Advanced Testing and Reliability Enhancements for RRAMs

Xun, Hanzhi

DOI
10.4233/uuid:58806228-60e2-48e9-9abc-2af03fa84006

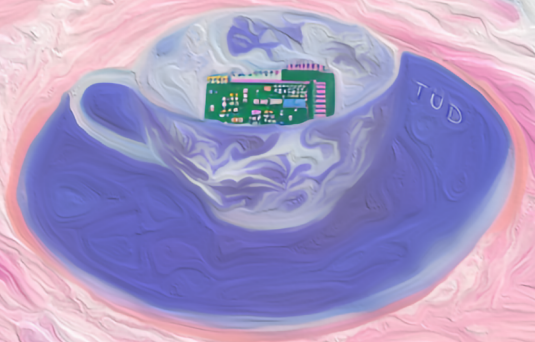
Publication date
2025

Document Version
Final published version

Citation (APA)
Xun, H. (2025). *Advanced Testing and Reliability Enhancements for RRAMs*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:58806228-60e2-48e9-9abc-2af03fa84006

Advanced Testing and Reliability Enhancements for RRAMs

Hanzhi Xun



INVITATION

It is my pleasure to invite you to attend the public defence of my PhD thesis titled:

Advanced Testing and Reliability Enhancements for RRAMs

on Thursday, September 11, 2025, 10:00 a.m. in the Senaatszaal of the TU Delft Aula at Mekelweg 5, Delft.

I will introduce my research briefly at 9:30 a.m.

You are also welcome to attend the reception that takes place after the defence.

Hanzhi Xun

Advanced Testing and Reliability Enhancements for RRAMs

ADVANCED TESTING AND RELIABILITY ENHANCEMENTS FOR RRAMS

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 11 September 2025 at 10:00 o'clock

by

Hanzhi XUN

Master of Science in Information, Production and Systems Engineering
Waseda University, Japan
Master of Engineering in Software Engineering
Xidian University, China
born in Xi'an, Shaanxi, China

This dissertation has been approved by the promoters.

Composition of the doctoral committee:

Rector Magnificus,	Chairperson
Prof.dr.ir. S. Hamdioui,	Delft University of Technology, promotor
Dr.ir. M. Fieback,	Delft University of Technology, copromotor
Dr.ir. M. Taouil,	Delft University of Technology
Independent members:	
Prof.dr. G. Q. Zhang,	Delft University of Technology
Prof.dr. H. W. Li,	ICT, Chinese Academy of Sciences, China
Dr.ir. M. Ottavi,	University of Twente, the Netherlands
Dr. B. Kruseman,	NXP Semiconductors, the Netherlands
Prof.dr.ir. G. N. Gaydadjiev,	Delft University of Technology, reserve member

Keywords: resistive RAM, memristor, defects, defect modeling, fault modeling, test development, design-for-testability, reliability, Edge-AI, neuromorphic computing

Printed by: Ipskamp Printing, The Netherlands

Front & Back: Cover artwork by Hanzhi Xun

ISBN 978-94-6473-918-3

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

Dedicated to:
my parents for raising me up and supporting me in education,
those who have used time to make me a better student, researcher, and human being.

CONTENTS

SUMMARY

This dissertation, conducted within the discipline of Electronic Science and Technology (specialization in Microelectronics and Solid-State Electronics), focuses on Resistive Random Access Memory (RRAM), an emerging non-volatile memory technology known for its high density and zero static power consumption. RRAM enables fast write and read operations in the nanosecond range and supports Computation-in-Memory (CIM), making it a strong candidate to replace Flash or even Dynamic Random Access Memory (DRAM). Recognizing its potential, both academic institutions and industry leaders have been actively developing RRAM prototypes, with some already reaching the commercial market. To ensure reliability, high-quality testing is essential for guaranteeing product quality.

This dissertation mainly focuses on developing effective test methodologies and robust designs for RRAMs. We begin by examining the RRAM manufacturing process and identifying potential physical defects at each stage through a comprehensive literature review and silicon measurements. To facilitate in-depth analysis, we develop a complete and systematic RRAM simulation platform, integrating a MATLAB-based simulation controller and fault analysis scripts integrated with a complete RRAM circuit design. The controller automates and manages all simulation procedures, while the circuit design comprises a 1T-1R memory array along with essential peripheral components such as write drivers and sense amplifiers. To achieve fast and accurate electrical simulations, we introduce two compact models for RRAMs. These models are optimized and calibrated using extensive measurement data from RRAM devices. We further calibrate the model with industrial measurements from ST Microelectronics. It enables robust device/circuit co-design, accounting for PVT variations and ensuring the reliability and efficiency of RRAM systems.

By using the simulation platform, we investigate RRAM testing for interconnects and contacts by using conventional defect/fault modeling and test designs. In this approach, physical defects are modeled as linear resistors (e.g., open, short, or bridge) and injected into the RRAM array netlist for fault analysis. Test development is then carried out based on the identified fault behaviors. Both intra-cell and inter-cell defects are analyzed, confirming the presence of not only single-cell faults but also multi-cell faults. While modeling interconnect defects as linear resistors is a well-established practice, this method has not been validated for defects inside the RRAM devices.

To tackle this challenge, we apply the Device-Aware Test (DAT) approach that can incorporate physical behaviors in RRAM devices. The DAT approach is applied to two observed unique defects, ion depletion and over RESET, as case studies in this dissertation. Based on this approach, accurate defect models, realistic fault models, and efficient Design-for-Testability (DfT) designs are provided.

While existing designs primarily focus on individual defects or a limited set of fault behaviors, this dissertation introduces a comprehensive Design-for-Testability (DfT)

framework tailored for RRAMs. The proposed framework is capable of effectively detecting both conventional and unique defects, ensuring broad defect coverage with optimized test cost. Furthermore, the framework is designed with robustness in mind, offering strong resilience against process variations that often compromise the effectiveness of traditional test solutions. By addressing both conventional and unique defects, the proposed DfT strategies establish a unified methodology that advances beyond existing fragmented approaches.

In addition to defect- and fault-oriented challenges, this dissertation also investigates time-dependent non-idealities inherent to RRAM technology. A key example is the Read Disturb Fault (RDF), which gradually degrades device reliability during repeated read operations. To overcome this limitation, we propose a bi-directional circuit architecture that can continuously monitor the device state and actively counteract RDF-induced degradation. By dynamically recovering the device state through controlled operations, this design-for-reliability scheme significantly extends the lifetime and stability of RRAM devices. Importantly, the proposed approach has been validated across both storage-class RRAM and CIM architectures, demonstrating its versatility and ensuring robust long-term system operation.

Finally, a diagnosis framework was presented that leverages accurate defect models and distinctive features to uniquely localize defects, enabling precise defect classification for yield learning and process optimization. Collectively, these contributions establish a holistic methodology for testing, reliability enhancement, and diagnosis of RRAMs, paving the way toward their dependable integration into future memory and computing platforms.

SAMENVATTING

Dit proefschrift, uitgevoerd binnen de discipline Electronic Science and Technology (specialisatie in Micro-elektronica en Vastestofelektronica), richt zich op Resistive Random Access Memory (RRAM), een opkomende niet-vluchtige geheugentechnologie die bekendstaat om haar hoge dichtheid en nul statisch stroomverbruik. RRAM maakt snelle schrijf- en leesbewerkingen in het nanosecondebereik mogelijk en ondersteunt Computation-in-Memory (CIM), waardoor het een sterke kandidaat is om Flash of zelfs Dynamic Random Access Memory (DRAM) te vervangen. Gezien het potentieel zijn zowel academische instellingen als industriële leiders actief bezig met de ontwikkeling van RRAM-prototypes, waarvan sommige al de commerciële markt hebben bereikt. Om de betrouwbaarheid te waarborgen, is hoogwaardige testmethodologie essentieel voor het garanderen van productkwaliteit.

Dit proefschrift richt zich voornamelijk op het ontwikkelen van effectieve testmethodologieën en robuuste ontwerpen voor RRAM's. We beginnen met het onderzoeken van het RRAM-productieproces en het identificeren van mogelijke fysieke defecten in elke fase via een uitgebreide literatuurstudie en siliciummetingen. Om diepgaande analyses mogelijk te maken, ontwikkelen we een volledig en systematisch RRAM-simulatieplatform, waarin een MATLAB-gebaseerde simulatiecontroller en foutanalysescripts geïntegreerd zijn met een compleet RRAM-circuitontwerp. De controller automatiseert en beheert alle simulatieprocedures, terwijl het circuitontwerp bestaat uit een 1T-1R-geheugenarray met essentiële perifere componenten zoals schrijfstuurprogramma's en sense amplifiers. Voor snelle en nauwkeurige elektrische simulaties introduceren we twee compacte RRAM-modellen. Deze modellen zijn geoptimaliseerd en gekalibreerd met uitgebreide meetgegevens van RRAM-apparaten, en verder verfijnd met industriële metingen van STMicroelectronics. Dit maakt robuust apparaat-/circuit-co-ontwerp mogelijk, waarbij PVT-variaties in rekening worden gebracht en de betrouwbaarheid en efficiëntie van RRAM-systemen worden gewaarborgd.

Met behulp van het simulatieplatform onderzoeken we RRAM-testen voor interconnecties en contacten, gebaseerd op conventionele defect-/foutmodellering en testontwerpen. In deze benadering worden fysieke defecten gemodelleerd als lineaire weerstanden (bijv. open, short of bridge) en geïnjecteerd in de RRAM-array netlist voor foutanalyse. Testontwikkeling wordt vervolgens uitgevoerd op basis van de geïdentificeerde foutgedragingen. Zowel intra-cellulaire als inter-cellulaire defecten worden geanalyseerd, wat bevestigt dat niet alleen single-cell fouten voorkomen, maar ook multi-cell fouten. Hoewel het modelleren van interconnect-defecten als lineaire weerstanden een gevestigde praktijk is, is deze methode nog niet gevalideerd voor defecten binnen de RRAM-apparaten zelf.

Om deze uitdaging aan te pakken, passen we de Device-Aware Test (DAT)-benadering toe, waarmee fysische gedragingen van RRAM-apparaten in rekening kunnen worden gebracht. Deze aanpak wordt toegepast op twee waargenomen unieke de-

fecten, ionendepletie en over-RESET, als casestudies in dit proefschrift. Op basis van deze aanpak worden nauwkeurige defectmodellen, realistische foutmodellen en efficiënte Design-for-Testability (DfT)-ontwerpen ontwikkeld.

Waar bestaande ontwerpen zich voornamelijk richten op afzonderlijke defecten of een beperkte set foutgedragingen, introduceert dit proefschrift een omvattend DfT-raamwerk dat specifiek is afgestemd op RRAM's. Het voorgestelde raamwerk kan zowel conventionele als unieke defecten effectief detecteren, wat resulteert in een brede defectdekking met geoptimaliseerde testkosten. Bovendien is het raamwerk ontworpen met robuustheid in gedachten, waardoor het een sterke weerstand biedt tegen procesvariaties die vaak de effectiviteit van traditionele testoplossingen ondermijnen. Door zowel conventionele als unieke defecten aan te pakken, vormen de voorgestelde DfT-strategieën een verenigde methodologie die verder gaat dan de bestaande gefragmenteerde benaderingen.

Naast defect- en foutgerichte uitdagingen onderzoekt dit proefschrift ook tijdafhankelijke niet-idealiteiten die inherent zijn aan RRAM-technologie. Een belangrijk voorbeeld hiervan is de Read Disturb Fault (RDF), die de betrouwbaarheid van het apparaat geleidelijk vermindert bij herhaalde leesbewerkingen. Om deze beperking te overwinnen, stellen we een bi-directionele circuitarchitectuur voor die de toestand van het apparaat continu kan monitoren en actief RDF-geïnduceerde degradatie tegengaat. Door de toestand van het apparaat dynamisch te herstellen via gecontroleerde operaties, verlengt dit design-for-reliability schema de levensduur en stabiliteit van RRAM-apparaten aanzienlijk. Belangrijk is dat de voorgestelde aanpak is gevalideerd voor zowel storage-class RRAM als CIM-architecturen, wat de veelzijdigheid aantoont en robuuste, langdurige systeemwerking waarborgt.

Tot slot wordt een diagnoseraamwerk gepresenteerd dat gebruikmaakt van nauwkeurige defectmodellen en onderscheidende kenmerken om defecten uniek te lokaliseren. Dit maakt een precieze defectclassificatie mogelijk, die waardevolle feedback biedt voor yield learning en procesoptimalisatie. Gezamenlijk vestigen deze bijdragen een holistische methodologie voor het testen, verbeteren van betrouwbaarheid en diagnosticeren van RRAM's, en banen zij de weg voor een betrouwbare integratie ervan in toekomstige geheugen- en computerplatformen.

ACKNOWLEDGEMENTS

After four years of exploration, effort, and learning, I am now standing at the finish line of my PhD journey. This moment is not just a personal milestone, but also a reminder of the support, inspiration, and kindness I have received from many people along the way. This thesis, although bearing my name, is the result of collective guidance, collaboration, and friendship. I am sincerely grateful to everyone who has walked alongside me.

First and foremost, I would like to express my deepest gratitude to my promotor and supervisor, **Prof.dr.ir. Said Hamdioui**. Thank you for offering me the opportunity to pursue my PhD under your guidance. From the very beginning, you welcomed me into your group and provided everything I needed to grow, including research freedom, access to essential resources, the opportunity to supervise students, and participation in academic activities such as summer schools. These experiences broadened my academic perspective and helped me mature as a researcher. Throughout the years, you taught me how to think critically, communicate effectively in a scientific context, and face challenges with confidence. I still remember the many hours spent revising papers, polishing presentations, and carefully preparing for deadlines. Those intense sessions taught me valuable lessons that will stay with me long after this PhD. You consistently provided insightful guidance on the strategic direction of my projects. Your ability to see the bigger picture and your industrial background helped me connect my research to real-world applications. At the very beginning of my PhD, you told me that you would closely follow my development, always aiming to help me grow step by step until I met the standard of a PhD graduate. I hope that I have now reached that level. Your sense of responsibility and strong work ethic also left a deep impression on me. I have often seen you working late in the office, and even during your holidays, you remained available and supportive. Your dedication and commitment to your students are qualities I admire deeply and hope to carry forward in my own career. Thank you for your trust, your time, and your mentorship throughout this journey.

Next, I am also deeply grateful to my copromotor, **Dr.ir. Moritz Fieback**. Moritz, you've been a steady guide throughout my PhD journey. I especially appreciate the encouragement you gave me during the most stressful moments in my first year. You reminded me to breathe, take breaks, and not take everything too seriously. Your kind feedback on my first ETS paper draft gave me confidence when I needed it most. Furthermore, you thoroughly reviewed every paper I wrote, often through several versions filled with precise and insightful comments. But more than helping me fix a paper, you were teaching me how to become a better writer and thinker. That made a lasting difference. I also appreciated your sense of humor and your relaxed but effective way of working. You showed me that it's not just about working hard, but also about working smart—and enjoying the process along the way. Thank you as well for trusting me with teaching responsibilities and sharing your views on how to grow as an educator. In addition, I would like to sincerely thank **Dr.ir. Mottaqiallah Taouil**. Although not listed as my

official copromotor, you have undoubtedly played the role of a true mentor throughout my academic life. From the early stages of my PhD, you have been a constant source of support, both academically and personally. Our meetings were always filled with fresh ideas and your confident, solution-driven mindset often helped push projects forward when I was stuck. Beyond research, you genuinely cared about how I was doing, and I deeply appreciate the attention you paid to my well-being. Thank you for your guidance, your creativity, and your kindness.

I am grateful to the committee members of this thesis, **Prof.dr. Guoqi Zhang**, **Prof.dr. Huawei Li**, **Dr.ir. Marco Ottavi**, **Dr. Bram Kruseman**, and **Prof.dr. Georgi Gaydadjiev**. Your insightful feedback has been crucial in enhancing my thesis, and I eagerly anticipate witnessing the influence of our future collaborative efforts.

I would also like to thank my academic collaborators. First, **Dr. Hassen Aziza**, thank you for introducing me to the fundamentals of RRAM design, for the thoughtful discussions, and for sharing important measurement data. I am especially grateful that you always provide your feedback very fast and helpfully. I truly appreciate your openness and generosity, and I hope we can continue this productive collaboration in the future. Next, I would like to thank **Dr. Leticia Bolzani Poehls** and **Dr. Thiago Copetti** for the work we have done together. We explored meaningful topics and laid a strong foundation for future research. It has been a pleasure working with you both. My thanks also go to **Dr. Riccardo Cantoro** for the enjoyable and insightful collaboration. I learned a lot from our technical discussions on compute-in-memory systems and on the contrasting characteristics of FeFET. Beyond external collaborations, I am grateful for the opportunity to work with talented researchers at TU Delft. Thank you, **Dr. Rajendra Bishnoi** and **Dr. Anteneh Gebregiorgis**, for your valuable input and kind support throughout different phases of my work.

Thanks as well to both my paranymphs, **Sicong Yuan** and **Erbing Hua**. Sicong, we started our PhD journeys around the same time and have worked closely ever since, not only because of our overlapping research topics, but also because of the friendship we've built along the way. Thank you for always being so generous—whether it was inviting us over for your home-cooked meals or sharing your thoughts and feedback on research ideas. I've learned a lot from our discussions and truly enjoyed your company over the years. Erbing, thank you for broadening my perspective by bringing in insights from materials science, and even beyond that—from social sciences and philosophy. Our conversations have often sparked new ways for me to think about research and its role. Both of you have been not just colleagues, but real companions throughout this journey. I will always treasure the memories we made, including our unforgettable road trip across Europe. I feel very lucky to have shared this PhD journey with you.

I would like to extend my thanks to all colleagues at the QCE department. Thanks to **Prof.dr. Rob Kooij** for your efforts in creating a nice working environment and supporting all different kinds of social events in the department. I also had the pleasure of sharing my workday with many other people in the department, including but not limited to **Dr. Abdullah**, **Dr. Abhairaj**, **Dr. Anteneh**, **Dr. Carlo**, **Dr. Guilherme**, **Dr. Heba**, **Dr. Lizhou**, **Dr. Mahdi**, **Dr. Micheal (Taha)**, **Dr. Moritz**, **Dr. Motta**, **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**. Thanks for the nice lunches that we had together, the daily coffee walks, and many more events. Amin, Asmae, and Mainak, thanks for organizing social events and sport activities. Francis, Helen, and Paul, thank you for keeping everything in order and for ensuring that my work could be performed effectively. Erik, Mark, and Ediz, thanks for your great help in setting up the simulation environments.

I also want to express my gratitude to some of my Chinese PhD fellows at TU Delft, and friends I have ever met. Thank you for always being there, whether it was for a spontaneous lunch, a late-night walk, or simply sharing thoughts on research and life. These everyday moments made the PhD journey lighter and more enjoyable. I also want to thank my friends in China. Even though we have been separated by thousands of miles, your care and support always reached me across the distance. The occasional messages, phone calls, and genuine concern reminded me that I was never alone on this journey. Your presence, near or far, has meant a lot to me.

Last but most importantly, I would like to express my deepest thanks to my family. To my parents, who have always been my pillars of strength and my guiding lights. Your unwavering and unconditional support has shaped the person I am today and enabled everything I have achieved. I owe this journey to you. To my grandparents, uncles, aunts, brother, and sister, thank you for your love, wisdom, and contribution to a lovely and warm big family. Your presence in my life means more than words can say. And finally, I want to thank myself for being brave, for holding on during the toughest moments, and for never giving up. This journey has not been easy, but I am proud that I made it here. Many thanks to all of you!

Hanzhi Xun
Delft, July, 2025

LIST OF FIGURES

List of Tables

1

INTRODUCTION

Resistive Random Access Memory (RRAM) 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 RRAM products. This dissertation focuses on investigating manufacturing defects introduced during RRAM fabrication, deriving corresponding fault models, and developing robust test solutions. The first chapter of this dissertation presents an introduction of the dissertation, including 1) a review of emerging non-volatile memory technologies; 2) an in-depth look at RRAMs, 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.

1

1.1. EMERGING MEMORY TECHNOLOGIES

This section introduces the emerging memory technology, which is the research target of this dissertation. First, we describe the memory hierarchy and the motivation to develop emerging memories. Then, we introduced various types of emerging memories individually and overview the potential applications.

1.1.1. MEMORY WALL AND HIERARCHY

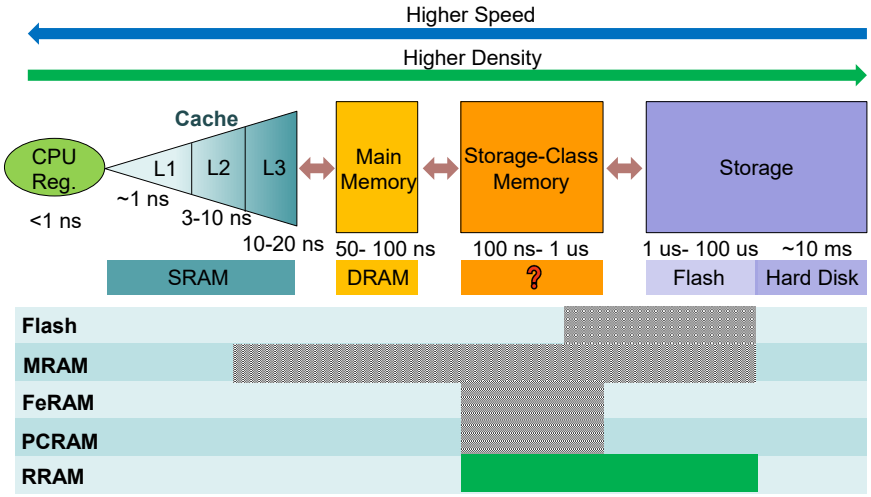


Figure 1.1: The memory hierarchy in computer systems.

The conventional digital computing hardware features physically separated processing and memory units [1]. The frequent and massive data movement between these units leads to time latency and power consumption penalties, the so-called Von Neumann bottleneck [2]–[4]. This is also known as the memory wall problem in von Neumann architecture due to the difference in the processing speed of the calculator and memory. Besides, the past development of digital electronics is primarily fueled by transistors' scaling, which is becoming increasingly cost-ineffective as a transistor's size approaches its physical limits [5], [6]. Therefore, fundamental changes to the paradigm of computing systems are critical.

A hierarchical memory structure, designed to leverage locality principles and optimize the balance between cost and performance across various memory technologies, serves as an efficient solution to mitigate the memory wall challenge and establish an ideal memory system. Programs often exhibit a pattern of reusing previously accessed data and instructions, a concept known as locality [7]. Fig. 1.1 presents a memory hierarchy, illustrating the placement of various memory technologies and the typical access speeds at each level. This hierarchy is generally divided into four main tiers: cache, main memory, storage-class memory, and mass storage, which are primarily realized using Static Random Access Memory (SRAM), Dynamic Random Access Memory (DRAM), and Hard Disk Drives (HDDs), respectively. Positioned closest to the CPU, the cache is a

high-speed, small-capacity, and relatively expensive memory layer designed to provide rapid data access. Following the cache is the main memory, which offers a balance between speed and cost, serving as an intermediary storage between the cache and mass storage. The storage-class memory has emerged as a promising technology that bridges the performance gap between conventional memories (DRAM, SRAM) and storage devices (NAND flash), offering non-volatility, high speed, and scalability. The mass storage layer, located furthest from the CPU, features the largest capacity but the slowest access speed. Data stored in layers further from the CPU typically includes all or a subset of the data found in closer layers. The purpose of this hierarchical structure is to achieve a memory system that closely matches the speed of the fastest tier while maintaining a cost per bit that approaches that of the most economical layer.

Non-Volatile Memories (NVMs) retain stored data even when power is removed, playing a crucial role in ensuring data persistence in computing systems [8], [9]. Traditional NVMs, such as Flash memory, have been widely adopted in applications ranging from consumer electronics to enterprise storage due to their high density and cost efficiency. However, as demand for faster, more energy-efficient, and highly durable memory solutions increases, emerging NVM technologies have garnered significant attention.

Among these next-generation memories, Resistive Random-Access Memory (RRAM) leverages resistive switching in metal oxides to achieve high scalability and low power consumption [10], [11]. Magnetoresistive Random-Access Memory (MRAM) utilizes Magnetic Tunnel Junctions (MTJs) with magnetization-switchable ferromagnetic layers, providing fast access speeds and excellent endurance [12], [13]. Ferroelectric Field-Effect Transistor (FeFET) memory incorporates a ferroelectric layer into the gate of a transistor, enabling low-power, non-volatile charge storage with CMOS compatibility [14], [15]. Ferroelectric Random-Access Memory (FeRAM) stores data through the polarization states of a ferroelectric material, offering fast read and write operations, low power consumption, and high endurance [16], [17]. Phase-Change Memory (PCRAM) exploits the reversible phase transition between amorphous and crystalline states in chalcogenide materials, delivering non-volatile storage with high-speed operation and good scalability [18], [19].

Each of these emerging NVM technologies presents unique trade-offs in terms of speed, power efficiency, endurance, and scalability. As the limitations of Flash memory become more evident, these novel memory solutions are being actively investigated for applications in embedded systems, AI accelerators, neuromorphic computing, and next-generation storage architectures. Moving forward, NVMs are expected to replace traditional memory technologies in computing systems [8], [9]. Advancements in fabrication processes and material engineering continue to improve the density, energy efficiency, and reliability of these memories, making them increasingly viable for future computing systems.

With the growing demand for high-speed, low-power, and non-volatile storage in artificial intelligence, edge computing, and cloud infrastructure, emerging NVMs are poised to play a critical role in reshaping the memory hierarchy. While challenges such as cost, scalability, and seamless integration with existing architectures remain, ongoing research and development efforts are accelerating their path toward commercial-

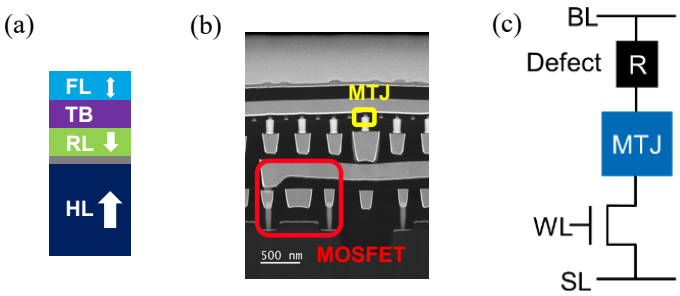


Figure 1.2: (a) MTJ device structure, (b) TEM of MTJ, (c) 1T-1MTJ cell.

ization. Ultimately, non-volatile memories hold the potential to redefine the balance between performance, energy efficiency, and data retention, driving innovation across a wide range of computing and storage applications.

1.1.2. EMERGING MEMORY TYPES

MAGNETORESISTIVE RANDOM-ACCESS MEMORY

MRAM is a non-volatile memory that utilizes the magnetization switching of ferromagnetic materials to store data [12], [13]. MRAM achieves high-speed operation, low power consumption, and virtually unlimited endurance, making it a promising candidate for next-generation memory applications in embedded systems, industrial electronics, and AI accelerators.

The mechanism of MRAM relies on the Tunneling Magneto Resistance (TMR) effect. An MTJ consists of two ferromagnetic layers separated by an insulating tunnel barrier. One layer has a fixed magnetization direction, while the other can switch between parallel and antiparallel orientations under an applied magnetic field or spin-polarized current. When the magnetization directions of both layers are parallel, the MTJ exhibits low resistance (0 state), whereas an antiparallel alignment results in high resistance (1 state) [20]. The resistance difference between these states enables reliable and fast read operations.

A standard MRAM cell is composed of an MTJ and a selection transistor, arranged in a crossbar memory array to enhance density and scalability. MRAM has evolved through multiple generations, including Toggle MRAM, Spin-Transfer Torque MRAM (STT-MRAM), and Voltage-Controlled MRAM (VC-MRAM). STT-MRAM, in particular, has gained commercial adoption due to its improved write efficiency, as it uses spin-polarized current rather than an external magnetic field to switch magnetization [21]; Fig. 1.2 presents the MTJ device structure, the Transmission Electron Microscopy (TEM) picture of the device, and its integration into a 1T-1MTJ cell. VC-MRAM further reduces power consumption by leveraging an electric field to modulate magnetic anisotropy, enhancing energy efficiency.

MRAM offers several advantages over conventional memory technologies [22]. It provides high-speed read and write operations, comparable to SRAM, while maintaining non-volatility. Additionally, MRAM exhibits exceptionally high endurance, withstanding virtually unlimited write cycles without degradation—a significant advantage

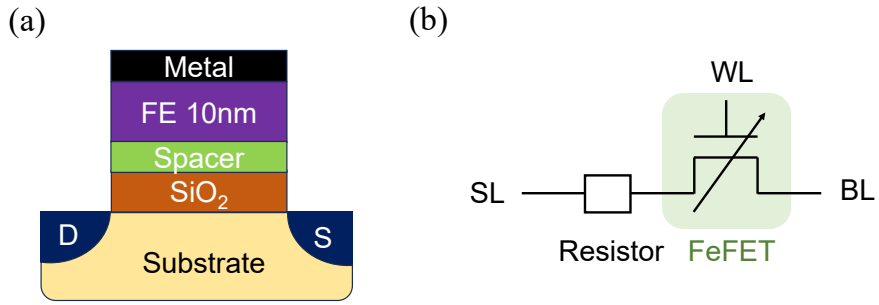


Figure 1.3: (a) Device structure of FeFET, (b) 1 FeFET-1 Resistor cell structure.

over Flash, which wears out over time. Furthermore, MRAM's low power consumption makes it well-suited for energy-efficient applications, particularly in battery-powered and always-on systems. However, challenges such as write energy reduction, scalability, and integration complexity remain areas of active research.

As MRAM technology continues to advance, it is expected to play a crucial role in cache memory, AI hardware accelerators, neuromorphic computing, and edge devices. With ongoing improvements in spintronic materials, fabrication processes, and power-efficient writing mechanisms, MRAM has the potential to complement or even replace conventional volatile and non-volatile memory solutions, driving innovation in next-generation computing architectures.

FERROELECTRIC FIELD-EFFECT TRANSISTOR

Ferroelectric Field-Effect Transistor (FeFET) Memory is a type of non-volatile memory that integrates ferroelectric materials into the gate stack of a conventional Field-Effect Transistor (FET) [16], [17]. The FeFET stores data by utilizing the polarization of the ferroelectric layer, allowing for bistable states that represent binary information, which achieves high-speed operation and long data retention.

The fundamental operating principle of FeFET relies on the stable polarization states of the ferroelectric material [16], [17]. The ferroelectric layer is made of multi-domain materials, which can be aligned in different orientations under an external electric field. When all domains are polarized positively (↑), the FeFET exhibits a high threshold voltage (V_{TH}), corresponding to a logic '0' state. Conversely, when the domains are polarized negatively (↓), the threshold voltage is lowered, representing a logic '1' state. These polarization states remain stable even after power is removed, enabling non-volatile data storage.

A typical FeFET structure consists of a FerroElectric (FE) layer integrated into the gate stack of a standard MOSFET. Fig. 1.3 illustrates the FeFET device structure along with the cell structure of 1 FeFET-1 Resistor[23]. The FE layer is typically composed of PZT (lead zirconate titanate) and $HfZrO_2$ (hafnium zirconium oxide), positioned between the gate electrode and the transistor's channel [24], [25]. Unlike conventional dielectric materials, the FE layer retains its polarization even after the external voltage is removed, enabling persistent data storage without the need for a refresh operation.

FeFETs offer several advantages over conventional memory technologies [16], [17].

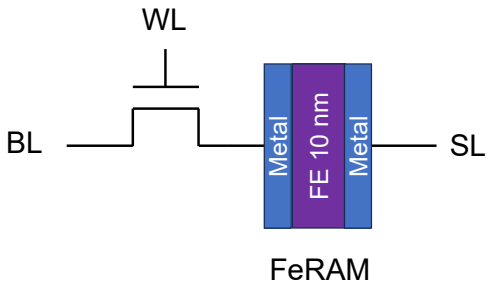


Figure 1.4: Structure of FeRAM cell and device.

They feature ultra-low power consumption, as state transitions require minimal energy, making them well-suited for energy-efficient applications. Additionally, FeFETs provide high-speed operation, with switching speeds comparable to DRAM. Their long data retention capability, exceeding ten years, surpasses that of traditional Flash memory. Moreover, FeFETs are fully compatible with CMOS fabrication processes, enabling seamless integration into existing semiconductor technology. However, challenges such as device variability, retention degradation, and scaling limitations remain areas of active research. With ongoing advancements in ferroelectric materials and fabrication techniques, FeFETs are expected to play a crucial role in the future of non-volatile memory, offering a balance between performance, power efficiency, and scalability.

FERROELECTRIC RANDOM ACCESS MEMORY

The FeRAM encodes information using the remanent polarization of a ferroelectric capacitor, allowing it to retain data even after power is removed [16], [17]. This mechanism enables fast read and write speeds, low energy consumption, and exceptional endurance, making FeRAM particularly attractive for low-power and embedded applications.

The working principle of FeRAM relies on the switchable polarization of ferroelectric materials [26]. When an external electric field is applied across the ferroelectric capacitor, its internal dipoles align, switching between two stable polarization states. These states correspond to binary values (0 and 1), enabling non-volatile data storage. However, FeRAM's read operation is destructive, meaning that each read disturbs the stored data, necessitating a rewrite afterward—similar to DRAM. Despite this, FeRAM boasts significantly faster write speeds compared to Flash and does not require high-voltage programming, making it well-suited for frequent data updates.

A standard FeRAM cell is based on a 1T-1C (one transistor, one capacitor) architecture, resembling DRAM [27]. Fig. 1.4 illustrates the FeRAM capacitor and cell structure; the capacitor is composed of ferroelectric materials such as PZT or $HfZrO_2$, sandwiched between two electrodes. The transistor serves as an access control mechanism, enabling selective reading and writing of the memory cell. Unlike traditional dielectrics, the ferroelectric layer retains its polarization state indefinitely, ensuring persistent data storage without requiring power.

FeRAM offers several compelling advantages. It consumes significantly less power

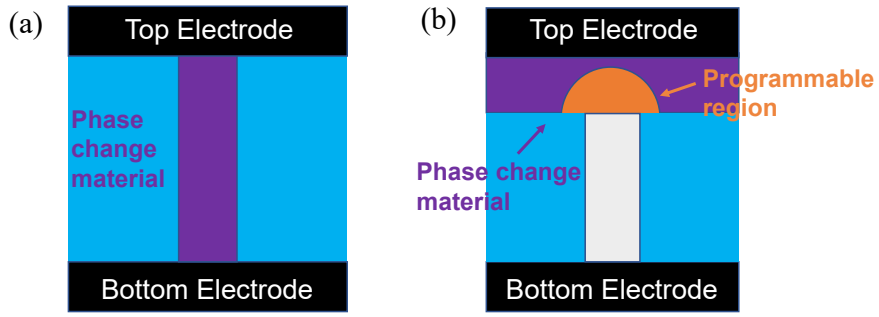


Figure 1.5: Structure of the PCRAM. (a) Confined cell structure, (b) Planar cell structure.

than Flash and DRAM, as it does not require periodic refresh cycles or high-voltage operations [28]. The high endurance of FeRAMs, capable of billions to trillions of write cycles, surpasses Flash memory while degrading over repeated writes. Additionally, FeRAM's access times are comparable to SRAM and DRAM, making it ideal for applications requiring real-time responsiveness. However, challenges remain, including limited storage density compared to Flash, which hinders FeRAM's scalability for high-capacity applications.

FeRAM is anticipated to be a key enabler in automotive electronics, smart cards, IoT applications, and industrial automation—fields that demand high reliability and minimal power consumption. Advances in materials and fabrication are poised to overcome FeRAM's present density limitations, paving the way for broader market penetration as a high-performance, energy-efficient substitute for established non-volatile memory technologies.

PHASE CHANGE RANDOM ACCESS MEMORY

Phase Change Random Access Memory (PCRAM) is a non-volatile memory technology that utilizes the phase transition of chalcogenide materials to store data. Unlike conventional charge-based memories, PCRAM operates by switching between the amorphous (high resistance) and crystalline (low resistance) states of a phase-change material, enabling reliable data storage with high endurance and fast switching speeds [18], [19]. Due to its scalability and compatibility with existing semiconductor fabrication processes, PCRAM is considered a promising alternative to traditional Flash memory.

The working principle of PCRAM is based on the controlled heating and cooling of a chalcogenide material, typically GeSbTe (germanium-antimony-tellurium, GST) or AgInSbTe (silver-indium-antimony-tellurium, AIST) [29], [30]. When a high-intensity, short-duration electrical pulse is applied, the material melts and rapidly quenches into an amorphous state, increasing its resistance (logic '0'). Conversely, a lower-intensity, longer-duration pulse allows the material to recrystallize, restoring its low-resistance state (logic '1'). This reversible phase transition enables PCRAM to store and retain information efficiently while providing fast read and write operations. Additionally, Multi-Level Storage (MLC) can be implemented by carefully controlling the intermediate resistance states of PCRAMs, further increasing the memory density.

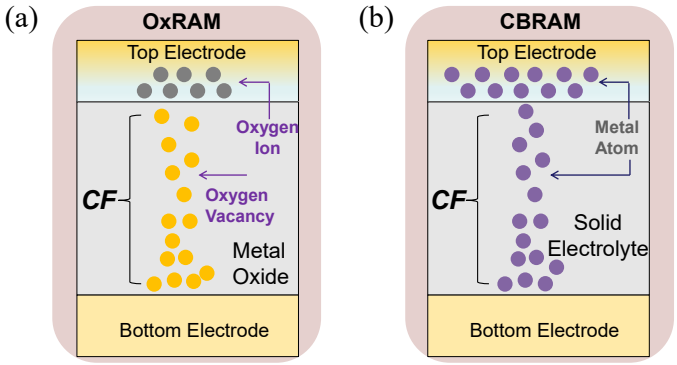


Figure 1.6: Structure of RRAMs. (a) OxRAM, (b) CBRAM.

Fig. 1.5 presents the confined and planar device structure of PCRAMs [31]. A typical PCRAM device consists of a phase-change layer, a heating electrode, and a bottom electrode. The heating electrode, usually made of titanium nitride (TiN) or other high-resistance materials, delivers precise electrical pulses to induce phase transitions.

PCRAM offers several advantages over conventional non-volatile memory technologies [18], [19]. It provides faster write speeds compared to Flash, high endurance, and better scalability due to its simple structure and material properties. However, it also faces some challenges, such as high power consumption during write operations, limited retention at high temperatures, and potential material degradation over repeated cycles. Researchers are actively working on improving phase-change materials and optimizing programming techniques to mitigate these drawbacks.

Looking ahead, PCRAM is expected to play a crucial role in future memory architectures, particularly in applications requiring high-speed, high-endurance, and non-volatile storage. Its potential integration into Storage-Class Memory (SCM), AI accelerators, and neuromorphic computing systems highlights its growing importance in next-generation computing. With continuous advancements in material science and device engineering, PCRAM could emerge as a viable replacement or complement to Flash memory, bridging the gap between volatile and non-volatile memory solutions.

RESISTIVE RANDOM ACCESS MEMORY

In Fig. 1.6, two RRAM types are shown. For example, two electrodes are connected by a filament in RRAMs. The resistance of devices is determined by the form of the filament, e.g., its width and length. By applying voltages to the electrodes, the filament's form may be modified. RRAM can be divided into two categories: Oxide Random-Access Memory (OxRAM) and Conductive-Bridge Random-Access Memory (CBRAM) because of the different CF types. The distinction between OxRAM and CBRAM is that OxRAM's CF is composed of oxygen vacancies in the metal oxide, while CBRAM's CF is built of metal atoms in the solid electrolyte [32]. Fig. 1.7(a) outlines the median initial Memory Window (MW) and endurance lifespan of OxRAM and CBRAM stacks. While CBRAM offers a considerably broader memory window, it comes at the cost of reduced endurance and higher switching energy requirements. Fig. 1.7(b) provides a clear benchmark of the re-

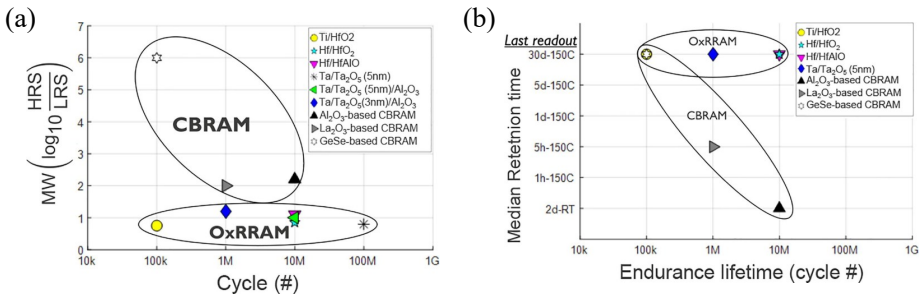


Figure 1.7: CBRAM vs. OxRAM. (a) Endurance and MW, (b) Endurance and retention.

Table 1.1: The benchmark of OxRAM vs CBRAM [35].

Performance	CBRAM Al_2O_3	CBRAM La_2O_3	CBRAM GeSe	OxRAM $Ta/Ta_2O_5/Al_2O_3$
Write time	5 μs	1 μs	1 ms	100 ns
Endurance	10^7 cyc.	10^6 cyc.	10^5 cyc.	$> 10^7$ cyc.
Retention	2 days @RT	5 hrs @150°C	5 days @150°C	>30 days @150°C
Initial MW	~100	~100	~ 10^6	~10
MW after 10^5 cycles	~50	~100	~2	~10

liability performance of OxRAM and CBRAM. A trade-off between endurance and retention affects CBRAM. The varying mobility of Cu atoms and ions within the switching layers is the cause of this phenomenon [33]. Short pulses may effectively program the device and sustain multiple switching cycles due to a high mobility of Cu^+. However, the rapid diffusion of Cu atoms within the solid electrolyte can lead to the degradation of the low resistance, ultimately causing retention failure [34]. Generally, OxRAM does not show the same trade-off as CBRAM, or if it does, the balance is altered in favor of a much longer retention time. Table 1.1 lists a comparison between OxRRAM and CBRAM according to a number of operating factors, which indicates the OxRAM devices' reduced switching voltage and better dependability [35].

Table 1.2 concludes a comparison of existing conventional and emerging memories. As depicted in the table, STT-MRAM and PCRAM occupy less space compared to SRAM. While STT-MRAM stands out with its low programming voltage, high endurance, and fast read/write speeds, PCM suffers from a notable drawback of high write latency. The FeRAM exhibits large area overhead and power consumption among emerging memory technologies. On the other hand, RRAM emerges as a promising alternative to Flash memory due to its lower programming voltage and superior write/read speed. Among the various next-generation memory technologies, RRAM presents a highly competitive solution for future digital storage applications. Its advantages include a straightforward fabrication process, a simple structure, exceptional scalability, nanosecond-level speed, prolonged data retention, and seamless integration with existing CMOS technology [37].

Table 1.2: Comparison of conventional and emerging memories [10], [36].

Memory Technology	SRAM	DRAM	NAND Flash	MRAM	FeFET	FeRAM	PCRAM	RRAM
Cell area	$>100F^2$	$6F^2$	$<4F^2$(3D)	$6{-}20F^2$	$\sim 0.05\,\mu m^2$	$10{-}20F^2$	$4{-}20F^2$	$<4F^2$(3D)
Cell element	6T	1T1C	1T	1(2)T1R	1T	1T1C	1T(D)1R	1T(D)1R
Voltage	$<1V$	$<1V$	$<10V$	$<2V$	3–5V	$<3V$	$<3V$	$<3V$
Read time	$\sim 1ns$	$\sim 10ns$	$\sim 10\,\mu s$	$<10ns$	$<200ns$	$<50ns$	$<10ns$	$<10ns$
Write time	$\sim 1ns$	$\sim 10ns$	$100\mu s{-}1ms$	$<5ns$	$10ns{-}1\mu s$	$<50ns$	$\sim 50ns$	$<10ns$
Write energy (J/bit)	$\sim fJ$	$\sim 10fJ$	$\sim 10fJ$	$\sim 0.1pJ$	~pJ	$\sim 10pJ$	$\sim 10pJ$	$\sim 0.1pJ$
Retention	N/A	$\sim 64ms$	$>10y$	$>10y$	$>10y$	$>10y$	$>10y$	$>10y$
Endurance	$>10^{16}$	$>10^{16}$	10^4	$>10^{15}$	$>5\times10^{10}$	10^{11}	10^9	$\sim 10^6{-}10^{12}$
Multibit capacity	No	No	Yes	Yes	Yes	No	Yes	Yes
Non-volatility	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Scalability	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

1.2. STATE OF THE ART IN MEMORY TESTING

Testing is a necessary means to verify the functionality and reliability of the chip. Generally, VLSI chips go through three stages from design to application delivery: the initial period, where the VLSI chips are manufactured; the time of system integration, in which the fabricated chips are loaded into an electronic system like a smartphone; and then the last and the longest period of purchase and use by end users. Because memory includes millions or even billions of states, memory testing is more challenging than logic testing. It is difficult to examine every situation that a memory chip might have. Consequently, specific testing methods must be created and used. In this section, the significant developments in conventional memory testing over the last several decades are first reviewed. After that, the state of the art in RRAM testing is examined.

1.2.1. CONVENTIONAL MEMORY TESTING

The development of memory testing has been a lengthy process. Because there are no dedicated fault models or improvements for the early memory tests (before the 1980s), they may be categorized as ad-hoc tests [38]. They take a long time to test and have a poor defect coverage, usually in the order of $O(n^2)$, where n is a count of memory chips. For example, Zero-One test, GALPAT test, and Walking 1/0 test in [38], [39] are divided into ah-hoc tests.

The emphasis of test development changed to examining potential memory defects in order to decrease test time and cost per memory chip with the exponential rise in memory capacity. This led to the introduction of many functional fault models in the early 1980s. The benefit of these models is that, although the test duration is often in the order $O(n)$: linear with the size of memory, the fault coverage of a particular test can be proven. Furthermore, address-decoder faults and stuck-at faults were two significant fault models that were introduced during that period [39], [40]. These are abstract fault models that are not grounded on any real defects or memory architecture.

Many identified defects could not be explained by the well-known fault models, according to experimental findings based on Defective Part Per Billion (DPPB) screening of many tests done on numerous memory chips in the late 1990s [41], [42]. This showed the presence of other defects and led to the development of unique static and dynamic fault models, such as read destructive faults, write disturb faults, transition coupling faults, and read destructive coupling faults, based on linear resistor defect injection and SPICE

simulation [43], [44].

In essence, the traditional linear-resistor-based test technique and the contemporary Cell-Aware Test (CAT) approach [45], [46] are quite similar. Defects inside library cells are becoming increasingly common as CMOS technology advances to more complex nodes. Conventional tests that target faults at the library cell ports were shown to identify fewer than 50% of these cell-internal problems [45]. In order to further lower the test escape rate, this has prompted the creation of CAT, which specifically targets cell-internal defects, within the last ten years. In the semiconductor sector, CAT has shown to be useful for newer technological nodes, such as 45 nm, 32 nm, and even 14 nm based on FinFET technology [46]. In early 2025, TSMC is expecting to produce 2 nm FinFET process technology in large quantities as the technology continues to downscale. Nonetheless, it is commonly acknowledged that fabrication-related defects and variations in device characteristics, as well as their effects on the system's overall quality and dependability, pose significant difficulties, particularly when high-quality levels are taken into account, such as those found in the range of DPPB [47]. The efficiency of CAT in identifying device-internal defects is still up for debate since it mainly targets resistive defects (e.g., opens and bridges) at the terminals and interconnects of devices.

1.2.2. RRAM TESTING

The test of three stages plays a big role in ensuring the chip quality, and the efficient test methods deserve to be studied. Memory testing is especially complex owing to the millions of states of memories. For conventional memories (e.g., SRAM), there are only three states: '0', 'U' (the undefined state), and '1' [43]. However, RRAM is an emerging memory technology that makes use of a resistive storage element. As a result, the continuous resistance range is separated into five states [48]. The evolving process technologies result in a large number of devices outside the specification: extremely low ('L') and extremely high ('H') conductance states. Here, 'L' and 'H' stand for resistance values of the RRAM device higher than a High-Resistive State (HRS) and lower than a Low-Resistive State (LRS), respectively. Besides, from the measurement data of RRAM, 'L' and 'H' need to be considered in RRAM states: $F \in \{H, 1, U, 0, L\}$ [49], [50]. The new fault will occur due to more complex resistance states and emerging manufacturing processes.

The RRAM testing approach consists of the study of defects, faults, and testing. Defects, faults, and errors are defined as follows by Bushnell and Agrawal [51]. An unintentional departure from the design in the actual chip that is fabricated, such as an open circuit existing in metal wires, is called a defect. This defect may cause the circuit to operate incorrectly (referred to as an error); e.g., a counter may provide the incorrect result for certain numbers. A fault serves as a high-level functional representation (abstraction) of a defect, such as a 'stuck-at fault'. Modeling the defect and comprehending the sensitized faults are necessary for defect detection. Once this understanding is established, a targeted testing approach can be created to detect these faults. In conventional memory, defects are typically modeled as linear resistors either within an interconnect or bridging two nodes [43]. Based on these defect models, sensitized faults are identified, and corresponding test procedures are designed to detect them. Researchers have taken a similar tack with RRAM. They verified the presence of conventional memory defects in RRAM and discovered distinct RRAM faults. Tests are then provided to detect these

faults.

Many works have investigated defect modeling and test generation for RRAMs. The majority of defect modeling is represented as linear resistors in (and at) the terminals of RRAM devices, i.e., bridges, shorts, and opens [52]–[56]. Several RRAM fault models such as stuck-at fault, deep faults, and coupling faults are derived and identified in these works [52], [54], [55]. In 2012, Haron and Hamdioui proposed a special Design-for-Testability (DfT) to detect undefined state faults [57]. In 2013, sneak-path testing for RRAMs was presented to reduce the test time [53]. In 2015, Chen and Li reported a dynamic write disturbance fault, and a March test to cover it by simulating in the netlist, and a March test was proposed to cover this fault [54]. Next year, Lin, Chen, Li, *et al.* offered a test method for finding the boundary currents of RRAMs in the production test phase [55]. In 2021, Liu, You, Wu, *et al.* developed a scheme for 3D hybrid RRAM array [56]. Nevertheless, all these works employ an inappropriate *linear resistor* in the surrounding interconnections rather than in the RRAM itself, which fails to accurately describe the defects within the naturally *non-linear* RRAM device, leading to test escapes and yield loss. To address the limitations of the conventional RRAM test method, the 'Device-Aware Test (DAT)' approach was proposed [58]–[60]. Finally, DfT works such as On-Chip Sensor [61], DFT-HR-ET-NOR [62], and Parallel-Multi-Reference Read [63] are employed to detect those unique faults.

Despite the achievements of previous works in RRAM testing, this field still faces several challenges that require innovative solutions. For instance, the physical switching mechanism of RRAM is not yet fully understood, making accurate defect modeling difficult. In addition, the intrinsic variability of RRAM devices further complicates fault modeling and detection. Future research in RRAM testing is expected to focus on developing more precise fault models and more efficient test methodologies. Potential solutions include leveraging machine learning techniques for test optimization and enhancing DfT strategies to achieve higher defect coverage while minimizing test overhead. As RRAM moves closer to commercialization, high-coverage and cost-effective testing solutions will be crucial to ensure reliable and scalable manufacturing.

1.3. RESEARCH TOPICS

Based on previous sections, RRAM has become a promising candidate for future memory technologies. Instead of focusing on just addressing issues in peripheral circuits, these tests must include the intrinsic physical mechanisms of RRAM faults. To achieve this, the research process must follow three key stages: 1) complete defect space and accurate defect modeling, 2) accurate fault modeling, validation, and analysis, and 3) high-quality test design. Thereafter, a diagnosis is also required to enhance the RRAM yield learning based on previous steps. Next, we discuss the research issues related to each step.

1.3.1. COMPLETE DEFECT SPACE AND DEFECT MODELING

The secret to getting optimized testing is to establish a complete defect space with accurate modeling. Establishing a comprehensive defect space ensures that all potential defects are systematically considered, while an accurate defect model captures the faulty

behavior, guaranteeing a realistic and reliable fault modeling. First, an analysis of the RRAM manufacturing process is required in order to identify any possible defects. Measurements and characterizations of these defects must then be made to determine their physical roots and the circuit impact. Additionally, this process improves knowledge of RRAM devices and may possibly reveal previously undiscovered defects. Hence, the comprehensive analysis of defect models is a benefit for yield learning.

1.3.2. FAULT MODELING, VALIDATION, AND ANALYSIS

The accurate fault modeling is based on the complete defect space and modeling. A systematic and automatic methodology is expected to conduct a fault validation. First, a fault space including various types of faults should be defined, especially taking those unique faults that only exist in RRAMs into account. After that, realistic faults can be validated based on the accurate defect models.

1.3.3. HIGH-QUALITY TEST DESIGN

Upon identifying the defect that sensitizes specific faults in the RRAM, high-quality testing designs could be developed for their detection. Two optimizations are required for the target. First, the designed test solution should be as complete as possible, i.e., it should be able to detect all RRAM conventional and unique faults that are observed or proven to be realistically existing. Next, the test solution should be efficient in terms of running time and costs, such as area and energy. To achieve this objective, procedures must be processed that avoid the test escapes.

1.3.4. DIAGNOSIS OF RRAM DEFECTS

The above testing designs aim to enhance the defect/fault coverage. Based on that, the diagnosis of RRAM defects is a further advancement to identify the defect and enhance the yield learning for RRAM productions. Therefore, a systematic approach of diagnosis needs to be developed for test optimization.

1.4. CONTRIBUTIONS OF THE THESIS

The contributions of this thesis correspond to all research topics outlined in the above section.

1.4.1. IDENTIFICATION OF CONVENTIONAL AND UNIQUE DEFECTS

We conducted a comprehensive analysis of the RRAM production process to identify all potential defect sources. Through this investigation, we determined that defects may arise in the transistors, interconnects, and the RRAM devices themselves. For each defect type, we provide a detailed explanation of its origin and its potential impact on RRAM behavior. As a result, we compiled a complete defect list that defines the full RRAM defect space for the given manufacturing flow. This defect set serves as a foundational reference for subsequent defect modeling and reliability analysis.

1.4.2. TESTING FOR CONVENTIONAL DEFECTS

Conventional defects are referred to as interconnects and contacts. Although previous works contributed to an understanding of the RRAM faulty behavior in the presence of the defects, they restricted the analysis to only faults involving one cell (i.e., victim-cell) or at most two cells (two-cell coupling faults). The potential impact of neighboring cells on the victim cell (i.e., Data Background (DB)) was ignored. This is a worthy aspect to investigate given the fact that extra paths can take place in such memories during read operations; this current is strongly DB dependent and, if high enough, may lead to incorrect operations.

This thesis advances the state-of-the-art by providing a systematic defect analysis and fault modeling for all possible interconnect and contact defects in RRAMs, while incorporating the impact of the DBs. It demonstrates that such DBs have an impact and could cause unique 3-cell and 4-cell faults; these have to be taken into consideration when developing test solutions, otherwise they lead to tests with low coverage resulting in escapes. We use a systematic approach to develop an optimal test algorithm, which detects all sensitized faults by interconnect/contact defects, including the 3-cell and 4-cell faults. Furthermore, we validate state-of-the-art RRAM tests in simulation and demonstrate the superiority of our solution. This has been published in [64].

1.4.3. TESTING FOR UNIQUE DEFECTS

Unique defects are those existing only in RRAMs due to the specific device structure and working mechanism of RRAMs. To overcome the constraints of the traditional RRAM test method, the 'Device-Aware Test (DAT)' method was introduced [58]–[60]. We further developed the DAT approach to obtain advanced and realistic defect models. We identify, characterize, and investigate two unique defects: Ion Depletion (ID) and Over RESET (OR). We develop the DA defect model for the ID and OR defect, which incorporates the intermittent behavior, and calibrate it with measurements to accurately model the defects. Furthermore, we utilize the DA defect model to develop appropriate fault models and subsequently optimal test solutions, validated through simulations. The proposed solutions are optimized for robustness against process variations, ensuring practical applicability and reliability. These have been published in [65], [66].

1.4.4. TEST ALGORITHM GENERATION AND DESIGN-FOR TESTABILITY

To address the challenges in RRAM faults that are observed from conventional and unique defects, we present test algorithms and a new DfT scheme for RRAM memories. We adopt a systematic methodology to develop an optimized test algorithm capable of detecting all sensitizable faults caused by interconnect and contact defects in RRAM arrays. The proposed algorithm is designed to maximize fault coverage while minimizing test cost and complexity. The DfT circuit replaces standard write drivers. It monitors the write current and compares it against multiple references simultaneously. The scheme can be used for manufacturing tests, diagnosis, yield learning, and even for in-field testing; it can detect defects that manifest themselves as a deviation in the write current. The DfT is adjustable such that e.g., the impact of process variability can be minimized. We implement and validate the DfT under process variations, and show that it outperforms the prior work. This has been published in [67].

1.4.5. RELIABILITY ANALYSIS AND DESIGN-FOR RELIABILITY

RRAM devices suffer from non-idealities that may cause functional errors during the deployment [68]–[70]. Such non-idealities consist of time zero and time dependent ones [71]. Time-zero non-idealities consist of device variation and wire parasitics, whereas time-dependent ones consist of endurance issues, device degradation, and resistance drift. Even if the RRAM chip passes the manufacturing test, it may suffer from in-field functional errors due to these non-idealities. Hence, it is essential to develop effective migration and recovery solutions to ensure the quality and reliability of RRAMs used as storage or as Computation in Memory (CIM) devices. We propose several approaches to address the read disturb failure, i.e., the HRS drift upon consecutive read cycles. We utilize the bi-directional read scheme, especially the read in the RESET direction to alleviate the resistance drift. The approaches can be developed for manufacturing tests, yield learning, and even for in-field detection/recovery; they can monitor the fault behavior in each read cycle. This has been published in [72].

1.4.6. DIAGNOSIS FOR RRAM YIELD LEARNING

A dedicated method of diagnosing unique defects is required to enhance the quality of RRAM fabrication further. We present a novel framework called Device-Aware Diagnosis (DA-Diagnosis), which goes one step further than DAT by utilizing the defect models obtained from DAT; the framework applies a systematic method to analyze physical manufacturing defects and extract their distinctive features for RRAM diagnosis. Thanks to the unique defect modeling of the DA method, unique and reliable fault signatures for each defect are obtained. DA-Diagnosis surpasses the limitation of existing memory diagnosis methodologies as it is a comprehensive approach that is based on defect origins and covers specific faults. This has been published in [73].

1.5. THESIS ORGANIZATION

The previously described contributions enhancing the state of the art in RRAM testing will be discussed in depth in the subsequent chapters of this thesis, which is structured as follows.

Chapter 2 establishes the basic background of RRAMs. First, the fundamental construction and switching concepts of an RRAM device are described. Subsequently, we provide an overview of RRAM device modeling and compare existing compact models. Then, we introduce the RRAM cell designs used in this thesis, following the peripheral circuits and the complete RRAM array.

Chapter 3 explains the manufacturing procedure, potential defects, and reliability issues of RRAMs. First, we overview the RRAM fabrication steps. Based on that, an analysis of potential defects and defect space is given. A part of these defects will be modeled later in this thesis. Next, the information of the RRAM structure and non idealities is used to explain the reliability issues.

Chapter 4 presents a systematic and automatic approach to test conventional defects in RRAMs. First, we define the complete interconnect and contact defect space. Then, we present the simulation methodology and analyze the resulting faults. Next, we propose a test solution, which is validated together with existing tests. A detailed com-

parison is then carried out to highlight the improvements.

Chapter 5 applies the DAT approach for ion depletion defects. We first establish experimental evidence, characterization, and underlying potential causes of the ID defect. Next, we illustrate the failure of the conventional defect models (linear resistors) and propose the Device-aware defect model that properly depicts the faulty behavior of a defective RRAM device. Based on the obtained model, we apply it to perform fault modeling and analysis. Finally, we develop the test solutions for the ID defect.

Chapter 6 analyzes another unique defect; i.e., over RESET. We apply DAT approach in a similar way as we did for ID defect.

Chapter 7 presents a DfT scheme. We first classify the targeted defects and faults in RRAM arrays. Subsequently, we discuss the limitations of existing test solutions. Finally, we validate DfT in Spectre simulators and provide the way to implement it.

Chapter 8 covers a design for reliability. First, we analyze one of the major reliability RRAM issues being read disturb failure, and provide measurements and simulations. Then, we present a bi-directional read mitigation scheme. Finally, we validate the approach and the impact on both RRAM as storage memory and CIM inference engine.

Chapter 9 presents the framework of the DA-diagnosis for conventional and unique defects in RRAMs, and applies it to targeted defects.

Chapter 10 puts a systematic approach for testing an industrial RRAM; it is compiled based on all findings generated in this thesis. First, we classify the completed defect space in RRAM arrays and peripherals. Then, we discuss the existence of conventional and unique faults in RRAMs. After that, overviews and evaluations of March algorithms and DfTs for RRAMs are provided. Finally, we present a case study that optimizes an ultimate design for industrial RRAM testing.

Chapter 11 concludes with final remarks and an outlook on potential future research directions.

2

BACKGROUND

This chapter establishes a background of RRAM technology and is organized into five parts. First, the fundamentals of RRAM device technology, including its principal structure, placement, switching mechanism, and stacked composition. Then, it reviews RRAM device models, with emphasis on multiple modeling levels and widely used compact models, several of which are employed in this thesis. Next, different design styles of RRAM cells are discussed. Then, it presents the architecture of RRAM systems, including a complete array that can be simulated in Spectre, as well as the peripheral circuits such as drivers, decoders, and sense amplifiers. Finally, the chapter reports on some existing prototypes and shows the potential of RRAM for commercialization and for serving different applications.

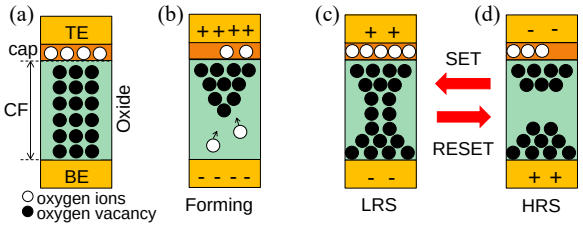


Figure 2.1: Evolution of the conductive filament. (a) RRAM stack, (b) Forming, (c) SET, (d) RESET.

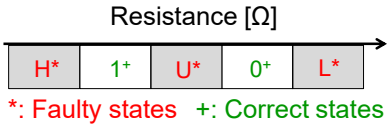


Figure 2.2: RRAM resistance states and corresponding logic values.

2.1. RRAM DEVICE TECHNOLOGY

This section explains the RRAM device technology. This section provides an overview of RRAM device technology, with a particular focus on OxRAM (also see Section 1.1.2.5), the subject of this thesis. We first introduce the general structure (placement) and switching mechanism of RRAM devices. Then, we highlight the advantages of OxRAM over other RRAM types and explain how resistance states determine logic levels.

2.1.1. OVERVIEW

An RRAM device is a Metal-Insulator-Metal (MIM) construction, as schematically shown in Fig. 2.1(a). In its stack organization, a middle metallic oxide (commonly TiO_x, HfO_x) is fabricated with an additional capping (cap) layer (commonly Hf, Ti, Ta), between two metal electrodes: the Top Electrode (TE), and the Bottom Electrode (BE) [74], [75]. Applying a forming voltage, localized oxygen deficiency leads to the formation of Conducting Filament (CF) between two electrodes. The forming process is regarded as soft-breakdown of insulator in which a reversible conducting path is formed in the dielectric. Its resistance state is determined by the length of CF in the middle layer. For instance, a longer and a shorter CF correspond to the low resistive state (logic '1') and high resistive state (logic '0'), respectively.

RAM devices can be categorized into different types based on their switching mechanism. CBRAM and OxRAM share similar electrical properties; however, OxRAM is preferred in this thesis due to its superior endurance and simpler fabrication process [10], [32], [35], [76]. Throughout this thesis, we will focus on OxRAM-based RRAM.

By applying specific programming voltages to an RRAM device, its resistance can be switched between different states. The (binary) RRAM is divided into 5 states as the resistance increases (see Fig. 2.2) [49], [50], [77]: 1) the extremely high conductance faulty state 'H', 2) the correct Low Resistive State (LRS) '1', 3) the undefined faulty state, also known as the intermediate range between the LRS and HRS, is represented by the state 'U' [77], 4) the correct High Resistive State (HRS) '0', and 5) the extremely low conduc-

tance faulty state 'L'. The CF length will increase due to the production of more Oxygen Vacancies (OV). Conversely, some oxygen ions migrate back from the capping layer, fill the vacancies, and shorten the CF. A general illustration describing the switching process of bipolar OxRAM is shown in Fig. 2.1.

