

Multi-Scale Reliability Analysis and AI-Based Automation for Advanced Semiconductor Circuits

Zur Erlangung des akademischen Grades eines
Doktors der Ingenieurwissenschaften
von der KIT-Fakultät für Informatik
des Karlsruher Instituts für Technologie (KIT)

genehmigte

Dissertation
von

Zhe Zhang

aus Tangshan, China

Tag der mündlichen Prüfung:

22. Juli 2026

1. Referent:

Prof. Dr. Mehdi B. Tahoori
Karlsruher Institut für Technologie (KIT)

2. Referent:

Prof. Dr. Hussam Amrouch
Technische Universität München (TUM)

Acknowledgement

First and foremost, I would like to express my profound gratitude to Prof. Dr. Mehdi Tahoori. His mentorship has been the foundation of my PhD journey. Through his encouragement, insightful feedback, and unwavering support, he guided me through every challenge and achievement along the way. His ability to combine scientific rigor with genuine care for his students has shaped not only my research, but also my growth as a scholar and as a person. I am truly thankful for the opportunity to work under his supervision and to learn from his vision, leadership, and expertise.

I am equally grateful to my colleagues, collaborators, and the broader research community who enriched this journey. I deeply appreciate the opportunity to work across institutions and countries, an experience that broadened both my technical and personal horizons. I would especially like to thank Dr. Sani Nassif, whose expertise, sharp intuition, and generosity in discussion have had a lasting influence on how I approach research. Collaborating with him has been both an honor and a privilege. My sincere appreciation also goes to Dr. Mahta Mayahinia, whose cooperation, enthusiasm, and technical skill made many demanding projects not only possible but also genuinely enjoyable. Working with her has been one of the most rewarding aspects of my research life. I would also like to extend my thanks to all colleagues and friends from CDNC, RPTU Kaiserslautern, and beyond, whose support, discussions, and friendship accompanied me throughout this journey and made these years truly memorable.

I would like to express my deepest love and gratitude to my grandparents, Shi Kun, Zhang Xiulian, Zhang Hui, and Li Shuqin, all of whom passed away during the course of my PhD journey. Their absence has been one of the most painful parts of these years. I only wish they could have witnessed this moment and known that I became the first “Dr.” in our family. Coming from a small mountain village, I have always been aware of how far this path has taken me. The distance between where I began and where I stand today feels immense, and I owe much of this journey to the people who believed in me long before I believed in myself. Everything I have achieved is built upon the values they passed on to our family: kindness, resilience, humility, and the belief that education can change a person’s life. This dissertation is dedicated to their memory.

My heartfelt thanks also go to my parents, Zhang Dongqiang and Shi Guohong, whose unwavering support carried me through the most difficult and uncertain moments of my life. Despite the distance, their love and trust always reached me, sometimes through words, sometimes through silence, but always through the reassurance that I was never alone. I am profoundly grateful for their sacrifices, encouragement, and belief in me. Without them, this achievement would never have been possible.

As I write these final words, I realize that my journey at KIT is coming to an end after nearly ten years. A decade is long enough for a place to become a home, for colleagues to become lifelong friends, and for experiences to become memories that I will cherish forever. My time in Karlsruhe shaped who I am not only as a researcher, but also as a person. I am deeply grateful for the countless moments of learning, growth, friendship, laughter, and support that I shared with the people around me during these years.

Although this dissertation marks the completion of one chapter, the spirit of KIT and the warmth of the people I met in Karlsruhe will always remain with me. I am proud to have been part of this community, and I will carry its influence with me wherever life leads next.

Hiermit erkläre ich an Eides statt, dass ich die von mir vorgelegte Arbeit selbstständig verfasst habe, dass ich die verwendeten Quellen, Internet-Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen – die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Karlsruhe, 18. Mai 2026
Zhe Zhang

Abstract

Modern computing systems rely on two fundamental components, logic and memory, whose co-optimization determines overall performance, energy efficiency, and reliability. Continuous technology scaling has significantly improved transistor performance and integration density across both front-end-of-the-line (FEOL) and back-end-of-the-line (BEOL) layers. Advanced technology nodes, such as Fully Depleted Silicon-on-Insulator (FDSOI) and Fin Field-Effect Transistor (FinFET), exemplify these advancements by enabling improved power, performance, and area (PPA).

However, continued scaling also increases the susceptibility of devices and interconnects to reliability degradation mechanisms, including Negative-Bias Temperature Instability (NBTI), electromigration (EM), and radiation-induced soft errors. These effects pose significant challenges to circuit lifetime and functional safety in modern electronic systems.

The interaction of these reliability mechanisms further complicates device behavior and circuit stability. Degradation in one domain can exacerbate failures in another; for example, FEOL transistor aging and BEOL interconnect degradation can jointly impact timing margins and noise immunity at the circuit level. As a result, reliability has evolved into a multi-scale challenge spanning device, circuit, layout, and system domains, requiring coordinated mitigation strategies across all abstraction levels. Addressing these challenges demands comprehensive modeling and simulation frameworks that establish a clear link between physical degradation mechanisms and circuit- and system-level performance.

This thesis focuses on three key reliability concerns: radiation effects, NBTI, and EM. The first objective is to develop accurate reliability evaluation and improvement methodologies for circuit-level lifetime estimation under these degradation mechanisms. To achieve this, a device-to-circuit co-simulation framework is proposed, combined with multi-scale co-optimization techniques to capture interactions between device- and circuit-level effects.

Electronic design automation (EDA) tools, ranging from device-level Technology Computer-Aided Design (TCAD) simulators to circuit-level SPICE models, play a crucial role in bridging semiconductor physics and design implementation, enabling multi-scale reliability analysis. However, as technology nodes advance, EDA workflows become increasingly complex and computationally expensive, limiting the efficiency of reliability evaluation. Within the context of Design–Technology Co-Optimization (DTCO) and System–Technology Co-Optimization (STCO), there is a growing need for efficient and automated multi-scale modeling frameworks.

The second objective of this thesis is to leverage machine learning techniques, including artificial neural networks (ANNs) and Large Language Models (LLMs), to accelerate reliability estimation and simulation, thereby improving the efficiency of reliability-aware design methodologies.

Zusammenfassung

Moderne Rechensysteme basieren auf zwei grundlegenden Komponenten – Logik und Speicher – deren gemeinsame Optimierung die Gesamtleistung, Energieeffizienz und Zuverlässigkeit bestimmt. Die kontinuierliche Technologieskalierung hat die Transistorleistung und Integrationsdichte sowohl in der Front-End-of-Line- (FEOL) als auch in der Back-End-of-Line- (BEOL) Ebene erheblich verbessert. Fortschrittliche Technologieknoten wie Fully Depleted Silicon-on-Insulator (FDSOI) und Fin Field-Effect Transistor (FinFET) veranschaulichen diese Entwicklungen und ermöglichen eine Optimierung von Leistung, Performance und Fläche (PPA).

Mit fortschreitender Skalierung steigt jedoch gleichzeitig die Anfälligkeit von Bauelementen und Interconnects gegenüber Zuverlässigkeitsmechanismen wie Negative-Bias-Temperature-Instability (NBTI), Elektromigration (EM) sowie strahlungsinduzierten Soft Errors. Diese Effekte stellen erhebliche Herausforderungen für die Lebensdauer und funktionale Sicherheit moderner elektronischer Systeme dar.

Das Zusammenspiel dieser Zuverlässigkeitsmechanismen erschwert zusätzlich das Verhalten von Bauelementen und die Stabilität von Schaltungen. Degradationsprozesse in einer Domäne können Ausfälle in einer anderen verstärken; beispielsweise können FEOL-Transistoralterung und BEOL-Interconnect-Degradation gemeinsam die Zeitreserven und Störfestigkeit auf Schaltungsebene beeinflussen. Dadurch hat sich Zuverlässigkeit zu einer domänenübergreifenden Herausforderung entwickelt, die Bauelement-, Schaltungs-, Layout- und Systemebenen umfasst und koordinierte Gegenmaßnahmen über alle Abstraktionsebenen hinweg erfordert. Die Bewältigung dieser Herausforderungen erfordert umfassende Modellierungs- und Simulationsansätze, die physikalische Degradationsmechanismen mit der Leistungsfähigkeit auf Schaltungs- und Systemebene verknüpfen.

Diese Arbeit konzentriert sich auf drei zentrale Zuverlässigkeitsaspekte: Strahlungseffekte, NBTI und EM. Das erste Ziel besteht darin, präzise Methoden zur Zuverlässigkeitsbewertung sowie zur Verbesserung der Schaltungslebensdauer unter Berücksichtigung dieser Degradationsmechanismen zu entwickeln. Hierzu wird ein Bauelement-zu-Schaltung-Co-Simulationsansatz vorgeschlagen, der mit Multi-Scale-Co-Optimierungsstrategien kombiniert wird, um die Wechselwirkungen zwischen Bauelement- und Schaltungsebene abzubilden.

Werkzeuge der Electronic Design Automation (EDA), die von bauelementbasierten Technology-Computer-Aided-Design-(TCAD)-Simulationen bis hin zu schaltungsbasierten SPICE-Modellen reichen, spielen eine zentrale Rolle bei der Verknüpfung von Halbleiterphysik und Schaltungsdesign und ermöglichen eine multiskalige Zuverlässigkeitsanalyse. Mit fortschreitenden Technologieknoten werden EDA-Workflows jedoch zunehmend komplex und rechenintensiv, was die Effizienz der Zuverlässigkeitsbewertung einschränkt. Im Kontext von Design–Technology Co-Optimization (DTCO) und System–Technology Co-Optimization (STCO) besteht daher ein wachsender Bedarf an effizienten und automatisierten Multi-Scale-Modellierungsansätzen.

Das zweite Ziel dieser Arbeit besteht darin, Methoden des maschinellen Lernens, einschließlich künstlicher neuronaler Netze (ANNs) und großer Sprachmodelle (LLMs), einzusetzen, um die Zuverlässigkeitsbewertung und Simulation zu beschleunigen und somit die Effizienz zuverlässigkeitsorientierter Entwurfsverfahren zu verbessern.

Contents

Abstract	iii
Zusammenfassung	v
List of own publications included in this thesis	xi
I. Preliminaries	1
1. Introduction	3
1.1. Contributions and Research Directions	4
1.1.1. Soft Error Analysis in Advanced Logic Structures	4
1.1.2. Aging and Radiation Reliability Analysis in Memory and Storage Structures . .	4
1.1.3. AI-Assisted EDA Automation	5
1.2. Dissertation Outline	5
2. Background	7
2.1. Technology Scaling Challenges	7
2.2. Pass-Transistor Logic (PTL)	8
2.2.1. Power-Performance Challenges in Conventional CMOS Logic	8
2.2.2. PTL Fundamentals and PPA Advantages	9
2.2.3. Reliability Challenges in PTL	10
2.3. Static Random Access Memory (SRAM) and Memory Structures	10
2.3.1. SRAM Architecture	10
2.3.2. Register Files and Latch Structures	11
2.3.3. Memory Stability and Scaling Challenges	11
2.4. Reliability Challenges in Advanced Technologies	12
2.4.1. Radiation-Induced Soft Errors	12
2.4.2. Bias Temperature Instability	13
2.4.3. Physical Concept of NBTI	13
2.4.4. NBTI Modeling	13
2.4.5. Electromigration	14
2.5. AI-Assisted Chip Design and Testing	15
2.5.1. AI-Assisted EDA Automation	15
2.5.2. LLM-Based Design and Testing Automation	15
II. Contributions	17
3. Machine Learning-Based Soft Error Rate Estimation for Pass-Transistor Logic	19
3.1. Challenges of SET Modeling in PTL	19
3.1.1. Irregular SET Propagation Behavior	20
3.1.2. Related Work	20

3.2.	Proposed Machine-Learning-Based SET Modeling Framework	21
3.2.1.	Overall Modeling Flow	21
3.2.2.	SET Propagation Model Based on Neural Network	23
3.2.3.	SET Classification Based on SVC	24
3.2.4.	Simulation Setup and Data Preparation for Training	24
3.2.5.	Evaluation of Regression and Classification Models	25
3.3.	Results and Discussion	25
3.3.1.	Simulation Setup and Test Scenario	25
3.3.2.	Accuracy Comparison	25
3.3.3.	Runtime Comparison	27
3.4.	Conclusion	28
4.	A Learning-Based Approach for Single Event Transient Analysis in Pass Transistor Logic . . .	29
4.1.	Preliminaries	29
4.1.1.	Challenges of SET Propagation in PTL	29
4.1.2.	Related Work	30
4.2.	Structural SET Propagation Modeling Framework	31
4.2.1.	Data Collection and Ground Truth Generation	33
4.2.2.	Modeling Regular SETs	33
4.2.3.	Modeling Irregular SETs	34
4.2.4.	Evaluation of Models	37
4.3.	Results and Discussion	39
4.3.1.	Simulation Setup for Validation	39
4.3.2.	Accuracy Comparison and Discussion	39
4.4.	Conclusion	40
5.	Impact of Radiation and Aging on DVFS in 22nm FDSOI Latch	41
5.1.	Preliminaries	42
5.1.1.	NBTI Models	42
5.1.2.	Heavy Ion Current Models	42
5.2.	Methodology	43
5.2.1.	TCAD Calibration	44
5.2.2.	Modelling NBTI Effect	44
5.2.3.	Modelling Radiation Current	45
5.2.4.	Setup Time Simulation	46
5.3.	Model Implementation and Analysis	46
5.3.1.	Simulation Setup	47
5.3.2.	Fresh Latch Analysis Excluding Radiation Impact	47
5.3.3.	Fresh Latch Analysis Including Radiation Impact	47
5.3.4.	Aged Latch Analysis Excluding Radiation Impact	48
5.3.5.	Aged Latch Analysis Including Radiation Impact	50
5.4.	Conclusion	51
6.	Addressing the Combined Effect of Transistor and Interconnect Aging in SRAM towards Silicon Lifecycle Management	53
6.1.	Related work	54
6.2.	Modeling Methodology	54
6.2.1.	TCAD Model Calibration	54
6.2.2.	Building the NBTI Model	55

6.2.3.	NBTI Fault Injection	55
6.2.4.	EM Fault Injection	55
6.2.5.	Model Exploitation	55
6.3.	Results and Discussion	56
6.3.1.	Simulation Setup	56
6.3.2.	Individual Effect of NBTI and EM on Writability	56
6.3.3.	Combined Effect of NBTI and EM on Writability	57
6.4.	Conclusion	58
7.	Testing for Aging in Advanced SRAM: From Front-End-of-The-Line Transistors to Back-End-of-The-Line Interconnects	61
7.1.	Related Work	61
7.2.	Aging Modeling	62
7.2.1.	TCAD Model Calibration	62
7.2.2.	Building the NBTI and EM Model	63
7.3.	Aging Testing	63
7.3.1.	NBTI Fault Injection	64
7.3.2.	EM Fault Injection	64
7.4.	Results and Discussion	65
7.4.1.	Critical Write Operation	65
7.4.2.	T_{switch} at the Nominal Operation Corner	65
7.4.3.	T_{switch} at the Test Operation Corner	66
7.4.4.	Earlier Detection of Write Failures	67
7.5.	Conclusion	68
8.	Effect of Transistor and Interconnect Scaling on Aging in Advanced SRAM	69
8.1.	Cross-Technology Aging Analysis and Mitigation Methodology	70
8.1.1.	NBTI and EM Fault Injection	70
8.1.2.	Sensitization and Voltage-Scaling Test Strategies	71
8.1.3.	Aging-Aware DVFS Strategy	71
8.2.	Results and Discussion	72
8.2.1.	T_{switch} at the Nominal Operation Corner	72
8.2.2.	Impact of Process Variation on T_{switch}	74
8.2.3.	Sensitization of Critical Write Operation with Test Vector	75
8.2.4.	Sensitization of Critical Write Operation with Test Corner	76
8.2.5.	Earlier Detection of Aging-Induced Write Failures	77
8.2.6.	Aging-Aware DVFS for Device Lifetime Extension	78
8.3.	Conclusion	81
9.	Analysis and Mitigation of Radiation Effects in SRAM-based Register Files	83
9.1.	Related Work	83
9.2.	Radiation Modeling and Testing	84
9.2.1.	TCAD Model Calibration	84
9.2.2.	Modeling Radiation-Induced Current	84
9.2.3.	Identifying Impactful Factors	84
9.2.4.	Calculating Error Probability	85
9.3.	Results and Discussion	85
9.3.1.	Simulation Setup	86
9.3.2.	Impact of RC Load	86

9.3.3. Impact of Read Word-Line Activation Pulse-Width	86
9.3.4. Impact of Read Operation	86
9.3.5. Impact of Cell Location	86
9.3.6. Impact of Radiation Energy	87
9.4. Technology-Dependent Mitigation Strategies	87
9.5. Conclusion	88
10. ChatTCAD: Leveraging Large Language Models for Automated TCAD Simulation File Generation	89
10.1. Preliminaries	90
10.1.1. Large Language Model in EDA	90
10.1.2. Machine Learning for TCAD Simulation	90
10.2. Methodology	90
10.2.1. Function Block Decomposition	91
10.2.2. Python Function Wrapper	91
10.2.3. Prompt Engineering for Single-Function Mode	92
10.2.4. Prompt Engineering for Complete-Device Mode	93
10.3. Results and Discussion	96
10.3.1. Evaluation of Dataset Generated for SFM	97
10.3.2. Similarity Evaluation	97
10.4. Conclusion	98
11. Conclusion and Perspectives	101
11.1. Conclusion	101
11.2. Future Perspectives	101
 III. Appendix	 103
Bibliography	105
List of Figures	113
List of Tables	117
Acronyms	119

List of own publications included in this thesis

- [1] Z. Zhang, J. Lappas, A. Chinazzo, C. Weis, Z. Wu, L. Ni, N. Wehn, and M. Tahoori, “Machine learning based soft error rate estimation of pass transistor logic in high-speed communication”, in *2022 IEEE European Test Symposium (ETS)*, IEEE, 2022, pp. 1–4.
- [2] Z. Zhang, Z. Wu, C. Weis, N. Wehn, and M. Tahoori, “A learning-based approach for single event transient analysis in pass transistor logic”, in *2023 IEEE 29th International Symposium on On-Line Testing and Robust System Design (IOLTS)*, IEEE, 2023, pp. 1–7.
- [3] Z. Zhang, C. Weis, N. Wehn, M. Tahoori, and S. Nassif, “Do radiation and aging impact dvfs? tcad-based analysis on 22 nm fdsoi latches”, in *2024 IEEE 30th International Symposium on On-Line Testing and Robust System Design (IOLTS)*, IEEE, 2024, pp. 1–6.
- [4] Z. Zhang, M. Mayahinia, C. Weis, N. Wehn, M. Tahoori, S. Nassif, G. Tshagharyan, G. Harutyunyan, and Y. Zorian, “Addressing the combined effect of transistor and interconnect aging in sram towards silicon lifecycle management”, in *2024 IEEE 42nd VLSI Test Symposium (VTS)*, 2024, pp. 1–5.
- [5] Z. Zhang, M. Mayahinia, C. Weis, N. Wehn, M. Tahoori, S. Nassif, G. Tshagharyan, G. Harutyunyan, and Y. Zorian, “Testing for aging in advanced sram: From front end of the line transistors to back end of the line interconnects”, in *2024 IEEE International Test Conference (ITC)*, IEEE, 2024, pp. 46–50.
- [6] Z. Zhang, M. Mayahinia, C. Weis, N. Wehn, M. Tahoori, S. Nassif, G. Tshagharyan, G. Harutyunyan, and Y. Zorian, “Effect of transistor and interconnect scaling on aging in advanced sram”, *IEEE Transactions on Device and Materials Reliability*, 2026.
- [7] Z. Zhang, S. Hemaram, M. Mayahinia, C. Weis, N. Wehn, M. Tahoori, S. Nassif, G. Tshagharyan, G. Harutyunyan, and Y. Zorian, “Analysis and mitigation of radiation effects in sram-based register files”, in *2025 IEEE European Test Symposium (ETS)*, IEEE, 2025, pp. 1–4.
- [8] Z. Zhang, S. Nassif, and M. Tahoori, “Chattcad: Leveraging large language models for automated tcad simulation file generation”, in *2025 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, IEEE, vol. 1, 2025, pp. 1–6.

