

STT-MRAM defect modeling and production testing techniques

Yuan, S.

DOI

[10.4233/uuid:0d4fee84-75c5-4f47-8b05-23199dc2f8d3](https://doi.org/10.4233/uuid:0d4fee84-75c5-4f47-8b05-23199dc2f8d3)

Publication date

2025

Document Version

Final published version

Citation (APA)

Yuan, S. (2025). *STT-MRAM defect modeling and production testing techniques*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:0d4fee84-75c5-4f47-8b05-23199dc2f8d3>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

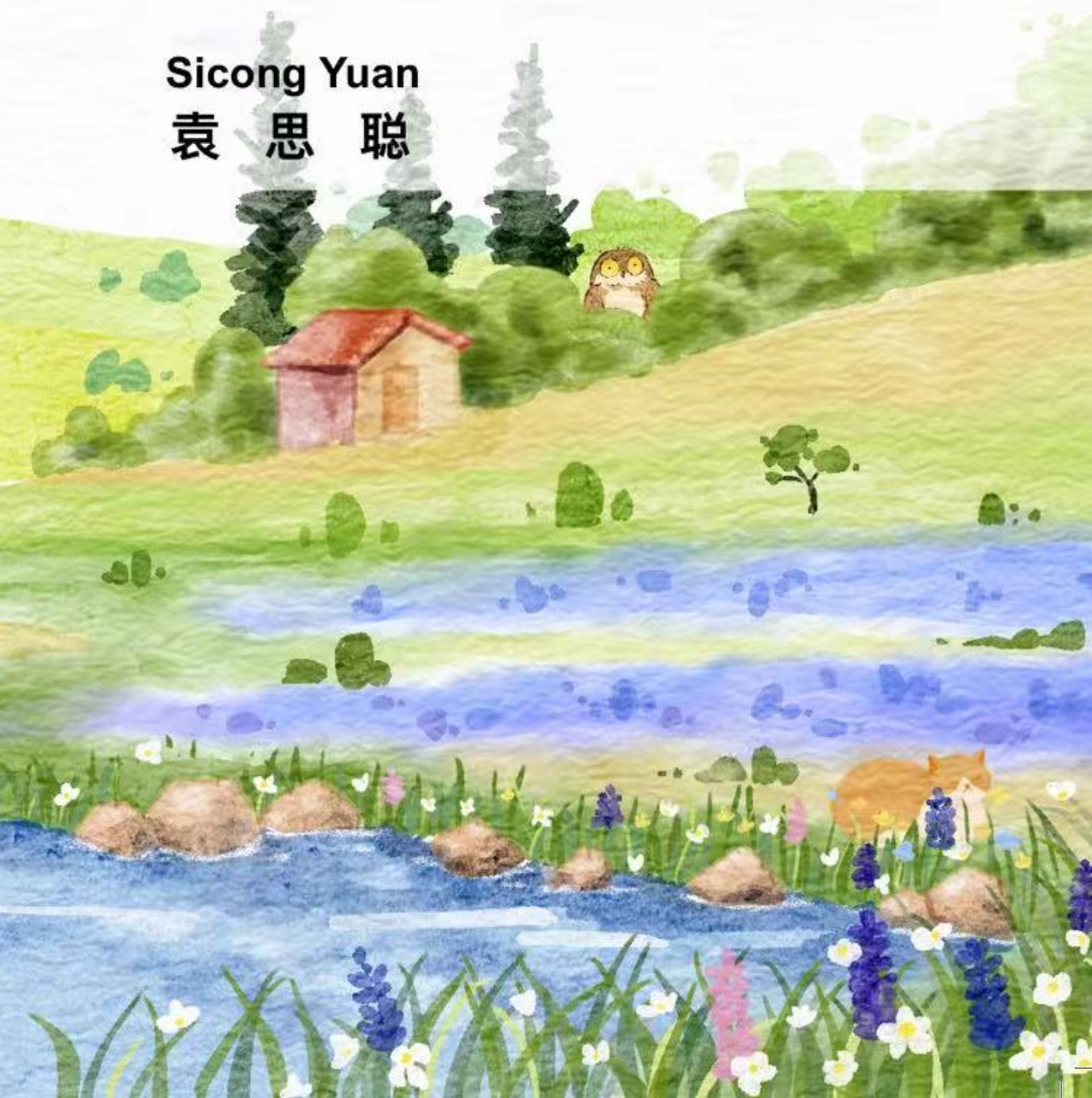
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

STT-MRAM

Defect Modeling and Production Testing Techniques

Sicong Yuan

袁 思 聪



**STT-MRAM DEFECT MODELING
AND PRODUCTION TESTING TECHNIQUES**

**STT-MRAM DEFECT MODELING
AND PRODUCTION TESTING TECHNIQUES**

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Thursday 30 October 2025 at 12:30

by

Sicong YUAN

Master of Science in Electronics Science and Technology
Zhejiang University, China
born in Zhejiang, China

This dissertation has been approved by the promotor.

Composition of the doctoral committee:

Rector Magnificus, Prof.dr.ir. S. Hamdioui, Dr.ir. M. Taouil,	Chairperson Delft University of Technology, promotor Delft University of Technology, copromotor
---	---

Independent members:

Prof.dr.ir. G. Gaydadjiev, Dr.ir. K. van Dijk, Prof.dr. J.-F. Li, Dr. S. Rao, Prof.dr. M.S. Reorda,	Delft University of Technology NXP Semiconductors National Central University IMEC Belgium PoliTO
---	---

Reserve member:

Prof.dr.ir. W.D. van Driel,	Delft University of Technology
-----------------------------	--------------------------------



Keywords: STT-MRAM, MTJ, testing, defect, magnetic field, fault modeling, test generation, design-for-testability, reliability, device-aware test, built-in self test

Printed by: Ipskamp Printing

Front & Back: Cover artwork by SicongYuan

Copyright © 2025 by S. Yuan

ISBN

An electronic version of this dissertation is available at
<http://repository.tudelft.nl/>.

*Dedicated to:
people who read this dissertation, and
those help make this work better.*

CONTENTS

Summary	xi
Acknowledgements	xiii
List of Figures	xvii
List of Tables	xxi
1 Introduction	1
1.1 Memory hierarchy	2
1.2 Emerging non-volatile memories	4
1.2.1 Resistive random-access memory	5
1.2.2 Phase change random access memory	6
1.2.3 Ferroelectric field-effect transistor	7
1.2.4 Ferroelectric field random-access memory	8
1.3 Introduction to MRAMs	9
1.3.1 MTJ and magnetoresistance	9
1.3.2 Switching mechanism and classification of MRAMs	10
1.3.3 Application of MRAMs	11
1.4 Memory testing philosophy	13
1.4.1 Traditional memory testing	13
1.4.2 Testing of emerging non-volatile memories	14
1.5 Research topics and contribution of the thesis	15
1.5.1 Testing of transistor, contact, and interconnect defects	15
1.5.2 Testing of MTJ defects	16
1.5.3 Testing STT-MRAMs: from defects to test chips	16
1.5.4 Diagnosis of MTJ defects	16
1.6 Thesis organization	17
2 STT-MRAM technology	19
2.1 MTJ structure and working mechanism	20
2.1.1 Device structure	20
2.1.2 Working mechanism	20
2.2 Structure of 1T-1MTJ cell	21
2.3 Structure of STT-MRAM array	22
2.4 Peripheral circuit	23
2.4.1 Address decoder	23
2.4.2 Write driver	24
2.4.3 Sense amplifier	26
2.5 STT-MRAM fabrication	26
2.6 Prospects and challenges of STT-MRAMs	27

3	MTJ compact model	31
3.1	Overview of existing MTJ modeling	32
3.2	Requirement for MTJ model	32
3.3	Proposed MTJ model	33
3.4	R-V model.	34
3.5	MTJ switching model	36
3.6	Write error rate	37
3.6.1	Intrinsic stochasticity mechanism in MTJs.	37
3.6.2	Write error rate measurement and fitting	38
3.7	Read error rate and modification	39
3.7.1	Conventional method	39
3.7.2	Limitation of the conventional method	39
3.7.3	Modification of the read error rate evaluation	40
3.8	Magnetic field impact.	43
3.8.1	Regular implementation of magnetic field impact	43
3.8.2	'Equivalent current' representation external magnetic field impact	44
3.8.3	Relationship between STT effect and external magnetic field impact	44
3.8.4	Stray field impact	46
4	State of the art in STT-MRAM testing	51
4.1	Framework of test design of STT-MRAMs	52
4.1.1	Conventional test development	52
4.1.2	Device-aware test methodology	54
4.2	Defect classification and modeling	56
4.2.1	Conventional defects and modeling	56
4.2.2	Unique defects and modeling	57
4.2.3	Summarize and overview of MTJ defects	62
4.3	Fault classification and modeling	63
4.3.1	Fault primitive notation	63
4.3.2	Fault classification	65
4.4	Test generation	66
4.4.1	March test generation	66
4.4.2	Design for test	67
5	Testing of transistor, contact, and interconnect defects in STT-MRAMs	69
5.1	Defect modeling and defect map	70
5.1.1	Defect classification	70
5.1.2	Space of open and short defects	71
5.1.3	Defect space of bridge defects	71
5.2	Fault modeling	72
5.2.1	Fault modeling methodology	72
5.2.2	Results for baseline simulation model	73
5.2.3	Results for magnetic coupling-only simulation model	74
5.2.4	Results for intrinsic stochasticity-only simulation model	78
5.2.5	Results for full-variability simulation model	80

5.3	Test generation	80
5.3.1	Permanent fault	80
5.3.2	Intermittent fault	88
6	Testing of MTJ defects in STT-MRAMs	91
6.1	Overview of defects in MTJs.	92
6.1.1	Testing of pinholes.	92
6.1.2	Testing of Synthetic Anti-Ferromagnet Flip (SAFF).	93
6.1.3	Testing of Intermediate State (IM)	94
6.2	Defect characterization of Back-hopping (BH)	94
6.2.1	Pulse I-V measurement	95
6.2.2	WER measurement	95
6.2.3	Write bias impact on defect occurrence	96
6.2.4	Definition of BH-defective and defect-free MTJs.	97
6.2.5	Physical mechanism	97
6.3	Defect modeling of BH	99
6.3.1	Physical modeling	99
6.3.2	Electrical modeling	101
6.3.3	Fitting and model optimization	103
6.4	Fault modeling of BH	105
6.5	Test generation of BH	106
6.5.1	Read-write based test solutions	106
6.5.2	Design for test	109
6.6	Application of external magnetic field in testing MTJ defects	112
6.6.1	SAFF test with external magnetic field	113
6.6.2	BH test with external magnetic field	116
6.6.3	Advantages and shortcomings	118
7	Testing STT-MRAMs: from defects to test chips	121
7.1	Targeted defect	122
7.1.1	Transistor, contact, and interconnect defects	122
7.1.2	Targeted MTJ defects.	122
7.2	Fault modeling	123
7.2.1	Simulation set up	123
7.2.2	Faults for interconnect and contact defects	123
7.2.3	Faults of unique defects	126
7.3	March test.	127
7.4	DfT methods	128
7.4.1	Classification	128
7.4.2	Evaluation	129
7.4.3	Selection	130
7.4.4	Implementation	131
7.4.5	Test flow	133
7.4.6	Integration.	134
7.5	Overall test flow.	134

8	Diagnosis of MTJ defects	137
8.1	Targeted defects	138
8.2	Device-aware diagnosis methodology	138
8.3	Application	140
8.3.1	Defect characterization	140
8.3.2	Defect modeling	140
8.3.3	Feature extraction	141
8.3.4	Distinctive feature identification	141
8.3.5	Diagnosis-pattern generation	143
8.4	Conclusion	144
9	Conclusion	147
9.1	Summary	147
9.2	Future research direction	149
	Bibliography	151
	Nomenclature	167
	Curriculum Vitae	175
	Main Publications	177
	Other Publications	179

SUMMARY

Spin-transfer torque magnetic random access memory (STT-MRAM) has attracted considerable attention owing to its competitive write performance, endurance, retention, and low power consumption [1]. Since the first commercial MRAM product was introduced in 2006, world-leading foundries and producers, such as TSMC, Samsung, Intel, and Everspin, have entered the market, leading to the single-chip storage capacity increasing from 4 MB to 1 GB. However, further development of the mass production of STT-MRAM technology still faces critical challenges, one of which is the vulnerability to manufacturing defects. Compared with the regular CMOS manufacturing, more manufacturing defects are introduced during the STT-MRAM manufacturing process, since it involves several additional steps for the magnetic tunnel junction (MTJ) fabrication and integration. Furthermore, the magnetic field and the spin-transfer torque (STT), which play a role in MTJ switching, have introduced a variety of unique defect types. Due to these additional defect types, directly transplanting test methods from conventional memories to STT-MRAMs causes a high test escape rate and/or high yield losses. To overcome this challenge, it is critical to design accurate defect models, derive appropriate fault models, and develop efficient test/diagnosis solutions, not only to speed up yield learning but also to provide high product quality.

This dissertation focuses on the development of effective test/diagnosis methodologies for STT-MRAMs. The testing methodology for STT-MRAMs follows the approach used for traditional memories, such as SRAMs, consisting of three steps: 1) defect modeling, 2) fault modeling, and 3) test generation.

The first step begins by analyzing the STT-MRAM fabrication process to identify potential physical defects, which can be classified into conventional memory defects in transistor, contact, and interconnect, and defects unique to MTJ. Literature and device characterization demonstrated the existence of these defects. Conventional defects are similar to those observed in traditional memories like SRAMs; these are typically modeled as linear resistors. This thesis uses the same model for STT-MRAM traditional defects, and classifies them into opens, shorts, and bridges. Unique defects within MTJs have been reported to cause irregular/non-linear faulty behaviors, rendering them unsuitable for modeling with linear resistors or parasitic capacitors. As a solution, a dedicated method named 'Device-Aware Test (DAT)' has been put forward, which designs compact models for defective MTJs by incorporating the impact of physical (manufacturing) defects on electrical and magnetic behavior. Previous works have applied the DAT method to pinhole, Synthetic Anti-Ferromagnet Flip, and Intermediate state. This thesis extends the application of the DAT method to a new defect unique to MTJs, viz., Back-hopping (BH).

In the second step, defect models are injected into the STT-MRAM array, and how these defects manifest themselves at the behavior level during actual operations is ana-

lyzed. Injecting conventional models is performed by adding an additional resistor into the array; read/write operations are then performed to incite potential erroneous functionality. Notice that the specific MTJ working mechanism effect must be included in this step to derive accurate fault models. This thesis incorporates the impact of MTJ intrinsic stochasticity, magnetic coupling, and process variation into the fault modeling, ensuring the robustness of the derived fault map for conventional defects. Unique defects are injected by replacing the regular MTJ model with the device-aware model for the defective one. Then, read/write operations are applied, and the corresponding faulty behaviors are analyzed. This thesis performs fault modeling for BH.

In the last step, we design test solutions that guarantee a high coverage of faults while maintaining a short test time. This thesis designs March tests that ensure full coverage of permanent faults that are sensitized in the presence of strong conventional defects and pinholes. However, applying March tests to detect intermittent faults and coupling faults that are introduced by weak conventional defects and other unique defects results in a high escape rate and/or low time-efficiency. To solve this challenge, dedicated Design-for-Test methodologies are put forward to improve the coverage and reduce test time.

In addition, this thesis presents a further advancement based on the DAT framework, introducing Device-Aware Diagnosis (DA-Diagnosis). This method is applied to diagnose and identify unique defects by employing a primitive march algorithm to extract distinctive features of each defect. Through measurement data from real chips, the DA-Diagnosis approach is demonstrated to be cost-effective, flexible, and practical for diagnosing unique defects in STT-MRAMs.

In short, this thesis conducts a structured testing on STT-MRAMs, tracing the process from defect analysis to test chip implementation. We begin by identifying defects, mapping them to corresponding fault behaviors, and evaluating potential test strategies, including March tests and Design-for-Testability (DfT) techniques. Notably, we provide a systematic assessment of various DfT approaches and, to the best of our knowledge, present the first implementations of a selected DfT strategy in an STT-MRAM array. This study encompasses circuit-to-layout integration, demonstrating the practical feasibility of the proposed methodology.

ACKNOWLEDGEMENTS

After four years of intensive exploration, persistent effort, and continuous learning, I now stand at the finish line of my PhD journey. This moment is far more than a personal milestone; it is a profound reminder of the support, inspiration, and kindness I've been privileged to receive along the way. This thesis, though bearing my name, is truly the result of collective guidance, steadfast collaboration, and enduring friendship. I am sincerely grateful to everyone who has walked this path alongside me.

First and foremost, I extend my deepest gratitude to my promotor, **Prof.dr.ir. Said Hamdioui**. Special thanks to you for granting me the opportunity to pursue my PhD under your guidance, an opportunity that was truly the start of everything. Your mentorship provided a crucial, dual focus: within the University, you cultivated my academic rigor and research independence, while simultaneously, through our collaboration with IMEC, you ensured I gained essential engineering insights into real-world challenges. This combined approach has been foundational to my growth as a researcher who bridges theory and application. I will never forget the refrain you repeated over these four years: "Sicong, I always give you much difficulty (in study and research), and you learn and make progress." Each difficulty you introduced — whether demanding a more rigorous proof or a clearer narrative — was a carefully placed stepping stone. I genuinely hope that I have made the significant progress you envisioned, and that your repeated encouragement was a sincere reflection of my growth. Thank you for always challenging me to aim higher and for believing in my capacity to rise to the occasion.

I am also deeply grateful to my copromotor and daily supervisor, Dr.ir. Moritz Fieback, Dr.ir. Mottaqiallah Taouil, for their invaluable support. During my challenging first year, when I was navigating a new environment and culture and struggling with language barriers in the Netherlands, you both provided essential anchors of stability and kindness. It was your initial warmth and guidance that allowed me to settle in and focus on my research. Beyond that foundational support, you continued to be indispensable throughout my studies: Moritz, your precise technical feedback and reassuring presence helped sharpen my writing and confidence, while Motta, your creative, solution-driven mindset and genuine personal care consistently propelled my projects forward, providing intellectual and emotional momentum when I needed it most. Your combined efforts not only made me a better researcher but also ensured I felt supported every step of the way.

My deep gratitude also goes to my supervisors at IMEC: **Dr. Sidharth Rao, Prof.dr.ir. Erik Jan Marinissen, and Dr. Woojin Kim**. Spending two full years in a combination environment of both industrial and academic at IMEC was an indispensable chapter of my PhD. You collectively played a vital role in transforming my academic research into tangible outcomes. Thank you for your essential guidance on industrial projects and chip design, helping me navigate the complexities of translating theory into robust silicon. Specifically, your expertise was crucial in areas where my background was less familiar,

particularly concerning magnetic devices, ensuring I could successfully integrate these technologies into my work. Your mentorship bridged the gap between fundamental research and high-level engineering, significantly enriching the scope of my thesis.

I am grateful to the committee members of this thesis, Prof.dr.ir. G. Gaydadjiev, Dr.ir. K. van Dijk, Prof.dr. J.-F. Li, Prof.dr. M.S. Reorda, and Prof.dr.ir. W.D. van Driel,. Your insightful feedback has been crucial in enhancing my thesis, and I look forward to the influence of our future collaborative efforts.

Thanks as well to both my Chinese fellows and my friends, Hanzhi Xun, Erbing Hua, Ziwei Zhang, Changhao Wang, Zhuo Chen, Leiming Du, Xinrui Ji, Kaiquan Fan, Chenlu Li, Hongwei Tang, and Fei Han. The journey would have been immeasurably harder and far less joyful without your presence. Thank you for the countless shared moments—from academic discussions that untangled complex problems to spontaneous dinners, late-night walks, and much-needed venting sessions. These daily acts of companionship, mutual support, and effortless understanding made the intensity of the PhD journey lighter and kept me grounded. Specific thanks go to Zhuo Chen, Hanzhi Xun, and Changhao Wang, for bravely enduring my highly ambitious and perhaps slightly "forced march" travel arrangement. Your good humor and resilience on those unforgettable trips provided essential breaks and created cherished memories that will last a lifetime. Thank you all for being not just colleagues, but true friends who always offered encouragement and care across every season of this demanding period.

I would like to thank all colleagues at the QCE department. Thank you, Prof.dr. Rob Kooij, for your efforts in fostering a welcoming work environment and supporting social events. I have had the pleasure of sharing my workday with many wonderful people, including (but not limited to) all those listed: Dr. Abdullah, Dr. Abhairaj, Dr. Anteneh, Dr. Carlo, Dr. Guilherme, Dr. Heba, Dr. Lizhou, Dr. Mahdi, Dr. Micheal (Taha), Dr. Paul, Dr. Rajendra, Dr. Sumit, Dr. Theofilos, Dr. Troya, Alexandra, Ali, Amin, Arne, Asmaa, Asmae, Bea, Christian, Ediz, Emmanouil-Anastasios, Emmanouil Arapidis, Erbing, Erik, Ethan Zhang, Folkert, Fouwad, Francis, Geerten, Helen, Ignacio, Ismail, Konstantinos, Mahmood, Mainak, Mark, Pantazis, Shadid, Shayesteh, Sicong, Yash, Yong, and Zacharia. Thank you for the enjoyable daily coffee walks, shared lunches, and social gatherings. Specifically, Amin, Asmae, and Mainak, thank you for organizing our social events and sports activities. Francis, Helen, and Paul, thank you for keeping everything orderly and ensuring my work could be performed effectively. Erik, Mark, and Ediz, thanks for your great help in setting up the simulation environments.

I would like to extend my thanks to all colleagues at IMEC. Thank you, Ankit Kumarm, Bob Vermeulen, Daniele Narducci, Diego Favaro, Domenico Giuliano, Vaishnavi Kateel, Maxwel Gama Monteiro Junior, Mohammed Aftab Baig, Van Dai Nguyen, Zhuo Chen, Wei Zheng, Jie Li, and Yiqun Wan. Thank you for creating such a welcoming, stimulating, and supportive environment during my two years at IMEC. I am profoundly grateful for your continuous technical support, insightful discussions, and camaraderie. Your daily help, whether through troubleshooting technical issues or simply sharing a coffee break, made the demanding environment of applied research enjoyable and highly productive.

Last but most importantly, I would like to express my deepest thanks to my family. To my parents, you have always been the pillars of my strength and my guiding lights. Your unwavering and unconditional support not only shaped the person I am today but also

enabled everything I have achieved; truly, I owe this journey to you. To my grandparents and other family members, thank you for your collective love, wisdom, and contribution to our warm extended family. Your presence in my life is a foundational gift that means more than words can fully express. And finally, to my partner, thank you for your patience, understanding, and love, which provided the calm and steady home I needed to complete this work.

Sicong Yuan
Delft, Oct., 2025

LIST OF FIGURES

1.1	The memory hierarchy.	2
1.2	Structure of RRAMs (a) OxRAM and (b) CBRAM.	6
1.3	Structure of PCRAM (a) planar cell structure, and (b) confined cell structure.	7
1.4	(a) Device structure of FeFET, (b) TEM of gate stack in FeFET.	8
1.5	(a) Structure of FE capacitor, (b) FeRAM cell structure.	8
1.6	(a) Perpendicular MTJ in P state; (b) Perpendicular MTJ in AP state; (c) In-plane MTJ in P state; (d) In-plane MTJ in AP state; (e) Tunneling current dependence on the magnetization in P state; (f) Tunneling current dependence on the magnetization in AP state.	10
1.7	The potential application of different MRAM types within the memory hierarchy.	12
2.1	(a) Structure of perpendicular MTJ; (b) MTJ switching mechanism.	20
2.2	(a) Structure of 1T-1MTJ cell and the related read/write operations; (b) TEM picture of the cell.	22
2.3	Bit-oriented STT-MRAM array.	23
2.4	Example of a row address decoder circuit.	24
2.5	(a) Circuit of the WD and the current of $w0$ and $w1$ operations.	24
2.6	(a) Circuit of the SA (b) I_{rd} in $1r1$ and $0r0$ operations.	25
2.7	Example of STT-MRAM fabrication steps [46].	27
3.1	Structure of proposed MTJ model.	34
3.2	R-V measurement and simulation.	34
3.3	Average switching time.	36
3.4	Simplified physical model of the MTJ switching.	38
3.5	WER measurement and simulation.	38
3.6	I_{rd} in $0r0$ and $1r1$ operations.	40
3.7	Example of numerical integration for (a) $0r0$ and (b) $1r1$ operation.	41
3.8	RER prediction by conventional and proposed methods for (a) $0r0$ operations and (b) $1r1$ operations.	42
3.9	H_{ext} impact on WER in (a) $1w0$ and (b) $0w1$	43
3.10	H_{ext} impact on WER in (a) $1w0$ and (b) $0w1$	45
3.11	Stray field in MTJs.	46
3.12	Illustration of of magnetic coupling in 4x4 STT-MRAM array.	47
3.13	MTJ behavior dependency on magnetic coupling effects.	48
3.14	Magnetic coupling effect on ΔI	49
4.1	Process of conventional test development.	52

4.2	Process of DAT.	54
4.3	Model of conventional defects: (a) open, (b) short, and (c) bridge.	57
4.4	(a) Defect mechanism of pinholes; (b) R-H measurement of pinhole-defective and defect-free MTJs; (c) Zoom in of the R-H measurement for the pinhole-defective MTJ.	58
4.5	(a) Defect mechanism of SAFF; (b) R-H measurement of defect-free MTJs; (c) R-H measurement of SAFF-defective MTJs.	59
4.6	(a) Defect mechanism of IM; (b) Extracting IM by repeating $w0, r0, w1, r1$	60
4.7	MTJ manufacturing process and the potential unique defects.	62
5.1	MTJ manufacturing process and the potential unique defects.	70
5.2	The basic fault map of OC2.	73
5.3	Illustration of the DBs for the fault modeling methodology for (a) single cell defects and (b)-(d) defects between adjacent cells.	75
5.4	Fault map of OS_w	76
5.5	Fault map of SS2.	76
5.6	Faults Venn diagram for all analyzed single cell defects.	77
5.7	Fault map of defect dBCC1.	78
5.8	Fault map of defect OC2.	79
5.9	DfT application for defect OS_w ; (a) Regular fault map, (b) Fault map by modified write pulse width.	87
5.10	(a) DfT methodology; (b) Circuit design of the DfT.	89
6.1	MTJ manufacturing process and the potential unique defects.	93
6.2	Plot of MTJ switching under the application of voltage pulses.	95
6.3	(a) WER measurement process; (b) WER measurement of $0w1$ operations; (c) WER measurement of $1w0$ operations; (d) WER extraction.	96
6.4	WER of (a) $1w0$ and $0w0$; (b) $0w1$ and $1w1$	97
6.5	(a) WER of multiple MTJs in the $0w1$ operation; (b) Selected WER comparison of defect-free and defective MTJs in the $0w1$ operation; (c) WER of multiple MTJs in $1w0$ operations; (d) Selected WER comparison of defect-free and defective MTJs in the $1w0$ operation.	98
6.6	Factors affecting RL instability.	99
6.7	The physical mechanism of BH in the $1w0$ operation.	100
6.8	(a) Magnetization switching during write operations; (b) BH fitting in the linear y-axis; (c) BH fitting in the log y-axis.	103
6.9	WER measurement and model fitting in (a) $1w1$ operation; (b) $0w0$ operation.	104
6.10	WER measurement of BH-defective and defect-free MTJs in (a) $0w1$ operation; (b) $1w0$ operation.	104
6.11	WER by varying H_p	105
6.12	(a) WER measurement by varying t_p in (a) $1w1$ and (b) $0w0$ operations.	108
6.13	Increase V_p for testing.	109
6.14	DfT for defecting the BH defect.	109
6.15	MTJ state and DfT output with 'OR' gate.	111
6.16	Comparison of conventional and H_{ext} -assisted test development	112

6.17	(a) SAFF mechanism; (b) P_{sw} distribution for SAFF-defective and defect-free MTJs; (c) Validation for the march algorithm Eq. 6.14; (d) Validation for the march algorithm Eq. 6.15.	113
6.18	H_{ext} impact on BH-induced WER with the (a) $0w0$ operation; (b) $1w1$ operation; (c) Comparison of H_{ext} impact on BH-induced WER between BH-defective and defect-free MTJs with the (c) $0w0$ operation; (d) $1w1$ operation.	116
7.1	The fault map of OC2: (a) baseline analysis, (b) static analysis with Stochastic-Only Model, and (c) static analysis with Full Variability Model.	125
7.2	DfT design (a) in [76], and (b) in [150].	129
7.3	DfT scheme.	132
7.4	Structure of WD and DfT block.	133
7.5	DfT integration to the STT-MRAM array.	135
7.6	Layout design.	136
7.7	Layout design zooming in for the DfT part.	136
8.1	Framework of STT-MRAM test and diagnosis development.	138
8.2	Characterization of defect-free device: (a) R-V, (b) R-H, (c) WER; Same characterization of SAFF-defective device: (d) R-V, (e) R-H, (f) WER.	140
8.3	(a) Unique feature identification for Pinholes, (b) Unique feature identification for BH, (c) SA structure.	142
8.4	Diagnosing for IM.	144

LIST OF TABLES

1.1	Summarization of emerging non-volatile memories.	5
1.2	Comparison of different MRAMs.	11
2.1	MTJ technology parameters	21
2.2	MTJ electrical parameters	22
3.1	Symbols in MTJ model.	35
4.1	Potential defects and the STT-MRAM fabrication steps.	55
4.2	MTJ defect mechanism and modeling method.	62
4.3	Fault primitive notation.	64
5.1	Fault map of each defect.	81
5.2	Example to DB-ILP table.	82
5.3	Selected DB and F for open and short defects.	83
5.4	Selected DB and S for bridge defects.	83
5.5	Selected DB and S in the DfT.	86
6.1	Key parameter calculations of defective MTJ model in the $1w0$ operation.	102
6.2	Fault modeling results of BH defect.	106
6.3	Truth table of DfT with the 'OR' gate.	110
6.4	Truth table of DfT with the 'AND' gate.	111
6.5	H_{ext} -assisted tests for different defect types in STT-MRAM testing.	118
7.1	Collapsed fault set for conventional defects.	123
7.2	Summarization of March tests.	126
7.3	Fault model for unique defects.	127
7.4	DfT method classification.	128
7.5	DfT evaluation.	130
7.6	Summarization of test solutions for each defect type.	131
7.7	DfT configuration	134
7.8	Input and output of the DfT	134
8.1	Feature-dictionary of defects	141
8.2	Conclusion of diagnosis for unique defects.	143

1

INTRODUCTION

Spin-transfer torque magnetic random access memory (STT-MRAM) has emerged as one of the leading non-volatile memory technologies, with commercial products coming after more than 40 years of research and development. As companies start ramping up efforts toward commercialization, the need for effective and cost-efficient testing solutions has become critical in ensuring the high quality of STT-MRAM products. This dissertation focuses on investigating manufacturing defects introduced during STT-MRAM fabrication, deriving corresponding fault models, and developing robust test solutions. This chapter presents an introduction to the dissertation, including 1) a review of emerging non-volatile memory technologies; 2) an in-depth look at MRAMs; 3) the principles and philosophies of memory testing; 4) the research objectives and contributions of this thesis; and 5) an outline of the dissertation structure.

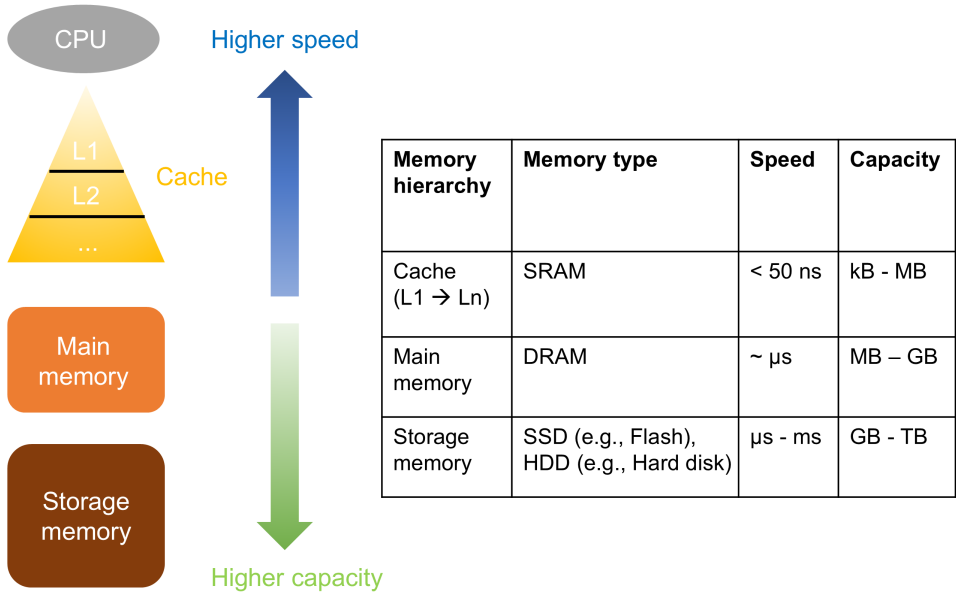


Figure 1.1: The memory hierarchy.

1.1. MEMORY HIERARCHY

In modern computer systems, memory plays an absolutely critical role, serving as the indispensable workspace for the Central Processing Unit (CPU) and the temporary or permanent storage for all data and instructions. The efficiency with which a computer system handles and accesses data is directly proportional to the effectiveness of its memory subsystem. However, designing a single type of memory that is both incredibly fast, infinitely large, and prohibitively cheap is physically impossible. This fundamental challenge leads to the sophisticated solution known as the memory hierarchy, an ingenious architectural design that balances these conflicting demands by employing multiple levels of memory, each with distinct characteristics, to optimize overall system performance.

The memory hierarchy is often visualized as a pyramid, as shown in Fig. 1.1, a layered structure that elegantly addresses the trade-offs between speed, capacity, and cost [2], [3]. At its apex, closest to the CPU, are the fastest, most expensive, and smallest memory types. As one descends the pyramid, the memory types become progressively slower, less expensive, and larger in capacity. This deliberate arrangement is underpinned by the principle of locality of reference, a fundamental observation that computer programs tend to access data and instructions that have either been recently accessed (temporal locality) or are located near recently accessed items in memory (spatial locality). By strategically placing frequently used data in faster, smaller memory levels, the overall performance and efficiency of the computing system are significantly enhanced, effectively creating the illusion of a single, large, and fast memory.

At the very pinnacle of this hierarchy, in intimate proximity to the CPU, lies Cache

memory. As illustrated in the provided in Fig. 1.1, cache memory is further subdivided into multiple levels, typically designated as L1, L2, and sometimes L3, or even L4, represented collectively as 'L1 → Ln' [2]. These cache levels are predominantly implemented using Static Random-Access Memory (SRAM). SRAM is celebrated for its exceptionally high speed and relatively low power consumption, largely because, unlike Dynamic Random-Access Memory (DRAM), it does not require constant refreshing to retain data. However, SRAM is also more complex and significantly more expensive to manufacture, which inherently limits its capacity. Consequently, cache memory possesses the highest speed within the entire memory hierarchy; L1 cache, being the closest to the CPU, offers the fastest access times, followed by L2, and then L3. This unparalleled speed is paramount because it directly dictates the CPU's ability to execute instructions without incurring idle wait states for data. A "cache hit" — where the required data is found within the cache—dramatically reduces data access time, whereas a "cache miss" — where the data is not present in the cache—necessitates fetching the data from a slower memory level, thereby imposing a performance penalty. Due to its high cost and the physical constraints of integrating it closely with the CPU, cache memory boasts the lowest capacity in the hierarchy. L1 cache typically ranges from mere tens to hundreds of kilobytes (KB), L2 from hundreds of KB to a few megabytes (MB), and L3 from several MB to perhaps tens of MB. Despite their seemingly small size, caches prove incredibly effective because they intelligently store the most frequently accessed data and instructions, ensuring their immediate availability to the CPU.

Moving down the hierarchy, just below the cache, we encounter Main memory, commonly referred to simply as Random-Access Memory (RAM). Main memory is predominantly constructed from DRAM. DRAM offers a compelling advantage over SRAM in terms of cost-effectiveness and density, enabling the realization of significantly larger capacities. However, its 'dynamic' nature means it requires constant electrical refreshing to preserve its data charge, which renders it inherently slower than SRAM. Despite this, Main memory is still considerably faster than any form of storage memory. Its speed is a critical determinant of overall system performance, as it serves as the primary working area for the CPU. All programs loaded for execution, as well as the operating system itself, reside in main memory. When a cache miss occurs, the CPU's next recourse is main memory to retrieve the necessary data. Main memory provides a medium capacity when compared to the extremes of cache and storage. Modern computer systems typically incorporate main memory capacities ranging from several gigabytes (GB) to tens or even hundreds of GB. This ample capacity allows the operating system, currently running applications, and their associated data to reside in memory, facilitating rapid and efficient access.

At the lowest tier of the active memory hierarchy lies Storage memory, also known as secondary storage or mass storage. This level encompasses a diverse array of non-volatile storage technologies, meaning they retain data even when power is removed, making them suitable for long-term data persistence. Fig. 1.1 specifically highlights "Flash, Hard disk," pointing to the two most prevalent types in contemporary computing: Hard Disk Drives (HDDs) and Solid State Drives (SSDs). HDDs store data magnetically on spinning platters, relying on mechanical components for operation, while SSDs leverage semiconductor flash memory, which operates entirely electronically. Con-

sequently, storage memory exhibits the lowest speed among the three primary levels. HDDs, with their mechanical nature, are inherently slower than electronic memory. While SSDs represent a substantial leap in speed over HDDs due to the absence of moving parts, they are still orders of magnitude slower than both DRAM and SRAM. Data transfer rates for storage are typically measured in megabytes per second (MB/s) or gigabytes per second (GB/s), which pale in comparison to the nanosecond-level access times of cache and main memory. Conversely, storage memory offers the highest capacity in the entire hierarchy. Modern HDDs and SSDs provide capacities ranging from hundreds of GB to several terabytes (TB), and in enterprise-level solutions, even petabytes (PB). This vast capacity is indispensable for storing the operating system, numerous applications, vast user files, and large multimedia content that are not actively being processed by the CPU but are required for long-term retention.

In essence, the memory hierarchy is an ingenious and indispensable architectural solution that orchestrates the flow of data within a computer system. By intelligently combining memory technologies with varying speeds, capacities, and costs, and by leveraging the principle of locality, the system effectively creates the illusion of a single, enormous, and exceptionally fast memory. This masterful design optimizes data access patterns, minimizes performance bottlenecks, and maximizes overall computational efficiency, thereby allowing the CPU to operate at its full potential and ensuring a responsive and powerful computing experience.

1.2. EMERGING NON-VOLATILE MEMORIES

Non-Volatile Memories (NVMs) are a type of memory that retains stored data even when power is removed, which plays a critical role in maintaining data integrity in computing systems [4], [5]. Traditional NVM, such as Flash, has been widely used in applications ranging from consumer electronics to enterprise storage due to its high density and cost-effectiveness. With the growing demand for faster, more energy-efficient, and highly durable memory solutions, emerging NVM technologies have gained significant attention.

Among these emerging memories, Magnetoresistive Random-Access Memory (MRAM) applies Magnetic Tunnel Junctions (MTJs) with magnetization-switchable thin ferromagnetic layer to store information, offering high speed and high endurance [6], [7]. Resistive Random-Access Memory (RRAM) utilizes the resistance switching mechanism in metal oxides by metal or oxide filaments, to achieve high scalability [8], [9]. Phase-Change Memory (PCRAM) relies on the reversible phase transition between amorphous and crystalline states in specific chalcogenide materials [10], [11]. Ferroelectric Random-Access Memory (FeRAM) uses the polarization of a Ferroelectric (FE) material to store data, offering fast read and write speeds, low power consumption, and high endurance [12], [13]. Ferroelectric Field-Effect Transistor (FeFET) memory employs a FE layer in the transistor gate to enable non-volatile charge storage with low power consumption [14], [15].

Each of these next-generation NVM technologies presents unique advantages and trade-offs in terms of speed, power efficiency, endurance, and scalability. As the limitations of traditional Flash memory become more pronounced, these novel memory solutions are being actively explored for applications in embedded systems, AI accel-

Table 1.1: Summarization of emerging non-volatile memories.

Technology	Mechanism	Work Voltage	Write time	Read time	Write energy	Retention	Endurance
MRAM	Magnetic tunnel	<2 V	<10 ns	<10 ns	0.1 pJ	>10y	>10 ¹⁴
RRAM	Resistive switching	<5 V	<10 μ s	<10 ns	10 pJ	>10y	>10 ⁴
PCRAM	Phase change material	<4 V	<100 ns	<10 ns	10 pJ	>10y	>10 ⁹
FeFET	Ferroelectrical material	<6 V	10 ns-10 μ s	<100 ns	1 pJ	>10y	>10 ⁷
FeRAM	Ferroelectrical material	<6 V	<30 ns	<100 ns	10 pJ	>10y	>10 ⁷

erators, neuromorphic computing, and next-generation storage architectures. Looking ahead, emerging NVMs are expected to complement and, in some cases, replace conventional memory technologies in various applications. With continuous advancements in fabrication processes and material engineering, these memories are anticipated to achieve higher densities, lower power consumption, and improved reliability, making them viable candidates for future computing systems. As the demand for high-speed, low-power, and non-volatile storage grows in areas such as artificial intelligence, edge computing, and cloud infrastructure, NVMs are poised to play a critical role in shaping the next generation of memory hierarchy. While challenges such as scalability, cost-effectiveness, and integration with existing architectures remain, ongoing research and development efforts are steadily bringing these technologies closer to widespread commercialization. Ultimately, non-volatile memories hold the potential to redefine the balance between performance, energy efficiency, and data retention, driving innovations across a broad spectrum of computing and storage applications.

Next in this section, RRAM, PCRAM, FeFET, and FeRAM are briefly described in sequence. This work focuses on MRAMs, which will be described in detail in the next section. The following table summarizes the mechanism and the parameters of emerging non-volatile memories.

1.2.1. RESISTIVE RANDOM-ACCESS MEMORY

A RRAM device features a Metal-Insulator-Metal (MIM) configuration. Fig. 1.2 displays two kinds of RRAM. Its structure consists of a central metal oxide layer—typically TiO_x or HfO_x - sandwiched between two metallic electrodes: the Top Electrode (TE) and Bottom Electrode (BE). Sometimes, a capping layer, often composed of Hf, Ti, or Ta, is integrated between the TE and oxide (not shown in the figure) as an oxygen reservoir, which manages oxygen ions to stabilize the formation and dissolution of the conductive filament, thereby improving the ON/OFF ratio and reproducibility, yet with a cost of reliability [16], [17].

The application of a forming voltage induces localized oxygen vacancies, leading to the creation of a Conducting Filament (CF) between the electrodes. This forming process, essentially a soft breakdown of the insulating layer, results in a reversible conductive pathway within the dielectric. The device's resistance state is dictated by the CF's length, with an extended CF corresponding to a low-resistance state (logic '1') and a shorter CF representing a high-resistance state (logic '0'). While Conductive Bridging Random Access Memory (CBRAM) and Oxide-based Resistive Random Access Memory (OxRAM) exhibit comparable electrical behavior, OxRAM stands out due to its su-

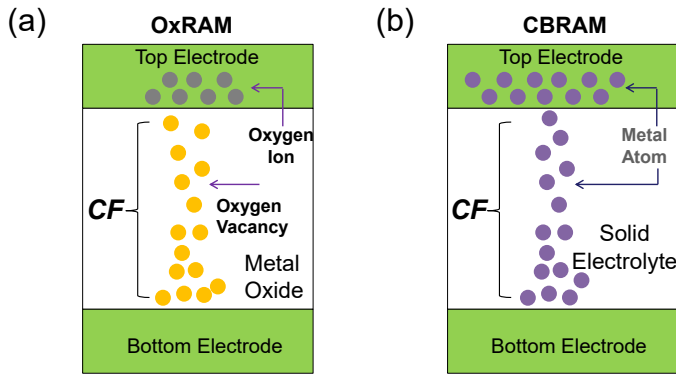


Figure 1.2: Structure of RRAMs (a) OxRAM and (b) CBRAM.

rior endurance and more straightforward fabrication process [18], [19].

It is possible to change an RRAM device's resistance between different states by applying certain programming voltages to it. As the resistance level rises, binary RRAM can be categorized into five states [20]–[22]: (1) the excessively high conductance faulty state 'H', (2) the correct Low Resistive State (LRS) '1', (3) an indeterminate faulty state, referred to as the intermediate range between LRS and HRS, labeled as 'U' [20], (4) the correct High Resistive State (HRS) '0', and (5) the extremely low conductance faulty state 'L'. The length of the CF increases when a positive voltage applied from TE to BE surpasses a specific threshold ($V_{TE} \geq V_{SET}$), leading to the formation of additional Oxygen Vacancies (OV). On the other hand, some oxygen ions shorten the CF by migrating back from the capping layer and filling the vacancies.

1.2.2. PHASE CHANGE RANDOM ACCESS MEMORY

PCRAM is a type of non-volatile memory that uses the unique properties of phase-change materials to store data. Unlike traditional memory technologies, PCRAM stores information by changing the phase of the material, typically from a crystalline to an amorphous state or vice versa [10], [11]. These phase transitions result in significant changes in the material's electrical resistance, which can be utilized to represent binary data. PCRAM combines the advantages of high-speed access and non-volatility, making it a promising candidate for future memory devices.

There are several key types of PCRAM, distinguished primarily by the phase-change materials used. The most common material is GeSbTe (germanium-antimony-tellurium), which undergoes a reversible phase transition between amorphous and crystalline states when applied with electrical pulses [23]. Other materials, such as AgInSbTe (silver-indium-antimony-tellurium) and Si-Ge alloys, are also explored for use in PCRAM due to their unique phase-change characteristics and potential for higher performance [24], [25].

Fig. 1.3 presents the planar and confined device structure of PCRAMs [23], where the phase change material is placed differently between these two structures. The typical PCRAM device structure consists of three main components: the top and bottom elec-

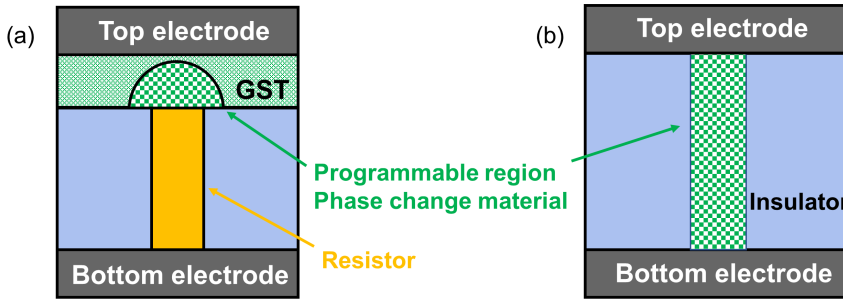


Figure 1.3: Structure of PCRAM (a) planar cell structure, and (b) confined cell structure.

trodes, which are applied with the electrical pulses needed to induce phase changes, and the phase-change material, which is the core of the memory cell. This material is sandwiched between the electrodes and is typically in a thin film form. When a high voltage is applied, it heats the material to a temperature where it can switch between its amorphous and crystalline states. A crystalline phase typically represents an LRS, while the amorphous phase corresponds to an HRS.

PCRAM operates based on the principle of material phase-change [26]: when a voltage pulse is applied, it causes localized heating of the phase-change material. A fast pulse induces a rapid transition to an amorphous state, while a slower, lower pulse allows the material to crystallize, changing the resistance. This resistance change is read by measuring the current passing through the material, with the two distinct resistance states corresponding to the binary values 0 and 1.

In conclusion, PCRAM offers several advantages, such as high-speed data access, low power consumption, and non-volatility, making it a strong candidate for applications in both storage and memory technologies [27]. Its ability to combine fast switching speeds with the benefits of non-volatile storage places PCRAM at the forefront of next-generation memory solutions.

1.2.3. FERROELECTRIC FIELD-EFFECT TRANSISTOR

FeFET is a type of non-volatile memory that incorporates FE materials into the gate stack of a conventional Field-Effect Transistor (FET) [12], [13]. Unlike traditional charge-based memory technologies, FeFET stores two-bit data based on the remanent polarization of FE materials. This enables advantages such as low power consumption, high-speed operation, and long data retention. Due to its compatibility with CMOS technology, FeFET is considered a promising candidate for next-generation embedded and standalone memory applications.

A typical FeFET consists of a FE layer integrated with a standard MOSFET structure [28]. Fig. 1.4 presents the device structure and the Transmission Electron Microscopy (TEM) picture of the gate stack [29]. The key component is the FE layer, commonly made of $HfZrO_2$ (hafnium zirconium oxide) or PZT (lead zirconate titanate), placed between the gate electrode and the transistor's channel [30], [31]. Unlike conventional dielectrics, this FE layer can retain its polarization state even after the applied voltage is removed,

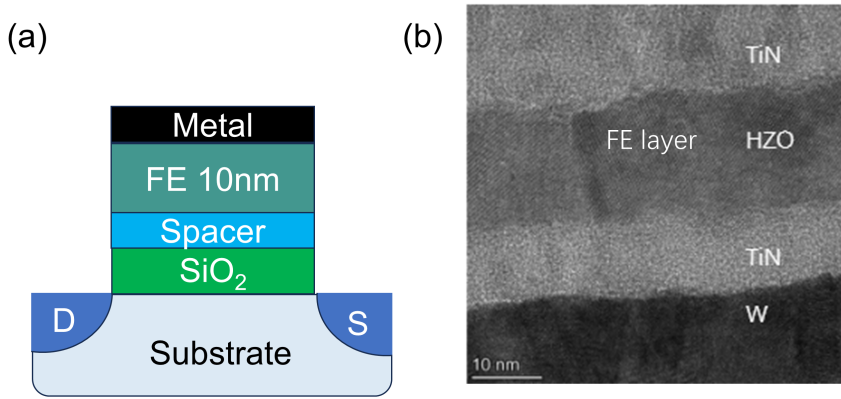


Figure 1.4: (a) Device structure of FeFET, (b) TEM of gate stack in FeFET.

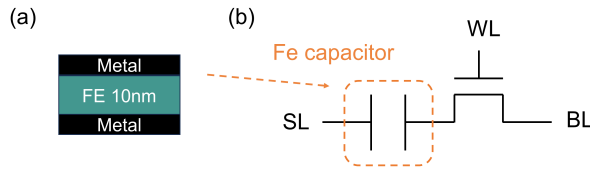


Figure 1.5: (a) Structure of FE capacitor, (b) FeRAM cell structure.

enabling non-volatile data storage.

FeFET operates based on the table polarization states of the FE material [32]. The FE layer consists of multiple domains; when all domains are polarized positively (\uparrow), the FeFET exhibits a high threshold voltage (V_{th}), also known as High Threshold Voltage (HVT) or logic 0'. Conversely, when all domains are polarized negatively (\downarrow), the FeFET demonstrates a Low Threshold Voltage (LVT), or logic '1'.

One of the key advantages of FeFETs is their ultra-low power consumption, as they require minimal energy to switch states, making them suitable for low-power applications. They also offer high-speed operation, with switching speeds comparable to DRAM. Additionally, FeFETs exhibit long data retention, capable of storing information for over a decade, outperforming traditional Flash memory. Furthermore, FeFETs are compatible with CMOS fabrication processes, allowing easy integration with existing semiconductor technology.

1.2.4. FERROELECTRIC FIELD RANDOM-ACCESS MEMORY

FeRAM is a type of non-volatile memory that utilizes the FE properties of certain materials to store data [12], [13]. Unlike traditional charge-based memories such as DRAM and Flash, FeRAM relies on the remanent polarization of a FE capacitor to represent binary data. This unique mechanism allows FeRAM to achieve fast read/write speeds, low power consumption, and high endurance, making it a promising alternative for embedded and low-power applications.

The core structure of FeRAM consists of an FE capacitor and a transistor-based access device (1T-1C), similar to DRAM [33]. Fig. 1.5 presents the FE capacitor and the FeRAM cell structure. The FE capacitor typically utilizes FE materials as its dielectric, which exhibit stable polarization states even after the applied voltage is removed. The capacitor is connected to a transistor that controls access to the memory cell. When a voltage is applied, the polarization direction of the FE material changes, corresponding to the storage of binary data (0 or 1).

FeRAM operates based on the hysteresis behavior of FE materials [34]. When an electric field is applied across the FE capacitor, its polarization switches between two stable states, each representing a distinct resistance level. Reading the stored data involves detecting this polarization state. However, the read process is destructive, meaning that the data must be rewritten after being read, similar to DRAM. Despite this, FeRAM offers much faster write speeds compared to Flash memory and does not require high-voltage programming.

One of FeRAM's key advantages is its low power consumption, as it requires significantly less energy per write cycle than Flash memory or DRAM [35]. It also features high-speed operation, with access times comparable to SRAM and DRAM. Additionally, FeRAM exhibits high endurance, capable of sustaining billions to trillions of write cycles without significant degradation, unlike Flash, which suffers from wear over time. However, FeRAM still faces challenges, such as its relatively lower storage density compared to Flash, which limits its scalability for high-capacity applications.

1.3. INTRODUCTION TO MRAMS

As an emerging non-volatile memory technology, MRAM utilizes magnetic properties to store data, distinguishing it from traditional charge-based memory systems, such as DRAM or Flash. Due to this specific mechanism, MRAM achieves non-volatility, high-speed, low power consumption, and high endurance, making it an attractive option for applications requiring both performance and reliability.

In this section, we introduce the fundamental principles of MRAM, including its core structure based on MTJs and the magnetoresistive effect that enables data storage. The classification of MRAM technologies is presented according to their switching mechanisms, highlighting key differences between different MRAM types. Finally, we discuss the potential applications of MRAM in various fields and its prospects as a next-generation non-volatile memory technology.

1.3.1. MTJ AND MAGNETORESISTANCE

The MTJ is the key component in MRAM, consisting of three main layers: a Reference Layer (RL), a Tunnel Barrier (TB), and a Free Layer (FL). Fig. 1.6 (a) and (b) present the structure of an in-plane MTJ in P state and AP state, respectively, in which the magnetization is within the plane of the two layers. Another type of MTJ is the perpendicular MTJ, in which the magnetization of the FL and the RL is perpendicular to the layer plane (out-of-plane), as presented in Fig. 1.6 (c) and (d). The RL has a stable magnetization direction, while the FL can change its magnetization in response to external stimuli (e.g., magnetic field). These two layers are separated by a thin TB, typically made of Magne-

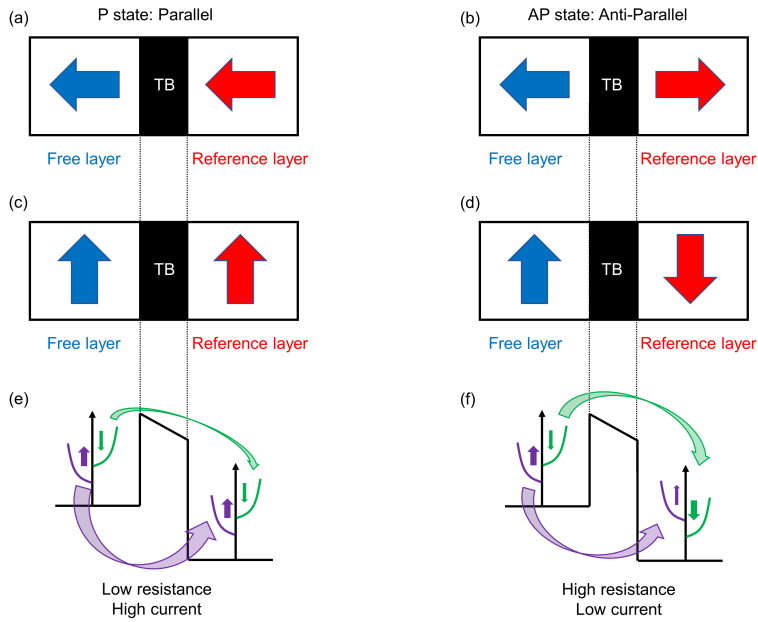


Figure 1.6: (a) Perpendicular MTJ in P state; (b) Perpendicular MTJ in AP state; (c) In-plane MTJ in P state; (d) In-plane MTJ in AP state; (e) Tunneling current dependence on the magnetization in P state; (f) Tunneling current dependence on the magnetization in AP state.

sium Oxide (MgO), which allows quantum mechanical electron tunneling. The electrical resistance of the MTJ depends on the relative alignment of the magnetization between the FL and RL, forming the basis of data storage in MRAM.

The Tunneling Magnetoresistance (TMR) effect describes how the resistance of an MTJ changes based on the relative magnetization of its two ferromagnetic layers [36]. According to Jullière's model, this resistance change is determined by the spin-dependent density of states in the FL and the RL [37]. When the magnetization of the two layers is in parallel (P), the majority spin electrons from one layer can easily tunnel through the barrier and occupy available majority spin states in the other layer, leading to high conductance and low resistance, as presented in Fig. 1.6 (e). In contrast, when the magnetization of the two layers is in antiparallel (AP), the majority spin electrons in one layer do not find enough available states in the other, reducing the tunneling probability and resulting in low conductance and high resistance, as presented in Fig. 1.6 (f). This significant difference in resistance between the two states enables reliable binary data storage in MRAM.

1.3.2. SWITCHING MECHANISM AND CLASSIFICATION OF MRAMS

The switching mechanism of MRAM is based on controlling the magnetization of the free layer in an MTJ to toggle between P and AP states, thereby altering resistance and storing binary data. Depending on the method used to switch the free layer's magne-

Table 1.2: Comparison of different MRAMs.

Feature	STT-MRAM	SOT-MRAM	VCMA-MRAM
Write Mechanism	STT by current [38]	SOT by injecting current [39]	VCMA by electric field [40]
Write Energy per Bit	100 fJ [41]	10–50 fJ [39]	<1 fJ [40]
Switching Speed	1–10 ns [42]	<1 ns [43]	<100 ps (potential) [40]
Endurance	10^{12} cycles [42]	$>10^{15}$ cycles [43]	$>10^{15}$ cycles (potential) [40]
Retention Time	>10 years [42]	>10 years [43]	Very limited [40]

tization, MRAM can be classified into Spin-Transfer Torque MRAM (STT-MRAM), Spin-Orbit Torque MRAM (SOT-MRAM), and Voltage-Controlled Magnetic Anisotropy MRAM (VCMA-MRAM) [44]. Table 1.2 presents the performance of different types of MRAMs.

STT-MRAM switches the magnetization of the free layer by applying a spin-polarized current directly through the MTJ [38]. The Spin-Transfer Torque (STT) generated by this current exerts a force on the free layer magnetization, causing it to flip between P and AP states. STT-MRAM offers high density, non-volatility, and compatibility with CMOS fabrication, making it a strong candidate for replacing SRAM and DRAM in various applications. However, it suffers from relatively high write energy consumption, as the switching current must be sufficiently large to overcome the energy barrier, leading to higher power dissipation and potential reliability concerns due to increased current stress on the MTJ.

Instead of passing current through the MTJ, SOT-MRAM injects current into an adjacent heavy metal layer (such as Pt, Ta, or W), where spin-orbit interaction generates spin current [39]. This spin current applies Spin-Orbit Torque (SOT) to the free layer, enabling magnetization switching with lower power and higher speed compared to STT. The advantages of SOT-MRAM include faster switching, higher endurance due to reduced stress on the MTJ, and improved reliability. However, its main drawback is that it requires an additional write electrode, increasing the overall cell size and limiting its scalability for high-density applications.

VCMA-MRAM takes a different approach by using an electric field rather than a spin-polarized current to overcome the energy barrier of the free layer [40]. By applying a voltage across the MTJ, the magnetic anisotropy of the free layer is altered, allowing low-energy and high-speed switching. However, VCMA-MRAM is still in the early stages of development, with challenges related to retention, reliability, and compatibility with CMOS technologies [44]. Despite these limitations, its potential for energy-efficient and high-density memory makes it a promising candidate for future low-power applications.

1.3.3. APPLICATION OF MRAMS

MRAM is emerging as a potential application across various domains, including embedded systems, Artificial Intelligence (AI), and standalone memory solutions [45]–[48]. In embedded systems, where power efficiency and reliability are critical, MRAM attracts attention due to the non-volatility and low energy consumption [47]. AI workloads, which demand high-speed memory access and endurance, can benefit from MRAM's fast read/write capabilities, particularly accelerator applications [49]. Additionally, MRAM is

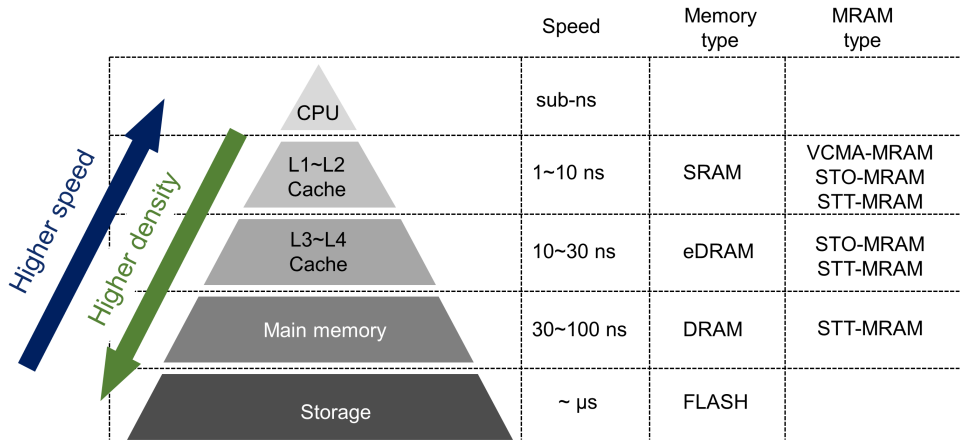


Figure 1.7: The potential application of different MRAM types within the memory hierarchy.

being explored as a standalone memory solution [48]. These diverse applications highlight the potential of MRAMs to reshape modern computing architectures by balancing speed, power efficiency, and data retention.