2.1.2. PLACEMENT STRUCTURE

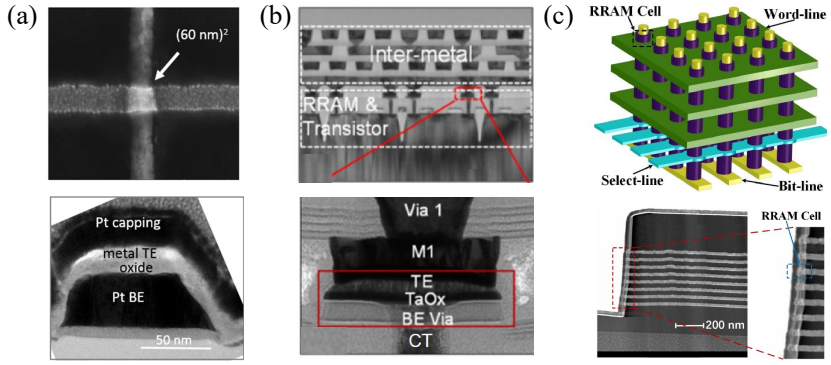


Figure 2.3: RRAM device placement structure. (a) Cross structure (reprinted from [78]), (b) Between metal layers (reprinted from [79]), (c) 3D vertical RRAM structure (reprinted from [80]).

RRAM devices are typically integrated above the transistor layer on the wafer. There are three possible configurations for their placement, as illustrated in Fig. 2.3: 1) a crossbar structure (Fig. 2.3(a) [78]), 2) integration between metal layers (Fig. 2.3(b) [79]), and 3) implementation within a dedicated 3D architecture (Fig. 2.3(c) [80]). Next, we provide a detailed discussion of each configuration.

CROSSBAR STRUCTURE

The crossbar structure is commonly employed in small-scale RRAM devices used for research purposes. Compared to RRAM integrated between metal layers or within dedicated 3D architectures, its fabrication process is relatively straightforward. These devices are primarily utilized for exploring new materials [81] or investigating specific switching behaviors [78], [82]. The crossbar design enables direct pad connections to the top and bottom electrodes, facilitating easy access for measurement probes and experimental characterization.

BETWEEN METAL LAYERS

Integrating RRAM devices directly between metal layers enables higher density compared to the crossbar structure. Additionally, this approach simplifies the integration of individual RRAM cells with transistors on the underlying wafer, facilitating the implementation of 1T-1R configurations. As a result, this placement method enhances both reliability and scalability, making it well-suited for high-density RRAM architectures [79], [83].

DEDICATED 3D STRUCTURE

The dedicated 3D architecture enables even higher memory density than RRAM devices integrated between metal layers. However, its fabrication process differs significantly from standard CMOS manufacturing [80], [84], introducing additional complexities that make production more challenging.

2.1.3. SWITCHING MECHANISM

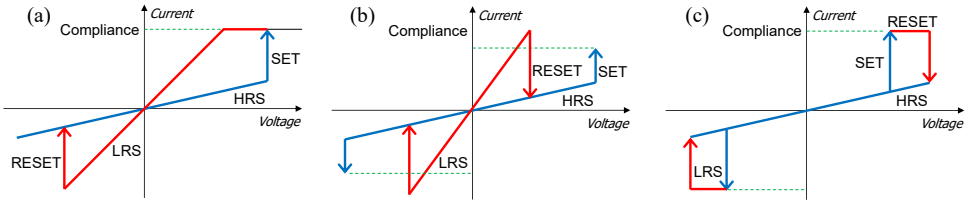


Figure 2.4: RRAM switching modes. (a) Bipolar switching, (b) Unipolar switching, (c) Complementary switching.

For OxRAM, there are three different modes for switching between high and low resistance states: unipolar, bipolar, and complementary [85]–[87]. Fig. 2.4 shows the simplified I-V electrical curve during the switching process. The defining characteristic of RRAM devices is their Resistive Switching (RS) behavior. Various mechanisms have been reported to explain the transition between HRS and LRS. These include the modulation of the Schottky barrier [88], Poole–Frenkel (P–F) emission [89], [90], Space Charge Limited Conduction (SCLC) [85], [91]–[95], charge trapping and detrapping [96], as well as the formation and rupture of CFs [97], among others. The Bipolar Switching (BS) mechanism is dependent on the production and dissolution of the CF as a result of *oxygen ion* migration [74], [85]. RRAM devices typically necessitate a forming process, which involves the application of a high positive voltage (V_{forming}) between the TE and BE to dissociate a portion of the metal-oxygen ionic bonds [85]. The positive anode attracts negatively charged oxygen ions (O^{2-}) from the lattice, which accumulate at the cap/oxide interface. A chain of positive charged OV known as *Conductive Filament (CF)* is generated in the insulator between two electrodes as a consequence of this localized deficiency, as illustrated in Fig. 2.1(b). The direction of switching is dependent on the polarity and amplitude of the applied voltage (V_{SET} and V_{RESET}) [85]. Resistance can be altered by applying specific programming voltages to an RRAM device, which allows it to transition between various states. The CF length will increase as the positive voltage from TE to BE (V_{TE}) exceeds the specified threshold ($V_{\text{TE}} \geq V_{\text{SET}}$) due to the generation of additional OV, as illustrated in Fig. 2.1(c) [98], [99]. This operation is referred to as a SET operation. The resistance in SET is denoted as R_{SET}. In contrast, a RESET operation begins when the negative voltage from TE to BE is less than the RESET threshold ($V_{\text{TE}} \leq V_{\text{RESET}}$), causing a small amount of O^{2-} to drift back into the bulk oxide and rupture the CF, as illustrated in Fig. 2.1(d) [98], [99]. The two switching mechanisms described above can be represented by the following chemical reaction equation:

$$O^{2-} \underset{RESET}{\overset{SET}{\rightleftharpoons}} OV^{2+} + \frac{1}{2} O_2 \uparrow .$$

In unipolar switching, the set and reset voltages have the same polarity but different amplitudes (V_{SET}>V_{RESET}) [100]. In complementary switching, the set and reset voltages have the same polarity as those in unipolar switching, but their magnitudes are inverted (V_{SET}<V_{RESET}). In this thesis, we will focus on the most common, bipolar switching devices [87].

2.1.4. STACKED COMPOSITION OF RRAM

An example of an RRAM device embedded between metal layers is illustrated in Fig. 2.3(b). The composition of the material stack plays a crucial role in defining the device's switching characteristics, determining whether it operates in bipolar, complementary, or unipolar mode. Next, we provide a detailed examination of the oxide layer, electrodes, and capping layer, highlighting their structural properties in relation to the bipolar switching mechanism that is the target of this thesis.

Bipolar switching in RRAM requires an asymmetric oxygen vacancy distribution, meaning one side of the device must have a stronger ability to attract oxygen vacancies than the other [87], [101], [102]. This can be achieved either by using a combination of a noble metal and a reactive metal as electrodes or by incorporating a capping layer [87], [103]. In the first approach, the reactive metal acts as an oxygen ion trap, whereas in the second, the *capping layer* serves this function. Fig. 2.5 provides an overview of metal materials used in the oxide layer as binary oxides (highlighted in yellow) and those serving as electrodes in the resistive switching process (highlighted in blue).

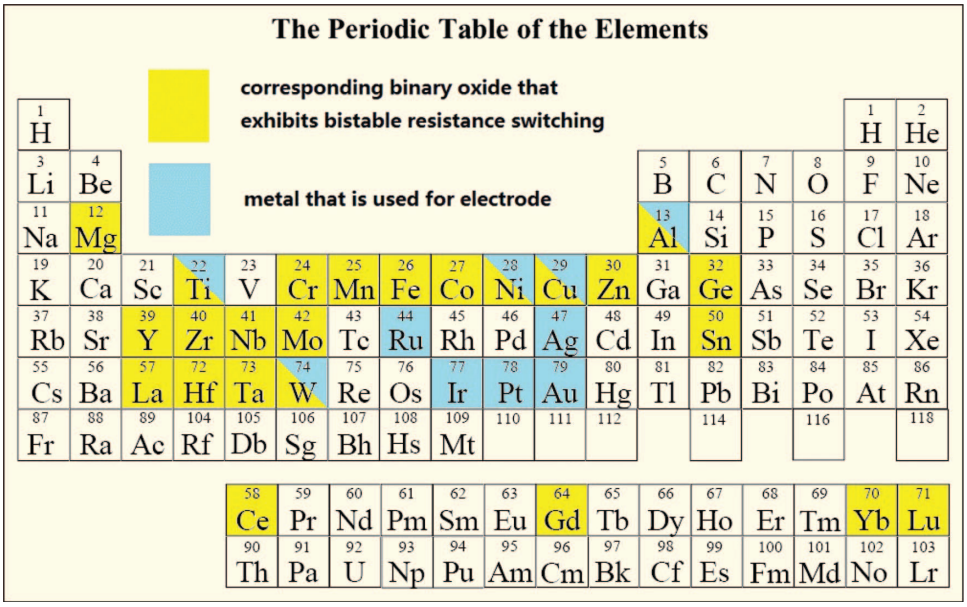


Figure 2.5: The periodic table of elements with host metals of RRAM (reprinted in [85]).

Typical examples of RRAM structures utilizing noble and reactive metal electrodes include (BE/oxide/TE): Pt/HfO_2/TiN [86] and Pt/TiO_2/TiN [104]. Meanwhile, struc-

tures that integrate a capping layer include (BE/oxide/capping layer/TE): Pt/HfO_2/Hf/Pt [105], TiN/HfO_2/Ti/TiN [106], and TiN/HfO_2/Hf/TiN [107].

2.2. RRAM DEVICE MODELS

This section overviews and compares different models that can simulate the RRAM behavior. We first present multiple levels of RRAM models. Then, we focus on compact models and provide an overview of a comparison among existing compact models.

2.2.1. MODELING LEVELS

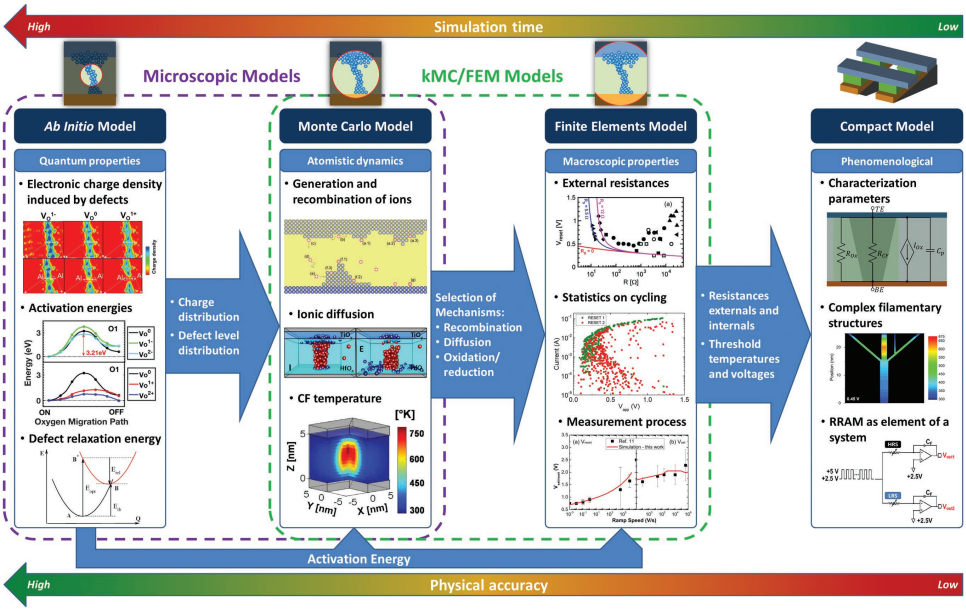


Figure 2.6: Multiple levels of RRAM models based on [108].

Fig. 2.6 depicts an overview of RRAM models that are grouped into four types according to their level of accuracy and simulation complexity [108]. Generally, increased modeling accuracy necessitates extended simulation duration. The classifications of models include: 1) microscopic models, 2) finite element models (kinetic Monte Carlo), and 3) compact models. These will be addressed thereafter.

MICROSCOPIC MODELS

This kind of model encompasses both ab initio and Monte Carlo approaches, both of which characterize the interactions among individual particles within an RRAM device. These interactions fundamentally dictate the device's electrical properties. For instance, such models facilitate predictions regarding the evolution and morphology of CFs across various materials. Ab initio models offer the highest level of physical accuracy as they capture quantum-level particle interactions. In contrast, Monte Carlo models, while

slightly less precise due to certain interaction simplifications, can incorporate a greater number of elements. By enabling the examination of even minute structural variations in RRAM devices, microscopic models provide valuable insights. However, their computational complexity leads to prolonged simulation runtimes, making them impractical for circuit-level simulations.

KINETIC MONTE CARLO OR FINITE ELEMENT MODELS

Kinetic Monte Carlo and finite element models offer a more direct approach to simulating the electrical behavior of RRAM devices while incorporating material properties. By simplifying and generalizing certain quantum interactions from microscopic models, these approaches significantly enhance computational efficiency. Kinetic Monte Carlo models account for the influence of individual particles, whereas finite element models employ mathematical approximations, such as differential equations, to represent the electric field within the device. A key advantage of kinetic Monte Carlo models is their ability to capture variation effects, whereas finite element models provide faster simulations. Although both methods trade some physical accuracy for computational speed compared to microscopic models, their simulation time remains a limiting factor, making them impractical for large-scale circuit simulations.

COMPACT MODELS

Compact models also characterize the electrical behavior of RRAM devices but further abstract the underlying physics to enhance simulation efficiency. These models typically assume the existence of a CF with a predefined shape that consistently evolves in response to similar voltage applications. By making such generalizations, compact models achieve faster computation, making them particularly suitable for circuit-level simulations. Due to their efficiency, these models are widely employed in circuit-level simulations. In this thesis, the focus remains on compact models, as they enable the circuit-level analyses necessary for test development.

2.2.2. OVERVIEW OF COMPACT MODELS

In [109], Hajri, Aziza, Mansour, *et al.* evaluate eight different RRAM compact models (BS type) across various criteria. Table 2.1 presents the findings from their study while incorporating an additional RRAM compact model, introduced later in [110], which was not available at the time of their publication. It is important to highlight that only the metrics pertinent to this thesis are considered in the comparison. The following provides a description of the metrics.

- Genericity: Whether the model is tailored for various memristor technologies, such as supply voltages and materials.
- Non-linearity: Whether the model equations incorporate this behavior and manifest it in the I-V characteristics. This nonlinearity is believed to stem from the nonlinear movement of ions, which is intensified by Joule heating [120].
- Hard SET/Soft RESET: Whether the model is able to fit with SET/RESET time difference, e.g., if the time during RESET is longer than the time during SET, it indicates a hard SET [121].

Table 2.1: Evaluation of various RRAM compact models based on [109].

Model	Linear [111]	Non-linear [112]	Simmons [113]	TEAM [114]	VTEAM [115]	SPICE [116]	IM2NP [117]	Stanford [118]	JART v1b [110]	JART var [119]
Genericity	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Non-linearity	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hard SET Soft RESET	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Forming	No	No	No	No	No	No	No	No	No	No
High frequency	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Switching thresholds	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pulse height dependence	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Pulse width dependence	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Temperature dependence	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Variability	No	No	No	No	No	No	Yes	Yes	No	Yes

- Forming: Whether the model description explicitly includes a mathematical equation that defines the electroforming process.
- High frequency: Whether the model is able to accommodate a broad range of operating frequencies to enable the simulation of newly developed devices with extremely rapid switching characteristics.
- Switching threshold: Whether the model supports the existence of a threshold voltage at which hysteresis manifests solely when the applied voltage surpasses a specific voltage.
- Pulse height dependence: Whether the model switching is dependent on applied voltage amplitude (pulse height).
- Pulse width dependence: Whether the model switching is dependent on the pulse width of applied voltage.
- Temperature dependence: Whether the temperature affects the model.
- Variability: Whether the model includes D2D and C2C variability.

As can be seen from Table 2.1, the IM2NP model, the Stanford model, the JART v1b model, and the JART var model exhibit the best performance with the listed criteria. Note that the JART var model is extended from the JART v1b model. Next, we will explain these models separately.

IM2NP MODEL

Fig. 2.7(a) illustrates the IM2NP model presented in [117] that represents the CF as a cylindrical structure whose radius dynamically changes in response to SET and RESET operations. When a positive voltage is applied, the filament radius expands, transitioning the device into the LRS. Conversely, applying a negative voltage contracts the CF radius, shifting the device into the HRS. In LRS, the device conductance is primarily

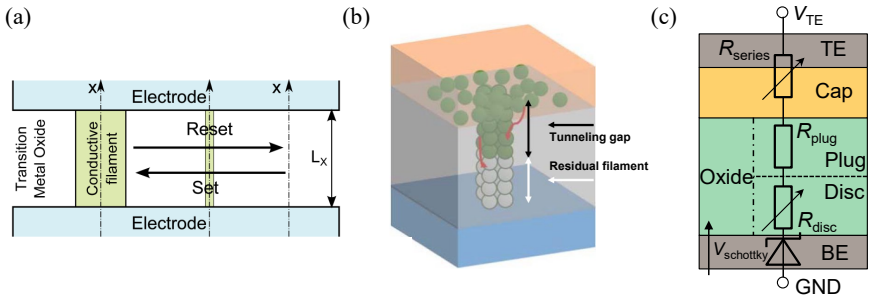


Figure 2.7: Basic illustration of three compact RRAM models (reprinted from [110], [117], [118]).

dictated by the ohmic resistance of the filament, whereas in HRS, the current flow is governed by leakage through the insulating oxide layer. At each simulation time step, the model computes the local temperature, the corresponding filament radius variation, and the resultant current passing through the device. To incorporate variability, stochastic elements are introduced into the filament radius parameters.

STANFORD MODEL

The Stanford model described in [118] represents the CF as a cylindrical structure responsible for current conduction, as illustrated in Fig. 2.7(b). During the SET process, the tunneling gap gradually narrows as the cylinder extends until full conduction is restored. Conversely, when a RESET operation is applied, the cylinder contracts, creating a tunneling gap that restricts current flow. At this point, the current is regulated by the dimensions of the filament. Throughout each simulation step, the model calculates the local temperature, variations in filament length, and the corresponding current flow. To account for variability, a random factor is incorporated into the rates of CF expansion and dissolution.

JART MODEL

The JART v1b model, introduced in [110], is designed to simulate the evolution of oxygen vacancies (OV) within the oxide layer of a Valence Change Mechanism (VCM)-based memory device. The equivalent circuit representation of this compact RRAM model is illustrated in Fig. 2.7(c), featuring a Schottky-like contact (V_{schottky}), two filament-related resistances (R_{plug} and R_{disc}), and a series resistance (R_{series}). In this model, the HfO_2 oxide layer is conceptually divided into two distinct zones: the disc region, where switching dynamics occur, and the plug region, which remains conductive and serves as an infinite reservoir of vacancies. The resistive switching mechanism is governed by the movement of OV during the SET and RESET operations, directly influencing the Schottky barrier and consequently altering the electrical conductivity of the device. To account for the concentration of oxygen vacancies in the disc region, the parameter N_{disc} is dynamically adjusted to capture the switching behavior and corresponding resistance changes in the RRAM model. The migration of vacancies is constrained within predefined boundaries, $N_{\text{disc,min}}$ and $N_{\text{disc,max}}$, ensuring that N_{disc} remains within this controlled range throughout the switching process. The JART var model [119] is extended from the v1b model but

includes both D2D and C2C variations by incorporating a random variation factor into the disc's radius, length, and N_{disc} constraints.

All the above RRAM models are physical-based and calibrated with measurements. In the rest of the thesis, we use the Stanford [118] and the JART [110], [119] models to model the RRAM device, calibrating with new silicon measurement data.

2.3. DESIGN STYLES OF RRAM CELLS

RRAM cells are generally not employed individually; instead, multiple cells are arranged together in array configurations for practical use as memory. Several common array configurations have been developed to facilitate this, including structures that integrate a resistor alone (1R), a diode coupled with a resistor (1D-1R), and a transistor combined with a resistor (1T-1R). These architectures provide various advantages depending on the specific application requirements and performance trade-offs.

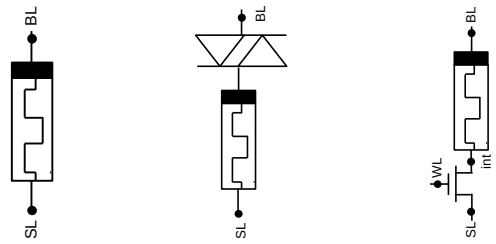


Figure 2.8: RRAM cell designs. (a) 1R, (b) 1D-1R, (c) 1T-1R.

The simplest cell structure is the 1R configuration, which integrates a single RRAM device per memory cell, as depicted in Fig. 2.8(a). Each cell is selected and accessed through two lines: the Bit Line (BL) and the Select Line (SL). One notable benefit of this configuration is its structural simplicity. However, it suffers from a significant drawback, known as sneak-path currents, as shown in Fig. 2.9. Sneak paths are unintended conduction routes within the memory array, allowing current to bypass the targeted cell. For example, when cell 1 is accessed via the BL and the SL (e.g., the corresponding BL is set to V_{DD} and the corresponding SL to GND to read the cell), the current $I_{primary}$ will flow through the cell to determine the state of the memristor. However, the sneak current (I_{sneak}) will also flow through the sneak paths formed by the other three cells. The total output current will be higher than $I_{primary}$. Thus, the sneak path restricts the read margin and may result in read errors. This increased overall current degrades the sensing accuracy and reduces the effective read margin, potentially causing read operation failures. The presence of sneak paths thus poses a major limitation on reliability for 1R memory architectures.

To mitigate the sneak path issue, a bidirectional diode can be placed within each memory cell at the cross-points of RRAM arrays, as can be seen in Fig. 2.8(b). This diode permits current flow in both directions selectively, thus minimizing undesired currents. However, the use of a bidirectional diode restricts device selection to unipolar RRAM

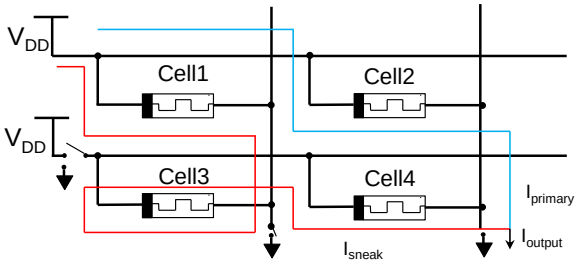


Figure 2.9: Sneak path in a 2 × 2 1R array.

technologies and necessitates higher operating voltages to effectively write data. Consequently, this solution increases the overall energy consumption of the memory cells and could pose substantial challenges, especially at advanced technology nodes where voltage scaling is critical.

Alternatively, the 1T-1R architecture integrates a transistor in series with each RRAM cell. This transistor serves as an access device, effectively eliminating sneak path currents by allowing selective access to each individual cell. Although the 1T-1R configuration significantly enhances control over bipolar RRAM operations, it compromises cell density due to the additional transistor requirement. The structure of the 1T-1R cell is demonstrated in Fig. 2.8(c). In this topology, a Word Line (WL) is employed to activate the transistor, granting selective access to the data stored in a particular row. Meanwhile, the BL connects directly to the TE of the RRAM, allowing read and write operations, while the transistor regulates the connection. The trade-off for this improvement in performance and reliability is a higher cell area overhead. Due to its effective management of sneak paths and improved reliability, the 1T-1R configuration has become the predominant choice for constructing practical and reliable RRAM arrays.

Because the 1T-1R design provides the best control, this cell design is the most common structure for the RRAM array and is adopted in this thesis.

2.4. RRAM System Architecture and Circuit Design

This section describes a typical RRAM architecture. First, an overall structure of the architecture is presented, followed by a detailed discussion of its components (peripheral circuits).

2.4.1. RRAM Array Overview

Fig. 2.10 illustrates an overview of an RRAM array architecture. The cell array is composed of four 1T-1R cells (see Fig. 2.8(c)), with each cell storing one bit of data. This particular design utilizes a single-port configuration, meaning it provides only one set of address inputs and control signals, allowing either a read or write operation at any given moment. Data storage in these cells is achieved by setting the device resistance to either a high (‘0’ state) or a low (‘1’ state) value, representing binary states. Accessing data within this structure is managed by external peripheral circuits. Specifically, cells aligned

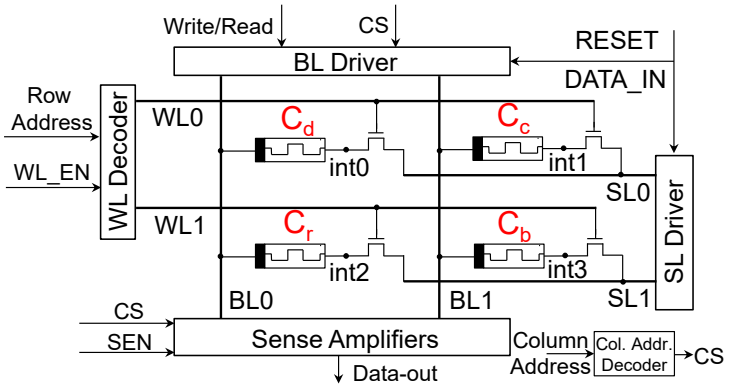


Figure 2.10: A 2 × 2 RRAM circuit architecture.

along a common row share the WL and SL, whereas cells within the same column share the BL. Peripheral circuitry includes essential components such as row and column Address Decoders (AD), a Write Driver (WD), and a Sense Amplifier (SA). Voltage pulses for addressing cells are generated and applied by these peripheral drivers. Fig. 2.10 explicitly demonstrates these necessary peripheral components, including WL, SL, and BL drivers, alongside the SA. During operation, the row decoder activates the appropriate WL, while the column decoder generates Column Select (CS) signals, enabling the read or write processes for targeted cells. Finally, the SA employed here functions similarly to those previously presented in the literature (e.g., [50]), effectively sensing and amplifying the stored resistance states of the selected RRAM cells.

2.4.2. PERIPHERAL CIRCUITS

Next, we explain the above peripheral circuits in detail, which include AD, SA, and WD. These circuits ensure effective execution of read and write processes.

ADDRESS DECODERS

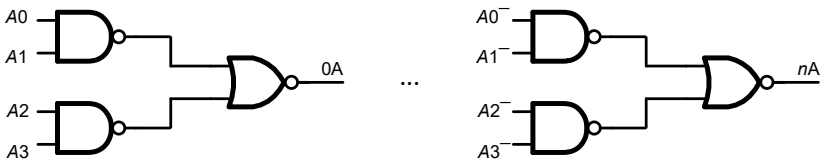


Figure 2.11: Address decoder.

The column address decoder and the WL decoder translate the provided addresses into their corresponding CS and WL signals. Fig. 2.11 illustrates an example of an address decoder, decoding a 4-bit address n, composed of bits $A0$, $A2$, $A3$, and $A4$, into an internal line designated as nA. For the column address decoder, this decoded line directly acts as the CS signal. For the WL decoder, the signal nA is combined with the enable signal WL_{EN} via an AND gate, generating the word line WL_n. Additionally, this decoded

signal is connected to the SL through the WL input line. The WL_{EN} signal ensures precise timing control over WL activation, preventing unintended selection of word lines during address transitions.

Write Drivers

The write drivers, which include the BL driver and the SL driver in RRAMs, are essential for managing the direction and magnitude of the write current, which determines the switching of the RRAM state from LRS to HRS and vice versa. Fig. 2.12 illustrates a typical configuration of a WD, comprising four transistors (Ctr_SET, $/Ctr\_SET$, Ctr_RESET, and $/Ctr\_RESET$). These transistors are sized relatively large to reduce the channel's resistance, thereby maximizing the voltage effectively applied across the MTJ and minimizing voltage losses within the driver circuit itself.

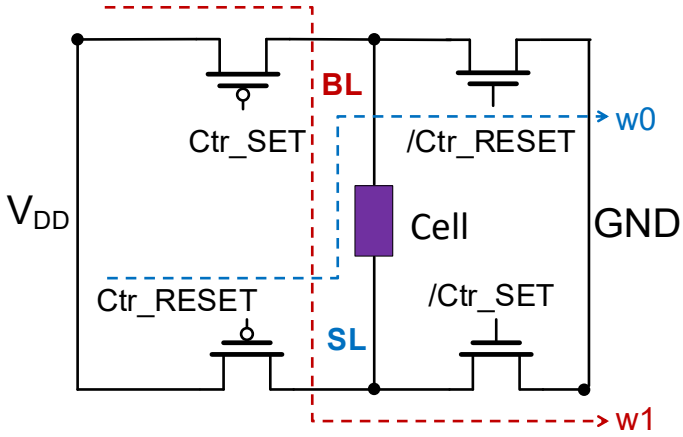


Figure 2.12: The schematic of circuit with write drivers and the currents of w1 and w0.

During the $w1$ operation, transistors Ctr_SET and $/Ctr\_SET$ are activated, while transistors Ctr_RESET and $/Ctr\_RESET$ remain deactivated. In this arrangement, a current path is created from BL to SL, causing electrons to flow from TE to BE. This is the SET operation to switch the RRAM, thus representing a stored bit of '1'. Conversely, for the $w0$ operation, transistors Ctr_RESET and $/Ctr\_RESET$ are activated, and Ctr_SET and $/Ctr\_SET$ are switched off. The direction of the current reverses, flowing from SL to BL, performing the RESET operation, representing a stored bit of '0'.

Transistor sizing in the WD is optimized to provide sufficient current amplitude for stable MTJ switching. Transistors with larger width-to-length (W/L) ratios are employed to decrease internal resistance, minimizing voltage losses within the transistor channels. This ensures an effective write voltage across the MTJ. Such optimization is crucial to balance between maintaining a reliable switching operation, reducing power consumption, and avoiding device degradation, as insufficient current can cause increased write errors and overly high current may negatively affect device lifetime.

In the memory array design, the write driver circuit is also controlled by inputs and address signals. Fig. 2.13(a) illustrates the design and operation of a BL driver, which is

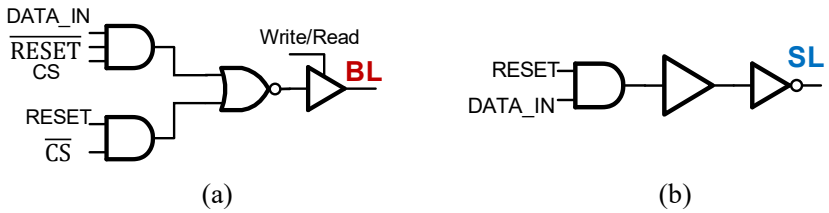


Figure 2.13: Write drivers. (a) The BL driver, (b) The SL driver.

applied in this thesis. During write mode, indicated by a low Write/Read signal, the tri-state buffer is activated, allowing data (DATA_IN) to be driven onto the BL. Conversely, in read mode, indicated by a high Write/Read signal, the tri-state buffer becomes inactive to prevent the BL driver from affecting the performing read operation. Similarly, Fig. 2.13(b) illustrates the operation of the SL Driver. Normally, it remains at a low logic level, except when both the RESET signal and the WL_In signal (controlled by the WL decoder) are simultaneously set to high.

SENSE AMPLIFIERS

The Sense Amplifier (SA) in RRAM detects the stored data by sensing the voltage difference caused by various resistance states. A regular voltage-based SA [122] architecture is depicted in Fig. 2.14(a). The SA requires two operating phases depending on the SEN (Sense Enable) control signal. When SEN is set to '0', the Q and $\overline{Q}$ nodes are pre-charged to VDD. When SEN signal is switched to '1', these pre-charged nodes start discharging. During a read operation, the SA determines the resistance of a cell by comparing it to a reference cell, achieved through the discharge of the bit line and Reference Line (RL) connected to the respective cell or reference cell. The reference cell is typically set in the middle of HRS and LRS. The values of discharged nodes Q and $\overline{Q}$ are defined as the read outputs, representing the stored data of the cell. Once SEN returns to low, nodes Q and $\overline{Q}$ are both restored to their pre-charge states.

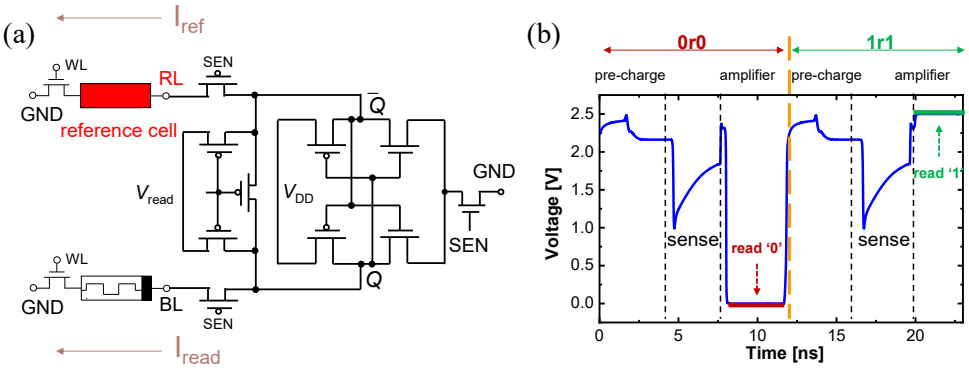


Figure 2.14: Sense amplifier design [122] and read waveform. (a) The SA design, (b) The waveform of 0r0 and 1r1 operations.

Next, we present the operation steps in detail. During sensing, the SA compares the current flowing through the memory cell against a reference path, measuring the difference between them. The SA has two output nodes, named as Q and its inverse $\overline{Q}$. Fig. 2.14(b) illustrates the voltage behavior of Q and $\overline{Q}$ during the 0r0 and 1r1 operations. At the start, the nodes Q and $\overline{Q}$ are both set to V_{read}. As the sensing operation progresses, the discharge rate of each node becomes dependent on the current in its corresponding branch, influenced by the resistance values of the RRAM cell and reference paths. Traditionally, the selected reference resistance falls between the RRAM resistance in the '0' and '1' states. During a 0r0 operation, the RRAM is in the '0' state, exhibiting higher resistance and, consequently, a lower read current ($I_{read} < I_{ref}$). This lower current slows down the discharge rate of the RRAM path relative to the reference path. Once the voltage difference $Q - \overline{Q}$ surpasses the MOSFET threshold voltage (V_{TH}), the node enters the recharging (amplifier) phase, and Q falls down toward V_{DD}, as illustrated by the amplifier phase and readout (the red line) in Fig. 2.14(b). In contrast, during a logical '1' read operation, the RRAM is in the LRS, which has lower resistance and consequently higher read current ($I_{read} > I_{ref}$). Under these conditions, the RRAM path discharges faster, causing the reference path to discharge comparatively slower, and allowing Q to steadily decrease toward GND, as illustrated in Fig. 2.14(b) (the green line).

To achieve faster sensing, a sufficiently large I_{read} is desirable to accelerate the charging and discharging processes at Q and $\overline{Q}$. However, I_{read} should not exceed a certain limit to prevent unintended device switching. Moreover, the charging requirement is determined by the MOSFET capacitances within the SA; larger MOS capacitances result in increased stored charge, prolonging the charge/discharge duration and thus decreasing read performance. Consequently, carefully balancing the SA circuit parameters and the I_{read} is essential for optimizing RRAM sensing and read performance.

REFERENCE CELL

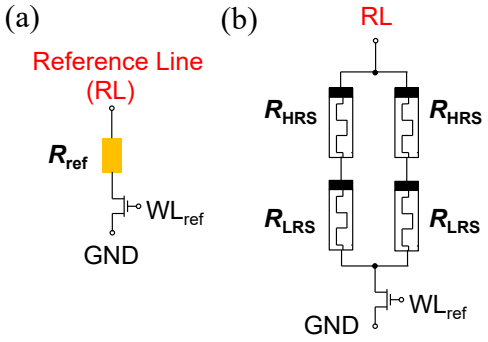


Figure 2.15: The design of reference cells. (a) Reference using a resistor, (b) Reference composed of RRAMs.

Fig. 2.15 illustrates two possible reference-cell designs suitable for use with the previously described sense amplifier. The first design, depicted in Fig. 2.15(a), employs a straightforward resistor (i.e., MOSFET channel resistance) whose resistance is set between Despite its simplicity, this approach occupies significant wafer area and lacks

adaptability to manufacturing variations. To address these limitations, an alternative design shown in Fig. 2.15(b) can be implemented. This configuration comprises two pairs of series-connected RRAM devices placed in parallel. Each pair contains one RRAM device programmed to LRS (R_{SET}) and another programmed to HRS R_{RESET}, resulting in an overall equivalent resistance precisely centered between these states. This design not only reduces cell size compared to the resistor-based reference but also ensures the reference cell experiences similar effects from process variations as the actual memory array cells.

2.5. PROTOTYPES AND COMMERCIALIZATIONS OF RRAM

An overview of current RRAM chips is provided in this section. First, we concentrate on both prototypes developed and commercially available RRAMs by research institutions and businesses. Next, we will provide a perspective overview of potential applications of RRAMs.

2.5.1. PROTOTYPES

Memristor technology, including RRAMs, has been extensively investigated and prototyped by both academia and industry (e.g., Samsung, Micron, Toshiba, and Sony), as shown in Fig. 2.16. Table 2.2 summarizes key parameters from various reported RRAM technologies, including details (if available) such as the publication year, developer/company, fabrication node, oxide (for oxide-based RRAM) or electrolyte (for CBRAM) materials, memory capacity, read/write latency (write latency is specifically listed as SET and RESET latency), cell architectures, endurance performance (i.e., how many cycles the cell can be written), and the resistance ratio (HRS/LRS). These parameters provide a comprehensive overview of the current state of RRAM technologies. We conclude the following points according to the table.

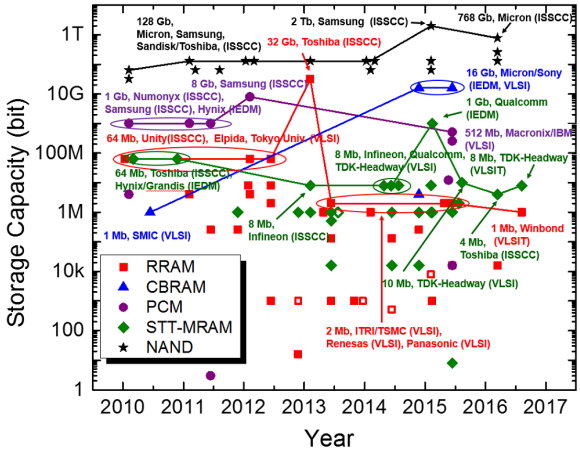


Figure 2.16: Trends of emerging device technologies (reprinted from [123]).

Technology node: Over time, prototypes have adopted increasingly smaller feature

Works	Year	Institution	Process	Material	Capacity	Arch.	Ratio	Write Latency SET/RESET	Read Latency	Endurance
[124]	2007	Qimonda	90 nm	Ag, CBRAM	1 Mb	1T-1R	10^7	9 ns	50 ns	10^7
[125]	2009	NTHU	90 nm	TiN/TiON		1T-1R	36	100 ns/10 μs		10^6
[126]	2009	NCU	180 nm	HfO_2	1 kB	1T-1R	10^3	5 ns	8.5 ns	10^8
[127]	2010	UnitySC	130 nm		64 MB	1R	10^3		100 μs	
[128]	2011	ITRI/NTHU	180 nm	HfO_2	4 MB	1T-1R	100	8 ns	8 ns	
[129]	2011	Sony	180 nm	CuTe, CBRAM	4 MB	1T-1R	100			
[130]	2011	Hynix	56 nm	TiO_x/Al_2O_3	256 kB	1T-1R	100			
[131]	2012	Panasonic	180 nm	TaO_x	8 MB	1T-1R	150		25 ns	
[132]	2012	Hynix	54 nm	Ta_2O_5/TiO_x	2 MB	1R	10	10 ns		
[133]	2012	NTHU	65 nm	TiON	4 MB	1T-1R	10		45 ns	
[134]	2012	NTHU	28 nm	TiON		1T-1R	10	500 ns/100 μs		10^6
[135]	2013	NTHU/TSMC	180 nm	TaO_x	512 kB	1T-1R				10^7
[136]	2013	Adesto	130 nm	CBRAM		1T-1R		250 ns	20 ns	10^5
[137]	2013	Sandisk/Toshiba	24 nm		32 GB	1D1R		230 μs	40 μs	
[138]	2014	NTHU/TSMC	28 nm	TiON	1 MB	1T-1R		500 ns/100 μs	6.8 ns	
[139]	2014	Macron/Sony	27 nm	Cu/CBRAM	16 GB	1T-1R				10^5
[140]	2015	NTHU/TSMC	16 nm	HfO_2	1 kB	1FinFET1R	>5			10^5
[141]	2015	Renesas	90 nm	$Ta_2O_5/TiO_x/Ru$			1.5			10^6
[142]	2017	Sony	180 nm	Cu/CBRAM	4 MB	2T 1S1R	100	100 ns		10^7
[143]	2017	Tsinghua Uni.	130 nm	HfO_2	16 MB	1T-1R				10^7
[144]	2017	UCAS	5 nm	TaO_x	10 MB		100	100 ns	300 ns	10^7
[145]	2018	TSMC	40 nm	HfO_x	11 Mb	1T-1R			9 ns	10^6
[146]	2019	Intel	22 nm		3.6 MB	1FinFET1R		10 μs	5 ns	
[147]	2020	UCAS	28 nm		1.5 MB	1T2R	50		3.3 ns	10^5
[83]	2021	CEA-Leti/UGA	28 nm	HfO_2	16 kB	1T(FDSOI)1R	10	1 μs		10^5
[148]	2023	Peking Uni.	40 nm	TaO_x	1 MB	3T2R	10			$> 1.2 \times 10^4$
[95]	2024	UCSD	180 nm	TiO_2		1R	$> 10^3$			2×10^5
[149]	2024	TSMC	12 nm	TMO	32 MB	1FinFET1R			5 ns	10^4
[150]	2025	TSMC	55 nm		16 MB	1T1R			18 ns	10^3@150 °C
[151]	2025	NTHU/TSMC	7 nm	HfO_2		1FinFET1R	> 5	50 ns/1 ms		$> 10^4$

Table 2.2: Summary of RRAM prototypes from different institutions and companies.

sizes, further demonstrating RRAM's compatibility with modern semiconductor technologies and its potential as a viable memory solution.

Memory capability: The table indicates that the highest recorded memory capacity was achieved in 2013, when a 32 GB prototype was introduced using 24 nm technology [137]. Since then, only smaller memory capacities have been reported. This decline may be attributed to the limited cycle endurance of larger memories or the relatively low write and read speeds of RRAM. Additionally, the manufacturing complexity of CBRAM compared to oxide-based RRAM could also be a contributing factor. Furthermore, RRAM is often utilized in embedded systems, where high memory capacity is not a primary requirement.

Cell design: The selector is applied widely for cell design, e.g., 1T-1R, 1D-1R. Interestingly, certain cells incorporate FinFET or FDSOI transistors in their fabrication process, demonstrating the compatibility of RRAM with these advanced semiconductor technologies.

HRS/LRS ratio: The RHRS/RLRS ratio has gradually decreased over time, which may indicate improvements in sensing accuracy. This enhancement contributes to increased chip reliability.

Unbalanced read & write: Read operations are typically faster than write operations. This is because reading only requires a brief sensing process, whereas writing involves the formation or rupture of the conductive filament, making it inherently slower. Additionally, SET operations tend to be faster than RESET operations due to the temperature feedback effect, which influences the dynamics of CF formation and rupture.

Endurance: Endurance was generally higher in older technology nodes but has consistently decreased in more advanced nodes. Coupled with the observed decline in the RHRS/RLRS ratio, this suggests that modern RRAM devices may be more prone to write failures, ultimately limiting their endurance.

2.5.2. COMMERCIALIZATIONS

Many companies are pushing the commercialization of RRAM devices. They can be classified into two types: 1) standalone RRAM devices serving as external memory in embedded systems, and 2) RRAM Intellectual Property (IP) modules designed for integration into custom chips. We list representative companies and introduce commercialized RRAM products below.

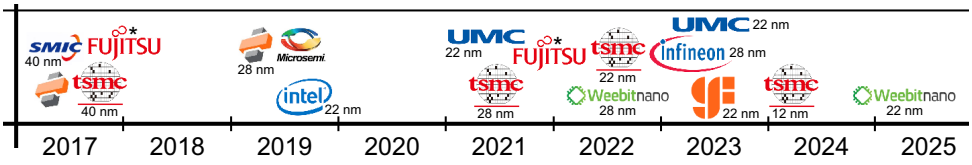


Figure 2.17: Timeline of recent technology demonstrations of eNVM with respective CMOS nodes (reprinted from [152]).

EMBEDDED RRAM

Fig. 2.17 demonstrates the timeline of recent technology for RRAM applied as Embedded NVM with respective technology nodes [152]. It is observed that they introduced one of their major RRAM product lines in 2022, with a previous significant launch occurring as early as 2016.

TSMC, as a leading semiconductor company, has made significant advancements in eNVM technology. By late 2017, their 40 nm RRAM technology entered risk production and obtained consumer-grade certification with a switching endurance (10^4 cycles) comparable to the lower range of NAND flash. By 2022, the 40 nm and 28 nm RRAM nodes had entered mass production, while the 22 nm node was fully prepared for manufacturing [153]. In 2024, TSMC demonstrated an RRAM-integrated macro using a 12 nm node, marking the smallest CMOS integration presented among the surveyed companies. Example performance parameters include a write endurance of cycles and a data retention capability of 10 years at 105 °C, the highest retention temperature reported in this comparison. The macro supports operation across a temperature range of –40 to 125 °C, supply voltages from 0.6 V to 0.9 V, word-line voltages exceeding 1.5 V, and read/write speeds around 21 nm [149].

Weebit Nano, as a prominent startup in the eNVM sector, has developed chips that are now production-ready on SkyWater's 130 nm CMOS platform. These chips feature a read voltage of 1.8 V, programming voltages of 1.8 V and 3.3 V, a read access time of under 20 nm, data retention of 10 years, and endurance of 10^4 cycles, all within an operating range of -40 to 125 °C. Efforts are underway to qualify this 130 nm architecture for enhanced performance, including operation at 150 °C, increased endurance of 10^5 cycles, and further optimizations as of 2024. Concurrently, Weebit is advancing its RRAM technology on Global Foundries' 22 nm platform [154] and has successfully demonstrated RRAM on CEA-Leti's 28 nm node in 2022, achieving switching endurance beyond 10^5 cycles and stable resistance for over 15 hours at 210 °C [155].

Infineon has embedded RRAM technology into their automotive-grade AURIX microcontrollers through a collaboration with TSMC [156]. These devices are specifically designed to meet stringent automotive standards, featuring the integration of RRAM within the microcontroller architecture. The embedded RRAM supports operation across a wide temperature range (−40 to 160 °C), and meets strict reliability criteria, including high endurance (more than 10^5 cycles) and data retention (1000 h), essential for automotive environments [157].

Fujitsu released details regarding their RRAM technology, highlighting features such as a wide operating voltage range (1.6–3.6 V), read currents optimized between 1.6 and 3.6 V, write endurance reaching 5×10^5 cycles, and unlimited endurance for read operations in 2024 [158]. Additionally, the reported specifications include compatibility with rapid read speeds, emphasizing robust reliability suitable for high-performance applications.

RRAM IP

Fig. 2.18 presents the number of IP applications for the four emerging memory technologies, which has consistently grown over recent years. Among them, RRAM and MRAM currently hold the largest share, while PCRAM and FeRAM maintain a stable presence, though with relatively fewer patent filings.

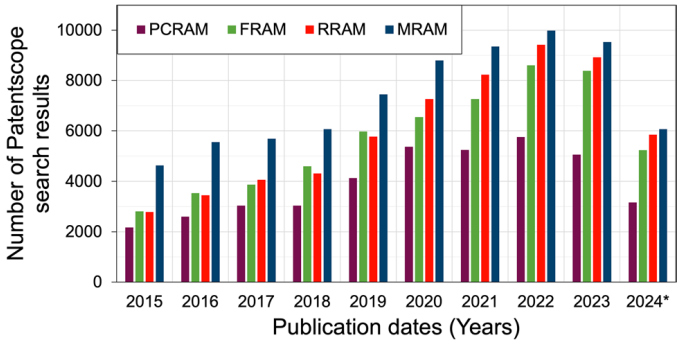


Figure 2.18: Word IP organization patent scope search results from 2015 to July 2024 (reprinted from [152]).

Dialog Semiconductor: They initially licensed its CBRAM technology to Global Foundries in 2020. After Renesas acquired Dialog in 2021, Global Foundries subsequently purchased the full rights to the CBRAM IP from Renesas in 2023.

Crossbar Inc.: They have commercialized the RRAM technology, employing it in applications such as mobile computing and AI. Rather than engaging in direct manufacturing, Crossbar appears focused on licensing its IP, particularly targeting advanced nodes like 2X nm and 1X nm. For instance, in 2018, Crossbar licensed its core 28 nm RRAM technology to Microsemi. This licensing model aligns with their ongoing strategy of IP commercialization, particularly aimed at advanced technology nodes, enabling further integration by external manufacturers.

2.5.3. RRAM APPLICATIONS

The primary goal of an eNVM is its effective utilization in memory storage solutions. RRAM technology shows promise as a replacement for existing memory devices, provided it meets critical performance criteria. However, widespread adoption depends not only on achieving superior specifications but also on overcoming additional barriers posed by existing memory solutions, such as mature manufacturing processes, established market presence, and competitive performance [159]. Furthermore, factors including cost competitiveness, compatibility with current infrastructure, and scalability must be addressed before new technologies can effectively penetrate the market. Basically, the RRAM devices are potentially to be used as follows.

LARGE STORAGE CLASS MEMORY

A significant performance gap exists between storage and memory technologies [7]. For example, NAND flash is nonvolatile but limited in speed and endurance, whereas DRAM offers high speed but sacrifices data retention due to volatility [4]. This gap can substantially impact overall system performance and cost. Consequently, developing a 'Storage Class Memory (SCM)' that effectively combines nonvolatility, speed, endurance, and retention remains a critical yet challenging goal [85]. Currently, RRAM devices have demonstrated significant potential, positioning them as leading candidates for use as SCM solutions [75], [160].

ENERGY EFFICIENT COMPUTING-IN-MEMORY ARCHITECTURES

Computing-in-Memory (CIM), which breaks the conventional separation between memory and processing units, has emerged as a promising paradigm to overcome the data-movement bottleneck of von Neumann architectures [161]–[165]. By directly performing operations such as vector–matrix multiplications within memory arrays, CIM can drastically improve energy efficiency and throughput. Among various device candidates, RRAM is particularly attractive due to its non-volatility, high density, and capability of supporting both digital and analog switching modes. In the analog mode, the conductance of RRAM cells can be finely tuned, enabling efficient implementation of multiply–accumulate operations central to artificial intelligence workloads. Furthermore, the compatibility of RRAM with CMOS technology makes it feasible to build compact, high-performance CIM systems. As demonstrated in several recent works, memristor-based CIM architectures are being applied not only in conventional AI acceleration but also in emerging domains such as taxonomic profiling and biomedical signal analysis [37], [166]–[170]. Therefore, incorporating analog-type RRAM devices into CMOS circuits provides a robust pathway toward scalable, energy-efficient computing-in-memory architectures for future AI and machine learning applications [171].

NEUROMORPHIC COMPUTING

Neuromorphic computing, also known as brain-inspired computing, represents a promising new application area beyond traditional memory [95], [172], [173]. Compared with conventional von Neumann architectures, the human brain efficiently executes sophisticated tasks such as pattern recognition and inference while consuming significantly less power. A key feature of neural networks is Spike-Time-Dependent Plasticity (STDP), a well-established learning principle based on synaptic weight adjustments. RRAM devices, which exhibit either digital or analog switching, become particularly valuable in neuromorphic computing if they can precisely modulate their conductance levels through analog switching [174]. Thanks to its compatibility with CMOS integration, RRAM technology can effectively address critical challenges in artificial intelligence applications [175]. Hence, incorporating analog-type RRAM devices into CMOS could provide a robust solution for developing compact, energy-efficient neural computing systems.

HARDWARE SECURITY

Although the inherent stochastic behaviors of eNVMs are typically undesirable for memory storage applications, these unpredictable variations can serve effectively as sources of entropy in security oriented implementations. Specifically, variations stemming from resistance states, which pose reliability concerns in memory applications, can be beneficially utilized as entropy sources for security. The intrinsic randomness present in eNVM devices, often seen through C2C and D2D variability of resistance states (LRS or HRS), can be leveraged in True Random Number Generators (TRNGs) [176]. This randomness is particularly valuable for security purposes, such as generating cryptographic keys. Thus, while stochasticity might pose reliability issues in conventional memory contexts, it provides significant opportunities for enhancing hardware security.

3

RRAM Manufacturing, Defects, and Non-Idealities

Manufacturing RRAM devices involves modifications to the standard CMOS fabrication procedures, as conventional processes alone are not sufficient. This is because the typical fabrication methods for CMOS devices don't inherently accommodate the materials and processes required by RRAM technology. As a result, specialized integration steps must be introduced. In this chapter, we thoroughly examine the RRAM device, starting with a comprehensive description of its physical structure and integration strategies. Subsequently, the detailed fabrication flow of RRAM is discussed, including critical phases such as Front-End-Of-Line (FEOL), Back-End-Of-Line (BEOL), and the filament-forming process. Alongside these steps, the unique considerations associated with each stage are explored. Finally, the potential defects arising throughout these manufacturing stages are identified and analyzed, providing insights into how each affects device performance and overall non-idealities (both time zero and time dependent).

3.1. MANUFACTURING PROCESS OVERVIEW

The fabrication process of integrated RRAM typically consists of three main stages: Front-End-Of-Line (FEOL), Back-End-Of-Line (BEOL), and filament formation. A schematic overview of this process flow is provided in Fig. 3.1 [85], [106], [107], [177]. Initially, transistors are fabricated on the silicon wafer during the FEOL phase. Following this, metal interconnects and memory devices are integrated during the BEOL phase. Finally, the conductive filament is established through a dedicated forming step, completing the manufacturing sequence. Next, we will introduce each step in detail.

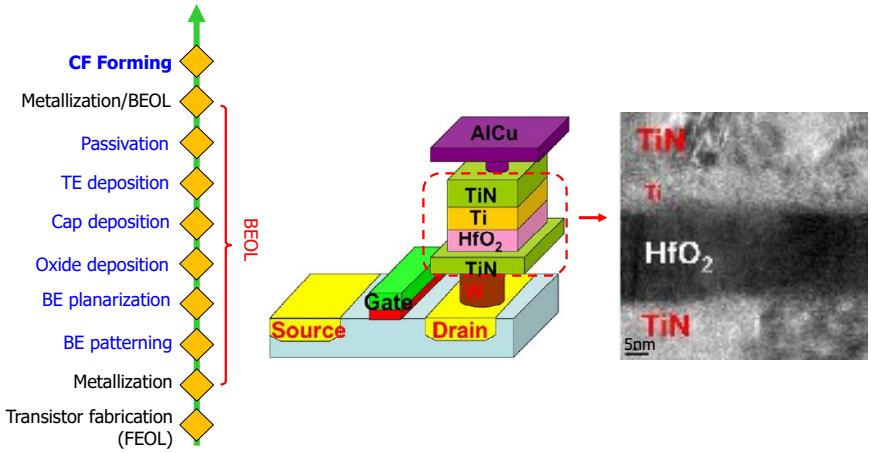


Figure 3.1: General manufacturing process of RRAM (right part reprinted from [85]).

3.1.1. FRONT-END-OF-LINE

Transistors are assembled on the wafer during the FEOL fabrication stage through a sequence of iterative processes, each of which contributes to the formation of distinct structural components. A material layer is deposited or grows on top of the substrate. This layer could be silicon oxide for insulation, a different type of oxide for the gate, or metal for connections. The process covers the whole surface of the substrate evenly. A photoresist layer is subsequently applied, and the desired patterns are defined through lithography. The underlying material is selectively exposed by selectively removing either the exposed or unexposed regions of the photoresist, depending on the process requirements. During this phase, the exposed regions are subjected to additional processing, including etching to modify the material's morphology or ion implantation to modify its electrical properties. The transistor's precise structural definition and functionality are guaranteed by the systematic repetition of these fabrication stages in the construction of various sections.

3.1.2. BACK-END-OF-LINE

The subsequent stage in manufacturing is the BEOL, which involves forming metal interconnects and integrating the RRAM devices. The fabrication of these layers utilizes de-

position and lithography techniques similar to those employed in the FEOL phase. Once these lower metal layers are established, fabrication of the RRAM device itself begins. Initially, the BE, which links to the underlying metal layer, is deposited and patterned. Following the formation of the BE, the oxide layer is deposited on its surface. To reduce variability in device characteristics, precise control over the oxide thickness and defect density—both within the oxide and at its interfaces—is essential. Consequently, Atomic Layer Deposition (ALD) is commonly preferred over Physical Vapor Deposition (PVD), as ALD allows superior control of these parameters [118], [177]. The next step in the fabrication process is the deposition of the capping layer. If a capping layer is not required, this step is omitted, and production proceeds directly to the TE. The individual memristive devices are subsequently separated by etching the layer of materials. Lastly, the memristive device is electrically isolated from surrounding structures, ensuring proper functionality. The fabrication then proceeds with the deposition of the next metal layer, following the standard BEOL workflow used in conventional CMOS manufacturing.

3.1.3. FORMING PROCESS

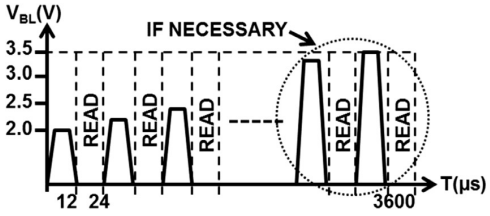


Figure 3.2: Controllable forming pulse and read verification scheme (reprinted from [178]).

A CF is formed within the oxide layer during the final stage of RRAM fabrication. The conditions under which this formation occurs significantly influence the geometry of the filament. Typically, the required forming voltage (V_{forming}) exceeds both the V_{SET} and V_{RESET} values [179]. While a higher V_{forming} accelerates CF formation, it also increases the risk of oxide breakdown [180]. To mitigate this risk, adaptive forming techniques can be employed, where V_{forming} is dynamically adjusted based on real-time CF development monitoring [178], [181]. Fig. 3.2 illustrates an example of such a method, in which short voltage pulses are applied sequentially, with resistance verification conducted after each pulse. If the target resistance has not yet been achieved, the voltage of the subsequent pulse is incrementally increased until the desired resistance level is attained.

In addition to V_{forming}, the forming current (I_{forming}) passing through RRAM devices is another critical factor influencing device characteristics. When I_{forming} is higher, CFs tend to be wider, which means lower resistance and less variation. When I_{forming} is lower, CFs tend to be thinner, which means higher resistance and more variation [106], [179], [182]–[184]. To ensure device stability, I_{forming} should be kept as consistent as possible, as fluctuations can introduce long-term resistance deviations [185].

Beyond forming conditions, device geometry and structure also play a significant role. Smaller devices tend to exhibit higher resistance due to a lower probability of CF formation, while larger devices have an increased probability of forming a stable CF

Table 3.1: Overview of RRAM manufacturing defects across different process stages.

FEOL		BEOL
Transistor	**Interconnection**	**RRAM Device**
Patterning proximity	Opens	Electrode roughness
Line roughness	Shorts	Polish variations
Polish variations	Line roughness	Varying defect density
Anneal	Irregular shapes	Dimensional variations
Strain	Big bubbles	Lower capping layer binding capacity
Gate granularity	Small particles	Material redeposition
Dielectric variations		Over/Under-forming
		Interface ion depletion

[107], [180], [184]. Finally, variations in structural properties, such as material composition and electrode configuration, can further impact the filament formation process and overall device performance.

3.2. RRAM DEFECTS AND CLASSIFICATIONS

Defects arising during the fabrication of RRAM can be categorized into three groups: transistor-related defects, interconnection defects, and defects specific to the RRAM device itself. Table 3.1 summarizes these defects based on this categorization. The subsequent sections provide an in-depth discussion of each defect type and its occurrence throughout the manufacturing process.

3.2.1. TRANSISTOR-RELATED DEFECTS

Transistor-related defects exist in the FEOL step, where various defects can arise. Kuhn, Giles, Becher, *et al.* classify into two main types: historical defects and emerging defects [186].

Historical defects, which have been present across multiple technology generations, include several key issues [187]:

- Patterning proximity effects, where unintended exposure to photons affects areas outside the intended pattern.
- Line-edge and line-width roughness, caused by photon fluctuations during lithography, leading to irregular pattern edges.
- Polish variations, which result in non-uniform layer thickness during planarization.
- Gate dielectric inconsistencies, where defects within the gate oxide material lead to variations in electrical properties.

On the other hand, **emerging defects** have become increasingly significant in advanced, smaller technology nodes [186]:

- Random dopant and anneal fluctuations, where minor variations in doping concentrations significantly impact transistor performance due to the reduced number of dopant atoms in the device.

- Mechanical strain, where nanoscale structures induce stress on adjacent regions of the wafer, potentially affecting reliability.
- Gate material granularity, where the gate is not perfectly uniform but exhibits structural variability, influencing electrical behavior.

3.2.2. RRAM CONVENTIONAL DEFECTS

Conventional defects are outside of the RRAM device, i.e., interconnects and contacts. In FEOL, conventional defects that may arise include semiconductor contamination, crystal lattice irregularities, big bubbles, and small particles [188]. Typically, they are studied and modeled by resistive bridges, opens, and shorts [189].

During the BEOL process, metallization procedures for the lower layers closely follow standard production processes, meaning common manufacturing defects, such as alignment inaccuracies or particle contamination, can still occur. Misalignment or particle contamination may cause degraded interconnect quality, leading to increased wire resistance. Additionally, lithography-related imperfections like line-edge roughness can lead to irregular shapes, affecting wire capacitance and resistance, and consequently degrading RRAM performance. Increased wire resistance diminishes the voltage reaching the RRAM device, constraining its switching operation, whereas higher capacitance may introduce undesired cross-talk between adjacent interconnects.

3.2.3. RRAM UNIQUE DEFECTS

Unique defects are intrinsic to the RRAM device and remain relatively unexplored. Both the BEOL and forming processes can cause them. We will summarize them according to each manufacturing step.

RRAM UNIQUE DEFECTS IN BEOL

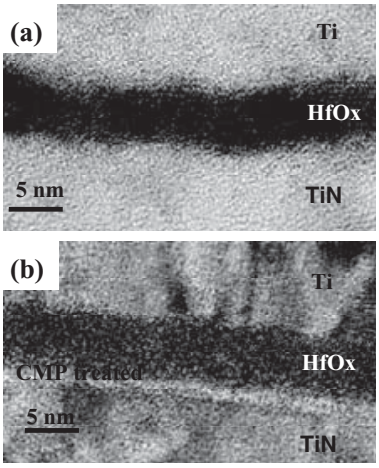


Figure 3.3: The comparison between electrodes (reprinted from [190]). (a) Unpolished electrode, (b) Polished electrode after post-meatal annealing.

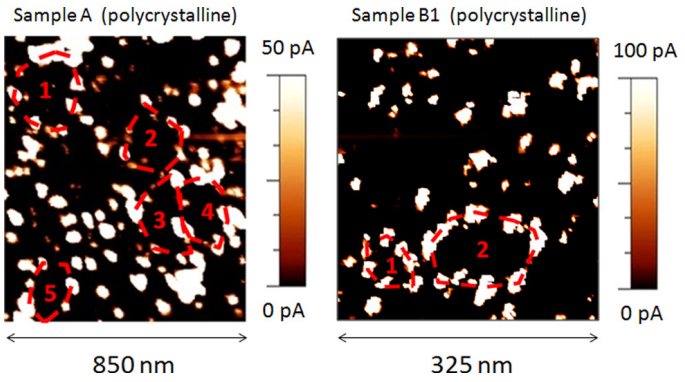


Figure 3.4: Grain boundaries conduct current in oxide (reprinted from [191]).

First, the deposition typically results in surface roughness of the BE. Such roughness creates numerous defects at the interface between the BE and the RRAM oxide layer, which increases device variability and heightens the risk of hard oxide breakdown. To mitigate these issues, a Chemical-Mechanical Polishing (CMP) step is performed to smoothen the BE surface, reducing electrode roughness as depicted in Fig. 3.3. Precise control during polishing is crucial to avoid introducing additional variations and defects.

Next, during the oxide deposition, the resulting oxide structure may be either polycrystalline or amorphous. Polycrystalline structures feature grain boundaries, which inherently possess a higher density of defects compared to crystal interiors. These boundaries serve as preferential paths for CF formation, as illustrated in Fig. 3.4, where electrical conduction through an HfO_2 layer predominantly occurs along grain boundaries [191], [192]. Variations in crystal sizes directly influence the density and arrangement of grain boundaries, resulting in broader resistance distributions for polycrystalline oxides than for amorphous ones [193]. On the other hand, amorphous oxides typically exhibit lower HRS/LRS ratios, making the resistance states more challenging to differentiate clearly [193].

