-P. Wang, and C. H. Kim, “A Scaling Roadmap and Performance Evaluation of In-Plane and Perpendicular MTJ Based STT-MRAMs for High-Density Cache Memory,” *IEEE J. Solid-State Circuits*, vol. 48, no. 2, pp. 598–610, Feb. 2013.
- [39] G. Yu, P. Upadhyaya, Y. Fan, J. G. Alzate, W. Jiang, K. L. Wong, S. Takei, S. A. Bender, L.-T. Chang, Y. Jiang, M. Lang, J. Tang, Y. Wang, Y. Tserkovnyak, P. K. Amiri, and K. L. Wang, “Switching of perpendicular magnetization by spin–orbit torques in the absence of external magnetic fields,” *Nat. Nanotechnol.*, vol. 9, no. 7, pp. 548–554, 2014.
- [40] Dajiang Yang, Qing Zhang, and Gang Chen, “A Comprehensive Study of Cobalt Salicide-Induced SRAM Leakage for 90-nm CMOS Technology,” *IEEE Trans. Electron Devices*, vol. 54, no. 10, pp. 2730–2737, Oct. 2007.

- [41] J. A. Mandelman, R. H. Dennard, G. B. Bronner, J. K. DeBrosse, R. Divakaruni, Y. Li, and C. J. Radens, "Challenges and future directions for the scaling of dynamic random-access memory (DRAM)," *IBM J. Res. Dev.*, vol. 46, no. 2.3, pp. 187–212, Mar. 2002.
- [42] L. Torres, R. M. Brum, L. V. Cargnini, and G. Sassatelli, "Trends on the application of emerging nonvolatile memory to processors and programmable devices," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*, 2013, pp. 101–104.
- [43] B. Jacob, S. Ng, and D. Wang, *Memory Systems: Cache, DRAM, Disk*. 2010.
- [44] S. A. Wolf, J. Lu, M. R. Stan, E. Chen, and D. M. Treger, "The Promise of Nanomagnetism and Spintronics for Future Logic and Universal Memory," *Proc. IEEE*, vol. 98, no. 12, pp. 2155–2168, Dec. 2010.
- [45] S. Raoux, G. W. Burr, M. J. Breitwisch, C. T. Rettner, Y.-C. Chen, R. M. Shelby, M. Salinga, D. Krebs, S.-H. Chen, H.-L. Lung, and C. H. Lam, "Phase-change random access memory: A scalable technology," *IBM J. Res. Dev.*, vol. 52, no. 4.5, pp. 465–479, Jul. 2008.
- [46] H.-S. P. Wong, S. Raoux, S. Kim, J. Liang, J. P. Reifenberg, B. Rajendran, M. Asheghi, and K. E. Goodson, "Phase Change Memory," *Proc. IEEE*, vol. 98, no. 12, pp. 2201–2227, Dec. 2010.
- [47] G. W. Burr, M. J. Breitwisch, M. Franceschini, D. Garetto, K. Gopalakrishnan, B. Jackson, B. Kurdi, C. Lam, L. A. Lastras, A. Padilla, B. Rajendran, S. Raoux, and R. S. Shenoy, "Phase change memory technology," *J. Vac. Sci. Technol. B Microelectron. Nanom. Struct.*, vol. 28, no. 2, p. 223, Mar. 2010.
- [48] S. Wang, H. Lee, F. Ebrahimi, P. K. Amiri, K. L. Wang, and P. Gupta, "Comparative Evaluation of Spin-Transfer-Torque and Magnetoelectric Random Access Memory," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 6, no. 2, pp. 134–145, Jun. 2016.
- [49] J. Yi, H. Choi, S. Song, D. Son, S. Lee, J. Park, W. Kim, M. Sung, S. Lee, J. Moon, C. Kim, J. Park, M. Joo, J. Roh, S. Park, S.-W. Chung, J. Jeong, S.-J. Hong, and S.-W. Park, "Requirements of bipolar switching ReRAM for 1T1R type high density memory array," in *Proceedings of 2011 International Symposium on VLSI Technology, Systems and Applications*, 2011, pp. 1–2.
- [50] Y. Fujisaki, "Review of Emerging New Solid-State Non-Volatile Memories," *Jpn. J. Appl. Phys.*, vol. 52, no. 4R, p. 40001, Apr. 2013.
- [51] T. Kishi, H. Yoda, T. Kai, T. Nagase, E. Kitagawa, M. Yoshikawa, K. Nishiyama, T. Daibou, M. Nagamine, M. Amano, S. Takahashi, M. Nakayama, N. Shimomura, H. Aikawa, S. Ikegawa, S. Yuasa, K. Yakushiji, H. Kubota, A. Fukushima, M. Oogane, T. Miyazaki, and K. Ando, "Lower-current and fast switching of a perpendicular TMR for high speed and high density spin-transfer-torque MRAM," in *2008 IEEE International Electron Devices Meeting*, 2008, pp. 1–4.