Part I.

Preliminaries

1. Introduction

Continuous scaling of semiconductor technologies has enabled remarkable improvements in computational performance, energy efficiency, and integration density over the past decades. Emerging transistor architectures such as FDSOI and FinFET technologies have significantly improved switching characteristics and power efficiency of modern integrated circuits. However, aggressive scaling of both FEOL devices and BEOL interconnects has simultaneously introduced substantial reliability challenges that increasingly limit the achievable performance and lifetime of modern semiconductor systems.

In deeply scaled technology nodes, reliability concerns are no longer restricted to isolated degradation phenomena. Instead, FEOL transistor aging, BEOL interconnect degradation, process variation, voltage scaling, and radiation-induced transient effects strongly interact across multiple abstraction levels, including device, circuit, layout, and architecture. Such interactions considerably complicate the reliability behavior of modern integrated circuits. For instance, transistor aging mechanisms such as NBTI gradually increase transistor threshold voltage and degrade transistor drive strength, while EM in scaled BEOL interconnects increases parasitic resistance. At the circuit level, these degradation mechanisms collectively increase signal delay and reduce timing margins. Consequently, long-term reliability prediction and Silicon Lifecycle Management (SLM) strategies become increasingly challenging. At the same time, reduced supply voltage and shrinking node capacitance increase the sensitivity of modern circuits to radiation-induced soft errors.

These reliability challenges strongly affect modern memory and logic structures widely used in deeply scaled computing systems. Memory structures such as Static Random Access Memory (SRAM) and register files, as well as sequential logic structures including latches, are extensively used in processors, graphics processing units (GPUs), and System-on-Chip (SoC) designs. As transistor dimensions and interconnect pitches continue to shrink, these structures become increasingly sensitive to FEOL and BEOL aging mechanisms, process variation, and radiation-induced transient effects. Reduced transistor drive strength, increased parasitic resistance, and tighter voltage margins significantly degrade timing behavior, operational lifetime, and overall reliability.

Beyond conventional CMOS logic styles, low-power circuit techniques further exacerbate reliability challenges in scaled technologies. Pass-Transistor Logic (PTL) structures achieve improved energy efficiency and reduced transistor count, leading to degraded voltage swing and weaker noise margins, making them increasingly susceptible to transient radiation effects and process variation.

Existing reliability studies mainly focus on individual degradation mechanisms or isolated abstraction levels. However, degradation at one abstraction level can significantly exacerbate failures at another level. For example, NBTI-induced threshold voltage shifts may worsen timing sensitivity to transient radiation effects, while EM-induced interconnect resistance increase can further aggravate SRAM delay degradation and write failures. Similarly, aggressive voltage scaling can improve energy efficiency while simultaneously reducing noise margins. These interactions demonstrate that modern reliability challenges are fundamentally multi-scale in nature. Therefore, comprehensive device-to-circuit reliability analysis and co-optimization methodologies are required to properly address reliability challenges in state-of-the-art semiconductor technologies without introducing unacceptable overheads in PPA.

Besides the reliability challenges themselves, accurate reliability evaluation in scaled semiconductor technologies has become increasingly computationally expensive. Reliability analysis often requires TCAD simulations, SPICE-level circuit analysis, and large-scale statistical methodologies such as Monte

Carlo (MC) simulation, resulting in substantial simulation overhead and modeling complexity. Therefore, efficient reliability-oriented EDA methodologies are becoming increasingly important.

Recently, Artificial Intelligence (AI)-assisted methodologies, particularly Machine Learning (ML)- and LLM-based approaches, have demonstrated strong potential for accelerating reliability analysis and improving the efficiency of EDA-related workflows. AI-assisted methodologies can significantly reduce simulation overhead and improve the automation of reliability-oriented semiconductor design flows. In particular, LLM-assisted automation can improve the productivity of TCAD and EDA workflows by automating simulation setup generation, parameter optimization, and reliability-oriented scripting tasks.

Motivated by these challenges, this thesis aims to establish multi-scale reliability analysis and aging-aware evaluation methodologies for advanced semiconductor memory and logic structures by jointly considering FEOL transistor degradation, BEOL interconnect aging, and radiation-induced transient effects under aggressive technology scaling. Furthermore, this work explores ML- and LLM-assisted approaches for improving the efficiency of reliability-oriented simulation and EDA automation.

1.1. Contributions and Research Directions

This thesis focuses on both transient reliability challenges and long-term degradation mechanisms in advanced semiconductor technologies. Transient reliability issues mainly originate from radiation-induced soft errors and signal integrity degradation in low-voltage logic structures. Long-term reliability concerns include transistor aging and interconnect degradation mechanisms such as NBTI and EM. These degradation mechanisms strongly affect the functionality, timing behavior, stability, and lifetime of modern memory and logic circuits.

In the scope of this thesis, transient fault propagation in nanoscale logic structures is first investigated under technology scaling. Secondly, the impact of aging and radiation effects on advanced memory and storage structures is analyzed by considering both FEOL and BEOL reliability challenges. Finally, AI-assisted methodologies are explored for improving the efficiency and automation of reliability-oriented EDA workflows.

1.1.1. Soft Error Analysis in Advanced Logic Structures

In advanced technology nodes, reduced supply voltage and shrinking node capacitance significantly increase the sensitivity of logic circuits to radiation-induced transient effects. In this thesis, the reliability of PTL structures under radiation-induced Single Event Transients (SETs) is investigated. Compared with conventional CMOS logic, PTL structures exhibit reduced voltage swing and degraded noise margins, making them more susceptible to transient pulse propagation and functional errors.

To accurately analyze the reliability of PTL circuits, SET propagation behavior is investigated under different logic styles and operating conditions. Furthermore, ML-assisted methodologies are developed for efficient prediction and estimation of SET behavior and accelerate the estimation of soft error rate (SER). The proposed approaches significantly reduce simulation complexity while maintaining acceptable accuracy compared with conventional MC-based simulations.

1.1.2. Aging and Radiation Reliability Analysis in Memory and Storage Structures

In scaled semiconductor technologies, the reliability of SRAM arrays, register files, and latch structures is increasingly affected by both FEOL transistor degradation and BEOL interconnect aging. In particular, transistor aging mechanisms such as NBTI degrade SRAM writability and read stability, while EM-induced interconnect degradation weakens voltage delivery and further aggravates memory reliability degradation in scaled BEOL structures.

Furthermore, this thesis explores aging-aware testing and sensitization methodologies for earlier detection of aging-induced SRAM failures. Detecting degradation before catastrophic in-field failures occur is essential for enabling effective SLM strategies and remaining lifetime estimation.

Additionally, the interaction between FEOL transistor scaling, BEOL interconnect scaling, and operating conditions is investigated to provide reliability-aware insights for SRAM design, testing, and mitigation under aggressive technology scaling.

In addition to aging-related reliability challenges, reduced node capacitance and aggressive voltage scaling considerably increase the susceptibility of memory and storage structures to radiation-induced soft errors. Therefore, this thesis also investigates radiation-induced transient faults in latches, SRAM arrays, and register files, together with technology-dependent mitigation methodologies for improving soft-error resilience.

1.1.3. AI-Assisted EDA Automation

Accurate reliability evaluation of modern semiconductor technologies requires computationally expensive TCAD simulations and SPICE-level circuit analysis. Therefore, improving the efficiency and automation of reliability-oriented EDA workflows is another research direction of this thesis.

In this work, LLM-assisted methodologies are explored for automating TCAD-related workflows and EDA scripting tasks. Modern semiconductor reliability analysis often requires extensive manual generation and optimization of simulation scripts, process configurations, and parameter setups. Therefore, AI-assisted automation frameworks can substantially improve productivity and reduce manual effort in reliability-oriented semiconductor design flows.

1.2. Dissertation Outline

This thesis is organized into the chapters as follows:

- Chapter 2 reviews the essential background required for this thesis, including technology scaling, PTL circuits, SRAM and storage structures, radiation-induced soft errors, NBTI, EM, and AI-assisted EDA methodologies.
- Chapter 3 proposes a machine-learning-based framework for fast gate-level SER estimation in PTL circuits by modeling SET propagation behavior.
- Chapter 4 presents a machine-learning-based structural SET modeling approach for PTL circuits, enabling reusable transistor-structure-level transient propagation analysis.
- Chapter 5 investigates the combined impact of radiation-induced transient effects and NBTI aging on latch timing behavior in 22 nm FDSOI technology.
- Chapter 6 analyzes the combined effects of FEOL transistor aging and BEOL interconnect degradation on SRAM writability in 22 nm FDSOI technology.
- Chapter 7 extends the combined FEOL/BEOL aging analysis to 12 nm FinFET SRAM and proposes an aging-aware testing methodology for earlier failure prediction and remaining lifetime estimation.
- Chapter 8 presents a unified analysis of transistor and interconnect scaling effects on SRAM aging behavior across FDSOI and FinFET technologies. Furthermore, aging-aware sensitization, voltage-scaling-based earlier detection, and Dynamic Voltage and Frequency Scaling (DVFS)-based mitigation methodologies are proposed for improving SRAM lifetime and reliability.

- Chapter 9 analyzes radiation-induced soft errors in SRAM-based register files and evaluates technology-dependent mitigation strategies.
- Chapter 10 presents ChatTCAD, an LLM-assisted framework for automated TCAD simulation file generation and reliability-oriented EDA workflow automation.
- Chapter 11 concludes this thesis with a summary of the main findings and possible future research directions.

2. Background

2.1. Technology Scaling Challenges

For decades, continuous transistor scaling driven by Moore's law [9] has enabled remarkable improvements in computational performance, energy efficiency, and integration density. However, as feature sizes shrink into deeply scaled nanometer regimes, conventional geometric scaling faces increasing challenges due to fundamental physical limitations, rising manufacturing complexity, and escalating fabrication cost [10]. Consequently, the semiconductor industry has gradually shifted from purely dimensional scaling toward advanced device architectures, evolving from planar CMOS to FinFET, Gate-All-Around (GAA), and complementary FET (CFET) structures [11], [12], heterogeneous integration technologies such as 2.5D/3D integration and chiplet-based architectures [13], [14], as well as novel material innovations including two-dimensional materials such as MoS_2 to sustain performance improvement [15] as shown in Fig. 2.1.

Classical Dennard scaling theory states that the circuit power density remains approximately constant when transistor dimensions, supply voltage (V_{DD}), and current scale in proportion [17]. In modern advanced technology nodes, however, this scaling behavior has significantly deviated from the original scaling assumptions. In particular, V_{DD} cannot be reduced proportionally with transistor dimensions because threshold voltage (V_{th}) scaling is fundamentally constrained by excessive subthreshold leakage current and static power consumption. At the same time, shrinking interconnect dimensions substantially increases parasitic resistance and capacitance, thereby aggravating interconnect-related Resistive-capacitive (RC) delay [18].

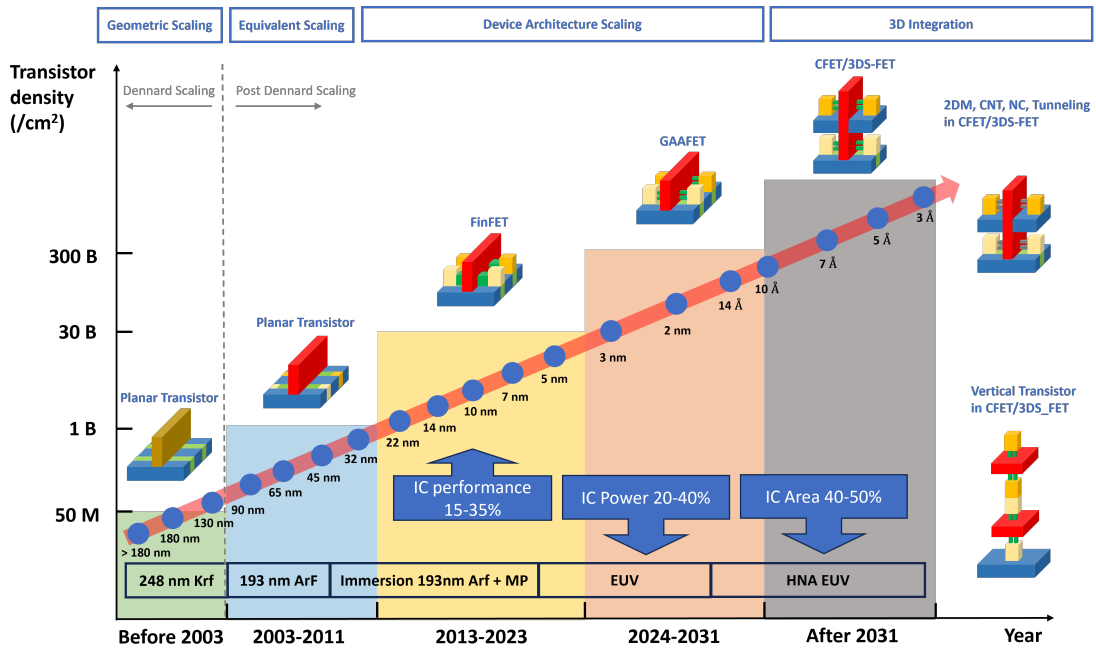


Figure 2.1.: Development history of and trends in MOSFETs with differing structures and scaling stages [16].

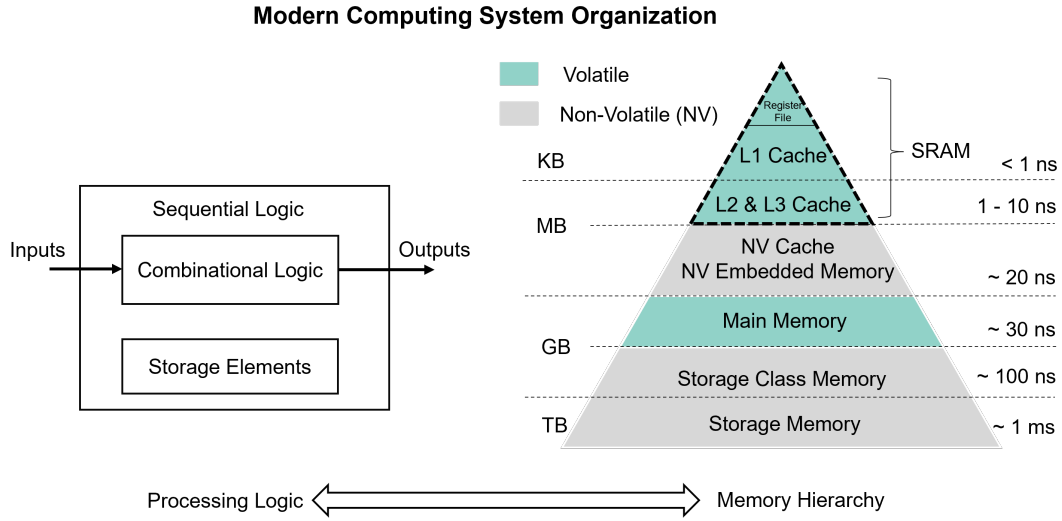


Figure 2.2.: Simplified organization of modern computing systems, including processing logic and memory hierarchy.

Consequently, the primary performance bottleneck of modern integrated circuits has gradually shifted from intrinsic transistor switching delay toward BEOL interconnect limitations. In advanced BEOL technologies, reduced interconnect cross-sectional area and increasing parasitic effects significantly degrade signal propagation delay and timing behavior. Furthermore, aggressive interconnect scaling considerably increases current density, thereby aggravating interconnect reliability concerns.