Then, when the capping layer is fabricated, the precise thickness of the capping layer is less critical, unlike the oxide layer, since it has significantly higher conductivity. As a result, deposition techniques with lower precision, such as PVD, are commonly used [107]. However, the capping layer plays a crucial role in oxygen ion binding. It must be sufficiently thick and appropriately doped to capture all oxygen ions released during the formation of the CF [103]. A potential issue in this stage is insufficient oxygen binding capacity, which can lead to device variability. Another example is the ion depletion in the capping layer, which causes the HRS degradation due to insufficient oxygen ions that can move back to the oxide [194]. Following this, the TE is deposited using a process similar to that of the BE, though in many cases, the final polishing step is not required.

Finally, the stack of materials is required to be etched. During this procedure, lithographic inconsistencies can once again introduce variations in device dimensions, ultimately impacting the performance of the final RRAM structure [68]. Additionally, the etching process may cause unintended material redeposition along the sidewalls of the

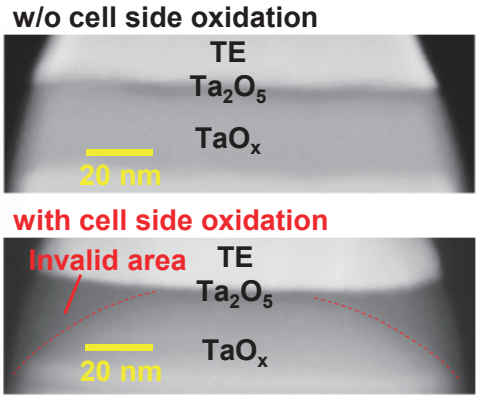


Figure 3.5: Partial invalidation of sidewall (reprinted from [195]).

device [196], [197]. This redeposition creates unwanted leakage paths, reducing the overall device resistance and leading to lower forming voltages and increased resistance variability [196]. To mitigate redeposition effects, the manufacturing process can be optimized by designing the via connecting to the bottom electrode to be smaller and completely covered by the electrode, preventing the incorporation of metallic contaminants [197]. Furthermore, optimizing the etching process to ensure a more vertical profile can further minimize atomic redeposition along the sidewalls [195]. Research by Hayakawa, Himeno, Yasuhara, et al. suggests that device uniformity can be improved by deactivating a strip adjacent to the sidewall, as shown in Fig. 3.5 [195]. This deactivated region can then be coated with a protective layer, effectively reducing the impact of sidewall leakage and redeposition effects.

RRAM UNIQUE DEFECTS IN FORMING PROCESS

During the forming process of RRAM, two defects can arise, influencing the device's performance relative to its intended function. One potential defect is the failure to establish a CF properly, resulting in the device being permanently stuck in an HRS or even extreme high resistance, known as an *under-forming* defect. This occurs when the conditions necessary for filament formation are insufficient. Conversely, excessive CF formation can occur due to factors such as uncontrolled forming current, causing the device's resistance to drop below its intended range, a phenomenon referred to as *over-forming* [49], [70]. In extreme cases, an excessively strong forming process can lead to oxide breakdown, permanently shorting the device into an LRS, preventing any switching back to HRS [180]. Fig. 3.6 illustrates an RRAM cell where the oxide layer has experienced breakdown, especially the formation of CF located at grain boundaries of HfO_2 crystals [192], [198].

3.3. NON-IDEALITY ISSUES IN RRAMS

RRAM devices suffer from non-idealities that may cause functional errors during the deployment [68]–[70]. Such non-idealities consist of *time-zero* and *time-dependent*

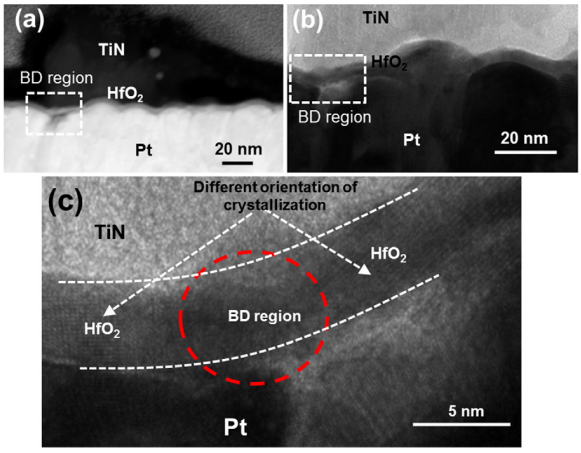


Figure 3.6: The breakdown region image in the TiN/HfO_2/Pt stack (reprinted from [192]).

ones [71]; the time-zero ones consist of variation, wire parasitic, etc, while the time-dependent ones consist of endurance, device degradation, resistance drift, etc. Even if the RRAM chip passes the manufacturing test, it may suffer from in-field functional errors due to these non-idealities. In this section, we discuss both time-zero and time-dependent non-idealities in RRAMs.

3.3.1. TIME-ZERO NON-IDEALITIES

- **Wire parasitics:** The inherent parasitic resistance and capacitance of interconnect wires introduce signal distortions, leading to delay mismatches and voltage attenuation, which can result in faults [199]. In logic operations, for example, discrepancies arise due to variations in propagation paths—causing reference and input signals to experience different delays before reaching sensing circuits such as sense amplifiers. Moreover, as signals travel along the wordline, degradation occurs, particularly in distant columns, reducing the associated current output and further impacting performance.

- **Variation:** Two main variations exist in RRAM devices due to an inherent stochastic process of CF growth and dissolution [182], [200]: 1) Device-to-Device (D2D) variations in the resistance among multiple devices, and 2) Cycle-to-Cycle (C2C) variations in multiple SET and RESET cycles within a single device. The D2D variation originates from inconsistencies during the fabrication of RRAM devices, for example, due to differences in the oxide's crystalline structure or fluctuations in forming current. C2C variations come from the random CF generation process in each cycle, e.g., variations in the length and radius of the CF. The movement of oxygen ions exhibits variations, which lead to different resistances of RRAM devices per write cycle. Typically, devices in the HRS exhibit higher variations than in the LRS, which can be explained by the wider CF in the LRS being less affected by single ions [106], [180]. Furthermore, there are traditional CMOS Process, Volt-

age, and Temperature (PVT) variations, which can also affect the memory storage and CIM implementation [201], [202].

Resistance variation in RRAMs refers to the deviation of the programmed resistance from its intended value, potentially causing faults in both storage memory and CIM implementations [62], [203]. This phenomenon primarily arises from fabrication inconsistencies and the inherent stochasticity of the device's physical processes. Furthermore, fluctuations in process parameters, operating voltage, and temperature (PVT variations) in standard CMOS manufacturing can exacerbate these deviations, further affecting the RRAM reliability. Variations in the production process and manufacturing defects can significantly amplify C2C fluctuations, further affecting device stability. For instance, the forming current plays a crucial role in determining the initial CF structure. A lower forming current tends to produce narrower CFs, making them more vulnerable to resistance variability [106], [179]. Additionally, inconsistencies in the polishing process can introduce defects at the oxide interface, disrupting the formation and dissolution dynamics of the filament, thereby increasing resistance variation [204]. In comparison, temperature has a relatively minor influence on C2C fluctuations [205].

3.3.2. TIME-DEPENDENT NON-IDEALITIES

- **Endurance:** RRAMs face endurance limitations due to the inherently disruptive nature of their programming mechanisms. In RRAM devices, resistive switching relies on the movement of oxygen ions, where the CF is repeatedly formed and ruptured during write operations. However, repeated write cycles progressively degrade the ON/OFF resistance ratio, ultimately leading to device failure over time. Furthermore, multiple writes can cause the device to either get stuck in HRS, or in LRS [204], [206]. The first issue arises because a RESET operation is excessively strong, leading to instability in the switching process. Conversely, the second occurs when each SET operation progressively enlarges the CF, making it increasingly difficult to rupture during subsequent RESET cycles.

- **Device degradation:** Over time, stress and aging contribute to the degradation of both CMOS peripheral circuits and RRAMs in circuits [207]. Fig. 3.7 illustrates two typical endurance degradation measurements and the potential mechanism [207]. These effects are further exacerbated by elevated operating voltages and high temperatures, accelerating wear-out mechanisms and impacting device reliability.

- **Conductance drift:** Over time, the conductance levels of memristors gradually shift, which may ultimately result in unintended bit errors or state changes [207].

- **Read disturb failure:** Read disturb is a phenomenon in which a memory cell initially returns the correct value during a read operation, but the process itself unintentionally alters the stored data, potentially causing a bit flip. An example of the read disturb issue is shown in Fig. 3.8.

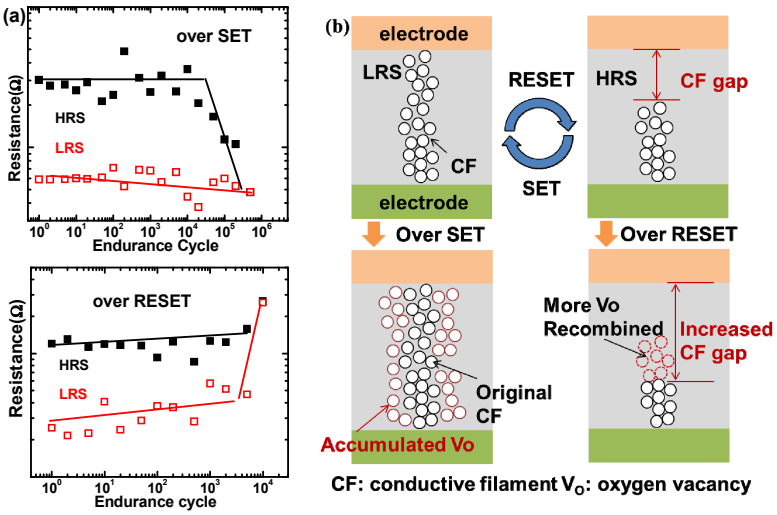


Figure 3.7: Two typical endurance degradation characteristics (reprinted from [207]). (a) Experiments. (b) Schematics of endurance degradation mechanisms.

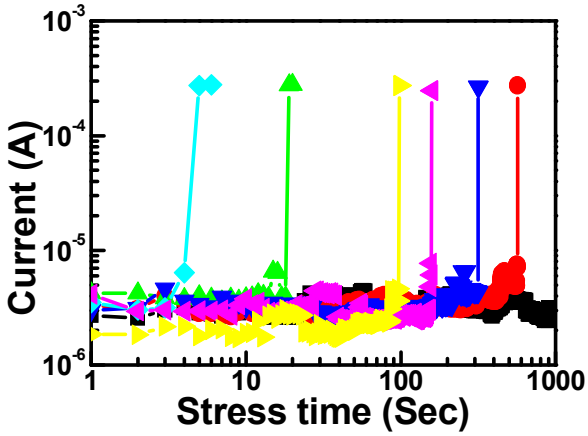


Figure 3.8: The resistance of the device in HRS decreases under stress test (reprinted from [208]).

4

TESTING FOR CONVENTIONAL DEFECTS

The commercialization of RRAM requires not only high yield, but also higher product quality and reliability. Hence, it is important to accurately model defects and faults for efficient test development. In this chapter, a holistic framework is demonstrated to model interconnect and contact defects, derive fault models, and generate test methods based on observed fault primitives. This chapter analyzes interconnect and contact defects in RRAMs, while considering the impact of the memory Data Background (DB), and proposes test solutions. The complete interconnect and contact defect space in a layout-independent RRAM design is defined. Exhaustive defect injection and circuit simulation are performed in a systematic manner to derive appropriate fault models, not only for single-cell and two-cell coupling faults, but also for multi-cell coupling faults where the DBs are important. The results show the existence of unique 3-cell and 4-cell coupling faults due to, e.g., the sneak path in the array induced by defects. These unique faults cannot be detected with traditional RRAM test solutions. Therefore, the chapter introduces a test generation method that takes into account the DB, which is able to efficiently detect all these faults; hence, further improving the fault/defect coverage in RRAMs.

Parts of this chapter were published in ETS'23 [64].

4.1. CONVENTIONAL RRAM TESTING APPROACH

Ensuring the reliability of commercial RRAM production relies heavily on efficient testing. As large-scale manufacturing advances, effective test methodologies become essential for identifying defects and maintaining optimal performance in practical applications. Due to the intricate nature of the fabrication process and the distinct operational characteristics of RRAMs, a comprehensive and systematic approach is crucial for developing robust testing solutions.

An effective test strategy must fulfill several key criteria to ensure both quality and efficiency in RRAM manufacturing. First, it should provide comprehensive defect coverage, capable of identifying a wide range of conventional and technology-specific defects. Insufficient detection can compromise product reliability and degrade overall quality. Second, test time optimization is essential for large-scale production. Lengthy testing procedures increase manufacturing costs and may negatively impact yield, making it crucial to strike a balance between thorough defect detection and time efficiency. Furthermore, an ideal test strategy should minimize yield loss, preventing defect-free chips from being mistakenly classified as faulty. Excessive yield loss leads to unnecessary waste and escalates production expenses. Finally, adaptability to different RRAM architectures is vital, ensuring the testing methodology remains effective across various manufacturing conditions and design variations. Next, we will provide a systematic test approach for RRAMs.

As illustrated in Fig. 4.1, the traditional RRAM test development process follows a structured three-step approach: defect modeling, fault modeling, and test development [39]. This framework is designed to systematically identify and address potential defects, enabling effective, high-coverage, and cost-efficient testing strategies.

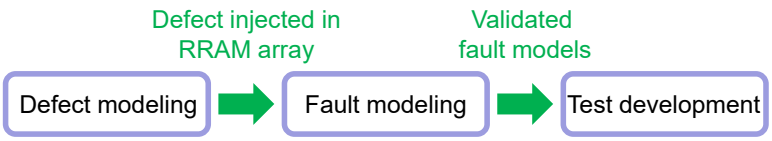


Figure 4.1: Process of conventional test development

DEFECT MODELING

Defect modeling is a fundamental step in the test development process for RRAM, as it translates physical manufacturing defects into their corresponding electrical representations [39]. During fabrication, RRAM devices may encounter defects stemming from lithography misalignment, etching imperfections, deposition inconsistencies, and other process variations. These defects can manifest in contacts, interconnects, switching layers, or peripheral circuits, leading to unintended electrical behaviors.

To systematically evaluate their impact, defects are modeled using electrical circuit elements. A widely adopted approach is to represent physical defects as resistors [43]. For example, metal contamination during fabrication may unintentionally create a conductive path between two normally isolated circuit nodes [209], [210]. This phenomenon is typically modeled as a bridge resistor, where its resistance value determines

the strength of the defect, i.e., a lower resistance indicates a stronger electrical impact, whereas a higher resistance suggests a weaker, condition-dependent effect.

Similarly, open defects can be represented as either high-resistance connections or complete disconnections between previously linked nodes. Such defects often result from incomplete etching or deposition errors, leading to floating nodes that disrupt memory operation [209], [210]. In this case, a higher resistance value correlates with a more severe defect.

By employing defect modeling, physical defects are mapped onto a circuit-level model that can be analyzed and simulated using tools like Spectre, enabling precise evaluation of their electrical impact and facilitating the development of effective test strategies.

FAULT MODELING

Once defects in RRAM are accurately modeled at the electrical level, fault modeling is performed to examine how these defects translate into faulty memory behaviors during actual operations [39]. This process involves injecting defect models into circuit simulations and analyzing their impact on memory functionality.

For instance, introducing a bridge resistor between two WLs can result in write disturb faults, where writing to one memory cell unintentionally alters the state of an adjacent cell due to unintended leakage paths [43]. Similarly, an open defect in the BL can lead to read failures, where the sensing circuit fails to accurately decide stored data due to excessive signal degradation, often caused by insufficient read current [40].

These faulty behaviors are systematically described using fault models, which employ specific fault notation to categorize different failure mechanisms [40]. Fault models are crucial for designing effective test solutions, as they help identify the types of memory errors that need to be detected. Traditional memory testing typically relies on stuck-at faults, transition faults, and coupling faults as standard fault models. However, in RRAM, additional fault mechanisms must be considered due to the unique resistive switching behavior of the device.

TEST DEVELOPMENT

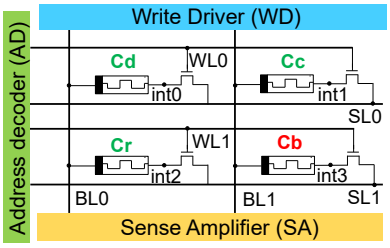
The final stage in RRAM test development involves designing efficient test solutions to identify faults detected in the previous step [39]. The goal of test generation is to create optimized test patterns that effectively target all identified faults while minimizing test execution time.

A widely used approach is March testing, where a series of write and read operations are performed on memory cells to verify if any failures occur. March tests, a common methodology in traditional memory testing [211], can be adapted for RRAM to detect conventional memory faults. The specific sequence of applied operations is defined by the March algorithm.

To enhance testing efficiency, Built-In-Self-Test (BIST) techniques are frequently employed [212]. A BIST system integrates dedicated test circuitry directly within the memory chip, enabling autonomous execution of March algorithms without requiring extensive external test equipment. This approach is particularly beneficial for large-scale RRAM manufacturing, where reducing test time is critical.

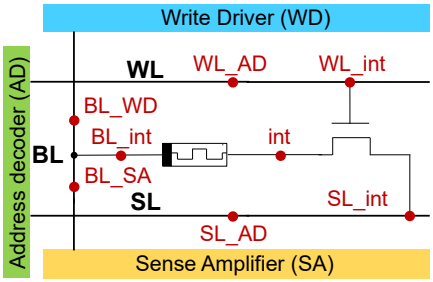
Additionally, Design-for-Testability (DfT) techniques can be incorporated to improve fault detection. These involve integrating specialized circuits that help identify defects that may not be effectively captured by standard March algorithms [212], [213]. By combining BIST and DfT methodologies, test coverage can be significantly enhanced, ensuring robust and efficient fault detection in RRAM devices.

(a)



Bridge	Location	Bridge	Location	Bridge	Location
BC1	BL1-int3	cBCC3	int1-SL1	rBCC2	BL0-int3
BC2	BL1-WL1	cBCC4	WL0-int3	rBCC3	int2-BL1
BC3	BL-V_{DD}	cBCC5	WL0-WL1	rBCC4	int2-int3
BC4	int3-WL1	cBCC6	WL0-SL1	dBCC1	int0-BL1
BC5	int3-SL1	cBCC7	SL0-int3	dBCC2	int0-int3
BC6	WL1-SL1	cBCC8	SL0-WL1	dBCC3	int0-WL1
cBCC1	int1-int3	cBCC9	SL0-SL1	dBCC4	int0-SL1
cBCC2	int1-WL1	rBCC1	BL0-BL1		

(b)



Shorts	Location	Opens	Location
SC1	int-V_{DD}	OC1	BL_int
SC2	int-GND	OC2	int
SB1	BL-V_{DD}	OC3	WL_int
SB2	BL-GND	OC4	SL_int
SW1	WL-V_{DD}	OW	WL_AD
SW2	WL-GND	OS	SL_AD
SS1	SL-V_{DD}	OB_w	BL_WD
SS2	SL-GND	OB_r	BL_SA

Figure 4.2: Conventional defect space. (a) Bridges that intra or inter cells, (b) Shorts and opens that intra the cell.

4.2. DEFECT MODELING

As the targeted defects in this chapter consist of interconnect and contact defects, it is appropriate to use *linear resistors* to model them [43], [52]; note that this is not the case for other unique defects in RRAMs; e.g., over-forming [59], [214]. Interconnect and contact defects can cause opens, shorts, and bridges [43], [52]. An *open* is defined as increased resistance in an existing connection, a *short* as an undesired resistive path between a node and a power node (V_{DD} or GND), and a *bridge* as a resistor between a pair of nodes different from the power nodes.

Before defining the total number of defects to be simulated, we need to define the simulation platform. To reduce the simulation time, we use the symmetrical nature of the memory array to derive a representative simulation platform while reducing the number of defects to be simulated. Each cell in the memory array (say base cell C_b) has at most 4 adjacent diagonal cells (C_d), two adjacent cells in the same column (C_c), and two adjacent cells in the same row (C_r). Given the symmetry, the simulation platform can be reduced to a 2×2 cell array as shown in Fig. 2.10; it presents a base cell C_b with a representative of each neighbor.

The symmetry can further help in reducing the number of defects to simulate within

2×2 arrays. For example, in Fig. 2.10, a bridge between int0 of C_d and int3 of C_b exhibits symmetry to a bridge between int1 of C_c and int2 of C_r; hence, only one of these needs to be simulated. Applying the symmetry to the simulation platform of Fig. 2.10 results in 8 opens, 8 shorts, and 23 bridges; these give the complete defect space that needs to be simulated in order to fully analyze all possible interconnect and contact defects within an RRAM array.

4.2.1. DEFINITION AND LOCATION OF BRIDGES

A bridge is a parallel resistor between a pair of connections. The resistance of a parallel resistor represents the strength of a bridge, e.g., a 100 Ω bridge is a stronger defect than a bridge of 100 kΩ. We only define all possible bridges within a cell or between two adjacent cells when considering their high occurrence probability in the real layout and circuit design [56], [215], [216]. For example, bridges between adjacent devices can be introduced by over-etching, as seen in CMOS technology [217]. First, the definition of bridges within a cell (BCs) is the bridge connections between two nodes in one cell. Each cell used in this chapter consists of four nodes n, $n \in \{int, BL, WL, SL\}$. Therefore, there are $\binom{4}{2} = \frac{4!}{2!(4-2)!} = 6$ bridges (named from BC1 to BC6) within a cell, as shown in Fig. 4.2(a).

Bridges may also occur between adjacent cells. The 2×2 memory array is adapted to establish all possible bridges between cells (BCCs). The location of BCCs can be further classified as BCCs in the same column (cBCCs), BCCs in the same row (rBCCs), and BCCs between diagonal cells (dBCCs). In this array, the adjacent cells in the same column share common BLs, while the adjacent cells in the same row share common WLs and SLs. There are 17 inter-cell bridges (denoted as $xBCC, x \in \{c, r, d\}$) as follows, as listed in Fig. 4.2(a). First, the adjacent cell in the same column (C_c) consists of four nodes: int_c, BL_c, WL_c, and SL_c. Since C_c shares a common BL with C_b, there are $(4-1)^2 = 9$ possible cBCCs numbered from cBCC1 to cBCC9. Secondly, possible rBCCs can be bridged between the adjacent cell in the same row (C_r) and C_b. They share both of the same WL and SL. Therefore, the possible number of rBCCs is $(4-2)^2 = 4$. The possible rBCCs, bridge combination of $n_b - n_r$; whereby $n_b \in \{int_b, BL_b\}$ and $n_r \in \{int_r, BL_r\}$ are named from rBCC1 to rBCC4. Finally, only four dBCCs (named from dBCC1 to dBCC4) remain to be considered. All other bridges between the diagonal cells are repeated with the above cBCCs, rBCCs, and BCs.

4.2.2. DEFINITION AND LOCATION OF SHORTS

A short is an undesired parallel resistive path between a node and power nodes (V_{DD} or GND). The smaller value of resistance represents a stronger short defect strength. In this chapter, we define all short defects as two types. The first type of shorts connects the inter node (between the RRAM device and access transistor) to V_{DD} and GND within a cell. The second type connects the BL, WL, or SL to the power supply. They are named a short in the BL, a short in the WL, and a short in the SL, respectively. SX is used to denote the shorts ($SX, X \in \{C, W, S, B\}$) Shorts also have spatial symmetry and do not need to be considered repeatedly. Names of considered all 8 shorts are shown in Fig. 4.2(b).

4.2.3. DEFINITION AND LOCATION OF OPENS

An open may occur owing to the missing material or broken nanowire. In such cases, there is increased resistance to an existing connection. Thus, the open defects can be modeled as series resistors with higher resistances. The larger value of resistance indicates a stronger open defect. Opens that cause broken connections in the memory array. OX is used to denote the opens (OX, X ∈ {C (inside the cell), W (in the WL), S (in the SL), $B_{w/r}$ (on the write/read side of BL)}) Fig. 4.2(b). shows the names of all 8 opens considered in this chapter.

4.3. FAULT MODELING

In this section, we will first define the fault space that is able to describe all theoretical faults and classify them. Then, the fault space validation methodology including the circuit simulation steps will be given.

4.3.1. FAULT SPACE DEFINITION AND CLASSIFICATION

A fault is an abstract functional faulty behavior description of the memory [51]. In other words, it specifies the defective behavior, comprises memory operations that can sensitize the fault, and includes the faulty logical value stored in the cell after sensitization [48]. A fault space is the defined set of all possible modeled faults that can occur in the circuit [218]. There is a systematic method, the Fault Primitive (FP) notation, to describe all faults that lead to an incorrect logical behavior [43]. Memory faults can be classified into multiple categories, based on a specific aspect [59]. Divided by the number of operations, static faults can be sensitized by applying up to one operation, while dynamic faults are sensitized by multiple operations. Divided by the number of cells, a distinction between single and coupling faults is that the latter involves more than one cell. In this thesis, we conduct the analysis of static and dynamic; single and coupling faults, which involve up to 4 RRAM cells. The FP notation is able to describe both single and coupling faults [219]. Here, we first explain single fault descriptions as an example. An FP can be denoted by a three-tuple ⟨S/F/R⟩, where:

- S (sensitizing sequence) denotes operation sequences of the cell. The form that S takes is $S = x_0 O_1 x_1 \dots O_m x_m \dots O_n x_n$, whereby $x_m \in \{0,1\}$, and $O_m \in \{w, r\}$. Here, '0' and '1' indicate the logic states of memory cells, while 'w' and 'r' denote writing and reading operations.

- F (faulty behavior) describes the stored value in the cell after a certain sensitizing S. For conventional memories (e.g., SRAM), there are only three states: '0', 'U' (the undefined state), and '1' [43]. For RRAM technology, two extreme low ('L') and extremely high ('H') conductance states are introduced. Thus, the faulty behavior in RRAM is: F ∈ {H, 1, U, 0, L}.

- R (read output) indicates the output of the read circuit. Here, R ∈ {0, 1, ?, −} where ? is a random read outcome (e.g., the sense amplifier voltage is an intermediate value due to the close sensing current and reference current); and − denotes the case where the last performed operation in the cell is not r.

The 2-cell faults can be extended and described by FP notation as $\langle S_a; S_v/F/R\rangle$. Here, S_a and S_v denote operation sequences of the aggressor cell (C_a) and the victim cell (C_v). Moreover, coupling faults can be specifically divided into three categories: 1) state coupling fault, 2) a-cell accessed coupling fault, and 3) v-cell accessed coupling fault. Note that C_a and C_v cannot perform sensitizing operations at the same time in the 1-port array. All static coupling faults that can occur in the RRAM are listed in Table 4.1 followed by the above notations. Besides, multi-cell faults involving n cells (n >2) can also be extended and denoted as $\langle S_{a_1}; \ldots; S_{a_{n-1}}; S_v/F/R\rangle$, where S_{ai} ($i \in [1, n-1]$) represents the state or sequence of the aggressor cell and S_v indicates the state or sensitizing sequence performed on the victim cell [220]. In the 2×2 memory array used in this chapter, we investigate up to 4-cell faults. Additionally, Table 4.1 can also be extended to describe dynamic faults and multi-cell coupling faults with the same logic.

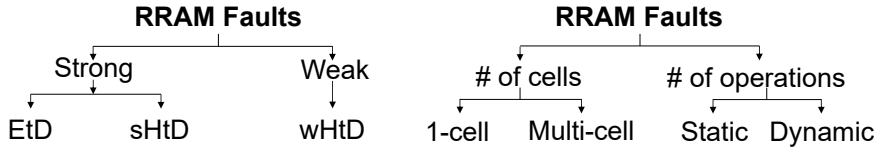


Figure 4.3: Fault space and classification.

There are several subclasses of memory faults. They can be categorized into two types based on the impact: strong (functional) and weak (parametric) faults [59], [221]. Strong faults are defined as they are guaranteed to be sensitized by applying a sequence of sensitized operations to the cell. They always cause the faulty state of the cell or read output, or both. Thus, strong faults can be denoted as FPs. Conversely, weak faults are parametric faults and do not lead to any functional errors (e.g., a voltage drop in the BL during a writing operation). The deviation needs to exceed a certain process variation specification (i.e., over 20% deviations from the defect-free case in the BL voltage), otherwise, it does not cause a weak fault. Note that weak faults cannot be denoted as FPs.

Depending on the ease of detection, faults can be further classified as 1) Easy-to-Detect (EtD) faults, and 2) Hard-to-Detect (HtD) faults. The EtD faults are guaranteed to be sensitized and detectable by regular memory operations (e.g., 1w0, 1r1) and thus detected by a March test. However, the HtD faults require additional detection solutions. Weak faults cannot be detected in a regular way, and are denoted as wHtD. Some strong faults (e.g., $\langle 1w0/U/-\rangle$) require more effort to be detected since the undefined states can cause random read outputs. These strong faults with all weak faults are called HtD faults. The HtD fault needs to be detected with more solutions, e.g., by using DfT schemes and stress tests. Fig. 4.3 shows the different levels of memory fault space and classification, based on specific aspects.

4.3.2. SIMULATION METHODOLOGY

The circuit in Fig. 2.10 is implemented in Cadence's Spectre simulator by using the Predictive Technology Model (PTM) 130-nm transistor library [222] and the RRAM compact model from [223]. The nominal supply voltage for the memory is 3 V. In order to accurately evaluate the circuit, capacitive loads are applied to BLs, SLs, and WLs in the

Table 4.1: Static coupling fault primitives.

#	Sa	Sv	F	R	Fault model	#	Sa	Sv	F	R	Fault model
1	x	0	L	-	State CF	39	x	1w1	U	-	Write Disturb CF
2	x	0	U	-		40	x	1w1	H	-	
3	x	0	1	-		41	x	0w1	L	-	Transition CF
4	x	0	H	-		42	x	0w1	0	-	
5	x	1	L	-		43	x	0w1	U	-	
6	x	1	0	-		44	x	0w1	H	-	
7	x	1	U	-		45	x	1w0	L	-	
8	x	1	H	-		46	x	1w0	U	-	
9	$xw\bar{x}$	0	L	-	Disturb CF	47	x	1w0	1	-	
10	$xw\bar{x}$	0	U	-		48	x	1w0	H	-	
11	$xw\bar{x}$	0	1	-		49	x	0r0	L	0	Deceptive Read CF
12	$xw\bar{x}$	0	H	-		50	x	0r0	U	0	
13	$xw\bar{x}$	1	L	-		51	x	0r0	1	0	
14	$xw\bar{x}$	1	0	-		52	x	0r0	H	0	
15	$xw\bar{x}$	1	U	-		53	x	1r1	L	1	
16	$xw\bar{x}$	1	H	-		54	x	1r1	0	1	
17	xwx	0	L	-		55	x	1r1	U	1	
18	xwx	0	U	-		56	x	1r1	H	1	
19	xwx	0	1	-		57	x	0r0	L	1	Read Destructive CF
20	xwx	0	H	-		58	x	0r0	U	1	
21	xwx	1	L	-		59	x	0r0	1	1	
22	xwx	1	0	-		60	x	0r0	H	1	
23	xwx	1	U	-		61	x	0r0	L	?	
24	xwx	1	H	-		62	x	0r0	U	?	
25	xrx	0	L	-		63	x	0r0	1	?	
26	xrx	0	U	-		64	x	0r0	H	?	
27	xrx	0	1	-		65	x	1r1	L	0	
28	xrx	0	H	-		66	x	1r1	0	0	
29	xrx	1	L	-		67	x	1r1	U	0	
30	xrx	1	0	-		68	x	1r1	H	0	
31	xrx	1	U	-		69	x	1r1	L	?	
32	xrx	1	H	-		70	x	1r1	0	?	
33	x	0w0	L	-	Write Disturb CF	71	x	1r1	U	?	
34	x	0w0	U	-		72	x	1r1	H	?	
35	x	0w0	1	-		73	x	0r0	0	1	Incorrect Read CF
36	x	0w0	H	-		74	x	0r0	0	?	
37	x	1w1	L	-		75	x	1r1	1	0	
38	x	1w1	0	-		76	x	1r1	1	?	

Note: x can be 0 or 1, and $\bar{x} = \sim x$

simulation. The defect-free circuit is verified for correct operations [48].

The simulation is conducted with the systematic approach, as shown in Fig. 4.4. For defect injection and circuit simulation of a 2×2 array, we consider three parameters: 1) defect strengths, 2) sensitizing sequences (S), and 3) Data-backgrounds (DBs). To perform simulation for different strengths of the defect, each resistive defect is swept from $1\,\Omega$ to $100\,M\Omega$ with 81 different defect strengths, distributed on a logarithmic scale. The applied sensitizing sequences consist of *up to three* consecutive read/write operations

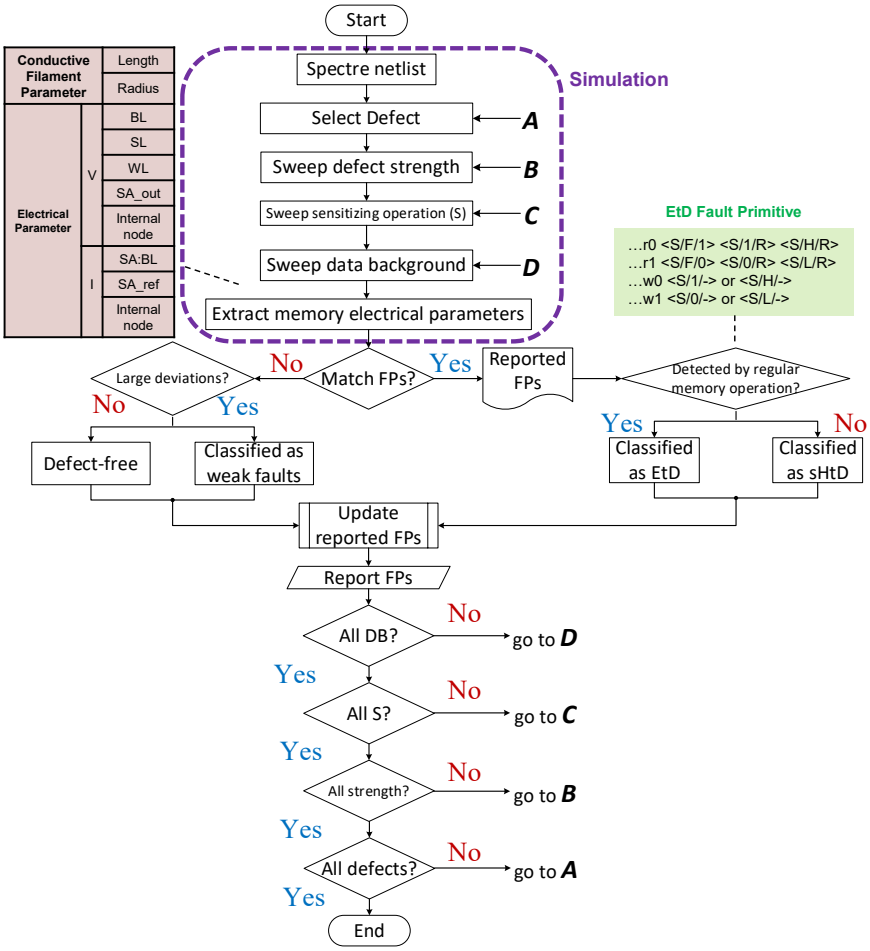


Figure 4.4: Fault space validation methodology.

(in total 80). Finally, a DB is established and defined as the pattern of ones and zeros as seen in an array of memory cells. Here, we use the most commonly known 4 DBs [40]: solid (all 0s and all 1s), row stripe (00.../11.../00.../11...), column stripe (01.../01...) and checkerboard (01.../10.../01.../10...); these are illustrated in colors in Fig. 4.5. We also represent the DB as states of 'C_d, C_c, C_r'. For example, '1, 1, 0' refers to the row stripe.

For each defect with a set strength injected in the 2×2 array, the sensitizing operations are applied, then the DB and its complement are established. After each DB and its complement, the state of each of the 2×2 cells is extracted before establishing the next DB. Once all four DBs are simulated, the *Ss* are changed and the process is repeated. Once all *Ss* are performed, the strength of the defect is changed and the process of applying *Ss* is repeated. Once all strengths are simulated, the next defect will be simulated using a similar process. This approach enables investigating the sensitization of single-

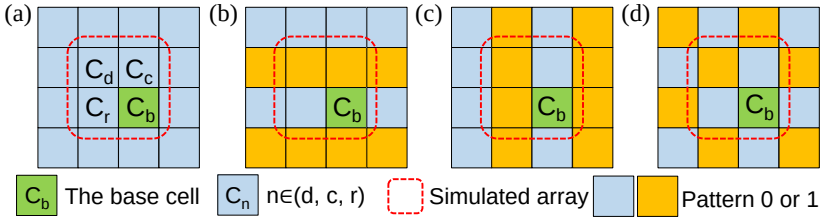


Figure 4.5: RRAM simulation setup with common DBs. (a) Solid, (b) Row stripe, (c) Column stripe, (d) Checkerboard.

Table 4.2: Fault map for dBCC1 defect. EtD sHtD Fault free

Defect: dBCC1													
Ss	DBs		Defect strength region										
	Cd	Cc Cr	1	20k	25k	40k	50k	63k	79k	159k	316k	398k	501k-100M
0w0	0	x x											
	1	x 0											
		x 1											
0w1	0	x x											
	1	x 0											
		x 1											
1w0	0	x x											
	1	x 0											
		x 1											
1w1	0	x x											
	1	x 0											
		x 1											
0r0	0	x x											
	1	x 0											
		x 1											
0w0r0	0	x x											
	1	x 0											
		x 1											
1w0r0	0	x x											
	1	x 0											
		x 1											
0r0r0	0	x x											
	1	x 0											
		x 1											

Note: $x \in \{0, 1\}$

cell faults and multi-cell faults at the same time. For example, if a fault in a cell occurs regardless of the states of other cells, then it is a single-cell fault; otherwise it is a multi-cell coupling fault. The number of involved cells determines whether the fault is a 2-cell, 3-cell, or 4-cell coupling fault.

4.3.3. FAULT MODELING RESULTS

We present the detailed validated fault results for two defects, followed by the combined results for all defects.

RESULTS FOR EXAMPLE DEFECTS (DBCC1, RBCC1)

Table 4.2 shows faults that are sensitized with sensitizing sequences, varying DBs, and defect strengths for the bridge defect dBCC1 (see Fig. 4.2(a)). In the table, we selectively list sequences due to limited space: 1) static sequences that sensitize faults, 2) dynamic sequences that sensitize faults in additional defect ranges. The DB consists of two parts: 1) the state of the C_d, 2) states of the two other neighboring cells C_c, C_r, where $x \in \{0, 1\}$. The gray shape indicates fault-free behavior, the green presents sensitized EtD faults,

and the orange sHtD faults. Note that several types of faults can be sensitized by even one defect strength and sequence. For example, both ⟨1w0;0/1/−⟩ and ⟨1w0/U/−⟩ are sensitized by 1w0 under the DB of 'C_d, C_c, C_r'='0, x, x' for the defect strength up to 159 kΩ; these details are not included in the table. If one of these is EtD, the box is green. The corresponding defect can be detected by at least sensitizing one EtD fault.

The table provides two insights for test development. First, the longest EtD range needs to be chosen for high test coverage; this is the case for Ss with a higher number of operations. E.g., a sequence 1w0r0 sensitizes more EtD faults and covers a wider defect range than 1w0 only, irrespective of the DBs. Hence, when designing a test, 1w0r0 should be selected over 1w0. Secondly, the DB does have an effect on the faulty behavior of RRAMs. For instance, the sequence 1w0r0 uniquely sensitizes EtD faults from 316 kΩ up to 398 kΩ under the DB of '1, x, 0'; i.e., the states of two neighboring cells are required. This case implies that we must incorporate the DB into the RRAM test development or risk missing EtD faults.

Table 4.3 illustrates the fault map for another defect rBCC1, containing three classes of validated faults (i.e., EtD, sHtD, wHtD).The green shape indicates at least one sensitized EtD fault, yellow only sHtD faults, and blue only wHtD faults. Several types of faults can be sensitized by even one defect strength because the accessed cell may be a victim or an aggressor. To illustrate different types of sensitized faults, the number of cells involved in faults is denoted in the table. For example, '2-HtD' presents 2-cell sHtD faults that are sensitized by corresponding defect strength and sequence. Our platform only presents a wHtD fault when no strong faults (i.e., can be described by FPs) occur with the defect strength. It is also observed that as the defect strength increases, EtD faults are not sensitized, and only sHtD faults take place (e.g., S = 0w1). For example, the static fault validation results of rBCC1 lead to unique fault classes, which depend on the defect strength. 1) The defect range rBCC1 ∈ [1 Ω, 251.2 Ω] sensitizes both EtD and sHtD faults, e.g., ⟨0w0;1/U/−⟩, ⟨0w1/0/−⟩, ⟨0r0/0/1⟩, ⟨1w0/U/−⟩, and ⟨1w1/U/−⟩. 2) If rBCC1 ∈ [316.2 Ω, 4 kΩ], except for one EtD fault ⟨0r0/0/1⟩ and one sHtD fault ⟨1w0/U/−⟩ remain, wHtD faults are sensitized by applying 0w1 and 1w0. 3) If the resistance ∈ [7.94 kΩ, 25.1 kΩ], an sHtD fault is sensitized: ⟨1w0/U/−⟩. As long as an EtD fault is sensitized, this defect at a certain strength can be easily detected by a regular test.

Table 4.3 also illustrates that as the length of the operation sequence increases, more faults are sensitized. For instance, performing an additional r0 operation after 1w0, and 0w1w0 sensitizes EtD faults with a higher defect strength range. It should be noted that the dynamic operation may operate the faulty cell back to the correct state. That means the fault activated by one operation can be masked by the subsequent operation and cannot be detected. For example, consider the row of fault map sensitized by 0w1w0 operation. From the table, there is a fault-free gap at defect strengths from 199.5 Ω to 251.2 Ω. It can be explained that a single fault ⟨0w1/0/−⟩ is first sensitized at the defect strengths, then a consecutive w0 will not write the cell to a faulty state. The dynamic operation of stimuli accumulates over time and leads to faulty behavior of the reading circuit when a longer sequence is performed. Similar results are also observed in both RRAM and STT-MRAM circuits [48], [216].

Table 4.3: Fault map for rBCC1 defect.

S	1	100	158.5	199.5	251.2	316.2	4k	5k	6.3k	7.9k	25.1k	316.2k	398.1k	1M-100M
0														
1														
0w0	2-HtD													
0w1	1-EtD,2-HtD			1-EtD		1-HtD								
0r0	1-EtD													
1w0	1,2-HtD					1-HtD								
1w1	1,2-HtD			1-HtD										
1r1														
0w0w0	2-HtD													
0w0w1	1-EtD,2-HtD			1-EtD		1-HtD								
0w0r0	1-EtD,2-HtD					1-EtD								
0w1w0	2-HtD					1-HtD								
0w1w1	1-EtD,2-HtD		1-EtD	1-HtD										
0w1r1	1-EtD,2-HtD			1-EtD		1-HtD								
0r0w0	2-HtD													
0r0w1	1-EtD,2-HtD			1-EtD		1-HtD								
0r0r0	1-EtD												2-HtD	
1w0w0	1,2-HtD					1-HtD								
1w0w1	1,2-HtD			1-HtD										
1w0r0	1-EtD,2-HtD					1-EtD							2,3-EtD	
1w1w0	1,2-HtD					1-HtD								
1w1w1	1,2-HtD			1-HtD										
1w1r1	1,2-HtD			1-HtD										
1r1w0	1,2-HtD					1-HtD								
1r1w1	1,2-HtD		1-HtD		2-HtD									
1r1r1														
0w0w0w0	2-HtD													
0w0w0w1	1-EtD,2-HtD			1-EtD		1-HtD								
0w0w0r0	1-EtD,2-HtD					1-EtD						2-EtD,3-HtD		
0w0w1w0	2-HtD					1-HtD								
0w0w1w1	1-EtD,2-HtD		1-EtD	1-HtD										
0w0w1r1	1-EtD,2-HtD			1-EtD		1-HtD								
0w0r0w0	2-HtD													
0w0r0w1	1-EtD,2-HtD			1-EtD		1-HtD								
0w0r0r0	1-EtD,2-HtD					1-EtD						3-EtD,3-HtD		
0w1w0w0	2-HtD					1-HtD								
0w1w0w1	1-EtD,2-HtD		1-EtD	1-HtD										
0w1w0r0	1-EtD,2-HtD					1-EtD							2-EtD,3-HtD	
0w1w1w0	2-HtD					1-HtD								
0w1w1w1	1-EtD,2-HtD		1-EtD	1-HtD										
0w1w1r1	1-EtD,2-HtD		1-EtD	1-HtD										
0w1r1w0	2-HtD					1-HtD								
0w1r1w1	1-EtD,2-HtD		1-EtD	1-HtD										
0w1r1r1	1-EtD,2-HtD			1-EtD		1-HtD								
0r0w0w0	2-HtD													
0r0w0w1	1-EtD,2-HtD			1-EtD		1-HtD								
0r0w0r0	1-EtD,2-HtD					1-EtD							3-HtD	
0r0w1w0	2-HtD					1-HtD								
0r0w1w1	1-EtD,2-HtD		1-EtD	1-HtD										
0r0w1r1	1-EtD,2-HtD			1-EtD		1-HtD								
0r0r0w0	2-HtD													
0r0r0w1	1-EtD,2-HtD			1-EtD		1-HtD								
0r0r0r0	1-EtD													
1w0w0w0	1,2-HtD				1-HtD									
1w0w0w1	1,2-HtD			1-HtD										
1w0w0r0	1-EtD,2-HtD					1-EtD								
1w0w1w0	1,2-HtD					1-HtD								
1w0w1w1	1,2-HtD			1-HtD										
1w0w1r1	1,2-HtD			1-HtD										
1w0r0w0	1,2-HtD					1-HtD								
1w0r0w1	1,2-HtD			1-HtD										
1w0r0r0	1-EtD,2-HtD					1-EtD								
1w1w0w0	1,2-HtD					1-HtD								
1w1w0w1	1,2-HtD			1-HtD										
1w1w0r0	1-EtD,2-HtD					1-EtD							2-EtD,3-HtD	
1w1w1w0	1,2-HtD					1-HtD								
1w1w1w1	1,2-HtD			1-HtD										
1w1w1r1	1,2-HtD			1-HtD										
1w1r1w0	1,2-HtD					1-HtD								
1w1r1w1	1,2-HtD			1-HtD										
1w1r1r1	1,2-HtD			1-HtD										
1r1w0w0	1,2-HtD					1-HtD								
1r1w0w1	1,2-HtD			1-HtD										
1r1w0r0	1-EtD,2-HtD					1-EtD							2-EtD,3-HtD	
1r1w1w0	1,2-HtD					1-HtD								
1r1w1w1	1,2-HtD			1-HtD										
1r1w1r1	1,2-HtD		1-HtD											
1r1r1w0	1,2-HtD					1-HtD								
1r1r1w1	1,2-HtD		1-HtD		2,3-HtD									
1r1r1r1														

EtD sHtD wHtD

OVERALL RESULT OVERVIEW

Fig. 4.6 gives the relative number of sensitized static faults for all simulated defects; *the number of faults* for each sequence is defined as *the number of defect strengths that sensitize faults*. For example, the number of (single static) faults sensitized by 1w0 for defect dBCC1 (see Table 4.2) is 2; note that there are faults, which are DB independent taking place for defect sizes of up to 20 kΩ. The total number of faults (single-cell) sensitized by 1w0 (when considering all defects) counts for 83% as shown in the orange bar of Fig. 4.6. Note that Fig. 4.6 is normalized to 1w0 because this sequence is the one sensitizing the maximum number of faults. This is because the reset process in an RRAM device is a negative feedback loop that is susceptible to defect-induced variations [74], [223]. Fig. 4.6 shows that both single-cell and multi-cell faults are sensitized. The number of single-cell and two-cell faults accounts for the majority (99%) of the total. However, a small number of 3-cell and 4-cell faults are observed when applying read operations, showing the importance of sensitizing multi-cell static faults if high product quality is targeted. The detailed analysis will be provided in the following.

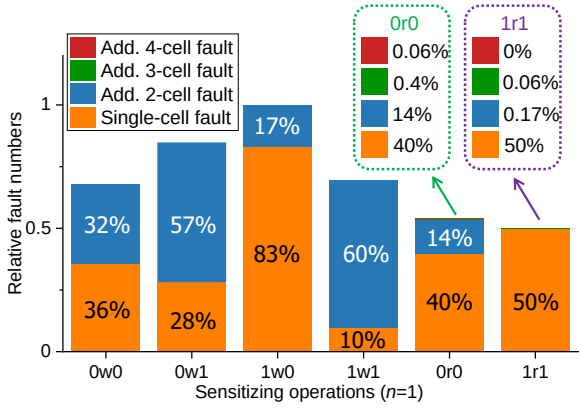


Figure 4.6: Relative number of faults sensitized by static sequences.

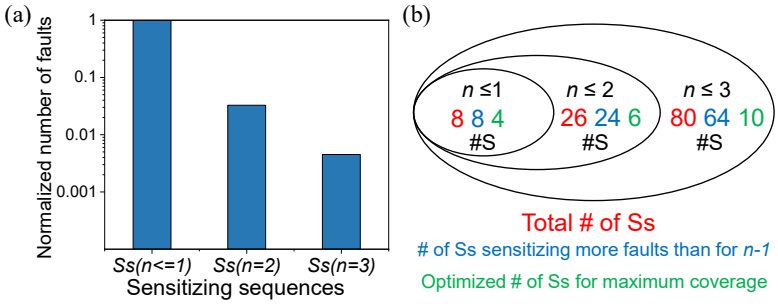


Figure 4.7: Dynamic EtD faults sensitized as n_{max} increases. (a) The normalized number of faults, (b) The number of sequences.

Fig. 4.7(a) shows the number of 2-operation and 3-operation sensitized dynamic EtD

faults as compared to static faults. The total number of *Ss* is: $\#S = \sum_{n=0}^{3} 2 \times 3^n = 80$, consisting of 8 1-operation *Ss*, 18 2-operation *Ss*, and 54 3-operation *Ss* [48]. The results of Fig. 4.7(a) are normalized to the total number of faults sensitized by the static analysis. As already mentioned, the number of faults is defined as the *number of defect strengths* that sensitize faults. The figure shows that although the total number of *Ss* increases exponentially, the expansion of defect strengths sensitizing new faults slows down. Clearly, increasing the number of operations (n_{max}) per *S* contributes to the sensitization of more faults/defect strengths.

In Fig. 4.7(b), three ellipses present faults sensitized by sequences with different lengths given by n. In each ellipse, there are three numbers from left to right: 1) the total number of sequences for the corresponding n (red), 2) the number of sequences for this n that can sensitize faults which were not sensitized for the case $\leq n-1$ (blue), and 3) the minimum numbers of sequences that need to be applied in order to cover the maximum defect coverage (green) for considered n. For example, for $n \leq 2$, there are 26 possible *Ss* (from which 8 are applied also for $n \leq 1$), only 16 *Ss* (when $n = 2$) can sensitize additional faults which were not observed for $n \leq 1$, and only 6 *Ss* are needed to maximum defect coverage; this optimization will be explained for the test development in Section 4.4. Fig. 4.7(b) indicates that not all dynamic sequences sensitize additional faults than static sequences, hence only a subset is needed to detect all defects. E.g., for $n \leq 3$, only 4 additional *Ss* are needed as compared with $n \leq 2$ to sensitize all validated faults.

ROOT AND ANALYSIS OF 3-CELL AND 4-CELL FAULTS

Table 4.2 shows the 3-cell fault takes place in the presence of dBCC1 for *S*= 1w0r0 and DB of '1, x, 0'. For the defect-free circuit, when performing a 1w0r0 operation, read current flows from BL1, into int3, and from SL1 to ground. For the defective circuit, extra read current also flows from BL1, through the defect, to int0, BL0, int2 (WL1 is active), and SL1 to ground. The current induced by the defect flows through neighboring cells (C_d and C_r) of the accessed cell C_b, which affects the discharge speed of the sensed node, leading to an incorrect read fault. The magnitude of the current depends on the resistance of the defect and the states of C_d and C_r. When the defect resistance is low, a read fault will occur that does not depend on the DB. However, with increasing defect resistance, the resistance of the neighboring cells becomes essential, and thus the DB starts to play a role in fault sensitizing. Hence, the 3-cell fault is sensitized. Furthermore, 4-cell faults can also be sensitized. The gate-drain capacitance of the transistor consumes a slight current. Hence, even though a particular cell (e.g., C_c) is neither in the path induced by defect nor is its WL conducting, the cell state has an impact on the magnitude of the charging current. Besides, the current induced by the defect still exists. In these special cases, the read current flows through all 4 cells and is affected by their states, leading to an incorrect 4-cell read fault.

4.4. TEST DEVELOPMENT

This section uses a specific test generation method to develop a test solution for targeted faults in this chapter. The test is validated and thereafter compared with prior work.

4.4.1. DATA BACKGROUND-AWARE TEST APPROACH GENERATION

To detect both single and multi-cell faults, a test solution needs to take the Data Background (DB) into account. There are two requirements: 1) to fully cover defect strengths that sensitize EtD faults, and 2) to minimize the test length. In [48], the test development is formulated as an Integer Linear Programming (ILP) problem that minimizes the number of applied sequences while maximizing the defect coverage. We apply a similar method. In addition, we aim at minimizing the number of DBs during the test development, as this makes the test implementation easier and even results in a shorter test length. Hence, a DB-ILP method is proposed.

Table 4.4: Example to solve the DB-ILP.

			DB_b, for $b:1\rightarrow B$							$\sum_{b=1}^{B}\sum_{z=1}^{Z} a_{d,r,z,b}$
			DB_1			$\cdots$			DB_B	
			$S_{z,b}$ for $z:1\rightarrow Z$			$\cdots$	$S_{z,b}$ for $z:1\rightarrow Z$			
			$S_{1,1}$	$\cdots$	$S_{Z,1}$	$\cdots$	$S_{1,B}$	$\cdots$	$S_{Z,B}$	
D_1	DS_1	1	1	$\cdots$	1	$\cdots$	0	$\cdots$	0	2
	DS_2	10	1	$\cdots$	0	$\cdots$	0	$\cdots$	1	2
						$\cdots$				
	$\cdots$	10M	1	$\cdots$	0	$\cdots$	0	$\cdots$	1	2
	DS_R	100M	0	$\cdots$	0	$\cdots$	0	$\cdots$	0	0
						$\cdots$				
D_D	DS_1	1	0	$\cdots$	0	$\cdots$	0	$\cdots$	1	1
	$\cdots$	10	1	$\cdots$	0	$\cdots$	0	$\cdots$	0	1
						$\cdots$				
	$\cdots$	10M	1	$\cdots$	0	$\cdots$	0	$\cdots$	0	1
	DS_R	100M	0	$\cdots$	0	$\cdots$	0	$\cdots$	1	1

We illustrate the use of DB-ILP approach in Table 4.4. Assume a four-dimensional binary matrix $a_{d,r,z,b} \in \mathbf{A}$, where '$d$' denotes the defects ($D_d$) from 1 to D, 'r' denotes the defect strengths (DS_r) for each defect from 1 to R, 'z' denotes the sequence ($S_{z,b}$) from 1 to Z, and 'b' denotes DBs (DB_b) from 1 to B. The total number of elements in $\mathbf{A}$ is $D \times R \times Z \times B$. Then, if at least one EtD fault is sensitized for defect strength r of the defect d with the corresponding sequence $S_{z,b}$ of the DB 'b', this element ($a_{d,r,z,b}$) in the matrix is set as '1', otherwise it is set to '0'. For example, when mapping to a single row of Table 4.4, the 'green' entities in Table 4.2 (for defect dBCC1) correspond to 1's, while the others correspond to 0's. The last column lists the sum of elements in every row, indicating the number of sequences that can sensitize a fault in the presence of a defect 'd' with a defect strength 'r'. This number is always greater than or equal to 1 as at least one $S_{z,b}$ should sensitize an EtD fault in the presence of D_d. For example, the sum of the elements of row D_D-DS_R is '1', hence we have to choose sequence $S_{Z,B}$ with DB_B for the test development since D_D-DS_R can be only detected by this single sensitizing operation. Now, the DB-ILP optimization can be mathematically denoted as:

$$\min \sum_{b=1}^{B} \left(\beta \cdot \mathrm{DB(sel)}_b \cdot \sum_{z=1}^{Z} \mathrm{S(sel)}_{z,b} \right)$$

$$\text{s.t.} \begin{cases} \text{For } (d: 1\ to\ D, r: 1\ to\ R): \\ \quad \text{if } \sum_{b=1}^{B} \sum_{z=1}^{Z} a_{d,r,z,b} \geq 1: \\ \quad\quad \sum_{b=1}^{B} \sum_{z=1}^{Z} a_{d,r,z,b} \cdot \mathrm{DB(sel)}_b \cdot \mathrm{S(sel)}_{z,b} \geq 1. \end{cases}$$

Here, $\mathrm{DB(sel)}_b$ and $\mathrm{S(sel)}_{z,b}$ are binary values, indicating whether the bth DB (DB_b) and the zth sensitizing sequence ($\mathrm{S}_{z,b}$) are selected (i.e., '1' is selected), meaning that $\mathrm{S}_{z,b}$ sensitizes an EtD fault. β is a parameter used to give higher weight/cost for changing DBs as compared with changing $\mathrm{S}_{z,b}$. We aim at having fewer DBs selected and set β=80, which is the maximum number of sequences for n=3 for a single DB. The minimization statement guarantees that we get the minimal number of DBs and the associated minimum number of $\mathrm{S}_{z,b}$. The constraints ensure that all sensitized defect strengths are covered.

We apply Python3's PuLP optimization package to solve the above DB-ILP problem for EtD faults [224]. The output provides the required minimum number of DBs and their associated sensitizing sequences $\mathrm{S}_{z,b}$ needed to sensitize all targeted EtD faults. Note that multiple solutions may exist; they all have the same cost (length). In our case, a minimum of three DBs, each with associated sensitizing sequences, are needed; they are:

DB1 = $all\ 0's$ for 1w1, 1r1, 1w1r1r1, 1w1w1w1
DB2 = $all\ 1's$ for 1w0, 0r0r0, 1w0r0, 0w0w0w0, 1r1w0r0
DB3 = 0000/1111/0000/1111/... for 0r0, 1w0r0

Note that DB1=(000)=(C_d, C_c, C_r) and DB2 represent solid DBs (denoted as S and $\overline{\mathrm{S}}$), while DB3 represents the row-stripe DB (denoted as R), see Fig. 4.5. Sequences with the same DB can be further combined to optimize the test cost. An additional read operation should be added after each sensitizing sequence to ensure the detection of the fault. Furthermore, 1w1 and 1w1r1r1 can be combined into 1w1r1r1. In this way, the March test algorithm (referred to as March-EtD) to detect all EtD faults is generated as follows:

$$\Big\{ \Updownarrow (\mathrm{wS}); \Updownarrow \left(\mathrm{w}\overline{\mathrm{S}}, \mathrm{w}\overline{\mathrm{S}}, \mathrm{w}\overline{\mathrm{S}}, \mathrm{w}\overline{\mathrm{S}}, \mathrm{r}\overline{\mathrm{S}}, \mathrm{r}\overline{\mathrm{S}}, \mathrm{r}\overline{\mathrm{S}}, \mathrm{wS} \right);$$
$$\Updownarrow \left(\mathrm{w}\overline{\mathrm{S}} \right); \Updownarrow \left(\mathrm{r}\overline{\mathrm{S}}, \mathrm{wS}, \mathrm{rS}, \mathrm{rS}, \mathrm{rS}, \mathrm{wS}, \mathrm{wS}, \mathrm{wS}, \mathrm{rS}, \mathrm{w}\overline{\mathrm{S}} \right);$$
$$\Updownarrow (\mathrm{wR}); \Updownarrow \left(\mathrm{w}\overline{\mathrm{R}}_{\mathrm{e}}, \mathrm{wR}_{\mathrm{e}}, \mathrm{rR}_{\mathrm{e}}, \mathrm{rR}_{\mathrm{e}} \right);$$
$$\Updownarrow (\mathrm{wR}); \Updownarrow \left(\mathrm{w}\overline{\mathrm{R}}_{\mathrm{u}}, \mathrm{wR}_{\mathrm{u}}, \mathrm{rR}_{\mathrm{u}}, \mathrm{rR}_{\mathrm{u}} \right) \Big\}.$$

The test has 8 march elements and uses march notation [40]. The first march element M_1 uses '⇕' addressing (indicates that addressing direction is irrelevant) to initialize the memory to S = solid DB = all 0's. M_2 ensures the application of all sensitizing sequences associated with S, and adds a read operation after sequences to guarantee the detection of faults, irrespective of whether they are destructive or deceptive [43]. M_3 and M_4 apply associted sensitizing sequences but then for $\overline{\mathrm{S}}$ = solid DB = all 1's. M_5 applies R = row stripe = 0000/1111/0000/1111/... to the memory. M_6 applies the associated sensitizing sequences to *even* rows consisting of 0's (e.g., wR_{e}) and 1's (e.g., $\mathrm{w}\overline{\mathrm{R}}_{\mathrm{e}}$); a read operation

is also applied after each sequence. M_7 writes the same R to the memory and M_8 does the same as M_6, but then for *uneven* rows. The test length of March-EtD is $15N_w$+$10N_r$, where N_w/N_r indicates the number of writes and reads, respectively. In conclusion, the March-EtD algorithm targets both single-cell and multi-cell faults, covering static as well as dynamic faults.

4.4.2. VALIDATION AND COMPARISON WITH EXISTING TESTS

Next, we further develop the platform and perform the march element sequences shown above to verify the correctness of the proposed March-EtD. First, we control and generate the timing voltage pulses applied to cells in the order of the addresses required by the test algorithm. Then, we apply the sequences to the defect-free circuit and check whether all the read operations output correctly. Finally, the March test sequences are performed to repeat the simulation using a similar approach shown in Section 4.3 by sweeping all injected defects with various strength ranges. The defective cells may lead to an incorrect readout when the defect is detectable by the applied March-EtD algorithm.

By applying the test elements in March-EtD designed order, each RRAM cell generates a 10-bit signature from the algorithm's 10 read operations. The 10-bit signature is extracted to verify the approach. When the March-EtD is performed to the defect-free circuit, the 10-bit sequence is expected to be '1110000000'. However, the extracted sequences from some defective circuits do not correspond to the defect-free circuit's read output sequence. For some defect ranges, show the same sequence of test results as the defect-free circuit because these defect ranges cannot sensitize EtD faults. Still taking the defect rBCC1 as an example, the March test is implemented on a defective circuit to detect the rBCC1 based on whether the readout is correct. Table 4.5 illustrates the detected rBCC1 defect strength range distribution. The green color indicates defect ranges that are detected. The maximum strength range of detected rBCC1 is consistent with its length that can sensitize at least one EtD fault, as shown in Table 4.3. The verification results show that all interconnect and contact defect strength ranges that can sensitize EtD faults are able to be detected by applying our March-EtD algorithm without test escapes.

Table 4.5: Test algorithm validation map for rBCC1 defect.

	1	10k	316k	398k	10M	100M
Cd						
Cc						
Cr						
Cb						

Finally, we compare the proposed March algorithm with existing RRAM test solutions that have been provided in literature [49], [54], [225]–[227]. The comparison is based on the validated test escape rate and test time. We use the same simulation setup as in Section 4.3. Each defect is still simulated with 81 strengths, logarithmically spaced. For each defect, the missing numbers of defect strengths are counted and divided by the total number of EtD defect strengths. The comparison result between March tests is summarized in Table 4.5. It can be concluded that all other March tests have test escapes. For example, defect dBCC1 ranging from 316 kΩ to 398 kΩ cannot be detected,

Table 4.6: Test comparison with existing RRAM tests.

Year	Name	Test escapes of EtD Defects	Test time	
			Write	Read
2013	March-MOM [225]	4.600%	5N	4N
2015	March-1T1R [54]	0.366%	(1+2a+2b)N	5N
2015	March C* [49]	0.653%	4N	6N
2016	March C*-1T1R [226]	0.340%	6N	6N
2017	March W-1T1R [227]	0.392%	9N	8N
2023	**Proposed (March-EtD)**	**0%**	**15N**	**10N**

although EtD faults are sensitized (see Table 4.2). Note that even a small number of test escapes is problematic towards meeting Defective Part Per Billion (DPPB)-level requirements for RRAMs. The table shows the fact that March-ETD targets and detects more faults than existing algorithms; they are new faults found in this work, such as multi-cell faults, which were not reported before. Besides, these existing test solutions may sensitize unrealistic faults as overkill, and lead to yield loss. The validation of DfT schemes for HtD faults is not included in this chapter.

