



AFRL-RY-WP-TR-2024-0136

**FRAMEWORK FOR NOVEL COMPUTE (FRANC)
Advanced Magnetic Tunnel Junctions (MTJs) For Computation In
And Near Random Access Memory**

**Jian-Ping Wang and Tony Low
University of Minnesota**

**Weigang Wang
Arizona University**

**Daniel Gopman and Brian Hoskings
NIST**

**William Taylor
Global Foundries**

**JULY 2024
Final Report**

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

See additional restrictions described on inside pages

STINFO COPY

**AIR FORCE RESEARCH LABORATORY
SENSORS DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7320
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with The Under Secretary of Defense memorandum dated 24 May 2010 and AFRL/DSO policy clarification email dated 13 January 2020. This report is available to the general public, including foreign nationals.

Copies may be obtained from the Defense Technical Information Center (DTIC)
(<http://www.dtic.mil>).

AFRL-RY-WP-TR-2024-0136 HAS BEEN REVIEWED AND IS APPROVED FOR
PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signature//

CHRISTOPHER A. BOZADA
Program Manager
Aerospace Components and Subsystems Division

//signature//

JOHN S. CETNAR (acting)
Deputy Chief, Aerospace Components &
Subsystems Technology Division
Sensors Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

*Disseminated copies will show “//Signature//” stamped or typed above the signature blocks.

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE July 2024		2. REPORT TYPE Final		3. DATES COVERED <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%;">START DATE 24 August 2018</td> <td style="width: 50%;">END DATE 31 August 2023</td> </tr> </table>		START DATE 24 August 2018	END DATE 31 August 2023
START DATE 24 August 2018	END DATE 31 August 2023						
4. TITLE AND SUBTITLE FRAMEWORK FOR NOVEL COMPUTE (FRANC) Advanced Magnetic Tunnel Junctions (MTJs) For Computation In And Near Random Access Memory							
5a. CONTRACT NUMBER FA8650-18-2-7868		5b. GRANT NUMBER N/A		5c. PROGRAM ELEMENT NUMBER 61101E			
5d. PROJECT NUMBER N/A		5e. TASK NUMBER N/A		5f. WORK UNIT NUMBER Y1UD			
6. AUTHOR(S) Jian-Ping Wang and Tony Low (University of Minnesota) Weigang Wang (Arizona University) Daniel Gopman and Brian Hoskings (NIST) William Taylor (Global Foundries)							
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Minnesota Minneapolis, MN 55455 Arizona University 1401 E University Blvd, Tucson, AZ 85721		NIST - National Institute of Standards and Technology 100 Bureau Dr Ste 1000, Gaithersburg, MD 20899 Global Foundries 400 Stonebreak Rd Ext, Malta, NY 12020		8. PERFORMING ORGANIZATION REPORT NUMBER			
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory Sensors Directorate Wright-Patterson Air Force Base, OH 45433-7320 Air Force Materiel Command United States Air Forces		Defense Advanced Research Projects Agency (DARPA/MTO) 675 North Randolph Street Arlington, VA 22203		10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/Ryd			
11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-RY-WP-TR-2024-0136							
12. DISTRIBUTION/AVAILABILITY STATEMENT DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.							
13. SUPPLEMENTARY NOTES This material is based on research sponsored by the Air Force Research Laboratory (AFRL) and the Defense Advanced Research Projects Agency (DARPA) under agreement number FA8650-18-2-7868. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory (AFRL), the Defense Advanced Research Projects Agency (DARPA), or the U.S. Government.							
14. ABSTRACT This effort theoretically and experimentally studied advanced magnetic tunnel junctions (MTJ) using new classes of magnetic materials, i.e. Weyl semimetals. The effort evaluated theoretically the potential for ultrahigh (~104%) tunnel magnetoresistance (TMR) in Weyl semimetal MTJs. Fabrication of MTJs with advanced materials, structures, and a novel switching mechanism, resulted in successful demonstration of magneto-crystalline perpendicular magnetic anisotropy (PMA) MTJs with a TMR of 65% and an ultralow switching current density down to 1.1x10 ⁵ A/cm ² , which is approximately 10x smaller than the current state-of-the-art (SOTA) results. In addition, by using a multi-interface free layer (MIFL) method, a strong PMA was achieved and up to 212% TMR with the mainstream CoFeB-MgO materials system and 100 ps switching delay, 260% TMR ratio, and 0.02 fJ switching energy (equivalent, projected to 10 nm), and MTJ size of 100 nm. The concept and unique features of computational random-access memory (CRAM) was developed.							
15. SUBJECT TERMS Magnetic junction tunnels, perpendicular magnetic anisotropy, synthetic antiferromagnetic free layers, computational random-access memory, magnetic Weyl semimetal tunnel junctions, interlayer exchange mechanisms							
16. SECURITY CLASSIFICATION OF:				17. LIMITATION OF ABSTRACT SAR			
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified		18. NUMBER OF PAGES 115			
19a. NAME OF RESPONSIBLE PERSON Christopher Bozada				19b. PHONE NUMBER (Include area code) N/A			

Table of Contents

Section	Page
List of Figures	iv
List of tables.....	vii
ABSTRACT.....	1
ACKNOWLEDGEMENTS	2
1 SUMMARY	3
2 INTRODUCTION	5
2.1 Fundamentals of Magnetic Tunnel Junctions (MTJs)	5
2.2 MTJs with Bulk Perpendicular Magnetic Anisotropy (PMA).....	5
2.2.1 MTJs with Magneto-crystalline PMA	5
2.2.2 MTJs composed of Co-based Multilayers	7
2.3 MTJs with Interfacial PMA	8
2.3.1 MTJs composed of Perpendicularly Magnetized CoFeB	9
2.4 Advanced MTJs	10
2.4.1 Synthetic Antiferromagnetic (SAF) Free Layers.....	10
2.5 Theoretical Explorations.....	10
2.5.1 Performance Evaluation of Magnetic Tunnel Junction based on Magnetic Weyl Semimetal	10
2.5.2 Interlayer Exchange Coupling on Magnetic Tunnel Junctions.....	11
2.5.3 Investigation of Unconventional SOT on Magnetization Switching.....	11
2.5.4 Domain Wall Motion in Asymmetric MTJ	12
2.5.5 Spin Hall Conductivity in Chiral Materials	12
2.6 Computational random-access memory (CRAM)	12
2.6.1 CRAM concept, working principle, and features	12
2.6.2 CRAM Benefits	15
2.6.3 Opportunity and Challenge for CRAM	16
2.6.4 NIST Test Vehicle and Technology Transfer Capability	20
3 METHODS, ASSUMPTIONS, AND PROCEDURES.....	23
3.1 Theoretical Explorations.....	23
3.1.1 Modeling Approach of Magnetic Weyl Semimetal Tunnel Junctions	23
3.1.2 Interlayer Exchange Coupling on Magnetic Tunnel Junctions.....	23
3.1.3 Modeling Approach of VCEC	23
3.1.4 Investigation of Unconventional SOT on Magnetization Switching.....	25
3.1.5 Domain Wall Motion in Asymmetric MTJ	25
3.1.6 Spin Hall Conductivity in Chiral Materials	25
3.2 MTJ Fabrications	25
3.2.1 MTJ Stack Development and Characterization	25
3.2.2 Micro- and nano-size MTJ device fabrication.....	29
3.3 MTJ Electrical Testing and Measurements.....	31
3.3.1 Infrastructures and Capabilities	31
3.3.2 Magnetostatic DC Electrical Measurements.....	31
3.3.3 Dynamic electrical measurements	33
3.3.4 Ultra-fast Switching Probability Measurements.....	34

Section	Page
3.4 CRAM.....	36
3.4.1 CRAM ‘Circuit-around-Die’ Demonstration and Experiments.....	36
3.4.2 Integration on CMOS Test Vehicle	37
3.4.3 Modeling and simulation of CRAM	40
4 RESULTS AND DISCUSSION	41
4.1 Theoretical Explorations	41
4.1.1 Gigantic TMR ($\sim 10^4\%$) in Magnetic Weyl Semimetal MTJs	41
4.1.2 Interlayer Exchange Coupling on Magnetic Tunnel Junctions.....	43
4.1.3 Magnitude and Sign Change of the IEC via VCEC Effect.....	43
4.1.4 Investigation of Unconventional SOT on Magnetization Switching.....	44
4.1.5 Domain Wall Motion in Asymmetric MTJ	45
4.1.6 Spin Hall Conductivity in Chiral Materials	46
4.2 Experimental Results of MTJs with Bulk PMA	48
4.2.1 VCEC Switching of MTJs with $L1_0$ FePd-based Dual-SAF Free Layers	48
4.2.2 A Substantial TMR Ratio of $L1_0$ FePd-based MTJs.....	50
4.2.3 High-quality $L1_0$ FePd and its SAF Prepared on Si Wafers.	52
4.3 Experimental Results of MTJs with Interfacial PMA.....	54
4.3.1 pMTJs with Multi-interface Free Layers and High TMR.....	54
4.3.2 pMTJs with Ultra-low Switching Energy	60
4.4 CRAM.....	68
4.4.1 CRAM Experimental Demonstration by the ‘Circuit-around-Die’ Method.....	68
4.4.2 Stochastic Computing-CRAM (SC-CRAM)	74
4.4.3 CRAM Error Rate and Manufacturability Studies	78
5 CONCLUSIONS.....	82
5.1 Theoretical Explorations.....	82
5.1.1 Performance Evaluation of Magnetic Tunnel Junction Based on Magnetic Weyl Semimetal	82
5.1.2 Interlayer Exchange Coupling on Magnetic Tunnel Junctions.....	82
5.1.3 Investigation of Unconventional SOT on Magnetization Switching.....	82
5.1.4 Domain Wall Motion in Asymmetric MTJ	83
5.1.5 Spin Hall Conductivity in Chiral Materials	83
5.2 MTJs with Bulk PMA.....	83
5.2.1 VCEC Switching of MTJs with $L1_0$ FePd-based Dual-SAF Free Layers	83
5.2.2 A substantial TMR Ratio of $L1_0$ FePd-based MTJs	83
5.2.3 High-quality $L1_0$ FePd and its SAF Prepared on Si Wafers	84
5.3 MTJs with Interfacial PMA	84
5.3.1 pMTJ with Multiple Interface Free Layers.....	84
5.3.2 pMTJs with Ultralow Switching Energy	84
5.4 CRAM.....	84
5.4.1 CRAM Experimental Demonstration by the ‘Circuit-around-die’ Method.....	85
5.4.2 SC-CRAM	85
5.4.3 CRAM Error Rate and Manufacturability Studies	85
6 REFERENCES	86

Section	Page
7 APPENDICES	99
LIST OF ACRONYMS, ABBREVIATIONS, AND SYMBOLS	103

List of Figures

Figure	Page
Figure 2.1 Thermal Stability Scales as a Function of Film Thickness and Device Diameter[10]..	6
Figure 2.2 Performance Benchmarking of Spintronics Memories	7
Figure 2.3 Schematic Representation of a pMTJ with Multiple-interface Free layers (MIFLs), where a Number of CoFeB Layers are coupled Through NM Layers to Function as a Single Magnetic Layer	9
Figure 2.4 Schematic Illustration of the Stack Structure of a L10 FePd-based VCEC-pMTJ Sample.....	10
Figure 2.5 Bilayer Heterostructure comprised of a Ferromagnet (top layer) and a Strong Spin-orbit-coupled NM Layer with a Single-crystal yz Mirror Plane.....	11
Figure 2.6 We Consider a Magnetic Tunnel Junction where an in-plane DW Displacement in the Free Layer is Induced by a Vertical Current Flow	12
Figure 2.7 Illustrations of CRAM Concept, Features, and Potential Applications[57].....	13
Figure 2.8 Illustrations of the CRAM Cell Architecture and the Working principle of CRAM Logic Operation [57].....	14
Figure 2.9 Illustration of SC Working Principles[122]	17
Figure 2.10 Illustration of Currant Paths in SC-CRAM Operations[122]	18
Figure 2.11 Illustrations of the Dependency of CRAM Logic Operation Error Rate on Logic Operation parameters[124]	20
Figure 2.12 NIST Test Vehicle shown with NIST Trial Test Integration	21
Figure 2.13 (a) Commercial 200 MHz Bandwidth High Frequency Switch Matrix	22
Figure 2.14 Hierarchical Diagrams of NIST Modeling Architecture and Depictions of NIST 75-channel DAC Board with Packaged memory Device Chip	22
Figure 3.1 Schematic of the Simulation Structure with an SAF stack of FePd/Ta/Co ₂ Fe ₆ (CoFe)/MgO.....	24
Figure 3.2 Some of the UHV Magnetron-sputtering Systems Employed for MTJ Stack Deposition at our Facilities	26
Figure 3.3 Example of the Thin Film Characterization Tools in our Team, (A) A PPMS Equipped with a VSM Module, (B) A Probe of the CIPT System	27
Figure 3.4 The Process Flow of MTJ Device Fabrication	30
Figure 3.5 (A) 100 KeV E-beam Lithography System (Ebeam Litho EBP5000+ Vistec).....	30
Figure 3.6 (a) Schematic Picture of the Four-probe Measurement Method, (b) Cascade GSG Probe, (c) GSG Probes on our Co-planar Waveguide Shape Electrodes.....	32
Figure 3.7 (a) A Typical Resistance versus Magnetic Field Loop Measured from an In-plane MTJ, (b) Home Build Magnetic Coils Testing Stage, (c) 3D Projected Magnet Testing Stage and (d) Big Electromagnet Testing Stage.....	32
Figure 3.8 (a) The Electrical Setup of the Frequency Domain Measurement, (b) The Measured Power Spectrum	34
Figure 3.9 Illustration of MTJ Stochastic Switching Pulse and State Read Timing[115].....	35
Figure 3.10 Illustration of Testing Setup[115]	36
Figure 3.11 CRAM Experimental Setup[57].....	37

Figure	Page
Figure 3.12 Schematic Depiction of Integrated CHIPS with Classes of Operations that can be Performed.....	38
Figure 3.13 (a) Diagram of each Cell on CMOS Substrates Patterned by NIST with Two Transistors and Two Metallic Contacts where Nonvolatile Memory (NVM) Units will be Patterned, (b) Diagram of Probe Connection Layout Relative to CDRAM Banks and (c) Image of Metallic Contacts Patterned on CMOS Substrate.....	39
Figure 3.14 Optical Images showing Current Progress of MTJ Fabrication on CMOS Substrates	39
Figure 3.15 Graphic Illustration of the Modeling Methods for CDRAM Manufacturability Evaluation by Considering Device-Device size VARIATION of MTJs[124]	40
Figure 4.1 (a) In-plane(left)and Out-of-plane(right)Magnetic Weyl Semimetal Tunnel Junctions, (b) Field-like Torque Efficiency T_x/J Exerted on the MWSM Interface	42
Figure 4.2 Critical Current Density as a Function of the Asymmetry Parameter δ	43
Figure 4.3 ΔE (IEC) vs E-field for the FePd/Ta/CoFe/MgO SAF Structure Calculated by our DFT	44
Figure 4.4 Bilayer Heterostructure Comprised of a Ferromagnet (top layer) and a Strong Spin-orbit-coupled NM Layer with a Single-crystal yz Mirror Plane.....	45
Figure 4.5 We Consider a Magnetic Tunnel Junction where an In-plane DW Displacement in the Free Layer is Induced by a Vertical Current Flow	46
Figure 4.6 Crystal Structures, Band Structures, SHCs, and SHAs of SG-198 Chiral Topological Semimetals.....	47
Figure 4.7 Bipolar E-field Switching in pMTJs with Ultralow Current Density[11].....	49
Figure 4.8 Robustness of Bipolar E-field Switching in pMTJs.....	50
Figure 4.9 (A) M-Hext Loop of the Unpatterned THIN films (red line and squares) and Major R-Hext Loop of a Patterned Device (blue line), (B) R-I loops of a L10 FePd pMTJ Device Under Different Hext	51
Figure 4.10 (A) Hysteresis Loops of (001)-Textured $L1_0$ FePd (solid symbols) and the Control Sample (empty symbols).....	53
Figure 4.11 (a) A Cross-sectional HRTEM Image of the FePd SAF, (b) A STEM-HAADF Image of the L10 FePd/Ir/L10 FePd Trilayer Structure, (c) NBD Patterns Collected within a Single Grain and their Overlap[10].....	54
Figure 4.12 Schematic Representation of a pMTJ with Multiple-interface Free Layer (MIFL) where a Number of FM Layers are Coupled Through NM Layers to Function as a Single Magnetic Layer	55
Figure 4.13 (a) Hysteresis Loops of [CoFeB(0.75 nm)/MgO(0.8 nm)]x3 Measured Under Different Magnetic Fields, (b) the Same Curves at the Low-field Region, (c) CoFeB Thickness Dependence of the TMR in pMTJs with MgO-MIFLs.....	56
Figure 4.14 (a) Hysteresis Loops of MgO/ CoFeB/Ta(1nm)/CoFeB/MgO Measured on Different Magnetic Fields, (b) the Same Curves at Low Field-region, (c) CoFeB Thickness Dependence of the TMR in pMTJs with Ta-MIFLs	57
Figure 4.15 (a) Hysteresis Loops of MgO/CoFeB/Mo(0.9 nm)/CoFeB/MgO/CoFeB /Mo Measured on Different Magnetic Fields, (b) the Same Curves at Low Field-region, (c) CoFeB Thickness Dependence of the TMR in pMTJs with Mo-MIFLs	59

Figure	Page
Figure 4.16 (a) Different MTJ Stacks for Sample 1, 2, 3 and 4, A thin Ir Layer is Inserted in the Different Position of MTJ for Sample 2, 3 and 4., (b) A Schematic of Measurement Setup, (c) Curves of TMR-H for Sample 2 with Different Thickness of Ir Insertion Layer, (d) Curves of TMR-H for MTJs with Different Positions of Ir Insertion Layer	62
Figure 4.17 (a) R-H Loop of Sample 1, The polar Angle (θ_H) of the Applied Magnetic Field is Fixed at 70 Degree, the Large (+/- 600 mV) and Small (-10 mV) Volage were Applied to the MTJ Nanodevice and the Resistance was Measured when Sweeping the Magnetic Field, (b) the Product of Switching Probabilities as a Function of Pulse Duration with Different Magnetic Fields Under a Fixed Voltage Pulse of 2.7V, (c) The Zoom-in (<1 ns) of b for the Region with Yellow-color Background, A Magnetic Switching with a Product of Switching Probability above 80% was Achieved with 150 ns Voltage Pulse Duration, which is Marked with Red Dash Circle, (d) R-H Loop of Sample 3, the Polar Angle of the Applied Field is Fixed at 50 Degree, the Large (+/- 440 mV) and Small (-10 mV) Volage were Applied to the MTJ Nanodevice and the Resistance was Measured when Sweeping the Magnetic Field, (e) Resistance Measured after the Application of Voltage Pulse of 3.0 V with 0.6 ns Duration, the Applied Magnetic Field is 320 Oe with a Polar Angle of 50 Degree, (f) the Product of Switching Probabilities as a Function of Pulse Duration with Different Voltage Magnitude Under External Magnetic Field of 320 Oe ...	64
Figure 4.18 (a) Coercivity for Sample 1, 2 and 3 Under Different Annealing Temperatures, (b) TMR Ratio for Sample 1, 2 and 3 Under Different Annealing Temperatures, the Product of Switching Probabilities as a Function of Pulse Voltage Understand the Annealing Temperature of (c) 300C, (d) 400C and (e) 500C, (f) the Peak of the Product of Switching Probabilities for Sample 1, 2 and 3 Under Different Annealing Temperatures	67
Figure 4.19 (a) High-angle Annular Dark-field Scanning Transmission Electron Microscopy (HAADF-STEM) Imaging of Sample 3 with Post-annealing of 400C, the Selected Region for the Further Analysis in (b) and (c) is Marked with Red Rectangle	68
Figure 4.20 Experimental Results for CRAM Logic Operation[57]	69
Figure 4.21 CRAM 1-bit Full Adder Demonstration Results[57]	71
Figure 4.22 CRAM Accuracy Projections[57]	74
Figure 4.23 Energy Consumption Breakdown[122]	77
Figure 4.24 Box Plot of NAND Gate Error Rate vs. TMR Ratio[124]	79
Figure 4.25 TMR ratio - σ_d/μ_d Combinations that Meet Specified Error Rate Requirements[124]	81

List of tables

Table	Page
Table 3.1 Electrical Components for the High-frequency Measurement	33
Table 4.1 Number of MTJs, Computation Steps, and Subarray Size for Various Arithmetic Functions in SC-CRAM[122]	76
Table 4.2 Performance Evaluation of SC-CRAM[122].....	76

ABSTRACT

This final report provides a comprehensive summary of the DARPA ERI FRANC project on advanced MTJs for computing. The major findings and achievements of the team, derived from both theoretical exploration and experimental demonstration, are detailed herein. Considering the new class of magnetic materials, i.e. Weyl semimetals, we evaluate theoretically the potential for ultrahigh ($\sim 10^4\%$) tunnel magnetoresistance (TMR) in Weyl semimetal magnetic tunnel junctions (MTJs). Experimentally, by fabricating MTJs with advanced materials, structures, and a novel switching mechanism, we demonstrated magneto-crystalline perpendicular magnetic anisotropy (PMA) MTJs with a TMR of 65% and an ultralow switching current density down to 1.1×10^5 A/cm², which is approximately 10x smaller than the current state-of-the-art (SOTA) results. In addition, by using a multi-interface free layer (MIFL) method, we achieved strong PMA and up to 212% TMR with the mainstream CoFeB-MgO materials system. We achieved 100 ps switching delay, 260% TMR ratio, and 0.02 fJ switching energy (equivalent, projected to 10 nm), and MTJ size of 100 nm. We introduce the concept and unique features of computational random-access memory (CRAM). Lastly, our CRAM-related experimental and numerical achievements are presented and discussed.

This final report starts with a summary of the entire project in **Section 1 Summary**. Then the fundamentals of MTJs, the advantages of MTJs with PMA, two origins of PMA (i.e., bulk and interfacial), synthetic antiferromagnet (SAF) free layer structures, and CRAM concept and features are introduced in **Section 2 Introduction**. The methods and approaches used are discussed in **Section 3 Methods, Assumptions, and Procedures**. After presenting our results and discussions in **Section 4 Results and Discussion**, we conclude this final report with **Section 5 Conclusions**. The references are given in **Section 6 References**.

ACKNOWLEDGEMENTS

We gratefully acknowledge the technical assistance provided by the University of Minnesota's research platforms, including the Minnesota Nano Center, Characterization Facility, and Institute of Rock Magnetism. We also extend our gratitude to Prof. Ulya Karpuzcu and Prof. Sachin Sapatnekar for their support on the computational random-access memory (CRAM) benchmarking and simulations, Prof. Paul Crowell for his support with the physical property measurement system (PPMS), Prof. Xiaojia Wang for her assistance with time-resolved magneto-optic Kerr effect (TR-MOKE) characterization, and Prof. Andre Mkhoyan for his expertise in scanning transmission electron microscopy (STEM) analysis.

1 SUMMARY

We have made significant advancements in the development of magnetic tunnel junctions (MTJs) and the related device applications in many aspects. Our team has developed MTJ stacks that are scalable down to sub-5 nm, compared to the state-of-the-art (SOTA) MTJs at 15 nm. Our research demonstrated an ultra-low switching current density, achieving a tenfold reduction compared to the SOA, in these advanced MTJs. We have demonstrated the feasibility of integrate advanced MTJs with CMOS. Crucially, we have successfully established a semiconductor industry-compatible fabrication process for these advanced MTJs using 200-mm multi-chamber multi-target sputtering system at the University of Minnesota, which is ready for lab to fab transition. We have established collaboration with Honeywell, Everspin, Western Digital and Global Foundries for the further development and applications of our advanced MTJs technology.

In the search for high efficiency MTJ, our theoretical model predicts a new chiral magnetic tunnel junction with a groundbreaking TMR (Tunnel Magnetoresistance) ratio ten times higher than the current SOTA. Additionally, we have developed a new MTJ structure with a large TMR and multi-interface free layers. This program has achieved a record-high TMR in voltage-controllable MTJs (260%) and a record-low switching energy of 3.3 fJ, which is equivalent to 0.01 fJ when MTJ is scaled down to 10 nm. We achieved 100 ps switching delay, 260% TMR ratio, and 0.02 fJ switching energy (equivalent, projected to 10 nm), and MTJ size of 100 nm. The switching energy exceeds the metric target of the project deliverable by 4 times.

At circuit and architecture level, our main efforts and achievements revolve around the MTJ-based computational random access memory (CRAM). We have experimentally demonstrated a hardware implementation of computational random access memory for the first time. The deliverables regarding CRAM demonstration are fulfilled. We invented and benchmarked the stochastic computing - CRAM (SC-CRAM), which transforms the way of operating the same CRAM array to realize stochastic computing. More excitingly, we have successfully integrated MTJs onto the pre-ordered CMOS dies, building a large, fully integrated CRAM array, and partially fulfilling the project deliverable. This effort is ongoing, and the process is being continuously optimized as we learn from each iteration.

Through this program, the FRANCScape MTJs have been developed and are now ready for lab-to-fab transfer for CMOS+X DoD technologies for rad-hard memory, computing and AI applications.

As a notable accomplishment, the advanced MTJ prototypes are being used as metrological samples for metrology work at NIST. These advanced MTJs are challenging for commercial MTJ measurement tools due to their unique demands for high measurement frequencies and large applied magnetic fields, which are necessary for benchmarking the non-volatile energy barrier and dynamic loss of MRAM bit cells. Similarly, emerging commercial MTJs are presenting comparable demands on measurement capacities. We are leveraging DARPA-funded advanced MTJ prototypes to meet these challenges in-house.

This project has established a foundation for potential collaborations to transfer the developed FRANCScape MTJs to U.S. semiconductor foundries and DoD customers (like Honeywell,

Raytheon, etc.), focusing on tapeout CRAM chips for new DoD programs in radiation-hard non-volatile memory and energy-efficient edge in-memory computing.

Discussions for an EEI program to facilitate the transfer of these technologies can be further explored.

2 INTRODUCTION

2.1 Fundamentals of Magnetic Tunnel Junctions (MTJs)

The core structure of MTJ contains a non-magnetic tunnel barrier sandwiched by two ferromagnetic electrodes. The magnetization directions of the electrodes can be parallel (P) or antiparallel (AP) with each other. The resistance of an MTJ is higher in the AP state and lower in the P state due to the spin-dependent tunneling of electrons[1], [2]. Due to its switchable junction resistance, an MTJ can function as a memory element (magnetic random access memory, MRAM) to store a binary bit. For example, AP stands for data '1' and P for data '0'. A large ON/OFF ratio, referred to as tunnel magnetoresistance (TMR) ratio for MTJs, is sought after for a higher signal-to-noise ratio and lower operating energy.

The switching between the two states, i.e., the write operation of an MTJ device, is achieved by reversing the magnetization of one MTJ electrode (referred to as the free layer), while maintaining the magnetization of the other electrode (thus, it is known as the reference layer or fixed layer). Conventionally, this switching can be performed by applying an external magnetic field that is strong enough to reverse the magnetization of the free layer but not the reference layer. The requisite magnetic field is produced by currents and thus a complex structure is necessary. Besides, the strength of this magnetic field is inversely proportional to the area of the free layer, inhibiting the miniaturization of the MTJ dimension. In addition, extra measures must be taken to prevent the selected memory element's switching field from interfering with unselected neighbors, particularly when densely arranged. Consequently, the scalability of magnetic field-driven MRAM is limited to approximately 90 nm[3] and fails to meet the increasingly high demands for memory storage density. As a result, multiple scalable MTJ switching mechanisms, such as spin-transfer torque (STT), spin-orbital torque (SOT), voltage-controlled magnetic anisotropy (VCMA), and innovation in new magnetic materials, have been developed to break through this size limit. STT-MRAM is a commercial product using 30 nm – 40 nm diameter MTJ features integrated with 22 nm CMOS as a discrete memory chip and is sampling as an add-on to 14 nm FinFET technologies. The other switching mechanisms are on the cusp of commercialization and are considered research hotspots within the research arms of many foundries and international research consortia.

Despite the variety of MTJ switching mechanisms, the energy barrier between the two states must be overcome to complete the switching. Reducing this energy barrier is highly preferable for ultrafast and energy-efficient MTJ switching. In this regard, perpendicular MTJs (pMTJs), whose electrodes manifest perpendicular magnetic anisotropy (PMA), are believed to outperform their in-plane counterparts (iMTJs) because the shape anisotropy term does not contribute to pMTJs' switching energy barrier[4]. Nowadays, pMTJs are considered of more technical relevance and can be classified based on the origin of PMA into two types: bulk PMA MTJs and interfacial PMA MTJs.

2.2 MTJs with Bulk Perpendicular Magnetic Anisotropy (PMA)

2.2.1 MTJs with Magneto-crystalline PMA

To maintain sufficient thermal stability for non-volatile data storage at downscaled technical nodes, a strong PMA is required. Additionally, a weak magnetic damping is essential to minimize high-frequency magnetic dissipation for low-energy operation. Bulk PMA materials with magneto-crystalline anisotropy, such as Mn-based Heusler alloys[5], [6] and $L1_0$ FePd, are promising candidates for satisfying these requirements. Compared with the Mn-based Heusler alloys, $L1_0$ FePd possesses very attractive properties including a large K_u ($1\text{--}2\text{ MJ/m}^3$)[7], a low α (~ 0.002)[8], and a low processing temperature down to $200\text{ }^\circ\text{C}$ [9]. With the large magneto-crystalline PMA, as shown in **Figure 2.1(A)**, $L1_0$ FePd-based pMTJs are expected to present good non-volatility even at a deeply downscaled device size of 5 nm .

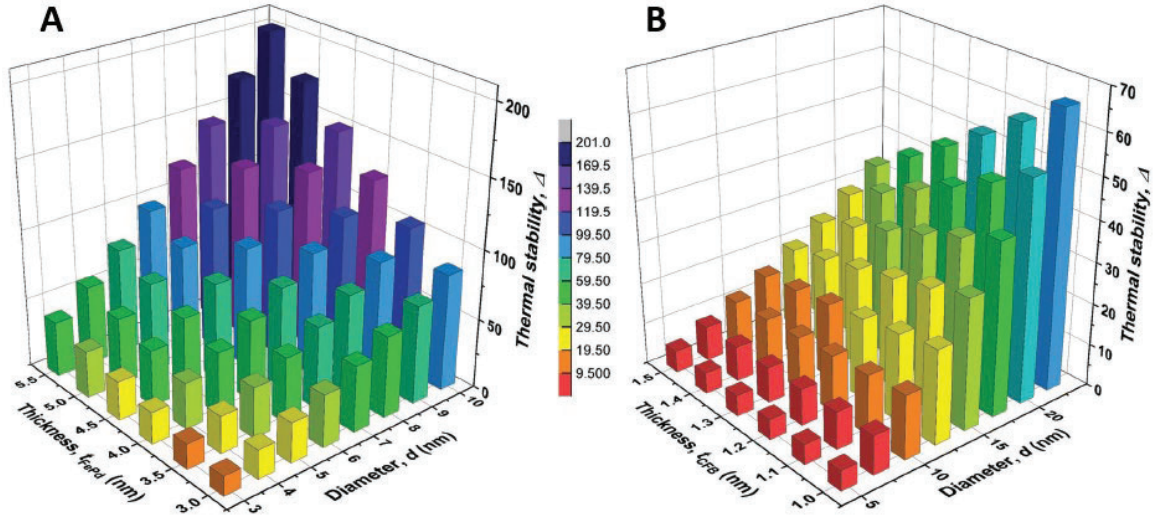


Figure 2.1 Thermal Stability Scales as a Function of Film Thickness and Device Diameter[10]

(A) $L1_0$ FePd can sustain a thermal stability of exceeding 60 at a moderate thickness of $\sim 5\text{ nm}$ while downscaling to a 5 nm device size. **(B)** In the thickness regime where interfacial PMA makes CoFeB perpendicularly magnetized ($\sim 1.3\text{ nm}$), thermal stability is always lower than 20 at $<10\text{ nm}$ device sizes. The parameters of CoFeB and $L1_0$ FePd are assumed to be $K_{i,CFB} = 1.5\text{ mJ m}^{-2}$, $K_{b,FePd} = 1.7\text{ MJ m}^{-3}$, and saturation magnetization $M_{S,CFB} = 1.2\text{ MA m}^{-1}$ and $M_{S,FePd} = 1.1\text{ MA m}^{-1}$, respectively.

Importantly, a novel MTJ switching mechanism, voltage-controlled exchange coupling (VCEC), has been reported[11] based on $L1_0$ FePd pMTJ devices. Compared to current-driven switching mechanisms like STT and SOT, VCEC is driven by voltage and thus the Joule heating is negligible, indicating an ultrahigh energy efficiency as illustrated in **Figure 2.2(A)**. In addition, as shown in **Figure 2.2(B)**, the layout area of VCEC-MRAM is smaller or comparable to the best STT-MRAM with only one associated transistor. In contrast to VCMA, another voltage-driven MTJ switching mechanism, whose switching process is completed in a dynamical manner, VCEC manifest the switching determinism necessary for practical memory and logic applications.

Despite the outstanding magnetic properties of $L1_0$ FePd and the promising VCEC switching mechanism reported with $L1_0$ FePd-based pMTJ devices, two major challenges must be overcome before $L1_0$ FePd-based MRAM could be commercially relevant: 1) Pd, as a diffusing

element, can easily diffuse into the adjacent functional layers upon post-annealing and decrease the TMR; 2) The PMA of L_{10} FePd relies on a (001) texture, which is usually hard to form well on industry-ready Si wafers.

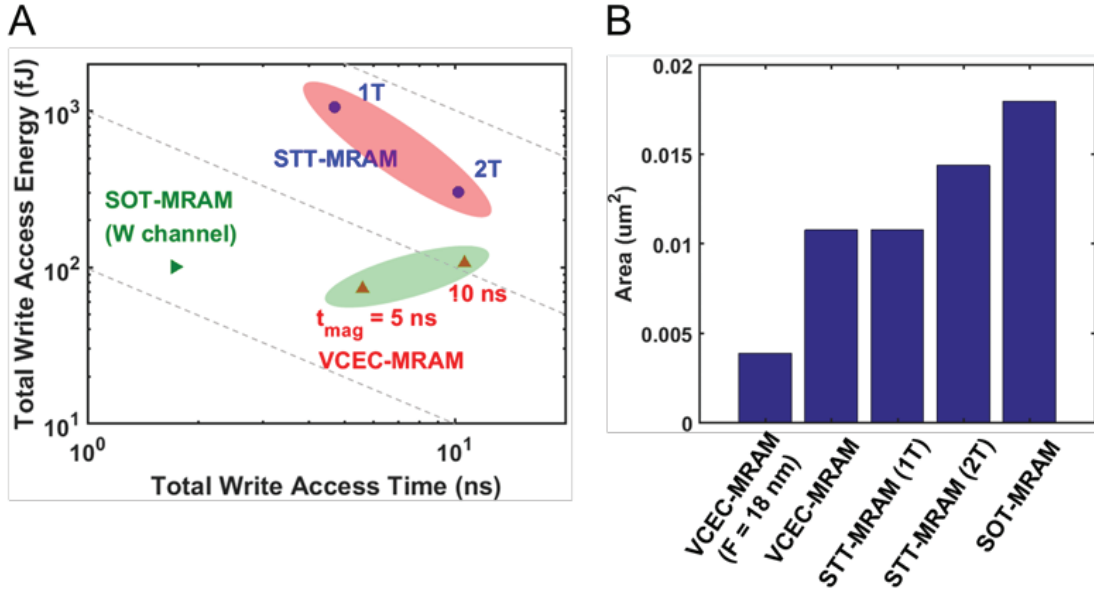


Figure 2.2 Performance Benchmarking of Spintronics Memories

(A) The write performance of the VCEC-MRAM compared to STT-MRAM and SOT-MRAM under a 1000x32 bit array. In the simulation, pMTJs are used for STT-MRAM simulation and iMTJs are used for SOT-MRAM. The magnetization switching time of VCEC-MTJ is hypothetically assumed to be 5 ns and 10 ns. (B) The comparison of the layout area for various memories devices including VCEC-MRAM, STT-MRAM, and SOT-MRAM. The half-metal pitch (F) of all the devices is 30 nm corresponding to the 15 nm CMOS technology node. For VCEC-MRAM, the case of $F = 18\text{ nm}$ (corresponding to the 7 nm CMOS technology node) is also considered because of its scalability due to the high PMA and low write currents of the pMTJ devices. For the STT-MRAM, the cases with one and two access transistors are considered.

2.2.2 MTJs composed of Co-based Multilayers

As mentioned in the last section, the PMA of L_{10} FePd relies on its (001) texture. To ensure a good L_{10} -phase crystallinity with desired (001) texture, only a few special substrates, such as MgO(001) and SrTiO₃(001)[8], are suitable. These substrates are costly, hard to obtain in large sizes, and most severely, incompatible to the monolithic integration with semiconductor microelectronics. The limit also applies to many other magneto-crystalline materials. As a promising alternative, Co-based multilayers that possess strong PMA, such as Co/Pd and Co/Pt multilayers, can be easily fabricated onto Si wafers and have already served in industry for decades as hard magnets. It is worth mentioning that the PMA of Co-based multilayers actually originates from the Co/Pd or Co/Pt interfaces[12]. However, as the repeats of these multilayers can be customized and the volume that contributes to PMA almost covers the entire multilayer stack, it is a common practice to take them as quasi-bulk PMA materials.

Though the fabrication of Co-based multilayers are much easier compared to magneto-crystalline PMA materials, such as L_{10} FePd, they also contain diffusing elements like Pd and Pt. In

addition, Co-based multilayers often present a (111) texture that is conflict with the (001) texture of the CoFe(B) electrodes and the MgO tunnel barrier. Therefore, it is necessary to insert an ultrathin spacing layer that can break the texture and diffusion from the multilayer but can still bridge its magnetic connections to the adjacent CoFeB layer.

2.3 MTJs with Interfacial PMA

The perpendicular magnetic anisotropy (PMA) can be found in bulk material (like in L1₀ ordered (Co, Fe)-Pt alloys) and ultrathin films. It is a kind of magnetocrystalline anisotropy (MCA) that has the easy axis pointing along a direction of low crystalline symmetry, which can be engineered to point out of the film's plane. PMA emerges the spin-orbit coupling (SOC) interaction

$$H_{SO} = \xi \mathbf{S} \cdot \mathbf{L}$$

In bulk PMA materials, specific crystal structures rearrange electron states (bands) with different orbital characteristics, making the state with PMA energetically favorable. The same effect can happen in ultrathin films. However, the main cause of PMA in the latter case is the reduced dimension that can make the band structure of orbitals go through dramatic changes.

The most studied and technologically relevant PMA in thin films comes from Fe/MgO or CoFeB/MgO systems. Theoretically, as SOC is much smaller than the band energy, the crystal field splitting, and exchange coupling, the MCA energy can be calculated using second-order perturbation theory[13]–[15],

$$E_{MCA} = E^{\parallel} - E^{\perp} = \frac{\xi^2}{4} \sum_{o,u} \frac{|\langle \psi_o | L_z | \psi_u \rangle|^2 - |\langle \psi_o | L_x | \psi_u \rangle|^2}{\varepsilon_u - \varepsilon_o},$$

where ψ_o and ψ_u are unperturbed wave functions for occupied and unoccupied states with energies ε_o and ε_u . E_{MCA} is defined as the energy difference between magnetization pointing in-plane and out-of-plane, a positive value indicates a perpendicular easy axis. We can also make the approximation to use only the minority-spin states as exchange splitting in Fe is quite large, and the mix between unoccupied minority-spin states to occupied majority-spin state is very small.

First principal calculation using the full-potential linearized augmented plane wave (FLAPW) method reveals elevation of energy in d_{z^2} orbital and the emergence of the O minority-spin peak at the same location. The enhancement of PMA is due to this orbital rearrangement and charge transfer between Fe and O.

Since the ultrathin Fe film can effectively avoid the restrictions put by the screening effect, a large electric field can be applied through the whole film and the charge transfer mechanism described in the last paragraph can be further enhanced. Hence, a voltage-controlled PMA mechanism can also be expected.

Perpendicular Fe/MgO, Fe/MgO/Fe MTJs or systems with similar structures have been successfully fabricated[16] and the voltage control of their magnetic property has been demonstrated by various groups[17]–[20].

From the discussion above, we can see the PMA in Fe/MgO system originates from the interface, a relatively thick FM layer will result in the easy axis change from perpendicular to in-plane, hence this magnetic anisotropy “interfacial perpendicular magnetic anisotropy” (iPMA). The

interface is of crucial importance as the PMA and the voltage control all depend heavily on its configuration. Most work conducted in this project is based on the CoFeB/MgO/CoFeB magnetic tunnel junctions with PMA. We have shown the interface plays an indispensable role in bringing out some of the most interesting phenomena in MTJs, including record low switching energy in our devices.

2.3.1 MTJs composed of Perpendicularly Magnetized CoFeB

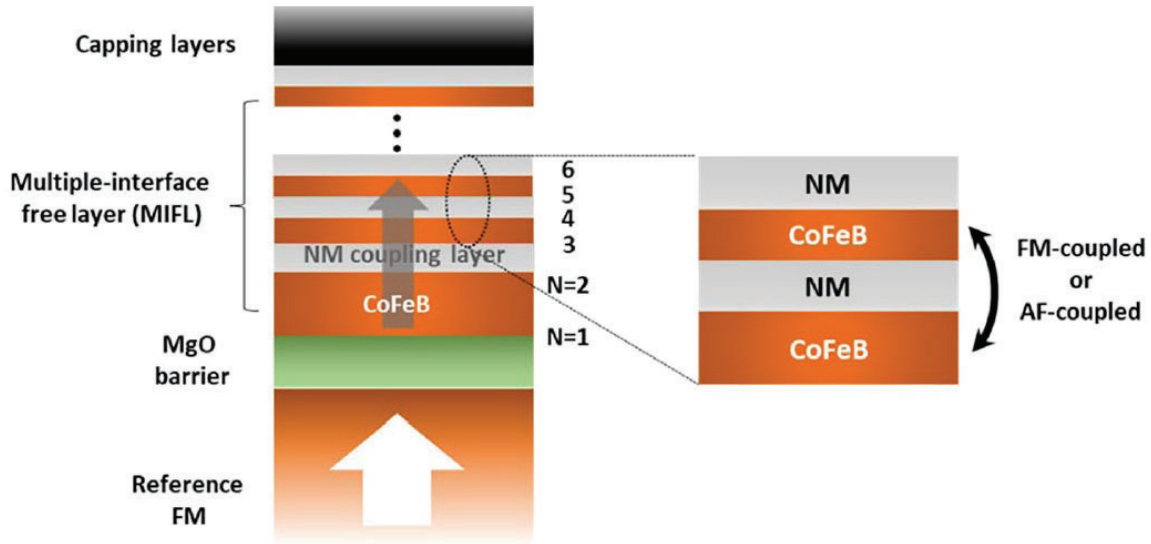


Figure 2.3 Schematic Representation of a pMTJ with Multiple-interface Free layers (MIFLs), where a Number of CoFeB Layers are coupled Through NM Layers to Function as a Single Magnetic Layer

The CoFeB-MgO materials system is dominant in the modern MTJ studies. As-deposited CoFeB is amorphous and can crystallize during post-annealing by solid-state epitaxy[21], forming sharp CoFe(B)-MgO interfaces and enabling large TMR via significant spin-dependent transmission of Δ_1 Bloch waves[22], [23]. Thus, the CoFeB-MgO system quickly became the mainstream and then dominated the development of pMTJs in the last decade after the discovery of its interfacial PMA[16]. Despite these encouraging advantages, the PMA of CoFeB stems from the Fe-O bonding at the CoFeB/MgO interface[24] and thus has an interfacial origin. The thickness of perpendicularly-magnetized CoFeB layers is limited to about 1.3 nm. Further increasing the CoFeB layer thickness can result in an in-plane magnetic anisotropy. In particular, as shown in **Figure 2.1(B)**, when the free layer of a pMTJ is a single CoFeB layer with interfacial PMA, the areal perpendicular energy density is usually 1–2 mJ/m², which can hardly provide enough retention when the junction size is below ~15 nm.

To break this size limit, as shown in **Figure 2.3**, one could fabricate multiple CoFeB/non-magnet (NM) interfaces to enhance the overall PMA of the free layer, where both the CoFeB and NM are limited to very thin thicknesses (~1 nm). However, as all the functional layers are ultrathin and can significantly influence each other via interlayer exchange coupling (IEC), stray field, roughness, texture, and inter-diffusion, the material selection and process optimization are non-trivial. Particularly, obtaining both strong PMA and high TMR is very challenging.

2.4 Advanced MTJs

2.4.1 Synthetic Antiferromagnetic (SAF) Free Layers

As previously mentioned, reducing the switching current density and increasing the switching speed are central to the development of next-generation MTJ-based memory and logic devices. In recent years, many researchers predicted that ultralow current density, ultrahigh-speed switching could be achieved with pMTJs that employ SAF free layers, or SAF-pMTJs [25]–[27]. In a SAF stack, two ferromagnetic layers are coupled antiferromagnetically through an NM spacer via the Ruderman-Kittel-Kasuya-Yosida (RKKY) interaction. Indeed, faster switching speeds of SAF-pMTJs were demonstrated in experiments in 2021[28]. Early in 2013, Yoshida *et al.* developed a SAF free layer composed of a CoFeB layer, Co/Pd multilayers, and a Ru/Ta bilayer spacer [29]. They demonstrated that, by utilizing the bilayer-mediated SAF free layer, the thermal stability could be improved by ~64% without increasing the switching current.

In addition to the performance gain in thermal stability and switching speed, a SAF free layer is also a pre-requisite of the VCEC switching mechanism introduced in 2.2.1. As shown in **Figure 2.4**, the $L1_0$ FePd layer and the bottom CoFeB layer are coupled through a bilayer spacer (colored by yellow and purple). By applying a voltage across the MTJ pillar, the IEC mediated by the spacer could be tuned from antiferromagnetic to ferromagnetic and vice versa. As a result, the magnetization direction of the bottom CoFeB layer can be switched back and forth without introducing Joule heating.