- [52] R. Dorrance, J. G. Alzate, S. S. Cherepov, P. Upadhyaya, I. N. Krivorotov, J. A. Katine, J. Langer, K. L. Wang, P. K. Amiri, and D. Markovic, "Diode-MTJ Crossbar Memory Cell Using Voltage-Induced Unipolar Switching for High-Density MRAM," *IEEE Electron Device Lett.*

vol. 34, no. 6, pp. 753–755, Jun. 2013.

- [53] T. Yamauchi, L. Hammond, and K. Olukotun, “The hierarchical multi-bank DRAM: a high-performance architecture for memory integrated with processors,” in *Proceedings Seventeenth Conference on Advanced Research in VLSI*, pp. 303–319.
- [54] S. Natarajan, M. Armstrong, M. Bost, R. Brain, M. Brazier, C.-H. Chang, V. Chikarmane, M. Childs, H. Deshpande, K. Dev, G. Ding, T. Ghani, O. Golonzka, W. Han, J. He, R. Heussner, R. James, I. Jin, C. Kenyon, S. Klopčic, S.-H. Lee, M. Liu, S. Lodha, B. McFadden, A. Murthy, L. Neiberg, J. Neiryneck, P. Packan, S. Pae, C. Parker, C. Pelto, L. Pipes, J. Sebastian, J. Seiple, B. Sell, S. Sivakumar, B. Song, K. Tone, T. Troeger, C. Weber, M. Yang, A. Yeoh, and K. Zhang, “A 32nm logic technology featuring 2nd-generation high-k + metal-gate transistors, enhanced channel strain and 0.171 μm^2 SRAM cell size in a 291Mb array,” in *2008 IEEE International Electron Devices Meeting*, 2008, pp. 1–3.
- [55] B. S. Haran, A. Kumar, L. Adam, J. Chang, V. Basker, S. Kanakasabapathy, D. Horak, S. Fan, J. Chen, J. Faltermeier, S. Seo, M. Burkhardt, S. Burns, S. Halle, S. Holmes, R. Johnson, E. McLellan, T. M. Levin, Y. Zhu, J. Kuss, A. Ebert, J. Cummings, D. Canaperi, S. Paparao, J. Arnold, T. Sparks, C. S. Koay, T. Kanarsky, S. Schmitz, K. Petrillo, R. H. Kim, J. Demarest, L. F. Edge, H. Jagannathan, M. Smalley, N. Berliner, K. Cheng, D. LaTulipe, C. Koburger, S. Mehta, M. Raymond, M. Colburn, T. Spooner, V. Paruchuri, W. Haensch, D. McHerron, and B. Doris, “22 nm technology compatible fully functional 0.1 μm^2 6T-SRAM cell,” in *2008 IEEE International Electron Devices Meeting*, 2008, pp. 1–4.
- [56] J. DeBrosse, T. Maffitt, Y. Nakamura, G. Jan, and P.-K. Wang, “A fully-functional 90nm 8Mb STT MRAM demonstrator featuring trimmed, reference cell-based sensing,” in *2015 IEEE Custom Integrated Circuits Conference (CICC)*, 2015, pp. 1–3.
- [57] Y. Lu, T. Zhong, W. Hsu, S. Kim, X. Lu, J. J. Kan, C. Park, W. C. Chen, X. Li, X. Zhu, P. Wang, M. Gottwald, J. Fatehi, L. Seward, J. P. Kim, N. Yu, G. Jan, J. Haq, S. Le, Y. J. Wang, L. Thomas, J. Zhu, H. Liu, Y. J. Lee, R. Y. Tong, K. Pi, D. Shen, R. He, Z. Teng, V. Lam, R. Annapragada, T. Torng, P.-K. Wang, and S. H. Kang, “Fully functional perpendicular STT-MRAM macro embedded in 40 nm logic for energy-efficient IOT applications,” in *2015 IEEE International Electron Devices Meeting (IEDM)*, 2015, p. 26.1.1-26.1.4.
- [58] Y. Cho, Y. Hwang, H. Kim, E. Lee, S. Hong, H. Chung, D. Kim, J. Kim, Y. Oh, H. Hong, G.-Y. Jin, and C. Chung, “Novel Deep Trench Buried-Body-Contact (DBBC) of 4F² cell for sub 30nm DRAM technology,” in *2012 Proceedings of the European Solid-State Device Research Conference (ESSDERC)*, 2012, pp. 193–196.
- [59] Hyunjin Lee, Dae-Young Kim, Bong-Ho Choi, Gyu-Seong Cho, Sung-Woong Chung, Wan-Soo Kim, Myoung-Sik Chang, Young-Sik Kim, Junki Kim, Tae-Kyun Kim, Hyung-Hwan Kim, Hae-Jung Lee, Han-Sang Song, S.-K. Park, Jin-Woong Kim, Sung-Joo Hong, and S.-W. Park, “Fully integrated and functioned 44nm DRAM technology for 1GB DRAM,” in *2008 Symposium on VLSI Technology*, 2008, pp. 86–87.