5

Device-Aware Test for Ion Depletion (ID) Defects

The previous chapter applies linear resistors to model interconnect and contact defects; fault models and test solutions are proposed while considering the impact of data backgrounds. This chapter presents the identification and characterization of a novel defect, called Ion Depletion (ID), utilizing silicon measurements. Our results show that 14 % *cycles of a defective device have intermittent decreases in the high resistance state and do not affect the low resistance state. Furthermore, this unique defect is non-linear and takes place inside the device itself; hence, the conventional defect and fault modeling, which rely on linear resistors beside the device, cannot be applied to model the ID defect. To tackle this issue, the Device-Aware (DA) defect modeling approach is used. A calibrated RRAM model representing the defective device is developed using measurement data to characterize precisely how the defect influences the electrical properties of the memory hardware. Subsequently, fault analysis is conducted based on the DA defect model, leading to the introduction of suitable fault models. These models demonstrate that the ID defect can trigger undefined state faults. Finally, dedicated test solutions specifically designed for the ID defect are proposed.*

The contents of this chapter were published in ITC'23 [65].

5.1. INTRODUCTION

Many works have investigated defect modeling and test generation for RRAMs. In 2009, Ginez, Portal, and Muller studied coupling faults by modeling bridge defects [52]. In 2012, Haron and Hamdioui proposed a special Design-for-Testability (DfT) to detect undefined state faults [57]. In 2013, sneak-path testing for RRAMs was presented to reduce the test time [53]. In 2015, Chen and Li reported a dynamic write disturbance fault, and a March test to cover it [54]. In 2021, Liu, You, Wu, *et al.* developed DfT scheme for 3D hybrid RRAM array [56]. Nevertheless, all these works employ an *linear resistor* to model defects. Although this model may be good enough for interconnects and contacts, it is certainly not accurate enough to describe unique defects in the RRAM device itself, as the device is *non-linear* by nature. To address the limitations of the conventional RRAM test method, the 'Device-Aware Test (DAT)' approach was proposed [58]–[60]. The DAT method properly models physical defects, enabling the exploration of realistic fault models and the development of high-quality test solutions. This has been shown to be very powerful for RRAM unique defects such as Intermittent Undefined State Fault (IUSF) [214]. Moreover, as RRAM is an immature technology, more unique defects induced by device miniaturization or new materials, are still to be discovered, understood, and modeled in order to develop optimal test solutions.

This chapter investigates and characterizes (based on silicon measurements) a new unique RRAM defect, referred to as an Ion Depletion (ID) defect. An ID defect occurs when insufficient oxygen ions stored near the capping layer/oxide interface can be released during the RESET process, which affects the on/off resistance ratio and leads to undefined state faults. The DAT approach is applied to accurately model the ID defect, derive realistic fault models, and thereafter develop test solutions. Based on measurements, the DAT defect model has been integrated with *intermittent behavior*. Moreover, the proposed test solution has been validated with process variation analysis. The primary contributions of this chapter are as follows:

- Identify the ID defect based on silicon measurements.
- Comprehensive characterization of RRAM devices exhibiting the ID defect, including an in-depth analysis of the underlying physics and its reliability issues across multiple cycles and devices.
- Develop the DA defect model for the ID defect, which incorporates the intermittent behavior, and calibrate it with measurements to accurately model this defect.
- Utilize the DA defect model to develop appropriate fault models and subsequently optimal test solutions, validated through simulations. The proposed solutions are optimized for robustness against process variations, ensuring practical applicability and reliability.

5.2. LIMITATIONS OF CONVENTIONAL DEFECT MODELS

The linear resistors are used in RRAM defect modeling by connecting them in either series or parallel with the device, as shown in Fig. 5.1. Varying resistance values are used in each of these linear defects to depict the dimensions and strength of the physical defect.

When series ohmic resistance increases or parallel ohmic resistance decreases, it means that there are stronger defects and the faulty behavior is more severe. Fig. 5.1(a) and Fig. 5.1(b) depict the behavior of the real defective device as well as the behavior of the model of the defective device based on linear resistors under different defect strengths. The red line represents the measurement of a unique defect, the Ion Depletion (ID)-defective device. The electrical characteristics of the ID-defective RRAM device, such as R_{SET}, R_{RESET}, V_{SET}, and V_{RESET}, are clearly influenced by different strengths of resistive defects. The parallel resistor may also cause the same behavior of R_{RESET} degradation as the ID-defective device (see the green dotted line in Fig. 5.1(b)). However, it is unable to be fitted with other electrical parameters such as V_{RESET} and R_{SET}; hence, it cannot accurately model the ID defect. If the linear resistor defect significantly affects the device, it may potentially result in the failure to transition between HRS and LRS states. Clearly, none of these two linear defect models can accurately describe the ID. Therefore, a more comprehensive defect modeling approach is necessary for accurately capturing the state drift in the RRAM RESET performance.

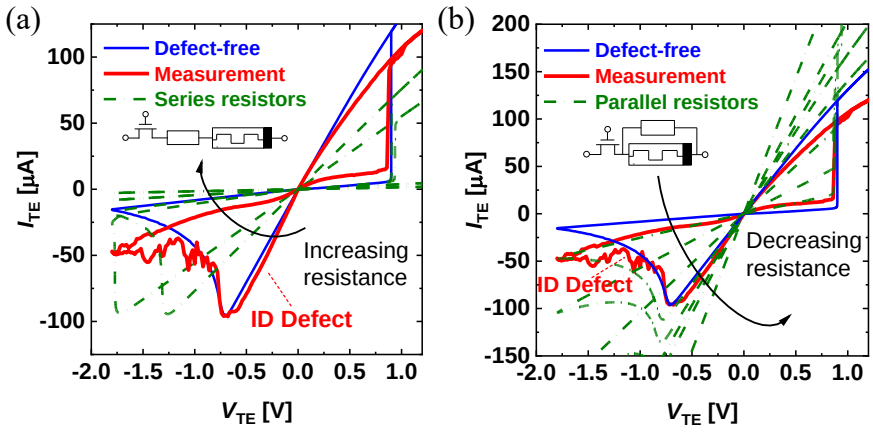


Figure 5.1: Linear resistor defect models. (a) Series resistors, (b) Parallel resistors.

5.3. DEVICE-AWARE TEST APPPROACH

Due to the distinct working principles and structural characteristics of RRAMs, unique defects such as the ID arise that differ from conventional defects. Unlike conventional defects, which can typically be modeled as simple resistive shorts or opens, these RRAM-specific defects exhibit non-linear and stochastic behaviors, making them challenging to model using standard resistors [58], [59]. Conventional defect modeling approach fails to accurately capture these complex fault characteristics, resulting in ineffective fault models and reduced test coverage, as can be seen in Fig. 5.1.

To overcome this limitation, the Device-Aware Test (DAT) methodology has been introduced. This approach integrates a deep understanding of the physical mechanisms governing defective RRAMs into the test development process, enabling more precise defect detection while optimizing test efficiency. The DAT framework follows a struc-

tured four-step process: defect characterization, defect modeling, fault modeling, and test development, ensuring a systematic and high-coverage testing strategy tailored to RRAM-based devices. Fig. 5.2 shows the DAT process. Next, we will explain each step of the DAT approach.

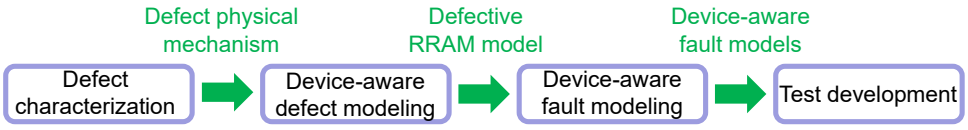


Figure 5.2: Process of DAT approach.

DEFECT CHARACTERIZATION

The first stage in the testing process is defect characterization, which involves analyzing the physical mechanisms underlying RRAM-specific defects [58]. This is conducted through a range of experimental techniques, including electrical measurements (such as current-voltage (I-V) analysis, write endurance tests, and resistance drift evaluations) and material inspections using methods like Transmission Electron Microscopy (TEM) and Scanning Electron Microscopy (SEM).

The primary objective of defect characterization is to determine how these unique defects influence the electrical properties and switching behavior of RRAM devices. By examining experimental data, researchers can identify defect mechanisms and establish common characteristics associated with specific defect types, such as oxygen vacancies and CF irregularities [59].

The insights gained from this analysis serve as the foundation for constructing accurate defect models, ensuring that the impact of these RRAM-specific defects is effectively captured in subsequent fault modeling and test development stages.

DEVICE-AWARE DEFECT MODELING

The second stage in the process is defect modeling, where the physical effects of RRAM-specific defects are incorporated into compact device models [60]. Unlike conventional approaches that simply model defects as resistive shorts or opens, device-aware modeling integrates the underlying physical mechanisms, enabling the development of more accurate defective RRAM models.

To achieve accurate defect modeling, insights from defect characterization are used to modify parameters in the original RRAM model, or introduce additional factors that influence its electrical behavior. For example, an OV-related defect, which occurs due to imperfections in the switching layer, can be modeled by introducing a parameter that quantifies the vacancy concentration (i.e., defect strength).

By incorporating these physical defect characteristics into compact RRAM models, circuit-level simulations can more accurately predict how faulty devices behave under standard read/write conditions. These optimized models form the basis for fault modeling, ensuring that simulated faults closely reflect realistic faulty behaviors and align with experimental measurement data from defective RRAM cells.

DEVICE-AWARE FAULT MODELING

The third phase, fault modeling, extends the analysis of defective RRAM behavior at the circuit level within a memory array. While similar to conventional fault modeling, this approach is refined to account for the nonlinear and stochastic characteristics of RRAM defects.

At this stage, identified defects are introduced into Spectre simulations by replacing the defect-free RRAM model with a defective version derived from the previous defect modeling step. These circuit-level simulations evaluate how defects influence read/write operations, reliability, and performance variations. The observed faulty behaviors are then systematically categorized using fault models to ensure an accurate representation of defects.

DEVICE-AWARE TEST DEVELOPMENT

The final stage in the testing process is test generation, where optimized strategies are developed to efficiently identify the faults characterized in previous steps [59]. Within the DAT framework, test patterns must be carefully designed to achieve high defect coverage while maintaining low test time and cost.

Traditional read/write-based tests, such as March algorithms, can still be applied to RRAM. For example, a March test incorporating a single 1r1 operation may detect certain pinhole defects. However, due to variations in defect severity, this approach has a high escape rate when testing for intermittent faults [214].

To enhance defect detection, additional techniques such as Design-for-Testability (DfT) with specialized circuits can be introduced. However, these solutions come with trade-offs, including increased circuit complexity and higher manufacturing costs.

To strike a balance between test efficiency and cost-effectiveness, adaptive test strategies must be employed, adjusting to the specific application requirements of RRAM devices.

Next, we will apply this DAT approach to the identified unique defects.

5.4. DEFECT CHARACTERIZATION

The electrical characteristics of a 7×7 1T-1R array, consisting of 49 BS RRAM devices, were measured across 936 RESET-SET cycles on a single wafer [214]. The device is HfO_2 based 1T-1R array fabricated at ST Microelectronics 130 nm technology. $(BE/oxide/cap/TE)$ =(TiN/10 nmHfO_2/10 nmTi/TiN) is the stack of the bipolar device. One ST Microelectronics high-voltage thick oxide transistor (W=0.8 µm, L=0.5 µm) is connected in series to regulate the current flowing through the device. The Keysight B1500 semiconductor parameter analyzer is connected to the measured 1T-1R device by a probe card, which is embedded in an 8-inch wafer. A 1 ms DC staircase voltage sweep with a 20 mV step is applied across BL and SL of the measured device for the measurement configuration, and the current flowing through the device is measured. V_{SET}, V_{RESET}, R_{SET}, and R_{RESET} are four critical electrical parameters that demonstrate the device's switching performance. From the measurement data, a device with a nominally defect-free switching is bipolar, where LRS is defined with R_{SET}<20 kΩ, and HRS with 100 kΩ<R_{RESET}<1 MΩ. The remaining range [20 kΩ, 100 kΩ] is defined as an undefined state ('U').

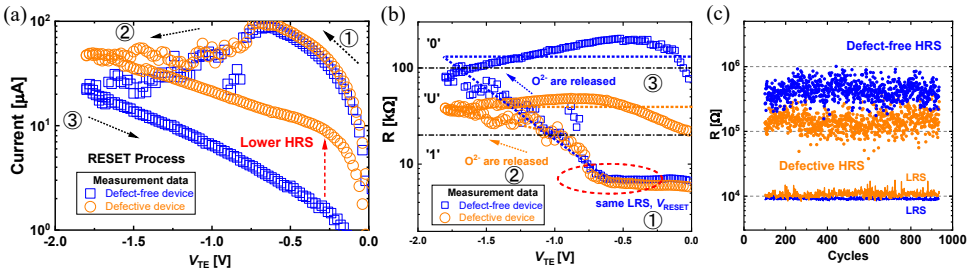


Figure 5.3: Comparing defect-free and defective devices. (a) Measured I-V curve in the logarithmic y-axis, (b) Measured R-V curve in the logarithmic y-axis, (c) Measured R_{RESET} and R_{SET} in multiple (836) cycles.

Upon examining the measurements, it is evident that certain devices have a defective RESET process. Specifically, their resistance states remain at 'U' regardless of the magnitude of the RESET voltage amplitude. The ID defect is characterized by comparing the typical measured 1T-1R log(I)-V and log(R)-V curves of defect-free and defective devices in the RESET process, as illustrated in Fig. 5.3(a) and Fig. 5.3(b). V_{TE} is applied in a sweeping motion from 0 V to −1.8 V (①, ②) and back to 0 V (③). The gradual RESET process with a significant current fluctuation is observed from our measurement, which can be explained by the thermodynamic process of the migration of Vo [99]. The RESET switch in the defect-free device starts at −0.7 V with a constant LRS and terminates at a very low current, approximately 20 μA. An evident transition is also observed at −0.7 V during the RESET switching of a defective device. Nevertheless, this transition fails at ② and remains in the 'U' state. The current through the device gradually decreases as the RESET voltage decreases further, but it remains greater than the typical RESET current (approximately 50 μA). The HRS of both devices remains nearly constant (③) as the applied negative voltage returns to 0 V. The following SET processes of defective devices are consistent with the defect-free devices.

Table 5.1: Occurrence probability of ID (P_{ID}) in % and maximal duration suffering from ID (D_{max}), during 536 cycles.

		WL 0	WL 1	WL 2	WL 3	WL 4	WL 5	WL 6
BL 0	P_{ID}	4.41	0	0	0.74	15.44	0	0.74
	D_{max}	1	0	0	1	3	0	1
BL 1	P_{ID}	0	1.47	0	1.47	1.47	0.74	0
	D_{max}	0	1	0	1	1	1	0
BL 2	P_{ID}	0	0	0	0	0	1.47	0
	D_{max}	0	0	0	0	0	1	0
BL 3	P_{ID}	0	9.56	0	0	13.24	11.03	20.59
	D_{max}	0	2	0	0	4	1	3
BL 4	P_{ID}	0	0.74	0	0	0.74	0.74	8.82
	D_{max}	0	1	0	0	1	1	1
BL 5	P_{ID}	0	2.94	2.94	0.74	0	18.38	5.15
	D_{max}	0	1	1	1	0	3	1
BL 6	P_{ID}	4.41	0.74	0	0	0	0	0.74
	D_{max}	1	1	0	0	0	0	1

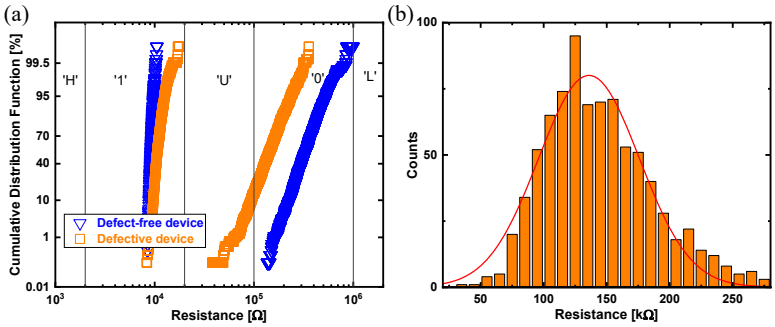


Figure 5.4: Resistance distribution. (a) Measured R_{RESET} and R_{SET} of defect-free and defective devices, (b) Histogram of measured R_{RESET} distribution.

Next, we characterize the ID defect in multiple RESET-SET cycles. In the initial cycles, devices exhibit erratic resistance states, a phenomenon also reported in [228]. Thus, we skip the initial cycles to obtain stable measurements. Table 5.1 summarizes the occurrence probability (P_{ID}) of the ID defect in the measured devices (using the WL and BL to indicate them). The table lists the percentage of cycles in which ID occurs given a total of 536 cycles after stabilization. Besides, the table counts the maximal duration (D_{max}) of the undesired 'U' state, expressing the maximal cycle the ID defect can last in 536 cycles for each device. For example, if the ID first occurs in cycles 100, 101, and 102, and later in cycle 110, then P_{ID}=4/536=0.74 % and D_{max}=3 cycles. The table indicates that the ID defect is present in 25 devices, with a maximum duration of 4 cycles and a probability of up to 20.59 % of the cycles. Additionally, the on/off resistance window is illustrated in Fig. 5.3(c) as a function of cycles for defect-free (WL=5, BL=0) and defective devices (WL=5, BL=3). This defective device only exhibits erratic resistance states in the first 100 cycles and is selected to analyze the ID defect in this paper. Following the initial few cycles, the resistance values exhibit minimal fluctuations. Fig. 5.4(a) shows the cumulative distributions of resistance for the defect-free and defective RRAM devices. The HRS distribution exhibits a more pronounced spread compared to the LRS, a finding that has also been reported in other papers [229]. Fig. 5.4(b) shows the histogram of defective measured R_{RESET} distribution, which follows the normal distribution. From Fig. 5.4, 14 % of RESET-SET cycles demonstrate the ID-defective HRS degradation. Hence, the defect displays *intermittent* behavior, which poses a significant concern for RRAM devices. Fig. 5.5 shows the correlation between the absolute value of R_{RESET} deviation and the increase in the number of cycles, plotted on a logarithmic scale. Again, we skip the first 100 cycles. The Person correlation coefficient of our 835 R_{RESET} deviation data is 0.162 (a Person correlation of 0 indicates no correlation) as a weak positive correlation coefficient. The p-value is 2.4×10^{-6}<0.05, indicating statistical significance. The red dotted line shows that the deviation increases as the cycles increase. This correlation indicates that the ID defect is not only in time 0 due to the imperfect fabrication, but also becomes stronger as the cycles increase. Hence, the ID defect probably raises concerns about reliability since the ID-defective device wears out and is irrecoverable.

Since both defective and defect-free devices show proper SET switching in the mea-

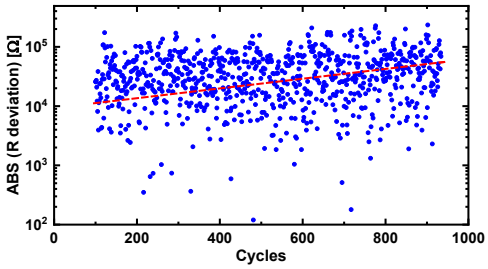


Figure 5.5: The absolute value of R_{RESET} deviation of the defective device as cycle increases.

surement data, we focus on the RESET behavior in the rest of the paper.

5.5. DEFECT MECHANISM

The concept of 'ion depletion' was introduced by Chen, Lu, Gao, *et al.* in 2011 to characterize the RRAM endurance degradation phenomenon that occurs as a result of a reduction in HRS [194]. The irreversible switching behavior observed after 10^5 cycles was the physical mechanism of this RESET failure. In 2012, the impact of the SET/RESET pulse on device reliability was examined [230]. It was asserted that the use of overly strong SET pulses induces an excessive quantity of OV at the interface, resulting in HRS failure after 10^6 cycles. Following that, these two earlier investigations demonstrated that the voltage amplitude, rather than the stress duration, has a substantial impact on the O^{2-} drift. In 2015, the thin oxygen-scavenging metal layer was argued to fail to form an Ohmic contact by Schönhals, Wouters, Marchewka, *et al.* This failure occurs when the capping layer is unable to release a sufficient amount of O^{2-} to recombine the OV in the RESET process [231]. In 2017, the HRS decrease was experimentally studied with the creation of rougher and thicker CFs following 10^4 switching cycles [232]. However, they all neglected to observe the HRS reduction caused by device defects during the initial hundreds of cycles, resulting in experimental results that were exclusively focused on endurance failures. In addition, they failed to establish a circuit-level defect model that would describe the degradation behavior of the HRS.

The BS mechanism is clarified due to the movement of O^{2-} to form and rupture the CF composed of OV, which is a *stochastic* process [98], [99]. During the SET operation, OV is generated, while O^{2-} drifts to the capping layer (Ti) and is stored at the interface between the capping layer and the HfO_2 [98], [99]. The rupture of CFs near the interface is caused by the migration of O^{2-} from Ti back to the bulk oxide and the subsequent recombination with OV under a reverse bias [98], [99]. The nearly constant LRS of both defect-free and defective devices (see Fig. 5.3(b), ①) suggests that the SET process generates an equal quantity of O^{2-}. The O^{2-} stored in the capping layer is necessary for the recombination of OV during the RESET process [98]. However, the oxygen reservoir of defective devices is depleted of O^{2-}. Only a limited amount O^{2-} can be released back to recombine the OV, resulting in an insufficient RESET, and hence a smaller resistance window as well as the reduced HRS. It is important to note that defect-free devices exhibit a slight decrease in HRS at low voltages (less than −1.5 V) in Fig. 5.3(b). This phenomenon

can be attributed to the effect of tunneling leakage currents [99]. Conversely, defective devices continue to release a small fraction of insufficient O^{2-} at high voltages in order to recombine the CF.

To better investigate the physical reason for the defect, we performed weak RESET (sweep the V_{stop} at −1 V) and compared the measurements with those of regular RESET (V_{stop} at −1.8 V), as shown in Fig. 5.6. For each cycle, the measurements are obtained as follows: regular RESET sweep, regular SET sweep, and weak RESET sweep. Fig. 5.6(a) shows that as the voltage decreases from −1 V to −1.8 V, the current slightly decreases since more recombination between O^{2-} and OV (i.e., higher R_{RESET}), which implies that the defect is affected by applied voltage. In Fig. 5.6(b), the R_{RESET} remains at a lower value at V_{stop} of −1 V than V_{stop} of −1.8 V, and the nearly constant R_{RESET} (V_{TE} from −1 V to 0 V) shows the defect is permanent once it occurs in the cycle.

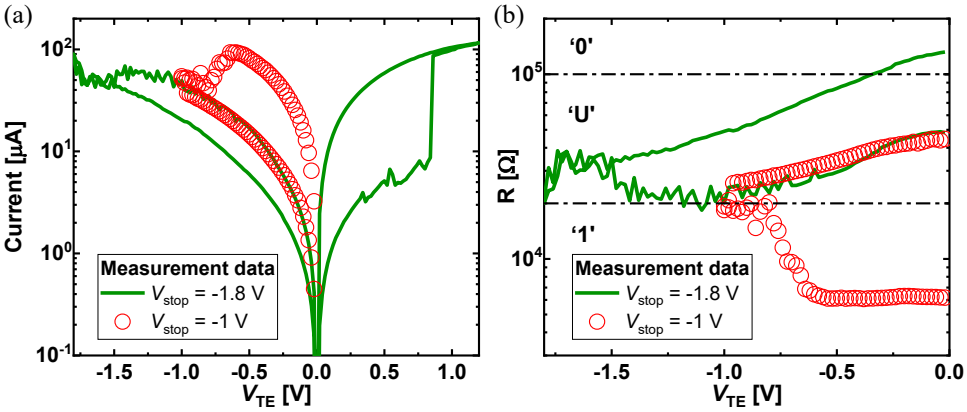


Figure 5.6: Regular and weak RESET of the defective device with different V_{stop}. (a) Measured I-V curve in the logarithmic y-axis, (b) Measured R-V curve in the logarithmic y-axis.

As indicated by the measurement data that was previously presented, the ion depletion defect is the likely cause of the defective RESET switching. The defect is likely caused by interface physical imperfections, which result in an elevated oxygen trap of the defective device. This leads to a reduced recombination rate between inadequate O^{2-} and OV [194], [230], [232], [233]. Hence, the reduced recovery of OV due to inadequate non-lattice O^{2-} stored in the capping layer induces lower HRS. Fig. 5.7 graphically depicts the degradation mechanism of the HRS and four potential resistance states of the device during forming, RESET, and SET processes: 1) R_{fresh}, pristine RRAM devices with minimal OV before forming, have an extremely high initial resistance state [106], 2) R_{SET}, O^{2-} and OV are generated in SET or forming process, 3) R_{RESET}, regular recombination between O^{2-} and OV results in the nominal HRS during RESET of defect-free devices, and 4) R_{RESET}(defective), in RESET of defective devices, less recombination of OV due to physical imperfections. As a result, the resistance values are as follows: $R_{fresh} > R_{RESET} > R_{RESET}$(defective) $> R_{SET}$.

The imperfections of the RRAM process have a significant impact on the test methods for its design. The RRAM defects are associated with its manufacturing process,

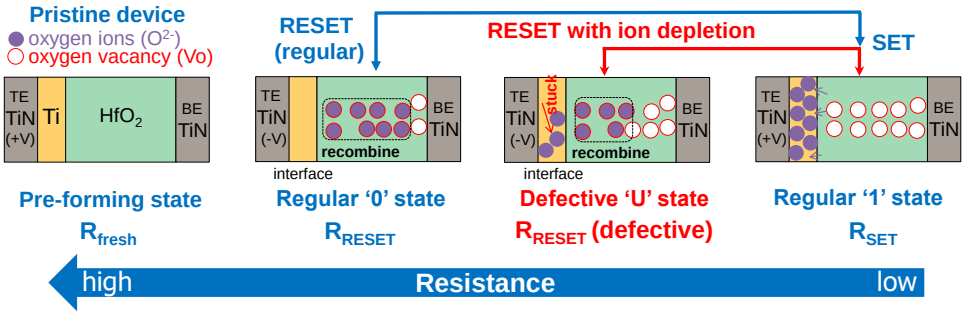


Figure 5.7: The schematic view of ion depletion mechanism during RESET process.

and the defects that are introduced during each stage of the manufacturing process are unique. The general manufacturing process is illustrated in Fig. 5.8 [106], [177], [234], [235]. Transistors are initially assembled on the wafer in the FEOL, which is equivalent to conventional CMOS technology. Then, the metal layers are deposited in the BEOL. Ultimately, the CF is generated by employing a high forming voltage.

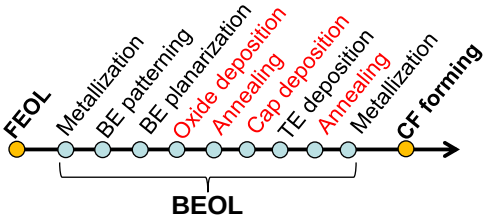


Figure 5.8: Flow of RRAM manufacturing process [106], [177], [234], [235].

In this section, we assume that the following imprecise fabrication processes (represented in red font in Fig. 5.8) are associated with the characterized ID defect. These procedures may impact the interface imperfections, thereby affecting the depletion of generated O^{2-}.

- **Thickness variations of deposition:** The capping layer is deposited at varying thicknesses. The stable BS is favored by the thicker capping layer due to its increased oxygen affinity and decreasing work function [231], [233]. Conversely, a thin capping layer captures a significant quantity of O^{2-}, making it challenging to fill the CF, which denotes a blocking contact formed at the interface [231]. The optimization of the proper barrier height at the interface between the cap and oxide is also influenced by the oxide layer [98]. Enhanced stability of RESET has been observed in thicker oxide [98], [236].

- **Annealing process:** An appropriate annealing treatment may enhance the crystalline quality of the oxide bulk and raise the oxygen concentration in the capping layer [234], [237]. The switching performance of RRAM is shown to be correlated

with variations in the concentration of hydroxyl groups [238]. Substantial dependency on annealing temperature is seen at HRS, as reported in [238]. The annealing temperature is also a determining factor in the surface roughness [238]–[240]. Therefore, the presence of nonuniform or unsuitable annealing temperatures might impact the interface activity, resulting in HRS degradation.

Considering the variability of the manufacturing process and the unpredictable nature of CF dissolving [229], the ID defect is not present on every device and every cycle (intermittent).

5.6. DEVICE-AWARE DEFECT MODELING OF ION DEPLETION

In this section, we apply a systematic device-aware approach to model the characteristic ion depletion precisely. By including the impact of technological parameters (e.g., the CF length and oxide thickness) on the electrical behavior of the defective devices, the device-aware approach is able to accurately represent the physical defects [59]. The following three phases comprise the systematic device-aware defect modeling approach: 1) analysis and modeling of physical defects, 2) electrical modeling of the defective device, and 3) model optimization and fitting. The DAT approach takes silicon data of defective devices and a physics-based device model as inputs. The output is a modified (parameterized) model of a defective device. Next, we will implement the DAT approach to design a physics-based model for the ID defect. To achieve this aim, our initial step is to model the impact of ID defects on the concentration of O^{2-} in defective devices. Following this, the electrical parameters of the RRAM, such as V_{RESET} and R_{RESET}, are used to map its impact. Finally, the measurements are employed to calibrate the ID-defective RRAM compact model.

5.6.1. PHYSICAL DEFECT ANALYSIS AND MODELING

The defect physical mechanism must be used as the foundation for the development of the ID defect model in order to assess the effects of each (key) technology parameter of the device. In Section 5.4, we characterize and analyze the ID defect, which is caused by the insufficient O^{2-} that can recombine OV in the bulk oxide. As a result, certain technological parameters that impact the concentration of insufficient O^{2-} (play a key role in OV recombination during the RESET process) should be modified from their ideal values to reflect defective behavior. The concentration of OV in the oxide bulk is directly proportional to the resistance state of the RRAM device. The physics-based HfO_2 RRAM model, JART VCM v1b, from [110], is employed to accurately model this defect. The compact model is intended to simulate the change of OV in the oxide in a Valence Change Mechanism (VCM)-based device. The equivalent circuit diagram of the compact RRAM model is depicted in Fig. 5.9(a). This diagram comprises a Schottky-like contact (V_{schottky}), two filament resistances (R_{plug} and R_{disc}), and a series resistance (R_{series}) [110]. In this model, the HfO_2 (oxide) layer is divided into two separate regions: the disc region and the plug region. The disc region is the spot where the switching occurs for simulation purposes, while the plug region is conductive and provides an infinite supply of vacancies. During the switching process, both the SET and RESET processes can be described as the ionic migration of OV. This migration affects the Schottky barrier and, as a result, the

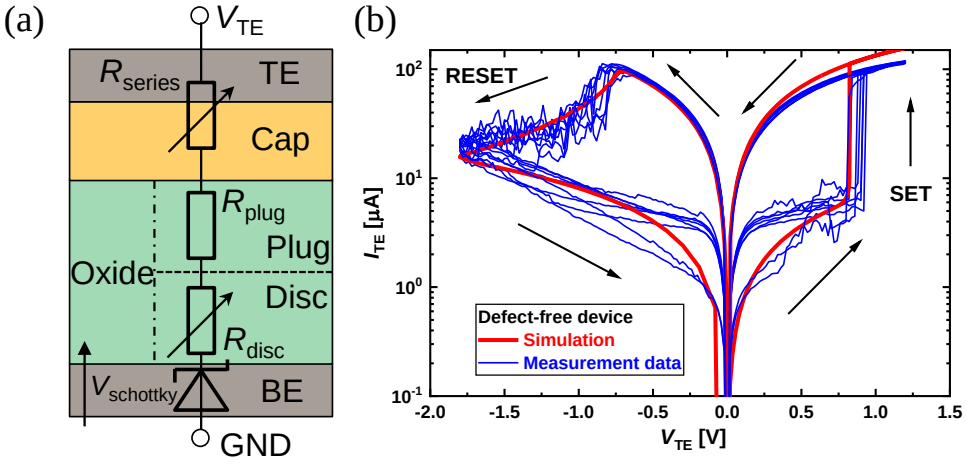


Figure 5.9: The Verilog-A compact model diagram and defect-free device simulation model. (a) Equivalent circuit diagram of the JART VCM v1b model [110], (b) Defect-free device fitting in the logarithmic y-axis.

5

electrical conductivity of the VCM device [110]. In order to calculate the concentration of OV in the disc region, the parameter N_{disc} is employed in the model simulation. The quantity of N_{disc} varies to reflect the switching characteristics and impact the resistance condition in the RRAM model. In switching processes, the limiting parameters $N_{disc,min}$ and $N_{disc,max}$ enable the maintenance of N_{disc} within the range of $N_{disc,min}$ and $N_{disc,max}$. Hence, we can tackle the undesired ID defect by adjusting the minimum oxygen vacancy concentration in the disc (denoted as $N_{disc,min}$). The presence of an ID defect restricts the recombination of OVs to a limited amount of O^{2-}. Consequently, a higher percentage of OVs remains after the RESET switching for the defective device. $N_{disc,min}$ is the residual number of OV in the disc after recombining the maximum number of OV. Reducing the negative RESET voltages can increase the production of O^{2-} to promote the recovery of the HRS and facilitate the migration of O^{2-} back to the oxide. Hence, the critical values of $N_{disc,min}$ for the parameter must be explicitly calculated considering the influence of the RESET voltage. Furthermore, ion movement into and out of the disc region is driven by a field-enhanced ionic hopping mechanism coupled with Trap-Assisted-Tunneling (TAT), both of which exhibit an exponential correlation [99], [110]. We employ the same exponential relationship function to formulate $N_{disc,min}$ for *numerical estimations* in the RESET procedure. The formula applied in the physical defect modeling phase is given by:

$$N_{disc,min} = exp\left[-\left(-V_{TE}/p_1\right)^{p_2}\right] (V_{TE} < -1\,V). \tag{5.1}$$

Here, p_1 and p_2 function as fitting parameters. The equation is only valid when its value remains below the actual N_{disc} of defect-free devices. In our case, this threshold is -1 V.

Due to the RRAM variation, the ID defects occur at intervals, usually irregular, and their occurrence probability is not 100 % during multiple cycles. To accurately design the ID defect model incorporating the intermittent behaviors, we conduct a statistical analysis based on the measurements.

The occurrence probability of undesired 'U' state of R_{RESET} in the measured defective device (WL=5, BL=3) during a total of 836 cycles (skip the first 100 cycles) is analyzed. Based on the measurements, we calculate the ID occurrence probability in consecutive n time cycles ($n = 1,2,3,4$). Among the 836 cycles, P_{D1}=117/836=14.00 % of them are defective (e.g., in cycles 110, 210, and 263), P_{D2}=29/(836-1)=3.47 % are sustained for 2 *consecutive* cycles (e.g., in cycles 107 and 108), P_{D3}=7/(836-2)=0.84 % are sustained for 3 *consecutive* cycles (e.g., in cycles 348, 349, and 350), P_{D4}=1/(836-3)=0.12 % are sustained for 4 *consecutive* cycles. The maximal duration of the undesired 'U' state is 4 cycles and only occurs once starting in cycle 763.

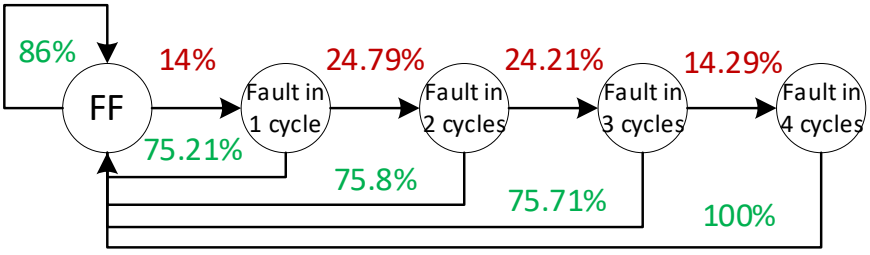


Figure 5.10: State transition diagram showing probabilities of faulty 'U' states over multiple cycles.

Next, we conduct the correlation analysis of consecutive cycles for the defective device. Assuming the undesired 'U' state occurring in each cycle are independent events from each other, the occurrence probability of the defect in 2 consecutive cycles is $P_2 = P_{\mathrm{D1}}^2 = 1.96\,\%$, which is smaller than the measured P_{D2} (3.47 %). Similarly, $P_3 = P_{\mathrm{D1}}^3 = 0.27\,\% < P_{\mathrm{D3}}$, and $P_4 = P_{\mathrm{D1}}^4 = 0.04\,\% < P_{\mathrm{D4}}$. Hence, there is a correlation between each cycle and the occurrence of undesired 'U' states. We can use the formula of conditional probability. First, we define 'the second cycle is defective' as event B, 'the third cycle is defective' as event C, and 'the fourth cycle is defective' as event D. Then, the conditional probability that the second cycle is defective (event B) given that the first cycle has already been defective (event A) is: $P(B|A) = P(A \cap B)/P(A) = P_{\mathrm{D2}}/P_{\mathrm{D1}} = 24.79\,\%$. Following the same formula, we can obtain $P(C|(A \cap B)) = P_{\mathrm{D3}}/P_{\mathrm{D2}} = 24.21\,\%$, and $P(D|(A \cap B \cap C)) = P_{\mathrm{D4}}/P_{\mathrm{D3}} = 14.29\,\%$. Fig. 5.10 represents a state transition diagram, illustrating the statistical probability associated with the faulty 'U' states occurring over multiple cycles in a system. Each state corresponds to a specific number of fault cycles (e.g., 'Fault in 1 cycle'), with transition probabilities indicating the statistical likelihood of moving to the next state or returning to a fault-free state ('FF'). For example, the diagram starts with an initial state, labeled 'FF' with a transition probability of 14 % leading to the 'Fault in 1 cycle" state, indicating that there is a 14 % chance of a fault occurring in the first cycle. Subsequent transitions between states show the probability of faults continuing across additional cycles, with 24.79 % leading to a fault in the second cycle, 24.21 % leading to a fault in the third cycle, and 14.29 % leading to a fault in the fourth cycle. Red percentages represent the chances of faults continuing, while green percentages show the likelihood of recovery or remaining fault-free. For instance, from the initial state 'FF', there is an 86 % chance of returning to the fault-free condition after one cycle and a 75.21 % chance after two cycles. Finally, once a fault occurs in all 4 cycles, it will

100 % recover to the 'FF' state since the maximal duration of the defect is 4 cycles based on the measurements.

5.6.2. ELECTRICAL DEFECT MODELING

Table 5.2: Model parameters for JART VCM v1b [110].

Symbol	Value	Symbol	Value
T_0	0.293 K	ν_0	2×10^{13} Hz
ϵ_s	11	ΔW_A	1.35 eV
$\epsilon_{\phi_{Bn0}}$	2.5	R_{th0}	15.72×10^6 KW^{-1}
ϕ_{Bn0}	0.18 eV	r_{det}	45 nm
ϕ_n	0.1 eV	l_{cell}	3 nm
μ_{n0}	4×10^{-6} m^2/(Vs)	l_{det}	0.45 nm
$N_{disc,max}$	6.5×10^{25} m^{-3}	$R_{theff,scaling}$	0.31
$N_{disc,min}$	5×10^{23} m^{-3}	$R_{series,ICL}$	650 Ω
N_{init}	5×10^{23} m^{-3}	R_0	1750 Ω
N_{plug}	6.5×10^{25} m^{-3}	$R_{th,line}$	90 471.5 WK^{-1}
a	0.25 nm	α_{line}	3.92×10^{-3} K^{-1}

Once the physical defect analysis is completed, the impacted physical parameters (e.g., the OV concentration) are integrated into the electrical parameters, including V_{RESET}, R_{RESET}, and I_{TE}. JART VCM v1b is a compact model that is written in Verilog-A and can be seamlessly integrated into the circuit-level (e.g., SPICE simulator) simulation. The model is capable of analyzing the impact of insufficient O^{2-} recombination on the switching electrical behavior (e.g., current, resistance) in the presence of the modeled defect by utilizing $N_{disc,min}$ as an input parameter. The temporal evolution of N_{disc} is presented in the defect-free model as [110].

$$\frac{dN_{disc}}{dt} = \frac{I_{ion}}{zVo \cdot e \cdot A \cdot l_{det}}. \tag{5.2}$$

Here, the cross-section and length of the filament are denoted by A and l_{det}, respectively. The ionic current in the device is denoted by I_{ion} and is calculated by the recommendations of Genreith-Schriever [241]. Thus, the ionic current alteration is directly proportional to the OV concentration (N_{disc}) and is equal to zero when either $N_{disc,max}$ or $N_{disc,min}$ is attained [110]. All other parameters are listed in Table 5.2. Hence, the electrical behavior of the defective device switching, which includes current variations, can be examined by connecting it to N_{disc} using model simulations.

Next, we incorporate the intermittent behavior into the electrical parameters. To include the intermittent behavior, we modified the defect model in the following steps: 1) count the number of cycles each time, 2) based on fault occurrences in previous cycles (see Fig. 5.10), determine the probability of a fault in the current cycle, 3) modify the $N_{disc,min}$ if an ID-defect occurs, and 4) update fault occurrences and cycle count. Reset the cycle counter to 0 after it reaches 4. Note that the $N_{disc,min}$ is modified for defects with varying defect strengths, which is fitted with defective measurements.

5.6.3. FITTING AND MODEL OPTIMIZATION

MATLAB R2021a is employed to perform the fitting in order to match the I-V measurements from Section 5.4. The fitting procedure is comprised of the subsequent two steps: 1) the fitting of defect-free device behavior, and 2) the fitting of the defective device behavior using defective parameter $N_{\text{disc,min}}$. The defect-free measurement data was calibrated using the JART VCM v1b model, with the fitted parameter values displayed in Table 5.2. In the simulation, the RRAM model is connected in series with a transistor of identical dimensions to those used for device characterization. The voltage (V_{TE}) is ramped from 0 V to −1.8 V, and back to 0 V for RESET; from 0 V, then to 1.2 V, and back to 0 V for SET. The fitting result of the BS device model simulation and the measurements of the defect-free device in various RESET-SET cycles are illustrated in Fig. 5.9(b). The model exhibits BS behavior, which is consistent with the measurements.

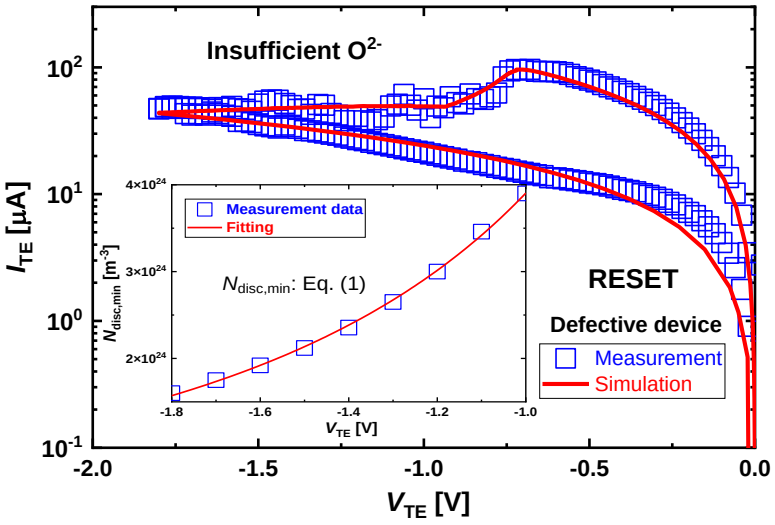


Figure 5.11: The ion depletion defective device fitting in the logarithmic y-axis.

Next, we calibrate the measurements for the defect. Fig. 5.11 illustrates the simulation outcome of the calibrated model and measurement data for an ID-defective device. We employ the same parameters as those in Table 5.2 but adjust $N_{\text{disc,min}}$ to the measurements by changing the values of parameters p_1 and p_2 using least squares in MathWorks's MATLAB R2021a. The parameters p_1 and p_2 are determined to be 0.0611 and 0.4209 to fit Equation 5.1, respectively. The inset of Fig. 5.11 demonstrates that a smaller value of $N_{\text{disc,min}}$ is induced by lower negative RESET voltages (although it remains greater than $N_{\text{disc,min}}$ for a defect-free device). This is in accordance with the actual RRAM mechanism observed in [194]. Fig. 5.12(a) illustrates the simulated I-V curves, while Fig. 5.12(b) compares the fitting results for both defect-free and defective devices through the simulated R-V curves under the RESET/SET voltage sweep. First, both devices exhibit comparable LRS behavior, increasing at an equivalent rate as the RESET voltage decreases (①). Then, the defective device encounters a delayed increase in resistance, which is indicative of an anomalous RESET (②). The R_{RESET} is defective and

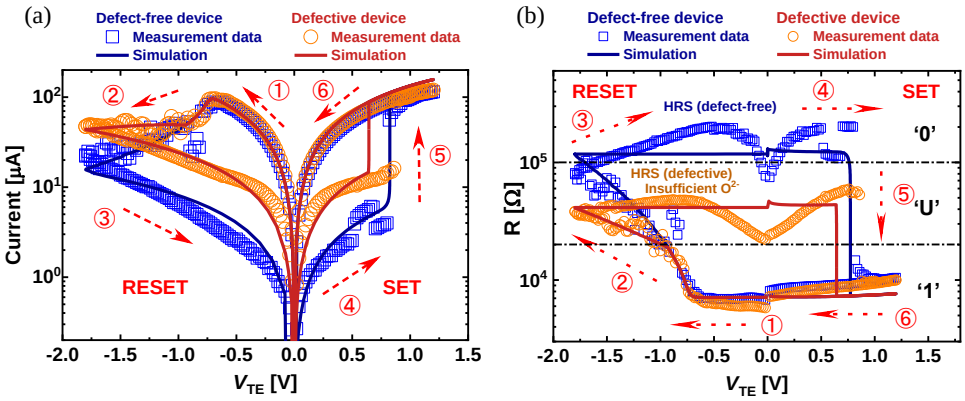


Figure 5.12: Defect-free and defective device simulation and measurements. (a) Simulation vs. measurements of the I-V curve, (b) Simulation vs. measurements of the R-V curve.

5

remains in an intermediate state (③, ④). However, it decreases in a normal manner with the positive SET voltage (⑤) and transitions to a normal LRS (⑥). The simulated HRS of the model at the beginning of SET is also greater than the HRS at RESET, as illustrated in Fig. 5.12(b) (④). This is elucidated by the fact that the Schottky contact at the BE/oxide interface is modeled as a diode in the RRAM model. Consequently, it can be categorized into forward (negative RESET voltage applied) and reverse (positive SET voltage applied) directions [110]. The potential barrier height of the diode is equivalent to the discrepancy between the electron affinity of the oxide and the work function of BE [110]. The SET process, particularly in the HRS, results in a higher device resistance and a small Schottky current due to the application of various electronic conduction mechanisms. The measurements are accurately simulated by the model, with the exception of the defective device's V_{SET} reduction (⑤). This is due to the fact that the model includes the initial resistance state-dependent thermoelectric coupling during the SET process, which results in a shorter delay time in transition for lower initial HRS [110]. Additionally, the final LRS is not dependent on the initial state; rather, it is based on the applied voltage amplitude, as demonstrated by both the model simulation and measurements [110]. Therefore, in the subsequent RESET cycle, both ID-defective and defect-free devices commence from the standard LRS. The model is capable of precisely describing the ID-defective RRAM device in a circuit simulator, as evidenced by the fitting results. Finally, the calibrated Verilog-A ID-defective RRAM compact model is the result of device-aware defect modeling. The fitting output indicates that the model is able to accurately describe the ID-defective RRAM device in a circuit simulator.

Then, we further develop the device-aware defective model to describe intermittent behavior and calibrate the model with the measurements. Fig. 5.13 presents ID-defective R_{RESET} frequency distributions of resistance values against the number of counts. For simulation purposes, we conduct 836 consecutive RESET-SET cycles and measure R_{RESET} after each cycle. Fig. 5.13 illustrates that the fitted model is capable of accurately describing the intermittent behavior as well as varying defect strengths. Furthermore, Fig. 5.14 shows the cumulative distributions of faulty resistance for the defec-

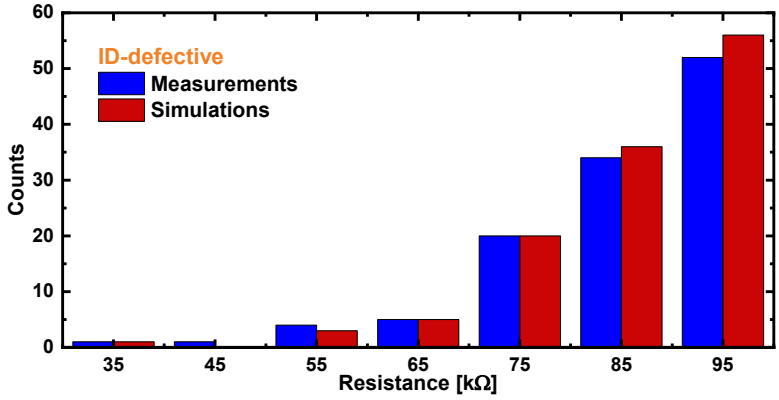


Figure 5.13: Histogram of defective R_{RESET} distributions, measurements vs. simulations.

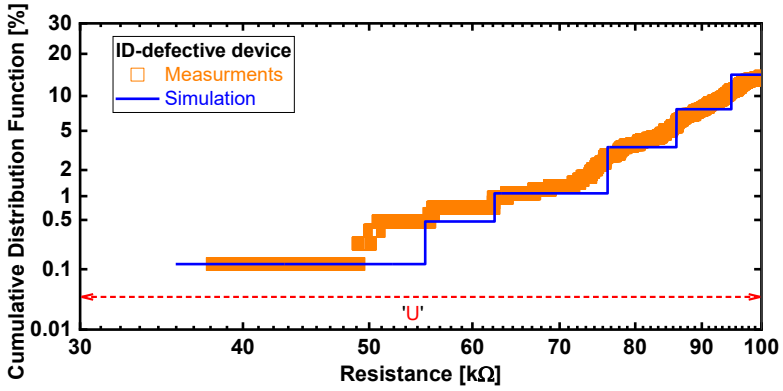


Figure 5.14: Cumulative probability of defective R_{RESET} distributions, measurements vs. simulations.

tive RRAM devices and the simulation results. Note that we only show measurements and simulation results that exhibit ID-defective 'U' states during cycles. Finally, the developed ID-defective model is compatible with circuit-level simulation, consistent with measurements, and significant for reliability analysis.

5.7. DEVICE-AWARE FAULT MODELING OF ION DEPLETION

5.7.1. SIMULATION SETUP

This step is derived from the simulation of the electrical model constructed in Section 5.6. The simulation of the ion depletion effect is conducted using Cadence Spectre, implemented with the TSMC-40 nm transistor library and the RRAM compact model from [110]. The structure consists of a 1T-1R cell (with identical dimensions as the devices used for characterization) and the required circuitry to provide the three control lines (BL, SL, and WL) with the appropriate voltages. Standard voltage-based SA is used for reading purposes [122]. Two experiments are conducted for the purpose of fault anal-

Table 5.3: Fault analysis results for ion depletion defect.

$N_{disc,min}$ [10^{26} m^{-3}]	R_{RESET} [kΩ]	Fault	Type
0.065	15.21	⟨1w0/1/−⟩	EtD
0.05	17.76	⟨1w0/1/−⟩	
0.0415	20.03	⟨1w0/U/−⟩	sHtD
0.035	22.51	⟨1w0/U/−⟩	
0.016	41.34	⟨1w0/U/−⟩	
0.0113	55.55	⟨1w0/U/−⟩	
0.0077	78.57	⟨1w0/U/−⟩	
0.006	99.64	⟨1w0/U/−⟩	
0.005	118.42	Fault free	

ysis. The first experiment is constructed to validate faults with different strengths of ID defects. The injection of ID defects is achieved by replacing the defect-free RRAM model with the model of the ID-defective RRAM device developed in Section 5.6. Various defect strengths of $N_{disc,min}$ are modified within the range of 0.005×10^{26} m^{-3} to 0.065×10^{26} m^{-3} to guarantee accurate model functioning and adherence to realistic physical constraints [110]. Consistent with the measurement setup, the voltage in the simulation sweeps via SL from 0 V to 1.8 V back to 0 V (in RESET hysteresis) to sensitize the fault. To determine the faulty behavior in the presence of the modeled defect, the simulation results are evaluated for the ultimate reached resistance of the defective devices (i.e., R_{RESET} after RESET operation). For example, F = U if the final R_{RESET} is outside of the HRS boundary (i.e., [100 kΩ, 1 MΩ]). The second experiment is established to validate the fault space when a linear resistor is present, either in series (R_s) or in parallel (R_p), with a device that is defect-free. The resistor defect sweeps from 1 Ω to 100 MΩ in strength.

5.7.2. FAULT MODELING AND RESULTS

The simulation results for varying defect strengths are shown in Table 5.3. Depending on the defect strengths, the table shows both Easy-to-Detect (EtD) (⟨1w0/1/−⟩) and strong Hard-to-Detect (sHtD) (⟨1w0/U/−⟩) faults. It can be shown that more faults get sensitized as $N_{disc,min}$ increases. While EtD faults are sensitized for $N_{disc,min}$ above 0.042×10^{26} m^{-3}, sHtD faults are sensitized for $N_{disc,min}$ between 0.006×10^{26} m^{-3} and 0.0415×10^{26} m^{-3}. Considering the defective measurement data shown in Fig. 5.12, $N_{disc,min} = 0.016\times10^{26}$ m^{-3} (bold in the table) is calibrated in this case. Greater space for OVs to be recombined in the disc region is given by higher $N_{disc,min}$, which increases the likelihood of undesired RESET failure in the device. The concentration of OV lowers as more O^{2-} are attracted toward the oxide bulk to recombine with it as the decrease in supplied negative RESET voltage. However, ID-defective devices are unable to release a sufficient amount of O^{2-} back into the oxide bulk, which raises the final OV concentration while lowering the R_{RESET}. As a result, faults can be sensitized at these voltages. Greater $N_{disc,min}$ indicates a higher defect strength and more O^{2-} depletion (inadequacy).

The static faults that are sensitized using the DAT as well as the traditional approach (linear resistor model) for all R_p, R_s, and $N_{disc,min}$ are then compared; these are shown

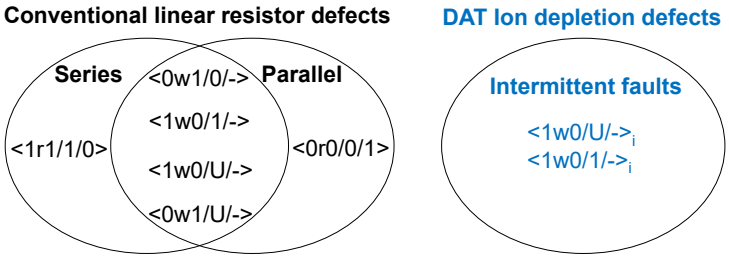


Figure 5.15: Static faults.

in Fig. 5.15. The figure illustrates the distinctions between the two approaches. Tests for 4 faults will result in needless yield loss since the traditional defect model approach triggers those that are unrealistic when modeling ID defects. Due to the intermittent behavior, ID will only sensitize intermittent faults, as shown in the right circle. We denote the corresponding fault primitive as $\langle 1w0/1/-\rangle_i$ and $\langle 1w0/U/-\rangle_i$, where the subscript i indicates the intermittent case. Clearly, conventional defect models based on either series or parallel resistors fail to capture the realistic faults sensitized by ID. Additionally, we are able to distinguish and diagnose ID defects from linear resistor defects due to the comparison result.

5.8. DEVICE-AWARE TEST DEVELOPMENT OF ION DEPLETION

The final stage of the DAT approach is to provide effective test solutions addressing the validated RESET faults in Section 5.7. In this section, we discuss and validate the test generations for ID defects.

5.8.1. MARCH ALGORITHMS

Table 5.3 illustrates how the ID defect could result in EtD and sHtd faults when the RESET procedure is performed. The March algorithm below is one straightforward test solution:

$$\text{March-ID} = \{\Updownarrow (w1); \Updownarrow (w0, r0)\}.$$

Here, '⇕' denotes an irrelevant addressing direction. The first march element sets all memory cells to state '1', with w0 indicating a RESET operation, and r0 signifying a read 0 operation for fault detection. This test algorithm ensures the detection of EtD faults, i.e., $N_{\text{disc,min}} > 0.0415 \times 10^{26}\,\text{m}^{-3}$ in our simulation. However, due to the *intermittent* nature of the ID, the proposed test algorithm detects the fault *probabilistically*. Hence, multiple iterations of 1w0 (for sensitization) and read (for detection) are needed to improve the probability of detection. Assuming the ID defect occurs with probability P_{ID}, then the detection probability can be expressed as: $P_{\text{d}}=1-(1-P_{\text{ID}})^k$, k indicates the repetition time of the applied sequence. In our case of P_{ID}=14 %, k=31 is necessary to achieve over 99 % fault coverage.

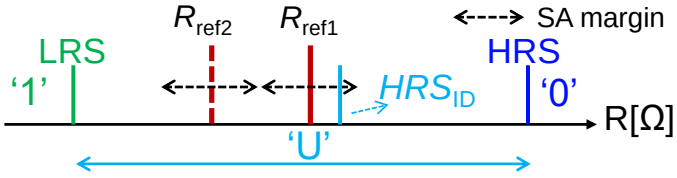


Figure 5.16: Resistances for faulty RRAM state.

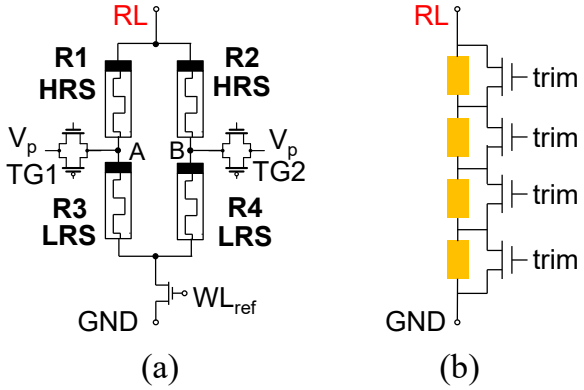


Figure 5.17: Reference resistance for DfT schemes. (a) Programmable reference cells, (b) Controllable reference trimming.

5.8.2. DFT SCHEMES

However, this March algorithm does not satisfy the fault coverage requirement for sHtD faults in the 'U' state, which can lead to random read outputs. Hence, DfT schemes specifically targeting these sHtD faults should be employed. It is possible to include different references in the SA so that it can distinguish resistance states between 'U'/'1' or 'U'/'0' instead of distinguishing only the regular '0'/'1' [61], [62]. In this section, we propose two low-cost re-configurable DfT schemes by modifying the references in SA to detect the intermediate HRS of ID-defective devices. Fig. 5.16 demonstrates the DfT concept involving various references to identify the 'U' state in ID-defective devices. R_{ref1} is set to (HRS + LRS) /2 for the defect-free device. R_{RESET} of the ID-defective device is referred to as HRS_{ID}, which is near R_{ref1} (within the SA margin) and cannot be reliably distinguished as logic '0'. To detect this faulty state, the reference can be set to R_{ref2} of (HRS_{ID} + LRS) /2 for the ID-defective device.

Next, we implement a 2 × 2 RRAM array circuit with read circuits and proposed write drivers to validate the following two DfT schemes. Note that here we evaluate the DfTs considering process variations but assuming that the ID defect occurs (not intermittent) to verify the detection capability.

REFERENCE TUNABLE SCHEME

The first scheme is an adaptable DfT method featuring a tunable reference design that remains compatible with standard read operations. The tunable reference cells com-

prise two series RRAM devices, paralleled to another two RRAM devices in series, with additional two transmission gates (TG1, TG2), as shown in Fig. 5.17(a). For instance, in a defect-free circuit design, one RRAM device is configured to LRS while the other is set to HRS within a single connection. The parallel configuration of the two links ensures that the equivalent cell resistance lies precisely between these two values (LRS and HRS). TG1 and TG2 function as switches to regulate the application of voltage pulses (i.e., V_P) at nodes A and B for programming the reference cells. For example, when TG1 is activated, a high voltage V_p is applied on node A, adjusting the resistance of R1. To implement the DfT scheme, an RRAM circuit including 2×2 1T-1R cells and SA (using TSMC transistor model and compact RRAM model [110]) is built. First, the defect-free circuit is validated. Under a 1w0 operation, a 1.8 V RESET voltage is applied to the SL, with the BL grounded, transitioning the resistance state from LRS to HRS. The fault-free HRS can then be read correctly by the used SA, with R_{ref1}=61.13 kΩ. Next, we replace the defect-free cell with the ID-defective device model including variations. After applying the same 1w0r0 operation, the faulty HRS (with the strength of $N_{disc,min}$=0.01×10^{26} m^{-3}) is incorrectly read because the BL/RL voltage difference ($\Delta V = V_{BL} - V_{RL}$) remains below the SA margin (ΔV_{min}). By tuning the original HRS in the reference cell to the HRS_{ID} (i.e., R_{ref2}=32.81 kΩ), the faulty HRS_{ID} can be correctly distinguished from the LRS by the readout value. Furthermore, we perform a sensitivity analysis to study the influence of process variation of both transistors and RRAM devices. For the transistor variations, we use the variation models that include the statistical mismatch from the TSMC 40 nm model library. For RRAM device variations, we set up the models as described in [119]. We incorporate both Device to Device (D2D) (from a truncated Gaussian distribution) and Cycle to Cycle (C2C) (change the variable parameters with confined step size, whose maximum is chosen to be 10 % of the current value) variations of the RRAM cells (including the reference cell). First, we implement and obtain the output of the normal SA by using the R_{ref1} to detect the injected ID defect. For every component combination, only 33.9 % of 5000 Monte Carlo (MC) simulations return the correct read output '0'. Then, we change to use the DfT by using the R_{ref2} (after adjusting the reference cell) and perform 5000 MC simulations, 99.34 % are read as expected '0' outputs. The first DfT scheme is resilient against process variations.

REFERENCE TRIMMING SCHEME

Although the first circuit proposed is validated to the ID defect, precisely controlling the programming of the RRAM-based reference resistor is challenging. Hence, we propose the second scheme, which involves the use of tunable resistors to modify the reference resistance. For example, this can be achieved through a serial connection of multiple resistors, each of which can be bypassed individually, as shown in Fig. 5.17(b). We use the minimal size (width=400 nm, length=270 nm) of the TSMC 40 nm library as the minimal resistor. It has an equivalent resistance of around 4 kΩ and is guaranteed to operate in the linear region. By controlling trim settings, the reference cell is set to R_{ref2}=40 kΩ(enable 10 transistors, design for defective 'U' cells), and R_{ref1}=60 kΩ (enable 15 transistors, design for defect-free cells). After injecting an ID defect with the strength of $N_{disc,min}$=0.01×10^{26} m^{-3}, the SA using R_{ref1} returns an incorrect '1', while the SA using the trimming R_{ref2} returns a correct '0'. Then we conduct the process variation analysis for both transistors and RRAMs. Fig. 5.18(a), (b), and (c) show histograms for the LRS,

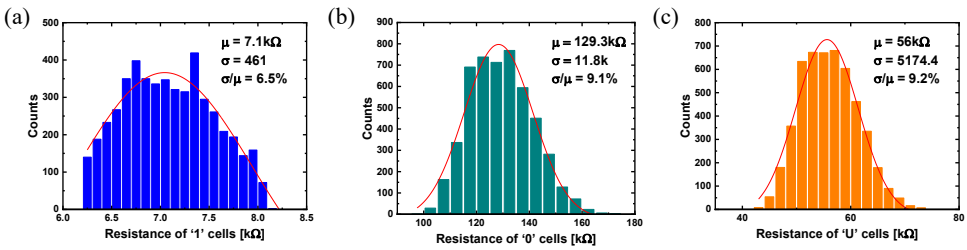


Figure 5.18: Histograms of the RRAM resistance distribution under variations. (a) Resistance of '1' cell (defect-free), (b) Resistance of '0' cell (defect-free), (c) Resistance of 'U' cell (ID-defective).