Beyond performance limitations, aggressive technology scaling also introduces substantial reliability challenges in both FEOL devices and BEOL interconnects. Process variation increasingly affects transistor and interconnect characteristics as device dimensions continue to shrink [19]. Furthermore, reduced supply voltage, shrinking node capacitance, and increasing current density considerably aggravate transistor aging, radiation-induced soft errors, and interconnect degradation mechanisms in advanced integrated circuits [20], [21], [22]. Consequently, reliability has become one of the major constraints in modern Very Large Scale Integration (VLSI) design despite the significant improvements in performance, power efficiency, and integration density enabled by advanced semiconductor technologies.

The impact of aggressive technology scaling extends beyond individual transistors and interconnects, considerably affecting the organization and reliability of modern computing systems. Modern computing architectures consist of both processing logic and hierarchical memory structures, each of which faces increasing challenges in performance, power efficiency, and reliability under advanced technology scaling. Fig. 2.2 illustrates the simplified organization of modern computing systems and the relationship between processing logic and memory hierarchy.

2.2. Pass-Transistor Logic (PTL)

2.2.1. Power-Performance Challenges in Conventional CMOS Logic

Despite the remarkable success of CMOS technology, conventional static CMOS logic increasingly faces severe power-performance challenges in deeply scaled integrated circuits. In large-scale digital systems, the total power consumption of CMOS circuits consists of both dynamic and static components and can be expressed as

$$P_{\text{total}} = P_{\text{dynamic}} + P_{\text{static}} = \alpha C_{\text{load}} V_{DD}^2 f + I_{\text{leak}} V_{DD}, \quad (2.1)$$

where α denotes the switching activity factor, C_{load} is the effective load capacitance, f is the operating frequency, and I_{leak} represents the total leakage current [23]. As transistor density and operating frequency

continue to increase, both dynamic and leakage power consumption have become major design constraints in advanced CMOS technologies.

In conventional CMOS logic, complex logic functions are typically implemented using complementary pull-up and pull-down transistor networks. Although such structures provide full voltage swing and strong noise immunity, they often require relatively large transistor counts for complex logic functions such as XOR/XNOR gates and multiplexers. Consequently, the associated switching capacitance, routing overhead, and dynamic power consumption increase significantly in highly integrated circuits.

Furthermore, aggressive technology scaling and continuous supply voltage reduction introduce additional challenges for conventional CMOS logic implementation. Although lowering the supply voltage and reducing switching capacitance are effective approaches for improving power efficiency, these optimizations considerably reduce noise margins and increase the sensitivity of integrated circuits to process variation, signal integrity degradation, and transient disturbances [24]. At the same time, increasing interconnect parasitics and transistor density further exacerbate timing degradation and power consumption challenges in large-scale integrated systems [25].

To improve PPA efficiency, various alternative low-power CMOS logic styles have been extensively investigated. Among these approaches, PTL has attracted considerable attention due to its capability to reduce transistor count, switching capacitance, and power consumption while maintaining relatively high-speed operation [26]. However, the same scaling-induced optimizations that improve PPA efficiency also make PTL structures more vulnerable to signal integrity and reliability challenges due to incomplete voltage swing, degraded signal restoration capability, and weaker noise margins [27].

2.2.2. PTL Fundamentals and PPA Advantages

PTL represents an alternative low-power logic style in which MOS transistors operate primarily as signal-passing switches rather than complementary pull-up and pull-down networks connected to the supply rails. Unlike conventional static CMOS logic, PTL propagates input signals directly through pass transistors to realize Boolean operations, enabling complex logic functions to be implemented with fewer transistors. The reduced transistor count lowers switching capacitance, routing overhead, and layout area, thereby improving PPA efficiency in highly integrated circuits. Since dynamic power consumption is directly proportional to switched capacitance, as shown in Equation 2.1, PTL structures can substantially improve power efficiency compared with conventional CMOS implementations.

A fundamental characteristic of PTL structures is the asymmetric signal transmission behavior of nMOS and pMOS transistors. An NMOS pass transistor efficiently propagates a strong logic '0' but suffers from threshold voltage loss when passing a logic '1', resulting in an output voltage approximately limited to $V_{DD} - V_{thn}$, where V_{thn} is the threshold voltage of the NMOS. Conversely, a PMOS transistor passes a strong logic '1' but a degraded logic '0'. To address this signal degradation problem, transmission-gate-based structures employ parallel nMOS and PMOS devices to achieve full rail-to-rail voltage swing at the cost of increased transistor count and routing complexity [28]. Other PTL styles employ alternative restoration mechanisms, such as level-restoring buffers or cross-coupled feedback structures, to improve signal integrity while preserving the transistor-count advantage of PTL.

PTL is particularly efficient for arithmetic and data-path-intensive logic functions such as XOR/XNOR gates, multiplexers, adders, and barrel shifters, where conventional CMOS implementations often require relatively large transistor counts. Various PTL styles have therefore been proposed to balance power efficiency and signal integrity. Representative approaches include Complementary Pass-Transistor Logic (CPL), which employs dual-rail nMOS pass networks together with output restoration circuitry to improve voltage swing and noise immunity [27]. Other PTL styles, including Dual-Value Logic (DVL) and Double Pass-Transistor Logic (DPL), introduce additional restoration mechanisms to mitigate threshold-voltage-induced signal degradation and improve output voltage swing [26]. Examples of PTL-based XOR gate implementations are shown in Fig. 2.3.

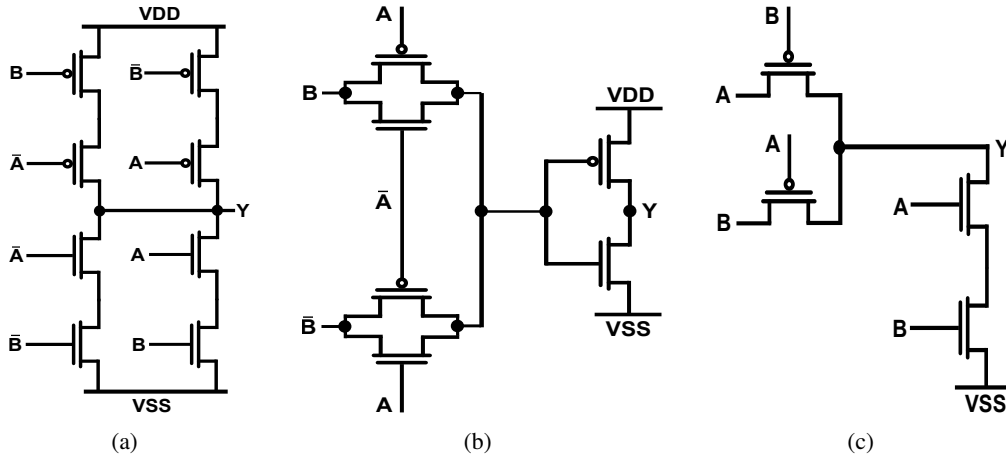


Figure 2.3.: Examples of DPL-based (b) and inverter-buffered PTL XOR circuits (c) adapted from [29], [30]. [29] reported that DPL-based XOR designs achieved up to 24% area reduction and 35% lower power consumption in synthesized multiplier circuits under low-voltage operation compared with complementary CMOS XOR implementations (a), whereas inverter-buffered single-pass-transistor structures (c) suffered from degraded voltage swing and failed to satisfy timing constraints at low supply voltages due to weak drivability.

2.2.3. Reliability Challenges in PTL

Despite its attractive PPA advantages, PTL introduces several reliability challenges in deeply scaled technologies. Threshold voltage-induced signal degradation and incomplete signal restoration reduce voltage swing, drive strength, and noise margins compared with conventional CMOS logic. These effects become more severe under aggressive voltage scaling and nanoscale process variation, where reduced supply voltage and threshold voltage fluctuations considerably affect signal propagation and timing behavior. Furthermore, the degraded restoration capability and reduced critical charge of PTL structures lead to different SET generation and propagation characteristics compared with static CMOS logic [31]. Consequently, ensuring reliable PTL operation under process variation and transient fault conditions remains a major challenge in advanced PTL-based circuit design.

2.3. Static Random Access Memory (SRAM) and Memory Structures

2.3.1. SRAM Architecture

SRAM is one of the most widely used on-chip memory technologies in modern integrated circuits due to its high operating speed and compatibility with standard CMOS fabrication. SRAM is extensively employed in cache hierarchies and register files in processors, GPUs, and SoC designs.

The conventional six-transistor (6T) SRAM cell in Fig. 2.4(a) consists of two cross-coupled inverters forming a bistable latch together with two access transistors connected to the word-line (WL) and complementary bit-lines (BL/BLB). The cross-coupled inverter pair enables the cell to retain stored data indefinitely as long as the supply voltage is maintained.

A complete SRAM array, illustrated in Fig. 2.4(b), requires not only the bit-cell itself but also multiple peripheral circuits, including row decoders, sense amplifiers, write drivers, precharge circuits, column multiplexers, and timing control units, all of which must be carefully co-designed to ensure reliable memory operation [32].

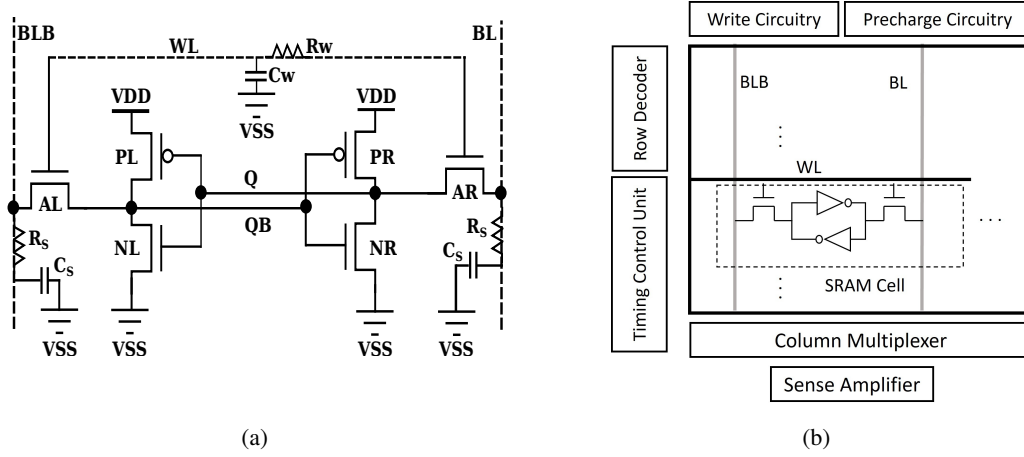


Figure 2.4.: (a) Conventional six-transistor (6T) SRAM cell with associated parasitic resistance and capacitance on WL and BL interconnects, (b) Simplified SRAM array architecture with the required peripheral circuitry.

2.3.2. Register Files and Latch Structures

Beyond conventional SRAM arrays, modern computing systems also rely on other state-holding elements including register files, latches, and flip-flops. Register files are small, high-speed storage arrays implemented with SRAM-like bit-cell structures and are widely used in processors and GPUs [33], [34]. Unlike large cache SRAMs that primarily target storage density, register files are designed for low-latency multi-ported access, allowing multiple execution units to perform simultaneous read and write operations. However, the additional transistors, bit-lines, and routing required for multi-port operation considerably increase power consumption, routing complexity, and timing closure challenges.

Latches and flip-flops, on the other hand, are the fundamental sequential building blocks of digital systems. They are used extensively for timing synchronization and pipeline stage boundaries. Because they continuously maintain internal circuit states across clock cycles, their reliability directly affects the functional correctness and timing behavior of the entire chip [35]. Hence, these characteristics are important in high-performance and safety-critical applications such as automotive microcontrollers.

2.3.3. Memory Stability and Scaling Challenges

Aggressive technology scaling has significantly improved SRAM density and operating speed. However, it has also introduced substantial stability and reliability challenges. Reduced noise margins and larger threshold voltage variations directly degrade SRAM stability during hold, read, and write operations [36]. In particular, lower supply voltage and degraded transistor drive strength considerably reduce the SRAM write margin, making reliable write operation increasingly difficult in advanced technologies.

BEOL interconnect scaling introduces additional challenges for dense SRAM arrays. Decreased metal pitches increase parasitic resistance and current density, which not only worsen timing but also increase EM-induced degradation over long-term operation. Furthermore, resistive voltage drop along scaled bit-line interconnects weakens the voltage delivered from the write drivers to the memory cell, thereby further degrading SRAM writability [37].

At the same time, shrinking node capacitance and reduced supply voltage lower the critical charge (Q_{crit}) required to flip a stored bit, making SRAM cells more susceptible to radiation-induced soft errors [38]. Similar challenges also affect register files, latches, and flip-flops. In register files, simultaneous access operations and low-voltage operation can increase transient fault sensitivity, while aging-induced parameter shifts in latches and flip-flops directly affect timing behavior and functional correctness.

Therefore, ensuring reliable operation of modern memory and storage structures requires reliability-aware analysis across both FEOL devices and BEOL interconnects.

2.4. Reliability Challenges in Advanced Technologies

As technology scaling continues, reduced dimensions, lower supply voltage, and higher operating frequency considerably aggravate multiple reliability mechanisms in modern integrated circuits. Among these mechanisms, Bias Temperature Instability (BTI) mainly affects FEOL transistor behavior, while EM primarily impacts BEOL interconnect reliability. Furthermore, these degradation mechanisms are strongly coupled with each other. For example, BTI-induced threshold voltage shifts can alter current density and timing behavior, thereby accelerating EM degradation and increasing circuit sensitivity to transient faults. Therefore, understanding these interactions is essential for accurate reliability analysis in advanced semiconductor technologies.

2.4.1. Radiation-Induced Soft Errors

Due to continued technology scaling, modern integrated circuits have become increasingly sensitive to radiation-induced faults. Reduced node capacitance and lower supply voltage decrease the critical charge (Q_{crit}) required to disturb a circuit node, while the reduced spacing between adjacent devices aggravates charge-sharing effects [39]. Consequently, particle strikes that were previously non-critical can now generate transient disturbances capable of corrupting circuit operation.

When an energetic particle penetrates silicon, it deposits energy through ionization and generates electron-hole pairs along its trajectory. The generated carriers are subsequently collected by nearby p-n junctions through drift and diffusion mechanisms, producing transient current pulses inside the device. Depending on the affected circuit structure, these transient effects may appear as SETs in combinational logic or single-event upsets (SEUs) in memory and sequential storage elements [40].

The shape and magnitude of radiation-induced transient currents strongly depend on device-related factors, including doping profile, electric-field distribution, and transistor architecture. Consequently, radiation behavior becomes increasingly technology-dependent in advanced nodes such as FDSOI and FinFET technologies [41]. For efficient circuit-level soft-error analysis, compact current-source models such as the double-exponential current model are widely used to approximate radiation-induced transient currents [42], [43]. However, such simplified models cannot fully capture the technology-dependent charge-collection behavior of advanced semiconductor devices.

To accurately model radiation-induced transient behavior, TCAD-based simulation approaches are widely employed [44]. TCAD enables physics-based analysis of particle interaction, carrier transport, and transient charge collection under different technology conditions. The extracted transient current waveforms can subsequently be injected into SPICE-level circuit simulations to evaluate the soft-error sensitivity of complex logic and memory structures. Consequently, TCAD-to-circuit simulation methodologies have become increasingly important for reliability analysis in advanced semiconductor technologies.

At the circuit level, accurate SER estimation additionally requires modeling the propagation behavior of SETs through combinational logic. Although MC and SPICE-based simulations can accurately capture SET propagation and masking effects, their computational complexity becomes prohibitive for large-scale circuit analysis and extensive design-space exploration. Consequently, simplified SET propagation models are widely employed for efficient SER estimation.

Among these approaches, trapezoidal pulse models are commonly used to represent SET waveforms using compact parameters such as pulse width and pulse amplitude [45]. Based on these simplified pulse representations, electrical masking effects during transient propagation can be efficiently evaluated at both

the logic and circuit levels. However, such compact pulse abstractions become increasingly inaccurate for advanced circuit structures exhibiting irregular transient behavior and nonlinear propagation effects.

2.4.2. Bias Temperature Instability

BTI is a time-dependent transistor degradation mechanism that mainly affects PMOS devices under NBTI, and in advanced technologies also impacts NMOS devices under Positive Bias Temperature Instability (PBTI) [46]. Under long-term electrical and thermal stress, the V_{th} gradually increases, resulting in reduced drive current and degraded switching performance.

BTI strongly depends on operating voltage, temperature, and signal activity, making the degradation highly workload-dependent. In logic circuits, BTI-induced degradation can introduce timing violations due to asymmetric degradation of pull-up and pull-down networks. In SRAM cells, threshold voltage shifts between the cross-coupled inverters degrade the static noise margin and can eventually lead to stability failures during read and write operations [47]. Furthermore, the stochastic nature of BTI degradation introduces additional device-to-device variability, complicating long-term reliability prediction.

2.4.3. Physical Concept of NBTI

From a physical perspective, NBTI degradation originates from interface trap generation and hydrogen-related reaction and diffusion mechanisms at the silicon-dielectric interface. The NBTI effect is commonly described as a multi-phase process consisting of reaction, diffusion, and recovery, as illustrated in Fig. 2.5.

Reaction: Dissociation of silicon-hydrogen (Si-H) bonds through electric-field-assisted bond dissociation [48], leaving behind silicon dangling bonds (Si^*). This process is accelerated at elevated temperatures and gate bias voltages.

Diffusion: Released hydrogen atoms diffuse, recombine to form H_2 , and subsequently diffuse away from the silicon-dielectric interface.

Recovery: Hydrogen atoms diffuse back to the interface, annealing the dangling bonds and partially restoring the Si-H bonds. This process is accelerated at lower temperatures and reduced gate bias voltages.

During this process, newly generated dangling bonds act as interface traps together with pre-existing defects, facilitating hole trapping and resulting in threshold voltage degradation (ΔV_{th}).