- [60] K. C. Huang, Y. W. Ting, C. Y. Chang, K. C. Tu, K. C. Tzeng, H. C. Chu, C. Y. Pai, A. Katoch, W. H. Kuo, K. W. Chen, T. H. Hsieh, C. Y. Tsai, W. C. Chiang, H. F. Lee, A. Achyuthan, C. Y. Chen, H. W. Chin, M. . Wang, C. J. Wang, C. S. Tsai, C. M. Oconnell, S. Natarajan, S. G.

- Wuu, I. F. Wang, H. Y. Hwang, and L. C. Tran, "A high-performance, high-density 28nm eDRAM technology with high-K/metal-gate," in *2011 International Electron Devices Meeting*, 2011, p. 24.7.1-24.7.4.
- [61] C. Pei, G. Wang, M. Aquilino, N. Arnold, B. Chandra, W. Chang, X. Chen, W. Davies, K. Hawkins, D. Jaeger, J. B. Johnson, O.-J. Kwon, R. Krishnasamy, W. Kong, J. Liu, X. Li, B. Messenger, E. Nelson, K. Nummy, K. Onishi, D. Poindexter, S. Rombawa, C. Sheraw, T. Tzou, X. Wang, M. Yin, G. Freeman, T. Kirahata, E. Maciejewski, J. Norum, N. Robson, S. Narasimha, P. Parries, P. Agnello, R. Malik, and S. S. Iyer, "0.026 μm^2 high performance Embedded DRAM in 22nm technology for server and SOC applications," in *2014 IEEE International Electron Devices Meeting*, 2014, p. 19.4.1-19.4.4.
- [62] P. Wang, W. Zhang, R. Joshi, R. Kanj, and Y. Chen, "A thermal and process variation aware MTJ switching model and its applications in soft error analysis," in *Proceedings of the International Conference on Computer-Aided Design - ICCAD '12*, 2012, p. 720.
- [63] G. D. Panagopoulos, C. Augustine, and K. Roy, "Physics-Based SPICE-Compatible Compact Model for Simulating Hybrid MTJ/CMOS Circuits," *IEEE Trans. Electron Devices*, vol. 60, no. 9, pp. 2808–2814, Sep. 2013.
- [64] G. D. Demin, E. E. Gusev, A. F. Popkov, P. A. Stepanov, and N. A. Djuzhev, "Compact HSPICE model of magnetic tunnel junction based on voltage-driven spin-transfer torque," in *2016 International Siberian Conference on Control and Communications (SIBCON)*, 2016, pp. 1–6.
- [65] A. S. Roy, A. Sarkar, and S. P. Mudanai, "Compact Modeling of Magnetic Tunneling Junctions," *IEEE Trans. Electron Devices*, vol. 63, no. 2, pp. 652–658, Feb. 2016.
- [66] S. Sharmin, A. Jaiswal, and K. Roy, "Modeling and Design Space Exploration for Bit-Cells Based on Voltage-Assisted Switching of Magnetic Tunnel Junctions," *IEEE Trans. Electron Devices*, vol. 63, no. 9, pp. 3493–3500, Sep. 2016.
- [67] W. Kang, Y. Ran, Y. Zhang, W. Lv, and W. Zhao, "Modeling and Exploration of the Voltage-Controlled Magnetic Anisotropy Effect for the Next-Generation Low-Power and High-Speed MRAM Applications," *IEEE Trans. Nanotechnol.*, vol. 16, no. 3, pp. 387–395, May 2017.
- [68] Y. Shiota, T. Nozaki, S. Tamaru, K. Yakushiji, H. Kubota, A. Fukushima, S. Yuasa, and Y. Suzuki, "Evaluation of write error rate for voltage-driven dynamic magnetization switching in magnetic tunnel junctions with perpendicular magnetization," *Appl. Phys. Express*, vol. 9, no. 1, p. 13001, Jan. 2016.
- [69] C. Grezes, F. Ebrahimi, J. G. Alzate, X. Cai, J. A. Katine, J. Langer, B. Ocker, P. Khalili Amiri, and K. L. Wang, "Ultra-low switching energy and scaling in electric-field-controlled nanoscale magnetic tunnel junctions with high resistance-area product," *Appl. Phys. Lett.*, vol. 108, no. 1, p. 12403, Jan. 2016.
- [70] L. D. Landau and E. M. Lifshits, "On the theory of the dispersion of magnetic permeability in ferromagnetic bodies," vol. 8, pp. 153–169, Jan. 1935.
- [71] 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.," *Nat. Mater.*, vol. 9, no. 9, pp. 721–4, Sep. 2010.
- [72] J. G. Alzate, P. K. Amiri, P. Upadhyaya, S. S. Cherepov, J. Zhu, M. Lewis, R. Dorrance, J. A. Katine, J. Langer, K. Galatsis, D. Markovic, I. Krivorotov, and K. L. Wang, "Voltage-induced switching of nanoscale magnetic tunnel junctions," in *2012 International Electron Devices Meeting*, 2012, p. 29.5.1-29.5.4.