HRS cells of the defect-free circuit, and the defective 'U' cells, for 5000 MC simulations. The mean values are 7.1 kΩ of '1' cells, 129.3 kΩ of '0' cells, and 56 kΩ of 'U' cells, all within the specification. Besides, the coefficient of variations (CV, σ/μ) are 6.5 %, 9.1 %, and 9.2 % for '1' cells, '0' cells, and 'U' cells, respectively. In statistics, a lower CV indicates more uniformity in the data, which is consistent with the physics of RRAMs [229]. Upon validation of read operations under the process variation in the defect-free case, 100 % of the 10000 iterations (5000 read '0' and 5000 read '1') result in the correct read outputs with R_{ref1}=60 kΩ. When the cell is defective in 'U' state, 2191/5000 iterations return to incorrect '1' read outputs with R_{ref1}=60 kΩ. After the reference trimming is set to R_{ref2}=40 kΩ, 100 % of the 5000 iterations return to correct '0' read outputs although the ID-defect is injected. In summary, the result shows that the proposed DfT is decently resilient against process variations.

To further reduce the test duration, dedicated DfTs are necessary, focusing on minimizing the occurrence of the ID defect by e.g., operating the device at low temperatures below 293K. This will lead to a slow migration of O^{2-} recombination [110], [194]. Then, the ID-defective device may fail to switch and stuck at '1' (during the RESET operation), being easily detected. A drawback of this approach is that extra cryogenic equipment is required and the electrical behavior of the transistors is also affected at low temperatures. The approach may involve increasing device current during SET over-forming the device, leading to a higher OV concentration and reduced final RESET resistance in ID-defective devices [74], [214]. This can be achieved by boosting the BL or WL in test mode. However, the downside of this method is a lower R_{SET}, which increases energy consumption and write latency.

5.9. COMPARISON OF EXISTING DFTS

There are multiple existing DfT schemes that are designed for HtD faults. In this section, we will give an analysis of their testability for the ID defect and compare them with our proposed DfTs.

Some DfTs only target reducing the test time but not to enhance defect/fault coverage, such as works in [225], [226], [242]. Thus, they are not able to detect the ID defects. Singh, Fieback, Bishnoi, *et al.* introduced DfT schemes utilizing multiple references in the SA during a single read to execute a logic NOR operation to determine RRAM device resistance [62]. However, this approach requires a significant area overhead and does

not account for process variations. Besides, this design is not adaptive to a conventional storage memory array; instead, it has a specific architecture and aims to be implemented in CIM. Similarly, the DfT in [243] is also designed for CIM applications, focusing solely on detecting read disturb faults. However, it lacks the capability to detect ID defects and incurs a significant area overhead, requiring 32 RRAM cells in parallel for testing. Both designs in [61], [63] modify and implement multi-reference read operations by monitoring currents, making them capable of detecting ID defects. However, the study in [61] does not include a variation analysis, while the work in [63] incorporates a current mirror, which introduces a voltage drop and results in high power consumption. Compared to these approaches, our DfT designs are highly adjustable, easy to implement, and robust against variations, ensuring reliable detection of ID defects.

6

DEVICE-AWARE TEST FOR OVER RESET (OR) DEFECTS

This chapter identifies and characterizes the over-RESET phenomenon based on silicon measurements. In our case study, 30% of the I-V switching cycles in measured devices suffered from an intermittent extremely high resistance state exceeding the high resistance state criteria. The chapter shows the limitations of conventional defect modeling based on linear resistors for testing over RESET. To address this challenge, the DA defect modeling method is applied; a model of the defective RRAM device is developed and calibrated using measurements to accurately describe the impact of the defect on the electrical behavior of the memory device. Afterward, fault analysis is performed based on the DA defect model, and appropriate fault models are introduced; they show that the DA defect model will sensitize deep (extremely high resistance) state faults. Finally, dedicated test solutions for over-RESET devices are proposed.

The contents of this chapter were published in ATS'23 [66].

6.1. INTRODUCTION

This chapter identifies and characterizes the Over RESET (OR) in RRAMs. The OR causes a deep state exceeding the High Resistance State (HRS) criterion intermittently, which is unstable and affects the subsequent switching mechanism. Furthermore, we notice that the tunneling leakage current plays a big role when OR occurs; hence we investigate and model the leakage current based on measurements. The DAT method is applied to accurately model the defects, derive realistic fault models, and thereafter develop efficient test solutions. The main contributions of this chapter are as follows:

- Discover an RRAM-specific intermittent OR based on measurements and analyze the physical roots.
- Incorporate the leakage current to improve the RRAM model and develop the DA defect modeling for defects.
- Perform DA fault modeling to develop realistic fault models and thereafter optimal test methods.

6.2. DEFECT CHARACTERIZATION

We use the same wafer and measurement set up with 5.4. While analyzing the measurements, we observe that some cycles of several devices suffer from faulty RESET processes, i.e., the extremely high resistance state after RESET. This occurs for approximately 30 % of the device's cycles. Fig. 6.1(a) represents the log(R)-V loop to characterize and compare defect-free and defective devices. The applied V_{TE} is swept from 0 V to −1.8 V (①, ②), and back to 0 V (③) for RESET; from 0 V to 1.2 V (④, ⑤) back to 0 V (⑥) for SET. For the defect-free device, the RESET transition starts from −0.75 V with a constant LRS and ends at a nominal HRS (around 165 kΩ). In the case of a defective device RESET switching, a similar transition performance is observed (①, ②). However, its resistance value increases to the extremely high resistance state at ③. Note that the resistance increases since the leakage current decreases as the applied voltage changes. The effect of leakage current is particularly non-negligible when the device is in an HRS (around one order of magnitude resistance increase, see Fig. 6.1(a)). The leakage current exists in both RESET and SET processes. The subsequent SET process follows the normal transition for both defect-free and defective devices, but the defective device shows a significantly increased V_{SET}.

Fig. 6.1(b) presents the on/off resistance window as a function of cycles for defect-free and defective devices. During the initial ≈ 100 cycles, devices exhibit erratic resistance states, also reported in [228]. After the initial few cycles, the variability of resistance values is low. It can be concluded that the extremely high HRS occurs in multiple cycles. Fig. 6.1(c) presents the cumulative distributions of resistance for the RRAM device with 1T-1R structure. The distributions contain the measured resistance values (read using 0.1 V) from 936 consecutive RESET-SET cycles. The HRS distribution spread is more pronounced than LRS, which is also reported in other papers [229]. Around 30 % cycles of the defective device exceed the HRS resistance criterion of 1 MΩ, exhibiting the extremely high resistance state. Hence, the faulty behavior is intermittent, and it is a serious concern for RRAM devices.

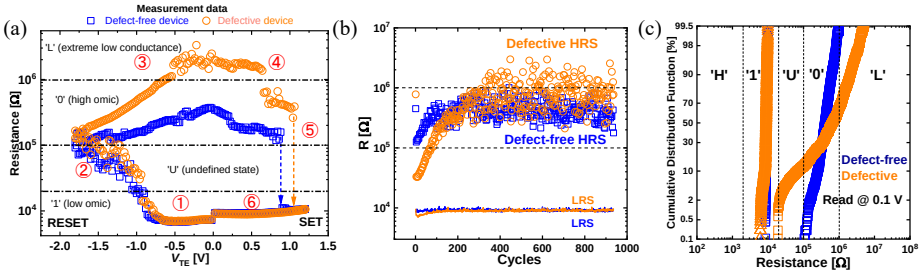


Figure 6.1: Comparing defect-free and defective devices. (a) Measured R-V curve in the logarithmic y-axis, (b) Measured R_{RESET} and R_{SET} in multiple (936) cycles, (c) Lognormal distributions of R_{RESET} and R_{SET} in multiple (936) cycles (read at 0.1 V).

Fig. 6.2 shows the dependence of measured HRS and subsequent V_{SET} during 936 cycles. For each V_{SET}, the mean of the HRS throughout numerous cycles is obtained in order to depict the trend clearly. It shows that a larger HRS usually corresponds to a larger V_{SET}. It can be explained that the V_{SET} is related to the amount of Oxygen Vacancy (OV) and affected chemical potential [244]. For example, one reported experiment shows that the low oxygen absorption rate of the insulating TiO_x capping layer leads to high V_{SET} than the metallic Ta capping layer [244]. Therefore, the extremely high resistance is risky in causing transition faults during the SET process, especially for low voltage (power) applications.

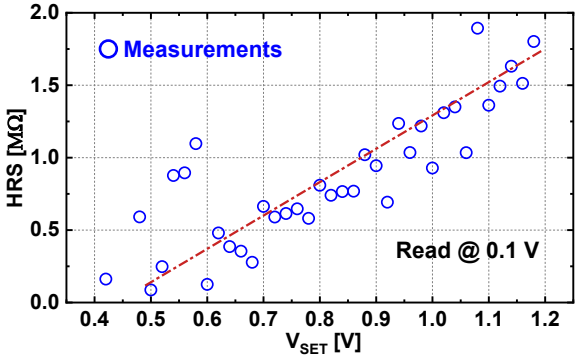


Figure 6.2: Dependence of V_{SET} and HRS of the device during 936 cycles.

6.3. DEFECT MECHANISM

The RRAM switching mechanism is explained as the movement of O^{2-} to form and rupture the CF composed of OV [99]. The same LRS values of defect-free and defective devices (see Fig. 6.1(a), ①, ⑥) indicate that the same amount of O^{2-} is generated during the SET process. Hence, the extremely high HRS value after RESET must involve a reduced O^{2-} storage capability in the capping layer of the device with respect to the nominal case. A probable cause of the fault is an increase in layer length or a decrease in

the cross-sectional area due to extreme process variation [245]. Furthermore, the doped oxygen leads to the high resistivity of the capping layer and provides more O^{2-}, which results in a deeper CF rupture [245]. Since the formation and rupture of a CF are *stochastic processes*, the precise number of O^{2-} fluctuates per write cycle, and hence the fault will not occur in every cycle [228]. Moreover, it is reported that the stochastic variability impacts HRS reliability [246], and hence the intermittent phenomenon.

6.4. DEVICE-AWARE DEFECT MODELING OF OVER RESET

In this Section, we first notice and incorporate the measured leakage current into the existing RRAM model to better calibrate the device behavior. Then, we analyze the OR-defective affected by the leakage current and apply the DA defect modeling approach to model the OR defect discussed in the previous section to calibrate with the measurements.

6.4.1. PHYSICAL DEFECT ANALYSIS AND MODELING

The physics of the defect must be investigated to comprehend its mechanism and determine its effect on one or more technology parameters of the RRAM device. There are two issues related to the OR: a) the leakage current that affects the device severely at extremely high resistance, and b) too many oxygen ions move to recombine the CF, resulting in the extremely high resistance state. Section 6.2 characterized the leakage current, which affects the current flowing through the device, hence the read resistance states. The impact of the leakage current is noticeable, especially for extremely high resistance states. However, most of the existing RRAM compact models assume the device switches as long as the applied voltage reaches the threshold [99], [110]. After that, the states (resistance) are constant unless the device switches again. Hence, the simulation of extremely high resistance states is less accurate when tunneling currents are not taken into account. Besides, the magnitude of the leakage current is related to the amount of OV and the resulting HRS, the root of which will be explained below.

In this section, we apply the physics-based RRAM model, JART VCM v1b, from [110] to incorporate the leakage current and appropriately model the defects. The compact model is designed as the change of OV (parameter N_{disc}) in the HfO_2 oxide layer. $N_{\text{disc,min}}$ and $N_{\text{disc,max}}$ are limiting parameters to keep N_{disc} between $N_{\text{disc,min}}$ and $N_{\text{disc,max}}$ in RESET and SET processes. To incorporate the leakage current into the model, we transform the leakage current to an equivalent current source, which is a function of applied voltages, and calibrate the current magnitude with measurements.

The leakage current between the cap and the oxide layer in the RRAM device is dominated by the Fowler-Nordheim tunneling mechanism, which is tunneling through an approximate triangular potential barrier. An expression for the current as a function of the applied voltage is [247]:

$$J_{\text{FN}} = \frac{q^3}{16\pi^2\hbar\phi_b} F_{ox}^2 exp\left[-\frac{4}{3}\frac{\sqrt{2m_{ox}^*}\phi_b^{\frac{2}{3}}}{\hbar q}\frac{1}{F_{ox}}\right] \tag{6.1}$$

where q is the electron charge, $\hbar$ is Planck's reduced constant, m_{ox}^* is the electron effective mass in the insulator (for HfO_2 of $0.1m_0$ [99]), ϕ_b is the barrier height at the interface,

and F_{ox} is the electric field cross the oxide (dependent on the V_{TE}).

Table 6.1: Model parameters for JART VCM v1b [110].

Symbol	Value	Symbol	Value
T_0	0.293 K	ν_0	1.6×10^{13} Hz
ϵ_s	10	ΔW_A	1.35 eV
$\epsilon_{\phi_{Bn0}}$	4.5	R_{th0}	6.5×10^6 KW^{-1}
ϕ_{Bn0}	0.18 eV	r_{det}	50 nm
ϕ_n	0.1 eV	l_{cell}	2 nm
μ_{n0}	3×10^{-6} m^2/(Vs)	l_{det}	0.65 nm
$N_{disc,max}$	5.3×10^{25} m^{-3}	$R_{theff,scaling}$	1
$N_{disc,min}$	2.5×10^{23} m^{-3}	$R_{series,ICL}$	2050 Ω
N_{init}	5.3×10^{25} m^{-3}	R_0	750 Ω
N_{plug}	5.3×10^{25} m^{-3}	$R_{th,line}$	90 471.5 WK^{-1}
a	0.54 nm	α_{line}	3.92×10^{-3} K^{-1}

As we analyzed in Section 6.2, the shortage of the free oxygen affinity causes over RESET in devices, which leads to the extremely high resistance fault. Therefore, we include the faulty final resistance by changing the model parameter that affects the O^{2-} storage capability: the minimal OV concentration in the oxide (disc). The lower this number, the higher the resistance value and thus the stronger extremely high resistance state faults.

6.4.2. ELECTRICAL DEFECT MODELING

Following the physical analysis, the electrical parameters (e.g., the OV concentration) comprise the impacted physical properties (e,g., R_{RESET}, I_{TE}) in this stage. The Verilog-A-written compact model JART VCM v1b can be directly integrated into the circuit-level simulation. Hence, by connecting it to $N_{disc,min}$ utilizing model simulations, the electrical behavior of the defective device switching comprising current changes can be examined.

6.4.3. FITTING AND MODEL OPTIMIZATION

The fitting is carried out to match the I (R)-V measurements from Section 6.2. The same parameter values listed in Table 6.1 other than $N_{disc,min}$ are used. In the simulation, the RRAM model is connected in series with a transistor that has the same dimensions as the devices used for characterization. Fig. 6.3(b) and Fig. 6.3(c) present the fitting results of I-V and R-V loops for both defect-free and defective devices. The applied voltage is ramped as in the characterization (① to ⑥), for a RESET-SET cycle. The circuit-level simulation results fit the measurements well.

The fitting of ϕ_b is carried out with the I-V measurements. We neglect the ionic hopping current caused by the minute amount of OV and approximate the current flowing through the RRAM as a tunneling current. Table 6.1 lists the parameter values used in Fig. 6.3(a) for the JART VCM V1b model to calibrate the measurement data. Fig. 5.12(a) shows the fitting result and implies that the model is able to accurately describe the R-V

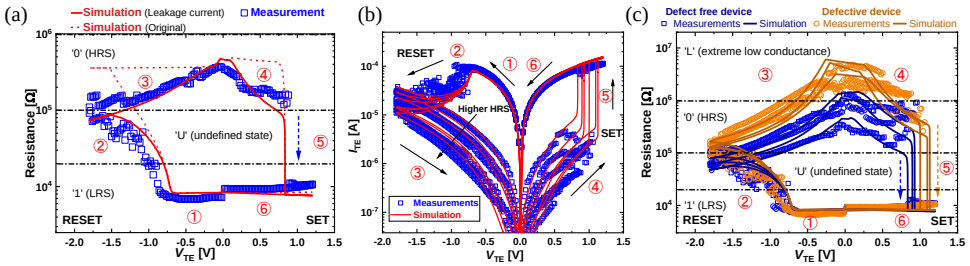


Figure 6.3: Defect-free and defective device simulation and measurements. (a) The leakage current incorporated model fitting of the R-V curve, (b) Simulation vs. measurements of the I-V curve, (c) Simulation vs. measurements of the R-V curve.

loop caused by leakage current.

Furthermore, we investigate the relationship between $N_{\text{disc,min}}$ and ϕ_b, which are two key parameters determining the extreme state fault and the magnitude of the tunneling leakage current, respectively. The ϕ_b is affected according to the concentration of OV. To obtain the results, we performed simulations under different fitted vacancy concentrations. For each case, the leakage current was calibrated by adjusting the barrier height parameter ϕ_b in the tunneling current equation until a good agreement with the measurement data was achieved. In this way, six values of ϕ_b were extracted, each corresponding to a specific OV concentration. These fitted barrier heights are then plotted as discrete data points, and a trend line is added to highlight the overall dependence, as shown in Fig. 6.4. In the figure, the extracted ϕ_b decreases with increasing OV concentration. This indicates that a higher OV concentration at the interface effectively reduces the tunneling barrier height, thereby enhancing the tunneling current, which is consistent with other reported results [248]. Therefore, ϕ_b is directly related to the OR defect strength, and its accurate fitting is crucial for reliable OR-defective modeling.

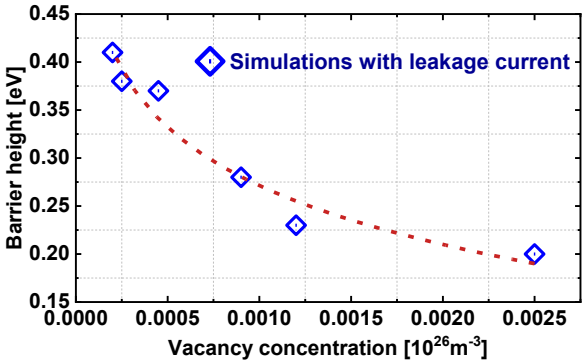


Figure 6.4: Reduced tunneling barrier height caused by OV concentration.

6.5. DEVICE-AWARE FAULT MODELING OF OVER RESET

6.5.1. SIMULATION SETUP

We adopt Cadence Spectre to establish the simulation by using the Predictive Technology Model (PTM) 130-nm transistor library [222] and the RRAM compact model from [110]. Regular voltage-based SA is applied for reading operations [122]. We perform two experiments to analyze fault results.

The first experiment is built to validate faults with varying OR defect strengths to analyze the extremely high resistance fault. The defect injection is carried out by replacing the defect-free RRAM model with the model of the defective RRAM device obtained in Section 6.2. The voltage sweep is applied through the SL from 0V to 1.8V back to 0V for RESET operation to sensitize the fault. The defect strength is governed by the varied $N_{\text{disc,min}}$ (from $0.0025 \times 10^{26}\,\text{m}^{-3}$ to $0.0001 \times 10^{26}\,\text{m}^{-3}$) after the RESET.

The second experiment is based on the traditional defect models to validate whether they can model the unique OR defect. Traditionally, RRAM defects are modeled using linear resistors [43]. There are only two options for defect modeling: the resistor can either be in parallel or in series with the defect-free device. Hence, we inject resistors both in parallel and in series to validate the sensitized faults. The strength of a resistor defect is swept from 1 Ω to 100 MΩ.

Table 6.2: Fault modeling results for OR defect.

$N_{\text{disc,min}}[10^{26}\,\text{m}^{-3}]$	R_{RESET}	Fault	Type
0.0025	358.3 kΩ	None	Fault free
0.0009	1 MΩ		
0.00025	3.7 MΩ	⟨1w0/L/−⟩	HtD
0.0002	4.7 MΩ		
0.0001	7.2 MΩ		

6.5.2. FAULT MODELING AND RESULTS

First, Table 6.2 shows the obtained results of the first experiment. We inspect the final resulting resistance (after RESET operation) to derive fault models. It can be seen that HtD faults are sensitized when the $N_{\text{disc,min}}$ decreases. A lower $N_{\text{disc,min}}$ (a larger ratio $N_{\text{disc,max}}/N_{\text{disc,min}}$) gives more room for OV to be recombined in the disc region, which makes it easier for the device to exhibit undesired RESET failure. Although no faults are sensitized by SET operation with OR defects, we still observe the V_{SET} delay as shown in Fig. 6.3(b), (c). That can be explained by the initial resistance state-dependent thermoelectric coupling during the SET process included in the used RRAM model, which reduces the transition delay time for lower initial HRS [110].

Next, we compare the static faults sensitized by conventional linear resistors and OR-defective models. Totally 6 static faults are sensitized by linear resistors: ⟨1w0/1/−⟩, ⟨1w0/U/−⟩, ⟨0w1/0/−⟩, ⟨0w1/U/−⟩, ⟨0r0/0/1⟩, ⟨1r1/1/0⟩. Clearly, neither parallel nor series resistors can sensitize the ⟨1w0/L/−⟩ fault, which indicates the need for the OR-defective model. Furthermore, linear resistors sensitize faults that are unrealistic when modeling OR defects; hence, tests for them will lead to unnecessary yield loss.

6.6. DEVICE-AWARE TEST DEVELOPMENT OF OVER RESET

As explained in Section 6.5, the targeted fault sensitized by the RESET operation is an HtD fault. Hence, a conventional March test cannot guarantee its detection. Reading this state will result in '0', the same as the normal HRS. To detect those HtD faults, DfT schemes can be applied. For example, we can design different references composed of RRAM devices, so that the SA can identify states between 'L'/'0' instead of identifying only the regular '0'/'1'. The shifted reference DfT concept is illustrated in Fig 6.5. For regular SA, R_{ref1} is typically set to (HRS + LRS) /2 to distinguish '1' and '0'. To detect the 'L' state (HRS_{EL}), another reference (R_{ref2}) needs to be set to the maximum value of the defined '0' range. Besides, due to the intermittent nature of the OR phenomenon, the proposed DfT scheme will only *probabilistically* detect the fault. Hence, repeated 1w0 (to sensitize) and read (to detect) operations are required to enhance the detection probability. If we assume the occurrence possibility of OR is P_{OR}, then the detection probability is: $P_{\text{d}}=1-(1-P_{\text{OR}})^{k}$, k indicates the number of times the sequence is applied. In our case of P_{OR}=30 %, k=13 is required to realize 99 % fault coverage.

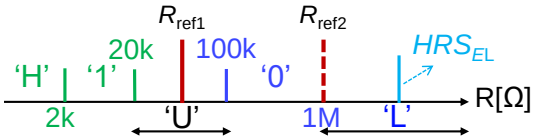
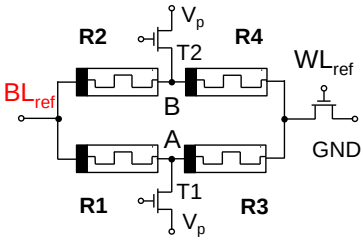


Figure 6.5: Resistances for faulty RRAM state.

Based on the concept, we develop an adaptable reference tunable structure that is compatible with normal read operations. Fig. 6.6 illustrates the two RRAM devices connected in series and parallel with two other RRAM devices connected in series, together with two extra transistors (T1, T2), for providing tunable reference cells. In Fig. 6.6, R1 to R4 are set to different states for regular (R_{ref1}) or DfT (R_{ref2}) purposes. In the regular case, R1, R2 are set to normal HRS while R3, R4 are set to normal LRS. In the DfT case, all 4 cells need to be set as the upper bound of the HRS region (1 MΩ) to detect 'L'. T1 and T2 are used as switches to control whether voltage pulses (i.e., V_{P}) can be applied at nodes A and B to program the reference cells. For example, when the T1 is conducting, V_{p} can be applied on node A, switching the state of R1.



	R1, R2	R3, R4
Regular	**100kΩ**	**20kΩ**
DfT	**1MΩ**	**1MΩ**

Figure 6.6: Modified reference cells.

The DfT scheme is implemented by establishing a Cadence Spectre-based simulation including a 1T-1R cell and SA. First, the defect-free circuit is validated to be correct.

Next, we replace the defect-free cell with the OR defective model. After applying the regular 0r0 operation using R_{ref1}, the faulty 'L' is read as an incorrect value '0'. By tuning the reference cell from R_{ref1} to R_{ref2}, the 'L' state can be correctly distinguished from the normal HRS by the readout value.

Other specific DfTs can be used to further reduce the test time. For example, a weak or fast write operation can be performed by the DfT scheme based on low write voltage or short write duration [57]. This will result in an insufficient RESET, which decreases the final RESET resistance of the faulty 'L' device to correct '0' state and be easily detected. The limitation of this scheme is a resulting lower R_{SET} for defect-free devices, which increases energy consumption.

6.7. COMPARISON OF EXISTING DFTS

In this section, we analyze the testability of existing works with respect to the OR phenomenon, which leads to the unique 'L' state write faults, and compare them against our proposed DfT schemes.

The approaches in [225], [226], [242] fail to detect OR faults, as they do not explicitly target enhanced fault coverage. The DfT proposed in [62] introduces a modified sense amplifier (SA) with multiple references; however, its specialized architecture is tailored for CIM implementations (e.g., scouting logic NOR operations) and is not adaptable to conventional memory arrays. Similarly, the DfT in [243] is also CIM-oriented and therefore unable to capture OR faults. More recent works in [61], [63] improve multi-reference read schemes by monitoring currents, thereby enabling OR detection. Nonetheless, the solution in [61] incurs significant area overhead, while [63] relies on a current mirror that introduces voltage drops and leads to high power consumption. In contrast, our proposed DfTs provide a more flexible and implementation-friendly solution, ensuring reliable detection of OR faults without excessive area or power penalties. More details will be discussed in Chapter 10.

7

Design-for-Testability Scheme for RRAM Faults

Due to the immature manufacturing process, RRAMs are prone to exhibit new failure mechanisms and faults, which should be efficiently detected for high-volume production. Some of those faults are hard-to-detect, and require specific Design-for-Testability (DfT) circuit design. This chapter proposes a DfT based on a parallel-reference write circuit that can detect all single-cell RRAM array faults: strong faults (directly causing logic errors) as well as weak faults (caused by parametric deviations). The scheme replaces the regular write driver, and enables the monitoring and comparison of the write current against multiple references during a single write operation. Hence, it serves as a DfT scheme and as a normal write circuit simultaneously. In addition, it enhances production testing speed and online fault detection while keeping the area overhead low. Furthermore, the DfT is configurable for efficient diagnosis and yield learning. The results of the simulations performed not only show that the DfT can detect single-cell conventional faults (due to interconnects and contacts) as well as unique RRAM faults (based on silicon data) that have been demonstrated to exist, but also that the DfT is robust to process variations.

The contents of this chapter were published in ITC'24 [67].

7.1. INTRODUCTION

Several works have focused on test solutions for RRAMs. These proposed solutions can be divided into two broad classes: March algorithms and specific Design-for-Testability (DfT) solutions. Examples of those March algorithms that involve specific sequences of memory operations are March-MOM [225], March W-1T1R [227], and March-CMOL [249]. While they are designed to optimize test time and enhance fault coverage, they only target interconnect and contact defects, and they are not designed to detect *unique defects* in RRAM devices (such as forming defects [59]) and Ion depletion [65]) causing *unique faults* such as undefined state faults [65]. Instead, DfT schemes, such as Weak Write operations [77], On-Chip Sensor [61], and DFT-HR-ET-NOR [62] are employed to detect those unique faults. However, they can only detect a part of this set of faults; they cannot guarantee the detection of the complete set of unique faults shown so far to exist. For example, the intermittent undefined state fault [214] only occurs *intermittently* during the write operation. Its detection cannot be guaranteed by Weak Write operations and DFT-HR-ET-NOR [62], [63]. The On-chip sensor may detect it; however, this DfT induces a large area overhead and requires hundreds of read operations to increase the detection probability. Also, the DfT in [63] fails to detect weak faults. Clearly, there is no test solution for RRAMs able to detect strong and weak faults, both due to conventional defects as well as unique defects in RRAMs.

This chapter provides a robust DfT scheme for RRAM memories targeting all studied RRAM defects and faults. The main contributions of this chapter are:

- Propose a multi-comparison write DfT scheme for the detection of RRAM (strong and weak) single-cell faults, in the presence of both conventional and unique defects.
- Implement and validate the DfT under process variations, and show that it outperforms the prior work. Silicon data measurements for unique RRAM defects are used.
- Demonstrate the reconfigurability of the DfT for different purposes, such as optimizing yield loss and minimizing the impact of variability.

7.2. TARGETED RRAM DEFECTS AND FAULTS

This section defines and classifies defects and faults in RRAM arrays, which are considered in this chapter. We target mainly single-cell faults.

7.2.1. RRAM CONVENTIONAL DEFECTS

Conventional defects consist of interconnect and contact defects in RRAM arrays, an example is a poorly placed contact [40], [48]. They are typically *modeled* by *linear resistors* [43], [52], and classified into three types: 1) a *bridge* being defined as a resistor between a pair of nodes different from the power nodes, 2) a *short* being an undesired resistive path between a node and a power node (V_{DD} or GND), and 3) an *open* being an increased resistance in an existing connection.

7.2.2. RRAM UNIQUE DEFECTS

These are defects that occur inside the RRAM device itself during the manufacturing; defects that have been shown to exist (based on silicon data) so far consist of: 1) an Over/Under Forming (O/UF) defect [49], [59], 2) a low-doping of the capping layer [214], 3) an Ion Depletion (ID) [65], [194], and 4) an Over RESET (OR) [66]. It has been shown that these defects cannot be accurately modeled with linear resistors; therefore, the Device-Aware Test (DAT) approach is used for their modeling [58], [59].

7.2.3. RRAM FAULTS

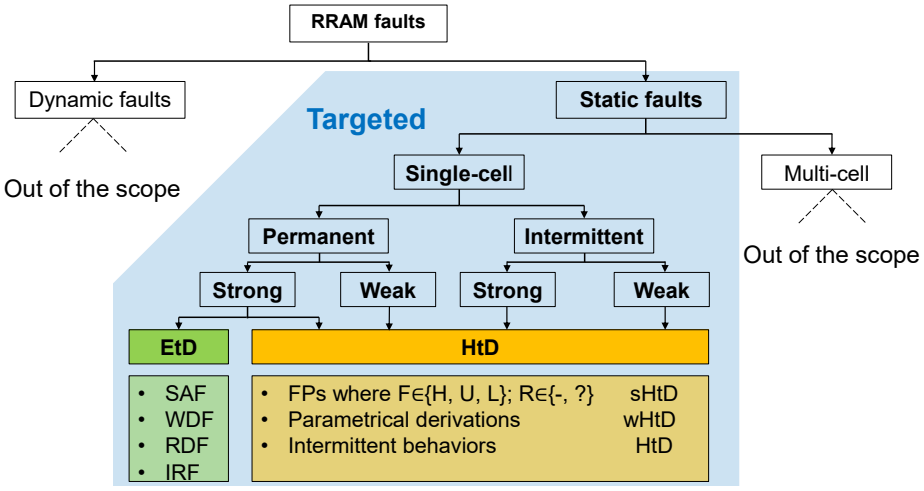


Figure 7.1: Fault classification depends on functionality and permanent nature.

In this chapter, we target static single-cell faults that have been shown to exist in RRAMs (either based on defect injection and circuit simulation or based on silicon data), which are either EtD or HtD, as shown in Fig. 7.1. The EtD static fault consists of [50], [54], [57], [250]:

- Stuck-at Faults (SAF): the RRAM cell is always in a certain state, e.g., ⟨0w1/0/−⟩.
- Write Destructive Faults (WDF): unintentional alteration of the state of the cell during a write operation, e.g., ⟨0w0/1/−⟩ and ⟨1w1/0/−⟩.
- Read Disturb Faults (RDF): a read operation switches the cell state, while the read value is correct, e.g., ⟨0r0/1/0⟩.
- Incorrect Read Fault (IRF) a read operation returns an incorrect output while the cell state is correct, e.g., ⟨0r0/0/1⟩ and ⟨1r1/1/0⟩.

The HtD static faults consist of [57], [214], [225], [251]:

- Deep Faults (DF): the RRAM cell falls into deep states, i.e., 'H', 'L', e.g., ⟨1w0/L/−⟩.

- Undefined Write Faults (UWF): a write operation leads to 'U' state, e.g., ⟨1w0/U/−⟩.
- Unkown Read Faults (URF): a read operation switches the cell to 'U' state (F = U) and/or returns random read outputs (R =?), e.g., ⟨1r1/U/?⟩.
- Weak faults: parametric deviations of RRAMs without functional errors.
- Intermittent Undefined State Faults (IUSF): RRAM switches into 'U' states during write operations intermittently.

Note that HtD faults are mainly unique to RRAMs; RRAM can store at most five states and thus detecting faults due to such states cannot be guaranteed with existing March tests. Due to the process variation and intrinsic stochasticity in RRAM switching, non-permanent faults occur intermittently, such as the IUSF. Weak faults in RRAMs are caused by degradation or extreme cycle-to-cycle variations [83], especially in the case of RRAM filaments, which have natural randomness in their formation and breakage.

7.3. LIMITATION OF EXISTING WORKS

Existing tests for RRAMs can be divided into March algorithms and DfT schemes. Table 7.1 summarizes all of these tests along with the types of RRAM faults (classified in 7.2.3) that can be detected. The DfT schemes generally have two targets: 1) reduce test time (RT), and 2) enhance test coverage (EC). The targets are listed in Table 7.1 for each of the DfTs. For example, the 'Divide and Conquer approach' DfT [252] is proposed to leverage upon the special current additive property, thus reducing test time. The DfT schemes in [225], [253] read multiple cells at once and thus reduce the number of read operations. However, [253] is not designed to enhance the FC. On the other hand, the DfT schemes in [61], [63] modify and apply multi-reference read operations to enhance the FC. Note that [63] is the only scheme that is able to partially detect intermittent faults since it can monitor the cell states continuously. In [243], an approach that can monitor the RRAM state ratio is presented for detecting read disturb faults; however, it requires testing 32 RRAM parallel cells at once. Furthermore, the DfT schemes in [57], [62], [242], [251] are designed to enhance the FC and reduce the test time by modifying the write or read operations. However, none of the existing DfT schemes can guarantee the detection of weak faults. Besides, some DfT approaches can only partially detect sHtD faults; i.e., they cannot cover faults that switch the cell into faulty states 'H', 'U', or 'L', resulting in test escapes. The state-of-the-art clearly shows that none of the existing tests detect all single-cell RRAM faults reliably and efficiently.

7.4. PROPOSED DFT METHODOLOGY

This section proposes the overall concept of the DfT, and demonstrates details on how it is implemented to efficiently detect all targeted single-cell RRAM faults in this chapter.

7.4.1. DFT CONCEPT

Based on the discussion from previous sections, we can derive that a high-quality DfT scheme for RRAMs should be able to detect specific faulty states ('H', 'U', and 'L' states)

Table 7.1: Targeted fault detection capabilities of existing tests for RRAMs.

Name	Type	Permanent			Intermittent		Target
		Strong		Weak	Strong	Weak	
		EtD	sHtD	wHtD			
March-MOM [225]	March	Y	N	N	N	N	-
March-1T1R [54]	March	Y	N	N	N	N	-
March C* [49]	March	Y	N	N	N	N	-
March C*-1T1R [226]	March	Y	N	N	N	N	-
March-CMOL [249]	March	Y	N	N	N	N	-
March W-1T1R [227]	March	Y	N	N	N	N	-
March-EtD [64]	March	Y	N	N	N	N	-
Divide and Conquer [252]	DfT	Y	N	N	N	N	RT
Sneak-path [225]	DfT	Y	P	N	N	N	RT
Weak-write [57]	DfT	N	Y	N	N	N	RT; EC
Fast-write [242]	DfT	Y	Y	N	N	N	RT; EC
Parallel March [253]	DfT	Y	P	N	N	N	RT
On-chip sensor [61]	DfT	Y	P	N	N	N	EC
Enhanced March [251]	DfT	Y	Y	N	N	N	RT; EC
DFT-HR-ET-NOR [62]	DfT	Y	Y	N	N	N	RT; EC
Read disturb fault detector [243]	DfT	Y	N	N	N	N	EC
PMRR [63]	DfT	Y	Y	N	Y	N	EC

Y: yes, N: no, P: partial, RT: reduce test time, EC: enhance test coverage

as well as weak faults (to enhance the chip's reliability). Furthermore, to detect aging degradation and intermittent behavior, the test should be performed during the runtime of the chip (e.g., online monitoring). For example, this is the case for ID defects reported in [65] which cause intermittent undefined write faults.

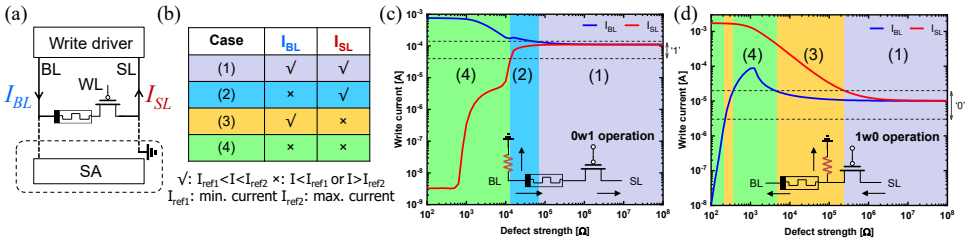


Figure 7.2: The concept of I_{BL} and I_{SL} conditions during the write operation. (a) The circuit high-level overview, (b) Four conditions of I_{BL} and I_{SL}, (c) I_{BL} and I_{SL} behavior as a function of one short defect located at the BL node, (d) I_{BL} and I_{SL} behavior as a function of one short defect located at the RRAM/transistor interconnection.

To detect the targeted RRAM faults, we design a new DfT based on monitoring and comparing two currents. We marginally modify the write drivers, and monitor the currents during the write operation; thus achieving a test without additional read operations. As explained in Chapter 2, the write operation is performed by write drivers. For instance, during the SET operation, and as shown in Fig. 7.2(a), currents from the write driver flow into the RRAM cell through BL (I_{BL}) and out of the cell through SL (I_{SL}). Depending on the magnitudes of such currents, four cases can be distinguished as shown in the table of Fig. 7.2(b). For a defect-free circuit, I_{BL} is expected to be the same as I_{SL}, both are within the specification (Case 1). However, the presence of defects (e.g., shorts,

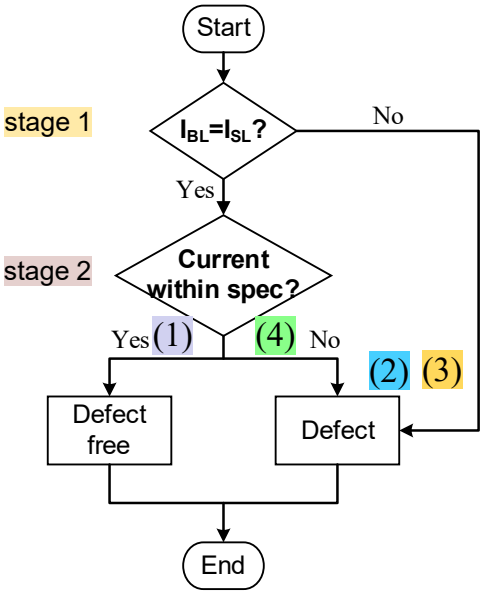


Figure 7.3: The flow chart of the proposed DfT process.

bridges) may cause one (Cases 2 and 3) or both (Case 4) of these currents to be out of the specification.

For example, Fig. 7.2(c) shows the amplitude of the two currents I_{BL} and I_{SL} during a write ($0w1$) operation in the presence of a short defect, as shown at the bottom of the same figure. The figure shows that the difference between the two currents is higher for smaller defect sizes (Case 4). As the BL is shorted to the ground in the presence of the defect, I_{BL} is higher than the correct value (out of the specification), and only a small current will flow through the RRAM device; hence I_{SL} is smaller than the correct value (out of the specification). As a consequence, the cell will fail to switch and it will remain in its initial state '0'. As the defect resistance increases, the shortcut current through the defect will reduce, resulting in a reduction of I_{BL} and an increase of I_{SL} following through the RRAM device. As the figure shows, in this case (Case 2), I_{BL} remains still larger than specification while I_{SL} reaches the correct value. When the defect resistance increases further, both I_{BL} and I_{SL} converge towards correct values (Case 1) as the impact of the defect becomes marginal. Note that Fig. 7.2(c) uses the same colors as the table in Fig. 7.2(b) to indicate the different regions/cases.

Fig. 7.2(d) presents another example where a write 0 transition operation (1w0) is performed in the presence of a short. Depending on the defect value, three cases ((1), (3), and (4)) are sensitized. Note that for Case 4, although the two currents I_{BL} and I_{SL} are quite the same, their value is possibly outside the specification (e.g., in the presence of open defects) and therefore it is a faulty case.

Fig. 7.3 gives a high overview of the DfT process. First, the two currents I_{BL} and I_{SL} are compared (i.e., stage 1); if they are not close to each other within a certain range (i.e., one

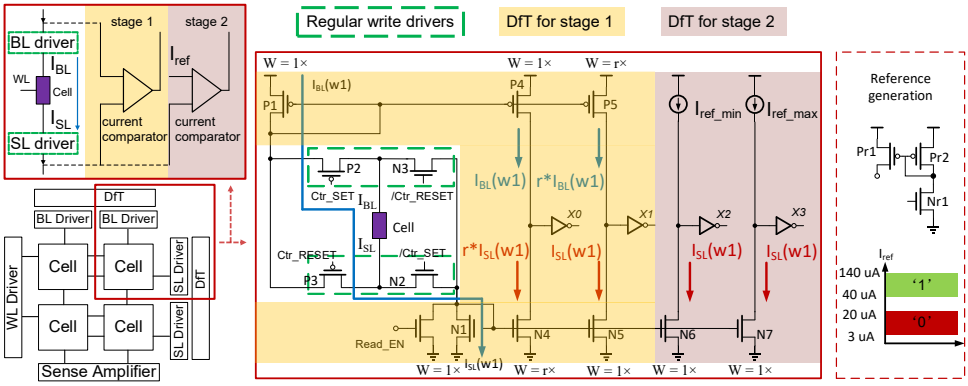


Figure 7.4: Modified write driver circuit acting as DfT.

of the two currents is outside the specification), then obviously there is a defect causing this deviation (Case 2 or Case 3). If the two currents are quite close to each other, we proceed to stage 2, where we verify whether the current value falls within the specified range. If it is, then clearly the design under test is defect-free (Case 1). However, if the current is out of specification, then the circuit is defective (Case 4).

7.4.2. POTENTIAL IMPLEMENTATIONS

The proposal DfT is based on a comparison of I_{BL} and I_{SL}. This concept of checking current differences and magnitude can be applied not only to write operations but also to read operations. For example, Gomez, Forero, Roy, *et al.* modifies the read operation to check the current difference in MRAMs [254]. The advantages of such an approach are low power consumption and detection of some read faults. However, it has many drawbacks: 1) the need to have the additional read operation for detection after writing the cell, 2) the slight read current is hard to sense, and 3) the ability to check (read) currents only in one direction. In comparison with checking the difference between read currents, checking the difference between write currents is much more favorable; i.e., it allows the real-time monitoring of the write current and test of the circuit without additional read operations, which is preferred for both defect and reliability testing. Therefore, we select that approach for the implementation of the proposed DfT concept.

7.4.3. SELECTED IMPLEMENTATION

Based on the above analysis, we propose to modify the basic write circuit (i.e., write drivers) to measure the difference between the write current flowing into and out of any cell. Besides, we check whether the current is within the specification during the write operation.

Fig. 7.4 shows the selected implementation of the proposed DfT. The figure shows the specific circuit design for one cell of the array and its modified write drivers (BL driver and SL driver). In the figure, we mark the $w1$ current path as a blue line. During the SET ($w1$) operation, the current flows through transistors P1, P2, and BL, via the RRAM device through the access transistor, into the SL (N2, N1). The current flowing into the

cell is $I_{BL}(w1)$ (via P1, P2), while the current out of the cell is $I_{SL}(w1)$ (via N2, N1).

The proposed implementation has two stages, as shown in the figure. Stage 1 checks if the difference between the two currents $I_{BL}(w1)$ and $I_{SL}(w1)$ is out of the specification, resulting in detection, e.g., of Case 2 with $X0X1 = 11$ if $I_{BL}(w1) < I_{SL}(w1)$ and with $X0X1 = 00$ if $I_{BL}(w1) > I_{SL}(w1)$. In case this current difference is small (resulting in $X0X1 = 10$), then stage 2 will be used to check if the magnitude is within the specification or not. In the defect-free case, $X2X3 = 10$; otherwise, $X2X3 = 11$ or 00 detect the fault. Note that stage 1 and stage 2 consist of two branches. $I_{BL}(w1)$ and $I_{SL}(w1)$ are mirrored (via P1 and N1) in each of these branches, which drives the four detection outputs $X0$ to $X3$.

In stage 1, as I_{BL} and I_{SL} may exhibit slight deviations as a result of process variation and non-idealities, the two branches are applied together to set a safe margin of the current difference and guarantee a stable output for the defect-free circuit. For example, the transistor sizes of P4 and N5 are the same as P1 and N1 (the width of P1 is 3 times larger than N1 to achieve the same driving capability); while the transistor sizes of P5 and N4 are r times that of P1 and N1. In this setting, P4 and P5 copy the current from P1 to $I_{BL}(w1)$ and $r * I_{BL}(w1)$; N4 and N5 copy the current from N1 to $r * I_{SL}(w1)$ and $I_{SL}(w1)$. Hence, the outputs of $X0$ and $X1$ are according to the current difference. For example, if I_{BL} is close to I_{SL} with the specification (i.e., $1/r * I_{SL} < I_{BL} < r * I_{SL}$), then $(X0X1)$ will be set to '10'; if $I_{BL} > r * I_{SL}$, then $(X0X1)$ will be set to '00'; and if $I_{BL} < 1/r * I_{SL}$, then $(X0X1)$ will be set to '11'. The choice of r value is based on the consideration of process variation, and it is assumed to be 1.2 in our design; the analysis and influence of r will be further discussed in the following. Note that the output $X0X1 = 10$ indicates the presence of Case 1 (defect-free) or Case 4; hence an additional check is needed.

In stage 2, an additional check is performed. The two branches are applied to compare the cell current to two specific reference currents (the minimum and maximum of the correct write current). N6 and N7 copy the current following through N1. If the mirrored current is within the correct boundaries, then outputs of $X2X3$ will be set to '10'. Otherwise, $X2X3$ will be set to '11' (when $I_{SL} > I_{ref-max}$) or to '00' (when $I_{SL} < I_{ref-min}$). The reference currents are generated using the controlled transistor (Nr1) and current mirror (Pr1, Pr2) as shown in the red dotted box of Fig. 7.4. Adjusting the width and length of Nr1 allows the selection of the right reference current. The correct current range is determined based on the five resistance state values introduced in Chapter 2.

DFT implementation for monitoring the $w0$ current, which flows in opposite directions as compared to $w1$ current, uses the same principles.

7.5. VERIFICATION OF DFT METHODOLOGY

This section shows the verification of the proposed DfT circuit. First, we briefly present the simulation setup of this chapter. Then, through simulation results, we illustrate the DfT of this chapter, its advantages compared with previous works. Finally, we analyze the impact of process variations and the robustness of the proposed circuit.

7.5.1. SIMULATION SETUP

To validate the proposed DfT, we implement a 2×2 RRAM array circuit with read circuits and proposed write drivers. We use the TSMC 40 nm 2.5 V transistor model, and the physics-based JART VCM v1b [110] RRAM compact model to implement the circuit. The RRAM model is designed for BS in a Valence Change Mechanism (VCM)-based device as the change of oxygen vacancies in the HfO_2 oxide layer. We applied the JART VCM v1b model to calibrate the defect-free measurement data. The measured (calibrated) 1T-1R device is fabricated by ST Microelectronics, with the stack of (BE/oxide/cap/TE) = (TiN/10 nm HfO_2/10 nm Ti/TiN) [65], [66]. The switching in a nominally defect-free device is bipolar. The spec of five resistance states is defined as the following: Logic '1' is represented by the LRS with $4\,k\Omega < R_{SET} < 20\,k\Omega$, and logic '0' by the HRS with $100\,k\Omega < R_{RESET} < 1\,M\Omega$. The remaining range $[20\,k\Omega, 100\,k\Omega]$ is referred to an undefined state ('U'). The fitting parameter values use the same setting as in Table 6.1 [66]. The circuit is simulated in Cadence's Spectre simulator. The nominal supply voltage for the memory is 2.5 V. In order to accurately evaluate the circuit, capacitive loads are applied to BLs, SLs, and WLs in the simulation. The defect-free circuit is validated by performing write and read operations within specified design parameters; no faults are sensitized during the applied operations.

We perform three experiments:

1) *Process Variation Analysis for Defect-free circuit*: to validate the feasibility of the DfT circuit, the impact of process variations on the functionality of the circuit is simulated. Here, we perform a sensitivity analysis to study the influence of both cell transistors and RRAM devices. For the transistor variations, we use the variation models that include the statistical mismatch from the TSMC 40 nm model library. For RRAM device variations, we set up the models as described in [119]. We incorporate both Device to Device (D2D) (from a truncated Gaussian distribution) and Cycle to Cycle (C2C) (change the variable parameters with confined step size, whose maximum is chosen to be 10 % of the current value) variations of the RRAM cells. For every component combination, we perform 10000 Monte Carlo (MC) simulations, in which all static operations are performed per iteration. For every MC iteration, we record three metrics: 1) write currents, 2) the number of strong faults (functional errors), and 3) the number of incorrect outputs of the DfT scheme.

2) *Detecting Conventional Defects*: we validate the DfT's defect-detecting capabilities by injecting resistive defects into the netlist, using a similar simulation platform in [64]. Conventional defects are modeled as linear resistances and injected in the circuit array, one defect at a time. The defect size ranges from 1 Ω up to 100 MΩ in 81 logarithmically spaced steps. In this chapter, we consider the complete intra-cell and inter-cell defect space of opens, shorts, and bridges. 8 opens, 8 shorts, and 23 bridges are injected in one cell or between at most two adjacent cells (the same as the conventional defect injection in Section 4.2), as listed in Fig. 4.2. We apply all static sensitizing sequences: 0w0, 0w1, 1w0, 1w1, 0r0, 1r1.

3) *Detecting Unique Defects*: we inject the following five RRAM unique defects known in the public domain: OF, UF [59], IUSF [214], ID [65], [194], and OR [66]. These defects are modeled using a *device-aware* defect modeling approach that incorporates the physical behaviors of the defective device [58], [59]. Fig. 7.5(a) presents the dependence

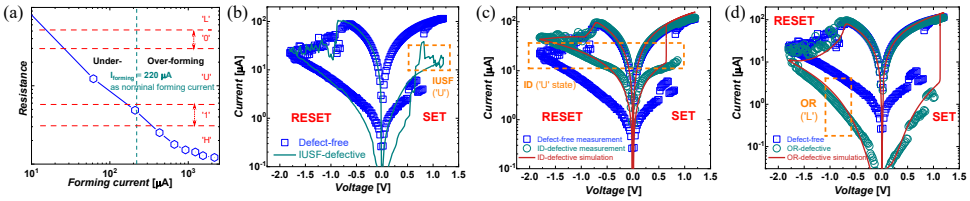


Figure 7.5: Targeted unique defects. (a) Characterization of forming defects [59], [106], (b) Comparison of IUSF-defective and defect-free devices [214], (c) Comparison of ID-defective and defect-free devices [65], (d) Comparison of OR-defective and defect-free devices [66].

between resistances and forming currents. Larger (OF) or smaller (UF) forming currents may make the cell switch into an incorrect state. Fig. 7.5(b) presents measurements of IUSF-defective and defect-free devices. The faulty switching behavior results in a 'U' state during SET operation. Fig. 7.5(c) presents measurements and simulations of ID-defective and defect-free devices. The faulty switching behavior results in a 'U' state during RESET operation. Fig. 7.5(d) presents measurements and simulations of OR-defective and defect-free devices. The faulty switching behavior results in an 'L' state during RESET operation. We inject unique defects in the memory cell by replacing the RRAM model with developed DAT models, one defect at a time. Those DAT models are calibrated with silicon data (the 1T-1R arrays fabricated by ST Microelectronics) and applied in this work. Specifically, we simulate the defect strength that can sensitize unique faults to validate the proposed DfT. In this chapter, the defect strength is represented by values of the following fitting parameters: radius of the filament (r_{det}) for O/UF defects, N_{max} for IUSF, and N_{min} for ID, and OR defects.

7

7.5.2. RESULTS

Next, we present the verification results. First, we validate the correctness of the defect-free circuit with process variation analysis. Second, we present the result for the detection of conventional defects. Third, we present the results for the detection of unique defects.

1) *Process Variation Analysis for Defect-free Circuit*: Fig. 7.6(a) and (b) show histograms for the SET and RESET current flowing through the cell of the defect-free circuit, for 1000 MC simulations. The mean value (μ) of the SET current is 111.31 µA, conforming to the 3 σ design specification (with standard deviation, σ, of 8.52). Similarly, the mean value (μ) of the RESET current is 10.57 µA, also aligning with the 3 σ design specification (with a standard deviation, σ, of 1.82). Currents are normalized to the mean values and shown in Fig. 7.6(a) and (b). The RESET current variation spread is more pronounced than SET, which is also reported in other papers such as [229]. Upon validation of write operations under the process variation, 100 % of the 40000 write operations result in the correct DfT output, demonstrating fault-free.

2) *Detecting Conventional Defects*: The validation of the detection capability of the DfT scheme is valued by the detected fault numbers and the defect coverage. Fig. 7.7 shows faults that are sensitized with related sensitizing sequences and defect strengths for one inter-cell bridge defect cBCC4 (see Fig. 4.2(a)). The figure illustrates faults that

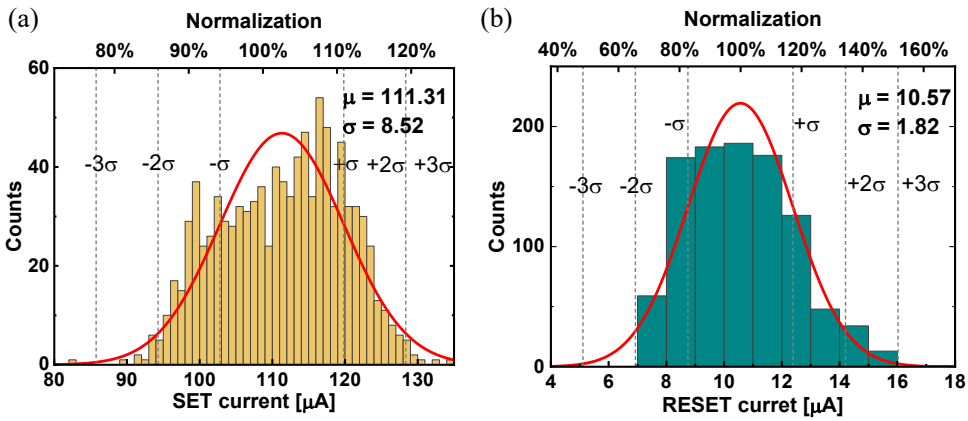


Figure 7.6: Histograms of the current distribution of defect-free circuit under variations. (a) SET ($0w1$) current, (b) RESET ($1w0$) current.

are sensitized by each applied sequence ('S'), together with the defect strength range. We only show write operations since the DfT is based on write current measurement. The green shapes indicate ranges for EtD faults, the orange shapes indicate ranges for sHtD faults, the blue shapes indicate ranges for weak faults, the gray shapes indicate ranges of fault-free cases, and the red shapes indicate ranges detectable by the proposed DfT scheme. It can be concluded that both strong faults and weak faults are sensitized. For example, both $\langle 0w1/0/-\rangle$ (EtD) and $\langle 0w1/U/-\rangle$ (sHtD) are sensitized by $0w1$. The standard March test is effective in detecting the defect strength corresponding to sensitized EtD faults. However, it may fail to detect sHtD faults (e.g., $\langle 0w1/U/-\rangle$) since the 'U' state may result in unstable read output, particularly under the process variation. Furthermore, weak faults are sensitized by $0w1$ when the defect range is from 7.9 kΩ to 158.5 kΩ. The proposed DfT provides incorrect values for this range, which indicates the detection of the defect ranges with corresponding weak faults. Note that the weak fault does not have functional errors but may damage the circuit's lifetime reliability. Similarly, $0w0$ and $1w0$ sensitize no strong faults, but the longest range of weak faults (improve the defect coverage from 7.9 kΩ to 158.5 kΩ). For high test coverage, the longest range of detected faults must be selected; in this case, both $0w0$ and $1w0$ can ensure the maximum defect coverage. Besides, we notice that the write currents of $0w0$ and $1w0$ are correct since this defect supports the RESET operation. Hence, test methods that simply measure the write current will cause test escapes.

We further conduct process variation analysis for the DfT detection robustness. The MC analysis is performed with 1000 iterations for each write sequence within each defect range. Table 7.2 shows the overall results of the MC analysis for defect cBCC4. Here, the result for two defect ranges (the maximum and second-largest defect ranges detectable by the DfT, see Fig. 7.7) are listed with the detection probability of the proposed DfT. There is better detection robustness under variations for $w1$ operations compared to $w0$ operations. It can be explained by: 1) the maximum covered defect strength for $w0$ (158 kΩ) is much larger than it for $w1$ (63 kΩ); hence the large defect range is hard to

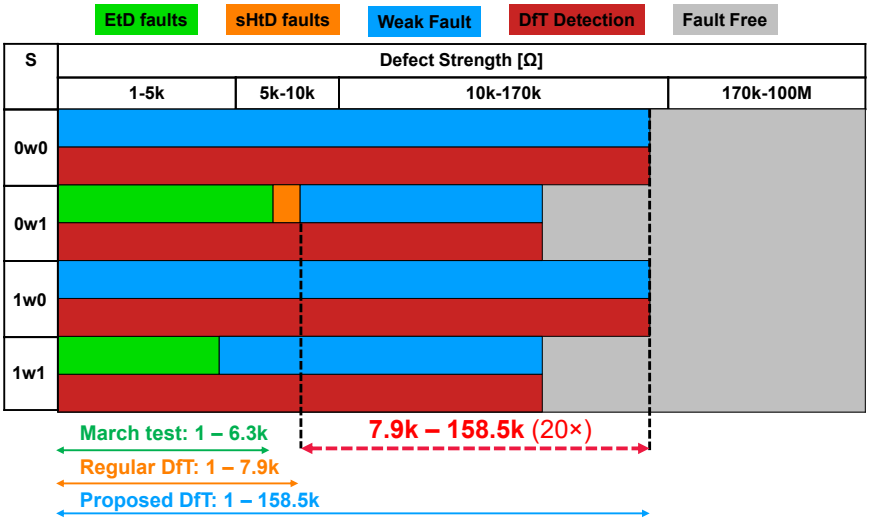


Figure 7.7: Fault map and detection range for defect cBCC4.

detect under variations, 2) the amplitude of $w0$ current is much smaller than the amplitude of $w1$ current (around 10 times); hence it is more sensitive to the variation. Fig. 7.8 presents the detection probability at different defect ranges of cBCC4 for $0w0$ and $1w0$. Compared with the detection at 170 kΩ of the defect range without process variations, the DfT can guarantee the detection until 85 kΩ with process variations. As the defect range increases, the detection probability decreases with process variations. Due to process variations, there are test escapes at the defect range between 85 kΩ and 170 kΩ and yield loss at the range between 170 kΩ and 300 kΩ. Note that no functional fault exists at these defect ranges, which indicates both the March test and existing DfT cannot detect all of them. Besides, it is observed that the process variation affects the sensitization of strong faults with a large defect range. For example, only 404 EtD faults and 90 sHtD faults (of 1000 MC iterations) are sensitized by $0w1$ when the defect range is 7.9 kΩ (strong faults are sensitized up to this range without variation). However, the defect range of 7.9 kΩ can be 100 % detected by the proposed DfT under variations. In summary, the result shows that the proposed DfT is decently resilient against process variations.

Table 7.2: Results of the MC analysis for defect cBCC4.

S	0w0		0w1		1w0		1w1	
MC iteration	1000		1000		1000		1000	
Defect range [kΩ]	126	158	50	63	126	158	50	63
Functional correct	1000		1000		1000		1000	
Detection rate [%]	90.6	62.9	100	96.9	90.5	53.3	100	96.5

Then we present the combined verification result for all conventional defects. We

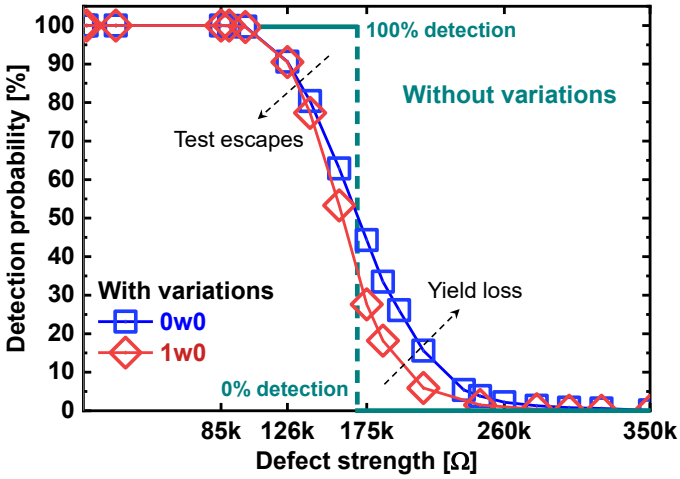


Figure 7.8: Variation analysis for defect cBCC4.

observed that the proposed DfT is able to sensitize a longer defect range (1323) than that by regular March tests (1083) and existing DfT schemes (1107); the proposed DfT improves the defect coverage with 22.16 % and 19.5 %, respectively. The increased defect coverage is due that the DfT can detect additional faults than existing works.

3) *Detecting Unique Defects*: The targeted unique defects can sensitize unique faults, as listed in Table 7.3. For example, the cell remains in the 'U' state after the RESET ($1w0$) in the presence of the ID defect [65]. Hence, the $\langle 1w0/U/-\rangle$ is sensitized at the defect strength of $N = 3 \times 10^{24}\,\mathrm{m}^{-3}$. Note that these unique faults sensitized by unique defects in Table 7.3 cannot be detected by the March test. Due to the real-time monitoring and detection of the writing current by the proposed DfT, the outputs ($X0, X1, X2, X3$) are '1,0,1,1' with the injected OF, '1,0,0,0' with the UF, '1,0,0,0' with the IUSF, '1,0,1,1' with the ID, and '1,0,0,0' with the OR, respectively. The correct output values of $X0, X1$ ('1,0') indicate that equal I_{BL} and I_{SL} flow through the cell since the unique defects are inside the RRAM cells. The incorrect output values of $X2, X3$ indicate that the write current is out of specification due to unique defects. For example, the 'U' state due to the IUSF results in a decreased SET current and thus incorrect output values of $X2, X3$. Note that unique defects such as IUSF, ID, and OR exhibit the *intermittent* behavior since they do not occur in every cycle. In this chapter, we assume the injected defect sensitizes faults in this cycle to verify the detectability of the proposed DfT in the worst case. However, the DfT can monitor the write current online and is guaranteed to detect intermittent faults as long as they are sensitized. Furthermore, the robustness of the DfT to detect unique defects is validated. For example, we perform 1000 MC iterations to the OF-defective circuit with the defect strength of $r_{det} = 80\,\mathrm{nm}$. It shows that the OF defect can be 100% detected under process variations. In conclusion, the proposed DfT has full defect coverage of targeted unique defects, avoiding test escapes from regular tests.

Table 7.3: The output and detection of proposed DfT for unique defects.