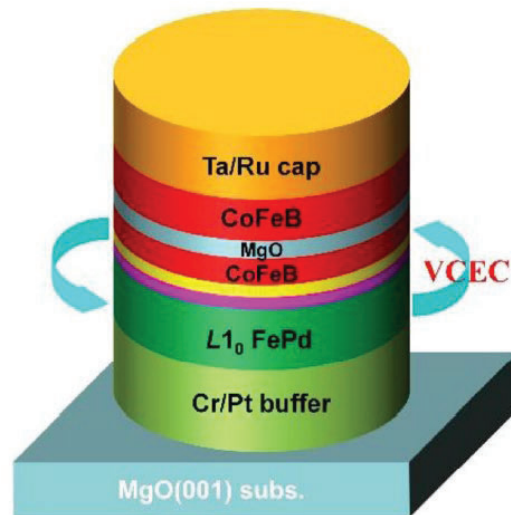


Figure 2.4 Schematic Illustration of the Stack Structure of a $L1_0$ FePd-based VCEC-pMTJ Sample

2.5 Theoretical Explorations

2.5.1 Performance Evaluation of Magnetic Tunnel Junction based on Magnetic Weyl Semimetal

The pursuit of dramatically higher MTJ performance in terms of energy-delay product may require new ideas in materials or devices. Our two works summarized here tackle this

fundamental problem through innovation in materials. We explored how the recently discovered new class of magnetic Weyl semimetals might offer advantages in terms of the magnitude of spin-transfer torques and the tunneling magnetoresistance ratio (TMR) in these novel devices[30], [31].

2.5.2 Interlayer Exchange Coupling on Magnetic Tunnel Junctions

The possibility of reducing the switching energy of perpendicular MTJ devices by means of an out-of-equilibrium interlayer exchange coupling interaction, as suggested by an experimental result from **J.-P. Wang's** group, has motivated the exploitation of the impact of such an effect whenever VCMA and regular spin-transfer torques are also present. The work summarized here tackles this problem[32].

2.5.3 Investigation of Unconventional SOT on Magnetization Switching

The lack of certain crystalline symmetries in strong spin-orbit-coupled nonmagnetic materials allows for the existence of unconventional spin Hall responses, with electrically generated transverse spin currents possessing collinear flow and spin directions. The injection of such spin currents into an adjacent ferromagnetic layer can excite magnetization dynamics via unconventional spin-orbit torques, leading to deterministic switching in ferromagnets with perpendicular magnetic anisotropy, as schematically shown in the left panel of **Figure 2.5**. Motivated by experimental results from **J.-P. Wang's** group and the potential of SOT for low energy magnetization switching, **T. Low's** group investigate the switching efficiency of SOT with multi-spin components. We numerically studied what it takes to improve switching efficiency of perpendicular MTJs by taking advantage of a new switching mechanisms based on SOT due to multi spin components. In summary, a systematic study of the simultaneous impact of conventional and unconventional SOTs on the dynamics of a perpendicular magnetization in ferromagnetic/non-magnetic bilayers is presented.

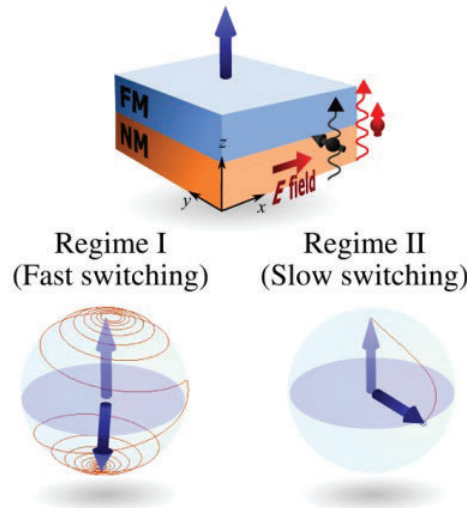


Figure 2.5 Bilayer Heterostructure comprised of a Ferromagnet (top layer) and a Strong Spin-orbit-coupled NM Layer with a Single-crystal yz Mirror Plane

An electric field applied along the x direction generates SOTs on the perpendicular magnetization due to conventional and unconventional spin Hall effects. The interplay between CSOT and USOT leads to the existence of distinctive dynamical regimes. The bottom figures illustrate dynamical regimes where full switching and full in-plane damping of the magnetization are expected.

2.5.4 Domain Wall Motion in Asymmetric MTJ

Control of magnetic domain wall (DW) motion in nanowire-like structures using current injection has been of interest for memory applications due to promising qualities of low power consumption, non-volatility, and neuromorphic device capability. **T. Low's** group has also dedicated efforts to study the domain wall motion induced by vertical current flow in asymmetric magnetic tunnel junctions, as shown in the top panel of **Figure 2.6**.

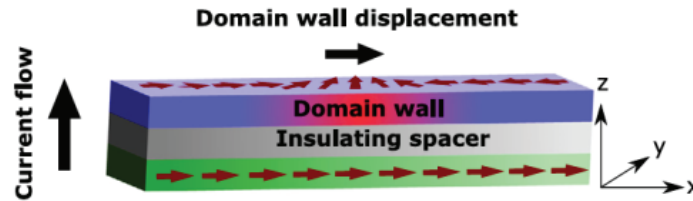


Figure 2.6 We Consider a Magnetic Tunnel Junction where an in-plane DW Displacement in the Free Layer is Induced by a Vertical Current Flow

2.5.5 Spin Hall Conductivity in Chiral Materials

The efficient generation of spin currents is a highly sought-after goal within the spintronics community. Of particular interest is the intrinsic spin Hall effect (SHE), which relies on the spin-orbit coupling (SOC) of a material, independent of any impurity scattering mechanisms. Motivated by experimental work, we studied new material systems supporting large spin Hall effects, focusing on topological chiral materials pertaining to space-group 198[33].

2.6 Computational random-access memory (CRAM)

2.6.1 CRAM concept, working principle, and features

Recent advances in machine intelligence[34], [35] for tasks such as recommender systems[36], speech recognition[37], natural language processing[38], and computer vision[39], have been placing growing demands on our computing systems, especially for implementations with artificial neural networks. A variety of platforms are used, from general-purpose CPUs and GPUs[40], [41], to FPGAs[42], to custom-designed accelerators and processors[43]–[46], to mixed- or fully- analog circuits[47]–[53]. Most are based on the Von Neumann architecture, with separate logic and memory systems. As shown in **Figure 2.7a**, the inherent segregation of logic and memory requires large amounts of data to be transferred between these modules. In data-intensive scenarios, this transfer becomes a major bottleneck in terms of performance, energy consumption, and cost[54]–[56]. For example, the data movement consumes about 200 times the energy used for computation when reading three 64-bit source operands from and writing one 64-bit destination operand to an off-chip main memory[54]. This bottleneck has long been

studied. Research aiming at connecting logic and memory more closely has led to new computation paradigms.

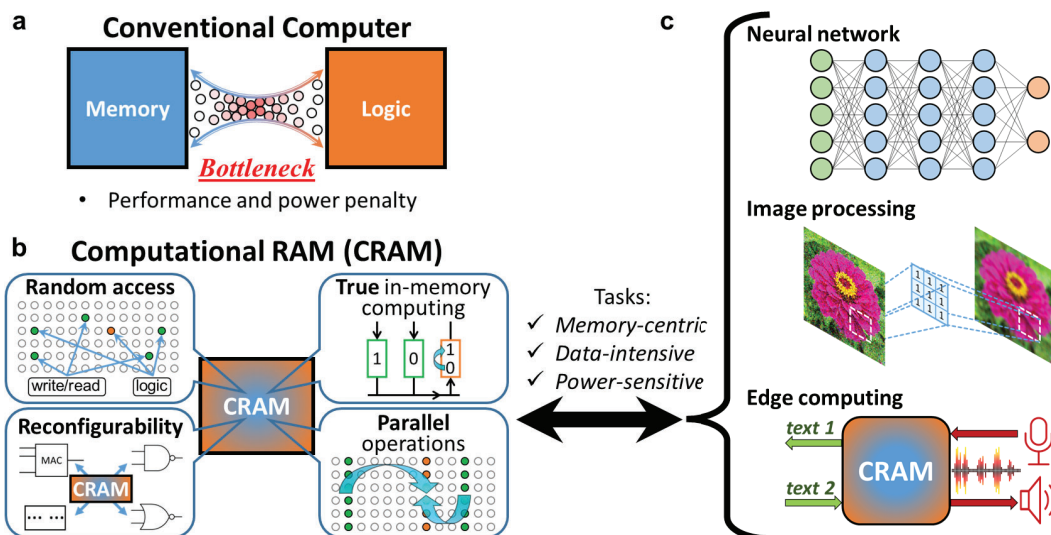


Figure 2.7 Illustrations of CRAM Concept, Features, and Potential Applications[57]

a, b Compared to a conventional computer architecture (a), which suffers from the memory-logic transfer bottleneck, CRAM (b) offers significant power and performance improvements. Its unique architecture allows for computation in memory, as well as, random access, reconfigurability, and parallel operation capability. c The CRAM could excel in data-intensive, memory-centric, or power-sensitive applications, such as neural networks, image processing, or edge computing (c).

Promising paradigms include “near-memory” and “in-memory” computing. Near-memory processing brings logic physically closer to memory by employing 3D-stacked architectures[58]–[63]. In-memory computing scatters clusters of logic throughout or around the memory banks on a single chip[14–20,30–35]. Yet another approach is to build systems where the memory itself can perform computation. This has been dubbed “true” in-memory computing[70]–[76]. The computational random-access memory (CRAM)[72], [74] is one of the true in-memory computing paradigms. Logic is performed natively by the memory cells; the data for logic operations never has to leave the memory (Figure 2.7b). Additionally, CRAM operates in a fully digital fashion, unlike most other reported in-memory computing schemes[47]–[53], which are partially or mostly analog. CRAM promises superior energy efficiency and processing performance for machine intelligence applications. It has unique additional features, such as random-access of data and operands, massive parallel computing capabilities, and reconfigurability of operations[72], [74]. Also note that although the transistor-less (crossbar) architecture employed by most of the previous true-in-memory computing paradigms[70], [71], [73], [76] allows for higher density, the maximum allowable size of the memory array is often severely limited due to the sneak path issues. CRAM includes transistors in each of its cells for better-controlled electrical accessibility and, therefore, a larger array size.

The CRAM was initially proposed based on the MTJ device[72], an emerging memory device that relies on spin electronics[77]. Such “spintronic” devices, along with other non-volatile

emerging memory devices, usually referred to as “X” for logic applications, have been intensively investigated over the past several decades for emerging memory and computing applications as “beyond-CMOS” and/or “CMOS+X” technologies. They offer vastly improved speed, energy efficiency, area, and cost. An additional feature that is exploited by CRAM is their non-volatility[78]. The MTJ device is the most mature of spintronic devices for embedded memory applications, based on endurance[79], energy efficiency[80], and speed[81]. We note that CRAM can be implemented based not only on spintronics devices, but also on other non-volatile emerging memory devices.

In its simplest form, an MTJ consists of a thin tunneling barrier layer sandwiched between two ferromagnetic (FM) layers. When a voltage is applied between the two layers, electrons tunnel through the barrier, resulting in a charge current. The resistance of the MTJ is a function of the magnetic state of the two FM layers, due to the tunneling magnetoresistance (TMR) effect [82]–[84]. An MTJ can be engineered to be magnetically bi-stable. Accordingly, it can store information based on its magnetic state. This information can be retrieved by reading the resistance of the device. The MTJ can be electrically switched from one state to the other with a current due to the spin-transfer torque (STT) effect [85], [86]. In this way, an MTJ can be used as an electrically operated memory device with both read and write functionality. A type of random-access memory, the STT-MRAM [87]–[90] has been developed commercially, utilizing MTJs as memory cells. A typical STT-MRAM consists of an array of bit cells, each containing one transistor and one MTJ. These are referred to as 1 transistor 1 MTJ (1T1M) cells.

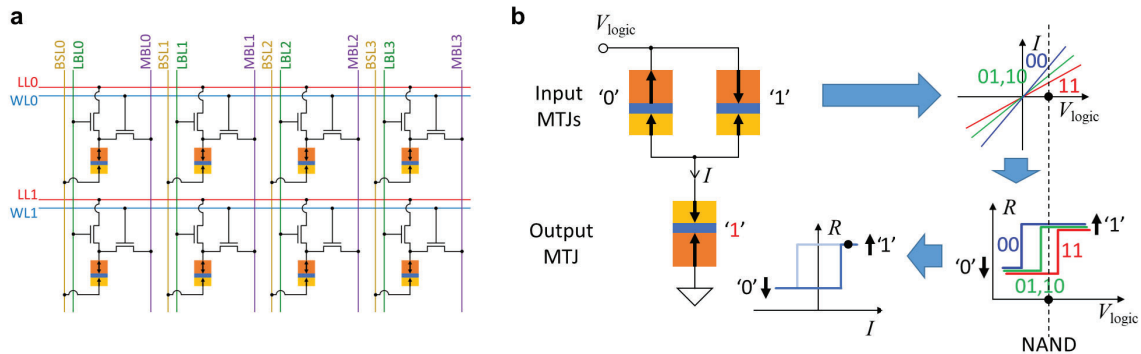


Figure 2.8 Illustrations of the CRAM Cell Architecture and the Working principle of CRAM Logic Operation [57]

a CRAM adopts the so-called 2 transistor 1 MTJ (2T1M) cell architecture. On top of the 1T1M cell architecture of STT-MRAM, an additional transistor, as well as the added logic line (LL) and logic bit line (LBL), allow the CRAM to perform logic operations. During a CRAM logic operation, the transistors and lines are manipulated to form an equivalent circuit, as shown in **b**.

Although CRAM can be built based on various emerging memory devices, we use MTJs and MTJ-based CRAM as an example for illustration purposes. **b** The working principle of CRAM logic operation, the VCL, utilizes the thresholding effect that occurs when switching an MTJ and the TMR effect of the MTJ. With an appropriate V_{logic} amplitude, the voltage is translated into different currents flowing through the output MTJ by the TMR effect of the input MTJs. Whether the output MTJ switches or not is dependent on the state of the input MTJs.

A typical CRAM cell design, as shown in **Figure 2.8a**, is a modification of the 1T1M STT-MRAM architecture[91]. The MTJ, one of the transistors, word line (WL), bit select line (BSL), and memory bit line (MBL) resemble the 1T1M cell architecture of STT-MRAM, which allows the CRAM to perform memory operations. In order to enable logic operations, a second transistor, as well as a logic line (LL) and a logic bit line (LBL), are added to each memory cell. During a logic operation, corresponding transistors and lines are manipulated so that several MTJs in a row are temporarily connected to a shared LL [74]. While the LL is left floating, voltage pulses are applied to the lines connecting to input MTJs, with that of the output MTJ being grounded. The logic operation is based on a working principle called voltage-controlled logic (VCL)[92], [93], which utilizes the thresholding effect that occurs when switching an MTJ and the TMR effect of MTJ. As shown in **Figure 2.8b**, when a voltage is applied across the input MTJs, the different resistance values result in different current levels. The current flows through the output MTJ, which may or may not switch its state, depending on the states of the input MTJs. In this way, basic bitwise logic operations, such as AND, OR, NAND, NOR, and MAJ, can be realized. A unique feature of VCL is that the logic operation itself does not require the data in the input MTJs to be read out through sense amplifiers at the edge of the array. Rather, it is used locally within the set of MTJs involved in the computation. This is fundamentally why the CRAM computation represents true-in-memory computing: the computation does not require data to travel out of the memory array. It is always processed locally by nearby cells. We note that this concept would also work with other two-terminal stateful passive memory devices, such as memristors. Accordingly, a CRAM could be implemented with such devices. A CRAM could also be implemented with three-terminal stateful devices, such as spin-orbit torque (SOT) devices. This could result in greater energy efficiency and reliability[94]. Although devices with progressive or accumulative switching behavior, such as spintronic domain wall devices[95], [96], can be adopted as well, CRAM would otherwise work best if adopting bi-stable memory devices with strong threshold switching behavior. As an oversimplified speculation, the performance comparison between CRAMs implemented by various emerging memory devices is expected to roughly follow the comparison between these for memory applications, since CRAM utilizes memory devices in similar manners as in memory applications. For example, a CRAM implemented based on MTJs should be expected to offer high endurance and high speed. Also, generally, a CRAM logic operation should consume energy comparable to the energy consumption of a memory write operation, for the same emerging memory device operating at the same speed. However, a careful case-by-case analysis is necessary for CRAMs implemented by each emerging memory device technology. Also note that we do not show a specific circuit design of CRAM peripherals because CRAM does not require significant circuit design change in sense amplifiers or peripherals compared to 1T1M STT-MRAM. And these in the STT-MRAM are already common and mature. Lastly, the true in-memory computing characteristic of CRAM is limited to within a continuous CRAM array: any computation that requires access to data across separate CRAM arrays will require additional data access and movement. The size of an array is ultimately limited by parasitic effects of interconnects[97]. However, these limitations are true for all other in-memory computing paradigms. CRAM is not at any disadvantage in this scenario.

2.6.2 CRAM Benefits

On top of the potential performance benefits that the emerging memory devices bring, at circuit and architecture level, CRAM fundamentally provides several benefits (**Figure 2.7b**): (1) the

elimination of the costly performance and energy penalties associated with transferring data between logic and memory; (2) random access of data for the inputs and outputs to operations; (3) the reconfigurability of operations, as any of the logic operations, AND, OR, NAND, NOR, and MAJ can be programmed; and (4) the performance gain of massive parallelism, as identical operations can be performed in parallel in each row of the CRAM array when data is allocated properly. Based on analysis and benchmarking, CRAM has the potential to deliver significant gains in performance and power efficiency, particularly for data-intensive, memory-centric, or power sensitive applications, such as bioinformatics[74], [98], [99], image[100] and signal[101] processing, neural networks[100], [102], and edge computing[103] (**Figure 2.7c**). For example, a CRAM-based machine-learning inference accelerator was estimated to achieve an improvement on the order of $1000\times$ over a state-of-art solution, in terms of the energy-delay product[104]. Another example shows that CRAM (at the 10 nm technology node) consumes 0.47 μJ and 434 ns of energy and time, respectively, to perform an MNIST handwritten digit classifier task. It is $2500\times$ and $1700\times$ less in energy and time, respectively, compared to a near-memory processing system at the 16 nm technology node[100].

2.6.3 Opportunity and Challenge for CRAM

CRAM experimental demonstration. While the advantages of CRAM, such as energy, delay, bandwidth, and scalability, are well established by multiple benchmarking works[74], [94], [106], [98]–[105], there lacks an experimental proof-of-concept demonstration. Such a demonstration is an essential step to incentivize both the research community and the industry to be more devoted to making fully integrated and larger CRAM arrays. Also, such an experimental demonstration will provide a realistic view of CRAM as well as a research platform for CRAM studies and explorations. Lastly, the error rate of CRAM is an application-critical metric. But it is yet to be investigated thoroughly. With such experimental demonstration, the error rate of CRAM can be realistically observed, studies, modelled, and projected, providing invaluable guidance for making larger CRAM arrays by CMOS integration.

With the support of this program, we took on this challenge and successfully demonstrated a fully functional CRAM array hardware. Detailed results and discussion can be found in **Section 4.4.1**.

Stochastic Computing-CRAM. Stochastic computing (SC) is a probabilistic scheme that has been studied intensively in recent years[107]. SC is an attractive method for several unconventional computing applications for two key reasons. One is that it is immune to the effects of thermal noise and random bit-flips, which allows for smaller device sizes[107]. The second is that complex functions that are common in numerous neuromorphic computing tasks, such as hyperbolic tangent, exponential functions, and square root, can be performed efficiently in SC using a small number of logic gates[107]–[111].

One of the major shortcomings of SC is the hardware area costs required for stochastic bit-stream generation (SBG). CMOS-based random number generators (RNGs) are typically linear shift feedback registers (LSFRs) or ring oscillators. When these circuits are used in SC-based networks, the hardware for SBG can consume up to 80% of the total circuit area and 80% of the total energy consumption [8]. Therefore, despite the low circuit area required for computation in SC, there may be little to zero reductions in the total circuit area and total energy consumption. One attractive solution to overcome these shortcomings of SC is to replace CMOS-based RNGs

with spintronics-based RNGs [112]–[117]. This is because a single magnetic tunnel junction (MTJ) can be used as a true RNG (TRNG) with a tunable probability[118]–[120]. While these solutions reduce the area and energy costs for SBG, the hardware and energy consumption for SBG still make up a majority of the total circuit area and total energy consumption. Furthermore, implementing spintronic-based RNGs in SC does not reduce the overall delay costs since SBG steps are still performed separately from computation steps.

In this study, we present an SC scheme that uses MTJ-based hardware where SBG and computation steps are completely embedded. Our method exploits the intrinsic stochasticity of MTJs for SBG and utilizes the computational random access memory (CRAM) array to efficiently implement various SC tasks. Previous studies have demonstrated that true in-memory computing can be achieved in the CRAM architecture, thus eliminating circuit area costs and energy consumption for transferring data between memory and computation circuits[74], [94], [100], [121]. However, performing neuromorphic applications that require processing large amounts of data in conventional CRAM (conv- CRAM) will still suffer large area and energy costs as well as noise sensitivity since the data are still encoded in binary bits. Our solution overcomes these shortcomings by implementing SC within the CRAM array, which we refer to as SC-CRAM. While conv-CRAM embeds memory and logic arrays, SC-CRAM embeds SBG and computation arrays. We demonstrate the effectiveness of SC-CRAM in four neuromorphic applications: local image thresholding, Bayesian inference for object location, Bayesian belief network for heart disaster prediction, and kernel density estimation.

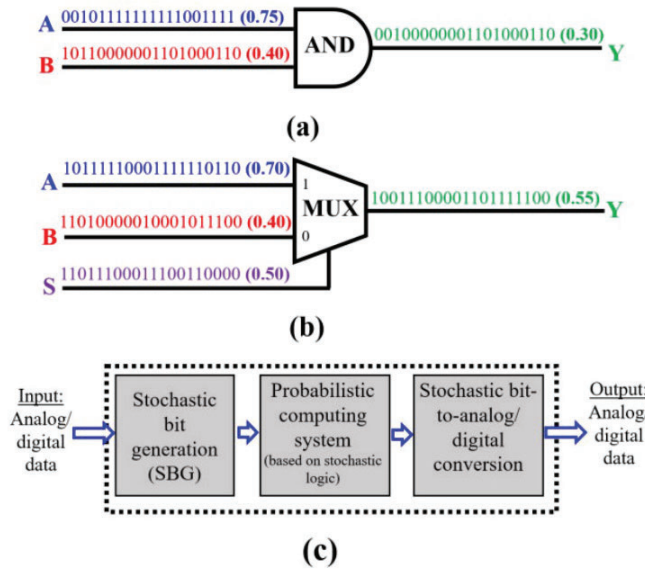


Figure 2.9 Illustration of SC Working Principles[122]

Illustration of (a) multiplication using AND logic and (b) scaled addition using MUX of stochastic bit-streams A and B. (c) Block diagram of the conventional stochastic computing scheme.

Figure 2.9c shows the overall scheduling of steps for conventional SC schemes. The first step is the SBG, where the input data (either digital or analog representation) are converted to stochastic bit-streams. This is typically done using comparators with the input data and the output of an

RNG with an output centered around 0.5 connected at the input terminals. At each clock cycle, a random number is generated and the comparator produces a “1” if the random number is greater than the input value; otherwise, it produces a “0.” The total number of clock cycles is dependent on the desired number of bits in the bit-streams (N_B). Consider a stochastic bit-stream A generated from an input value of X_A . The numeric representation of A is the probability of a bit in A being “1,” which is directly correlated with X_A . In this study, we represent bit-streams using unipolar formatting, meaning that they lie within a $[0, 1]$ range.

Data computation is performed in the second step, where various arithmetic functions, determined by the desired application, are performed on the stochastic bit-streams. It is in this step where we see the advantages of SC best exemplified. For example, **Figure 2.9a** shows how multiplication of bit-streams A and B can be performed using a single AND gate and **Figure 2.9b** shows how a single MUX can perform scaled addition on A and B . For scaled addition, a third bit-stream, S , is applied at the selector terminal of the MUX and the output is expressed as $Y \approx S \times A + (1-S) \times B$. For most applications, S is fixed at 0.5, therefore $Y \approx 0.5 \times (A+B)$. Note that scaled addition can also be done using a MAJ3 gate if one of the inputs is fixed at 0.5. More advanced arithmetic functions, such as scaled division, exponential, and hyperbolic tangent, can also be effectively implemented in SC[108], [111].

In the third and final steps, the numeric representations of the output bit-streams are determined. This is typically done using an N_B -bit digital counter. As with the input data, the output data can be represented in either analog or digital format. In this article, we will not focus on this third step since it is the same for our proposed method and conventional SC schemes.

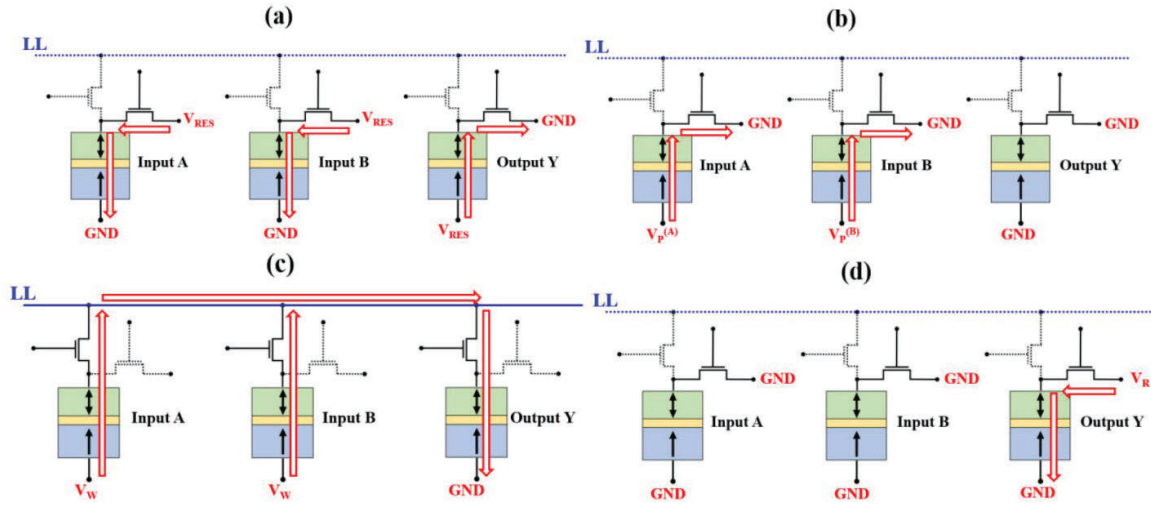


Figure 2.10 Illustration of Current Paths in SC-CRAM Operations[122]

Current paths for (a) reset, (b) perturb, (c) logic, and (d) read steps for multiplication of bit-streams A and B using AND logic in SC-CRAM.

Consider the example shown in **Figure 2.10**, where AND logic in CRAM is used to multiply inputs A and B . **Figure 2.10a** shows the current paths during the reset step, where V_{RES} initializes the input MTJs (A and B) to the P-state and the output MTJ (Y) to the AP-state. Note that Y is initialized to the AP-state to meet the output preset criteria for AND logic in CRAM.

The perturb pulses, shown in **Figure 2.10b**, are applied along the MBL only for input cells A and B , which cause the MTJs to switch probabilistically, with probabilities dependent on $V^{(A)}_P$ and $V^{(B)}_P$. Note that the values for $V^{(A)}_P$ and $V^{(B)}_P$ can be determined in three ways depending on the desired task: 1) in machine learning applications, one of the input cells can perform a synaptic function, and therefore, V_P is dependent on the synaptic weight determined using learning algorithms; 2) other tasks may require multiplication of a fixed constant, and therefore, V_P is determined to produce a probability equal to the desired constant; and 3) the most likely scenario is that V_P is dependent on the input data. For example, in image processing applications, $V^{(A)}_P$ corresponds to intensity at pixel A . It should be noted that V_P only switches the MTJ from the P-state to the AP-state since the reset pulse switches the MTJ back to the P-state each cycle.

The logic step, shown in **Figure 2.10c**, replicates the logic operation performed in conv-CRAM. The voltage pulses applied at the BSL at A and B (V_W) follow the same criteria as in conv-CRAM. During the perturb step, A and B switch probabilistically, whereas Y switches deterministically during the logic step.

The final state of Y is read at the end of each cycle with V_R , as shown in **Figure 2.10d**. V_R should be set so that it does not affect the state of Y . Alternatively, for tasks involving multiple computation stages, the steps at each computation stage can be overlapped so that the read steps at each stage prior to the final stage can perform the logic steps in the following stages. By adding a logic step in each cycle, SBG and computation are embedded within the same circuit.

CRAM Error Rate and Device-to-device Variations. One major realistic mechanism to consider is the probabilistic nature of MTJ switching. With a given current or voltage amplitude and duration, the output MTJ's final state, or whether it will switch in any instance, is uncertain. The probability of switching, however, can be measured and modeled. **Figure 2.11a** shows the statistical average of the binary final state of the output MTJ ($\langle D_{out} \rangle$), considering a two-input logic operation attempting to switch the output MTJ from '0' state to '1' state. Coincidentally in this particular case, $\langle D_{out} \rangle$ is the same as the switching probability of the output MTJ. For a given input state combination, $\langle D_{out} \rangle$ vs. V_{logic} shows a sigmoidal shape. When the input state combination changes, the corresponding curve remains in a similar shape, but is shifted and scaled differently. The margins for NOR and NAND Boolean logic are highlighted in blue and red, respectively. For example, if V_{logic} is appropriately set within the NAND margin, the logic operation will be equivalent to a NAND gate. Because input '11' will result in output of nearly '0' while the rest of input combinations will result in output of nearly '1'. And this behavior of such logic operation is what to be expected of a NAND gate. Then when taking a closer look at the $\langle D_{out} \rangle$ curves, one will see that $\langle D_{out} \rangle$ slightly deviates from being perfectly '0' or '1' for various input cases, simply due to the fact that each $\langle D_{out} \rangle$ curve is a smooth sigmoidal shape, instead of being, for example, a sharp step function. This means that there is a finite chance that the logic operation yields an incorrect result. This brings the metric of focus of this work, the accuracy or error rate of CRAM logic operation. The accuracy (error rate) of a CRAM logic operation can be simply understood as the chance of such an operation yield a correct (incorrect) result. Note that error rate exists and remains being finite above zero in every aspect of modern electronic hardware, such as digital systems of computing, memory and storage, and communications. Ideally, accuracy (error rate) of 1 (0) is desired. But in reality, accuracy (error rate) will be a value slightly less than 1 (greater than 0). From the input state dependent $\langle D_{out} \rangle$ curves shown in **Figure 2.11a**, for a particular Boolean logic, e.g., NAND, one

can obtain the error rate vs. V_{logic} , as shown in **Figure 2.11b**. The minimum of the curve is where the V_{logic} should be set at to achieve the lowest (best) error rate. To achieve a better (lower) accuracy (error rate), mainly two strategies will help significantly. Firstly, the $\langle D_{\text{out}} \rangle$ vs. V_{logic} curves are input state dependent due to the effect of tunneling magnetoresistance (TMR)[82], [84], [123] of input MTJs. Therefore, increasing the TMR ratio of MTJs is an effective way to improve error rate. Secondly, making the $\langle D_{\text{out}} \rangle$ of the output MTJ relatively sharper or narrower will improve the error rate. This is clearly shown in **Figure 2.11b** of the error rate with various pulse duration or speed, provided that the switching probability curve gets relatively narrower or sharper with increased (decreased) speed (pulse duration). Part of results and discussion in **Section 4.4.1** reveals and address this challenge in part.

Detailed results and discussion regarding the evaluation of CRAM manufacturability can be found in **Section 4.4.3**.

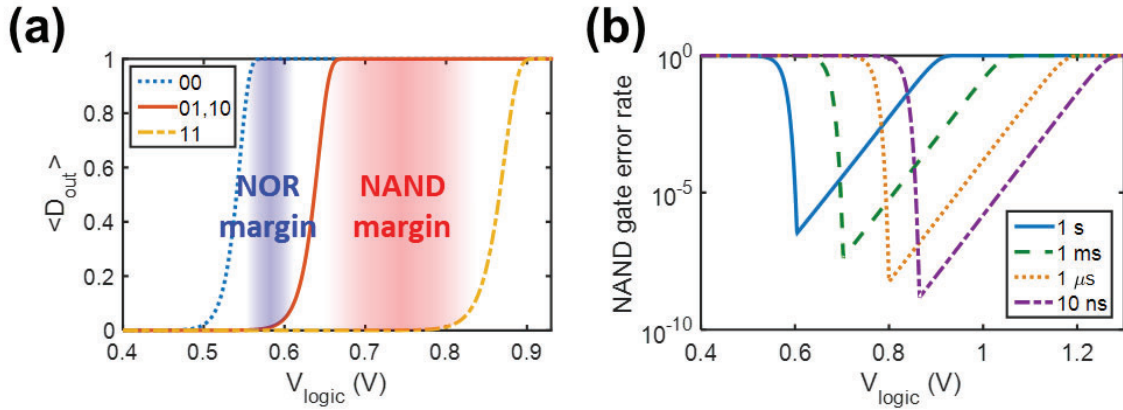


Figure 2.11 Illustrations of the Dependency of CRAM Logic Operation Error Rate on Logic Operation parameters[124]

(a) the average state of output cell vs. logic operation voltage, V_{logic} . (b) CRAM logic operation error rate of a NAND gate (operation) with various logic voltage pulse width (duration). Plots are for illustration purposes only.

Further CRAM Improvement Via Device Breakthroughs. The most fundamental or even revolutionizing level of CRAM performance improvements are rooted in advances in memory device developments. The development of high performance Spintronic memory and logic devices is the main effort of this program. It aligns with the demand of CRAM for further development. With the results and discussions presented in **Sections 4.1, 4.1.1, and 4.3**, it would be clearer that the technological relevance of CRAM is and will be present and enhanced.

2.6.4 NIST Test Vehicle and Technology Transfer Capability

Fab-Friendly Integration Interface. Researchers at NIST have been developing a test capability to rapidly prototype new types of emerging 2-terminal memories[125]–[127]. The vehicle has been developed using a commercial 180-nm foundry and received mid-production so as to facilitate direct integration in the back-end-of-line of novel memory devices in either the wafer level form factor or the die-level form factor. The interface to the underlying CMOS has a 5 micron feature size so that research facilities can best characterize their own integrations.

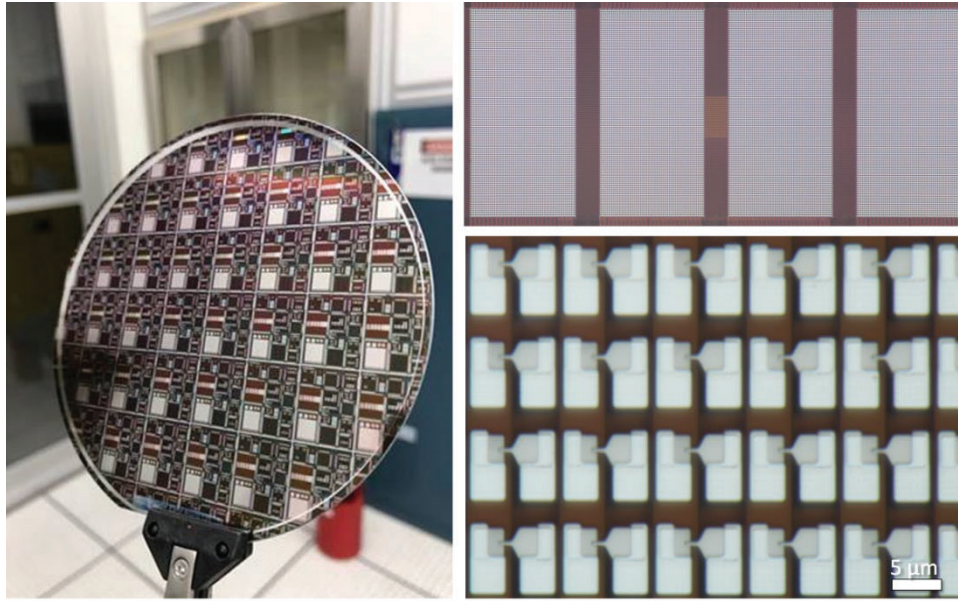


Figure 2.12 NIST Test Vehicle shown with NIST Trial Test Integration
The structure has a 5-micron feature size for contacts to the underlying CMOS so it can be easily used in research facilities. The shown 20,000 memory devices can be accessed through the CMOS.

The chip is intended to be a 20,000 device multiplexer system accessed through over 400 pins enabling both individual or parallel cell access. For this reason, the platform can be used to demonstrate CRAM operations using large numbers of individual cells.

Probe Station Level Testing. Before moving to packaging the die, NIST developed the interface and software for a sophisticated high-frequency probing interface. The new probe card iteration has over 150 50-ohm impedance high frequency connections networked through a parallelized commercial 200MHz bandwidth RF-Switch matrix. This setup allows for both parametric test of individual devices as well as parallelized testing of CRAM operations through the CMOS layers.

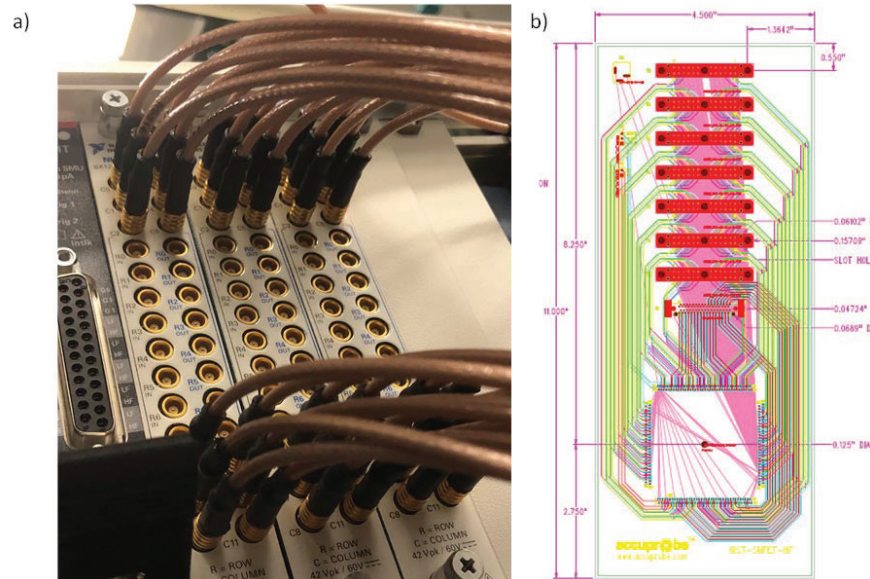


Figure 2.13 (a) Commercial 200 MHz Bandwidth High Frequency Switch Matrix
(b) Schematic design of the custom probe card assembly with 50 ohm impedance connections from the probecard assembly to the underlying chip.

Packaged chip Testing. NIST has been actively developing its packaged chip infrastructure, including hiring a dedicated engineer for the lab-to-fab testing. The current hardware, developed by NIST and collaborators, includes a 75-DAC channel board with a custom embedded Linux distribution running on a Zynq development board. NIST implemented some software updates to support CRAM operations should an appropriate set of tests identify strong candidates prior to packaging.

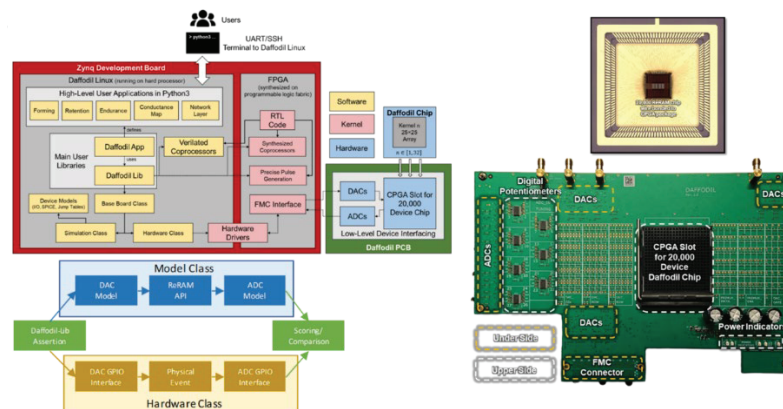


Figure 2.14 Hierarchical Diagrams of NIST Modeling Architecture and Depictions of NIST 75-channel DAC Board with Packaged memory Device Chip

3 METHODS, ASSUMPTIONS, AND PROCEDURES

3.1 Theoretical Explorations

3.1.1 Modeling Approach of Magnetic Weyl Semimetal Tunnel Junctions

We theoretically investigated the TMR in MTJs comprised of Weyl semimetal contacts. A standard four-band lattice model that describes a 3D magnetic Weyl semimetal constructed on a cubic lattice is employed to calculate the band structure of Weyl semimetal-based MTJs. The tunneling of chiral massless Weyl fermions through an insulating barrier is modelled within the context of Bardeen's transfer Hamiltonian approach[128]. In this system, electrons can tunnel through the barrier with conserved in-plane momentum. This because the translational invariance in the film plane. Following the Bardeen approach, we assume that solutions from opposite electrodes are approximately orthogonal everywhere with an exponential tail within the barrier region characterized by an energy independent decay constant.

We make no assumptions about the shape of the barrier potential, only that it varies slowly compared to the carrier wavelength and that it leads to solutions that are well-localized in the regions of the electrodes. The in-plane translational symmetry allows us to expand our solutions in a basis of Bloch states. Finally, we compute the conductance of the MTJ by means of the nonequilibrium Green's function technique. Both the pMTJ and iMTJ cases are considered and calculated.

The methodology employed in this work relied on the numerical calculation of quantum spin transport in MTJ devices. We utilized the non-equilibrium Green's function approach to capture coherent spin tunneling across a tunnel barrier. The numerical calculations were implemented in MatLab to be flexible enough to accommodate any combination of magnetic contacts or insulating layer. The electronic structure of all layers was simulated by means of tight-binding approaches that were previously shown to capture well the essential features of these exotic magnetic Weyl materials. The TMR and Spin-transfer torque calculations were processed in the Minnesota Supercomputing Institute (MSI).

3.1.2 Interlayer Exchange Coupling on Magnetic Tunnel Junctions

The methodology employed in this work relied on the numerical calculation of quantum spin transport in MTJ devices and the subsequent spin dynamics simulations. We utilized the non-equilibrium Green's function approach to capture coherent spin tunneling across a tunnel barrier. The magnetization dynamics in the presence of interlayer exchange coupling, spin-transfer torque and VCMA was captured through the Landau-Lifshitz-Gilbert equation. All numerical calculations were implemented in MatLab. The electronic structure of all layers was simulated by means of tight-binding approaches that were previously shown to capture well the essential features of these exotic magnetic Weyl materials. The TMR and Spin-transfer torque calculations were processed in the Minnesota Supercomputing Institute (MSI).

3.1.3 Modeling Approach of VCEC

The IEC of the SAF structure as a function of E-field was calculated using first-principles methods based on density functional theory (DFT). As implemented in the Vienna *ab initio* simulation package (VASP) code[129], the generalized gradient approximation (GGA) exchange-correlation potentials plus the projector augmented wave (PAW) method for the electron-ion interaction was used[130]. We constructed a SAF structure with the stack of MgO (6)/Co₂Fe₆ (8)/Ta (6)/FePd (11), as shown in **Figure 3.1**, to study IEC between Co₂Fe₆ and FePd layers through a Ta spacer, where the numbers in the parentheses indicate the atomic thickness for each layer, which corresponds to a thickness of 10.7, 8.5, 10.8, 15.7 Å for MgO, Co₂Fe₆, Ta, and FePd, respectively. This SAF structure is employed to understand the VCEC effect through DFT calculations and a quantum interference model proposed by Bruno[131]. The Co₂Fe₆ composition was used because it most closely reflects the stoichiometry of the Co₂Fe₆B₂ electrode after post-annealing.

All self-consistent calculations were performed with a plane-wave cutoff of 400 eV. The geometric optimizations were carried out without any constraint until the force on each atom is less than 0.02 eV/Å and the change of total energy per cell is smaller than 10⁻⁴ eV. The Brillouin zone *k*-point sampling was set with a 9 × 9 × 1 Γ -centered Monkhorst-Pack grids. A vacuum layer thicker than 15 Å was applied along the *z*-direction to eliminate the interaction between slabs. The lattice constant and the atomic arrangement of the SAF structure were first relaxed without considering spin. The relaxed structure with a lattice constant of 2.92 Å was then applied to study the magnetic interaction between the Co₂Fe₆ and FePd FM layers. Spin-orbit coupling effects were considered to study the magnetic anisotropy. The IEC was calculated by comparing the energy difference between ferromagnetic and antiferromagnetic couplings aligned of Co₂Fe₆ and FePd layers. An electric field is applied along the *z*-direction with the dipole corrections performed to avoid interactions between the periodically repeated images.

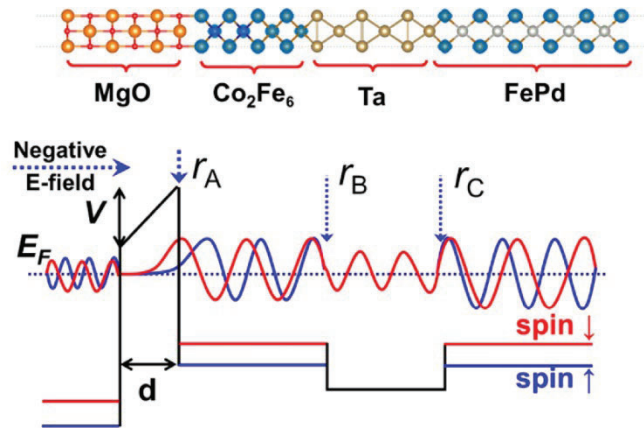


Figure 3.1 Schematic of the Simulation Structure with an SAF stack of FePd/Ta/Co₂Fe₆ (CoFe)/MgO

Here, we define the interfacial reflection coefficient r_A , r_B , and r_C corresponding to the three interfaces: CoFe/MgO, Ta/CoFe, and FePd/Ta, respectively. The ferromagnetic layers are characterized by an exchange splitting, which separates the majority and minority spin bands. The voltage is effectively dropped inside the MgO tunnel barrier, changing the reflection coefficient of electron wave functions at this interface and modulating the IEC.