2.4.4. NBTI Modeling

To evaluate BTI-induced degradation in advanced semiconductor technologies, physics-based TCAD approaches are widely employed to model threshold voltage degradation and its impact on device and circuit behavior [49], [50]. In this work, the multi-state configuration (MSC) framework presented in [49] is employed to model the ΔV_{th} degradation caused by NBTI by considering the reaction, diffusion,

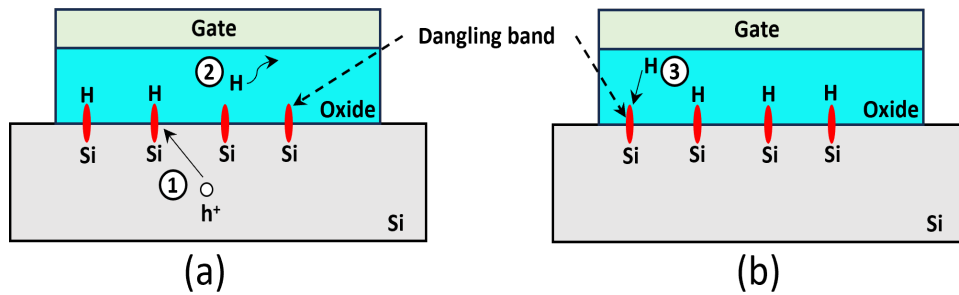


Figure 2.5.: Conceptual illustration of the reaction-diffusion (R/D) model for NBTI degradation, including Si-H bond dissociation, hydrogen diffusion, and recovery mechanisms.

recovery, and hole trapping/de-trapping mechanisms described above. The proposed ΔV_{th} model is used to investigate NBTI degradation under different operating voltage and temperature stress conditions. Detailed initial settings for Si-H bond concentration and trap concentration follow the framework presented in [49].

2.4.5. Electromigration

EM is a BEOL reliability degradation mechanism caused by atomic migration in metal interconnects under high current density. As electrons flow through narrow metal lines, momentum transfer from the conducting electrons gradually displaces metal atoms along the current direction. Over long-term operation, this process can lead to void formation, resistance increase, and eventual interconnect degradation.

Aggressive interconnect scaling considerably aggravates EM-related reliability challenges. In advanced technologies, minimum-width metal interconnects and increasing current density substantially increase resistance drift and local Joule heating effects. Dense SRAM arrays are particularly sensitive to such degradation because their bit-line and word-line interconnects operate under highly scaled metal pitches and resistance-dominated interconnect conditions. Consequently, EM-induced resistance degradation can significantly affect interconnect delay, signal integrity, and SRAM write operation reliability over the operational lifetime of the circuit.

To analyze EM-induced degradation behavior, physics-based modeling and TCAD-assisted methodologies are commonly employed to evaluate current-density distribution, stress evolution, and resistance degradation under different operating conditions [51]. The resulting interconnect degradation can subsequently be incorporated into circuit-level simulations to evaluate timing degradation and long-term reliability behavior in advanced integrated circuits.

2.4.5.1. Physical Concept of EM

The ultimate impact of EM is the increase of the parasitic resistance of the interconnect. However, from the physical point of view, the EM phenomenon is a multi-phase process: 1. void nucleation, 2. void incubation, and 3. resistance increase.

Void nucleation: Electron movement in the metallic interconnect increases hydrostatic stress until it reaches a critical value at nucleation time (t_{nucl}), leading to void formation in the metal wire and resetting the stress to zero.

Void incubation: The size of the freshly nucleated void keeps incubating in the presence of the electrical current until it reaches the critical size; i.e., covers the cross-sectional area of the interconnect at time t_{incub} .

Resistance increase: Before the existence of the fully grown void, the parasitic resistance of the interconnect does not show a significant increase. However, after t_{incub} , the interconnect's barrier shunts the current, which is the reason for the rather sharp rate of resistance increase associated with EM. In the extended circuit-level aging analysis, the amount of resistance increase leading to in-field failure is considered the criterion for EM-induced failure.

2.4.5.2. EM Modeling

In this thesis, we use physical modeling based on these three phases to model the EM phenomenon by fully capturing all the involved parameters. The resistance-increase phase affects the SRAM functionality at the circuit level. In the case of copper (Cu) metal, surrounded by a tantalum (Ta/Ta-N) barrier, the transient rate of the resistance increase (ΔR) is calculated as Equation 2.2. In Equation 2.2, the resistivity of Cu and Ta are shown by ρ_{Ta} and ρ_{Cu} respectively. The h_{Ta} , H , and W are the dimensional parameters

and correspond to the barrier thickness, height, and width of the interconnect, respectively. The physical correspondence of the other parameters is as follows: D_a is the atomic diffusivity, K is Boltzmann's constant, T is the temperature, e is the electron charge, Z is the effective valence charge, and finally j stands for the current density [52].

$$\Delta R = \left[\frac{\rho_{Ta}}{h_{Ta}(2H + W)} - \frac{\rho_{Cu}}{HW} \right] \cdot \frac{D_a}{KT} \cdot e \cdot Z \cdot \rho_{Cu} \cdot j \quad (2.2)$$

2.5. AI-Assisted Chip Design and Testing

2.5.1. AI-Assisted EDA Automation

Modern semiconductor design flows require extensive manual effort for simulation setup generation, parameter tuning, physical-design optimization, result analysis, and testing workflow configuration. Recently, AI-assisted methodologies have been increasingly explored to improve the efficiency and automation of EDA-related workflows.

Machine learning techniques have been applied to various stages of the RTL-to-GDS design flow, including design-space exploration, timing and power prediction, placement and routing optimization, congestion estimation, and reliability analysis. For instance, [53] proposes a deep reinforcement learning framework for analog circuit design automation that efficiently explores the design space and accelerates post-layout analog circuit optimization. Furthermore, [54] proposes a physics-based ANN compact modeling approach for emerging FET technologies, overcoming the convergence limitations of conventional lookup-table (LUT)-based methodologies.

In reliability-oriented semiconductor analysis, AI-assisted approaches can considerably reduce simulation overhead and improve workflow automation. These methodologies are particularly useful for computationally expensive reliability evaluation flows involving TCAD simulation and circuit-level analysis. For example, [55] proposes a graph neural network (GNN)-based methodology for efficient prediction of aging- and process-variation-induced timing degradation without requiring computationally expensive static timing analysis (STA).

2.5.2. LLM-Based Design and Testing Automation

Recent advances in LLMs have further expanded the application of AI in chip design and testing automation. Emerging LLMs such as GPT [56] and DeepSeek [57] have demonstrated strong capabilities in natural-language understanding, reasoning, and code-generation tasks. Unlike conventional machine-learning approaches that mainly focus on numerical optimization tasks, LLMs can assist in script generation, workflow automation, test-case generation, and EDA tool interaction through natural-language instructions.

LLM-assisted methodologies have recently been explored for RTL generation, hardware verification, debugging assistance, and automated generation of simulation and testing scripts [58], [59], [60], [61]. Furthermore, LLM-based agents have shown strong potential for automating complex EDA workflows across multiple stages of the RTL-to-GDS flow. In reliability-oriented design flows, such approaches can improve the automation of TCAD setup generation, parameter configuration, and reliability analysis workflows, thereby reducing manual engineering effort and improving productivity.

Part II.

Contributions

3. Machine Learning-Based Soft Error Rate Estimation for Pass-Transistor Logic

Efficient gate-level SER estimation in PTL circuits is challenging due to degraded voltage restoration and irregular transient propagation behavior. Conventional fast SER estimation methodologies commonly rely on simplified SET propagation models based on trapezoidal pulse propagation characterized by pulse width and pulse amplitude. However, PTL circuits frequently exhibit distorted SET waveforms and long-tail transient propagation behavior that cannot be accurately represented using such simplified pulse models.

Although MC- and SPICE-based simulations can accurately evaluate SET propagation behavior, their computational complexity becomes prohibitive for large-scale circuits and extensive design-space exploration. Therefore, this chapter presents a machine-learning-based framework for efficient gate-level SER estimation in PTL circuits [1]. The main contributions of this chapter are summarized as follows:

- Modeling SET propagation in PTL circuits using a data-driven framework that captures both pulse-width and pulse-amplitude characteristics.
- Developing a hybrid machine-learning framework combining multi-layer feed-forward neural networks and support vector classifiers (SVCs) for SET prediction and irregular transient classification.
- Applying Bayesian optimization based on Gaussian Processes for efficient hyperparameter tuning.
- Extending the proposed methodology from 1-bit full adders (FAs) to n -bit ripple-carry adders (RCA) to enable scalable SER estimation.
- Demonstrating substantial runtime reduction while maintaining high prediction accuracy compared with MC-based circuit-level simulations.

The remainder of this chapter is organized as follows. Section 3.1 reviews related work and discusses the challenges of SET modeling in PTL circuits. Section 3.2 presents the proposed machine-learning-based framework, including the training data generation methodology. Section 3.3 evaluates the accuracy and runtime performance of the proposed approach and compares it with MC-based simulations. Finally, Section 3.4 concludes the chapter.

3.1. Challenges of SET Modeling in PTL

The proposed SER estimation methodology is applied to an RCA. Adders are fundamental building blocks in digital systems and are widely used in communication and data-processing applications, such as Low Density Parity Check (LDPC) decoders, digital signal processors, and arithmetic logic units (ALUs) [62]. In advanced technology nodes, reduced supply voltages and shrinking noise margins increase the susceptibility of combinational logic circuits to radiation-induced soft errors. Since transient faults occurring in adders can propagate through arithmetic paths and affect system-level functionality, accurate SER evaluation is essential for reliable operation. From an industrial perspective, adder design is additionally constrained by PPA requirements. As modern processors integrate a large number of arithmetic units, achieving an optimal trade-off between PPA and reliability has become increasingly important in advanced digital circuit design.

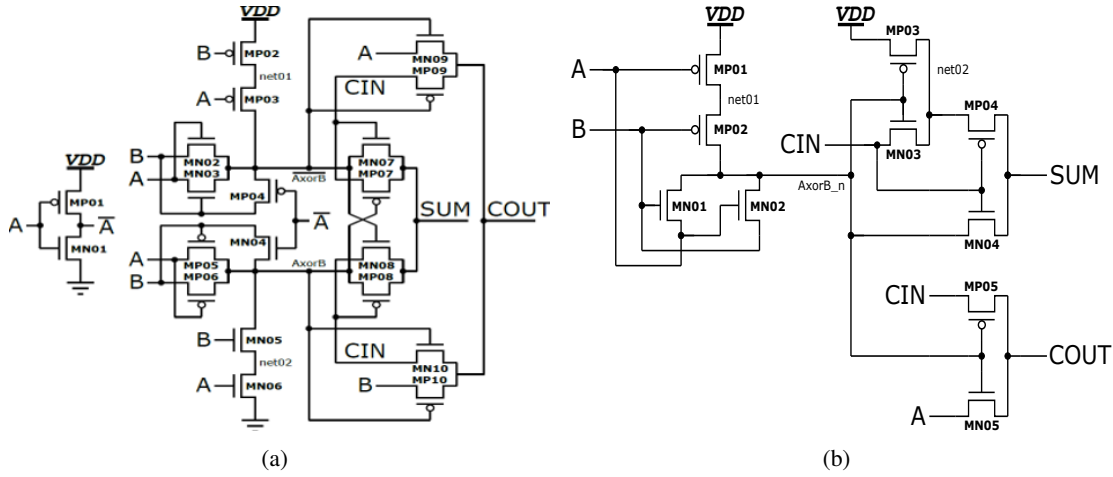


Figure 3.1.: Two of the four PTL-based full adders investigated in this chapter: (a) NASERI 20T full adder [63], (b) BUI02A9 (BA) full adder [64].

3.1.1. Irregular SET Propagation Behavior

SET propagation behavior in PTL circuits with open-drain structures differs significantly from that of conventional CMOS logic. To illustrate this behavior, transient simulations are performed for the BUI02A9 (BA) 1-bit full adder [64]. As shown in Fig. 3.1(b), the COUT node is directly driven by the “CIN” node rather than by the supply voltage (VDD). Consequently, the transient response strongly depends on charge-transfer dynamics within the PTL structure.

To investigate SET propagation characteristics, trapezoidal voltage pulses with varying pulse widths and amplitudes are injected into the CIN node. The input nodes A and B are fixed at logic values of 800 mV and 0 V, respectively. The resulting SET responses observed at the COUT node are shown in Fig. 3.2. Unlike conventional CMOS circuits, the transient voltage at COUT does not fully recover to 0 V within the simulation window of 3 ns, resulting in a long-tail transient waveform.

Such long-tail SET behavior presents a major challenge for pulse-width modeling, since the second crossing point at 50% of the pulse amplitude cannot be accurately extracted during transient simulation. Consequently, conventional SET characterization approaches based on standard pulse-width definitions become insufficient for accurately modeling SET propagation in PTL circuits.

3.1.2. Related Work

Adders are fundamental building blocks in modern computing systems, since arithmetic summation forms the basis of numerous digital operations [62]. PTL-based adder architectures have attracted considerable research interest due to their reduced transistor count and improved power efficiency compared with conventional CMOS implementations.

A partially PTL-based adder tree for convolutional neural network acceleration is presented in [65], where an interleaved structure combining CMOS 28T full adders and CHIH 14T (C14) PTL full adders achieves approximately 30% reduction in power-delay product (PDP). Furthermore, a comparative analysis of multiple 1-bit full-adder implementations is reported in [66], demonstrating that PTL-based adder circuits can achieve up to 63% reduction in energy-delay product (EDP) compared with conventional CMOS 28T full adders.

As technology scaling continues, radiation-induced soft errors have become an increasingly important reliability concern in advanced VLSI systems due to reduced supply voltages, shrinking noise margins, and higher operating frequencies [67]. Accurate SER estimation is therefore essential for ensuring reliable

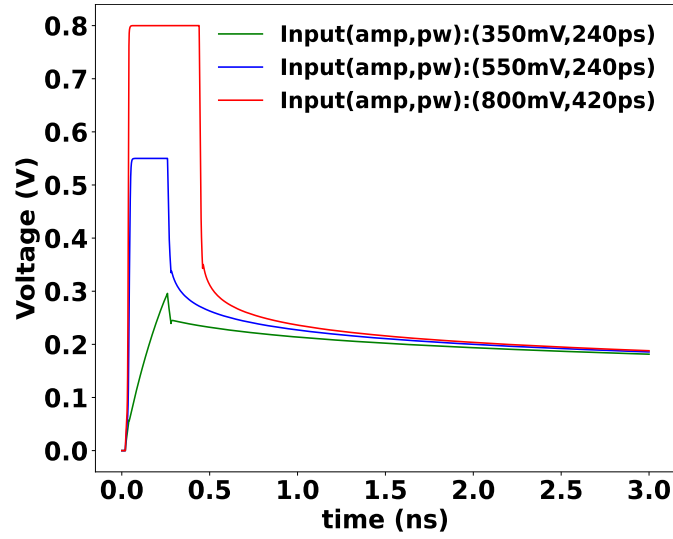


Figure 3.2.: SET propagation behavior in the BA PTL full adder under different injected pulse conditions. Long-tail transient responses are observed due to incomplete voltage recovery at the output node.

operation in modern digital circuits, particularly in high-performance and communication-oriented applications. Several approaches have been proposed for SER estimation in combinational logic circuits. The most straightforward approach relies on circuit-level transient simulations using randomly generated input patterns [67], [68]. Although this method can provide high accuracy, its computational cost becomes prohibitive for large-scale circuits and extensive Monte Carlo analysis. To improve efficiency, probabilistic SER estimation models focusing on logical masking effects have been proposed in [69]. In [70], a trapezoidal waveform approximation approach is introduced to model electrical masking effects during SET propagation. More recently, machine-learning-based reliability modeling approaches have also been investigated. A data-driven machine learning regression discriminator is proposed in [71] for SRAM reliability analysis. The model is trained using TCAD-generated datasets, where charge deposition information from secondary ion strikes is utilized as the model input to predict soft-error behavior.

Despite these advances, existing SER estimation approaches mainly focus on conventional CMOS circuits and do not adequately capture the unique SET propagation behavior observed in PTL structures, particularly the irregular long-tail transient responses discussed in Section 3.1. These challenges motivate the machine-learning-based SER estimation framework proposed in this chapter.

3.2. Proposed Machine-Learning-Based SET Modeling Framework

This section presents the proposed machine learning-based methodology for modeling SET propagation in combinational logic circuits. Monte Carlo simulations are first employed to generate datasets for training neural network regression models and SVC classifiers under fixed PVT conditions. The trained models are then utilized to predict SET pulse propagation and soft error behavior in larger circuits, such as ripple carry adders, while also handling long-tail transient signals observed in PTL-based designs. Finally, the accuracy of the proposed approach is evaluated using regression and classification metrics on multiple full-adder implementations.

3.2.1. Overall Modeling Flow

Fig.3.3 shows the flow of our proposed SET modeling method. The two main steps are:

- **Model building.** Building regression models for SET propagation through the logic cell elements, and classification models for special long-tail signal cases.
- **Model exploitation.** Obtain the SER of combinational logic circuits that consist of the logic cell elements in the presence of SETs using models.