Within the memory hierarchy, different types of MRAM potentially serve specific roles. Fig. 1.7 presents a general memory hierarchy on the computer and the potential application of different MRAM types [50]–[52]. At the L1 and L2 cache level, STT-MRAM, SOT-MRAM, and VCMA-MRAM are considered potential replacements for traditional SRAM caches due to their super high speed switching (ns level) and energy efficiency [53], [54]. These technologies provide fast access times with lower power consumption, making them ideal for High-Performance Computing (HPC) and AI accelerators where quick data retrieval is essential.

For L3 and L4 caches, SOT-MRAM and STT-MRAM offer a compelling alternative due to their relatively high scalability and lower power consumption. These caches require speeds within 10 ns to maintain system performance while reducing overall power consumption and chip area [53], [54]. By integrating MRAM at these cache levels, system architects can balance performance and energy efficiency while reducing standby power losses.

At the main memory level, the STT-MRAM is being actively explored as a potential replacement for DRAMs [55]. While DRAM still holds advantages in terms of bandwidth and latency, its volatility and high power consumption limit further improvement in performance. The non-volatile nature of MRAMs means that data can be retained even when power is lost, reducing refresh energy costs and improving power efficiency. However, current STT-MRAM technology still faces challenges such as the relatively low density, though research is focusing on optimizing materials and architectures to make it a viable alternative. Main memory typically operates at tens of nanoseconds, making MRAM a promising but developing option in this space.

Beyond main memory, MRAM is also a candidate for storage-class memories, bridg-

ing the gap between DRAM and non-volatile storage like SSDs. Persistent MRAM solutions could enable faster and more energy-efficient secondary storage, reducing the performance bottlenecks associated with NAND flash memory. As MRAM technology continues to mature, it is expected to revolutionize computing architectures by offering a unified memory solution that balances speed, power efficiency, and data persistence, ultimately leading to more efficient and energy-conscious computing systems.

1.4. MEMORY TESTING PHILOSOPHY

Testing is essential for ensuring chip functionality and reliability. After their design, Very Large Scale Integration (VLSI) chips typically go through three key phases: manufacturing, where the chips are produced; system integration, where the fabricated chips are placed into products such as smartphones; and finally, deployment in which end users purchase and utilize the product. Because memory can contain millions or even billions of states, it is considerably more difficult to test than logic. Exhaustively checking every possible scenario for a memory chip is infeasible, so specialized testing approaches are required. This section begins by reviewing the major developments in traditional memory testing over the past several decades, then explores the latest advances in STT-MRAM testing.

1.4.1. TRADITIONAL MEMORY TESTING

The evolution of memory testing has been protracted. Prior to 1980s, there were no dedicated fault models or enhancements for memory tests, and thus they tended to be considered ad-hoc [56]. They typically demanded lengthy execution times and offered limited defect coverage, often on the order of $O(n^2)$, where n denotes the number of memory chips. Examples of such ad-hoc tests include the Zero-One test, the Galpat test, and the Walking 1/0 test, as discussed in [56], [57].

With the exponential growth in memory capacity, test development began focusing on potential memory defects to reduce both testing time and per-chip costs. This shift resulted in the introduction of numerous functional fault models in the early 1980s. Although tests using these models typically run in $O(n)$ time - linear with the memory size - they provide provable fault coverage. Additionally, address decoder faults and stuck-at faults were two pivotal models introduced in that era [57], [58]. Both are abstract representations, not directly tied to specific physical defects or memory architectures.

Experimental results from Defective Part Per Billion (DPPB) screening, conducted on numerous memory chips in the late 1990s, revealed that many detected defects could not be accounted for by the prevalent fault models [59], [60]. This discovery pointed to additional defect mechanisms and spurred the creation of novel static and dynamic fault models. Examples include read destructive faults, write disturb faults, transition coupling faults, and read destructive coupling faults, which were derived through linear resistor defect injection and SPICE simulation [61], [62].

Essentially, the classical linear-resistor-based test method shares notable similarities with the modern Cell-Aware Test (CAT) technique [63], [64]. As CMOS technology evolves to more intricate process nodes, internal defects within library cells have become increasingly prevalent. Conventional testing—aimed primarily at faults at cell

ports—was found to detect under 50% of these intra-cell issues [63]. Over the past decade, industry has responded by introducing CAT, a methodology specifically designed to address cell-internal defects, which has proved valuable for emerging process nodes such as 45 nm, 32 nm, and even FinFET-based 14 nm technology [64]. In 2022, TSMC plans to commence large-scale production of 3 nm FinFET technology as node scaling continues. Nevertheless, fabrication-induced defects and device characteristic variations, and their impact on overall system quality and reliability, pose major challenges, especially when higher quality levels (down to DPPB ranges) are required [65]. Whether CAT can effectively detect internal device defects remains open to debate, as it primarily targets resistive faults such as bridges, shorts, and opens at device terminals and interconnects.

1.4.2. TESTING OF EMERGING NON-VOLATILE MEMORIES

Testing of emerging NVMs primarily involves detecting and diagnosing (i.e., identifying defect types if possible) defects that are unintendedly introduced during the fabrication process. NVMs are more vulnerable to defects compared with conventional memories like SRAMs; this is due to the additional fabrication steps involved in NVM manufacturing, and unique defects are introduced into NVMs due to their specific structure and working mechanism [29], [66], [67]. Hence, it is critical to develop dedicated test solutions for NVMs. Defects in NVMs can be classified into two categories: 1) conventional defects that also exist in traditional memory technologies such as SRAMs and DRAMs, and 2) unique defects that arise from the novel physical mechanisms underlying these memories [68].

Conventional defects in emerging NVMs are similar to those found in traditional memory technologies, including interconnect and contact defects, which are modeled as opens, shorts, and bridges [61], [69]. For example, open defects can cause cells to get stuck permanently at logic ‘0’ or ‘1’, while short defects may lead to unintended state switching [70]. All conventional defects can be modeled as linear resistors, injected into the memory array, and analyzed using standard fault modeling techniques. Common approaches for detecting these defects involve conventional memory test methodologies, including March tests and Design-for-Testability (DfT) techniques [71].

Beyond conventional defects, unique defects are observed in emerging NVMs, which arise due to their distinct physical mechanisms and structures [29], [66], [67]. Some of these unique defects include pinholes in STT-MRAM and ion depletion in RRAMs; these defects are associated with certain NVM types, and typically will not appear in other memory types [66], [72]. These unique defects within NVMs have been reported to cause irregular faulty behaviors, which cannot be modeled with linear resistors. Therefore, understanding and detecting these defects requires specialized defect modeling, fault models, and testing techniques tailored to the specific characteristics of each memory technology. To address this challenge, Device-Aware Test (DAT) methodology is applied, targeting these unique defects by incorporating physical mechanisms into the compact models of memory devices and designing defective compact models. The DAT method has been proven powerful in testing unique defects.

Despite the achievements of the previous works in NVM testing, this field still faces several challenges that require innovative solutions. For example, the working mech-

anism of some NVMs is not completely understood, and the complete defect space is not totally set, causing difficulty in defect modeling. Future research in NVM testing aims to identify new defects, develop more accurate fault models, and more efficient test methods. Solutions include exploring machine learning techniques in testing and enhancing DfT strategies to improve defect coverage while minimizing test overhead. As emerging NVMs move toward commercialization, high coverage and cost-effective testing methodologies will be critical to ensure their manufacturing.

1.5. RESEARCH TOPICS AND CONTRIBUTION OF THE THESIS

Playing a key role in STT-MRAMs, the MTJ also poses additional challenges in achieving high-quality test solutions, such as the high escape rate in detecting unique defects [66]. This thesis focuses on defect modeling and manufacturing testing of STT-MRAMs, a crucial area for ensuring the yield and reliability of STT-MRAM technology. The research in this thesis covers:

- Testing of transistor, contact, and interconnect defects in STT-MRAMs, which addresses conventional defects arising from the CMOS fabrication process.
- Testing of MTJ defects in STT-MRAMs, focusing on unique defects within MTJs arising from the MTJ fabrication process.
- A case study of testing design in STT-MRAM chips, providing practical insights into test implementation for an industry STT-MRAM chip.
- Diagnosing MTJ defects in STT-MRAMs, enabling distinction of unique defects and improving the yield learning.

1.5.1. TESTING OF TRANSISTOR, CONTACT, AND INTERCONNECT DEFECTS

Defects in STT-MRAMs can be classified into two types: (1) conventional defects outside MTJs, like opens and shorts [73]–[78]; (2) unique defects inside MTJs, like pinholes, the Intermediate state, and the back-hopping [79]–[82]. This part focuses on testing conventional defects, i.e., transistor, contact, and interconnect defects.

To detect conventional defects, researchers extend what has been done on traditional memories, like DRAMs and SRAMs [61], [69], to STT-MRAMs [73]–[78]. They inject defects modeled as linear resistors, and perform fault modeling by observing MRAM faulty behavior in SPICE simulations. However, this conventional approach fails to deliver high-quality tests due to ignoring the impact of the MTJ specific mechanism on the fault modeling of conventional defects (e.g., magnetic coupling) [80], [83]. These effects, intrinsic to the nature of MTJs, introduce complexities in faulty behaviors in the presence of conventional defects [61], [69], [84], leading to inaccurate fault models and the potential escapes of the generated test solutions. As a result, insufficient test coverage reduces the quality and reliability of STT-MRAM production. Therefore, a more comprehensive fault modeling and testing approach is required. This work aims to develop enhanced fault models that integrate the impact of the following MTJ mechanisms:

Magnetic coupling: We show that in the presence of interconnect and contact defects, both electrical effects (e.g., unwanted current paths by bridge defects [61]) and

magnetic coupling may foster the occurrence of such coupling faults for STT-MRAMs. A proper fault modeling methodology is put forward that includes both electrical effects and magnetic coupling. Thus, accurate fault models are generated. Two test methods are proposed, one is based on regular write-read operations (e.g., March tests), and the other applies the DfT method, which modifies the write pulse width.

MTJ stochasticity: The inherent property ‘intrinsic switching stochasticity’ in MTJ switching causes unavoidable random write errors [41]. This stochasticity is modeled and included in fault modeling to obtain accurate fault models. In the precise of conventional defects, intermittent faults are generated with the MTJ stochasticity. A DfT methodology dedicated to the detection of intermittent faults due to conventional defects (e.g., small interconnect and contact defects); it is based on the monitoring of write currents. The DfT provides a higher fault coverage for intermittent faults compared to functional test methods (e.g., March tests).

1.5.2. TESTING OF MTJ DEFECTS

This part focuses on the testing of MTJ defects (i.e., unique defects), which are introduced due to the specific device structure and working mechanism of MTJs [66]. In the presence of unique defects, the defective STT-MRAM behaves irregularly; hence, these defects cannot be modeled by linear resistors. To solve this challenge, DAT is used, in which compact models for defective MTJs are designed by incorporating the impact of physical (manufacturing) defects on electrical and magnetic behavior [67], [79], [85]. The fault modeling is followed by simulating with defective MTJs by Spice and observing the corresponding faulty behaviors. Dedicated test solutions, including March tests and DfTs, are proposed to catch these faults. DAT has been proven powerful in detecting unique defects in STT-MRAMs [67], [79], [85]. Previous works have applied the DAT method to unique defects of pinholes, Synthetic Anti-Ferromagnet Flip (SAFF), and Intermediate State (IM) defects [66], [86], [87]. This thesis extends the DAT method to another identified unique defect Back-Hopping (BH). The BH is characterized, which favors the following defect modeling and fault modeling.. Both read/write-based tests (e.g., March test) and DfT with dedicated circuit design are proposed for BH detection. Additionally, several test methods are proposed targeting a variety of unique defects.

1.5.3. TESTING STT-MRAMS: FROM DEFECTS TO TEST CHIPS

This part presents a structured testing approach for an industrial STT-MRAM chip by combining the findings of the whole thesis. Targeting both conventional and unique defects, we analyze their corresponding faults and evaluate potential test solutions, including March tests and DfT. Notably, we systematically assess various DfT techniques (to the best of our knowledge) and present a case study implementing a selected DfT in an industrial STT-MRAM array, covering circuit-to-layout integration to demonstrate practical feasibility.

1.5.4. DIAGNOSIS OF MTJ DEFECTS

The diagnosis of MTJ defects is a further advancement based on the DAT methods, introducing the Device-Aware Diagnosis (DA-Diagnosis) method. This method comprises two steps: a) defining distinctive features of each unique defect by characterization and

physical analysis of defective MTJs, and b) utilizing March algorithms to extract distinctive features. This DA-Diagnosis facilitates the development of diagnosis algorithms that can efficiently distinguish the different unique features of the different defects. The effectiveness of the proposed approach is validated in an industrial setting with real devices and data measurement.

1.6. THESIS ORGANIZATION

This thesis explores defect modeling and manufacturing testing of STT-MRAMs, addressing key challenges in ensuring their reliability and scalability. Chapters in this thesis are organized as follows.

Chapter 2 introduces STT-MRAM technology. It begins with an overview of the MTJ, explaining its working mechanisms. Next, it discusses the STT-MRAM cell and array, focusing on their integration and functionality. Peripheral circuits, such as sense amplifiers and write drivers, are also covered. Then, the chapter reviews the manufacturing process of the STT-MRAM and potential defects affecting device performance. Finally, the modeling methodology of MTJ is presented, detailing how to simulate specific mechanisms such as intrinsic stochasticity.

Chapter 3 introduces the compact model for MTJs. It begins with an overview of the conventional MTJ models, including the resistance part and the switching part. Then, this chapter incorporates intrinsic stochasticity into the model, which leads to unavoidable write errors and read errors. Furthermore, the chapter analyzes the impact of magnetic fields on MTJ switching.

Chapter 4 reviews existing testing approaches for STT-MRAMs. It first discusses conventional testing techniques, including resistor-based defect modeling and fault modeling. The limitations of these methods in detecting unique defects within MTJs are evaluated. Then, the chapter introduces the DAT method, which integrates the MTJ mechanism into the MTJ compact, and designs a specific defective MTJ model.

Chapter 5 covers testing transistor, contact, and interconnect defects in STT-MRAMs. It starts by presenting a complete defect map based on the bit-oriented STT-MRAM array. Fault modeling is developed with the consideration of specific MTJ properties, like the magnetic coupling (i.e., stray field). March tests are designed to efficiently detect derived faults. Additionally, DfT methodologies are proposed to improve defect coverage and reduce test time.

Chapter 6 introduces the new MTJ defect called BH. DAT is applied to tests of BH, consisting of defect characterization, defect modeling, fault modeling, and test generation. Additionally, new DfT strategies are put forward to improve defect coverage and reduce test time. These techniques aim to enhance the overall manufacturability and reliability of STT-MRAMs.

Chapter 7 combines all the previous knowledge & studies to generate a structured testing approach for STT-MRAMs, and applies it to an industrial STT-MRAM chip. A customized test solution is proposed, addressing various defects with a combination of March tests and DfT techniques. The approach ensures a high defect detection while optimizing test efficiency. Simulation results validate the effectiveness of the methodology, and the chip is sent to tape out.

Chapter 8 focuses on diagnosing MTJ defects using DA-Diagnosis approach. It ex-

plores techniques to accurately identify MTJ defects based on electrical responses. Advanced fault localization methods are introduced to improve diagnosis accuracy. The effectiveness of these methods is demonstrated through simulations and experimental results. This approach enhances defect analysis, aiding yield learning.

Chapter 9 summarizes the key contributions of this work in STT-MRAM defect modeling and testing. It highlights the advancements made in defect analysis & modeling, fault detection, and test generation. The potential future research directions are outlined.

2

STT-MRAM TECHNOLOGY

The STT-MRAM is on the way to commercialization, thanks to its competitive write performance, endurance, retention, and low power consumption [88]. Since the initial product in 2006, world-leading foundries have entered the MRAM market, and STT-MRAMs are applied in a wide range of applications, like embedded systems, computing-in-memory systems, and are considered potential replacements for SRAMs. This chapter presents the fundamental concepts of STT-MRAMs, covering their key components and operating principles. It begins with an overview of the MTJ structure and its switching mechanism, followed by a discussion of the 1T-1MTJ memory cell and its corresponding read/write operations. The organization of STT-MRAM arrays is then introduced, along with the structure of peripheral circuits. Additionally, the modeling of MTJ is described. These topics provide a foundation for understanding the characteristics and functionality of STT-MRAM in various applications.

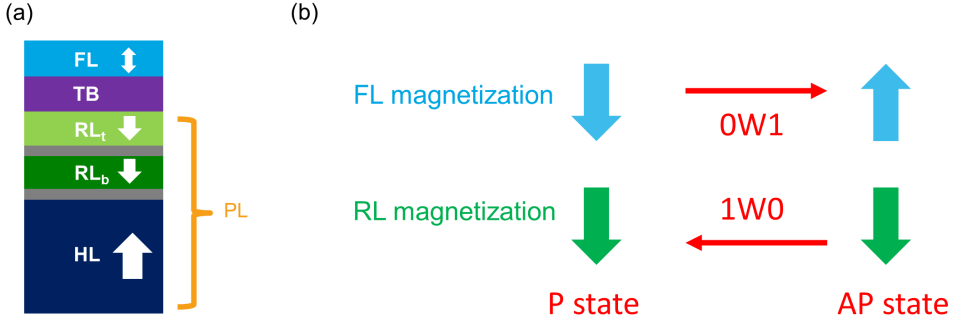


Figure 2.1: (a) Structure of perpendicular MTJ; (b) MTJ switching mechanism.

2.1. MTJ STRUCTURE AND WORKING MECHANISM

2.1.1. DEVICE STRUCTURE

The fundamental data-recording element in STT-MRAMs is the perpendicular MTJ; it demonstrates the one-bit data storage by encoding two bi-stable resistance states. For convenience, the ‘MTJ’ mentioned in the following of this thesis all refers to the perpendicular MTJ. Fig. 2.1 (a) presents a simplified schematic of an MTJ [89].

Typically, the MTJ consists of an ultra-thin dielectric TB sandwiched between the FL and the Pinned Layer (PL). The FL is a thin ferromagnetic layer (e.g. 1 nm), commonly made of CoFeB, which exhibits high spin polarization and compatibility with MgO-based tunnel barriers [90], [91] Its magnetization can be switched through write operations, enabling data storage. The TB is a thin insulator made of MgO, the thickness of which is around 1 nm. MgO is chosen due to its ability to provide a high tunneling TMR ratio, which enhances the read signal contrast. In the meantime, MgO exhibits coherent tunneling, allowing for low Resistance-Area (RA) products [90].

The PL is, in itself, a multi-layer stack consisting of 1) a Top Reference Layer (RL_t), 2) a thin metal spacer, 3) a Bottom Reference Layer (RL_b), 4) another thin metal spacer, 5) a thick Hard Layer (HL). The magnetization of all the layers in the PL is stable, and their magnetization never switches. The RL_b is anti-ferromagnetically coupled to the HL through the metal spacer, resulting in their opposite magnetization directions. The magnetization direction of two reference layers (i.e., RL_t and RL_b) is the same, with the ferromagnetic coupling in between. The RL is conventionally made of the same material as the FL (e.g., CoFeB) to guarantee strong spin polarization between FL and RL [91]. The HL is thick (e.g., >5 nm) and stable, which ensures the stability of the RL through coupling [92].

2.1.2. WORKING MECHANISM

Fig. 2.1 (b) presents magnetization switching of $0w1$ and $1w0$ in the MTJ. The magnetization switching of MTJ in STT-MRAMs relies on the STT effect; the STT is a quantum mechanical effect where the angular momentum of spin-polarized electrons is transferred to a magnetic layer, exerting a torque on its magnetization [93], [94]. In an MTJ,

when a current flows through the device, spin-polarized electrons originating from the RL interact with the FL. If the current exceeds a critical threshold, the STT effect can switch the FL magnetization between two stable states: parallel or anti-parallel to that of the RL [95], [96].

The MTJ resistance is determined by the relative magnetization orientation of the FL and RL. When the magnetization of the two layers is in parallel, the MTJ exhibits low resistance (R_P), described as P state or '0' state. Conversely, when the magnetization of the two layers is in anti-parallel, the MTJ enters a high-resistance state (R_{AP}), referred to as the AP state or '1' state. This resistance difference arises from the TMR effect, which is explained in Chapter 1. The TMR can be calculated by [41]:

$$TMR = \frac{R_{AP} - R_P}{R_P} \quad (2.1)$$

TMR is a crucial parameter in STT-MRAM, as it determines the read margin and overall performance of the memory cell. A high TMR enhances the distinction between the '0' and '1' states of the MTJ, improving read stability and avoiding read errors. High-performance MTJs typically require TMR values greater than 1 (i.e., ON/OFF ratio larger than 2) to ensure robust signal detection, which is usually demonstrated exceptionally with MgO-based MTJs [41], [90], [97].

A summary of the technology parameters in MTJs is described in the 2.1, where the values refer to the MTJs applied in this thesis.

Table 2.1: MTJ technology parameters

Symbol	Definition	Value
A	Area	around 3600 nm ²
t_{FL}	FL thickness	around 1 nm
t_{RL}	RL thickness	around 2 nm
t_{MgO}	MgO barrier thickness	around 1 nm
M_s	Saturation magnetization	around 10 kOe
H_k	Magnetic anisotropy field	around 1.5 kOe

2.2. STRUCTURE OF 1T-1MTJ CELL

Fig. 2.2 (a) presents the 1 Transistor - 1 MTJ (1T-1MTJ) bit cell structure and the related read/write operations. This structure is widely used in STT-MRAMs due to its simplicity (i.e., low circuit area) and compatibility with CMOS technology [98]. Fig. 2.2 (b) presents the related TEM cross-section picture of the 1T-1MTJ bit cell (applied in this thesis), illustrating the physical integration of the MTJ and the MOSFET selector.

The 1T-1MTJ cell consists of an MTJ, where the bottom electrode is connected to an N-type Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET) selector. This configuration enables efficient read and write access to the memory cell. The cell includes three terminals: the Bit Line (BL), the Source Line (SL), and the Word Line (WL), which work together to control read and write operations.

During write operations, the voltage is applied to the WL to turn on the MOSFET and select the target cell. The voltage difference between the BL and the SL determines the current direction, which flows either from the FL to the PL or vice versa. When BL is

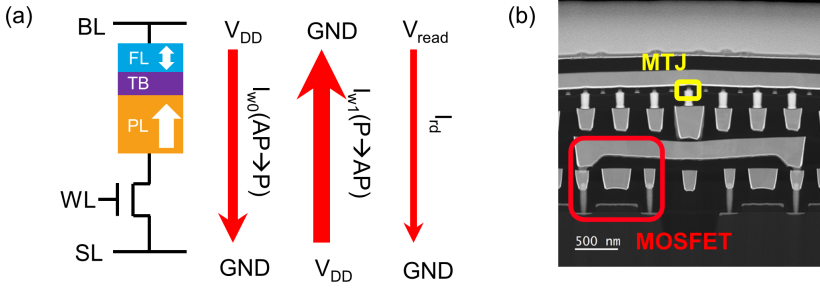


Figure 2.2: (a) Structure of 1T-1MTJ cell and the related read/write operations; (b) TEM picture of the cell.

connected to Supply Voltage (V_{DD}) and SL is grounded, the current flows from FL to PL, performing the ‘1w0’ operation (i.e., I_{w0} in Fig. 2.2 (a)). The spin-polarized electron flow induces STT on the FL, switching its magnetization from anti-parallel to parallel to that of RL, resulting in the AP to P transition.

Conversely, when BL is grounded and SL is connected to V_{DD} , the ‘0w1’ operation is applied, where the current I_{w1} flows in the opposite direction, causing the FL magnetization to switch from parallel to anti-parallel to that of RL. To achieve reliable switching, the write current I_w must exceed the Critical Current (I_c), ensuring a low Write Error Rate (WER) [41]. The gap $I_w - I_c$ is inversely proportional to the average switching time t_w , meaning that a higher write current leads to faster switching.

For read operations, a smaller Read Current (I_{rd}) is applied, significantly lower than I_c , to prevent unintended switching of the FL due to read disturbs and to reduce power consumption [99]. This I_{rd} is compared with a Reference Current I_{ref} to detect the resistance of the MTJ.

A summary of the electrical parameters in MTJs is described in Table 2.2, where the values refer to the MTJs applied in this thesis.

Table 2.2: MTJ electrical parameters

Symbol	Definition	Value
I_w	Write current, including w1 and w0	around hundreds of μA
I_{rd}	Read current, including r1 and r0	around tens of μA
R_{AP}	MTJ resistance in AP state (high resistance)	around 8 k Ω
R_P	MTJ resistance in P state (low resistance)	around 1.5 k Ω
TMR	Magnetoresistance value	larger than 1
I_c	Critical current	around 100 μA

2.3. STRUCTURE OF STT-MRAM ARRAY

Fig. 2.3 presents a bit-oriented STT-MRAM array/sub-array [100]. In this architecture, cells in the same row share a common WL, while those in the same column share the same BL and SL. This structure, like the NOR-FLASH structure [101], enables efficient memory access and reduces peripheral circuit complexity. The peripheral circuit consists of the Write Driver (WD), the address decoder, and the Sense Amplifier (SA).

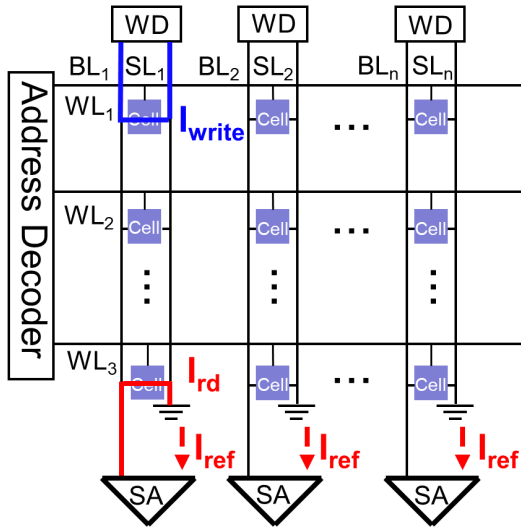


Figure 2.3: Bit-oriented STT-MRAM array.

The address decoder plays a crucial role in selecting memory cells for read and write operations by decoding input addresses and applying V_{DD} to the corresponding WL, ensuring accurate data access

The WD is responsible for generating an appropriate write pulse. During a write operation, the WD supplies a current through the selected STT-MRAM cell, attempting to switch the MTJ state. The current direction determines whether the FL magnetization switches from parallel to anti-parallel or vice versa (illustrated by the blue line in Fig. 2.3).

The SA is applied to sense the resistance of the MTJ in different states. In read operations, the SA applies a read pulse to the selected STT-MRAM cell and compares the resulting I_{rd} with a reference current. This comparison determines the resistance state of the MTJ: if $I_{read} > I_{ref}$, the MTJ is in the low-resistance state (i.e., '0'); otherwise, it is in the high-resistance state (i.e., '1') (illustrated by the red lines in Fig. 2.3).

2.4. PERIPHERAL CIRCUIT

The performance and reliability of STT-MRAMs significantly depend on their peripheral circuits, which include SA, WD, and the control logic (e.g., the address decoder). These circuits are responsible for enabling efficient read and write operations. This section introduces the address decoder, the WD, and the SA, which play the major role in the STT-MRAM read/write operations.

2.4.1. ADDRESS DECODER

Address decoders play a crucial role in memory arrays by selecting the memory cells for read and write operations. In general, a row address decoder processes the row address to activate a designated WL, while a column address decoder deciphers the column ad-

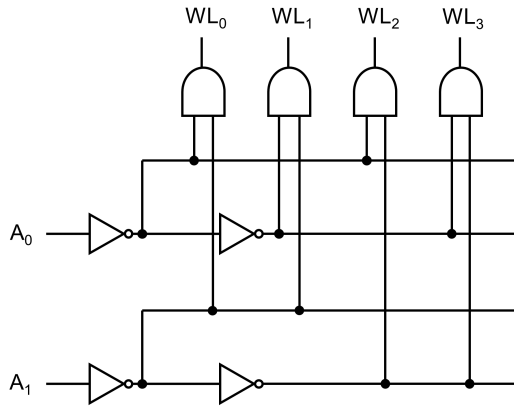


Figure 2.4: Example of a row address decoder circuit.

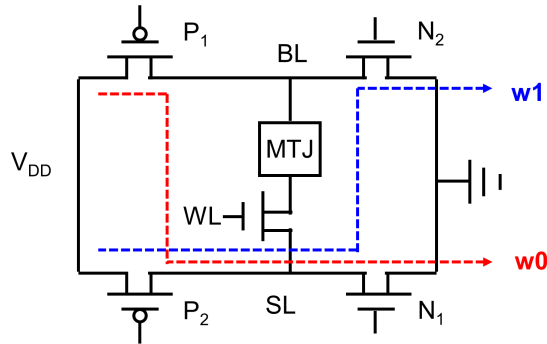


Figure 2.5: (a) Circuit of the WD and the current of w_0 and w_1 operations.

dress to enable a corresponding BL and SL. This mechanism ensures precise access to individual cells within the memory array. To optimize simulation efficiency, a compact STT-MRAM array (e.g., 3×3) was designed with essential peripheral circuits. Fig. 2.4 provides an example of a row address decoder, which demonstrates a basic design utilizing NOT and AND gates. This configuration includes two input address lines and four WL outputs, offering a straightforward approach to row selection. For example, if the inputs of the address decoder (i.e., A_0 and A_1) are both '0', the WL_0 is selected (i.e., '1'), and all other outputs are '0'.

2.4.2. WRITE DRIVER

The WD in STT-MRAMs plays a crucial role in controlling the bidirectional write current that switches the MTJ state between P and AP. The structure of the WD commonly used is shown in Fig. 2.5, where it consists of four large-sized MOSFETs (P_1 , N_1 , P_2 , and N_2) designed to minimize channel resistance and ensure that most of the applied voltage is utilized across the MTJ rather than being dissipated in the driver circuit.

During the w_0 operation, the transistors P_1 and N_1 are turned on, while P_2 and N_2

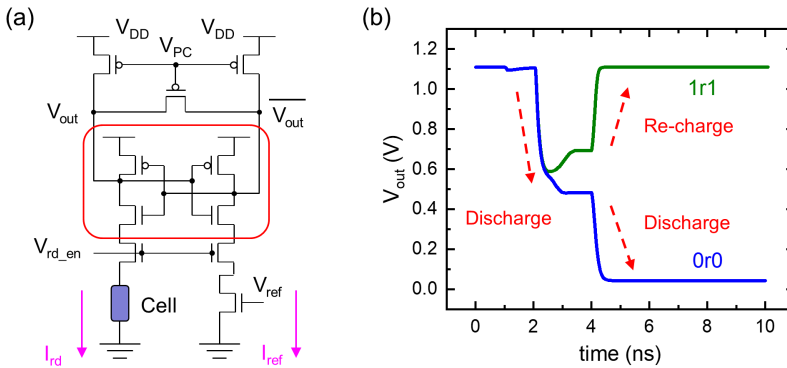


Figure 2.6: (a) Circuit of the SA (b) I_{rd} in 1r1 and 0r0 operations.

remain turned off. This configuration establishes a current path from the BL to SL, and the electrons flow from RL to FL, inducing STT that forces the MTJ into the P state, encoding a stored '0'. Conversely, during the $w1$ operation, P_2 and N_2 are switched on while P_1 and N_1 are turned off. This reverses the current direction, flowing from SL to BL, thereby switching the MTJ to the AP state, encoding a stored '1'.

The sizes of the transistors in the WD are carefully designed to ensure reliable switching by maintaining a sufficient write current amplitude. Larger MOSFETs with a high Width-to-Length (W/L) ratio are used to reduce their resistance, minimizing voltage drops that could otherwise reduce the effective write voltage applied to the MTJ. This optimization is critical for ensuring a robust and energy-efficient write operation, as insufficient current can lead to increased write error rates, while excessive current may degrade device reliability while increasing the power consumption.

One important consideration in WD design is the asymmetry in MTJ resistance between the P and AP state. Since the MTJ exhibits a lower resistance in the P state and a higher resistance in the AP state, maintaining identical sizes for N_1 and N_2 as well as for P_1 and P_2 can result in significantly different I_w for the $0w1$ and $1w0$ operations. Specifically, the write current during a $0w1$ operation tends to be much larger than that in a $1w0$ operation, leading to an imbalance in the MTJ average switching time in between. This discrepancy complicates precise control of write performance and can negatively impact MTJ reliability over time.

To address this issue, some designs adjust MOSFET sizes to equalize the write currents in both directions. By carefully tuning the W/L of P_1 , P_2 , N_1 , and N_2 , designers can ensure that the effective resistance of the MOSFETs compensates for the intrinsic asymmetry in MTJ resistance. This approach helps maintain the same MTJ average switching time for both operations, simplifying write control circuitry. A balanced write current also reduces the risk of excessive Joule heating or device degradation, ultimately contributing to the longer MTJ endurance.

2.4.3. SENSE AMPLIFIER

The SA in STT-MRAMs works for distinguishing the resistance of MTJs to determine stored data. Fig. 2.6 (a) presents the structure of a current-based SA [102]. The read operation begins with all signals at their default values, where the node of pre-charge control voltage (V_{PC}) is set to V_{DD} , and other signals are initialized to zero. After the recharging process (around 1 ns), the read enable signal (V_{rd_en}), reference control voltage (V_{ref}) and the voltage to WD are activated to V_{DD} , selecting the target memory cell. At this point, V_{PC} transitions to zero, initiating the sensing phase.

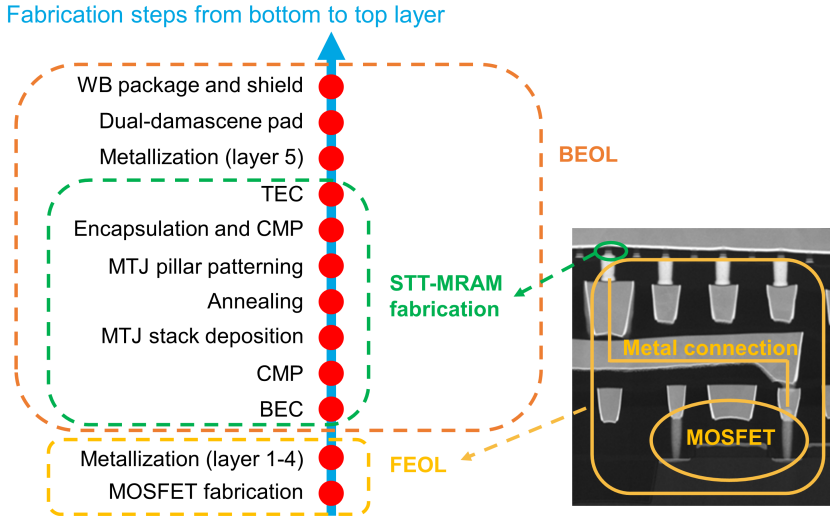
During sensing, the SA measures the current difference between two paths, one with the cell, and the other with the references. The output point of the SA is named as the V_{out} and the $\overline{V_{out}}$; $\overline{V_{out}}$ is the reverse of V_{out} . Fig. 2.6 (b) presents the voltage of V_{out} in 1r1 and 0r0 operations. Initially, both V_{out} and $\overline{V_{out}}$ are pre-charged to V_{DD} . When the discharge process begins, the speed at which each node discharges depends on the current through the corresponding path, determined by the resistance of the MTJ and the reference. Conventionally, the selected reference resistance (i.e., MOSFET channel resistance) is between the channel resistance of MTJ in P and AP state. In the case of a 1r1 operation, the MTJ remains in the AP state, meaning it has a higher resistance and a lower read current ($I_{rd} < I_{ref}$). This causes the MTJ path to discharge more slowly than the reference path. Once the voltage difference $V_{out} - \overline{V_{out}}$ exceeds the threshold voltage (V_{th}) of the MOSFET, the recharging process starts, pulling the V_{out} back toward V_{DD} (i.e., the re-charge process in Fig. 2.6 (b)). Conversely, when reading a logical '0', the MTJ is in the P state, resulting in lower resistance and higher read current ($I_{rd} > I_{ref}$). This makes the reference path discharge more slowly, and the V_{out} continues to discharge to zero, as presented in Fig. 2.6 (b),

To enhance read speed, a relatively I_{rd} is required to accelerate the charge/discharge process at V_{out} and $\overline{V_{out}}$. Yet I_{rd} cannot be too high to avoid unexpected switching of the MTJ. Additionally, the total charge required for this process depends on the capacitance of the MOSFETs in the SA. Larger MOS capacitors store more charge, necessitating a longer charge/discharge cycle for a given current, ultimately reducing read speed. Therefore, optimizing both the SA circuit design and the I_{rd} is crucial to achieving high-performance STT-MRAM read operation.

2.5. STT-MRAM FABRICATION

Fig. 2.7 presents the fabrication steps of STT-MRAMs in [103]. The STT-MRAM manufacturing process consists of two main parts: the Front-End-of-Line (FEOL) process and the Back-End-of-Line (BEOL) process.

- The FEOL stage primarily involves the fabrication of the MOSFETs and the lower metallization layers (e.g., metal layer 1 – 4), which establish the foundation for device operation. This phase includes standard CMOS fabrication steps, ensuring that the underlying transistors and interconnects are properly formed to support the MTJ read/write operations.
- In the BEOL phase, the MTJ devices are integrated into the metal interconnect layers, typically between metal layer 4 and 5 [103], [104]. As shown in Fig. 2.7, the MTJ



TEC: top electrode contact; BEC: bottom electrode contact; CMP: chemical mechanical polishing

Figure 2.7: Example of STT-MRAM fabrication steps [46].

fabrication follows a bottom-up sequence, starting with the deposition of the back electrode contact and the MTJ stack. This is followed by annealing to enhance the magnetic properties of the MTJ layers. Next, the MTJ pillars are patterned, defining individual memory cells. Chemical Mechanical Polishing (CMP) is performed to ensure a smooth and uniform surface, preventing defects in subsequent processing steps. Encapsulation is then applied to protect the MTJ structures from contamination. Finally, the top electrode contact and dual-damascene interconnects are deposited to establish electrical connections between the MTJ and the upper metallization layers, culminating in the final metallization and packaging.

This hierarchical integration of MTJs into the BEOL allows STT-MRAM to be seamlessly integrated with conventional CMOS fabrication processes, ensuring high compatibility. However, precise control of fabrication steps, such as MTJ pillar etching and interfacial engineering, remains critical for achieving high reliability and scalability in advanced process nodes.

2.6. PROSPECTS AND CHALLENGES OF STT-MRAMS

STT-MRAM has emerged as a highly promising memory technology, offering non-volatility, high endurance, and high read/write speed. STT-MRAM holds significant potential for applications in the following areas:

- **Replacement of SRAMs:** STT-MRAM presents a strong potential for replacing SRAM in certain applications, like caches, where power consumption, circuit area, and leakage current are major concerns. Unlike SRAM, which suffers from large

circuit area and significant leakage due to its use of six-MOSFET cells, STT-MRAM provides a more efficient alternative with the 1T-1MTJ cell [105], [106]. The high speed, combined with non-volatility properties in STT-MRAMs, enables it to serve as an effective cache memory, reducing power consumption and circuit area without compromising performance [41].

- **Embedded memories:** STT-MRAM is well-suited for embedded memory applications within a single chip. Its key advantages—non-volatility, low power consumption, and high speed—make it ideal for integrating with logic circuits, such as in microcontrollers and system-on-a-chip designs. By eliminating the need for separate external memory chips, STT-MRAM can reduce system size, simplify board design, and improve overall energy efficiency. This is particularly valuable for portable and Internet of Things devices where battery life and form factor are critical.
- **New application scenario:** STT-MRAM is not only a memory but also a building block for new computing paradigms. Its non-volatile nature and ability to be integrated directly with logic make it an excellent candidate for novel systems like Computing-in-Memory (CIM). In a CIM architecture, computation is performed directly within the memory array, reducing the energy and time costs associated with data movement between the processor and memory. This is particularly beneficial for AI and machine learning applications, which are data-intensive. For example, the resistive nature of MTJs can be used to perform analog matrix-vector multiplication, a fundamental operation in neural networks, directly within the memory array. This approach promises to significantly improve the efficiency of AI hardware.

However, despite its advantages, STT-MRAM faces several challenges [99]:

- **Device performance:** The STT-MRAM shows two critical limitations referring to the single device performance: 1) the low ON/OFF ratio (around 2), and 2) the high write voltage (> 0.8 V) [41].
- **Scalability:** The scalability of MTJs is currently limited by the magnetic mechanism and fabrication process; Especially, 3-D integration like FLASH structures [101] is not applied due to the ferromagnetic material layer properties, which must be physically separated to prevent mutual interference.
- **Yield** Achieving high yield in STT-MRAM fabrication remains challenging due to the vulnerability to defects and the relatively high variability in MTJ properties, such as resistance variation and stochastic switching behavior [66].
- **Fabrication cost:** The complex fabrication process of STT-MRAM, particularly the deposition and patterning of MTJs, requires additional processing steps compared to conventional CMOS-based memories [99]. These additional steps increase manufacturing costs, making STT-MRAM less competitive in cost-sensitive applications.

In conclusion, STT-MRAM holds great promise for various application fields. Continued advancements in material science, device engineering, and circuit-level optimizations will be essential for realizing the full potential of STT-MRAM and enabling its widespread application in the future.

3

MTJ COMPACT MODEL

The MTJ compact model is essential for accurate simulation and effective testing of STT-MRAMs. This chapter begins by providing a comprehensive overview of the MTJ fundamental compact model, which primarily focuses on its key electrical and magnetic characteristics, such as the TMR and switching time. To achieve a more realistic and predictive model, this chapter extends the traditional approach by incorporating the device's inherent intrinsic stochasticity. This crucial addition is a direct response to the random nature of the MTJ switching, which leads to unavoidable write errors and read errors. This chapter further refines the model by analyzing the complex impact of stray magnetic fields on MTJ switching. Based on all analyses above, we propose a simplified and highly efficient model that not only captures these complex physical phenomena but also makes it practical for use in large-scale circuit and system-level simulations, thereby facilitating more accurate defect modeling and test generation for STT-MRAMs.

3.1. OVERVIEW OF EXISTING MTJ MODELING

For the efficient design and optimization, as so for fault modeling, of STT-MRAMs using computer-aided design tools, an accurate yet computationally efficient MTJ model is crucial. Over the years, researchers have developed various MTJ models, which can be broadly classified into three categories: micromagnetic models, macro models, and compact (behavioral) models [108].

1. Micromagnetic model: it provides the highest simulation accuracy by capturing the movement of individual magnetic moments through specialized micromagnetic simulation tools such as OOMMF [109]. These models are valuable for analyzing the detailed physical switching behavior of an MTJ at a fundamental level. However, their computational complexity makes them impractical for large-scale STT-MRAM simulations, where thousands or even millions of MTJs are required to be evaluated simultaneously.

2. Macro model: such as the one presented in [110], it represents MTJ behavior using fundamental circuit components like resistors, capacitors, and voltage sources. Their advantage lies in their direct compatibility with circuit-level simulations. However, as the complexity of MTJ switching dynamics increases, these models require an increasing number of circuit elements, which in turn affects their computational complexity.

3. Compact model: including those proposed in [111], it is typically written in hardware description languages such as Verilog-A. These models offer an effective trade-off between simulation accuracy and computation complexity. By abstracting MTJ behavior into mathematical equations rather than relying on detailed physical representations, compact models achieve faster simulations while remaining sufficiently accurate for practical circuit design. Additionally, their high compatibility with circuit-level simulations makes them suitable for large-scale STT-MRAM design and validation.

Given the trade-offs among these modeling approaches, selecting the appropriate MTJ model depends on the specific requirements of a given design task. For fundamental research and material-level studies, micromagnetic models provide invaluable insight. In contrast, macro and compact models are more suitable for circuit-level and system-level simulations, balancing accuracy and computation complexity. As STT-MRAM technology advances, further refinements in modeling techniques will be essential to enable more accurate and scalable simulations, ultimately facilitating the widespread adoption of MRAM in future memory architectures.

This thesis applies the MTJ compact model, which is built upon previous works in [89], [112], [113] while advancing a further step in explaining the relationship between the STT effect and the impact of the external magnetic field. A well-calibrated compact model is essential for accurately simulating MRAM behaviors at the circuit level, enabling precise evaluation of MTJ resistance and switching.

3.2. REQUIREMENT FOR MTJ MODEL

A well-calibrated compact model is essential for accurately simulating MRAM behaviors at the circuit level, enabling precise evaluation of MTJ resistance and switching [112]. A complete single MTJ model requires including the following five components:

- **R-V Calculation:** This is the foundational component for DC (i.e., static) simulations. The R-V calculation is responsible for accurately capturing the MTJ resis-

tance and its TMR. These properties are fundamental to how the device operates under static conditions, making this part of the model indispensable for understanding the basic behavior of the MTJ.

- **Basic Switching:** This component is crucial for simulating how the MTJ state changes during a write operation (i.e., a voltage applied). It is essential for predicting the success or failure of write operations.
- **Intrinsic Stochasticity at High Voltage** With a high voltage, the STT that causes switching isn't perfectly deterministic. This inherent randomness, or intrinsic stochasticity, can cause a state to fail to switch correctly, leading to write errors. This component determines if this type of flip occurs at a certain high voltage.
- **Intrinsic Stochasticity at Low Voltage:** At low voltages, thermal fluctuations can cause the MTJ state to randomly flip. This type of intrinsic stochasticity (at low voltage) can lead to read errors. This component determines if this type of flip occurs at a certain low voltage.
- **Magnetic Field Impact:** The presence of an external magnetic field can significantly influence the MTJ switching characteristics. This component of the model accounts for the impact of a magnetic field on the STT-induced switching, as the field can either assist or hinder the process. By including this effect, the model can more accurately simulate MTJ behaviors in a magnetic environment, which is crucial for applications where stray fields or field-assisted switching are a factor.

Conventional MTJ compact models excel at representing the first two components [112], but they lack the latter three. In this chapter, we address this limitation by incorporating these three crucial parts into the MTJ compact model.

3.3. PROPOSED MTJ MODEL

Fig. 3.1 presents the structure of the proposed MTJ model, which enables two types of simulations, i.e., the DC (static) simulation and the transient simulation.

The DC (static) simulation, shown in the red box, focuses solely on the device behavior at a fixed voltage. It includes the 'R-V calculation' component, which precisely determines the MTJ resistance and TMR at a given voltage. The data from this static analysis is then used as a foundational input for the more complex transient simulation.

The green box outlines the transient simulation, which models the MTJ dynamic behavior over time to determine if its state will switch. Once a voltage is applied, the 'Basic Switching' and 'Magnetic Field Impact' components work together to provide the fundamental information for a potential state change. The model then evaluates the magnitude of the applied voltage. If the voltage is large, the simulation incorporates the effects of 'Intrinsic Stochasticity at High Voltage' to account for the randomness of STT-induced switching. Conversely, if the voltage is low, the model considers 'Intrinsic Stochasticity at Low Voltage' to capture the random thermal fluctuations that could lead to unexpected state changes. This comprehensive approach ensures that the model can accurately determine whether the MTJ state eventually switches or remains unchanged.

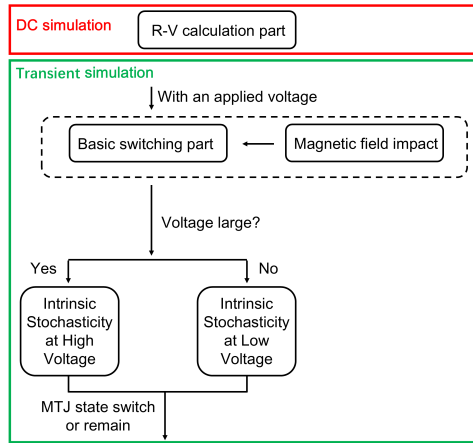


Figure 3.1: Structure of proposed MTJ model.

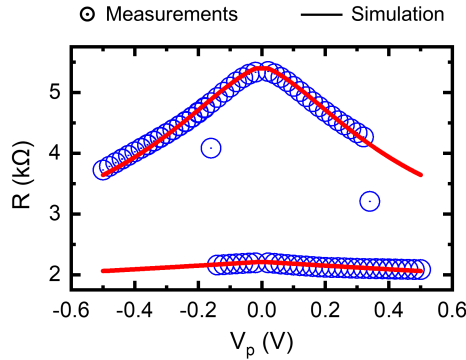


Figure 3.2: R-V measurement and simulation.

Before presenting the details of the MTJ model, the symbols used in the equations of this section are described in the Table 3.1.

3.4. R-V MODEL

The resistance-voltage (R-V) characteristics of MTJs are fundamental to the electrical behavior of STT-MRAM devices. The resistance of an MTJ varies with the V_p ; As the absolute value of V_p increases, the MTJ resistance decreases. As demonstrated by the R-V hysteresis loop measured for an MTJ with a critical dimension (CD) of 60 nm (see Fig. 3.2), the measurement is performed by V_p sweeping from -0.5 V to 0.5 V in step of 0.02 V, and the corresponding MTJ resistance is recorded at each step for both the P and AP states.

According to the physical model presented in, the resistance of an MTJ is primarily influenced by the thickness of the MgO barrier and interfacial effects between the barrier

Table 3.1: Symbols in MTJ model.

R_0	TB thickness	s	Fitting coefficient
V_p	Biased voltage	t_{ox}	TB thickness
F	Fitting coefficient	$\bar{\phi}$	Potential barrier height of TB
TMR(0)	TMR at 0 V	V_h	Bias voltage when TMR(V_h) = 0.5 · TMR(0)
b	Fitting coefficient	t_0	Characteristic relaxation time
C	Euler constant	Δ	Thermal stability
e	unit charge	m	FL magnetization
μ_B	Bohr magneton	η	STT efficiency
I_w	Write current	I_{rd}	Read current
V_p	Pulse height	t_p	Pulse width
I_c	Critical current	α	Damping factor
\hbar	Reduced Planck constant	t_{FL}	Thickness the FL
A	MTJ cross area	M_s	Saturation magnetization
H_k	Anisotropy magnetic field	H_{ext}	External magnetic field H_{ext}
H_s	Stray field	k_B	Boltzmann constant
T	Temperature	μ_0	Vacuum permeability
γ	Gyromagnetic ratio	τ_0	Inverse of the attempt frequency
t_w	Average switching time	m_{FL}	FL magnetization
θ	Angle between m_{FL} and easy-axis	θ_0	Initial angle between m_{FL} and easy-axis
T	Total input torque	E_B	Energy barrier between P and AP state
coef	Fitting coefficient		

and adjacent CoFeB [114]. The relationship of the MTJ resistance in P state (R_p) and the biased voltage can be described as follows [114]:

$$R_p(V_p) = \frac{R_0}{1 + s \cdot |V_p|} \quad (3.1)$$

$$R_0 = \frac{t_{ox}}{F \cdot \sqrt{\bar{\phi}} \cdot A} \exp\left(\text{coef} \cdot t_{ox} \cdot \sqrt{\bar{\phi}}\right)$$

One of the key observations in MTJ behavior is the decrease in TMR as the bias voltage increases. This relationship is captured by the following equation from [115], which models the voltage-dependent decline of TMR:

$$TMR(V_p) = \frac{TMR(0)}{1 + \frac{V_p^2}{V_h^2} + b \cdot V_p^{\frac{4}{3}}} \quad (3.2)$$

To improve accuracy, a fitting term ‘b’ is incorporated into the denominator, yielding a better fit result compared to the original equation in [115]. The R_{AP} at a given bias voltage can be derived through the two equations above and Eq. 2.1. This R-V model provides an accurate representation of MTJ resistance variations under different V_p , which is essential for developing precise circuit simulations and optimizing the performance of STT-MRAM devices. Fig. 3.2 presents the model fitting result of the R-V measurement.

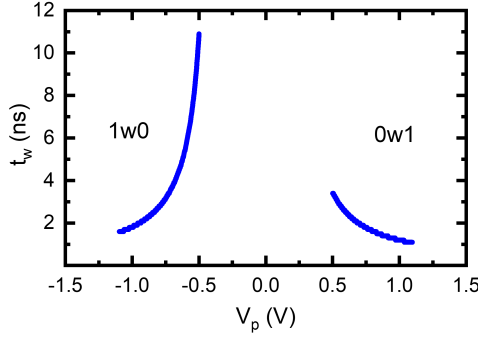


Figure 3.3: Average switching time.

3.5. MTJ SWITCHING MODEL

The MTJ switching mode can be classified into three regimes according to the average switching time (t_w): the precessional regime for $t_w < 10$ ns, the dynamic regime for $10 \text{ ns} < t_w < 100$ ns, and the thermal activation regime for $t_w > 100$ ns [41]. Since SPICE simulations are time-resolved (transient), the determination of the MTJ switching regime relies on the average switching time or critical current [95]:

$$I_c = \frac{1}{\eta} \cdot \frac{\alpha \cdot e}{\hbar} \cdot t_{FL} \cdot A \cdot M_s \cdot H_k \quad (b) \quad (3.3)$$

The critical current serves as a key parameter in estimating the average switching time (t_w). When the applied write pulse width exceeds t_w , the MTJ successfully switches states. For example, in a 1w0 operation, the MTJ is initialized in ‘1’. A write current is flowing through the MTJ, trying to switch the MTJ state to ‘0’. If the write pulse width exceeds the t_w , which relies on I_c , the MTJ is successfully switched to ‘0’. Notice that the calculation of t_w varies in different switching regimes, and each switching regime is governed by distinct physical principles, influencing how the magnetization evolves over time.

In the precessional regime, the switching occurs rapidly due to coherent magnetization precession induced by STT [41]. When the applied current exceeds I_c to a certain extent (in our model, $> 1.1 \cdot I_c$), the FL magnetization undergoes a motion and then stabilizes in the opposite state. This regime offers the fastest switching speed but is highly sensitive to current variations and thermal noise [95]. The calculation of t_w is presented as follows [112]:

$$t_w = \frac{\left(C + \ln\left(\frac{\pi^2}{4} \cdot \Delta\right) \right) \cdot e \cdot m}{4 \cdot \mu_B \cdot \eta \cdot (I - I_c)} \quad (3.4)$$

In the dynamic regime, the switching process is governed by a combination of precessional motion and damping effects [116]. The MTJ switching is in the dynamic regime when the applied current is around I_c . Unlike the precessional regime, where switching is ultra-fast, the dynamic regime exhibits a slower speed and larger variability due to

the combined effect of STT and thermal energy. It is difficult to simply calculate the t_w following physics, and an empirical representation is applied, where the specific I_{c1} is introduced that the MTJ switches in the dynamic regime when $I_{c1} < I < 1.1 \cdot I_c$ [113]:

$$t_w = \delta \cdot \left[\frac{I - I_{c1}}{I_c} + \frac{1}{\Delta} \right]^{-1} \quad (3.5)$$

where:

$$\begin{aligned} I_{c1} &= I_c \left[1 - \left(\frac{1}{\Delta} \right) \ln \left(\left(\frac{\delta}{\tau_0} \right) \cdot \Delta \right) \right] \\ \theta_0 &= \sqrt{\frac{1}{2\Delta}} \\ \delta &= \frac{\ln \left(\frac{\pi}{2\theta_0} \right)}{\alpha \cdot \gamma \cdot H_k} \end{aligned} \quad (3.6)$$

In the thermal activation regime, the switching is primarily driven by thermal fluctuation assisted transitions [117]. In the compact model, when the applied current is much lower than I_c ($I < I_{c1}$), the MTJ switches in this regime. The calculation of t_w in the thermal activation regime is presented as follows [112], [117]–[119]:

$$t_w = \tau_0 \cdot \exp \left(\Delta \cdot \left(1 - \frac{I}{I_c} \right) \right) \quad (3.7)$$

Fig. 3.3 presents the calculated t_w , which will be validated by the WER measurement in the next subsection.

3.6. WRITE ERROR RATE

The WER in STT-MRAMs refers to the probability of a write operation failing to switch the MTJ into the intended resistance state. Ensuring a low WER is essential for achieving high-quality commercial STT-MRAMs. In the MTJ compact model, WER is primarily considered in the precessional regime under high write current, while unintended switching at low current mainly contributes to the RER.

3.6.1. INTRINSIC STOCHASTICITY MECHANISM IN MTJS

The WER arises from the intrinsic stochasticity in MTJs. Fig. 3.4 presents a simplified physical model of the MTJ switching [41]. The switching process is viewed as the m_{FL} rotation. Initially, the angle θ between m_{FL} and easy-axis is around 0° . In write operations, m_{FL} rotates under the spin transfer torque effect, and θ increases. θ will be aligned to the easy-axis with $\theta \approx 180^\circ$ in successful switching; the MTJ switching time refers to the time that θ alters from $\approx 0^\circ$ to $\approx 180^\circ$. On the other hand, θ will be aligned back to the easy-axis with $\theta \approx 0^\circ$ in case switching fails.

The stochasticity in the switching is attributed to the thermal fluctuation, which affects both the initial MTJ state and the switching process [41]. In the temperature of absolute zero, θ_0 keeps exactly 0° , and the MTJ can never be switched by STT [99]. In room temperature, the initial θ (i.e., θ_0) follows the Maxwell–Boltzmann distribution

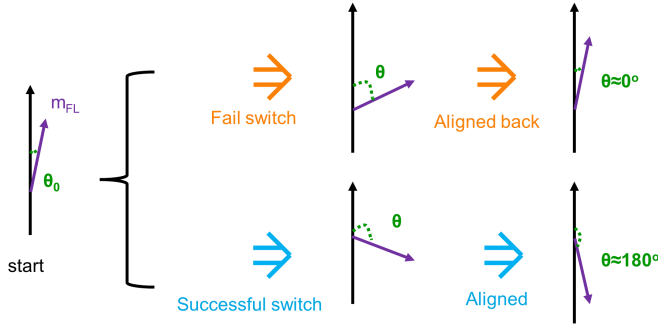


Figure 3.4: Simplified physical model of the MTJ switching.

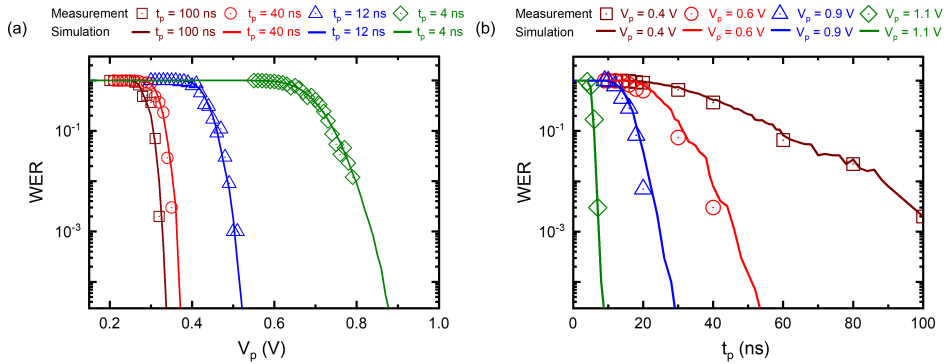


Figure 3.5: WER measurement and simulation.

[99], and the stochasticity of the distribution introduces the variation of MTJ switching time. Besides, the MTJ switching depends on both θ at the write pulse ending and the thermal fluctuation. For example, if write pulses end with $\theta=90^\circ$, the switching probability $P_{sw}=50\%$; if $\theta\approx 180^\circ$ at the write pulse ending, $P_{sw}\approx 100\%$, yet with a remaining small probability of switching failure. This switching stochasticity is an inherent property that cannot be avoided; even in an ideal situation (e.g., temperature is kept constant at 22°C), the stochasticity still exists. This stochasticity causes cycle-to-cycle random write errors [41].

3.6.2. WRITE ERROR RATE MEASUREMENT AND FITTING

The ‘intrinsic switching stochasticity’ can impact the MTJ electrical performance; it leads to *unavoidable random write errors*. Fig. 3.5 (a) and 3.5 (b) present WER for $1w0$ (i.e., apply write 0 to a cell initialized to 1) by varying the pulse height V_p (i.e., the voltage between to FL and HL of MTJs) and pulse width t_p . As the figure shows, the WER can never be reduced to 0 regardless of the value V_p or t_p . Hence, additional efforts are needed to make the STT-MRAM design fault tolerant; e.g., Error Correcting Code (ECC) is necessary for high-quality MRAM systems [120].

The calculation of WER and write pulses applied in the MTJ compact model is presented as follows [41]:

$$WER(t) = 1 - \exp\left(-\frac{\Delta \cdot \pi^2 \cdot j}{4 \cdot i \cdot (\exp(2(t/t_0) \cdot j) - 1)}\right) \quad (3.8)$$

where $i = I/I_c$, and $j = i - 1$

During the transient SPICE simulations, the probability that a write error occurs at a certain time point 't' is presented as this $WER(t)$. The MTJ compact model fitted with the WER measurement is presented in Fig. 3.5

3.7. READ ERROR RATE AND MODIFICATION

The RER in STT-MRAMs refers to the probability of the unintended MTJ switching during read operations [41]. One of the potential applications of STT-MRAMs is as a replacement for SRAMs, which demands high-frequency read operations performed with short pulses [121]. In such scenarios, achieving a low RER is critical. This switching arises from the thermal fluctuations of MTJs; hence, the RER is only considered in the thermal activation regime in the MTJ compact model.

In this subsection, we present the conventional RER evaluation method, indicating the limitations while applying this method to circuit-level simulations (i.e., the irregular read pulse provided by SA). Then, an alternative RER prediction model is proposed, addressing the limitations of the traditional one, and serves as a generalized solution that accommodates all types of read pulse shapes. We integrate this modeling method into the MTJ compact model using Verilog-A, ensuring compatibility with circuit-level simulations.