3.1.4 Investigation of Unconventional SOT on Magnetization Switching

The methodology employed in this work relied on the numerical calculation of the dynamics of a magnetization subjected to multispin components SOTs. The magnetization dynamics was described through the Landau-Lifshitz-Gilbert equation. All numerical calculations were implemented in MatLab and were processed in the Minnesota Supercomputing Institute (MSI).

3.1.5 Domain Wall Motion in Asymmetric MTJ

The methodology employed in this work relied on the numerical calculation of quantum spin transport in MTJ devices and the subsequent micromagnetic simulations through the opensource software MuMax3. We utilized the non-equilibrium Green's function approach to capture coherent spin tunneling across a tunnel barrier and compute the spin transfer torque magnitudes. This information has been fed to MuMax3 from which the domain wall motion properties have been investigated. The quantum transport calculations were implemented in MatLab and were processed in the Minnesota Supercomputing Institute (MSI). The micromagnetic simulations were performed in T. Low's group servers.

3.1.6 Spin Hall Conductivity in Chiral Materials

In order to calculate transport coefficients of interest, a tight-binding description based on a basis set of maximally localized Wannier functions is extracted from our DFT results via Wannier interpolation. Our DFT calculations of the electronic structure are based on a plane-wave basis set and were performed using the Quantum Espresso (QE) suite. To ensure computational efficiency and accuracy, the projector augmented-wave (PAW) technique is used in conjunction with a fully-relativistic Perdew–Burke–Ernzerhof (PBE) exchange–correlation functional. Spin-orbit effects are turned on within QE. The k-mesh densities are chosen based on results from total ground state energy convergence tests. Bands deemed relevant for electronic transport are fitted via Wannier interpolation using the Wannier 90 (W90) program. Spin transport properties are calculated using the W90 post-processing module.

3.2 MTJ Fabrications

3.2.1 MTJ Stack Development and Characterization

At the University of Minnesota, the $L1_0$ FePd thin films and MTJ stacks of our team are deposited by using multiple ultrahigh vacuum (UHV) magnetron-sputtering systems. These sputters have their own unique features and can be used together or solely to prepare various sample thin films. As shown in **Figure 3.2(A)**, the OLV sputter includes one etching chamber (PM1) and two sputtering process chambers (PM2 and PM3) for thin film deposition on wafer substrate of 6" or 8" diameter. This system is provided with a load lock chamber (LL), which has 12 wafer slots and is used to load and unload wafers. It also has a vacuum transfer chamber (TM), which is always kept under vacuum during the normal operation. The PM1 can be used for wafer surface cleaning that has base pressure of 10^{-7} Torr. The PM2 is used for metal thin films deposition with base pressure down to 10^{-9} Torr. The PM3 is used for metal and oxide deposition that also has base pressure down to 10^{-9} Torr. Beside the hardware, the OLV sputtering system is controlled by a software and fully automatic during the normal operation and allowed to process the thin film deposition in the different chamber at the same time. By using different sputtering

powers, DC, DC pulse, and RF sputtering are capable with the OLV sputter. Co-sputtering can be easily done using the targets in the same sputtering chamber. During the deposition, the substrate temperature can be elevated up to 300°C.

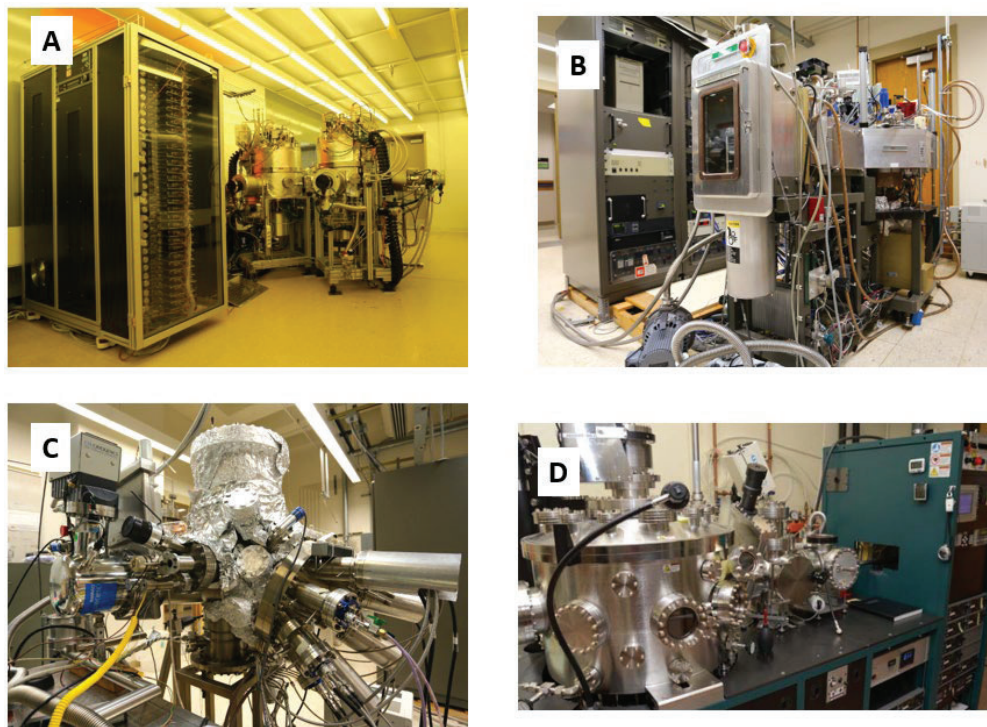


Figure 3.2 Some of the UHV Magnetron-sputtering Systems Employed for MTJ Stack Deposition at our Facilities

(A) The new cluster tool UNIVEX MULTICHAMBER (“OLV” sputter) located in the University of Minnesota’s class-100 cleanroom. (B) The Shamrock sputtering system. (C) A homemade sputter called “MBE” sputter because it was made using an old MBE chamber. (D) An AJA sputter with 12 sources.

Figure 3.2(B) shows UMN’s Shamrock sputter system, which is a special sputter tool that can be used to deposit high-quality films, including MTJ stacks. It is an automatic tool with three chambers, one load-lock chamber (CM) with 12 slots and base pressure about 10^{-7} Torr, one transfer chamber (TM) with base pressure about 10^{-7} Torr and one high vacuum process chamber (PM) with five 3-inch DC Guns and one 4-inch RF Gun and base pressure $< 5 \times 10^{-8}$ Torr. It can prepare thin films up to six-inch wafers.

Our 'MBE' sputtering system is presented in **Figure 3.2(C)**. It is an eight target sputtering system modified from an old MBE system. The system has an in-situ Pt heater, which can produce substrate temperatures up to 400 °C during deposition. The system reaches an ultra-high vacuum level of 3×10^{-8} Torr. The system design allows for both DC and RF sputtering as well as the co-sputtering of multiple targets at once.

Figure 3.2(D) shows our 12-source AJA deposition system capable of metal, oxide and nitride deposition with a base vacuum in the low 10^{-9} Torr range. Combinatorial growth of compounds with continuous concentration change and superlattices with continuous thickness change are capable. It accepts 1.5" and 2" targets with wafers up to 4". A load lock wafer transfer system is equipped. Reactive sputtering with Ar+O₂ or Ar+N₂ is also capable.

Though patterning an MTJ stack into devices is the ultimate way to verify its performance, it is possible to characterize the magnetic and electrical properties of an MTJ stack at thin film level. Our team employs multiple thin film characterization techniques to guide our research without spending time and funding in the device patterning process. For example, as shown in **Figure 3.3(A)**, we have a physical property measurement system (PPMS) equipped with a vibrating-sample magnetometer (VSM) module located in UMN facilities. By conducting VSM measurements, we can collect the magnetization response of a sample thin film (like an MTJ stack) to external magnetic fields. It is worth mentioning that our team also has multiple VSM tools in UA and NIST facilities, which significantly saved us time in sample shipment.

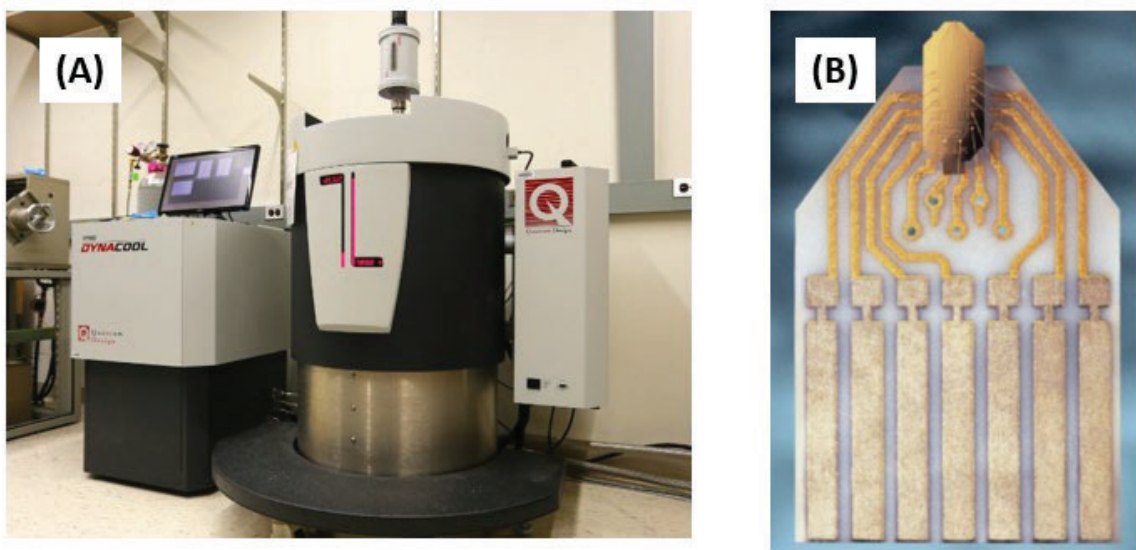


Figure 3.3 Example of the Thin Film Characterization Tools in our Team, (A) A PPMS Equipped with a VSM Module, (B) A Probe of the CIPT System

*In addition to the magnetic properties, we can also characterize the electrical properties, such as the TMR, of unpatterned MTJ thin films by using the current-in-plane tunneling (CIPT) technique[132]. By landing a CIPT probe, as shown **Figure 3.3(B)**, onto the surface of the MTJ stack, we can measure the TMR ratio, resistance-area product, major and minor R-H loops at the thin film level.*

At the University of Arizona, there are three UHV magnetron sputtering systems in the W. Wang's lab, all of them are capable of combinatorial growth of complex materials. **Chamber-A** is an AJA ACT-2600 UHV deposition tool, features a 26" chamber with 12 sources for conventional metals and oxides, with a base vacuum in the low 10^{-9} Torr. All source heads are tiltable, capable of co-deposition and it accepts 1.5" and 2" targets and wafers up to 4" in diameter. It has a loadlock wafer transfer system and 12 DC power supplies and 2 RF power supplies, capable of reactive sputtering with Ar+O₂ or Ar +N₂. **Chamber-B** is an AJA Orien-5

tool, with 5 magnetron sources configured in the sputtering down configuration. It accepts 2" targets and wafer up to 4". It also has a loadlock wafer transfer system. RF and DC biasing during deposition is possible for conventional sputtering or reactive sputtering with Ar+O₂ or Ar+N₂. **Chamber-C** is a customized 3-source UHV deposition system with two separate deposition chambers: one for high vapor pressure compounds and one for oxides, each chamber has three sources, capable co-deposition. The magnetrons are equipped with interchangeable magnet arrays for 1.5", 1.75" and 2" targets. The substrate temperature of system-C range can be from 300 to 1000 K. In addition, there is a **thermal evaporator** in W. Wang's Lab for growing conventional contact materials such as Ti and Au.

W. Wang's lab is equipped with a **Microdiffraction XRD system** (Bruker D8 Discover). It has a microfocus XRD tube that requires no cooling water, which produces a bright radiation with the X-ray spot size tunable from a 1mm down to 25 μ m. Coupled with a 2D detector, a regular θ -2 θ can be completed in minutes. This system accepts substrates 1" to 4" on a centric 1/4-Circle Eulerian Cradle and it can be programmed to perform automatic measurement over the whole wafer, providing rapid mapping for the desired properties of the films (such as crystal structure, epitaxy, thickness, and roughness). The aligned X-ray, laser and CCD camera enables point-and-shoot measurement. Another major characterization tool is a Quantum Design **DYNACOOOL 9T PPMS**, which operates from 1.9 to 400 K, with the resistivity option, high resolution horizontal sample rotator, vibrating sample magnetometer, heat capacity measurement system. It is equipped with the option of Dilution Refrigerator, enabling reaching temperature as low as 50 mK for transport and thermal measurement. For regular room temperature or high temperature (up to 1000K) magnetic measurement, the W. Wang's Lab utilizes a **Mirosense EZ9 Vibrating Sample Magnetometer (VSM)**. It is ideal for ultra-thin magnetic films and nanostructures with a 0.75×10^{-7} emu noise floor. Fast characterization can be performed by a customized Magneto-optical Kerr effect (**MOKE**) system. It has a Photoelastic modulator coupled with a lock-in amplifier, enabling monolayer detecting sensitivity down to ~ 100 μ rad, under perpendicular magnetic field > 2 T and in-plane magnetic field > 1 T. It is equipped with an XY sample holder for automatic probe of whole wafers under perpendicular field.

The MTJs are mainly fabricated in two of our magnetron sputtering systems. The systems are equipped with residue gas analyzers (RGA) to monitor the partial pressures of different gas molecules. There are three inlets to inject gases into the chamber for reactive sputtering. It has one load lock chamber separated by an isolation valve which allows us to transfer the samples without breaking the ultra-high vacuum (UHV) of the main chamber. The load lock chamber is connected to the turbo pump which can go down to 10^{-5} Torr in about 40 mins. To obtain the high-quality samples, the chamber's total pressure and the Ar gas flow need to be precisely controlled. 15 sccm Ar flow and 2-2.5 mTorr pressure are used for metals deposition and 20 sccm Ar flow and 1-1.5 mTorr pressure are used for oxides deposition. The water vapor inside the chamber is detrimental especially for the MgO barrier and one of the methods of removing it is Ta gettering. X-ray photoelectron spectroscopy (XPS) and high-resolution Rutherford backscattering (HRBS) revealed that the Ta gettering enhances the barrier quality by lowering the interstitial defect [133]. Ta has lower enthalpy of formation with oxygen, so it is a good oxygen absorber. On the other hand, it has higher enthalpy of formation with Mg, so it doesn't poison the MgO target. Turning on Ta plasma for about 30 mins before the deposition of MgO barrier was found to be effective in decreasing the H₂O concentration inside the chamber.

At the National Institute of Standards and Technology

The Functional Nanostructured Materials Group (<https://www.nist.gov/mml/materials-science-and-engineering-division/functional-nanostructured-materials-group>) within the NIST Materials Science and Engineering Division (<https://www.nist.gov/mml/materials-science-and-engineering-division>) is host to a comprehensive suite of tools necessary for the synthesis, processing and characterization of inorganic thin films and devices with controlled functional properties. Relevant to the synthesis of magnetic tunnel junction (MTJ) films, we used the custom-built physical vapor deposition facility called the Magnetic Engineering Research Facility (MERF). Up to 14 magnetron sputtering guns are disposed in the MERF (base pressure 2×10^{-10} Torr) for depositing MTJ films onto substrates ranging from small pieces up to 2" wafers. Substrate heating (800° C), dc/rf substrate biasing, continuous azimuthal rotation and tilt were critical for forming the L1₀-phase FePd films and magnetic tunnel junction devices with bulk perpendicular magnetic anisotropy. Two *in-situ* mass spectrometers provide feedback on the in-vacuum sample environment. Surface science studies are available with in-situ low energy electron diffraction, x-ray photoemission (XPS) and Auger spectroscopies. Through-thickness XPS can be carried out with in-situ ion milling. *Ex-situ* processing includes multiple annealing stations, including a vacuum annealing station with the ability to apply magnetic fields (up to 1 T) at arbitrary angles to sample films and rapid thermal annealing (nominally zero field). The MERF was a key component to realizing the gains in TMR for the L1₀-FePd based MTJs and the integration of L1₀-FePd on semiconductor compatible Si wafers, leading to several publications and one submitted patent at USPTO.

Magnetic characterization tools available to this project included a CapresTech Current-in-plane-Tester of Magnetic Tunnel Junction films with in-plane and out-of-plane field capability; a custom broadband ferromagnetic resonance spectrometer (2 T, 110 GHz) for in-plane and out-of-plane spectroscopy of single layer and MTJ films; electrical probe stations for quasi-static and continuous-wave excitation of MTJ devices; a BH-looper; two SQUID magnetometers (7 T, < 10⁻⁸ emu sensitivity) and a vibrating sample magnetometer; and a magneto-optic imaging facility with variable 2 T magnet capable of Kerr and Faraday microscopy as well as Magneto-Optic Imaging Film microscopy.

Relevant structural characterization tools available to this work include x-ray diffractometers (Bruker D8 area detector and Rigaku X-ray Diffractometer with Rotating Anode Source); atomic force and magnetic force microscopy; two scanning electron microscopes, two transmission electron microscopes (includes ability for Lorentz and holographic imaging, and on-staff support for TEM specimen preparation). In collaboration with other Divisions we have access to time-of-flight secondary ion mass spectrometry and atom probe tomography to probe lateral- and through-thickness MTJ compositional deviations.

3.2.2 Micro- and nano-size MTJ device fabrication

At the University of Minnesota, despite our capacities in characterizing an MTJ stack's performance in the thin film level, patterning the stack in MTJ devices is still the only way to prove its functionalities. By using conventional photolithography and Ar⁺ ion milling, we can easily fabricate micro-size MTJ devices following the process flow shown in **Figure 3.4**, where step (A) is the first step and step (F) shows the completed MTJ device.

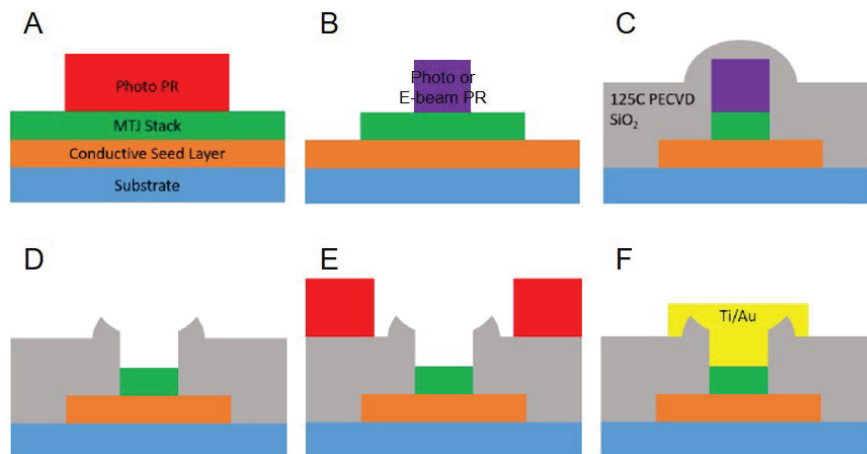


Figure 3.4 The Process Flow of MTJ Device Fabrication

Side views of the MTJ devices at different fabrication procedures are listed from (A) to (F), in which (A) shows unpatterned thin film and (F) represents completed MTJ devices. To micro-size devices, regular photo resist is used in (B). However, to fabricate nano-size MTJ devices, E-beam lithography required to define the device pillars, and thus E-beam resist is used in (B).

To demonstrate the functionality of our MTJ devices at advanced technical nodes, it is necessary to fabricate nano MTJ devices and thus the resolution of our photolithography ($\sim 1\text{ }\mu\text{m}$) can no longer meet the requirement in this scenario. Instead, E-beam lithography is employed by our team and a deeply downscaled device size of $\sim 80\text{ nm}$ has been demonstrated. E-beam resist must be utilized in the step (B) of **Figure 3.4**. The alignment and writing of the E-beam lithography is performed by using Ebeam Litho EBP5000+ Vistec system, as shown in **Figure 3.5(A)**. It is noteworthy that Ar^+ ion milling process must be preciously controlled during the MTJ device fabrication, especially for nano-size devices. Our ion mill etcher, as shown in **Figure 3.5(B)**, can control the beam angle to effectively remove the re-deposited sidewall[134] to avoid device from being short. Moreover, a secondary-ion mass spectrometry (SIMS) system is installed to control the etching endpoint.

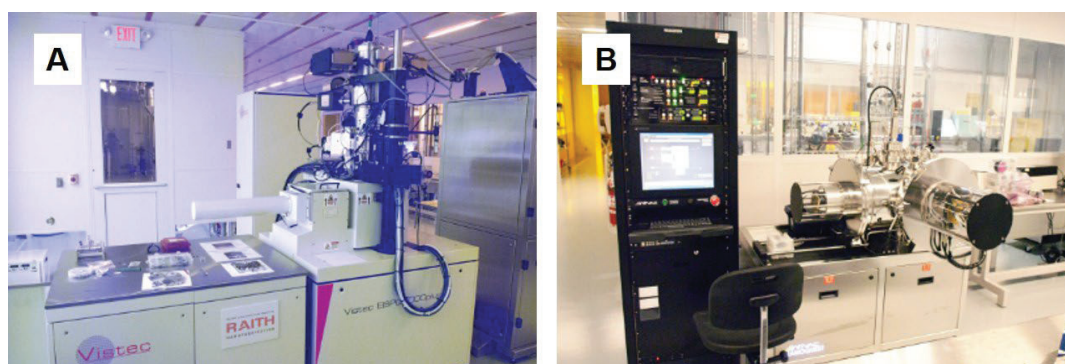


Figure 3.5 (A) 100 KeV E-beam Lithography System (Ebeam Litho EBP5000+ Vistec) Varying sizes allowed, from pieces, all the way up to 4 inch wafers. **(B) Intivac ion mill etcher.** It is used for non-chemical etching of thin films. A SIMS system is installed to precisely control the etching endpoint.

At the University of Arizona, W. Wang's lab has two small in-house clean rooms (class-1000) for photolithography, nanofabrication and wafer preparation. Inside the clean rooms are a Neutronix-7400 photolithography Mask Aligner and nanolithography test station. The mask aligner features a true Split field view with two digital microscopes and IR back side alignment. It accepts 4" masks and wafer with size of 1"-3". W. Wang's Lab is equipped with an Ion-beam Etching system with gridded DC Ion source with UHV flange, used for etching and substrate cleaning. It supplies a stable beam intensity at $0.5 \sim 10 \text{ mA/cm}^2$ and operates with rotating substrates. It can accept wafer up to 3 inches

3.3 MTJ Electrical Testing and Measurements

3.3.1 Infrastructures and Capabilities

At the University of Minnesota, there are three testing stages in our lab equipped with different electromagnets (**Figure 3.7**). The home build magnetic coils stage can supply a maximum of 220 Oe (without iron core) and 780 Oe (with iron core) in-plane field. This stage can also use an alternative perpendicular magnetic coil with 1800 Oe maximum field. The 3D projected magnet can apply the maximum field of 3000 Oe and 2000 Oe in the in-plane direction and perpendicular direction, respectively. The big electromagnet can generate 4000 Oe maximum field with rotational sample holder. The resistance of the MTJ is measured by Keithley 2182, Keithley 6221 or Keithley 2400. For RF measurement, the probe model we used (ACP40) has a bandwidth of DC to 40 GHz. Bias Tee (Picosecond 5542) and power combiner (Picosecond 5331) were also used in the testing circuit. The high frequency magnetic dynamic behavior requires analyzing the data collected by the broadband spectrum analyzer or the storage oscilloscope Tektronix DPO72004C. The pulses are generated by HP 8110A and Picosecond 10070A. All the measurements are done in a computer-controlled program written by LabView.

At the University of Arizona, the DC and RF characterization of transport properties are formed on one of the three probe-station in W. Wang's Lab. **Prober-1** can be configured for both DC and RF measurement. It has a GMW projection magnet providing both in-plane (360-degree rotatable) and out-of-plane magnetic field. Max. in-plane field 4000 Oe and out-of-plane field 2000 Oe can be achieved. It features Keithley 6220 & 2182A meters for fast low noise dI/dV measurement, as well as a picosecond pulse generator for dynamic switching experiment down to 100ps. Prober-2 is a customized Signatone S1007 station. It is equipped with both perpendicular and in-plane magnetic field and a Stanford S830 lockin amplifier and Keithley 2400 sources meter. Currently it is mostly used for training and REU programs. **Prober-3** is a second-generation probe station designed and fabricated by the W. Wang's Lab. It has all the features of probe-one, now with the extended magnetic field strength to 1.5 T. It can access wafer up to 4". For measurement needs large field, W. Wang's Lab utilizes a RT **high field station**, which produce field > 2.5 Tesla with water cooled magnet, for both in-plane and out-of-plane orientation. There is also a home-built low temperature transport system with cryogen-free closed cycle refrigerator, which operates in the 10~300 K range under field up to 1 T.

3.3.2 Magnetostatic DC Electrical Measurements

The resistance of the MTJ device is characterized by the four-probe method as shown in **Figure 3.6a**. The current is applied on one side of the electrodes and voltage is measured on another side

of the electrodes. We use Cascade's DC probe and RF probe for the testing. The DC probe is a simple tungsten-rhenium needle. The RF probe is designed as an air coplanar waveguide with three contacts standing for Ground Source and Ground (GSG). The model we used (ACP40) has a bandwidth of DC to 40 GHz.

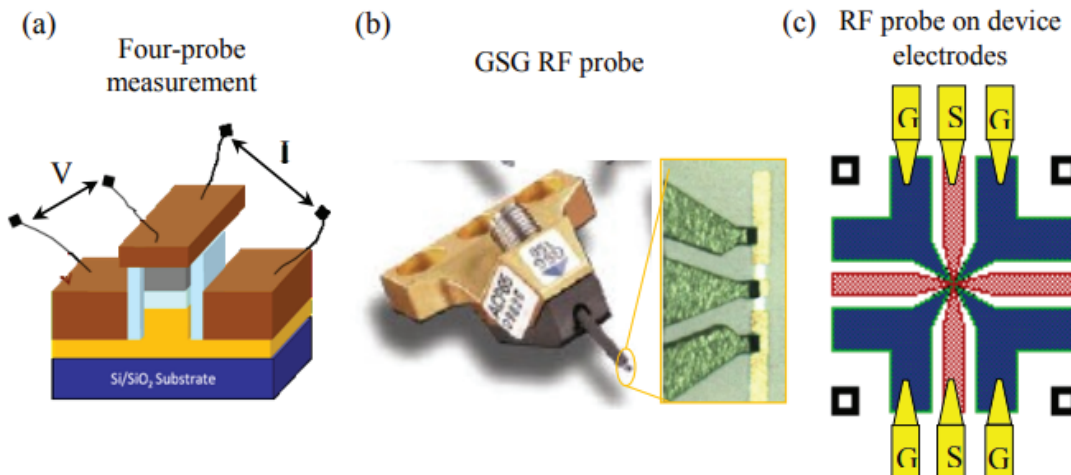


Figure 3.6 (a) Schematic Picture of the Four-probe Measurement Method, (b) Cascade GSG Probe, (c) GSG Probes on our Co-planar Waveguide Shape Electrodes

The TMR ratio is measured by testing the device resistance as a function of the magnetic field. **Figure 3.7(a)** shows a typical resistance-field (R-H) loop.

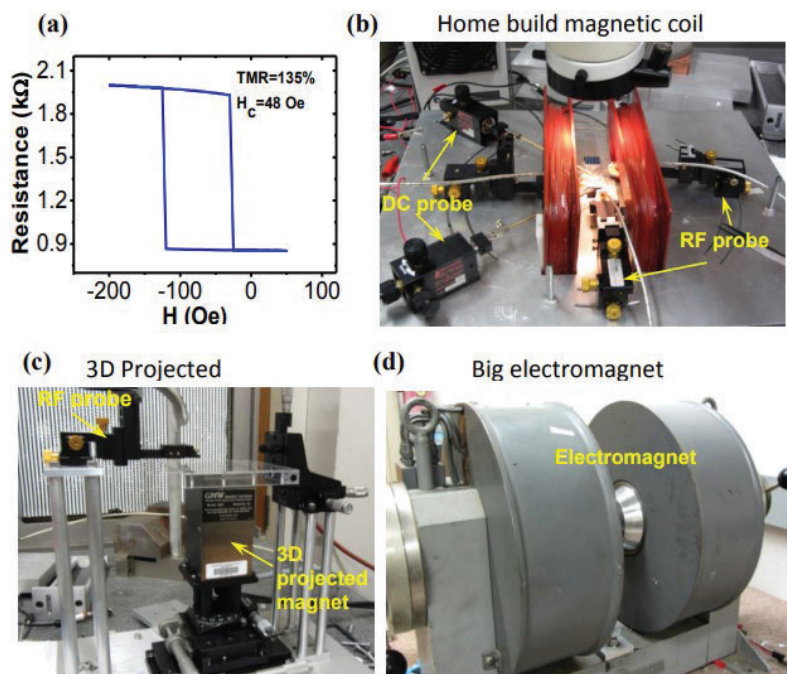


Figure 3.7 (a) A Typical Resistance versus Magnetic Field Loop Measured from an In-plane MTJ, (b) Home Build Magnetic Coils Testing Stage, (c) 3D Projected Magnet Testing Stage and (d) Big Electromagnet Testing Stage

3.3.3 Dynamic electrical measurements

The magnetic dynamic behavior characterization for the device is a magnetoresistance-based electrical signal measurement. So the measurement system is mainly built based on electrical components suitable for high-frequency capabilities listed in Table 2-2. The ground-signal-ground (GSG) probes are used to connect the device contact pad to the 2.92mm SMA cable (18GHz) based measurement circuits. The GSG probe's pitch size must fit the pitch size of the lithography pattern. The connection for the GSG probe is usually K-type (40GHz) which is electrically and physically suitable for the SMA cable when the interested band is below 18GHz, however, one K-type to SMA adapter is suggested to protect the K-type connector from physical damage if any stress will be applied to the connection. Bias-T is the key RF passive component to separate the DC current and high-frequency AC signal. It includes three ports DC, AC and DC+AC. The DC port can pass the DC which is supplied by a current source. The AC port only passes the signal with frequency between 80kHz/10kHz and 16GHz/40GHz to the spectrum analyzer or oscilloscope. The DC+AC port goes to the device under test. If the signal amplitude/power is too low, one or more amplifiers can be added between the bias T and spectrum analyzer or oscilloscope. Detailed models and specifications of the electrical test components are listed in **Table 3.1**.

Table 3.1 Electrical Components for the High-frequency Measurement

Equipment	Model	Function
Probe	Cascade GSG	Pick up signal from CPW contact pads, up to 40GHz
SMA cable	Picosecond Pulse Lab	DC-18GHz
Bias Tee	Picosecond Pulse Lab 5541A/5542	Separate DC and AC signal with 80kHz – 16GHz (5541A) or 10kHz – 50GHz (5542)
Amplifier	Picosecond Pulse Lab 5867	Provide 15dB gain for signal from 10kHz to 15GHz
Source Meter	Keithly 2400	DC current source and voltage meter
Spectrum Analyzer	HP8593A	Detect power with frequency range 9kHz to 22GHz and power range -110dBm to 30dBm (50 ohm Load)
Oscilloscope	Tektronix DPO72004C	Record the signal by 50GS/s sampling rate (20GHz band, 50 ohm load)

The frequency domain measurement is often used to measure the spin torque oscillation. The test electrical setup is shown in **Figure 3.8(a)**. The device under test is connected to the AC+DC port of Bias-T by a SMA coaxial cable to make sure that the DC current can be sent into the device and the AC oscillation signal can be transferred out of the device. The DC source-meter is connected to the DC port of Bias-T and the spectrum analyzer is connected to the AC port of Bias-T. When the current source drives the device and excites the spin torque oscillation, the spectrum analyzer collects the AC signal amplified by the broadband amplifier. A typical power spectrum is shown in **Figure 3.8 (b)**. The raw data points collected by the spectrum analyzer are the power at each specified frequency with certain bandwidth known as resolution bandwidth (RBW). The frequency sweep range (span) should cover the interested band. Span should be

chosen smaller than the maximum data point multiple RBW. So power has to be divided by the set RBW and normalized by the amplifier's gain. The data points can be fit by the Lorentz function to find out the key parameters as frequency, peak amplitude, and full width at half maximum (FWHM, or linewidth). The current and applied field dependencies of these key metrics describe the spin torque oscillation behavior of each specific device.

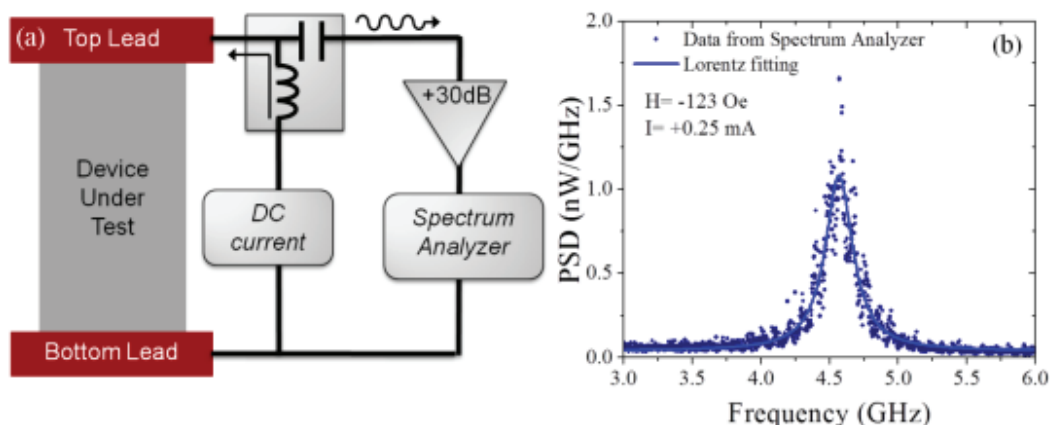


Figure 3.8 (a) The Electrical Setup of the Frequency Domain Measurement, (b) The Measured Power Spectrum

3.3.4 Ultra-fast Switching Probability Measurements

To measure the switching probability of MTJs under certain conditions, electrical pulses are applied to the devices to attempt switching. **Figure 3.9** shows the sequence of each trial. A strong and wide negative pulse is applied first to reset the device to AP state with an almost 100% successful rate. Then a perturb pulse is applied to attempt to switch. The final state of MTJ is random and needs to be read and recorded by applying a weak sensing pulse and measuring the voltage.

Figure 3.10 shows the experimental setup with extra components to allow for a new way of operation. Two pulse generators are used to generate reset and perturb pulses. An ADC (NI-PCI-6221) is used to read the final state of MTJ after each trial.

In this new setup, the microcontroller (MCU) takes a digital signal from the DAQ card as an enable signal. The software controls when the MCU is active through the enable signal. When the MCU is active, it simply generates square wave trigger signals to the two pulse generators and the DAQ card's ADC. The trigger signals are 25 KHz and have a proper delay between each other. The MCU is also responsible for outputting a 5V pulse to bias the MTJ through a resistor facilitating the ADC read-out. A current source, Keithley 6221, is also added to the setup to apply bias DC current on the MTJ if needed.

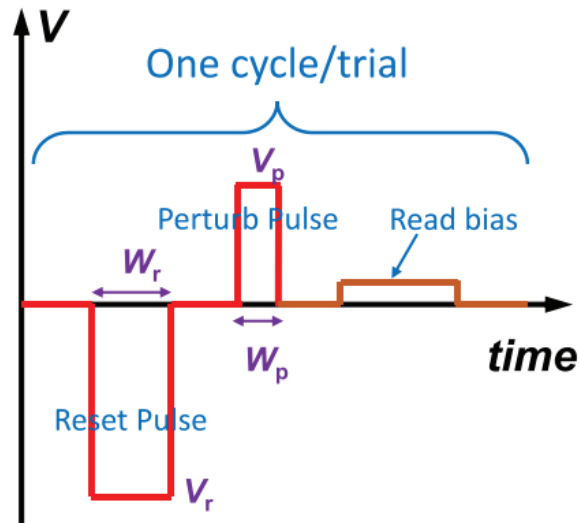


Figure 3.9 Illustration of MTJ Stochastic Switching Pulse and State Read Timing[115]

At the beginning of each measurement, field and bias DC are applied and two pulse generators are configured with appropriate pulse parameters and external trigger mode. Then a digital high is output to the MCU's enable terminal to enable the generation of trigger and bias signals. An N-sample ADC acquisition task is initiated. In this task, a total N samples are requested, and N is equal to the intended number of cycles of trials. The ADC's conversion rate is set to a maximum 250KHz while the sampling clock is set as external active on each falling edge. This sample clock is from the MCU at 25 KHz. Therefore, each sampling of MTJ's voltage will occur exactly at the right moment of each cycle: after the bias voltage on MTJ is turned on. The hardware timing of sampling has negligible delay and the rate of measurement is determined by the clock frequency generated by the MCU. Theoretically, the measurement can run almost as fast of the maximum conversion rate. In practice, to leave an adequate safety margin and due to the limitations of the MCU used, measurements are run at 25000 cycles per second. In comparison, the software in the previous setup that handles the timing of each sampling costs a few ms each cycle and drags the speed down to 100 cycles per second. The new setup is improved by 250 times in terms of measurement speed making the data collection 250 times more efficient.

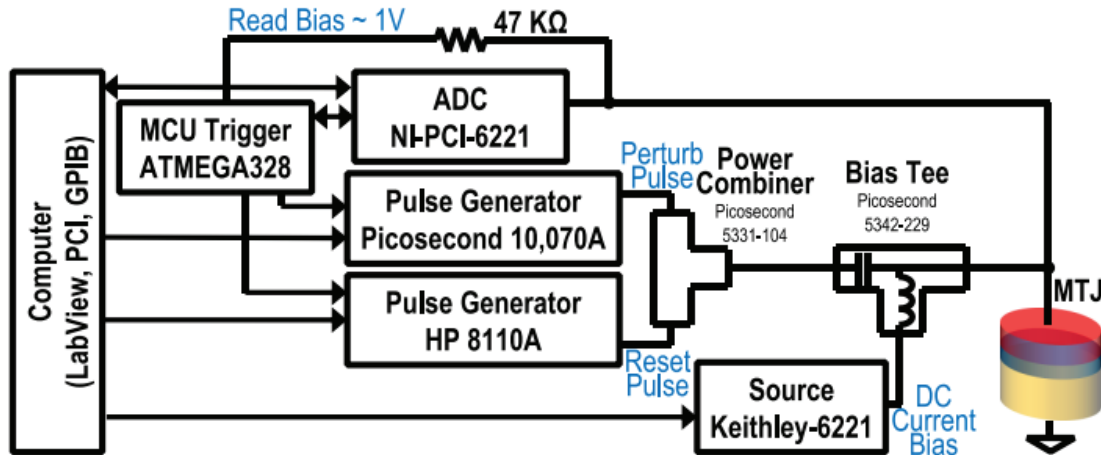


Figure 3.10 Illustration of Testing Setup[115]

3.4 CRAM

3.4.1 CRAM ‘Circuit-around-Die’ Demonstration and Experiments

Figure 3.11 shows the experimental setup, consisting of both hardware and software. The hardware is built with a so-called ‘circuit-around-die’ approach[135]: semiconductor circuitry is built with commercially available components around the MTJ dies. This approach offers a more rapid development cycle and flexibility needed for exploratory experimental studies on CRAM arrays and potential new MTJ technologies, while the major foundries lack the specific process design kit available for making a CRAM array fully integrated with CMOS. The hardware is a 1×7 CRAM array, with the design of cells taken from the 2T1M CRAM cells[72], [74], modified for simplified memory access. Software on a PC controls the operation. It communicates with the hardware with basic commands: ‘open/close transistors’; ‘apply voltage pulses’ to perform write and logic operations; and ‘read cell resistance’. The software collects real-time measurements of the data associated with CRAM operations for analysis and visualization. All aspects of the 1×7 CRAM array are functional: memory write, memory read, and logic operations[57].

MTJs with perpendicular interfacial anisotropy are used in the CRAM. They exhibit low resistance-area (RA) product and high TMR ratio—approximately 100%—when sized at 100 nm in diameter.

An individual bias magnetic field was implemented for each of the seven MTJs on the connection board by positioning a permanent magnet at a certain distance from the MTJ devices. The bias magnetic field was used to compensate for intrinsic magnetic exchange bias and stray fields in the MTJ devices, thereby restoring the balance between the P and AP states. Additionally, slight rotation of bias field in the device plane was used to effectively adjust the switching voltage of each MTJ.

The connection board with seven MTJs was connected to the main board. On the main board, necessary active and passive electronic components were populated on the custom-designed PCB. The CRAM demo hardware circuit implemented a 1×7 CRAM array with a modified

architecture to emphasize logic operations while compromising on memory operations bandwidth for simplicity. It was modified from the full-fledged 2T1M[74] architecture. It was equivalent to a 2T1M CRAM in logic mode, but it only had serial access to all cells for memory read and write operations. The hardware was powered by a battery and communicated with the controller PC wirelessly via Bluetooth®. In this way, the entire hardware was electrically isolated from the environment so that the risk of ESD to these sensitive MTJs was minimized.

The experiment control software running on a PC was implemented using National Instruments' LabVIEW™. It was responsible for real-time measurements and control of the experiments, as well as necessary visualizations. Certain results were further analyzed post-experiment.

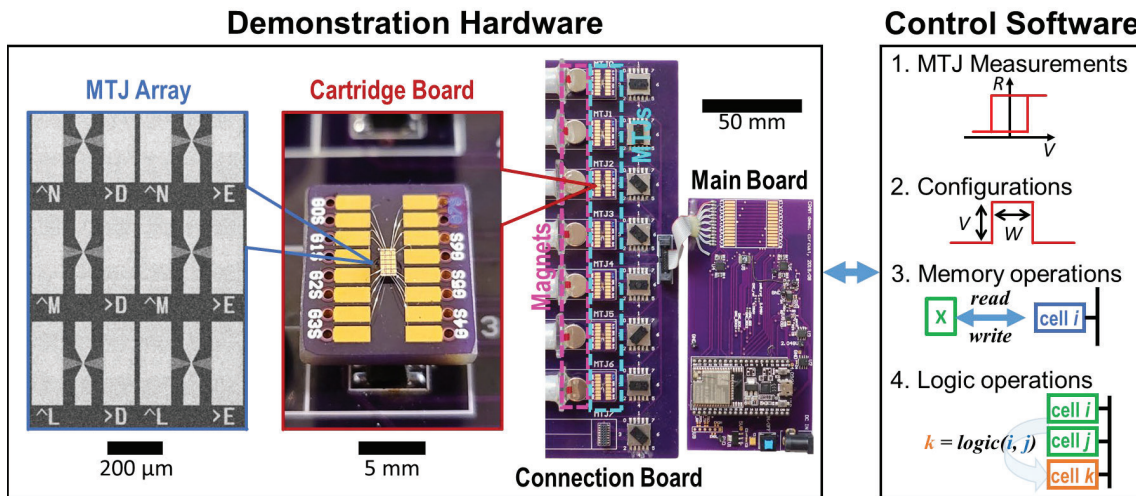


Figure 3.11 CRAM Experimental Setup[57]

The setup consists of custom-built hardware and a suite of control software. It demonstrates a fully functioning 1×7 CRAM array. The hardware consists of a main board hosting all necessary electronics except for the MTJ devices; a connection board on which passive switches, connectors, and magnetic bias field mechanisms are hosted; and multiple cartridge boards that each have an MTJ array mounted and multiple MTJ devices that are wire bonded. The gray-scale scanning electron microscopy image shows the MTJ array used. The color optical photographs show the cartridge board and the entire hardware setup. The software is responsible for real-time measurements of the MTJs; configuration and execution of CRAM operations: memory write, memory read, and logic; and data collection. It is run on a PC, which communicates wirelessly with the main board.

3.4.2 Integration on CMOS Test Vehicle

Since the NIST test vehicle does not perfectly map to a traditional CRAM architecture, it is necessary to copy the outputs of CRAM operations to a new column in between operations. However, the test structure still allows for all the desired CRAM operations to happen in parallel, still allowing for a useful CRAM set of operations. In principle, up to 50 operations could be performed in parallel using the NIST test vehicle as a demo platform. To fully utilize the capabilities of the chip and the described above test interface, we proposed to use a parallel

prefix-adder as the test case for the system. To accommodate the non-ideal setup, the outputs would be copied into a new row before propagating to the next row.

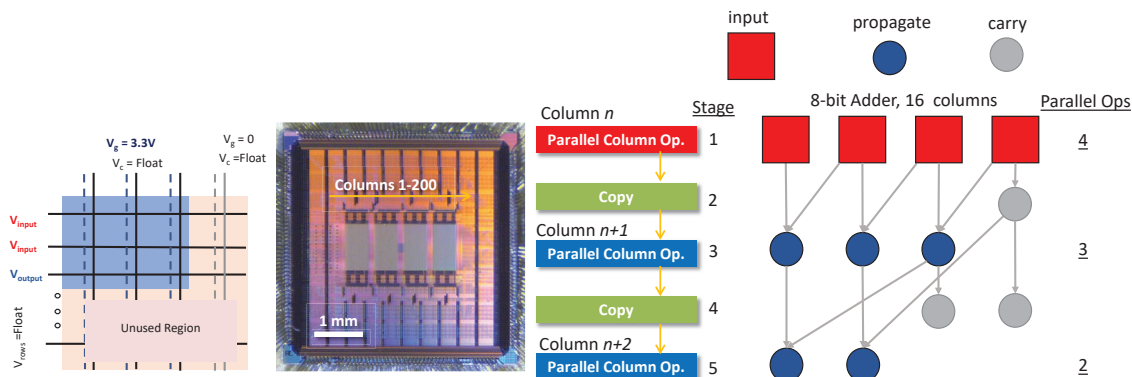


Figure 3.12 Schematic Depiction of Integrated CHIPS with Classes of Operations that can be Performed

It's possible to propagate a single operation across 50 columns at once.