- [73] T. Nozaki, Y. Shiota, M. Shiraishi, T. Shinjo, and Y. Suzuki, "Voltage-induced perpendicular magnetic anisotropy change in magnetic tunnel junctions," *Appl. Phys. Lett.*, vol. 96, no. 2, p. 22506, 2010.
- [74] T. Nozaki, K. Yakushiji, S. Tamaru, M. Sekine, R. Matsumoto, M. Konoto, H. Kubota, A. Fukushima, and S. Yuasa, "Voltage-Induced Magnetic Anisotropy Changes in an Ultrathin FeB Layer Sandwiched between Two MgO Layers," *Appl. Phys. Express*, vol. 6, no. 7, p. 73005, Jul. 2013.
- [75] J. P. Velez, S. S. Jaswal, and E. Y. Tsybal, "Multi-ferroic and magnetoelectric materials and interfaces.," *Philos. Trans. A. Math. Phys. Eng. Sci.*, vol. 369, no. 1948, pp. 3069–97, Aug. 2011.
- [76] S. E. Barnes, J. Ieda, and S. Maekawa, "Rashba spin-orbit anisotropy and the electric field control of magnetism.," *Sci. Rep.*, vol. 4, p. 4105, Jan. 2014.
- [77] J. Z. Sun, "Spin angular momentum transfer in current-perpendicular nanomagnetic junctions," *IBM J. Res. Dev.*, vol. 50, no. 1, pp. 81–100, Jan. 2006.
- [78] D. V. Berkov, "Fast switching of magnetic nanoparticles: simulation of thermal noise effects using the Langevin dynamics," *IEEE Trans. Magn.*, vol. 38, no. 5, pp. 2489–2495, Sep. 2002.
- [79] R. Ahmed and R. H. Victora, "Possible Explanation for Observed Effectiveness of Voltage-Controlled Anisotropy in CoFeB/MgO MTJ," *IEEE Trans. Magn.*, vol. 51, no. 11, pp. 1–4, Nov. 2015.
- [80] H. Lee, C. Grezes, S. Wang, F. Ebrahimi, P. Gupta, P. K. Amiri, and K. L. Wang, "Source Line Sensing in Magneto-Electric Random-Access Memory to Reduce Read Disturbance and Improve Sensing Margin," *IEEE Magn. Lett.*, vol. 7, pp. 1–5, 2016.
- [81] J. J. Nowak, R. P. Robertazzi, J. Z. Sun, G. Hu, J.-H. Park, J. Lee, A. J. Annunziata, G. P. Lauer, R. Kothandaraman, E. J. O Sullivan, P. L. Trouilloud, Y. Kim, and D. C. Worledge, "Dependence of Voltage and Size on Write Error Rates in Spin-Transfer Torque Magnetic Random-Access Memory," *IEEE Magn. Lett.*, vol. 7, pp. 1–4, 2016.
- [82] S. Wang, H. C. Hu, H. Zheng, and P. Gupta, "MEMRES: A Fast Memory System Reliability Simulator," *IEEE Trans. Reliab.*, vol. 65, no. 4, pp. 1783–1797, Dec. 2016.
- [83] C. Grezes, H. Lee, A. Lee, S. Wang, F. Ebrahimi, X. Li, K. Wong, J. A. Katine, B. Ocker, J. Langer, P. Gupta, P. Khalili, and K. L. Wang, "Write Error Rate and Read Disturbance in Electric-Field-Controlled MRAM," *IEEE Magn. Lett.*, pp. 1–1, 2016.

- [84] H. Noguchi, K. Ikegami, K. Abe, S. Fujita, Y. Shiota, T. Nozaki, S. Yuasa, and Y. Suzuki, "Novel voltage controlled MRAM (VCM) with fast read/write circuits for ultra large last level cache," in *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016, p. 27.5.1-27.5.4.
- [85] K. Ikegami, H. Noguchi, S. Takaya, C. Kamata, M. Amano, K. Abe, K. Kushida, E. Kitagawa, T. Ochiai, N. Shimomura, D. Saida, A. Kawasumi, H. Hara, J. Ito, and S. Fujita, "MTJ-based "normally-off processors" with thermal stability factor engineered perpendicular MTJ, L2 cache based on 2T-2MTJ cell, L3 and last level cache based on 1T-1MTJ cell and novel error handling scheme," in *2015 IEEE International Electron Devices Meeting (IEDM)*, 2015, p. 25.1.1-25.1.4.
- [86] L. Liu, O. J. Lee, T. J. Gudmundsen, D. C. Ralph, and R. A. Buhrman, "Current-Induced Switching of Perpendicularly Magnetized Magnetic Layers Using Spin Torque from the Spin Hall Effect," *Phys. Rev. Lett.*, vol. 109, no. 9, p. 96602, Aug. 2012.