Defect	Defect range	Unique fault	DfT outputs
OF	$r_{\text{det}} \in [70,100]$, nm	⟨1w1/H/−⟩	1, 0, 1, 1
UF	$r_{\text{det}} \in [10,30]$, nm	⟨0w0/L/−⟩	1, 0, 0, 0
IUSF	$N_{\text{max}} \in [1.5,4.5], 10^{24}\,\text{m}^{-3}$	⟨0w1/U/−⟩	1, 0, 0, 0
ID	$N_{\text{min}} \in [1.5,4.5], 10^{24}\,\text{m}^{-3}$	⟨1w0/U/−⟩	1, 0, 1, 1
OR	$N_{\text{min}} \in [1,10], 10^{22}\,\text{m}^{-3}$	⟨1w0/L/−⟩	1, 0, 0, 0

7.6. TEST DEVELOPMENT

Next, we develop tests using the proposed DfT that detect as many defect sizes as possible while minimizing test time for conventional and unique defects. Since multiple sequences may sensitize the same fault for a given defect strength, only one representative sequence is required to design the test and ensure full coverage. By analyzing the fault maps obtained from simulation, we identify the minimal set of static operations that still achieves maximum defect coverage across the complete conventional defect space. As a result, the sensitizing sequence set (all static operations) is reduced to: $S_{conv} \in \{0w0, 0w1, 1w0, 0r0, 1r1\}$. These sequences can be combined in a March test as follows:

$$\text{March-Conv} = \{\Updownarrow (\text{w1}); \Updownarrow (\text{w0}, \text{w0}, \text{r0}, \text{w1}, \text{r1})\}.$$

Here, $\Updownarrow$ represents an irrelevant addressing direction, wy, $y \in \{0,1\}$ represents the specific write operation using the proposed DfT, and ry, $y \in \{0,1\}$ represents a regular read operation using the SA. Similarly, the following set of sequences is obtained for the targeted unique defects: $S_{uniq} \in \{0w0, 0w1, 1w0, 1w1\}$. This results in the following:

$$\text{March-Uniq} = \{\Updownarrow (\text{w1}); \Updownarrow (\text{w0}, \text{w0}, \text{w1}, \text{w1})\}.$$

8

Design for Reliability for Read Disturb Faults

Addressing non-idealities in Resistive Random Access Memories (RRAMs) is crucial for their successful commercialization. For example, the inherent resistance drift that occurs during consecutive read operations can induce Read Disturb Faults (RDF), leading to functional errors. This paper analyzes and characterizes the resistance drift and the RDF based on data measurements and presents a physics-based RRAM compact model that incorporates these non-idealities. Additionally, an in-field mitigation scheme is proposed, leveraging bidirectional read operations to balance the resistance. The scheme is implemented and validated through circuit simulations, both for RRAM used as memory and for RRAM-based computation-in-memory microarchitectures for deep neural networks. The results demonstrate that RRAM without any mitigation scheme can start failing after 8,000 consecutive reads, while our mitigation scheme ensures that the memory remains functional even after 10^6 *consecutive reads. Furthermore, the results indicate that using the MNIST dataset as a case study, the accuracy can drop significantly from 86% to as low as 12.5% without any mitigation scheme. In contrast, the proposed mitigation scheme prevents the degradation of accuracy below 84.2%.*

The contents of this chapter were published in ETS'25 [72].

8.1. INTRODUCTION

Research on reliability improvements and mitigation of the impact of non-idealities in RRAMs used as 'memory' or as 'computing device (CIM)' is still in its infancy stage. For RRAM as memories, Error Correction Codes (ECC) were proposed [255], [256] for the detection or recovery of hard and soft errors. In addition, an incremental RESET and verify technique was proposed to enhance both variations and reliability in RRAM memories [257]. However, these works only focus on non-idealities caused by write operations and have a high area or time overhead. The work in [258] applied a Monte Carlo-based methodology to investigate the read disturb mechanism, but it fails to develop a solution to mitigate it. The impact of non-idealities in RRAM-based CIM has been widely studied [71], [259], with mitigation approaches proposed at both the software (mapping) [260], [261] and hardware [71], [243], [262] levels. However, all these studies individually analyze non-idealities from a single perspective, lacking a comprehensive treatment that addresses all related issues. Moreover, many of them are highly vulnerable to process variations, limiting their robustness in practical scenarios. Clearly, there is still a need for robust and efficient approaches to deal with non-idealities in RRAM-based systems.

This chapter focuses on resistance drift and read disturb fault, and proposes a mitigation scheme to address them. The main contributions of this chapter are:

- Analyze the resistance drift and RDF, providing measurements and simulations at both device and circuit levels.
- Propose test and mitigation schemes based on bi-directional read.
- Implement and validate the effectiveness of the different implementations both for RRAM as a memory as well as for an RRAM-based CIM architecture.

8.2. PHYSICS AND CHARACTERIZATION OF RDFS

This section provides the device physical properties behind Read Disturb Fault (RDF), and gives the characterization of RDF both at the device as well as the circuit level. The introduction of RDF can be seen in Section 3.3.2.

8.2.1. PHYSICAL PHENOMENON BEHIND RESISTANCE DRIFT

The RDF is a time-dependent phenomenon where reading a memory cell unintentionally alters its stored data, potentially causing a bit flip. The *resistance drift* of the device is behind this phenomenon [65], [74], [223], [259], [262]–[264]. Read operations utilize a small read current to avoid affecting the programmed state. However, extensive repeated read operations still stress the device, especially for applications that require multiple read operations, switching the state. The likelihood of unintended switching during read is affected by several aspects, including device variability, polarity of the read current, and device state. The RDF should be evaluated for both voltage polarities (SET and RESET). Specifically, reading in the SET direction poses greater risks due to broader voltage distributions and increased variability, potentially leading to unintended state changes. Conversely, applying read voltages in the RESET direction to devices initially in LRS has demonstrated comparatively limited drift, which can be explained by the hard RESET

nature of RRAM devices [74], [223]. Hence, we target the resistance drift of HRS due to the read disturbance in this paper. Furthermore, resistance drift itself is non-linear and state-dependent (low resistance state is hard to SET further), influenced significantly by applied voltage amplitude and pulse duration (even at ns level) accumulated to the device.

8.2.2. DEVICE-LEVEL RDF CHARACTERIZATION

Next, we will provide the characterization of RDFs at a device level. 32 RRAM devices with a (BE/oxide/cap/TE) = (30 nm Pt/5 nm ZrO_2/20 nm Ta/30 nm Pt) stack are fabricated and characterized by performing different voltages in SET and RESET polarity [259]. From the measurements, a device with a nominally defect-free switching is bipolar, where LRS is defined with [2 kΩ, 7 kΩ], and HRS with [15 kΩ, 30 kΩ]; each device is read by a voltage pulse with 20k cycles.

Fig. 8.1(a) shows the resulting resistance distributions of measurements, as a Cumulative Distribution Function (CDF) under the standard normal distribution σ, for different pulse width ranges from 10^{-5} to 10^{-2}s applied in a SET polarity (i.e., a forward read). Cells are in an initial HRS from 15 kΩ to 25 kΩ. The resistance decreases as the applied voltage pulse becomes higher (e.g., from 0.6 to 0.8 V) and wider (e.g., from 10^{-5} to 10^{-2} s), indicating a gradual shift from HRS towards LRS. Clearly, read in SET direction can cause unintended switching, e.g., more than 50% cells switch to LRS by using read pulse of 0.8 V height and 10^{-4}s width.

Fig. 8.1(b) shows the readout results for devices in which four LRS cells and four HRS cells are connected in parallel, under the RESET direction (i.e., a backward read). As observed, applying a read voltage in the RESET direction up to 0.8 V can increase the resistance slightly. Under the same applied voltage, the impact of RESET-direction reading on cell resistance is smaller compared to SET-direction reading, as rupturing CFs is inherently more difficult than forming them.

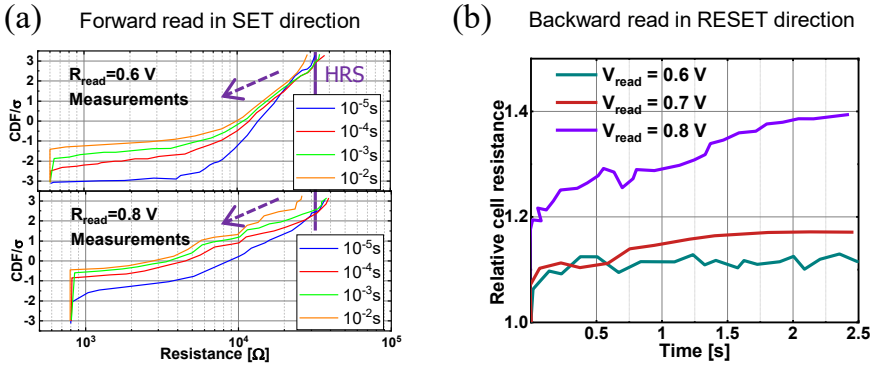


Figure 8.1: Experiments [259]. (a) Forward read in SET direction, (b) Backward read in RESET direction.

8.2.3. CIRCUIT-LEVEL RDF CHARACTERIZATION

We set up a simulation model consisting of a 1T-1R cell and a pre-charge based SA, see Fig. 2.14. We use the JART V1b var model for the RRAM device [119] and calibrate the model with the measurement presented in the previous section.

We applied 200 Monte Carlo (MC) iterations, while performing 4500 forward read operations (defined as read in SET direction) to the device initialized in HRS with a mean value of 22 kΩ, and followed with 4500 backward read operations (defined as read in RESET direction).

Fig. 5.18 shows that cells in HRS exhibit the resistance decrease by applying forward read operations. This resistance drift is state-dependent and variable CF shapes of RRAMs induce the observed variation [119]. Hence, the same initialized resistance states exhibit different degrees of resistance increase or decrease with repeated read operations. For each iteration, the resistance decreases at a nonlinear rate with forward reading; it goes down faster at first and then slower.

Fig. 5.18 also shows that the resistance of cells is able to increase during backward read operations. Note that the 1T-1R cell has an asymmetry between SET and RESET directions (reading in SET direction induces a voltage drop). Consequently, all cells consistently exhibit a gradual increase in HRS when repeatedly read in the RESET direction.

Therefore, applying backward read operations can help mitigate the resistance reduction, which provides a promising opportunity for bi-directional read approaches.

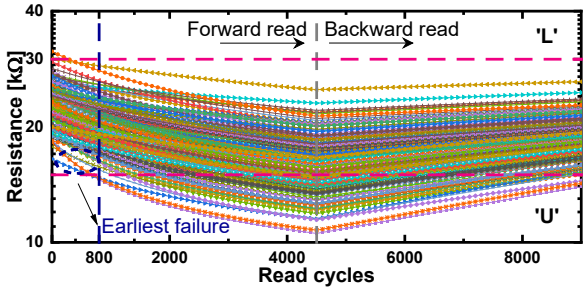


Figure 8.2: The resistance changes when performing forward and backward read operations.

8.3. THE BI-DIRECTIONAL READ APPROACH

This section proposes a mitigation scheme for the RDF; it is based on a bi-directional read of the RRAM devices. Different implementations are provided, together with a trade-off analysis of such implementations.

8.3.1. CONCEPT AND POTENTIAL IMPLEMENTATIONS

The resistance of bipolar RRAM either decreases (causing the RDF) or increases (recovering) by applying positive or negative voltages. Thus, a reverse read direction can boost device resistance and recover it from the RDF. Fig. 8.3 illustrates the concept of bi-directional read for RRAMs. It allows changing the reading current directions through the cell by selecting the control signal D. When D=0, a normal read (i.e., forward) is per-

formed, causing a shift or decrease in the RRAM resistance. When D=1, the reverse read is performed, enabling the recovery of the RRAM resistance (see also Fig. 5.18). There are three schemes for such an approach: 1) a *static* method based on applying bi-directional reads with a fixed ratio, 2) an *enhanced static* method that can reduce costs, and 3) a *dynamic* method based on online monitoring of the RRAM state (using dedicated sensors to switch the read direction at the right time).

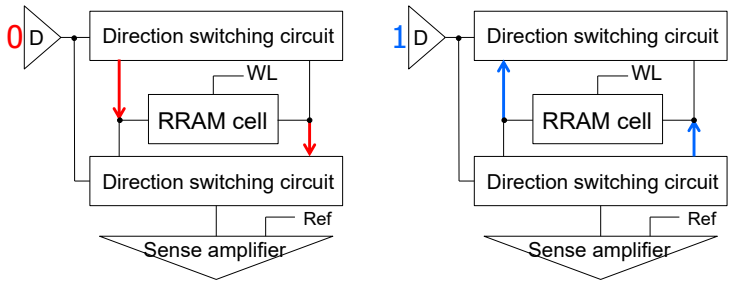


Figure 8.3: The bi-directional read circuit with alternate current paths shown in red and blue.

STATIC APPROACH

The direct approach involves applying static implementations, i.e., bi-directional read operations with a *fixed* forward-to-backward ratio, typically determined through measurements to balance RRAM resistance change. These measurements are essential for characterization at both the device and circuit levels. However, it has some limitations; e.g., it cannot accommodate the wide variability RRAM is inherently suffering from in HRS, and it is relatively hard to implement when considering the management of the access of each address.

ENHANCED STATIC APPROACH

This approach is designed to improve the aforementioned approach by accounting for the unequal distribution of workloads and data access frequency across the array. It has been reported that memory-cell accesses follow a power-law distribution, with approximately 84% of total accesses concentrated in the top 30% most frequently accessed cells [265]. For instance, Fig. 8.4 shows the results of a high level simulation (C++) we performed using RRAM array model of a Deep Neural Network (DNN) for the MNIST classification task. The figure shows a heatmap depicting the distribution of read cycles across individual memory cells. The color bar represents the number of read accesses for each cell, with the gradient increasing from light blue to dark blue. Due to the inherent sparsity of neural networks, read operations for rows that correspond to zero-valued inputs are skipped, leading to different access frequencies per row. The heatmap illustrates read access patterns, indicating that certain parts of the array are accessed more frequently due to inputs of the DNN model. To efficiently track frequently accessed cells, we propose to keep a dynamically updated list of recently accessed addresses with the number of accesses per address. If the address is not on the list, it will be added on the top with the address count set to one. If an address is already on the list, it gets put on the top with

its increased address count. If the list is full, and another address needs to be added, the bottom address will be removed from the list. In this way, the recovery scheme will only be applied to frequently accessed cells.

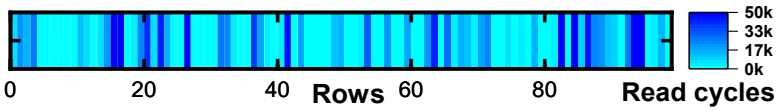


Figure 8.4: Read cycle distribution across DNN array.

DYNAMIC APPROACH

This approach utilizes an in-field, sensor-based solution to monitor the resistance drift and adjust the read direction, at the right time. Fig. 8.5(a) shows the high-level implementation, including the core array (C0-C3), dummy cells (D0 and D1), and peripheral circuits. The dummy cells are placed in a single column, and share WL's and SL's drivers, as well as registers, per row. They act as sensors, tracking maximum read cycles and triggering read direction switching. Whenever a row cell is accessed, the corresponding dummy cell is simultaneously read, and its resistance is recorded; hence, it presents the worst-case scenario. Upon switching the dummy cell resistance state, the control signal D is triggered to execute the read operation in the reverse manner.

Furthermore, we consider the worst-case situation if the array cells are over-repaired by backward reads, as shown in Fig. 8.5(b). Consider a cell with an initial HRS (R_{int0}) that is read in the forward direction until the resistance reaches its minimum range (R_{min}). After a write operation, the cell returns to its original R_{int0}, while the dummy cell remains inactive, maintaining its resistance state. Following consecutive read operations, the dummy cell triggers a reverse read operation due to falling below the HRS threshold. Subsequent backward reads on the cell increase its resistance value until the dummy cell resistance reaches the R_{max}. To ensure proper reference resistance, the upper resistance should be less than R_{max}, as $R_{\text{max}} - \Delta, \Delta = R_{int0} - R_{\text{min}}$. This setting can guarantee that the array cells cannot switch over R_{max} with backward reads.

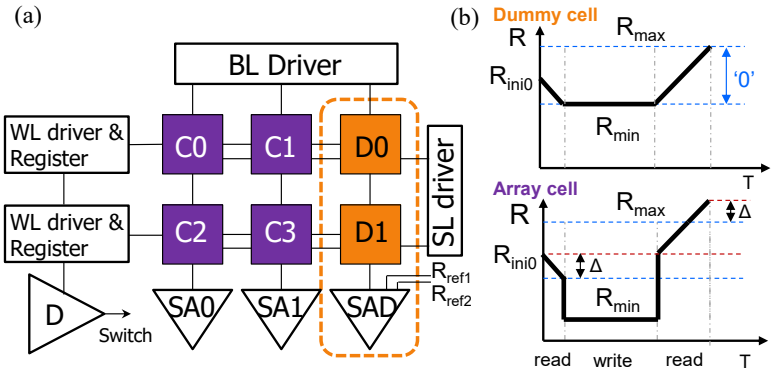


Figure 8.5: The high-level implementation of the in-field test approach.

8.3.2. TRADE-OFF FOR APPROACH SELECTION

Applying a static read ratio requires detailed measurements. Cell-based counters need to be used to implement the static approach. Depending on workloads, the scrolling list can help further apply the approach only to frequently accessed cells. However, the blind approaches are limited by inadequate consideration of variations. Thus, including the real-time sensor to switch read direction properly when measurements are insufficient is more reliable and robust to variations, as opposed to using a dummy cell that typically assumes worst-case conditions, leading to possible over-repair.

8.4. VALIDATION

This section shows the validation of the bi-directional read approaches. First, we briefly present the simulation setup. Then, through simulation results, we demonstrate that the proposed approaches are effective in addressing the RDF. Finally, we analyze the feasibility of the bipolar read approach for improving the accuracy of CIM implementation.

8.4.1. SIMULATION SET UP

To estimate the read disturb behavior, the RRAM resistance drift under periodic read was simulated for HRS. The RRAM array circuit in Fig. 8.5 is built in Cadence's Spectre simulator by using the TSMC 40 nm transistor library and the physics-based JART VCM v1b [110], [119] RRAM compact model, which is calibrated with measurements in [259]. The pre-charged SA performs the bi-directional read operations with the control signal D. The positive read is performed by the current discharge from the pre-charged SA to the SL, while the negative read is performed by the current discharge from the pre-charged SA to the BL. The supply voltage V_{DD} = 2.5 V and read voltage V_{read} = 1.1 V. The V_{read} is selected to sufficiently surpass the transistor voltage drop (V_{th} is around 0.8 V) for current sensing. The read period is 12 ns for both positive and negative read operations.

8.4.2. VALIDATIONS FOR STORAGE MEMORY ARRAY

VALIDATION OF THE STATIC APPROACH

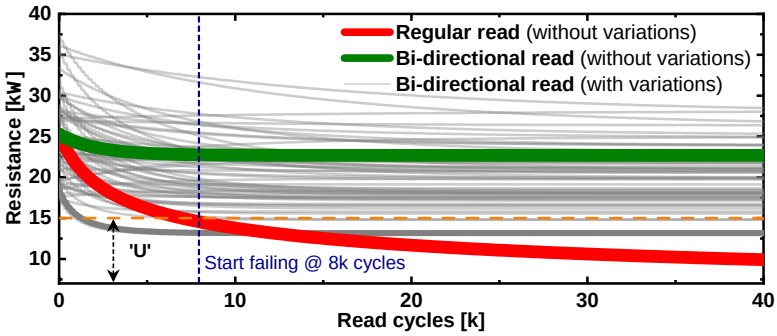


Figure 8.6: Resistance drifts for bi-directional read compared to regular read.

To validate the approach, a $m:n$ switching ratio is implemented by performing m

forward reads, followed by n backward reads. we set the ratio to $r = 1 = \frac{n}{m}$. Fig. 8.6 shows the final resistance states after 40k read cycles for three cases: a) by performing regular read operations (i.e., only forward read operations), b) by performing bi-directional read with ratio 1 assuming the ideal scenario (i.e., without variations), and c) the same as in (b), but taking the variability into consideration. The figure shows that applying regular reads will cause HRS to drift to 10 kΩ after 40k cycles, and become undefined already after 8k reads. On the other hand, bi-directional read is quite effective to mitigate RDF. However, in the extreme variety cases, the scheme may fail; the bold gray line in the figure for bi-directional read (with variations) shows that the cell enters an undefined state.

VALIDATION OF THE ENHANCED STATIC APPROACH

For our experiment, we assume 1M read accesses to a 256 × 256 RRAM array; 80 % of these reads are randomly applied to cells inside a core subarray of size $s \times s$ (which are the ones to likely cause RDF), while 20 % are applied to other cells [265]. We examined the dependency of the list size, in terms of entries E, on the percentage of the recorded cells exhibiting the RDF. The results are shown in Fig. 8.7 for five cases; note that for s=10, two options are reported (with and without considering variability). In the figure, each solid line shows the probability of recording cells starting to fail; as shown in the previous section, the RRAM cells start failing after 8k read cycles (see also Fig. 8.6). As expected, the minimum E to achieve 100% recording of RDF cells increases gradually as the core subarray size grows; however, E increases much slower than the increase in the core subarray size. For example, the minimum E needed in case $s = 100$ is about 50× larger than that needed when $s = 10$. It is also worth noting that when considering the variability, the minimum required E is relatively smaller than in the mean case. The dashed red line in Fig. 8.7 shows the probability of recording cells (for $s = 10$) that have over 700 read access cycles, which are the worst case with variations (see Fig. 5.18). Cells can be 100 % recorded with $E = 400$, smaller than $E = 480$ for the mean case. Limitations of this approach include: 1) power and area overhead for scrolling the update list, and 2) test escapes due to insufficient E when cell workload is evenly distributed.

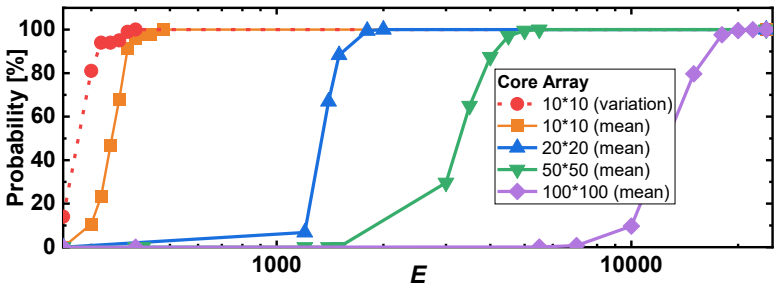


Figure 8.7: The probability of counting cells with RDF.

VALIDATION OF THE DYNAMIC IMPLEMENTATION

This approach is validated in the circuit by switching to perform backward reads and thus repairing memory cells when the sensor detects the RDFs of dummy cells. Fig. 8.8

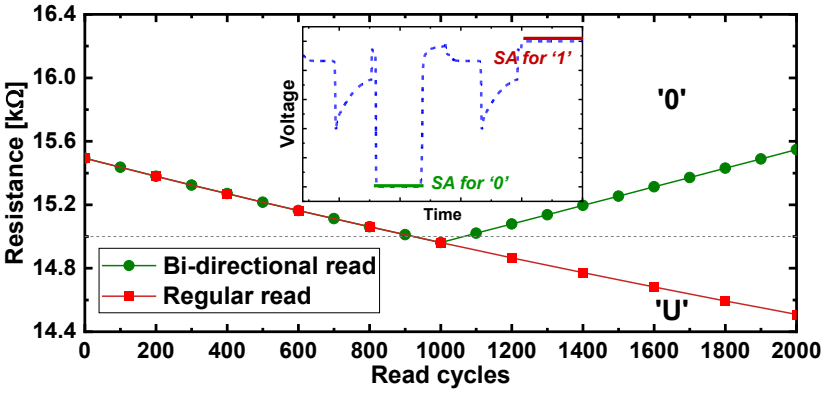


Figure 8.8: The effectiveness of the in-field test and repair approach.

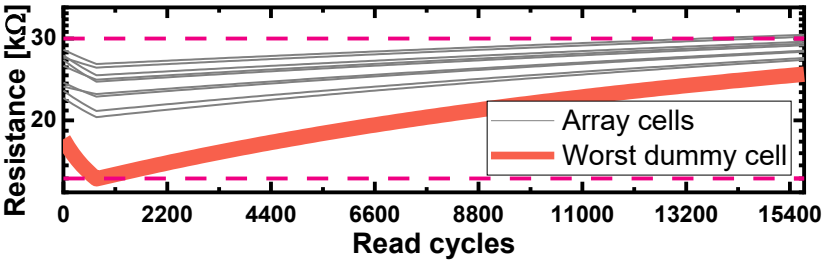


Figure 8.9: The worst case of the dummy cell setting with variations.

shows how the resistance of the RRAM device degrades over time when performing regular reads and how bi-directional read enables the recovery after 1000 cycles; note that the reference value is set to $R_{\text{ref1}} = 15\,\text{k}\Omega$ (see section 8.3). The resistance decreases from 15.5 kΩ to 14.5 kΩ after 2000 consecutive regular read operations. With the dynamic approach, the output of the SA changes from GND (read output for '0') to V_{DD} (read output for '1') after 1000 regular reads, thus triggering the backward read direction to repair the cell. Hence, it can be repaired to 15.5 kΩ by performing 1000 backward read operations after 1000 forward read operations. The approach provides correct in-field testing and repair with the application of a bi-directional read scheme. Furthermore, the robustness is validated by assuming the worst case with variations: the dummy cell resistance decreases very fast with forward reads until the boundary (700 cycles, see Fig. 5.18) and triggers cells to perform backward reads. As shown in Fig. 8.9, by performing 700 forward reads and 13k backward reads on other cells that decrease the least, they are still within the HRS specification.

In conclusion, the dynamic approach can easily ensure the robustness of RRAM devices against RDFs; the exact resistance value of the RRAM does not matter as long as we guarantee that it is within the specification of the '0' state.

8.4.3. VALIDATIONS FOR SYSTEM-LEVEL CIM INFERENCE

As the impact of RDF intensifies with frequent read tasks, we validate the proposed approaches using an RRAM-based CIM inference engine [266]. The applied platform is developed to emulate the online learning classification task with an MNIST handwritten dataset in a 2-layer (400 × 100 × 10) DNN-based RRAM array. Each weight is mapped to one RRAM cell. We simulate the RDF as resistance drift according to read cycles only for hardware inference, as the hardware training is assumed to be fault-free. The accuracy is determined as the maximal value from 30 epochs in each inference. As shown in Fig. 8.10, regular reads up to 27k cycles lead to a significant accuracy drop, from 86% to 12.5% without variations, and from 48.32% to 10.27% under worst-case variation conditions.

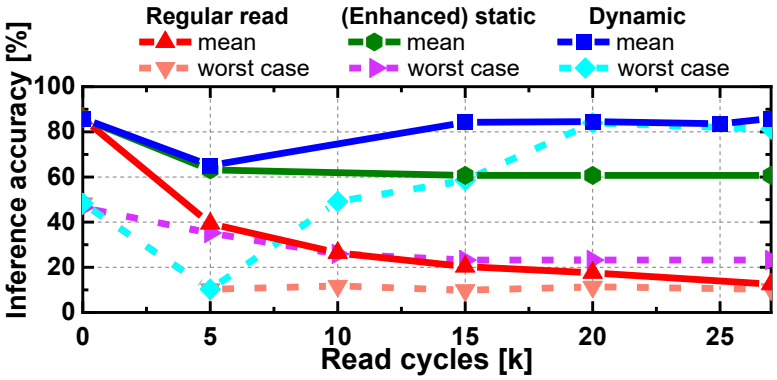


Figure 8.10: Inference accuracy enhancement by bi-directional read approaches.

VALIDATION OF THE STATIC AND ENHANCED STATIC APPROACHES

The green line in Fig. 8.10 shows that applying the static approach improves the inference accuracy from 12.5 % to 60.69 % after 27k read cycles, compared with applying regular reads. Considering the worst case with variations, the accuracy starts at 46.47 % and drops to 23.25 % after 27k read cycles, which is still higher than the accuracy using regular reads without variations. The enhanced static approach can be applied to the static approach, achieving the same accuracy in a more cost-effective manner.

VALIDATION OF THE DYNAMIC APPROACH

Fig. 8.10 shows the validation of the dynamic approach. The reference values are set to $R_{\text{ref1}} = 20\,\text{k}\Omega$, $R_{\text{ref2}} = 30\,\text{k}\Omega$. When the read cycles reach up to 5k, the accuracy first drops and recovers to 65.09 % since the resistance decreases and triggers the sensor. Then, the accuracy is enhanced to 84.19 % when the resistance is recovered. After that, the resistance ($> 25\,\text{k}\Omega$) keeps increasing until $30\,\text{k}\Omega$, resulting in a slight drop in accuracy but still larger than 80 % up to 25k cycles. Finally, the sensor is triggered again, recovering the accuracy to 86%, after which the process repeats. Furthermore, the worst-case is considered as all array cells drop fast (orange line in Fig. 8.9) but the dummy cell drops and triggers the read direction slowly. The accuracy first drops until the dummy cell

triggers the sensor. Then, the accuracy is enhanced to 80.5 % up to 27k cycles, which shows the robustness against the variation.

The in-field dynamic approach is robust to variations and validated to enhance the CIM accuracy. Further enhancements are possible by fine-tuning the reference range, enabling early repair of cells with small margins.

8.5. DISCUSSION

The proposed bi-directional read scheme offers several advantages. Most importantly, it effectively mitigates resistance drift faults (RDF) in RRAM, thereby improving read reliability and sustaining inference accuracy in CIM applications. Compared with conventional read operations, the scheme provides enhanced robustness against both time- and state-dependent variations. In addition, it can be integrated with existing DfT techniques, such as multi-reference trimming, to further improve fault coverage and overall system reliability.

Despite these strengths, the scheme also has limitations. The static approach requires cell-level counters to control the ratio of forward and backward reads, leading to a large area overhead that is impractical for large arrays. Although the enhanced static and dynamic approaches reduce this overhead, they still introduce additional circuit complexity and power consumption. Another limitation is the dependency of effectiveness on workload characteristics; for lightly accessed arrays, the benefits may be less pronounced.

From an implementation perspective, the dynamic approach with row-based dummy cells is more hardware-efficient and easier to scale compared to the static approach. This makes it particularly attractive for CIM arrays, where energy and area are critical constraints. For conventional storage RRAM arrays, the scrolling list recorder and enhanced static approaches may be more suitable, as they can balance accuracy and area overhead depending on access frequency.

Moreover, the schemes can be further improved by optimizing control logic to minimize overhead, introducing adaptive mechanisms to automatically adjust read ratios based on real-time drift behavior, and exploring co-optimization with error correction or redundancy strategies. These improvements would enhance both the practicality and the efficiency of the proposed designs across diverse application scenarios.

9

DIAGNOSIS FOR CONVENTIONAL AND UNIQUE DEFECTS

Due to the immature manufacturing process, RRAMs are prone to exhibit unique defects, which should be efficiently identified for high-volume production. Hence, obtaining diagnostic solutions for RRAMs is necessary to facilitate yield learning, and improve RRAM quality. Recently, the Device-Aware Test (DAT) approach has been proposed as an effective method to detect unique defects in RRAMs. However, the DAT focuses more on developing defect models to aid production testing but does not focus on the distinctive features of defects to diagnose different defects. This chapter proposes a Device-Aware Diagnosis (DA-diagnosis) method; it is based on the DAT approach, which is extended for diagnosis. The method aims to efficiently distinguish unique defects and conventional defects based on their features. To achieve this, we first define distinctive features of each defect based on physical analysis and characterizations. Then, we develop efficient diagnosis algorithms to extract electrical features and fault signatures for them. The simulation results show the effectiveness of the developed method to reliably diagnose all targeted defects.

The contents of this chapter were published in DATE'24 [73].

9.1. INTRODUCTION

It is generally recognized that variations and defects in device characterization throughout the manufacturing process, create significant challenges [59], [68], [267]. Hence, there is a pressing need for a comprehensive understanding of manufacturing defects and the development of high-quality test solutions. Moreover, the seamless combination of testing with diagnosis is paramount in achieving a holistic approach to improving RRAM manufacturing processes and yield learning. Therefore, effective diagnosis methods are as important as defect detection to enable high-volume detection [267], [268].

Although there are several works on test and diagnosis of RRAMs, most of the work focuses on detection rather than diagnosis. Compared to testing, works of RRAM diagnosis are still limited. In 2017, Li, Bi, Jing, *et al.* leveraged judicious control of the sneak-paths to design a fault isolation technique and diagnosis algorithm [269]. In 2022, a March diagnosis algorithm was designed for distinguishing conventional RRAM defects [267]. To address non-linear defects inside the RRAM, the DAT approach has demonstrated its power in modeling and testing RRAM-related unique defects; however, it is not practical to diagnose these defects, as its primary optimization focuses on *defect detection* instead of *defect diagnosis*. For instance, one test solution proposed in [48] for forming defects could detect other defects as well, thus it is inefficient for diagnosis. Therefore, a dedicated method of diagnosing unique defects is required to enhance the quality of RRAM fabrication further.

This chapter focuses on the diagnosis of both conventional and unique defects in RRAMs. The main contributions of this chapter are:

- Present the framework of the DA-Diagnosis method for conventional and unique defects in RRAMs.
- Analyze and define distinctive features of all targeted RRAM defects, especially unique defects.
- Develop diagnosis algorithms based on the distinctive features.
- Demonstrate the effectiveness of developed diagnosis algorithms for targeted conventional and unique defects.

9.2. DEVICE-AWARE DIAGNOSIS APPROACH

Fig. 9.1(a) shows the DAT approach; it consists of four steps: defect characterization, defect modeling, fault modeling, and test development [59], [60]. DAT has been shown to be very powerful in modeling and testing unique defects in emerging memories such as RRAMs [65], [214] and STT-MRAMs [58]. It focuses on maximizing the fault/defect coverage while optimizing the test time. However, such an approach cannot be applied straightforwardly to diagnosis as it cannot uniquely identify which defect causes the detected fault. For example, the March test and DfT designed for Ion Depletion (ID) defects in RRAMs in [65] can detect other conventional defects; hence they fail to provide the distinctiveness for defects. Therefore, the DAT approach needs to be adapted for diagnosis.

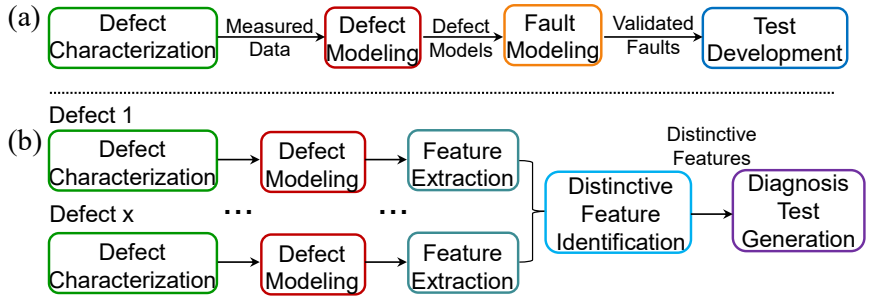


Figure 9.1: Framework of RRAM test and diagnosis development: (a) DAT process, (b) DA-diagnosis process.

Fig. 9.1(b) shows the Device-Aware Diagnosis (DA-diagnosis) approach; it consists of five steps. Note that steps 1, 2, and 3 have to be repeated for each targeted defect (to be diagnosed); the results will be then processed to create a 'feature dictionary', which will be the basis to generate diagnosis test algorithms. Note also that steps 1 and 2 for DAT as well as for DA-diagnosis are the same. The platform was first applied for Spin-Transfer Torque Magnetic RAMs (STT-MRAMs) in [270], and adapted to apply for RRAMs in this work. In the rest of this section, we will discuss the five steps of the DA-diagnosis approach.

1. **Defect Characterization:** In this step, the targeted defects are determined. Defect models and characterization data of calibrated defects are used to develop a library of defects and their features.

2. **Defect Modeling:** This step appropriately models defects. The conventional defects are modeled as linear resistors [43], while unique defects require DA defect modeling, which describes the impact of the defect on the device technology parameters and thereafter on electrical behavior and obtains a defective device model [59]. DA defect models have been shown to be accurate defect modeling of unique RRAM defects [59], [65], [66], [214].

3. **Feature Extraction:** This step extracts features of each defective device based on the way the defect manifests itself in the functional behavior of the memory. The features of RRAM refer to the electrical parameters of the memory device. For RRAM, these are V_{SET}, V_{RESET}, R_{SET}, and R_{RESET}; they can be extracted from the simulation or measurement. Combining all extracted features for all targeted defects results in an initial feature dictionary.

4. **Distinctive Feature Identification:** The distinctive features of each defect are selected from the feature dictionary. Here, a distinctive feature indicates a unique behavior caused by one or more electrical parameters due to the presence of a defect; such behavior is uniquely associated with that defect. For example, a cell enters 'H' state only if a certain defect (i.e., over forming) is presented in the memory; no other defects can cause such behavior.

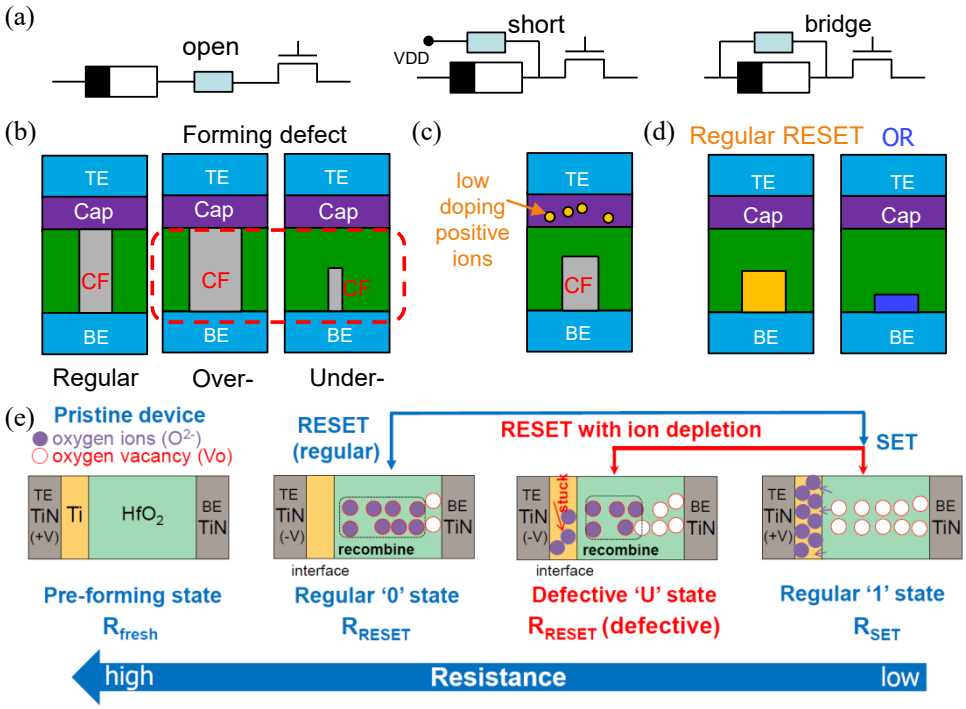


Figure 9.2: Physical mechanism of defects. (a) Conventional defects [43], [52], (b) Forming defects [59], (c) IUSF [214], (d) OR [66], (e) ID [65].

5. **Diagnosis Test Generation:** Based on distinctive features, this step generates a set of diagnosis test algorithms that are able to distinguish which defect is causing certain faults; i.e., if one or more tests from the set fail, then we can derive (based on the signature created) which defect the memory suffers from. Note that in our analysis, we assume a single defect at a time. Other diagnosis methods based on Physical Failure Analysis (PFA) are not considered since they require expensive, time-consuming, and destructive analysis [271]. Finally, the output of this step is the DA-Diagnosis algorithm for each defect.

9.3. APPLICATION OF THE METHODOLOGY FOR RRAMS

In this section, we follow the DA-Diagnosis framework to design diagnosis methods for targeted defects.

9.3.1. TARGETED DEFECTS AND THEIR CHARACTERIZATIONS

The targeted defects consist of conventional defects and unique RRAM defects.

Conventional defects [43], [52]: These consist of well-known interconnect and contact defects, which can cause opens, shorts, and bridges. They can be modeled accurately by *linear resistors* [43], [52], see Fig. 9.2(a). Both write and read faults can be sensi-

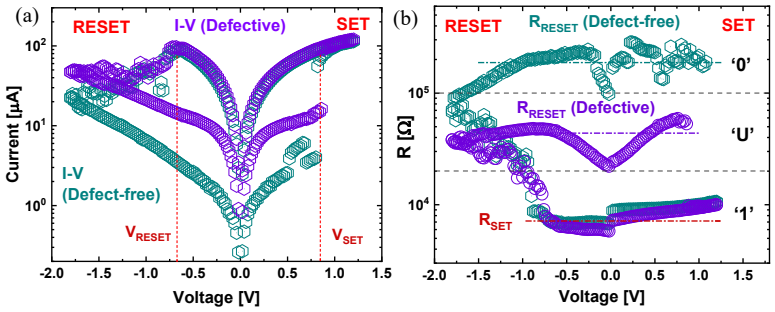


Figure 9.3: Characterization of defect-free and ID-defective device: (a) I-V measurement, (b) R-V measurement.

tized by these defects.

Unique RRAM defects: These consist of defects that are found inside the RRAM device due to fabrication imperfection. As of now, there are five unique defects known in the public domain; they are explained next.

- **Over/under forming (O/UF) [59], [106]:** The forming defect is introduced by an inappropriate forming current. The external tester usually controls the forming current; hence an unstable voltage source may affect the forming process [49]. The forming defect may come in two variants: over (under) forming when a too-large (small) CF is generated, as presented in Fig. 9.2(b), initializing to a resistance state that is much lower (higher) than the nominal LRS.
- **Intermittent Undefined State Fault (IUSF) [214]:** As presented in Fig. 9.2(c), the IUSF is related to imperfect capping layer doping or OF, which causes the RRAM device to *intermittently* change its switching mechanism from bipolar to complementary switching, leading to an undefined state during SET process.
- **Ion Depletion (ID) [65], [194]:** As presented in Fig. 9.2(e), the ID defect is caused by interface physical imperfections, which involve an increased oxygen trap of the defective capping layer, resulting in a lower recombination rate between insufficient O^{2-} and OV.
- **Over RESET (OR) [66]:** As presented in Fig. 9.2(d), the OR phenomenon is caused by a reduced O^{2-} storage capacity (oxygen affinity), and extra ions come back to recombine the CF, leading to an *intermittent* extremely high resistance state exceeding the HRS criteria.

For the characterizations, we perform I-V, R-V, and R_{SET}-I_{forming} measurements for the above defects. For conventional defects, we characterize them with the RRAM compact model in [223] in Cadence's Spectre simulation, following the process in [48], [64]. For unique defects, we have reused the characterization results produced in [59], [65], [66], [106], [214]. Measurements are produced on a 7×7 1T-1R array. The I-V curves are performed by the Keysight B1500 parameter analyzer and extracted by Python. After that, the R-V can be converted from I-V measurements. For example, Fig. 9.3(a) and

(b) present measurements and compare those for defect-free and ID-defective devices. A current increase and R_{RESET} decrease can be observed for I-V and R-V curves of the defective device. Besides, we also investigate the relationship between resistance and forming currents (R_{SET}-$I_{forming}$) based on HfO_2-RRAM measurements in [106]. Not all measurements are shown in this paper due to space limitations.

9.3.2. DEFECT MODELING

In this step, we directly use linear resistance for conventional defects [43], [64] and DA-defect models of the 5 targeted unique defects reported in previous works in [59], [65], [66], [214].

9.3.3. FEATURE EXTRACTION

In this step, we summarize the feature dictionary for targeted defects through measurements and simulation in Table 9.1. The symbols used in the table are listed as notations. For example, in the presence of IUSF, there is a certain probability that the R_{SET} increases intermittently. Especially, we present possible faulty states ('H', 'U', 'L') that are sensitized by each defect when R_{SET} and R_{RESET} are affected by defects as features.

Table 9.1: Feature dictionary of targeted defects.

Defects		Electrical parameters					
		R_{SET}		R_{RESET}		$\|V_{SET}\|$	$\|V_{RESET}\|$
Conventional	Interconnects & Contacts	↑, ↓, NA	U L	↑, ↓, NA	U L	↑, ↓, NA	↑, ↓, NA
Unique	OF	↓	H	NA		↓	↑
	UF	↑	U	↑	L	↑	↓
	IUSF	↑ (IB)	U	NA		NA	NA
	ID	NA		↓ (IB)	U	NA	NA
	OR	NA		↑ (IB)	L	↑ (IB)	NA

Note: ↑: increase; ↓: decrease; NA: not affected; (IB): intermittent behavior; H/U/L: faulty states

9.3.4. DISTINCTIVE FEATURE IDENTIFICATION

This step aims to identify distinctive features of each defect. If a feature is unique in each column of Table 9.1, it is considered a distinctive feature. Consequently, distinctive features of targeted defects can be directly derived from Table 9.1. For example, the *intermittently (IB)* increased R_{RESET} is identified as a distinctive feature for the OR defect; yet the permanent increased R_{RESET} is not a distinctive feature, since the same feature can also be found in the presence of conventional defects and UF. For the defects that have multiple distinctive features, we can select those that facilitate the most effective diagnosis. For the defects without distinctive features, we can also consider combining features to make them distinctive. Next, we present distinctive features of each targeted defect.

- **Conventional defects:** The impact on features of conventional defects also depends on the precise type (bridges, shorts, or opens), range, and location of the

defect. For example, we observed that bridges in parallel with the cell affect both R_{SET} and R_{RESET}; while some shorts affect only one of the resistive states [48], [64]. Hence, conventional defects lack distinctive features, which will be discussed later.

- **OF:** As presented in Table 9.1, the 'H' state of R_{SET} can be considered as the distinctive feature because other defects cannot make the cell switch to 'H'. Besides, $|V_{RESET}|$ is another distinctive feature compared with other *unique defects*. We can even weaken the RESET write pulse to better identify the unique $|V_{RESET}|$.
- **UF:** As presented in Table 9.1, there is no ensured distinctive feature of UF. However, $|V_{RESET}|$, R_{SET}, and R_{RESET} are distinctive features compared with other unique defects. Especially, the 'L' state R_{RESET} can be combined with other features to facilitate diagnosis, which will be discussed later.
- **IUSF:** As presented in Table 9.1, R_{SET} is the distinctive feature of IUSF. Especially, this defect occurs intermittently with different cycles owing to the intrinsic stochasticity. The intermittent behavior indicates that the defect has a probability of occurring rather than exhibiting a constant resistance in different cycles, like the conventional defects.
- **ID:** As presented in Table 9.1, an intermittently reduced R_{RESET} is the distinctive feature. Other defects, such as conventional defects, can also lead to reduced R_{RESET} but without the intermittent behavior.
- **OR:** As presented in Table 9.1, intermittently increased $|V_{SET}|$ and R_{RESET} are distinctive features of OR.

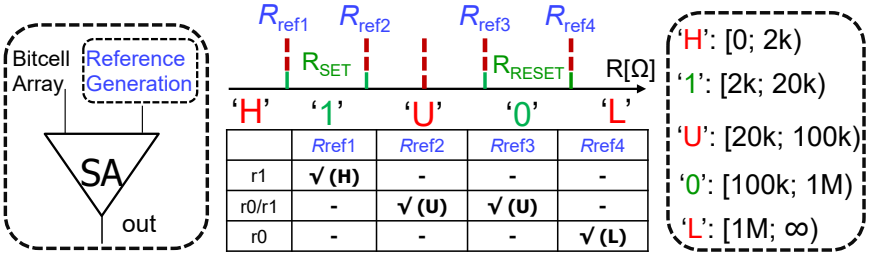
In conclusion, among those unique defects, OF sensitizes unique 'H' state faults; IUSF, ID, and OR all exhibit intermittent faulty behavior owing to the intrinsic stochasticity [99]. These specific faulty behaviors are impossible to occur in the presence of conventional defects. Hence, these unique defects will not be mixed with conventional defects. Unfortunately, UF may sensitize faults that are also sensitized by conventional defects; hence, it is critical to distinguish them.

To achieve this, we propose a process based on the comparison of faults that are sensitized by different defects. The assumption is that every defect sensitizes a different set of faults. Hence, comparing these fault sets of each defect can facilitate the diagnosis. The whole process is: 1) obtain a set of sensitized faults in the presence of all defects, 2) compare obtained fault sets of the targeted unique defects with the other fault sets; diagnose the defect as either a conventional or an unknown defect when there is no overlap in the comparison. If there is an overlap, proceed to the third step for further diagnosis, and 3) further compare fault sets of unique defects with the set of faults that are sensitized by all conventional defects. If there is no overlap in this comparison, diagnose the defect as either a conventional or a unique defect. Otherwise, a diagnosis cannot be guaranteed. Another benefit of this process is that once we complete the first step, we will be able to identify the location of faulty cells, allowing us to save time by applying the algorithm only to the faulty cells. In our case study, all faults sensitized by UF do not overlap with conventional defects of any strength [48], [64]; hence, it is possible to distinguish them.

9.3.5. DIAGNOSIS TEST GENERATION

This step generates DA-Diagnosis methods for all RRAM-targeted defects. The diagnosis design first aims to distinguish between conventional and unique defects and then applies patterns to further distinguish which unique defect is present.

From the distinctive features obtained in Table 9.1, it is necessary to detect not only regular '0' and '1' states, but also faulty 'H', 'U', and 'L' states. In this chapter, the diagnosis methods are based on March algorithms with adjustable read operations. A regular SA design [122] with multiple yield learning references is applied. Fig. 9.4 shows the implementation of the reference setting for three specific read operations. Those references are set according to the defined resistance states and with a checkmark to indicate the reference used. For example, the reference (R_{ref1}) equal to the minimum R_{SET} is applied to detect the 'H' state. Two references (R_{ref2}, R_{ref3}) are applied to detect 'U' from '0' and '1'. The reference (R_{ref4}) equal to the maximum R_{RESET} is applied to detect the 'L' state. Next, we apply the above process and present the diagnosis designs for each defect. Table 9.2 summarizes the final result of this section.



	Rref1	Rref2	Rref3	Rref4
r1	√ (H)	-	-	-
r0/r1	-	√ (U)	√ (U)	-
r0	-	-	-	√ (L)

Figure 9.4: Reference settings and resistance state specifications.

Table 9.2: Overview of diagnosis for unique defects.

Defect	Mechanism	Related steps in RRAM fabrication	Test method	Diagnosis algorithm	Notation
OF	Higher forming current	Forming step	$\Updownarrow(w0,w1,w0)$;	$\Updownarrow(w1,r1)$	Multiple references
UF	Lower forming current	Forming step	$\Updownarrow(r0)$; $\Updownarrow(w1,r1)$ [48]	$\Updownarrow(w0,r0,w1,r1)$	
IUSF	Reduced oxygen storage capability	(1) Over forming (2) Capping layer doping, Annealing	$\Updownarrow(w0,w1,r1)^i$ [214]	$\Updownarrow(w0,w1,r1)^i$	Multiple references Algorithm repetitions
ID	O^{2-} depletion, interface imperfection	HfO_2 or TiO_2 deposition, Annealing	$\Updownarrow(w1)$; $\Updownarrow(w0,r0)$ [65]	1. $\Updownarrow(w1)$; $\Updownarrow(w0,r0)$; $\Updownarrow(w1,r1)$ 2. $\Updownarrow(w1,w0,r0)^j$	
OR	Reduced oxygen storage capability	TiO_2 deposition, Annealing	$\Updownarrow(w1)$; $\Updownarrow(w0,r0)$ [66]	1. $\Updownarrow(w1,r1)$; $\Updownarrow(w0,r0)$ 2. $\Updownarrow(w1,w0,r0)^k$	

- **Conventional defects:** We follow the above process (compare sensitized faults) to distinguish conventional defects from unique defects, further diagnosis for types and locations of conventional defects is out of the scope of this paper.
- **OF:** R_{SET} serves as a distinctive feature of OF. Hence, we design the diagnosis method by reading the 'H' state (with equivalent R_{H} shown in Fig. 9.5) of the faulty cell to identify fault origins in the presence of OF as follows: $\{\Updownarrow(w1,r1)\}$. Here, '$\Updownarrow$' indicates that operations are performed with irrelevant addressing direction; $w1$ denotes write 1 and $r1$ read 1 (using R_{ref1}). In case of the OF defect, $r1$ returns 'H'.

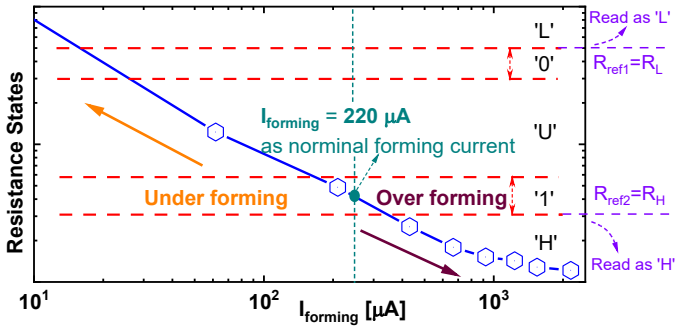


Figure 9.5: Diagnosis for forming defects.

- **UF:** UF may cause the defective cell into an 'L' state (see Fig. 9.5). Therefore, we design the March algorithm: $\{\Updownarrow(w0,r0,w1,r1)\}$. Here, the $r0$ operation is performed by applying R_{ref4}. If the returned values are (L, 0), the cell has a UF defect. If there are (L, 1), the cell may have an OR defect.

- **IUSF:** The intermittent R_{SET} is the distinctive feature of IUSF. The affected R_{SET} exhibits an intermittent 'U' state during the SET process with a low probability (up to 1.068 %) [214]. To distinguish IUSF from other defects, it is sufficient to detect R_{SET} with the following repeated algorithm: $\{\Updownarrow(w0,w1,r1)^i\}$ (i.e., repeating $w0$, $w1$ and $r1$ operations for 'i' times). The $r1$ here applies R_{ref2} and R_{ref3} to detect the 'U' state. The probability of IUSF occurrence P_{IUSF} can result in the diagnostic probability: P_{d}=1−(1 − P_{IUSF})^i. From the measurement, $i = 644$ can achieve a 99 % Fault Coverage (FC). If the values obtained from the algorithm are oscillating (U), the defective cell suffers from an IUSF.

- **ID:** The ID-defective device is stuck at the intermediate state 'U' (14 % cycles), which requires multiple reference resistors for detection and diagnosis. For diagnosis, we design the algorithm:

  $\{\Updownarrow(w1);\Updownarrow(w0,r0);\Updownarrow(w1,r1)\}$.

  This algorithm consists of three elements: 1) apply $w1$ for initialization, 2) apply $1w0$ and $r0$ using R_{ref2}, R_{ref3} to detect 'U', and 3) apply $\Updownarrow(w1,r1)$ to check whether the R_{SET} is in a nominal '1' state (using R_{ref1}). Note that the third element is necessary to distinguish ID from OF (i.e., OF may switch a cell into 'H'). If read values return (U, 1), the ID defect is diagnosed (assume it occurs at this cycle). To target intermittent behavior, another algorithm: $\{\Updownarrow(w1,w0,r0)^j\}$ can be applied ($r0$ applies R_{ref2}, R_{ref3} to detect 'U'). The second algorithm is expensive but guaranteed for ID diagnosis (an FC=99 % requires j=31).

- **OR:** The proposed diagnosis algorithm is: $\{\Updownarrow(w1,r1);\Updownarrow(w0,r0)\}$. The first element is to initialize the cell to the '1' state and read the state. The second element is applied to identify whether the cell switched to 'L' owing to the OR. The $r0$ here applies R_{ref4} to detect the 'L' state. If the read values are (1, L), the OR is

diagnosed (assume OR occurs at this time). Again, we design another algorithm: $\{\Updownarrow (w1, w0, r0)^k\}$ to guarantee the high probability of diagnosis for OR (an FC=99 % requires k=13).

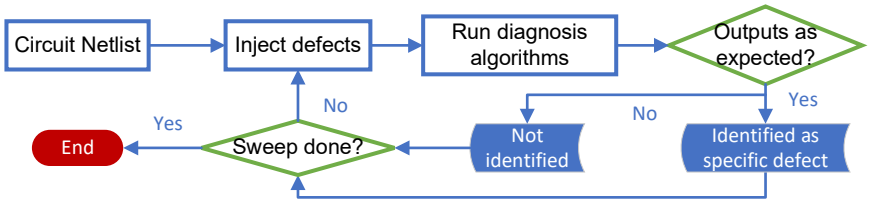


Figure 9.6: Flow chart of diagnosis algorithm verification.

9.4. VALIDATION OF DIAGNOSIS APPROACH

Finally, we validate the diagnosis algorithms in Table 9.2 through simulation, following the process in Fig. 9.6. For each targeted defect injected in the circuit netlist, the designed diagnosis algorithms are applied. After that, the outputs of algorithms are extracted and compared with expected outputs to validate whether the algorithm can identify the defect. Once all algorithms are simulated, the next defect will be injected and the process will be repeated. Note that we focus more on applying algorithms designed for unique defects; the identification of conventional defects can be achieved by comparing different sets of faults that are sensitized by different defects.

Table 9.3: Validation with OF-defective cell as an example.

Defect	Expected output	Actual output	Diagnosed?
OF	H	H	Yes
UF	L, 0	1, 1	No
IUSF	U	1	No
ID	U, 1	1, H	No
OR	1, L	1, 1	No

Table 9.3 shows the validation result for the OF defect as an example. In the simulation setup, we follow the above process. All designed diagnosis algorithms are applied in the presence of OF defect. Table 9.3 shows expected and actual outputs for each applied algorithm. Only if the actual read output is the same as expected, the corresponding defect is diagnosed. For example, both the actual and expected outputs of the algorithm designed for OF are 'H', which means the defect is identified as OF. However, the actual outputs (1, 1) of the algorithm designed for OR are different from its expected outputs (1, L), which indicates the defect is not OR. Similar to the example, other algorithms designed for corresponding unique defects can generate distinctive readout patterns that enable their identification.

9.5. DISCUSSION

While the proposed DA-diagnosis framework demonstrates strong effectiveness and applicability in detecting both conventional and unique RRAM defects, certain limitations remain that highlight potential directions for improvement.

1. **Applicability:** The framework can be applied to currently unknown RRAM defects and other memory technologies as well. Once new RRAM defects are identified, they need to be modeled using the DAT approach. Based on the resulting models and the measurements of the defect, distinctive features of this defect can be identified as well. Next, a diagnosis solution can be developed that uniquely identifies this defect. A similar argumentation holds for other memory technologies: first, defects need to be modeled, then distinctive features can be identified, and finally, effective diagnosis solutions can be developed.

2. **Limitations:** First, the framework heavily relies on the availability of accurate defect models. The development of such models is often time-consuming and requires extensive measurement data as well as an in-depth physical understanding of defect behavior. This dependency may delay the diagnosis of newly discovered defects, especially when characterization data is scarce or incomplete. Second, although the framework is demonstrated to be extendable to other memory technologies, its effectiveness in such cases depends on the maturity of defect modeling methodologies for those technologies, which may not always be as advanced as in RRAM research. Finally, the framework primarily focuses on time-zero defects. Dynamic or time-dependent defects (e.g., aging-induced variations, temperature sensitivity, or workload-dependent degradation) are not fully addressed, which may limit the robustness of the diagnosis under real operational conditions.

3. **Future improvements:** To address these limitations, several research directions can be pursued. First, the defect modeling process can be accelerated and improved by leveraging data-driven approaches, such as machine learning or physics-informed neural networks, to complement traditional physics-based modeling. This could reduce the dependence on exhaustive measurements while still capturing the essential defect behavior. Second, the framework can be extended to incorporate adaptive diagnosis strategies that dynamically adjust to changing defect characteristics over time, thus enhancing reliability under aging and varying environmental conditions. Third, cross-technology studies may help establish generalized defect modeling and diagnosis methodologies that are not strictly tied to one memory type, enabling broader applicability. Finally, integrating in-field monitoring data into the framework could enable continuous refinement of defect models and diagnosis solutions, bridging the gap between laboratory characterization and real-world deployment.

10

Putting All Together: Complete Industrial Test Development for RRAMs

In this chapter, we present a comprehensive and structured framework to address various aspects in RRAM testing; the framework can be used to develop high-quality test solutions for any industrial RRAM. The framework consists of 3 steps: defect modeling, fault modeling, and test development. We begin by providing an overview of defects in the memory array and peripherals, which can be classified into conventional defects and unique defects that require special attention. Then, we conduct systematic fault modeling, taking into account the impact of Data Backgrounds (DBs), sensitizing operations (i.e., static versus dynamic), and variations. Next, we compare existing March algorithms as well as different Design-for-Test (DfT) designs based on their capability to detect RRAM faults. Finally, an optimized test solution is provided. Overall, this chapter presents a structural and complete framework to improve the RRAM testing, enabling a robust and comprehensive development for industrial deployments.

10.1. INTRODUCTION

Test development of RRAMs is rapidly developing. However, each of these studies targets isolated aspects, lacking a systematic and comprehensive framework capable of delivering a complete, high-quality, and efficient test solution that addresses all defect types. Hence, a comprehensive test strategy is needed that enables engineers to develop appropriate test programs for industry RRAMs.

In this chapter, we introduce a comprehensive testing framework that bridges the gap from defect characterization to chip-level realization. It encompasses defect mechanisms, corresponding models, fault models, March algorithms, and DfT methodologies. A test design that is highly optimized for RRAMs is provided to illustrate the practical approach. The main contributions of this chapter are as follows:

- Develop a complete framework of testing designs for industry RRAMs, from the analysis of defects and faults to the design of March tests and DfT approaches.
- Analyze interconnect and contact defects in the RRAM array and derive fault models by considering the impact of *data backgrounds (DBs),* analysis of static & dynamic fault, RRAM & transistor variation, and develop appropriate test solutions.
- Report on all unique RRAM defects in the RRAM array that have been shown to exist, including their associated fault models and test solutions.
- Provide all interconnect and contact defects in peripheral circuits and develop proper test solutions.
- Propose a comprehensive overview and evaluation for the March test and DfT designs in RRAMs to assess their quality (e.g., defect coverage) and associated costs (e.g., area overhead).
- Provide a complete test development framework for industrial RRAMs that is highly optimized.

10.2. DEFECT AND DEFECT MODELING

This section provides a condensed summary of the key insights from Chapter 3 and Chapter 4, which should be considered for the industry. Both conventional and unique defects are found to exist in RRAMs. First, a defect space is given. Then, we overview the defect modeling for RRAMs.

10.2.1. CONVENTIONAL DEFECTS AND THEIR MODELING IN RRAM ARRAY

Conventional defects include interconnects and contacts, which are also studied in other memory technologies. Typically, they are out of the RRAM device and can be appropriately modeled by *linear resistors* to model them [43], [52]. Interconnect and contact defects can cause bridges, shorts, and opens [43], [52]. A *bridge* is modeled as a resistor between a pair of nodes different from the power nodes, a *short* as an undesired resistive path between a node and a power node (V_{DD} or GND), and an *open* is defined as increased resistance in an existing connection.

Conventional defects must be considered depending on intra-cell or inter-cell defects in a memory array [64]. Note that the inter-cell defects involve up to two adjacent cells at a time and three location relationships for each cell (the base cell C_b), i.e., cells in the same row (C_r), cells in the same column (C_c), or diagonal cells (C_d). Considering the symmetry nature, a 2×2 cell array effectively studies the complete defect injection for RRAM array design [64].

10.2.2. UNIQUE DEFECTS AND THEIR MODELING IN RRAM ARRAY

Unique defects are specific and inside the RRAM devices due to imperfect fabrication. Each unique defect in RRAMs exhibits different physical roots and electrical behaviors with conventional defects, e.g., the intermittent RESET failure in ID-defective devices. Hence, they cannot be modeled by linear resistors like conventional defects. Instead, the DAT approach is developed to integrate the impact of the defective devices into the RRAM compact model and appropriately model the defective behavior with non-linearity. So far, there are five unique defects studied in the public domain, as overviewed in Table 10.1: Over/Under Forming [59], [106], Intermittent Undefined State Fault (IUSF) [214], Ion Depletion (ID) [66], [194], and Over RESET (OR) [66].

Table 10.1: Overview of unique defects.

Defect	Mechanism	Defective performance	Defect modeling method
OF	High forming current	Extremely low resistance	Induce forming current
UF	Low forming current	Extremely high resistance	Induce forming current
IUSF	Reduced oxygen storage capability	RRAM intermittently enters 'U' state during SET	Induce complementary switching model
ID	O^{2-} depletion, interface imperfection	RRAM intermittently enters 'U' state during RESET	Change O^{2-} concentration during RESET process
OR	Reduced oxygen storage capability	RRAM intermittently enters 'L' state during RESET	Change O^{2-} concentration, induce leakage current

OF: Over Forming; UF: Under Forming; IUSF: Intermittent Undefined State Fault; ID: Ion Depletion; OR: Over RESET

10.2.3. DEFECTS AND THEIR MODELING IN RRAM PERIPHERAL CIRCUITS

Defects can also exist in peripheral circuits. In this section, we consider interconnect and contact defects between each pair of nodes of the peripherals, i.e., decoders, drivers, and the sense amplifier. Such defects are modeled using linear resistors, which capture the electrical impact of bridges, shorts, and opens. To cover all possible defect locations, we systematically inject these defects into the WL decoders, BL drivers, SL drivers, and the SA. For the WL decoders, BL drivers, and SL drivers, defects are modeled at the gate level between circuit nodes, while for the SA a more detailed transistor-level modeling is applied. The number of nodes in each peripheral determines the complete defect space: 7 nodes in the WL decoder, 9 nodes in the BL driver, 7 nodes in the SL driver, and 13 nodes in the SA. Hence, it results in a complete defect space for each peripheral circuit: 1) $\binom{7}{2} = \frac{7\times6}{2} = 21$ bridges, $7 \times 2 = 14$ shorts, and 7 opens for the WL decoder, 2) $\binom{9}{2} = \frac{9\times8}{2} = 36$ bridges, $9 \times 2 = 18$ shorts, and 9 opens for the BL driver, 3) $\binom{7}{2} = \frac{7\times6}{2} = 21$ bridges, $7 \times 2 = 14$ shorts, and 7 opens for the SL driver, and 4) $\binom{13}{2} = \frac{13\times12}{2} = 78$ bridges, $13 \times 2 = 26$ shorts, and 13 opens for the SA.