3.7.1. CONVENTIONAL METHOD

A common spec of RER is $RER_{spec} < 10^{-9}$ for STT-MRAMs with ECC, and $RER_{spec} < 10^{-18}$ for STT-MRAMs without ECC [41], [122]. However, accurately extracting RER at this low level is time-consuming. For example, assuming $RER = 10^{-15}$ and the read operation period is 10 ns, performing 10^{16} times of read operations to extract RER require at least $10 \text{ ns} \cdot 10^{16} \approx 1157$ days. Due to this impracticality, most studies rely on analytical prediction models.

One of the most popular analytical approaches to predict the RER for STT-MRAM design is given in Eq. 3.9 [41], [123]; it is mainly driven by the physics.

$$RER = 1 - \exp\left(\frac{-t_p}{\tau_0 \cdot \exp\left(\Delta \cdot \left(1 - \frac{I_{rd}}{I_c}\right)\right)}\right) \quad (3.9)$$

3.7.2. LIMITATION OF THE CONVENTIONAL METHOD

Fig. 3.6 (a) and (b) present the I_{rd} changing with time during 0r0 and 1r1 operations. The irregularity in the I_{rd} curves arises from various factors, the primary one being the

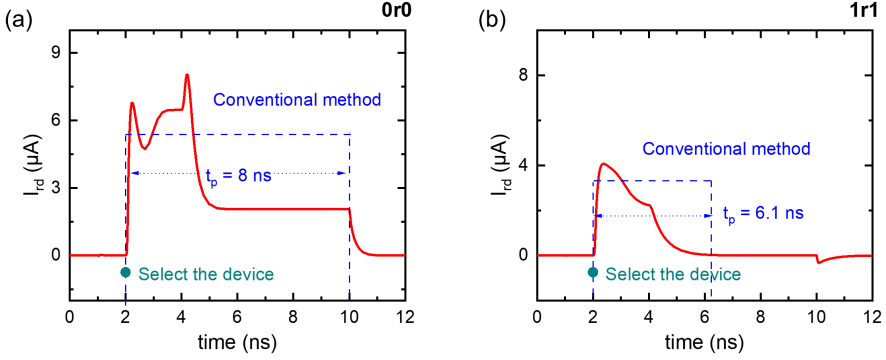


Figure 3.6: I_{rd} in 0r0 and 1r1 operations.

non-uniform discharging/charging process of the SA nodes. Additionally, parasitic capacitors and resistances also contribute to this phenomenon. Despite the variety of circuit designs and read operation strategies, the I_{rd} curves inherently exhibit an irregular form. For example, another common type of SA is the voltage-based SA [124], which also relies on the discharging/charging of MOSCAPs, hence also resulting in irregular I_{rd} curves. To obtain the accurate I_{rd} curves for both SA types, circuit-level simulations are required.

Since the I_{rd} varies over time, it is difficult to define a constant I_{rd} value for Eq. 3.9. A straightforward solution is to set a rectangle based on the simulated I_{rd} curve and define t_p & I_{rd} by this rectangle [123], shown as the blue square in Fig. 3.6. Nevertheless, the accuracy of this approach is affected by two factors: 1) According to Eq. 3.9, the RER is highly sensitive to I_{rd} , even minor fluctuations in I_{rd} significantly affect the final RER results. 2) There is no standard approach to define the t_p & I_{rd} . To optimize the STT-MRAM design, an accurate RER prediction model is needed.

3.7.3. MODIFICATION OF THE READ ERROR RATE EVALUATION

The physical analysis of the probability of MTJ flipping due to a small current (i.e. RER) is presented as [125]:

$$\text{RER} = \int_0^{t_p} (1 - \langle m_z \rangle) \cdot \tau_{AP \rightarrow P}^{-1} + \langle m_z \rangle \cdot \tau_{P \rightarrow AP}^{-1} dt$$

where:

$$\begin{aligned} \tau_{P \rightarrow AP} &= \tau_0 \cdot \exp \left(\Delta \cdot \left(1 - \frac{I_{rd}(t)}{I_c} \right) \right) \\ \tau_{AP \rightarrow P} &= \tau_0 \cdot \exp \left(\Delta \cdot \left(1 + \frac{I_{rd}(t)}{I_c} \right) \right) \end{aligned} \quad (3.10)$$

Here, $\langle m_z \rangle$ represents the initial MTJ state, which takes only two binary values; $\langle m_z \rangle = 1$ refers to the P state, and $\langle m_z \rangle = 0$ refers to the AP state. I_{rd} can be either

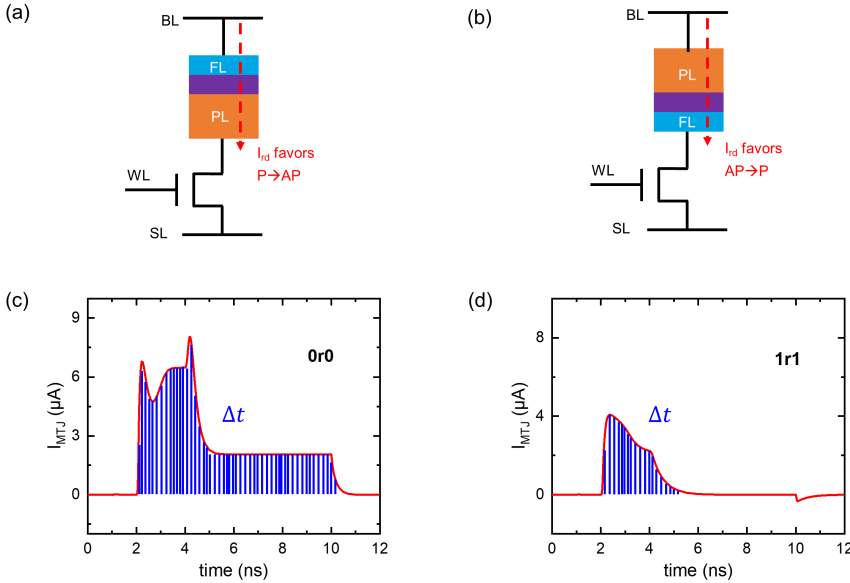


Figure 3.7: Example of numerical integration for (a) 0r0 and (b) 1r1 operation.

negative or positive, depending on the current direction. In this work, we define the positive I_{rd} favoring P to AP switching, and the negative I_{rd} favoring AP to P switching.

The RER is largely affected by the direction of the read current, which either flows through the MTJ from FL to HL or in the opposite direction. This direction is defined by the MTJ cell design, as illustrated in Fig. 3.7 (a) and (b). Notice that although the design in Fig. 3.7 (a) is the most commonly used [126], we consider both designs in this work.

In the design of Fig. 3.7 (a), the FL connects to the BL, the HL connects to the SL, and the read current flows from FL to HL, which favors the AP to P switch. Here, $I_{rd} < 0$, $\tau_{AP \rightarrow P}^{-1} \gg \tau_{P \rightarrow AP}^{-1}$, and the probability of AP to P switching is much larger than that of P to AP switching. On the other hand, in the design of Fig. 3.7 (b) the FL connects to the HL and the BL connects to the SL, suggesting $I_{rd} > 0$ and favoring P to AP switching. Here, $\tau_{P \rightarrow AP}^{-1} \gg \tau_{AP \rightarrow P}^{-1}$, and the probability of P to AP switching is much larger than that of AP to P switching. In this work, we only focus on the worst case. When investigating RER in 1r1 operation, the design of Fig. 3.7 (a) is applied, and when investigating RER in 0r0 operation, the design of Fig. 3.7 (b) is applied.

If a standard rectangular pulse is assumed to be applied to the MTJ (i.e., I_{rd} remains constant), then an analytical solution can be derived from Eq. 3.10, which is equal to Eq. 3.9. However, since the actual I_{rd} curves are irregular (see Fig. 3.6), the analytic solution is unattainable for Eq. 3.10 in real scenarios. To solve this challenge, we carry out a numerical integration process of the extracted curve, presented as:

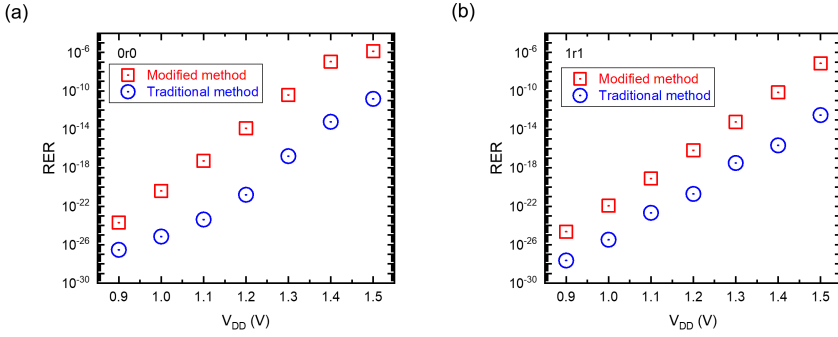


Figure 3.8: RER prediction by conventional and proposed methods for (a) 0r0 operations and (b) 1r1 operations.

$$RDR = \sum_0^T \frac{d\langle m_z \rangle}{dt} \cdot \Delta t$$

where:

$$\frac{d\langle m_z \rangle}{dt} = (\langle m_z \rangle \cdot (\tau_{P \rightarrow AP}^{-1}(t) + \tau_{P \rightarrow AP}^{-1}(t + \Delta t)) + (1 - \langle m_z \rangle) \cdot (\tau_{AP \rightarrow P}^{-1}(t) + \tau_{AP \rightarrow P}^{-1}(t + \Delta t))) / 2 \quad (3.11)$$

The $\tau_{P \rightarrow AP}^{-1}$ (or $\tau_{AP \rightarrow P}^{-1}$) as a function of t is given in Eq. 3.10. Here, Δt represents the time step used in the numerical integration process. Fig. 3.7 illustrates the example of discretizing the I_{rd} curve for numerical integration. At each Δt , the I_{rd} is applied to Eq. 3.11 to calculate $\frac{d\langle m_z \rangle}{dt}$. A smaller Δt improves the accuracy of the RER prediction. Although the I_{rd} curve varies considerably in different circuit designs and read operation strategies, Eq. 3.11 can be applied to calculate the RER as long as the curve can be extracted.

The aim is to integrate Eq. 3.11 into circuit-level simulations. Using Spice simulations, Δt is automatically defined, which varies for each time step. When I_{rd} changes faster over time, then Δt becomes smaller; when I_{rd} changes slowly over time, then Δt becomes larger. In addition, we define the largest value Δt as 0.1 ns.

The equation of $\frac{d\langle m_z \rangle}{dt}$ (see Eq. 3.11) is implanted into the MTJ compact model [112] by Verilog-A. At each time step, $\frac{d\langle m_z \rangle}{dt}$ is calculated and accumulated. The final accumulated value at the end of the read operation is RER. We also apply Python to extract the I_{rd} curves from the Spice simulations and calculate the RER, which shows the same values as those given by the Spice simulations.

Varying the supply voltage V_{DD} and the RER using both traditional and proposed modeling methods is extracted, as presented in Fig. 3.8. The results reveal a high discrepancy exceeding 10^2 between the RER values of the two modeling methods for both 0r0 and 1r1 operations. Given the limitations of the traditional method, the actual discrepancy varies across different circuit designs, the read strategies, the setting of I_{MTJ}

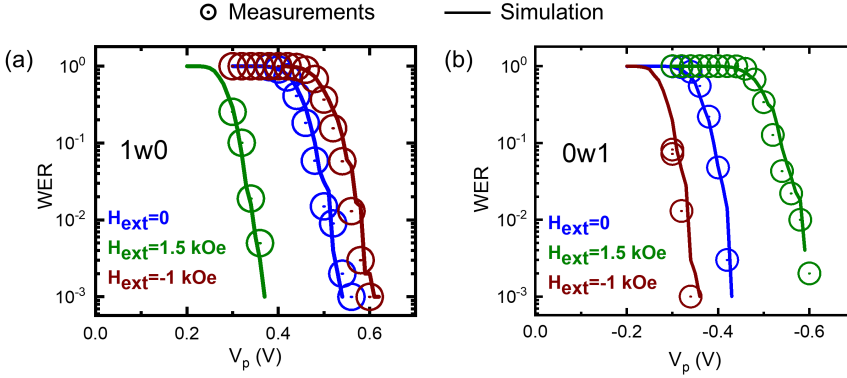


Figure 3.9: H_{ext} impact on WER in (a) $1w0$ and (b) $0w1$.

and t_p ; in this study, the RER predicted by the traditional method is found to be underestimated. Furthermore, both Eq. 3.11 and Eq. 3.9 demonstrate that the RER is highly sensitive to I_{MTJ} . Consequently, V_{DD} has a significant impact on RER. The simulation results show that as V_{DD} varies from 0.9 V to 1.5 V, the RER increases by 10^{20} . Obviously, a higher V_{DD} reduces RER, but it increases energy consumption.

The accuracy of the RER prediction is critical to the STT-MRAM design. If the RER is underestimated, commercial STT-MRAM products experience a higher-than-expected rate of read error events, leading to reduced product quality. On the other hand, when the RER is overestimated, more circuit area (i.e., more bits) for ECC is configured, resulting in unnecessary costs.

3.8. MAGNETIC FIELD IMPACT

3.8.1. REGULAR IMPLEMENTATION OF MAGNETIC FIELD IMPACT

In 2020, Wu *et al.* incorporated the magnetic field impact into the MTJ model; they adjusted the calculation of I_c with the incorporation of the magnetic field parameters [89], and kept the calculation of t_w and WER unchanged. The adjusted I_c calculation is presented as follows:

$$I_c = \frac{\alpha \cdot e \cdot M_s \cdot A \cdot t_{FL}}{\hbar \cdot \eta} \cdot (H_k - H_{ext}) \quad (3.12)$$

The t_w is derived by applying the I_c calculation shown above into Eq. 3.4. The H_{ext} includes all potential external magnetic field, like the magnetic field provided by Automatic Testing Equipment (ATE) and H_s [127], [128]). Notice that whether H_{ext} impedes or facilitates the MTJ switching depends on its direction and the switching type (either $1w0$ or $0w1$). In this thesis, we assume that the positive magnetic field favors AP to P switching, and the negative magnetic field favors P to AP switching. Hence, H_{ext} has opposite impacts on t_w and I_c for P to AP versus P to AP switching. For example, for AP to P switching, the higher H_{ext} , the smaller I_c and t_w .

Fig. 3.9 presents the simulated WER in the presence of H_{ext} from -1 kOe to 1.5 kOe.

The model is calibrated with the measurement data, which is also presented in the figure.

3.8.2. 'EQUIVALENT CURRENT' REPRESENTATION EXTERNAL MAGNETIC FIELD IMPACT

Although the MTJ compact model successfully incorporates the stray field effect, it is inconvenient to be directly applied to Spice simulations when the external magnetic field fluctuates (i.e., changes time by time); This model is based on the calculation of the magnetic field, which cannot be flexibly adjusted during the circuit-level simulation process. For example, when the magnetic field changes in the environment (i.e., H_{ext} changed), yet in Spice, changing parameters in the model requires stopping simulations and changing the model code. In order to accurately model the MTJ performance with the H_{ext} changing during the simulation process, the H_{ext} impact needs to be incorporated at the circuit level, rather than the model level.

Here, we propose a method to extract the H_{ext} impact out of the MTJ model and introduce an equivalent circuit representation that dynamically updates time by time. A specific 'equivalent (virtual) current source' model is designed through Verilog-A to represent the H_{ext} impact: the input is referred to the magnetic field impact, and the output is the current (I_{ext}) applied to an 'equivalent input' of the MTJ model. By applying this virtual current source, the key electrical parameters I_c and t_w of the MTJ model are adjusted to be:

$$\begin{aligned} t_w &= \frac{\left(C + \ln\left(\frac{\pi^2}{4} \Delta\right)\right) \cdot e \cdot m}{4 \cdot \mu_B \cdot \eta \cdot (\sum(I_{ext}) + I - I_c)} \\ I_c &= \frac{\epsilon}{\eta} \cdot (-H_{ext}) \end{aligned} \quad (3.13)$$

Here, the parameter of switchable stray field (H_{ext}) is avoided in the model, and a new related parameter (I_{ext}) is introduced. I_{ext} refers to the magnetic field impact, which can be flexibly controlled by Verilog-A code. For example, I_{ext} increases time by time to dynamically represent the magnetic field rising over time. In this way, the MTJ model is set in the whole simulation process, while the influence of the time-varying H_{ext} is dynamically introduced at the circuit level.

3.8.3. RELATIONSHIP BETWEEN STT EFFECT AND EXTERNAL MAGNETIC FIELD IMPACT

This subsection analyzes the relationship between STT and the magnetic field from an alternative perspective. Notice that the alternative perspective does not change the calculation of the original MTJ model shown above. First, we the energy barrier E_B between P and AP states can be presented as [129]:

$$E_B = \frac{\left(C + \ln\left(\frac{\pi^2}{4} \cdot \Delta\right)\right) \cdot e \cdot m}{4 \cdot \mu_B} \quad (3.14)$$

Secondly, the total STT torque caused by *electron flowing* and *magnetic field* during write operations can be presented as [41], [126], [129]:

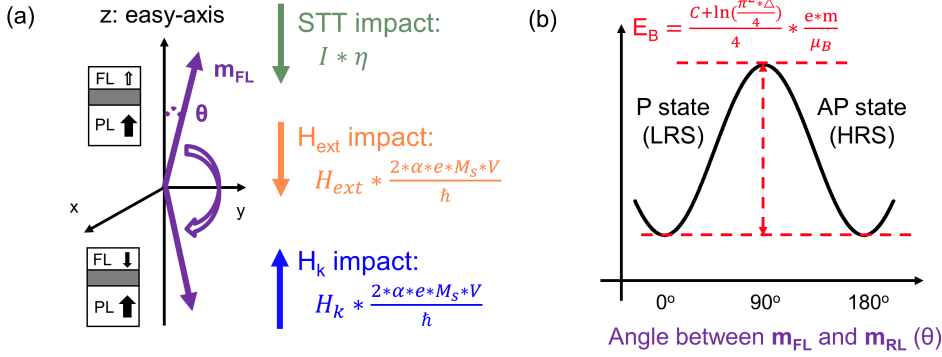


Figure 3.10: H_{ext} impact on WER in (a) $1w0$ and (b) $0w1$.

$$T = t_p \cdot \left(I \cdot \eta + (H_{ext} - H_k) \cdot \frac{\alpha \cdot e \cdot M_S \cdot A \cdot t_{FL}}{\hbar} \right) \quad (3.15)$$

Enough torque must be input into the MTJ to overcome the energy barrier and achieve successful switching (i.e., $T > E_B$). Three components are involved in Eq. 3.15:

- **STT effect:** $I \cdot \eta$ [129]. The STT is provided by the current (i.e. the electrons) through MTJs.
- **Magnetic field impact:** $(H_{ext}) \cdot \frac{\alpha \cdot e \cdot M_S \cdot A \cdot t_{FL}}{\hbar}$. Here, H_{ext} refers to the external magnetic fields from all possible sources. H_{ext} can be either positive or negative. Notice that $H_{ext} = 0$ when MTJs work in an ideal environment. H_s is the stray field described in the previous section.
- **Anisotropy magnetic field impact:** $H_k \cdot \frac{\alpha \cdot e \cdot M_S \cdot A \cdot t_{FL}}{\hbar}$. H_k is a constant material parameter aligned with the easy axis, maintaining the same direction as the FL magnetization. It switches along with the MTJ, consistently opposing the switching process. [129].

Fig. 3.10 (a) presents an example of the MTJ switching process. The STT effect and the magnetic field either accelerate or impede the switching as shown in Eq. 3.15. When the total impact of STT and magnetic field exceeds that of H_k , the MTJ switches in the precessional regime [41]. Once the input torque T overcomes E_B , the FL magnetization (m_{FL}) switches (see Fig. 3.10 (b)); the angle θ between m_{FL} and easy-axis alters from closing 0° to closing 180° [129]. Notice that after the switching, the H_k is reversed as well, which is not shown in the figure, impeding MTJ AP to P switching.

In the original model, the impact of the magnetic field is integrated into the calculation of I_c . Although this integration may simplify model design, it complicates assessing the magnetic field impact on MTJ switching. In this alternative perspective, however, the impact of the STT (i.e., current) and the magnetic field on the MTJ switching is disentangled. Consequently, we can analyze the contributions of STT and the magnetic

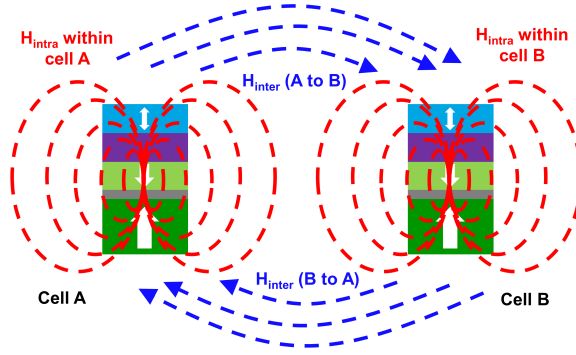


Figure 3.11: Stray field in MTJs.

field to the torque T independently, and facilitate an in-depth understanding of the MTJ switching behavior.

3.8.4. STRAY FIELD IMPACT

INTRODUCTION OF STRAY FIELD

Even when the STT-MRAM works in an ideal environment without any external magnetic field, extra magnetic field introduced by the MTJ ferromagnetic layers, named the ‘stray field’, still exists [73]. As shown in Fig. 3.11, the stray field consists of the *intra-cell* stray field originating from the PL of the cell itself, and the *inter-cell* stray field generated by the FL and the PL of neighboring cells [89]. Note that *magnetic coupling* between cells is attributed to the inter-cell stray field [73], [89]. Since PL is stable and never switches, the PL-introduced intra-cell stray field never changes, thus its contribution to the magnetic coupling is constant. However, the FL-introduced inter-cell stray field is *cell-state dependent*; when the MTJ state is switched, the stray field is also reversed (i.e., switched).

MAGNETIC COUPLING IMPLEMENTATION OF THE STT-MRAM ARRAY

Here, we apply the ‘equivalent current’ to represent the switchable stray field, i.e., FL-introduced inter-cell stray field. The total H_s exercised on a cell in the STT-MRAM array can be calculated by simply adding up the Intra-Cell Stray Field ($H_{s-intra}$) and the sum of all PL-/FL-introduced inter-cell stray field. $H_{s-inter}$ introduced by all neighboring cells consists of $H_{s-inter-PL}$ and $H_{s-inter-FL}$. H_s is presented as:

$$H_s = H_{s-intra} + \sum (H_{s-inter-PL} \pm H_{s-inter-FL}) \quad (3.16)$$

Note that, $H_{s-intra}$ and $H_{s-inter-PL}$ are constant, while $H_{s-inter-FL}$ switches with the neighboring cell state switching. Here, we apply the modeling in Chapter 3 to flexibly represent the switchable property of $H_{s-inter-FL}$, as presented:

$$I_{s-FL} \cdot \eta = H_{s-inter-FL} \cdot \epsilon \quad (3.17)$$

The ‘equivalent current source’ I_{s-FL} is designed through Verilog-A to represent the $H_{s-inter-FL}$ impact: the input is the state of one cell (e.g., the MTJ resistance), and the

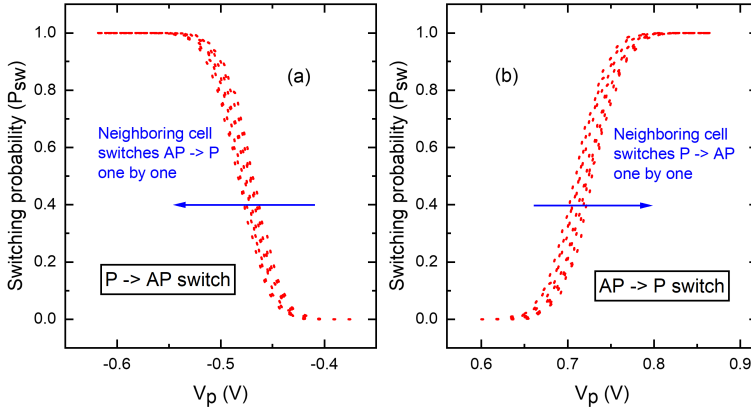


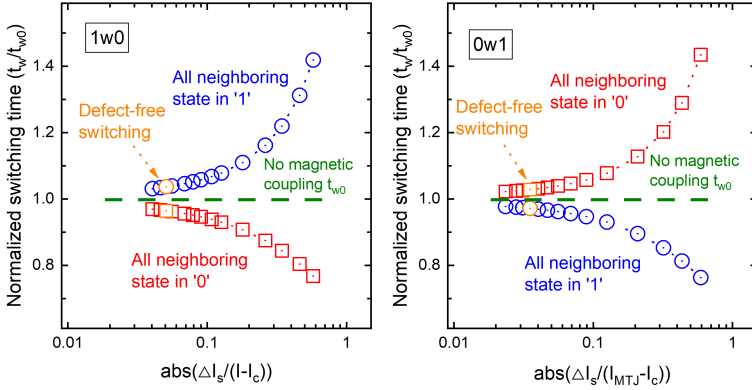
Figure 3.13: MTJ behavior dependency on magnetic coupling effects.

evaluate the magnetic coupling effect through circuit simulations.

To completely include the magnetic coupling, we perform all the circuit-level simulations based on the 4×4 1T-1MTJ STT-MRAM array with the addition of ‘equivalent current sources’ between neighboring cells, as presented in Fig. 3.12. Here, the purple dashed lines are the ‘equivalent (virtual) current sources’, which are determined only by the state of the devices (as given in Eq. 3.17), and modeled with Verilog-A. As a result, each MTJ controls eight current sources to all its neighboring cells to model the inter-cell stray field impact.

Fig. 3.13 presents the magnetic coupling impact on the MTJ writing performance. We extract the MTJ Switching Probability P_{sw} by varying the write pulse height V_p with constant pulse width t_p of 10 ns. Fig. 3.13 (a) shows that when the eight neighboring cell states are switched from ‘1’ to ‘0’ one by one, the $P_{sw} - V_p$ line moves horizontally to the left and the MTJ requires a higher V_p for P→AP switching because the neighboring cell states in ‘0’ favoring the $w0$ operation. A complementary phenomenon for AP→P switching is shown in Fig. 3.13 (b).

The model design in this section enables us to qualitatively analyze the magnetic coupling effect on STT-MRAM write behaviors. Note that the magnetic coupling always plays a role in the MTJ switching performance, regardless of the defects; it should have a negligible impact on the defect-free MTJ switching. For the reliable and robust STT-MRAM design, $\sum(I_{s-FL}) \ll (I - I_c)$ [84]. However, in the presence of defects, the write current I may be reduced and come close to I_c , resulting in a stronger impact of I_{s-FL} (see Eq. 3.18) on t_w which could cause a potential write fault. Here, we consider the two extreme situations: a) induced current $I_{s-FL(P)}$ when all the neighboring cells are in P state, and b) induced current $I_{s-FL(AP)}$ when all the neighboring cells are in AP state. Notice that $I_{s-FL(P)}$ and $I_{s-FL(AP)}$ are opposite due to the reversed $H_{s-inter-FL}$. Here we define the max I_{s-FL} gap between the two extreme cases as ΔI_s :

Figure 3.14: Magnetic coupling effect on ΔI .

$$\begin{aligned}
 \Delta I_s &= \sum_{i=1}^8 |I_{s-FL}(P) - I_{s-FL}(AP)| \\
 &= 16 \cdot \frac{\epsilon}{\eta} \cdot |H_{s-inter-FL}|
 \end{aligned} \tag{3.19}$$

Fig. 3.14 shows the dependency of the magnetic coupling effect on $\Delta I_s/(I - I_c)$, where ‘8’ is the number of neighboring cells. As $\Delta I_s/(I - I_c)$ increases due to a defect, the normalized switching time (t_w/t_{w0}) gap between two extreme cases becomes larger, suggesting a defect may amplify the magnetic coupling effects on the STT-MRAM writing operation. It is worth noting that magnetic coupling has a negligible impact on read operations; applying a small read voltage V_{rd} in the presence of the stray field will never cause the MTJ state to switch, as $(I + I_{s-FL}) \ll I_c$.

4

STATE OF THE ART IN STT-MRAM TESTING

A high-quality test solution with low escape rates and high time efficiency is essential for STT-MRAM manufacturing. The testing methodology for STT-MRAMs basically follows the approach used for traditional memories, such as SRAMs, consisting of three steps: 1) defect modeling, 2) fault modeling, and 3) test generation. However, due to the unique mechanisms and structure of MTJs, additional efforts are required to ensure comprehensive test coverage. Fault modeling for conventional defects must consider the physical impact of MTJs, like magnetic coupling. Besides, unique defects such as pinhole, Synthetic Anti-Ferromagnet Flip (SAFF), and Intermediate State (IM) cannot be accurately modeled using linear resistors. To address this, the DAT is applied by designing specialized compact models for defective MTJs. This chapter provides a state-of-the-art review of STT-MRAM testing, covering the test methodologies for both conventional and unique defects.

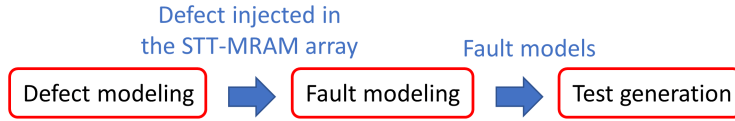


Figure 4.1: Process of conventional test development.

4.1. FRAMEWORK OF TEST DESIGN OF STT-MRAMS

Testing is a critical step in ensuring the quality of commercial STT-MRAM manufacturing. As STT-MRAM moves toward large-scale production, robust test strategies are required to detect defects and guarantee high performance under real-world conditions. Given the complexity of its fabrication process and the unique working mechanisms associated with MTJs, a well-structured framework is essential to develop high-quality test solutions.

A high-quality test strategy must meet several requirements. First, it should achieve high defect coverage, meaning it must be capable of detecting a variety of defect types, including both conventional and unique defects. Inadequate coverage can lower overall product quality and lead to reliability issues. Second, test time must be minimized to ensure cost efficiency in testing for large-scale manufacturing. Lengthy test procedures increase production costs and may reduce yield; hence, it is crucial to balance between test coverage and time efficiency. Additionally, an effective test must minimize yield loss, which occurs when defect-free chips are incorrectly diagnosed as faulty and discarded. High yield loss can lead to unnecessary waste and increased production costs. Lastly, test methodologies should be adaptable to different memory STT-MRAM architectures, ensuring they remain effective across different manufacturing conditions.

This section presents a framework for test design for STT-MRAMs, introducing two primary approaches: 1) conventional testing approach for defects out of MTJs (i.e., conventional defects like transistor, contact, and interconnect defects) and 2) DAT approach for defects within MTJs (i.e., unique defects). The first approach extends traditional memory test techniques to STT-MRAM, targeting defects that are also found in traditional memories. The second approach is dedicated to emerging NVMs, targeting defects that only exist in certain NVM types; in this thesis, unique defects in STT-MRAMs. By discussing these two methodologies, this section provides a comprehensive overview of how test solutions can be developed to improve STT-MRAM quality while maintaining efficiency in manufacturing.

4.1.1. CONVENTIONAL TEST DEVELOPMENT

As presented in Fig. 4.1, the conventional test development process for STT-MRAM consists of three key steps: defect modeling, fault modeling, and test generation [57]. These steps form a structured methodology to ensure comprehensive defect coverage, allowing for efficient and cost-effective testing methodologies.

Defect modeling serves as the foundation of the test development process by translating physical manufacturing defects into equivalent electrical representations (i.e., models) [57]. During the fabrication of STT-MRAM, various defects may arise due to im-

perfections in lithography, etching, deposition, and other manufacturing steps. These defects can occur in contacts, interconnects, MOSFETs, or peripheral circuits, leading to unintended electrical behaviors. To systematically analyze the impact of these defects, they are modeled by electrical circuit elements. One common approach is to represent physical defects in terms of resistors [61]. For instance, a metal contamination during the fabrication process may unintentionally connect two previously isolated circuit nodes [130], [131]. This defect can be modeled as a bridge resistor, which introduces an unintended conductive path between these two nodes. The resistance value of this bridge depends on the severity of the contamination. Lower resistance values indicate a stronger electrical impact, while higher resistance values suggest a weaker connection that might only affect specific operating conditions. Hence, the defect strength can be represented by the resistance value; a lower value indicates higher strength. Similarly, open defects can be modeled as high resistance connections or complete disconnections in originally connected nodes. This could occur due to incomplete etching or material deposition, resulting in floating nodes that impact memory functionality [130], [131]. In this case, the higher resistance value implies higher defect strength. Through defect modeling, the physical properties of these unintended defects are translated into circuit-level representations that can be simulated through SPICE.

Once defects are accurately modeled at the electrical level, fault modeling is performed to analyze how these defects manifest in the form of memory faulty behaviors during actual memory operations [57]. Fault modeling involves injecting the defect models into circuits and studying their impact on memory functionality. For example, injecting a bridge resistor between two WL may cause write disturb faults, where writing to one memory cell unintentionally alters the state of another cell due to the unintended conductive path [132]. Similarly, an open defect in the BL can lead to read faults, where the sensing circuit is unable to correctly detect the stored data due to signal degradation (i.e., too low read current) [61]. These faulty behaviors are described by fault models using fault notations [132]. Fault models are essential for developing targeted test solutions, as they provide insight into the potential memory behaviors that need to be detected. In conventional memory testing, stuck-at faults, transition faults, and coupling faults are commonly used fault models. However, in STT-MRAM, additional fault models must be considered due to the presence of MTJs.

The final step in conventional test development is generating efficient test solutions to detect the faults identified in the previous step [57]. The primary target of test generation is to design test patterns that can sensitize all validated faults while maintaining low test time. One of the most common test methods is the read/write-based tests, where a sequence of write and read operations is applied to memory cells and verifies if these operations fail. March tests, which are widely used in traditional memory testing [133], can be extended to STT-MRAMs to detect conventional faults. The sequence of applied write and read operations is described by the ‘March algorithm’. The Built-In Self-Test (BIST) techniques are frequently employed to enhance the testing efficiency of March test [71]. A BIST system integrates a dedicated test circuit within the memory chip, allowing it to autonomously execute March algorithms without requiring extensive external test equipment. BIST techniques can detect a wide range of faults and are particularly beneficial for large-scale manufacturing, where reducing test time is crucial.

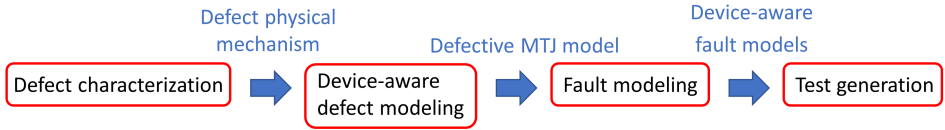


Figure 4.2: Process of DAT.

Additionally, a strategy for improving test coverage is the use of DfT techniques, which involve applying dedicated circuits to detect faults that cannot be caught by March tests [71], [134].

4

4.1.2. DEVICE-AWARE TEST METHODOLOGY

Unique defects are introduced due to the specific working mechanism and structure of MTJs. Unlike conventional defects, which can be represented by resistive shorts or opens, these unique defects within MTJs cannot be accurately modeled using simple linear resistors [66]; MTJs exhibit non-linear and stochastic behaviors in the presence of unique defects. Traditional defect models fail to capture these complex faulty behaviors, leading to inaccurate fault models and, consequently, low-coverage test solutions. To address this challenge, the Device-Aware Test (DAT) methodology is introduced [85], [135]. This methodology incorporates a deeper understanding of the physical mechanisms of defective MTJs into the test development process, ensuring high defect coverage with optimized test efficiency. The device-aware test approach consists of four key steps, as presented in Fig. 4.2: defect characterization, defect modeling, fault modeling, and test generation.

The first step is defect characterization, which involves studying the physical mechanisms of these MTJ defects [66]. This is achieved through a variety of experimental analyses, including the electrical/magnetic measurement (e.g., current-voltage (I-V) measurements, write error rate tests, and thermal stability analysis) on defective MTJs and material inspections (e.g., TEM, and Scanning Electron Microscopy). Defect characterization aims to understand how these unique defects impact the electrical and magnetic properties of MTJs. By analyzing experimental data, researchers can conclude the mechanisms and common properties for certain defect types, such as done for pinholes [66]. The insights obtained from defect characterization are then used to develop accurate defect models that capture the impact of these unique defects on the behaviors of MTJs.

The second step is defect modeling, where the physical impacts of defects are incorporated into compact MTJ models [66]. Unlike conventional defect models, which simply model defects as resistors, device-aware defect models integrate the physical mechanism of defects and design defective MTJ models. To achieve accurate defect modeling, the insights gained from defect characterization are used to change the parameters in the original MTJ model or add more parameters that affect the MTJ behaviors. For instance, a pinhole defect, which is generated due to the imperfection of the TB layer, is modeled by adding a specific parameter that represents the pinhole area (i.e., defect strength) and reduces the MTJ resistance. This change in the MTJ model further alters the TMR ratio, causing deviations in read signals [66]. By embedding these physical defect character-

Table 4.1: Potential defects and the STT-MRAM fabrication steps.

FEOL		BEOL	
Transistor and contact & interconnect	Contact & interconnect	Contact & interconnect	MTJ device
Material impurity	Open vias/contacts	Open vias/contacts	Pinhole
Gate oxide breakdown	Irregular shapes	Irregular shapes	SAFF
Shifting of dopants	Big bubbles	Big bubbles	IM
Patterning proximity	Small particles	Small particles	BH
etc.	etc.	etc.	etc.

istics into MTJ compact models, circuit-level simulations can more accurately predict the behavior of defective MTJs under regular read/write conditions. These models serve as the foundation for fault modeling, ensuring that fault representations reflect the real behavior in simulations that fit the measurement data of defective MTJs.

The third step, fault modeling, builds upon the defective MTJ behavior in the STT-MRAM array at the circuit level [66]. This process is similar to conventional fault modeling but is enhanced to accommodate the irregular (e.g., non-linear) MTJ behaviors in the presence of unique defects. In this stage, defects are injected into SPICE simulations of the memory array by replacing the defect-free MTJ model with the defective model derived from the last step (i.e., defect modeling). Circuit-level simulations are performed to examine how defects affect read/write operations, and other potential factors like reliability. The resulting faulty behaviors are observed and recorded with dedicated fault models. Compared with the conventional fault models, these derived fault models for unique defects may require some adjustments to describe the special behaviors. For example, the pinhole may cause the MTJ resistance to be super low; the ‘L’ is applied to describe the low-resistance state of MTJs.

The final step is test generation, where optimized test solutions are designed to efficiently detect the faults derived from the previous steps [66]. In the DAT approach, test patterns must be carefully developed to ensure high coverage while minimizing test time and cost. Read/Write-Based Tests, like traditional March tests, may remain applicable. For instance, the March algorithm, consisting of one single $1r1$ operation, may detect some pinhole defects; yet this March algorithm shows a high escape rate for pinhole testing considering the varying defect strength [66]. Hence, additional efforts, like the DfT with specific circuits or applying an external magnetic field, are required to generate higher-coverage test solutions; yet these efforts imply higher cost, like extra circuit area. To ensure a balance between high test coverage and low test cost, adaptive test strategies are required depending on the actual application of the STT-MRAMs.

The device-aware test methodology provides a more accurate and efficient approach for testing unique defects in STT-MRAMs by incorporating a deep understanding of MTJ-specific defects into the test process. Unlike conventional methods, which rely on simplified linear resistor models, device-aware testing applies specific defective MTJ models to capture the complex, irregular, and stochastic behaviors. Through the four steps mentioned above, this approach ensures high coverage while minimizing test escapes, hence enhancing the quality of STT-MRAMs in manufacturing.

4.2. DEFECT CLASSIFICATION AND MODELING

The section then introduces the classification of STT-MRAM defects, highlighting their mechanisms and modeling approaches. Defects in STT-MRAMs arise from various sources during the fabrication process, impacting device functionality and reliability. Table 4.1 presents the potential defects and their associated fabrication steps. The corresponding fabrication process of STT-MRAMs is shown in Fig. 2.7. Defects in STT-MRAMs can be broadly classified into conventional defects and unique defects. Conventional defects, i.e., transistor, contact, and interconnect defects, are defects outside MTJs. These defects are well understood and can often be analyzed using established defect models as done in traditional memories like SRAMs and DRAMs [61], [132]. In contrast, unique defects, i.e., MTJ defects, are defects within MTJs stemming from their special structure and physical mechanisms. In the presence of unique defects, MTJs exhibit non-linear behaviors; the DAT is applied to detect these defects

To ensure effective defect detection, defect modeling plays a crucial role in bridging the gap between physical defects and their electrical manifestations. A robust defect model must accurately capture the physical properties of defects and translate them into circuit elements like resistors [61]. By systematically analyzing these defects and defect models, this discussion provides a foundation for accurate fault modeling and effective test development.

4.2.1. CONVENTIONAL DEFECTS AND MODELING

This subsection focuses on the conventional defects in STT-MRAMs, which are outside of MTJs. These defects have been extensively studied in traditional memories like SRAMs and DRAMs [57], [61], [70], [132], [133]. Understanding these defects is essential for fault modeling and developing effective test strategies.

As presented in Table 4.1, conventional defects arise from various fabrication steps involving contacts, interconnects, and MOSFETs [46]. For example, during the MOSFET fabrication in the FEOL process, physical defects may emerge like material impurities, gate oxide breakdown, dopant misplacement, and patterning proximity. Similarly, during the fabrication of contact and interconnect, potential defects like open vias, irregular shapes, and small particles may be introduced. These defects can significantly affect the electrical behavior of STT-MRAM cells, failing read/write operations, and reducing memory reliability.

Despite a large variety of potential conventional defects that exist in STT-MRAMs, the models of these defects can be broadly classified into three categories: open, short, and bridge.

- Open defect model represents cases when an intended electrical connection is partially or completely broken, leading to increased resistance or a complete disconnection. For example, open vias or contacts that occur in WL, SL, or BL can be modeled by this open model. As shown in Fig. 4.3 (a), an open via in the bit line creates an unintended resistive break, preventing proper current flow through the MTJ; this defect is modeled by injecting an extra resistor between BL and the MTJ. Depending on the severity of the open defect, the impact can range from increased read/write latency to a complete functional failure of the memory cell.

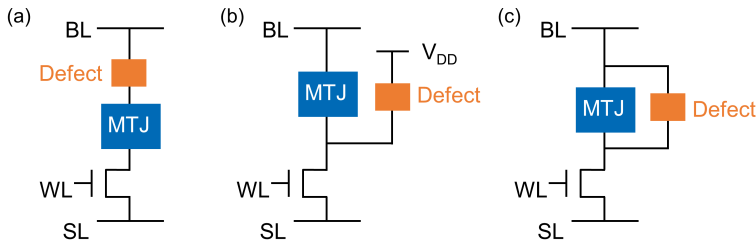


Figure 4.3: Model of conventional defects: (a) open, (b) short, and (c) bridge.

The severity of the open defect, i.e., defect strength, is indicated by the resistance value.

- Short defect model shows an unintended connection between one node and V_{DD} or ground. For example, due to the metal contamination, a memory node between the MTJ and the MOSFET is unintentionally connected to the V_{DD} , as illustrated in Fig. 4.3 (b). This defect leads to failing in MTJ switching, erroneous read operations, and static power leakage. In extreme cases, short defects can cause excessive current draw, leading to thermal damage and further degradation of the whole system. Similar to open defects, the resistance value indicates the defect strength.
- Bridge defect model demonstrates itself as two otherwise disconnected nodes unintentionally connected through an additional resistive path. An example is shown in Fig. 4.3 (c), where gate oxide breakdown prevents the MOSFET from turning off properly; this defect is modeled by a resistor connecting BL and the node between the MTJ and the MOSFET, and the defect strength is represented by the resistance value. Other physical defects that can be represented by a bridge model include lithography misalignment and metal contamination during deposition. Bridge defects can alter the expected resistance of circuit elements, introducing unpredictable voltage shifts and logic errors in STT-MRAM read/write operations.

4.2.2. UNIQUE DEFECTS AND MODELING

Unique defects in STT-MRAM arise from the intrinsic properties of MTJs, which impact the MTJ switching, reliability, and retention of memory cells, leading to irregular faulty behaviors. These defects cannot be accurately modeled using traditional resistor-based approaches due to their complex physical mechanisms. This subsection summarizes previous works on unique defects, showing their mechanisms and modeling methods. Understanding these defects is essential for developing accurate fault models and effective test strategies tailored to the specific characteristics of MTJs.

PINHOLES

Pinhole defects in MTJs are among the critical reliability concerns in fabricating STT-MRAM. These defects originate from imperfections in the tunnel barrier, leading to the formation of unintended conductive paths that affect the device electrical properties. Pinhole defects can significantly degrade the resistance-area (RA) product and the TMR

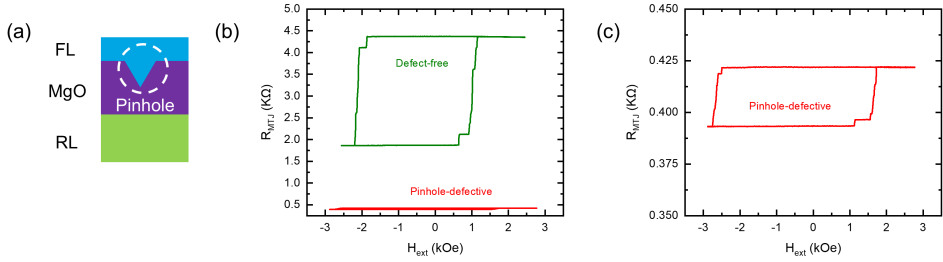


Figure 4.4: (a) Defect mechanism of pinholes; (b) R-H measurement of pinhole-defective and defect-free MTJs; (c) Zoom in of the R-H measurement for the pinhole-defective MTJ.

4

ratio, influencing the read/write performance and long-term reliability of MTJs. Understanding the underlying mechanism and impact of pinhole defects is crucial for improving the yield and device quality of STT-MRAM. Here we discuss the pinhole formation mechanism, its effect on device performance, and the modeling approach used to evaluate its impact.

Defect mechanism

As illustrated in Fig. 4.4 (a) and Fig. 4.7, pinhole defects arise during the multi-layer deposition of MTJs, primarily in the MgO. These defects can result from unoptimized fabrication processes, including contamination, non-uniform deposition, or diffusion of impurities. The formation of pinholes creates localized conductive pathways that ‘short-circuit’ the TB, significantly altering the electron transport properties of the TB. If not detected during initial testing, pinhole defects can lead to premature breakdown of the MTJ, affecting the overall reliability of STT-MRAMs.

Defective MTJ performance

The impact of pinhole defects on MTJ performance is evident in resistance-field (R-H) measurements, as depicted in Fig.4.4 (b). A defect-free MTJ exhibits a regular resistance switching behavior, reflecting the expected TMR characteristics. In contrast, a pinhole-defective MTJ shows significantly reduced resistance in both P and AP states, leading to a lower TMR ratio. The zoomed-in view in Fig.4.4 (c) highlights the reduced dynamic range of resistance for the pinhole-defective MTJ. Although a switching behavior is still observed in the figure, the low resistance and the TMR (0.05 for this case) demonstrate its diminished ability to differentiate between the two resistance states. This degradation in the resistance and the switching behavior adversely affects the read margin of STT-MRAM cells.

Defect modeling

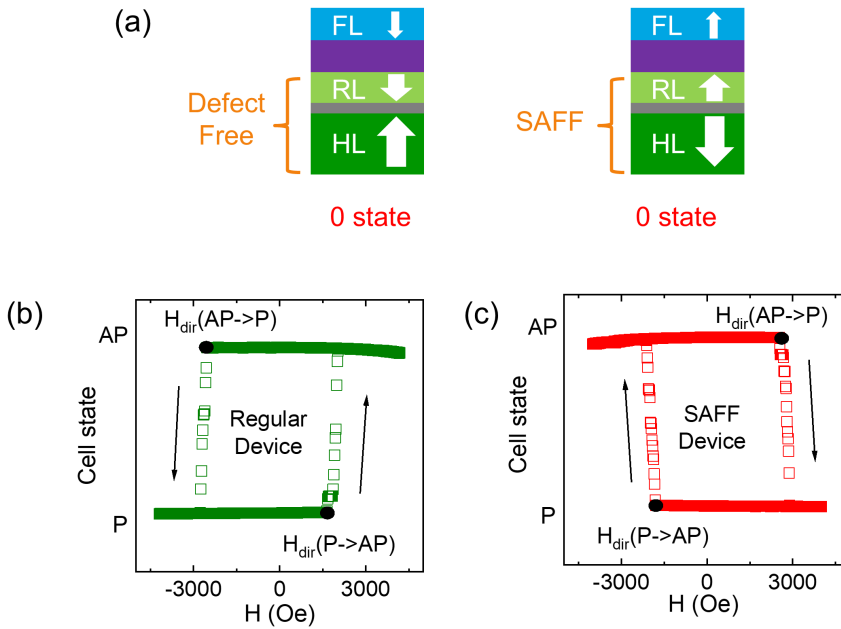


Figure 4.5: (a) Defect mechanism of SAFF; (b) R-H measurement of defect-free MTJs; (c) R-H measurement of SAFF-defective MTJs.

To analyze the impact of pinhole defects on MTJ behavior, an effective modeling approach is adopted. The defect is characterized in terms of its relative area, i.e., ‘pinhole area’, denoted as A_{ph} , normalized to the total cross-sectional area A_0 of the MTJ. The effective resistance-area product and TMR ratio of a defective MTJ are modeled as functions of A_{ph} , allowing for a quantitative assessment of pinhole-induced degradation. By integrating this model into an MTJ compact model by Verilog-A, the electrical behavior of pinhole-defective MTJs can be accurately evaluated, aiding in the development of testing for STT-MRAMs.

SYNTHETIC ANTI-FERROMAGNET FLIP DEFECT

The SAFF defect arises due to an unexpected reversal of the HL magnetization, leading to further unintended changes in the magnetization of RL and FL. Such defects can impact the switching characteristics of MTJs, altering the overall device behavior in write operations. Understanding and accurately modeling the SAFF defect is required to analyze its impact on STT-MRAM performance and develop high-quality test solutions.

Defect mechanism

The SAFF defect is primarily caused by an unintentional flipping of the HL magnetization during the device fabrication process. Due to the strong anti-ferromagnetic coupling between the HL and the RL, it also leads to the reversal of the RL magnetiza-

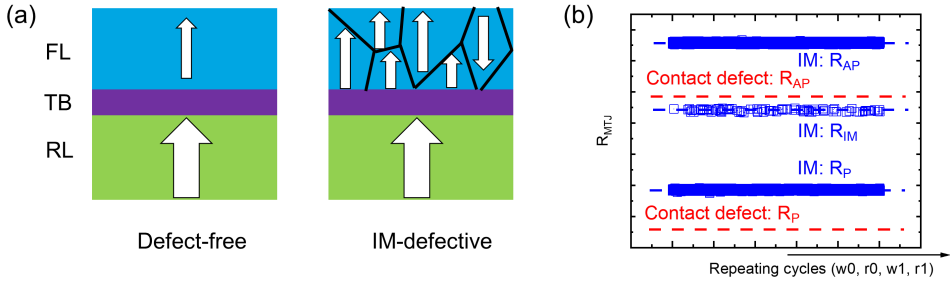


Figure 4.6: (a) Defect mechanism of IM; (b) Extracting IM by repeating w_0, r_0, w_1, r_1 .

tion. Fig. 4.5 (a) illustrates the mechanism of the SAFF defect. For the defective MTJ, the magnetization of HL is upward (\uparrow), and the magnetization of RL is downward (\downarrow). For the SAFF-defective MTJ, the magnetization of HL and RL is reversed compared with the defect-free one.

Defective MTJ performance

The reversal of the RL and HL magnetization leads to a flipped R-H characteristic compared to defect-free devices, as shown in Fig. 4.5 (b) and (c). While the resistance values in P and AP state of MTJs remain unchanged, the defect reverses the effect of the magnetic field. For example, Fig. 4.5 (b) and (c) exhibit that a positive magnetic field switches the defect-free MTJ from P to AP, yet switches the SAFF-defective MTJ from AP to P. Although the SAFF defect has no impact on the electrical switching of MTJs, the functionality While the SAFF defect impacts the magnetic switching behavior, it does not affect the switching behavior in write operations of the MTJ. In the absence of an external magnetic field, the SAFF defect does not influence MTJ functionality. However, in practical scenarios where magnetic fields such as stray fields exist, the SAFF defect alters MTJ switching performance, especially considering the WER [83], [84].

Defect modeling

In the modeling approach, the effect of the magnetic field on the critical current I_c calculation (see Eq. 3.3) is reversed compared to the defect-free case (i.e., positive contributions become negative and vice versa). This adjustment ensures an accurate representation of the SAFF defect impact on both the magnetic and electrical switching behaviors. This correction allows for a precise depiction of how the magnetization reversal of the RL and the HL affects the switching behavior of FL, while not involving the simulations of the two layers, reducing the cost of simulations. Consequently, the model accounts for the observed deviations in MTJ switching and provides a more reliable prediction of the impact of SAFF defects on MTJ performance in the presence of magnetic field.

INTERMEDIATE STATE DEFECT

In MTJs, there are typically two stable magnetic states, i.e., P and AP. The P state exhibits a stable, relatively low resistance (R_P), while the AP state shows another stable high resistance (R_{AP}). These two distinct states correspond to one bit of data in STT-MRAM cells. However, in the presence of IM defects, the MTJ may manifest as a third resistance state between R_P and R_{AP} , disrupting normal binary data storage and increasing the RER and WER. Understanding the mechanism and accurately modeling IM defect is essential for developing high coverage testing for this defect.

Defect mechanism

Several studies have explored IM states in MTJs through both experiments and simulations. The occurrence of the IM defect is usually attributed to the multi-structure of the FL. Especially, the large device size contributes to the phenomenon. Fig. 4.6 (a) presents the mechanism of the IM state. In a defect-free MTJ, the FL is uniform, and the magnetization is consistent across the entire layer, ensuring stable switching between the P and AP states. However, in an IM-defective MTJ, the FL is multi-domain, where the magnetization of one domain may be opposite to that of other domains. This results in a variation in MTJ resistance, as the different domains contribute differently to the tunneling current, leading to intermediate resistance states between the typical P and AP states. This variation in resistance can cause instability in the read/write operations, making it less reliable for memory applications.

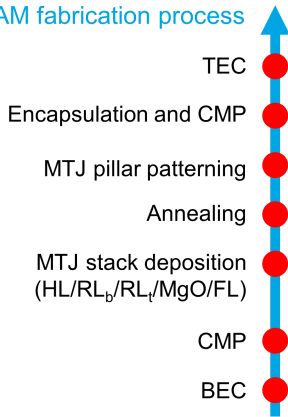
Defective MTJ performance

Fig. 4.6 (b) presents the characterization of the IM-defective MTJ. By repeatedly performing a sequence of operations consisting of $w0, r0, w1, r1$ for multiple cycles (10^4 for this figure), the IM resistance state appears in part of read operations. The intermittent appearance of the IM state suggests that the defects in the MTJ occur sporadically, rather than in a constant manner. This intermittent phenomenon is caused by the fact that some domains have a small probability of not switching during regular write operations. Besides, the resistance remains consistent in every IM defect appearing; the problematic domain (e.g., the domain failing to switch) may be the same one.

Defect modeling

To model the IM defect in MTJs, the measurements for IM-defective MTJs are performed to determine the resistance when the IM state occurs, which occurs intermittently, and the value of which is between the typical R_P and R_{AP} . Next, the probability of the IM occurring is analyzed. This probability has been found to follow a normal distribution, reflecting the intermittent and unpredictable nature of this defect. Once the probability distribution and the corresponding resistance values for the IM state are established, they can be incorporated into the MTJ compact model through Verilog-A, to accurately capture the behavior of the MTJ under conditions where the IM state may arise. This IM-defective MTJ model allows for the simulation of MTJ performance under

STT-MRAM fabrication process



Possible fabrication step introducing MTJ defects

Defect	Fabrication process
Pinhole	MTJ stack deposition (RL _t /MgO/FL) Annealing Encapsulation and CMP
SAFF	MTJ stack deposition (HL) Annealing
IM	MTJ stack deposition (MgO) Annealing
BH	MTJ stack deposition (HL/RL _b /RL _t) Annealing

TEC: top electrode contact; BEC: bottom electrode contact;
CMP: chemical mechanical polishing

Figure 4.7: MTJ manufacturing process and the potential unique defects.

Table 4.2: MTJ defect mechanism and modeling method.

Defect	Mechanism	Defective MTJ performance	Defect modeling method
Pinhole	TB deterioration [136]	MTJ resistance reduction [136]	Introduce breakdown area in TB [79]
SAFF	HL flip [87]	Magnetic field impact on MTJ reversed [87]	Flip FL magnetization [87]
IM	Multi-domain FL [137]	MTJ intermittently enter undefined state [137]	Introduce undefined MTJ state, introduce switching probability [86]
BH	RL instability [80]	MTJ state oscillation [138]	Add pinning effect for RL, form RL switching process [82]

realistic conditions where defects may lead to intermediate resistance states, enabling a more reliable prediction of MTJ behaviors

4.2.3. SUMMARIZE AND OVERVIEW OF MTJ DEFECTS

Fig. 4.7 presents the MTJ fabrication steps and the potential unique defects that may be introduced by each step. Table 4.2 summarizes the mechanism and modeling method of each defect. The unique defect of BH is one of the major contributions of this thesis, which will be explained in chapter 6.

Fig. 4.7 presents the MTJ fabrication steps and the potential unique defects that may be introduced by each step. The identification of which MTJ fabrication step introduces which defect is based on reasonable assumptions derived from the defect mechanisms. Due to the complexity of the manufacturing process, it is difficult to precisely pinpoint the exact source of defects in real-world scenarios. TABLE 4.2 summarizes the mechanism and modeling method for each defect. Among these, the unique defect of BH is one of the major contributions of this thesis, which will be explained in detail in chap-

ter 6.

4.3. FAULT CLASSIFICATION AND MODELING

The fault model is applied to describe the faulty behaviors of STT-MRAMs in the presence of various manufacturing defects. Fault modeling involves injecting defects into circuits and studying their impact on memory functionality. Conventional defects are modeled as linear resistors and directly injected into the circuit. For unique MTJ defects, the DAT method is applied to design a specific defective MTJ model, which represents the faulty behavior caused by the corresponding defect. The defect injection is performed by replacing the defect-free MTJ model with the defective model. Then, read/write operations are applied to memories to observe the potential faulty behaviors.

Fault modeling helps develop test strategies by providing insights into how defects can affect the performance and reliability of a memory device under various conditions. Through a proper fault modeling approach, designers can identify potential weaknesses in the STT-MRAM prior to physical testing, thereby focusing testing on the key cases of the potential STT-MRAM failing, ensuring the testing process is efficient and thorough. For example, if a fault model predicts that certain defects lead to intermittent resistance states, test strategies can be designed to specifically check for these conditions under realistic operational scenarios.

In this section, we first define fault notation, which standardizes the representation of faults in models, making it easier to analyze. We then discuss fault classification, grouping faults based on different aspects, like their impact and the property. Understanding fault classification aids in developing efficient detection strategies for certain defect types.

4.3.1. FAULT PRIMITIVE NOTATION

Table 4.3 presents the fault notations applied to describe the memory faults [61], [79]: $\langle S/F/R \rangle$, S describes the sensitizing sequence, F describes the faulty effect, and R describes the readout value. For example, $\langle 0r0/0/1 \rangle$ denotes a $r0$ operation on a cell that holds '0' ($S=0r0$), where the cell remains in its correct state '0' ($F=0$) yet the read output returns to '1' ($R=1$) instead of the expected '0'. Next, we have a detailed description of 'S', 'F', and 'R' separately.

S: SENSITIZING SEQUENCE

S describes the operation sequences of the STT-MRAM cell, which can be presented as $S = x_0 O_1 x_1 \dots O_m x_m \dots O_n x_n$, where $x_m \in \{0, 1\}$, and $O_m \in \{w, r\}$. Here, '0' and '1' indicate the state of the STT-MRAM cell, and 'w' and 'r' denote write and read operations, respectively.

The sequence (i.e., operations) can be further classified as the static operation and the dynamic operation [139]. The static operation is defined where at most one operation is performed within the sequence. Notice that no operation, e.g., just keeping the STT-MRAM cell in state '1', also belongs to this 'static operation' type. Dynamic operation, on the other hand, involves at least two actions within the sequence [139]. It is important to note that the number of dynamic operations significantly exceeds the number of static operations. While dynamic operations may expose more faults, they also incur

Table 4.3: Fault primitive notation.

$\langle S/F/R \rangle$	Explanation	Value	
S	Sensitizing sequence	static operations: $0, 1, 0w0, 0w1, 1w0, 1w1, 0r0, 1r1$ dynamic (two) operations: $0w0w0, 0w0w1, 0w0r0, 0w1w0, 0w1w1, 0w1r1,$ $1w0w0, 1w0w1, 1w0r0, 1w1w0, 1w1w1, 1w1r1,$ $0r0w0, 0r0w1, 0r0r0, 1r1w0, 1r1w1, 1r1r1$	
F	Faulty effect	L, 0, U, 1, H	
R	Readout value	0, 1, ?, -	
note in 'F':			
'L'	MTJ extreme low state	'0'	MTJ normal low state
'U'	MTJ undefined state	'1'	MTJ normal high state
'H'	MTJ extreme high state		
note in 'R':			
'0'	Readout low state	'1'	MTJ Readout high state
'?'	Readout random state	'-'	Readout not applicable

a higher time cost [70]. All of the static operations and dynamic operations (with two operations) are shown in Table 4.3.