- [87] S. Manipatruni, D. E. Nikonov, and I. A. Young, "Voltage and Energy-Delay Performance of Giant Spin Hall Effect Switching for Magnetic Memory and Logic," p. 16, Jan. 2013.
- [88] M. Cubukcu, O. Boulle, M. Drouard, K. Garello, C. Onur Avci, I. Mihai Miron, J. Langer, B. Ocker, P. Gambardella, and G. Gaudin, "Spin-orbit torque magnetization switching of a three-terminal perpendicular magnetic tunnel junction," *Appl. Phys. Lett.*, vol. 104, no. 4, 2014.
- [89] C.-F. Pai, L. Liu, Y. Li, H. W. Tseng, D. C. Ralph, and R. A. Buhrman, "Spin transfer torque devices utilizing the giant spin Hall effect of tungsten," *Appl. Phys. Lett.*, vol. 101, no. 12, p. 122404, Sep. 2012.
- [90] K.-S. Lee, S.-W. Lee, B.-C. Min, and K.-J. Lee, "Threshold current for switching of a perpendicular magnetic layer induced by spin Hall effect," *Appl. Phys. Lett.*, vol. 102, no. 11, p. 112410, Mar. 2013.
- [91] P. Khalili Amiri, P. Upadhyaya, J. G. Alzate, and K. L. Wang, "Electric-field-induced thermally assisted switching of monodomain magnetic bits," *J. Appl. Phys.*, vol. 113, no. 1, p. 13912, 2013.
- [92] Jinn-Shyan Wang, Chao-Ching Wang, and Chingwei Yeh, "TCAM for IP-Address Lookup Using Tree-style AND-type Match Lines and Segmented Search Lines," in *2006 IEEE International Solid State Circuits Conference - Digest of Technical Papers*, 2006, pp. 577–586.
- [93] A. Igor, C. Trevis, and A. Sheikholeslami, "A ternary content-addressable memory (TCAM) based on 4T static storage and including a current-race sensing scheme," *IEEE J. Solid-State Circuits*, vol. 38, no. 1, pp. 155–158, Jan. 2003.
- [94] S. Matsunaga, A. Katsumata, M. Natsui, T. Endoh, H. Ohno, and T. Hanyu, "Design of a Nine-Transistor/Two-Magnetic-Tunnel-Junction-Cell-Based Low-Energy Nonvolatile Ternary Content-Addressable Memory," *Jpn. J. Appl. Phys.*, vol. 51, no. 2, p. 02BM06, Feb. 2012.
- [95] J. Li, R. K. Montoye, M. Ishii, and L. Chang, "1 Mb 0.41 μm^2 2T-2R Cell Nonvolatile TCAM With Two-Bit Encoding and Clocked Self-Referenced Sensing," *IEEE J. Solid-State Circuits*, vol. 49, no. 4, pp. 896–907, Apr. 2014.

- [96] L. Zheng, S. Shin, and S.-M. S. Kang, "Memristors-based Ternary Content Addressable Memory (mTCAM)," in *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2014, pp. 2253–2256.
- [97] K. L. W. P. K. Amiri, R. Dorrance, D. Marković, "Read-disturbance-free nonvolatile content addressable memory (cam)," US 20140071728 A1, 13-Mar-2014.
- [98] M. Ben-Romdhane, C. Rebai, A. Ghazel, P. Desgreys, and P. Loumeau, "Pseudorandom clock signal generation for data conversion in a multistandard receiver," in *2008 3rd International Conference on Design and Technology of Integrated Systems in Nanoscale Era*, 2008, pp. 1–4.
- [99] M. Ben-Romdhane, C. Rebai, K. Grati, A. Ghazel, G. Hechmi, P. Desgreys, and P. Loumeau, "Non-Uniform Sampled Signal Reconstruction for Multistandard WiMAX/WiFi Receiver," in *2007 IEEE International Conference on Signal Processing and Communications*, 2007, pp. 181–184.
- [100] G. Roa, T. Le Pelleter, A. Bonvilain, A. Chagoya, and L. Fesquet, "Designing ultra-low power systems with non-uniform sampling and event-driven logic," in *Proceedings of the 27th Symposium on Integrated Circuits and Systems Design - SBCCI '14*, 2014, pp. 1–6.
- [101] T. Beyrouthy, L. Fesquet, and R. Rolland, "Data sampling and processing: Uniform vs. non-uniform schemes," in *2015 International Conference on Event-based Control, Communication, and Signal Processing (EBCCSP)*, 2015, pp. 1–6.
- [102] P. S. Wyckoff, "Frequency estimator performance analysis with compressive sensing or non-uniform sampling," in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2014, pp. 674–678.
- [103] E. Masry, "Random sampling and reconstruction of spectra," *Inf. Control*, vol. 19, no. 4, pp. 275–288, 1971.
- [104] M. A. Davenport, J. N. Laska, J. R. Treichler, and R. G. Baraniuk, "The Pros and Cons of Compressive Sensing for Wideband Signal Acquisition: Noise Folding versus Dynamic Range," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4628–4642, Sep. 2012.