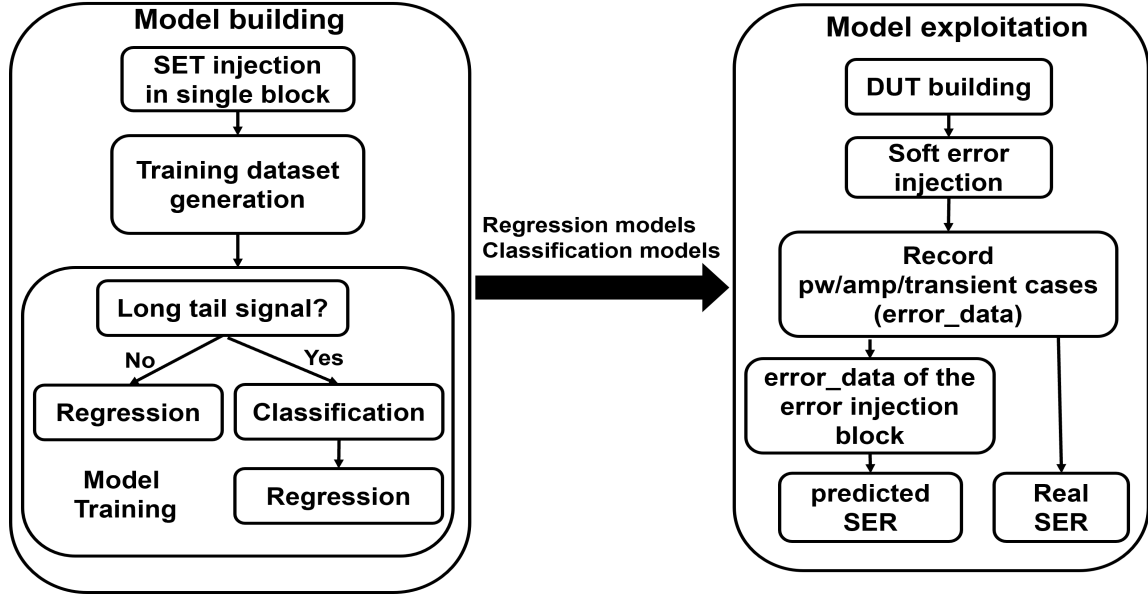


Figure 3.3.: Proposed method

During the first step, we run MC-based simulations to generate datasets for building the SET propagation models. Simulation results are collected under constant PVT conditions. Similar to [72], the trapezoidal signal is used in our work to approximate how a glitch propagates through a gate. We inject trapezoidal signals with different pulse amplitudes (amp) and pulse widths (pw) into the input port (In_{amp}, In_{pw}), and record the pulse amplitude and pulse width of the output (Out_{amp}, Out_{pw}). Pulse width is defined as the time difference between the 50% amplitudes of the rising and falling edges, while pulse amplitude can be either the most positive amplitude or the most negative amplitude.

As mentioned in Section 3.1.1, we observe long-tail signals in simulations for 1-bit FA based on BA and BUI02 SERF (BS) PTL designs [64]. Our proposed solution to model such SETs with long tails is to set the pulse width of such a long-tail pulse to a value that is larger than the transient simulation time (3ns), for instance 4ns, and train support vector classification (SVC) classifiers to recognize whether such a long tail can be caused by the given (In_{amp}, In_{pw}) or not, namely two categories, "Irregular" and "Regular". Samples of the "Irregular" category are not used in building regression models because their assigned pulse widths do not reflect the physical relationship between (In_{amp}, In_{pw}) and Out_{pw} .

Then, feed-forward neural networks are trained with these datasets to model how (Out_{amp}, Out_{pw}) is related to (In_{amp}, In_{pw}). Each machine learning model has two predictor variables and one target variable since we build models for pulse width and pulse amplitude prediction separately. These models are also specified for each input vector combination (X). In this step, we obtain models that represent the following relationship for the tested single block:

$$SET_{in}(pw, amp, X) \rightarrow SET_{out}(pw, amp, X) \quad (3.1)$$

In the second step, we build up the simulation device under test (DUT), for instance, an 8-bit ripple carry adder. Double exponential sources [42] are considered to approximate the radiation-induced transient

Table 3.1.: Hyperparameters

Hyperparameters	space
Regression	
number of hidden layers	[1, 10]
neurons per layer	[2, 36]
activation function	ReLU, Tanh, Sigmoid
optimizer	SGD, Adam
learning rate	[0.0001, 0.11]
batch size	2,4,8,16,32,64
Classification	
C	[1e - 6, 100.0]
Kernel	Linear, Poly, RBF, Sigmoid
Gamma	[1e - 6, 100.0]
Kernel	1,2,3,4,5

current waveform. Current sources connected to device nodes are randomly activated to simulate a random particle strike. Then, (Out_{amp}, Out_{pw}) of the errors-injected block are recorded.

For a block except the error-injected block in the DUT, (In_{amp}, In_{pw}) is equivalent to (Out_{amp}, Out_{pw}) of its Fan-in block. Similarly, $SET_{i,out}$ is equivalent to $SET_{i-1,out}$, i stands for the i^{th} block in the chain. SVCs are first called to recognize whether $SET_{i,out}(pw)$ is infinite because of the long tail or not. If yes, then $SET_{i,out}(pw)$ is set to 4ns. Otherwise, neural network models are loaded to predict $SET_{i,out}$ of all blocks in the DUT. In this step, Equation 3.1 is extended for the chain of blocks:

$$SET_{i,in}(pw, amp, X) \rightarrow SET_{i,out}(pw, amp, X) \quad (3.2)$$

Finally, all Out_{amp} are compared with a pre-defined error margin to determine whether soft errors can be caused by SETs. In this paper, the error margin is specified to 0.4V, which means that when the absolute difference between the erroneous amplitude and the error-free output is higher than 0.4V, a soft error occurs.

The adder circuit is investigated to experimentally evaluate our proposed method. SET propagation models for 1-bit FA with different adder types, as mentioned in Section 3.1.2, are first obtained. Afterwards, an 8-bit ripple carry adder is considered as the DUT to test the performance of the generated SET propagation models.

3.2.2. SET Propagation Model Based on Neural Network

Multilayer feed-forward neural networks, trained with a back-propagation learning algorithm, are widely used because of their ability to model complex nonlinear relationships between dependent and independent variables, and to detect all possible interactions between predictor variables. We use feed-forward neural networks, which contain two input layers, one output layer, and several hidden layers to build models for propagated SET pulse width and amplitude prediction. To tune hyperparameters, we select Bayesian optimization using the Gaussian Processes tuning algorithm [74]. All considered hyperparameters and their spaces are listed in Table 3.1.

Table 3.2.: Simulation setup and adder types

Sizing of transistors	minimal sizing, W/L = 80nm/20nm
PVT	TT (typical-typical), 0.8V, 25°C
Threshold voltage	LVT (low threshold voltage)
Simulation time	3ns
Adder types	CMOS, BA [64], BS [64], C14 [73], N20 [63], N22 [63]
Amplitude of injected pulses	(0V,0.8V)
Width of injected pulses	(30ps, 1ns)
Hardware resources setting	1 K80 GPU, 1 Xeon 2.20GHz CPU

Table 3.3.: Performance of SVC and NN models

Circuit	Classification					
	Irregular			Regular		
	Precision	Recall	F1	Precision	Recall	F1
BA	0.99	0.995	0.99	1.0	0.995	0.995
BS	0.988	0.995	0.993	1.0	0.993	0.998
Circuit	Regression					
	MSE					
CMOS	0.0005					
BA	0.0021					
BS	0.00359					
C14	0.00252					
N20	0.001167					
N22	0.00118					

3.2.3. SET Classification Based on SVC

Although regression models can efficiently capture regular SET propagation behavior, PTL circuits frequently exhibit irregular long-tail transients that cannot be accurately represented using conventional pulse-width metrics. Therefore, an additional classification mechanism is introduced. The SVC classifier is well known because of its effectiveness in high-dimensional spaces and low memory usage. The training setup is the same as training neural network models. All hyperparameters and their spaces are listed in Table 3.1.

3.2.4. Simulation Setup and Data Preparation for Training

The purpose of the circuit simulations is to detect how different 1-bit FA types respond to transient pulses. To collect the training dataset, Virtuoso simulations are set up. We use GlobalFoundries FDSOI 22 nm technology transistor models to simulate all circuits. Simulation setup and all considered adders are listed in Table 3.2. A trapezoidal voltage source is connected to the carry-in port to approximate a single-event transient. To ensure the randomness of trapezoidal pulses, we use an MC-based simulation (50000 points) to generate pulses with random width and amplitude.

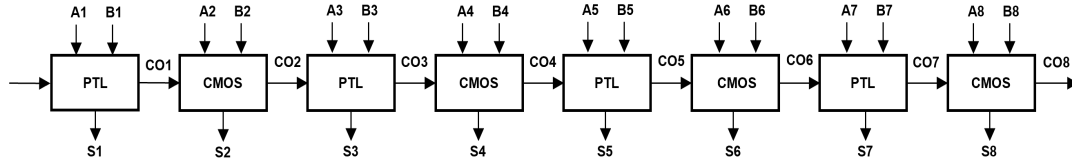


Figure 3.4.: Structure of the evaluated 8-bit RCA. Input and output ports are connected to X1 load buffers, while SET injection is performed at the first full-adder stage (LSB).

3.2.5. Evaluation of Regression and Classification Models

As introduced in Section 3.2.2, test data sets contain 10000 samples for both regression and classification. These test data are utilized to evaluate the performance of trained regression and SVC models. In the test, we use Mean Squared Error (MSE) as the metric for regression and precision, recall, and F1-score as the metrics for classification. All mean results are presented in Table 3.3. Irregular SETs are only observed in BA and BS. Hence, SVC is only needed for these two circuits.

3.3. Results and Discussion

3.3.1. Simulation Setup and Test Scenario

To validate the proposed method, we first run MC-based transient simulations (10000 transient-event injections) in Virtuoso to obtain a stochastic SER of three types of 8-bit ripple-carry adders. We then use the neural network models together with a support vector classifier to estimate SER and compare the results. Simulation setup and hardware resources setting are shown in Table 3.2.

Compared to pure CMOS design, energy efficiency is reported in [65] to be enhanced if one stage of PTL is followed by CMOS in an adder tree. To model SET generation and propagation in such structures, we considered the following test scenario with respect to the SET generation block: the DUT consists of one type PTL FA and one CMOS FA, each PTL FA is followed by a CMOS FA, the error-injection FA is a PTL FA, as shown in Fig. 3.4.

3.3.2. Accuracy Comparison

The prediction accuracy of the proposed ML-based methodology was evaluated by comparing its results with MC simulations for all outputs, as shown in Fig. 3.5. Overall, the proposed model demonstrates high prediction accuracy for all adder topologies. The CMOS-based adder exhibits the best agreement with MC simulation, while the maximum observed difference (Diff) is approximately 0.014 for the C14 topology.

Regarding the SER analysis, the conventional CMOS RCA shows the strongest robustness against transient fault propagation, as shown in Fig. 3.6 and Fig. 3.7. For the COUT1 node, the CMOS implementation achieves the lowest SER of approximately 14%. In contrast, the C14 PTL-based topology exhibits a significantly higher SER of around 36%, corresponding to nearly 2.5× higher SER compared to CMOS. Among the hybrid PTL-CMOS structures, N20 achieves the lowest COUT1 SER, approximately 18%.

A similar trend is observed for the SUM1 node. The CMOS RCA again provides the lowest SER, approximately 25.6%, indicating superior stability of the sum generation path. The BA PTL-based topology shows the highest degradation, with a SUM1 SER of around 48%, corresponding to nearly 1.9× higher SER than CMOS. Among the PTL-CMOS hybrid designs, N20 achieves the best reliability performance, with a SUM1 SER of approximately 32%.

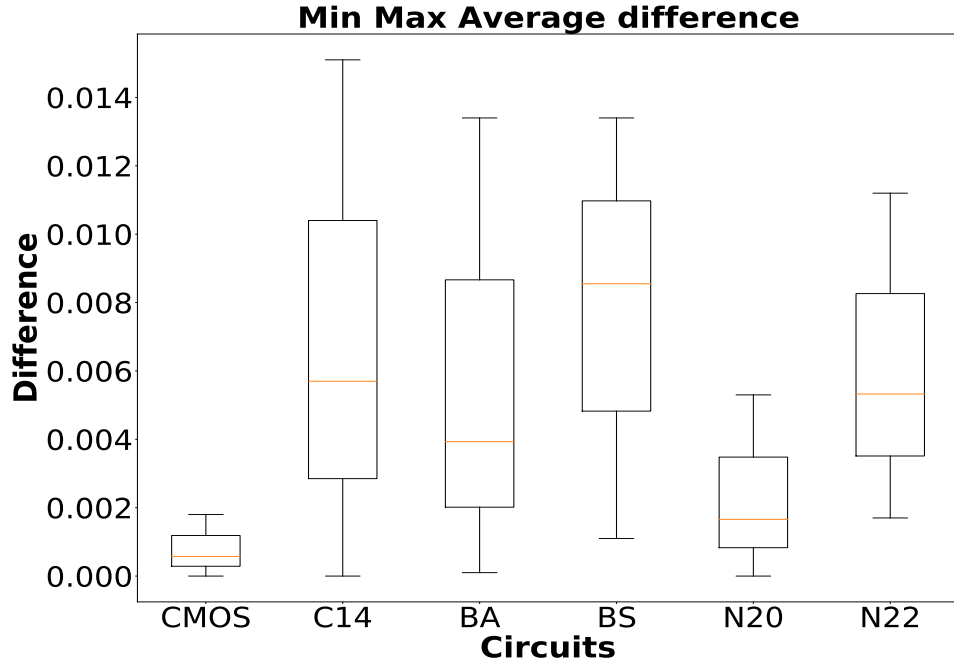


Figure 3.5.: SET injection in the first PTL stage, SER comparison over all output bits (all sum bits and final carry out).

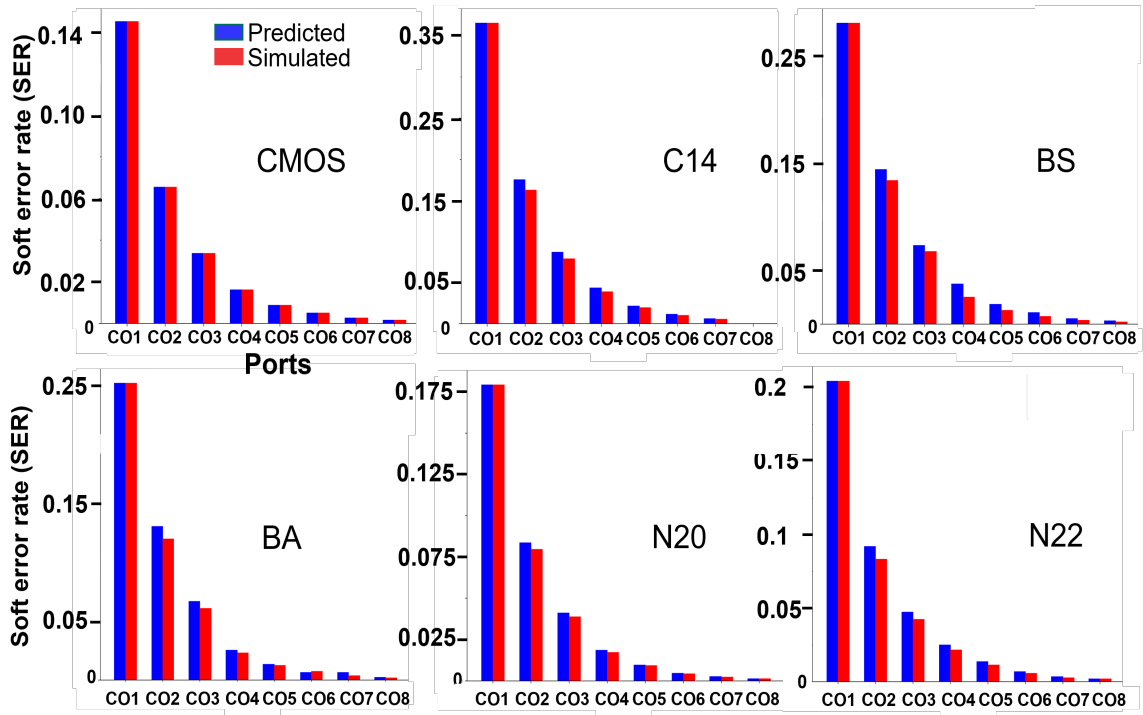


Figure 3.6.: Comparison between simulated and predicted SERs at the carry-output ports of different RCA implementations.

Overall, the results indicate that the conventional CMOS implementation provides the highest resilience against soft-error propagation, whereas PTL-based designs generally exhibit increased SER vulnerability due to their reduced logic-level robustness.

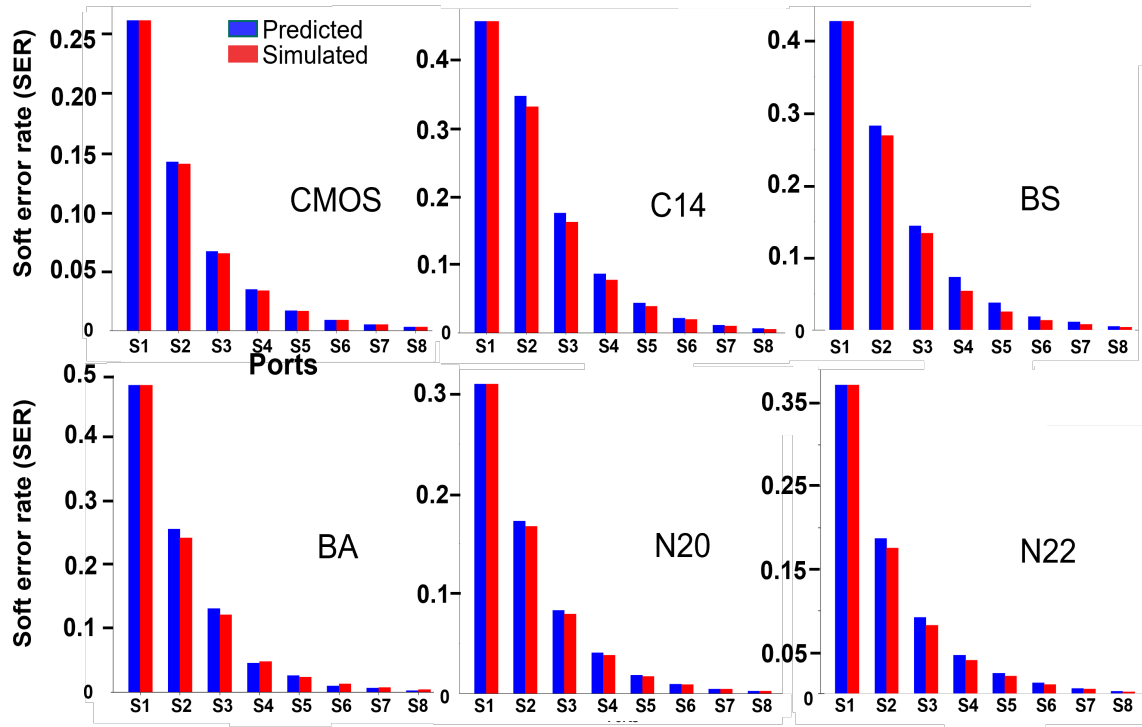


Figure 3.7.: Comparison between simulated and predicted SERs at the sum-output ports of different RCA implementations.