F: FAULTY EFFECT

F denotes the stored value (i.e., states) in the STT-MRAM cell after a certain sequence S. The possible states rely on different memory types. For conventional memories, like SRAMs and DRAMs, there are only three states: '0', 'U' (the undefined state), and '1' [57]. For STT-MRAM technology, the possible resistance states of the MTJ vary in the presence of conventional defects or unique defects. In the presence of conventional defects, the MTJ is defect-free and has only two stable resistance states, i.e., '1' and '0' [41]. In the presence of unique defects, the defective MTJ may exhibit additional resistance states, i.e., 'L', 'U', and 'H' [137]. Here, U' refers to the undefined state, where the MTJ resistance is in an intermediate range between that in the regular '0' and '1' states. 'L' refers to an extremely low resistance state, where the MTJ resistance is lower than that in the '0' state, and H' refers to an extremely high resistance state, where the MTJ resistance is higher than that in the '1' state. [137].

R: READOUT VALUE

R refers to the output of the read circuit (i.e., SA). Here, $R \in \{0, 1, ?, -\}$. '0' is the regular output when the MTJ is in state '0'. '1' is the regular output when the MTJ is in state '1'. '?' is a random read outcome (e.g., the output voltage of the SA is an intermediate value

between V_{DD} and ground). ‘-’ denotes the case where the last operation of S is not a read operation, hence no output from SA.

4.3.2. FAULT CLASSIFICATION

These faults can be classified based on different perspectives, such as their origins, properties, and impact on the system. Next, we show some common classification methods [67].

STRONG/WEAK

If a fault always causes functional errors, it is defined as a strong fault; if a fault cannot always cause a functional error, it is defined as a weak fault [67]. For example, in the context of STT-MRAMs, a strong fault could be a permanent short-circuit at the node of SL, which prevents the MTJ switching. This would lead to a functional error, and can be represented by the fault primitive $\langle 1w0/1/- \rangle$.

PERMANENT/INTERMITTENT

If a fault is permanently present in the memory irrespective of the time of access, it is defined as a permanent fault; if a fault is sensitized intermittently (i.e., from time to time), it is defined as an intermittent fault [67]. For example, a permanent fault is sensitized when a short defect exists at SL, preventing the MTJ from receiving sufficient energy to switch its state [86]. As a result, the MTJ remains stuck in one resistance state. In contrast, an IM defect sensitizes an intermittent fault because it does not always affect the switching behavior of MTJs. Instead, due to the nature of the defect, the IM-defective MTJ occasionally exhibits the ‘U’ state, rather than a stable 0 or 1. This issue occurs unpredictably during certain write operations, making it challenging to design high-coverage testing solutions. A specific subscript ‘i’ is applied to represent the intermittent fault. For example, $\langle 1w1/0_i/- \rangle$ means that by applying the $1w1$ operation, the MTJ state remains in ‘0’ intermittently.

SINGLE-CELL/MULTI-CELL

If a fault is not affected by states of neighboring cells, it is defined as a single-cell fault; if a fault sensitization is affected by states of neighboring cells, it is defined as a multi-cell fault [67], [140]. For example, in the presence of a bridge defect between two neighboring cells, the write current of one cell may rely on the state of the other cell by a sneak current path through this defect [132]; hence, sensitizing a multi-cell fault. On the other hand, the MTJ behavior will not be affected by neighboring cells in the presence of the defect in Fig. 4.3 (a); hence, this defect only sensitizes a single-cell fault.

STATIC/DYNAMIC

If a fault can be sensitized by performing at most one operation, it is defined as a static fault; if a fault can only be sensitized by performing more than one operation (i.e., dynamic operations), it is defined as a dynamic fault. Dynamic faults are common in conventional memories and other NVMs [139], [141]; however, we show that dynamic faults are rare for STT-MRAMs in this thesis.

HARD-TO-DETECT/EASY-TO-DETECT

If a fault is guaranteed to be sensitized and detected by regular read/write operation, it is defined as an easy-to-detect fault; if a fault cannot always be sensitized and detected by regular read/write operation, it is defined as a hard-to-detect fault. For example, in the presence of an IM defect, the MTJ is in 'U' state, yet the state may not be caught by the SA; the fault relating to this defect is categorized as a type of hard-to-detect defect. On the other hand, the SA readout may get stuck at '0' in the presence of the defect in Fig. 4.3 (a) (e.g., too low read current), and the corresponding defect belongs to the 'easy-to-detect' type.

4.4. TEST GENERATION

Testing is a critical step in ensuring the functionality and reliability of STT-MRAMs, which are susceptible to both conventional and unique defects. Due to the non-volatile nature and unique switching mechanism of STT-MRAM cells, dedicated test strategies are required to detect faults effectively. This section presents two key approaches: March tests and DfT techniques.

March tests are widely used algorithmic test methods that apply systematic read and write operations to detect various faults in memory arrays. To further enhance testability, DfT techniques have been developed to improve fault coverage and reduce test time. By combining March tests with DfT techniques, manufacturers can achieve a more comprehensive and efficient test strategy, ensuring the high reliability of STT-MRAM devices.

4.4.1. MARCH TEST GENERATION

March testing is a widely used read/write-based approach for detecting faults in memory systems, including STT-MRAMs [57], [58], [61], [69], [70], [76]. It consists of a sequence of read and write operations applied to each memory cell in a specific order. The goal of a March test is to efficiently identify different types of faults by systematically stressing the memory with various operations and observing the resulting behavior. Due to its simplicity and effectiveness, March testing is commonly used in BIST architectures for high-quality memory applications [71].

A typical March test consists of multiple March elements, where each element represents a series of operations performed on all memory cells in a specific address order [57], [133]. The order in which these operations are applied can be increasing (from lower to higher memory addresses), decreasing (from higher to lower addresses), or both. The sequence is carefully designed to guarantee a high coverage of a variety of faults.

One of the key advantages of March tests is their ability to detect faults efficiently with a linear time complexity, making them practical for large-scale memory arrays [57], [133]. Different March algorithms exist, such as March C-, March Y, and March B, each optimized for detecting specific fault types [142], [143]. For example, March C- is commonly used for detecting common stuck-at faults, while March B is designed for more complex situations [143], [144].

For STT-MRAMs, traditional March tests are also a good solution that can be applied to detect a variety of faults sensitized by manufacturing defects [76]. However, limita-

tions exist if only applying March tests to STT-MRAMs [66]:

- **High escape rate:** Some faults, such as intermittent faults and unique faults (introduced by unique defects), can not be 100% sensitized by regular March tests [81]. Hence, a high escape rate is observed.
- **Low time efficiency:** To detect some faults, the march algorithm is set to be long, resulting in a low time efficiency.

To address these challenges, March tests may need modifications like increasing the number of test iterations, introducing adaptive test patterns.

4.4.2. DESIGN FOR TEST

DfT is a set of techniques used to improve the testability of memory systems by enhancing fault detection, enabling diagnosis, and reducing yield loss. In STT-MRAMs, DfT strategies are essential in the presence of both conventional defects and unique defects. Without effective DfT solutions, applying conventional read/write-based tests to detect faults sensitized by these defects exhibits a high escape rate [66].

Several DfT techniques have been developed specifically for STT-MRAMs, targeting different aspects of fault detection and defect mitigation. Besides, the BIST can also be viewed as a widely used DfT technology, integrating test circuitry within the memory chip to perform self-testing [71]; BIST allows efficient execution of test patterns, such as March tests, without the need for external test equipment, reducing test time and cost. A variety of DfT designs have been proposed dedicated to STT-MRAMs [82], [100], [107], [145]–[150], each targeting certain defects.

DfT plays a crucial role in enhancing the testability and reliability of STT-MRAMs by addressing both conventional and unique defects. As STT-MRAM technology continues to evolve, optimizing testing solutions remains essential for ensuring high-quality products. By integrating effective DfT solutions, manufacturers can enhance fault coverage, minimize production yield loss, and improve the overall robustness of the STT-MRAM system.

In order to apply DfTs effectively in industrial STT-MRAM chips, the following aspects must be carefully addressed:

- **Integration:** The DfT circuit must be integrated to STT-MRAM peripheral circuits, and the control signals must be compatible with STT-MRAM write/read control signals. For example, the DfT output can be shared with the SA output,
- **Defect coverage:** The DfT must achieve high coverage for faults that cannot be reliably sensitized by traditional March tests, such as those unique faults [83].
- **Time efficiency:** The time required to perform the DfT must be short to avoid significant overhead in the testing process, ensuring it is practical for high-volume manufacturing.
- **Circuit area cost:** The DfT circuit should have a low area overhead to minimize its impact on the overall chip size and manufacturing cost.

A detailed investigation and evaluation of existing DfTs for STT-MRAMs will be presented in Chapter 7.

5

TESTING OF TRANSISTOR, CONTACT, AND INTERCONNECT DEFECTS IN STT-MRAMs

To advance STT-MRAM manufacturing, it is critical to obtain high-quality test solutions. This chapter presents a comprehensive framework for modeling transistor, interconnect, and contact defects in STT-MRAMs, deriving corresponding fault models, and generating test methods based on observed fault models. These defects have already been observed in traditional memories like SRAMs, where mature test methods have been developed. Due to the specific working mechanism of MTJs, such as magnetic coupling and intrinsic stochasticity, it is not appropriate to straightforwardly apply regular march algorithms to STT-MRAMs. To solve this challenge, it is essential to integrate the impacts of MTJ working mechanisms into fault modeling, which is the key contribution of this chapter. The test solutions are then generated based on accurate fault models, including the March test and dedicated DfT solutions to detect derived faults.

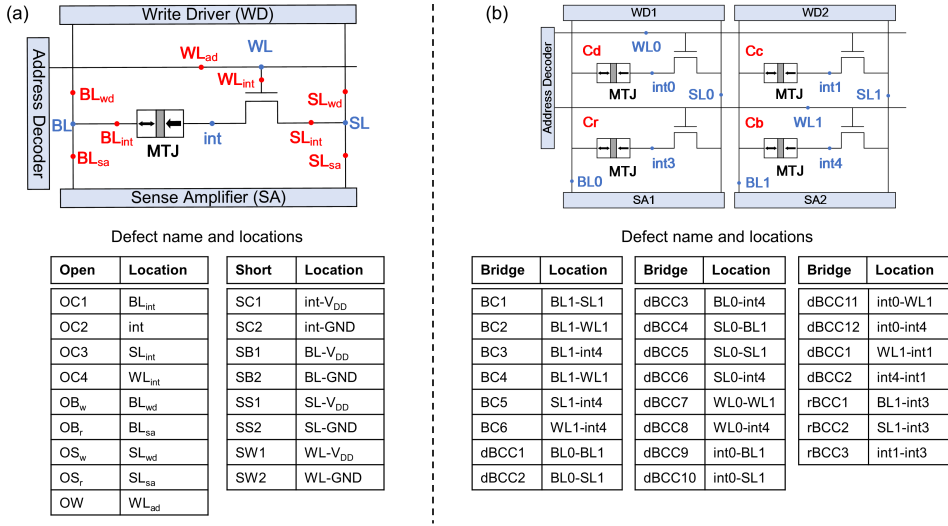


Figure 5.1: MTJ manufacturing process and the potential unique defects.

5.1. DEFECT MODELING AND DEFECT MAP

In this section, the complete space of defect models of transistor, contact, and interconnect defects is defined for STT-MRAMs. Here, the defect is modeled as resistors, which are commonly used [61]. As already introduced in the previous chapter, traditional defects can be modeled as: 1) open, an extra resistance within the connection; 2) short, an undesired resistive path between the node and power supply; and 3) bridge, an extra parallel resistance between two disconnected nodes. The defect strength is described by the resistance value ranging from $1\ \Omega$ to $10\ \text{M}\Omega$. Note that whether all defined defects are realistic or not is strongly chip layout dependent [74]. Such information is barely published, and therefore, in our analysis, we will cover the whole space. Next, we design the defect space of each defect type.

5.1.1. DEFECT CLASSIFICATION

Conventional defects can be classified based on different perspectives. Here we introduce some classification methods that will be applied in this thesis [67].

INTRA- AND INTER- CELL DEFECTS

Depending on the location, defects can be classified as *intra-cell* and *inter-cell* defects [151], [152]. Intra-cell defects occur within a single STT-MRAM cell. These defects include open defects, short defects, and bridge defects that only connect nodes within one cell (e.g., BC3). Inter-cell defects, on the other hand, involve multiple memory cells and create a sneak current path between them, potentially leading to coupling faults. These defects typically manifest as bridging defects connecting nodes between adjacent cells. Examples include dBCC1, dBCC2, and rBCC3 in Fig. 5.1 (b).

WEAK AND STRONG DEFECTS

Based on their impact on memory functionality and fault sensitization conditions, defects can be classified as *weak* and *strong* defects. A strong defect is defined as one that always results in a functional fault when it is present. In contrast, a weak defect does not lead to functional faults, but rather causes parametric deviations such as a high WER. This classification relates to the defect strength. For example, the open defect in OC2 can permanently prevent the MTJ from switching states when it shows a high strength (i.e., high resistance value), hence categorized as a ‘strong defect’. However, with low strength, OC2 may just reduce the write current a bit and not lead to write failures. In this case, the defect is categorized as a ‘weak defect’.

5.1.2. SPACE OF OPEN AND SHORT DEFECTS

Although the next step, ‘fault modeling’, will be performed by simulating a small STT-MRAM array, we only inject the short and open defects in one cell for simplification, as presented in Fig. 5.1 (a). Due to the symmetry, defects in different cells cause similar faults. For instance, the open defects in different cells will cause these cells to fail in the same way.

OPENS

Fig. 5.1 (a) demonstrates the complete space of the open and short defects. As presented in the blue words in the figure, we show all four ‘nodes’ in the netlist of the 1T-1MTJ cell: SL, WL, BL, int. The method of defining the defect is as follows: The open defect is presented as an extra resistance between the connected node and the device, or between the connected node and the peripheral circuit. Regarding the location, the open defects can be further classified as 1) open within the cell (OC), 2) open at BL (OB), 3) open at SL (OS), and 4) open at WL (OW). For example, the defect OS_r is modeled as an extra resistance between the node SL and the peripheral circuit SA. Hence, there are nine open defects in total.

SHORTS

The short defect refers to the unexpected resistive path between a node and Power/GND. This defect type can be further classified as 1) short within the cell (SC), 2) short at BL (SB), 3) short at SL (SS), and 4) short at WL (SW). For example, SC1 is modeled as an extra pass connecting the node int and GND. With four nodes, each node can be connected to either Power or GND to introduce short defects; hence, there are a total of eight short defects.

5.1.3. DEFECT SPACE OF BRIDGE DEFECTS

To simplify the simulation process, the bridge defects are only injected in the 2×2 array, which includes all three possible locations of adjacent cells: cells in the same row, cells in the same column, and cells in the same diagonal. Regarding the defect location, bridge defects can be further classified as 1) bridge within the cell (BC), 2) bridge between cells in the same diagonal (dBCC), 3) bridge between neighboring cells in the same column (cBCC), and 4) bridge between neighboring cells in the same row (rBCC). Due to the symmetry, faults sensitized by the bridge defects in the 2×2 array represent the faults

sensitized by all the possible bridge defects in the large-scale array [67]. Because the large-scale $m \times n$ array can be treated as the repetition of 2×2 arrays, the same defect in different 2×2 arrays actually results in the same types of faults.

Fig. 5.1 (b) demonstrates the complete space of the bridge defects. The method of defining the bridge defect is as follows. We present all ten nodes in the 2×2 array (the blue words in Fig. 5.1 (b)). Since the bridge indicates an extra parallel resistance between two disconnected nodes, we first define all the possible defects between every two of the ten nodes. However, not all defects between the ten nodes need to be considered due to symmetry. For instance, with the existence of defect BC3 (BL1-int4), the bridge defect between BL0-int0 is unnecessary to be considered. Therefore, we delete the repeating ones, and eventually get a total of 23 bridge defects as shown in Fig. 5.1 (b).

5.2. FAULT MODELING

5.2.1. FAULT MODELING METHODOLOGY

Although previous studies have made significant contributions to fault modeling for interconnect and contact defects in STT-MRAMs, they show some limitations. Some papers simply inject the defect and apply operations to STT-MRAMs and observe the corresponding faults [74], [100], [153]. However, due to the lack of in-depth understanding of the MTJ mechanism, they sometimes introduced inaccuracies, e.g., some works do not take into consideration magnetic coupling in the fault modeling. Some studies have conducted in-depth investigations into MTJ mechanisms and incorporated them into STT-MRAM testing. Yet, they typically focus on a single aspect rather than presenting a comprehensive testing methodology [74], [83], [107], [122]. For instance, [107] explores the impact of MTJ stochasticity on fault modeling but does not propose a complete testing scheme. To solve this challenge, this chapter performs fault modeling on four different platforms. This analysis helps us understand how all MTJ-specific mechanisms impact fault modeling.

SIMULATION SET-UP

In circuit simulations, Cadence Spectre is adopted, and the library of 40 nm TSMC is applied. The 4×4 sub-array is applied with peripheral circuits presented in chapter 2. The supply voltages are set to $V_{DD} = 0.9V$. The MTJ model is also described in chapter 2; the model is calibrated with measurement data of MTJs with the Critical Diameter (eCD) of 60 nm, and the pitch (distance between neighboring cells in the same row or the same column) is 90 nm. The complete defect space is presented in Fig. 5.1. Each time, we inject one defect, varying the defect strength (i.e., the resistance) from 1 to 10 M Ω , performing operations as presented in Table 4.3, and observing the potential faulty behaviors.

SIMULATION PLATFORM SETTING

While significant work has been reported on fault modeling for conventional defects in STT-MRAMs [74], [83], [100], [107], [122], [153], these studies often overlook the effects that originate from MTJ working mechanisms on fault modeling, including the magnetic coupling and intrinsic stochasticity [107]. To address this challenge, we set up four platforms, which help distinguish how each effect in the MTJ working mechanism affects the

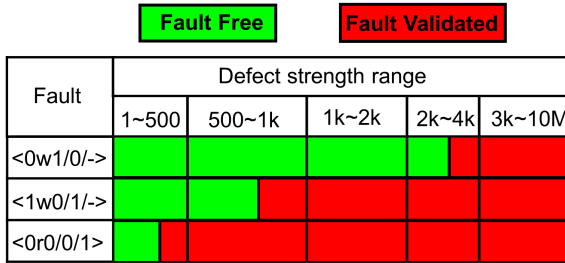


Figure 5.2: The basic fault map of OC2.

fault modeling; then, we show the fault modeling results in which all effects are included. The four platforms with four different model settings are presented as follows:

- 1 *Baseline Model*: in this platform, no magnetic coupling or intrinsic stochasticity is considered, two types of data backgrounds (DBs) of all neighboring cells are all set to 0. The model applied in this platform only includes the R-V part and basic switching part (see Fig. 3.1).
- 2 *Magnetic Coupling-Only Model*: this platform considers the impact of DB compared with the 'Basic platform'. Especially, the magnetic coupling is included in the MTJ model. Compared with the Baseline Model, the model applied in this platform includes the part of the magnetic field impact (see Fig. 3.1).
- 3 *Stochastic-Only Model*: this platform considers the intrinsic stochasticity of MTJs compared with the 'Basic platform'. Compared with the Baseline Model, the model applied in this platform includes the part of the intrinsic stochasticity at both low and high voltage (see Fig. 3.1).
- 4 *Full-Variability Model*: this platform considers both DB and MTJ intrinsic stochasticity compared with the 'Basic platform'. The model applied in this platform includes all parts in Fig. 3.1.

The involved MTJ modeling work has been introduced in Chapter 3; our model can open or close the function of components about *Magnetic Field Impact* and *Intrinsic Stochasticity*, which supports the four platforms above. Through these four platforms, we illustrate the mechanism and root cause of certain fault types. The results of the complete fault map in our study will be shown at the end of this section.

5.2.2. RESULTS FOR BASELINE SIMULATION MODEL

In this subsection, we apply the model only with components of the R-V calculation and the basic switching, and close all other components. The fault modeling is performed, and a fault map is generated for each defect.

FAULT MAP ANALYSIS

An example of the fault map is presented in Fig. 5.2 of the defect OC2. The fault map consists of (from left to right) the fault and whether the faults are sensitized, which depends

on the defect strength range. The ‘green’ boxes indicate defect strength ranges where faults are sensitized, and the ‘green’ ones indicate that no faults are sensitized. Three faults are sensitized, as shown in the figure, the $\langle 0w1/0/- \rangle$, $\langle 1w0/1/- \rangle$, $\langle 0r0/0/1 \rangle$. To maximize the test coverage, we need to select the fault that sensitizes the longest defect strength. For OC2, the $\langle 0r0/0/1 \rangle$ is selected. A similar fault map is generated for each defect, and a similar process is performed on each fault map to select the proper fault, which will not be presented here.

FINDINGS

The major findings of this subsection are:

- Only static permanent faults are observed, like the three faults shown in Fig. 5.2.
- No dynamic faults are observed.
- No intermittent faults are observed.
- Since we only apply one DB, no coupling DB is observed.

Notice that, in our work, no dynamic faults are found (not only in this subsection, but also in the other three subsections). While we do not entirely rule out their existence, particularly in complex peripheral circuits [154], our findings suggest that dynamic faults are rare in STT-MRAMs compared to other memory technologies, which can be attributed to two key reasons:

- **Simple structure:** The STT-MRAM cell is in the 1T-1MTJ structure, which is much simpler compared with conventional memories (e.g., SRAM consists of 6 SRAMs [155] and DRAMs).
- **Stable resistance states of MTJs:** Defect-free MTJ has only two stable resistance states, and never shows intermediate states due to the anisotropy magnetic field [41]. Other non-volatile memories, like Flash, RRAMs, FeFETs, and PCRAMs, all have intermediate resistance states in their defect-free devices, which introduce dynamic faults [156]. For example, in the presence of a bridge defect, a relatively large current may flow across the RRAM device during one read operation. This voltage cannot completely switch the device, but changes the device state a bit. The effects of multiple read operations will accumulate and eventually switch the device, introducing a dynamic fault [141]. Yet this accumulation will never occur in STT-MRAMs.

5.2.3. RESULTS FOR MAGNETIC COUPLING-ONLY SIMULATION MODEL

In this subsection, we apply the model with components of the R-V calculation, the basic switching, and the magnetic field impact, closing the components of stochasticity. We aim to investigate the DB impact on fault modeling. Two factors are involved once we apply different DBs: the magnetic coupling introduced by the stray field [68], [84], [89], and the electrical effect introduced by the sneak current path [61], [67]. The MTJ model work incorporated with the magnetic coupling is introduced in Chapter 3. The results in this subsection will be compared with those from the subsection with the baseline simulation model. Next, we introduce how we set up different DBs.

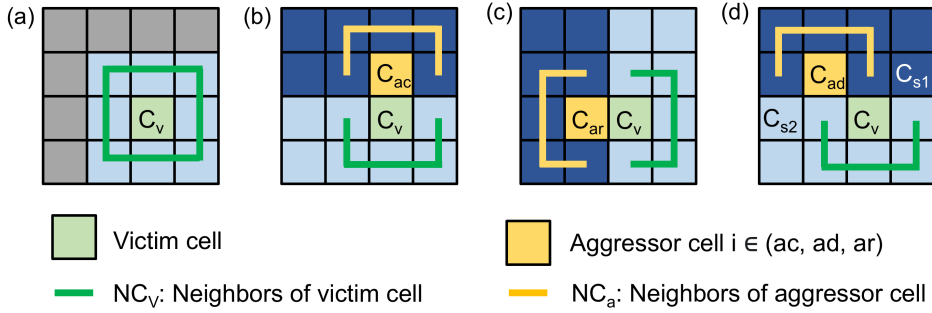


Figure 5.3: Illustration of the DBs for the fault modeling methodology for (a) single cell defects and (b)-(d) defects between adjacent cells.

DATA BACKGROUND SET-UP

In this subsection, the fault modeling is performed for two cases: 1) all defects in a single cell (see Fig. 5.3 (a)), and 2) all defects between two adjacent cells (see Fig. 5.3 (b)-(d)). For both cases, the victim-cell C_v is presented as the green square, in the middle of the array. The neighboring cells of C_v are defined as NC_v .

The DB setting of case 1) is presented in Fig. 5.3 (a). Here, we consider only two extreme cases for the DBs: all the eight neighboring cells are set to 0 or to '1' (i.e., NC_v to be all '0' or '1' state), meaning using *solid* DB patterns [132].

The DB setting of case 2) is presented in Fig. 5.3 (b) - (d). Here, defects are between C_v and another neighboring cell; this cell connected with C_v by the defect is called the Aggressor-Cell C_a , $a \in \{ac, ar, ad\}$; the state of C_a may affect the fault modeling. Depending on the defect location, Fig. 5.3 (b)-(d) present three types of C_a : 1) C_{ac} , the same column of C_v , 2) C_{ar} , the same row of C_v , 3) C_{ad} , the same diagonal of C_v . The behaviors of both C_v and C_a are observed while applying the operations on C_v to perform the fault modeling. The neighboring cells of C_a are defined as NC_a . Because both NC_v and NC_a may have an impact on coupling faults, we need to set them separately, with a total of four cases: $(NC_v, NC_a) = (0, 0), (0, 1), (1, 0), (1, 1)$. Next, we combine our methodology with certain DB patterns that are applied in the industry test design [132]. Since the regular 'Row Stripe' and 'Column Stripe' patterns in [132] do not directly match our methodology, we use the 'Double-Row Stripe' and 'Double-column Stripe' patterns, as presented in Fig. 5.3. The 'Double-Row Stripe' pattern is applied for defects between C_v and C_{ac} , and the 'Double-Column Stripe' pattern is applied for defects between C_v and C_{ar} (Fig. 5.3 (b) - (c)). For cells that are neighboring both C_v and C_a , how to set them depends on the actual situation. For instance, in Fig. 5.3 (b), the cell on the diagonal of C_v and the same row of C_{ac} is treated as the NC_a , since the stray field from this cell has a stronger impact on C_{ac} than C_v . However, it is difficult to combine the DB pattern with cases when defects are between C_v and C_{ad} , since no proper DB pattern can properly distinguish the NC_v and NC_a cells. For example, Fig. 5.3 (d) presents the application of the 'Double-Row Stripe' pattern in this case, there are two specific cells C_{s1} and C_{s2} that cannot be grouped well: C_{s1} should belong to NC_v , yet in this pattern, it is the same state of to NC_a , the similar case occurs for C_{s2} . In this work, we temporarily ignore the two

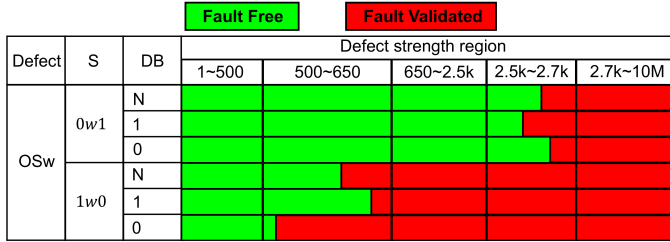
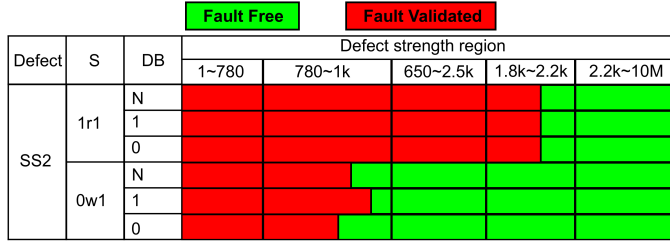
Figure 5.4: Fault map of OS_w .

Figure 5.5: Fault map of SS2.

cells, and simply apply the ‘Double-Row Stripe’ in Fig. 5.3 (d). Next section, we will apply the operations on C_v for all three cases for fault modeling first, then analyze the impact of the two cells, and show that it can be ignored.

ANALYSIS OF FAULT MAP

Fig. 5.4 and Fig. 5.5 show two examples of fault maps summarizing the results of the fault analysis performed for the defects OS_w , respectively, SS2 (see the table in Fig. 5.1). The fault map consists of (from left to right) the fault, DB used for the eight physical neighbor cells, and whether the faults are sensitized, depending on the defect strength range. Here, ‘DB=N’ for DB indicates No Magnetic Coupling considered (i.e., the ‘baseline model’ where I_{s-FL} is not incorporated during the simulation), while ‘DB=1’ (‘DB=0’) indicates that magnetic coupling is incorporated with all the neighboring cells of the C_v set to 1 or 0. The ‘red’ boxes indicate ranges where faults are sensitized, and the ‘green’ ones indicate that no faults are sensitized. To demonstrate the magnetic coupling impact and electrical effect, we make additional simulations to zoom in on the range around the border of the red and green boxes.

To maximize the test coverage, we need to consider the DB pattern and associated operations that can sensitize faults for the longest defect strength range; this has to be done for each defect. For example, Fig. 5.4 reveals that the operation $1w0$ sensitizes the longest defect strength range of OS_w for DB=0, which is attributed to the magnetic coupling. When the neighboring cells are in state ‘1’, the $0w1$ operation is favored since the magnetic coupling contributes to the switching process; it is modeled as the magnetic coupling contributes with a current I_{s-FL} going in the same direction as that of I . On the other hand, when the neighboring cells are in state ‘0’, the $0w1$ operation is disfavored

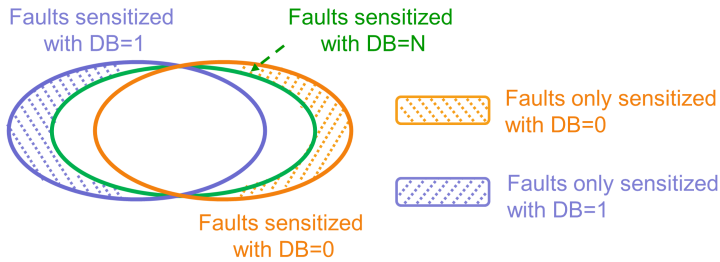


Figure 5.6: Faults Venn diagram for all analyzed single cell defects.

since I and I_{s-FL} are in opposite directions; hence, $DB=1$ enables more $1w0$ operations to fail for smaller defect sizes. A similar explanation can be given for the operation $0w1$. The defect coverage for the $DB=N$ case is always between that of the two extreme cases of $DB=0$ and $DB=1$. Hence, not considering the magnetic coupling will never provide maximal test coverage.

Fig. 5.5 shows the magnetic coupling effect for $SS2$. For the $0w1$ operation, the longer defect strength range is sensitized for $DB=1$. However, the longest defect strength range is sensitized by the $1r1$ operation, and it is the same for all the cases $DB=N$, $DB=1$, and $DB=0$; this indicates that read operation is not affected by magnetic coupling, which is consistent with our analysis above. Clearly that to maximize the defect coverage for $SS2$, the $1r1$ operation has to be used regardless of the DB pattern.

Fig. 5.6 gives the Venn diagram that summarizes the results found when analyzing all single-cell defects for the three cases ($DB=N$, $DB=1$, and $DB=0$). As shown in the figure, performing fault modeling without incorporating magnetic coupling will result in a reduced validated fault space. Building test solutions based on this will obviously not lead to high-quality test solutions. Including the impact of magnetic coupling in the fault analysis fosters the sensitization of additional faults; some of these are DB -dependent. As the figure shows, there are some faults that take place only at $DB=0$, and others only at $DB=1$. Hence, to obtain accurate fault models, it is crucial to incorporate magnetic coupling. Note that the results for the $DB=N$ case can be seen as a ‘middle state’ between two extreme cases involving magnetic coupling.

Fig. 5.7 shows a fault map example for defects involving two cells; it reports the results of the defect $dBCC1$ which is a bridge between the victim cell C_v and C_{ac} being the aggressor cell in the same column (see also the table of Fig. 5.1 (a)). The same notation is used as for Fig. 5.4; the added column ‘State C_{ac} ’ provides the state of the aggressor cell. Note that the DB column now consists of five cases per state of C_{ac} : $DB=N$, and four extreme cases providing magnetic coupling the DB patterns for both NC_v (5 neighbor cells of the victim cell) and NC_a (5 neighbor cells of the aggressor cell); see Fig. 5.3. The ‘-’ indicates that the fault model is DB -independent.

As presented in Fig. 5.7, the longest defect strength range is sensitized only by the $1r1$ operation with the C_{ac} state in ‘1’; the electrical effect (unexpected current path) plays a role in this case. Besides, this fault is independent of the DB pattern, as the magnetic coupling never affects the read operation. On the other hand, while the faults with the

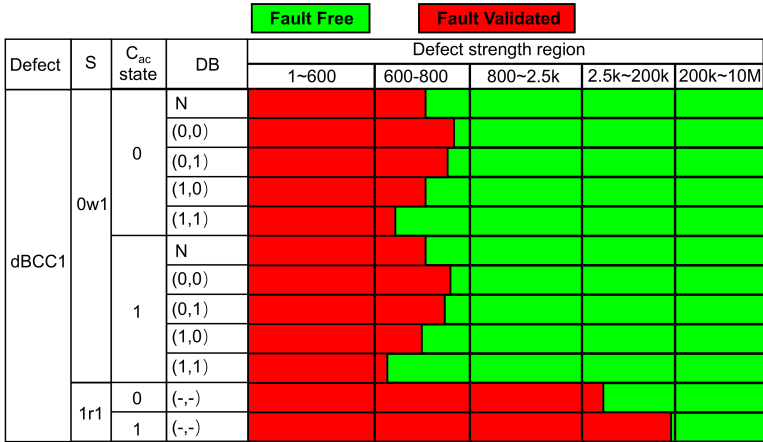


Figure 5.7: Fault map of defect dBCC1.

5

'0w1' operation relate to the magnetic coupling, it is uncertain if the unexpected current path is also involved unless we make additional efforts to analyze the detailed circuit (it is involved in this case). Note that to detect this defect, only the 1r1 operation is required, irrespective of the DB pattern. Hence, the magnetic coupling is not involved in detecting dBCC1. Additionally, we present how to deal with cells C_{s1} and C_{s2} in Fig. 5.3 (d). If read operations sensitize a much longer defect strength range than write operations, we ignore the impact of the two cells. Otherwise, extra simulations are carried out by switching the state of C_{s1} and C_{s2} . Here, our simulation reveals that these two cells have no impact on fault modeling.

FINDINGS

The major findings of this subsection are:

- Compared with the baseline simulation model, no additional faults are observed in this subsection, yet the sensitized defect strength range is affected by these coupling issues.
- For all faults that relate to MTJ switching (either unexpected switching or failing to switch), the sensitized defect strength range is always affected by the magnetic coupling impact, and may be affected by the electrical effect.
- For faults that do not relate to MTJ switching, the sensitized defect strength range may be affected by the electrical effect.

5.2.4. RESULTS FOR INTRINSIC STOCHASTICITY-ONLY SIMULATION MODEL

In this subsection, we apply the model with components of the R-V calculation, the basic switching, and the stochasticity with both high and low voltage, closing the component of the magnetic field impact. This subsection aims to investigate how the intrinsic stochasticity in MTJs affects fault modeling.

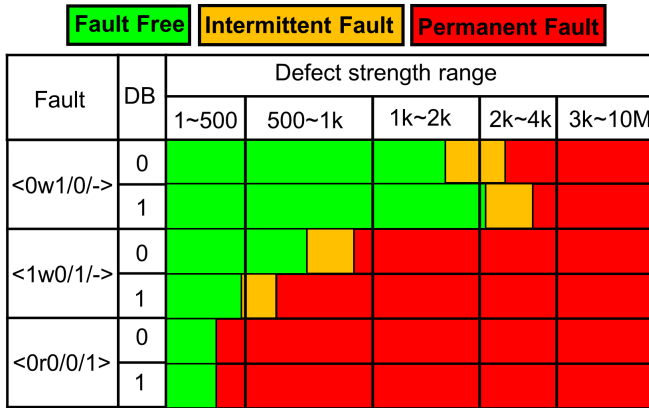


Figure 5.8: Fault map of defect OC2.

STOCHASTICITY AND INTERMITTENT FAULT

Due to this stochasticity, random read/write errors are unavoidable, and the WER and RER of MTJs can never reach zero [41]. Therefore, WER and RER specifications are introduced, i.e., WER_{spec} and RER_{spec} . A common specification for WER is $WER_{spec} < 10^{-3}$ for typical application as a storage memory, and $WER_{spec} < 10^{-9}$ with ECC or $WER_{spec} < 10^{-18}$ without ECC for demanding working memory [41]. A common specification for RER is $RER_{spec} < 10^{-9}$ with ECC and $RER_{spec} < 10^{-18}$ without ECC for demanding working memory [41]. In the presence of a defect, there are three cases [107]:

- Fault-free case where the defect strength is very low, $WER < WER_{spec}$ and $RER < RER_{spec}$.
- Intermittent faults, where defect strength is increased, and $WER_{spec} < WER < 1$ or $RER_{spec} < RER < 1$.
- Permanent faults with defect strength continue to increase, $WER = 1$ or $RER = 1$.

FAULT MAP ANALYSIS

During fault modeling, to assess the WER and RER for MTJs in the presence of defects, we have two methods: 1) perform Monte Carlo simulations for each defect at a certain strength, and 2) calculate RER and WER based on read/write currents (the calculation methods are presented in chapter 3). Considering that performing the Monte Carlo simulation process takes too long a time, method 2) is selected.

Fig. 5.8 presents the fault map of the defect OC2, where the ‘yellow’ box indicates the sensitization of intermittent faults. In the presence of defect OC2, the write current becomes relatively smaller, which does not completely prevent the MTJ switching but causes it to occur intermittently. In this work, the result shows that intermittent faults are observed when the faulty behavior involves MTJ switching. The intermittent fault is described by adding a subscript ‘i’ [86]; for example, $<1w0/1_i/->$ indicates the MTJ state remaining in ‘1’ from time to time in $1w0$ operations.

FINDINGS

The major findings of this subsection are:

- Compared with the subsection of the baseline model, where only permanent faults are sensitized, intermittent faults are sensitized in this subsection.
- Intermittent faults are observed at all faults that relate to MTJ switching.
- When intermittent faults occur, they are always between the fault-free box and the permanent box.
- Applying march tests for intermittent faults results in low time-efficiency and high escape rate; we suggest applying specific DfT to detect these defects.

5.2.5. RESULTS FOR FULL-VARIABILITY SIMULATION MODEL

In this subsection, we open all model components in Fig. 3.1, and show Table 5.1 of all fault maps for all defects. In the table, the fault with i means intermittent faults are observed, and the fault with * means it sensitized the longest defect strength range for this defect. If DB impacts the defect strength range sensitized by a certain fault, the bracket after the fault indicates in which DB this fault sensitizes the longest defect strength range. Similarly, if the state of C_a affects the sensitized defect strength range, it will be described in this bracket. Yet, if the two issues above do not impact the sensitized defect strength range, they will not appear.

5.3. TEST GENERATION

In this section, we introduce test methods designed to detect validated faults that are sensitized by transistor, contact, and interconnect defects in the previous section. Given the differing characteristics of permanent and intermittent faults, we analyze them separately. Permanent faults, which involve certain functionality failing consistently, can be fully detected using March tests, which systematically apply read and write operations to identify faulty behavior. Some DfT methods can be applied to reduce the test time for permanent faults. In contrast, intermittent faults, which occur unpredictably, pose a greater challenge for detection. While some intermittent faults may be detected by March tests, additional DfT techniques may be required to improve fault coverage and/or reduce the test time.

5.3.1. PERMANENT FAULT

For permanent faults, we only need to consider the intra/inter-cell static faults; no dynamic faults are found. For example, in Fig. 5.4, the longest defect strength length is sensitized by the $1w0$ operation with $DB = 0$.

Next, we introduce the ‘DB-Based Integer Linear Programming (DB-ILP)’, which enhances the primary ILP approach by incorporating DB patterns to optimize fault selection and test generation [67]. This method helps in selecting the DBs and faults to optimize March tests. Based on the faults and DBs selected through DB-ILP, the March algorithm is generated with full coverage of defects and low test time. Additionally, we

Table 5.1: Fault map of each defect.

Defect	Fault
OC1	$\langle 1w0/1/- \rangle_i$ (DB=0), $\langle 0w1/0/- \rangle_j$ (DB=1), $\langle 0r0/0/1 \rangle^*$
OC2	$\langle 1w0/1/- \rangle_i$ (DB=0), $\langle 0w1/0/- \rangle_j$ (DB=1), $\langle 0r0/0/1 \rangle^*$
OC3	$\langle 1w0/1/- \rangle_i$ (DB=0), $\langle 0w1/0/- \rangle_j$ (DB=1), $\langle 0r0/0/1 \rangle^*$
oc4	$\langle 1w0/1/- \rangle_i$ (DB=0), $\langle 0w1/0/- \rangle_j$ (DB=1), $\langle 0r0/0/1 \rangle^*$
OB _w	$\langle 1w0/1/- \rangle_i^*$ (DB=0), $\langle 0w1/0/- \rangle_j$ (DB=1)
OB _r	$\langle 0r0/0/1 \rangle^*$
OS _w	$\langle 1w0/1/- \rangle_i^*$ (DB=0), $\langle 0w1/0/- \rangle_j$ (DB=1)
OS _r	$\langle 0r0/0/1 \rangle^*$
OW	$\langle 1w0/1/- \rangle_i$ (DB=0), $\langle 0w1/0/- \rangle_j$ (DB=1), $\langle 0r0/0/1 \rangle^*$
SC1	$\langle 1w1/0/- \rangle_i$ (DB=1), $\langle 1w0/1/- \rangle_j$ (DB=0), $\langle 0r0/0/1 \rangle^*$, $\langle 0r0/1/1 \rangle_i$ (DB=0)
SC2	$\langle 1w0/1/- \rangle_i$ (DB=0), $\langle 0w1/0/- \rangle_j$ (DB=1), $\langle 1r1/1/0 \rangle^*$, $\langle 1r1/0/0 \rangle_i$ (DB=1), $\langle 0w0/1/1/- \rangle_j$ (DB=0)
SS1	$\langle 1w0/1/- \rangle_i$ (DB=0), $\langle 0r0/0/1 \rangle^*$, $\langle 0r0/1/1 \rangle_i$ (DB=0)
SS2	$\langle 0w1/0/- \rangle_j$ (DB=1), $\langle 1r1/1/0 \rangle^*$, $\langle 1r1/0/0 \rangle_i$ (DB=1)
SB1	$\langle 0w1/0/- \rangle_j$ (DB=1), $\langle 0r0/0/1 \rangle^*$, $\langle 0r0/1/1 \rangle_i$ (DB=0)
SB2	$\langle 1w0/1/- \rangle_i$ (DB=0), $\langle 1r1/1/0 \rangle^*$, $\langle 1r1/0/0 \rangle_i$ (DB=1)
SW1	$\langle 1w1/0/- \rangle_i$ (DB=1), $\langle 0w1/0/- \rangle_j$ (DB=1), $\langle 1r1/1/0 \rangle^*$, $\langle 1r1/0/0 \rangle_i$ (DB=1)
SW2	$\langle 1w0/1/- \rangle_i$ (DB=0), $\langle 0w1/0/- \rangle_j$ (DB=1), $\langle 0r0/0/1 \rangle^*$, $\langle 0r0/1/1 \rangle_i$ (DB=0)
BC1	$\langle 1w0/1/- \rangle_i$ (DB=(0,0), C _{ar} =0), $\langle 0w1/0/- \rangle_j$ (DB=(1,1), C _{ar} =1), $\langle 0r0/0/1 \rangle^*$ (C _{ar} =1)
BC2	$\langle 1w0/1/- \rangle_i$ (DB=(0,0), C _{ar} =0), $\langle 0w1/0/- \rangle_j$ (DB=(1,1), C _{ar} =1), $\langle 0r0/0/1 \rangle^*$ (C _{ar} =0), $\langle 1r1/1/0 \rangle^*$ (C _{ar} =1),
BC3	$\langle 1w0/1/- \rangle_i$ (DB=(0,0), C _{ar} =0), $\langle 0w1/0/- \rangle_j$ (DB=(1,1), C _{ar} =1), $\langle 1r1/1/0 \rangle^*$ (C _{ar} =1), $\langle 1r1/0/0 \rangle_i$ (DB=(1,1), C _{ar} =1)
BC4	$\langle 1w0/1/- \rangle_i$ (DB=(0,0), C _{ar} =0), $\langle 0w1/0/- \rangle_j$ (DB=(1,1), C _{ar} =1), $\langle 0r0/0/1 \rangle^*$ (C _{ar} =1), $\langle 0r0/1/1 \rangle_i$ (DB=(0,0), C _{ar} =0)
BC5	$\langle 0w1/0/- \rangle_j$ (DB=(1,1), C _{ar} =1), $\langle 1w1/0/- \rangle_i$ (DB=(1,1), C _{ar} =1), $\langle 1r1/1/0 \rangle^*$ (C _{ar} =1), $\langle 1r1/0/0 \rangle_i$ (DB=(1,1), C _{ar} =1)
BC6	$\langle 1w0/1/- \rangle_i$ (DB=(0,0), C _{ar} =0), $\langle 0w1/0/- \rangle_j$ (DB=(1,1), C _{ar} =1), $\langle 1r1/1/0 \rangle^*$ (C _{ar} =1), $\langle 1r1/0/0 \rangle_i$ (DB=(1,1), C _{ar} =1)
dBCC1	$\langle 0r0/0/1 \rangle^*$
dBCC2	$\langle 0r0/0/1 \rangle^*$
dBCC3	$\langle 0r0/0/1 \rangle^*$
dBCC4	$\langle 0r0/0/1 \rangle^*$
dBCC5	$\langle 0r0/0/1 \rangle^*$
dBCC6	$\langle 0r0/0/1 \rangle^*$
dBCC7	$\langle 1w0/1/- \rangle_i$ (DB=(0,0), C _{ad} =0), $\langle 0w1/0/- \rangle_j$ (DB=(1,1), C _{ad} =1), $\langle 0r0/0/1 \rangle^*$
dBCC8	$\langle 0w1/0/- \rangle_j$ (DB=(1,1), C _{ad} =1), $\langle 1w1/0/- \rangle_i$ (DB=(1,1), C _{ad} =1), $\langle 1r1/1/0 \rangle^*$, $\langle 0r0/0/1 \rangle^*$
dBCC9	$\langle 0r0/0/1 \rangle^*$
dBCC10	$\langle 0r0/0/1 \rangle^*$
dBCC11	$\langle 0w1/0/- \rangle_j$ (DB=(1,1), C _{ad} =1), $\langle 1w1/0/- \rangle_i$ (DB=(1,1), $\langle 1r1/1/0 \rangle^*$), $\langle 0r0/0/1 \rangle^*$
cBCC1	$\langle 0w1/0/- \rangle_j$ (DB=(1,1), C _{ac} =1), $\langle 1w1/0/- \rangle_i$ (DB=(1,1), C _{ac} =1), $\langle 0r0/0/1 \rangle^*$
cBCC2	$\langle 1w0/1/- \rangle_i$ (DB=(0,0), C _{ac} =0), $\langle 1w1/0/- \rangle_j$ (DB=(1,1), C _{ac} =1), $\langle 0w1/0/- \rangle_i$ (DB=(1,1), C _{ac} =1), $\langle 1r1/0/1 \rangle_i$ (DB=(1,1), C _{ac} =1), $\langle 1r1/0/1 \rangle^*$ (C _{ac} =1)
rBCC1	$\langle 0r0/0/1 \rangle^*$ (C _{ar} =1)
rBCC2	$\langle 0r0/0/1 \rangle^*$ (C _{ar} =1)
Notation	$_i$: intermittent faults; * : longest defect strength range sensitized; DB: DB=(N _{cv}) or DB=(N _{cv} ,N _{ca}) in Fig. 5.3

introduce DfT methods to further reduce test time, enhancing the overall efficiency of the testing process while maintaining the same defect coverage.

DATA BACKGROUND BASED INTEGER LINEAR PROGRAMMING

To obtain the optimal test solution, the DB-ILP is applied. The DB-ILP is considered a generalized method that offers test solutions to STT-MRAM testing with DB patterns. The target of the DB-ILP method is to generate a March test algorithm that can detect all validated faults while minimizing the March test length, by selecting the appropriate DB patterns and faults. This is achieved through mathematical optimization with two steps: 1) Establish the DB-ILP table with the data of the fault modeling; 2) Formulate and solve the DB-ILP equation.

The example of the DB-ILP table is presented in Table 5.2. On the top of the table, it presents all the possible DB patterns (DB_m for $m \in [1, M]$), and the fault ($F_{m,n}$ for $m \in$

Table 5.2: Example to DB-ILP table.

		DB pattern, DB_m , for $m:1 \rightarrow M$												$\sum_{m=1}^M \sum_{n=1}^N a_{k,l,m,n}$		
		DB ₁				DB ₂				DB _M						
		$F_{1,n}$, for $n:1 \rightarrow N$				$F_{2,n}$, for $n:1 \rightarrow N$				$F_{M,n}$, for $n:1 \rightarrow N$						
		$F_{1,1}$	$F_{1,2}$...	$F_{1,N}$	$F_{2,1}$	$F_{2,2}$...	$F_{2,N}$...	$F_{K,1}$	$F_{K,2}$...		$F_{K,N}$	
D_1	DS_1	1	$a_{1,1,1,1}$	$a_{1,1,1,2}$...	$a_{1,1,1,N}$	$a_{1,1,2,1}$	$a_{1,1,2,2}$...	$a_{1,1,2,N}$...	$a_{1,1,M,1}$	$a_{1,1,M,2}$...	$a_{1,1,M,N}$	
	DS_2	10	$a_{1,2,1,1}$	$a_{1,2,1,2}$...	$a_{1,2,1,N}$	$a_{1,2,2,1}$	$a_{1,2,2,2}$...	$a_{1,2,2,N}$...	$a_{1,2,M,1}$	$a_{1,2,M,2}$...	$a_{1,2,M,N}$	
	DS_L	100M	$a_{1,L,1,1}$	$a_{1,L,1,2}$...	$a_{1,L,1,N}$	$a_{1,L,2,1}$	$a_{1,L,2,2}$...	$a_{1,L,2,N}$...	$a_{1,L,M,1}$	$a_{1,L,M,2}$...	$a_{1,L,M,N}$	
D_K	DS_1	1	$a_{K,1,1,1}$	$a_{K,1,1,2}$...	$a_{K,1,1,N}$	$a_{K,1,2,1}$	$a_{K,1,2,2}$...	$a_{K,1,2,N}$...	$a_{K,1,M,1}$	$a_{K,1,M,2}$...	$a_{K,1,M,N}$	
	DS_2	10	$a_{K,2,1,1}$	$a_{K,2,1,2}$...	$a_{K,2,1,N}$	$a_{K,2,2,1}$	$a_{K,2,2,2}$...	$a_{K,2,2,N}$...	$a_{K,2,M,1}$	$a_{K,2,M,2}$...	$a_{K,2,M,N}$	
	DS_L	100M	$a_{K,L,1,1}$	$a_{K,L,1,2}$...	$a_{K,L,1,N}$	$a_{K,L,2,1}$	$a_{K,L,2,2}$...	$a_{K,L,2,N}$...	$a_{K,L,M,1}$	$a_{K,L,M,2}$...	$a_{K,L,M,N}$	

$[1, M]$ and $n \in [1, N]$). On the left of the table, it presents the defect strength (D_k for $k \in [1, K]$) in each defect (DS_l for $l \in [1, L]$). In the center of the table, a four-dimensional binary matrix $a_{k,l,m,n} \in A$ is defined, where 'k' labels the defect name, 'l' labels the defect strength, 'm' labels the DB pattern, and 'n' labels the faults. The total number of $a_{k,l,m,n}$ is $K \times L \times M \times N$. When the defect 'k' with strength 'l' can be sensitized by fault 'n' under DB pattern 'm', $a_{k,l,m,n} = 1$, otherwise $a_{k,l,m,n} = 0$. On the right of the table, it presents the sum of $a_{k,l,m,n}$ in this row ($\sum_{m=1}^M \sum_{n=1}^N a_{k,l,m,n}$). We populate this DB-ILP table with the fault modeling information.

The optimization process of selecting proper DB patterns and faults while minimizing the march can be mathematically denoted as the DB-ILP equation:

$$\min \sum_{m=1}^M \left(\beta \cdot DB(\text{sel})_m \cdot \sum_{n=1}^N \cdot F(\text{sel})_{m,n} \right)$$

$$\text{s.t.} \begin{cases} \text{For } (k: 1 \text{ to } K, l: 1 \text{ to } L): \\ \text{if } \sum_{m=1}^M \sum_{n=1}^N a_{k,l,m,n} \geq 1: \\ \sum_{n=1}^M \sum_{n=1}^N a_{k,l,m,n} \cdot DB(\text{sel})_m \cdot F(\text{sel})_{m,n} \geq 1. \end{cases}$$

The DB-ILP equation consists of two elements: the minimization statement and the constraint. The minimization statement (first line of the equation) guarantees that appropriate DB patterns and their corresponding faults are selected to achieve the minimal cost. Here, $DB(\text{sel})_m$ and $F(\text{sel})_{m,n}$ are binary values (be either '0' or '1'), indicating whether DB_m and $S_{m,n}$ is selected in the test (i.e., '1' is selected). The β represents the cost ratio between changing the DB pattern and adding one more fault, which is mainly determined by the circuit design and the size of the array. Here we simply assume $\beta = 100$, to make reducing the number of DB patterns a priority. The second statement (the constraint) guarantees that all validated faults are sensitized. The 'For' loop checks every D_k with DS_l , if it is detectable by regular March test methods (i.e., $\sum_{m=1}^M \sum_{n=1}^N a_{k,l,m,n} > 0$), it must be detected by the $DB(\text{sel})_m$ and $F(\text{sel})_{m,n}$. Python3's PuLP optimization package is applied to solve the DB-ILP equation [157]. Notice that there may exist DB-ILP equations with multiple solutions that have the same cost. Choosing which solution depends on the actual situation.

TEST PATTERN GENERATION

The target of this DfT is to reduce the test time while keeping the same full defect coverage. Next, we show how to apply the DB-ILP. First, we populate the date of fault maps into Table 5.2 As an example, we consider the defect OS_w (see Fig. 5.4), and assume D_1 refers to OS_w . The defect strength DS_l is defined by fault modeling methodology; in this work DS_l ranges from 1 to 10 M Ω . There are two DBs in Fig. 5.4, hence $M = 2$, $BD_1 = 0$, and $BD_2 = 1$. Eight faults are considered in the fault modeling methodology, implying $N = 8$. If faults are sensitized by the fault at the certain defect strength and DB (e.g., $F = \langle 1w0/1/- \rangle$, $BD = 1$, $DS = 500\Omega$ for OS_w), the related $a_{k,l,m,n} = 1$; otherwise, $a_{k,l,m,n} = 0$.

Here we classify all defects into two groups: 1) open and short defects, where two DBs are applied, and 2) bridge defects, where four DBs are applied. The same process as done for OS_w is carried out for defects in each group separately, and the table is generated for each group. Eq. 5.3.1.1 is applied for each table.

For the group of open and short defects, there are 17 defects, and $L = 17$. The output of this group is presented as follows:

Table 5.3: Selected DB and F for open and short defects.

DB(sel)	F(sel)
DB=0	$\langle 1w0/1/- \rangle$, $\langle 0r0/0/1 \rangle$, $\langle 1r1/1/0 \rangle$

For the group of open and short defects, there are 23 defects, and $L = 23$. For this case, the DB(sel) is in the form of (Ca state, NC_v , NC_a); hence, there are 8 DB situations. The output of this group is presented as follows:

Table 5.4: Selected DB and S for bridge defects.

C_i	DB(sel)	F(sel)
C_{ac}	(1, -, -)	$\langle 0r0/0/1 \rangle$, $\langle 1r1/1/0 \rangle$
C_{ar}	(1, -, -)	$\langle 0r0/0/1 \rangle$, $\langle 1r1/1/0 \rangle$
C_{ad}	(-, -, -)	$\langle 0r0/0/1 \rangle$

To obtain the optimal March test algorithm, we need to analyze the outputs of the DB-ILP first. Sensitizing all faults when defects within one cell require the DB pattern to be $NC_v = 0$, due to the magnetic coupling effect. On the other hand, sensitizing all faults when defects are between cells requires setting the C_{ac} and C_{ar} to be '1', regardless of the DB pattern (NC_v , NC_a); this implies that the read operation always sensitizes the longest defect strength range and we can ignore the magnetic coupling impact, like what has been shown in Fig. 5.7. Considering all the DB-ILP method outputs, we can simply apply two solid DB patterns: all cell states in '0' to include the magnetic coupling, and all cell states in '1' to include the electrical effect. Notice that this is just for our case; the STT-MRAM test design depends on the actual circuit design, the device performance, the stray field intensity, and so on. Once the write operation sensitizes the longest defect strength range when defects are between cells, we need more complex DB patterns. The March test algorithm is as follows:

$$\text{March-MRAM-1} = \{ \Downarrow (w0); \Downarrow (r0, w1, r1, w0, r0); \\ \Downarrow (w1); \Downarrow (w0, r0, w1); \Downarrow (r1) \}. \quad (5.1)$$

Here, the ‘ \Downarrow ’ indicates that the addressing direction is irrelevant. The first march element initializes the STT-MRAM array to be the first solid DB pattern with all cells in ‘0’. The second march element firstly checks whether the initialization has succeeded, to avoid the previously operated cell state disturbing the operations on the later cell and leading to escapes. Then, operations are applied to sensitize all validated faults when defects are within one cell. The third march element initializes the STT-MRAM array to be the second solid DB pattern with all cells in ‘1’. The fourth march element firstly checks the second initialization, then applies operations to sensitize all validated faults when defects are between cells. The final march element checks the final state of the device. The algorithm length is $11N$. It is verified that this March test algorithm can detect all validated faults by Spice simulation.

DESIGN FOR TEST METHODOLOGIES

The DfT method, as an alternative to the conventional March tests, is discussed to detect conventional defects. The target of the DfT is to reduce the test time (i.e., reduce the march algorithm length) while keeping the same full fault coverage. As presented in Eq. 3.15, there are three components independently affecting the MTJ switching: the STT effect, the magnetic field impact, and the H_k impact. The magnetic field impact further consists of the stray field impact and the external magnetic field impact. The march algorithm in Eq. 5.1 is lengthened due to the stray field impact (i.e., magnetic coupling). This impact is possibly compensated (i.e., balanced) by adjusting the other components and reducing the algorithm length of Eq. 5.1. To adjust the STT effect, we can modify the write pulse width or height. To adjust the magnetic field impact, we can apply the external magnetic field by the specific Automatic Testing Equipment (ATE). The H_k impact is not adjustable, since H_k is the material property. Next, we analyze the feasibility of these methods:

- **Apply external magnetic field:** Applying a proper external magnetic field that exactly equals the stray field yet in a reversed direction can perfectly compensate for the stray field. The ATE, which can perform regular write/read-based tests and apply the external magnetic field simultaneously, is still immature [127]. However, the time required to switch between different magnetic fields (e.g., several microseconds according to [127]) is much longer compared to the time of write/read operations (e.g., 10 ns in this work). For example, if Eq. 5.1 can be adjusted to:

$$\text{March} = \{ \Downarrow (w0_{H1}); \Downarrow (r0, w1_{H2}, r1, w0_{H1}, r0) \}. \quad (5.2)$$

In the second march element, it performs $w0$ with the external magnetic field H_1 , and performs $w1$ with H_2 . The $w1_{H2}$, $r1$, $w0_{H1}$ are performed in sequence, each operation takes 10 ns. However, the time of switching H_1 to H_2 requires μs , which the test time. Therefore, we do not consider this method in this paper. Since both methods detect the longest strength range for each defect, they have the same defect coverage.

- **Adjust pulse height:** This is not a desired way to compensate for the stray field, since the defect also impacts the current flow through the MTJ. Therefore, it is difficult to evaluate how the adjusted pulse height will affect the write current in the presence of defects.
- **Adjust pulse width:** This method is a perfect way to balance the stray field. The pulse width is always constant and not affected by the defect. Besides, modifying the pulse width can be easily realized by applying weak write DfT circuits in [158], [159], which will not be described here.

Subsequently, we will demonstrate that by applying different write pulse widths, the stray field impact can be exactly compensated, ensuring the same operation sensitizes the same defect strength in different DBs. However, to reduce the cost of circuit designing, it is critical to configure the DfT with fewer types of write pulse widths. Next, we will illustrate the DfT configuration process.