- [105] M. Osama, L. Gaber, and A. Hussein, "Design of high performance Pseudorandom Clock Generator for compressive sampling applications," in *2016 33rd National Radio Science Conference (NRSC)*, 2016, pp. 257–265.
- [106] R. Z. Bhatti, K. M. Chugg, and J. Draper, "Standard cell based pseudo-random clock generator for statistical random sampling of digital signals," in *2007 50th Midwest Symposium on Circuits and Systems*, 2007, pp. 1110–1113.
- [107] D. Bellasi, L. Bettini, T. Burger, Q. Huang, C. Benkeser, and C. Studer, "A 1.9 GS/s 4-bit sub-Nyquist flash ADC for 3.8 GHz compressive spectrum sensing in 28 nm CMOS," in *2014 IEEE 57th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2014, pp. 101–104.
- [108] S. Kose, E. Salman, Z. Ignjatovic, and E. G. Friedman, "Pseudo-random clocking to enhance

- signal integrity,” in *2008 IEEE International SOC Conference*, 2008, pp. 47–50.
- [109] P. Khalili Amiri, J. G. Alzate, X. Q. Cai, F. Ebrahimi, Q. Hu, K. Wong, C. Grezes, H. Lee, G. Yu, X. Li, M. Akyol, Q. Shao, J. A. Katine, J. Langer, B. Ocker, and K. L. Wang, “Electric-Field-Controlled Magnetoelectric RAM: Progress, Challenges, and Scaling,” *IEEE Trans. Magn.*, vol. 51, no. 11, pp. 1–7, Nov. 2015.
- [110] S. Kanai, M. Yamanouchi, S. Ikeda, Y. Nakatani, F. Matsukura, and H. Ohno, “Electric field-induced magnetization reversal in a perpendicular-anisotropy CoFeB-MgO magnetic tunnel junction,” *Appl. Phys. Lett.*, vol. 101, no. 12, p. 122403, Sep. 2012.
- [111] K. L. Wang, J. G. Alzate, and P. Khalili Amiri, “Low-power non-volatile spintronic memory: STT-RAM and beyond,” *J. Phys. D. Appl. Phys.*, vol. 46, no. 7, p. 74003, Feb. 2013.
- [112] A. Djemouai, M. Sawan, and M. Slamani, “New circuit techniques based on a high performance frequency-to-voltage converter,” in *ICECS’99. Proceedings of ICECS ’99. 6th IEEE International Conference on Electronics, Circuits and Systems (Cat. No.99EX357)*, vol. 1, pp. 13–16.
- [113] Y. Gao, K. Abouda, R. Beges, and P. Besse, “Study of Frequency-to-Voltage Converter immunity to fast transient pulses,” in *2016 Asia-Pacific International Symposium on Electromagnetic Compatibility (APEMC)*, 2016, pp. 849–851.
- [114] M. Bucci, L. Germani, R. Luzzi, A. Trifiletti, and M. Varanonuovo, “A high-speed oscillator-based truly random number source for cryptographic applications on a smartcard IC,” *IEEE Trans. Comput.*, vol. 52, no. 4, pp. 403–409, Apr. 2003.
- [115] P. Xu, Y. L. Wong, T. K. Horiuchi, and P. A. Abshire, “Compact floating-gate true random number generator,” *Electron. Lett.*, vol. 42, no. 23, p. 1346, 2006.
- [116] M. Degaldo-Restituto, F. Medeiro, and A. Rodríguez-Vázquez, “Nonlinear switched-current CMOS IC for random signal generation,” *Electron. Lett.*, vol. 29, no. 25, p. 2190, 1993.
- [117] W. T. Holman, J. A. Connelly, and A. B. Dowlatabadi, “An integrated analog/digital random noise source,” *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.*, vol. 44, no. 6, pp. 521–528, Jun. 1997.
- [118] C. S. Petrie and J. A. Connelly, “A noise-based IC random number generator for applications in cryptography,” *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.*, vol. 47, no. 5, pp. 615–621, May 2000.
- [119] A. T. Marketos and S. W. Moore, “The Frequency Injection Attack on Ring-Oscillator-Based True Random Number Generators,” Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [120] M. Alawad and mingjie lin, “Survey of Stochastic-based Computation Paradigms,” *IEEE Trans. Emerg. Top. Comput.*, pp. 1–1, 2016.
- [121] Won Ho Choi, Yang Lv, Jongyeon Kim, A. Deshpande, Gyuseong Kang, Jian-Ping Wang, and C. H. Kim, “A Magnetic Tunnel Junction based True Random Number Generator with

conditional perturb and real-time output probability tracking,” in *2014 IEEE International Electron Devices Meeting*, 2014, p. 12.5.1-12.5.4.

- [122] S. Oosawa, T. Konishi, N. Onizawa, and T. Hanyu, “Design of an STT-MTJ based true random number generator using digitally controlled probability-locked loop,” in *2015 IEEE 13th International New Circuits and Systems Conference (NEWCAS)*, 2015, pp. 1–4.