3.3.3. Runtime Comparison

We compare the execution runtimes of MC-based Virtuoso Spectre simulation and the ML-based prediction method, and then present the average speedup (ASU). As shown in Table 3.4, the simulation results indicate that the proposed method is much faster than Virtuoso MC simulation. The smallest speed-up is 8.9 $\times$, which is achieved for BA, reaching up to 28.6 $\times$ for CMOS. The average speed-up is around 13.8 $\times$.

Table 3.4.: Runtime comparison of ML and MC methods. ML: proposed Machine Learning approach, MC: Monte Carlo simulation, ASU: Average Speed-up

Circuit	First scenario			Second scenario		
	MC	ML	ASU	MC	ML	ASU
CMOS	31m21s	1m6s	28.6x	31m21s	1m6s	28.6x
BA	37m56s	4m17s	8.9x	32m21s	2m47s	11.6x
BS	36m49s	3m12s	11.5x	32m27s	1m24s	23.2x
C14	49m10s	4m8s	11.9x	33m10s	1m12s	27.6x
N20	45m35s	2m1s	22.6x	40m3s	1m16s	31.6x
N22	41m33s	2m51s	14.6x	36m29s	1m16s	28.8x
Average	40m24s	2m56s	13.8x	34m18s	1m30s	22.9x

3.4. Conclusion

PTL emerges in advanced high-speed communication systems because of its high operating speed and potentially lower power consumption. Conventional trapezoidal SET propagation modeling for rapid estimation of the SER becomes insufficient for PTL, and the proposed framework enables scalable SER estimation under nonlinear transient propagation behavior. In this chapter, we developed a data-driven SET propagation framework for efficient gate-level SER estimation in PTL circuits. Furthermore, the accuracy and run time of the proposed method are compared with extensive Monte Carlo simulations. Our proposed models are data-driven, easy to reuse, and much faster than MC simulation.

4. A Learning-Based Approach for Single Event Transient Analysis in Pass Transistor Logic

Although the gate-level ML-based SET estimation framework presented in Chapter 3 can efficiently predict soft error behavior for specific PTL circuits, its generalization capability to unseen PTL topologies remains limited. In practice, newly designed PTL structures often require additional simulation data and model retraining, which reduces the scalability and reusability of black-box prediction approaches.

To address this limitation, SET propagation behavior can instead be modeled using reusable transistor-level propagation primitives. Since complex PTL circuits are typically constructed from a limited number of fundamental PTL structures, transient propagation behavior in large circuits can be analyzed compositionally by combining the propagation characteristics of basic PTL building blocks.

However, accurate propagation modeling of PTL structures remains challenging due to degraded voltage restoration, distorted transient propagation behavior, and long-tail SET waveforms. Conventional compact pulse models developed for CMOS circuits cannot fully capture these effects during PTL transient propagation.

Therefore, this chapter presents a learning-based structural SET propagation framework for transistor-level modeling in PTL circuits, based on the work published in [2]. The proposed methodology combines reusable transistor-level propagation models with curve-fitting and machine-learning techniques to enable scalable compositional SET analysis for complex PTL and CMOS circuits while significantly reducing simulation complexity compared with conventional SPICE-based approaches.

The remainder of this chapter is organized as follows. Section 4.1 introduces PTL fundamentals, SET propagation challenges, and related work. Section 4.2 presents the proposed learning-based SET modeling framework, including simulation setup and regression modeling methodology. Section 4.3 evaluates the accuracy of the proposed approach and compares it with circuit-level simulations. Finally, Section 4.4 concludes the chapter.

4.1. Preliminaries

This section discusses the fundamental challenges of SET propagation modeling in PTL circuits. Due to the charge-transfer-based operation of PTL structures, transient propagation behavior differs significantly from conventional CMOS logic, making accurate SET characterization considerably more difficult. Furthermore, the structural characteristics of PTL circuits and related work on regression-based SET modeling approaches are reviewed to motivate the proposed structural learning-based methodology.

4.1.1. Challenges of SET Propagation in PTL

PTL circuits rely on pass-transistor-based charge transfer between internal nodes rather than conventional pull-up/pull-down switching behavior. As shown in Fig. 4.1(a), a PTL 1-bit FA [64] can be implemented using significantly fewer transistors compared with conventional CMOS implementations. However, degraded voltage restoration, threshold-voltage loss, and charge-sharing effects considerably complicate transient propagation analysis in PTL circuits.

In conventional CMOS circuits, SET propagation is typically characterized using pulse height (PH) and pulse width (PW), as illustrated in Fig. 4.2. However, such characterization becomes challenging in PTL

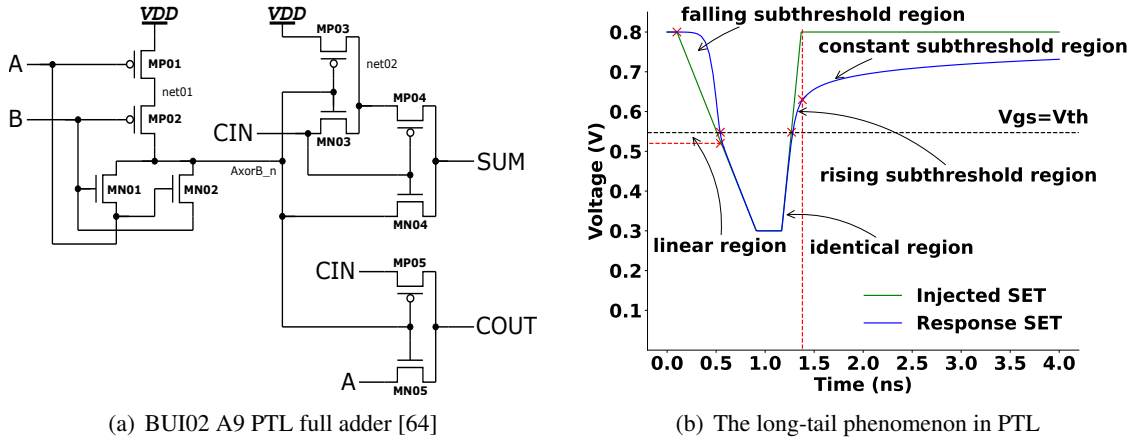


Figure 4.1.: (a) BUI02 A9 PTL full adder [64], and (b) long-tail transient behavior observed during SET propagation in PTL circuits.

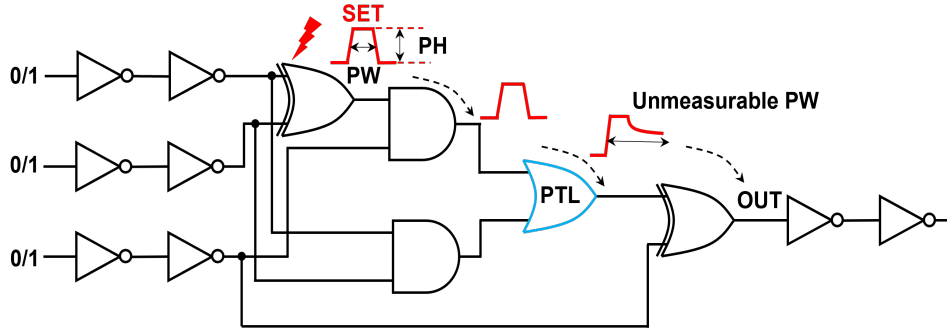


Figure 4.2.: SET generation and propagation through a combinational logic chain.

circuits due to distorted transient propagation behavior. As shown in Fig. 4.1(b), PTL circuits may exhibit long-tail transient waveforms in which the output voltage does not fully recover within the transient simulation window. This behavior is mainly caused by weak restoration capability and charge-sharing effects in PTL structures. Consequently, conventional pulse-width extraction based on the second 50% crossing point becomes difficult or even impossible.

Despite the complexity of transient propagation behavior, many PTL circuits are constructed from a limited number of fundamental pass-transistor structures. Therefore, transient propagation behavior in complex PTL circuits can potentially be analyzed compositionally by combining reusable propagation characteristics of basic PTL primitives. Such a structural modeling methodology could improve the scalability and reusability of SET propagation analysis compared with conventional gate-level black-box approaches.

4.1.2. Related Work

Curve-fitting and regression techniques are widely used in MOSFET circuit analysis. A study investigating the temperature dependence of MOSFET drain current using curve fitting is presented in [75], where a simplified delay-time model is derived to predict the linear temperature coefficient of CMOS circuit delay. Inspired by the fact that inverter delay is a complicated function of several process parameters, such as transistor sizing and load capacitance, a propagation delay model for CMOS inverters is developed in [76], demonstrating high prediction accuracy compared with SPICE simulations.

Several studies have also investigated SET modeling in recent years. A graphical derivation methodology is presented in [77], where the shape of SET voltage pulses is estimated using transient current data obtained at different drain voltages. Moreover, an analytical approach is proposed in [78] to explain electrical masking effects of SETs at transistor level. Effective capacitance and SET currents are modeled to capture nonlinear transistor behavior and approximate SET waveform propagation.

More recently, data-driven approaches have attracted increasing attention in SET modeling and soft error analysis. Machine-learning-based regression models trained using SET current data extracted from TCAD simulations are proposed in [79]. These approaches can predict SET-induced circuit behavior without explicitly modeling complex physical mechanisms.

Despite these advances, existing SET modeling approaches mainly focus on conventional CMOS circuits and do not adequately capture the distorted transient propagation behavior frequently observed in PTL structures, particularly long-tail transient waveforms. Furthermore, existing data-driven approaches are primarily developed at the gate level and therefore require retraining for different circuit implementations. In addition, most existing methodologies do not exploit the structural compositionality of PTL circuits for reusable transient propagation analysis. These limitations motivate the structural learning-based SET propagation framework proposed in this work.

4.2. Structural SET Propagation Modeling Framework

In this section, the proposed structural learning-based framework for SET propagation modeling is presented. Unlike gate-level black-box prediction approaches, the proposed methodology models SET propagation behavior at the transistor-structure level using reusable PTL propagation primitives. Since complex PTL circuits are typically composed of a limited number of fundamental pass-transistor structures, SET propagation in large PTL circuits can be analyzed compositionally by combining the propagation characteristics of basic PTL building blocks. As illustrated in Fig. 4.3, the proposed framework consists of four main phases. The major steps include:

- *Data Collection*: Circuit-level simulations are performed for each PTL structure, and the corresponding SET waveforms are recorded.
- *Ground Truth Generation*: For regular pulses, the shape parameters are derived from waveforms. For irregular pulses, curve fitting is performed for each region, and the fitting parameter values are recorded.
- *Model Building*: SET propagation models, including pulse-shape and propagation-delay models, are trained for each PTL structure.
- *Model Exploitation*: The trained models are used to estimate the propagated SET shape and delay in larger combinational circuits constructed from the considered PTL elements.

Unlike conventional CMOS logic, transient recovery in PTL circuits can require a significantly longer time. This phenomenon is illustrated by the example shown in Fig. 4.1(b). As presented in [80], the mechanism of production of radiation-induced SETs can be modeled by the injection of a current pulse, while the voltage pulse resulting from the current injection can be approximated by a trapezoidal or triangular pulse. In this chapter, regression-based functions are proposed to model the propagation of trapezoidal SET waveforms through PTL structures. These functions can effectively characterize the variations in the voltage pulse shapes as they propagate through various structures in the PTL chain.

The main limitation of our methodology is the requirement to measure the equivalent output capacitance of each net in a tested circuit by circuit-level simulation under various load conditions. However, this challenge can be mitigated with the idea of ParaGraph [81], which leverages a graph neural network

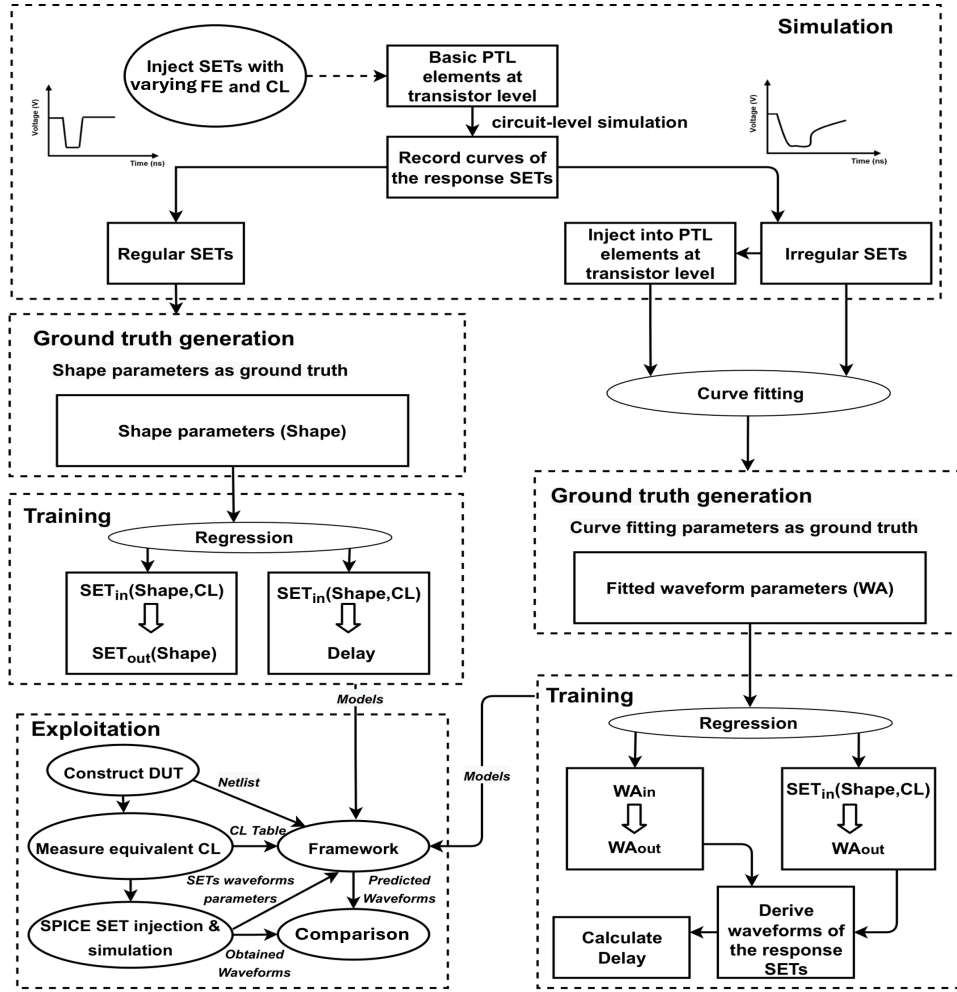


Figure 4.3.: Overall workflow of the proposed learning-based SET propagation modeling framework.

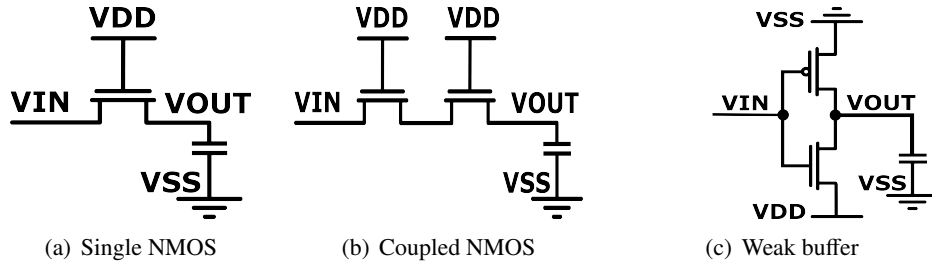


Figure 4.4.: Common PTL structures considered in the proposed SET propagation modeling framework.

model to estimate net parasitic by converting circuit netlists into graphs, thereby enabling our models to operate without running exhaustive circuit-level simulations for model building.

We inject a high-to-low SET into the VIN port of the structure shown in Fig. 4.4(a). After the recovered output exceeds the minimal ON-state gate-source threshold voltage, the NMOS PTL stops conducting. Hence, the propagated SET in PTL can have a very long tail as illustrated in Fig. 4.1(b), which makes SET modeling challenging. Based on whether PTL elements recover towards the initial value quickly, we classify the propagated SETs as follows:

- Irregular pulse (long-tail): Output recovers very slowly towards the initial value and has a long tail.

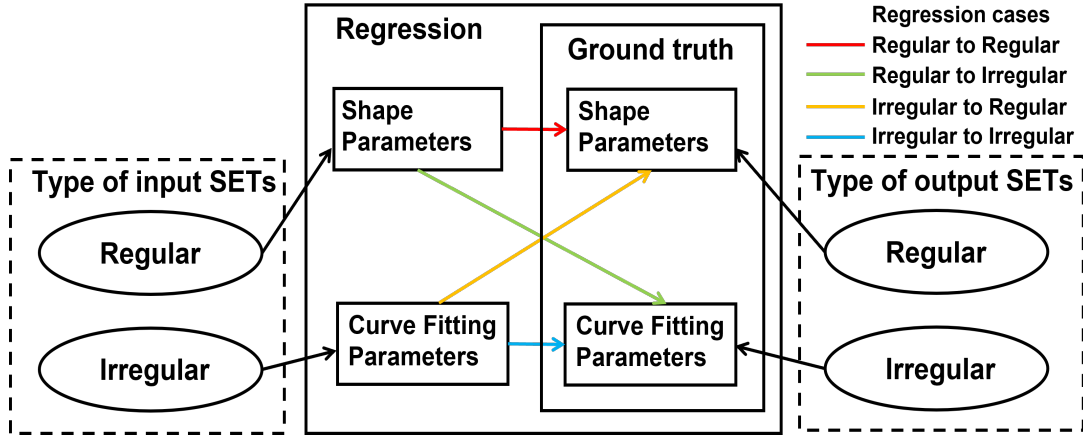


Figure 4.5.: Regression cases and their ground truth.