Here, we first configure the DfT for the defect OS_w as an example, then extend the process for all defects. Fig. 5.9 (a) presents the original fault map for the defect OS_w , in which the case of NMC is not shown as the magnetic coupling is inevitable in reality. We assume that $1w0$ is applied to the MTJ cell in the presence of defect OS_w , considering two DBs with a specific defect strength: 1) in DB=1, the fault is exactly sensitized, and the input torque $T_{DB=1} = E_B$; and 2) in DB=0, no fault is sensitized, and the input torque $T_{DB=0} > E_B$ (i.e., a successful switching). By applying Eq. 3.15 to the two cases above, $T_{DB=1}$ and $T_{DB=0}$ are presented as:

$$\begin{aligned}
 T_{DB=0} &> E_B = t_p \cdot (H_k \cdot \epsilon + I \cdot \eta + \epsilon \cdot (H_{s-inter-PL} + H_{s-intra} + H_{s-inter-FL})) \\
 T_{DB=1} &= E_B = t_p \cdot (H_k \cdot \epsilon + I \cdot \eta + \epsilon \cdot (H_{s-inter-PL} + H_{s-intra} - H_{s-inter-FL}))
 \end{aligned} \tag{5.3}$$

where $\epsilon = \frac{\alpha \cdot e \cdot M_s \cdot A \cdot t_{FL}}{\hbar}$

Notice that, with the same defect strength, the current I through the MTJ is the same. Yet different defects or different defect strengths will affect I . Subtracting $T_{DB=1}$ and $T_{DB=0}$, we obtain the gap (T_{gap}) between the input torque of the two cases:

$$T_{gap} = |2 \cdot H_{s-inter-FL} \cdot \epsilon| \cdot t_p \tag{5.4}$$

It is found that T_{gap} is proportional to the stray field intensity. The larger $H_{s-inter-FL}$ (e.g., smaller pitch) results in a larger T_{gap} , implying a more significant impact of magnetic coupling on the testing. The target of the DfT is to exactly compensate for the magnetic coupling, thereby reducing T_{gap} to 0. This can be achieved by adjusting the write pulse width t'_p in DB=0 while keeping the original write pulse width t_p in DB=1. The input torque in these two cases is re-calculated and presented as:

$$\begin{aligned}
T_{DB=0} &= E_B = t'_p \cdot (H_k \cdot \epsilon + I \cdot \eta + \epsilon \cdot \\
&\quad (H_{s-inter-PL} + H_{s-intra} - H_{s-inter-FL})) \\
T_{DB=1} &= E_B = t_p \cdot (H_k \cdot \epsilon + I \cdot \eta + \epsilon \cdot \\
&\quad (H_{s-inter-PL} + H_{s-intra} + H_{s-inter-FL}))
\end{aligned} \tag{5.5}$$

The t'_p can be calculated by the two equations above:

$$t'_p = t_p \cdot \frac{E_B}{E_B + |2 \cdot H_{s-inter-FL} \cdot t_p \cdot \epsilon|} \tag{5.6}$$

In our case, t'_p is 14% lower than the regular t_p . By applying $1w0$ with t'_p in DB=0 and t_p in DB=1, T_{gap} is always 0. Therefore, the magnetic coupling is exactly compensated, and the same defect strength range is sensitized in different DBs. We re-perform fault modeling with the calculated t'_p in DB=0, and Fig. 5.9 (b) presents the new fault map. It is shown that the $1w0$ operation sensitizes the same longest defect strength range under both DB=1 and DB=0. A similar process can be done for the $0w1$ operation, and the obtained fault models are also presented in Fig. 5.9 (b).

It is worth noticing that in Eq. 5.6, all the parameters to calculate t'_p are constant, and especially I is not involved. Since interconnect and contact defects only affect I while not affecting other parameters in Eq. 5.6, the defect type or the defect strength has no impact on the t'_p value. If we apply the same process from Eq. 5.3 to 5.6 for other defects, the calculation of t'_p is always the same. We perform the fault modeling for all interconnect and contact defects, and the result verifies that $1w0$ always sensitizes the same defect strength under with t'_p in DB=0 and t_p in DB=1 (details not shown here to save space). In summary, by a proper DfT configuration shown in this section, it is demonstrated that we only require one additional t_p for the DfT.

Next, we generate the test solution. First, we select the proper DB based on Eq. 5.1, which considers both the stray field impact and the electrical effect. Since the stray field impact can be exactly compensated by adjusting the write pulse width, we only need to consider the electrical effect. The DB of DB=1 in Table 5.4 is selected solely due to the magnetic coupling, hence we only need to include the selected sequences in the final test solution, yet ignore the the DB. On the other hand, the DB of DB=0 in Table 5.4 is selected due to the electrical effect, therefore both the DB and the S need to be considered. Combining the two tables, we obtain:

Table 5.5: Selected DB and S in the DfT.

DB(sel)	F(sel)
DB=0	$1w0$, $0r0$, $1r1$

Secondly, we design the pulse width for different write operations. Since we selected the DB of DB=1, $0w1$ always sensitizes the longer defect strength range. On the other hand, $1w0$ always sensitizes the shorter defect strength range (as described in the last section). Therefore, we keep the original pulse width t_p for $0w1$, and adjust the pulse

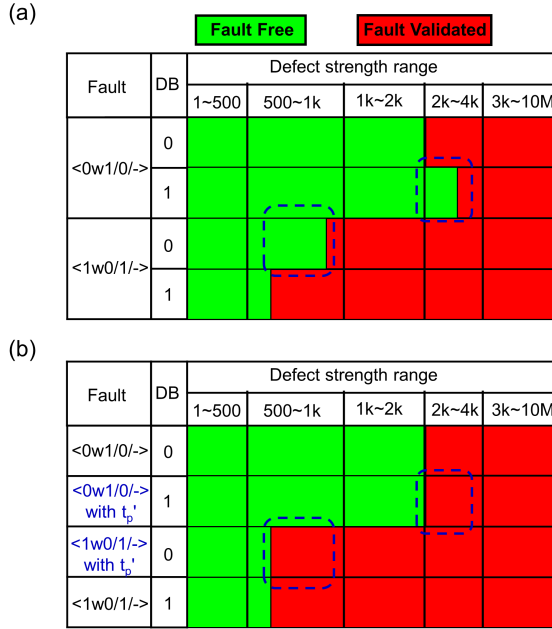


Figure 5.9: DfT application for defect OS_w ; (a) Regular fault map, (b) Fault map by modified write pulse width.

width to t'_p for $1w0$, where t'_p is presented in Eq. 5.6. Finally, we design the march algorithm based on Table 5.5, presented as:

$$March - MRAM - 2 = \{\uparrow (w0^l); \uparrow (r0, w1, r1, w0^l, r0)\}. \quad (5.7)$$

Here, $w0^l$ applies a shorter write pulse width t'_p , and $w1$ applies the original pulse width t_p . The first element initializes all cells to state '0'. The second element applies write or read operations (i.e., selected sequences in Table 5.5) to sensitize faults. It is verified by Spice simulation that this march algorithm can detect all validated faults as Eq. 5.1. The new march algorithm has the same effect as Eq. 5.1.

The regular march algorithm is described in Eq. 5.1, and the DfT test algorithm is described in Eq. 5.7. Next, we compare the two test methods regarding defect coverage, test time, and cost:

- **Defect coverage:** Eq. 5.1 selects the proper DB to sensitize the longest strength range for each defect. Eq. 5.1 applies the DfT method to guarantee that in different DBs, the longest defect strength range is always sensitized. Since both methods detect the longest strength range for each defect, they have the same defect coverage.
- **Test time:** The march algorithm length of Eq. 5.1 is $11N$, while that of Eq. 5.7 is $6N$. If ignoring the difference between t_p and t'_p , the time-efficiency is improved

by $(11 - 6)/11 = 45\%$. The reduced t'_p can further reduce the test time for the DfT method. Hence, the DfT method has a shorter test time.

- **Cost:** The regular march test has no additional cost. However, the DfT requires additional area for circuits to properly adjust the pulse width (e.g., the weak write scheme in [158]).

5.3.2. INTERMITTENT FAULT

Designing tests for intermittent faults presents challenges due to their unpredictable nature. A natural starting point is the March test, which is widely used for detecting permanent faults. However, applying conventional March tests to intermittent faults often results in a high escape rate, as these faults may not be consistently sensitized by regular read/write operations. To address this limitation, DfT techniques are explored. By incorporating specialized DfT methods, such as dedicated circuits, the detection of intermittent faults can be significantly improved.

MARCH TEST

The DB-ILP method is performed again to select the proper faults, and the march test is designed. The details are not shown here. The optimized March algorithm is presented as follows:

$$\text{March-MRAM-2} = \{\uparrow(w1); \uparrow(w0, r0, w1)^n\}. \quad (5.8)$$

Only one DB is required and initialized by the first march element. The second element is repeated 'n' times due to the intermittent nature of these faults. Even with repetition, the escape and yield loss cannot be avoided.

Let us assume that we apply $1w0$ for n times to a defective cell, and the defect of strength $R_d=R_1$ causes an intermittent fault with $WER=WER(R_d)$. Then the Fault Sensitization Probability (FSP) can be estimated as:

$$FSP(R_1) = 1 - (1 - WER(R_1))^n \quad (5.9)$$

We assume that the defect causes intermittent faults for the range $R_1 \leq R_d < R_2$, then the total FSP for all intermittent faults due to the defect can be obtained by applying an integration as:

$$FSP = \frac{\int_{R_1}^{R_2} 1 - (1 - WER(R_d))^n dR_d}{\int_{R_1}^{R_2} 1 dR_d} \quad (5.10)$$

$WER(R_d)$ is the function of how WER changes with R_d . For example, for the fault $<0w1/0/->$ in DB=1, $R_1=1.75\text{ k}\Omega$ and $R_2=2.1\text{ k}\Omega$ for $1w0$ can be extracted from Fig. 5.8 (b). The calculation will give $FSP=1.3\%$ for $n=1$.

Next, we will calculate Fault Detection Probability (FDP) assuming the application of Eq. 5.8; the focus here is only on intermittent faults. We assume the same defect of strength $R_d=R_1$, which causes faults with $WER=WER_{0w1}(R_1)$ for $0w1$, and $WER_{1w0}(R_1)$ for $1w0$, then:

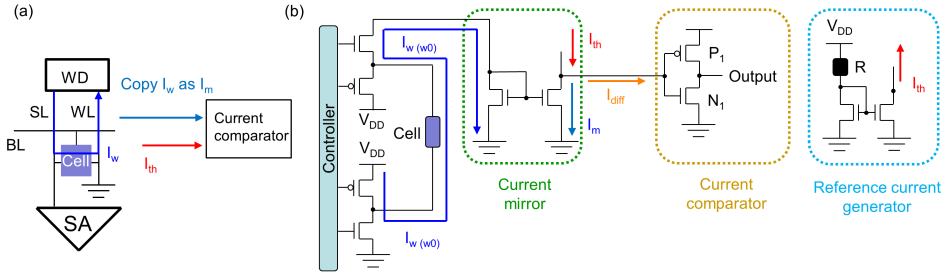


Figure 5.10: (a) DfT methodology; (b) Circuit design of the DfT.

$$FDP(R_1) = 1 - [(1 - WER_{1w0}(R_1))(1 - WER_{0w1}(R_1))]^n \quad (5.11)$$

Similar to Eq. 5.10, FPD for the whole range of the defect strength, being only able to sensitize intermittent faults for the targeted defect, can be derived:

$$FDP = \frac{\int_{R_1}^{R_2} 1 - [(1 - WER_{1w0}(R))(1 - WER_{0w1}(R))]^n dR}{\int_{R_1}^{R_2} 1 dR} \quad (5.12)$$

The Escape Probability (EP) of the above defect can then be calculated as: $EP=1-FDP$. If we apply the march test of Eq. 5.8, $EP=99.82\%$ for $n=1$, and $EP=91\%$ for $n=1000$. Even with a very high n , EP is still very high.

Finally, we performed circuit simulations to verify the above estimation. E.g., for the same defect, we selected 15 different values of R_d within the range $[0.9\text{k}\Omega, 4.5\text{k}\Omega]$. The test of Eq. 5.8 is simulated in the presence of each of the 15 R_d values, and for each R_d value, we perform the test for $n=1000$ times. We assumed to have 1000 similar defective chips of each of the 15 R_d values (resulting in the representation of 15000 defective chips in total); hence, we repeated the process of applying the test with $n=1000$ for 15000 times. The fault is assumed to be detected if any read operation within the test returns a wrong value. The simulation results show that only in 2133 simulations (among the total of $15 * 1000$) the fault was detected (about 14.2%), which is quite in line with our estimation.

DESIGN FOR TEST METHODOLOGIES

A common DfT methodology to detect intermittent faults is by monitoring the read/write current [146]–[150]. Here, we show a DfT example by monitoring the write current.

Fig. 5.10 (a) presents the concept of the proposed DfT, which is based on: a) copying/mirroring the write current I_w to I_m , and b) comparing the copied current with a reference current I_{th} which is the write current derived at the borderline of ‘fault-free case’ and ‘intermittent case’. I_{th} is a constant value, since the relationship between WER and I_w is constant for defect-free MTJs. Any defect leading to $I_w < I_{th}$ suggests the cell is faulty.

It is important to note that there are four options to select I_{th} , i.e., based on I_w for $0w1$, $1w0$, $0w0$, or $1w1$. Selecting I_{th} for $0w1$ or $1w0$ will not do the targeted job; be-

cause of the intermittent behavior of the cell, the MTJ may or may not switch, hence I_w will change accordingly. As a consequence, comparing the changing I_w with a reference I_{th} will not work. In addition, we need to select the comparison of I_w for $1w1$ or for $0w0$; the idea is to select the case resulting in the maximal current deviation ΔI_w (in the presence of the defect) as compared with I_{th} . The differences in the current in the two cases can be given as:

$$\Delta I_w(0w0) \propto \frac{V_{DD}}{R_{M_low}} - \frac{V_{DD}}{R_{M_low} + R_d} \quad (5.13)$$

$$\Delta I_w(1w1) \propto \frac{V_{DD}}{R_{M_high}} - \frac{V_{DD}}{R_{M_high} + R_d} \quad (5.14)$$

Here, R_{M_low} and R_{M_high} refer to the low resistance, respectively, high resistance states of the MTJ cell. The analysis of the above equations reveals that the maximum current deviation is always obtained for the case of $0w0$ for lower R_d ; hence this will be used for the DfT. The cell should be initialized to 0 before applying the DfT.

Fig. 5.10 (b) presents the DfT circuit design consisting of three parts: the current mirror, the current comparator, and the reference current generator. At time 0 (i.e., the start of the testing), the current mirror copies I_w to I_m , and the reference current generator provides I_{th} . Then $I_{diff} = I_{th} - I_m$ is generated; it can be either positive or negative. It will drive the input of the current comparator (i.e., inverter consists of MOSFETs P_1 and N_1). If $I_m > I_{th}$, then $I_{diff} < 0$, and the inverter reports '0' at 'Output' (fault-free). Conversely, if $I_m < I_{th}$, then $I_{diff} > 0$, and the 'Output' of the inverter reports '1' (fault detected).

The DfT is simulated while injecting different R_d values for our case study. The results confirm the superiority of the DfT in guaranteeing the detection of the targeted intermittent faults. Moreover, the proposed DfT also detects all permanent faults. Hence, it covers both permanent and intermittent faults!

The proposed DfT is superior compared with any functional march test (even with repeating march elements). The fault coverage of the march test is just a subset of that of DfT. Hence, the proposed DfT offers an outstanding alternative to increase fault coverage by considering intermittent faults. Although we cannot claim that all intermittent faults in the STT-MRAM memory can be detected, the proposed DfT has added value as it also covers intermittent faults and permanent faults that are caused by defects in *the write path*. A similar philosophy can be used to develop another DfT to cover intermittent faults caused by defects in the read path.

Nevertheless, the DfT comes at some additional cost; it requires only 7 MOSFETs per column and some minor design effort. The larger the size of the memory array (in terms of rows), the smaller the overall area overhead.

6

TESTING OF MTJ DEFECTS IN STT-MRAMs

Testing MTJ (i.e., unique) defects in STT-MRAMs, such as pinhole, Synthetic Anti-Ferromagnet Flip (SAFF), Intermediate State (IM), and Back-hopping (BH), is challenging due to the unique behaviors of MTJs in the presence of these defects. For example, the IM defect can cause the MTJ to present an undefined resistance state. These defects cannot be modeled as linear resistors, and hence, traditional test methods struggle to detect such defects effectively. To address this issue, the Device-aware Test (DAT) is applied, which incorporates the physical impact of these defects into the MTJ compact model and designs a specific defective MTJ model. DAT has been demonstrated to be highly effective in detecting MTJ defects. This chapter provides an overview of unique defect testing and extends the DAT methodology to the BH defect by developing a BH-defective MTJ model, performing fault modeling, and proposing dedicated test solutions.

6.1. OVERVIEW OF DEFECTS IN MTJS

In the presence of MTJ defects, MTJs behave irregularly (e.g., intermittently entering an undefined ('U') state in the presence of IM). Thus, conventional linear resistor-based models cannot accurately model unique defects. To address this, the DAT method is applied to design specific defective MTJ models, to further derive accurate fault models and dedicated test solutions. In this section, we review previous studies on testing MTJ defects, i.e., pinhole, SAFE, and IM [66], [81], [87].

6.1.1. TESTING OF PINHOLES

Pinhole defects occur due to imperfections in the tunnel barrier, forming unintended conductive pathways that degrade electrical properties involving localized short circuits in the MgO layer caused by non-uniform deposition, contamination, or diffusion. The impact of the pinhole is as follows [66]:

- Lower resistance.
- Reduced TMR ratio.
- Reliability issues: the vulnerable MgO makes the MTJ easy to break down.

FAULT MODELING

Fault types sensitized by pinholes rely on the defect strength:

- High strength (i.e., large pinhole): EtD faults $\langle 1w1/0/- \rangle$, $\langle 0w1/0/- \rangle$, $\langle 1r1/0/0 \rangle$, $\langle 1w1/L/- \rangle$, $\langle 0w1/L/- \rangle$, $\langle 1r1/L/0 \rangle$.
- Medium strength (i.e., middle pinhole): HtD easy to detect faults $\langle 1w1/U/- \rangle$, $\langle 0w1/U/- \rangle$, $\langle 1r1/U/? \rangle$.
- Low strength (i.e., small pinhole): no fault sensitized with small pinholes, yet they cause reliability issues. The TB in MTJs with a minor pinhole is more vulnerable, hence easier to break down.

TEST METHODOLOGY

March test: To detect high-strength pinholes, the following March algorithm is applied [66]:

$$\text{March - pinhole - EtD} = \{\uparrow(w1); \uparrow(r1)\} \quad (6.1)$$

To detect medium-strength pinholes, we repeat $r1$ operations for 'n' times with the hope of sensitizing the fault $\langle 1r1/U/? \rangle$:

$$\text{March - pinhole - HtD} = \{\uparrow(w1); \uparrow(r1)^n\} \quad (6.2)$$

To detect low-strength pinholes, we repeat $w1$ operations for 'n' times to degrade TB, and lower the MTJ resistance; the value of 'n' depends on the device material and fabrication, and $n=20$ by our previous works [135]. The march algorithm is shown below [161], which can also detect pinholes with medium and high strength:

$$\text{March - pinhole - Rel.} = \{\uparrow(w1)^n; \uparrow(r1)\} \quad (6.3)$$

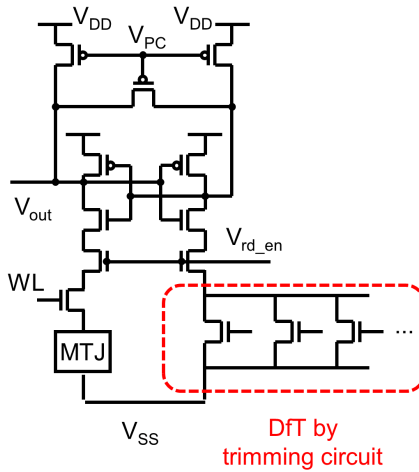


Figure 6.1: MTJ manufacturing process and the potential unique defects.

Design for test Considering the property that the pinhole reduces the MTJ resistance, DfTs monitoring read currents are applied; a typical DfT makes use of trimming circuits [162]. Fig. 6.1 presents the circuit of trimming-attached SA. By selecting proper references, the MTJ resistance is compared with different reference resistors (i.e., MOSFETs), allowing fine-tuned detection of resistance variations [76]. If the MTJ resistance deviates slightly from the expected range in the presence of pinholes, the trimming circuit enables the identification of defective cells.

6.1.2. TESTING OF SYNTHETIC ANTI-FERROMAGNET FLIP (SAFF)

SAFF defects arise from an unintended reversal of the HL magnetization, which in turn flips the RL due to strong anti-ferromagnetic coupling. The concept of ‘defect strength’ is not included in SAFF, as this defect only includes one situation, i.e., the magnetization of ferromagnetic layers all reversed. The impact of the SAFF is as follows [87]

- The impact of H_{ext} on the MTJ switching characteristics is reversed.
- A higher WER.

FAULT MODELING

Passive neighborhood pattern sensitive faults are sensitized by SAFF, where faults are sensitized with specific DB patterns (setting of the DB in this thesis is presented in Fig. 5.3) [87]:

- Passive neighborhood pattern sensitive faults:
 - < 1; 1; 1; 1; 1; 1; 1; 1 w0/1_i/- >, i.e., a down write transition will fail intermittently when all 8 neighboring cells are in state ‘1’ (DB=1)
 - < 0; 0; 0; 0; 0; 0; 0; 0 w1/0_i/- >, i.e., an up write transition will fail intermittently when all 8 neighboring cells are in state ‘0’ (DB=0)

TEST METHODOLOGY

March test: To detect intermittent faults, the march algorithm is applied [87]:

$$\text{March - SAFF} = \{\uparrow (w1); \uparrow (w0, r0, w1)^n\} \quad (6.4)$$

However, due to the stochastic nature of these intermittent faults, escapes cannot be avoided. The calculation method of the escape rate is presented in chapter 5

Design for test: Currently, no DfT has been proposed to detect SAFF. In this thesis, we propose that applying an external magnetic field can fully detect this defect, which will be presented in this section.

6.1.3. TESTING OF INTERMEDIATE STATE (IM)

The IM defect causes MTJs to exhibit an unintended 'U' resistance state between P and AP, disrupting binary storage. The impact of the SAFF is as follows [87]:

- The MTJ intermittently presents a stable 'U' state in write operations.

FAULT MODELING

The defect strength of IM is defined as the probability of MTJ entering specific 'U' states, and IM only sensitizes intermittent HtD faults: [87]:

- HtD faults (intermittent faults): $\langle 1 w0/U_i / - \rangle$, $\langle 0 w1/U_i / - \rangle$.

TEST METHODOLOGY

March test: March tests to detect 'U' state require repeating read operations, and the march algorithm is:

$$\text{March - IM} = \{\uparrow (w0, r0, w1, r1)^n\} \quad (6.5)$$

Design for test: To detect the 'U' defect, DfTs monitoring read currents are applied. The DfT with trimming circuits in Fig. 6.1 can be applied here. By selecting a proper reference resistor, the 'U' state can be detected. A March algorithm is applied:

$$\text{March - IM - DfT} = \{\uparrow (w1); \uparrow (w0, rU^*, w1)^n\} \quad (6.6)$$

Here, the operation rU^* refers to detecting the MTJ state by the DfT to read the specific 'U' state. However, due to the stochastic nature of IM appearing, the escape ratio can never reach 0.

6.2. DEFECT CHARACTERIZATION OF BACK-HOPPING (BH)

Previous works have applied the DAT method to MTJ defects such as pinhole, SAFF, and IM [66], [81], [87], as discussed above. This thesis focuses on a new MTJ defect - 'BH' - in STT-MRAMs. The BH defect is defined as MTJ state oscillations during write operations, which can lead to potential write failures [80]. The DAT methodology is applied following four steps: defect characterization, defect modeling, fault modeling, and test generation [85]. We evaluate the BH behavior on several MTJ devices using electrical pulses of varying duration in standard I-V sweeps and WER test schemes. Thereafter, the definition of 'BH defect' is proposed; and it is observed that the BH-defective MTJ shows a higher WER than defect-free MTJs.

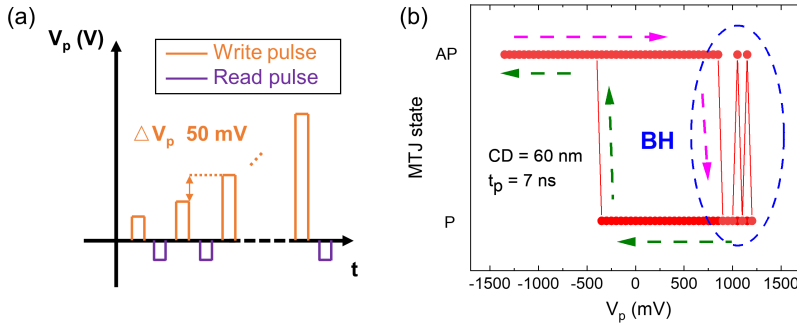


Figure 6.2: Plot of MTJ switching under the application of voltage pulses.

6.2.1. PULSE I-V MEASUREMENT

Fig. 6.2 (a) presents the method of ‘pulse I-V’ measurement. The MTJ device is initialized to the P state. Thereafter, a sequence of write-read operations is performed. The write operations are a series of voltage pulses with a constant $t_p = 7$ ns and a stair-case V_p ; V_p initially sweeps from -1.3 V to 1.4 V with a stair-gap $\Delta V_p = 50$ mV, and then sweeps back to -1.3 V with $\Delta V_p = -50$ mV (the sweeping not shown in Fig. 6.2). After every write operation, a read operation with $V_p = -50$ mV is performed to detect the MTJ state.

Fig. 6.2 (b) shows the measurement data. During the positive-oriented V_p sweep, the MTJ state is expected to stay at the P state after the first successful $1w0$ operation. However, it demonstrates that the MTJ undergoes an oscillation between AP and P with a strong stress (i.e., a large pulse height V_p). This oscillation behavior is attributed to the BH [138], [163], which indicates unexpected write errors may occur during regular write operations. Hence, it is critical to investigate the impact of BH on the WER for MTJs.

6.2.2. WER MEASUREMENT

Fig. 6.3 (a) presents the method of extracting WER with $1w0$ operations. Fig. 6.3 (b) and (c) present the second read results of the 1000 ‘initialization-read-write-read’ cycles for $0w1$ and $1w0$ operations separately, where write errors are observed. In Fig. 6.3 (b), the MTJ is expected in the AP state after the $0w1$ operations, yet some of the final states are in the P state. Similarly, in Fig. 6.3 (c), successful $1w0$ operations are expected to switch the MTJ from AP to P state, yet the final state after part of the $1w0$ operations is still in AP.

The extracted WER of $0w1$ and $1w0$ operations with different pulse heights is presented in Fig. 6.3 (d). For both types of switching, write errors are initially observed under weak stress (i.e., low V_p), in which insufficient spin torque causes unsuccessful switches. The WER is reduced to 0 with a larger V_p inputting more spin torque. Conventionally, a further strengthening of the stress is expected to keep the WER at 0. However, the figure presents a high WER under strong stress, which implies the unexpected MTJ state flipping after the first successful switch. Therefore, while stressing the MTJ devices with overdrive conditions was the standard approach [41], the BH effect implies an increase in WER if V_p exceeds a certain threshold.

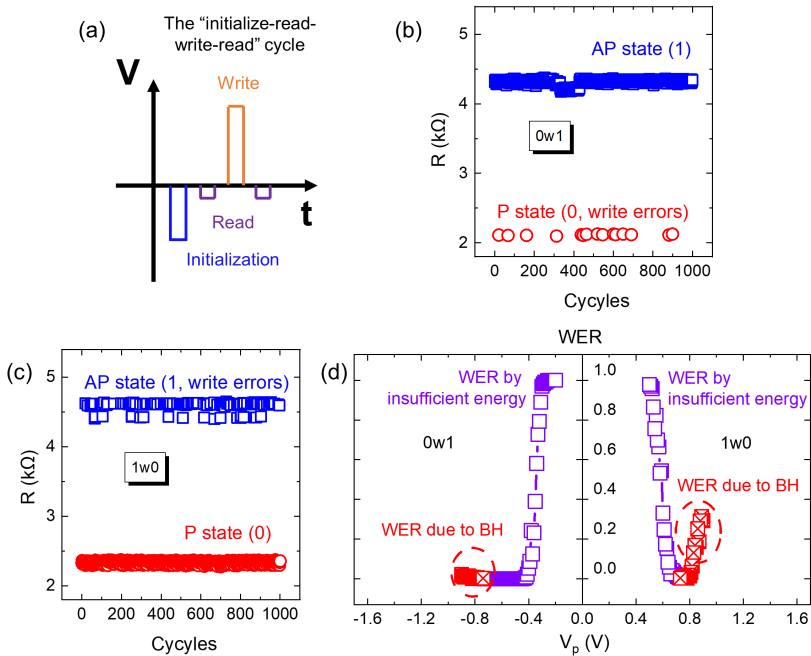


Figure 6.3: (a) WER measurement process; (b) WER measurement of $0w1$ operations; (c) WER measurement of $1w0$ operations; (d) WER extraction.

Furthermore, Fig. 6.3 (d) shows that the BH-induced WER is higher in the $1w0$ operation than in the $0w1$ operation. This is in-line with the observation of no BH for $0w1$ operations in Fig. 6.2.

6.2.3. WRITE BIAS IMPACT ON DEFECT OCCURRENCE

Fig. 6.4 presents the WER of all four write types (i.e., $1w0$, $0w1$, $1w1$, $0w0$). Note that the figure has a logarithmic y-axis, when WER=0 (i.e., no write error occurs in) In $1w0$ operations, the WER decreases with increasing V_p bias, as expected from STT theory, but increases again once BH is activated for $V_p > 0.9$ V (see Fig. 6.4 (a)). In $0w0$ operations, the WER is initially 0 as no switching is expected to occur under the applied write bias (i.e. device is already in the targeted final state). However, an increase in the WER is observed with higher V_p , implying the BH occurrence. Similarly, in Fig. 6.4 (b), the WER with $0w1$ operations is initially reduced and then raised; the WER $1w1$ operations is initially 0 and then raised.

Fig. 6.4 demonstrates that the BH-induced WER performance (i.e., WER with high V_p) is perfectly matching for $1w0$ and $0w0$ operations, and also for $0w1$ and $1w1$ operations. On the other hand, at the same $|V_p|$, the BH-induced WER is higher in $1w0$ and $0w0$ operations than that in $0w1$ and $1w1$ operations. Hence, the initial state has little impact on the BH-induced WER performance, while the write type (i.e., the write current direction) has a major impact.

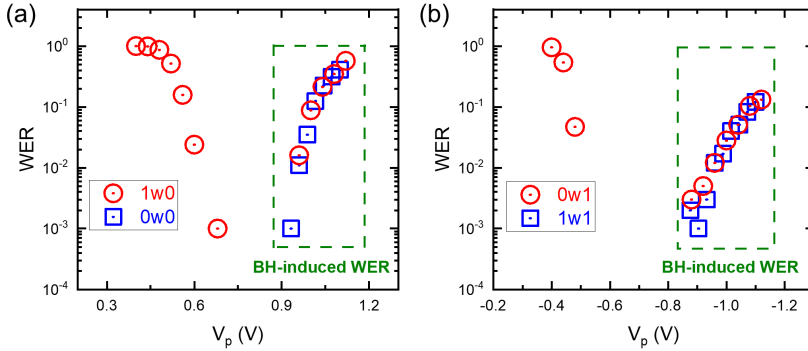


Figure 6.4: WER of (a) $1w0$ and $0w0$; (b) $0w1$ and $1w1$.

6.2.4. DEFINITION OF BH-DEFECTIVE AND DEFECT-FREE MTJs

The WER is an important practice in evaluating the STT-MRAM writing performance. The requirement on the WER (i.e., the WER spec, WER_{spec}) varies depending on different applications of the STT-MRAM. This work sets a large WER_{spec} of 10^{-3} to save time for measurements. We expect a defect-free MTJ with WER lower than WER_{spec} by the presence of regular write operations.

Fig. 6.5 (a) presents the WER of multiple MTJs with $0w1$ operations. It is observed that the WER performance due to the insufficient spin torque input is similar in all devices. However, the BH-induced WER of two devices is 5x-10x higher than that the other devices in the sample-set. Fig. 6.5 (b) extracts two lines from Fig. 6.5 (a), one with a higher BH-induced WER and the other with a normal BH-induced WER. Here, we set $|V_p| = 0.8V$ in the standard write operation, as presented in the green vertical line in the picture. In Fig. 6.5 (b), the BH-induced WER of one MTJ is higher than WER_{spec} (i.e., 10^{-3}) at this V_p , which is defined as BH-defective; the BH-induced WER of the other MTJ is lower than WER_{spec} , which is defined as defect-free.

A similar case is presented in Fig. 6.5 (c) and (d) with $1w0$ operations. The BH-induced WER of two devices is higher than that of others, as shown in Fig. 6.5 (c). Fig. 6.5 (d) compares the WER of two lines extracted from Fig. 6.5 (c), in which one device is with a BH-induced WER is higher than WER_{spec} at $|V_p| = 0.8V$, while the BH-induced WER of the other falls below it. Consequently, one MTJ is classified as BH-defective, and the other as defect-free. Similar works have been performed for the $1w1$ and the $0w0$ operation, which will not be presented here.

6.2.5. PHYSICAL MECHANISM

RELATED WORK

In 2009, the phenomenon that the MTJ state oscillates permanently under strong stress was observed by J. Z. Sun *et al.* [138], which is then named as ‘back-hopping’. The physical mechanism of BH was initially a puzzle with different possible theories, one of which is the low thermal instability [164]. In 2016, W. Kim *et al.* pointed out that the low stability of the reference layer is the key reason for the oscillating phenomenon, and experimen-

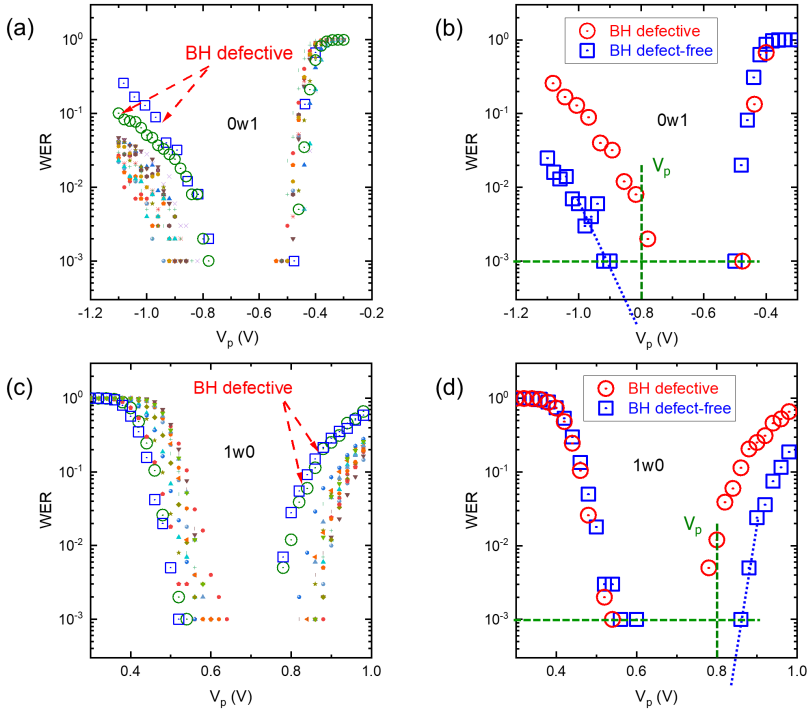


Figure 6.5: (a) WER of multiple MTJs in the $0w1$ operation; (b) Selected WER comparison of defect-free and defective MTJs in the $0w1$ operation; (c) WER of multiple MTJs in $1w0$ operations; (d) Selected WER comparison of defect-free and defective MTJs in the $1w0$ operation.

tally validated their theory [80].

The BH is thought to be one of the key challenges in the STT-MRAM design. To ensure a high switching rate, enough spin torque must be input to the MTJ. However, too much spin torque input may ‘over-drive’ the device and cause BH-induced write errors. Hence, a compromise is required by selecting a proper pulse height and width to reduce WER in STT-MRAMs [165].

PHYSICAL ANALYSIS

The BH occurs due to the instability of RL. Fig. 6.6 presents the factors affecting the RL_t stability. To elucidate the mechanism, symbols of ‘ \uparrow ’ and ‘ \downarrow ’ are applied to reflect the magnetization direction of different ferromagnetic layers and the direction of STT. When electrons flow from RL_t to FL, the STT of ‘ \downarrow ’ is offered from RL_t to FL; conversely, FL provides the STT of equal magnitude but opposite direction (i.e., ‘ \uparrow ’) back to RL_t . Additionally, the RL_t is influenced by the pinning effect from the RL_b , which forces the RL_t monetization to be parallel (i.e., the same) to that of RL_b through the ferromagnetic coupling [166]. In brief, while the pinning effect enhances the RL_t stability, the STT effect reduces it. Competition between these two factors determines the stability of RL_t : the RL_t becomes unstable if the STT effect exceeds the pinning effect. Since mag-

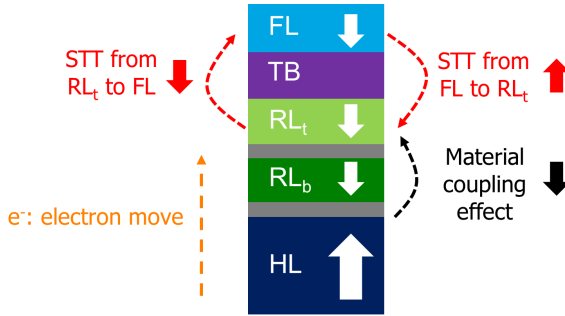


Figure 6.6: Factors affecting RL instability.

netic coupling is constant and STT is proportional to the current (i.e., the total amount of electrons flowing from RL_t to FL per second) [96], the RL_t may be switched when a sufficiently high V_p is applied. Once the RL_t is switched, the STT directed \uparrow is sent to FL (with the same electron direction), forcing the magnetization switch ' \uparrow ' and causing write errors.

It is important to note that when BH occurs, all thin ferromagnetic layers (i.e., FL, RL_t , and RL_b) may lose stability. Furthermore, when any one of these layers is switched, the STT applied to the other layers reverses direction. Consequently, ferromagnetic layers undergo continuous switching during the write operations, causing the MTJ state to oscillate between P and AP.

6.3. DEFECT MODELING OF BH

Due to the intrinsic nonlinear characteristics, regular defect models with linear resistors are inappropriate for representing BH. To model irregular defects, Wu *et al.* demonstrated a systematic device-aware defect modeling approach with three steps [68]: 1) physical defect analysis and modeling, 2) electrical defect modeling, 3) model optimization. In this section, the BH defect model is designed following these steps.

6.3.1. PHYSICAL MODELING

To physically model BH, three simplifications are proposed as preconditions: 1) For all layers of the PL stack in Fig. 2.1, only the RL_t are switchable as it is the most vulnerable layer (i.e., the thinnest) [80]. 2) At each time, only one layer (i.e., either FL or RL_t) is unstable, and the other is stable [167], [168]. 3) Ignore interface effects [168], since we find that in our experiment, including them in the compact model introduces a large amount of calculation without improving accuracy.

Fig. 6.7 shows the AP-to-P (i.e., $1\omega 0$) switching process with BH [80], [167], [168]. The write current flows from FL to RL_t , meaning the electrons flow from RL_t to FL. A four-phase loop is formed with BH; at each phase, either FL or RL_t will switch, and other layers (e.g., the RL_b and the hard layer) are stable. The magnetization direction of both FL is ' \uparrow ' and RL_t is ' \downarrow ' before the write operation. The four phases are modeled as follows:

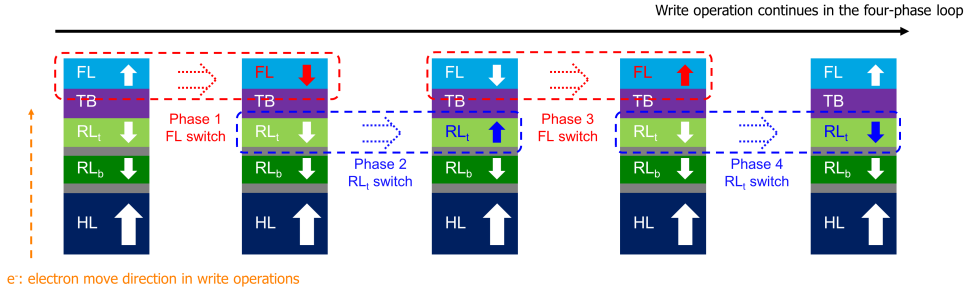


Figure 6.7: The physical mechanism of BH in the $1w0$ operation.

- **Phase 1:** The FL magnetization is unstable, and may be switched from ‘↑’ to ‘↓’. The STT effect is responsible for the switching. This is a common AP-to-P switching in $1w0$ operations, which will occur in every successful $1w0$ operation.
- **Phase 2:** The RL_t is unstable, and may be switched from ‘↓’ to ‘↑’. Two issues affect the stability of RL_t : 1) the STT effect from the FL, which facilitates the switching, and 2) the pinning effect from the RL_b , which impedes the switching. When the STT effect is stronger than the pinning effect, this phase may occur. For defect-free MTJs, the pinning effect is strong, and Phase 2 will occur with an extremely low probability during regular write operations. However, for BH-defective MTJs, the pinning effect is weak, and Phase 2 may occur even with a relatively low V_p , leading to Phase 3 and Phase 4 following in sequence. To model this pinning effect, an effective pinning magnetic field H_p is introduced [166]. The H_p is an effective magnetic field that only works on the RL_t , yet not affecting other layers; a low H_p refers to a weak pinning effect implying a BH-defective MTJ, and a high H_p refers to a strong pinning effect implying the MTJ defect-free from BH.
- **Phase 3:** The FL is unstable and may be switched from ‘↓’ to ‘↑’. This phase is similar to Phase 1, and the STT from RL_t accounts for this switching. Notice that both the RL_t magnetization, the FL magnetization, and the direction of the STT effect are reversed compared with Phase 2, hence the switching is also reversed.
- **Phase 4:** The RL_t is unstable, and may be switched from ‘↑’ to ‘↓’. Both factors of the STT effect and the magnetic coupling favor the switching. The end of Phase 4 indicates a new start of Phase 1; hence, the four phases form a complete loop, and the MTJ state oscillates permanently in this loop.

Since the MTJ state oscillates within this loop during write operations, whether the write error occurs depends on which phase the MTJ is in at the end of the write operation. Because of the magnetic coupling, the RL_t will be switched back to ‘↓’ when the write current is removed. Hence, the final MTJ state after write operations only relies on the FL magnetization. If one write operation stops during Phase 1 or Phase 4, the MTJ state ends in AP, and the write error occurs. On the other hand, the MTJ state ends in P when the write operation stops during Phase 2 or Phase 3, and this operation succeeds. When the write operation is long enough, multiple loops will occur. Due to the

stochasticity in the duration of each phase, the MTJ can never be guaranteed to stay during one certain phase at the end of the write operation. Hence, a high BH-induced WER is observed.

The physics of BH in $0w0$, $0w0$, and $0w0$ operations are similar to that in $1w0$ operations. In P-to-P switching, the current flow direction is the same as that in the $1w0$ operation (see Fig. 6.7). Phase 1 in P-to-P switching corresponds to Phase 2 in the $1w0$ operation, with both phases having similar durations, and so forth. Write errors occur when the write operation stops at the FL magnetization ‘↑’, meaning Phase 3 and Phase 4 (which refer to Phase 1 and Phase 4 in the $1w0$ operation). This explains the phenomenon in Fig. 6.5, that the BH-induced WER performance is closing in $1w0$ & $0w0$ operations at certain V_p . On the other hand, the current flow direction is reversed in $0w1$ & $1w1$ operations compared with $1w0$ operations. As a result, the occurrence of the four phases proceeds in the opposite direction (i.e., the loop progresses from right to left in Fig. 6.5). For example, Phase 1 in the $0w1$ operation refers to the reversed Phase 1 in the $1w0$ operation; the duration of the two phases is different at the same V_p considering different MTJ resistances in P and AP states. Therefore, the BH-induced WER in $0w1$ & $1w1$ operations is different in $1w0$ & $0w0$ operations at a certain V_p .

6.3.2. ELECTRICAL MODELING

Following the obtained four-phase loop physical defect model, the electrical modeling can be realized by calculating the critical parameters of I_c and t_w in each phase, which are denoted as I_{c1} , t_{w1} , I_{c2} , t_{w2} , I_{c3} , t_{w3} , I_{c4} , t_{w4} separately. Here, no external magnetic field is applied, and $H_{ext} = 0$. Besides, the stray field is ignored, since the STT effect is much larger than the stray field effect under large V_p . Table. 6.1 presents the calculation of I_c and t_w in each phase, explained as follows:

- **Phase 1:** the FL is switched, and the calculation of I_{c1} and t_{w1} is presented the same in Eq. 3.8.
- **Phase 2:** the RL_t is switched. Since the FL and RL_t are made of the same material, the parameters of calculating I_{c2} and t_{w2} are basically the same as applied in I_{c1} and t_{w1} with several exceptions. 1) the thickness of RL_t is 1.5 nm compared with that of FL is 1 nm; 2) Due to the FL switching in Phase 1, the MTJ resistance is switched from R_{AP} to R_P , hence the write current I is increased at the same V_p ; 3) The STT efficiency is changed from η_P to η_{AP} because the magnetization of FL and RL_t is in parallel after Phase 1; 4) The pinning effect parameter H_p is applied to the calculation of I_{c2} , which is preventing the switching.
- **Phase 3:** the FL is switched, and the calculation of I_{c3} and t_{w3} is presented the same as in Eq. 3.8. Both magnetization of FL and RL_t are switched; they are in anti-parallel. Hence, the write current I and STT efficiency η are the same as that in Phase 1.
- **Phase 4:** the RL_t is switched, and the calculation of I_{c4} and t_{w4} is similar to that in Phase 2. The H_p has the same effect as in Phase 2, which forces the magnetization of RL_t to be parallel to that of RL_b . Hence, the H_p in this phase favors the switching, which is presented as reducing the critical current in the I_{c4} calculation.

Table 6.1: Key parameter calculations of defective MTJ model in the $1w0$ operation.

Phase 1	I_{c1}	$\frac{1}{\eta_{AP}} \cdot \frac{\alpha \cdot e}{\hbar} \cdot A \cdot t_{FL} \cdot M_s \cdot H_k$
	t_{w1}	$\frac{(C + \ln(\frac{\pi^2}{4} \Delta)) \cdot e \cdot m}{4 \cdot \mu_B \cdot \eta_{AP} \cdot (I - I_{c1})}$
Phase 2	I_{c2}	$\frac{1}{\eta_P} \cdot \frac{\alpha \cdot e}{\hbar} \cdot A \cdot t_{RLtop} \cdot M_s \cdot (H_k + H_p)$
	t_{w2}	$\frac{(C + \ln(\frac{\pi^2}{4} \Delta)) \cdot e \cdot m}{4 \cdot \mu_B \cdot \eta_P \cdot (I - I_{c2})}$
Phase 3	I_{c3}	$\frac{1}{\eta_{AP}} \cdot \frac{\alpha \cdot e}{\hbar} \cdot A \cdot t_{FL} \cdot M_s \cdot H_k$
	t_{w3}	$\frac{(C + \ln(\frac{\pi^2}{4} \Delta)) \cdot e \cdot m}{4 \cdot \mu_B \cdot \eta_{AP} \cdot (I - I_{c3})}$
Phase 4	I_{c4}	$\frac{1}{\eta_P} \cdot \frac{\alpha \cdot e}{\hbar} \cdot A \cdot t_{RLtop} \cdot M_s \cdot (H_k - H_p)$
	t_{w4}	$\frac{(C + \ln(\frac{\pi^2}{4} \Delta)) \cdot e \cdot m}{4 \cdot \mu_B \cdot \eta_P \cdot (I - I_{c4})}$

The electrical defect model for the defective MTJ switching in other operations (i.e., $0w1$, $0w0$, and $1w1$) can be approached in a similar manner. At the start of the write operation, the model checks the initial MTJ state and the write current direction, then sets the four-phase loop. The WER calculation is applied to t_w at each phase, to present the stochasticity. When the time elapsed in each phase exceeds the associated t_w , this phase ends, the corresponding ferromagnetic layer (i.e., FL or RL_t) is switched, and the next phase starts. At the start of each phase, the model checks the magnetization of FL and RL_t : if they are in parallel, the MTJ demonstrates low resistance; otherwise, if they are in anti-parallel, the MTJ demonstrates high resistance.

During the four-phase loop, the magnetization of RL_t may be switched to be anti-parallel to that of RL_b due to the high write current. However, once the write operation ends, the magnetization of RL_t is forced to be parallel to that of RL_b by the model. We assume that the time for this magnetization switching (in nanoseconds, which can be evaluated in the same way as calculating I_{c4} and t_{w4} with $I = 0$) is much shorter than the gap between different read/write operations. Hence, the final MTJ resistance after the write operation may differ from that in the four-phase loop. H_p is demonstrated as the

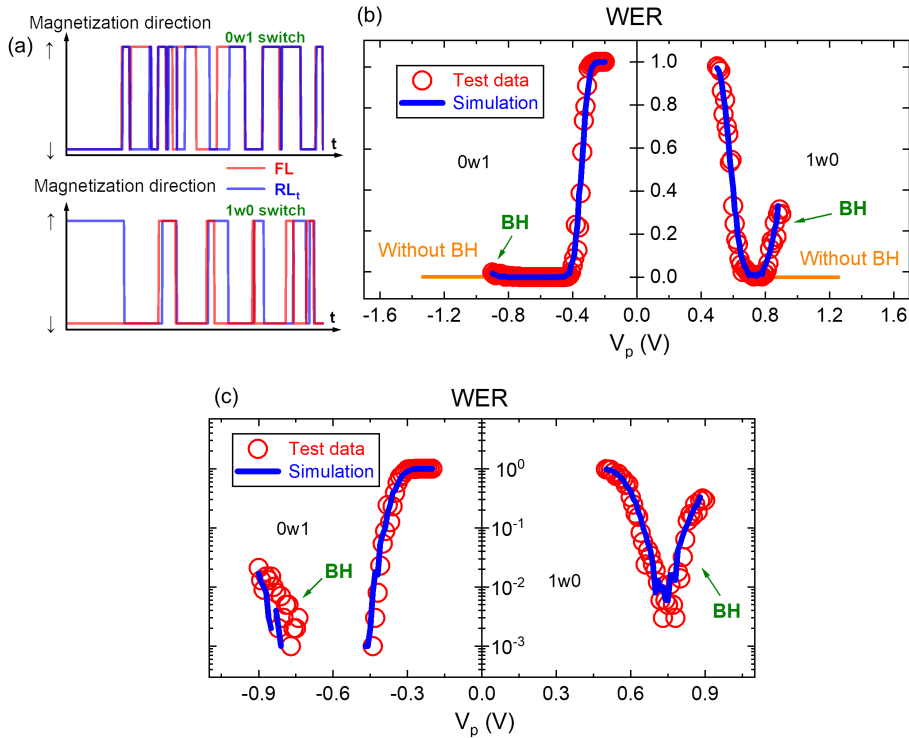


Figure 6.8: (a) Magnetization switching during write operations; (b) BH fitting in the linear y-axis; (c) BH fitting in the log y-axis.

defect strength. A high H_p indicates the device to be defect-free; conversely, a low H_p refers to a strong BH defect and a high BH-induced WER.

6.3.3. FITTING AND MODEL OPTIMIZATION

The MTJ compact model is first realized by Python for the convenience of model optimization. Fig. 6.8 (a) shows the magnetization oscillation of FL and RL_t during write operations. When the magnetization of the FL and RL_t is in anti-parallel (i.e., one in '↑' and the other in '↓'), the MTJ is in high resistance; if the magnetization of the two layers is parallel (i.e., both '↑' or '↓'), the MTJ is low resistance. When the write operation ends, the magnetization of the FL remains unchanged, and the magnetization of the RL_t is pinned to be '↓' immediately.

Next, we optimize and fit the model with measurement data. The fitting follows three steps: 1) The fitting of the R-V measurement. 2) The fitting of the regular switching performance by setting H_p to be $+\infty$, and varying H_k and M_s . 3) The fitting of BH-induced WER performance by varying the defect strength H_p . The fitting needs to be carried out in $1w0$ and $0w1$ operations; in $0w0$ and $1w1$ operations, no normal switching occurs, hence step 2 cannot be performed.

Fig. 6.8 (b) presents the fitting result of a severe BH-defective MTJ with the linear

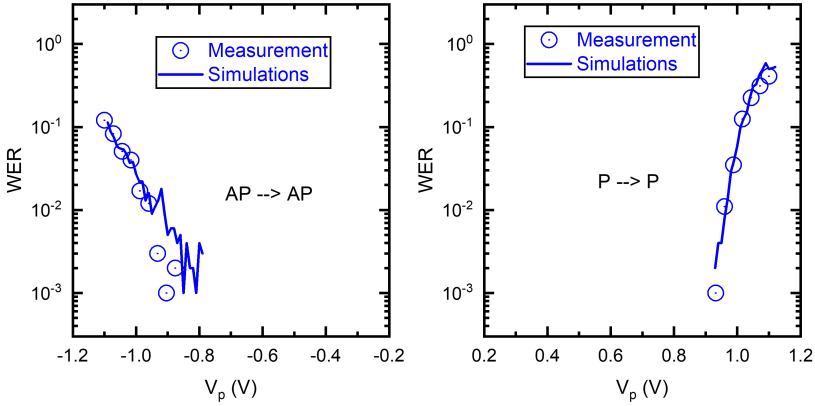


Figure 6.9: WER measurement and model fitting in (a) $1w1$ operation; (b) $0w0$ operation.

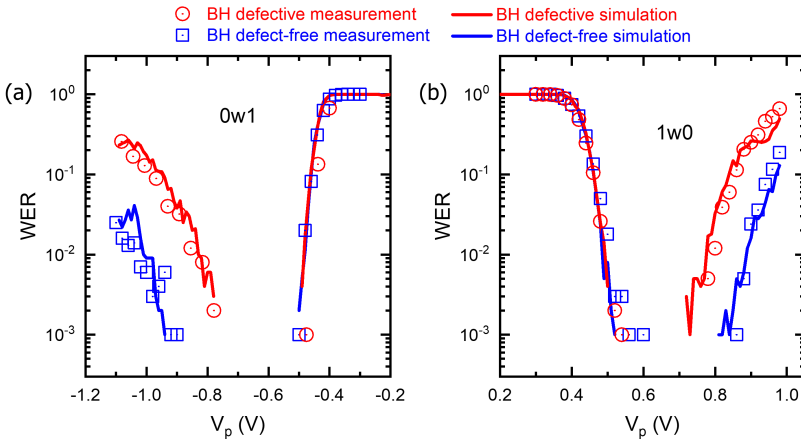
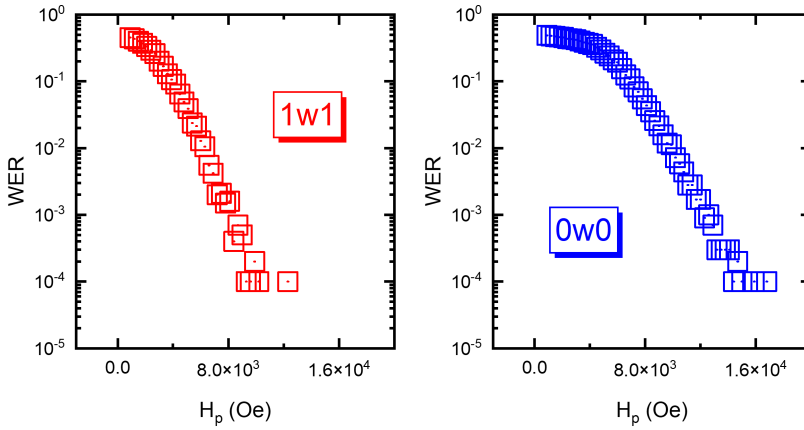


Figure 6.10: WER measurement of BH-defective and defect-free MTJs in (a) $0w1$ operation; (b) $1w0$ operation.

y-axis. Here, the defect strength is H_p 3.7 kOe. When the BH model is not integrated into the MTJ model, the WER is 0 with a large V_p , as demonstrated in the orange line in Fig. 6.8 (b). Fig. 6.8 (c) shows the same fitting result as that in Fig. 6.8 (b), yet with the logarithmic y-axis, to better highlight the accuracy of this model. Fig. 6.9 presents the simulation results with $1w1$ and $0w0$ operations, which fits to the measurement result

Fig. 6.10 presents the fitting for a defect-free and a BH-defective MTJ. The fitting of the two devices applies the same H_k and M_s , hence showing a similar normal switching performance. The only difference between the two models is the H_p : for the defect-free MTJ, $H_p = 15.5$ kOe, and for the BH-defective MTJ, $H_p = 6$ kOe.

Fig. 6.11 presents the simulation results of how the defect strength H_p affects the BH-induced WER performance in $0w0$ and $1w1$ switching at a certain V_p . As the H_p increases (i.e., the defect strength is reduced), the BH-induced WER is reduced.

Figure 6.11: WER by varying H_p .

After the verification by Python, the model is moved to Verilog-A to make it compatible with circuit-level simulations.

6.4. FAULT MODELING OF BH

Table 6.2 presents simulation results; here, F of FP notation $\langle S/F/R \rangle$ is extended with a new symbol ‘~’ to describe the STT-MRAM oscillation state during write operations. This oscillation is an irregular, faulty behavior, which can be described only by device-aware defect models, not by linear resistance defect models. Besides, this oscillation can only be observed in Spice simulation; in commercial STT-MRAM arrays, the MTJ state switching during write operations cannot be directly extracted. In total four FPs are derived based on the simulation results: $\langle 0w0/\sim/- \rangle$, $\langle 0w1/\sim/- \rangle$, $\langle 1w0/\sim/- \rangle$, $\langle 1w1/\sim/- \rangle$. The two write destructive faults ‘ $w0DF\sim$ ’ and ‘ $w1DF\sim$ ’ indicate the inadvertent state oscillations during $1w1$ or $0w0$ operations. The two write transition faults $w1TF\sim$ and $w0TF\sim$ demonstrate the state oscillations of transition write operations $0w1$ and $1w0$. We name these faults as ‘Write Oscillating Faults’. Since the specific ‘oscillating’ state cannot be detected by regular testing methods, all faults are classified as hard-to-detect (HtD) faults, as defined in [61].

Table 6.2 suggests that BH has a stronger effect on $w0$ operations, and $0w0$ is the most sensitive operation to BH; this operation will cause the MTJ to oscillate for all BH defect sizes and result in a fault. During $w0$ operations, the MTJ stays in the low resistance state in Phase 2. With a certain $|V_p|$, the RL_t suffers a larger write current, and becomes easier to switch. Besides, the defect strength range sensitized by $1w0$ and $0w0$ operations are close, which is 12.7 kOe and 13.2 kOe separately as demonstrated in Table 6.2; this corresponds to the case that the BH-induced WER performance is close in $0w0$ and $1w1$ operations, as presented in Fig. 6.4 (a).

Table 6.2: Fault modeling results of BH defect.

Defect strength	Sensitized FP	FP name
$H_p \in (0, 5.2 \text{ kOe})$	$\langle 0w0/\sim/- \rangle$	Write 0 oscillating fault: w0DF~
	$\langle 1w0/\sim/- \rangle$	Write 0 oscillating fault: w0TF~
	$\langle 1w1/\sim/- \rangle$	Write 1 oscillating fault: w1DF~
	$\langle 0w1/\sim/- \rangle$	Write 1 oscillating fault: w1TF~
$H_p \in (5.2, 5.9 \text{ kOe})$	$\langle 0w0/\sim/- \rangle$	Write 0 oscillating fault: w0DF~
	$\langle 1w0/\sim/- \rangle$	Write 0 oscillating fault: w0TF~
	$\langle 1w1/\sim/- \rangle$	Write 1 oscillating fault: w1DF~
$H_p \in (5.9, 12.7 \text{ kOe})$	$\langle 0w0/\sim/- \rangle$	Write 0 oscillating fault: w0DF~
	$\langle 1w0/\sim/- \rangle$	Write 0 oscillating fault: w0TF~
$H_p \in (12.7, 13.2 \text{ kOe})$	$\langle 0w0/\sim/- \rangle$	Write 0 oscillating fault: w0DF~
$H_p \in (13.2 \text{ kOe}, +\infty)$	No fault	

6.5. TEST GENERATION OF BH

6.5.1. READ-WRITE BASED TEST SOLUTIONS

Following the fault modeling, we design the tests for BH, which include 1) the read/write-based tests, like the march test; and 2) Dedicated Design-for-Testability circuits. In this section, we introduce the read/write-based tests for BH, evaluate their time efficiency, and propose methods for improvement.

REGULAR MARCH TESTS

Here, we select the proper faults for the march test design. Considering the physical mechanism and Table 6.2, we select the fault: $\langle 0w0/\sim/- \rangle$ (with the *). This fault sensitizes the longest defect strength, in which H_p ranges from 0 to 13.2 kOe. Besides, during the $0w0$ operation, no normal switching occurs; this avoids the write errors that are attributed to not enough spin torque input rather than BH. Afterwards, we design the march test the algorithm as presented:

$$\text{March} - \text{BH} = \{\updownarrow(w0); \updownarrow(w0, r0)^n\} \quad (6.7)$$

The ' \updownarrow ' indicates that addressing direction is irrelevant. The first element initializes the state of all MTJs to be '0'. The second element tries to sensitize the fault by $w0$ and reads the state by $r0$. Here, ' n ' refers to the number of times the march element should be repeated. Assuming that the total number of incorrect feedback provided by $r0$ is i , the WER is $WER = i/N$. Then, we compare the extracted WER with the WER_{spec} ; if $WER > WER_{spec}$, the MTJ is BH-defective.

There are two shortcomings of this march algorithm. 1) Low time efficiency; 2) Unavoidable escape. For example, $WER_{spec} = 10^{-3}$ as designed in this work, $n \geq 1000$ is required. Assuming $n = 1000$, the detected faulty result $i \geq 1$ refers to the device being BH-defective. The march test length is calculated as:

$$\text{Length} = (1 + 2 \cdot n) \cdot s = 2001 \cdot s \quad (6.8)$$

Where s is the array size. This is a large length, indicating a long test time. Besides, due to the stochasticity of BH, escapes cannot be avoided. For instance, if the BH-induced WER of one defective MTJ is exactly at 10^{-3} , the probability that no write error occurs during the 1000 repeating 'w0r0' is:

$$Probability = (1 - 10^{-3})^{1000} = 0.368 \quad (6.9)$$

This probability refers to the escape rate of this device. Increasing 'n' can reduce the escape rate, yet it requires a longer test time. Yet anyway, the escape rate can never be reduced to 0.