The layout of the CMOS substrates patterned by NIST is shown in **Figure 3.13a-c**. The substrate consists of 4 banks, each of which consists of 5000 cells arranged in 50 columns and 100 rows (20,000 cells total). **Figure 2.13a** shows that each cell contains two transistors, where the gates can be accessed via the enable-bar lines and enable lines and the drain and source of the transistors can be accessed via the out-row and in-column lines, respectively. **Figure 2.13b** shows that the 4 banks are surrounded by 400 probe contacts that can access each line within the cells. The probes on the right- and left-hand side of the substrate access the out-row lines, the probe contacts on the top of the substrate access the enable and enable-bar lines of the transistors, and the probes on the bottom of the substrate access the in-column lines of each cell. It should be noted that the cells in **Figure 2.13a** are slightly different than those shown in **Figure 2.8a**. However, these cells will operate in a similar manner with proper cell selection and multiplexing from the probes. The one remaining task to complete the construction of these substrates is to fabricate NVM memory units in the regions between the sources of the two transistors and the in-column lines, as shown in **Figure 2.13c**.

Fabricating NVM units on substrates with existing CMOS features has several challenges that fabricating on bare Si substrates does not. These challenges include depositing the MTJ stack on selected regions between the rows of the metallic contacts, proper alignment for defining all features during the patterning process and etching all patterns without damaging critical existing features. Note that the fabrication recipes for MRAM follows a process flow which involves 1) depositing the base layers for the NVM cells, 2) patterning the bottom contacts, 3) defining the nanopillars, and 4) patterning the top contacts.

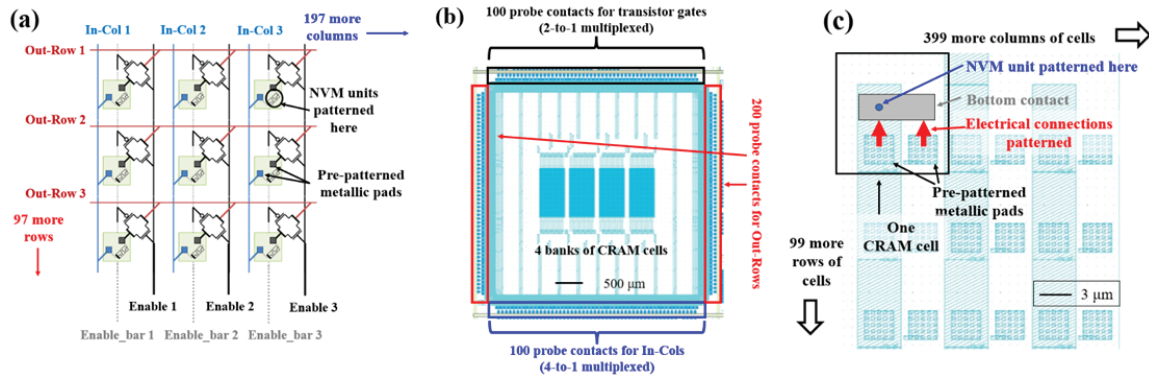


Figure 3.13 (a) Diagram of each Cell on CMOS Substrates Patterned by NIST with Two Transistors and Two Metallic Contacts where Nonvolatile Memory (NVM) Units will be Patterned, (b) Diagram of Probe Connection Layout Relative to CRAM Banks and (c) Image of Metallic Contacts Patterned on CMOS Substrate

Over the past year, we have developed a fabrication recipe at the Minnesota Nano Center for integrating MTJs within each cell. The optical images in **Figure 3.14a-c** illustrate three of the four key steps in the proposed fabrication process. These steps include 1) depositing the base layers for the NVM cells, 2) patterning the bottom contacts, 3) defining the nanopillars, and 4) patterning the top contacts. While these steps are very similar to the ones described in **Section 3.2.2**, it should be noted that fabricating MTJs on substrates with existing CMOS features has several challenges that fabricating MTJs on bare Si substrates does not. These challenges include depositing the MTJ stack on selected regions between the rows of the metallic contacts, proper alignment for defining all features during the patterning process, and etching all patterns without damaging critical existing features.

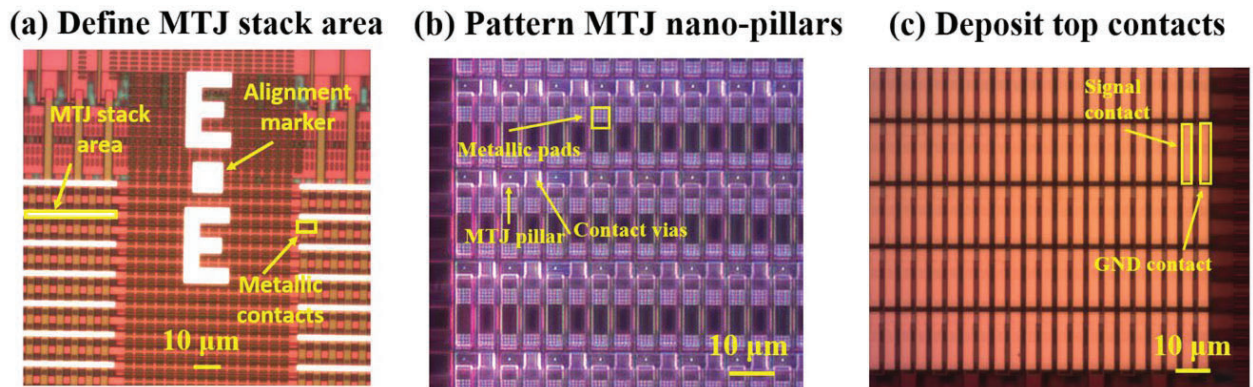


Figure 3.14 Optical Images showing Current Progress of MTJ Fabrication on CMOS Substrates

The fabrication of MTJs was successfully accomplished at the Minnesota Nano Center. However, the current performance of the MTJs is not adequate for CRAM operations. Across 5000 MTJs, the device-to-device variation is around 20% and the average TMR is ~25%. Optimization of stack deposition and fabrication parameters are ongoing in order to improve the average TMR to above 150% and to minimize device-to-device variation.

3.4.3 Modeling and simulation of CRAM

The simulation studies of accuracy as well as error origination, accumulation, and propagation began with a simple probabilistic model of each NAND logic operation. A probabilistic truth table was used to describe the expected statistical average of the output logical state. Then, the 1-bit full adder designs and operations were simulated using the Monte Carlo method with assumed probabilistic truth tables for each of the logic steps.

The experiment-based physics modeling and calculations for obtaining the projected CRAM logic operation accuracies began with an MTJ resistance-voltage model[136], which was fit to the experimental data of TMR vs. bias voltage. The coefficients of this model were scaled accordingly to model projected TMR ratios higher than those observed experimentally. Then, a thermal activation model[4], [137] of MTJ switching probability was fit to experimental data and was used to calculate the switching probability of the output MTJ cell under various bias voltages. Finally, the average of the output state, $\langle D_{out} \rangle$, could be calculated under various V_{logic} , and the optimal NAND accuracies could be obtained in a manner similar to that discussed with **Figure 4.20**.

Further simulation studies of a ripple-carry adder, a systolic multiplier, and the dot-product operation of a matrix multiplication for various numbers of bits as well as matrix sizes were carried out using the same methods.

On top of the error rate calculation models, for the study on CRAM manufacturability (**Section 4.4.3**), the device-device variation of MTJ size or diameter is modeled. The size of each MTJ is modeled to vary following a truncated normal distribution. The distribution is truncated beyond $\pm 3\sigma$ to simulate a purposely engineered population of MTJs with certain means quality control in place, where σ is standard deviation. Then the Monte Carlo method is used. As shown in **Figure 3.15**, each time, a logic operation error rate is obtained with a random sample of MTJ sizes. With adequate sample size, the statistics of the error rate can be obtained with respect to device-device size variation of MTJs.

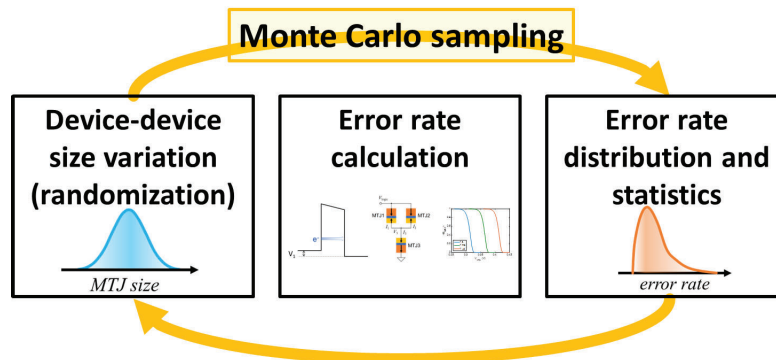


Figure 3.15 Graphic Illustration of the Modeling Methods for CRAM Manufacturability Evaluation by Considering Device-Device size VARIATION of MTJs[124]

The simulations of SC-CRAM, whose results are shown in **Section 4.4.2**, are carried out by using SPICE.

4 RESULTS AND DISCUSSION

4.1 Theoretical Explorations

4.1.1 Gigantic TMR ($\sim 10^4\%$) in Magnetic Weyl Semimetal MTJs

We first studied the current-induced torques in asymmetric magnetic tunnel junctions containing a conventional ferromagnet and an MWS contact. The Weyl semimetal hosts chiral bulk states and topologically protected Fermi arc surface states which were found to govern the voltage behavior and efficiency of current-induced torques. We reported how the bulk chirality dictates the sign of the nonequilibrium torques acting on the ferromagnet and discuss the existence of large field-like torques acting on the magnetic Weyl semimetal which exceed the theoretical maximum of conventional magnetic tunnel junctions, as shown in **Figure 4.1b**. The latter are derived from the Fermi arc spin texture and display a counterintuitive dependence on the separation of the Weyl nodes. Our results shed light on the new physics of multilayered spintronic devices composed of magnetic Weyl semimetals, which might open doors for new energy-efficient spintronic devices[30].

We also proposed a new chiral magnetic tunnel junction with a TMR at least ten times higher than the current state-of-the-art. The new working principle is based on the exotic properties of electrons in a new class of topological magnetic materials known as MWS. In these systems, the interplay between electron chirality and magnetization in MWSs gives rise to a mechanism to controllably suppress tunneling currents in tunneling junctions in the anti-parallel configuration. It is shown that such mechanism enables a performance improvement of at least one order of magnitude over conventional MTJ devices based on typical transition metal ferromagnets[31]. An IP was filed for this idea.

We display in **Figure 4.1c** the angular dependence of the conductance in units of e^2/h for both pMTJ and iMTJ with in-plane magnetization rotation. We have assumed $\beta = 500$ meV and a doping level of $\mu = 250$ meV in these calculations. For the iMTJ case (red square symbols), the conductance in the antiparallel configuration is observed to be orders of magnitude smaller than in the parallel configuration and is limited by the finite electron damping. Hence, electrons can only transmit if chirality is conserved as elucidated by the Bardeen model.

It is instructive to also consider the pMTJ case, where both magnetizations are in the out-of-plane orientation. In this case, the bulk Fermi surface around different Weyl nodes are projected on the top of one another in the $k_{||}$ space. As a result, quasiparticles with a given $k_{||}$ are in coherent superpositions of right- and left- handed states and, therefore, have no well-defined chirality. The full angular dependence of conductance for pMTJs is shown in **Figure 4.1c** (blue triangles). Contrary to the iMTJ case, the antiparallel conductance of pMTJs is only an order of magnitude smaller than its parallel configuration counterpart. The antiparallel resistances of these two distinct situations, iMTJs and pMTJs, differ by several orders of magnitude, allowing for experimental validation of the proposed physics.

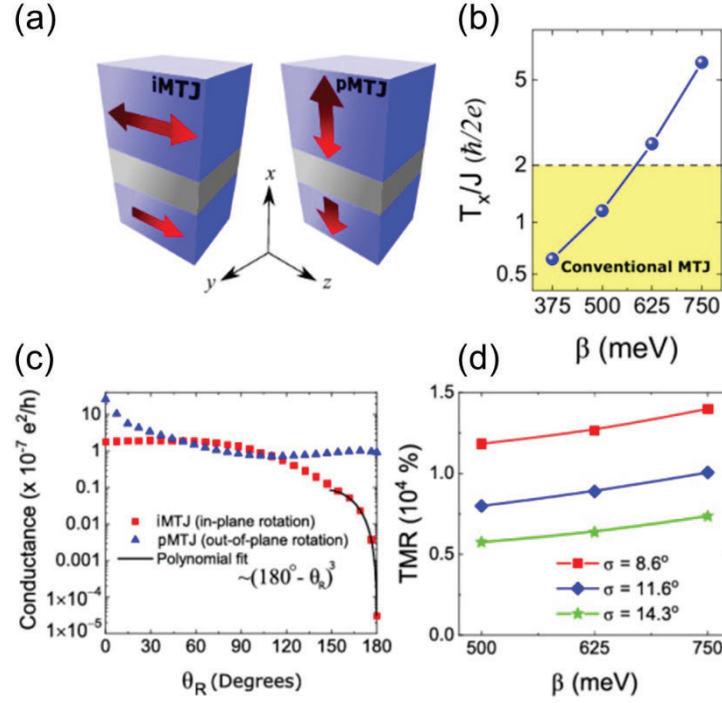


Figure 4.1 (a) In-plane(left)and Out-of-plane(right)Magnetic Weyl Semimetal Tunnel Junctions, (b) Field-like Torque Efficiency T_x/J Exerted on the MWSM Interface Torque efficiency in conventional MTJs is limited to the shaded region. (a) Angular dependence of the transmission probability (quantum conductance per e^2/h) for pMTJs with out-of-plane rotation (blue triangles) and for iMTJs with in-plane rotation (red squares). The solid black line is a polynomial fit for the iMTJ case at $\theta_R \sim 180^\circ$, revealing that conductance of Weyl metal MTJs decays faster than in trivial MTJs. (b) TMR ratio as a function of exchange field strength β . Resistances for parallel (anti-parallel) configuration were obtained by convoluting the zero-temperature resistance with a Gaussian kernel of spread σ and center at $\theta_R = 0^\circ$ ($\theta_R = 180^\circ$), respectively. The spread can be view as due to small thermal fluctuations.

From the analysis above, an ideally infinite TMR ratio should be expected for Weyl iMTJs, as in ideal half-metallic MTJs[138], [139]. However, **Figure 4.1c** suggests the conductance exhibits a stronger angular dependence around $\theta_R = 180^\circ$. In order to quantify how fast the conductance reaches its minimum value, we have fitted the iMTJ curve of **Figure 4.1c** with a polynomial function of the small angular deviations θ_{AP} around $\theta_R = 180^\circ$. We have found that the best fit goes as θ_{AP}^3 [See black solid curve in **Figure 4.1c**], which is faster than the θ_{AP}^2 for trivial MTJs[140]. In fact, a realistic estimation of the size of the TMR should include thermal effects which give rise to an angular window around which the magnetization fluctuates. This can be estimated by convoluting the angular dependent conductance with a Gaussian kernel of spread σ centered around $\theta_R = 0^\circ$ and $\theta_R = 180^\circ$ for parallel and antiparallel alignments, respectively. **Figure 4.1d** shows the TMR obtained from this procedure at different β values for Weyl iMTJs. We observed an additional enhancement with β in all cases. Most importantly, gigantic TMR ratios of $\sim 10^{40}\%$ are obtained for the range of experimentally relevant σ [141].

4.1.2 Interlayer Exchange Coupling on Magnetic Tunnel Junctions

We found that asymmetric MTJs, with dissimilar ferromagnetic contacts, give rise to a linear-in-voltage out-of-equilibrium interlayer exchange interaction. When contrasted with regular spin-transfer torque switching, we found that the linear-in-voltage dependence of the previous interaction reduces the total switching current, as shown in **Figure 4.2**. The combined effects of spin transfer torque, voltage modulation of interlayer exchange coupling, and magnetic anisotropy on the switching behavior of perpendicular magnetic tunnel junctions (p-MTJs) analyzed through this work has revealed the potential in attaining more energy efficient MTJ switching through state-of-the-art effects, including

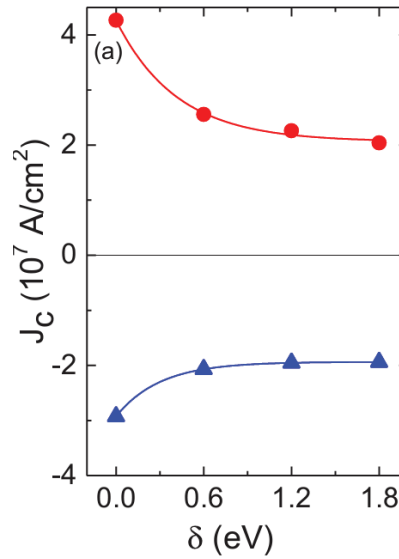


Figure 4.2 Critical Current Density as a Function of the Asymmetry Parameter δ
The red circles (blue triangles) show the trend for P to AP (AP to P) switching.

4.1.3 Magnitude and Sign Change of the IEC via VCEC Effect

Based on our calculation, when applying an E-field ranging from -1.0 to $+1.0 \text{ V/nm}$, a clear transition between the ferromagnetic and antiferromagnetic exchange coupling configurations emerges as shown in **Figure 4.3**. By analyzing the results, we found that an E-field can induce the energetic shifts of the d -orbitals and change the magnetic moment of the CoFe layer. Most importantly, the calculated results reveal that there is an intrinsic E-field (intrinsic dipole field), which originates from the interaction between CoFe and MgO at the interface and is verified by the change in magnetic moment between interfacial and bulk Fe atoms.

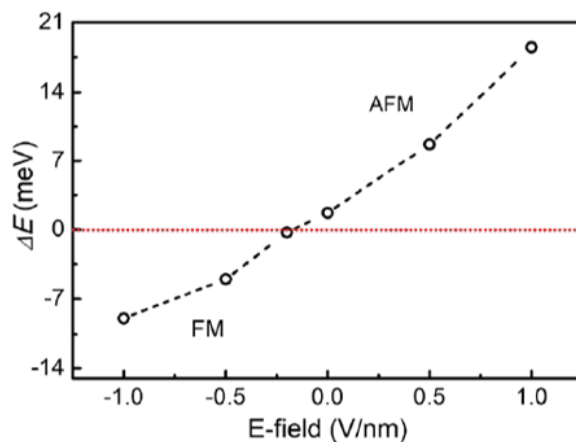


Figure 4.3 ΔE (IEC) vs E-field for the FePd/Ta/CoFe/MgO SAF Structure Calculated by our DFT

As the E-field changes from negative to positive, the IEC gradually changes from ferromagnetic to antiferromagnetic.

We further comprehend the physical mechanism of the VCEC switching. We developed a theoretical scheme based on the quantum interference Bruno's model, in which the IEC between two ferromagnetic layers depends on the spin-dependent reflection coefficients at the interfaces. The evolution of IEC energy ΔE as a function of E-field shows that the sign change of the IEC between FePd and CoFe layers is in consistent with our DFT calculations. Additionally, we verified that the phase modulation (VCEC effect) is suppressed when the MgO tunnel barrier becomes thicker and becomes vanishingly small in the limit of a semi-infinite barrier, where the total phase for the spin-up and spin-down electrons is $-\pi$, as expected.

4.1.4 Investigation of Unconventional SOT on Magnetization Switching

The presence of unconventional spin-orbit torques was shown to have a great impact on the state diagram of the perpendicular magnetization and might enable energy-efficient switching with several orders of magnitude benefit[142], as shown in **Figure 4.4a** and **b**. Contrary to common belief, we found that there exists a critical conventional spin Hall angle, beyond which deterministic magnetization switching transitions to a precessional or pinned state. Conversely, our results indicated that larger unconventional spin Hall angles are generally beneficial for deterministic switching and can lead to several orders of magnitude switching efficiency improvements, as shown by our calculation results presented in **Figure 4.4b**. An approximate expression that qualitatively describes the state-diagram boundary between the full deterministic switching and precessional states was derived. This finding gave rise to a criterion for searching symmetry-broken spin Hall materials in order to maximize switching efficiency, offering invaluable guidance to experimental research.

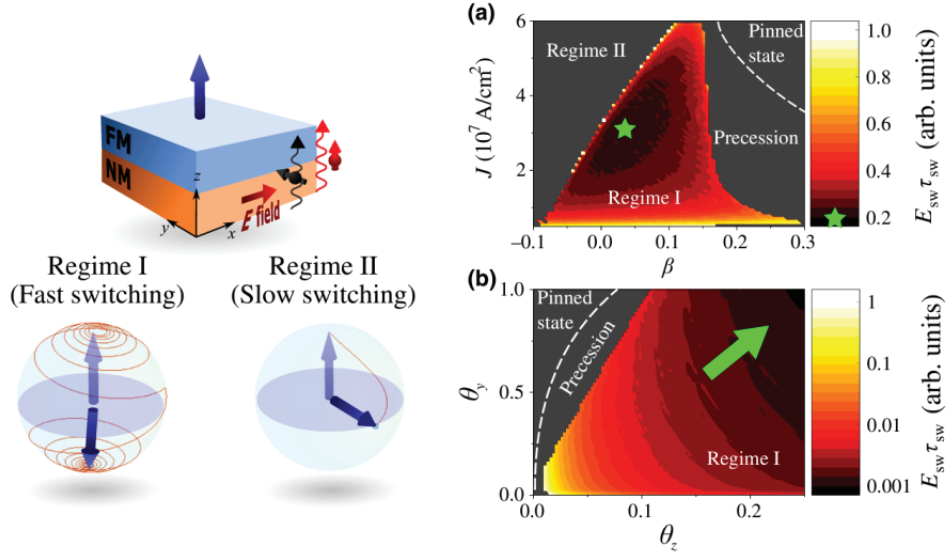


Figure 4.4 Bilayer Heterostructure Comprised of a Ferromagnet (top layer) and a Strong Spin-orbit-coupled NM Layer with a Single-crystal yz Mirror Plane

An electric field applied along the x direction generates SOTs on the perpendicular magnetization due to conventional and unconventional spin Hall effects. The interplay between CSOT and USOT leads to the existence of distinctive dynamical regimes. The bottom figures illustrate dynamical regimes where full switching and full in-plane damping of the magnetization are expected. We note that switching to the $-z$ direction in regime I is faster than switching in regime II in the small intrinsic damping limit. Current density versus field-like to damping-like torque ratio (J versus β) diagram, at fixed $\theta_y = 0.6$ and $\theta_z = 0.05$. The energy-delay product is normalized by that of $J = 0.5 \times 10^7$ A/cm² and $\beta = 0.15$. The green star highlights the optimal energy-delay product. (b) θ_y versus θ_z diagram, at fixed $J = 3 \times 10^7$ A/cm² and $\beta = +0.2$, where we normalize the energy-delay product by that of the case where $\theta_y = 0$ and $\theta_z = 0.015$. The green arrow indicates the direction of best energy-delay product. The darker regions correspond to dynamical regimes other than regime I.

4.1.5 Domain Wall Motion in Asymmetric MTJ

We have shown that as one increases the MTJ asymmetry; by considering dissimilar ferromagnetic contacts, a linear-in-voltage field-like (FL) torque behavior is accompanied by an enhancement in the domain wall displacement efficiency and a higher degree of bidirectional propagation[143]. The domain wall velocity with the vertical current flow as well as with the degree of MTJ asymmetry are shown in **Figure 4.5a** and **b**, respectively. This is due to the fact that DW propagation with this geometry is dominated by the FL torque; as the asymmetry of the MTJ is increased, the voltage dependence of the FL torque acting on the free layer is altered. In analyzing the DW velocity as a function of current density for several asymmetries, there is a clear general trend of increased DW mobility with increased asymmetry, along with bidirectional control of the DW, as shown in **Figure 4.5c**.

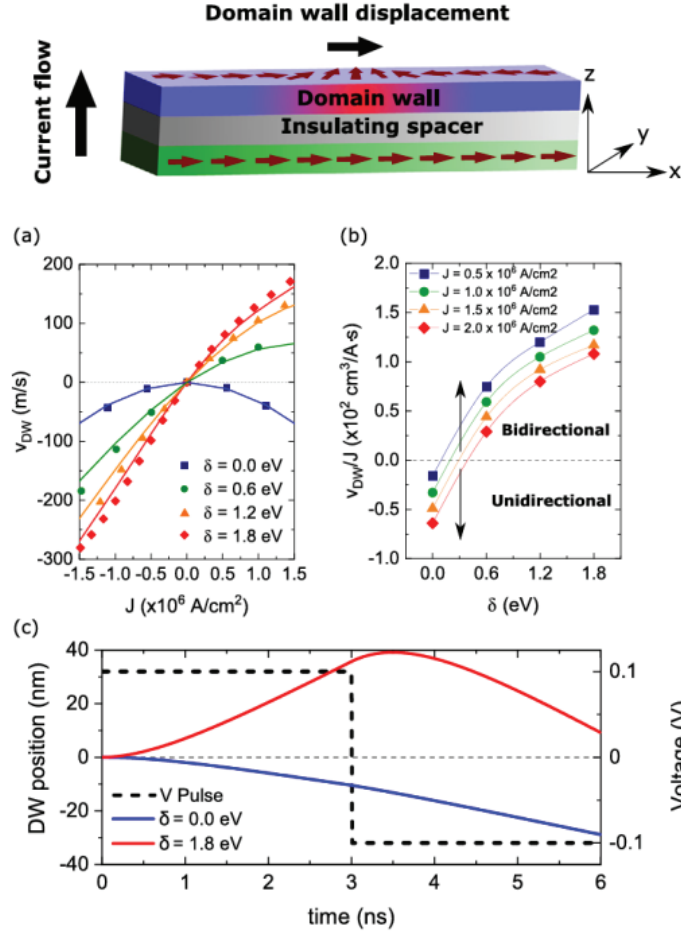


Figure 4.5 We Consider a Magnetic Tunnel Junction where an In-plane DW Displacement in the Free Layer is Induced by a Vertical Current Flow

(a) DW velocity as a function of the average current density for different asymmetries δ s. The symbols were obtained numerically by solving the LLG equation for the rectangular nanowire with current-induced torques obtained from the quantum transport simulations. The solid lines are the analytical toy model results. (b) Asymmetry-dependent DW velocity per applied current density. Different symbols/curves were obtained by considering different applied current densities. Unidirectional and bidirectional highlighted regions for negative and positive v_{DW}/J , respectively, refers to the directionality of DW motion for each asymmetry case. (c) DW position for two cases $\delta = 0.0$ eV and $\delta = 1.8$ eV as a function of time. The voltage pulse is also shown as a solid black curve.

4.1.6 Spin Hall Conductivity in Chiral Materials

We also proposed a spin Hall effect (SHE) enhancement mechanism due to Kramers-Weyl point (KWP) alignment in chiral topological semimetals with high Chern numbers. The crystal and electronic structure of these chiral crystals are shown in **Figure 4.6a-c**, where we also highlight the giant spin Hall conductivity of this material class. We identify enhancements in the intrinsic spin Hall conductivity (SHC) and the spin Hall angle (SHA), which were supported by our first principles calculations is a series of materials. Such enhancements, attributed to a unique high

Chern number KWP energetic alignment and a high degree of band nesting induced by spin-orbit coupling, strongly depend on orbital-orbital interactions. This represents a mechanism to enhance SHE, differing from the spin-orbit induced anti-crossing mechanism in gapped systems. Guided by this principle, we corroborate our results by means of first-principles calculations and reveal multiple realistic materials with large intrinsic SHCs and even larger SHAs than Pt. Our intensive first principles calculations, summarized in **Figure 4.6d**, also identified the best materials for application in highly efficient SOT generation, of which BiTePd and BiPdSe display the largest spin Hall angles.

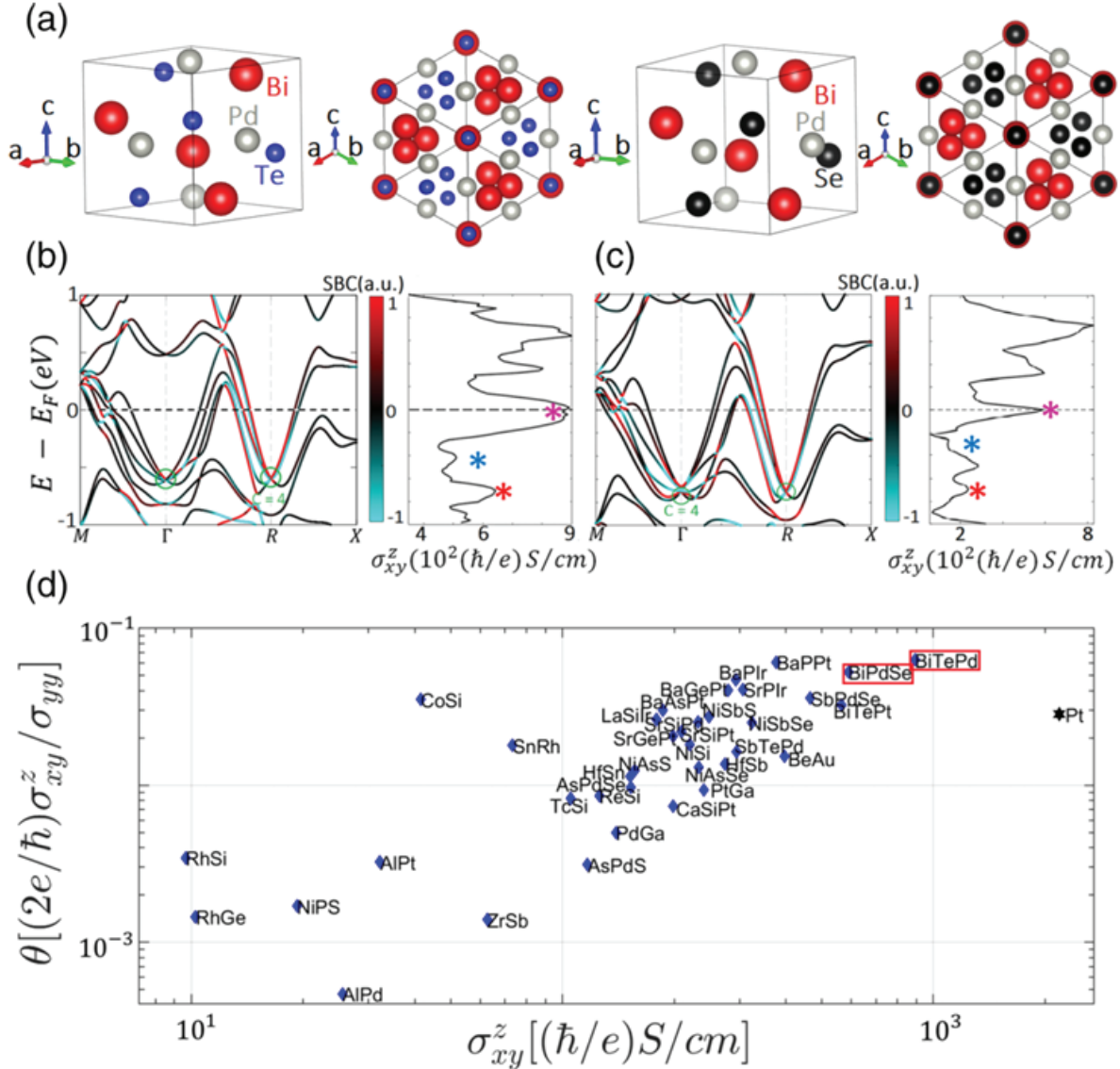


Figure 4.6 Crystal Structures, Band Structures, SHCs, and SHAs of SG-198 Chiral Topological Semimetals

(a) Crystal structures of BiTePd and BiPdSe. (b) and (c) Band structure and SHC of BiTePd and BiPdSe, respectively. (d) SHAs (θ), and SHCs of 37 SG-198 materials at Fermi level with representatives highlighted in red boxes. The numerically determined SHA of Pt (black star) is included for comparison.

4.2 Experimental Results of MTJs with Bulk PMA

4.2.1 VCEC Switching of MTJs with $L1_0$ FePd-based Dual-SAF Free Layers

To demonstrate the predicted VCEC switching of pMTJs, we designed and fabricated a pMTJ stack with an FePd dual-SAF free layer structure (FePd/Ru/FePd/Ta/CFB), as shown in **Figure 4.7(A)**. In our MTJ stack, the FePd (3 nm)/Ru (1.1 nm)/FePd (3 nm) perpendicular SAF stack was prepared on a single crystal (001) MgO substrate with a Cr (15 nm)/Pt (4 nm) seed layer, keeping the substrate temperature at 350 °C. The two FePd layers are coupled antiferromagnetically and this FePd SAF structure is then coupled with a CoFeB (1.3 nm) layer through a Ta (0.8 nm) spacer with weak antiferromagnetic coupling. The rest of the layers of the FePd SAF p-MTJ structures with a stack of MgO (2.0)/CFB (1.3)/Ta (0.7)/[Pd (0.7)/Co (0.3)]₄/Pd (5)/capping layer (where the numbers in parentheses are thicknesses in nanometers) were grown after the substrate was cooled to room temperature.

The deposited FePd dual-SAF pMTJ stack was annealed at 350 °C and then characterized by the high-resolution cross-sectional scanning transmission electron microscopy (STEM). The stack shows a highly textured layer structure, as seen in **Figure 4.7(B)**. To investigate its spin transport properties, the FePd dual-SAF pMTJ stack was patterned into nanosized MTJ pillars (see **Figure 4.7(C)**). Subsequently, the magnetization switching of these FePd dual-SAF pMTJ nanopillars was tested using a conductive atomic force microscope (C-AFM) setup without applying external magnetic field (H_{ext}). In all of the C-AFM measurements, the tip was grounded, and the bias voltage (V_{bias}), either positive or negative, was applied at the bottom electrode. The positive (negative) V_{bias} corresponds to a positive (negative) E-field. **Figure 4.7(D)** shows the current vs bias voltage (I - V_{bias}) curves of the 100-nm FePd dual-SAF p-MTJ pillars with a 2.0-nm MgO tunnel barrier. Here, V_{bias} is swept from -0.6 to $+0.6$ V. A bidirectional magnetization switching is observed in the FePd dual-SAF pMTJs (see the insets of **Figure 4.7(D)**) with bipolar V_{bias} (parallel-to-antiparallel state at $V_{\text{bias}} \sim -0.45$ V and antiparallel-to-parallel state at $V_{\text{bias}} \sim +0.48$). The switching current density, J_c , is calculated to be $\sim 1.1 \times 10^5$ A/cm² based on the switching V_{bias} , which is one order of magnitude lower than that of the best reported STT devices. In addition, we took the VCEC switching measurement above in the different sizes of the MTJ devices. They present a similar switching behavior.

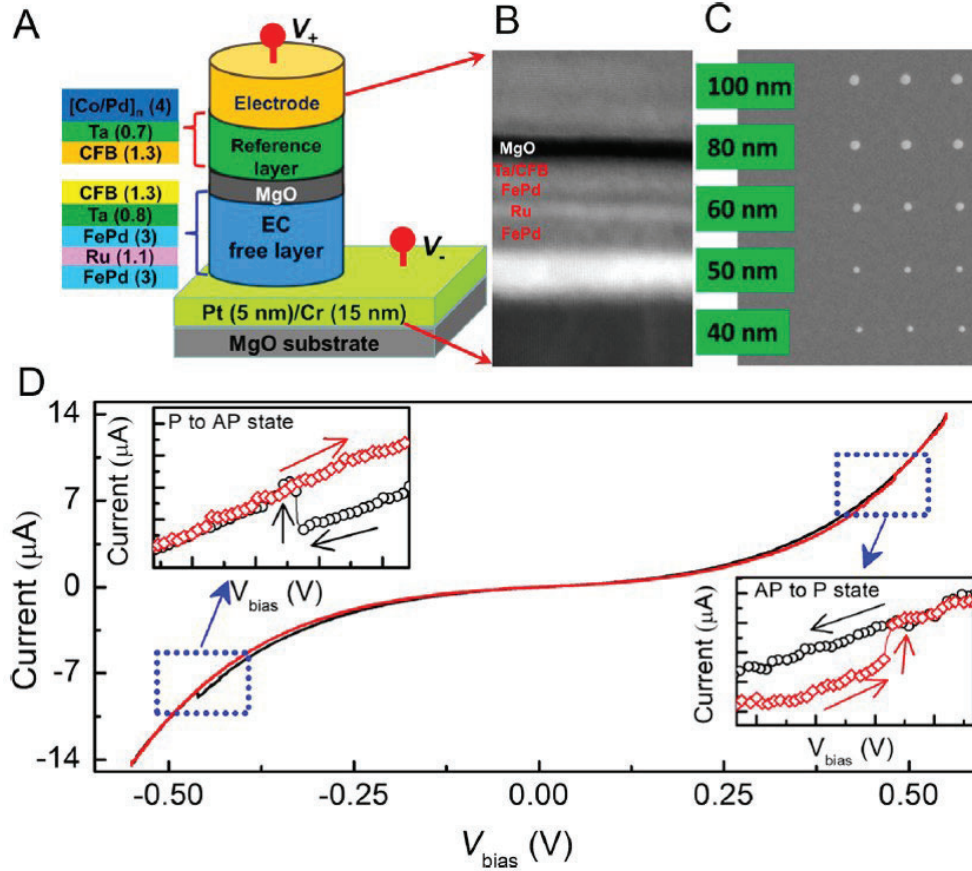


Figure 4.7 Bipolar E-field Switching in pMTJs with Ultralow Current Density[11]
 (A) Schematic of the FePd dual-SAF pMTJ stack. Two FePd layers couple antiferromagnetically through a Ru spacer and then couple with CFB via a Ta spacer, forming the bottom SAF free layer. The CFB/Ta/[Co/Pd]_n stack is the top reference layer. (B, C) Cross-sectional STEM image and the SEM image of FePd dual-SAF MTJ, respectively. (D) Current vs bias voltage ($I-V_{\text{bias}}$) curve for a 100 nm FePd dual-SAF pMTJ device measured by C-AFM without applying H_{ext} . Sharp magnetization switching with the parallel (P)-to-antiparallel (AP) state and AP-to-P state is observed by applying a negative $V_{\text{bias}} \sim -0.45$ V and a positive $V_{\text{bias}} \sim +0.48$ V with the switching current density $\sim 1.1 \times 10^5$ A/cm², respectively.

To demonstrate the robustness of magnetization switching in FePd dual-SAF pMTJs, the current vs time trace was measured while applying voltage pulses on the same pMTJ device without H_{ext} . The results are plotted in **Figure 4.8**. During the testing, ± 0.85 V voltage pulses were applied to *write*, and a small *read* voltage of -0.1 V was utilized to monitor the changes in current values in-between the *write* pulses, as shown in the top panel of **Figure 4.8**. It is clearly observed that the positive/negative voltage pulses can lead to the low- and high-resistance states toggling back and forth, demonstrating bidirectional magnetization switching via VCEC, as depicted in the bottom panel of **Figure 4.8**. In our devices, the IEC strength of the SAF free layer is about 0.056 erg/cm² ($\sim 1 \times 10^5$ erg/cm³ normalized by the thickness of the free layer) without an E-field. This value is smaller than the interfacial PMA ($K_u = 3 \times 10^5$ erg/cm³) of the CFB layer. When applying the negative E-field, the antiferromagnetic to ferromagnetic exchange coupling transition happens due to a dramatic enhancement of the IEC strength of the FePd-CFB

SAF layer and the decrease of the magnetic anisotropy of the CFB layer. When removing the negative E-field, the interfacial PMA of the CoFeB layer goes back to the initial value. At the same time, the SAF free layer will prefer to recover back to the initial state (antiferromagnetic exchange coupling). However, because the interfacial PMA of the CoFeB layer is larger than the interlayer exchange coupling energy between the FePd and CoFeB layers, the SAF free layer will keep the ferromagnetic state when removing the E-field. For the positive E-field case, the SAF free layer always stays in the antiferromagnetic state. As a result, the VCEC switching of our FePd dual-SAF pMTJ devices is demonstrated to non-volatile.

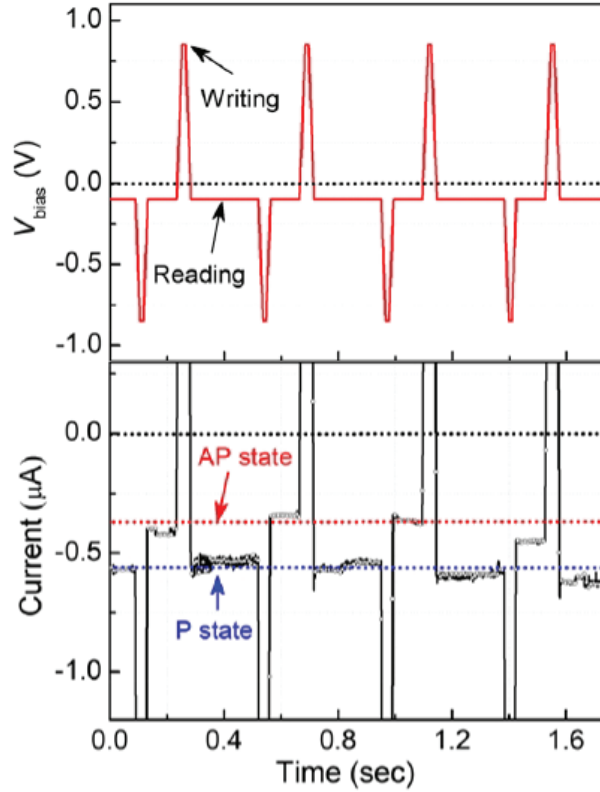


Figure 4.8 Robustness of Bipolar E-field Switching in pMTJs

(Top) Negative and positive 0.85 V voltage pulses were applied for a write operation, and a negative 0.1 V was used to read the current levels to check the device state (parallel or antiparallel). *(Bottom)* Current vs time trace measured for the same device without H_{ext} . Clear bidirectional magnetization switching between parallel and antiparallel states with bipolar V_{bias} [11].

4.2.2 A Substantial TMR Ratio of $L1_0$ FePd-based MTJs

In the last section, our experimental results have demonstrated the VCEC switching with $L1_0$ FePd-based pMTJs. However, as we previously mentioned in **Section 2.2**, the diffusion of Pd atoms can significantly reduce the TMR of the device. In the FePd dual-SAF pMTJ embodiment, the TMR is merely 25%. By using an ultrathin Ru/Mo bilayer spacer between FePd and CoFeB, we achieved a decent TMR of 65%. The $L1_0$ FePd pMTJ thin films were prepared on MgO(001) substrates by DC and RF magnetron sputtering. The stack is MgO subs./Cr (7.5)/Pt (8)/ $L1_0$ FePd (5)/Ru (0.5)/Mo (0.5–0.9)/Co₂Fe₆B₂ (1)/MgO (1.2)/Co₂Fe₆B₂ (1.5)/Ta (3)/Ru (8), in which the

number in parentheses are the layer thickness in nm. In prior to the deposition, we annealed the substrate at 600 °C for 60 min in vacuum and then 30 min in 0.8 mTorr of N₂. The sample was then cooled down to 350 °C in the N₂ environment. The N₂ treatment is performed to heal the Oxygen vacancies generated on MgO surface during vacuum annealing [144]. Then, the Cr, Pt, and FePd layers were all grown at 350 °C. After cooling the sample down to room temperature, we deposited a Ru (0.5) cap on top of the FePd layer. Given the constraints of our equipment with limited cathode sites, the vacuum was interrupted after depositing Ru (0.5) to transition the sample between sputtering systems mentioned in **Section 3.2.1**. After the completion of the layer deposition, the sample underwent post-annealing at 300 °C for 10 min to induce the PMA of the CoFeB layers. The M - H_{ext} loop, taken from a specimen cut from the sample's central region (with a Mo thickness of ~ 0.7 nm), was assessed under perpendicular fields using VSM and is presented by the red squares and curve in **Figure 4.9(A)**. The stack's PMA is evident, with a distinct two-step magnetization switching pattern emerging. Evaluating the magnetization and switching field for both events, it is plausible to infer that the low field switching (occurring around $\mu_0 H_{\text{sw}} \sim 5$ mT) is attributed to the magnetization reversal of the top CoFeB layer. Conversely, the high field switching (initiating at ~ 20 mT) signifies a synchronous reversal of both the bottom CoFeB and the FePd layers.

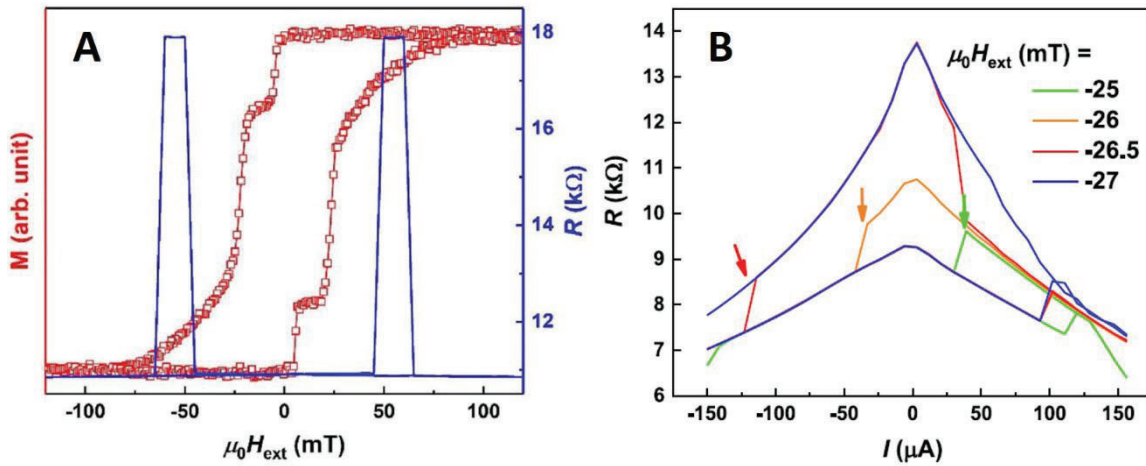


Figure 4.9 (A) M - H_{ext} Loop of the Unpatterned THIN films (red line and squares) and Major R - H_{ext} Loop of a Patterned Device (blue line), (B) R - I loops of a L10 FePd pMTJ Device Under Different H_{ext}

The AP-to-P JSW values are marked by arrows[145].