- [123] I. Kuon and J. Rose, “Measuring the Gap Between FPGAs and ASICs,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 26, no. 2, pp. 203–215, Feb. 2007.
- [124] S. Masui, T. Ninomiya, M. Oura, W. Yokozeki, K. Mukaida, and S. Kawashima, “A ferroelectric memory-based secure dynamically programmable gate array,” *IEEE J. Solid-State Circuits*, vol. 38, no. 5, pp. 715–725, May 2003.
- [125] D. Suzuki, M. Natsui, T. Endoh, H. Ohno, and T. Hanyu, “Six-input lookup table circuit with 62% fewer transistors using nonvolatile logic-in-memory architecture with series/parallel-connected magnetic tunnel junctions,” *J. Appl. Phys.*, vol. 111, no. 7, p. 07E318, Feb. 2012.
- [126] W. Zhao, E. Belhaire, C. Chappert, F. Jacquet, and P. Mazoyer, “New non-volatile logic based on spin-MTJ,” *Phys. status solidi*, vol. 205, no. 6, pp. 1373–1377, Jun. 2008.
- [127] Y. Y. Liauw, Z. Zhang, W. Kim, A. El Gamal, and S. S. Wong, “Nonvolatile 3D-FPGA with monolithically stacked RRAM-based configuration memory,” in *2012 IEEE International Solid-State Circuits Conference*, 2012, pp. 406–408.
- [128] H. Lee, J. G. Alzate, R. Dorrance, X. Q. Cai, D. Markovic, P. Khalili Amiri, and K. L. wang, “Design of a Fast and Low-Power Sense Amplifier and Writing Circuit for High-Speed MRAM,” *IEEE Trans. Magn.*, vol. 51, no. 5, pp. 1–7, May 2015.
- [129] J. Song, J. Kim, S. H. Kang, and S. Yoon, “Sensing margin trend with technology scaling in MRAM,” no. March 2010, pp. 313–325, 2011.
- [130] H. Koike and T. Endoh, “A New Sensing Scheme with High Signal Margin Suitable for Spin-Transfer Torque RAM,” pp. 5–6, 2011.
- [131] B. R. Gaines and B. R., “Stochastic computing,” in *Proceedings of the April 18-20, 1967, spring joint computer conference on - AFIPS '67 (Spring)*, 1967, p. 149.
- [132] W. J. Poppelbaum, C. Afuso, and J. W. Esch, “Stochastic computing elements and systems,” in *Proceedings of the November 14-16, 1967, fall joint computer conference on - AFIPS '67 (Fall)*, 1967, p. 635.
- [133] B. R. Gaines, “Stochastic Computing Systems,” in *Advances in Information Systems Science*, Boston, MA: Springer US, 1969, pp. 37–172.
- [134] S. Sharifi Tehrani, S. Mannor, and W. J. Gross, “Fully Parallel Stochastic LDPC Decoders,” *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5692–5703, Nov. 2008.
- [135] X. Li, W. Qian, M. D. Riedel, K. Bazargan, and D. J. Lilja, “A reconfigurable stochastic

architecture for highly reliable computing,” in *Proceedings of the 19th ACM Great Lakes symposium on VLSI - GLSVLSI '09*, 2009, p. 315.

- [136] S. Sharifi Tehrani, A. Naderi, G.-A. Kamendje, S. Hemati, S. Mannor, and W. J. Gross, “Majority-Based Tracking Forecast Memories for Stochastic LDPC Decoding,” *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4883–4896, Sep. 2010.
- [137] P. Li and D. J. Lilja, “Using stochastic computing to implement digital image processing algorithms,” in *2011 IEEE 29th International Conference on Computer Design (ICCD)*, 2011, pp. 154–161.
- [138] W. Qian, X. Li, M. D. Riedel, K. Bazargan, and D. J. Lilja, “An Architecture for Fault-Tolerant Computation with Stochastic Logic,” *IEEE Trans. Comput.*, vol. 60, no. 1, pp. 93–105, Jan. 2011.
- [139] P. Li, D. J. Lilja, W. Qian, K. Bazargan, and M. D. Riedel, “Computation on Stochastic Bit Streams Digital Image Processing Case Studies,” *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 22, no. 3, pp. 449–462, Mar. 2014.
- [140] J. D. Alves and J. Lobo, “Remote lab for Stochastic Computing using reconfigurable logic,” in *2015 3rd Experiment International Conference (exp.at'15)*, 2015, pp. 173–174.
- [141] A. Alaghi, C. Li, and J. P. Hayes, “Stochastic circuits for real-time image-processing applications,” in *Proceedings of the 50th Annual Design Automation Conference on - DAC '13*, 2013, p. 1.
- [142] N. Onizawa, D. Katagiri, W. J. Gross, and T. Hanyu, “Analog-to-stochastic converter using magnetic-tunnel junction devices,” in *2014 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*, 2014, pp. 59–64.