Similarly, the yield loss cannot be avoided due to the stochasticity of the BH. For example, the WER of an MTJ is 5×10^{-4} , and the probability of at least one write error occurring is:

$$Probability = 1 - (1 - 5 \times 10^{-4})^{1000} = 0.393 \quad (6.10)$$

This probability implies the yield loss, which cannot be reduced to 0 either. Likewise, increasing 'n' can reduce yield loss, yet it requires a longer test time.

ENHANCED MARCH TESET

One way to reduce the test time is to increase the BH-induced WER [169]. There are four possible approaches that can be applied: 1) Raise the temperature; 2) Apply the external magnetic field; 3) Increase the write pulse width t_p ; 4) Increase the write pulse height V_p ; Method 1) has severe limitations. The MTJ electrical performance (e.g., R_{AP}) is changed with increasing temperature, causing unexpected inaccuracy. Besides, a high temperature may damage the device. Method 2) has been described in [128], and will not be extended.

Next, we will present a detailed analysis of the methods 3) and 4) with measurement data. The implementation circuits of these two methods have been discussed in [158], [159], [170].

Increase pulse width:

Fig. 6.12 (a) presents the WER measurement data by varying t_p and keeping $|V_p| = 1.1V$ with 1w1 operations. The BH-induced WER initially increases with t_p at $t_p < 20ns$, then remains constant as t_p continues to increase. This initial rise may occur because, at very low t_p , the BH is in the first loop (see Fig. 6.7); thus, the WER is primarily determined by the probability of RL_t switching within this duration. Once multiple loops occur with a longer t_p , the WER instead relies on the proportion of the duration in each phase and hence becomes independent of t_p . Fig. 6.12 (b) shows a similar phenomenon with 1w1 operations. Fig. 6.12 demonstrates that increasing t_p only raises the BH-induced WER when t_p is very short, while further increasing t_p has a limited effect. Therefore, this approach offers little benefit in reducing test time.

Increase pulse height:

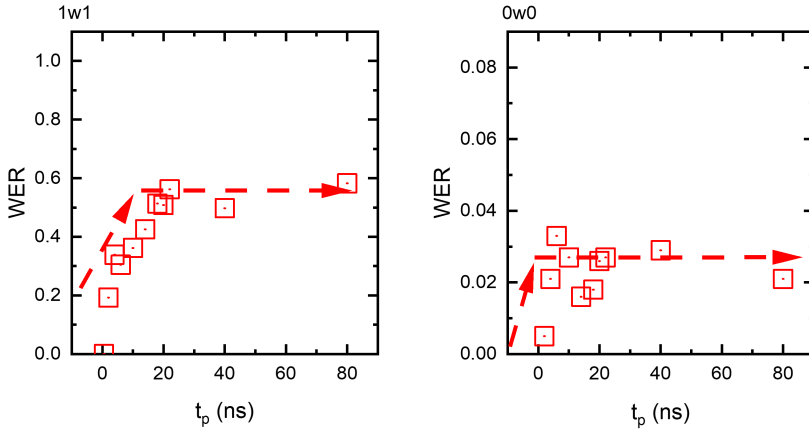


Figure 6.12: (a) WER measurement by varying t_p in (a) $1w1$ and (b) $0w0$ operations.

Here, we extract the data from Fig. 6.10 to explain the case of increasing pulse height to raise test time-efficiency. Although increasing V_p has raised the BH-induced WER, it has such an impact on all MTJs. If we apply a larger V_p in the test as the same method as presented in Sec. 6.5.1.1, a high yield loss occurs since the WER of defect-free MTJ is increased. Hence, a post-calibration process is required. Here, we assume that the BH-induced WER exactly equals WER_{spec} at the standard V_p (i.e., $WER = 10^{-3}$ at $|V_p|$ of 0.8 V in this work) for one device, we can extract the WER of this device at a different V'_p by simulations. The extracted WER is set to be the WER'_{spec} . At this V'_p , if the WER of one MTJ is larger than WER'_{spec} , it is classified as BH-defective; otherwise, it is BH defect-free.

An example is presented in Fig. 6.10 (a) for the $0w1$ operation. In the regular march test, $WER_{spec} = 10^{-3}$ and $V_p = -0.8V$. In the picture, one MTJ is BH-defective and the other is defect-free. When we increase the V_p to $V'_p = -1.04V$, the $WER'_{spec} = 8\%$. In the test, if the extracted WER' is larger than 8% with V'_p , the MTJ is BH-defective. With the new spec, the MTJ originally classified as 'BH-defective' is still defective, and the other one, originally defect-free, is still free from BH. Fig. 6.10 (b) presents another example for the $0w1$ operation, with $V'_p = 0.92V$, the $WER'_{spec} = 10\%$.

The test process is presented as follows:

- Calibration process:
 - Simulation set-up with a BH-defective model, where $WER = WER_{spec}$ at V_p .
 - Select a proper V'_p .
 - Extract the WER'_{spec} at the V'_p by simulation.
- Perform $\{\uparrow (w0r0)^n\}$ at V'_p .
- Extract the WER, and compares it with WER'_{spec} .

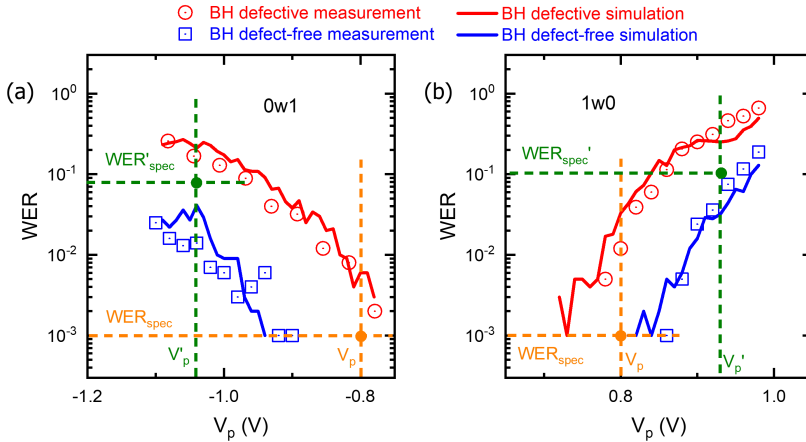


Figure 6.13: Increase V_p for testing.

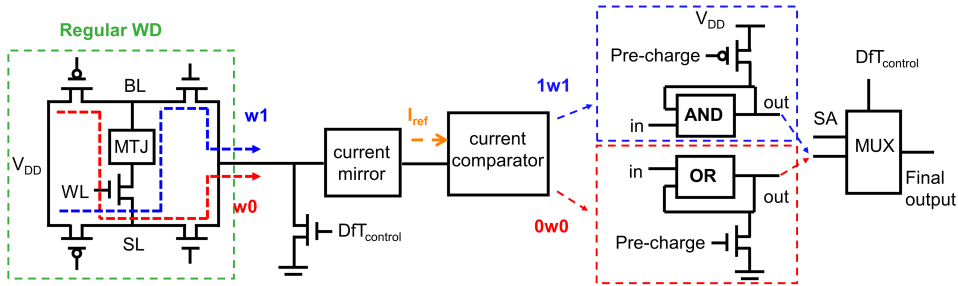


Figure 6.14: DFT for defecting the BH defect.

Here, ‘ n ’ is evaluated by WER'_{spec} . For example, when $WER'_{spec} = 10\%$, $n \geq 10$. Compared with the original march test in Sec. 6.5.1.1, where $WER_{spec} = 0.1\%$ and $n \geq 1000$, the pulse height enhanced test method has a smaller ‘ n ’, indicating a shorter test time. Yet the cost is the additional circuit to adjust the pulse height and additional effort to set up V'_p & WER_{spec} . Besides, although time efficiency is improved, escape and yield loss can still not be avoided due to the intrinsic stochasticity of BH.

6.5.2. DESIGN FOR TEST

Due to the nature of oscillation in BH-defective MTJs, all test methodologies described above cannot guarantee full coverage of this defect type. Therefore, we propose a new DfT as presented in Fig. 6.14. The original WD is shown in the green square, which is expected to directly connect to the ground, rather than the current mirror. An N-MOSFET is set to connect the WD and the ground, with a signal $DfT_{control}$. When $DfT_{control}$ is set to 1 (i.e., V_{DD}), the N-MOSFET is ‘ON’, the WD directly connects to the ground, and the DfT is off. Otherwise, when $DfT_{control}$ is set to ‘0’ (i.e., to the ground), the N-MOSFET is ‘OFF’, the WD connects to the current mirror, and it is the testing mode. The

current mirror copies the write current and connects to a current comparator (i.e., the inverter in the picture). The current comparator compares the write current and an I_{ref} . An example of the reference current generator is presented in [171], and this generator can be shared by multiple DfT blocks. If I_{ref} is larger than the write current (i.e., MTJ in '1' state), the inverter outputs 1; otherwise (i.e., MTJ in '0' state), it outputs 0. Next, we divide the DfT into two types, depending on whether the inverter connects to an 'OR' gate, or connects to an 'AND' gate, corresponding to testing BH with $0w0$ or $1w1$ operations, respectively. Although the $0w0$ operation is chosen to test BH, both potential DfT options are introduced in this section. The two DfT options can be implemented simultaneously, or a single option can be selected to save circuit area, depending on the requirements of the practical design. The output of the DfT connects to a 2×1 multiplexer (MUX); the $DfT_{control}$ signal controls the final output to be the read result from SA or the DfT output.

In the first option, the inverter connects to the 'input' node of the 'OR' gate. A feedback line connects to the 'out' node and another input of the gate. AN N-MOSFET is set to connect the ground and the 'out' node, with a signal 'Pre-charge'. The testing of the BH consists of four steps:

- **Initialization:** close the DfT, apply $\{\downarrow(w0)\}$ to set all MTJs in '0' state.
- **Pre-charge:** set Pre-charge to 1, the N-MOSFET is open, and out is forced to be 0.
- **Testing:** set Pre-charge to 0, the N-MOSFET is closed; open testing mode with $DfT_{control}$ signal and apply $\{\downarrow(w0)\}$. The logic of the 'OR' gate is presented in Table 6.3. If no BH occurs during the write operation, the MTJ state remains in '0', and the output of the 'OR' gate is always 0. Once the MTJ state hops to '1', the current comparator provides 1, and the 'out' node of the 'OR' gate hops to 1. Due to the feedback line, the associate input of the gate is set to 1, and the 'out' node is locked to 1. Even if the MTJ state hops back to '0' after the first hop (i.e., the output of the current comparator hops back to 0), the 'out' node is permanently set to '1'.
- **Result check:** apply $\{\downarrow(r0)\}$ in testing mode, the final output is the DfT output. If the final output is 1, the MTJ is BH defective.

Table 6.3: Truth table of DfT with the 'OR' gate.

Step	Pre-charge	in	out
Initialization	1	0	0
$0w0$ starts	0	0	0
BH to 1	0	1	1
BH to 0	0	0	1

Fig/ 6.15 presents the MTJ state and DfT output with the 'OR' gate. During the initialization from 0 ns to 3 ns, the Pre-charge is set to V_{DD} , and the output is forced to 0. At 3 ns, the $0w0$ operation starts and the Pre-charge is set to '0'. The output remains at '0' as long as the MTJ state remains at '0'. When BH occurs, the MTJ switches to '1', locking

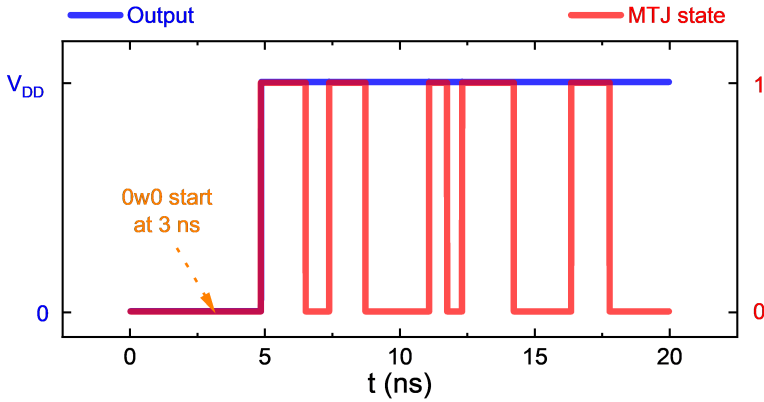


Figure 6.15: MTJ state and DfT output with 'OR' gate.

the output at '1'. Despite the MTJ oscillating between '0' and '1' due to BH, the output is locked to '1'.

Similar to the first type with the 'OR' gate, the second DfT type connects an 'AND' gate to the current comparator and configures a P-MOSFET. The testing is also composed of four steps.

- **Initialization:** close the DfT, apply $\{\hat{\uparrow}(w1)\}$ to set all MTJs in '1' state.
- **Pre-charge:** set Pre-charge to 0, the P-MOSFET is open, and out is forced to be V_{DD} .
- **Testing:** set Pre-charge to 1, the P-MOSFET is closed; open testing mode with $DfT_{control}$ signal and apply $\{\hat{\uparrow}(w1)\}$. The logic of the 'AND' gate is presented in Table 6.4. If no BH occurs during the write operation, the output of the gate keeps 1; if BH occurs, the output of the gate is locked to 0.
- **Result check:** apply $\{\hat{\uparrow}(r1)\}$ in testing mode, the final output is the DfT output. If the final output is 0, the MTJ is BH defective.

Table 6.4: Truth table of DfT with the 'AND' gate.

Step	Pre-charge	in	out
Initialization	0	-	1
1w1 starts	0	1	1
BH to 0	0	0	0
BH to 1	0	1	0

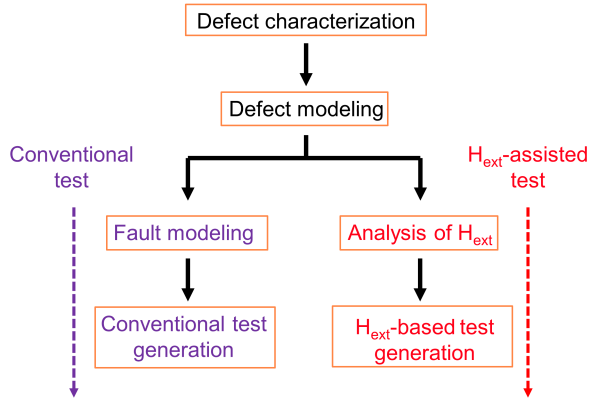


Figure 6.16: Comparison of conventional and H_{ext} -assisted test development

6.6. APPLICATION OF EXTERNAL MAGNETIC FIELD IN TESTING MTJ DEFECTS

Test methodologies for memories are usually classified into two types: 1) functional tests (e.g., march tests) and 2) DfT. However, both types of tests show some shortcomings when applied to STT-MRAMs, e.g., the long test time when applying march tests to intermittent faults and the cost of circuit area for DfTs [78], [89], [172], as described above in this section. Further improvement of STT-MRAM manufacturing requires better test solutions with a low escape rate and low test cost. Due to the unique magnetic-related working mechanism of MTJ switching, the external magnetic field presents itself as a potential tool that can be applied to STT-MRAM testing to improve defect coverage or reduce test time.

Fig. 6.16 presents a generic framework of both conventional test (utilizing the DAT method as an example [68]) and H_{ext} -assisted test development [83]. The first two steps are the same for the two test types. The first step, ‘defect characterization’, measures defective MTJ behaviors and studies the physical mechanism of the defect. The second step, ‘defect modeling’, incorporates the impact of physical defects into the technology parameters of the MTJ device and thereafter in the electrical parameters, and then develops a compact model of the defective MTJ. For conventional tests, the third step, ‘fault modeling’, describes the STT-MRAM faulty behaviors in the presence of defects through circuit simulations; the fourth step, ‘conventional test generation’, generates test solutions to sensitize faults (here, we only consider functional tests, like march tests). For H_{ext} -assisted test development, the third step ‘analysis of H_{ext} ’ performs the physical analysis on how H_{ext} impacts the components in Eq. 3.14 and Eq. 3.15 for defective MTJs; the fourth step ‘ H_{ext} -assisted test generation’ generates H_{ext} -assisted test solutions to detect defects.

The framework in Fig. 6.16 is applied to design H_{ext} -assisted tests for all known defects in the public domain, and compare these two approaches for each defect type. The results are summarized in Table 6.5 (see the last of this section), in which H_{ext} -assisted tests are advantageous in defect coverage and/or time-efficiency for most defects.

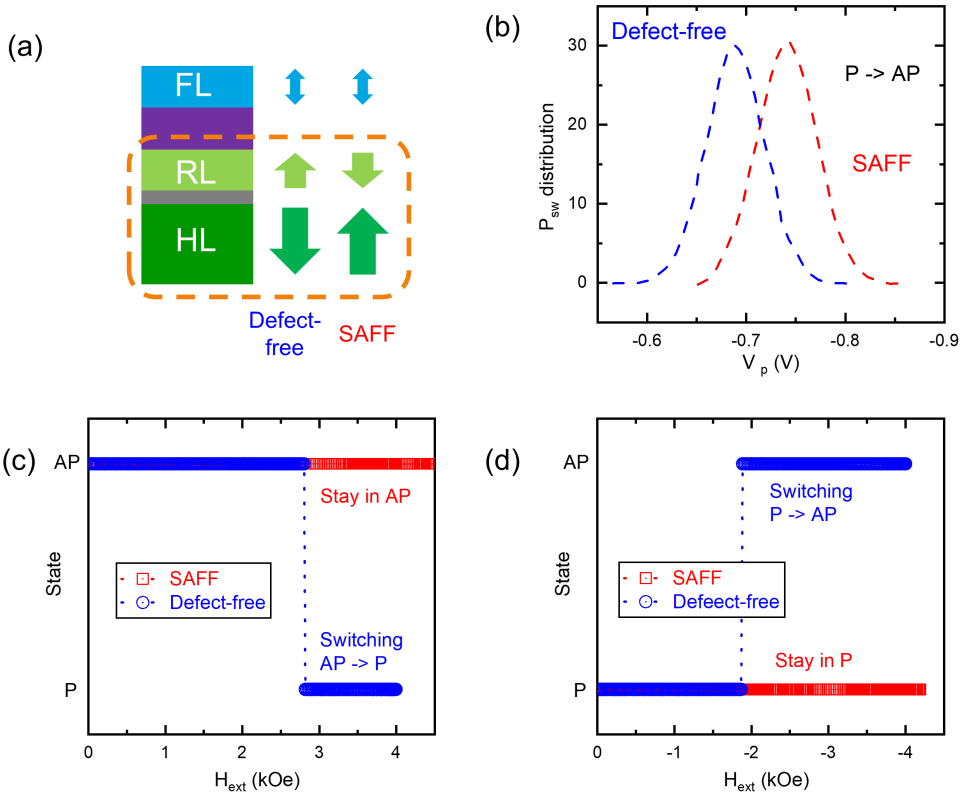


Figure 6.17: (a) SAFV mechanism; (b) P_{sw} distribution for SAFV-defective and defect-free MTJs; (c) Validation for the march algorithm Eq. 6.14; (d) Validation for the march algorithm Eq. 6.15.

Next, we will design H_{ext} -assisted tests to detect two defects ‘SAFF’ [68] and ‘BH’ [82] as illustrative examples. The framework can be applied to other defects, which this paper will not discuss.

6.6.1. SAFV TEST WITH EXTERNAL MAGNETIC FIELD

Next, we follow the steps in Fig. 6.16 to design the H_{ext} -assisted test for SAFV. First, we analyze the defect impact on the four components in Eq. 3.14 and Eq. 3.15. Since the FL magnetization is flipped in the presence of SAFV, the effect of the magnetic field on the FL is reversed as compared with defect-free MTJs. Hence, components relating to magnetic fields (i.e., H_s and H_{ext}) become opposite to their original directions (see Eq. 3.15), yet other components of (i.e., E_B , $H_k \cdot \epsilon$, and $I \cdot \eta$) remain unchanged, since the FL property is not affected by the defect. Notice that H_k is a material parameter, not the real magnetic field.

If we apply the same $1w0$ operation on both SAFV-defective and defect-free MTJs, the total input energy can be presented as follows:

$$\begin{aligned} E_{\text{defect-free}} &= t_p \cdot (-H_k \cdot \epsilon + I \cdot \eta + (+H_{\text{ext}} + H_s) \cdot \epsilon) \\ E_{\text{SAFF}} &= t_p \cdot (-H_k \cdot \epsilon + I \cdot \eta + (-H_{\text{ext}} - H_s) \cdot \epsilon) \end{aligned} \quad (6.11)$$

The differences between the two formulas above are the reversed H_s and H_{ext} . If $H_{\text{ext}} = 0$, H_s is the only origin of the gap between the input energy, which results in the different P_{sw} behavior in Fig. 6.17. Subtracting the two formulas in Eq. 6.11 results into the gap of the Input Energy (E_{gap}), presented as:

$$E_{\text{gap}} = |2 \cdot (H_{\text{ext}} + H_s) \cdot \epsilon \cdot t_p| \quad (6.12)$$

To distinguish SAFF-defective MTJs from defect-free ones, it is critical to enlarge E_{gap} to guarantee switching only occurs in one device type (i.e., only defective MTJs or only defect-free MTJs switch). A preferable way to realize this is by increasing the input energy for the defect-free case while decreasing it for the defective case. Given the fact that H_{ext} has opposite impacts on defect-free and defective case (increasing versus decreasing the input energy), it can be easily used to realize the purpose. For example, by applying a large positive H_{ext} , it achieves $E_{\text{SAFF}} \ll E_B$ while $E_{\text{defect-free}} \gg E_B$. Thus, defect-free MTJs switch while defective ones stay constant.

Furthermore, if we keep $I = 0$ (i.e., do not perform any write operation) and try to switch the MTJ purely by H_{ext} , the total energy input for defective and defect-free devices can be presented as:

$$\begin{aligned} E_{\text{defect-free}} &= t_p \cdot (-H_k \cdot \epsilon + H_{\text{ext}} + H_s) \cdot \epsilon > E_B \\ E_{\text{SAFF}} &= t_p \cdot (-H_k - H_{\text{ext}} - H_s) \cdot \epsilon < E_B \end{aligned} \quad (6.13)$$

To guarantee a successful switching for defect-free MTJs, it requires $E > E_B > 0$, meaning $H_{\text{ext}} > (H_k - H_s)$. On the other hand, the positive H_{ext} actually stabilizes the FL state for the SAFF-defective MTJ (i.e. it guarantees $E < 0$), and the device will never switch. Therefore, we can apply a positive H_{ext} and keep it for several microseconds; the actual value of H_{ext} is inconsequential as long as it exceeds a certain threshold, ensuring only defect-free MTJs switch.

The H_{ext} -assisted test for the SAFF defect can be described by the following three steps:

- Perform $\{\uparrow(w1)\}$.
- Apply a positive $H_{\text{ext}} = H_{1+}$ to the whole chip for several microseconds, then remove H_{ext} .
- Perform $\{\uparrow(r0)\}$.

The first step initializes all MTJs to '1' state. The second step applies the H_{ext} , which tries to switch the defect-free MTJ to state '0' while stabilizing the SAFF-defective MTJ state in '1'. Notice that the H_{ext} can be applied to devices on the whole chip simultaneously. The third step checks the final state of MTJs, and the SAFF is detected if MTJs stay in '1'. Here, the H_{ext} -assisted test is presented by the following march algorithm:

$$\{\uparrow(w1), H_{1+}, \uparrow(r0)\} \quad (6.14)$$

Fig. 6.17 (c) presents the R-H loop measurement data for both defective and defect-free MTJs with the initial state '1'. When $H_{1+} > 2.8 \text{ kOe}$, the defect-free device is switched to '0', yet the defective device still stays in '1'.

A similar test can be applied by initializing MTJ state to '0' and performing a large negative $H_{ext} = H_{1-}$. The march algorithm is presented as:

$$\{\uparrow(w0), H_{1-}, \uparrow(r1)\} \quad (6.15)$$

Fig. 6.17 (d) validates this algorithm. Devices are initialized to '0'. When $H_{1-} < -1.9 \text{ kOe}$ is applied for several microseconds, the defect-free device is switched to '1', yet the defective device still stays in '0'.

COMPARISON OF FUNCTION AND H_{ext} -ASSISTED TESTS

The march algorithm based on the conventional approach is presented in Eq. 6.4; the march algorithm of the H_{ext} -assisted approach is presented in Eq. 6.14 or Eq. 6.15. Next, we compare the two test methods in terms of defect coverage and test time.

- **Defect coverage:** The conventional functional tests in Eq. 6.4 can never guarantee 100% coverage of the SAFF defects. It relies on comparing the difference of WER between defective and defect-free MTJs; this difference is caused by E_{gap} in Eq. 6.12, which can be small at $H_{ext}=0$. Besides, the write error occurs stochastically. Consequently, testing SAFF purely by the conventional approach results in escapes. On the other hand, the H_{ext} -assisted test applies H_{ext} to enlarge E_{gap} ; it switches defect-free MTJs while stabilizing the defective ones. Hence, this approach can guarantee 100% coverage of the SAFF defect.
- **Test time:** Eq. 6.4 only performs read and write operations, and the conventional march test has a length of $(1 + 3i) \cdot N$, where ' i ' can be very large considering the low WER_{spec} . The test time of Eq. 6.14 or Eq. 6.15 can be divided into two parts: 1) the time for performing write/read operations, with a longer length of which is $2N$; and 2) the time of applying H_{ext} , which takes several microseconds (e.g. 10 us in our experiments). Here we compare the test time of the two tests by an example: when the pulse width of read and write operation is both 10 ns, the test time applying Eq. 6.4 to a 256×256 STT-MRAM array is calculated as:

$$256 \cdot 256 \cdot 4 \cdot 10 \text{ ns} \approx 2621 \text{ us}, \quad \text{when } i = 1 \quad (6.16)$$

The test time applying Eq. 6.15 is calculated as:

$$256 \cdot 256 \cdot 2 \cdot 10 \text{ ns} + 10 \text{ us} \approx 1321 \text{ us} \quad (6.17)$$

Since ' i ' is much larger in reality, the H_{ext} -assisted test realizes much higher time-efficiency than the regular functional test.

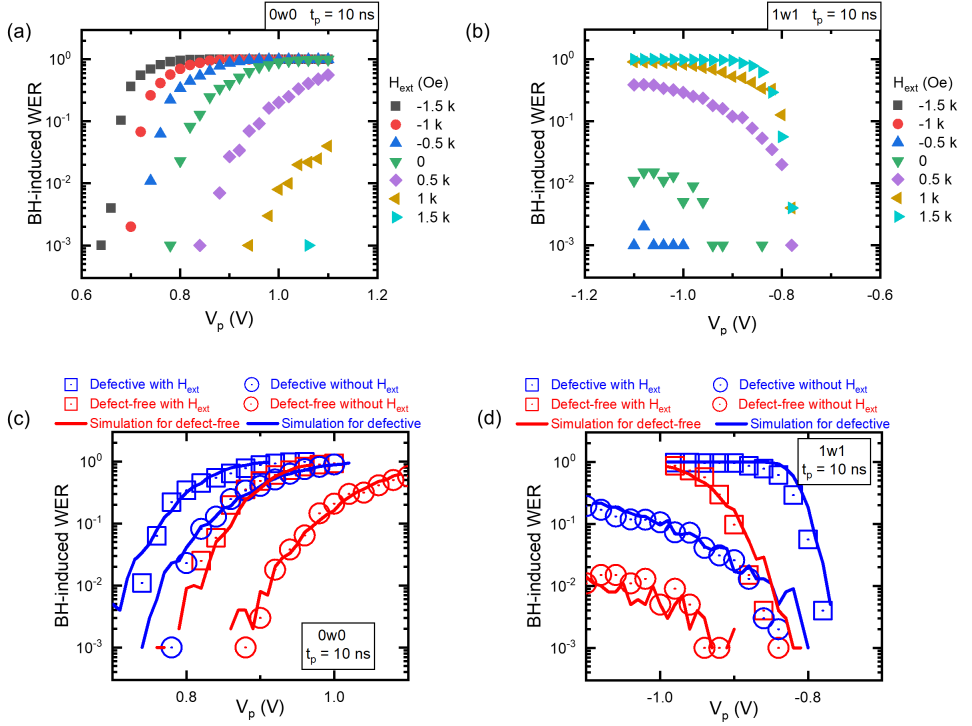


Figure 6.18: H_{ext} impact on BH-induced WER with the (a) $0w0$ operation; (b) $1w1$ operation; (c) Comparison of H_{ext} impact on BH-induced WER between BH-defective and defect-free MTJs with the (c) $0w0$ operation; (d) $1w1$ operation.

6.6.2. BH TEST WITH EXTERNAL MAGNETIC FIELD

Next, we follow the steps in Fig. 6.16, to design H_{ext} -assisted test for BH. First, we analyze the BH impact on four components in Eq. 3.14 and Eq.3.15. Since the defect does not affect the material property of either FL or RL, it does not affect the four components. However, due to the four-phase loop of BH occurring, we must consider the input energy E for RL and FL switching in each phase separately to analyze the H_{ext} impact on BH-induced WER. Here, we apply Eq. 3.15 to each phase in the $0w0$ operation, presented as follows:

$$\begin{aligned}
 P1: t_{w,1} \cdot (I \cdot \eta + (-H_{ext} - H_s - H_{k,RL} - H_p) \cdot \epsilon_{RL}) &= E_{B,RL} \\
 P2: t_{w,2} \cdot (I \cdot \eta + (-H_{ext} - H_s - H_{k,FL}) \cdot \epsilon_{FL}) &= E_{B,FL} \\
 P3: t_{w,3} \cdot (I \cdot \eta + (+H_{ext} + H_s - H_{k,RL}) \cdot \epsilon_{RL}) &= E_{B,RL} \\
 P4: t_{w,4} \cdot (I \cdot \eta + (+H_{ext} + H_s - H_{k,FL}) \cdot \epsilon_{FL}) &= E_{B,FL}
 \end{aligned} \tag{6.18}$$

Here, $H_{k,FL}$ & $H_{k,RL}$, $E_{B,FL}$ & $E_{B,RL}$, ϵ_{FL} & ϵ_{RL} refer to H_k , E_B , and ϵ of FL and RL separately; H_p is introduced in the model, which is the only parameter affected by BH (i.e., represents the defect strength) $t_{w,i}$, $i = (1, 2, 3, 4)$ indicates the t_w for each phase in Fig. 6.7 (a), which will occur sequentially to form the loop. For example, the MTJ state is

initially in 'P', and the RL is switched after t_{w_1} . Subsequently, the FL is switched further after t_{w_2} , and so forth. After t_{w_4} , the MTJ returns to the initial 'P' state, and the loop is restarted. In a single write operation, each phase may occur multiple times. To improve the time-efficiency of detecting BH, it is critical to increase the BH-induced WER, which is determined by the distribution of t_{w_i} . For example, if t_{w_1} and t_{w_2} become larger, t_{w_3} and t_{w_4} become smaller, BH-induced WER is reduced.

For the convenience of the following analysis, we define:

$$\gamma_i = \frac{E_{B_RL \text{ or } B_FL}}{t_{w_i}}, i = (1, 2, 3, 4) \quad (6.19)$$

$$\Gamma = \gamma_1 + \gamma_2 - \gamma_3 - \gamma_4 \quad (6.20)$$

Increasing BH-induced WER implies enlarging Γ , which can be achieved by applying the H_{ext} with a certain direction. When we apply a certain $H_{ext} = H_1$, the impact of H_{ext} on Γ is defined as $\Delta\Gamma(H_1) = \Gamma(H_{ext} = H_1) - \Gamma(H_{ext} = 0)$:

$$\Delta\Gamma(H_1) = -2 \cdot H_1 \cdot (\epsilon_{FL} + \epsilon_{RL}) \quad (6.21)$$

For the $0w0$ operation, if $H_1 < 0$, $\Delta\Gamma(H_1) > 0$, Γ increases, and BH-induced WER increases. Fig. 6.18 (a) presents the H_{ext} impact on BH-induced WER in the $0w0$ operation. With the same V_p , a higher BH-induced WER is observed with the negative H_{ext} ; BH-induced WER increases with the $|H_{ext}|$ increasing. Notice that when we apply a large positive H_{ext} , few BH-induced write errors occur because the H_{ext} stabilizes the RL and avoids Phase 1 taking place; hence, related lines are not shown in the figure. A similar analysis can be performed on the BH occurring in the $1w1$ operation. Fig. 6.18 (b) presents the H_{ext} impact on BH-induced WER in the $1w1$ operation, where a positive H_{ext} incites a higher WER at the same V_p .

Fig. 6.18 (c) compares the H_{ext} impact between the BH-defective and defect-free MTJs in the $0w0$ operation. At the same V_p , BH-induced WER of both MTJs increases under the same negative H_{ext} . A large gap is observed between the two MTJs with or without H_{ext} at a proper V_p . For example, the BH-induced WER of defective MTJ is one order of magnitude higher than that of the defect-free one at $V_p = 0.8V$, either with or without H_{ext} . A similar phenomenon is presented in Fig. 6.18 (d) with the $1w1$ operation. The simulation result, fitted with the measurement data, is also presented in figures.

Based on the analysis above, the H_{ext} -assisted test for BH follows three steps:

- Pre-calibration:
 - Find the threshold between defect-free and BH-defective MTJs, i.e., BH-induced WER = WER_{spec} in the $0w0$ operation.
 - Select the proper H_{ext} . Here, a negative $H_{ext} = H_{1-}$ is applied.
 - Find the new threshold between defect-free and BH-defective MTJs under H_{1-} , i.e., BH-induced WER = WER'_{spec} in $0w0$ operations with H_{1-} .
- Perform $\{\uparrow (w0r0)^n\}$ with H_{1-} .

The first step selects the proper H_{1-} with associated WER'_{spec} , which can be achieved by simulations. For example, here we select $H_{1-} = -0.5\text{kOe}$ and $WER'_{spec} = 30\%$, as presented in Fig. 6.18 (c). The second step extracts the BH-induced WER under H_{1-} , and compares it with WER'_{spec} ; if the extracted WER $> WER'_{spec}$, BH is detected. A simplified march algorithm is presented as follows:

$$\{\Downarrow (w0, r0)_{H_{1-}}^n\} \quad (6.22)$$

The march length of the march algorithms is $(2 \cdot i) \cdot N$, where ' n ' is determined by WER'_{spec} . A similar process can be performed with $1w1$ operations. A positive $H_{ext} = H_{1+}$ is applied, and the march algorithm is presented as follows:

$$\{\Downarrow (w1, r1)_{H_{1+}}^n\} \quad (6.23)$$

COMPARISON OF FUNCTION AND H_{ext} -ASSISTED TESTS

The march algorithm of the conventional test is presented in Eq. 6.7; the march algorithm of the H_{ext} -assisted test is presented in Eq. 6.22 and Eq. 6.23. Next, we compare the two test methods regarding defect coverage and test time.

- **Defect coverage:** Due to the intrinsic stochastic mechanism of write errors, both two methods cannot guarantee 100% coverage of the BH defect.
- **Test time:** The march length of the march algorithms in both methods is $(2 \cdot n) \cdot N$, where ' n ' is determined by either WER_{spec} or WER'_{spec} , as presented in Eq. 5.9. Assuming both testing methods are required to reach the same defect coverage (e.g., the same P_d in Eq. 5.9), since $WER'_{spec} > WER_{spec}$, a much shorter test time can be realized by H_{ext} -assisted tests. For example, if $WER_{spec} = 10^{-4}$, $WER'_{spec} = 0.3$, to guarantee $P_d = 99.9\%$, $n = 70000$ by functional tests, and $n = 13$ is required by H_{ext} -assisted tests. The test time is reduced by 99.98% by H_{ext} -assisted tests.

6.6.3. ADVANTAGES AND SHORTCOMINGS

Table 6.5: H_{ext} -assisted tests for different defect types in STT-MRAM testing.

Defect	Functional tests	H_{ext} -assisted tests	Advantages of H_{ext} -assisted test
Pinhole [173]	$\{\Downarrow (w1)^n; \Downarrow (r1)\}$ [161]	Unavailable	\
SAFF [87]	$\{\Downarrow (w1); \Downarrow (w0, r0, w1)^n\}$ [87]	$\{\Downarrow (w1), H_{1+}, \Downarrow (r0)\}$	Improve defect coverage, reduce test time
Intermediate States [137]	$\{\Downarrow (w0, r0, w1, r1)^n\}$ [81]	$\{\Downarrow (w0, r0, w1, r1)_{H_{1-}}^n\}$	Reduce test time
BH [80]	$\{\Downarrow (w1, r1)^n\}$ [83]	$\{\Downarrow (w0, r0)_{H_{1-}}^n\}$	Reduce test time

Table 6.5 summarizes the H_{ext} -assisted tests for all known HtD faults. In the presence of some strong interconnect and contact defects, $E < E_B$ must occur, implying functional tests present full coverage and high time-efficiency in detecting these defects [75], [100]. For Pinholes, the defect mechanism (i.e., MgO breaking-down process) is not fully understood, hence the third step in the H_{ext} -assisted test framework cannot be performed.

Advantages of H_{ext} -assisted tests are stated as follows:

- **Improve defect coverage and/or reduce test time without costs of chip area:** Table 6.5 demonstrates that H_{ext} -assisted tests can realize a high defect coverage and a short test time without costs of additional chip area. Especially, since H_{ext} can be applied to devices in the whole chip simultaneously, which further reduces the test time (see the case of detecting SAFF). To realize the same defect coverage or time-efficiency, DfTs with specific circuit designs are usually applied [76], [82], [147], [149], which requires additional chip area. For example, the DfT for BH in [82] requires an additional current mirror and a NOR gate for each column. Notice that, in large-scale manufacturing, a small saving on each chip (e.g., saving chip area) suggests a significant overall cost saving.
- **Flexible in test design** The H_{ext} can be continuously applied by ATE [127], which offers high flexibility in the test design; any H_{ext} value (not exceeding the maximum) can be applied in tests for STT-MRAM. On the other hand, DfTs with specific circuits have relatively low flexibility [158], [159]. For example, the DfT ‘weak write’ can offer only a few choices of different write pulse heights [159].

Shortcomings and costs of H_{ext} -assisted tests are stated as follows:

- **Cost of equipment:** To perform write/read operations and apply H_{ext} simultaneously, specific ATE is required [127], implying additional cost for the test equipment.
- **Time-compatibility:** Switching between different magnetic fields requires a longer time (e.g., several microseconds for current ATE [127]) than applying operations (e.g., 10 ns in this work), hence some H_{ext} -assisted tests may cause low time-efficiency. For example, the march algorithm $\{\uparrow (w0_{H1}, w1_{H2})\}$ applies $w0$ with $H_{ext} = H_1$, then immediately applies $w1$ under $H_{ext} = H_2$. This algorithm necessitates a long waiting period of several microseconds for switching H_1 to H_2 , despite the write operation requiring only 10 ns; hence, it indicates low time efficiency. On the other hand, the march algorithm $\{\uparrow (w0_{H1}, w1_{H1})\}$ applies both write operations with the same H_1 , avoiding switching H_{ext} in the whole test process, hence with higher time-efficiency.

7

TESTING STT-MRAMs: FROM DEFECTS TO TEST CHIPS

STT-MRAM has shown significant potential for commercialization, yet its susceptibility to defects during manufacturing remains a critical barrier to large-scale production. A high-quality test solution is crucial for the development of STT-MRAM manufacturing. Previous works have significantly contributed to STT-MRAM testing, but often focus on specific defects without providing a comprehensive framework. This chapter summarizes and integrates previous chapters, proposing a structured testing methodology for an industrial STT-MRAM chip, i.e., a case study. Starting from defects, we analyze their corresponding faults and evaluate potential test solutions, including March tests and DfT. Notably, we systematically assess various DfT techniques (to the best of our knowledge) and present a design implementing the selected DfT in an STT-MRAM array, covering circuit-to-layout integration to demonstrate practical feasibility.

7.1. TARGETED DEFECT

Testing is crucial for ensuring the quality and reliability of STT-MRAMs, as defects can significantly impact their performance and yield. Designing an effective STT-MRAM testing strategy begins with identifying the target defects. These defects can be categorized into two types, i.e., conventional defects and unique defects, each requiring different considerations due to their distinct origins and fabrication processes. Conventional defects typically stem from CMOS fabrication steps, whereas unique defects arise from the additional MTJ fabrication process, which involves specialized materials and techniques. Considering these differences, this section discusses conventional and unique defects separately and identifies the targeting defect types.

7.1.1. TRANSISTOR, CONTACT, AND INTERCONNECT DEFECTS

Conventional defects, such as transistor, contact, and interconnect defects, have been widely observed in traditional memories [61]. Therefore, it is reasonable to expect their presence in STT-MRAM as well. The exact locations of these defects depend on the circuit design and layout of the STT-MRAM array. For example, in the structure of word-oriented memory systems, the location of bridge defects may differ from those in bit-oriented memories [133], [174]. In this chapter, we apply the bit-oriented STT-MRAM array, as presented in Fig. 2.3, while all potential defect locations are illustrated in Fig. 5.1. Notice that we only consider defects in the STT-MRAM array; defects in peripheral circuits are out of the scope for this thesis.

7.1.2. TARGETED MTJ DEFECTS

Determining the targeted MTJ defects is more challenging than conventional defects, as it involves the complexities of the MTJ fabrication process, which is still far less mature compared with CMOS fabrication, and the intricate working mechanisms of MTJs. Unlike conventional defects, which can be inferred based on well-established testing of traditional memories, MTJ defects require a deeper understanding of their mechanism, origin, and impact on MTJ performance.

To identify the relevant MTJ defects, we first perform characterization on prototype MTJs, i.e., single MTJs that allow flexible electrical and magnetic measurements. These measurements are presented in chapter 4 and chapter 6. Through these characterizations, we confirm the presence of four previously studied defects: pinhole, Synthetic Anti-Ferromagnet Flip (SAFF), Intermediate State (IM), and Back-hopping (BH). However, determining which defects should be considered further requires additional analysis. Especially, by utilizing the defect models (see chapter 4) and incorporating peripheral circuits (see chapter 2), we investigate the behavior of defective MTJs during regular read/write operations.

Our analysis reveals that pinhole, SAFF, and BH may impact the MTJ functionality, making them critical defects to address. However, the IM defect behaves differently. This defect primarily occurs in large-area MTJs (eCD larger than 60 nm) [86]; Given that fact that the MTJ size in industry tends to become smaller than 60 nm, the likelihood of IM occurring can be ignored. Moreover, studies show that the probability of IM is negatively correlated with the portion of the input torque exceeding the barrier energy, following a normal distribution [81]. In the designed WD and write operations used in this

work (which are shown in chapter 2), which produce an overdrive to guarantee successful write operations, the IM-induced WER remains below the specified WER spec. As a result, IM is only discussed in the part of fault modeling, while ignored in the part of test design, since faults sensitized by IM can be tolerated based on the peripheral circuit design. Therefore, this chapter focuses on three MTJ defects: pinhole, SAFE, and BH.

7.2. FAULT MODELING

7.2.1. SIMULATION SET UP

The simulation platform consists of 4×4 STT-MRA array, as shown in Fig. 2.3, with Fig. 2.5 and Fig. 2.6 illustrating the WD and SA used in this study with $V_{DD} = 0.9V$, as describe in Chapter 2. The pulse width of both write and read operations is 10 ns. Cadence Spectre is adopted for circuit simulations, and the library of 40 nm TSMC is applied. The magnetic coupling is incorporated into the array by transferring the impact of the stray field to a virtual current source, as described in Chapter 5 [83]. The MTJ compact model in [112] is applied, which is calibrated with measurement data of MTJs with the eCD of 60 nm and the pitch of 90 nm, as presented in Chapter 2. The pulse width of both write and read operations is 10 ns, with V_{DD} of 0.9V

In this work, part of the simulations involve process variation, which is incorporated in both MTJs and MOSFETs. Process variation in MTJs is introduced through the eCD, which follows a normal random function with 10% away from its nominal value at 3σ corners. The process variation in MOSFETs is introduced through V_{th} , assuming it also follows a normal random function with 10% away from its nominal value at 3σ corners.

Table 7.1: Collapsed fault set for conventional defects.

Experiment	Baseline analysis		Static analysis		Dynamic analysis	
	Baseline	Stochastic-Only	Full Variability	Stochastic-Only	Full Variability	
Platform (model)						
Collapsed fault set	$\langle 0r0/0/1 \rangle, \langle 1r1/1/0 \rangle,$ $\langle 1c; 0r0/0/1 \rangle, \langle 1c; 1r1/1/0 \rangle,$ $\langle 1c; 0r0/0/1 \rangle, \langle 1c; 1r1/1/0 \rangle,$ $\langle 1c; 1w0/1/- \rangle$	$\langle 0r0/0/1 \rangle, \langle 1r1/1/0 \rangle,$ $\langle 1c; 0r0/0/1 \rangle, \langle 1c; 1r1/1/0 \rangle,$ $\langle 1c; 0r0/0/1 \rangle, \langle 1c; 1r1/1/0 \rangle,$ $\langle 1c; 1w0/1/- \rangle$	$\langle 0r0/0/1 \rangle, \langle 1r1/1/0 \rangle,$ $\langle 1c; 0r0/0/1 \rangle, \langle 1c; 1r1/1/0 \rangle,$ $\langle 1c; 0r0/0/1 \rangle, \langle 1c; 1r1/1/0 \rangle,$ $\langle 1c; 1w0/1/- \rangle, \langle 1c; 1w0/1/- \rangle, \langle 1c; 1w0/1/- \rangle$	None	None	
Notation	$\langle S_a; S_r/F/R \rangle$ denotes a coupling fault involving two cells; an aggressor cell C_a and a victim cell C_v ; this is an extension of the fault primitive notation in Section IV.A. C_a can be in the same column, same row or same diagonal as C_v ; hence $a \in \{c, r, d\}$. E.g., $\langle 0; 1r1/1/0 \rangle$ means that when the adjacent cell in the same column is 0, applying $r1$ operation to the victim cell (which is in 1) will not lead the cell to flip, but the read operation will return a wrong value 0. $\langle S_c; S_r/F/R \rangle$ denotes a multi-cell fault involving the victim cell and its eight neighboring cells. E.g., $\langle 1c; 1w0/1/- \rangle$ means that when all neighboring cells are 1, applying $w0$ operation to the victim cell (which is in 1) will fail. $'/'$ associated with fault effect F denotes the intermittent behavior. E.g., $\langle 1w0/1/- \rangle$ means applying the transition $w0$ operation will fail from time to time.					

7.2.2. FAULTS FOR INTERCONNECT AND CONTACT DEFECTS

Although quite some work has been reported on fault modeling for conventional defects in STT-MRAMs [74], [83], [100], [107], [122], [153], they failed to consider the effect of MTJ intrinsic stochasticity and process variability, combined with different DBs in the presence of a defect. In order to address these shortcomings and illustrate the dependency of fault modeling on accurate simulation platforms incorporating the inherent properties of STT-MRAMs, we will perform three experiments while considering three simulation platforms (i.e., models of MOSFETs and MTJs). The three simulation platforms are:

- *Baseline Model*: The 4x4 STT-MRAM simulation setup is described as above, without incorporating the MTJ intrinsic stochasticity or process variation.
- *Stochastic-Only Model*: The same as Baseline Model, augmented by incorporating the intrinsic stochasticity in the MTJ compact model. The model does not account for process variation.
- *Full Variability Model*: The same as Baseline Model, incorporated with the intrinsic stochasticity and process variation in the MTJ compact model; the MOSFET is also incorporated with process variation.

The three experiments are:

1. *Baseline analysis*: simulation based on static fault analysis (i.e., sensitizing operations consist of one operation at most) using the Baseline Model. This step shows the simplest case of fault modeling. Each time we inject only one defect, perform static operations, and observe the corresponding faulty behaviors. The impact of DB (including magnetic coupling) is considered.
2. *Static analysis*: Static fault analysis is applied using both Stochastic-Only Model and Full Variability Model, while also considering different DBs
3. *Dynamic analysis*: Dynamic fault analysis (i.e., sensitizing operations consist of two consecutive operations) is applied using both Stochastic-Only Model and Full Variability Model, while also considering different DBs Sequences in this experiment include static operation and two-operation, as presented in Table 4.3. Then, static faults are excluded in this experiment to focus on the analysis of dynamic faults.

Each of the above three experiments was performed for interconnect and contact defects by modeling them as linear resistances. We simulated all possible opens, shorts, and bridges while swiping the value of the resistance from $1\ \Omega$ to $1\ \text{M}\Omega$. In total, we simulated 23 resistive defects within a cell (9 opens, 8 shorts, and 6 bridges), and 17 bridges between cells (including adjacent cells in the same column, row, and diagonal).

Table 7.1 gives a summary of *collapsed fault set*, being the *minimum set* of faults that have to be detected to guarantee the detection of all possible resistive defects. Notice that the same defect type may sensitize multiple faults; the selected fault is the one that covers the maximum defect coverage (range of resistance). It is worth noting that there are EtD faults as well as HtD faults in the list.

In the rest of this section, we will illustrate results for the open defect OC2 (see Fig 5.1, which is a contact defect between the MOSFET and the MTJ) for three experiments as an example.

BASELINE ANALYSIS

Fig. 7.1 (a) shows the fault map of the open defect OC2. Six faults are sensitized (when considering the dependency on data-background DB); the 'red' boxes indicate defect strength ranges where faults are sensitized, and the 'green' ones indicate that no faults are sensitized. The column 'DB' indicates the case of DB, being the eight neighbors of

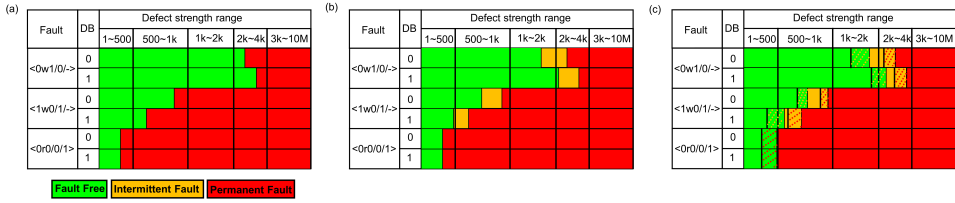


Figure 7.1: The fault map of OC2: (a) baseline analysis, (b) static analysis with Stochastic-Only Model, and (c) static analysis with Full Variability Model.

the defective cell; they are all 0 or all 1. The figure clearly shows three key points. First, many faults are sensitized to a single defect OC2. Second, although it is about a single cell defect, the DB has an impact on the defect coverage when faults are sensitized by write operations; for instance, when DB=1, the $1w0$ fails for a wide range of resistance values as compared with when DB=0. Third, the incorrect read fault $<0r0/0/1>$ is SB-independent, and its detection will maximize the defect coverage for the defect under consideration. Hence, it is best to include it in the collapsed fault set.

STATIC ANALYSIS

The static analysis is performed twice: once with the Stochastic-Only Model, and once with the Full Variability Model.

STOCHASTIC-ONLY MODEL

Fig. 7.1 (b) presents the fault map of the defect OC2 in this case. Three observations can be made. First, the same fault class is sensitized as in Fig. 7.1(a). Second, the range of undetected resistance defects reduces (green boxes). Write faults are also sensitized around the boundary resistance between fault-free cases and permanent fault cases, represented by the yellow boxes. However, the nature of such additional faults is *intermittent* due to the stochastic nature of MTJs. E.g., the fault denoted as $<1\Box; 1w0/1; ->$ and referred to as ‘Passive Neighborhood Pattern Sensitive Fault’, indicates that the $w0$ operation fails intermittently (i.e., from time to time) when all eight neighbors are set to 1; note that this takes place for relatively low defect strength. Third, the faults related to read operations are not affected by the stochastic nature of MTJs, as it does not require any switching.

FULL VARIABILITY MODEL

Fig. 7.1 (c) presents the fault map of the open defect OC2 in this case. The dashed yellow boxes indicate that, for such a resistance range, sometimes permanent faults are sensitized, and sometimes intermittent faults are sensitized; the green dashed boxes for write faults indicate that sometimes no faults take place, and sometimes intermittent faults occur; the green dashed boxes for read faults indicate that sometimes no faults take place, and sometimes permanent faults occur. The figure shows that at the boundary regions, the process variability affects the nature of the faults that will be sensitized. Nevertheless, the fault class generated in this case is the same as in Fig. 7.1 (a) and Fig. 7.1 (b) Note that the variability simultaneously lengthens or shortens the range of resistances

Table 7.2: Summarization of March tests.

Defect type	Defect	Fault	March name	March algorithm
Conventional	Interconnect and contact defect (IC)	EtD	IC-EtD	$\{\downarrow(w0); \downarrow(r0, w1, r1, w0, r0); \downarrow(w1); \downarrow(w0, r0, w1); \downarrow(r1)\}$
		HtD	IC-HtD	$\{\downarrow(w1); \downarrow(w0, r0, w1)^n\}$
Unique	Pinhole (PH)	EtD	PH-EtD	$\{\downarrow(w1)^{10}; \downarrow(r1)\}$ [161]
	SAFF	HtD	SAFF-HtD	$\{\downarrow(w1); \downarrow(w0, r0, w1)^n\}$ [87]
	IM	HtD	IM-HtD	$\{\downarrow(w0, r0, w1, r1)^n\}$
	BH	HtD	BH-HtD	$\{\downarrow(w1, r1)^n\}$ [82]

for which the faults are sensitized in the same way for all faults. For example, if the V_{th} of MOSFETs changes, causing a reduction in read/write currents, and all faults in Fig. 7.1 (c) will sensitize longer resistance ranges. Eventually, the selected fault for the collapsed set remains $\langle 0r0/0/1 \rangle$.

DYNAMIC ANALYSIS

The dynamic analysis is performed twice, once with the Stochastic-Only Model and once with the Full Variability Model, by simulating only resistance ranges that did not cause any fault in the static analysis (i.e., the green box region in Fig. 7.1). No single dynamic fault is observed in these regions. While we do not entirely rule out their existence, our findings suggest that dynamic faults are rare in STT-MRAMs compared to other memory technologies [70], [141]; this can be attributed to stable resistance states of MTJs. Defect-free MTJ has only two stable resistance states, and never shows intermediate states due to the anisotropy magnetic field [41]. Other non-volatile memories, like Flash, RRAMs, FeFETs, and PCRAMs, all have intermediate resistance states in their defect-free devices, which may introduce dynamic faults [156]. For example, when a bridge defect is present, a relatively large current may flow through the RRAM device during a read operation, which, while insufficient to fully switch the device state, slightly alters its resistance. The effects of multiple read operations accumulate and eventually switch the device, introducing a dynamic fault [141]. Yet this accumulation will never occur in STT-MRAMs [155].

Notice that previous works have demonstrated dynamic intermittent faults, like $\langle 0r0r0r0/1_i/1 \rangle$ in [74]. We verified the existence of these faults, yet none of them are in the green box region, hence not selected for the collapsed fault set. For example, the fault $\langle 0r0r0r0/1_i/1 \rangle$ occurs in the presence of a short defect; however, $\langle 0r0/0/1 \rangle$ sensitizes a longer strength range for this defect and is therefore selected for the collapsed set.

7.2.3. FAULTS OF UNIQUE DEFECTS

As mentioned in previous chapters, modeling unique MTJ defects using linear resistors fails to accurately capture corresponding faulty behaviors, leading to unrealistic fault models [79], [87]. We therefore introduced and have used the Device-Aware modeling approach for these unique defects; this approach incorporates the impact of the physical defect into the technology parameters of the device and thereafter into its electrical parameters to design the compact model of the defective MTJ. Such models are used to

Table 7.3: Fault model for unique defects.

Defect	Fault modeling
Pinhole [79]	Stuck at 0 fault: $\langle 1/0/- \rangle$
	Stuck at L fault: $\langle 1/L/- \rangle$
	Write disturb fault: $\langle 1w1/0/- \rangle$
	Undefined read disturb fault: $\langle 1r1/U/0 \rangle$
	Dynamic fault: $\langle 1w1w1w1w1w1/L/- \rangle$
SAFF [87]	Intermittent passive neighborhood pattern sensitive fault: $\langle 1\Box; 1w0/1_i/- \rangle$
IM [86]	Intermittent undefined write transition fault: $\langle 1w0/U_i/- \rangle$
BH [82]	Oscillating write fault $\langle 1w1/\sim/- \rangle$
	where ' \sim ' means MTJ state oscillating

replace defect-free MTJ models in the simulation platform, and the circuit simulation is performed while changing the strength of the defect through the change of a dedicated parameter in the model.

The results of our fault modeling for all unique defects discussed are summarized in Table 7.3; only the collapsed fault set per defect is presented [79], [82], [86], [87]. There are four key observations from the table. First, both static and dynamic faults take place in the presence of unique defects. Second, unique defects cause both permanent as well as intermittent faults. Third, unique defects also cause faults that bring the state of the cell into an undefined state or extremely low state (see Chapter 6). Fourth, the sensitized faults consist of both EtD faults (e.g., caused by pinhole) as well as HtD faults (e.g., caused by SAFF, IM, and BH).

7.3. MARCH TEST

In this section, March tests are designed for each targeted defect type. For the transistor, contact, and interconnect defects, we classify faults sensitized by these defects into two types: permanent faults and intermittent faults, as described in chapter 5. March tests are designed for these two types of faults separately. For permanent faults:

$$\begin{aligned} \text{March permanent} = \{ & \uparrow\downarrow (w0); \uparrow\downarrow (r0, w1, r1, w0, r0); \\ & \uparrow\downarrow (w1); \uparrow\downarrow (w0, r0, w1); \uparrow\downarrow (r1) \}. \end{aligned} \quad (7.1)$$

This march algorithm is expected to have full coverage of all sensitized faults.

For intermittent faults:

$$\text{March intermittent} = \{ \uparrow\downarrow (w1); \uparrow\downarrow (w0, r0, w1)^n \}. \quad (7.2)$$

Due to the stochastic nature of intermittent faults, coverage by applying this march algorithm can never reach 100%. Details of March tests for MTJ defects are described in chapter 6. Here, applying DfTs for HtDs (including intermittent) is the suggested testing method.

7.4. DFT METHODS

To further increase the probability of detecting HtD faults and/or reduce test time, researchers have been exploring two other alternatives than just using March tests, as shown in Table 7.4:

- Enabling ATE [128]: This approach explores the integration of magnets into ATE; hence, enabling the application of the external magnetic field (to perform write operations) during manufacturing test; this has been shown to guarantee the detection of some HtD faults such as those caused by SAFF defects, while significantly reducing the test time. To the best of our knowledge, there is only one work that reported this approach.
- DFT [82], [100], [107], [145]–[150]: This approach, which is more popular, integrates additional circuits to improve the coverage of HtD faults while optimizing the test time.

Next, we will first classify the proposed DfTs and provide a quantitative comparison considering different metrics (e.g., fault coverage). This will help us select the best DfTs to be adapted in our test chip.

Table 7.4: DfT method classification.

Enabling ATE	External magnetic field	Writing operation	[128]
DfT	Current Monitoring (CM)	During Normal Write (CM-NW)	[82], [107]
		During Normal Read (CM-NR)	[146]–[150], [162]
		During Specific Test Mode (CM-SM)	[100]
	Voltage Monitoring (VM)	During Normal Write (VM-NW)	
		During Normal Read (VM-NR)	[145]
		During Specific Test Mode (VM-SM)	

7.4.1. CLASSIFICATION

Based on their working mechanisms, proposed DfTs can be classified into two main classes, as presented in Table 7.4):

- Current monitoring (CM) based DfTs: the monitoring of the current can take place during normal write (NW) operation, during normal read (NR) operation, or during a special test mode (SM), resulting in three subclasses (CM-NW, CM-NR, CM-SM).

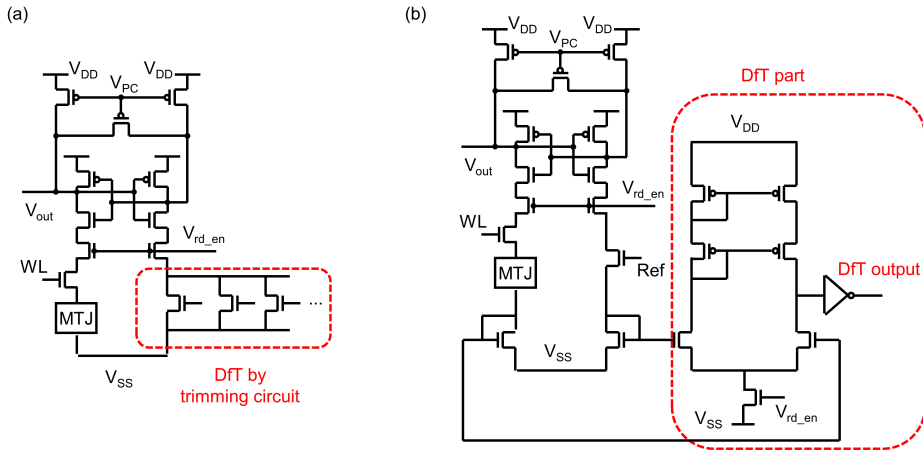


Figure 7.2: DFT design (a) in [76], and (b) in [150].

- Voltage monitoring (VM) based DfTs: the monitoring of the voltage can take place during normal write (NW) operation, during normal read (NR) operation, or during a special test mode (SM), resulting in three subclasses (VM-NW, VM-NR, VM-SM).

Note that Table 7.4 reveals that the DfT based on CM-NR is more explored than other classifications, since applying the SA can cheaply monitor the read current (e.g., the trimming-based DfT [148], [149]). On the other hand, monitoring the voltage or applying specific test modes is more expensive and unreliable, hence less explored. For example, the DfT in [100] applies specific clock signals, which significantly increase the cost of circuit design.