To study the magneto-transport properties of L1₀ FePd pMTJs, we patterned the sample thin films into nanopillar devices by utilizing the combination of photo and E-beam lithography techniques and Ar⁺ ion milling. These devices have a pillar diameter of 300 nm. Subsequent contact pads were fashioned through a liftoff process. As illustrated by the blue line in **Figure 4.9(A)**, the R - H_{ext} loop reveals a TMR ratio of $\sim 65\%$, marking a notable enhancement compared to previous results. In this context, the ultrathin Ru/Mo bilayer is leveraged as a diffusion barrier for Pd atoms, contributing to the enhanced TMR. It is worth noting that the switching fields evident in the R - H_{ext} loop are considerably greater than those in the M - H_{ext} loop. This discrepancy can be attributed to alterations in dipolar interaction and shape anisotropy.

To delve deeper into the electrical characteristics of $L1_0$ FePd pMTJ devices, we recorded resistance as a function of current under varying offset fields. As depicted by the red line in **Figure 4.9(B)**, pronounced bi-directional switching events emerge when $\mu_0 H_{\text{ext}}$ is set to -26.5 mT. Remarkably, we observed a switching current of ~ 100 μA , yielding an ultralow switching current density (J_{sw}) of around 1.4×10^5 A/cm², very close to our previous results with VCEC switching.

4.2.3 High-quality $L1_0$ FePd and its SAF Prepared on Si Wafers.

Though FePd plays an important role in VCEC pMTJ devices, its PMA depends on the crystallinity of $L1_0$ phase and (001) texture. In order to seed a strong (001)-textured $L1_0$ order, only a few special substrates, such as MgO(001) and SrTiO₃(001), are suitable for $L1_0$ FePd growth. These substrates are costly, hard to obtain in large sizes, and most severely, incompatible to the monolithic integration with semiconductor microelectronics. As a result, applying $L1_0$ FePd to extend the scalability of spintronic devices remains impractical. To advance the development of $L1_0$ FePd-based pMTJ, we demonstrate the success of synthesizing $L1_0$ FePd with (001) texture on industry-ready thermally-oxidized Si wafers by magnetron sputtering.

Our stack containing a Cr/Pt bilayer buffer and the FePd layer were DC magnetron sputtered onto Si(001)/SiO₂ (300 nm, thermally oxidized) wafers with/without an RF sputtered MgO seed layer, followed by a capping layer. Except for the 5 nm MgO that was deposited in an auxiliary chamber (the OLV sputter, base vacuum better than 3×10^{-9} Torr) at room temperature, all the layers were prepared in the NIST Magnetic Engineering Research Facility ultrahigh vacuum chamber with base pressure better than 5×10^{-10} Torr. Before the deposition of Cr, the MgO-coated wafers were baked at 600 °C for 1 h in vacuum and held for an additional 0.5 h in an 0.8 mTorr N₂ environment to heal any oxygen vacancies known to form on the surface of MgO during vacuum annealing. The Cr (30 nm)/Pt (4 nm) buffer and 8 nm FePd layer were grown at 350 °C with a 10 min *in situ* anneal at 500 °C performed after the Cr layer deposition for stress relief. Finally, a Ru (2 nm)/Ta (3 nm) capping layer was grown after cooling the substrate down to room temperature. The sputtering was carried out under an Ar pressure of 2, 3, and 6 mTorr for MgO, Cr/Pt, and FePd, respectively.

As shown in **Figure 4.10(A)**, despite the amorphous SiO₂ surface, FePd shows strong PMA with the insertion of MgO, indicating its (001)-textured $L1_0$ phase. By integrating over the in-plane curve, we estimate the uniaxial anisotropy energy density to be 1.13 MJ m^{-3} , comparable with results of FePd formed on MgO(001) or SrTiO₃(001) substrates [8]. The control sample (i.e., without the MgO seed layer) shows a significant in-plane component of the easy axis. As the PMA follows the orientation of the (001)-textured $L1_0$ phase, the VSM measurements imply that the MgO seed layer is key to engineer the correct FePd microstructure. The Gilbert damping constant is characterized to be 0.012 ± 0.007 , which is close to our best results obtained with MgO(001) substrates[146].

In addition, we also prepared FePd-based SAF structures on Si/SiO₂ wafers by inserting a sub-nm Ir spacer (t_{Ir} is wedged from 4.5 to 6.0 Å) between two $L1_0$ FePd layers. The samples were cleaved along the wedge direction and indexed from 1 to 12. To quantitatively measure the strength of their IEC ($-J_{\text{ex}}$), ultrafast time-resolved magneto-optical Kerr effect (TR-MOKE) technique was employed to study the SAF samples as a function of t_{Ir} . In **Figure 4.10(B)**, $-J_{\text{ex}}$ is

plotted as a function of sample index and nominal t_{Ir} . Clearly, $-J_{\text{ex}}$ decreases as t_{Ir} increases, indicating the coverage of t_{Ir} by the first (and strongest) antiferromagnetic range of Ir's Ruderman–Kittel–Kasuya–Yosida (RKKY) oscillation. Particularly, sample 1 shows a $-J_{\text{ex}}$ of $3.48 \pm 1.09 \text{ mJ m}^{-2}$, which is approximately 34% larger compared to the former result[147].

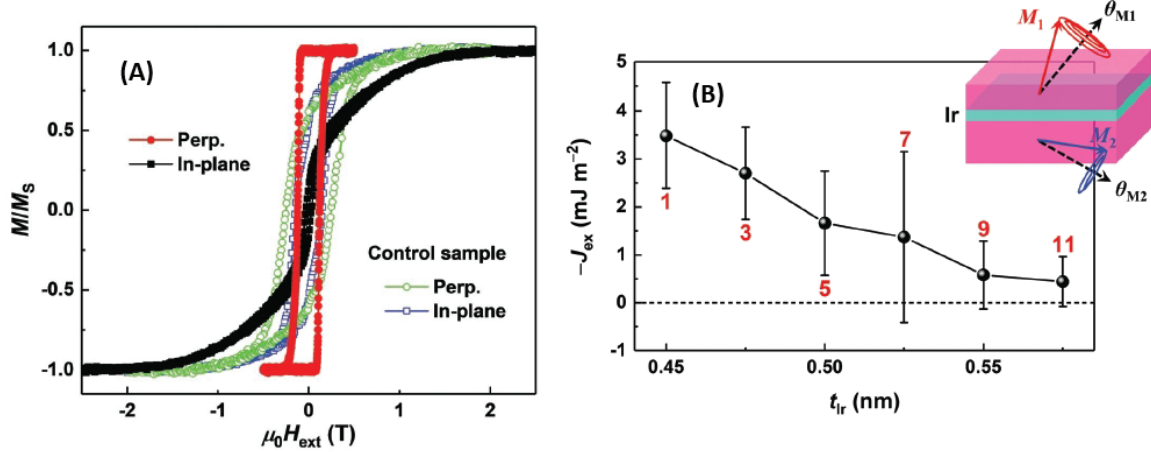


Figure 4.10 (A) Hysteresis Loops of (001)-Textured $L1_0$ FePd (solid symbols) and the Control Sample (empty symbols)

(B) $-J_{\text{ex}}$ of FePd-based SAF samples vs sample index (numbers in red) and nominal t_{Ir} . Inset: Schematic of high-fidelity TR-MOKE metrology developed to study perpendicular SAFs. M_1 , M_2 , θ_{M1} , and θ_{M2} denote the magnetization of the top and bottom $L1_0$ FePd layers and their equilibrium angles under H_{ext} , respectively[10].

To elucidate the high performance of $L1_0$ FePd/Ir/ $L1_0$ FePd SAF stacks, we performed high-resolution transmission electron microscopy (HRTEM) and STEM to investigate their microstructure. As shown in **Figure 4.11(a)**, the high crystallinity of both the top and bottom FePd layers is confirmed. Though the Ir spacer is ultrathin, it is still observable. By using the STEM-HAADF mode, the Ir spacer can be imaged with better contrast as shown in **Figure 4.11(b)**, where the two FePd slabs are separated by a continuous Ir band. To convincingly illustrate the epitaxial relationship between each layer, especially the two FePd layers, we conducted nanobeam diffraction (NBD) measurements to a single grain. The results, as presented in **Figure 4.11(c)**, show a grain-on-grain heteroepitaxy through the Ir layer. Specifically, the SAF trilayer is textured as $L1_0$ FePd [100](001)/Ir/ $L1_0$ FePd [100](001). Such an epitaxial growth, together with the sharp and flat interfaces between each layer, enables the considerable improvement in $-J_{\text{ex}}$.

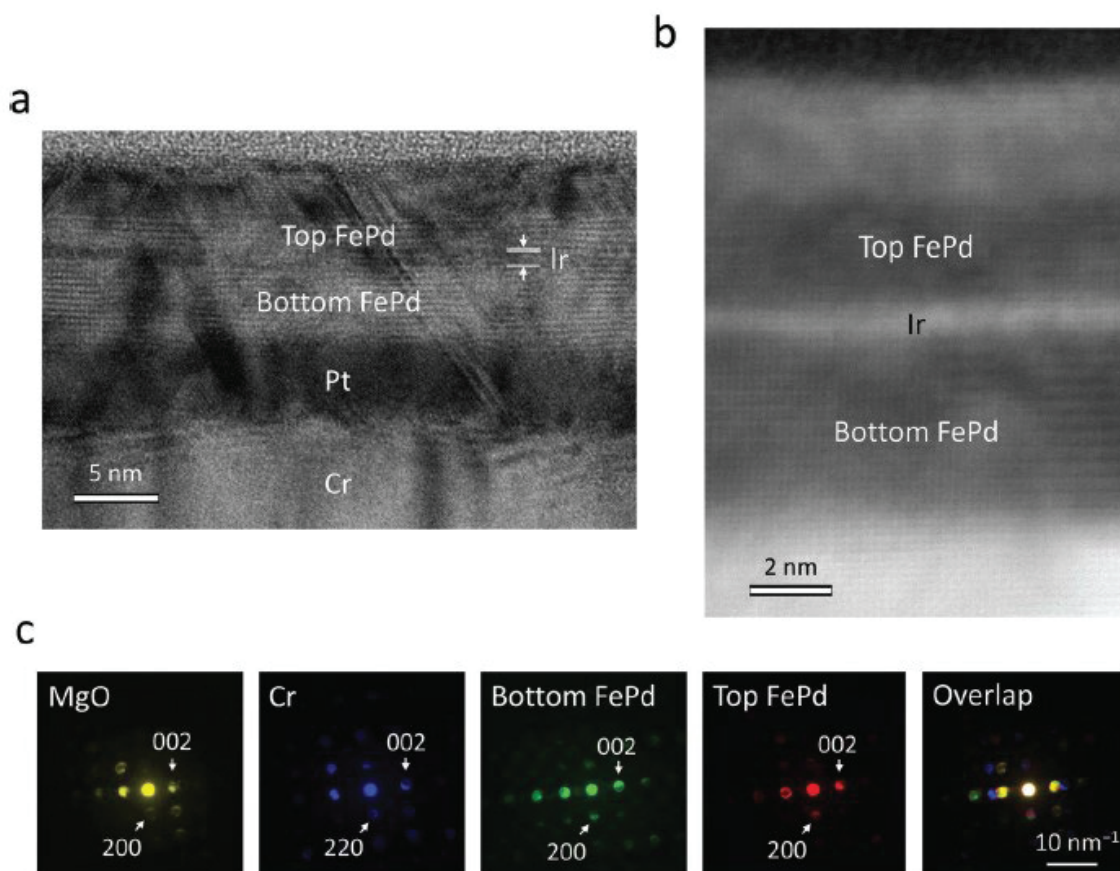


Figure 4.11 (a) A Cross-sectional HRTEM Image of the FePd SAF, (b) A STEM-HAADF Image of the L10 FePd/Ir/L10 FePd Trilayer Structure, (c) NBD Patterns Collected within a Single Grain and their Overlap[10]

4.3 Experimental Results of MTJs with Interfacial PMA

4.3.1 pMTJs with Multi-interface Free Layers and High TMR

The schematic structure of a pMTJ with multi-interface free layer (MIFL) is shown in **Figure 4.12**. Multiple FM layers are coupled through the NM layers to behave like a single free layer of the pMTJ. When FM layers are thin (≈ 1 nm), the hybridization of $3d$ orbitals of FM with the $2p$ orbitals of O,[148] or with the $5d$ orbitals of heavy metals,[13], [149] leads to the interfacial PMA. When properly designed, the PMA from each FM/NM interface can add up to each other therefore providing a sufficiently large Δ for small pMTJ cells. The coupling between the FM layers can be either FM or AF, depending on the needs of a particular application, which can be controlled by the materials and thickness of the FM and NM layers, as well as post-fabrication processes such as thermal annealing. AF-coupled FMs may provide technologically superior performance as it has been predicted that the switching speed can be increased in AF-coupled FMs.[25] This structure is similar to those used in perpendicular spin valves,[150] and pMTJ,[151] except now it combines FM with high spin-polarization with multiple FM/NM interfaces that can provide at the same time a large TMR and strong retention after high temperature annealing at 400 °C or above. CoFeB alloys are a promising candidate for the first

FM in the MIFL, owing to the symmetry-conserved tunneling effect at the Fe/MgO interface.[152] However, for the other layers in the MIFL, a combination of different materials could be used to maximize the PMA and interlayer coupling. Ideally, for any pMTJ at a given lateral size, the number of FM layers and therefore the total number of interfaces (N) contributing to PMA energy density could be increased until the desired Δ is reached.

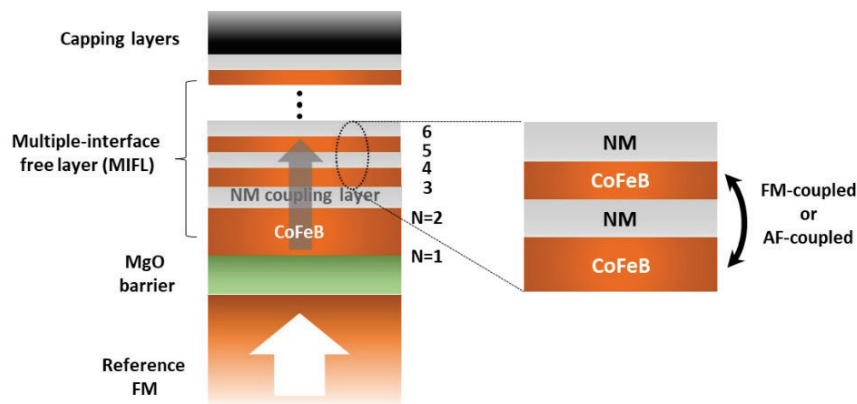


Figure 4.12 Schematic Representation of a pMTJ with Multiple-interface Free Layer (MIFL) where a Number of FM Layers are Coupled Through NM Layers to Function as a Single Magnetic Layer

We first investigated MIFLs with MgO as the coupling layer. In addition to the larger TMR, the CoFeB/MgO interface also provides a strong interfacial PMA.[16], [153] Interlayer exchange coupling in epitaxial Fe/MgO structures has been observed previously, where AF coupling has been observed when the MgO thickness is less than 0.8 nm.[154], [155] However, most of these studies were performed with thick Fe layers where the magnetic easy axis lies within the plane. For sputtered MgO that is sandwiched between two perpendicularly magnetized Co layers, AF coupling was also found even when MgO was as thick as 1.3 nm.[156] For the MIFL depicted in **Figure 4.12**, it is desirable to have a MgO coupling layer that is thick enough to support a strong coupling, but thin enough to contribute only minimal additional series resistance to the overall resistance of the pMTJ – which may be satisfied if current can conduct across pinholes within the thin MgO layer. The magnetic properties of the MIFL with MgO was first investigated in a sample with the structure of [CoFeB(0.75 nm)/MgO(0.8 nm)]₃ by a vibrating sample magnetometer (VSM). The film exhibits PMA as shown in **Figure 4.13a**, where an in-plane anisotropy field larger than 1 T (10 kOe) can be observed. The MH loops in the low field region is shown in **Figure 4.13b**, where the remanent magnetic moment is about 20 nA·m² (20 μ emu) which is about one-third of the saturation moment (60 nA·m²), indicating the three CoFeB layers are AF-coupled.

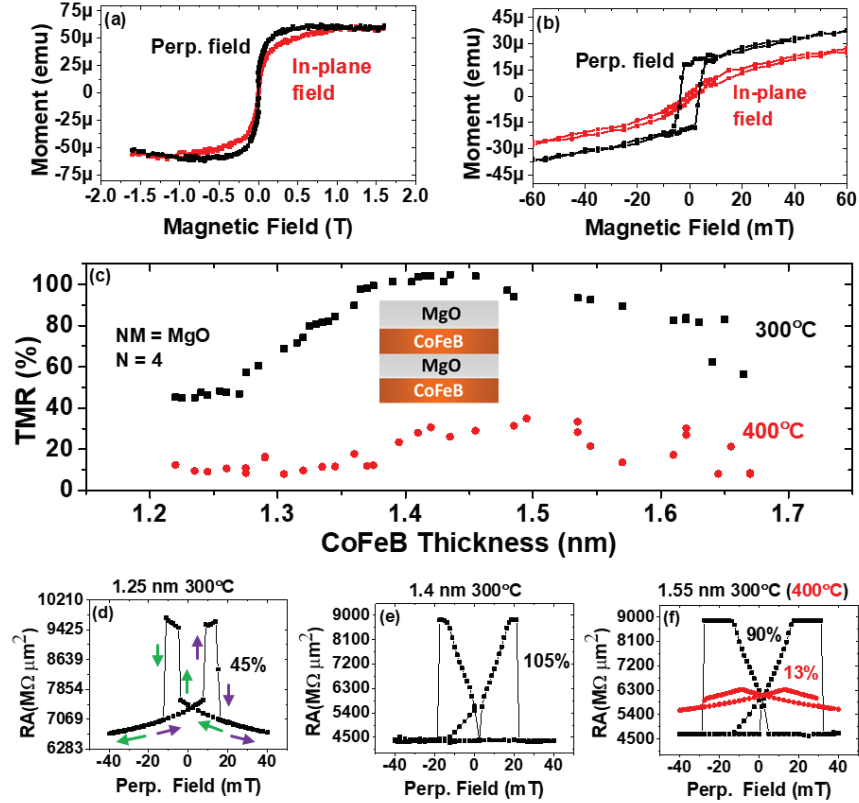


Figure 4.13 (a) Hysteresis Loops of [CoFeB(0.75 nm)/MgO(0.8 nm)]x3 Measured Under Different Magnetic Fields, (b) the Same Curves at the Low-field Region, (c) CoFeB Thickness Dependence of the TMR in pMTJs with MgO-MIFLs

Inset: schematic of MgO-MIFL. The samples were first annealed at 300°C for 10 min, then 400°C for another 10 min. (d-f) representative TMR curves of the pMTJs after the 300°C annealing. The arrows in (d) show the representative magnetic field sweeping direction [purple (green) towards the positive(negative) field direction]. The red curve in (f) is the TMR of the same sample after the 400°C annealing.

Next, pMTJs with MIFLs of the structure of CoFeB (1.2 nm - 1.7 nm)/ MgO (0.9 nm)/CoFeB (1.3 nm) were fabricated. These pMTJs are denoted as $N=4$ because the total PMA originates from three CoFeB/MgO interfaces and one CoFeB/Mo interface (recall the capping layers are Mo/Ta/Ru). Since the first CoFeB in the MIFL is the one contributing to both TMR and PMA, it is critical to study the thickness dependence of this layer. The TMR of these junctions after annealing at 300°C for 10 min is plotted in **Figure 4.13c**. The TMR is about 45% when CoFeB is 1.25 nm thick (**Figure 4.13d**), which increases to more than 100% when CoFeB is 1.45 nm thick. Further increase of CoFeB thickness beyond this point leads to a slight decrease of TMR (**Figure 4.13f**). Note MTJs located on the edge of the wafer were not included due to poor patterning as a result of edge bead formation. The TMR is, however, reduced across the entire thickness series when the same pMTJs were annealed again (after testing at RT) at 400°C for 10 min. A number of processes simultaneously occur during the annealing process, most importantly the formation of the CoFe(001)/MgO(001) epitaxial structure with the B diffusing out of CoFeB, and the reduction of interfacial oxidation which eventually leads to proper hybridization of Oxygen $2p$ orbitals and Fe/Co $3d$ orbitals that is required for a strong PMA.

Typically the parallel state resistance (R_P) of the junction momentarily drops at the beginning of the annealing, resulting from the initial establishment of the highly conductive Δ_1 channel, followed by a steady increase due to the gradual deterioration of that channel when other atomic species inevitably diffuse into the barrier.[157], [158] Despite the increase of R_P , the TMR may continue to increase at 400°C for up to a few hours of annealing, provided that increases in the anti-parallel state resistance (R_{AP}) due to the reduction of the Δ_2 and Δ_5 conduction channels outpaces of the increase of R_P . [157], [159] The comparison of the TMR curves from the same junction after annealing at 300°C and 400°C is presented in **Figure 4.13f**. In addition to the increase of R_P , we note that R_{AP} is decreased as shown by the red TMR curve. The decrease of R_{AP} in **Figure 4.13f** is accompanied by the disappearance of the sharp switching in the TMR curve, which is likely due to the reduction of PMA of the MIFL, instead of a more transport-intrinsic reason that is usually only expected when the annealing is much longer.[157] The reduction of PMA might be a result of Boron aggregation at the CoFeB/MgO interface in the absence of any “Boron-absorbing” layer adjacent to the CoFeB layer.

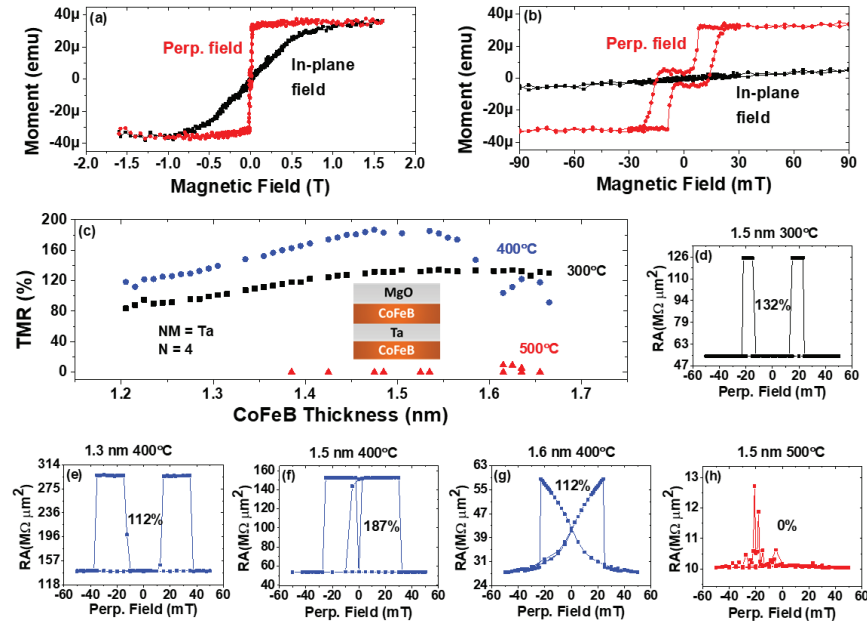


Figure 4.14 (a) Hysteresis Loops of MgO/ CoFeB/Ta(1nm)/CoFeB/MgO Measured on Different Magnetic Fields, (b) the Same Curves at Low Field-region, (c) CoFeB Thickness Dependence of the TMR in pMTJs with Ta-MIFLs
Inset: schematic of Ta-MIFL. The samples were successive annealed at each temperature for 10 min. (d) TMR curve of the 1.5 nm sample after the 300°C annealing. (e-g) TMR curves of three pMTJs after the 400°C annealing. (h) TMR of the 1.5 nm sample after the 500°C annealing.

We also investigated pMTJs with the MIFL of the structure of CoFeB (1.2 nm-1.7 nm)/ Ta(1 nm)/CoFeB (1 nm)/MgO (0.8 nm), which is denoted by $N=4$ with Ta in **Figure 4.14** since now there are two CoFeB/MgO interfaces and two CoFeB/Ta interfaces that each contribute to PMA. This structure is similar to what was used previously,[160]–[163] except here we chose the Ta at 1nm. Maximum AF-coupling was observed in Co/Ta superlattices when Ta is near 0.7nm.[164] For CoFeB/Ta/CoFeB, a sizable AF coupling was obtained when the thickness of Ta is around 1nm,[165] which is in agreement with our VSM results shown in **Figure 4.14a** and **b**. After the

300°C annealing for 10 min, maximal TMR of 135 % was obtained as shown in **Figure 4.14c**, which is considerably better in those shown in **Figure 4.13**. Reasonably high TMR (> 100 %) is present in pMTJs across a wide range of CoFeB thickness. A representative TMR curve under this annealing condition is shown in **Figure 4.14d**, with sharp transitions between states and a clear AP state. This improved TMR behavior is likely related to the fact that the Ta coupling layer more readily absorbs Boron diffusing out from the CoFeB layer, compared to a MgO coupling layer. Subsequent annealing of these pMTJs at 400°C for 10 min substantially increased the maximum TMR to above 180 % as shown in **Figure 4.14f**, which is noticeably higher than previous reports.[160]–[163] However, the TMR starts to fall off after the CoFeB thickness exceeds 1.5 nm and exhibits large fluctuations when CoFeB is thicker than 1.6nm. This reduction in the TMR is attributed to the loss of the AP state as shown in **Figure 4.14g**. Usually, the strength and sign of the interlayer coupling is sensitively depended on the thickness of the FM and NM layers.[131], [156] Here another complexity is involved, which is the PMA of MIFL. Due to the relatively small formation energy between Ta and Fe, it is known that the PMA of MgO/CoFeB/Ta is not stable when annealed at 400°C,[166] which leads to the deterioration of PMA of the MIFL stack. Further annealing of these pMTJs at 500°C leads to a dramatic reduction of TMR to nearly zero, consistent with previous studies.[166]

It has been shown that pMTJs with Mo as the heavy metal layer exhibited much higher TMR than that of Ta[166]–[168]. Therefore, MIFLs with Mo as the coupling layer may provide larger TMR as well. Another benefit of Mo is that its interlayer exchange coupling energy is larger compared to that of Ta.[164] It was also shown that Mo can substantially enhance damping.[169] pMTJs with three CoFeB layers in the MIFL have been fabricated. The MIFL stack structure is CoFeB (1.2 nm-1.7 nm)/Mo (0.9 nm)/ CoFeB (1 nm)/MgO(0.9 nm)/ CoFeB (1.3 nm). These samples are denoted as $N=6$ with Mo as plotted in **Figure 4.15**. The VSM results show the strong AF coupling for Mo(0.9nm) are presented in **Figure 4.15a** and **b**. The TMR ratios of the samples after annealing at 300°C for 10 min (red curve) are presented in **Figure 4.15c**. The maximum TMR in pMTJs with Mo-MIFL is similar to that of Ta-MIFL under this annealing condition. However, the TMR starts to decay in pMTJs with Mo-MIFL when CoFeB is thicker than 1.5 nm. By comparison, the TMR with Ta-MIFL gains a slight increase over 1.5 nm to 1.6 nm under the same annealing condition as shown in **Figure 4.14c**. This feature of Mo-MIFL becomes more pronounced after the annealing at 400°C for 10 min as shown in the black curve of **Figure 4.15c**, where TMR drops sharply when CoFeB is thicker than 1.4 nm. When the pMTJs are annealed at 400°C for one hour, the overall TMR has been increased (blue curve), with maximum TMR reaching 212 % as shown in **Figure 4.15e**, which is even higher than the TMR in pMTJs we obtained previously with a single CoFeB layer as the free layer.[167] However, TMR quickly drops when CoFeB thickness exceeds 1.42 nm. The reduction of TMR is again related to the disappearance of the AP state as shown in the **Figure 4.15f**, which is likely due to the loss of PMA of the MIFL. These results highlight the very sensitive dependence of TMR on the first CoFeB layer thickness in the MIFL. The AF coupling peak with Mo varies in different reports, ranging from 0.5 nm,[164] to 0.8 nm.[170] In another series of pMTJ where the Mo in the MIFL is slightly thicker (1 nm), a more pronounced AF coupling of the free layer can be seen as shown in the inset of **Figure 4.15c**. The free layer switching fields are obviously not symmetric about the zero magnetic field, which is a signature of the AF coupling of the CoFeB in the MIFL. The behavior of these pMTJs is plotted in the green curve in **Figure 4.15c**. Interestingly, the range of CoFeB thicknesses where high TMR ratio is observed is extended by nearly 0.1 nm, as evident from the shift of the green curve relative to the black one.

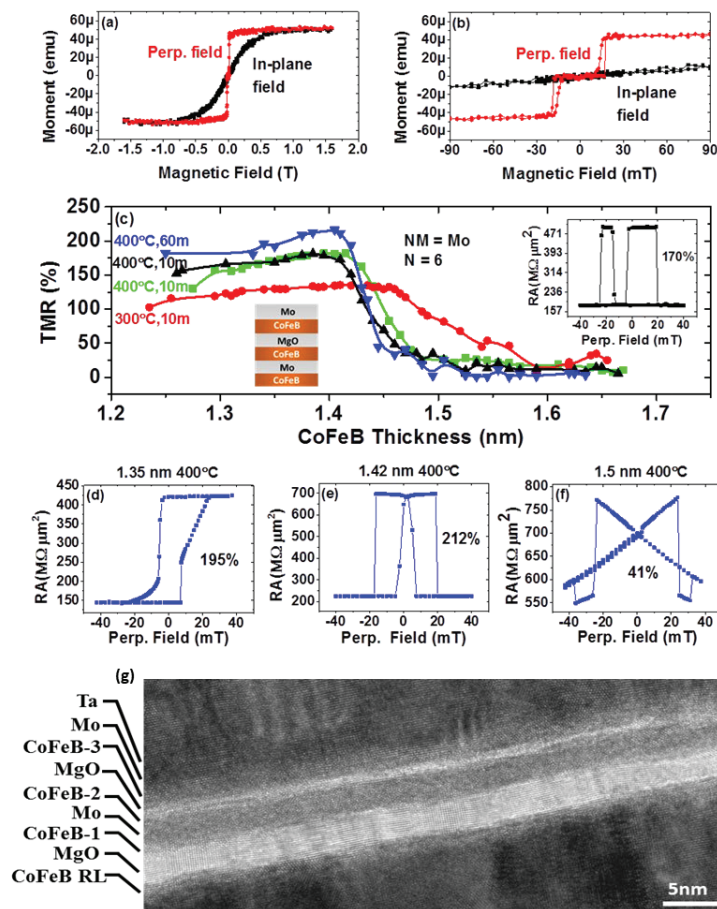


Figure 4.15 (a) Hysteresis Loops of MgO/CoFeB/Mo(0.9 nm)/CoFeB/MgO/CoFeB/Mo Measured on Different Magnetic Fields, (b) the Same Curves at Low Field-region, (c) CoFeB Thickness Dependence of the TMR in pMTJs with Mo-MIFLs

Inset: schematic of Mo-MIFL. The samples with Mo-0.9nm coupling layer were successive annealed at 300°C for 10min (red dots), 400°C for 10min (black up-pointing triangles), then 400°C for another 50min (blue down-pointing triangles, where the total annealing time at 400°C is 60min). The green square data points are the TMR values of the samples with the Mo-1nm coupling layer, annealed at 400°C for 10min. The lines are for guiding eyes only. Inset shows the TMR curve of a pMTJ with the Mo-1nm coupling layer. (d-f) Representative TMR curves of three pMTJ after the 400°C annealing for 60min. (g) HRTEM image of the pMTJ with Mo-MIFL.

The microstructure of pMTJ with Mo(0.6 nm) -MIFL was investigated by TEM and is presented in **Figure 4.15g**. The MgO tunnel barrier exhibits good (001) crystalline structure throughout the specimen. The CoFeB reference layer beneath the MgO tunnel barrier and the first CoFeB in the MIFL show predominantly (001) texture, indicating successful solid-state-epitaxy from the MgO barrier outward during the 400°C anneal. The successful recrystallization of the CoFeB/MgO/CoFeB complex is critical for the high TMR ratios observed within these pMTJs. The second and the third CoFeB layers in the MIFL, however, are only partially crystallized. The MgO layer in the MIFL is mostly continuous and exhibits partial crystallization from the thermal processing. Note in MIFLs the amorphous second (and third) CoFeB and some pinholes in the

MgO could potentially be advantageous, as they may help to reduce the Gilbert damping and the resistance-area product of the devices, respectively. Note for Ta-MIFL with $N=6$, the maximum TMR is only 180%, again demonstrating the advantage of Mo as a better HM layer.

These results suggest the first CoFeB thickness must be precisely controlled in order to achieve the largest TMR in pMTJs with MIFL. Though the drop of TMR after a certain threshold thickness of the first CoFeB is a common feature observed in all three types of MIFLs in this study, the decay in pMTJs with Mo-MIFL is most pronounced. Obviously, this phenomenon is related to the reduced PMA of the first CoFeB layer when its thickness is getting larger. However the fast drop of TMR with Mo-MIFL cannot solely be explained by PMA, since PMA is usually stronger in pMTJs with Mo compared to those with Ta.[166] The presence of interlayer coupling makes the situation more complicated, which can certainly have a large influence on the switching behavior of the CoFeB layers in the MIFL. In particular, the coupling strength may vary with the thickness of the FM layer, which in the Bruno model is due to the Fabry-Perot-type interferences of the electron wave functions through multiple reflections in FM layers.[131] Meanwhile, the magnetic properties of the CoFeB itself at a given thickness is under constant change (such as crystallization and redistribution of O at the MgO/CoFeB interface) when annealed at different conditions, which will in turn impact the interlayer magnetic coupling. A more detailed study is needed to understand the evolution of interlayer coupling in these pMTJs. The difference of the two sets of Mo-MIFL samples (0.9 nm vs 1 nm) also indicates the range of CoFeB thickness that gives rise to high TMR may be expanded if the AF coupling is enhanced.

4.3.2 pMTJs with Ultra-low Switching Energy

In this project, we systematically studied the influence of Ir doping in CoFeB-MgO based MTJ nanodevices. Our experimental results indicate that the both the Ir doping gradient near the CoFeB-MgO interface plays an important role in determining several critical MTJ performance, including TMR, PMA and switching probabilities of voltage-driven precessional switching. Our results demonstrate that an ultra-low writing energy <3.2 fJ/bit can be achieved in sub-ns regime for MTJ nanodevice with TMR over 160% and post-annealing temperature over 400C.

Figure 4.16a shows the four different series of MTJ stacks prepared in this study (from the substrate): Mo(1nm)/CoFeB(0.85nm)/MgO(2.5nm)/CoFeB(1.5nm)/Mo(1.5nm) (Sample 1), Mo(1nm)/CoFeB(0.85nm)/MgO(2.5nm)/Ir(0.5A, 1A, 1.5 A)/CoFeB/Mo (Sample 2), Mo(1nm)/CoFeB(0.85nm)/MgO(2.5nm)/CoFeB(1.5nm)/Ir(1A)/Mo(1.5nm) (Sample 3) and Mo(1nm)/CoFeB(0.85nm)/MgO(2.5nm)/CoFeB(0.75nm)/Ir(1A)/CoFeB(0.75nm)/Mo(1.5nm) (Sample 4). There is no Ir insertion in Sample 1 while an ultrathin Ir layer was inserted into the interface of one of the three following configurations: at the interface of MgO and CoFeB (Sample 2), at the interface of CoFeB and Mo (Sample 3), and in the middle of CoFeB layer (Sample 4). Note that Sample 1 can be treated as the one of the extreme conditions in the following: (1) when the thickness of Ir insertion layer approaches zero at the interface of MgO and CoFeB (Sample 2); (2) when the position of Ir insertion layer is sufficiently far away from the interface of MgO and CoFeB. All the samples are firstly deposited on Si/SiO₂ wafers and then patterned into MTJ nanodevices of 160nm in diameter by optical lithography, nanosphere lithography (NSL) technique[171] and ion beam etching. Then the patterned MTJ nanodevices were annealed at fixed temperatures from 300C to 500C under N₂ atmosphere.

Figure 4.16b schematically illustrates the microwave circuit used for evaluating the TMR ratio and switching probability of voltage-driven magnetization switching. The MTJ nanodevices were connected into the microwave circuit via an RF probe. A pulse generator (PICOSECOND PULSE LABS 10070A) was used to generate sub-nanosecond voltage pulses with various amplitudes and duration times, and these voltage pulses were fed into the MTJ nanodevices through the RF port of the bias tee. The actual pulse voltages applied on the MTJ nanodevices are almost doubled due to the impedance mismatch between MTJ and the $50\ \Omega$ coaxial tip and transmission line[172]. The DC port of the bias tee was connected to a DC current source and a voltmeter to monitor the parallel (P) and anti-parallel (AP) configuration of magnetization of MTJ. The TMR ratio is defined as $(R_{AP}-R_P)/R_P$, where R_P and R_{AP} are the MTJ resistance in P and AP configuration, respectively. Tilted external magnetic fields with different magnitudes and polar angles (θ_H) were applied to MTJ nanodevices. The magnetization switching probability is determined by measuring the resistance of MTJ nanodevices 500 times after identical voltage pulse application. All the measurements are performed at room temperature.

Figure 4.16c, d and e show the influence of Ir insertion with different position and thickness on TMR ratio. All the samples shown in **Figure 4.16c, d and e** were annealing under the temperature of 300C. **Figure 4.16c** shows the TMR-field (TMR-H) curves of MTJ nanodevice with different thickness of Ir insetion layer at the interface of tunnelling barrier MgO and the free layer CoFeB (Sample 2). As shown in **Figure 4.16c**, the MTJ nanodevice with thinner Ir insertion layer gets larger TMR. **Figure 4.16d** shows the TMR-H curves of MTJ nanodevice with different location of Ir insertion layer. As shown in **Figure 4.16d**, a larger TMR is obtained when no Ir layer was inserted (Sample 1) or 1Å Ir layer was inserted at the interface of free layer CoFeB and capping layer Mo (Sample 3) rather than in the middle of free layer CoFeB (Sample 4). **Figure 4.16e** summarizes the influence of Ir insertion layer with different locations and varied thickness on TMR using the data from **Figure 4.16c and d**. As shown in **Figure 4.16e**, on one hand, the TMR shows a clear decrease with the increase of inserted Ir thickness at MgO/CoFeB interface (Sample 1, 2), which indicates the deterioration of Fe-O bonding[148] (which is critical to the TMR ratio) could happen when Ir atoms diffuse into the MgO/CoFeB interface even at a moderate annealing temperature as 300C. On the other hand, TMR shows an obvious increase when the location of inserted Ir layer gradually gets far away from the MgO/CoFeB interface while keeping the inserted Ir thickness as constant $\sim 1\text{\AA}$ (Sample 2, 3 and 4). It is worth to mention that, as shown in **Figure 4.16d**, there is no obvious TMR difference between Sample 1 (no Ir insertion) and Sample 3 (1Å Ir insertion at CoFeB/Mo interface), which indicates that the negative influence of Ir insertion (i.e., TMR loss) can be fully eliminated by reducing the Ir doping density near MgO/CoFeB interface.

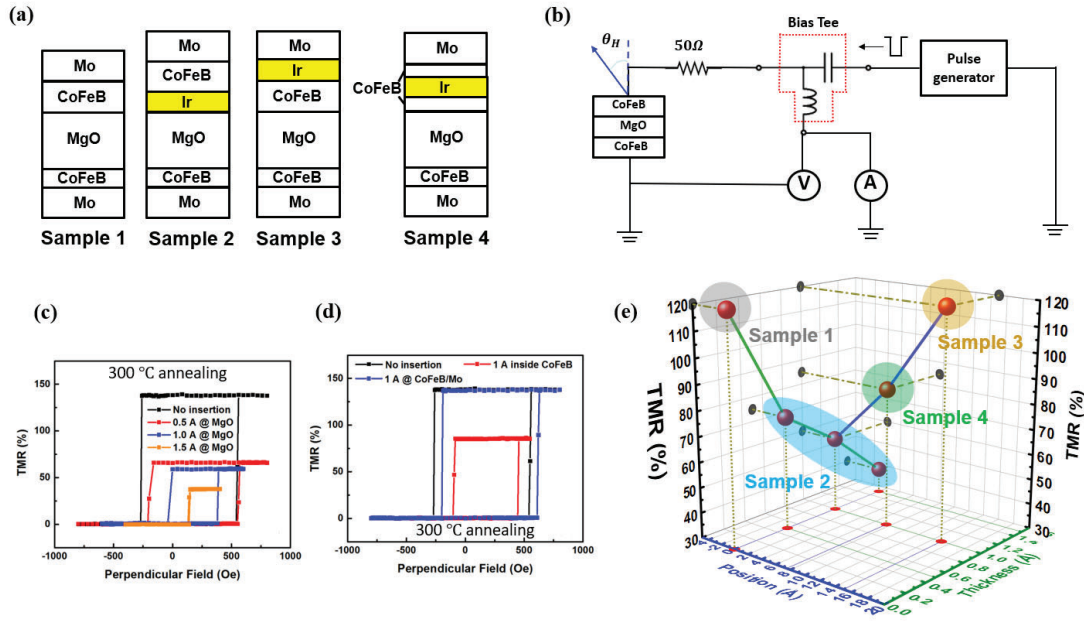


Figure 4.16 (a) Different MTJ Stacks for Sample 1, 2, 3 and 4, A thin Ir Layer is Inserted in the Different Position of MTJ for Sample 2, 3 and 4., (b) A Schematic of Measurement Setup, (c) Curves of TMR-H for Sample 2 with Different Thickness of Ir Insertion Layer, (d) Curves of TMR-H for MTJs with Different Positions of Ir Insertion Layer

The black curve (no Ir insertion), blue curve (1A Ir insertion layer) and red curve (1A Ir insertion layer) refer to Sample 1, Sample 3 and Sample 4, respectively. (e) TMR evolves with different thickness of Ir insertion layer and different position of Ir insertion layer. The data points of Sample 1, Sample 2, Sample 3 and Sample 4 are marked as the background colors of grey, blue, green and yellow, respectively.

Then the voltage-driven precessional switching behaviour for both Sample 1 (no Ir insertion) and sample 3 (1A Ir insertion at CoFeB/Mo interface) were carefully investigated. All the samples shown in **Figure 4.17** were annealing under the temperature of 400C. **Figure 4.17a(d)** shows the measurements of resistance R versus the magnetic field H for Sample 1 (Sample 3) at three different dc bias voltage: -10 mV and ± 600 mV (± 400 mV) with a fixed tilted polar angle (θ_H) ~ 70 (50) degree. The coercivity kept the same or slight increased when the DC voltage changes from -10 mV to a larger positive voltage (i.e., +600 mV or +400 mV), while the coercivity decreased dramatically to nearly zero when the applied DC voltage changed from -10 mV to a larger negative voltage (i.e., -600 mV or -400 mV), which is known as voltage-control magnetic anisotropy (VCMA) effect.

When the PMA is completely eliminated by a short voltage pulse (negative voltage pulse in our case) applied on MTJ nanodevice, the magnetization of free layer CoFeB starts to precess around the in-plane component of the external magnetic field. The precession frequency is determined by the amplitude of the in-plane component of the external magnetic field. If the voltage pulse stops at the timing of half turn precession, the magnetization can be switched to the opposite direction and then stabilized through relaxing. Therefore, the switching time t_{sw} equals to the half of the precession period and can be further expressed as

$$t_{sw} \sim \frac{\pi(1-\alpha^2)}{\gamma\mu_0 H_{in-plane}}, \quad eq. 4.1$$

where α , γ and μ_0 are the Gilbert damping factor, the gyromagnetic ratio and the permeability of vacuum, respectively[173]. More details are discussed in Supplementary Information-1. The precessional dynamics of the MTJ magnetization are reflected in the following two aspects: firstly, the oscillation of switching probability (P_{sw}) as a function of pulse width, as shown in **Figure 4.17b** (Sample 1) and **f** (Sample 3). Secondly, the toggle switching behaviour for the resistance of MTJ nanodevice with successive voltage pulses of same voltage magnitude and pulse duration, as shown in **Figure 4.17e** (Sample 3).

In **Figure 4.17b** (Sample 1), a series of voltage pulses with the fixed magnitude ~ 2.7 V were applied to the MTJ nanodevices. The polar angles (θ_H) of external magnetic field changed from 40, 50, 60 and 70 degree, while the magnitude of field changed from -258, -301, -442 to -630 Oe. Note that the perpendicular component of external magnetic field keeps constant to compensate the stray field from reference layer in MTJ, while the varied in-plane component serves as the in-plane axis for Lamor precession for magnetization [Supplementary Information-1]. With the increase of in-plane component of magnetic field, the oscillating curves shift towards left, which is in a consistence with **eq. 4.1**. **Figure 4.17c** (Sample 1) show the zoom-in of **Figure 4.17b** (Sample 1) for pulse duration below 1 ns. As shown in **Figure 4.17c** (Sample 1), a product of switching probabilities ($P_{sw}=P_{01}*P_{10}$, P_{01} is the switching probability from parallel to antiparallel and P_{10} is the switching probability from antiparallel to parallel) above 80% was achieved with a pulse duration ~ 150 ps, showing the great potential for ultra-fast voltage-driven switching within pico-second scale. **Figure 4.17e** (Sample 3) shows an example of successful back-and-forth precessional magnetization switching measured by the change of R after each identical voltage pulse application of 3.0 V with 0.6 ns duration. The high (AP) and low (P) resistance states appear alternately under a fixed magnetic field 320 Oe with the polar angle of 50 degree, indicating the reversible magnetization switching of free layer happens with each successive voltage pulse. As shown in **Figure 4.17f** (Sample 3), the product of switching probabilities oscillates with the increase of pulse duration, and the product of switching probabilities also increases with increasing voltage pulse magnitude. The product of switching probabilities above 80% can be achieved at voltage pulse magnitude ~ 3.0 V with the pulse duration of 0.6 ns, at which the ultra-low switching energy is evaluated to be 3.2 fJ/bit. Note that the dimension of our MTJ nanodevice in this work is around 160 nm, which is limited by the nanosphere lithography in our fabrication process. The switching energy can be further reduced by shrinking the size of MTJ nanopillar due to the fact that RA (Resistance-Area product) is constant for MTJ thin films. With that saying, the normalized writing energy for our voltage-driven MTJ nanodevices at a dimension of 10 nm can be at ~ 0.01 fJ/bit, which is 1 order smaller than the current typical CMOS switching energy ~ 0.1 fJ/bit[174], showing the great potential of ultra-low energy consumption and good scalability for MTJ nanodevice with voltage-induced scheme.