10.3. RRAM FAULT MODELING

This section focuses on RRAM faults that should be investigated for the industry, especially those unique faults that only exist in RRAM devices. First, we provide faults that are sensitized by conventional defects in the RRAM array. Then, faults that are specifically sensitized by unique defects in the RRAM array are reported. Finally, we provide an overview of faults that can be sensitized by conventional defects in peripherals.

10.3.1. FAULTS FOR CONVENTIONAL DEFECTS IN RRAM ARRAY

Although quite some work has been reported on fault modeling for conventional defects in RRAMs [48], [50], [59], [64], they failed to consider the effect of RRAM intrinsic stochasticity and process variability, dynamic operations, and combined with different DBs in the presence of a defect. In order to address these shortcomings and illustrate the dependency of fault modeling using accurate simulation platforms incorporating the inherent properties of RRAMs, we will perform three experiments while considering three simulation platforms (i.e., models of transistors and RRAMs). The two simulation platforms are:

1. *Baseline Model*: The 2 × 2 RRAM simulation setup is described as above, without incorporating the RRAM variation and process variation.

2. *Full Variability Model*: The same as Baseline Model, incorporated with the RRAM variation (both cycle to cycle and device to device) in the RRAM compact model; the transistor 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. The impact of Data Background (DB) is considered.

2. *Dynamic analysis without variations*: Dynamic fault analysis (i.e., sensitizing operations consist of up to three consecutive operations) is applied using Baseline Model, while considering different DBs.

3. *Static and dynamic analysis with variations*: Both static and dynamic fault analysis are applied using Full Variability Model, while also considering different DBs.

Each of the above three experiments was performed for interconnect and contact defects by modeling them as linear resistances. We simulated all possible bridges, shorts, and opens while swiping the value of the resistance from 1 Ω to 100 MΩ. In total, we simulated 22 resistive defects within a cell (6 bridges, 8 shorts, and 8 opens) and 17 bridges between cells (including adjacent cells in the same column, row, and diagonal).

Table 10.2 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). The table lists both EtD and HtD faults. It is worth noting that all listed faults are permanent without variations, but

Table 10.2: Collapsed fault set for conventional defects in RRAM array.

Experiment	Static analysis	Dynamic analysis
Collapsed fault set (EtD)	$\langle 0r0/0/1\rangle$, $\langle 1r1/1/0\rangle$, $\langle 1w0/1/-\rangle$, $\langle 1w1;0/1/-\rangle$, $\langle 1_c;0r0/0/1\rangle$	$\langle 1w0r0/0/1\rangle$, $\langle 1w0r0/U/1\rangle$, $\langle 0r0r0/0/1\rangle$, $\langle 1r1w0r0/U/1\rangle$, $\langle 1_c;1w0r0/0/1\rangle$, $\langle 1_c;0w0r0/U/1\rangle$, $\langle 1_c;0r0r0/0/1\rangle$, $\langle 1_r;0r0r0/0/1\rangle$, $\langle 1_c;1_r;1w0r0/U/1\rangle$, $\langle 1_d;0_r;1w0r0/0/1\rangle$, $\langle 0w0w0w0;1/0/-\rangle$, $\langle 1w1r1r1;0/1/-\rangle$, $\langle 1w1w1w1;0/1/-\rangle$
Collapsed fault set (HtD)	$\langle 0w0/L/-\rangle$, $\langle 1w0/U/-\rangle$, $\langle 0w1;0/U/-\rangle$, $\langle 1w0;0/U/-\rangle$	$\langle 1w0r0/U/0\rangle$, $\langle 0w0w0w0/L/-\rangle$, $\langle 1w0w0;0/U/-\rangle$, $\langle 0w1w0;0/U/-\rangle$
Notation	$\langle S_a;S_v/F/R\rangle$ denotes a coupling fault involving two cells; An aggressor cell C_a and a victim cell C_v. C_a can be in the same column, same row or same diagonal as C_v; hence $a \in \{c,r,d\}$. E.g., $\langle 1_c;0r0/0/1\rangle$ means that when the adjacent cell in the same column is '1', applying $r0$ operation to the victim cell (which is in '0') will not lead the cell to flip, but the read operation will return a wrong value 1.	

they can become *intermittent* within certain defect ranges when variations are present, which will be analyzed in the subsequent section.

In the rest of this section, we will illustrate results for one bridge defect cBCC3, which is an interconnect defect between the SL and the RRAM, for three experiments.

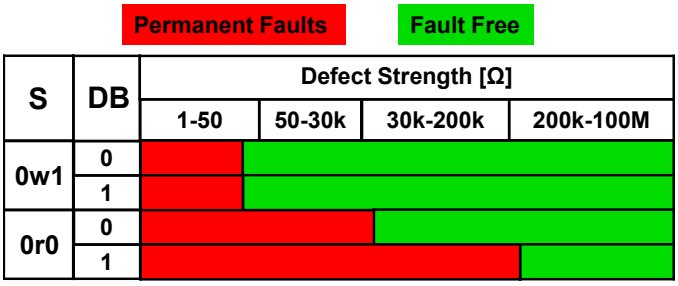


Figure 10.1: Static fault analysis.

BASELINE ANALYSIS

Fig. 10.1 shows the fault map of the bridge defect cBCC3. Four static EtD faults are sensitized (when considering the dependency on DB); the 'red' boxes indicate defect ranges where faults are sensitized, and the 'green' ones indicate that no faults are sensitized. The column 'DB' indicates the value of the data background, being the three neighbors of the defective cell; they are all '0' or all '1'. The figure clearly shows three key points. First, many faults are sensitized for a single defect cBCC3. Second, transition fault $\langle 0w1/0/-\rangle$ is DB-independent. Third, since it is about an inter-cell defect, the DB has an impact on the defect coverage when faults are sensitized by read operations; for instance, when DB=1, the $0r0$ fails for a wide range of resistance values as compared with when DB=0. The detection of this read fault with DB=1 will maximize the defect coverage for the defect under consideration. Hence, it is best to include it in the collapsed fault set.

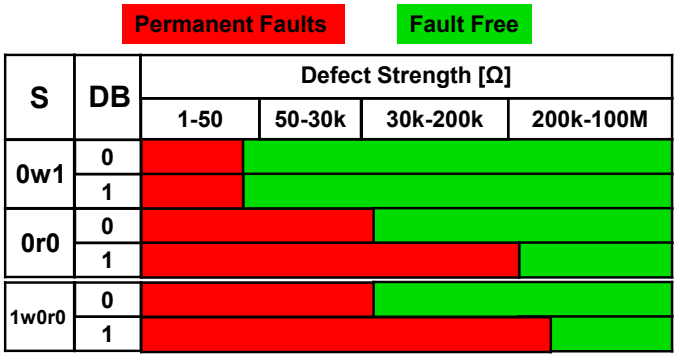


Figure 10.2: Dynamic fault analysis without variations.

DYNAMIC ANALYSIS WITHOUT VARIATIONS

Fig. 10.2 presents the fault map of the defect cBCC3 in this case. The presented dynamic faults, as an example, can cover the maximum defect range with up to three operations. Two observations can be made. First, dynamic operations can sensitize more faults than static operations. The presence of dynamic faults can be attributed to the fact that RRAMs, by their analog device nature, have multiple states other than '0' and '1', and that the movement of oxygen vacancy (affecting resistance) is cumulative. Hence, consecutive read and write operations may change the state of the RRAMs. For example, the 'U' state can further switch to '1' or even 'H' after multiple 0w1 operations. Besides, even a small read voltage across the device can accumulate and eventually switch the device [259]. Second, although dynamic operations may sensitize more faults, the number of dynamic operations is exponentially increasing (the cost of simulation increases). We find that the number of new faults is decreasing as the number of operations increases, hence the need to balance the number of operations chosen against the cost. On the other hand, the simulation can be optimized in ways such as increasing the number of operations each time by targeting only the defect strengths that were not sensitized by previous operations [48], [64]. Other emerging memories, such as STT-MRAM, exist in only two stable resistance states, and never show intermediate states due to the anisotropy magnetic field [272]; hence lack of validation of dynamic faults.

STATIC AND DYNAMIC ANALYSIS WITH VARIATIONS

Both static and dynamic analyses are then performed with the Full Variability Model. Fig. 10.3 presents the fault map of the bridge defect cBCC3 in this case. For every component combination, we perform 1000 Monte Carlo (MC) simulations in which all boundary defect strengths are performed per iteration. First, the same fault class is sensitized as in Fig. 10.2. Second, the range of undetected resistance defects reduces (green boxes). Both write and read faults are sensitized around the boundary resistance between fault-free cases and permanent fault cases, represented by the yellow boxes. The nature of such additional faults is *intermittent* due to the stochastic nature of RRAMs and process variation, i.e., the variation affects write and read currents. For example, the fault denoted as $\langle 1_c; 1w0r0/0/1 \rangle$ indicates that the $1w0r0$ operation fails intermittently (i.e.,

from time to time) when the neighbor in the same column is set to '1'; note that this takes place for relatively low defect strength. Third, the figure shows that in the boundary regions, variability in both RRAM devices and transistors can alter the fault behavior. Specifically, faults that were permanently sensitized at certain defect strengths may become intermittent, while some cases that were originally fault-free can also turn into intermittent faults. It can be concluded that the variation impacts faults sensitized with defect strengths since the fluctuation in RRAM CF impacts the device switching stability, leading to intermittent faults at the boundary. Consequently, variation analysis is essential to accurately capture the complete fault space of RRAMs.

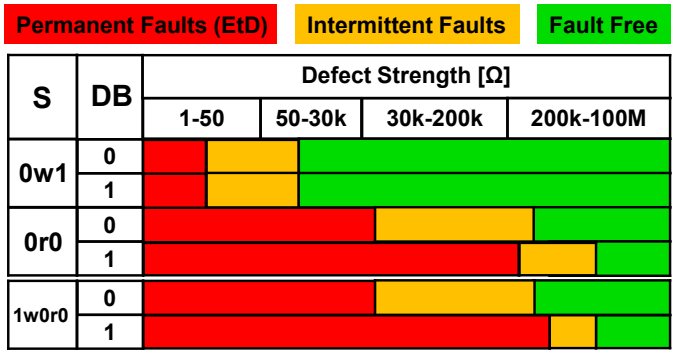


Figure 10.3: Dynamic fault analysis with full variability model.

10.3.2. FAULTS FOR UNIQUE DEFECTS IN RRAM ARRAY

In our recent work, we introduced many unique defects by characterizing industrial RRAM chips. As already mentioned, modeling these unique RRAM defects using linear resistors fails to accurately capture corresponding faulty behaviors, leading to unrealistic fault models [59], [65], [66], [214]. 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 RRAM. Such models are used to replace defect-free RRAM 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.

Table 10.3: Overview fault set for unique defects.

	Collapsed fault set
EtD	⟨0w1/0/−⟩, ⟨1w0/1/−⟩
HtD	⟨1w0/U/−⟩, ⟨1w1/H/−⟩, ⟨1w1/U/−⟩, ⟨1w0/L/−⟩, ⟨0w1/U/−⟩

The results of our fault modeling for all unique defects discussed are summarized in Table 10.3; only the *collapsed fault set* per defect is presented [59], [65], [66], [214]. There

are four key observations from the table. First, unique defects cause both permanent as well as intermittent faults. That is mainly due to the C2C variation and device stochasticity. For example, we characterize the occurrence probabilities of IUSF, ID, and OR as 1.068%, 14%, and 30% in devices switched in multiple cycles [65], [66], [214], which indicates that faults caused by unique defects do not exist permanently. Second, unique defects also cause faults that bring the state of the cell into an *undefined state* or *extremely low state*. Third, the sensitized faults consist of both EtD faults as well as HtD faults (e.g., $\langle 0w1/H/-\rangle$ caused by the OF). Fourth, the unique faults that have been investigated tend to be write faults because unique defects and imperfect processes tend to have a direct impact on RRAM switching, while read faults are typically associated with circuit design (e.g., SA margins).

10.3.3. FAULTS FOR DEFECTS IN PERIPHERALS

Table 10.4: Collapsed fault set for interconnect and contact defects in peripherals.

	Collapsed fault set
EtD	$\langle 0w1/0/-\rangle$, $\langle 1w0/1/-\rangle$, $\langle 1r1/1/0\rangle$, $\langle 0r0/0/1\rangle$, $\langle 1w0;0/1/-\rangle$, $\langle 0_c;0r0/0/1\rangle$, $\langle 1_c;1_r;1r1/1/0\rangle$
HtD	$\langle 1r1/1/?\rangle$, $\langle 0r0/0/?\rangle$, $\langle 0w1/U/1\rangle$
Notation:	$\langle S_a;S_v/F/R\rangle$ denotes a coupling fault involving two cells; An aggressor cell C_a and a victim cell C_v. C_a can be in the same column, same row or same diagonal as C_v; hence $a \in \{c, r, d\}$. E.g., $\langle 1_c;0r0/0/1\rangle$ means that when the adjacent cell in the same column is '1', applying $r0$ operation to the victim cell (which is in '0') will not lead the cell to flip, but the read operation will return a wrong value 1.

Table 10.4 summarizes the collapsed fault set of injected interconnect and contact defects in peripherals. In this work, we only perform static analysis for defects in the intra-peripherals. There are key observations from the results. First, both EtD and HtD faults are validated in peripherals. For example, the write driver cannot provide enough write current due to defects, causing 'U' state faults. Besides, the random read fault (e.g., $\langle 0r0/0/?\rangle$), which can be attributed to the SA read delay, is validated in the presence of conventional defects. Here, the random readout is defined as an intermediate value of the SA output, whereas it should normally resolve to either V_{DD} or GND. Such faults typically occur when the sensed signal is too close to the reference level, falling within the read margin window, which leads to incorrect or delayed resolution in the SA. Second, both single-cell and coupling faults are observed.

10

10.4. TEST DEVELOPMENT

This section gives an overview and analysis of existing test designs in the context of the validated fault models. We first review the March algorithms that have been applied to RRAM testing and analyze their effectiveness and limitations. Then, we summarize the current DfT techniques and their corresponding evaluation methods, highlighting their strengths and gaps with respect to RRAM-specific defect coverage.

10.4.1. MARCH ALGORITHM DESIGN

The fault modeling results show that the collapsed fault set includes both *EtD* and *HtD* faults. While *EtD* faults can be reliably detected through standard write and read operations (e.g., March tests), this is not always the case for *HtD* faults. Their detection by a March test is not guaranteed, and therefore, additional methods are required to enhance or ensure their coverage. In this work, we analyze the March algorithm designs for both conventional and unique defects.

MARCH ALGORITHMS FOR CONVENTIONAL DEFECTS IN RRAM ARRAY

Optimized and high-quality March algorithms require a high coverage of all defects, strengths, and data backgrounds with a minimal test length. For example, if we consider the fault map in Fig. 10.1, the fault with 1r1 rather than 0w1 should be selected. Here, we design the test algorithms for observed EtD permanent faults, including single-cell, multi-cell, static, and dynamic faults. In our previous works, the test development is formulated as an Integer Linear Programming (ILP) problem that minimizes the number of applied sequences and established DBs while maximizing the defect coverage [48], [64]. The March algorithm is explained in Chapter 4.

Compared with existing March algorithms designed for RRAMs, we found that they all target specific fault models (e.g., [49], [226] proposed March-1T1R and March C* for obtained dynamic write faults in their cases). Another March W-1T1R [227] focused on the whole EtD fault models known at the time, including coupling faults, and is the most powerful compared to previous work. However, it lacks detection for multi-cell faults and dynamic faults. To solve this challenge, the DB-based March-EtD [64] integrated DB impact and dynamic operations to achieve high fault coverage.

Table 10.5 overviews existing March algorithms for RRAMs and compares the detection capability for EtD faults. These EtD faults are explained in Section 7.2.3, the green 'Y', orange 'P', and red 'N' indicate guaranteed detection, partial detection, and no detection, respectively. The test time is evaluated as the cost of write and read operations (N represents the number of memory cells). Note that typically, a read operation is faster than a write operation in RRAM design; hence, a read-intensive March algorithm design is preferred. The fault coverage (defined as covered EtD defect strengths) of all those March algorithms is validated in our platform by injecting all possible conventional defects.

Table 10.5: EtD fault detection capability of existing March tests for RRAMs.

										Test time		
Year	**Name**	**SAF**	**TF**	**WDF**	**IRF**	**RDF**	**CFst**	**DWDF**	**NPSF**	**Write**	**Read**	**Coverage**
2013	March-MOM [225]	Y	Y	P	N	N	P	N	N	5N	4N	95.4%
2015	March-1T1R [54]	Y	Y	Y	N	N	Y	Y	P	(1+2a+2b)N	5N	99.634%
2015	March C* [49]	Y	Y	N	Y	Y	Y	Y	P	4N	6N	99.347%
2016	March C*-1T1R [226]	Y	Y	N	Y	Y	Y	Y	P	6N	6N	99.66%
2017	March W-1T1R [227]	Y	Y	Y	Y	Y	Y	N	P	9N	8N	99.608%
2023	March-EtD [64]	Y	Y	Y	Y	Y	Y	Y	Y	15N	10N	100%
Y: yes; N: no; P: partial												

MARCH ALGORITHMS FOR UNIQUE DEFECTS IN RRAM ARRAY

The March algorithm of each unique defect is investigated in [59], [65], [66], [214]. March-forming, designed to detect forming defects, is: $\Updownarrow(w0, w1, w0); \Updownarrow(r0); \Updownarrow(w1, r1)$. March-IUSF, designed to detect IUSF defects, is: $\Updownarrow(w0, w1, r1)^i$. March-ID, designed to detect ID defects, is: $\Updownarrow(w1)^j; \Updownarrow(w0, r0)^j$. March-OR, designed to detect OR defects, is: $\Updownarrow(w1)^k; \Updownarrow(w0, r0)^k$. Here, we note that faults sensitized by unique defects are all single-cell faults, thus requiring simple March algorithms, i.e., only one DB needs to be established. However, due to the intermittent nature of unique defects, the designed March algorithms for IUSF, ID, and OR have to be repeated i, j, and k times. The numbers i, j, and k depend on the occurrence probability of intermittent faults sensitized by defects. A higher fault occurrence probability requires a smaller i, j, and k to achieve the same detection rate. Specifically, with occurrence probabilities of 1.068% and 30%, values of $i = 1291$ and $k = 13$, respectively, are necessary to reach a 99% detection rate for IUSF and OR [66], [214]. Nevertheless, even with multiple repetitions, test escapes cannot be fully avoided, highlighting the need for enhanced DfT approaches, which will be discussed in the following.

MARCH ALGORITHMS FOR DEFECTS IN PERIPHERALS

The March algorithm can be easily obtained based on the collapsed fault set. Defects in peripherals cause both EtD and HtD faults. The detection of EtD faults can be guaranteed with a test algorithm of a test length of $8N$ (where N is the total memory addresses):

$$\text{March-P} = \{\Updownarrow(\text{w0}); \Updownarrow(\text{r0,w1,r1,w0,r0}); \Updownarrow(\text{w1}); \Updownarrow(\text{r1})\}.$$

The first march element of the March-P initiates the memory and guarantees that all three neighbors of the victim cell are set to '1' as the fault model indicates. The second march element aims to detect the fault by performing corresponding operations. The third and fourth march elements aim to detect the read fault when the DB is set to '0'. The detection of HtD faults may be detected by March-P; DfT is required to increase the detection coverage of HtD faults.

10.4.2. DESIGN-FOR-TESTABILITY DESIGN

Multiple DfT designs exist for RRAM testing to detect unique faults that cannot be caught by March algorithms. Here, we classify those existing designs based on their methods and targets. After that, we compare and review those existing DfTs and present a generalized evaluation methodology according to various aspects.

DFT CLASSIFICATION

Since the March algorithms can not guarantee the detection of HtD and intermittent faults, various DfT studies have been proposed, as shown in Table 10.6 and categorized according to their methodological purpose. First, DfTs are analyzed in terms of whether they are write operations or read operations that implement the test. For example, designs in [61], [63] modify and apply multi-reference read operations based on monitoring currents. DfTs that apply write operations can be further divided as voltage-based or current-based (e.g., [67] monitors the potential write current change). Besides, they are also divided by online or offline detections. Note that the online design makes it powerful to detect intermittent faults with assurance (e.g., the design in [67]). The design in

Table 10.6: DfT classification based on the applied operation (read or write) and the mode of application.

Works	DfT type		Applied mode
None	Write-based	voltage-based	online
[57], [62]			offline
[67]		current-based	online
[242]			offline
[61], [63], [243]	Read-based	current-based	online
[225], [226], [253]			offline
[252]	Independent		

[252] is independent of write or read operations. Instead, it applies an algorithm to measure the current sum of the entire crossbar and iterates the measurements of divided equal halves.

Table 10.7: DfT classification depending on targets.

Works	Fault coverage improvement	Speed improvement
Magic NOR 1T1R [226]	No	Yes
Divide and conquer [252]	Yes	No
Sneak-path testing [225]	No	Yes
Weak write [57]	Yes	No
Fast write [242]	No	Yes
On-chip sensor [61]	Yes	No
Enhanced March [251]	Yes	No
DFT-HR-ET-NOR [62]	Yes	Yes
RDF detector [243]	Yes	No
PMRR [63]	Yes	Yes
In-field monitor [67]	Yes	Yes
RDF: Read Disturb Fault; PMRR: Parallel-Multi-Reference Read		

There are two targets of DfT designs in RRAM testing: 1) to enhance the defect or fault coverage and 2) to improve the whole test speed and cost. Table 10.7 concluded those DfT designs according to the target. As shown in the table, only designs in [62], [67], [243] are able to both enhance the test coverage and reduce test time. All of them applied an optimized operation to parallel read or without read. However, works in [62], [243] are only designed and compatible with CIM architectures, requiring huge area consumption (e.g., [243] needs to test 32 RRAM parallel cells at once).

EVALUATION OF DFTS

The evaluation scheme is based on analyzing various aspects of the DfTs. Note that the DfT performance is also dependent on peripheral circuit designs, array architectures, and applied technologies of RRAMs and transistors; this work tries to provide insights for DfT selection in RRAMs. Table 10.8 overviews existing DfT designs in RRAMs based on the above evaluation. The evaluation is based on the following aspects:

Table 10.8: Overview and evaluation of existing DfT designs for RRAMs.

Works	EtD	HtD					Time		Compatibility	
		UWF URF	DF	Intermittent	Weak	Area overhead	Write	Read	March algorithm	Circuit integration (modified)
Divide and conquer [252]	Y	P	P	N	P	-	Increased exponentially		No	AD, SA
Sneak-path testing [225]	P	P	Y	N	N	28+26R	7N	5N/3	No	AD, SA
Weak write [57]	N	Y	Y	N	N	24+18R	-	-	Yes	WD
Fast write [242]	P	Y	Y	N	P	50+18R	(4T+1 +X)N	6N	Yes	WD
Parallel March [253]	P	N	P	N	N	-	4(N+1)	5N+R	Yes	-
On-chip sensor [61]	P	Y	N	N	N	20RC	-	-	Yes	SA
Enhanced March [251]	Y	Y	N	N	N	26	8N	6N	Yes	SA
DfT-HR-ET-NOR [62]	Y	Y	Y	N	N	$2\log_2 R$ $+96+4C$	6N	$8\log_2 N$ $+8$	No	SA
PMRR [63]	Y	Y	Y	Y	N	13C	5N	4N	Yes	Read circuit
RDF detector [243]	Y	N	N	Y	N	6R+16C	-	-	No	AD, SA
Write monitor [67]	Y	Y	Y	Y	Y	10R+7C	5N	2N	Yes	WD

N: total number of memristors; X: ratio between w0 and w1 time; R: number of rows; C: number of columns
Address Decoder (AD); Sense Amplifier (SA); Write Driver (WD)

- **Coverage:** The evaluation on coverage is considered in terms of fault coverage. The defect can be detected when the corresponding fault occurs. For example, one unique fault such as $\langle 0w1/H/-\rangle$ is guaranteed to be detected by the design in [67] directly since this scheme is able to monitor the write current and detect if the cell is in the 'H' state. Therefore, the only unique defect that can sensitize this fault, an over forming defect, will be able to be detected by [67]. Another example is the detection capability for intermittent faults, which requires in-field tests such as designs in [63], [67]. Most of the works are designed to detect strong faults but only the scheme in [67] is able to monitor write currents and detect the current deviation even without logical errors. Note that repeated March tests designed for unique defects cost a longer test time and cannot achieve 100% fault coverage, compared with in-field DfT schemes.

- **Test time:** The test time of DfTs are evaluated according to their implementation time. Most of them are compatible with a March algorithm but perform special write/read operations. For example, a scheme in [63] applies parallel read operations with four references at once to run the March algorithm, with the test length of $5N_w$+$5N_r$, where N_w/N_r denotes the number of writes and reads, respectively. Moreover, certain DfTs adjust the duration of individual read or write operations. For example, the fast-write scheme [242] uses only half of the nominal write period to sensitize the 'U' state. By varying the operation time, such schemes can detect a wider range of unique faults within a shorter test time.

- **Area overhead:** The additional cost is evaluated by counting the additional transistors of the DfTS. The works in [252], [253] utilised new memory architecture but normal peripheral circuits; therefore, their area cost is not considered in our evaluation. Note that we already include the inputs/outputs cost for some DfTs, as well as the switchings of test modes.

- **Compatibility of impact on regular operations:** The regular write or read operations can be affected by the integrated DfTs. Note that the operations are affected depending on the modified peripheral circuits. For example, 'Divide and

conquer' [252] and 'Sneak-path testing' [225] both modified address decoders and references in the SA, thus affecting regular write/read operations. Designs in [57], [67], [242], [253] modified write drivers and only affected write operations. Designs in [61]–[63], [251] modified the read circuits (especially the reference) and affected read operations. Among them, designs in [63], [67] introduced voltage drops by using current mirrors with diode-like transistors to copy the currents. Hence, it requires high voltage to drive the normal write/read operations.

- **Compatibility of circuit integration:** One indication of compatibility is the ease with which DfT can be integrated into normal, fault-free circuits. Works in [225], [252] modified address decoders and read circuits to achieve specific functions (e.g., measure the sneak-path current), which is not compatible with regular circuit integration. Instead, the design in [67] modified the write driver and can still perform regular write operations. Hence, this design is easy to integrate into RRAM arrays.

- **Compatibility of March algorithms:** If the DfT design is compatible with *regular write or read operations*, then it is adaptive to March algorithms. For example, the DfTs in [61], [251] can be applied for March algorithms directly with modified read circuits that can detect more faulty states rather than '0' and '1'. Similarly, approaches in [57], [242] are available to apply either fast write or weak write pulses for detection and thus are easy to apply for March algorithms. On the other hand, works in [62], [225] claimed a March algorithm, but it is impossible be applied directly for other memory designs.

10.5. OPTIMIZING TEST SOLUTIONS

Fig. 10.4 shows the RRAM test development approach, based on a complete combination; it is called Device-Aware-Test and consists of three steps: 1) *defect modeling* where defects are modeled in an appropriate way, 2) *fault modeling* where the defective device is replaced with a representative compact model during circuit simulation to analyze the memory behavior, and 3) *test development* targeting validated faults in step 2.

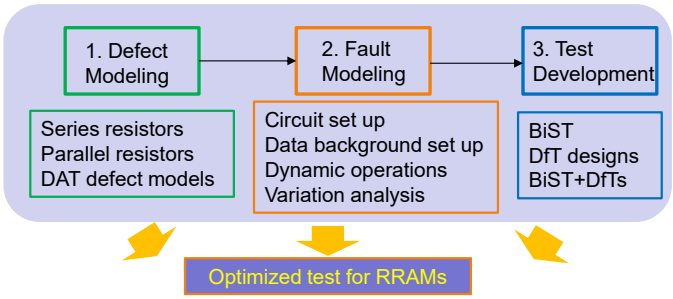


Figure 10.4: Systematic and complete approach for the RRAM testing.

In order to develop a comprehensive RRAM test design, each of these steps should be considered and implemented in as much detail as possible (those considerations are

shown in previous sections). Moreover, there is also a need to optimize the final solution not only by combining aspects in each step together but also by analysing the trade-off among various cases and presenting an optimized solution in an efficient manner. Here, we provide a case study: an RRAM array (at least be 2×2) with peripheral circuits needs to be tested. Considering all aspects, both EtD and HtD faults exist in the presence of both conventional and unique defects. The simplest test solution is to use Built-in-Self Test (BiST), e.g., the March algorithm in [225]. This algorithm takes the shortest time and can already achieve more than 95% coverage (see Table 10.5). However, if we recognise the existence of coupling-cell even multi-cell faults in RRAMs, a more expensive BiST is required, such as the test in [227]. Then, it can be concluded that from our fault modeling, DfT is essential because of the wide presence of faulty states in RRAM; hence, cheap DfT can be combined with simple BiST to achieve the fault coverage of most EtD and HtD faults. Besides, we can still integrate more expensive DfT into March algorithms for higher coverage, depending on RRAM quality requirements.

The optimized test solution that is highly optimized is to apply the March algorithm in [64] with integrated DfT in [67] to replace all write operations in the algorithm. This solution provides complete coverage of all known RRAM faults as well as defects. The DfT is efficient and compatible, especially for those intermittent weak faults (they are neglected by all other existing works), although the March algorithm is long due to the establishment of DBs and dynamic operations.

11

Conclusion

11.1. SUMMARY

Chapter 1: Introduction

This chapter is an introduction to this thesis. First, it provides an overview of emerging memory technologies, memory hierarchy, and different types of emerging memories. Then, we discuss the state-of-the-art in memory testing, covering both conventional memory testing methodologies and RRAM-specific testing techniques. Next, we outline key research topics, including defect space modeling, fault analysis, high-quality test design, and RRAM defect diagnosis. Additionally, it highlights the contributions of the thesis, which include testing strategies for both conventional and unique defects, as well as design for reliability, testability, and RRAM-specific diagnosis. Finally, the chapter concludes with an overview of the thesis organization.

Chapter 2: Background

This chapter provides an overview of the fundamental concepts of RRAM. We explore three primary switching modes of RRAM devices: unipolar switching, bipolar switching, and complementary switching. We conduct a broad comparison of several compact RRAM models, followed by an in-depth analysis of three specific models. These models are applicable to the RRAM architecture, which is subsequently introduced. We detail the structure and functionality of its key components. Lastly, we review potential applications of RRAMs and existing RRAM chips, including both prototypes and commercially available implementations.

Chapter 3: RRAM Manufacturing, Defects, and Non-Idealities

This chapter explores the production process of RRAM, potential defects, and associated reliability challenges. We start by the three key stages of the manufacturing process: Front-End-Of-Line (FEOL), Back-End-Of-Line (BEOL), and CF forming. Each step is examined in detail, highlighting the types of defects that may arise during fabrication. Finally, we discussed the major non-ideality concerns of RRAM and their correlation with the manufacturing process.

Chapter 4: Testing for Conventional Defects

This chapter has identified the presence of unique 3-cell and 4-cell coupling faults in RRAMs, alongside conventional single-cell and two-cell coupling faults, arising due to interconnect and contact defects. These faults are DB-dependent and require careful consideration; otherwise, they may go undetected, leading to test escapes. We provide a systematic method to develop defect modeling, fault modeling, and test generation.

Chapter 5: DAT for Ion Depletion (ID) Defects

This chapter emphasizes the effectiveness of the DAT approach in accurately identifying and modeling unique RRAM-related defects. Additionally, the complexity and additional manufacturing stages in RRAMs, particularly as devices scale down, may introduce unique defects that have yet to be thoroughly explored. This underscores the need for advanced fault modeling and testing solutions, such as the DAT approach, along with a deeper understanding of emerging defect mechanisms.

Chapter 6: DAT for Over RESET (OR) Defects

This chapter applies the DAT approach again in accurately identifying and modeling the intermittent (30% cycles) OR phenomenon; the defect causes extremely high state faults. Moreover, the measurement data implies the impact of tunneling leakage current, which requires a device model to incorporate it. Hence, we propose DA-defect modeling, fault modeling, and the test solution for the unique defect.

Chapter 7: Design-for-Testability Scheme for RRAM Faults

This chapter proposes a novel write current monitor DfT that guarantees full coverage of validated faults in the RRAM array. We design a specific write driver that performs write operations and measures the write currents simultaneously, allowing fast and efficient detection of faults not only during production testing but also in the field. Our proposed DfT implementations are validated with process variations and detect both conventional defects as well as unique defects (based on silicon data). Results demonstrate the superiority of our design compared to the state-of-the-art. Furthermore, the circuit can be reconfigured to optimize the yield process.

Chapter 8: Design for Reliability for Read Disturb Failures

This chapter proposes a bi-directional read system that detects and resolves the read disturb failure in the RRAM array. We investigate the resistance drift referred to as the RDF with device-level analysis and measurements. We propose and implement bi-directional read approaches at a circuit level. The approaches that alleviate the RDF are validated to compare with the regular read operations. Furthermore, the inference accuracy drop due to the RDF in the RRAM-based CIM is investigated and enhanced by the proposed approach, which shows the effectiveness of our designs.

Chapter 9: Diagnosis for Conventional and Unique Defects

This chapter introduces the DA-diagnosis framework to detect conventional and unique defects in RRAMs. First, it identifies distinctive features for every defect based on physical analysis and characterization. Next, efficient mechanisms are developed that detect these distinctive features and thus efficiently diagnose the defects.

Chapter 10: Put All Together: Complete Industrial Test Development for RRAMs

In this chapter, a comprehensive and systematic testing strategy for industrial RRAMs is presented to target both conventional and unique defects. The whole methodology is proposed by developing a complete defect space, systematic and realistic fault models, an overview of designed March tests, evaluating various DfT solutions, and optimizing an ultimate test solution for industrial RRAMs. The practical integration of certain DfTs with a complete March algorithm is illustrated in a case study, demonstrating how well they work to enhance testability while preserving compatibility with RRAM arrays. Those insights point out that various aspects in terms of cost, coverage, and robustness need to be considered and optimized in order to ensure RRAM dependability for high-quality industrial production.

11.2. FUTURE RESEARCH DIRECTIONS

In this thesis, we highlighted the systematic and complete methodology in developing high-quality test and diagnosis strategies for RRAMs. Furthermore, it is essential to develop effective migration and recovery solutions to ensure the quality and reliability of RRAMs. Future research should prioritize the following key areas:

Characterization and modeling for more RRAM defects

As discussed in Chapter 3, numerous unique RRAM defects remain unmodeled despite their significant impact on device performance. Notable examples include electrode roughness and material redeposition, both of which strongly influence the interface engineering and, consequently, the overall performance of fabricated RRAM devices.

Experimental analysis and testing for peripheral circuits

A memory chip can be functionally divided into three main blocks: address decoders, memory cell arrays, and read/write circuits. In RRAM testing, both theoretical and experimental research, including this dissertation, has primarily focused on faults within RRAM arrays. However, faults arising in address decoders and read/write circuits remain unexplored in published literature. Therefore, investigating faulty behaviors caused by defects in these two circuit blocks, developing appropriate fault models, and incorporating them into test programs are essential steps toward comprehensive RRAM testing.

Broad-ranging applications of DAT

This thesis has focused on the DAT approach and extended it to DA diagnosis. DAT, which is grounded in the underlying physical mechanisms of defects and accurately incorporates the intrinsic nonlinearity of RRAMs, has proven effective for high-quality test development. Looking ahead, several promising research directions can be envisioned: 1) Built-In Self-Test (BIST): integration of DAT into on-chip self-test schemes to enable efficient, low-overhead in-field testing, 2) Built-In-Self-Repair (BISR): coupling DAT with repair mechanisms to enhance yield and extend device usability, 3) Diagnosis: extending the diagnostic capability of DAT for more precise fault localization and classification, and 4) Reliability and lifetime management: employing DAT as a continuous monitoring tool to support adaptive reliability strategies and lifetime prediction.

BIBLIOGRAPHY

[1] A. Farmahini-Farahani, J. H. Ahn, K. Morrow, and N. S. Kim, "NDA: Near-DRAM Acceleration Architecture Leveraging Commodity DRAM Devices and Standard Memory Modules", in *IEEE Int. Symp. High-Performance Computer Architecture*, 2015, pp. 283–295. DOI: 10.1109/HPCA.2015.7056040.

[2] R. Hameed, W. Qadeer, M. Wachs, *et al.*, "Understanding sources of inefficiency in general-purpose chips", in *IEEE/ACM Int. Symp. Comput. Archit.*, 2010, pp. 37–47. DOI: 10.1145/1815961.1815968.

[3] M. Horowitz, "1.1 Computing's Energy Problem (and What We Can Do about It)", in *IEEE Int. Solid-State Circuits Conf.*, 2014, pp. 10–14. DOI: 10.1109/ISSCC.2014.6757323.

[4] H.-S. P. Wong and S. Salahuddin, "Memory Leads the Way to Better Computing", *Nat. Nanotechnol.*, vol. 10, no. 3, pp. 191–194, 2015. DOI: 10.1038/nnano.2015.29.

[5] H. Tsai, S. Ambrogio, P. Narayanan, R. M. Shelby, and G. W. Burr, "Recent Progress in Analog Memory-based Accelerators for Deep Learning", *J. Phys. D: Appl. Phys.*, vol. 51, no. 28, p. 283 001, 2018. DOI: 10.1088/1361-6463/aac8a5.

[6] M. A. Zidan, J. P. Strachan, and W. D. Lu, "The Future of Electronics Based on Memristive Systems", *Nat. Electron.*, vol. 1, no. 1, pp. 22–29, 2018. DOI: 10.1038/s41928-017-0006-8.

[7] J. L. Hennessy and D. A. Patterson, *Computer Architecture: A Quantitative Approach*, 5th. Morgan Kaufmann, 2011. [Online]. Available: https://www.elsevier.com/books/computer-architecture/hennessy/978-0-12-383872-8.

[8] R. Bez and A. Pirovano, "Non-Volatile Memory Technologies: Emerging Concepts and New Materials", *Mater. Sci. Semicond. Process.*, vol. 7, no. 4-6, pp. 349–355, 2004. DOI: 10.1016/j.mssp.2004.09.127.

[9] S.-T. Han, Y. Zhou, and V. Roy, "Towards the Development of Flexible Non-Volatile Memories", *Adv. Mater.*, vol. 25, no. 38, pp. 5425–5449, 2013. DOI: 10.1002/adma.201301361.

[10] 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 Res. Lett.*, vol. 15, no. 1, 2020. DOI: 10.1186/s11671-020-03299-9.

[11] 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: 10.3390/nano10081437.

[12] Y. Huai *et al.*, "Spin-Transfer Torque MRAM (STT-MRAM): Challenges and Prospects", *AAPPS Bull.*, vol. 18, no. 6, pp. 33–40, 2008.

[13] S. Tehrani, B. Engel, J. Slaughter, *et al.*, "Recent Developments in Magnetic Tunnel Junction MRAM", *IEEE Trans. Magn.*, vol. 36, no. 5, pp. 2752–2757, 2000. DOI: 10.1109/20.908581.

[14] S. Dünkel, M. Trentzsch, R. Richter, *et al.*, "A FeFET-Based Super-Low-Power Ultra-Fast Embedded NVM Technology for 22nm FDSOI and Beyond", in *IEEE Int. Electron Devices Meet.*, 2017, pp. 19–7. DOI: 10.1109/IEDM.2017.8268425.

[15] J. Y. Kim, M.-J. Choi, and H. W. Jang, "Ferroelectric Field-Effect Transistors: Progress and Perspective", *APL Mater.*, vol. 9, no. 2, 2021. DOI: 10.1063/5.0035515.

[16] T. Mikolajick, C. Dehm, W. Hartner, *et al.*, "FeRAM Technology for High-Density Applications", *Microelectron. Reliab.*, vol. 41, no. 7, pp. 947–950, 2001. DOI: 10.1016/S0026-2714(01)00049-X.

[17] D. Takashima, "Overview of FeRAMs: Trends and Perspectives", in *IEEE Annu. Non-Volatile Memory Technol. Symp.*, 2011, pp. 1–6. DOI: 10.1109/NVMTS.2011.6137107.

[18] X. Dong, N. Muralimanohar, N. Jouppi, R. Kaufmann, and Y. Xie, "Leveraging 3D PCRAM Technologies to Reduce Checkpoint Overhead for Future Exascale Systems", in *ACM Conf. High Perform. Comput. Netw. Storage Anal.*, 2009, pp. 1–12. DOI: 10.1145/1654059.1654117.

[19] X. Dong, N. P. Jouppi, and Y. Xie, "PCRAMsim: System-Level Performance, Energy, and Area Modeling for Phase-Change RAM", in *ACM/IEEE Int. Conf. Comput.-Aided Des.*, 2009, pp. 269–275. DOI: 10.1145/1687399.1687449.

[20] S. Bhatti, R. Sbiaa, A. Hirohata, H. Ohno, S. Fukami, and S. Piramanayagam, "Spintronics-Based Random Access Memory: A Review", *Mater. Today*, vol. 20, no. 9, pp. 530–548, 2017. DOI: 10.1016/j.mattod.2017.07.007.

[21] D. Apalkov, A. Khvalkovskiy, S. Watts, *et al.*, "Spin-Transfer Torque Magnetic Random Access Memory (STT-MRAM)", *ACM J. Emerg. Technol. Comput. Syst.*, vol. 9, no. 2, pp. 1–35, 2013. DOI: 10.1145/2463585.2463589.

[22] S. Tehrani, J. Slaughter, E. Chen, M. Durlam, J. Shi, and M. DeHerren, "Progress and Outlook for MRAM Technology", *IEEE Trans. Magn.*, vol. 35, no. 5, pp. 2814–2819, 2002. DOI: 10.1109/20.800991.

[23] C. Wang, S. Yuan, H. Xun, *et al.*, "Defects, Fault Modeling, and Test Development Framework for FeFETs", in *IEEE Int. Test Conf.*, 2024, pp. 91–95. DOI: 10.1109/ITC51657.2024.00026.

[24] 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", *Appl. Phys. Lett.*, vol. 112, no. 22, 2018. DOI: 10.1063/1.5029324.

[25] C. Besleaga, R. Radu, L.-M. Balescu, *et al.*, "Ferroelectric Field-Effect Transistors Based on PZT and IGZO", *IEEE J. Electron Devices Soc.*, vol. 7, pp. 268–275, 2019. DOI: 10.1109/JEDS.2019.2895367.

[26] S.-C. Chang, N. Haratipour, S. Shivaraman, *et al.*, "FeRAM Using Anti-Ferroelectric Capacitors for High-Speed and High-Density Embedded Memory", in *IEEE Int. Electron Devices Meet.*, 2021, pp. 33–2. DOI: 10.1109/IEDM19574.2021.9720510.

[27] 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 *IEEE Int. Electron Devices Meet.*, 2021, pp. 8–1. DOI: 10.1109/IEDM19574.2021.9720535.

[28] T. Francois, J. Coignus, A. Makosiej, *et al.*, "16kbit HfO_2: Si-Based 1T-1C FeRAM Arrays Demonstrating High-Performance Operation and Solder Reflow Compatibility", in *IEEE Int. Electron Devices Meet.*, 2021, pp. 33–1. DOI: 10.1109/IEDM19574.2021.9720640.

[29] 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: 10.1016/j.ijleo.2018.11.100.

[30] C. Bonafos, M. Carrada, G. Benassayag, *et al.*, "Si and Ge Nanocrystals for Future Memory Devices", *Mater. Sci. Semicond. Process.*, vol. 15, no. 6, pp. 615–626, 2012. DOI: 10.1016/j.mssp.2012.09.004.

[31] P. Guo, A. M. Sarangan, and I. Agha, "A Review of Germanium-Antimony-Telluride Phase Change Materials for Non-Volatile Memories and Optical Modulators", *Appl. Sci.*, vol. 9, no. 3, p. 530, 2019. DOI: 10.3390/app9030530.

[32] I. Valov, R. Waser, J. R. Jameson, and M. N. Kozicki, "Erratum: Electrochemical Metallization Memories – Fundamentals, Applications, Prospects", *Nanotechnology*, vol. 22, no. 28, 2011. DOI: 10.1088/0957-4484/22/28/289502.

[33] J. Radhakrishnan, A. Belmonte, T. Witters, *et al.*, "On the Key Impact of Composition of Ge-Te and Ge-Se Electrolytes on CBRAM Properties", in *IEEE Int. Mem. Workshop*, 2018, pp. 1–4. DOI: 10.1109/IMW.2018.8388842.

[34] C. Nail, G. Molas, P. Blaise, *et al.*, "Understanding RRAM Endurance, Retention and Window Margin Trade-Off Using Experimental Results and Simulations", in *IEEE Int. Electron Devices Meet.*, 2016, pp. 4–5. DOI: 10.1109/IEDM.2016.7838346.

[35] A. Belmonte, G. Reale, A. Fantini, *et al.*, "Effect of the Switching Layer on CBRAM Reliability and Benchmarking Against OxRAM Devices", *Solid-State Electron.*, vol. 184, p. 108 058, 2021. DOI: 10.1016/j.sse.2021.108058.

[36] T. Mikolajick, M. H. Park, L. Begon-Lours, and S. Slesazeck, "From Ferroelectric Material Optimization to Neuromorphic Devices", *Adv. Mater.*, vol. 35, no. 37, p. 2 206 042, 2023. DOI: 10.1002/adma.202206042.

[37] S. Hamdioui, H. A. Du Nguyen, M. Taouil, *et al.*, "Applications of Computation-in-Memory Architectures Based on Memristive Devices", in *ACM/IEEE Design, Autom. Test Eur. Conf.*, 2019, pp. 486–491. DOI: 10.23919/DATE.2019.8715020.

[38] M. A. Breuer and A. D. Friedman, *Diagnosis & Reliable Design of Digital Systems* (Digital System Design Series). Computer Science Press, 1976. [Online]. Available: https://link.springer.com/book/9783642954269.

[39] A. J. Van de Goor, *Testing Semiconductor Memories: Theory and Practice*. John Wiley & Sons, Inc., 1991.

[40] S. Hamdioui, R. Wadsworth, J. D. Reyes, and A. J. Van De Goor, "Memory Fault Modeling Trends: A Case Study", *J. Electron. Test.*, vol. 20, no. 3, pp. 245–255, 2004. DOI: 10.1023/B:JETT.0000029458.57095.bb.

[41] I. Schanstra and A. Van De Goor, "Industrial Evaluation of Stress Combinations for March Tests Applied to SRAMs", in *Int. Test Conf.*, 1999, pp. 983–992. DOI: 10.1109/TEST.1999.805831.

[42] A. van de Goer and J. de Neef, "Industrial Evaluation of DRAM Tests", in *ACM/IEEE Design, Autom. Test Eur. Conf. Exhib.*, 1999, pp. 623–630. DOI: 10.1109/DATE.1999.761194.

[43] S. Hamdioui and A. Van De Goor, "An Experimental Analysis of Spot Defects in SRAMs: Realistic Fault Models and Tests", in *IEEE Asian Test Symp.*, 2000, pp. 131–138. DOI: 10.1109/ATS.2000.893615.

[44] E. I. Vatajelu, A. Bosio, L. Dilillo, *et al.*, "Analyzing Resistive-Open Defects in SRAM Core-Cell under the Effect of Process Variability", in *IEEE Eur. Test Symp.*, 2013, pp. 1–6. DOI: 10.1109/ETS.2013.6569373.

[45] F. Hapke, W. Redemund, J. Schloeffel, *et al.*, "Defect-Oriented Cell-Internal Testing", in *IEEE Int. Test Conf.*, 2010, pp. 1–10. DOI: 10.1109/TEST.2010.5699229.

[46] Z. Gao, S. Malagi, M.-C. Hu, *et al.*, "Application of Cell-Aware Test on an Advanced 3nm CMOS Technology Library", in *IEEE Int. Test Conf.*, 2019, pp. 1–6. DOI: 10.1109/ITC44170.2019.9000164.

[47] S. Borkar, "Microarchitecture and Design Challenges for Gigascale Integration", in *IEEE Int. Symp. Microarchitecture*, 2004, pp. 3–3. DOI: 10.1109/MICRO.2004.24.

[48] M. Fieback, G. C. Medeiros, L. Wu, *et al.*, "Defects, Fault Modeling, and Test Development Framework for RRAMs", *ACM J. Emerg. Technol. Comput. Syst.*, vol. 18, no. 3, pp. 1–26, 2022. DOI: 10.1145/3510851.

[49] 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", *IEEE Trans. Comput.*, vol. 64, no. 1, pp. 180–190, 2015. DOI: 10.1109/TC.2014.12.

[50] M. Fieback, M. Taouil, and S. Hamdioui, "Testing Resistive Memories: Where Are We and What Is Missing?", in *IEEE Int. Test Conf.*, 2018, pp. 1–9. DOI: 10.1109/TEST.2018.8624895.

[51] M. L. Bushnell and V. D. Agrawal, *Essentials of Electronic Testing for Digital Memory & Mixed-Signal VLSI Circuits*. Springer, 2000, ISBN: 978-0-7923-7991-1. DOI: 10.1007/978-0-306-47040-4. [Online]. Available: https://link.springer.com/book/10.1007/978-0-306-47040-4.

[52] O. Ginez, J.-M. Portal, and C. Muller, "Design and Test Challenges in Resistive Switching RAM: An Electrical Model for Defect Injections", in *IEEE Eur. Test Symp.*, 2009, pp. 61–66. DOI: 10.1109/ETS.2009.23.

[53] S. Kannan, J. Rajendran, R. Karri, and O. Sinanoglu, "Sneak-Path Testing of Memristor-Based Memories", in *IEEE Int. Conf. VLSI Des.*, 2013, pp. 386–391. DOI: 10.1109/VLSID.2013.219.

[54] Y.-X. Chen and J.-F. Li, "Fault Modeling and Testing of 1T1R Memristor Memories", in *IEEE VLSI Test Symp.*, 2015, pp. 1–6. DOI: 10.1109/VTS.2015.7116247.

[55] T.-Y. Lin, Y.-X. Chen, J.-F. Li, C.-Y. Lo, D.-M. Kwai, and Y.-F. Chou, "A Test Method for Finding Boundary Currents of 1T1R Memristor Memories", in *IEEE Asian Test Symp.*, 2016, pp. 281–286. DOI: 10.1109/ATS.2016.44.

[56] P. Liu, Z. You, J. Wu, and M. Elimu, "Defect Analysis and Parallel Testing for 3D Hybrid CMOS-Memristor Memory", *IEEE Trans. Emerg. Top. Comput.*, vol. 9, no. 2, pp. 745–758, 2021. DOI: 10.1109/TETC.2020.2982830.

[57] N. Z. Haron and S. Hamdioui, "DfT Schemes for Resistive Open Defects in RRAMs", in *ACM/IEEE Design, Autom. Test Eur. Conf.*, 2012, pp. 799–804. DOI: 10.1109/DATE.2012.6176603.

[58] L. Wu, M. Taouil, S. Rao, E. J. Marinissen, and S. Hamdioui, "Electrical Modeling of STT-MRAM Defects", in *IEEE Int. Test Conf.*, 2018, pp. 1–10. DOI: 10.1109/TEST.2018.8624749.

[59] M. Fieback, L. Wu, G. C. Medeiros, *et al.*, "Device-Aware Test: A New Test Approach Towards DPPB Level", in *IEEE Int. Test Conf.*, 2019, pp. 1–10. DOI: 10.1109/ITC44170.2019.9000134.

[60] M. Taouil and S. Hamdioui, *Device Aware Test for Memory Units*, Google Patents, US Patent App. 17/639,386, Nov. 2022.

[61] T. S. Copetti, T. Gemmeke, and L. M. Bolzani Poehls, "Validating a DFT Strategy's Detection Capability Regarding Emerging Faults in RRAMs", in *IFIP/IEEE Int. Conf. Very Large Scale Integr.*, 2021, pp. 1–6. DOI: 10.1109/VLSI-SoC53125.2021.9606993.

[62] A. Singh, M. Fieback, R. Bishnoi, *et al.*, "Accelerating RRAM Testing with a Low-Cost Computation-in-Memory-Based DFT", in *IEEE Int. Test Conf.*, 2022, pp. 400–409. DOI: 10.1109/ITC50671.2022.00085.

[63] M. Fieback, F. Bradarić, M. Taouil, and S. Hamdioui, "Online Fault Detection and Diagnosis in RRAM", in *IEEE Eur. Test Symp.*, 2023, pp. 1–6. DOI: 10.1109/ETS56758.2023.10174113.

[64] 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.

[65] H. Xun, S. Yuan, M. Fieback, H. Aziza, M. Taouil, and S. Hamdioui, "Device-Aware Test for Ion Depletion Defects in RRAMs", in *IEEE Int. Test Conf.*, 2023, pp. 246–255. DOI: 10.1109/ITC51656.2023.00041.

[66] H. Xun, M. Fieback, S. Yuan, *et al.*, "Characterization and Test of Intermittent Over RESET in RRAMs", in *IEEE Asian Test Symp.*, 2023, pp. 1–6. DOI: 10.1109/ATS59501.2023.10317990.

[67] H. Xun, M. Fieback, M. A. Yaldagard, *et al.*, "Robust Design-for-Testability Scheme for Conventional and Unique Defects in RRAMs", in *IEEE Int. Test Conf.*, 2024, pp. 374–383. DOI: 10.1109/ITC51657.2024.00059.

[68] D. Niu, Y. Chen, C. Xu, and Y. Xie, "Impact of Process Variations on Emerging Memristor", in *ACM/IEEE Design Autom. Conf.*, 2010, pp. 877–882. DOI: 10.1145/1837274.1837495.

[69] A. Grossi, D. Walczyk, C. Zambelli, *et al.*, "Impact of Intercell and Intracell Variability on Forming and Switching Parameters in RRAM Arrays", *IEEE Trans. Electron Devices*, vol. 62, no. 8, pp. 2502–2509, 2015. DOI: 10.1109/TED.2015.2442412.

[70] E. I. Vatajelu, P. Prinetto, M. Taouil, and S. Hamdioui, "Challenges and Solutions in Emerging Memory Testing", *IEEE Trans. Emerg. Top. Comput.*, vol. 7, no. 3, pp. 493–506, 2019. DOI: 10.1109/TETC.2017.2691263.

[71] A. Gebregiorgis, A. Singh, S. Diware, R. Bishnoi, and S. Hamdioui, "Dealing with Non-Idealities in Memristor Based Computation-in-Memory Designs", in *IFIP/IEEE Int. Conf. Very Large Scale Integr.*, 2022, pp. 1–6. DOI: 10.1109/VLSI-SoC54400.2022.9939618.

[72] H. Xun, M. Fieback, S. Yuan, *et al.*, "In-Field Monitoring and Preventing Read Disturb Faults in RRAMs", in *IEEE Eur. Test Symp.*, 2025, pp. 1–6. DOI: 10.1109/ETS63895.2025.11049639.

[73] H. Xun, M. Fieback, S. Yuan, H. Aziza, M. Taouil, and S. Hamdioui, "Device-Aware Diagnosis for Yield Learning in RRAMs", in *ACM/IEEE Design, Autom. Test Eur. Conf.*, 2024. DOI: 10.23919/DATE58400.2024.10546660.

[74] 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 Trans. Electron Devices*, vol. 59, no. 9, pp. 2461–2467, 2012. DOI: 10.1109/TED.2012.2202320.

[75] S. Yu and P.-Y. Chen, "Emerging Memory Technologies: Recent Trends and Prospects", *IEEE Solid-State Circuits Mag.*, vol. 8, no. 2, pp. 43–56, 2016. DOI: 10.1109/MSSC.2016.2546199.

[76] 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.

[77] S. Hamdioui, M. Taouil, and N. Z. Haron, "Testing Open Defects in Memristor-Based Memories", *IEEE Trans. Comput.*, vol. 64, no. 1, pp. 247–259, 2015. DOI: 10.1109/TC.2013.206.

[78] H. Zhang, S. Yoo, S. Menzel, *et al.*, "Understanding the Coexistence of Two Bipolar Resistive Switching Modes with Opposite Polarity in Pt/TiO_2/Ti/Pt Nanosized ReRAM Devices", *ACS Appl. Mater. Interfaces*, vol. 10, no. 35, pp. 29 766–29 778, 2018. DOI: 10.1021/acsami.8b09068.

[79] Y. Liu, Q. Hu, Q. Wu, *et al.*, "Probabilistic Circuit Implementation Based on P-Bits Using the Intrinsic Random Property of RRAM and P-Bit Multiplexing Strategy", *Micromachines*, vol. 13, no. 6, p. 924, 2022. DOI: 10.3390/mi13060924.

[80] Z. Chen, R. Song, Q. Huo, *et al.*, "Analysis of Leakage Current of HfO_2/TaO_x-Based 3-D Vertical Resistive Random Access Memory Array", *Micromachines*, vol. 12, no. 6, p. 614, 2021. DOI: 10.3390/mi12060614.

[81] W. Kim, S. Menzel, D. J. Wouters, R. Waser, and V. Rana, "3-Bit Multilevel Switching by Deep Reset Phenomenon in $Pt/W/TaO_X/Pt$-ReRAM Devices", *IEEE Electron Device Lett.*, vol. 37, no. 5, pp. 564–567, 2016. DOI: 10.1109/LED.2016.2542879.

[82] S. Mei, M. Bosman, K. Shubhakar, N. Raghavan, L. Ming, and K. Pey, "3D Characterization of Hard Breakdown in RRAM Device", *Microelectron. Eng.*, vol. 216, p. 111 042, 2019. DOI: 10.1016/j.mee.2019.111042.

[83] L. Grenouillet, N. Castellani, A. Persico, *et al.*, "16kbit 1T1R OxRAM Arrays Embedded in 28nm FDSOI Technology Demonstrating Low BER, High Endurance, and Compatibility with Core Logic Transistors", in *IEEE Int. Mem. Workshop*, 2021, pp. 1–4. DOI: 10.1109/IMW51353.2021.9439607.

[84] K.-S. Li, C. Ho, M.-T. Lee, *et al.*, "Utilizing Sub-5 nm Sidewall Electrode Technology for Atomic-Scale Resistive Memory Fabrication", in *IEEE Symp. VLSI Technol.*, 2014, pp. 1–2. DOI: 10.1109/VLSIT.2014.6894402.

[85] H.-S. P. Wong, H.-Y. Lee, S. Yu, *et al.*, "Metal–Oxide RRAM", *Proc. IEEE*, vol. 100, no. 6, pp. 1951–1970, 2012. DOI: 10.1109/JPROC.2012.2190369.

[86] F. Nardi, S. Balatti, S. Larentis, D. C. Gilmer, and D. Ielmini, "Complementary Switching in Oxide-Based Bipolar Resistive-Switching Random Memory", *IEEE Trans. Electron Devices*, vol. 60, no. 1, pp. 70–77, 2013. DOI: 10.1109/TED.2012.2226728.

[87] Y. Chen, "ReRAM: History, Status, and Future", *IEEE Trans. Electron Devices*, vol. 67, no. 4, pp. 1420–1433, 2020. DOI: 10.1109/TED.2019.2961505.

[88] T. Fujii, M. Kawasaki, A. Sawa, H. Akoh, Y. Kawazoe, and Y. Tokura, "Hysteretic Current-Voltage Characteristics and Resistance Switching at an Epitaxial Oxide Schottky Junction $SrRuO_3/SrTi_{0.99}Nb_{0.01}O_3$", *Appl. Phys. Lett.*, vol. 86, no. 1, 2005. DOI: 10.1063/1.1845598.

[89] M.-J. Lee, C. B. Lee, D. Lee, *et al.*, "A Fast, High-Endurance and Scalable Non-Volatile Memory Device Made from Asymmetric Ta_2O_{5-x}/TaO_{2-x} Bilayer Structures", *Nat. Mater.*, vol. 10, no. 8, pp. 625–630, 2011. DOI: 10.1038/nmat3070.

[90] C.-Y. Lin, S.-Y. Wang, D.-Y. Lee, and T.-Y. Tseng, "Electrical Properties and Fatigue Behaviors of ZrO_2 Resistive Switching Thin Films", *J. Electrochem. Soc.*, vol. 155, no. 8, H615, 2008. DOI: 10.1149/1.2946430.

[91] B. Cho, S. Song, Y. Ji, T.-W. Kim, and T. Lee, "Organic Resistive Memory Devices: Performance Enhancement, Integration, and Advanced Architectures", *Adv. Funct. Mater.*, vol. 21, no. 15, pp. 2806–2829, 2011. DOI: 10.1002/adfm.201100686.

[92] H.-D. Kim, H.-M. An, E. B. Lee, and T. G. Kim, "Stable Bipolar Resistive Switching Characteristics and Resistive Switching Mechanisms Observed in Aluminum Nitride-Based ReRAM Devices", *IEEE Trans. Electron Devices*, vol. 58, no. 10, pp. 3566–3573, 2011. DOI: 10.1109/TED.2011.2162518.

[93] Q. Liu, W. Guan, S. Long, R. Jia, M. Liu, and J. Chen, "Resistive Switching Memory Effect of ZrO_2 Films with Zr^+ Implanted", *Appl. Phys. Lett.*, vol. 92, no. 1, 2008. DOI: 10.1063/1.2832660.

[94] K. Kinoshita, T. Tamura, M. Aoki, Y. Sugiyama, and H. Tanaka, "Bias Polarity Dependent Data Retention of Resistive Random Access Memory Consisting of Binary Transition Metal Oxide", *Appl. Phys. Lett.*, vol. 89, no. 10, 2006. DOI: 10.1063/1.2339032.

[95] J. Park, A. Kumar, Y. Zhou, *et al.*, "Multi-level, Forming and Filament Free, Bulk Switching Trilayer RRAM for Neuromorphic Computing at the Edge", *Nat. Commun.*, vol. 15, no. 1, p. 3492, 2024. DOI: 10.1038/s41467-024-46682-1.

[96] W. Guan, S. Long, R. Jia, and M. Liu, "Nonvolatile Resistive Switching Memory Utilizing Gold Nanocrystals Embedded in Zirconium Oxide", *Appl. Phys. Lett.*, vol. 91, no. 6, 2007. DOI: 10.1063/1.2760156.

[97] K. Szot, W. Speier, G. Bihlmayer, and R. Waser, "Switching the Electrical Resistance of Individual Dislocations in Single-Crystalline $SrTiO_3$", *Nat. Mater.*, vol. 5, no. 4, pp. 312–320, 2006. DOI: 10.1038/nmat1614.

[98] B. Gao, S. Yu, N. Xu, *et al.*, "Oxide-Based RRAM Switching Mechanism: A New Ion-Transport-Recombination Model", in *IEEE Int. Electron Devices Meet.*, 2008, pp. 1–4. DOI: 10.1109/IEDM.2008.4796751.

[99] S. Yu, X. Guan, and H.-S. P. Wong, "On the Stochastic Nature of Resistive Switching in Metal Oxide RRAM: Physical Modeling, Monte Carlo Simulation, and Experimental Characterization", in *IEEE Int. Electron Devices Meet.*, 2011, pp. 17–3. DOI: 10.1109/IEDM.2011.6131572.

[100] W.-Y. Chang, Y.-C. Lai, T.-B. Wu, S.-F. Wang, F. Chen, and M.-J. Tsai, "Unipolar Resistive Switching Characteristics of ZnO Thin Films for Nonvolatile Memory Applications", *Appl. Phys. Lett.*, vol. 92, no. 2, 2008. DOI: 10.1063/1.2834852.

[101] X. Chen, W. Hu, Y. Li, S. Wu, and D. Bao, "Complementary Resistive Switching Behaviors Evolved from Bipolar TiN/HfO_2/Pt Device", *Appl. Phys. Lett.*, vol. 108, no. 5, 2016. DOI: 10.1063/1.4941287.

[102] G. Bersuker, D. Gilmer, D. Veksler, *et al.*, "Metal Oxide RRAM Switching Mechanism Based on Conductive Filament Microscopic Properties", in *IEEE Int. Electron Devices Meet.*, 2010, pp. 19–6. DOI: 10.1109/IEDM.2010.5703394.

[103] Z. Fang, X. P. Wang, J. Sohn, *et al.*, "The Role of Ti Capping Layer in HfO_x-Based RRAM Devices", *IEEE Electron Device Lett.*, vol. 35, no. 9, pp. 912–914, 2014. DOI: 10.1109/LED.2014.2334311.

[104] Y. H. Do, J. S. Kwak, Y. C. Bae, *et al.*, "TiN Electrode-Induced Bipolar Resistive Switching of TiO_2 Thin Films", *Curr. Appl. Phys.*, vol. 10, no. 1, e71–e74, 2010. DOI: 10.1016/j.cap.2009.12.017.