7.4.2. EVALUATION

To appropriately compare the proposed DfTs, we use the following metrics (see also Table 7.4):

- Fault coverage: This refers to the ability to detect the targeted defect or fault. It can be categorized as 'Full' (100%) or 'High', or 'Low' compared with March tests (if the March algorithm requires repeating n times, we assume $n = 100$). Using the proposed DfT is supposed to achieve a better fault coverage than March tests for the targeted HtD faults.
- Test time: the symbol '*' is used to show a qualitative comparison. For example, a DfT with two '*' requires a longer test time than a DfT with only one '*'.
- Flexibility: a flexible DfT allows for post-manufacturing calibration and tuning, i.e., enabling the DfT plug-in after completing the design of peripheral circuits.

- Design complexity: this indicates the effort required to design the DfT and its associated controller. The same '*' notation as used for test time is applied here for comparison.
- DfT impact: this includes the effect on area overhead, power consumption, and (read/write) latency. The same '*' notation as used for test time is applied here for comparison. '0' is used to indicate no effect.

Table 7.5 presents the result of our DfT comparison. Four observations can be made. First, among all defects that cannot be fully detected by March tests, only BH can be guaranteed by the dedicated DfT [82]; other defects, like IM, still cannot be fully detected. Second, some DfTs realize better coverage with a shorter test time. For example, the DfT in [148] enables a high coverage of weak contact defects with a short time (e.g., 20 ns), while applying March tests in Table 7.2 requires a long test time (e.g., 2 μ s) yet only realize a low coverage. Third, some DfTs enable a tuning post manufacturing (i.e., flexibility), which is preferred in chip design. For example, in the DfT of Fig. 7.2 (a), the trimming circuit is realized by easily adding additional MOSFETs in parallel to the reference current path. However, in the DfT of Fig. 7.2 (b), the current mirror (in the green square) is within the read path, making the design of the SA must include the impact of this current mirror. Fourth, some DfTs cause additional power overhead in regular read/write functions. An example, in the DfT design of Fig. 7.2 (b), the MOSFET of the current mirror consumes some power even in regular read operations; to realize reliable read operations, V_{DD} must be increased to enable a high enough read current.

Next, we will select the proper DfTs based on Table 7.4 and Table 7.5, and inject them into the chip

Table 7.5: DfT evaluation.

DfT	Classification	Target	Coverage	Test time	Flexibility	Design complexity	Impact		
							Area	Power	Latency
[146]	CM-NR	Weak bridge and short	High	*	No	*****	****	***	0
[147]	CM-NR	Weak bridge and short	High	*	No	*****	****	****	0
[148], [149], [162]	CM-NR	Weak contact, bridge, and short	High	***	Yes	*	*	0	0
		Pinhole	Low	**					
		IM	Low	****					
[150]	CM-NR	Weak bridge and short	High	*	No	*	**	****	0
[100]	CM-SM	Low retention MTJs	Low	*****	No	*****	***	*	0
[145]	VM-NR	Low TMR MTJs	Low	**	Yes	*	**	0	****
		Weak contact	Low	**					
[107]	CM-NW	Weak contact, bridge, and short	High	**	No	**	**	*	0
		Pinhole	Low	**					
[82]	CM-NW	BH	Full	**	No	**	**	**	0

'*' means the relative level of each criteria. E.g., in test time criteria, '*' means a very short test time while '*****' means a very long test time

7.4.3. SELECTION

We focus on selecting minimum DfT(s) to maximize the coverage of HtD faults and/or faults requiring longer test time; these are faults caused by the unique defects (i.e., SAFE, IM, and BH), as well as weak/small defects (i.e., weak bridges/shorts, small pinholes).

The content in Table 7.5 and 7.4 has led to the selection of DfTs in [82], [107], and [76]. To detect BH, the DfT in [82] is the only choice, which guarantees full coverage of

Table 7.6: Summarization of test solutions for each defect type.

	March test		DfT	
	March algorithm	Coverage	Method	Coverage
Strong conventional defects	Eq. 7.1	Full	Not necessary	
Weak conventional defects	Eq. 5.8	Larger 'n', higher coverage	Monitor read/write current [107], [145]–[150]	See Tab. 7.5
Pinhole	$\{\emptyset (w1)^n; \emptyset (r1)\}$ [161]	Full	Monitor read/write current [107], [148], [149], [162]	Low
SAFF	$\{\emptyset (w1); \emptyset (w0, r0, w1)^n\}$ [87]	Low, depends on 'n'	External magnetic field [128]	Full
IM	Not available	0	Monitor read current, and apply $\{\emptyset (w1, r1, w0, r0)^n\}$ [148], [149], [162]	Low, depends on 'n'
BH	$\{\emptyset (w1, r1)^n\}$ [82]	Larger 'n', higher coverage	Monitor write current, and add a flip-flop [82]	Full

the faults caused by BH. To detect IM, the DfT in [76] is the only choice, even though it does not fully cover this defect. This DfT also ensures a high coverage of weak conventional defects; however, it is based on monitoring the read current (see Table 7.4), hence only targeting weak conventional defects in the read path. To further include weak conventional defects in the write path, we include the DfT reported in [107], which monitors the write current. Notice that while applying March tests for SAFF results in a low coverage, no DfT has been reported for the detection of this defect. For our use case, we will test SAFF by using the proposed approach in [128] (i.e., 'enabling ATE' of Table 7.4, [127]); by applying a pre-defined external magnetic field to perform write operations, defective cells will flip while healthy cells will remain in their state.

In our test chip, we included only the DfT functionality as described in [82] and [107]. Due to the tight deadline for the test chip tape-out, we were unable to incorporate the DfT design from [76]. Since both integrated DfTs belong to the CM-NW class (see Table 7.4), we decided to combine them into an optimized, flexible DfT design that retains the functionality of both while reducing the circuit area. The details of this design will be explained next. The complete DfT consists of the functionality of DFT_1 (from [107]) and DFT_2 (from [82]).

7.4.4. IMPLEMENTATION

In brief, the two DfTs are combined and attached to each WD, they require three digital control signals for the DfT and n-bit digital configuration signals to select the reference and provide two independent digital outputs. The DfT scheme is depicted in Fig. 7.3. Details of the WD and DfT block are presented in Fig. 7.4. Fig. 7.6 shows the layout design of one page, and Fig. 7.7 zooms in to show the layout design of the DfT part. The DfT consists of four parts: 1) *Cont.*: the control part, 2) *WD + DfT*: the DfT is attached to each WD, 3) *Ref. gen.*: the reference current generator, and 4) *Output selector*.

Next, we explain the main function of each part of the DfT scheme as shown in Fig. 7.3.

CONT.

The *Cont.* provides four types of control signals:

- The V_{ci} connects to the *Ref. gen.*, selecting the proper reference current. The structure of the reference current generator is similar to that in [175]. An n-bit current generator can provide 2^n different reference currents. The resolution (i.e., preci-

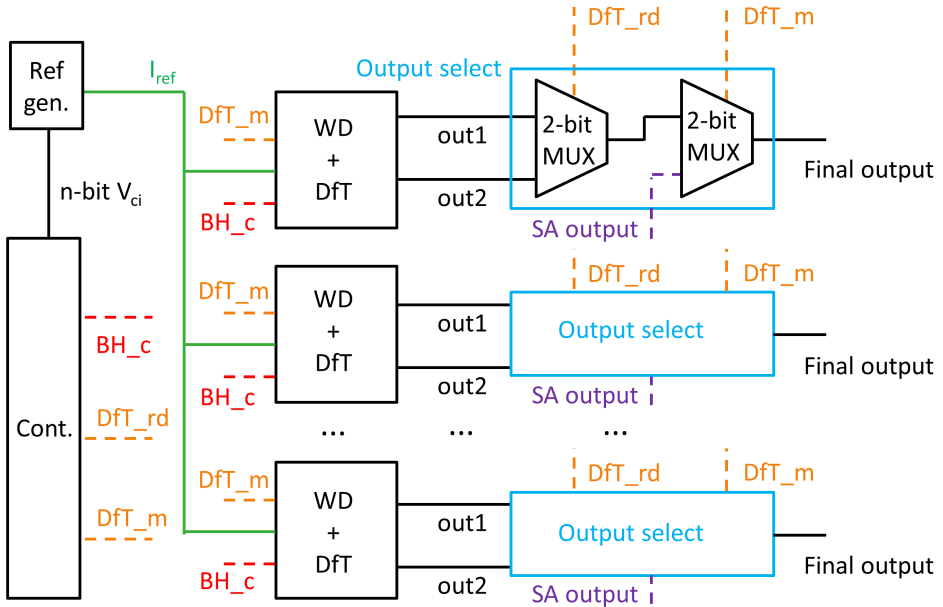


Figure 7.3: DfT scheme.

7

sion) of the current generator depends on the state-of-the-art manufacturing process for MOSFETs, in which the process variation must be considered.

- The DfT_m connects to the $WD + DfT$ and the *Output selector*, switching between test mode and regular mode. When this signal is '1', the DfT is closed, as the switch MOSFET (see Fig. 7.4) is turned on. Due to the channel resistance of the switch MOSFET being much smaller than that in the current mirror, the WD in this case can be viewed as directly connected to the ground, rather than the WD. Simultaneously, the 2-bit Multiplexer (MUX) selects the signal from the SA as the final output. In contrast, when this signal is '0', the switch MOSFET is turned off, and the STT-MRAM works in test mode.
- The DfT_rd connects to the *Output selector*, selecting which output of the two DfTs is used as the final output. When this signal is '0', it selects 'out1'; when this signal is '1', it selects 'out2'.
- The BH_c connects to the $WD + DfT$, which is the initialization for BH testing, details see chapter 6.

WD + DfT

The $WD + DfT$ is the combination of the WD and the two DfTs [82], [107], which is presented in Fig. 7.4 consists of five parts

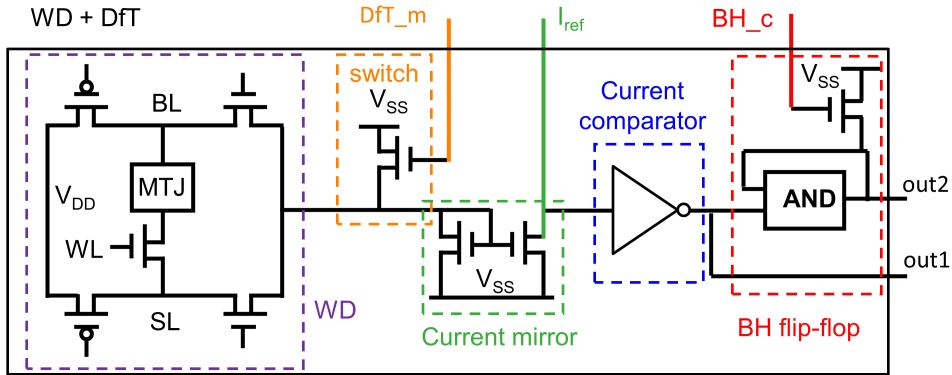


Figure 7.4: Structure of WD and DfT block.

- The basic WD, which is connected to the V_{SS} without DfT integration. With DfT integration, the write current of both w_0 and w_1 flows to the DfT.
- The switch (i.e., a large-size MOSFET), which is controlled by DfT_m . In regular mode, the DfT_m is '1', and the WD directs connect to V_{SS} ; In test mode, the DfT_m is '0', the WD connects to the MOSFET is closed, and the write current flows to the current mirror.
- The current mirror, which copies the write current.
- The current comparator, which compares the write current and a reference current. The weak contact defect is detected if the write current is smaller than a certain value.
- The BH flip-flop, which detects the MTJ state hopping during write operations

REF. GEN.

The *Ref. gen.* generates the reference current for the current comparator in *WD + DfT*, the circuit design of which is shown in [175]. The reference current is shared by all DfT blocks for each page, as presented in Fig. 7.6.

OUTPUT SELECTOR

The *Output selector* consists of 2 MUX, and connects to the DfT_m and the DfT_rd . Table 7.7 presents the relationship between the control signal and the final output:

7.4.5. TEST FLOW

As shown in Fig. 7.4, a current mirror copies the write current, which is then compared with a reference current I_{ref} . When the write current larger than I_{ref} , the 'out1' ends in V_{DD} (i.e., 1), otherwise the 'out1' ends in 0. If a $0w_0$ operation is applied, the 'out1' is expected to end in '1'. However, if a defect, such as the contact defect OC2 (see Fig. 5.2), reduces the write current, the 'out1' ends in '0'. On the contrary, a defect increases the

Table 7.7: DfT configuration

DfT_m	DfT_rd	Final output
0 (test mode)	0	out1
0 (test mode)	1	out2
1 (regular mode)	0	SA_output
1 (regular mode)	1	SA_output

write current, causing the ‘out1’ ends in ‘1’ with a $1w1$ operation, which ends in ‘0’ for the defect-free case.

To detect BH, the pre-charge signal BH_c is set to 0, forcing ‘out2’ to be ‘1’. A $0w0$ operation is applied, with ‘out1’ in ‘1’, then setting BH_c to ‘1’ to turn off the corresponding MOSFET. If the MTJ stays in ‘0’, ‘out2’ always keeps in ‘1’. However, if BH occurs and the MTJ hops to ‘1’, ‘out1’ turns to ‘0’, switching ‘out2’ to ‘0’. Due to the feedback loop between ‘out2’ and the input of the AND gate, ‘out2’ will stuck at ‘0’ even if the MTJ hops back to ‘0’.

The following table presents the test function of the DfT:

Table 7.8: Input and output of the DfT

Operation	out_1	out_2	Result
0w0	0	-	Detect weak short & bridge defect
	1	-	Pass
	-	0	Detect BH
	-	1	Pass
1w1	0	-	Pass
	1	-	Detect weak open & bridge defect and small pinhole
‘-’ means unimportant			

7.4.6. INTEGRATION

Fig. 7.5 presents the integration of the DfT, where cells of each column share the same WD, SA, DfT block, and the output select. The corresponding layout design is shown in Fig. 7.6. The peripheral circuit is on the left, the DfT & the *output selector* is in the middle, and the STT-MRAM array is on the right. The reference generator is included in the square of ‘DfT’ in the picture The BUS is at below, which connects the control part and the peripheral circuit/DfT. The chip has been sent out for tape out.

7.5. OVERALL TEST FLOW

In conclusion, the test flow of this chip follows three steps:

1. Applying March tests of Table 7.2 to detect EtD faults and potentially some HtD faults.

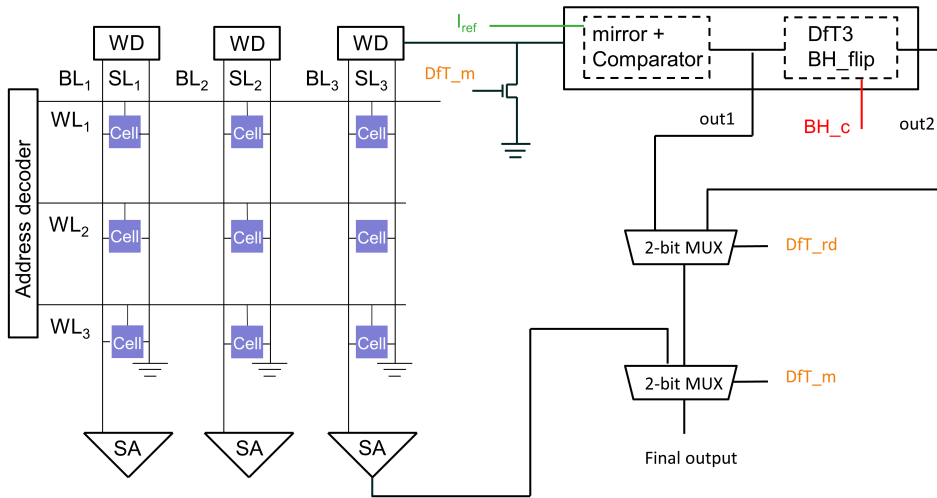
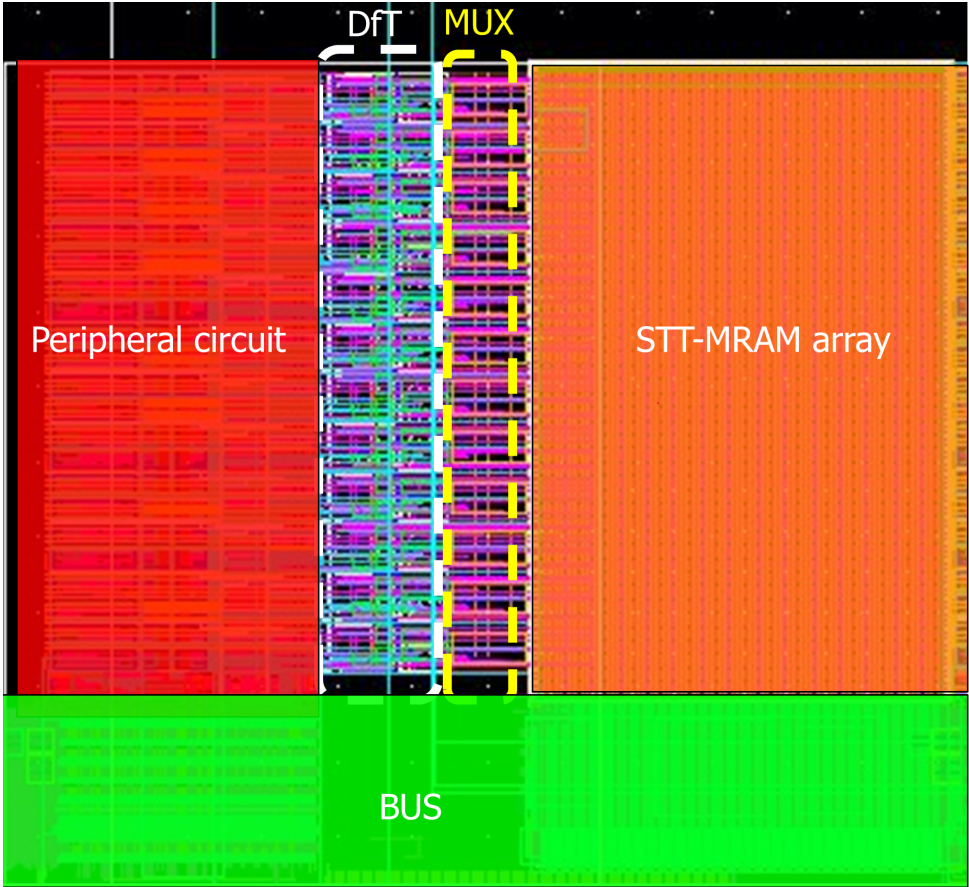


Figure 7.5: DfT integration to the STT-MRAM array.

2. Performing DfT to detect all HtD faults caused by BH, and increase the fault coverage of HtD faults by pinhole, weak conventional defects that escaped from step 1.
3. Applying ATE enabled by the external magnetic field to ensure the detection of HtD faults by SAFF that escaped from step 1.



7

Figure 7.6: Layout design.

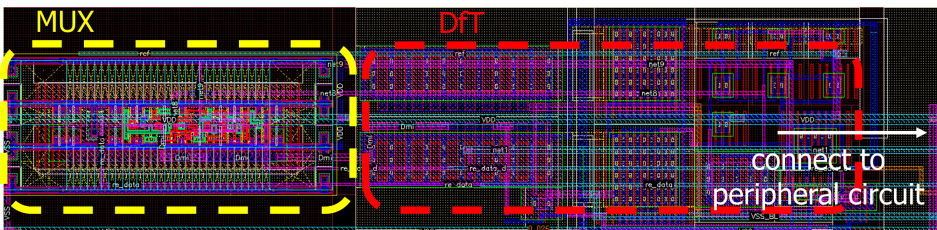


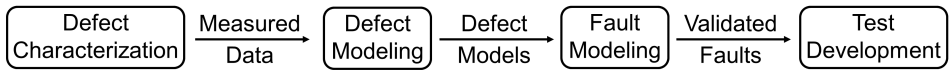
Figure 7.7: Layout design zooming in for the DfT part.

8

DIAGNOSIS OF MTJ DEFECTS

Obtaining high-quality diagnosis solutions for STT-MRAMs is challenging due to the existence of unique defects in MTJs. While the DAT method has been put forward as an effective approach mainly for detecting unique defecting STT-MRAMs, this chapter proposes a further advancement based on the DAT framework, introducing the Device-Aware Diagnosis (DA-Diagnosis) method. This method comprises two steps: a) defining distinctive features of each unique defect by characterization and physical analysis of defective MTJs, and b) utilizing march algorithms to extract distinctive features. The effectiveness of the proposed approach is validated in an industrial setting with real devices and data measurements.

(a) DA-Testing process



(b) DA-Diagnosis process

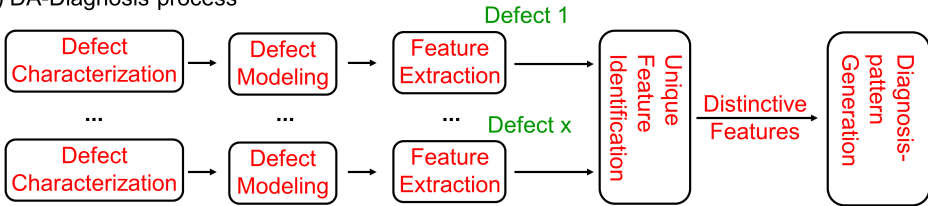


Figure 8.1: Framework of STT-MRAM test and diagnosis development.

8.1. TARGETED DEFECTS

This chapter focuses on diagnosing unique defects in MTJs, i.e., pinhole, Synthetic Anti-Ferromagnet Flip (SAFF), Intermediate State (IM), and Back-hopping (BH) [66], [81], [82], [87]. Beyond CMOS technology, which has undergone extensive refinement over decades, the fabrication of STT-MRAMs introduces additional steps specific to MTJs, as presented in Fig. 2.7. These include steps like deposition and annealing of magnetic materials, which remain less mature compared to well-established CMOS processes. As a result, MTJ fabrication is more susceptible to various factors like temperature, leading to unique defects that differ from conventional CMOS defects. Since these MTJ defects are closely tied to these fabrication steps (see Fig. 4.7), diagnosing them effectively can aid in yield learning and guide improvements in fabrication quality, ultimately enhancing the reliability and performance of the STT-MRAM system.

In the presence of unique MTJ defects, certain performance degradations may resemble those caused by conventional defects such as transistor, interconnect, or contact defects. For instance, the resistance reduction in defective MTJs may be caused by pinholes or open defects. This faulty behavior overlap makes it a challenge to distinguish MTJ defects from conventional defects. Any fault that could potentially be sensitized by conventional defects cannot serve as a reliable indicator for diagnosing unique MTJ defects. Instead, a more tailored diagnosis approach is required, one that accounts for the distinctive characteristics of MTJ defects and their specific impact on device behavior.

To address this challenge, we propose the DA-Diagnosis methodology, specifically designed for MTJ defects. Unlike traditional fault diagnosis, which primarily relies on standard fault models [177], DA-Diagnosis integrates the physical mechanisms of MTJ defects into the diagnostic process. By leveraging knowledge of MTJ fabrication steps, defect mechanisms, and the corresponding electrical and magnetic faulty behaviors, this method enables a more accurate identification of unique defects.

8.2. DEVICE-AWARE DIAGNOSIS METHODOLOGY

The DAT is optimized to detect faults, yet does not provide the distinctiveness for defects. For example, the March test algorithm designed for detecting Pinholes in [79] may detect

other defects such as BH. In this work, we introduce the DA-Diagnosis process, a new approach that identifies each defect through unique features and provides march-based diagnosis solutions. As presented in Fig. 8.1 (b), the DA-Diagnosis method follows five steps:

DEFECT CHARACTERIZATION

This step involves characterizing MTJs. In this work, we apply three common measurement methods: R-V measurement, R-H measurement, and WER extraction. For example, measurement data for defect-free devices are presented in Fig. 8.2 (a) - (c), compared with those for SAFF-defective devices in Fig. 8.2 (d) - (f).

DEFECT MODELING

This step designs models for manufacturing defects. In this work, we directly adopt defect models from previous works [79], [81], [82].

FEATURE EXTRACTION

This step extracts the features of defective devices based on the characterization and physical analysis. Here, features refer to the electrical or magnetic parameters that can be extracted from the measurement data. All possible features from the various measurement methods are extracted, as summarized in Table 8.1 in the next section. The first three steps will be repeated for the four unique defects and all possible conventional defects (by simulation).

UNIQUE FEATURE IDENTIFICATION

This step forms the feature dictionary by concluding all the features of both conventional and unique defects. Distinctive features are then extracted for each unique defect. Here, a distinctive feature refers to the parameter behaving uniquely to a specific defect, thus allowing for the diagnosis. For example, the reversed R-H hysteresis loop direction (reversed H_{dir}) extracted from the R-H measurement is the only distinctive feature of SAFE, as presented in Fig. 8.2 (b) and (e); while other measurements show the same features for both SAFF-defective and defect-free STT-MRAM cells.

DIAGNOSIS PATTERN GENERATION

This step generates the diagnosis solutions to recognize distinctive features. Here, we only apply the march-based diagnosis method, which is time-efficient and practical for STT-MRAM arrays. Other methods, like device characterization and Physical Failure Analysis (PFA), are not considered. Characterization methods for prototype MTJs are usually not practical for *STT-MRAM arrays*, such as the R-V measurement, where it is infeasible to directly extract the MTJ resistance from the structure shown in Fig. 2.3. PFA methods, like TEM, are usually costly, time-consuming, and destructive [178]. The final output of the DA-Diagnosis method is the march algorithm for diagnosing each unique defect. If regular march algorithms cannot diagnose certain types of unique defects, additional approaches are applied, like the external magnetic field. For example, the march-based diagnosis for SAFF requires the external magnetic field, which is presented as: $\{\uparrow(w0); \uparrow(w0_{Hext}); \uparrow(r0)\}$. This march algorithm consists of three steps: a) apply $w0$ to all devices for initialization, b) apply an external magnetic field for the whole chip to

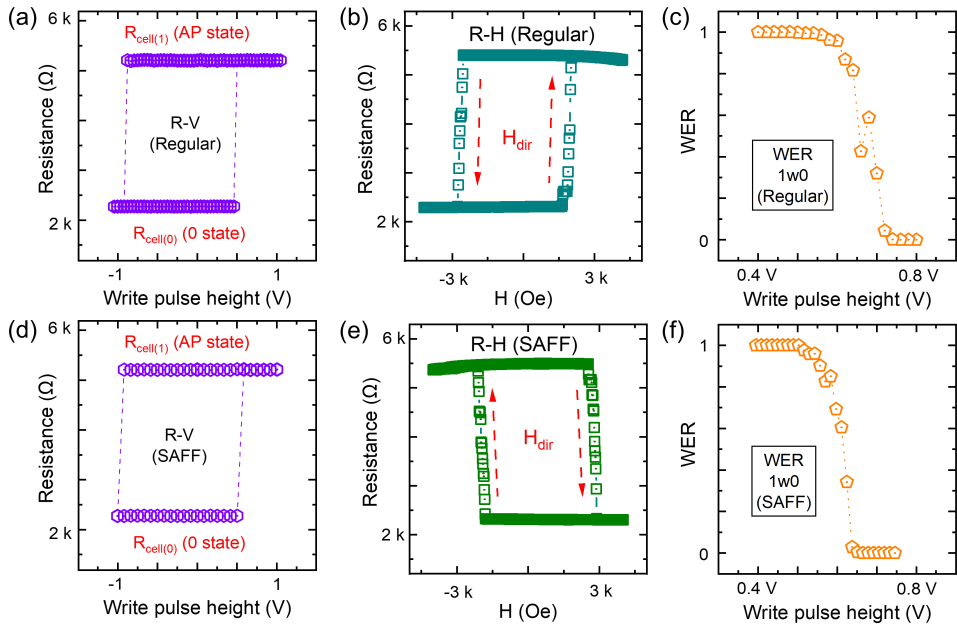


Figure 8.2: Characterization of defect-free device: (a) R-V, (b) R-H, (c) WER; Same characterization of SAFF-defective device: (d) R-V, (e) R-H, (f) WER.

perform the $w0$ operation, c) apply $r0$ operation to all devices (Details will be discussed later).

Notice that the process variation does not affect unique features of defects, hence, no impact on the diagnosis method.

8.3. APPLICATION

This section applies DA-Diagnosis for four targeted unique defects. The order we present the four unique defects depends on their frequency of occurrence in our measurement/test data (i.e., Pinhole, BH, SAFF, IM).

8.3.1. DEFECT CHARACTERIZATION

In this step, we first perform the three measurement methods (i.e., R-V, R-H, and WER extraction) on all the devices of 1 MB STT-MRAM cells; an example of results is shown in Fig. 8.2 for SAFF. Other measurement data is not shown to save space. On the other hand, we extract the MTJ faulty behaviors in the presence of all possible interconnect and contact defects through spice simulations, following the process in [141].

8.3.2. DEFECT MODELING

As device-aware defect models of the four targeted unique defects are reported in our previous works [79], [81], [82], [87], we therefore make use of them directly.

8.3.3. FEATURE EXTRACTION

In this step, we summarize all the features we extracted from the R-V, R-H, and WER measurement data for the four unique defects as well as the interconnect & contact defects; the results are reported in Table 8.1.

Table 8.1: Feature-dictionary of defects

	R-V			R-H	WER
	R_{cell-0}	R_{cell-1}	$R_{cell-IM}$	H_{dir}	
Pinhole	↓	↓	NA	NA	NA
SAFF	NA	NA	NA	RD	↓
IM	NA	NA	IB	NA	↑
BH	NA	NA	NA	NA	↑
Interconnects & contacts	↓, ↑, NA	↓, ↑, NA	NA	NA	↓, ↑, NA

Where the symbols used in the table are described below:

R_{cell-0}	Cell resistance in state '0'
R_{cell-1}	Cell resistance in state '1'
$R_{cell-IM}$	Cell resistance in intermediate state
H_{dir}	R-H hysteresis loop direction
WER	Write error rate
↑	Increase
↓	Decrease
RD	Reversed direction
NA	Not affected
↑	Higher or lower depending on neighboring cells
IB	Intermittent behavior

The last row presents all possible situations with the presence of interconnect and contact defects; for instance, $R_{cell}(0)$ may increase, decrease, or stay constant depending on the nature of the interconnect and contact defects. As an example, Fig. 8.2 shows the results for SAFE, from which the corresponding features are extracted, and included in Table 8.1.

8.3.4. DISTINCTIVE FEATURE IDENTIFICATION

This step extracts the distinctive features of each unique defect. A feature is considered distinctive if it behaves uniquely within the column of Table 8.1. For example, the IB of the $R_{cell-IM}$ is identified as the distinctive feature for IM; yet the WER increasing (↑) is not a distinctive one, since this behavior is also observed in the presence of BH and conventional defects. Consequently, the distinctive features of SAFF and IM can be directly derived from Table 8.1; H_{dir} is the distinctive feature of SAFF (i.e., RD), and and

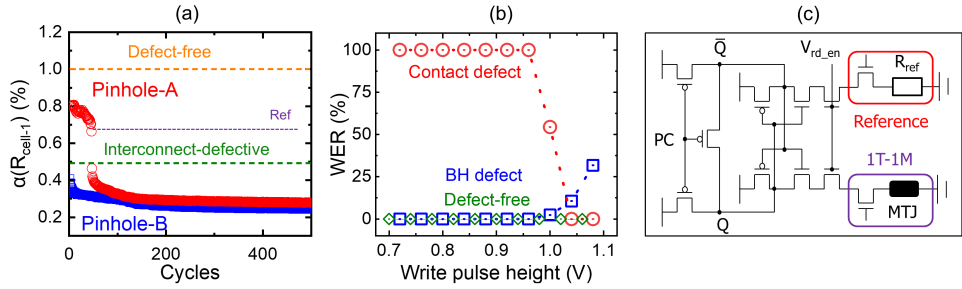


Figure 8.3: (a) Unique feature identification for Pinholes, (b) Unique feature identification for BH, (c) SA structure.

the $R_{cell-1M}$ is the distinctive feature of IM (i.e., IB). However, for Pinhole and BH, there is no distinctive feature that can be directly extracted from Table 8.1. Hence, there is a need for ‘secondary features’ extraction to ensure distinctive features(s) for each unique defect. Next, we present how we do this for Pinhole and BH:

PINHOLES

As presented in Table 8.1, the major feature of Pinhole is the reduced R_{cell-0} and R_{cell-1} . However, other defects may also exhibit the same feature, like an interconnect defect between BL and SL of the cell. To identify the distinctive feature of Pinhole, we rely on the aid of physical analysis. It has been demonstrated that the MgO of the Pinhole-defective MTJ is vulnerable. When applying the stress test by repeating write operations, the MgO of Pinhole-defective devices experiences further damage, thus further lowering the MTJ resistance [161]. In contrast, the stress test has a negligible impact on the MgO of defect-free MTJs, and on the MTJs with other defects. Therefore, we define the distinctive ‘secondary feature’ as $\alpha(R_{cell-1}) = R_{cell(measured)} / R_{cell(defect-free)}$, being the ratio of the measured R_{cell} compared with that of a defect-free R_{cell} when MTJ in the ‘1’ state; α refers to the ratio. Fig. 8.3 (a) compares $\alpha(R_{cell-1})$ behaviors under the stress test for pinhole-defective cells, a defect-free cell, and a cell with an interconnect defect. Clearly, $\alpha(R_{cell-1})$ of the Pinhole-defective cell decreases with more write cycles, yet $\alpha(R_{cell-1})$ of cells with other defects remains constant.

BACK-HOPPING

As presented in Table 8.1, a high WER is the major irregular feature for BH, yet this is not a distinctive one. Other defects, such as a contact defect between BL and the MTJ, can also cause a high WER. Hence, the ‘secondary feature’ is necessary. Here, we extract the WER with different write pulse heights. The ‘secondary feature’ is ‘ $\alpha(WER)$ ’, defined as how WER changes with the write pulse height increasing. For BH-defective MTJs, the range of $\alpha(WER)$ is limited considering the MTJ state oscillation; when the write pulse height increases, the WER undergoes only slight changes, without exhibiting extremely high or low values. In contrast, the WER of MTJs with other defects will either approach 100% or 0% with the increase of write pulse height. Fig. 8.3 (b) compares $\alpha(WER)$ behaviors of a BH-defective cell, a defect-free cell, and a cell with a contact defect. Clearly $\alpha(R_{cell-1})$

increases a bit for the BH-defective cell, but it decreases dramatically and eventually closes to 0 for the cell with a contact defect. Notice how WER changes with write pulse depends on various factors like the BH defect strength [82].

8.3.5. DIAGNOSIS-PATTERN GENERATION

This step generates the diagnosis patterns. The final output of this step is a set of march algorithms that can be practically performed on industrial STT-MRAM chips. Next, we present how to design the algorithm for each targeted unique defect. Table 8.2 summarizes the result of this section, which will be explained next.

Table 8.2: Conclusion of diagnosis for unique defects.

Defect	Mechanism	Related steps in STT-MRAM fabrication	Unique feature	Diagnosis patterns
Pinholes	Physical imperfections in MgO or FL/MgO interface [179]	MgO or FL deposition, annealing [179]	$\alpha(R_{cell-1})$	$\{\uparrow(w1, r1)^i\}$
SAFF	PL magnetization reversed[87]	Unknown	H_{dir}	$\{\uparrow(w0); \uparrow(w0_{Hex}); \uparrow(r0)\}$
IM	Non-unified FL [180]	Unknown	P_{IM}	$\{\uparrow(w0, r0, w1, r1)^i\}$
BH	RL instability [80]	PL deposition [80]	$\alpha(WER)$	<ol style="list-style-type: none"> 1. $\{\uparrow(w0, r0)^i\}$ 2. adjust write pulse 3. $\{\uparrow(w0, r0)^i\}$

PINHOLES

While $\alpha(R_{cell-1})$ serves as the distinctive feature of Pinhole, the STT-MRAM cell resistance cannot be directly extracted through the STT-MRAM array. For example, we utilize the march algorithm: $\{\uparrow(w1, r1)^i\}$ to diagnose Pinhole, indicating repeating $w1$ and $r1$ operations for 'i' times, where $i \geq 30$ according to our measurement. The effectiveness of this march algorithm depends on the SA structure and the defect strength (i.e., Pinhole area [79]). Our work applies the regular SA as given in Fig. 8.3 (c). If $R_{cell} \geq R_{ref}$, then SA gives '1', otherwise '0'. For the two Pinhole-defective devices depicted in Fig. 8.3 (a), only device-A can be diagnosed; the read operations initially provide correct results, followed by incorrect results, with repeating write operations. Yet, if the Pinhole-defective R_{cell} is initially lower than R_{ref} (e.g., device-B), read operations initially provide wrong results, making it impossible to distinguish between Pinholes and other defects. Applying multiple reference resistors in the SA, such as in [161], may diagnose a broader range of Pinholes, but it never guarantees to diagnose all Pinholes.

BACK-HOPPING

$\alpha(WER)$ is the distinctive feature of BH; it can be extracted from the STT-MRAM array in three steps: a) Set up the write pulse height V_p properly (i.e. spec); b) Apply march algorithm $\{\uparrow(w0, r0)^i\}$, where 'i' depends on the targeted WER; c) Repeat the two steps above yet with $V'_p = 2V_p$. The BH is diagnosed if read operations provide wrong results in both steps b) and c). Notice that we can also adjust t_p rather than V_p for BH diagnosis.

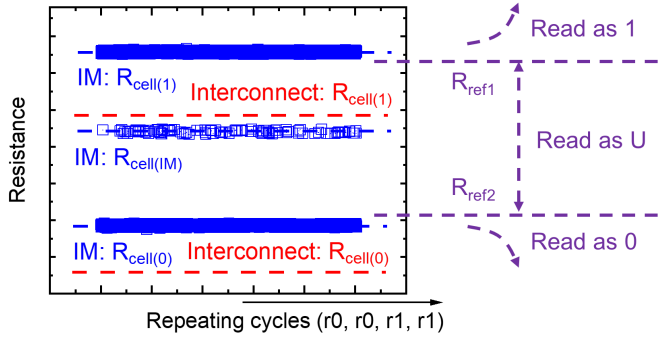


Figure 8.4: Diagnosing for IM.

SYNTHETIC ANTI-FERROMAGNET FLIP

H_{dir} is recognized as the distinctive feature of SAFE. However, the R-H measurement cannot be applied to regular STT-MRAM arrays, since the STT-MRAM cell resistance cannot be directly extracted. Hence, we apply the following algorithm to observe H_{dir} : $\{\hat{\uparrow}(w0); \hat{\uparrow}(w0_{Hext}); \hat{\uparrow}(r0)\}$. This march algorithm consists of three steps: a) apply $w0$ for initialization, b) apply an external magnetic field, and c) apply $r0$ to detect the final state. In the presence of SAFE, the state of the faulty cell will flip to '1'; hence diagnosis is guaranteed.

INTERMEDIATE STATE

$R_{cell-IM}$ is recognized as the distinctive feature of IM. However, detecting $R_{cell-IM}$ is limited by the SA circuit. The SA applied in this work (Fig. 8.3 (c)) can only detect two states, and $R_{cell-IM}$ will be read either as '0' or as '1'. Hence, IM can never be detected or diagnosed. To overcome this limitation, multiple reference resistors are required for the SA. For example, the results presented in Fig. 8.4 are obtained by such SA utilizing two reference resistors to detect three states: '0', '1', and 'U', representing the intermediate state. However, it is important to note that other defects, such as the interconnect defect between BL and SL, may also cause the defective STT-MRAM cell state to be located in the 'U' range (see Fig. 8.4). Therefore, diagnosing IM requires the detection of all three states. To achieve this, we employ the march algorithm: $\{\hat{\uparrow}(w0, r0, w1, r1)^i\}$, where $i \geq 10$ according to our measurement. Notice IM has an intermittent behavior. In the presence of contact or interconnect defects, the read operations will result permanently in 'U, 1' or in '0, U'. However, in the presence of IM, the read operations will result most of the time in '0, 1', and intermittently in 'U, 1' or '0, U'. Hence, the difference between read values will distinguish contact & interconnect defects from IM.

8.4. CONCLUSION

DA-diagnosis provides an effective method for accurately identifying most of the unique defects. Once the defect type is determined, it becomes possible to trace it back to specific fabrication steps, aiding in yield learning and process optimization.

However, some defects remain challenging to identify. For instance, pinhole defects

with high defect strength exhibit no unique features. Additionally, the diagnosis process can be time-consuming for certain defects. For example, diagnosing a BH defect requires at least extracting WER twice, which is significantly more time-intensive compared to a standard March test (e.g., Eq. 7.1). To balance accuracy and efficiency, a two-step approach is recommended: first, detecting defects and providing an initial evaluation of possible defect types; then, DA-diagnosis is applied selectively to precisely identify the defect type. This strategy significantly improves diagnostic efficiency while maintaining accuracy.

9

CONCLUSION

9.1. SUMMARY

Chapter 1: Introduction

This chapter serves as an introduction to this thesis. First, it provides an overview of emerging memory technologies, introducing different types of emerging memories. Then, it focuses on MRAMs, briefly describing the MRAM mechanism and different types of MRAMs, followed by an introduction to memory testing philosophy covering both traditional memories and emerging memories. Next, we outline the key research topics of this thesis, including defect modeling, fault analysis, high-quality test design, and MRAM defect diagnosis. Additionally, it highlights the contributions of the thesis, which include testing strategies for both conventional and unique defects, as well as the STT-MRAM dedicated diagnosis. Finally, the chapter concludes with an overview of the thesis organization.

Chapter 2: STT-MRAM technology

This chapter provides a comprehensive introduction to STT-MRAM technology, beginning with a detailed explanation of the MTJ structure and its working mechanism, which form the foundation of data storage and switching behavior in STT-MRAMs. It then introduces the structure of the 1T-1MTJ cell as the fundamental block of STT-MRAM arrays, followed by an exploration of the array organization and architecture, discussing how memory cells are interconnected and accessed. The chapter further examines the design and functionality of peripheral circuits, which are essential for enabling efficient read and write operations while ensuring memory performance and reliability. Afterward, it delves into the fabrication process of STT-MRAM. The discussion then extends to MTJ compact modeling, explaining how accurate models are developed and applied in conjunction with peripheral circuits to optimize memory design and performance evaluation. Finally, the chapter provides an outlook on the future prospects of STT-MRAMs.

Chapter 3: State of the art in STT-MRAM testing

This chapter presents the state of the art in STT-MRAM testing, providing a comprehensive overview of existing methodologies and advancements. It begins by introducing the

framework of test design in STT-MRAMs, outlining the key steps and strategies involved in ensuring high-coverage test solutions. Next, it discusses defect classification, distinguishing between conventional and unique defects while exploring their modeling approaches. The chapter then examines fault classification and modeling, addressing both conventional and unique faults and their corresponding faults (i.e., impact on memory performance). Finally, it delves into test generation techniques, highlighting how to design a high-efficiency test methodology while ensuring fault coverage.

Chapter 4: Testing of transistor, contact, and interconnect defects in STT-MRAMs

This chapter explores the testing of transistor, contact, and interconnect defects, which are considered conventional defects in STT-MRAMs. It begins with defect modeling, providing a complete defect map to categorize various defect types. Next, it examines fault modeling, analyzing the impact of data background, distinguishing between dynamic and static faults, and considering the effects of MTJ intrinsic stochasticity and process variation. The analysis reveals that static and intermittent faults are observed, while no dynamic faults are present. Finally, the chapter discusses test generation strategies aimed at minimizing test time while achieving high fault coverage, noting that static faults can be fully covered using March tests, whereas dynamic faults necessitate the use of DfT techniques.

Chapter 5: Testing of MTJ defect in STT-MRAMs

This chapter provides a thorough examination of the testing of MTJ defects in STT-MRAMs, starting with an overview of the existing unique defects that occur in MTJ structures and their corresponding test solutions. The chapter then focuses on the introduction of a new unique defect, BH, and explores the defect and fault modeling techniques used to analyze the faulty MTJ behavior in the presence of BH. Afterward, it presents test development for BH, including March tests and DfT approaches. The last part of this chapter is dedicated to the application of external magnetic fields in testing MTJ defects, demonstrating how this technique enhances defect coverage for SAFF testing and reduces test time for BH testing, thereby improving overall testing efficiency.

Chapter 6: Testing STT-MRAMs: from defects to test chips

This chapter presents a comprehensive test development approach for STT-MRAMs, spanning from defect analysis to test generations, ultimately creating a robust test structure for industry-grade STT-MRAM chips. It begins by targeting defects that need to be tested in STT-MRAM, followed by examining the faults associated with these targeted defects. Following this, it explores the use of March tests for detecting derived faults. A significant portion of the chapter is dedicated to reviewing existing DfT techniques, classifying them, and evaluating the effectiveness of each DfT. Finally, the chapter discusses how to select and integrate the most suitable DfTs into the STT-MRAM array to ensure efficient and reliable testing at the industry level.

Chapter 7: Diagnosis of MTJ defects

This chapter explores the diagnosis of MTJ defects, introducing the DA-Diagnosis method as an effective approach for identifying and analyzing MTJ defects in STT-MRAMs. It begins by presenting the methodology, which consists of three key steps: defect characterization and modeling to understand defect behavior, feature extraction to identify distinguishing signatures, and diagnosis-pattern generation to enable efficient defect detection. Finally, the chapter demonstrates the application of the DA-Diagnosis

method to four MTJ defects, presenting the generated diagnosis patterns and discussing their effectiveness.

9.2. FUTURE RESEARCH DIRECTION

Ensuring the quality and reliability of commercial STT-MRAMs requires effective solutions for defect detection. In this thesis, we presented a systematic and comprehensive methodology for developing high-quality test and diagnosis strategies for STT-MRAMs. Future research should focus on the following key directions:

ADVANCED CHARACTERIZATION AND MODELING OF MRAM DEFECTS

Future research is crucial to move beyond the current subset of known STT-MRAM defects. The focus must be on developing advanced, high-resolution physical and electrical characterization techniques to discover and analyze all possible unique MTJ defects. This requires not only finding new failure mechanisms but also precisely modeling their impact on device performance, variability, and long-term reliability. Research should leverage advanced statistical analysis and machine learning to correlate manufacturing process variations with specific defect signatures. The ultimate goal is to create a comprehensive defect catalog and corresponding physics-based compact models that allow for highly accurate circuit-level simulation and complete test coverage to guarantee the highest outgoing product quality.

EXPANDING THE APPLICATIONS OF DAT

While the Device-Aware Test (DAT) approach has shown impressive versatility across various emerging memories (STT-MRAMs, SOT-MRAMs, FeFETs, RRAMs), future work should concentrate on formalizing the device-aware framework into a standardized methodology. This involves developing systematic procedures for translating the intrinsic physical characteristics and failure modes of any new memory technology into effective test and diagnosis patterns. Research should explore applying the DAT principle to system-level reliability assessment, not just cell-level testing, by integrating device-aware fault models into system-on-chip test structures.

DEFECT REPAIR

Future work should focus on developing effective and efficient defect repair mechanisms to enhance STT-MRAM yield and lifetime. This involves exploring both physical redundancy techniques at the chip level, such as utilizing spare rows and columns that can be switched in upon defect detection, and architectural or algorithmic approaches like ECC that can mask a limited number of faults. A significant challenge is integrating these repair capabilities seamlessly without overly impacting performance or power consumption, especially given the density and unique failure modes of MRAM cells. Research should also investigate in-situ repair or annealing techniques that could potentially mitigate certain types of reversible defects post-manufacturing.

READ/WRITE SCHEME IMPROVEMENT

Future work should focus on developing effective and efficient defect repair mechanisms to enhance STT-MRAM yield and lifetime. This involves exploring both physical

redundancy techniques at the chip level, such as utilizing spare rows and columns that can be switched in upon defect detection, and architectural or algorithmic approaches like error-correcting codes (ECC) that can mask a limited number of faults. A significant challenge is integrating these repair capabilities seamlessly without overly impacting performance or power consumption, especially given the density and unique failure modes of MRAM cells. Research should also investigate in-situ repair or annealing techniques that could potentially mitigate certain types of reversible defects post-manufacturing.

RELIABILITY

A major direction for future research is a deep dive into the long-term reliability challenges of STT-MRAMs. This includes comprehensive studies on wear-out mechanisms such as time-dependent dielectric breakdown in the tunnel barrier, cycling endurance limits, and retention degradation under various operating and storage conditions. Research needs to develop accelerated testing methodologies that accurately predict device lifetime in real-world applications. Furthermore, the impact of process variations and scaling on reliability needs to be thoroughly characterized, necessitating the development of advanced physics-based models that can simulate and predict failure rates across different technology nodes.

YIELD LEARNING

Future research on yield learning should focus on closing the loop between defect detection/diagnosis and manufacturing optimization. This involves developing advanced data analytics and machine learning techniques to process the massive amounts of data generated from the developed test and diagnosis strategies. The goal is to rapidly identify the root causes of yield loss, linking specific defect signatures (e.g., as provided by the DAT approach) back to particular process steps or equipment issues in the fabrication line. This effort requires creating a holistic yield learning framework that provides actionable insights, enabling quicker iteration cycles for process refinement, thereby accelerating the time-to-market for high-quality commercial STT-MRAM products.

AI-DRIVEN MRAM TESTING

Future research must focus on integrating Artificial Intelligence and Machine Learning techniques to revolutionize STT-MRAM testing and diagnosis. The sheer complexity and volume of test data generated by advanced nodes, especially concerning statistical variations and unique defect modes, make traditional rule-based testing increasingly inefficient. Research should concentrate on developing smart, adaptive test algorithms that can learn from historical failure data to optimize test pattern selection, significantly reducing test time and cost. Furthermore, AI models should be used for predictive diagnosis—identifying subtle defect signatures that precede catastrophic failure—and for root cause analysis, linking complex electrical test results directly back to specific process variations with greater accuracy than human experts or conventional methods. This AI-driven approach will lead to higher test coverage with fewer test vectors, marking a paradigm shift toward truly autonomous and high-quality MRAM manufacturing.

BIBLIOGRAPHY

- [1] L. Wei, J. G. Alzate, U. Arslan, *et al.*, “13.3 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 *International Solid-State Circuits Conference*, 2019, pp. 214–216. DOI: [10.1109/ISSCC.2019.8662444](https://doi.org/10.1109/ISSCC.2019.8662444).
- [2] J. L. Hennessy and D. A. Patterson, *Computer architecture: a quantitative approach*. Elsevier, 2011.
- [3] W. Stallings, *Computer organization and architecture: designing for performance*. Pearson Education India, 2016.
- [4] R. Bez and A. Pirovano, “Non-volatile memory technologies: emerging concepts and new materials,” *Materials Science in Semiconductor Processing*, vol. 7, no. 4-6, pp. 349–355, 2004. DOI: <https://doi.org/10.1016/j.mssp.2004.09.127>.
- [5] S.-T. Han, Y. Zhou, and V. Roy, “Towards the development of flexible non-volatile memories,” *Advanced Materials*, vol. 25, no. 38, pp. 5425–5449, 2013. DOI: <https://doi.org/10.1016/j.mssp.2004.09.127>.
- [6] Y. Huai *et al.*, “Spin-transfer torque MRAM (STT-MRAM): Challenges and prospects,” *AAPPS bulletin*, vol. 18, no. 6, pp. 33–40, 2008.
- [7] S. Tehrani, B. Engel, J. Slaughter, *et al.*, “Recent developments in magnetic tunnel junction MRAM,” *IEEE Transactions on magnetics*, vol. 36, no. 5, pp. 2752–2757, 2000. DOI: [10.1109/20.908581](https://doi.org/10.1109/20.908581).
- [8] F. Zahoor, T. Z. Azni Zulkifli, and F. A. Khanday, “Resistive random access memory (RRAM): an overview of materials, switching mechanism, performance, multi-level cell (MLC) storage, modeling, and applications,” *Nanoscale research letters*, vol. 15, pp. 1–26, 2020. DOI: <https://doi.org/10.1186/s11671-020-03299-9>.
- [9] Z. Shen, C. Zhao, Y. Qi, *et al.*, “Advances of RRAM devices: Resistive switching mechanisms, materials and bionic synaptic application,” *Nanomaterials*, vol. 10, no. 8, p. 1437, 2020. DOI: <https://doi.org/10.3390/nano10081437>.
- [10] X. Dong, N. Muralimanohar, N. Jouppi, R. Kaufmann, and Y. Xie, “Leveraging 3D PCRAM technologies to reduce checkpoint overhead for future exascale systems,” in *Proceedings of the conference on high performance computing networking, storage and analysis*, 2009, pp. 1–12. DOI: [10.1109/20.908581](https://doi.org/10.1109/20.908581).
- [11] X. Dong, N. P. Jouppi, and Y. Xie, “PCRAMsim: System-level performance, energy, and area modeling for phase-change RAM,” in *Proceedings of the 2009 International Conference on Computer-Aided Design*, 2009, pp. 269–275. DOI: <https://doi.org/10.1145/1687399.1687449>.

- [12] T. Mikolajick, C. Dehm, W. Hartner, *et al.*, “FeRAM technology for high density applications,” *Microelectronics Reliability*, vol. 41, no. 7, pp. 947–950, 2001. DOI: [https://doi.org/10.1016/S0026-2714\(01\)00049-X](https://doi.org/10.1016/S0026-2714(01)00049-X).
- [13] D. Takashima, “Overview of FeRAMs: Trends and perspectives,” in *2011 11th Annual Non-Volatile Memory Technology Symposium Proceeding*, IEEE, 2011, pp. 1–6. DOI: [10.1109/NVMTS.2011.6137107](https://doi.org/10.1109/NVMTS.2011.6137107).
- [14] S. Dünkler, M. Trentzsch, R. Richter, *et al.*, “A FeFET based super-low-power ultra-fast embedded NVM technology for 22nm FDSOI and beyond,” in *2017 IEEE International Electron Devices Meeting (IEDM)*, IEEE, 2017, pp. 19–7. DOI: [10.1109/IEDM.2017.8268425](https://doi.org/10.1109/IEDM.2017.8268425).
- [15] J. Y. Kim, M.-J. Choi, and H. W. Jang, “Ferroelectric field effect transistors: Progress and perspective,” *APL Materials*, vol. 9, no. 2, 2021. DOI: <https://doi.org/10.1063/5.0035515>.
- [16] F. Nardi, S. Larentis, S. Balatti, D. C. Gilmer, and D. Ielmini, “Resistive Switching by Voltage-Driven Ion Migration in Bipolar RRAM—Part I: Experimental Study,” *IEEE Transactions on Electron Devices*, vol. 59, no. 9, pp. 2461–2467, 2012. DOI: [10.1109/TED.2012.2202319](https://doi.org/10.1109/TED.2012.2202319).
- [17] S. Yu and P.-Y. Chen, “Emerging Memory Technologies: Recent Trends and Prospects,” *MSSC*, vol. 8, no. 2, pp. 43–56, 2016. DOI: [10.1109/MSSC.2016.2546199](https://doi.org/10.1109/MSSC.2016.2546199).
- [18] Y. Xie and Y. Zhao, “Emerging memory technologies,” *IEEE Micro*, vol. 39, no. 1, pp. 6–7, 2019, ISSN: 19374143. DOI: [10.1109/MM.2019.2892165](https://doi.org/10.1109/MM.2019.2892165).
- [19] F. Zahoor, T. Z. Azni Zulkifli, and F. A. Khanday, “Resistive Random Access Memory (RRAM): an Overview of Materials, Switching Mechanism, Performance, Multilevel Cell (mlc) Storage, Modeling, and Applications,” *Nanoscale Research Letters*, vol. 15, no. 1, 2020. DOI: [10.1186/s11671-020-03299-9](https://doi.org/10.1186/s11671-020-03299-9).
- [20] S. Hamdioui, M. Taouil, and N. Z. Haron, “Testing Open Defects in Memristor-Based Memories,” *TC*, vol. 64, no. 1, pp. 247–259, 2015. DOI: [10.1109/TC.2013.206](https://doi.org/10.1109/TC.2013.206).
- [21] C. Y. Chen, H. C. Shih, C. W. Wu, *et al.*, “RRAM Defect Modeling and Failure Analysis Based on March Test and a Novel Squeeze-Search Scheme,” *TC*, vol. 64, no. 1, pp. 180–190, 2015, ISSN: 00189340. DOI: [10.1109/TC.2014.12](https://doi.org/10.1109/TC.2014.12). [Online]. Available: <http://ieeexplore.ieee.org/document/6725492/>.
- [22] M. Fieback, M. Taouil, and S. Hamdioui, “Testing Resistive Memories: Where Are We and What Is Missing?” In *International Test Conference*, IEEE, 2018. DOI: [10.1109/TEST.2018.8624895](https://doi.org/10.1109/TEST.2018.8624895).
- [23] P. Guo, A. M. Sarangan, and I. Agha, “A review of germanium-antimony-telluride phase change materials for non-volatile memories and optical modulators,” *Applied sciences*, vol. 9, no. 3, p. 530, 2019. DOI: <https://doi.org/10.3390/app9030530>.

- [24] F. Zhai, K. Yang, D. Wang, *et al.*, “Bipolar resistive switching effect in laser-induced crystallization AgInSbTe films,” *Optik*, vol. 180, pp. 271–275, 2019. DOI: <https://doi.org/10.1016/j.ijleo.2018.11.100>.
- [25] C. Bonafos, M. Carrada, G. Benassayag, *et al.*, “Si and Ge nanocrystals for future memory devices,” *Materials Science in Semiconductor Processing*, vol. 15, no. 6, pp. 615–626, 2012. DOI: <https://doi.org/10.1016/j.mssp.2012.09.004>.
- [26] M. K. A. Qureshi, S. Gurumurthi, and B. Rajendran, *Phase change memory: From devices to systems*. Morgan & Claypool Publishers, 2012, vol. 18.
- [27] S. W. Fong, C. M. Neumann, and H.-S. P. Wong, “Phase-change memory—towards a storage-class memory,” *IEEE Transactions on Electron Devices*, vol. 64, no. 11, pp. 4374–4385, 2017. DOI: [10.1109/TED.2017.2746342](https://doi.org/10.1109/TED.2017.2746342).
- [28] H. Ishiwara, “Ferroelectric random access memories,” *Journal of nanoscience and nanotechnology*, vol. 12, no. 10, pp. 7619–7627, 2012. DOI: <https://doi.org/10.1166/jnn.2012.6651>.
- [29] C. Wang, S. Yuan, H. Xun, *et al.*, “Defects, Fault Modeling, and Test Development Framework for FeFETs,” in *2024 IEEE International Test Conference (ITC)*, IEEE, 2024, pp. 91–95. DOI: <https://doi.org/10.1145/3510851>.
- [30] T. Ali, P. Polakowski, S. Riedel, *et al.*, “Silicon doped hafnium oxide (HSO) and hafnium zirconium oxide (HZO) based FeFET: A material relation to device physics,” *Applied Physics Letters*, vol. 112, no. 22, 2018. DOI: <https://doi.org/10.1063/1.5029324>.
- [31] C. Besleaga, R. Radu, L.-M. Balescu, *et al.*, “Ferroelectric field effect transistors based on PZT and IGZO,” *IEEE Journal of the Electron Devices Society*, vol. 7, pp. 268–275, 2019. DOI: [10.1109/JEDS.2019.2895367](https://doi.org/10.1109/JEDS.2019.2895367).
- [32] H. Ishiwara, M. Okuyama, and Y. Arimoto, *Ferroelectric random access memories: fundamentals and applications*. Springer Science & Business Media, 2004, vol. 93.
- [33] C. Liu, Q. Wang, W. Yang, *et al.*, “Multiscale modeling of Al 0.7 Sc 0.3 N-based FeRAM: the steep switching, leakage and selector-free array,” in *2021 IEEE International Electron Devices Meeting (IEDM)*, IEEE, 2021, pp. 8–1. DOI: [10.1109/IEDM19574.2021.9720535](https://doi.org/10.1109/IEDM19574.2021.9720535).
- [34] S.-C. Chang, N. Haratipour, S. Shivaraman, *et al.*, “FeRAM using anti-ferroelectric capacitors for high-speed and high-density embedded memory,” in *2021 IEEE International Electron Devices Meeting (IEDM)*, IEEE, 2021, pp. 33–2. DOI: [10.1109/IEDM19574.2021.9720510](https://doi.org/10.1109/IEDM19574.2021.9720510).
- [35] T. Francois, J. Coignus, A. Makosiej, *et al.*, “16kbit HfO₂: Si-based 1T-1C FeRAM arrays demonstrating high performance operation and solder reflow compatibility,” in *2021 IEEE International Electron Devices Meeting (IEDM)*, IEEE, 2021, pp. 33–1. DOI: [10.1109/IEDM19574.2021.9720640](https://doi.org/10.1109/IEDM19574.2021.9720640).
- [36] J. Inoue and S. Maekawa, “Theory of tunneling magnetoresistance in granular magnetic films,” *Physical Review B*, vol. 53, no. 18, R11927, 1996. DOI: <https://doi.org/10.1103/PhysRevB.53.R11927>.