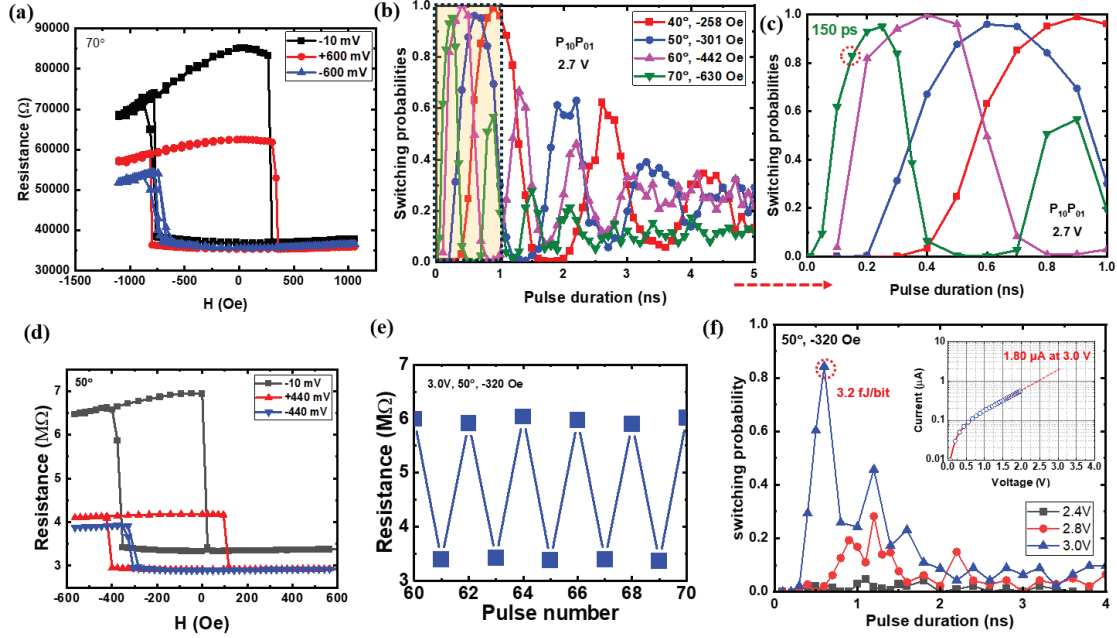


Figure 4.17 (a) R-H Loop of Sample 1, The polar Angle (θ_H) of the Applied Magnetic Field is Fixed at 70 Degree, the Large (+/- 600 mV) and Small (-10 mV) Volage were Applied to the MTJ Nanodevice and the Resistance was Measured when Sweeping the Magnetic Field, (b) the Product of Switching Probabilities as a Function of Pulse Duration with Different Magnetic Fields Under a Fixed Voltage Pulse of 2.7V, (c) The Zoom-in (<1 ns) of b for the Region with Yellow-color Background, A Magnetic Switching with a Product of Switching Probability above 80% was Achieved with 150 ns Voltage Pulse Duration, which is Marked with Red Dash Circle, (d) R-H Loop of Sample 3, the Polar Angle of the Applied Field is Fixed at 50 Degree, the Large (+/- 440 mV) and Small (-10 mV) Volage were Applied to the MTJ Nanodevice and the Resistance was Measured when Sweeping the Magnetic Field, (e) Resistance Measured after the Application of Voltage Pulse of 3.0 V with 0.6 ns Duration, the Applied Magnetic Field is 320 Oe with a Polar Angle of 50 Degree, (f) the Product of Switching Probabilities as a Function of Pulse Duration with Different Voltage Magnitude Under External Magnetic Field of 320 Oe

The red dash circle marks the data points with the writing energy as low as 3.2 fJ/bit with voltage pulse magnitude ~ 3.0 V and pulse duration ~ 0.6 ns. The inset shows how to estimate the switching current through linear fitting of current by voltage.

Then we systematically investigate the Sample 1, Sample 2 (1A Ir insertion only) and Sample 3 for the influence of Ir insertion on Coercivity (**Figure 4.18a**), TMR ratio (**Figure 4.18b**) and switching probabilities (**Figure 4.18c-f**) under different annealing temperatures. Firstly, as shown in **Figure 4.18a**, the coercivity (H_c) of all three samples reduces with the increasing of annealing temperature from 300C to 500C, indicating the appearance of PMA deterioration due to the diffusion of Ir atoms into CoFeB/MgO interface[175]. Secondly, Sample 1 and 3 shows a larger H_c than Sample 2 in the whole arrange of annealing temperature from 300C to 500C. For the annealing temperature at 300C and 350C, Sample 1 and Sample 3 have almost identical H_c , which is almost twice of the H_c of Sample 2. This result indicates that the PMA in Sample 2 can be degraded even at a moderate annealing temperature (i.e., 300C) if Fe-O and/or Co-O bonds at CoFeB/MgO interface was interrupt by Ir insertion[148], [176]. For annealing temperature at

400C, the H_c of Sample 3 decreases by 1/3 but still 4 times larger than that of Sample 2, while Sample 1 shows nearly no decrease of H_c . This result indicates that Sample 3 is more thermal robust of PMA than Sample 2 at annealing temperature of 400C, which is required for compatibility with back end of line (BEOL) process of CMOS technology[166]. For annealing at 450C and 500C, the no PMA can be measured for Sample 1 while H_c of Sample 3 keeps low comparing to that of Sample 1. Therefore, we can see that the insertion of Ir has negative influence on PMA, but it can be largely improved by depressing the Ir atoms diffusion, as in Sample 3. Another concern is that lower PMA could limit the switching speed of MTJ. As previously shown in **eq. 4.1**, the switching speed of voltage-driven magnetization switching is proportional to the in-plane component of external magnetic field. Therefore, the ultra-fast switching can be facilitated by a larger $H_{in-plane}$. However, for MTJs with PMA, the increase of $H_{in-plane}$ will low down the energy barrier between P and AP states given by $\Delta K_B T$, where

$$\Delta = \Delta_0 \left(1 - \frac{H_{in-plane}}{H_k} \right)^2, \quad \text{eq. 4.2}$$

and H_k is the anisotropy field and Δ is the thermal stability factor [177]. Since the decrease of thermal stability will lead to the instability of information storage, a large PMA is essential for the ultra-fast precessional switching in MTJ. And that is the reason why the fastest switching speed as 150 ps has been only achieved in Sample 1 in our study, as shown in **Figure 4.18b** and **c**.

Figure 4.18b shows the TMR evolution with the increase of annealing temperatures. As the annealing temperature increasing from 300C to 400C, all three samples show the increase of TMR. This TMR increasement can be explained by the crystallinity enhancement for the CoFeB layer as more B atoms migrates out with higher temperature[84], [157], [178], [179]. Sample 2 shows a lower TMR comparing to Sample 1 and 3 at each temperature, which is in a consistence with the results shown previously in **Figure 4.16d** and **Figure 4.16e**. When annealing temperature increases to 450C and 500C, the sample 2 lose its TMR due to the loss of its PMA[157], which has been indicated from **Figure 4.18a**. TMR of Sample 3 starts to gradually decreases due to the migration of Ir into CoFeB and MgO, while TMR of Sample 1 firstly increase at 450C and then decrease at 500C, due to competition between the migration of B atoms out of CoFeB[180] and the diffusion of Mo atoms into CoFeB and MgO[166]. These results shows that the insertion of Ir is also not preferred by TMR, but similar to PMA, Sample 3 shows better TMR than Sample 2 due to its successfully depression of Ir atom migrates towards CoFeB/MgO interface, especially at higher annealing temperature.

Figure 4.18c, d and **e** show the evolution of the product of switching probabilities with increasing pulse voltages from 1.4V to 2.3 V under an annealing temperature of 300C, 400C and 500C, respectively. The applied voltages here is relatively small in order to protect the MTJ nanodevices from breakdown. As shown in **Figure 4.18c**, for annealing temperature at 300C, with the increase of pulse voltages, all three samples show the similar trends [172]: P_{sw} firstly remain very low when pulse voltages (V) are below a critical threshold pulse voltage (V_{ct}), and then P_{sw} gradually increases to a higher level when V becomes larger than V_{ct} . When $V < V_{ct}$, the switching behavior is mostly dominated by thermal activation [181] since the VCMA effect is not sufficiently low down the energy barrier of MTJ. When $V = V_{ct}$, the magnetization starts to process around the effective magnetic field H_{eff} , which is the sum of the following three components: magnetic anisotropy field under VCMA effect, external magnetic field and stray field from reference layer of MTJ. Therefore, a smaller PMA and a larger VCMA coefficient

result in a lower V_{ct} . That is why Sample 2 has the lowest V_{ct} and then Sample 3, while Sample 1 has the highest V_{ct} . For pulse voltage $V > V_{ct}$, the precessional switching will gradually dominate the switching behavior and the P_{sw} will keep increasing until precessional switching overwhelms the thermal activation and fully dominates the switching behavior. No more obvious increase of P_{sw} will be observed when PMA is fully eliminated by voltage pulses and the magnetization starts to only process around the in-plane component of the external magnetic field [177]. (P_{sw} will be only slightly increase with even higher voltage pulse due to the switching errors induced by thermal fluctuation is further depressed.) As shown in **Figure 4.18c**, Sample 2 shows a larger P_{sw} at higher pulse voltages, indicating its higher VCMA coefficient than Sample 3 and Sample 1 at annealing temperature of 300C [182], [183].

As shown in **Figure 4.18d**, for the annealing temperature at 400C, Sample 1 shows a very similar behavior of P_{sw} with the increasing pulse voltage at 300C (**Figure 4.18c**), while Sample 2 and Sample 3 present distinct behavior: on one hand, the P_{sw} remains very low even with higher pulse voltage for Sample 1, but on the other hand, a remarkable increase of P_{sw} has been observed for Sample 3. This behavior can be understood by the possible hypotheses of the tailoring of doping gradient of Ir atoms near MgO/CoFeB interface in MTJ. When the annealing temperature increasing from 300C to 400C, the higher temperature enables a limited amount of Ir atoms to diffuse towards the MgO/CoFeB interface, forming a doping gradient of Ir inside MTJ structure: the Ir doping density is much lower at MgO/CoFeB interface in Sample 3 comparing to Sample 2. The low doping density of Ir atoms at MgO/CoFeB interface can bring two obvious advantages: Firstly, comparing to Sample 1, a relatively high of PMA is observed for Sample 3 (**Figure 4.18a**) at higher annealing temperature at 400C, which helps to avoid the severe P_{sw} degradation due to the loss of PMA, as Sample 2 (**Figure 4.18a, d**); Secondly, a relatively high VCMA coefficient can be induced even with low doping density of Ir atoms at MgO/CoFeB interface, which helps to obtaining a high P_{sw} of Sample 3 at higher pulse voltages (**Figure 4.18d**). Therefore, by carefully tailoring the Ir doping density near MgO/CoFeB interface (Sample 3), a sweet point of trade-off between high PMA (Sample 1) and high VCMA coefficient (Sample 2) can be obtained. Note that even for Sample 3, the high Ir doping density near MgO/CoFeB interface could ruin the P_{sw} easily due to the rapid decrease of PMA, which is the case for higher annealing temperature of 500C, as shown in **Figure 4.18e**. **Figure 4.18f** shows a summary of **Figure 4.18c, d** and **e** that how the peak of P_{sw} changes under different annealing temperatures. Sample 3 shows the highest P_{sw} at 400C among all three samples, together with its high TMR ratio $\sim 160\%$ (**Figure 4.18b**). This is the reason why the ultra-low writing energy ~ 3.2 fJ/bit can be achieved in Sample 3 with the product of switching probabilities above 80%, as shown as **Figure 4.18f**.

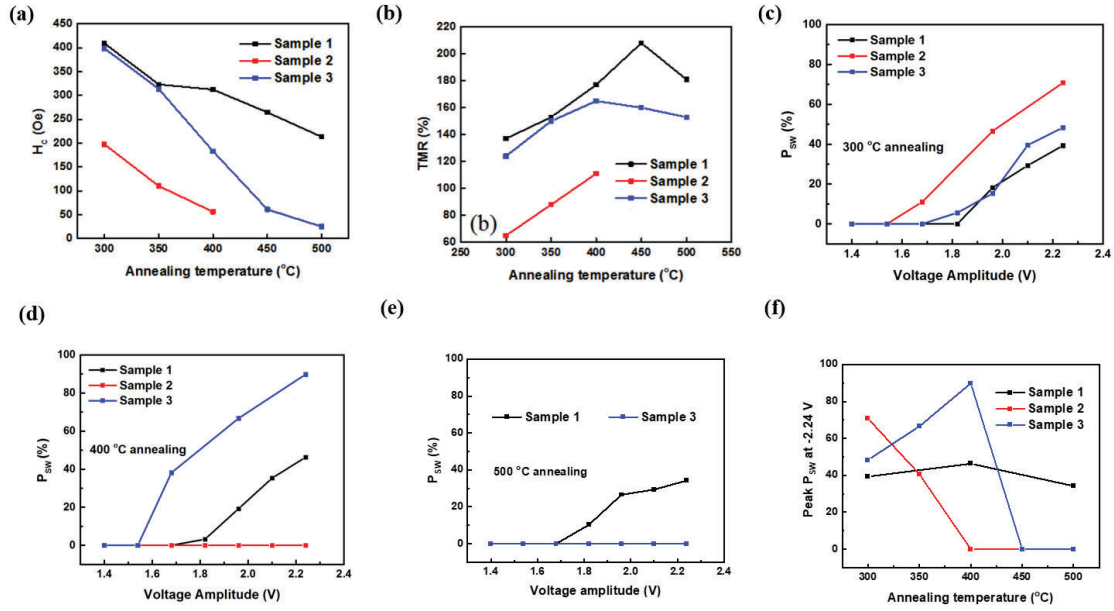


Figure 4.18 (a) Coercivity for Sample 1, 2 and 3 Under Different Annealing Temperatures, (b) TMR Ratio for Sample 1, 2 and 3 Under Different Annealing Temperatures, the Product of Switching Probabilities as a Function of Pulse Voltage Understand the Annealing Temperature of (c) 300C, (d) 400C and (e) 500C, (f) the Peak of the Product of Switching Probabilities for Sample 1, 2 and 3 Under Different Annealing Temperatures

In order to further verify the possible hypotheses of the doping gradient of Ir atoms near MgO/CoFeB interface in Sample 3 with 400C post-annealing, the cross-sectional energy dispersive X-ray spectroscopy (EDX) mapping and the elemental composition analysis across the section of high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) imaging were investigated. **Figure 4.19a** shows HAADF-STEM image of Sample 3 with 400C post-annealing. The selected region used for EDX mapping and elemental composition analysis are marked with red rectangle, which has the structure of Mo(1nm)/CoFeB(0.85nm)/MgO(2.5nm)/CoFeB(1.5nm)/Ir(1Å)/Mo(1.5nm). As shown in **Figure 4.19b**, the dispersion of Ir atoms into CoFeB layer as well as CoFeB/MgO interface is clearly confirmed in the EDX mapping of Co, Fe, Ir, Mg and O elements. This result is further clarified in the elemental composition analysis profile shown in **Figure 4.19c**, where a clear doping gradient of Ir near MgO/CoFeB interface is presented, with obvious lower Ir doping density at MgO/CoFeB interface than the peak of Ir signal. The results above confirm that a doping gradient of Ir element has been successfully created by carefully MTJ stack engineering (i.e., Ir insertion between CoFeB/cap layer) followed by rapid-annealing treatment at 400C. We can also conclude that keeping low Ir doping density at MgO/CoFeB interface not only can achieve the good VCMA effect, but also can avoid the deterioration of PMA, which is critical for high thermal stability, fast switching speed, and the low writing energy with high switching probability.

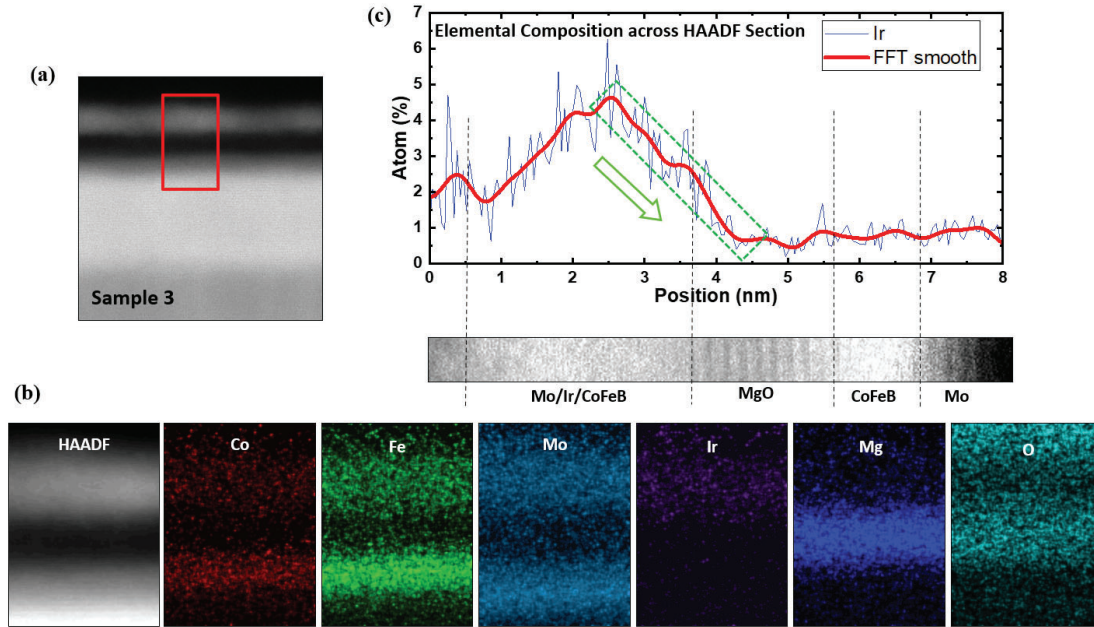


Figure 4.19 (a) High-angle Annular Dark-field Scanning Transmission Electron Microscopy (HAADF-STEM) Imaging of Sample 3 with Post-annealing of 400C, the Selected Region for the Further Analysis in (b) and (c) is Marked with Red Rectangle (b) Cross-sectional energy dispersive X-ray spectroscopy (EDX) maps of Co, Fe, Mo, Ir, Mg and O elements for Mo(1nm)/CoFeB(0.85nm)/MgO(2.5nm)/CoFeB(1.5nm)/Ir(1A)/Mo(1.5nm) structure in Sample 3. (c) Elemental composition scan of Ir element across the HAADF section and its corresponding HAADF-STEM image. The FFT smooth method was used to generate the curve for eye-guide.

4.4 CRAM

4.4.1 CRAM Experimental Demonstration by the ‘Circuit-around-Die’ Method

CRAM Logic Operations. Two-input CRAM logic operations are studied first. The output cell is first initialized by writing ‘0’ to it. Then two input cells are connected to the output cell through the LL by turning on the corresponding transistors. Voltage pulses of amplitude of V_{logic} , V_{logic} , and 0, are simultaneously applied to the two input cells and the output cell, respectively. This is the same as grounding the output cell while applying a voltage pulse of V_{logic} to the two input cells. Then depending on the input cells’ states, the output cell will have a certain probability of being switched from ‘0’ to ‘1’. Such a cycle of operations is repeated n times, and the statistical mean of the output logic state, $\langle D_{\text{out}} \rangle$, is obtained. The entire process is repeated for different V_{logic} values and input states. The basis for logic operations in the CRAM is the state-dependent resistance of the input cells. These shift and displace the output cell’s switching probability transfer curve. As a result, the output cell switches state based on specific input states, thereby implementing a logic function such as AND, OR, NAND, NOR, or MAJ. A specific initial state of the output cell and V_{logic} value corresponds to one of these logic gates[100]. The time duration or pulse width of the voltage pulse applied during a logic operation is expected to contribute to most of the time required to complete a logic operation. In the following, we use the term logic

speed to generally refer to the speed of a logic operation. Logic speed is approximately inversely proportional to the time duration of the voltage pulse used during a logic operation.

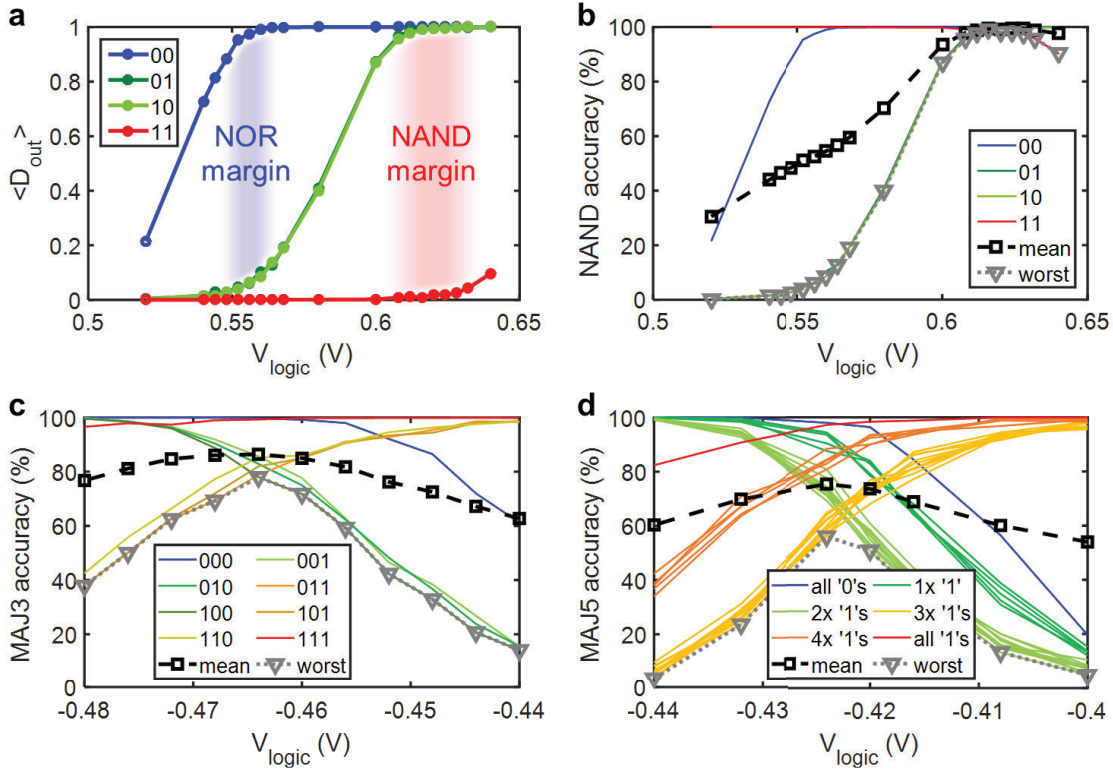


Figure 4.20 Experimental Results for CRAM Logic Operation[57]

a Output logic average, D_{out} , vs. logic voltage, V_{logic} . In a 2-input logic operation, two input cells and one output MTJ cell are involved. The output cell's terminal is grounded, while the common line is left floating. A logic operation voltage pulse is applied to the two input cells' terminals for a fixed duration (pulse width) of 1 ms. Before each logic operation, input data is written to the input cells. After each logic operation, the output cell's state is read. Each curve corresponds to a specific input state. Each data point represents the statistical average of the output cell's logic state, $\langle D_{out} \rangle$, sampled by 1000 repeats ($n = 1000$) of the operations. The separation between the $\langle D_{out} \rangle$ curves indicates the margins for NOR or NAND operation, highlighted in blue and red, respectively. *b* Accuracy of 2-input NAND operation vs. logic voltage, V_{logic} . The results in *a* can be converted into a more straightforward metric, accuracy, for the NAND truth table. The curve labeled 'mean' and 'worst' indicates the average and the worst-case accuracy across all input states, respectively. So, for NAND operation, the optimal logic voltage is indicated in such a plot where the mean or worst accuracy is maximized. *c*, *d* Accuracy of MAJ3 (*c*) and MAJ5 (*d*) logic operations vs. logic voltage, V_{logic} . Each curve corresponds to an input state or a group of input states. And each data point represents the statistical average of the output MTJ logic state sampled by $n = 1000$ and $n = 250$, for *c* and *d*, respectively.

Generally, for a given input state, $\langle D_{out} \rangle$ increases with increasing V_{logic} . The $\langle D_{out} \rangle$ response curves are input state-dependent. The four input states can be divided into three groups:

- The ‘00’ input state yields the lowest resistance at the two input cells, so the output cell switches from ‘0’ to ‘1’ first (with the lowest V_{logic}).
- The ‘11’ input state yields the highest resistance at the two input cells, so the output cell switches from ‘0’ to ‘1’ last (with the highest V_{logic}).
- The ‘01’ and ‘10’ input states both yield resistance that falls in between that of ‘00’ and ‘11’ so that the output cell’s response curve falls in between that of ‘00’ and ‘11’.

Figure 4.20a shows the experiment results. The two regions highlighted in blue and red that fall in between the three groups of response curves are suitable for NOR and NAND operations, respectively. For example, in the red region, the ‘11’ input has a high probability of yielding a ‘0’ output, while the other three input states have a high probability of yielding a ‘1’ output. This matches the expected truth table for a NAND logic gate. Therefore, if V_{logic} is chosen carefully—within the red region for the CRAM 2-input logic operation—the operation performed is highly likely to be NAND.

The experimental results of $\langle D_{\text{out}} \rangle$ can be converted into a straightforward format representing the accuracy for a specified logic function. This translation can be computed by simply subtracting $\langle D_{\text{out}} \rangle$ from 1 for those input states where a ‘0’ output is expected in the truth table of the logic function. **Figure 4.20b** shows the NAND accuracy of the same 2-input CRAM logic operation. The ‘mean’ and ‘worst’ plots are based on the average value and minimum value of the accuracy, respectively, across all input state combinations at a fixed value for V_{logic} . Based on the experimental results, if $V_{\text{logic}} = 0.624$ or 0.616 V, the CRAM delivers a NAND operation with a best mean and a worst-case accuracy of about 99.4% and 99.0%, respectively. From a circuit perspective, both increasing the effective TMR ratio of input cells and/or making the output cell’s response curve steeper would increase the vertical separation of these input-state-dependent curves, resulting in higher accuracy. For example, a higher effective TMR ratio of input cells results in a larger contrast of current in the output cell between different input states. Therefore, there is more ‘horizontal’ room to separate the $\langle D_{\text{out}} \rangle$ curves associated with different input states so that for the inputs with which the output is expected to be ‘0’ or ‘1’, the $\langle D_{\text{out}} \rangle$ of the output cell is closer to the expected value (‘0’ or ‘1’). Also note that for a logic operation, the ‘accuracy’ and ‘error rate’ are essentially two quantities describing the same thing: how true the logic operation is, statistically. By definition, the sum of accuracy and error rate is always 1. The higher or closer to 1 the accuracy is, the better. The lower or closer to 0 the error rate is, the better. Lastly, to facilitate better visualization of how the resistance changes of different input cell states are translated into voltage differences on the output cell resulting in it being switched or unswitched, we list the equivalent resistance of the two input cells combined in parallel and the resulting voltage on the output cell as follows: With $V_{\text{logic}} = 0.620$ V, the equivalent resistance of input cells and the resulting voltage on the output cell are 0.4133 V and 1120 Ω , 0.3753 V and 1461 Ω , and 0.3248 V and 2037 Ω , for input states ‘00’, ‘01’ or ‘10’, and ‘11’, respectively. Note that these values are estimated by the experiment-based modeling, which is introduced in the later part of this paper.

With more input cells, we also studied 3-input and 5-input majority logic operations. **Figure 4.20c** shows the accuracy of a 3-input MAJ3 logic operation. At $V_{\text{logic}} = -0.464$ V, both the optimal mean and the worst-case accuracy are observed to be 86.5% and 78.0%, respectively. Similarly, for a 5-input MAJ5 logic operation, shown in **Figure 4.20d**, both the optimal mean

and the worst-case accuracy are observed to be 75% and 56%, respectively. As expected, comparing 2-input, 3-input, and 5-input logic operations, the accuracy decreases with an increasing number of inputs.

CRAM Full Adder. Having demonstrated fundamental elements of CRAM—memory write operations, memory read operations, and logic operations—we turn to more complex operations. We demonstrate a 1-bit full adder. This device takes two 1-bit operands, A and B, as well as a 1-bit carry-in, C, as inputs, and outputs a 1-bit sum, S, and a 1-bit carry-out, C_{out}. A variety of implementations exist. We investigate two common designs: (1) one that uses a combination of majority and inversion logic gates, which we will refer to as a ‘MAJ + NOT’ design; and (2) one that uses only NAND gates, which we will refer to as an ‘all-NAND’ design. **Figure 4.21a, b** illustrate these designs.

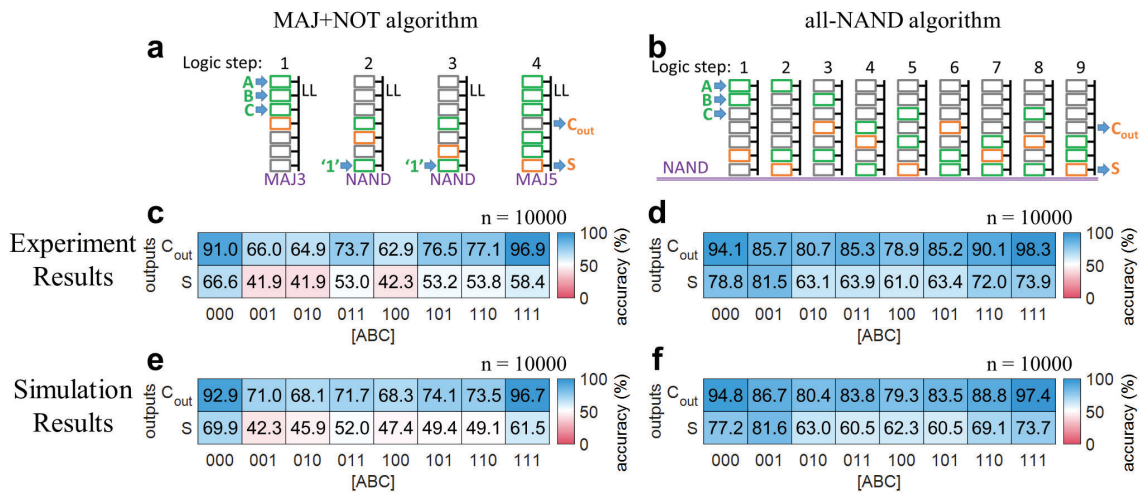


Figure 4.21 CRAM 1-bit Full Adder Demonstration Results[57]

a, b Illustrations of the ‘MAJ + NOT’ and ‘all-NAND’ 1-bit full adder designs. Green and orange letter symbols indicate input and output data for the full adder, respectively. From left to right, numbered by ‘logic step,’ each drawing shows the intended input (green rectangle) and output (orange rectangle) cells involved in the logic operation. The text in purple under each drawing indicates the intended function of the logic operation (MAJ3, NAND, or MAJ5). c–f Experimental (c, d) and simulation (e, f) results of the output accuracy of 1-bit full adder operations by CRAM with the MAJ + NOT (c, e) and all-NAND (d, f) designs. The CRAM adder’s outputs, S and C_{out}, are assessed against the expected values, i.e., their truth table, for all input states of A, B, and C. The accuracy of each result for each input state is shown by the numerical value in black font, as well as, represented by the color of the box with red (or blue) indicating wrong (or correct), or accuracy of 0% (100%). The accuracy is calculated based on the statistical average of outputs obtained by repeating the full adder execution n times, for n = 10,000. The experimental results for the MAJ + NOT (c) and all-NAND (d) designs are obtained by repeatedly executing the operation for all input states and observing the output states. The simulation results for the MAJ + NOT (e) and all-NAND (f) designs are obtained with probabilistic modeling, using Monte Carlo methods. The accuracy of individual logic operations is set to what was observed experimentally.

Figure 4.21c, e show the experimental and simulation results for the MAJ + NOT and the all-NAND designs, respectively. Each plot is a colormap that lists the accuracy of the output bits S and C_{out} , with each input state coded as [ABC]. The blue (red) indicates good/desired (bad/undesired) accuracy. In the boxes of colormap, results in saturated blue are the most desirable. The numerical values of accuracy are also labeled accordingly. From the experimental results for the MAJ + NOT design full adder shown in **Figure 4.21c**, we make two observations:

- The accuracy of C_{out} is generally higher than that of S. This is because C_{out} is directly produced by the first MAJ3 operation from inputs A, B, and C, while S is produced after additional logic operations. We also note that since C_{out} is produced earlier than S, it is less impacted by error propagation and accumulation during each step; and the MAJ5 involved in producing S is inherently less accurate than the MAJ3.
- Both C_{out} and S have higher accuracy when the input [ABC] = 000 or 111 than in the other cases. This is expected since the input states of all '0's and all '1's yield higher accuracy than those with mixed numbers of '0's and '1's for both MAJ3 and MAJ5.

The experimental results for the all-NAND design are shown in **Figure 4.21d**. The same observations regarding accuracy vs. inputs as the MAJ + NOT design apply. However, it is clear that the accuracy of the all-NAND full adder, at 78.5%, is higher than that of the MAJ + NOT full adder, at 63.8%. This is likely due to the fact that 2-input NAND operations are inherently more accurate than MAJ3 and MAJ5 operations. This offsets the impact of the additional steps required in the all-NAND design. We note that the accuracy of all computation blocks will improve as the underlying MTJ technology evolves. Accordingly, the relative accuracy of the all-NAND versus the MAJ + NOT designs may change[100].

Modeling and Analysis of CRAM Logic Accuracy. To understand the origin of errors, how they accumulate, and how they propagate, we performed numerical simulations of the full adder designs. These are based on probabilistic models of logic operations, implemented using Monte Carlo methods. **Figure 4.21e, f** shows the simulation results for the MAJ + NOT and all-NAND designs, respectively. In these simulations, the accuracy of individual logic operations was set to match what was experimentally observed. The simulation results for the overall designs of the full adders correspond well to what was observed experimentally for these, confirming the validity of the proposed probabilistic models.

We note that beyond the inherent inaccuracy of logic operations, other factors such as device drift and device-to-device variation in MTJ devices will contribute to error in a CRAM. Specifically, drifts in temperature, external magnetic field, MTJ anisotropy, and MTJ resistance can lead to drift of the response curve, $\langle D_{out} \rangle$. Most likely, any such drift will result in a reduction (increase) of accuracy (error rate).

On the other hand, the accuracy of logic operations will significantly benefit from improvements in TMR ratio as MTJ technology evolves. To project the future accuracy of CRAM operations, we employ various types of physical modeling informed by existing experimental results.

Three sets of assumptions on the accuracies (or error rates) of NAND logic operations underlie the following studies.

- The ‘experimental’ assumptions are based on the best accuracy experimentally observed among the 9 NAND steps involved with the all-NAND 1-bit full adder. These are adjusted linearly to ensure that the error for inputs ‘01’ and ‘10’ equals that for input ‘11’. In reality, as supported by the experimental results shown in **Figure 4.20a**, such a condition can be reached by properly tuning the V_{logic} . Therefore, assuming the gate-level error rate is already optimized by tuning the V_{logic} , then the per-input-state NAND accuracies can be further simplified so that an error rate, δ ($0 \leq \delta \leq 1$), can be used to characterize the error, accuracy, and probabilistic truth table of NAND operations in a CRAM. The NAND accuracy is $[1, 1-\delta, 1-\delta, 1-\delta]$, and the NAND probabilistic truth table is $[1, 1-\delta, 1-\delta, \delta]$, both being a function of δ . Through the above-mentioned modeling and calculations, the ‘experimental’ assumptions yield $\delta = 0.0076$, which corresponds to a TMR ratio of approximately 109%, based on experiments.
- Two additional sets of assumptions, labeled as ‘production’ and ‘improved’, assume MTJ TMR ratios of 200% and 300%, respectively. These two assumptions yield $\delta = 2.1 \times 10^{-4}$, and $\delta = 7.6 \times 10^{-6}$, respectively, based on modeling and calculations. The ‘production’ assumptions represent the current industry-level TMR ratios developed for STT-MRAM technologies. The ‘improved’ assumptions present reasonable expectations for near-future MTJ developments.

CRAM NAND error rates vs. TMR ratio with various logic voltage pulse widths are shown in **Figure 4.22a**. Higher TMR ratios and faster logic speed—so shorter V_{logic} pulse widths—lead to smaller error rates. Also included is an analysis of error rates vs. effective TMR ratio, which is independent of the specific TMR modeling. Note that, for all subsequent results, we will use the NAND error rate at the assumed TMR ratios, with pulse widths of 1 ms. This is more conservative but is consistent with the pulse widths used in the experimental results reported above.

Analysis of CRAM Multi-bit Adder, Multiplier, and Matrix Multiplier. With these defined sets of assumptions, we provide projections of CRAM accuracy at a larger scale for meaningful applications. First, we evaluate ripple-carry adders and array multipliers[184] operating on scalar operands, with up to 6 bits. To evaluate the results, we adopt the normalized error distance (NED) metric[185] to represent the error of these primitives, as it has been shown to be more suitable for arithmetic primitives in the presence of computational error. We will refer to the error for a given primitive as ‘NED error’. We also define a complementary metric of ‘NED accuracy’ as the NED subtracted from 1 and then multiplied by 100%, to facilitate a more intuitive visualization of the error values. While the ‘experimental’ assumptions with a TMR ratio of 109% yield good overall accuracy for adders and multipliers, as the TMR ratio increases, the ‘production’ assumption with a TMR ratio of 200%, and the ‘improved’ assumption with a TMR ratio of 300%, yield significantly better or higher accuracies. Specifically, a 4-bit adder produces NED error of 2.8×10^{-2} , 8.6×10^{-4} , and 3.3×10^{-5} , or NED accuracy of 97.2%, 99.914%, and 99.9967%, for the ‘experimental’, ‘production’, and ‘improved’ assumptions, respectively. A 4-bit multiplier produces NED error of 5.5×10^{-2} , 1.8×10^{-3} , and 6.6×10^{-5} , or NED accuracy of 94.5%, 99.82%, and 99.9934%, for the three sets of assumptions, respectively. It is expected that when comparing the adder to the multiplier, since the latter is more complex and involves more

gates, its accuracy is generally lower than that of the adder. Similarly, as the bit width of the adder or multiplier increases, their accuracy decreases.

Then, using these primitives, we evaluate dot-product operations, which form the basis of matrix multiplication. They are heavily employed in many applications in both conventional domains and machine intelligence. Dot products consist of element-wise multiplication of two unsigned integer vectors, followed by addition. We perform additions with binary trees to maintain smaller circuit depth. **Figure 4.22b** shows the NED error of a 4-bit 4×4 dot-product matrix multiplier with respect to various TMR ratio assumptions. Like the adders and multipliers, as the TMR ratio increases, the NED error decreases, or the NED accuracy improves. Specifically, a 4-bit 4×4 dot-product matrix multiplier produces an NED error of 0.11, 3.4×10^{-3} , and 1.2×10^{-4} , or NED accuracy of 89%, 99.66%, and 99.988%, for the ‘experimental’, ‘production’, and ‘improved’ assumptions, respectively. Also, when comparing different input sizes (e.g., 1×1 to 4×4), as expected, the NED error is larger for larger input sizes due to the increased number of gates involved.

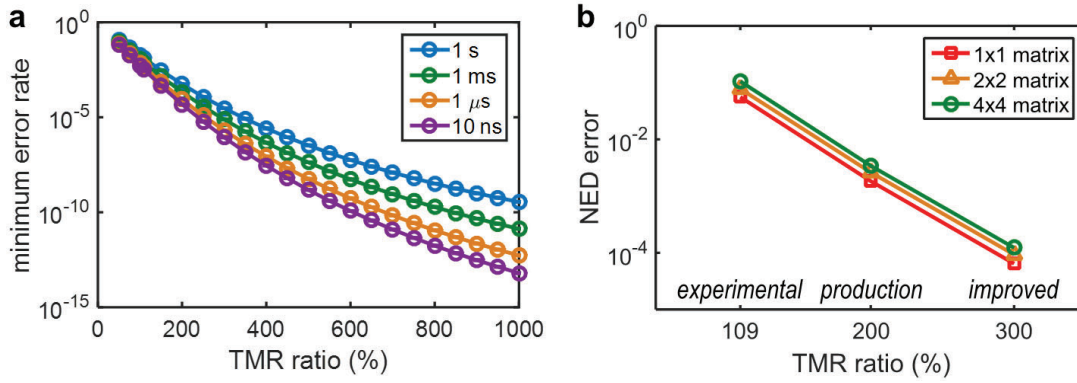


Figure 4.22 CRAM Accuracy Projections[57]

a NAND gate minimum error rate vs. MTJ TMR ratio with various Vlogic pulse widths. For a given TMR ratio, the error rate is a function of Vlogic. So, the ‘minimum error rate’ represents the minimum error rate achievable with an appropriate Vlogic value. All subsequent studies use the error rates observed with 1 ms pulse widths (to be consistent with the earlier experimental studies) at assumed TMR ratios. *b* The NED error of a 4-bit dot-product matrix multiplier vs.

TMR ratio. TMR ratios of 109%, 200%, and 300% are adopted for the ‘experimental,’ ‘production,’ and ‘improved’ assumptions, respectively. The size of the input matrix is indicated in the legend of the plot.

4.4.2 Stochastic Computing-CRAM (SC-CRAM)

Table 4.1 shows the number of MTJs (N_{MTJ}), the number of computation steps (N_{C}), and the subarray size for each arithmetic function described in Sections III-A and III-B in SC-CRAM. The values for N_{MTJ} and N_{C} are shown relative to conv-CRAM. Furthermore, calculations for each of these performance metrics compared computation on numbers with 8-bit resolution. For conv-CRAM, calculations were done on 8-bit digital numbers, and for SC-CRAM, calculations were performed on bit-streams with $N_{\text{B}} = 2^8 = 256$ bits. When calculating performance metrics in conv-CRAM, we assumed that addition and multiplication were performed using an 8-bit carry ripple adder (CRA) and an 8-bit Wallace tree multiplier (WTM), respectively. We also assumed

that subtraction and division in conv-CRAM were performed using full subtractor (FS) circuits and nonrestoring array dividers (NAD), respectively. Each of these circuits comprises NAND-based full adders (FAs). Finally, we assumed that square root in conv-CRAM was performed using three cycles of the Newton–Raphson (N–R) method.

The values in **Table 4.1** show that SC-CRAM uses significantly less MTJs for all functions than conv-CRAM. The reduction in N_{MTJ} in SC-CRAM is the most significant for multiplication and square root functions but the least significant for scaled addition and scaled subtraction functions. This is due to the large number of FAs needed for 8-bit WTM circuits and for the N–R method, which is not needed in SC-CRAM. However, N_C is larger in SC-CRAM than in conv-CRAM for all functions, except square root. This result is not unique to SC-CRAM and is observed when comparing any SC schemes to conventional deterministic schemes on binary numbers. The increase in N_C in SC-CRAM is most noticeable for scaled addition and absolute valued subtraction but least significant for multiplication and scaled division. Most notably, N_C is smaller for SC-CRAM than for conv-CRAM for square root. This is likely because we assumed that conv-CRAM repeated the N–R method three times to approximate the square root function, which increases N_C three times.

The subarray dimensions shown in **Table 4.1** reveal that multiple subarray rows are required for multiplication, division, and square root in conv-CRAM, whereas only one subarray row is required for all functions in SC-CRAM. A low number of subarray rows are very important in applications where a large number of arithmetic functions are performed in parallel since subarrays with a large number of rows are susceptible to parasitic effects such as a large IR drop and low noise margins. As we will see in the following, the large number of subarray rows in conv-CRAM makes it an unfeasible solution for certain applications.

The calculations shown in **Table 4.1** indicate several key features of SC-CRAM. One is that SC-CRAM is ideal for solving functions involving a large number of multiplication and scaled division functions (such as high-order polynomials and Maclaurin expansions of more complicated functions). Second is that both total circuit area and computation delay may decrease in SC-CRAM for more complex functions, as seen in the calculations for the square root function. While SC-CRAM is not well suited for tasks involving rapid, high-precision calculations, the results in **Table 4.1** suggest that SC-CRAM is ideal for applications that involve rough approximations to detect statistical anomalies in large datasets.

Table 4.1 Number of MTJs, Computation Steps, and Subarray Size for Various Arithmetic Functions in SC-CRAM[122]

Function	Gates used for SC-CRAM	Number of MTJs	Number of Computation steps	Sub-array (rows x columns)	
				SC-CRAM	Conv-CRAM
Scaled Addition	MUX	0.080X	14.3X	1 x 7	
	MAJ3	0.045X	10.7X	1 x 4	1 x 88
Multiplication	AND	0.006X	5.1X	1 x 4	16 x 161
Absolute valued Subtraction	XOR	0.090X	22.5X	1 x 8	1 x 90
Scaled Division	JK flip-flop	0.013X	2.0X	1 x 13	8 x 130
Square Root	AND-OR-OR	0.002X	0.49X	1 x 10	32 x 1413

**Assume 8-bit resolution for values shown.*

Then, we describe the three applications that were performed in this study: local image thresholding for character recognition, Bayesian inference for object location, and Bayesian belief network for heart disaster prediction. We evaluated the performance of SC-CRAM for each of these applications. For each application, the execution time, circuit area, energy usage, and noise margin are compared with those of conv-CRAM. The performance of SC-CRAM and conv-CRAM was evaluated in terms of circuit area, computation delay, and total energy consumption (E_{TOT}).