- [143] L. A. de Barros Naviner, H. Cai, Y. Wang, W. Zhao, and A. Ben Dhia, “Stochastic computation with Spin Torque Transfer Magnetic Tunnel Junction,” in *2015 IEEE 13th International New Circuits and Systems Conference (NEWCAS)*, 2015, pp. 1–4.
- [144] N. Onizawa, D. Katagiri, W. J. Gross, and T. Hanyu, “Analog-to-Stochastic Converter Using Magnetic Tunnel Junction Devices for Vision Chips,” *IEEE Trans. Nanotechnol.*, vol. 15, no. 5, pp. 705–714, Sep. 2016.
- [145] K. L. Wang, H. Lee, and P. Khalili Amiri, “Magnetoelectric Random Access Memory-Based Circuit Design by Using Voltage-Controlled Magnetic Anisotropy in Magnetic Tunnel Junctions,” *IEEE Trans. Nanotechnol.*, vol. 14, no. 6, pp. 992–997, Nov. 2015.
- [146] Q. Hao, W. Chen, and G. Xiao, “Beta (β) tungsten thin films: Structure, electron transport, and giant spin Hall effect,” *Appl. Phys. Lett.*, vol. 106, no. 18, p. 182403, May 2015.
- [147] Q. Hao and G. Xiao, “Giant Spin Hall Effect and Switching Induced by Spin-Transfer Torque in a W / Co 40 Fe 40 B 20 / MgO Structure with Perpendicular Magnetic Anisotropy,” *Phys. Rev. Appl.*, vol. 3, no. 3, p. 34009, Mar. 2015.
- [148] J. Seo, B. Brezzo, Y. Liu, B. D. Parker, S. K. Esser, R. K. Montoye, B. Rajendran, J. A. Tierno,

- L. Chang, D. S. Modha, and D. J. Friedman, "A 45nm CMOS neuromorphic chip with a scalable architecture for learning in networks of spiking neurons," in *2011 IEEE Custom Integrated Circuits Conference (CICC)*, 2011, pp. 1–4.
- [149] C.-S. Poon and K. Zhou, "Neuromorphic silicon neurons and large-scale neural networks: challenges and opportunities.," *Front. Neurosci.*, vol. 5, p. 108, Jan. 2011.
- [150] 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 :," vol. 0, no. c, 2005.
- [151] W. F. Brown, "Thermal Fluctuations of a Single-Domain Particle," *Phys. Rev.*, vol. 130, no. 5, pp. 1677–1686, Jun. 1963.
- [152] A. Kalitsov, P.-J. Zermatten, F. Bonell, G. Gaudin, S. Andrieu, C. Tiusan, M. Chshiev, and J. P. Velev, "Bias dependence of tunneling magnetoresistance in magnetic tunnel junctions with asymmetric barriers.," *J. Phys. Condens. Matter*, vol. 25, no. 49, p. 496005, Dec. 2013.
- [153] S. Kanai, Y. Nakatani, M. Yamanouchi, S. Ikeda, H. Sato, F. Matsukura, and H. Ohno, "Magnetization switching in a CoFeB/MgO magnetic tunnel junction by combining spin-transfer torque and electric field-effect," *Appl. Phys. Lett.*, vol. 104, no. 21, p. 212406, May 2014.
- [154] W. Kang, Y. Ran, W. Lv, Y. Zhang, and W. Zhao, "High-Speed, Low-Power, Magnetic Non-Volatile Flip-Flop With Voltage-Controlled, Magnetic Anisotropy Assistance," *IEEE Magn. Lett.*, vol. 7, pp. 1–5, 2016.
- [155] H. Lee, F. Ebrahimi, P. K. Amiri, and K. L. Wang, "Low-Power and High-Density Spintronic Programmable Logic (SPL) Using Voltage-Gated Spin Hall Effect in Magnetic Tunnel Junctions," *IEEE Magn. Lett.*, vol. PP, no. 99, pp. 1–1, 2016.
- [156] M. Fukuda, K. Higuchi, and K. Takeuchi, "Non-volatile Random Access Memory and NAND Flash Memory Integrated Solid-State Drives with Adaptive Codeword Error Correcting Code for 3.6 Times Acceptable Raw Bit Error Rate Enhancement and 97% Power Reduction," *Jpn. J. Appl. Phys.*, vol. 50, no. 4, p. 04DE09, Apr. 2011.
- [157] Joohee Kim and M. C. Papaefthymiou, "Constant-Load Energy Recovery Memory for Efficient High-speed Operation," *Low Power Electronics and Design, 2004. ISLPED '04. Proceedings of the 2004 International Symposium on*. pp. 240–243, 2004.
- [158] P. Knag, W. Lu, and Z. Zhang, "A Native Stochastic Computing Architecture Enabled by Memristors," *IEEE Trans. Nanotechnol.*, vol. 13, no. 2, pp. 283–293, Mar. 2014.