- Regular pulse: The output recovers quickly.

The shape and delay of regular and irregular SETs are modelled separately in the following subsections.

4.2.1. Data Collection and Ground Truth Generation

During this step, we run electrical-level simulations to generate datasets. All datasets are generated using 12nm FinFET technology. SETs with varying shape parameters are injected into each structure in Fig. 4.4 that can be found in the implementation of PTL circuits, with different capacitive load cases (CL), to simulate the SET propagation in a combinational logic chain as shown in Fig. 4.2. Each dataset contains 30000 samples, which are divided randomly into a training set and a test set with the ratio 70:30. The simulation setup can be found in Table 4.1.

In this chapter, we mainly explain NMOS PTL since PMOS PTL behavior is complementary to NMOS PTL behavior and can be derived accordingly. In our work, discrete values of the waveform are used to approximate the transient pulse. A time step (Δt) is defined, and at each time step, the voltage value is sampled. In this way, the waveforms of output SET voltage pulses of each tested structure are recorded.

Ground truth of all related regression cases is illustrated in Fig. 4.5. Shape parameters and curve fitting parameters are either predictors or targets, depending on the regression cases.

4.2.2. Modeling Regular SETs

This subsection describes the proposed shape and propagation delay modeling method for regular SETs, i.e., the case "Regular to Regular". First, we derive the shape parameters of regular SETs from waveforms. Consistent with existing SET modeling approaches [82], the shape parameters are: rise time (RT), fall

Table 4.1.: Simulation setup

Transistor sizing	Minimal sizing, $W/L = 80 \text{ nm}/20 \text{ nm}$
Process, Voltage, and Temperature (PVT)	TT (typical-typical), 0.8 V, 25°C
Threshold voltage	SLVT (super low threshold voltage)
Load capacitance range	(200 aF, 2000 aF)
Pulse parameters	$(RT, FT, PW) = (50 \text{ ps}, 500 \text{ ps}), PH \in (0 \text{ V}, 0.8 \text{ V})$

time (FT), pulse width (PW), and pulse height (PH), which can inherently define the corresponding waveform. The obtained datasets represent the following relationship for each structure:

$$SET_{in}(Waveforms, Load) \rightarrow SET_{in}(Shape) \quad (4.1)$$

The shape parameters of propagated regular SETs depend solely on the shape parameters of injected SETs and load. Hence, regression is utilized to establish the following relationship:

$$SET_{in}(Shape, Load) \rightarrow SET_{out}(Shape) \quad (4.2)$$

Then, neural network (NN) models are trained to predict the propagation delay, which indicates the relationship:

$$SET_{in}(Shape, Load) \rightarrow Delay \quad (4.3)$$

To simplify the modeling process, we utilize neural networks due to our observation of the complex dependency of delay on shape parameters and load, which includes both linear and non-linear relationships. Hyperparameters are listed in Table 4.2.

The following cases are categorized as regular SET propagation cases, and all the rest of the cases are irregular pulses:

- High-to-low SETs propagate through PMOS
- Low-to-high SETs propagate through NMOS
- Regular pulses propagate through inverter.
- Irregular pulses propagate through inverter.
- Regular pulses propagate through a weak buffer.

4.2.3. Modeling Irregular SETs

This subsection explains the basic concept of modeling the shape of the irregular SETs. The shape modeling method for the regular SETs is not appropriate for irregular SETs since the output may not reach the 50% PH, and PW is thus not measurable. Hence, an analytical fitting-based method is proposed to approximate the waveforms of irregular SETs. Then, regression is needed to estimate how such irregular waveforms propagate through several structures. Modeling propagation delay is not essential for irregular SETs since it can be estimated from the waveforms.

As illustrated in Fig. 4.1(b), as the input voltage falls, the NMOS PTL starts operating in the subthreshold region, and shortly thereafter it moves to the saturation and then the linear region, respectively. After the

Table 4.2.: Hyperparameters

Hyperparameters	Search space
Number of hidden layers	[1, 10]
neurons per layer	[2, 36]
activation function	ReLU, Tanh, Sigmoid
optimizer	SGD, Adam
learning rate	[0.0001, 0.11]
batch size	2,4,8,16,32,64

output voltage is equal to the input voltage, they remain identical until the input rises above a value, which can be safely assumed to be the minimum ON-state gate-source voltage ($V_{gs(th)}(min)$) of the NMOS transistor according to the simulation results. Finally, the output rises extremely slowly, since the NMOS PTL operates again in the subthreshold region.

Hence, the operating regions of the NMOS PTL during SET propagation can be classified as follows and be fitted separately:

- Falling subthreshold region: Output voltage is higher than $V_{gs(th)}(min)$ while input voltage is falling.
- Identical region: Output and input voltages are identical.
- Transition region: The region between the falling subthreshold and the identical region.
- Rising subthreshold region: Output voltage is higher than $V_{gs(th)}(min)$ while input voltage is rising.
- Constant subthreshold region: Output voltage is higher than $V_{gs(th)}(min)$ while input voltage is constant.

The objective of curve fitting is to model the irregular output voltage using concise mathematical expressions. We first present the proposed method of fitting the irregular output voltage and calculating the fitting parameters for each region. The output voltage in the falling subthreshold region can be analytically calculated as a function of time (t):

$$y = -e^{\lambda \cdot (t - \epsilon)} + constant \quad (4.4)$$

with λ and ϵ as fitting parameters. The following relationship is obtained between the falling slope of injected SET, load, and these two fitting parameters:

$$SET_{in}(FS, CL) \rightarrow \lambda, \epsilon \quad (4.5)$$

where FS stands for the falling slope, which can be derived using TF and PH. CL stands for load capacitance. This relationship can be modelled by linear regression functions with incorporated non-linear terms:

$$\log(\lambda) = a \cdot FS + b \cdot FS^c + d \cdot CL^e \cdot FS^f + constant \quad (4.6)$$

$$\epsilon = a \cdot FS + b \cdot CL^c \cdot FS^d + constant \quad (4.7)$$

It should be noted that the parameters like a and $constant$, appearing in all equations, are not identical, implying that they are local variables that are specific to each equation.

In the transition region, the output voltage decreases linearly with time and can be accurately expressed as:

$$y = a \cdot t + constant \quad (4.8)$$

where the fitting parameter a is correlated with the falling slope and the load. The non-linear relationship between the fitting parameters of these two regions, the falling slope and the CL, is identified using a feedforward neural network.

After the output voltage rises into the rising subthreshold region, its slope decreases linearly with the increment of voltage. Hence, the curve fitting can be done as follows: given the output voltage at time t , the output voltage at time $(t + \Delta t)$ is :

$$y_{(t+\Delta t)} - y_t = a \cdot \Delta t \quad (4.9)$$

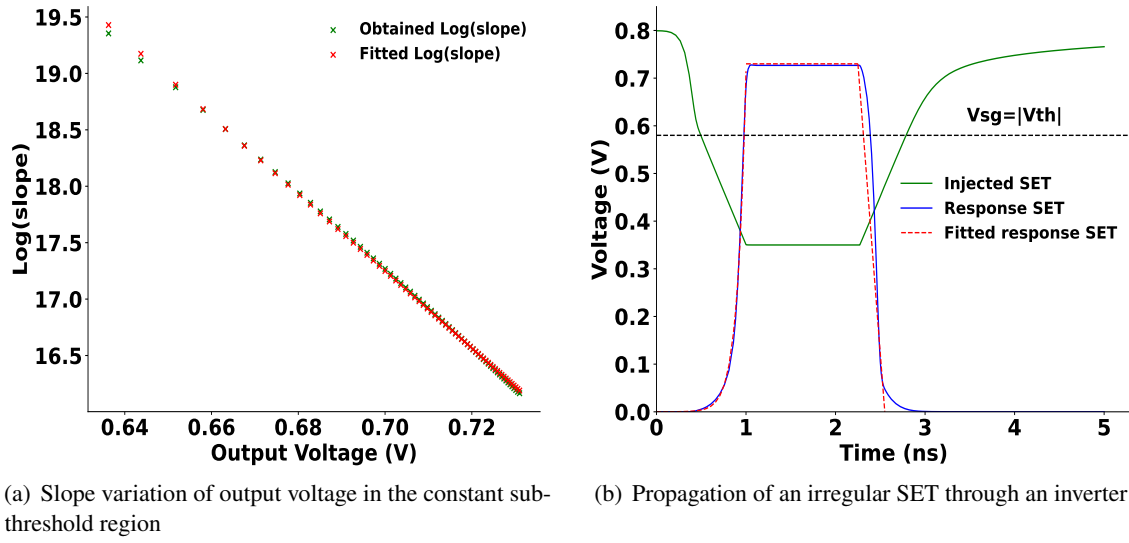


Figure 4.6.: Curve-fitting results for irregular SET propagation modeling as well as an example of irregular SETs.

where Δt is the time step and a is the slope parameter. Equation. 4.6 is used to calculate the fitted slope parameter, in which the falling slope is replaced by the rising slope. Similar to the rising subthreshold region, we fit the log-transformed slope change for the constant subthreshold region. However, the fitted slope parameter of this region depends only on the load. An example of this fitting method's outcome for the constant subthreshold region is illustrated in Fig. 4.6(a), with the green marker the obtained $\log(\text{slope})$, red marker the fitted $\log(\text{slope})$.

We thus fit the irregular waveform and illustrate the calculation of the fitting parameters. To further our analysis, we investigate how such irregular pulses propagate through an NMOS PTL or an inverter, which are the most relevant cases in PTL designs.

When irregular SETs propagate through an NMOS PTL, the output SET's fitting parameters show a strong dependency on the input SET's fitting parameters and CL. For instance, the parameters λ of the propagated SETs can be computed by Equation 4.6, and the attenuation effect can be modeled by a linear function F which includes incorporated polynomial terms:

$$SET_{out}(\log(\lambda)) = F(SET_{in}(\log(\lambda))^c, CL^f) \quad (4.10)$$

Similarly, analysis of the dependency of the fitting parameters on propagation is performed for each region. When irregular SETs propagate through an inverter, an exponential function similar to Equation. 4.4 and a linear function are demonstrated as appropriate fits for the rising and falling regions, respectively.

Algorithm 1 summarizes the proposed estimation procedure for propagated irregular SETs. We observe a direct relationship between the input PH and the maximal output PH, which can be well approximated by a SoftSign function:

$$PH_{out} = a \cdot b \cdot (PH_{in} - c) / (1 + |b \cdot (PH_{in} - c)|) + constant \quad (4.11)$$

where PH stands for pulse height, and the others are fitting parameters. Thus, the output waveform estimation, which includes piecewise fitting for three regions, can be accurately performed. A neural network is configured to identify the relationship between the input irregular fitting parameters, the output fitting parameters, and the load to simplify the modeling process. An example of the outcome of this method is shown in Fig. 4.6(b), with the green line the input SET, the blue line the response output SET, and the red dashed lines the region-based fitted output SET. Hyperparameters are listed in Table 4.2.

Algorithm 1 Estimation of shape parameters for irregular SETs propagated through an inverter.

Require: $SET_{out}(waveforms)$, $SET_{in}(waveforms)$, T

```

1:  $Rising, Falling, Plateau \leftarrow SET_{out}(waveforms)$ 
2: // Split the waveforms into three regions
3:  $OUT\_fit \leftarrow Piecewise\_fitting(Rising, Falling)$ 
4: // Fit the Rising and Falling regions
5:  $IN\_fit \leftarrow Curve\_fitting(SET_{in}(waveforms))$ 
6: // Fit the irregular SETs according to the mentioned steps
7:  $Reg\_model \leftarrow (OUT\_fit, IN\_fit, CL)$ 
8: // Perform regression on fitting parameters and load
9:  $IN\_wave \leftarrow Curve\_fitting(IN\_wave)$ 
10: // Given an irregular input SET, at first perform curve fitting
11: while  $t \in T$  do
12:   // For time point  $t$  in the time interval  $T$ 
13:    $Rising(t), Falling(t) \leftarrow Reg\_model(IN\_wave, CL)$ 
14:   if  $Rising(t) \geq PH_{out}$  then
15:     Conduct  $Plateau(t)$ 
16:   end if
17: end while
18:  $Out\_wave \leftarrow Concat(Rising(t), Plateau(t), Falling(t))$ 
19: Derive shape parameters from  $Out\_wave$ 

```

Table 4.3.: Performance of regression models for SET propagation through NMOS structures evaluated on test datasets using RMSE and R^2 metrics

Parameters	Regular through NMOS		Irregular through NMOS		Value interval
	RMSE	R^2	RMSE	R^2	
Falling $\log(\lambda)$	0.069	0.985	0.128	0.933	$(2-3) \times 10^1$
Falling ϵ	6.48×10^{-12}	0.994	8.44×10^{-12}	0.990	$(1-40) \times 10^{-10}$
Transition a	1.57×10^6	0.972	2.31×10^7	0.961	$(3.5-15) \times 10^{11}$
Rising a	2.44×10^7	0.991	7.34×10^7	0.975	
Constant a	2.12×10^7	0.982	5.23×10^7	0.984	

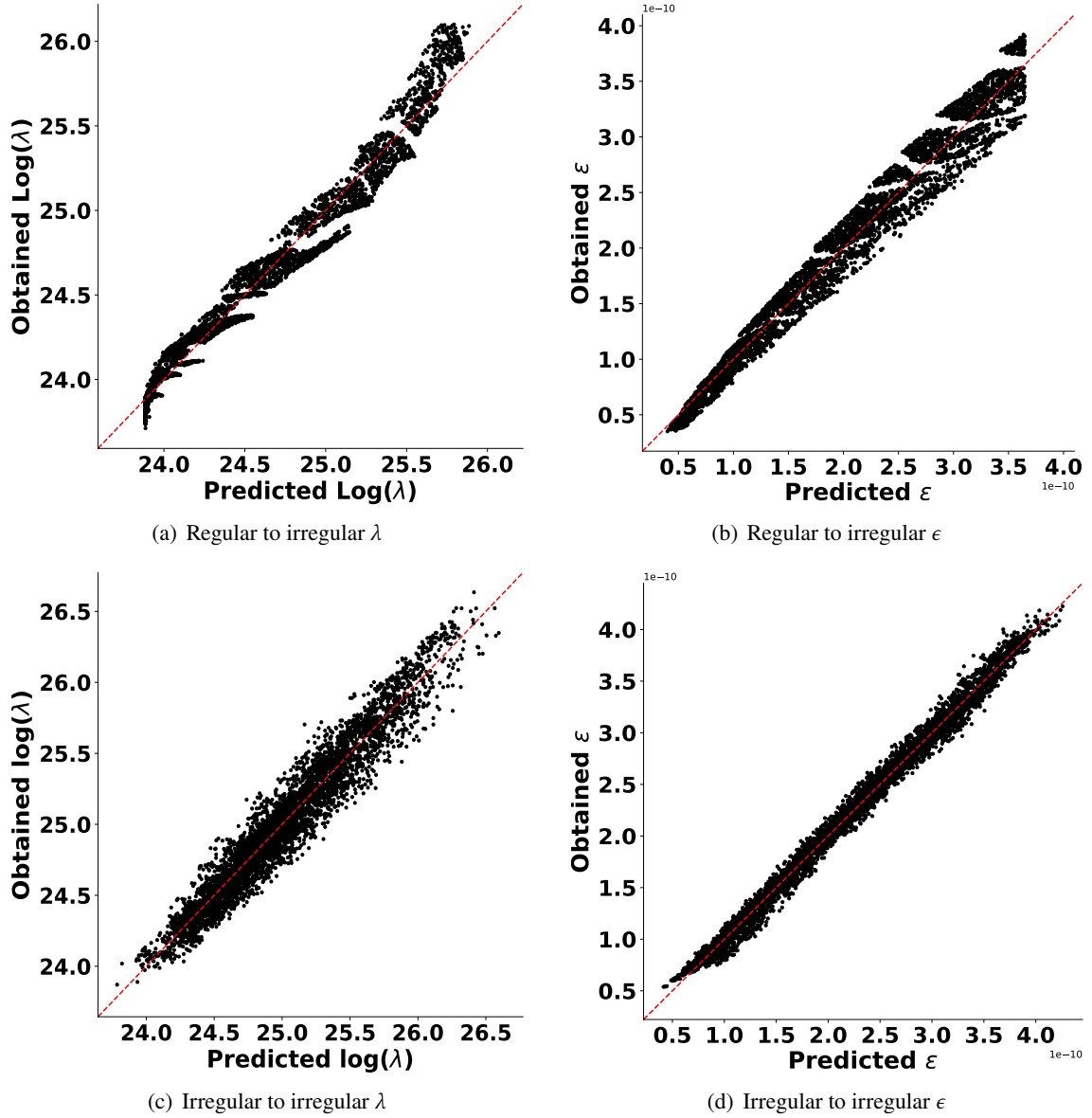
4.2.4. Evaluation of Models

Table 4.3 and Table 4.4 illustrate the Root Mean Square Errors (RMSEs) and R-squared scores obtained from the test sets of all regression cases. The first column of the table displays the fitting parameters of each region, e.g., $Falling_log(\lambda)$ is the parameter $\log(\lambda)$ of the falling region of the irregular pulses introduced in Eq. 4.6. A regression model is developed to estimate $Falling_log(\lambda)$ after regular pulses propagate through an NMOS, based on the shape parameters (PW, PH, RT, FT) of the input regular pulses and the load. In the case of irregular pulses, $Falling_log(\lambda)$ is estimated using the $\log(\lambda)$ value of the input irregular pulses.