Table 4.2 Performance Evaluation of SC-CRAM[122]

Application	Sub-array size (rows x columns)		Number of CRAM cells	Number of Computation steps	Energy consumption	Noise margin (feasible?)	
	SC-CRAM	Conv-CRAM				SC-CRAM	Conv-CRAM
Local image thresholding	128 x 128	2048 x 4096	0.048X	0.377X	12.9X	2% (Yes)	-196% (No)
Object location	1 x 16	16 x 1024	0.005X	1.016X	1.12X	10% (Yes)	9% (Yes)
Heart disaster prediction	8 x 32	128 x 1024	0.005X	0.662X	1.18X	9% (Yes)	1% (Yes)
Kernel density estimation	32 x 64	512 x 2048	0.022X	0.674X	5.37X	8% (Yes)	-117% (No)

**Assume 8-bit resolution for values shown.*

Table 4.2 shows all of the calculations for N_{CELLS} , N_C , and E_{TOT} . In addition, **Table 4.2** shows the number of rows and the number of columns in the CRAM subarray. These data show that SC-CRAM uses approximately 20–180 times fewer CRAM cells than conv-CRAM for all four applications. Furthermore, SC-CRAM also has about 1/3–2/3 fewer computation steps than conv-CRAM for each application except the Bayesian inference system, where N_C is nearly equal between SC-CRAM and conv-CRAM. This shows that, for three of the four applications, the SC-CRAM circuit has less computation delay than the conv-CRAM circuit. This may seem counterintuitive since each step needs to be repeated 256 times in SC-CRAM, and however, the smaller N_C values in SC-CRAM can be attributed to two features of SC-CRAM. The first is the small number of cells required to perform both simple and complex arithmetic functions and the second is the ability to overlap the logic steps in some CRAM cells with the perturb and/or reset steps in other CRAM cells.

These data also show that the total energy consumption is approximately equal between SC-CRAM and conv-CRAM for the Bayesian inference system and the Bayesian belief network.

However, E_{TOT} is larger for SC-CRAM for local image thresholding and kernel density approximation by $12.9\times$ and $5.4\times$, respectively. It should be noted that conv-CRAM has nearly $40\times$ less energy consumption than modern near-memory processing systems [31]; therefore, SC-CRAM is still a competitive solution since E_{TOT} in SC-CRAM is on the same order of magnitude as E_{TOT} in conv-CRAM.

The subarrays for SC-CRAM have significantly less rows than conv-CRAM, which leads to significantly less IR drop [39], leading to an increase in noise margin. The noise margin represents the size of the range of voltages where a given logic operation can be performed. For example, a noise margin of 10% means that there is a 10% difference between the minimum and maximum voltage that can be applied during the logic operation while maintaining an acceptable output accuracy. Any noise margin below zero means that there are no voltage values that exist that can reliably produce an accurate output. **Table 4.2** shows that the noise margin is too low for local image thresholding and kernel density estimation to be feasible in conv-CRAM due to the large number of subarray rows. This shows that, despite the reduction in E_{TOT} in conv-CRAM, SC-CRAM is the preferred approach for local image thresholding and kernel density estimation because the larger noise margin in SC-CRAM makes it more reliable. The noise margin for the Bayesian inference system and the Bayesian belief network is large enough to be feasible in both SC-CRAM and conv-CRAM. However, the noise margin in SC-CRAM is still noticeably larger, and therefore, it is still the preferred method for these two applications.

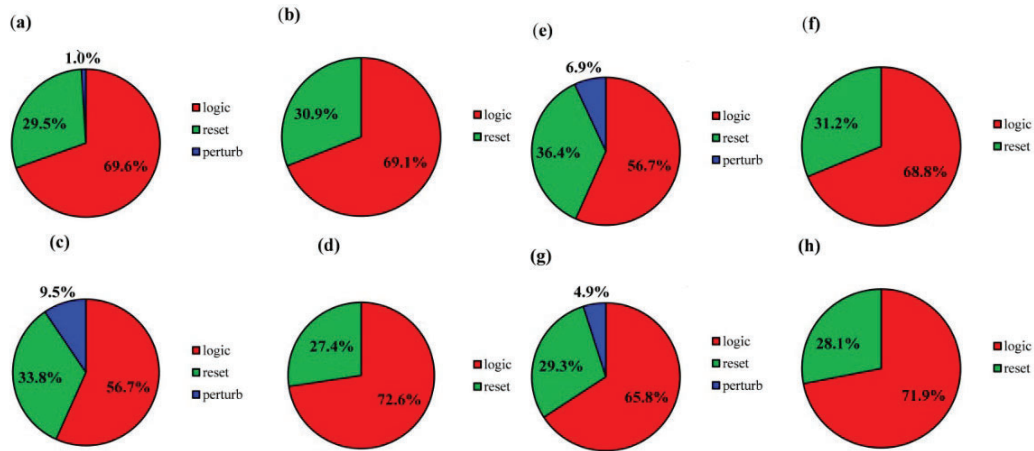


Figure 4.23 Energy Consumption Breakdown[122]

The breakdown of energy consumption in SC-CRAM and conv-CRAM for (a) and (b) local image thresholding, (c) and (d) object location via Bayesian inference, (e) and (f) heart disaster prediction via a Bayesian belief network, and (g) and (h) kernel density estimation.

The pie charts in **Figure 4.23** show the portion of E_{TOT} being consumed by the logic, reset, and perturb steps for SC-CRAM and conv-CRAM. Note that conv-CRAM does not use a perturb step, and therefore, only the logic and reset steps are considered. For conv-CRAM, the distribution of energy consumption between the reset and logic steps is nearly the same for all four applications, where approximately 25%–30% of E_{TOT} is consumed during the reset steps and 70%–75% of E_{TOT} is consumed during the logic steps. For SC-CRAM, the reset steps account for around 25%–30% of E_{TOT} . Most importantly, these charts show that adding the perturb step in SC-CRAM only accounts for a minority of the total energy consumption. It

should be noted that the Bayesian inference system consumes the most portion of E_{TOT} for the perturb step, and however, it still only consumes 9.5% of E_{TOT} . Conversely, generating stochastic bit-streams with conventional CMOS-based methods could increase the total energy consumption by nearly 80% [8]. These results indicate that SC-CRAM may outperform CMOS-based SC methods in terms of E_{TOT} .

4.4.3 CRAM Error Rate and Manufacturability Studies

In this section, the symbol d represents the MTJ device diameter, and the mean and standard deviation of d , is μ_d and σ_d , respectively. While μ_d is always set to be 100 nm, relative standard deviation of d , σ_d/μ_d , is used to represent the device-device size variation. And it will be referred to as ‘variation’ in short. The symbols V_{logic} and τ_{logic} , represent the amplitude and duration of the voltage pulse for a logic operation, respectively. Duration of the voltage pulse is also referred to as ‘speed’ since application of the voltage pulse for the duration specified by τ_{logic} does constitute most of time consumed in a logic operation of CRAM.

Figure 4.24 shows the statistics of NAND gate error rate vs. TMR ratio. The plot is made in standard box chart style: the middle red line indicates the median value; the top and bottom edges of the blue box indicate the 75th and 25th percentiles, respectively; and the ends of the extended black bars indicate the maximum and minimum values excluding outliers. The τ_{logic} and σ_d/μ_d parameter values for each subplot are listed on the plot. Firstly, if we focus on the median or minimum value of error rate, it can be seen from all three plots that the error rate decreases with increasing TMR ratio. As discussed previously in the Introduction section, increased TMR ratios result in more dramatic resistance contrast between different states of input MTJs. This leads to a larger voltage margin, separating the switched state from the non-switched state of the output MTJ, overcoming the stochastic switching nature, and making the gate more deterministic. Secondly, we can look further at the distribution of error rate (vertical size or spread of the box plot for each given TMR ratio value). If the MTJ devices are all identical (device-device size variation, σ_d/μ_d , is set to zero), the error rate would have followed the minimum values of the box chart for each TMR ratio value. Then if finite device-device size variation (σ_d/μ_d) is included, the error rate still decreases but it will exhibit its distributions. The spread of its distribution increases as the TMR increases. The distribution typically spreads several orders of magnitude for larger TMR ratios. This is mostly due to the nearly exponential dependence of switching probability on voltage, which is modulated by the device-device size variation approximately in first order. As a result, the error rate is sensitive to device-device size variation.

Furthermore, by comparing **Figure 4.24a** to **Figure 4.24b**, it is clear that faster speed (lower τ_{logic}) yields better (lower) error rate. When the thermal activation switching regime, one can plot the switching probability curves at various speeds on a normalized voltage axis, on which each curve is scaled horizontally to its own voltage that yields 50% switching probability. Then it will be clear that these switching probability curves on a normalized voltage axis becomes sharper (steeper) with faster speed. And as previously discussed, sharper switching probability curved will result in lower error rate. Another observation is that faster speed, however, also results in larger (worse) spread of error rate distribution for a given TMR ratio.

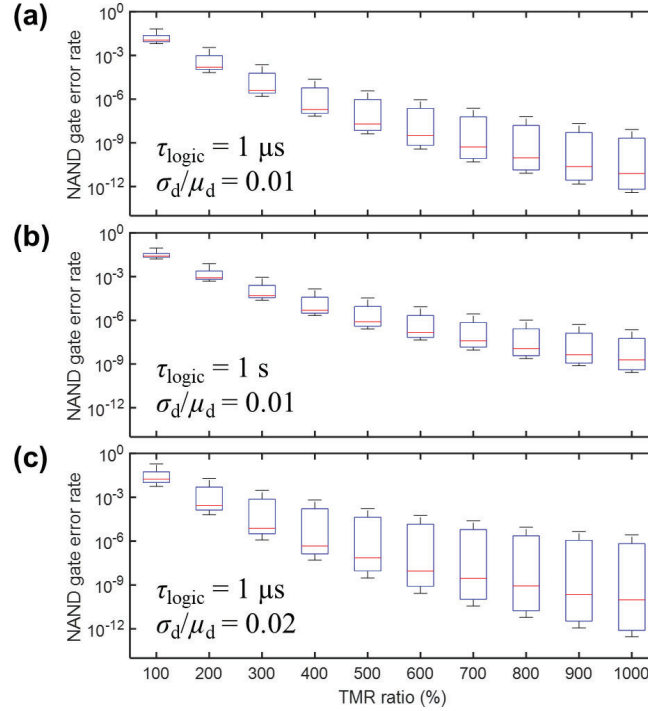


Figure 4.24 Box Plot of NAND Gate Error Rate vs. TMR Ratio[124]

The logic voltage duration, τ_{logic} , and relative device-device size variation, σ_d/μ_d , are $1\ \mu s$ and 0.01 , $1\ s$ and 0.01 , and $1\ \mu s$ and 0.02 , for (a), (b), and (c), respectively. Note that all plots share the same horizontal axis. Therefore, duplicate horizontal axes are omitted for simplicity.

Lastly, we use scattered point plots (**Figure 4.25**) to outline acceptable ranges or combinations of TMR ratio and σ_d/μ_d values given a specific set of error rate requirements. In each subfigure of **Figure 4.25**, two requirements are set for the both the median (labeled as ‘50th’) and the 99th percentile (labeled as ‘99th’) to be less than a specific value listed in the legend. In the plot, if a symbol is present, it means that the corresponding TMR ratio — σ_d/μ_d combination satisfies the specific error rate requirement. The blue circle and orange cross symbols represent the median and 99th percentile requirement, respectively. The median requirement represents the general or average requirement of error rate, while the 99th percentile requirement simulates the requirement for error rate in occasional ‘worst cases’. Different intended applications may require different percentile or error rate to be less than different thresholds. In this work and the following discussions, we use the 50th (median) and 99th percentile as an example of a set of requirements. And we will mostly discuss the area or margin on the TMR ratio — σ_d/μ_d plane where both requirements are satisfied.

The first observation is that the 50th and 99th percentile requirements outline areas on the TMR ratio — σ_d/μ_d plane of different shapes. This is why it is important to set more than one requirement to specify the statistical requirements on error rate in more detail. Secondly, the error rate requirements appear to translate into fairly rigid upper and lower thresholds for TMR ratio and σ_d/μ_d , respectively. And there is some tradeoff between TMR ratio and σ_d/μ_d . The third observation is made by comparing **Figure 4.25a** and **Figure 4.25b**, or **Figure 4.25c** and **Figure 4.25d** where the speed (τ_{logic}) is different. It is clear that faster speed opens up more margin on the TMR ratio — σ_d/μ_d plane. This echoes with the previous observation that faster speed leads

to lower error rate. The last observation can be made by comparing **Figure 4.25a** and **Figure 4.25c**, or **Figure 4.25b** and **Figure 4.25d**, where the error rate requirements are eased for the latter. Expectedly, when the requirements for error rate are eased, more areas on the TMR ratio — σ_d/μ_d plane become acceptable for the requirements.

The case shown in **Figure 4.25d** can be viewed as a more favorable yet still realistic case with faster speed and lower requirements for error rate. For example, the error rate requirements can be met with TMR ratio and σ_d/μ_d being brought to 350% and 0.005, respectively. Regarding the 350% TMR ratio target, although it is generally challenging to bring TMR ratio beyond 150% for low-RA product STT-operated MTJs with academic tools and processes, it is still possible to achieve TMR ratio beyond 600% at room temperature[186] if certain tradeoffs and optimizations are made. Whether the 0.005 target for σ_d/μ_d is viable is highly dependent on many factors, such as stability of repeatability of tools and processes, level of optimizations of tools and processes, yield, cost, etc. Note that the σ_d/μ_d of 0.005 is equivalent to the relative standard deviation of MTJ resistance of 0.01 if only the variation of device size is considered and is fully converted to variation of device resistance. If $\pm 3\sigma$ is considered, it is then equivalent to peak-to-peak relative variation of 0.06. In terms of whether the listed error rate requirement is adequate for a specific application, we believe that it should be carefully evaluated case-by-case. Generally, more complex applications impose more stringent (lower) requirements on the error rate. However, for some novel applications, such as neural network-based classifiers, the requirement on gate-level error rate may be much lower than that for traditional digital deterministic application, because the application accuracy is already expected to be less than perfect due to the limitations of the algorithm itself. As a reference point, it is found that an error rate of approximately 10^{-6} is adequate for a neural network based MNIST[187] handwritten digit classifier to yield useful and fairly accurate results [57]. We believe that the manufacturability or CRAM in this particular case may exist but could be challenging.

However, if some error correction techniques can be implemented in the CRAM logic operations, the error rate requirements can be further eased (increase) significantly. This will bring the target for TMR ratio and σ_d/μ_d to much more manageable ranges, increasing the likelihood of the manufacturability. Also new or novel MTJ switching mechanisms could help make the switching probability curve sharper and result in improvement of error rate. Certain self-correction mechanisms could also be a direction to investigate whether the device-device size variation can be improved.

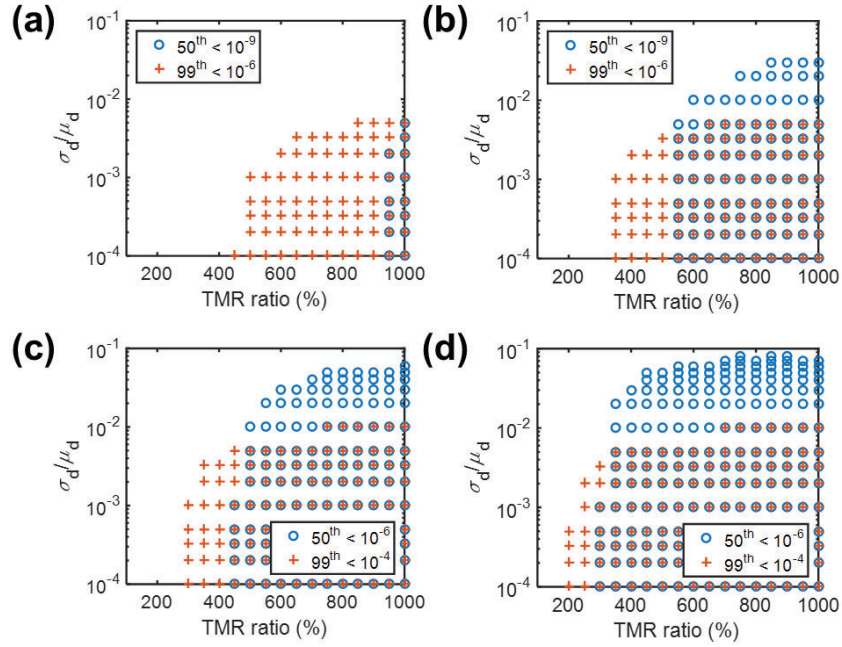


Figure 4.25 TMR ratio - σ_d/μ_d Combinations that Meet Specified Error Rate Requirements[124]

The presence of a blue circle or an orange cross symbol indicates that the corresponding combination meets the requirement of the 50th (median) or 99th percentile error rate, respectively. The specific thresholds of error rate requirements are listed in the legend of each plot. The pulse duration, τ_{logic} , for (a) and (c) is 1 s, and for (b) and (d) is 10 ns. Note that τ_{logic} of 10 ns is at the boundary of validity of the thermal activation switching model. Therefore, the results are expected to have larger deviation from reality.

5 CONCLUSIONS

As a conclusion, we successfully developed a set of advanced MTJ devices. We calculated and predicted a chiral MTJ structure with 10x higher TMR ratio ($\sim 10^4\%$) and 10x fast switching vs SOTA MTJs. We experimentally demonstrated the MTJ stacks that can be scalable down to 5 nm, substantially improved from the 15 nm of SOTA MTJs. This advanced scalability is achieved through two technical pathways: 1) utilizing bulk PMA materials; and 2) preparing MIFL stacks. Based on the $L1_0$ FePd SAF structure, we reported a novel switching mechanism VCEC, which is energy-efficient and deterministic. A decent TMR of 65% has been reached by optimizing the SAF spacer. An ultralow switching current density of $\sim 1.1 \times 10^5$ A/cm², one order of magnitude lower than the SOTA STT-MTJ devices, is reported with VCEC-MTJs. High quality $L1_0$ FePd and its SAF structures have been formed onto Si wafers, indicating a huge large-scale application potential of VCEC-pMTJ devices in semiconductor industry.

5.1 Theoretical Explorations

5.1.1 Performance Evaluation of Magnetic Tunnel Junction Based on Magnetic Weyl Semimetal

We have proposed magnetic Weyl semimetals (MWS) as candidates for advanced magnetic tunnel junctions (**Figure 4.1a**). It has been shown that the Fermi arc states, a topological property of magnetic Weyl semimetals, gives rise to a giant spin-transfer torque efficiency in asymmetric MTJ with a Weyl semimetal contact. The chirality of electrons in magnetic Weyl semimetals have been shown to give rise to a new mechanism to enable giant TMR ratios in MTJs, at least one order of magnitude larger than state-of-the-art. These exotic materials have been demonstrated, through these works, to enable unprecedented improvements on the TMR ratios and Spin-transfer torque efficiency in MTJ devices, opening up opportunities toward next generation spintronic devices.

5.1.2 Interlayer Exchange Coupling on Magnetic Tunnel Junctions

We have studied the simultaneous impact of VCMA effect, spin transfer torque and voltage-controlled interlayer exchange interaction in asymmetric MTJs. The results indicate an enhancement in energy efficiency in these systems and the possibility of full magnetization reversal through non-equilibrium interlayer exchange interaction alone.

5.1.3 Investigation of Unconventional SOT on Magnetization Switching

We have studied the simultaneous impact SOTs from multi-spin components on the dynamics of a perpendicular magnetization in bilayer heterostructures in the small intrinsic damping limit. Our results reveal that nonmagnets displaying large conventional spin Hall angles, θ_y , might be detrimental to full deterministic switching in this system. We also establish that highly efficient deterministic switching is attained at larger unconventional spin Hall angles, θ_z , within a specific dynamical regime constrained by the ratio θ_z/θ_y . Our work clarifies the physics and performance metric necessary for highly energy-efficient switching of a perpendicular magnetization in symmetry-broken bilayer heterostructures

5.1.4 Domain Wall Motion in Asymmetric MTJ

We demonstrate that DW propagation can be made efficient in a vertical current injection MTJs with in-plane magnetization by increasing the asymmetry of the ferromagnetic contacts. These results have promising implications in improving speed and efficiency for applications like spintronic interconnects, where vertical current injection has already been shown to be effective, as well as in cases where 2-terminal devices are advantageous. The current density efficiency results are also significant for neuromorphic computing applications, where synaptic updates done by displacing the domain wall must be done for arrays of hundreds of thousands of devices at a time.

5.1.5 Spin Hall Conductivity in Chiral Materials

We systematically investigated the intrinsic spin Hall effect (SHE) in topological semimetallic systems with high Chern numbers, which suggests that SHE enhancement can occur under the energetic alignment of high Chern number Kramers-Weyl points (KWPs). Our work indicates the existence of a mechanism the results in large spin Hall angles in these systems. Through intensive first principles calculations, we identified BiTePd and BiPdSe as being a promising spin Hall channel candidate to provide highly energy-efficient magnetization switching via SOTs.

5.2 MTJs with Bulk PMA

5.2.1 VCEC Switching of MTJs with $L1_0$ FePd-based Dual-SAF Free Layers

We experimentally demonstrated and theoretically confirmed the bipolar E-field switching of the p-MTJ devices with an SAF free layer. The E-field can modulate the sign of the IEC of the SAF structure and also induce a linear IEC torque, which can switch the p-MTJ devices bidirectionally. The p-MTJ devices switched by the VCEC show an order of magnitude lower energy per write operation compared to that of reported STT-switched MTJ devices, and $5.5\times$ higher densities than those for SOT-switched MTJ devices are possible, based on a benchmarking simulation for a 7-nm node architecture. By further optimizing the spacer, ferromagnetic layer, and MTJ stack, VCEC-MTJ could provide many functionalities for realizing the ultra-low-energy consumption and ultrafast speed memory and logic applications.

5.2.2 A substantial TMR Ratio of $L1_0$ FePd-based MTJs

We have successfully fabricated nano-sized $L1_0$ FePd PMA-MTJ devices, achieving an enhanced TMR of up to 65%, a record for such an advanced MTJ. Higher TMR up to 200% is expected, considering the usage of CoFeB/MgO/CoFeB interfaces, through the further optimization of the deposition and patterning processes, and especially through pinning of the top CoFeB layer to a hard reference layer. VSM results confirmed the PMA of the stack and showed two switching steps, which respectively correspond to the top CoFeB layer and the composite of the bottom CoFeB layer and the $L1_0$ FePd layer. Electrical-driven bi-directional switching was observed with an ultralow J_{sw} of $\sim 1.4\times 10^5$ A/cm², which is about one order of magnitude lower than that of typical STT-MTJs. To ensure that the switching is associated with magnetization reversal, we verified the Hext-sensitivity of the switching characteristics and thus possibility of resistive

switching should be excluded. Based on the sample stack structure and device performance, two possible switching mechanisms were proposed. One is the VCEC-driven switching of the bottom CoFeB layer, in which its IEC with the L1₀ FePd layer can be modified via voltage. The other is the STT-driven switching of the L1₀ FePd–CoFeB composite, in which the two layers switch together and serve as the free layer. These results shed light on the application of L1₀ FePd as an important material choice in PMA MTJs and offer promising directions for the development of spintronics-based memory and logic devices with ultralow energy consumption.

5.2.3 High-quality L1₀ FePd and its SAF Prepared on Si Wafers

In summary, high-quality L1₀-FePd(001) thin films were deposited on industry-compatible Si/SiO₂ wafers. Without expensive MgO(001) or SrTiO₃(001) substrates, the FePd thin films still possess strong PMA and low α with suitable seed and buffer layers. Furthermore, by adding an Ir spacer, p-SAFs with markedly improved IEC were also achieved. The high performance of both the L1₀-FePd single layer and its SAF indicates high crystallinity, strong (001) texture, as well as an epitaxial growth on amorphous SiO₂ surface. Comprehensive characterizations were performed to reveal the lattice details of the MgO seed layer, Cr/Pt bilayer buffer, L1₀-FePd, and its SAF and the epitaxial relationship throughout the stack. As MgO, the crucial seed layer material in this study, has been widely used by the spintronics community and is deposited by conventional RF sputtering, our results removed a key obstacle of applying L1₀-FePd for deeply-scalable spintronic devices.

5.3 MTJs with Interfacial PMA

5.3.1 pMTJ with Multiple Interface Free Layers

We have developed pMTJs with MIFL that can simultaneously afford high TMR, strong retention and efficient switching for MRAM cells that are scaled down to small dimensions. Different nonmagnetic materials have been explored as the coupling layer in the MIFL. The TMR of the pMTJ exhibits a strong dependence on the thickness of the first CoFeB layer. Large TMR above 200% has been achieved after 400°C annealing in pMTJs where three CoFeB layers are incorporated in the MIFL.

5.3.2 pMTJs with Ultralow Switching Energy

We have demonstrated the voltage-driven MTJ nanodevices with high TMR > 160%, ultra-fast switching speed <100 ps and ultra-low switching energy <3.2 fJ/bit (100 nm). Through the systematically investigating the PMA, TMR and the switching probabilities for different Ir-inserted MTJ structures under a series of annealing temperatures, we found that a low Ir doping density near MgO/CoFeB interface which can be introduced by a doping gradient of Ir element inside MTJ structure is critical to obtain high TMR, low PMA deterioration and good VCMA coefficient. The existence of Ir doping gradient was confirmed by EDX mapping and elemental composition analysis. An effective method to create such doping gradient of Ir element is via MTJ stack engineering followed by rapid-annealing treatment. These voltage-driven MTJ nanodevices with high thermal stability, fast switching speed, and the low switching energy provide potential chances for next generation magnetoresistance random access memory.

5.4 CRAM

5.4.1 CRAM Experimental Demonstration by the ‘Circuit-around-die’ Method

In summary, this work serves as the first step in experimentally demonstrating the viability, feasibility, and realistic properties of MTJ-based CRAM hardware. Through modeling and simulation, it also lays out the foundation for a coherent view of CRAM, from the device physics level up to the application level. Prior work had established the potential of CRAM through numerical simulation only. Accordingly, there had been considerable interest in the unique features, speed, power, and energy benefits of the technology. This study puts the earlier work on a firm experimental footing, providing application-critical metrics of gate-level accuracy or error rate and linking it to the application accuracy. It paves the way for future work on large-scale applications, in conventional domains as well as new ones emerging in machine intelligence. It also indicates the possibility of making competitive large-scale CMOS-integrated CRAM hardware.

5.4.2 SC-CRAM

In this study, we introduced a method in which CRAM can be used to generate and compute stochastic bit-streams called SC-CRAM. In SC-CRAM, the SBG and computation steps are embedded within the same circuit block, meaning zero cost in the total circuit area and computation delay to generate stochastic bit-streams. We compared the performance metrics between SC-CRAM and conv-CRAM for local image thresholding, object location, heart disaster prediction, and kernel density estimation. We determined that SC-CRAM is very efficient at computing complicated functions, which reduces the total number of CRAM cells required to perform these four applications without sacrificing the total computation delay. Furthermore, the energy consumption was on the same order of magnitude for SC-CRAM and conv-CRAM except for local image thresholding. Finally, SC-CRAM had a significantly larger noise margin for local image thresholding and kernel density estimation, making it a much more feasible solution for these applications.

5.4.3 CRAM Error Rate and Manufacturability Studies

The accuracy or error rate of CRAM logic operations is an application-critical performance metric. This work steps on the experiment-based modeling of CRAM logic operations from the previous work and expands it to account for the effects of device-device size variation. It is found that the CRAM logic error rate is sensitive to both the TMR ratio and device-device size variation. When the device-device size variation increases, the error rate increases. And the high-end tail of error rate distribution spreads several orders of magnitude. In order for CRAM to become a viable solution to some meaningful applications, its manufacturability needs to be assessed with the awareness of device-device size variation. Certain error rate requirements are imposed and the allowable margins for both the TMR ratio and device-device size variation are investigated. It is found that both must reach certain thresholds. However, there is some tradeoff possible between TMR ratio and device-device size variation. This work serves as the first step to investigate the manufacturability of CRAM hardware. It calls for developments of advanced MTJ devices, new switching mechanisms, and CRAM error correction techniques.

6 REFERENCES

- [1] T. Miyazaki and N. Tezuka, “Spin polarized tunneling in ferromagnet/insulator/ferromagnet junctions,” *J. Magn. Magn. Mater.*, vol. 151, no. 3, pp. 403–410, Dec. 1995, doi: 10.1016/0304-8853(95)00563-3.
- [2] 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, 1995, doi: 10.1103/PhysRevLett.74.3273.
- [3] R. Sbiaa, H. Meng, and S. N. Piramanayagam, “Materials with perpendicular magnetic anisotropy for magnetic random access memory,” *Phys. status solidi – Rapid Res. Lett.*, vol. 5, no. 12, pp. 413–419, Dec. 2011, doi: 10.1002/PSSR.201105420.
- [4] Z. Diao *et al.*, “Spin-transfer torque switching in magnetic tunnel junctions and spin-transfer torque random access memory,” *J. Phys. Condens. Matter*, vol. 19, no. 16, p. 165209, Apr. 2007, doi: 10.1088/0953-8984/19/16/165209.
- [5] S. Mizukami *et al.*, “Long-lived ultrafast spin precession in manganese alloys films with a large perpendicular magnetic anisotropy,” *Phys. Rev. Lett.*, vol. 106, no. 11, p. 117201, Mar. 2011, doi: 10.1103/PHYSREVLETT.106.117201/FIGURES/3/MEDIUM.
- [6] S. Mizukami, A. Sugihara, S. Iihama, Y. Sasaki, K. Z. Suzuki, and T. Miyazaki, “Laser-induced THz magnetization precession for a tetragonal Heusler-like nearly compensated ferrimagnet,” *Appl. Phys. Lett.*, vol. 108, no. 1, p. 12404, Jan. 2016, doi: 10.1063/1.4939447/30775.
- [7] S. Iihama *et al.*, “Low precessional damping observed for L10-ordered FePd epitaxial thin films with large perpendicular magnetic anisotropy,” *Appl. Phys. Lett.*, vol. 105, no. 14, p. 142403, Oct. 2014, doi: 10.1063/1.4897547.
- [8] S. Iihama, A. Sakuma, H. Naganuma, M. Oogane, S. Mizukami, and Y. Ando, “Influence of L10 order parameter on Gilbert damping constants for FePd thin films investigated by means of time-resolved magneto-optical Kerr effect,” *Phys. Rev. B*, vol. 94, no. 17, p. 174425, Nov. 2016, doi: 10.1103/PHYSREVB.94.174425/FIGURES/10/MEDIUM.
- [9] S. Iihama *et al.*, “Magnetization Dynamics and Damping for L1 0-FePd Thin Films with Perpendicular Magnetic Anisotropy,” *J. Magn. Soc. Japan*, vol. 39, no. 2, pp. 57–61, 2015.
- [10] D. Lyu *et al.*, “Sputtered L10-FePd and its Synthetic Antiferromagnet on Si/SiO₂ Wafers for Scalable Spintronics,” *Adv. Funct. Mater.*, vol. 33, no. 18, p. 2214201, May 2023, doi: 10.1002/ADFM.202214201.
- [11] D. Zhang *et al.*, “Bipolar Electric-Field Switching of Perpendicular Magnetic Tunnel Junctions through Voltage-Controlled Exchange Coupling,” *Nano Lett.*, vol. 22, no. 2, pp. 622–629, Jan. 2022, doi: 10.1021/acs.nanolett.1c03395.
- [12] J.-H. Park, C. Park, T. Jeong, M. T. Moneck, N. T. Nufer, and J.-G. Zhu, “CoPt multilayer based magnetic tunnel junctions using perpendicular magnetic anisotropy,” *J. Appl. Phys.*, vol. 103, no. 7, p. 07A917, 2008, doi: 10.1063/1.2838754.
- [13] P. Bruno, “Tight-binding approach to the orbital magnetic moment and magnetocrystalline anisotropy of transition-metal monolayers,” *Phys. Rev. B*, vol. 39, no. 1, pp. 865–868, 1989.
- [14] G. Van Der Laan, “Microscopic origin of magnetocrystalline anisotropy in transition metal thin films,” *J. Phys. Condens. Matter*, vol. 10, no. 14, pp. 3239–3253, 1999, doi: 10.1088/0953-8984/10/14/012.

- [15] J. Zhang, P. V. Lukashev, S. S. Jaswal, and E. Y. Tsymbal, “Model of orbital populations for voltage-controlled magnetic anisotropy in transition-metal thin films,” *Phys. Rev. B*, vol. 96, no. 1, p. 014435, 2017, doi: 10.1103/PhysRevB.96.014435.
- [16] S. Ikeda *et al.*, “A perpendicular-anisotropy CoFeB-MgO magnetic tunnel junction,” *Nat. Mater.*, vol. 9, no. 9, pp. 721–724, Sep. 2010, doi: 10.1038/nmat2804.
- [17] 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, 2012, doi: 10.1038/nmat3171.
- [18] T. Maruyama *et al.*, “Large voltage-induced magnetic anisotropy change in a few atomic layers of iron,” *Nat. Nanotechnol.*, vol. 4, no. 3, pp. 158–61, Mar. 2009, doi: 10.1038/nnano.2008.406.
- [19] 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, doi: 10.1038/nmat3172.
- [20] 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, 2012, doi: 10.1063/1.4753816.
- [21] Z. Wang *et al.*, “Atomic-Scale Structure and Local Chemistry of CoFeB-MgO Magnetic Tunnel Junctions,” *Nano Lett.*, vol. 16, no. 3, pp. 1530–1536, Mar. 2016, doi: 10.1021/ACS.NANOLETT.5B03627/ASSET/IMAGES/LARGE/NL-2015-03627B_0005.JPEG.
- [22] W. Butler, X.-G. Zhang, T. Schulthess, and J. MacLaren, “Spin-dependent tunneling conductance of Fe|MgO|Fe sandwiches,” *Phys. Rev. B*, vol. 63, no. 5, p. 054416, Jan. 2001, doi: 10.1103/PhysRevB.63.054416.
- [23] 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, doi: 10.1103/PhysRevB.63.220403.
- [24] A. Hallal, H. X. Yang, B. Dieny, and M. Chshiev, “Anatomy of perpendicular magnetic anisotropy in Fe/MgO magnetic tunnel junctions: First-principles insight,” *Phys. Rev. B - Condens. Matter Mater. Phys.*, vol. 88, no. 18, p. 184423, Nov. 2013, doi: 10.1103/PHYSREVB.88.184423/FIGURES/5/MEDIUM.
- [25] K. Y. Camsari, A. Z. Pervaiz, R. Faria, E. E. Marinero, and S. Datta, “Ultrafast Spin-Transfer-Torque Switching of Synthetic Ferrimagnets,” *IEEE Magn. Lett.*, vol. 7, pp. 1–5, 2016, doi: 10.1109/LMAG.2016.2610942.
- [26] S. Noh, D. H. Kang, and M. Shin, “Simulation of Strain-Assisted Switching in Synthetic Antiferromagnetic Free Layer-Based Magnetic Tunnel Junction,” *IEEE Trans. Magn.*, vol. 55, no. 4, Apr. 2019, doi: 10.1109/TMAG.2019.2898947.
- [27] K. Yamada, K. Kubota, and Y. Nakatani, “Reduction of the switching current in perpendicularly magnetized nanomagnets using an antiferromagnetic coupling structure,” *J. Appl. Phys.*, vol. 127, no. 13, p. 133906, Apr. 2020, doi: 10.1063/5.0005472/156745.
- [28] E. Raymenants *et al.*, “Nanoscale domain wall devices with magnetic tunnel junction read and write,” *Nat. Electron. 2021* 46, vol. 4, no. 6, pp. 392–398, Jun. 2021, doi: 10.1038/s41928-021-00593-x.
- [29] C. Yoshida *et al.*, “Enhanced thermal stability in perpendicular top-pinned magnetic tunnel junction with synthetic antiferromagnetic free layers,” *IEEE Trans. Magn.*, vol. 49,

- no. 7, pp. 4363–4366, 2013, doi: 10.1109/TMAG.2013.2248350.
- [30] D. J. P. De Sousa, F. Xue, J. P. Wang, P. M. Haney, and T. Low, “Current-induced torques in magnetic Weyl semimetal tunnel junctions,” *Phys. Rev. B*, vol. 103, no. 24, p. L241103, Jun. 2021, doi: 10.1103/PHYSREVB.103.L241103/FIGURES/4/MEDIUM.
 - [31] D. J. P. De Sousa, C. O. Ascencio, P. M. Haney, J. P. Wang, and T. Low, “Gigantic tunneling magnetoresistance in magnetic Weyl semimetal tunnel junctions,” *Phys. Rev. B*, vol. 104, no. 4, p. L041401, Jul. 2021, doi: 10.1103/PHYSREVB.104.L041401/FIGURES/3/MEDIUM.
 - [32] D. J. P. De Sousa, P. M. Haney, D. L. Zhang, J. P. Wang, and T. Low, “Bidirectional switching assisted by interlayer exchange coupling in asymmetric magnetic tunnel junctions,” *Phys. Rev. B*, vol. 101, no. 8, p. 081404, Feb. 2020, doi: 10.1103/PhysRevB.101.081404.
 - [33] C. O. Ascencio, W. Jiang, D. J. P. De Sousa, S. Lee, J. P. Wang, and T. Low, “Enhanced spin Hall response from aligned Kramers-Weyl points in high Chern number semimetals,” *Phys. Rev. B*, vol. 108, no. 20, p. L201404, Nov. 2023, doi: 10.1103/PHYSREVB.108.L201404/FIGURES/4/MEDIUM.
 - [34] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
 - [35] M. I. Jordan and T. M. Mitchell, “Machine learning: trends, perspectives, and prospects,” *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 2015, doi: 10.1126/science.aaa8415.
 - [36] G. Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions,” *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, Jun. 2005, doi: 10.1109/TKDE.2005.99.
 - [37] G. Hinton *et al.*, “Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups,” *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, 2012, doi: 10.1109/MSP.2012.2205597.
 - [38] R. Collobert, J. Weston, J. Com, M. Karlen, K. Kavukcuoglu, and P. Kuksa, “Natural language processing (almost) from scratch,” *J. Mach. Learn. Res.*, vol. 12, pp. 2493–2537, 2011, doi: 10.5555/1953048.2078186.
 - [39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
 - [40] K. S. Oh and K. Jung, “GPU implementation of neural networks,” *Pattern Recognit.*, vol. 37, no. 6, pp. 1311–1314, Jun. 2004, doi: 10.1016/j.patcog.2004.01.013.
 - [41] D. Strigl, K. Kofler, and S. Podlipnig, “Performance and scalability of GPU-based convolutional neural networks,” in *2010 18th Euromicro Conference on Parallel, Distributed and Network-based Processing*, 2010, pp. 317–324, doi: 10.1109/PDP.2010.43.
 - [42] E. Nurvitadhi, D. Sheffield, J. Sim, A. Mishra, G. Venkatesh, and D. Marr, “Accelerating binarized neural networks: comparison of FPGA, CPU, GPU, and ASIC,” in *2016 International Conference on Field-Programmable Technology (FPT)*, May 2017, pp. 77–84, doi: 10.1109/FPT.2016.7929192.
 - [43] J. Sawada *et al.*, “TrueNorth ecosystem for brain-inspired computing: scalable systems, software, and applications,” in *SC ’16: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, Jul. 2016, pp. 130–141, doi: 10.1109/SC.2016.11.

- [44] N. P. Jouppi *et al.*, “In-datacenter performance analysis of a tensor processing unit,” in *Proceedings of the 44th Annual International Symposium on Computer Architecture*, 2017, pp. 1–12, doi: 10.1145/3079856.3080246.
- [45] Y. H. Chen, T. Krishna, J. S. Emer, and V. Sze, “Eyeriss: an energy-efficient reconfigurable accelerator for deep convolutional neural networks,” *IEEE J. Solid-State Circuits*, vol. 52, no. 1, pp. 127–138, Jan. 2017, doi: 10.1109/JSSC.2016.2616357.
- [46] S. Yin *et al.*, “A high energy efficient reconfigurable hybrid neural network processor for deep learning applications,” *IEEE J. Solid-State Circuits*, vol. 53, no. 4, pp. 968–982, Apr. 2018, doi: 10.1109/JSSC.2017.2778281.
- [47] J. Borghetti, G. S. Snider, P. J. Kuekes, J. J. Yang, D. R. Stewart, and R. S. Williams, “Memristive switches enable stateful logic operations via material implication,” *Nature*, vol. 464, no. 7290, pp. 873–876, Apr. 2010, doi: 10.1038/nature08940.
- [48] P. Chi *et al.*, “PRIME: a novel processing-in-memory architecture for neural network computation in ReRAM-based main memory,” in *2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA)*, Jun. 2016, pp. 27–39, doi: 10.1145/3007787.3001140.
- [49] A. Shafiee *et al.*, “ISAAC: a convolutional neural network accelerator with in-situ analog arithmetic in crossbars,” in *2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA)*, Jun. 2016, pp. 14–26, doi: 10.1145/3007787.3001139.
- [50] M. Hu *et al.*, “Dot-product engine for neuromorphic computing: programming 1T1M crossbar to accelerate matrix-vector multiplication,” in *2016 53rd ACM/EDAC/IEEE Design Automation Conference (DAC)*, Jun. 2016, pp. 1–6, doi: 10.1145/2897937.2898010.
- [51] V. Seshadri *et al.*, “Ambit: in-memory accelerator for bulk bitwise operations using commodity DRAM technology,” in *2017 50th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*, 2017, pp. 273–287, doi: 10.1145/3123939.3124544.
- [52] P. Yao *et al.*, “Fully hardware-implemented memristor convolutional neural network,” *Nature*, vol. 577, no. 7792, pp. 641–646, Jan. 2020, doi: 10.1038/s41586-020-1942-4.
- [53] S. Jung *et al.*, “A crossbar array of magnetoresistive memory devices for in-memory computing,” *Nature*, vol. 601, no. 7892, pp. 211–216, Jan. 2022, doi: 10.1038/s41586-021-04196-6.
- [54] S. W. Keckler, W. J. Dally, B. Khailany, M. Garland, and D. Glasco, “GPUs and the future of parallel computing,” *IEEE Micro*, vol. 31, no. 5, pp. 7–17, Sep. 2011, doi: 10.1109/MM.2011.89.
- [55] K. Bergman *et al.*, “ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems,” 2008. Accessed: Sep. 09, 2022. [Online]. Available: www.cse.nd.edu/Reports/2008/TR-2008-13.pdf.
- [56] M. Horowitz, “Computing’s energy problem (and what we can do about it),” in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2014, pp. 10–14, doi: 10.1109/ISSCC.2014.6757323.
- [57] Y. Lv *et al.*, “Experimental demonstration of magnetic tunnel junction-based computational random-access memory,” *arXiv Prepr.*, p. 2312.14264, Dec. 2023, [Online]. Available: <https://arxiv.org/abs/2312.14264>.
- [58] D. Kim, J. Kung, S. Chai, S. Yalamanchili, and S. Mukhopadhyay, “Neurocube: a programmable digital neuromorphic architecture with high-density 3D memory,” in *2016*

- ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA)*, Jun. 2016, pp. 380–392, doi: 10.1145/3007787.3001178.
- [59] J. Huang *et al.*, “Active-routing: compute on the way for near-data processing,” in *2019 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, Mar. 2019, pp. 674–686, doi: 10.1109/HPCA.2019.00018.
 - [60] R. Nair *et al.*, “Active memory cube: a processing-in-memory architecture for exascale systems,” *IBM J. Res. Dev.*, vol. 59, no. 2/3, pp. 17:1–17:14, Mar. 2015, doi: 10.1147/JRD.2015.2409732.
 - [61] J. T. Pawlowski, “Hybrid memory cube (HMC),” in *2011 IEEE Hot Chips 23 Symposium (HCS)*, May 2011, pp. 1–24, doi: 10.1109/HOTCHIPS.2011.7477494.
 - [62] M. Gao, G. Ayers, and C. Kozyrakis, “Practical near-data processing for in-memory analytics frameworks,” in *2015 International Conference on Parallel Architecture and Compilation (PACT)*, 2015, pp. 113–124, doi: 10.1109/PACT.2015.22.
 - [63] M. Gao, J. Pu, X. Yang, M. Horowitz, and C. Kozyrakis, “TETRIS: scalable and efficient neural network acceleration with 3D memory,” *SIGARCH Comput. Arch. News*, vol. 45, no. 1, pp. 751–764, Apr. 2017, doi: 10.1145/3037697.3037702.
 - [64] S. Matsunaga *et al.*, “Fabrication of a Nonvolatile Full Adder Based on Logic-in-Memory Architecture Using Magnetic Tunnel Junctions,” *Appl. Phys. Express*, vol. 1, p. 091301, Aug. 2008, doi: 10.1143/APEX.1.091301.
 - [65] T. Hanyu *et al.*, “Standby-power-free integrated circuits using MTJ-based VLSI computing,” *Proc. IEEE*, vol. 104, no. 10, pp. 1844–1863, Oct. 2016, doi: 10.1109/JPROC.2016.2574939.
 - [66] S. Li, C. Xu, Q. Zou, J. Zhao, Y. Lu, and Y. Xie, “Pinatubo: a processing-in-memory architecture for bulk bitwise operations in emerging non-volatile memories,” in *2016 53rd ACM/EDAC/IEEE Design Automation Conference (DAC)*, Jun. 2016, pp. 1–6, doi: 10.1145/2897937.2898064.
 - [67] S. Aga, S. Jeloka, A. Subramaniyan, S. Narayanasamy, D. Blaauw, and R. Das, “Compute caches,” in *2017 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, May 2017, pp. 481–492, doi: 10.1109/HPCA.2017.21.
 - [68] M. A. Zidan, J. P. Strachan, and W. D. Lu, “The future of electronics based on memristive systems,” *Nat. Electron. 2017 11*, vol. 1, no. 1, pp. 22–29, Jan. 2018, doi: 10.1038/s41928-017-0006-8.
 - [69] K. Jeon, J. J. Ryu, D. S. Jeong, and G. H. Kim, “Dot-Product Operation in Crossbar Array Using a Self-Rectifying Resistive Device,” *Adv. Mater. Interfaces*, vol. 9, no. 20, p. 2200392, Jul. 2022, doi: 10.1002/ADMI.202200392.
 - [70] S. Kvatinsky, G. Satat, N. Wald, E. G. Friedman, A. Kolodny, and U. C. Weiser, “Memristor-based material implication (IMPLY) logic: Design principles and methodologies,” *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 22, no. 10, pp. 2054–2066, Oct. 2014, doi: 10.1109/TVLSI.2013.2282132.
 - [71] S. Kvatinsky *et al.*, “MAGIC—memristor-aided logic,” *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 61, no. 11, pp. 895–899, Nov. 2014, doi: 10.1109/TCSII.2014.2357292.
 - [72] J.-P. Wang and J. D. Harms, “General structure for computational random access memory (CRAM),” US9224447B2, 2015.
 - [73] S. Gupta, M. Imani, and T. Rosing, “FELIX: fast and energy-efficient logic in memory,” in *2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, Nov.