- [37] M. Julliere, "Tunneling between ferromagnetic films," *Physics letters A*, vol. 54, no. 3, pp. 225–226, 1975. DOI: [https://doi.org/10.1016/0375-9601\(75\)90174-7](https://doi.org/10.1016/0375-9601(75)90174-7).
- [38] A. D. Kent and D. C. Worledge, "A new spin on magnetic memories," *Nature nanotechnology*, vol. 10, no. 3, pp. 187–191, 2015. DOI: [10.1038/nnano.2015.24](https://doi.org/10.1038/nnano.2015.24).
- [39] A. Manchon, J. Železný, I. M. Miron, *et al.*, "Current-induced spin-orbit torques in ferromagnetic and antiferromagnetic systems," *Reviews of Modern Physics*, vol. 91, no. 3, p. 035 004, 2019. DOI: <https://doi.org/10.1103/RevModPhys.91.035004>.
- [40] 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," *Nature Materials*, vol. 11, no. 1, pp. 39–43, 2012. DOI: <https://doi.org/10.1038/nmat3172>.
- [41] A. Khvalkovskiy, D. Apalkov, S. Watts, *et al.*, "Basic principles of STT-MRAM cell operation in memory arrays," *Journal of Physics D: Applied Physics*, vol. 46, no. 7, p. 074 001, 2013. DOI: [10.1088/0022-3727/46/7/074001](https://doi.org/10.1088/0022-3727/46/7/074001).
- [42] S. Bhatti, R. Sbiaa, A. Hirohata, H. Ohno, S. Fukami, and S. Piramanayagam, "Spintronics based random access memory: A review," *Materials today*, vol. 20, no. 9, pp. 530–548, 2017. DOI: <https://doi.org/10.1016/j.mattod.2017.07.007>.
- [43] G. Prenat, K. Jabeur, P. Vanhauwaert, *et al.*, "Ultra-fast and high-reliability SOT-MRAM: From cache replacement to normally-off computing," *IEEE Transactions on Multi-Scale Computing Systems*, vol. 2, no. 1, pp. 49–60, 2015. DOI: [10.1109/TMSCS.2015.2509963](https://doi.org/10.1109/TMSCS.2015.2509963).
- [44] W. Kang, L. Chang, Y. Zhang, and W. Zhao, "Voltage-controlled MRAM for working memory: Perspectives and challenges," in *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017*, IEEE, 2017, pp. 542–547. DOI: [10.23919/DATE.2017.7927047](https://doi.org/10.23919/DATE.2017.7927047).
- [45] L. Wu, M. Taouil, S. Rao, E. J. Marinissen, and S. Hamdioui, "Survey on STT-MRAM testing: Failure mechanisms, fault models, and tests," *arXiv preprint arXiv:2001.05463*, 2020. DOI: <https://doi.org/10.48550/arXiv.2001.05463>.
- [46] Y.-J. Song, J. H. Lee, H. Shin, *et al.*, "Highly functional and reliable 8Mb STT-MRAM embedded in 28nm logic," in *2016 IEEE International Electron Devices Meeting (IEDM)*, IEEE, 2016, pp. 27–2. DOI: [10.1109/IEDM.2016.7838491](https://doi.org/10.1109/IEDM.2016.7838491).
- [47] K. Korgaonkar, I. Bhati, H. Liu, *et al.*, "Density tradeoffs of non-volatile memory as a replacement for SRAM based last level cache," in *2018 ACM/IEEE 45th Annual International Symposium on Computer Architecture (ISCA)*, IEEE, 2018, pp. 315–327. DOI: [10.1109/ISCA.2018.00035](https://doi.org/10.1109/ISCA.2018.00035).
- [48] R. Nebashi, N. Sakimura, H. Honjo, *et al.*, "A 90nm 12ns 32Mb 2T1MTJ MRAM," in *2009 IEEE International Solid-State Circuits Conference-Digest of Technical Papers*, IEEE, 2009, pp. 462–463. DOI: [10.1109/ISSCC.2009.4977508](https://doi.org/10.1109/ISSCC.2009.4977508).

- [49] B. Behin-Aein, D. Datta, S. Salahuddin, and S. Datta, "Proposal for an all-spin logic device with built-in memory," *Nature nanotechnology*, vol. 5, no. 4, pp. 266–270, 2010. DOI: <https://doi.org/10.1038/nnano.2010.31>.
- [50] V. Krizakova, M. Perumkunnil, S. Couet, P. Gambardella, and K. Garello, "Spin-orbit torque switching of magnetic tunnel junctions for memory applications," *Journal of Magnetism and Magnetic Materials*, vol. 562, p. 169 692, 2022. DOI: <https://doi.org/10.1016/j.jmmm.2022.169692>.
- [51] H.-S. P. Wong and S. Salahuddin, "Memory leads the way to better computing," *Nature nanotechnology*, vol. 10, no. 3, pp. 191–194, 2015. DOI: <https://doi.org/10.1038/nnano.2015.29>.
- [52] T. Endoh, H. Koike, S. Ikeda, T. Hanyu, and H. Ohno, "An overview of non-volatile emerging memories—Spintronics for working memories," *IEEE journal on emerging and selected topics in circuits and systems*, vol. 6, no. 2, pp. 109–119, 2016. DOI: [10.1109/JETCAS.2016.2547704](https://doi.org/10.1109/JETCAS.2016.2547704).
- [53] A. J. Smith, "Cache memories," *ACM Computing Surveys (CSUR)*, vol. 14, no. 3, pp. 473–530, 1982. DOI: <https://doi.org/10.1145/356887.356892>.
- [54] S. Kumar and P. Singh, "An overview of modern cache memory and performance analysis of replacement policies," in *2016 IEEE International Conference on Engineering and Technology (ICETECH)*, IEEE, 2016, pp. 210–214. DOI: [10.1109/ICETECH.2016.7569243](https://doi.org/10.1109/ICETECH.2016.7569243).
- [55] V. Cuppu, B. Jacob, B. Davis, and T. Mudge, "A performance comparison of contemporary DRAM architectures," in *Proceedings of the 26th annual international symposium on Computer architecture*, 1999, pp. 222–233. DOI: <https://doi.org/10.1145/300979.300998>.
- [56] M. A. Breuer, A. D. Friedman, *et al.*, *Diagnosis & Reliable Design of Digital Systems*. Springer, 1976, vol. 19.
- [57] A. J. Van de Goor, *Testing Semiconductor Memories: Theory and Practice*. John Wiley & Sons, Inc., 1991.
- [58] S. Hamdioui, A. J. Van De Goor, J. Delos Reyes, and M. Rodgers, "Memory test experiment: Industrial results and data," *IEE Proceedings-Computers and Digital Techniques*, vol. 153, no. 1, pp. 1–8, 2006. DOI: <https://doi.org/10.1049/ip-cdt:20050104>.
- [59] I. Schanstra and A. Van De Goor, "Industrial Evaluation of Stress Combinations for March Tests Applied to SRAMs," in *International Test Conference (ITC)*, 1999, pp. 983–992. DOI: [10.1109/TEST.1999.805831](https://doi.org/10.1109/TEST.1999.805831).
- [60] A. van de Goer and J. de Neef, "Industrial Evaluation of DRAM Tests," in *Design, Automation and Test in Europe Conference and Exhibition (DATE)*, 1999, pp. 623–630. DOI: [10.1109/DATE.1999.761194](https://doi.org/10.1109/DATE.1999.761194).
- [61] S. Hamdioui and A. J. Van De Goor, "An experimental analysis of spot defects in srams: Realistic fault models and tests," in *Proceedings of the Ninth Asian Test Symposium*, IEEE, 2000, pp. 131–138. DOI: [10.1109/ATS.2000.893615](https://doi.org/10.1109/ATS.2000.893615).

- [62] E. I. Vatajelu, A. Bosio, L. Dilillo, *et al.*, “Analyzing Resistive-Open Defects in SRAM Core-Cell under the Effect of Process Variability,” in *European Test Symposium (ETS)*, 2013, pp. 1–6. DOI: [10.1109/ETS.2013.6569373](https://doi.org/10.1109/ETS.2013.6569373).
- [63] F. Hapke, W. Redemund, J. Schloeffel, *et al.*, “Defect-Oriented Cell-Internal Testing,” in *International Test Conference*, 2010, pp. 1–10. DOI: [10.1109/TEST.2010.5699229](https://doi.org/10.1109/TEST.2010.5699229).
- [64] Z. Gao, S. Malagi, M.-C. Hu, *et al.*, “Application of Cell-Aware Test on an Advanced 3nm CMOS Technology Library,” in *International Test Conference (ITC)*, 2019, pp. 1–6. DOI: [10.1109/ITC44170.2019.9000164](https://doi.org/10.1109/ITC44170.2019.9000164).
- [65] S. Borkar, “Microarchitecture and Design Challenges for Gigascale Integration,” in *International Symposium on Microarchitecture*, 2004, pp. 3–3. DOI: [10.1109/MICRO.2004.24](https://doi.org/10.1109/MICRO.2004.24).
- [66] L. Wu, M. Taouil, S. Rao, E. J. Marinissen, and S. Hamdioui, “Electrical Modeling of STT-MRAM Defects,” in *International Test Conference*, 2018. DOI: [10.1109/TEST.2018.8624749](https://doi.org/10.1109/TEST.2018.8624749).
- [67] M. Fieback, G. C. Medeiros, L. Wu, *et al.*, “Defects, fault modeling, and test development framework for rrams,” *ACM Journal on Emerging Technologies in Computing Systems*, vol. 18, no. 3, pp. 1–26, 2022. DOI: <https://doi.org/10.1145/3510851>.
- [68] L. Wu, S. Rao, M. Taouil, E. J. Marinissen, G. S. Kar, and S. Hamdioui, “MFA-MTJ Model: Magnetic-Field-Aware Compact Model of pMTJ for Robust STT-MRAM Design,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 41, no. 11, pp. 1–1, 2022. DOI: [10.1109/TCAD.2021.3140157](https://doi.org/10.1109/TCAD.2021.3140157).
- [69] A. J. van de Goor and I. B. Tlili, “Disturb neighborhood pattern sensitive fault,” in *Proceedings. 15th IEEE VLSI Test Symposium (Cat. No. 97TB100125)*, IEEE, 1997, pp. 37–45. DOI: [10.1109/VTEST.1997.599439](https://doi.org/10.1109/VTEST.1997.599439).
- [70] S. Hamdioui, Z. Al-Ars, and A. J. Van de Goor, “Testing static and dynamic faults in random access memories,” in *Proceedings 20th IEEE VLSI Test Symposium (VTS 2002)*, IEEE, 2002, pp. 395–400. DOI: [10.1109/VTS.2002.1011170](https://doi.org/10.1109/VTS.2002.1011170).
- [71] H. T. Nagle, S. Roy, C. Hawkins, M. McNamer, and R. Fritzscheier, “Design for testability and built-in self test: A review,” *IEEE Transactions on Industrial Electronics*, vol. 36, no. 2, pp. 129–140, 1989. DOI: [10.1109/41.19062](https://doi.org/10.1109/41.19062).
- [72] H. Xun, S. Yuan, M. Fieback, H. Aziza, M. Taouil, and S. Hamdioui, “Device-Aware Test for Ion Depletion Defects in RRAMs,” in *International Test Conference*, 2023. DOI: [10.1109/ITC51656.2023.00041](https://doi.org/10.1109/ITC51656.2023.00041).
- [73] J. Azevedo, A. Virazel, A. Bosio, *et al.*, “A complete resistive-open defect analysis for thermally assisted switching MRAMs,” in *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, IEEE, 2014, pp. 2326–2335. DOI: [10.1109/TVLSI.2013.2294080](https://doi.org/10.1109/TVLSI.2013.2294080).

- [74] S. M. Nair, R. Bishnoi, M. B. Tahoori, G. Tshagharyan, H. Grigoryan, and G. Harutyunyan, "Defect injection, fault modeling and test algorithm generation methodology for STT-MRAM," in *International Test Conference*, 2018, pp. 1–10, ISBN: 9781538683828. DOI: [10.1109/TEST.2018.8624725](https://doi.org/10.1109/TEST.2018.8624725).
- [75] A. Chintaluri, A. Parihar, S. Natarajan, H. Naeimi, and A. Raychowdhury, "A model study of defects and faults in embedded spin transfer torque (stt) mram arrays," in *2015 IEEE 24th Asian test symposium (ATS)*, IEEE, 2015, pp. 187–192. DOI: [10.1109/ATS.2015.39](https://doi.org/10.1109/ATS.2015.39).
- [76] C. Münch, J. Yun, M. Keim, and M. B. Tahoori, "Mbist-based trim-search test time reduction for stt-mram," in *2022 IEEE 40th VLSI Test Symposium (VTS)*, IEEE, 2022, pp. 1–7. DOI: [10.1109/VTS52500.2021.9794178](https://doi.org/10.1109/VTS52500.2021.9794178).
- [77] I. Yoon and A. Raychowdhury, "Modeling and analysis of magnetic field induced coupling on embedded stt-mram arrays," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 37, no. 2, pp. 337–349, 2017. DOI: [10.1109/TCAD.2017.2697963](https://doi.org/10.1109/TCAD.2017.2697963).
- [78] S. M. Nair, R. Bishnoi, M. S. Golanbari, F. Oboril, F. Hameed, and M. B. Tahoori, "Vaet-stt: Variation aware stt-mram analysis and design space exploration tool," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 37, no. 7, pp. 1396–1407, 2017. DOI: [10.1109/TCAD.2017.2760861](https://doi.org/10.1109/TCAD.2017.2760861).
- [79] L. Wu, M. Taouil, S. Rao, E. J. Marinissen, and S. Hamdioui, "Electrical Modeling of STT-MRAM Defects," in *International Test Conference*, 2018, pp. 1–10. DOI: [10.1109/TEST.2018.8624749](https://doi.org/10.1109/TEST.2018.8624749).
- [80] W. Kim, S. Couet, J. Swerts, *et al.*, "Experimental Observation of Back-Hopping With Reference Layer Flipping by High-Voltage Pulse in Perpendicular Magnetic Tunnel Junctions," *IEEE Transactions on Magnetics*, vol. 52, no. 7, pp. 1–4, 2016. DOI: [10.1109/TMAG.2016.2536158](https://doi.org/10.1109/TMAG.2016.2536158).
- [81] L. Wu, S. Rao, M. Taouil, E. J. Marinissen, G. S. Kar, and S. Hamdioui, "Characterization, modeling, and test of intermediate state defects in stt-mrams," *IEEE Transactions on Computers*, vol. 71, no. 9, pp. 2219–2233, 2021. DOI: [10.1109/TC.2021.3125228](https://doi.org/10.1109/TC.2021.3125228).
- [82] S. Yuan, M. Taouil, M. Fieback, *et al.*, "Device-Aware Test for Back-Hopping Defects in STT-MRAMs," in *2023 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2023, pp. 1–6. DOI: [10.23919/DAT56975.2023.10137071](https://doi.org/10.23919/DAT56975.2023.10137071).
- [83] S. Yuan, M. Taouil, M. Fieback, *et al.*, "Magnetic Coupling Based Test Development for Contact and Interconnect Defects in STT-MRAMs," in *International Test Conference*, 2023, pp. 1–10. DOI: [10.1109/ITC51656.2023.00040](https://doi.org/10.1109/ITC51656.2023.00040).
- [84] Y.-H. Wang, S.-H. Huang, D.-Y. Wang, *et al.*, "Impact of stray field on the switching properties of perpendicular MTJ for scaled MRAM," in *IEDM*, 2012, pp. 29.2.1–29.2.4. DOI: [10.1109/IEDM.2012.6479127](https://doi.org/10.1109/IEDM.2012.6479127).
- [85] M. TAOUIL and S. Hamdioui, *Device aware test for memory units*, US Patent 11,869,612, 2024.

- [86] L. Wu, S. Rao, M. Taouil, E. J. Marinissen, G. S. Kar, and S. Hamdioui, "Characterization and fault modeling of intermediate state defects in stt-mram," in *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, IEEE, 2021, pp. 1717–1722. DOI: [10.23919/DATE51398.2021.9473999](https://doi.org/10.23919/DATE51398.2021.9473999).
- [87] L. Wu, S. Rao, M. Taouil, E. J. Marinissen, G. S. Kar, and S. Hamdioui, "Characterization, modeling and test of synthetic anti-ferromagnet flip defect in stt-mrams," in *2020 IEEE International Test Conference (ITC)*, 2020, pp. 1–10. DOI: [10.1109/ITC44778.2020.9325258](https://doi.org/10.1109/ITC44778.2020.9325258).
- [88] H. Cai, Z. Bian, Y. Hou, *et al.*, "33.4 A 28nm 2Mb STT-MRAM computing-in-memory macro with a refined bit-cell and 22.4-41.5 TOPS/W for AI inference," in *International Solid-State Circuits Conference*, 2023, pp. 500–502. DOI: [10.1109/ISSCC42615.2023.10067339](https://doi.org/10.1109/ISSCC42615.2023.10067339).
- [89] L. Wu, S. Rao, M. Taouil, E. J. Marinissen, G. S. Kar, and S. Hamdioui, "Impact of magnetic coupling and density on stt-mram performance," 2020. DOI: <https://doi.org/10.48550/arXiv.2011.11349>.
- [90] S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, and K. Ando, "Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions," *Nature materials*, vol. 3, no. 12, pp. 868–871, 2004. DOI: <https://doi.org/10.1038/nmat1257>.
- [91] S. Ikeda, K. Miura, H. Yamamoto, *et al.*, "A perpendicular-anisotropy CoFeB–MgO magnetic tunnel junction," *Nature materials*, vol. 9, no. 9, pp. 721–724, 2010. DOI: <https://doi.org/10.1038/nmat2804>.
- [92] M. Konoto, H. Imamura, T. Taniguchi, *et al.*, "Effect of MgO cap layer on Gilbert damping of FeB electrode layer in MgO-based magnetic tunnel junctions," *Applied Physics Express*, vol. 6, no. 7, p. 073002, 2013. DOI: [10.7567/APEX.6.073002](https://doi.org/10.7567/APEX.6.073002).
- [93] J. C. Slonczewski, "Current-driven excitation of magnetic multilayers," *Journal of Magnetism and Magnetic Materials*, vol. 159, no. 1-2, pp. L1–L7, 1996. DOI: [https://doi.org/10.1016/0304-8853\(96\)00062-5](https://doi.org/10.1016/0304-8853(96)00062-5).
- [94] L. Berger, "Emission of spin waves by a magnetic multilayer traversed by a current," *Physical Review B*, vol. 54, no. 13, p. 9353, 1996. DOI: <https://doi.org/10.1103/PhysRevB.54.9353>.
- [95] J. Z. Sun, "Spin-current interaction with a monodomain magnetic body: A model study," *Physical Review B*, vol. 62, no. 1, p. 570, 2000. DOI: <https://doi.org/10.1103/PhysRevB.62.570>.
- [96] D. C. Ralph and M. D. Stiles, "Spin transfer torques," *Journal of Magnetism and Magnetic Materials*, vol. 320, no. 7, pp. 1190–1216, 2008. DOI: <https://doi.org/10.1016/j.jmmm.2007.12.019>.
- [97] S. S. Parkin, C. Kaiser, A. Panchula, *et al.*, "Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers," *Nature materials*, vol. 3, no. 12, pp. 862–867, 2004. DOI: <https://doi.org/10.1038/nmat1256>.

- [98] “45nm low power CMOS logic compatible embedded STT MRAM utilizing a reverse-connection 1T/1MTJ cell, author=Lin, CJ and Kang, SH and Wang, YJ and Lee, K and Zhu, X and Chen, WC and Li, X and Hsu, WN and Kao, YC and Liu, MT and others,” in *2009 IEEE International Electron Devices Meeting (IEDM)*, IEEE, 2009, pp. 1–4. DOI: [10.1109/IEDM.2009.5424368](https://doi.org/10.1109/IEDM.2009.5424368).
- [99] T. Kawahara, K. Ito, R. Takemura, and H. Ohno, “Spin-transfer torque ram technology: Review and prospect,” *Microelectronics Reliability*, vol. 52, no. 4, pp. 613–627, 2012. DOI: <https://doi.org/10.1016/j.microrel.2011.09.028>.
- [100] I. Yoon, A. Chintaluri, and A. Raychowdhury, “Emacs: Efficient mbist architecture for test and characterization of stt-mram arrays,” in *2016 IEEE International Test Conference (ITC)*, IEEE, 2016, pp. 1–10. DOI: [10.1109/TEST.2016.7805834](https://doi.org/10.1109/TEST.2016.7805834).
- [101] R. Bez, E. Camerlenghi, A. Modelli, and A. Visconti, “Introduction to flash memory,” *Proceedings of the IEEE*, vol. 91, no. 4, pp. 489–502, 2003. DOI: [10.1109/JPROC.2003.811702](https://doi.org/10.1109/JPROC.2003.811702).
- [102] E. Seevinck, P. J. van Beers, and H. Ontrop, “Current-mode techniques for high-speed VLSI circuits with application to current sense amplifier for CMOS SRAM’s,” *IEEE Journal of Solid-State Circuits*, vol. 26, no. 4, pp. 525–536, 1991. DOI: [10.1109/4.75050](https://doi.org/10.1109/4.75050).
- [103] D. Shum, D. Houssameddine, S. T. Woo, *et al.*, “CMOS-embedded STT-MRAM arrays in 2x nm nodes for GP-MCU applications,” in *2017 Symposium on VLSI Technology*, IEEE, 2017, T208–T209. DOI: [10.23919/VLSIT.2017.7998174](https://doi.org/10.23919/VLSIT.2017.7998174).
- [104] L. Tillie, E. Nowak, R. Sousa, *et al.*, “Data retention extraction methodology for perpendicular STT-MRAM,” in *2016 IEEE International Electron Devices Meeting (IEDM)*, IEEE, 2016, pp. 27–3. DOI: [10.1109/IEDM.2016.7838492](https://doi.org/10.1109/IEDM.2016.7838492).
- [105] L. Chang, D. M. Fried, J. Hergenrother, *et al.*, “Stable SRAM cell design for the 32 nm node and beyond,” in *Digest of Technical Papers. 2005 Symposium on VLSI Technology, 2005.*, IEEE, 2005, pp. 128–129. DOI: [10.1109/.2005.1469239](https://doi.org/10.1109/.2005.1469239).
- [106] Z. Guo, S. Balasubramanian, R. Zlatanovici, T.-J. King, and B. Nikolić, “FinFET-based SRAM design,” in *Proceedings of the 2005 international symposium on Low power electronics and design*, 2005, pp. 2–7. DOI: <https://doi.org/10.1145/1077603.1077607>.
- [107] S. Yuan, M. A. Yaldagard, H. Xun, *et al.*, “Design-for-Test for Intermittent Faults in STT-MRAMs,” in *2024 IEEE European Test Symposium (ETS)*, 2024, pp. 1–6. DOI: [10.1109/ETS61313.2024.10567702](https://doi.org/10.1109/ETS61313.2024.10567702).
- [108] H. Lim, S. Lee, and H. Shin, “A survey on the modeling of magnetic tunnel junctions for circuit simulation,” *Active and Passive Electronic Components*, vol. 2016, no. 1, p. 3858621, 2016. DOI: <https://doi.org/10.1155/2016/3858621>.
- [109] S. S. Mahalingam, B. Manikandan, and S. Arockiaraj, “Review–micromagnetic simulation using OOMMF and experimental investigations on nano composite magnets,” in *Journal of Physics: Conference Series*, IOP Publishing, vol. 1172, 2019, p. 012070. DOI: [10.1088/1742-6596/1172/1/012070](https://doi.org/10.1088/1742-6596/1172/1/012070).

- [110] G. D. Panagopoulos, C. Augustine, and K. Roy, "Physics-based SPICE-compatible compact model for simulating hybrid MTJ/CMOS circuits," *IEEE Transactions on Electron Devices*, vol. 60, no. 9, pp. 2808–2814, 2013. DOI: [10.1109/TED.2013.2275082](https://doi.org/10.1109/TED.2013.2275082).
- [111] Y. Zhang, W. Zhao, Y. Lakys, *et al.*, "Compact modeling of perpendicular-anisotropy CoFeB/MgO magnetic tunnel junctions," *IEEE transactions on Electron devices*, vol. 59, no. 3, pp. 819–826, 2012. DOI: [10.1109/TED.2011.2178416](https://doi.org/10.1109/TED.2011.2178416).
- [112] Y. Wang, Y. Zhang, E. Deng, J.-O. Klein, L. A. Naviner, and W. Zhao, "Compact model of magnetic tunnel junction with stochastic spin transfer torque switching for reliability analyses," *Microelectronics Reliability*, vol. 54, no. 9-10, pp. 1774–1778, 2014. DOI: <https://doi.org/10.1016/j.microrel.2014.07.019>.
- [113] H. Lim, S. Lee, and H. Shin, "Unified analytical model for switching behavior of magnetic tunnel junction," *IEEE Electron Device Letters*, vol. 35, no. 2, pp. 193–195, 2013. DOI: [10.1109/LED.2013.2293598](https://doi.org/10.1109/LED.2013.2293598).
- [114] W. Brinkman, R. Dynes, and J. Rowell, "Tunneling conductance of asymmetrical barriers," *Journal of applied physics*, vol. 41, no. 5, pp. 1915–1921, 1970. DOI: <https://doi.org/10.1063/1.1659141>.
- [115] Y. Wang, "Reliability analysis of spintronic device based logic and memory circuits," Ph.D. dissertation, Télécom ParisTech, 2017.
- [116] Z. Li and S. Zhang, "Thermally assisted magnetization reversal in the presence of a spin-transfer torque," *Physical Review B*, vol. 69, no. 13, p. 134416, 2004. DOI: <https://doi.org/10.1103/PhysRevB.69.134416>.
- [117] 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," *Journal of Applied Physics*, vol. 109, no. 7, 2011. DOI: <https://doi.org/10.1063/1.3562136>.
- [118] W. Wernsdorfer, E. B. Orozco, K. Hasselbach, *et al.*, "Experimental evidence of the Néel-Brown model of magnetization reversal," *Physical Review Letters*, vol. 78, no. 9, p. 1791, 1997. DOI: <https://doi.org/10.1103/PhysRevLett.78.1791>.
- [119] J. Sun, "Current-driven magnetic switching in manganite trilayer junctions," *Journal of Magnetism and Magnetic Materials*, vol. 202, no. 1, pp. 157–162, 1999. DOI: [https://doi.org/10.1016/S0304-8853\(99\)00289-9](https://doi.org/10.1016/S0304-8853(99)00289-9).
- [120] X. Guo, M. N. Bojnordi, Q. Guo, and E. Ipek, "Sanitizer: Mitigating the impact of expensive ECC checks on STT-MRAM based main memories," *IEEE Transactions on Computers*, vol. 67, no. 6, pp. 847–860, 2017. DOI: [10.1109/TC.2017.2779151](https://doi.org/10.1109/TC.2017.2779151).
- [121] H. Kumar and V. Tomar, "A review on performance evaluation of different low power SRAM cells in nano-scale era," *Wireless Personal Communications*, vol. 117, no. 3, pp. 1959–1984, 2021. DOI: <https://doi.org/10.1007/s11277-020-07953-4>.

- [122] J. Song, H. Dixit, B. Behin-Aein, C. H. Kim, and W. Taylor, "Impact of process variability on write error rate and read disturbance in STT-MRAM devices," *IEEE Transactions on Magnetics*, vol. 56, no. 12, pp. 1–11, 2020. DOI: [10.1109/TMAG.2020.3028045](https://doi.org/10.1109/TMAG.2020.3028045).
- [123] S. M. Nair, R. Bishnoi, and M. B. Tahoori, "Mitigating read failures in stt-mram," in *2020 IEEE 38th VLSI Test Symposium (VTS)*, IEEE, 2020, pp. 1–6. DOI: [10.1109/VTS48691.2020.9107605](https://doi.org/10.1109/VTS48691.2020.9107605).
- [124] B. Wicht, T. Nirschl, and D. Schmitt-Landsiedel, "Yield and speed optimization of a latch-type voltage sense amplifier," *IEEE journal of solid-state circuits*, vol. 39, no. 7, pp. 1148–1158, 2004. DOI: [10.1109/JSSC.2004.829399](https://doi.org/10.1109/JSSC.2004.829399).
- [125] M. Pakala, Y. Huai, T. Valet, Y. Ding, and Z. Diao, "Critical current distribution in spin-transfer-switched magnetic tunnel junctions," *Journal of applied physics*, vol. 98, no. 5, 2005. DOI: <https://doi.org/10.1063/1.2039997>.
- [126] D. Apalkov, A. Khvalkovskiy, S. Watts, *et al.*, "Spin-transfer torque magnetic random access memory (STT-MRAM)," *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, vol. 9, no. 2, pp. 1–35, 2013. DOI: <https://doi.org/10.1145/2463585.2463589>.
- [127] <https://www.hprobe.com/>.
- [128] S. Yuan, H. Xun, W. Kim, *et al.*, "Testing STT-MRAMs: Do We Need Magnets in our Automated Test Equipment?" In *International Test Conference*, 2024, pp. 364–373. DOI: [10.1109/ITC51657.2024.00058](https://doi.org/10.1109/ITC51657.2024.00058).
- [129] D. Worledge, G. Hu, D. W. Abraham, *et al.*, "Spin torque switching of perpendicular TaCoFeB MgO-based magnetic tunnel junctions," *Appl. Phys. Lett.*, vol. 98, no. 2, 2011.
- [130] S. J. Fonash, "An overview of dry etching damage and contamination effects," *Journal of The Electrochemical Society*, vol. 137, no. 12, p. 3885, 1990. DOI: [10.1149/1.2086322](https://doi.org/10.1149/1.2086322).
- [131] L. Ephrath and R. Bennett, "RIE contamination of etched silicon surfaces," *Journal of the Electrochemical Society*, vol. 129, no. 8, p. 1822, 1982. DOI: [10.1149/1.2124301](https://doi.org/10.1149/1.2124301).
- [132] S. Hamdioui, R. Wadsworth, J. D. Reyes, and A. J. Van De Goor, "Memory fault modeling trends: A case study," *J. Electron. Testing*, vol. 20, no. 3, pp. 245–255, 2004. DOI: <https://doi.org/10.1023/B:JETT.0000029458.57095.bb>.
- [133] A. J. Van de Goor and I. B. Tlili, "March tests for word-oriented memories," in *Proceedings Design, Automation and Test in Europe*, IEEE, 1998, pp. 501–508. DOI: [10.1109/DATE.1998.655905](https://doi.org/10.1109/DATE.1998.655905).
- [134] E. J. McCluskey, "Built-in self-test structures," *IEEE Design & Test of Computers*, vol. 2, no. 2, pp. 29–36, 1985. DOI: [10.1109/MDT.1985.294857](https://doi.org/10.1109/MDT.1985.294857).
- [135] L. Wu, S. Rao, G. C. Medeiros, *et al.*, "Pinhole defect characterization and fault modeling for STT-MRAM testing," in *IEEE European Test Symposium (ETS)*, 2019, pp. 1–6. DOI: [10.1109/ETS.2019.8791518](https://doi.org/10.1109/ETS.2019.8791518).

- [136] J. Teixeira, J. Ventura, F. Carpinteiro, *et al.*, “The effect of pinhole formation/growth on the tunnel magnetoresistance of MgO-based magnetic tunnel junctions,” *Journal of Applied Physics*, vol. 106, no. 7, 2009. DOI: <https://doi.org/10.1063/1.3236512>.
- [137] X. Yao, R. Malmhall, R. Ranjan, and J.-P. Wang, “Observation of intermediate states in magnetic tunnel junctions with composite free layer,” *IEEE Transactions on Magnetics*, vol. 44, no. 11, pp. 2496–2499, 2008.
- [138] J. Sun, M. Gaidis, G. Hu, *et al.*, “High-bias backhopping in nanosecond time-domain spin-torque switches of mgo-based magnetic tunnel junctions,” *Journal of Applied Physics*, vol. 105, no. 7, 2009. DOI: <https://doi.org/10.1063/1.3058614>.
- [139] S. Hamdioui, Z. Al-Ars, and A. van de Goor, “Testing Static and Dynamic Faults in Random Access Memories,” in *Proceedings 20th IEEE VLSI Test Symposium (VTS 2002)*, IEEE, 2002, pp. 395–400. DOI: [10.1109/VTS.2002.1011170](https://doi.org/10.1109/VTS.2002.1011170).
- [140] A. van de Goor and Z. Al-Ars, “Functional Memory Faults: a Formal Notation and a Taxonomy,” in *VTS, 2000*, pp. 281–289. DOI: [10.1109/VTEST.2000.843856](https://doi.org/10.1109/VTEST.2000.843856).
- [141] H. Xun, M. Fieback, S. Yuan, Z. Zhang, M. Taouil, and S. Hamdioui, “Data Background-Based Test Development for All Interconnect and Contact Defects in RRAMs,” in *IEEE Eur. Test Symp.*, 2023, pp. 1–6. DOI: [10.1109/ETS56758.2023.10174106](https://doi.org/10.1109/ETS56758.2023.10174106).
- [142] N. A. Zakaria, W. Hasan, I. Halin, R. Sidek, and X. Wen, “Fault detection with optimum March test algorithm,” in *2012 Third International Conference on Intelligent Systems Modelling and Simulation*, IEEE, 2012, pp. 700–704. DOI: [10.1109/ISMS.2012.88](https://doi.org/10.1109/ISMS.2012.88).
- [143] A. J. Van De Goor and Y. Zorian, “Effective march algorithms for testing single-order addressed memories,” *Journal of Electronic Testing*, vol. 5, pp. 337–345, 1994. DOI: <https://doi.org/10.1007/BF00972518>.
- [144] Suk and Reddy, “A march test for functional faults in semiconductor random access memories,” *IEEE Transactions on Computers*, vol. 100, no. 12, pp. 982–985, 1981. DOI: [10.1109/TC.1981.1675739](https://doi.org/10.1109/TC.1981.1675739).
- [145] Y. Zhou, H. Cai, M. Zhang, L. Naviner, and J. Yang, “A novel BIST for monitoring aging/temperature by self-triggered scheme to improve the reliability of STT-MRAM,” *Microelectron. Reliab.*, vol. 114, p. 113 735, 2020. DOI: <https://doi.org/10.1016/j.microrel.2020.113735>.
- [146] V. Champac, A. Gomez, F. Forero, and K. Roy, “Analysis of bridge defects in STT-MRAM cells under process variations and a robust DFT technique for their detection,” in *VLSI-SoC*, Springer, 2019, pp. 207–231.
- [147] S. Taghipour, M. Kamal, R. N. Asli, A. Afzali-Kusha, and M. Pedram, “CD-DFT: A current-difference design-for-testability to detect short defects of STT-MRAM under process variations,” *IEEE Transactions on Device and Materials Reliability*, vol. 21, no. 3, pp. 436–443, 2021. DOI: [10.1109/TDMR.2021.3104764](https://doi.org/10.1109/TDMR.2021.3104764).

- [148] J. Yun, B. Nadeau-Dostie, M. Keim, C. Dray, and M. Boujamaa, "MBIST support for reliable eMRAM sensing," in *ETS*, 2020, pp. 1–6. DOI: [10.1109/ETS48528.2020.9131564](https://doi.org/10.1109/ETS48528.2020.9131564).
- [149] Z.-W. Pan and J.-F. Li, "DFT-Enhanced Test Scheme for Spin-Transfer-Torque (STT) MRAMs," in *International Test Conference*, 2022, pp. 489–493. DOI: [10.1109/ITC50671.2022.00058](https://doi.org/10.1109/ITC50671.2022.00058).
- [150] R. Bishnoi, M. Ebrahimi, F. Oboril, and M. B. Tahoori, "Read disturb fault detection in STT-MRAM," in *International Test Conference*, 2014, pp. 1–7. DOI: [10.1109/TEST.2014.7035342](https://doi.org/10.1109/TEST.2014.7035342).
- [151] P. Girard, A. Bosio, A. Ladhar, and A. Virazel, "Defect Diagnosis Techniques for Silicon Customer Returns," in *Frontiers of Quality Electronic Design (QED) AI, IoT and Hardware Security*, Springer, 2023, pp. 641–676.
- [152] X. Xhafa, E. Faehn, P. Girard, and A. Virazel, "A Structural Testing Approach for SRAM Address Decoders Using Cell-Aware Methodology," in *2024 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)*, IEEE, 2024, pp. 1–4. DOI: [10.1109/DFT63277.2024.10753545](https://doi.org/10.1109/DFT63277.2024.10753545).
- [153] C.-L. Su, C.-W. Tsai, C.-W. Wu, *et al.*, "Write disturbance modeling and testing for MRAM," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 16, no. 3, pp. 277–288, 2008. DOI: [10.1109/TVLSI.2007.915402](https://doi.org/10.1109/TVLSI.2007.915402).
- [154] A. Hossam-Eldeen, B. Wang, P. Pandey, M. G. Bardon, and F. G. Redondo, "DTCO for Fast STT-MRAM Periphery Operation," in *2024 20th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD)*, IEEE, 2024, pp. 1–4. DOI: [10.1109/SMACD61181.2024.10745436](https://doi.org/10.1109/SMACD61181.2024.10745436).
- [155] C. Wann, R. Wong, D. J. Frank, *et al.*, "SRAM cell design for stability methodology," in *IEEE VLSI-TSA International Symposium on VLSI Technology, 2005. (VLSI-TSA-Tech)*, IEEE, 2005, pp. 21–22. DOI: [10.1109/VTSA.2005.1497065](https://doi.org/10.1109/VTSA.2005.1497065).
- [156] S. Yu and P.-Y. Chen, "Emerging memory technologies: Recent trends and prospects," *IEEE Solid-State Circuits Magazine*, vol. 8, no. 2, pp. 43–56, 2016. DOI: [10.1109/MSSC.2016.2546199](https://doi.org/10.1109/MSSC.2016.2546199).
- [157] S. Mitchell, M. O'Sullivan, and I. Dunning, "PuLP: A Linear Programming Toolkit for Python. 2011," URL <https://code.google.com/p/pulp-or/>, 2011.
- [158] I. E. Ebong and P. Mazumder, "Self-controlled writing and erasing in a memristor crossbar memory," *IEEE Transactions on Nanotechnology*, vol. 10, no. 6, pp. 1454–1463, 2011. DOI: [10.1109/TNANO.2011.2166805](https://doi.org/10.1109/TNANO.2011.2166805).
- [159] S. Wang, H. Lee, C. Grezes, P. Khalili, K. L. Wang, and P. Gupta, "MTJ variation monitor-assisted adaptive MRAM write," in *Proceedings of the 53rd Annual Design Automation Conference*, 2016, pp. 1–6. DOI: <https://doi.org/10.1145/2897937.2897979>.

- [160] S. Yuan, M. Taouil, M. Fieback, *et al.*, “Device-Aware Test for Back-Hopping Defects in STT-MRAMs,” in *IEEE/ACM Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Antwerp, Belgium, 2023, pp. 1–6. DOI: [10.23919/DATE56975.2023.10137071](https://doi.org/10.23919/DATE56975.2023.10137071).
- [161] S. B. Mamaghani, C. Münch, J. Yun, M. Keim, and M. B. Tahoori, “Smart Hammering: A practical method of pinhole detection in MRAM memories,” in *2023 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2023, pp. 1–6. DOI: [10.23919/DATE56975.2023.10137267](https://doi.org/10.23919/DATE56975.2023.10137267).
- [162] S. B. Mamaghani, J. Yun, M. Keim, and M. Tahoori, “Multi-Level Reference for Test Coverage Enhancement of Resistive-Based NVM,” in *2024 IEEE 42nd VLSI Test Symposium (VTS)*, IEEE, 2024, pp. 1–7. DOI: [10.1109/VTS60656.2024.10538585](https://doi.org/10.1109/VTS60656.2024.10538585).
- [163] S. Rao, W. Kim, S. Van Beek, *et al.*, “Stt-mram array performance improvement through optimization of ion beam etch and mtj for last-level cache application,” in *2021 IEEE International Memory Workshop (IMW)*, IEEE, 2021, pp. 1–4. DOI: [10.1109/IRPS.2017.7936318](https://doi.org/10.1109/IRPS.2017.7936318).
- [164] M. Frankowski, W. Skowroński, M. Czapkiewicz, and T. Stobiecki, “Backhopping in magnetic tunnel junctions: Micromagnetic approach and experiment,” *Journal of Magnetism and Magnetic Materials*, vol. 374, pp. 451–454, 2015. DOI: <https://doi.org/10.1016/j.jmmm.2014.08.056>.
- [165] M. P. Komalan, C. Tenllado, J. I. G. Pérez, F. T. Fernández, and F. Catthoor, “System level exploration of a STT-MRAM based level 1 data-cache,” in *2015 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2015, pp. 1311–1316. DOI: [10.7873/date.2015.0551](https://doi.org/10.7873/date.2015.0551).
- [166] J. Nogués and I. K. Schuller, “Exchange bias,” *Journal of Magnetism and Magnetic Materials*, vol. 192, no. 2, pp. 203–232, 1999. DOI: [https://doi.org/10.1016/S0304-8853\(98\)00266-2](https://doi.org/10.1016/S0304-8853(98)00266-2).
- [167] C. Abert, H. Sepehri-Amin, F. Bruckner, C. Vogler, M. Hayashi, and D. Suess, “Back-Hopping in Spin-Transfer-Torque Devices: Possible Origin and Countermeasures,” *Phys. Rev. Applied*, vol. 9, p. 054010, 5 2018. DOI: [10.1103/PhysRevApplied.9.054010](https://doi.org/10.1103/PhysRevApplied.9.054010). [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevApplied.9.054010>.
- [168] C. Safranski and J. Z. Sun, “Interface moment dynamics and its contribution to spin-transfer torque switching process in magnetic tunnel junctions,” *Phys. Rev. B*, vol. 100, p. 014435, 1 2019. DOI: [10.1103/PhysRevB.100.014435](https://doi.org/10.1103/PhysRevB.100.014435). [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.100.014435>.
- [169] J. Tan, J. Lim, J. Kwon, V. Naik, N. Raghavan, and K. Pey, “Role of temperature, MTJ size and pulse-width on STT-MRAM bit-error rate and backhopping,” *Solid-State Electronics*, vol. 183, p. 108032, 2021. DOI: <https://doi.org/10.1016/j.sse.2021.108032>.
- [170] M. Fieback, L. Wu, G. C. Medeiros, *et al.*, “Device-Aware Test: A New Test Approach Towards DPPB Level,” in *International Test Conference*, 2019. DOI: [10.1109/ITC44170.2019.9000134](https://doi.org/10.1109/ITC44170.2019.9000134).

- [171] M. Fieback, G. C. Medeiros, A. Gebregiorgis, H. Aziza, M. Taouil, and S. Hamdioui, "Intermittent undefined state fault in rrams," in *2021 IEEE European Test Symposium (ETS)*, IEEE, 2021, pp. 1–6. DOI: [10.1109/ETS50041.2021.9465401](https://doi.org/10.1109/ETS50041.2021.9465401).
- [172] H. Xun, M. Fieback, M. A. Yaldagard, *et al.*, "Robust design-for-testability scheme for conventional and unique defects in rrams," in *International Test Conference (ITC)*, 2024, pp. 374–383. DOI: [10.1109/ITC51657.2024.00059](https://doi.org/10.1109/ITC51657.2024.00059).
- [173] X. Chen and R. Victora, "Effect of pinholes in magnetic tunnel junctions," *Appl. Phys. Lett.*, vol. 91, no. 21, 2007. DOI: <https://doi.org/10.1063/1.2814038>.
- [174] A. J. Van de Goor and I. B. Tlili, "A systematic method for modifying march tests for bit-oriented memories into tests for word-oriented memories," *IEEE Transactions on Computers*, vol. 52, no. 10, pp. 1320–1331, 2003. DOI: [10.1109/TC.2003.1234529](https://doi.org/10.1109/TC.2003.1234529).
- [175] M. Fieback, F. Bradarić, M. Taouil, and S. Hamdioui, "Online Fault Detection and Diagnosis in RRAM," in *2021 IEEE European Test Symposium (ETS)*, 2023, pp. 1–6. DOI: [10.1109/ETS50041.2021.9465401](https://doi.org/10.1109/ETS50041.2021.9465401).
- [176] A. Aouichi, S. Yuan, M. Fieback, *et al.*, "Device Aware Diagnosis for Unique Defects in STT-MRAMs," in *ATS*, 2023, pp. 1–6. DOI: [10.1109/ATS59501.2023.10317952](https://doi.org/10.1109/ATS59501.2023.10317952).
- [177] C.-W. Wang, K.-L. Cheng, J.-N. Lee, *et al.*, "Fault Pattern Oriented Defect Diagnosis for Memories," in *International Test Conference*, Citeseer, 2003, pp. 29–38. DOI: [10.1109/TEST.2003.1270847](https://doi.org/10.1109/TEST.2003.1270847).
- [178] S.-Y. Liu, Y.-C. Hou, C.-C. Chang, and J.-C. Lin, "Sige profile inspection by using dual beam fib system in physical failure analysis," in *Proceedings of the 20th IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*, IEEE, 2013, pp. 490–492. DOI: [10.1109/IPFA.2013.6599207](https://doi.org/10.1109/IPFA.2013.6599207).
- [179] W. Zhao, X. Zhao, B. Zhang, *et al.*, "Failure analysis in magnetic tunnel junction nanopillar with interfacial perpendicular magnetic anisotropy," *Materials*, vol. 9, no. 1, p. 41, 2016. DOI: <https://doi.org/10.3390/ma9010041>.
- [180] T. Devolder, A. Le Goff, and V. Nikitin, "Size dependence of nanosecond-scale spin-torque switching in perpendicularly magnetized tunnel junctions," *Physical Review B*, vol. 93, no. 22, p. 224432, 2016. DOI: <https://doi.org/10.1103/PhysRevB.93.224432>.
- [181] **S. Yuan**, C. Wang, M. Fieback, *et al.*, "Device-Aware Test for Anomalous Charge Trapping in FeFETs," in *30th Asia and South Pacific Design Automation Conference (ASP-DAC)*, Tokyo, Japan, 2025, pp. 635–641. DOI: <https://doi.org/10.1145/3658617.3697755>.
- [182] **S. Yuan**, H. Xun, W. Kim, *et al.*, "Testing STT-MRAMs: Do We Need Magnets in our Automated Test Equipment?" In *IEEE International Test Conference (ITC)*, San Diego, CA, USA, 2024, pp. 364–373. DOI: [10.1109/ITC51657.2024.00058](https://doi.org/10.1109/ITC51657.2024.00058).
- [183] **S. Yuan**, M. A. Yaldagard, H. Xun, *et al.*, "Design-for-Test for Intermittent Faults in STT-MRAMs," in *IEEE European Test Symposium (ETS)*, The Hague, Netherlands, 2024, pp. 1–6. DOI: [10.1109/ETS61313.2024.10567702](https://doi.org/10.1109/ETS61313.2024.10567702).

- [184] A. Aouichi, **S. Yuan**, M. Fieback, *et al.*, “Device Aware Diagnosis for Unique Defects in STT-MRAMs,” in *2023 IEEE 32nd Asian Test Symposium (ATS)*, IEEE, 2023, pp. 1–6.
- [185] **S. Yuan**, Z. Zhang, M. Fieback, *et al.*, “Magnetic Coupling Based Test Development for Contact and Interconnect Defects in STT-MRAMs,” in *IEEE International Test Conference (ITC)*, Anaheim, CA, USA, 2023, pp. 236–245. DOI: [10 . 1109/ITC51656 . 2023 . 00040](https://doi.org/10.1109/ITC51656.2023.00040).
- [186] C. Wang, **S. Yuan**, H. Xun, *et al.*, “Defects, Fault Modeling, and Test Development Framework for FeFETs,” in *IEEE International Test Conference (ITC)*, San Diego, CA, USA, 2024, pp. 91–95. DOI: [10 . 1109/ITC51657 . 2024 . 00026](https://doi.org/10.1109/ITC51657.2024.00026).
- [187] H. Xun, M. Fieback, M. A. Yaldagard, *et al.*, “Robust Design-for-Testability Scheme for Conventional and Unique Defects in RRAMs,” in *IEEE International Test Conference (ITC)*, San Diego, CA, USA, 2024, pp. 374–383. DOI: [10 . 1109 / ITC51657 . 2024 . 00059](https://doi.org/10.1109/ITC51657.2024.00059).
- [188] H. Xun, M. Fieback, M. A. Yaldagard, *et al.*, “Online Detection of Unique Faults in RRAMs,” in *IEEE European Test Symposium (ETS)*, The Hague, Netherlands, 2024, pp. 1–2. DOI: [10 . 1109/ETS61313 . 2024 . 10567352](https://doi.org/10.1109/ETS61313.2024.10567352).
- [189] H. Xun, M. Fieback, **S. Yuan**, H. Aziza, M. Taouil, and S. Hamdioui, “Device-Aware Diagnosis for Yield Learning in RRAMs,” in *IEEE/ACM Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Valencia, Spain, 2024, pp. 1–6. DOI: [10 . 23919/DAT58400 . 2024 . 10546660](https://doi.org/10.23919/DAT58400.2024.10546660).
- [190] H. Xun, M. Fieback, **S. Yuan**, *et al.*, “Characterization and Test of Intermittent Over RESET in RRAMs,” in *IEEE 32nd Asian Test Symposium (ATS)*, Beijing, China, 2023, pp. 1–6. DOI: [10 . 1109/ATS59501 . 2023 . 10317990](https://doi.org/10.1109/ATS59501.2023.10317990).
- [191] H. Xun, **S. Yuan**, M. Fieback, H. Aziza, M. Taouil, and S. Hamdioui, “Device-Aware Test for Ion Depletion Defects in RRAMs,” in *IEEE International Test Conference (ITC)*, Anaheim, CA, USA, 2023, pp. 246–255. DOI: [10 . 1109 / ITC51656 . 2023 . 00041](https://doi.org/10.1109/ITC51656.2023.00041).
- [192] H. Xun, M. Fieback, **S. Yuan**, Z. Zhang, M. Taouil, and S. Hamdioui, “Data Background-Based Test Development for All Interconnect and Contact Defects in RRAMs,” in *2023 IEEE European Test Symposium (ETS)*, IEEE, 2023, pp. 1–6.

NOMENCLATURE

ACRONYMS

RL_b	Bottom Reference Layer
RL_t	Top Reference Layer
AI	Artificial Intelligence
ATE	Automatic Testing Equipment
BE	Bottom Electrode
BEOL	Back-End-of-Line
BH	Back-Hopping
BIST	Built-In Self-Test
BL	Bit Line
CAT	Cell-Aware Test
CBRAM	Conductive Bridging Random Access Memory
CF	Conducting Filament
CIM	Computing-in-Memory
CM	Current Monitoring
CM-NR	Current Monitoring Normal Read
CM-NW	Current Monitoring Normal Write
CM-SM	Current Monitoring Specific Test Mode
CMP	Chemical Mechanical Polishing
CPU	Central Processing Unit
DA-Diagnosis	Device-Aware Diagnosis
DAT	Device-Aware Test
DB	Data Background
DB-ILP	DB-Based Integer Linear Programming

DfT	Design-for-Testability
DPPB	Defective Part Per Billion
DRAM	Dynamic Random-Access Memory
ECC	Error Correcting Code
EtD	Easy-to-Detect
F	Faulty Effect
FDP	Fault Detection Probability
FE	Ferroelectric
FeFET	Ferroelectric Field-Effect Transistor
FEOL	Front-End-of-Line
FeRAM	Ferroelectric Random-Access Memory
FET	Field-Effect Transistor
FL	Free Layer
FSP	Fault Sensitization Probability
GB	gigabyte
HDD	Hard Disk Drive
HPC	High-Performance Computing
HRS	High Resistive State
HtD	Hard-to-Detect
HVT	High Threshold Voltage
IB	Intermittent Behavior
IM	Intermediate State
KB	kilobyte
LRS	Low Resistive State
LVT	Low Threshold Voltage
MB	megabyte
MIM	Metal-Insulator-Metal
MOSFET	Metal-Oxide-Semiconductor Field-Effect Transistor

MRAM	Magnetoresistive Random-Access Memory
MTJ	Magnetic Tunnel Junction
MUX	Multiplexer
NA	Not affected
NVM	Non-Volatile Memory
OV	Oxygen Vacancies
OxRAM	Oxide-based Resistive Random Access Memory
PB	petabyte
PCRAM	Phase-Change Memory
PFA	Physical Failure Analysis
PL	Pinned Layer
R	Readout Value
RA	Resistance-Area
RAM	Random-Access Memory
RD	Reversed Direction
RL	Reference Layer
RRAM	Resistive Random-Access Memory
S	Sensitizing Sequence
SA	Sense Amplifier
SAFF	Synthetic Anti-Ferromagnet Flip
SL	Source Line
SOT	Spin-Orbit Torque
SOT-MRAM	Spin-Orbit Torque MRAM
SRAM	Static Random-Access Memory
SSD	Solid State Drive
STT	Spin-Transfer Torque
STT-MRAM	Spin-Transfer Torque MRAM
TB	Tunnel Barrier

TB	terabyte
TE	Top Electrode
TEM	Transmission Electron Microscopy
TMR	Tunneling Magnetoresistance
VCMA-MRAM	Voltage-Controlled Magnetic Anisotropy MRAM
VLSI	Very Large Scale Integration
VM	Voltage Monitoring
VM-NR	Voltage Monitoring Normal Read
VM-NW	Voltage Monitoring Normal Write
VM-SM	Voltage Monitoring Specific Test Mode
W/L	MOSFET Width-to-Length
WD	Write Driver
WER	Write Error Rate
WL	Word Line

SYMBOLS

α	Damping Factor
$\bar{\phi}$	Potential Barrier Height of TB
Δt	Time Step in RER calculation
Δ	Thermal Stability
η	STT Efficiency
γ	Gyromagnetic Ratio
\hbar	Reduced Planck Constant
$\langle m_z \rangle$	Initial MTJ State in RER calculation
μ_0	Vacuum Permeability
μ_B	Bohr Magneton
τ_0	Inverse of the Attempt Frequency
θ	Angle between m_{FL} and easy-axis
θ_0	Initial Angle between m_{FL} and easy-axis
C	Euler Constant
C_a	Aggressor-Cell
E_B	Energy Barrier between P and AP state
E_{gap}	Input Energy
F	Fitting Coefficient
H_{dir}	R-H Hysteresis Loop Direction
H_{ext}	External Magnetic Field
H_k	Anisotropy Magnetic Field
H_p	Effective Magnetic Field for coupling effect
$H_{s-inter-FL}$	Inter-Cell Stray Field by FL
$H_{s-inter-PL}$	Inter-Cell Stray Field by PL
$H_{s-inter}$	Inter-Cell Stray Field
$H_{s-intra}$	Intra-Cell Stray Field
H_s	Stray Field

I_c	Critical Current
I_{rd}	Read Current
I_{ref}	Reference Current
I_{s-FL}	Equivalent Current by $H_{s-inter-FL}$
I_{w0}	Write Voltage in $w0$ operations
I_{w1}	Write Voltage in $w1$ operations
I_w	Write Voltage in write operations
k_B	Boltzmann Constant
m_{FL}	FL Magnetization
M_s	Saturation Magnetization
P_{sw}	Switching Probability
R_0	TB Thickness
R_{AP}	MTJ Resistance in AP state
R_{cell-0}	Cell Resistance in state '0'
R_{cell-1}	Cell Resistance in state '1'
$R_{cell-IM}$	Cell Resistance in intermediate state
R_p	MTJ Resistance in P state
RER_{spec}	Spec of RER
s	Fitting Coefficient
T	Temperature
T	Total Input Torque
t_0	Characteristic Relaxation Time
t_w	Average Switching Time
t_{FL}	Thickness the FL
t_{ox}	TB Thickness
t_p	Pulse Width
$TMR(0)$	TMR at 0 V
V_h	Bias Voltage when $TMR(V_h) = 0.5 \cdot TMR(0)$

V_{DD}	Supply Voltage
V_p	Pulse Height
V_{th}	Threshold Voltage
WER_{spec}	Spec of WER
A	Cross-Section Area
ATE	Automatic Testing Equipment
b	Fitting Coefficient
coef	Fitting Coefficient
e	Unit Charge
eCD	Critical Diameter
m	FL Magnetization
t	WER at a certain time point 't'

CURRICULUM VITÆ

Sicong YUAN

PERSONAL INFO.

Emails: sicongyuan@outlook.com or s.yuan-4@tudelft.nl

EDUCATION

2021–2025	Ph.D. degree in Computer Engineering Delft University of Technology, the Netherlands <i>Thesis title:</i> <i>Supervisors:</i> Prof. dr. ir. S. Hamdioui, Dr. ir. M. Taouil
2017–2020	M. Eng. degree in Electronic Engineering Zhejiang University, China
2013–2017	B. Eng. degree in Information Science Zhejiang University, China

MAIN PUBLICATIONS

6. **S. Yuan**, C. Wang, M. Fieback, H. Xun, M. Taouil, X. Li, L. Wang, D. Chen, N. Bellarmino, R. Cantoro, and S. Hamdioui, “Device-Aware Test for Anomalous Charge Trapping in Fe-FETs,” in *30th Asia and South Pacific Design Automation Conference (ASP-DAC)*, Tokyo, Japan, 2025, pp. 635–641. DOI: <https://doi.org/10.1145/3658617.3697755>
5. **S. Yuan**, H. Xun, W. Kim, S. Rao, E. J. Marinissen, S. Couet, M. Fieback, M. Taouil, and S. Hamdioui, “Testing STT-MRAMs: Do We Need Magnets in our Automated Test Equipment?” In *IEEE International Test Conference (ITC)*, San Diego, CA, USA, 2024, pp. 364–373. DOI: [10.1109/ITC51657.2024.00058](https://doi.org/10.1109/ITC51657.2024.00058)
4. **S. Yuan**, M. A. Yaldagard, H. Xun, M. Fieback, E. J. Marinissen, W. Kim, S. Rao, S. Couet, M. Taouil, and S. Hamdioui, “Design-for-Test for Intermittent Faults in STT-MRAMs,” in *IEEE European Test Symposium (ETS)*, The Hague, Netherlands, 2024, pp. 1–6. DOI: [10.1109/ETS61313.2024.10567702](https://doi.org/10.1109/ETS61313.2024.10567702)
3. A. Aouichi, **S. Yuan**, M. Fieback, S. Rao, W. Kim, E. J. Marinissen, S. Couet, M. Taouil, and S. Hamdioui, “Device Aware Diagnosis for Unique Defects in STT-MRAMs,” in *2023 IEEE 32nd Asian Test Symposium (ATS)*, IEEE, 2023, pp. 1–6
2. **S. Yuan**, Z. Zhang, M. Fieback, H. Xun, E. J. Marinissen, G. S. Kar, S. Rao, S. Couet, M. Taouil, and S. Hamdioui, “Magnetic Coupling Based Test Development for Contact and Interconnect Defects in STT-MRAMs,” in *IEEE International Test Conference (ITC)*, Anaheim, CA, USA, 2023, pp. 236–245. DOI: [10.1109/ITC51656.2023.00040](https://doi.org/10.1109/ITC51656.2023.00040)
1. **S. Yuan**, M. Taouil, M. Fieback, H. Xun, E. J. Marinissen, G. S. Kar, S. Rao, S. Couet, and S. Hamdioui, “Device-Aware Test for Back-Hopping Defects in STT-MRAMs,” in *IEEE/ACM Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Antwerp, Belgium, 2023, pp. 1–6. DOI: [10.23919/DATE56975.2023.10137071](https://doi.org/10.23919/DATE56975.2023.10137071)

OTHER PUBLICATIONS

7. C. Wang, **S. Yuan**, H. Xun, C. Li, M. Taouil, M. Fieback, D. Chen, X. Li, L. Wang, R. Cantoro, C. Yin, and S. Hamdioui, “Defects, Fault Modeling, and Test Development Framework for FeFETs,” in *IEEE International Test Conference (ITC)*, San Diego, CA, USA, 2024, pp. 91–95. DOI: [10.1109/ITC51657.2024.00026](https://doi.org/10.1109/ITC51657.2024.00026)
6. H. Xun, M. Fieback, M. A. Yaldagard, **S. Yuan**, E. Hua, H. Aziza, M. Taouil, and S. Hamdioui, “Robust Design-for-Testability Scheme for Conventional and Unique Defects in RRAMs,” in *IEEE International Test Conference (ITC)*, San Diego, CA, USA, 2024, pp. 374–383. DOI: [10.1109/ITC51657.2024.00059](https://doi.org/10.1109/ITC51657.2024.00059)
5. H. Xun, M. Fieback, M. A. Yaldagard, **S. Yuan**, H. Aziza, M. Taouil, and S. Hamdioui, “Online Detection of Unique Faults in RRAMs,” in *IEEE European Test Symposium (ETS)*, The Hague, Netherlands, 2024, pp. 1–2. DOI: [10.1109/ETS61313.2024.10567352](https://doi.org/10.1109/ETS61313.2024.10567352)
4. H. Xun, M. Fieback, **S. Yuan**, H. Aziza, M. Taouil, and S. Hamdioui, “Device-Aware Diagnosis for Yield Learning in RRAMs,” in *IEEE/ACM Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Valencia, Spain, 2024, pp. 1–6. DOI: [10.23919/DATE58400.2024.10546660](https://doi.org/10.23919/DATE58400.2024.10546660)
3. H. Xun, M. Fieback, **S. Yuan**, H. Aziza, M. Heidekamp, T. Copetti, L. B. Poehls, M. Taouil, and S. Hamdioui, “Characterization and Test of Intermittent Over RESET in RRAMs,” in *IEEE 32nd Asian Test Symposium (ATS)*, Beijing, China, 2023, pp. 1–6. DOI: [10.1109/ATS59501.2023.10317990](https://doi.org/10.1109/ATS59501.2023.10317990)
2. H. Xun, **S. Yuan**, M. Fieback, H. Aziza, M. Taouil, and S. Hamdioui, “Device-Aware Test for Ion Depletion Defects in RRAMs,” in *IEEE International Test Conference (ITC)*, Anaheim, CA, USA, 2023, pp. 246–255. DOI: [10.1109/ITC51656.2023.00041](https://doi.org/10.1109/ITC51656.2023.00041)
1. H. Xun, M. Fieback, **S. Yuan**, Z. Zhang, M. Taouil, and S. Hamdioui, “Data Background-Based Test Development for All Interconnect and Contact Defects in RRAMs,” in *2023 IEEE European Test Symposium (ETS)*, IEEE, 2023, pp. 1–6