[105] A. Hardtdegen, C. La Torre, F. Cüppers, S. Menzel, R. Waser, and S. Hoffmann-Eifert, "Improved Switching Stability and the Effect of an Internal Series Resistor in HfO_2/TiO_x Bilayer ReRAM Cells", *IEEE Trans. Electron Devices*, vol. 65, no. 8, pp. 3229–3236, 2018. DOI: 10.1109/TED.2018.2849872.

[106] A. Grossi, E. Nowak, C. Zambelli, *et al.*, "Fundamental Variability Limits of Filament-Based RRAM", in *IEEE Int. Electron Devices Meet.*, 2016, pp. 4.7.1–4.7.4. DOI: 10.1109/IEDM.2016.7838348.

[107] B. Govoreanu, G. Kar, Y.-Y. Chen, *et al.*, "10×10 nm^2 Hf/HfO_x Crossbar Resistive RAM with Excellent Performance, Reliability and Low-Energy Operation", in *IEEE Int. Electron Devices Meet.*, 2011, pp. 31.6.1–31.6.4. DOI: 10.1109/IEDM.2011.6131652.

[108] M. Lanza, H.-S. P. Wong, E. Pop, *et al.*, "Recommended Methods to Study Resistive Switching Devices", *Adv. Electron. Mater.*, vol. 5, no. 1, p. 1 800 143, 2019. DOI: 10.1002/aelm.201800143.

[109] B. Hajri, H. Aziza, M. M. Mansour, and A. Chehab, "RRAM Device Models: A Comparative Analysis with Experimental Validation", *IEEE Access*, vol. 7, pp. 168 963–168 980, 2019. DOI: 10.1109/ACCESS.2019.2954753.

[110] F. Cüppers, S. Menzel, C. Bengel, *et al.*, "Exploiting the Switching Dynamics of HfO_2-Based ReRAM Devices for Reliable Analog Memristive Behavior", *APL Mater.*, vol. 7, no. 9, p. 091 105, 2019. DOI: 10.1063/1.5108654.

[111] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The Missing Memristor Found", *Nature*, vol. 453, no. 7191, pp. 80–83, 2008. DOI: 10.1038/nature06932.

[112] D. B. Strukov and R. S. Williams, "Exponential Ionic Drift: Fast Switching and Low Volatility of Thin-Film Memristors", *Appl. Phys. A*, vol. 94, no. 3, pp. 515–519, 2009. DOI: 10.1007/s00339-008-4975-3.

[113] M. D. Pickett, D. B. Strukov, J. L. Borghetti, *et al.*, "Switching Dynamics in Titanium Dioxide Memristive Devices", *J. Appl. Phys.*, vol. 106, no. 7, 2009. DOI: 10.1063/1.3236506.

[114] S. Kvatinsky, E. G. Friedman, A. Kolodny, and U. C. Weiser, "TEAM: Threshold Adaptive Memristor Model", *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 60, no. 1, pp. 211–221, 2012. DOI: 10.1109/TCSI.2012.2215714.

[115] S. Kvatinsky, M. Ramadan, E. G. Friedman, and A. Kolodny, "VTEAM: A General Model for Voltage-Controlled Memristors", *IEEE Trans. Circuits Syst. II, Express Briefs*, vol. 62, no. 8, pp. 786–790, 2015. DOI: 10.1109/TCSII.2015.2433536.

[116] Y. V. Pershin and M. Di Ventra, *SPICE Model of Memristive Devices with Threshold*, Available: https://arxiv.org/abs/1204.2600, 2012. DOI: 10.48550/arXiv.1204.2600.

[117] M. Bocquet, H. Aziza, W. Zhao, *et al.*, "Compact Modeling Solutions for Oxide-Based Resistive Switching Memories (OxRAM)", *J. Low Power Electron. Appl.*, vol. 4, no. 1, pp. 1–14, 2014. DOI: 10.3390/jlpea4010001.

[118] Z. Jiang, S. Member, Y. Wu, *et al.*, "A Compact Model for Metal–Oxide Resistive Experiment Verification", *IEEE Trans. Electron Devices*, vol. 63, no. 5, pp. 1884–1892, 2016. DOI: 10.1109/TED.2016.2545412.

[119] C. Bengel, A. Siemon, F. Cüppers, *et al.*, "Variability-Aware Modeling of Filamentary Oxide-Based Bipolar Resistive Switching Cells Using SPICE Level Compact Models", *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 12, pp. 4618–4630, 2020. DOI: 10.1109/TCSI.2020.3018502.

[120] S. Menzel, M. Waters, A. Marchewka, U. Böttger, R. Dittmann, and R. Waser, "Origin of the Ultra-Nonlinear Switching Kinetics in Oxide-based Resistive Switches", *Adv. Funct. Mater.*, vol. 21, no. 23, pp. 4487–4492, 2011. DOI: 10.1002/adfm.201101117.

[121] D. B. Strukov, J. L. Borghetti, and R. S. Williams, "Coupled Ionic and Electronic Transport Model of Thin-Film Semiconductor Memristive Behavior", *Small*, vol. 5, no. 9, pp. 1058–1063, 2009. DOI: 10.1002/smll.200801323.

[122] W. Zhao, M. Moreau, E. Deng, *et al.*, "Synchronous Non-Volatile Logic Gate Design Based on Resistive Switching Memories", *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 2, pp. 443–454, 2014. DOI: 10.1109/TCSI.2013.2278332.

[123] H.-S. P. Wong, C. Ahn, J. Cao, *et al.*, *Stanford Memory Trends*, https://nano.stanford.edu/stanford-memory-trends, Accessed: 26 Aug 2025.

[124] S. Dietrich, M. Angerbauer, M. Ivanov, *et al.*, "A Nonvolatile 2-Mbit CBRAM Memory Core Featuring Advanced Read and Program Control", *IEEE J. Solid-State Circuits*, vol. 42, no. 4, pp. 839–845, 2007. DOI: 10.1109/JSSC.2007.892207.

[125] Y. H. Tseng, C.-E. Huang, C.-H. Kuo, Y.-D. Chih, and C. J. Lin, "High Density and Ultra Small Cell Size of Contact ReRAM (CR-RAM) in 90nm CMOS Logic Technology and Circuits", in *IEEE Int. Electron Devices Meet.*, 2009, pp. 1–4. DOI: 10.1109/IEDM.2009.5424408.

[126] S.-S. Sheu, P.-C. Chiang, W.-P. Lin, *et al.*, "A 5ns Fast Write Multi-Level Non-Volatile 1 K Bits RRAM Memory with Advance Write Scheme", in *Symp. VLSI Circuits*, Available on IEEE Xplore. ISBN: 978-1-4244-3307-0, 2009, pp. 82–83.

[127] C. J. Chevallier, C. H. Siau, S. F. Lim, *et al.*, "A 0.13 μm 64mb multi-layered conductive metal-oxide memory", in *IEEE Int. Solid-State Circuits Conf.*, 2010, pp. 260–261. DOI: 10.1109/ISSCC.2010.5433945.

[128] S.-S. Sheu, M.-F. Chang, K.-F. Lin, *et al.*, "A 4Mb Embedded SLC Resistive-RAM Macro with 7.2 ns Read-Write Random-Access Time and 160 ns MLC-Access Capability", in *IEEE Int. Solid-State Circuits Conf.*, 2011, pp. 200–202. DOI: 10.1109/ISSCC.2011.5746281.

[129] W. Otsuka, K. Miyata, M. Kitagawa, *et al.*, "A 4Mb Conductive-Bridge Resistive Memory with 2.3 GB/s Read-Throughput and 216 MB/s Program-Throughput", in *IEEE Int. Solid-State Circuits Conf.*, 2011, pp. 210–211. DOI: 10.1109/ISSCC.2011.5746286.

[130] J. Yi, H. Choi, S. Lee, *et al.*, "Highly Reliable and Fast Nonvolatile Hybrid Switching ReRAM Memory Using Thin Al_2O_3 Demonstrated at 54nm Memory Array", in *IEEE Symp. VLSI Technol.*, Available on IEEE Xplore. ISBN: 978-1-4244-9949-6, 2011, pp. 48–49.

[131] A. Kawahara, R. Azuma, Y. Ikeda, *et al.*, "An 8 Mb Multi-Layered Cross-Point ReRAM Macro with 443 MB/s Write Throughput", *IEEE J. Solid-State Circuits*, vol. 48, no. 1, pp. 178–185, 2012. DOI: 10.1109/JSSC.2012.2215121.

[132] H. D. Lee, S. Kim, K. Cho, *et al.*, "Integration of $4F^2$ Selector-Less Crossbar Array 2Mb ReRAM Based on Transition Metal Oxides for High Density Memory Applications", in *IEEE Symp. VLSI Technol.*, 2012, pp. 151–152. DOI: 10.1109/VLSIT.2012.6242506.

[133] M.-F. Chang, C.-W. Wu, C.-C. Kuo, *et al.*, "A 0.5 V 4Mb Logic-Process Compatible Embedded Resistive RAM (ReRAM) in 65nm CMOS Using Low-Voltage Current-Mode Sensing Scheme with 45ns Random Read Time", in *IEEE Int. Solid-State Circuits Conf.*, 2012, pp. 434–436. DOI: 10.1109/ISSCC.2012.6177079.

[134] W. C. Shen, C. Y. Mei, Y.-D. Chih, *et al.*, "High-K Metal Gate Contact RRAM (CRRAM) in Pure 28nm CMOS Logic Process", in *IEEE Int. Electron Devices Meet.*, 2012, pp. 31–6. DOI: 10.1109/IEDM.2012.6479146.

[135] A. Kawahara, K. Kawai, Y. Ikeda, *et al.*, "Filament Scaling Forming Technique and Level-Verify-Write Scheme with Endurance over 10^7 Cycles in ReRAM", in *IEEE Int. Solid-State Circuits Conf.*, 2013, pp. 220–221. DOI: 10.1109/ISSCC.2013.6487708.

[136] J. Jameson, P. Blanchard, C. Cheng, *et al.*, "Conductive-Bridge Memory (CBRAM) with Excellent High-Temperature Retention", in *IEEE Int. Electron Devices Meet.*, 2013, pp. 30–1. DOI: 10.1109/IEDM.2013.6724721.

[137] T. Liu *et al.*, "A 130.7 mm^2 2-Layer 32 Gb ReRAM Memory Device in 24 nm Technology", in *IEEE Int. Solid-State Circuits Conf.*, 2013. DOI: 10.1109/ISSCC.2013.6487703.

[138] M.-F. Chang, J.-J. Wu, T.-F. Chien, *et al.*, "19.4 Embedded 1Mb ReRAM in 28nm CMOS with 0.27-to-1V Read Using Swing-Sample-and-Couple Sense Amplifier and Self-Boost-Write-Termination Scheme", in *IEEE Int. Solid-State Circuits Conf.*, 2014, pp. 332–333. DOI: 10.1109/ISSCC.2014.6757457.

[139] S. Sills, S. Yasuda, J. Strand, *et al.*, "A Copper ReRAM Cell for Storage Class Memory Applications", in *IEEE Symp. VLSI Technol.*, 2014, pp. 1–2. DOI: 10.1109/VLSIT.2014.6894368.

[140] H. W. Pan, K. P. Huang, S. Y. Chen, *et al.*, "1Kbit FinFET Dielectric (FIND) RRAM in Pure 16nm FinFET CMOS Logic Process", in *IEEE Int. Electron Devices Meet.*, 2015, pp. 10–5. DOI: 10.1109/IEDM.2015.7409670.

[141] M. Ueki, K. Takeuchi, T. Yamamoto, *et al.*, "Low-Power Embedded ReRAM Technology for IoT Applications", in *IEEE Symp. VLSI Technol.*, 2015, T108–T109. DOI: 10.1109/VLSIT.2015.7223640.

[142] S. Yasuda, K. Ohba, T. Mizuguchi, *et al.*, "A Cross-Point Cu-ReRAM with a Novel OTS Selector for Storage Class Memory Applications", in *IEEE Symp. VLSI Technol.*, 2017, T30–T31. DOI: 10.23919/VLSIT.2017.7998189.

[143] Z. Chen, H. Wu, B. Gao, *et al.*, "Performance Improvements by SL-Current Limiter and Novel Programming Methods on 16MB RRAM Chip", in *IEEE Int. Mem. Workshop*, 2017, pp. 1–4. DOI: 10.1109/IMW.2017.7939097.

[144] Q. Luo, X. Xu, T. Gong, *et al.*, "8-Layers 3D Vertical RRAM with Excellent Scalability Towards Storage Class Memory Applications", in *IEEE Int. Electron Devices Meet.*, 2017, pp. 2–7. DOI: 10.1109/IEDM.2017.8268315.

[145] C.-C. Chou, Z.-J. Lin, P.-L. Tseng, *et al.*, "An N40 256K×44 Embedded RRAM Macro with SL-Precharge SA and Low-Voltage Current Limiter to Improve Read and Write Performance", in *IEEE Int. Solid-State Circuits Conf.*, 2018, pp. 478–480. DOI: 10.1109/ISSCC.2018.8310392.

[146] P. Jain, U. Arslan, M. Sekhar, *et al.*, "13.2 A 3.6 Mb 10.1 Mb/mm^2 Embedded Non-Volatile ReRAM Macro in 22nm FinFET Technology with Adaptive Forming/Set/Reset Schemes Yielding Down to 0.5 V with Sensing Time of 5 ns at 0.7 V", in *IEEE Int. Solid-State Circuits Conf.*, 2019, pp. 212–214. DOI: 10.1109/ISSCC.2019.8662393.

[147] J. Yang, X. Xue, X. Xu, *et al.*, "A 28nm 1.5 Mb Embedded 1T2R RRAM with 14.8 Mb/mm^2 Using Sneaking Current Suppression and Compensation Techniques", in *IEEE Symp. VLSI Circuits*, 2020, pp. 1–2. DOI: 10.1109/VLSICircuits18222.2020.9163035.

[148] Q. Wang, Y. Yang, Z. Wang, *et al.*, "A Logic-Process Compatible RRAM with 15.43 Mb/mm^2 Density and 10 Years@150°C Retention Using STI-Less Dynamic-Gate and Self-Passivation Sidewall", in *IEEE Int. Electron Devices Meet.*, 2023, pp. 1–4. DOI: 10.1109/IEDM45741.2023.10413885.

[149] Y.-C. Huang, S.-H. Liu, H.-S. Chen, *et al.*, "15.7 A 32Mb RRAM in a 12nm FinFET Technology with a 0.0249μm^2 Bit-Cell, a 3.2GB/s Read Throughput, a 10K-Cycle Write Endurance and a 10-Year Retention at 105 °C", in *IEEE Int. Solid-State Circuits Conf.*, vol. 67, 2024, pp. 288–290. DOI: 10.1109/ISSCC49657.2024.10454367.

[150] K. Santo, M. Nakayama, R. Mochida, *et al.*, "A Prototype 16Mbit RRAM on 55nm BCD with 56% Compact-Area Wordline Driver and Constant Write-Current Scheme for Automotive 150°C Operation", in *IEEE Symp. VLSI Technol.*, 2025, pp. 1–3. DOI: 10.23919/VLSITechnologyandCir65189.2025.11075037.

[151] P.-H. Shih, W.-H. Lin, Y.-D. Chih, *et al.*, "High Density 7nm FinFET Dielectric RRAM in Embedded Memory Applications", in *IEEE Symp. VLSI Technol.*, 2025, pp. 1–3. DOI: 10.23919/VLSITechnologyandCir65189.2025.11075153.

[152] M. Hellenbrand, I. Teck, and J. L. MacManus-Driscoll, "Progress of Emerging Non-Volatile Memory Technologies in Industry", *MRS Commun.*, pp. 1–14, 2024. DOI: 10.1557/s43579-024-00660-2.

[153] TSMC, *TSMC Investor Report 2022 (5 Operational Highlights)*, Accessed: 11 Mar. 2025, 2022. [Online]. Available: %5Curl%7Bhttps://investor.tsmc.com/static/annualReports/2022/english/pdf/2022_tsmc_ar_e_ch5.pdf%7D.

[154] Weebit Nano, *Weebit Nano to Demo Its ReRAM Technology on GlobalFoundries 22FDX Platform*, Accessed: 11 Mar. 2025, 2024. [Online]. Available: %5Curl%7Bhttps://www.weebit-nano.com/news/press-releases/weebit-nano-to-demo-its-reram-rram-embedded-technology-on-globalfoundries-22fdx-platform/%7D.

[155] G. Molas, G. Piccolboni, A. Bricalli, *et al.*, "High Temperature Stability Embedded ReRAM for 2x nm Node and Beyond", in *IEEE Int. Mem. Workshop*, 2022, pp. 1–4. DOI: 10.1109/IMW52921.2022.9779293.

[156] Infineon, *Infineon and TSMC to Introduce RRAM Technology for Automotive AURIX™ TC4x Product Family*, Accessed: 11 Mar. 2025, 2022. [Online]. Available: %5Curl%7Bhttps://www.infineon.com/cms/en/about-infineon/press/market-news/2022/INFATV202211-031.html%7D.

[157] A. Grossi, M. Coppetta, S. Aresu, A. Kux, T. Kern, and R. Strenz, "28nm Data Memory with Embedded RRAM Technology in Automotive Microcontrollers", in *IEEE Int. Mem. Workshop*, 2023, pp. 1–4. DOI: 10.1109/IMW56887.2023.10145951.

[158] RAMXEED Limited, *RAMXEED Represents Our Company That Continues to Grow and Co-Create in the Pursuit of Infinite Possibilities Based on Random Access Memory (RAM) Technology*, Accessed: 11 Mar 2025, 2024. [Online]. Available: %5Curl%7Bhttps://www.ramxeed.com/assets/images/downloads/RAMXEED_memory_catalog_en_202409.pdf%7D.

[159] W. Banerjee, "Challenges and Applications of Emerging Nonvolatile Memory Devices", *Electronics*, vol. 9, no. 6, p. 1029, 2020. DOI: 10.3390/electronics9061029.

[160] J. J. Yang, D. B. Strukov, and D. R. Stewart, "Memristive Devices for Computing", *Nature Nanotechnology*, vol. 8, no. 1, pp. 13–24, 2013. DOI: 10.1038/nnano.2012.240.

[161] S. Hamdioui, L. Xie, H. A. Du Nguyen, *et al.*, "Memristor Based Computation-in-Memory Architecture for Data-Intensive Applications", in *ACM/IEEE Des. Autom. Test Eur. Conf. Exhib.*, 2015, pp. 1718–1725.

[162] S. Hamdioui, S. Kvatinsky, G. Cauwenberghs, *et al.*, "Memristor for Computing: Myth or Reality?", *ACM/IEEE Des. Autom. Test Eur. Conf. Exhib.*, pp. 722–731, 2017. DOI: 10.23919/DATE.2017.7927083.

[163] A. Singh, S. Diware, A. Gebregiorgis, *et al.*, "Low-Power Memristor-Based Computing for Edge-AI Applications", in *IEEE Int. Symp. Circuits Syst.*, 2021, pp. 1–5. DOI: 10.1109/ISCAS51556.2021.9401226.

[164] A. Gebregiorgis, A. Singh, A. Yousefzadeh, *et al.*, "Tutorial on Memristor-Based Computing for Smart Edge Applications", *Memories-Mater. Devices, Circuits and Syst.*, vol. 4, p. 100 025, 2023. DOI: https://doi.org/10.1016/j.memori.2023.100025.

[165] A. B. Gebregiorgis, H. A. Du Nguyen, J. Yu, *et al.*, "A Survey on Memory-centric Computer Architectures", *ACM J. Emerg. Technol. Comput. Syst.*, vol. 18, no. 4, pp. 1–50, 2022. DOI: 10.1145/3544974.

[166] R. Bishnoi, L. Wu, M. Fieback, *et al.*, "Special Session – Emerging Memristor Based Memory and CIM Architecture: Test, Repair and Yield Analysis", in *IEEE VLSI Test Symp.*, 2020, pp. 1–10. DOI: 10.1109/VTS48691.2020.9107595.

[167] C.-X. Xue, W.-H. Chen, J.-S. Liu, *et al.*, "Embedded 1-Mb ReRAM-Based Computing-in- Memory Macro With Multibit Input and Weight for CNN-Based AI Edge Processors", *IEEE J. Solid-State Circuits*, vol. 55, no. 1, 2020. DOI: 10.1109/JSSC.2019.2951363.

[168] T. Shahroodi, M. Zahedi, A. Singh, S. Wong, and S. Hamdioui, "KrakenOnMem: A Memristor-Augmented HW/SW Framework for Taxonomic Profiling", in *ACM Int. Conf. Supercomputing*, 2022, pp. 1–14. DOI: 10.1145/3524059.3532367.

[169] S. S. Diware, S. Dash, A. B. Gebregiorgis, *et al.*, "Severity-Based Hierarchical ECG Classification Using Neural Networks", *IEEE Trans. on Biomed. Circuits and Syst.*, vol. 17, no. 1, pp. 77–91, 2023. DOI: 10.1109/TBCAS.2023.3242683.

[170] S. Diware, K. Chilakala, R. V. Joshi, S. Hamdioui, and R. Bishnoi, "Reliable and Energy-Efficient Diabetic Retinopathy Screening Using Memristor-Based Neural Networks", *IEEE Access*, vol. 12, pp. 47 469–47 482, 2024. DOI: 10.1109/ACCESS.2024.3383014.

[171] P. Dang, H. Li, and W. Wang, "A Fully Hardware Implemented Accelerator Design in ReRAM Analog Computing without ADCs", in *IEEE Int. Symp. Circuits Syst.*, 2025, pp. 1–5. DOI: 10.1109/ISCAS56072.2025.11043697.

[172] S. Liu, L. Liu, J. Tang, B. Yu, Y. Wang, and W. Shi, "Edge Computing for Autonomous Driving: Opportunities and Challenges", *Proc. IEEE*, vol. 107, no. 8, 2019. DOI: 10.1109/JPROC.2019.2915983.

[173] S. A. El-Sayed, T. Spyrou, L. A. Camuñas-Mesa, and H.-G. Stratigopoulos, "Compact functional testing for neuromorphic computing circuits", *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 42, no. 7, pp. 2391–2403, 2023. DOI: 10.1109/TCAD.2022.3223843.

[174] N. K. Upadhyay, H. Jiang, Z. Wang, S. Asapu, Q. Xia, and J. Joshua Yang, "Emerging Memory Devices for Neuromorphic Computing", *Adv. Mater. Technol.*, vol. 4, no. 4, p. 1 800 589, 2019. DOI: 10.1002/admt.201800589.

[175] A. Ruospo, M. S. Reorda, R. Mariani, and E. Sanchez, "An Effective Iterative Statistical Fault Injection Methodology for Deep Neural Networks", *IEEE Trans. Comput.*, vol. 74, no. 7, pp. 2431–2444, 2025. DOI: 10.1109/TC.2025.3566863.

[176] H. Jiang, D. Belkin, S. E. Savel'ev, *et al.*, "A Novel True Random Number Generator Based on a Stochastic Diffusive Memristor", *Nat. Commun.*, vol. 8, no. 1, p. 882, 2017. DOI: 10.1038/s41467-017-00869-x.

[177] G. Kar, A. Fantini, Y.-Y. Chen, *et al.*, "Process-Improved RRAM Cell Performance and Reliability and Paving the Way for Manufacturability and Scalability for High Density Memory Application", in *IEEE Symp. VLSI Technol.*, 2012, pp. 157–158. DOI: 10.1109/VLSIT.2012.6242509.

[178] A. Grossi, C. Zambelli, P. Olivo, *et al.*, "Electrical Characterization and Modeling of Pulse-Based Forming Techniques in RRAM Arrays", *Solid-State Electron.*, vol. 115, pp. 17–25, 2016. DOI: 10.1016/j.sse.2015.10.003.

[179] C. Nguyen, C. Cagli, G. Molas, *et al.*, "Study of Forming Impact on 4Kbit RRAM Array Performances and Reliability", in *IEEE Int. Mem. Workshop*, 2017, pp. 1–4. DOI: 10.1109/IMW.2017.7939105.

[180] N. Raghavan, "Performance and Reliability Trade-offs for High-κ RRAM", *Microelectron. Reliab.*, vol. 54, no. 9-10, pp. 2253–2257, 2014. DOI: 10.1016/j.microrel.2014.07.135.

[181] P. Lorenzi, R. Rao, and F. Irrera, "Forming Kinetics in HfO_2-Based RRAM Cells", *IEEE Trans. Electron Devices*, vol. 60, no. 1, pp. 438–443, 2012. DOI: 10.1109/TED.2012.2227324.

[182] A. Fantini, L. Goux, R. Degraeve, *et al.*, "Intrinsic Switching Variability in HfO_2 RRAM", in *IEEE Int. Mem. Workshop*, 2013, pp. 30–33. DOI: 10.1109/IMW.2013.6582090.

[183] Y.-S. Chen, W.-H. Liu, H.-Y. Lee, *et al.*, "Impact of Compliance Current Overshoot on High Resistance State, Memory Performance, and Device Yield of HfO_x Based Resistive Memory and Its Solution", in *IEEE Int. Symp. VLSI Technol., Syst. Appl.*, 2011, pp. 1–2. DOI: 10.1109/VTSA.2011.5872252.

[184] H.-L. Chang, H.-C. Li, C. Liu, F. Chen, and M.-J. Tsai, "Physical Mechanism of HfO_2-Based Bipolar Resistive Random Access Memory", in *IEEE Int. Symp. VLSI Technol., Syst. Appl.*, 2011, pp. 1–2. DOI: 10.1109/VTSA.2011.5872253.

[185] A. Kalantarian, G. Bersuker, D. Gilmer, *et al.*, "Controlling Uniformity of RRAM Characteristics through the Forming Process", in *IEEE Int. Rel. Phys. Symp.*, 2012, pp. 6C–4. DOI: 10.1109/IRPS.2012.6241874.

[186] K. J. Kuhn, M. D. Giles, D. Becher, *et al.*, "Process Technology Variation", *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2197–2208, 2011. DOI: 10.1109/TED.2011.2121913.

[187] K. Kuhn, C. Kenyon, A. Kornfeld, *et al.*, "Managing Process Variation in Intel's 45nm CMOS Technology.", *Intel Technol. J.*, vol. 12, no. 2, 2008, ISSN: 1535-864X.

[188] Y. G. Fedorenko, *Ion-Beam-Induced Defects in CMOS Technology: Methods of Study*. IntechOpen, 2017. [Online]. Available: https://www.intechopen.com/chapters/53462.

[189] M. Sachdev and J. Pineda de Gyvez, *Defect-Oriented Testing for Nano-Metric CMOS VLSI Circuits* (Frontiers in Electronic Testing), 2nd. Springer, 2007, vol. 34, ISBN: 978-0-387-46546-3. DOI: 10.1007/0-387-46547-2. [Online]. Available: https://link.springer.com/book/10.1007/0-387-46547-2.

[190] P.-S. Chen, Y.-S. Chen, H.-Y. Lee, *et al.*, "Impact of Flattened TiN Electrode on the Memory Performance of HfO_2 Based Resistive Memory", *Electrochem. Solid-State Lett.*, vol. 15, no. 4, H136, 2012. DOI: 10.1149/2.001205esl.

[191] M. Lanza, K. Zhang, M. Porti, *et al.*, "Grain Boundaries as Preferential Sites for Resistive Switching in the HfO_2 Resistive Random Access Memory Structures", *Appl. Phys. Lett.*, vol. 100, no. 12, 2012. DOI: 10.1063/1.3697648.

[192] K. Shubhakar, S. Mei, M. Bosman, *et al.*, "Conductive Filament Formation at Grain Boundary Locations in Polycrystalline HfO_2-Based MIM Stacks: Computational and Physical Insight", *Microelectron. Reliab.*, vol. 64, pp. 204–209, 2016. DOI: 10.1016/j.microrel.2016.07.145.

[193] A. Grossi, E. Perez, C. Zambelli, P. Olivo, and C. Wenger, "Performance and Reliability Comparison of 1T-1R RRAM Arrays with Amorphous and Polycrystalline HfO_2", in *EuroSOI-ULIS*, 2016, pp. 80–83. DOI: 10.1109/ULIS.2016.7440057.

[194] B. Chen, Y. Lu, B. Gao, *et al.*, "Physical Mechanisms of Endurance Degradation in TMO-RRAM", in *IEEE Int. Electron Devices Meet.*, 2011, pp. 12.3.1–12.3.4. DOI: 10.1109/IEDM.2011.6131539.

[195] Y. Hayakawa, A. Himeno, R. Yasuhara, *et al.*, "Highly Reliable TaO_x ReRAM with Centralized Filament for 28-nm Embedded Application", in *IEEE Symp. VLSI Technol.*, 2015, T14–T15. DOI: 10.1109/VLSIT.2015.7223684.

[196] K. Beckmann, J. Holt, W. Olin-Ammentorp, Z. Alamgir, J. Van Nostrand, and N. Cady, "The Effect of Reactive Ion Etch (RIE) Process Conditions on ReRAM Device Performance", *Semicond. Sci. Technol.*, vol. 32, no. 9, p. 095 013, 2017. DOI: 10.1088/1361-6641/aa7eed.

[197] J. H. Nickel, J. P. Strachan, M. D. Pickett, *et al.*, "Memristor Structures for High Scalability: Non-Linear and Symmetric Devices Utilizing Fabrication Friendly Materials and Processes", *Microelectron. Reliab.*, vol. 103, pp. 66–69, 2013. DOI: 10.1016/j.mee.2012.09.007.

[198] Y. Li, S. Long, Q. Liu, H. Lv, and M. Liu, "Resistive Switching Performance Improvement via Modulating Nanoscale Conductive Filament, Involving the Application of Two-Dimensional Layered Materials", *Small*, vol. 13, no. 35, p. 1 604 306, 2017. DOI: 10.1002/smll.201604306.

[199] Y. Jeong, M. A. Zidan, and W. D. Lu, "Parasitic Effect Analysis in Memristor-Array-Based Neuromorphic Systems", *IEEE Trans. Nanotechnol.*, vol. 17, no. 1, pp. 184–193, 2017. DOI: 10.1109/TNANO.2017.2784364.

[200] R. Degraeve, A. Fantini, N. Raghavan, *et al.*, "Causes and Consequences of the Stochastic Aspect of Filamentary RRAM", *Microelectron. Eng.*, vol. 147, 2015. DOI: 10.1016/J.MEE.2015.04.025.

[201] A. Singh, M. Zahedi, T. Shahroodi, *et al.*, "CIM-Based Robust Logic Accelerator Using 28 nm STT-MRAM Characterization Chip Tape-Out", in *IEEE Int. Conf. Artif. Intell. Circuits Syst.*, 2022, pp. 451–454. DOI: 10.1109/AICAS54282.2022.9869993.

[202] M. Fieback, C. Münch, A. Gebregiorgis, *et al.*, "PVT Analysis for RRAM and STT-MRAM-Based Logic Computation-in-Memory", in *IEEE Eur. Test Symp.*, 2022, pp. 1–6. DOI: 10.1109/ETS54262.2022.9810436.

[203] J.-H. Lee, D.-H. Lim, H. Jeong, H. Ma, and L. Shi, "Exploring Cycle-to-Cycle and Device-to-Device Variation Tolerance in MLC Storage-Based Neural Network Training", *IEEE Trans. Electron Devices*, vol. 66, no. 5, pp. 2172–2178, 2019. DOI: 10.1109/TED.2019.2906249.

[204] H.-Y. Lee, Y. Chen, P. Chen, *et al.*, "Evidence and Solution of Over-RESET Problem for HfO_x Based Resistive Memory with Sub-ns Switching Speed and High Endurance", in *IEEE Int. Electron Devices Meet.*, 2010, pp. 19–7. DOI: 10.1109/IEDM.2010.5703395.

[205] H. García, J. Boo, G. Vinuesa, *et al.*, "Influences of the Temperature on the Electrical Properties of HfO_2-Based Resistive Switching Devices", *Electronics*, vol. 10, no. 22, p. 2816, 2021. DOI: 10.3390/electronics10222816.

[206] Y. Y. Chen, R. Degraeve, S. Clima, *et al.*, "Understanding of the Endurance Failure in Scaled HfO_2-Based 1T1R RRAM through Vacancy Mobility Degradation", in *IEEE Int. Electron Devices Meet.*, 2012, pp. 20–3. DOI: 10.1109/IEDM.2012.6479079.

[207] P. Huang, B. Chen, Y. J. Wang, *et al.*, "Analytic Model of Endurance Degradation and its Practical Applications for Operation Scheme Optimization in Metal Oxide Based RRAM", in *IEEE Int. Electron Devices Meet.*, 2013, pp. 22.5.1–22.5.4. DOI: 10.1109/IEDM.2013.6724685.

[208] Y. Chen, H. Lee, P. Chen, *et al.*, "Highly Scalable Hafnium Oxide Memory with Improvements of Resistive Distribution and Read Disturb Immunity", in *IEEE Int. Electron Devices Meet.*, 2009, pp. 1–4. DOI: 10.1109/IEDM.2009.5424411.

[209] S. J. Fonash, "An Overview of Dry Etching Damage and Contamination Effects", *J. Electrochem. Soc.*, vol. 137, no. 12, p. 3885, 1990. DOI: 10.1149/1.2086322.

[210] L. Ephrath and R. Bennett, "RIE Contamination of Etched Silicon Surfaces", *J. Electrochem. Soc.*, vol. 129, no. 8, p. 1822, 1982. DOI: 10.1149/1.2124301.

[211] A. J. Van de Goor and I. B. Tlili, "March Tests for Word-Oriented Memories", in *ACM/IEEE Des. Autom. Test Eur. Conf. Exhib.*, 1998, pp. 501–508. DOI: 10.1109/DATE.1998.655905.

[212] H. T. Nagle, S. Roy, C. Hawkins, M. McNamer, and R. Fritzemeier, "Design for Testability and Built-In Self Test: A Review", *IEEE Trans. Ind. Electron.*, vol. 36, no. 2, pp. 129–140, 1989. DOI: 10.1109/41.19062.

[213] E. J. McCluskey, "Built-in Self-Test Techniques", *IEEE Des. Test Comput.*, vol. 2, no. 2, pp. 21–28, 1985. DOI: 10.1109/MDT.1985.294856.

[214] M. Fieback, G. C. Medeiros, A. Gebregiorgis, H. Aziza, M. Taouil, and S. Hamdioui, "Intermittent Undefined State Fault in RRAMs", in *IEEE Eur. Test Symp.*, 2021, pp. 1–6. DOI: 10.1109/ETS50041.2021.9465401.

[215] S. Hamdioui and A. J. Van De Goor, "Efficient Tests for Realistic Faults in Dual-Port SRAMs", *IEEE Trans. Comput.*, vol. 51, no. 5, pp. 460–473, 2002. DOI: 10.1109/TC.2002.1004586.

[216] S. M. Nair, R. Bishnoi, M. B. Tahoori, *et al.*, "Defect Injection, Fault Modeling and Test Algorithm Generation Methodology for STT-MRAM", in *IEEE Int. Test Conf.*, IEEE, 2019, pp. 1–10. DOI: 10.1109/TEST.2018.8624725.

[217] L. M. Poehls, M. C. Fieback, S. Hoffmann-Eifert, *et al.*, "Review of Manufacturing Process Defects and Their Effects on Memristive Devices", *J. Electron. Test.*, vol. 37, no. 4, pp. 427–437, 2021. DOI: 10.1007/s10836-021-05968-8.

[218] A. J. van de Goor, *Testing Semiconductor Memories: Theory and Practice*. New York: John Wiley & Sons Inc, 1991, ISBN: 978-0471925866. [Online]. Available: https://www.amazon.com/Testing-Semiconductor-Memories-Theory-Practice/dp/0471925861.

[219] S. Hamdioui, A. J. Van De Goor, and M. Rodgers, "March SS: A Test for All Static Simple RAM Faults", in *IEEE Int. Workshop Mem. Technol., Des. Test.*, 2002, pp. 95–100. DOI: 10.1109/MTDT.2002.1029769.

[220] A. van de Goor and Z. Al-Ars, "Functional Memory Faults: A Formal Notation and a Taxonomy", in *IEEE VLSI Test Symp.*, 2000, pp. 281–289. DOI: 10.1109/VTEST.2000.843856.

[221] G. C. Medeiros, G. S. Member, M. Fieback, *et al.*, "Hard-to-Detect Fault Analysis in FinFET SRAMs", *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 29, no. 6, pp. 1271–1284, 2021. DOI: 10.1109/TVLSI.2021.3071940.

[222] ASU, *Predictive Technology Model (PTM)*, 2012.

[223] H. Li, P. Huang, B. Gao, B. Chen, X. Liu, and J. Kang, "A SPICE Model of Resistive Random Access Memory for Large-Scale Memory Array Simulation", *IEEE Electron Device Lett.*, vol. 35, no. 2, pp. 211–213, 2014. DOI: 10.1109/LED.2013.2293354.

[224] S. Mitchell, M. O'Sullivan, and I. Dunning, "PuLP: A linear programming toolkit for Python", *Optimization Online*, 2011. [Online]. Available: https://optimization-online.org/2011/09/3178/.

[225] S. Kannan, J. Rajendran, R. Karri, and O. Sinanoglu, "Sneak-Path Testing of Crossbar-Based Nonvolatile Random Access Memories", *IEEE Trans. Nanotechnol.*, vol. 12, no. 3, pp. 413–426, 2013. DOI: 10.1109/TNANO.2013.2253329.

[226] P. Liu, Z. You, J. Kuang, Z. Hu, H. Duan, and W. Wang, "Efficient March Test Algorithm for 1T1R Cross-Bar with Complete Fault Coverage", *Electron. Lett.*, vol. 52, 18 2016. DOI: 10.1049/el.2016.1693.

[227] Y. Luo, X. Cui, M. Luo, and Q. Lin, "A High Fault Coverage March Test for 1T1R Memristor Array", in *IEEE Int. Conf. Electron Devices Solid-State Circuits*, 2017, pp. 1–2. DOI: 10.1109/EDSSC.2017.8126415.

[228] N. Xiao, M. A. Villena, B. Yuan, *et al.*, "Resistive Random Access Memory Cells with a Bilayer TiO_2/SiO_x Insulating Stack for Simultaneous Filamentary and Distributed Resistive Switching", *Adv. Funct. Mater.*, vol. 27, no. 33, p. 1 700 384, 2017. DOI: 10.1002/adfm.201700384.

[229] H. Aziza, J. Postel-Pellerin, and M. Moreau, "STATE: A Test Structure for Rapid and Reliable Prediction of Resistive RAM Endurance", *IEEE Trans. Device Mater. Reliab.*, vol. 22, no. 4, pp. 500–505, 2022. DOI: 10.1109/TDMR.2022.3213191.

[230] Y. Y. Chen, B. Govoreanu, L. Goux, *et al.*, "Balancing SET/RESET Pulse for > 10^{10} Endurance in HfO_2 / Hf 1T1R Bipolar RRAM", *IEEE Trans. Electron Devices*, vol. 59, no. 12, pp. 3243–3249, 2012. DOI: 10.1109/TED.2012.2218607.

[231] A. Schönhals, D. Wouters, A. Marchewka, *et al.*, "Critical ReRAM Stack Parameters Controlling Complimentary Versus Bipolar Resistive Switching", in *IEEE Int. Mem. Workshop*, 2015, pp. 1–4. DOI: 10.1109/IMW.2015.7150281.

[232] C. Wang, H. Wu, B. Gao, T. Zhang, Y. Yang, and H. Qian, "Conduction Mechanisms, Dynamics and Stability in ReRAMs", *Microelectron. Eng.*, vol. 187, pp. 121–133, 2018. DOI: 10.1016/j.mee.2017.11.003.

[233] M. Ismail, M. K. Rahmani, S. A. Khan, *et al.*, "Effects of Gibbs Free Energy Difference and Oxygen Vacancies Distribution in a Bilayer ZnO/ZrO_2 Structure for Applications to Bipolar Resistive Switching", *Appl. Surf. Sci.*, vol. 498, p. 143 833, 2019. DOI: 10.1016/j.apsusc.2019.143833.

[234] Y. Y. Chen, M. Komura, R. Degraeve, *et al.*, "Improvement of Data Retention in HfO_2/Hf 1T1R RRAM Cell under Low Operating Current", in *IEEE Int. Electron Devices Meet.*, 2013, pp. 10.1.1–10.1.4. DOI: 10.1109/IEDM.2013.6724598.

[235] A. Grossi, E. Perez, C. Zambelli, *et al.*, "Impact of the Precursor Chemistry and Process Conditions on the Cell-to-Cell Variability in 1T-1R Based HfO_2 RRAM Devices", *Sci. Rep.*, vol. 8, no. 1, p. 11 160, 2018. DOI: 10.1038/s41598-018-29548-7.

[236] T. Diokh, E. Le-Roux, S. Jeannot, *et al.*, "Investigation of the Impact of the Oxide Thickness and RESET Conditions on Disturb in HfO_2-RRAM Integrated in a 65 nm CMOS Technology", in *IEEE Int. Rel. Phys. Symp.*, 2013, 5E.4.1–5E.4.4. DOI: 10.1109/IRPS.2013.6532043.

[237] A. Regoutz, I. Gupta, A. Serb, *et al.*, "Role and Optimization of the Active Oxide Layer in TiO_2-Based RRAM", *Adv. Funct. Mater.*, vol. 26, no. 4, pp. 507–513, 2016. DOI: 10.1002/adfm.201503522.

[238] Z. Shen, Y. Qi, I. Z. Mitrovic, *et al.*, "Effect of Annealing Temperature for Ni/AlO_x/Pt RRAM Devices Fabricated with Solution-Based Dielectric", *Micromachines*, vol. 10, no. 7, p. 446, 2019. DOI: 10.1016/S0254-0584(02)00455-8.

[239] M. Boronat, A. Corma, F. Illas, J. Radilla, T. Ródenas, and M. J. Sabater, "Mechanism of Selective Alcohol Oxidation to Aldehydes on Gold Catalysts: Influence of Surface Roughness on Reactivity", *J. Catal.*, vol. 278, no. 1, pp. 50–58, 2011. DOI: 10.1016/j.jcat.2010.11.013.

[240] A. Allahbakhsh, F. Sharif, and S. Mazinani, "The Influence of Oxygen-Containing Functional Groups on the Surface Behavior and Roughness Characteristics of Graphene Oxide", *Nano*, vol. 8, no. 4, p. 1 350 045, 2013. DOI: 10.1142/S1793292013500458.

[241] A. Genreith-Schriever and R. De Souza, "Field-Enhanced Ion Transport in Solids: Reexamination with Molecular Dynamics Simulations", *Phys. Rev., B*, vol. 94, no. 22, p. 224 304, 2016. DOI: 10.1103/PhysRevB.94.224304.

[242] S. N. Mozaffari, S. Tragoudas, and T. Haniotakis, "More Efficient Testing of Metal-Oxide Memristor–Based Memory", *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 36, no. 6, pp. 1018–1029, 2017. DOI: 10.1109/TCAD.2016.2608863.

[243] M. A. Yaldagard, S. Diware, R. V. Joshi, S. Hamdioui, and R. Bishnoi, "Read-Disturb Detection Methodology for RRAM-Based Computation-in-Memory Architecture", in *IEEE Int. Conf. Artif. Intell. Circuits Syst.*, 2023, pp. 1–5. DOI: 10.1109/AICAS57966.2023.10168638.

[244] W. Y. Park, W. Ju, Y. S. Ko, *et al.*, "Improvement of Sensing Margin and Reset Switching Fail of RRAM", *Solid-State Electron.*, vol. 156, pp. 87–91, 2019. DOI: 10.1016/j.sse.2019.02.010.

[245] C.-W. Zhong, W.-H. Tzeng, K.-C. Liu, *et al.*, "Effect of ITO Electrode with Different Oxygen Contents on the Electrical Characteristics of HfO_x RRAM Devices", *Surf. Coat. Technol.*, vol. 231, pp. 563–566, 2013. DOI: 10.1016/j.surfcoat.2012.07.039.

[246] N. Raghavan, R. Degraeve, A. Fantini, *et al.*, "Stochastic Variability of Vacancy Filament Configuration in Ultra-Thin Dielectric RRAM and Its Impact on OFF-State Reliability", in *IEEE Int. Electron Devices Meet.*, 2013, pp. 21.1.1–21.1.4. DOI: 10.1109/IEDM.2013.6724674.

[247] Z. Weinberg, "On Tunneling in Metal-Oxide-Silicon Structures", *J. Appl. Phys.*, vol. 53, no. 7, pp. 5052–5056, 1982. DOI: 10.1063/1.331336.

[248] J. J. Yang, M. D. Pickett, X. Li, D. A. Ohlberg, D. R. Stewart, and R. S. Williams, "Memristive Switching Mechanism for Metal/Oxide/Metal Nanodevices", *Nat. Nanotechnol.*, vol. 3, no. 7, pp. 429–433, 2008. DOI: 10.1038/nnano.2008.160.

[249] P. Liu, J. Wu, Z. You, M. Elimu, W. Wang, and S. Cai, "Defect Analysis and Parallel March Test Algorithm for 3D Hybrid CMOS-Memristor Memory", in *IEEE Asian Test Symp.*, 2018, pp. 25–29. DOI: 10.1109/ATS.2018.00016.

[250] S. Hamdioui, Z. Al-Ars, and A. van de Goor, "Testing Static and Dynamic Faults in Random Access Memories", in *IEEE VLSI Test Symp.*, 2002, pp. 395–400. DOI: 10.1109/VTS.2002.1011170.

[251] P. Liu, Z. You, J. Wu, B. Liu, Y. Han, and K. Chakrabarty, "Fault Modeling and Efficient Testing of Memristor-Based Memory", *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 68, no. 11, pp. 4444–4455, 2021. DOI: 10.1109/TCSI.2021.3098639.

[252] V. A. Hongal, R. Kotikalapudi, Y.-B. Kim, and M. Choi, "A Novel 'Divide and Conquer' Testing Technique for Memristor Based Lookup Table", in *IEEE Int. Midwest Symp. Circuits Syst.*, 2011, pp. 1–4. DOI: 10.1109/MWSCAS.2011.6026406.

[253] P. Liu, Z. You, J. Kuang, Z. Hu, and W. Wang, "Logic Operation-Based DFT Method and 1R Memristive Crossbar March-Like Test Algorithm", *IEICE Trans. Electron.*, vol. 12, no. 23, pp. 20 150 839–20 150 839, 2015. DOI: 10.1587/elex.12.20150839.

[254] A. F. Gomez, F. Forero, K. Roy, and V. Champac, "Robust Detection of Bridge Defects in STT-MRAM Cells under Process Variations", in *IFIP/IEEE Int. Conf. Very Large Scale Integr.*, 2018, pp. 65–70. DOI: 10.1109/VLSI-SoC.2018.8645022.

[255] C.-L. Su, Y.-T. Yeh, and C.-W. Wu, "An Integrated ECC and Redundancy Repair Scheme for Memory Reliability Enhancement", in *IEEE Int. Symp. Defect Fault Toler. VLSI Syst.*, 2005, pp. 81–89. DOI: 10.1109/DFTVS.2005.18.

[256] G. Mayuga, Y. Yamato, T. Yoneda, M. Inoue, and Y. Sato, "An ECC-Based Memory Architecture with Online Self-Repair Capabilities for Reliability Enhancement", in *IEEE Eur. Test Symp.*, 2015, pp. 1–6. DOI: 10.1109/ETS.2015.7138734.

[257] A. Baroni, C. Zambelli, P. Olivo, E. Pérez, C. Wenger, and D. Ielmini, "Tackling the Low Conductance State Drift through Incremental Reset and Verify in RRAM arrays", in *IEEE Int. Integr. Rel. Workshop*, 2021, pp. 1–5. DOI: 10.1109/IIRW53245.2021.9635613.

[258] N. Raghavan, D. D. Frey, M. Bosman, and K. L. Pey, "Monte Carlo Model of RESET Stochastics and Failure Rate Estimation of Read Disturb Mechanism in HfO_x RRAM", in *IEEE Int. Rel. Phys. Symp.*, 2015, 5B.2.1–5B.2.9. DOI: 10.1109/IRPS.2015.7112743.

[259] C. Bengel, J. Mohr, S. Wiefels, *et al.*, "Reliability Aspects of Binary Vector-Matrix-Multiplications Using ReRAM Devices", *Neuromorph. Comput. Eng.*, vol. 2, no. 3, p. 034 001, 2022. DOI: 10.1088/2634-4386/ac6d04.

[260] D. Gaol, G. L. Zhang, X. Yin, B. Li, U. Schlichtmann, and C. Zhuo, "Reliable Memristor-based Neuromorphic Design Using Variation- and Defect-Aware Training", in *IEEE/ACM Int. Conf. Comput.-Aided Des.*, 2021, pp. 1–9. DOI: 10.1109/ICCAD51958.2021.9643468.

[261] S. Diware, A. Gebregiorgis, R. V. Joshi, S. Hamdioui, and R. Bishnoi, "Mapping-Aware Biased Training for Accurate Memristor-based Neural Networks", in *IEEE Int. Conf. Artif. Intell. Circuits Syst.*, 2023, pp. 1–5. DOI: 10.1109/AICAS57966.2023.10168661.

[262] W. Shim, Y. Luo, J.-S. Seo, and S. Yu, "Investigation of Read Disturb and Bipolar Read Scheme on Multilevel RRAM-Based Deep Learning Inference Engine", *IEEE Trans. Electron Devices*, vol. 67, no. 6, pp. 2318–2323, 2020. DOI: 10.1109/TED.2020.2985013.

[263] V. Mahboubi, Á. Gómez, A. Calomarde, D. Arumí, R. Rodríguez, and S. Manich, "On the Fine Tuning of RRAM Resistance Under Variability Using Current Pulses at SET", in *IEEE Int. Symp. On-Line Testing Robust Syst. Design*, 2024, pp. 1–7. DOI: 10.1109/IOLTS60994.2024.10616054.

[264] J. Park and S. Kim, "Improving Endurance and Reliability by Optimizing the Alternating Voltage in Pt/ZnO/TiN RRAM", *Results Phys.*, vol. 39, p. 105 731, 2022. DOI: 10.1016/j.rinp.2022.105731.

[265] H. Kim, H. Ye, T. Mudge, R. Dreslinski, and N. Talati, "RecPIM: A PIM-Enabled DRAM-RRAM Hybrid Memory System For Recommendation Models", in *IEEE/ACM Int. Symp. Low Power Electron. Design*, 2023, pp. 1–6. DOI: 10.1109/ISLPED58423.2023.10244420.

[266] P.-Y. Chen, X. Peng, and S. Yu, "NeuroSim+: An Integrated Device-to-Algorithm Framework for Benchmarking Synaptic Devices and Array Architectures", in *IEEE Int. Electron Devices Meet.*, 2017, pp. 6–1. DOI: 10.1109/IEDM.2017.8268337.

[267] S.-C. Hung, A. Chaudhuri, S. Banerjee, and K. Chakrabarty, "Fault Diagnosis for Resistive Random-Access Memory and Monolithic Inter-tier Vias in Monolithic 3D Integration", in *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 32, 2024, pp. 1336–1349. DOI: 10.1109/TVLSI.2024.3380549.

[268] R. Adapa, S. Tragoudas, and M. K. Michael, "Accelerating Diagnosis via Dominance Relations between Sets of Faults", in *IEEE VLSI Test Symp.*, 2007, pp. 219–224. DOI: 10.1109/VTS.2007.10.

[269] T. Li, X. Bi, N. Jing, X. Liang, and L. Jiang, "Sneak-Path Based Test and Diagnosis for 1R RRAM Crossbar Using Voltage Bias Technique", in *ACM/IEEE Design Autom. Conf.*, 2017, pp. 1–6. DOI: 10.1145/3061639.3062318.

[270] A. Aouichi, S. Yuan, M. Fieback, *et al.*, "Device Aware Diagnosis for Unique Defects in STT-MRAMs", in *IEEE Asian Test Symp.*, 2023, pp. 1–6. DOI: 10.1109/ATS59501.2023.10317952.

[271] 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 *IEEE Int. Symp. Phys. Failure Anal. Integr. Circuits*, 2013, pp. 490–492. DOI: 10.1109/IPFA.2013.6599207.

[272] A. V. Khvalkovskiy, D. Apalkov, S. Watts, *et al.*, "Basic Principles of STT-MRAM Cell Operation in Memory Arrays", *J. Phys. D: Appl. Phys*, vol. 46, no. 13, p. 139 601, 2013. DOI: 10.1088/0022-3727/46/13/139601.

NOMENCLATURE

ACRONYMS

AD	Address Decoders
ALD	Atomic Layer Deposition
BE	Bottom Electrode
BEOL	Back-End-Of-Line
BISR	Built-In-Self-Repair
BIST	Built-In-Self-Test
BL	Bit Line
BS	Bipolar Switching
C2C	Cycle-to-Cycle
CAT	Cell-Aware Test
CBRAM	Conductive-Bridge Random-Access Memory
CF	Conducting Filament
CIM	Computation in Memory
CMP	Chemical-Mechanical Polishing
CS	Column Select
D2D	Device-to-Device
DA-Diagnosis	Device-Aware Diagnosis
DAT	Device-Aware Test
DB	Data Background
DfT	Design-for-Testability
DPPB	Defective Part Per Billion
DRAM	Dynamic Random Access Memory
ECC	Error Correction Codes

EtD	Easy-to-Detect
FC	Fault Coverage
FE	Ferroelectric
FeFET	Ferroelectric Field-Effect Transistor
FEOL	Front-End-Of-Line
FeRAM	Ferroelectric Random-Access Memory
FET	Field-Effect Transistor
FP	Fault Primitive
HDDs	Hard Disk Drives
HRS	High-Resistive State
HtD	Hard-to-Detect
ID	Ion Depletion
ILP	Integer Linear Programming
IP	Intellectual Property
IUSF	Intermittent Undefined State Fault
LRS	Low-Resistive State
MC	Monte Carlo
MIM	Metal-Insulator-Metal
MLC	Multi-Level Storage
MRAM	Magnetoresistive Random-Access Memory
MTJs	Magnetic Tunnel Junctions
MW	Memory Window
NVMs	Non-Volatile Memories
OR	Over RESET
OV	Oxygen Vacancies
OxRAM	Oxide Random-Access Memory
PCRAM	Phase Change Random Access Memory
PTM	Predictive Technology Model

PVD	Physical Vapor Deposition
PVT	Process, Voltage, and Temperature
RL	Reference Line
RRAM	Resistive Random-Access Memory
RS	Resistive Switching
SA	Sense Amplifier
SCLC	Space Charge Limited Conduction
SCM	Storage-Class Memory
SEM	Scanning Electron Microscopy
SL	Select Line
SRAM	Static Random Access Memory
STDP	Spike-Time-Dependent Plasticity
STT-MRAM	Spin-Transfer Torque MRAM
TE	Top Electrode
TEM	Transmission Electron Microscopy
TMR	Tunneling Magneto Resistance
TRNGs	True Random Number Generators
VC-MRAM	Voltage-Controlled MRAM
VCM	Valence Change Mechanism
WD	Write Driver
WL	Word Line

SYMBOLS

$\hbar$	Planck's reduced constant
C_a	agressor cell
C_b	the base cell
C_c	cell in the same column
C_d	adjacent diagonal cell
C_r	cell in the same row
C_v	victim cell
GND	ground voltage
V_{RESET}	RESET voltage
V_{SET}	SET voltage
μ	the mean value
ϕ_b	barrier height
σ	standard deviation
$\Updownarrow$	addressing order irrelevant
$D_{\max}$	maximal duration suffering from ion depletion
F_{ox}	electric field cross the oxide
I_{BL}	current flowing into BL
I_{forming}	forming current
I_{ion}	the ionic current
I_{primary}	the primary current in the path
I_{SL}	current flowing into SL
I_{sneak}	the sneak current in the path
I_{TE}	current through top electrode to bottom electrode
m^*_{ox}	electron effective mass in the insulator
$N_{\text{disc,max}}$	maximum oxygen vacancy concentration in the disc
$N_{\text{disc,min}}$	minimum oxygen vacancy concentration in the disc
N_{disc}	the concentration of oxygen vacancies in the disc region

$O(n^2)$	oxygen ions
P_{ID}	occurrence probability of ion depletion
P_{IUSF}	occurrence probability of intermittent undefined state fault
P_{OR}	occurrence probability of over RESET
q	electron charge
R_{disc}	resistance in disc
R_{plug}	resistance in plug
R_{RESET}	resistance in RESET
R_{series}	series resistance
R_{SET}	resistance in SET
V_{DD}	supply voltage
V_{read}	read voltage
V_{RESET}	RESET voltage
V_{schottky}	Schottky-like contact
V_{SET}	SET voltage
V_{stop}	the terminal voltage of an I–V sweep
V_{TH}	thereshold voltage
V_{forming}	forming voltage
V_{TE}	voltage from TE to BE

Curriculum Vitæ

Hanzhi Xun

Personal Info.

Emails: hzxun_chn@163.com or xunhanzhi@gmail.com

29-07-1996 Born in Xi'an, Shaanxi, China

Education

2021–2025 Ph.D. degree in Electronic Science and Technology
Delft University of Technology, the Netherlands
Thesis title: Advanced Testing and Reliability Enhancements for RRAMs
Supervisors: Prof.dr.ir. S. Hamdioui, Dr.ir. M. Fieback Dr.ir. M. Taouil

2019–2021 M.Sc. degree in Information, Production and Systems Engineering
Waseda University, Japan
Thesis title: Fast Analog Neuromorphic In-memory Classifier Based on Novel Flash Memory Array Architecture
Supervisor: Prof.dr. T. Ohsawa

2018–2021 M.Eng. degree in Software Engineering
Xidian University, China
Thesis title: Research on a Deep Learning Neural Network Circuit Design Based on Floating Gate Transistor
Supervisor: Prof.dr. H. Y. Hu

2014–2018 B.Sc. degree in Microelectronic Science and Engineering
Xidian University, China

Awards

2024 Ph.D. Forum Award at ETS'24

2020 ISIPS Conference Best Presentation Award

2019 Japanese Ministry of Education Honours scholarship

2019 Kitakyushu Science and Research Park scholarship

MAIN PUBLICATIONS

8. **H. Xun**, M. Fieback, S. Yuan, H. Aziza, M. Taouil, and S. Hamdioui, "Ion Depletion Defects in RRAMs: Characterization, Modeling, and Test", *IEEE Transactions on Computers*, pp. 1–14, 2025, (Under review, 2025 impact factor = 3.6)

7. **H. Xun**, M. Fieback, S. Yuan, C. Wang, E. Hua, L. B. Poehls, H. Aziza, R. Cantoro, B. Rajendra, M. Taouil, and S. Hamdioui, "In-Field Monitoring and Preventing Read Disturb Faults in RRAMs", in *IEEE European Test Symposium (ETS)*, Tallinn, Estonia, 2025, pp. 1–6. DOI: 10.1109/ETS63895.2025.11049639

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

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

4. **H. Xun**, M. Fieback, S. Yuan, H. Aziza, M. Taouil, and S. Hamdioui, "Device-Aware Diagnosis for Yield Learning in RRAMs", in *ACM/IEEE Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Valencia, Spain, 2024, pp. 1–6. DOI: 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

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

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 *IEEE European Test Symposium (ETS)*, Venezia, Italy, 2023, pp. 1–6. DOI: 10.1109/ETS56758.2023.10174106

OTHER PUBLICATIONS

9. E. A. Serlis, **H. Xun**, M. Taouil, S. Hamdioui, and M. Fieback, "Structural Testing of a RRAM-based AI Accelerator Core", in *IEEE European Test Symposium (ETS)*, Tallinn, Estonia, 2025, pp. 1–6. DOI: 10.1109/ETS63895.2025.11049610

8. 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: 10.1145/3658617.3697755

7. H. Aziza, M. Fieback, S. Hamdioui, **H. Xun**, and M. Taouil, "Conductance Variability in RRAM and its Implications at the Neural Network Level", *Microelectronics Reliability*, vol. 166, p. 115 594, 2025, ISSN: 0026-2714. DOI: https://doi.org/10.1016/j.microrel.2025.115594. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0026271425000071

6. 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

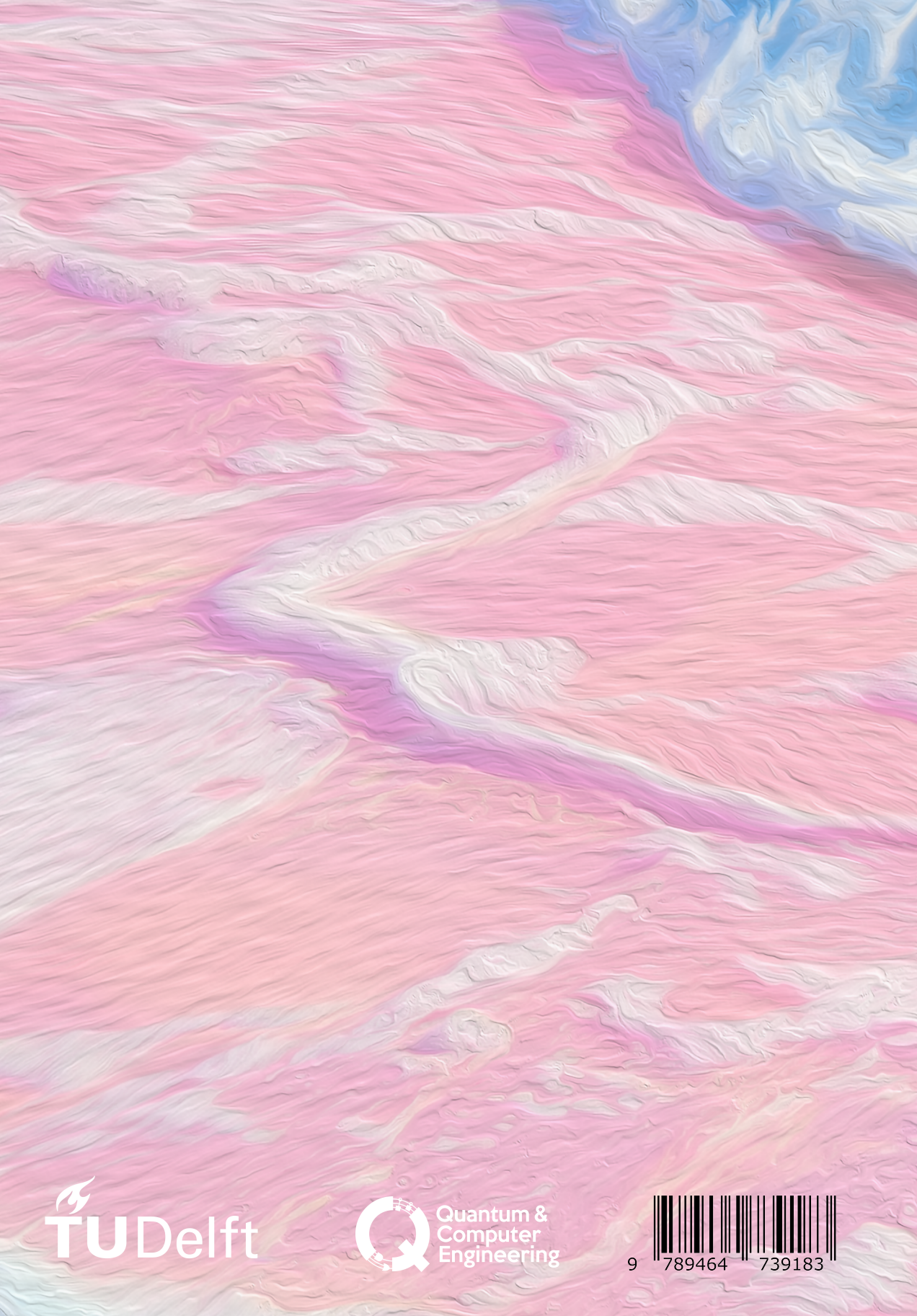
5. C. Wang, S. Yuan, **H. Xun**, C. Li, M. Taouil, M. Fieback, D. Chen, X. Li, L. Wang, R. Cantoro, C. Yin, and S. Hamdioui, "Defects, Fault Modeling, and Test Development Framework for FeFETs", in *IEEE International Test Conference (ITC)*, San Diego, CA, USA, 2024, pp. 91–95. DOI: 10.1109/ITC51657.2024.00026

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

3. H. Aziza, J. Postel-Pellerin, M. Fieback, S. Hamdioui, **H. Xun**, M. Taouil, K. Coulié, and W. Rahajandraibe, "Analysis of Conductance Variability in RRAM for Accurate Neuromorphic Computing", in *IEEE 25th Latin American Test Symposium (LATS)*, 2024, pp. 1–5. DOI: 10.1109/LATS62223.2024.10534620

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

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


TUDelft
Quantum &
Computer
Engineering
9 789464 739183