- 2018, pp. 1–7, doi: 10.1145/3240765.3240811.
- [74] Z. Chowdhury *et al.*, “Efficient in-memory processing using spintronics,” *IEEE Comput. Archit. Lett.*, vol. 17, no. 1, pp. 42–46, Jan. 2018, doi: 10.1109/LCA.2017.2751042.
 - [75] F. Gao, G. Tziantzioulis, and D. Wentzlaff, “ComputeDRAM: in-memory compute using off-the-shelf DRAMs,” in *2019 52nd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*, Oct. 2019, pp. 100–113, doi: 10.1145/3352460.3358260.
 - [76] M. S. Q. Truong *et al.*, “RACER: Bit-pipelined processing using resistive memory,” in *MICRO-54: 54th Annual IEEE/ACM International Symposium on Microarchitecture*, Oct. 2021, pp. 100–116, doi: 10.1145/3466752.3480071.
 - [77] I. Žutić, J. Fabian, and S. Das Sarma, “Spintronics: fundamentals and applications,” *Rev. Mod. Phys.*, vol. 76, no. 2, pp. 323–410, Apr. 2004, doi: 10.1103/RevModPhys.76.323.
 - [78] D. E. Nikonov and I. A. Young, “Benchmarking of beyond-CMOS exploratory devices for logic integrated circuits,” *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 1, pp. 3–11, Dec. 2015, doi: 10.1109/JXCDC.2015.2418033.
 - [79] T. Y. Lee *et al.*, “World-most energy-efficient MRAM technology for non-volatile RAM applications,” in *2022 International Electron Devices Meeting (IEDM)*, 2022, pp. 10.7.1-10.7.4, doi: 10.1109/IEDM45625.2022.10019430.
 - [80] G. Jan *et al.*, “Demonstration of ultra-low voltage and ultra low power STT-MRAM designed for compatibility with 0x node embedded LLC applications,” in *2018 IEEE Symposium on VLSI Technology*, Oct. 2018, pp. 65–66, doi: 10.1109/VLSIT.2018.8510672.
 - [81] H. Zhao *et al.*, “Sub-200 ps spin transfer torque switching in in-plane magnetic tunnel junctions with interface perpendicular anisotropy,” *J. Phys. D. Appl. Phys.*, vol. 45, no. 2, p. 025001, Jan. 2012, doi: 10.1088/0022-3727/45/2/025001.
 - [82] M. Julliere, “Tunneling between ferromagnetic films,” *Phys. Lett. A*, vol. 54, no. 3, pp. 225–226, 1975, doi: 10.1016/0375-9601(75)90174-7.
 - [83] S. S. P. Parkin *et al.*, “Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers,” *Nat. Mater.*, vol. 3, no. 12, pp. 862–867, Oct. 2004, doi: 10.1038/nmat1256.
 - [84] S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, and K. Ando, “Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions,” *Nat. Mater.*, vol. 3, no. 12, pp. 868–871, Dec. 2004, doi: 10.1038/nmat1257.
 - [85] L. Berger, “Emission of spin waves by a magnetic multilayer traversed by a current,” *Phys. Rev. B*, vol. 54, no. 13, pp. 9353–9358, 1996.
 - [86] J. C. Slonczewski, “Current-driven excitation of magnetic multilayers,” *J. Magn. Magn. Mater.*, vol. 159, no. 1–2, pp. L1–L7, Jun. 1996, doi: 10.1016/0304-8853(96)00062-5.
 - [87] L. Wei *et al.*, “A 7Mb STT-MRAM in 22FFL FinFET technology with 4ns read sensing time at 0.9V using write-verify-write scheme and offset-cancellation sensing technique,” in *2019 IEEE International Solid- State Circuits Conference - (ISSCC)*, Mar. 2019, pp. 214–216, doi: 10.1109/ISSCC.2019.8662444.
 - [88] W. J. Gallagher *et al.*, “22nm STT-MRAM for reflow and automotive uses with high yield, reliability, and magnetic immunity and with performance and shielding options,” in *2019 International Electron Devices Meeting (IEDM)*, Dec. 2019, pp. 2.7.1-2.7.4, doi: 10.1109/IEDM19573.2019.8993469.
 - [89] Y. Der Chih *et al.*, “A 22nm 32Mb embedded STT-MRAM with 10ns read speed, 1M cycle write endurance, 10 years retention at 150°C and high immunity to magnetic field

- interference,” in *2020 IEEE International Solid-State Circuits Conference - (ISSCC)*, Feb. 2020, pp. 222–224, doi: 10.1109/ISSCC19947.2020.9062955.
- [90] D. Edelstein *et al.*, “A 14 nm embedded STT-MRAM CMOS technology,” in *2020 International Electron Devices Meeting (IEDM)*, Dec. 2020, pp. 11.5.1–11.5.4, doi: 10.1109/IEDM13553.2020.9371922.
- [91] 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, 2013, doi: 10.1109/JSSC.2012.2224256.
- [92] D. J. Lilja, J.-P. Wang, A. P. Lyle, S. R. Patil, J. D. Harms, and X. Yao, “Systems and methods for direct communication between magnetic tunnel junctions,” May 18, 2014.
- [93] A. Lyle, J. Harms, S. Patil, X. Yao, D. J. Lilja, and J. P. Wang, “Direct communication between magnetic tunnel junctions for nonvolatile logic fan-out architecture,” *Appl. Phys. Lett.*, vol. 97, no. 15, p. 152504, Oct. 2010, doi: 10.1063/1.3499427.
- [94] M. Zabihi *et al.*, “Using spin-Hall MTJs to build an energy-efficient in-memory computation platform,” in *20th International Symposium on Quality Electronic Design (ISQED)*, Apr. 2019, pp. 52–57, doi: 10.1109/ISQED.2019.8697377.
- [95] J. A. Currivan-Incorvia *et al.*, “Logic circuit prototypes for three-terminal magnetic tunnel junctions with mobile domain walls,” *Nat. Commun.*, vol. 7, no. 1, pp. 1–7, Jan. 2016, doi: 10.1038/ncomms10275.
- [96] M. Alamdar *et al.*, “Domain wall-magnetic tunnel junction spin-orbit torque devices and circuits for in-memory computing,” *Appl. Phys. Lett.*, vol. 118, no. 11, p. 112401, Mar. 2021, doi: 10.1063/5.0038521.
- [97] M. Zabihi *et al.*, “Analyzing the effects of interconnect parasitics in the STT CRAM in-memory computational platform,” *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 6, no. 1, pp. 71–79, Jun. 2020, doi: 10.1109/JXCDC.2020.2985314.
- [98] Z. I. Chowdhury *et al.*, “A DNA read alignment accelerator based on computational RAM,” *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 6, no. 1, pp. 80–88, Jun. 2020, doi: 10.1109/JXCDC.2020.2987527.
- [99] Z. I. Chowdhury *et al.*, “CRAM-Seq: accelerating RNA-Seq abundance quantification using computational RAM,” *IEEE Trans. Emerg. Top. Comput.*, vol. 10, no. 4, pp. 2055–2071, 2022, doi: 10.1109/TETC.2022.3153613.
- [100] M. Zabihi, Z. I. Chowdhury, Z. Zhao, U. R. Karpuzcu, J. P. Wang, and S. S. Sapatnekar, “In-memory processing on the spintronic CRAM: from hardware design to application mapping,” *IEEE Trans. Comput.*, vol. 68, no. 8, pp. 1159–1173, Aug. 2019, doi: 10.1109/TC.2018.2858251.
- [101] H. Cilasun *et al.*, “CRAFFT: High resolution FFT accelerator in spintronic computational RAM,” in *2020 57th ACM/IEEE Design Automation Conference (DAC)*, Jul. 2020, pp. 1–6, doi: 10.1109/DAC18072.2020.9218673.
- [102] S. Resch *et al.*, “PIMBALL: Binary neural networks in spintronic memory,” *ACM Trans. Archit. Code Optim.*, vol. 16, no. 4, p. 41, Oct. 2019, doi: 10.1145/3357250.
- [103] Z. I. Chowdhury *et al.*, “CAMEleon: reconfigurable B(T)CAM in computational RAM,” in *Proceedings of the 2021 on Great Lakes Symposium on VLSI*, Jun. 2021, pp. 57–63, doi: 10.1145/3453688.3461507.
- [104] S. Resch *et al.*, “MOUSE: inference in non-volatile memory for energy harvesting applications,” in *2020 53rd Annual IEEE/ACM International Symposium on*

- Microarchitecture (MICRO)*, Oct. 2020, pp. 400–414, doi: 10.1109/MICRO50266.2020.00042.
- [105] M. Zabihi *et al.*, “True in-memory computing with the CRAM: from technology to applications,” in *Proceedings of the 2019 on Great Lakes Symposium on VLSI*, 2019, p. 379, doi: 10.1145/3299874.3319451.
 - [106] H. Cilasun *et al.*, “Seeds of SEED: H-CRAM: in-memory homomorphic search accelerator using spintronic computational RAM,” in *2021 International Symposium on Secure and Private Execution Environment Design (SEED)*, 2021, pp. 70–75, doi: 10.1109/SEED51797.2021.00018.
 - [107] A. Alaghi, W. Qian, and J. P. Hayes, “The Promise and Challenge of Stochastic Computing,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 37, no. 8, pp. 1515–1531, Aug. 2018, doi: 10.1109/TCAD.2017.2778107.
 - [108] Y. Liu and K. K. Parhi, “Computing hyperbolic tangent and sigmoid functions using stochastic logic,” *Conf. Rec. - Asilomar Conf. Signals, Syst. Comput.*, pp. 1580–1585, Mar. 2017, doi: 10.1109/ACSSC.2016.7869645.
 - [109] P. Li, W. Qian, D. J. Lilja, K. Bazargan, and M. D. Riedel, “Case Studies of Logical Computation on Stochastic Bit Streams,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7606 LNCS, pp. 235–244, 2013, doi: 10.1007/978-3-642-36157-9_24.
 - [110] J. M. De Aguiar and S. P. Khatrri, “Exploring the viability of stochastic computing,” *Proc. 33rd IEEE Int. Conf. Comput. Des. ICCD 2015*, pp. 391–394, Dec. 2015, doi: 10.1109/ICCD.2015.7357131.
 - [111] 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, 2014, doi: 10.1109/TVLSI.2013.2247429.
 - [112] 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.
 - [113] 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, 2016, doi: 10.1109/TNANO.2015.2511151.
 - [114] X. Jia *et al.*, “Spintronics based Stochastic Computing for Efficient Bayesian Inference System,” Nov. 2017, doi: 10.1109/TED.2014.2372475.
 - [115] Y. Lv and J.-P. Wang, “A single magnetic-tunnel-junction stochastic computing unit,” in *2017 International Electron Devices Meeting (IEDM)*, Dec. 2017, pp. 36.2.1–36.2.4, doi: 10.1109/IEDM.2017.8268504.
 - [116] Y. Shao *et al.*, “Implementation of Artificial Neural Networks Using Magnetoresistive Random-Access Memory-Based Stochastic Computing Units,” *IEEE Magn. Lett.*, vol. 12, 2021, doi: 10.1109/LMAG.2021.3071084.
 - [117] E. Becle, G. Prenat, P. Talatchian, L. Anghel, and I. L. Prejbeanu, “A Fast, Energy Efficient and Tunable Magnetic Tunnel Junction Based Bitstream Generator for Stochastic Computing,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 69, no. 8, pp. 3251–3259, Aug. 2022, doi: 10.1109/TCSI.2022.3173030.
 - [118] W. H. Choi *et al.*, “A Magnetic Tunnel Junction based True Random Number Generator with conditional perturb and real-time output probability tracking,” in *2014 International Electron Devices Meeting (IEDM)*, 2014, pp. 12.5.1–12.5.4, doi:

- 10.1109/IEDM.2014.7047039.
- [119] 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 *Conference Proceedings - 13th IEEE International NEW Circuits and Systems Conference, NEWCAS 2015*, 2015, no. 1, pp. 3–6, doi: 10.1109/NEWCAS.2015.7182089.
 - [120] W. H. Choi, Y. Lv, H. Kim, J.-P. Wang, and C. H. Kim, "An 8-bit Analog-to-Digital Converter based on the voltage-dependent switching probability of a Magnetic Tunnel Junction," in *2015 Symposium on VLSI Technology (VLSI Technology)*, Jun. 2015, pp. T162–T163, doi: 10.1109/VLSIT.2015.7223662.
 - [121] H. Cilasun *et al.*, "Spiking Neural Networks in Spintronic Computational RAM," *ACM Trans. Archit. Code Optim.*, vol. 18, no. 4, Sep. 2021, doi: 10.1145/3475963.
 - [122] B. R. Zink *et al.*, "A Stochastic Computing Scheme of Embedding Random Bit Generation and Processing in Computational Random Access Memory (SC-CRAM)," *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 9, no. 1, pp. 29–37, Jun. 2023, doi: 10.1109/JXCDC.2023.3266136.
 - [123] S. S. P. Parkin *et al.*, "Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers," *Nat. Mater.*, vol. 3, no. 12, pp. 862–867, Dec. 2004, doi: 10.1038/nmat1256.
 - [124] Y. Lv and J. P. Wang, "Manufacturability Evaluation of Magnetic Tunnel Junction-Based Computational Random Access Memory," *IEEE Trans. Magn.*, 2023, doi: 10.1109/TMAG.2023.3323935.
 - [125] W. A. Borders *et al.*, "Measurement-driven neural-network training for integrated magnetic tunnel junction arrays," *Phys. Rev. Appl.*, vol. 21, no. 5, p. 054028, May 2024, doi: 10.1103/PHYSREVAPPLIED.21.054028/FIGURES/9/MEDIUM.
 - [126] O. Yousuf *et al.*, "Layer Ensemble Averaging for Improving Memristor-Based Artificial Neural Network Performance," Apr. 2024, Accessed: Jun. 19, 2024. [Online]. Available: <https://arxiv.org/abs/2404.15621v1>.
 - [127] B. Hoskins *et al.*, "A System for Validating Resistive Neural Network Prototypes," *ACM Int. Conf. Proceeding Ser.*, vol. 5, Jul. 2021, doi: 10.1145/3477145.3477260.
 - [128] J. Bardeen, "Tunnelling from a Many-Particle Point of View," *Phys. Rev. Lett.*, vol. 6, no. 2, p. 57, Jan. 1961, doi: 10.1103/PhysRevLett.6.57.
 - [129] G. Kresse and J. Furthmüller, "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set," *Phys. Rev. B*, vol. 54, no. 16, p. 11169, Oct. 1996, doi: 10.1103/PhysRevB.54.11169.
 - [130] G. Kresse and D. Joubert, "From ultrasoft pseudopotentials to the projector augmented-wave method," *Phys. Rev. B*, vol. 59, no. 3, p. 1758, Jan. 1999, doi: 10.1103/PhysRevB.59.1758.
 - [131] P. Bruno, "Theory of interlayer magnetic coupling," *Phys. Rev. B*, vol. 52, no. 1, p. 411, 1995.
 - [132] D. C. Worledge and P. L. Trouilloud, "Magnetoresistance measurement of unpatterned magnetic tunnel junction wafers by current-in-plane tunneling," *Appl. Phys. Lett.*, vol. 83, no. 1, pp. 84–86, Jul. 2003, doi: 10.1063/1.1590740.
 - [133] Y. S. Choi *et al.*, "Effect of Ta getter on the quality of MgO tunnel barrier in the polycrystalline CoFeB/MgO/CoFeB magnetic tunnel junction," *Appl. Phys. Lett.*, vol. 90, no. 1, pp. 12503–12505, 2007.
 - [134] Y. Ohsawa *et al.*, "Precise Damage Observation in Ion-Beam Etched MTJ," *IEEE Trans.*

- Magn.*, vol. 52, no. 7, Jul. 2016, doi: 10.1109/TMAG.2015.2512588.
- [135] Y. Lv, R. P. Bloom, and J.-P. Wang, “Experimental demonstration of probabilistic spin logic by magnetic tunnel junctions,” *IEEE Magn. Lett.*, vol. 10, pp. 1–5, Sep. 2019, doi: 10.1109/LMAG.2019.2957258.
 - [136] J. Kim, A. Chen, B. Behin-Aein, S. Kumar, J. P. Wang, and C. H. Kim, “A technology-agnostic MTJ SPICE model with user-defined dimensions for STT-MRAM scalability studies,” in *2015 IEEE Custom Integrated Circuits Conference (CICC)*, Nov. 2015, pp. 1–4, doi: 10.1109/CICC.2015.7338407.
 - [137] R. Heindl, W. H. Rippard, S. E. Russek, M. R. Pufall, and A. B. Kos, “Validity of the thermal activation model for spin-transfer torque switching in magnetic tunnel junctions,” *J. Appl. Phys.*, vol. 109, no. 7, p. 073910, Apr. 2011, doi: 10.1063/1.3562136.
 - [138] M. Bowen *et al.*, “Nearly total spin polarization in La₂/3Sr₁/3MnO₃ from tunneling experiments,” *Appl. Phys. Lett.*, vol. 82, no. 2, pp. 233–235, Jan. 2003, doi: 10.1063/1.1534619.
 - [139] Y. Sakuraba *et al.*, “Giant tunneling magnetoresistance in Co₂MnSi/Al-O/Co₂MnSi magnetic tunnel junctions,” *Appl. Phys. Lett.*, vol. 88, no. 19, p. 192508, May 2006, doi: 10.1063/1.2202724/567699.
 - [140] J. C. Slonczewski, “Conductance and exchange coupling of two ferromagnets separated by a tunneling barrier,” *Phys. Rev. B*, vol. 39, no. 10, p. 6995, Apr. 1989, doi: 10.1103/PhysRevB.39.6995.
 - [141] T. Aoki, Y. Ando, D. Watanabe, M. Oogane, and T. Miyazaki, “Spin transfer switching in the nanosecond regime for CoFeB/MgO/CoFeB ferromagnetic tunnel junctions,” *J. Appl. Phys.*, vol. 103, no. 10, p. 103911, May 2008, doi: 10.1063/1.2930873/370875.
 - [142] D. J. P. De Sousa, P. M. Haney, J. P. Wang, and T. Low, “Field-Free-Switching State Diagram of Perpendicular Magnetization Subjected to Conventional and Unconventional Spin-Orbit Torques,” *Phys. Rev. Appl.*, vol. 18, no. 5, p. 054020, Nov. 2022, doi: 10.1103/PHYSREVAPPLIED.18.054020/FIGURES/7/MEDIUM.
 - [143] S. Liu, D. J. P. de Sousa, M. Sammon, J. P. Wang, and T. Low, “Efficient domain wall motion in asymmetric magnetic tunnel junctions with vertical current flow,” *J. Magn. Magn. Mater.*, vol. 549, p. 168949, May 2022, doi: 10.1016/J.JMMM.2021.168949.
 - [144] I. Suzuki, S. Kubo, H. Sepehri-Amin, and Y. K. Takahashi, “Dependence of the Growth Mode in Epitaxial FePt Films on Surface Free Energy,” *ACS Appl. Mater. Interfaces*, vol. 13, no. 14, pp. 16620–16627, Apr. 2021, doi: 10.1021/ACSAMI.0C22510/ASSET/IMAGES/LARGE/AM0C22510_0010.JPEG.
 - [145] D. Lyu *et al.*, “L10 FePd-based perpendicular magnetic tunnel junctions with 65% tunnel magnetoresistance and ultralow switching current density,” *AIP Adv.*, vol. 14, no. 2, Feb. 2024, doi: 10.1063/9.0000818.
 - [146] D. Zhang *et al.*, “Low Gilbert damping and high thermal stability of Ru-seeded L10-phase FePd perpendicular magnetic thin films at elevated temperatures,” *Appl. Phys. Lett.*, vol. 117, no. 8, p. 82405, Aug. 2020, doi: 10.1063/5.0016100/13546325/082405_1_ACCEPTED_MANUSCRIPT.PDF.
 - [147] D. L. Zhang *et al.*, “L10 Fe-Pd Synthetic Antiferromagnet through an fcc Ru Spacer Utilized for Perpendicular Magnetic Tunnel Junctions,” *Phys. Rev. Appl.*, vol. 9, no. 4, p. 044028, Apr. 2018, doi: 10.1103/PhysRevApplied.9.044028.
 - [148] H. X. Yang, M. Chshiev, B. Dieny, J. H. Lee, A. Manchon, and K. H. Shin, “First-principles investigation of the very large perpendicular magnetic anisotropy at Fe|MgO

- and Co|MgO interfaces,” *Phys. Rev. B*, vol. 84, no. 5, p. 054401, 2011, doi: 10.1103/PhysRevB.84.054401.
- [149] D. Weller *et al.*, “Microscopic Origin of Magnetic Anisotropy in Au/Co/Au Probed with X-Ray Magnetic Circular Dichroism,” *Phys. Rev. Lett.*, vol. 75, no. 20, pp. 3752–3755, 1995.
- [150] S. Mangin, D. Ravelosona, J. a. Katine, M. J. Carey, B. D. Terris, and E. E. Fullerton, “Current-induced magnetization reversal in nanopillars with perpendicular anisotropy,” *Nat. Mater.*, vol. 5, no. 3, pp. 210–215, Feb. 2006, doi: 10.1038/nmat1595.
- [151] E. Liu *et al.*, “[Co/Ni]-CoFeB hybrid free layer stack materials for high density magnetic random access memory applications,” *Appl. Phys. Lett.*, vol. 108, no. 13, p. 132405, 2016, doi: 10.1063/1.4945089.
- [152] W. H. Butler, X.-G. Zhang, T. C. Schulthess, and J. M. MacLaren, “Spin-dependent tunneling conductance of Fe/MgO/Fe sandwiches,” *Phys. Rev. B*, vol. 63, no. 5, p. 054416, 2001, doi: 10.1103/PhysRevB.63.054416.
- [153] D. C. Worledge *et al.*, “Spin torque switching of perpendicular Ta|CoFeB|MgO-based magnetic tunnel junctions,” *Appl. Phys. Lett.*, vol. 98, no. 2, p. 022501, 2011, doi: 10.1063/1.3536482.
- [154] J. Faure-Vincent *et al.*, “Interlayer Magnetic Coupling Interactions of Two Ferromagnetic Layers by Spin Polarized Tunneling,” *Phys. Rev. Lett.*, vol. 89, no. 10, p. 107206, Aug. 2002, doi: 10.1103/PhysRevLett.89.107206.
- [155] T. Katayama, S. Yuasa, J. Velez, M. Y. Zhuravlev, S. S. Jaswal, and E. Y. Tsymbal, “Interlayer exchange coupling in Fe/MgO/Fe magnetic tunnel junctions,” *Appl. Phys. Lett.*, vol. 89, no. 11, p. 112503, 2006.
- [156] L. E. Nistor, B. Rodmacq, S. Auffret, A. Schuhl, M. Chshiev, and B. Dieny, “Oscillatory interlayer exchange coupling in MgO tunnel junctions with perpendicular magnetic anisotropy,” *Phys. Rev. B*, vol. 81, no. 22, p. 220407, Jun. 2010, doi: 10.1103/PhysRevB.81.220407.
- [157] W. G. Wang *et al.*, “Understanding tunneling magnetoresistance during thermal annealing in MgO-based junctions with CoFeB electrodes,” *Phys. Rev. B*, vol. 81, no. 14, p. 144406, 2010.
- [158] W. G. Wang *et al.*, “In situ characterization of rapid crystallization of amorphous CoFeB electrodes in CoFeB/MgO/CoFeB junctions during thermal annealing,” *Appl. Phys. Lett.*, vol. 95, no. 24, p. 242501, 2009, doi: 10.1063/1.3273397.
- [159] W.-G. Wang *et al.*, “Rapid thermal annealing study of magnetoresistance and perpendicular anisotropy in magnetic tunnel junctions based on MgO and CoFeB,” *Appl. Phys. Lett.*, vol. 99, no. 10, p. 102502, 2011, doi: 10.1063/1.3634026.
- [160] H. Sato, M. Yamanouchi, S. Ikeda, S. Fukami, F. Matsukura, and H. Ohno, “Perpendicular-anisotropy CoFeB-MgO magnetic tunnel junctions with a MgO/CoFeB/Ta/CoFeB/MgO recording structure,” *Appl. Phys. Lett.*, vol. 101, no. 2, p. 022414, 2012, doi: 10.1063/1.4736727.
- [161] H. Sato *et al.*, “Properties of magnetic tunnel junctions with a MgO/CoFeB/Ta/CoFeB/MgO recording structure down to junction diameter of 11 nm,” *Appl. Phys. Lett.*, vol. 105, no. 6, p. 062403, Aug. 2014, doi: 10.1063/1.4892924.
- [162] S.-E. Lee, J.-U. Baek, and J.-G. Park, “Highly Enhanced TMR Ratio and Δ for Double MgO-based p-MTJ Spin-Valves with Top Co2Fe6B2 Free Layer by Nanoscale-thick Iron Diffusion-barrier,” *Sci. Rep.*, vol. 7, no. 1, p. 11907, Dec. 2017, doi: 10.1038/s41598-017-

10967-x.

- [163] J. H. Kim *et al.*, “Ultrathin W space layer-enabled thermal stability enhancement in a perpendicular MgO/CoFeB/W/CoFeB/MgO recording frame,” *Sci. Rep.*, vol. 5, no. October, pp. 1–7, 2015, doi: 10.1038/srep16903.
- [164] S. S. P. Parkin, “Systematic variation of the strength and oscillation period of indirect magnetic exchange coupling through the 3d, 4d, and 5d transition metals,” *Phys. Rev. Lett.*, vol. 67, no. 25, pp. 3598–3601, 1991, doi: 10.1103/PhysRevLett.67.3598.
- [165] C. W. Cheng, T. I. Cheng, C. H. Shiue, C. L. Weng, Y. C. Tsai, and G. Chern, “Synthetic antiferromagnetic MgO/CoFeB/Ta(x)/CoFeB/MgO structures with perpendicular magnetic anisotropy,” *IEEE Trans. Magn.*, vol. 49, no. 7, pp. 4433–4436, 2013, doi: 10.1109/TMAG.2013.2239971.
- [166] H. Almasi *et al.*, “Enhanced tunneling magnetoresistance and perpendicular magnetic anisotropy in Mo/CoFeB/MgO magnetic tunnel junctions,” *Appl. Phys. Lett.*, vol. 106, no. 18, p. 182406, 2015, doi: 10.1063/1.4919873.
- [167] H. Almasi, M. Xu, Y. Xu, T. Newhouse-Illige, and W. G. Wang, “Effect of Mo insertion layers on the magnetoresistance and perpendicular magnetic anisotropy in Ta/CoFeB/MgO junctions,” *Appl. Phys. Lett.*, vol. 109, p. 032401, 2016, doi: 10.1063/1.4958732.
- [168] T. Liu, Y. Zhang, J. W. Cai, and H. Y. Pan, “Thermally robust Mo/CoFeB/MgO trilayers with strong perpendicular magnetic anisotropy,” *Sci. Rep.*, vol. 4, no. 5, p. 5895, Jan. 2014, doi: 10.1038/srep05895.
- [169] S. Husain *et al.*, “Spin pumping in ion-beam sputtered Co₂FeAl/Mo bilayers: Interfacial Gilbert damping,” *Phys. Rev. B*, vol. 97, no. 6, pp. 2–9, 2018, doi: 10.1103/PhysRevB.97.064420.
- [170] X. Zhang, Y. Zhang, and J. W. Cai, “Antiferromagnetically coupled perpendicular magnetic anisotropic CoFeB/MgO films across a Mo spacer with high thermal stability,” *J. Appl. Phys.*, vol. 118, no. 14, 2015, doi: 10.1063/1.4933183.
- [171] W. G. Wang *et al.*, “Parallel fabrication of magnetic tunnel junction nanopillars by nanosphere lithography,” *Sci. Rep.*, vol. 3, p. 1948, Jan. 2013, doi: 10.1038/srep01948.
- [172] C. Grezes *et al.*, “Ultra-low switching energy and scaling in electric-field-controlled nanoscale magnetic tunnel junctions with high resistance-area product,” *Appl. Phys. Lett.*, vol. 108, p. 012403, 2016, doi: 10.1063/1.4939446.
- [173] T. Nozaki *et al.*, “Recent progress in the voltage-controlled magnetic anisotropy effect and the challenges faced in developing voltage-torque MRAM,” *Micromachines*, vol. 10, no. 5, 2019, doi: 10.3390/mi10050327.
- [174] K. Hinton, G. Raskutti, P. M. Farrell, and R. S. Tucker, “Switching energy and device size limits on digital photonic signal processing technologies,” *IEEE J. Sel. Top. Quantum Electron.*, vol. 14, no. 3, pp. 938–945, 2008, doi: 10.1109/JSTQE.2008.916242.
- [175] P. J. Chen, Y. L. Iunin, S. F. Cheng, and R. D. Shull, “Underlayer Effect on Perpendicular Magnetic Anisotropy in Co₂₀Fe₆₀B₂₀/MgO Films,” *IEEE Trans. Magn.*, vol. 52, no. 7, pp. 18–21, 2016, doi: 10.1109/TMAG.2015.2511662.
- [176] S. Peng *et al.*, “Origin of interfacial perpendicular magnetic anisotropy in MgO/CoFe/metallic capping layer structures,” *Sci. Rep.*, vol. 5, pp. 3–8, 2015, doi: 10.1038/srep18173.
- [177] T. Yamamoto *et al.*, “Thermally Induced Precession-Orbit Transition of Magnetization in Voltage-Driven Magnetization Switching,” *Phys. Rev. Appl.*, vol. 10, no. 2, p. 024004,

- 2018, doi: 10.1103/PhysRevApplied.10.024004.
- [178] 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, 2007, doi: 10.1088/0022-3727/40/21/R01.
 - [179] S. Ikeda *et al.*, “Tunnel magnetoresistance of 604% at 300 K by suppression of Ta diffusion in CoFeBMgOCoFeB pseudo-spin-valves annealed at high temperature,” *Appl. Phys. Lett.*, vol. 93, no. 8, pp. 1–4, 2008, doi: 10.1063/1.2976435.
 - [180] J. Sinha *et al.*, “Influence of boron diffusion on the perpendicular magnetic anisotropy in Ta|CoFeB|MgO ultrathin films,” *J. Appl. Phys.*, vol. 117, no. 4, p. 043913, Jan. 2015, doi: 10.1063/1.4906096.
 - [181] P. K. Amiri *et al.*, “Electric-Field-Controlled Magnetoelectric RAM: Progress, Challenges, and Scaling,” *IEEE Trans. Magn.*, vol. 51, no. 11, 2015, doi: 10.1109/TMAG.2015.2443124.
 - [182] T. Nozaki *et al.*, “Highly efficient voltage control of spin and enhanced interfacial perpendicular magnetic anisotropy in iridium-doped Fe/MgO magnetic tunnel junctions,” *NPG Asia Mater.*, vol. 9, no. 12, p. e451, 2017, doi: 10.1038/am.2017.204.
 - [183] T. Nozaki *et al.*, “Enhancement in the interfacial perpendicular magnetic anisotropy and the voltage-controlled magnetic anisotropy by heavy metal doping at the Fe/MgO interface,” *APL Mater.*, vol. 6, no. 2, 2018, doi: 10.1063/1.5018162.
 - [184] S. Subathradevi and C. Vennila, “Systolic array multiplier for augmenting data center networks communication link,” *Cluster Comput.*, vol. 22, no. 6, pp. 13773–13783, Nov. 2019, doi: 10.1007/s10586-018-2092-4.
 - [185] J. Liang, J. Han, and F. Lombardi, “New metrics for the reliability of approximate and probabilistic adders,” *IEEE Trans. Comput.*, vol. 62, no. 9, pp. 1760–1771, 2013, doi: 10.1109/TC.2012.146.
 - [186] T. Scheike, Z. Wen, H. Sukegawa, and S. Mitani, “631% room temperature tunnel magnetoresistance with large oscillation effect in CoFe/MgO/CoFe(001) junctions,” *Appl. Phys. Lett.*, vol. 122, no. 11, p. 112404, Mar. 2023, doi: 10.1063/5.0145873.
 - [187] Y. LeCun, C. Cortes, and C. J. C. Burges, “The MNIST database of handwritten digits,” 1998. <http://yann.lecun.com/exdb/mnist/>.

7 APPENDICES

• Appendix A: List of Patents

1. CRAM for Stochastic Computing, filed in 2021 (supported by DARPA FRANC)
2. Semiconductor Industry Compatible Process for Scalable MTJs (down to 5 nm), filed in 2022 (supported by DARPA FRANC)
3. Perpendicular magnetic tunnel junction with multi-interface Free layer for magnetoelectric devices, filed in 2022 (supported by DARPA FRANC and SRC)
4. Preparation of Bulk-Perpendicular Magnetic Anisotropy Iron-Palladium Alloy and its Synthetic Antiferromagnets on Si Wafers, filed in 2023 (supported by DARPA FRANC)

• Appendix B: List of Publications

1. M. Xu, et al., “Voltage-controlled antiferromagnetism in magnetic tunnel junctions,” *Phys. Rev. Lett.*, 124, 187701 (2020)
2. P. Khanal, *et al.*, “Perpendicular magnetic tunnel junctions with multi-interface free layer,” *Appl. Phys. Lett.*, 119, 242404 (2021)
3. D. Sousa, *et al.*, “Gigantic tunneling magnetoresistance in magnetic Weyl semimetal tunnel junctions,” *Phys. Rev. B (Letts)*, 104, L041401(2021)
4. B. R. Zink, *et al.*, “Review of magnetic tunnel junctions for stochastic computing,” *IEEE J. Explor. Solid-State Comput. Devices*, 8(2), 173-184 (2022).
5. Y. Lv, and J.-P. Wang, “Manufacturability Evaluation of Magnetic Tunnel Junction-Based Computational Random Access Memory,” *IEEE Trans. Magn.*, 60(5), 1-7 (2024).
6. B. R. Zink, *et al.*, “A comparison study of spin-transfer torque- and spin-orbit torque-based stochastic computing using computational random access memory (SC-CRAM),” *IEEE Trans. Magn.*, 60(5), 1-15 (2024).
7. D. Lyu, *et al.*, “Sputtered L10-FePd and its synthetic antiferromagnet on Si/SiO₂ wafers for scalable spintronics,” *Adv. Funct. Mater.*, 33, 2214201 (2023).
8. B. R. Zink, *et al.*, “A stochastic computing scheme of embedding random bit generation and processing in computational random access memory (SC-CRAM),” *IEEE J. Explor. Solid-State Comput. Devices*, 9(1), 29-37 (2023).
9. Y. Lv, *et al.*, “Experimental demonstration of magnetic tunnel junction-based computational random-access memory,” *arXiv:2312.14264*, (2023).
10. B. R. Zink, *et al.*, “Influence of total ionizing dose on magnetic tunnel junctions with perpendicular anisotropy,” *IEEE Trans. Nucl. Sci.*, 68(5), 748-755 (2020).
11. X. Wang, *et al.*, “Effect of oblique versus normal deposition on the properties of perpendicularly magnetized L10 FePd thin films,” *IEEE Magn. Lett.*, 11, 2504105 (2020).
12. D. Zhang, *et al.*, “Low Gilbert damping and high thermal stability of Ru-seeded L1-phase FePd perpendicular magnetic thin films at elevated temperatures,” *Appl. Phys. Lett.*, 117(8), 082405 (2020).

13. X. Wang, *et al.*, “Buffer layer engineering of L10 FePd thin films with large perpendicular magnetic anisotropy,” *AIP Adv.*, 11(2), 025106 (2021).
14. M. Huang, *et al.*, “Voltage control of ferrimagnetic order and voltage-assisted writing of ferrimagnetic spin textures,” *Nat. Nanotech.*, 16(9), 981-988 (2021).
15. D. Huang, *et al.*, “Temperature-dependent perpendicular anisotropy and Gilbert damping of L10–FePd films: Role of noble-metal buffer layers,” *Phys. Rev. Mater.*, 6(11), 113402 (2022).
16. B. R. Zink, *et al.*, “Ultralow current switching of synthetic-antiferromagnetic magnetic tunnel junctions via electric-field assisted by spin–orbit torque,” *Adv. Electron. Mater.*, 8(10), 2200382 (2022).
17. D. Sousa, *et al.*, “Field-free-switching state diagram of perpendicular magnetization subjected to conventional and unconventional spin-orbit torques,” *Phys. Rev. Appl.*, 18(5), 054020 (2022).
18. D. Lyu, *et al.*, “Ferromagnetic resonance and magnetization switching characteristics of perpendicular magnetic tunnel junctions with synthetic antiferromagnetic free layers,” *Appl. Phys. Lett.*, 120(1), (2022).
19. D. Lyu, *et al.*, “Sub-ns switching and cryogenic-temperature performance of Mo-based perpendicular magnetic tunnel junctions,” *IEEE Electron Device Lett.*, 43(8), 1215-1218 (2022).
20. B. Zhou, *et al.*, “Perpendicular magnetic anisotropy, tunneling magnetoresistance and spin-transfer torque effect in magnetic tunnel junctions with Nb layers,” *Sci. Rep.*, 13(1), 3454 (2023).
21. D. Huang, *et al.*, “Magnetization dynamics in synthetic antiferromagnets with perpendicular magnetic anisotropy,” *Phys. Rev. B*, 107(21), 214438 (2023).

• **Appendix C: Selected Invited Talks**

1. JP Wang, “From Magnetic Random Access Memory to Computational Random Access Memory (CRAM) for AI”, *MRAM Forum / IEDM*, San Francisco, Dec, 2019
2. JP Wang, “From Computing using Non-volatile Memory Cells to Computational Random Access Memory (CRAM) for AI”, **KEYNOTE SPEECH**, 19th Non-Volatile Memory Technology Symposium (NVMTS), Durham, Oct, 2019
3. JP Wang, “Computational Random Access Memory (CRAM)”, 78th Device Research Conference, June 21-24, 2020 (converted to virtual conference now).
4. JP Wang, “Computational Random Access Memory for AI”, DARPA ERI Summit, Workshop-Autonomy and Processing – Technologies and Non-Volatile Memory for Multifunctional Memory and Computing, Aug, 2020
5. JP Wang, “Computational Random Access Memory: Computing using Memory Cells and What We Should Learn from Brain”, The 3rd Symposium for The Core Research Clusters for Materials Science and Spintronics, Sendai, Japan, Feb, 2020
6. W-G Wang, *et. al.*, "Voltage-Controlled Antiferromagnetism in Magnetic Tunnel Junctions", *MMM 2020* (Invited talk)

7. JP Wang, “Energy Efficient Spintronics-based Computing - From Materials to Devices to Systems”, DOE DOD Workshop on Energy Consequences of Information, Feb 18, 2022
8. JP Wang, “Computational Random Access Memory (CRAM): Recent Progress and Applications”, 60th Design Automation Conference (DAC), San Francisco July 16, 2022
9. JP Wang, “Bipolar electric-field switching of perpendicular magnetic tunnel junctions through voltage-controlled exchange coupling”, The Magnetic Recording Conference (TMRC), August, 2022
10. JP Wang, “Computational Random Access Memory (CRAM): Recent Progress and Applications”, Non-volatile Memory Technology Symposium, San Francisco, August, 2022
11. JP Wang, “From New Physics, Novel Materials to Super-Performance and Energy Efficient Devices (SPEED) and Enabled Extremely Energy-Efficient Systems”, SRC-SIA Webinar on Collaboration towards Decadal Plan goals, Sept, 29, 2022
12. W-G Wang, et. al., "Energy efficient switching in magnetic tunnel junctions", MMM 2022 (Invited talk)
13. Jian-Ping Wang, *et al.*, “Perpendicular Magnetic Tunnel Junctions with Crystalline Anisotropy and Energy efficient Switching of Synthetic Antiferromagnet pMTJs by voltage”, MMM 2022 (invited talk)
14. Yang Lv, *et. al.*, “Manufacturability Evaluation of Magnetic Tunnel Junction-Based Computational Random Access Memory”. TMRC 2023 (Invited talk)
15. Brandon Zink, *et al.*, “A Stochastic Computing Scheme Using Computational Random Access Memory (SC-CRAM)”. TMRC 2023 (Invited talk)
16. Jenae E. Shoup, *et al.*, “High Perpendicular Anisotropy in L10-FePd and Its Synthetic Antiferromagnet on Si/SiO₂ Wafers for Spintronics Devices”. TMRC 2023 (Invited talk)

Note: MMM (Annual Conference on Magnetism and Magnetic Materials) and TMRC (The Magnetic Recording Conference) are the most prestigious conferences for magnetic recording, memory and computing.

• **Appendix D: List of Awards**

1. Prof. Jian-Ping Wang received the 2024 Achievement Award of the IEEE Magnetics Society, the highest honor in Magnetics. The citation reads: “For pioneering contributions and inventions to magnetic materials, devices, and applications in information storage and computation, particularly on magnetic media, MRAM, and in-memory computing”.
2. *Best Paper Award* received in 2022 GOMACTech, Daniel Gopman, William Taylor, Weigang Wang, Tony Low, and Jian-Ping Wang, “Advanced Perpendicular Magnetic Tunnel Junctions for Computation in Random Access Memory”. (only one among more than 300 papers)
3. *Student Travel Grant* received in 2023 INTERMAG, Deyuan Lyu, Jenae E. Shoup, Dingbin Huang, Xiaojia Wang, Daniel Gopman, and Jian-ping Wang, “FePd Single Layer and Synthetic Antiferromagnet with Low Damping and Crystalline Perpendicular Magnetic Anisotropy on Amorphous Si/SiO₂ Wafers”.

4. *IEEE Electron Device Society Fellowship*, Duarte Sousa from Tony Low group received this prestigious award (only one award to North America each year).

- **Appendix E: Additional Outcomes**

- 65 nm Tapeout of CRAM Chip:
 - Initialized and kicked off a new joint effort for the tapeout of both MTJ based and RRAM based CRAM chips on the SUNY's 65 nm node platform by a newly hired U of Minnesota colleague, Prof. Kevin Cao, based on his previous collaboration with SUNY team.
 - Initialized and kicked off new NIST test vehicles based on Skywater 130 nm CMOS technology and a CRAM demonstration was planned by our FRANC team member, Dr. Brian Hoskins/NIST.
- Upcoming Sky130 Test Vehicles:
 - New NIST Test Vehicles for Academic Researchers
 - Planarized manufacturing Interface
 - Even easier to use than previous versions (larger feature sizes)
 - 4-Terminal device version more compatible with CRAM type design
 - Compatible with NIST developed FPGA control systems

LIST OF ACRONYMS, ABBREVIATIONS, AND SYMBOLS

2T1M	two-transistor–one-memristor
AC	alternating current
AP	antiparallel
CRA	carry ripple adder
CMOS	complementary metal-oxide semiconductor
CPUs	central Processing Unit
CRAM	computational random-access memory
EDX	energy dispersive X-ray spectroscopy
DC	direct current
FFT	fast Fourier transform
FPGAs	field programmable gate arrays
FAs	full adders
FM	ferromagnetic material
FWHM	full width at half maximum
FLAPW	full-potential linearized augmented plane wave
GPUs	graphics processing unit
GSG	ground source and ground
HAADF-STEM	High-angle annular dark-field scanning transmission electron microscopy
LSFRs	linear shift feedback registers
LBL	logic bit line
LL	logic line
MTJs	magnetic tunnel junctions
MCA	magneto crystalline anisotropy
MIFL	multiple interface free layers
N–R	Newton–Raphson
NED	normalized error distance
NM	nonmagnetic material
PMA	perpendicular magnetic anisotropy
pMTJs	perpendicular magnetic tunnel junction
RNGs	random number generators
R-H	resistance-field
RBW	resolution bandwidth
SAF	synthetic antiferromagnetic
SMA	Radio Frequency connector, Subminiature, type A
SOC	spin-orbit coupling
SHC	spin Hall conductivity
SHE	spin hall effect
SHA	spin hall angle
SOTA	state-of-the-art
SBG	stochastic bit-stream generation
SC-CRAM	stochastic computing - CRAM

SC	stochastic computing
STT	spin transfer torque
SOT	Spin orbit torque
TMR	tunnel Magnetoresistance
VCMA	voltage controlled magnetic anisotropy
VCEC	voltage controlled exchange coupling
VCL	voltage-controlled logic
WTM	Wallace tree multiplier
WL	word line

INDEX

Number	Title/subtitle (sorted alphabetically)
2.4	Advanced MTJs
2.6	Computational random-access memory (CRAM)
5	Conclusions
3.4	CRAM
4.4	CRAM
5.4	CRAM
4.2	Experimental results of MTJs with bulk PMA
4.3	Experimental results of MTJs with interfacial PMA
2.1	Fundamentals of magnetic tunnel junctions (MTJs)
2	Introduction
3	Methods, Assumptions, and Procedures
3.3	MTJ electrical testing and measurements
3.2	MTJ fabrications
2.2	MTJs with bulk perpendicular magnetic anisotropy
5.2	MTJs with bulk PMA
5.3	MTJs with interfacial
2.3	MTJs with interfacial PMA
6	References
4	Results and Discussion
1	Summary
2.5	Theoretical explorations
3.1	Theoretical explorations
4.1	Theoretical explorations
5.1	Theoretical explorations