The RMSE values are small compared with the corresponding value intervals, indicating strong predictive capability of the proposed models. Moreover, the proposed functions accurately fit all regions, as indicated by the average RMSE and R-squared values of 0.015 and 0.969, respectively. Even the worst-performing region exhibits respectable values of 0.028 and 0.945 for RMSE and R-squared. Predicted against actual values of several cases are shown in Fig. 4.7, which shows that the predicted values have good agreement with the actual values.

Table 4.4.: Performance of regression models for SET propagation through inverter structures evaluated on test datasets using RMSE and R^2 metrics

Parameters	RMSE	R^2	Value interval
Inverter $\log(\lambda)$	0.088	0.979	$(1-1.7) \times 10^1$
Inverter ϵ	7.62×10^{-12}	0.975	$(1-28) \times 10^{-10}$
Inverter a	4.33×10^7	0.986	$(3-12) \times 10^{11}$

**Figure 4.7.:** Predicted versus actual fitting parameters in the falling subthreshold region. (a)-(b) correspond to regular-to-irregular propagation cases, while (c)-(d) correspond to irregular-to-irregular propagation through a single NMOS structure.

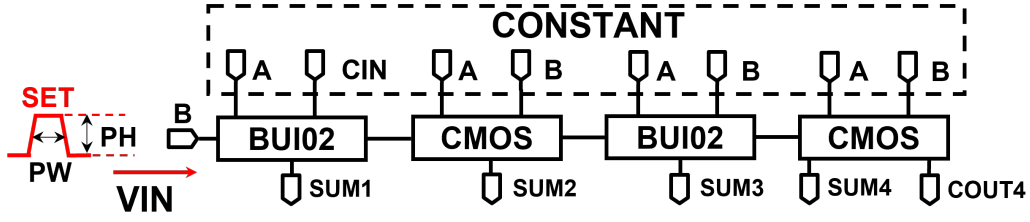


Figure 4.8.: 4-bit RCA used as the DUT for validation.

4.3. Results and Discussion

4.3.1. Simulation Setup for Validation

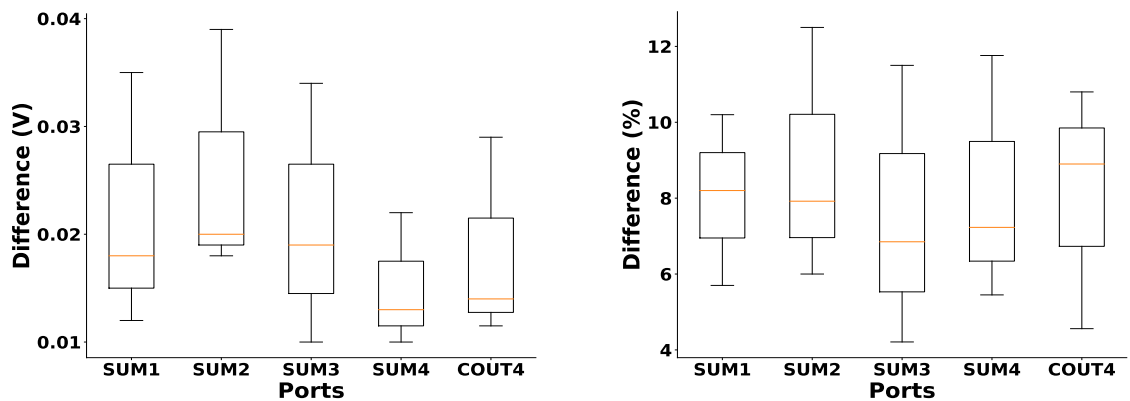
A 4-bit RCA consisting of BUI02 PTL full adders FAs [64] and CMOS full adders is considered as the DUT, as shown in Fig. 4.8. Each PTL FA stage is followed by a CMOS FA stage. We run circuit-level simulation to inject 10000 regular transient SETs with randomly generated shape parameters into the "B" port of the least significant bit BUI02 FA, which is selected as the error-injection FA. An input combination is selected such that the injected SETs can propagate through the entire RCA without logical masking. The output waveforms and propagation delays at all ports are recorded and compared with the results predicted by the proposed framework to verify the accuracy. The simulation setup used for validation is identical to the setup used during training dataset generation, as summarized in Table 4.1.

4.3.2. Accuracy Comparison and Discussion

In this paper, waveforms are represented as a function of time. The mean absolute error (MAE) is utilized to characterize the difference between each waveform predicted by the models and obtained from the simulation, which is defined as :

$$MAE = \frac{\sum_{t=1}^n |Wave_{pre}(t) - Wave_{sim}(t)|}{n} \quad (4.12)$$

where n is the number of discrete time points, $Wave_{pre}(t)$ and $Wave_{sim}(t)$ are the predicted and simulated voltage values of a waveform at a time point t , respectively. The minimum, maximum, and average



(a) Minimum, maximum, and average waveform prediction MAE.

(b) Minimum, maximum, and average delay prediction MAE.

Figure 4.9.: Comparison between predicted and simulated waveform and delay errors.

waveform prediction errors at each output port are illustrated in Fig. 4.9(a), and the average waveform prediction MAE remains around 15 mV. Moreover, the maximum propagation-delay prediction error is approximately 12.5%, while the average error remains around 8%.

The prediction error is mainly attributed to the approximation of MOSFET loads as equivalent capacitive loads, as well as inaccuracies introduced by curve fitting and regression modeling. In addition, the current framework does not explicitly model SET reconvergence scenarios, where multiple transient pulses simultaneously propagate through different transistor paths. Overall, the proposed framework demonstrates strong agreement with circuit-level simulations while significantly reducing the complexity of SET propagation analysis in PTL circuits.

4.4. Conclusion

In this chapter, a learning-based structural SET propagation framework for PTL circuits was proposed. Unlike conventional gate-level black-box prediction approaches, the proposed methodology models transient propagation behavior using reusable transistor-level PTL building blocks, enabling compositional SET analysis for complex PTL structures. By combining curve-fitting techniques with machine-learning-based prediction models, the proposed framework can efficiently capture distorted transient propagation behavior and long-tail SET waveforms frequently observed in PTL circuits. The accuracy of the proposed methodology was validated through extensive circuit-level simulations while significantly reducing simulation complexity compared with conventional SPICE-based approaches.

5. Impact of Radiation and Aging on DVFS in 22nm FDSOI Latch

FDSOI technology has emerged as an important alternative to conventional bulk CMOS technologies for safety-critical automotive applications due to its improved electrostatic control, reduced variability, and enhanced low-voltage operation. These characteristics are particularly beneficial for advanced digital systems that contain large numbers of sequential storage elements, such as latches and flip-flops, where reliability and energy efficiency are critical.

Previous chapters focused primarily on SET propagation behavior in combinational logic. However, modern computing systems also contain large numbers of sequential storage elements, where radiation-induced transients can propagate to storage nodes and result in SEUs or timing failures. Since the susceptibility of sequential circuits to radiation-induced faults strongly depends on latch timing margins, accurate functional correctness analysis becomes increasingly important for reliability evaluation. The maximum operating frequency of a circuit is determined by the delay along its critical path. This delay consists of the logic cloud delay and the setup time (T_{setup}) of the capture latch. Setup time, defined as the interval during which data must remain stable at the latch input before clock arrival, plays a pivotal role in ensuring correct circuit functionality. Consequently, understanding and optimizing T_{setup} is fundamental for improving the performance and reliability of sequential circuits utilizing FDSOI technology. Such reliability challenges must be carefully addressed as part of SLM infrastructures, particularly in safety-critical applications such as automotive systems, where strict functional safety requirements and near-zero defective parts per million (DPPM) targets are mandated by standards such as ISO 26262 [83]. At the same time, accurately modeling radiation-induced transient currents in advanced FDSOI technologies remains challenging because charge collection behavior strongly depends on technology-specific device properties such as doping profile, electric-field distribution, and device geometry. Therefore, physics-based TCAD simulations become increasingly important for accurately capturing technology-dependent transient current behavior and enabling reliable circuit-level radiation analysis.

In advanced circuits, NBTI also poses a significant reliability concern due to its impact on transistor V_{th} . This degradation mechanism mainly affects PMOS transistors and progressively degrades circuit performance over time. In latch structures, NBTI-induced threshold voltage shifts can alter switching behavior and timing margins, thereby modifying the vulnerability window in which radiation-induced transients may cause functional failures.

Consequently, aging-induced timing degradation and radiation-induced transient faults cannot be analyzed independently in advanced sequential circuits, since both mechanisms jointly determine the timing vulnerability of sequential storage elements. While numerous studies have investigated the effects of NBTI and radiation individually, their interaction in FDSOI latch structures remains insufficiently explored. Therefore, this chapter presents a detailed analysis of widely used latch designs under NBTI and radiation stress in 22 nm FDSOI technology, based on the work published in [3]. By calibrating TCAD models and developing corresponding circuit-level models, this work investigates how NBTI-induced threshold voltage shifts and radiation-induced transient currents affect latch timing margins. Furthermore, voltage, sizing, and frequency adjustments are explored as potential strategies to improve latch robustness under aging and radiation. The main contributions of this chapter are summarized as follows:

1. Calibration of TCAD models to accurately capture the characteristics of GlobalFoundries' 22 nm FDSOI transistors, followed by NBTI simulations for PMOS devices and development of a corresponding ΔV_{th} model.
2. Development of radiation current models that enable more accurate incorporation of low- and high-energy radiation effects into circuit-level simulations.
3. Integration of the NBTI model with the radiation current model in circuit-level simulations to perform detailed reliability analysis of latch structures, with particular focus on T_{setup} as a key timing metric.
4. Comprehensive analysis of latch robustness under variations in supply voltage (V_{DD}) and transistor sizing, together with design insights for improving latch reliability through strategic adjustment of sizing and V_{DD} .

The remainder of this chapter is organized as follows. Section 5.1 introduces existing NBTI and radiation current models. Section 5.2 presents the proposed methodology, including TCAD calibration, NBTI simulation, radiation simulation, and model development. Section 5.3 evaluates latch T_{setup} under NBTI, radiation, and combined reliability stress, as well as the influence of supply voltage and transistor sizing on latch vulnerability. Finally, Section 5.4 summarizes the chapter.

5.1. Preliminaries

This section briefly reviews the background and related work on NBTI degradation modeling and radiation-induced heavy-ion current modeling in advanced FDSOI technologies. Since both NBTI-induced threshold voltage degradation and radiation-induced transient currents strongly affect latch timing behavior and reliability, accurate physical modeling of these effects is essential for the combined reliability analysis presented in this chapter.

5.1.1. NBTI Models

NBTI poses a major challenge to the reliability of p-channel MOSFETs. It leads to a progressive accumulation of positive charges in the gate insulator when the gate is negatively biased at high temperatures, which leads to changes in the ΔV_{th} , transconductance, and subthreshold slope over time [84]. Several studies have been conducted on NBTI modeling. In [85], a thermionic double-well model is presented. The change of the interface trap (ΔN_{IT}) is calculated according to a first-order thermionic reaction between the H-bound ground state and the H-transport state. A framework based on an MSC model is proposed [49], which integrates passivation–depassivation of bonds via a reaction-diffusion (RD) model and trapping and detrapping of holes. We use this framework to perform NBTI simulations in TCAD. Details will be explained in the next section.

5.1.2. Heavy Ion Current Models

When a heavy ion hits a transistor, this leads to a charge accumulation and thus to a transient current at the affected circuit node. Several models have been developed to simulate the processes of charge generation, charge transport, and charge accumulation and to accurately calculate the electrical response. The double exponential current model [21] has been proven to be an efficient model that models the current as a process of prompt charge accumulation and diffusion charge accumulation sequentially. However, this model does not account for the bipolar amplification effect in FDSOI. Therefore, the author introduces a TCAD-based current model in [86] that comprehensively integrates the effects of the lowering of the

barrier and thus facilitates the flow of charge carriers from the source to the drain. This process leads to the generation of an additional current, identified as bipolar amplification. However, the implementation of this model requires a high level of expertise in semiconductor physics. Therefore, we derive the current model from TCAD simulations and use it in the circuit-level simulation to accurately represent the radiation current induced while reducing the complexity of the modeling process.

5.2. Methodology

In this work, a detailed analysis of four widely used latches [87], [88], [89] as shown in Fig. 5.1, is presented, specifically considering the combined effects of NBTI and radiation. In this section, we introduce our proposed method for TCAD-based NBTI and radiation analysis. The major steps include:

- *TCAD Calibration:* The FDSOI models are constructed using TCAD and calibrated against the process design kit (PDK) model. This calibration process includes a detailed comparison of the

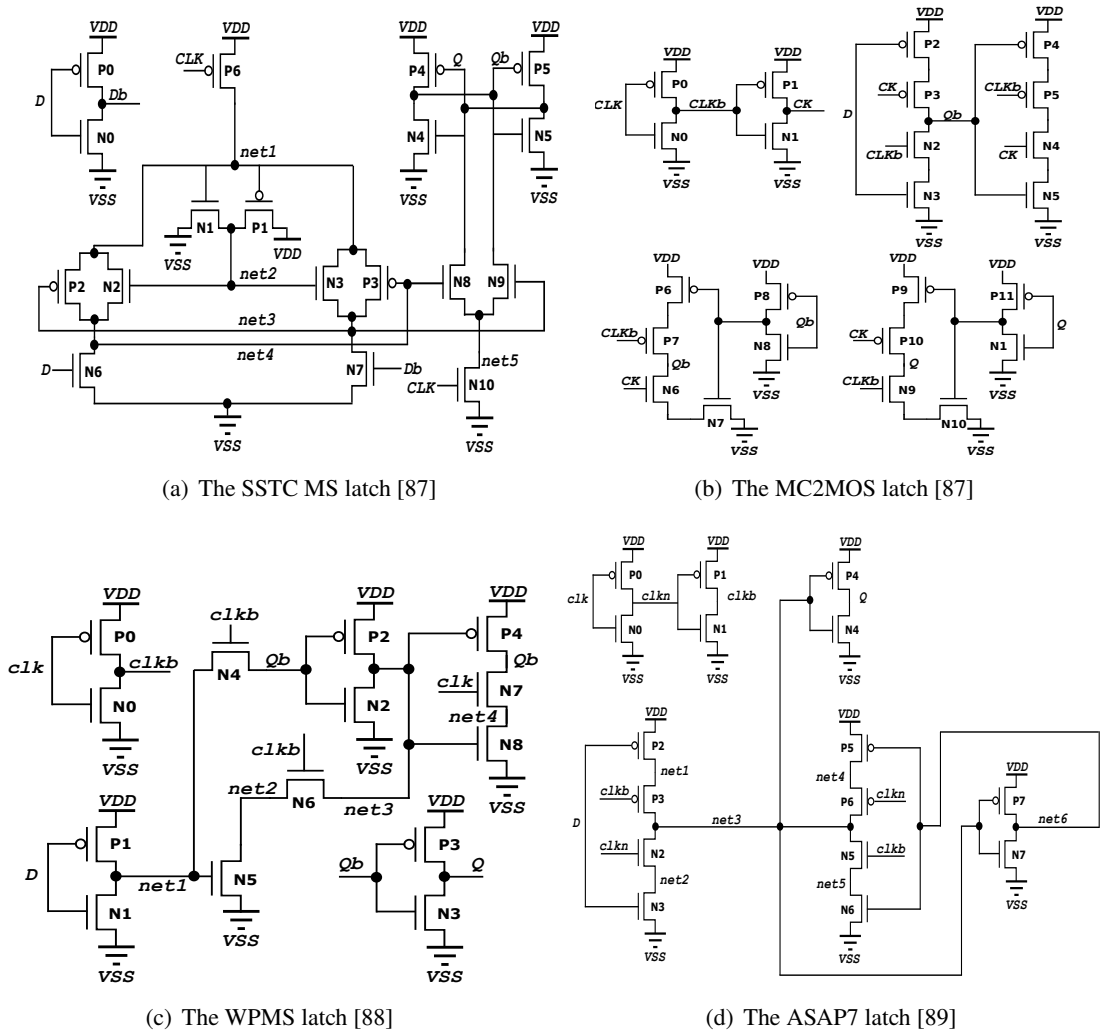


Figure 5.1.: Schematics of the considered latches under test. SSTC MS: static single-transistor-clocked master-slave latch, WPMS: write-port master-slave latch, ASAP7: latch in Arizona State University's 7-nm FinFET predictive PDK, MC2MOS: modified clocked CMOS logic latch.

drain current versus drain voltage ($I_D - V_D$) and drain current versus gate voltage ($I_D - V_G$) curves to ensure accuracy and consistency with the PDK's specifications.

- *Building the NBTI Model:* We model the ΔV_{th} due to NBTI over time under constant stress, examining the impact of V_{dd} and temperature on ΔV_{th} .
- *Building the Radiation Current Model:* First, we perform TCAD simulations to identify the MOSFET's radiation-sensitive region by varying strike locations, angles, and operating conditions. These results are then converted into a current source for circuit-level simulations.
- *Model Exploitation:* We perform tests on latches to assess the impacts of NBTI and heavy ion radiation on T_{setup} , aiming to comprehend their combined effect on latches' reliability and to strategically adjust size and V_{dd} .

5.2.1. TCAD Calibration

In this work, NMOS and PMOS devices are calibrated utilizing the Sentaurus Structure Editor, and later simulated via the Sentaurus Device Simulator [90]. This process aims to align their specifications with those outlined in GlobalFoundries' PDK [91]. The calibration is conducted at a temperature of 300 K, with a transistor width-to-length (W/L) ratio of 80 nm/20 nm, which represents the minimum size characterized within the PDK, ensuring the devices' compliance with industry standards. Moreover, the calibration between the TCAD simulations and the PDK data is performed by adjusting various geometrical and doping parameters to achieve a qualitative match suitable for digital logic applications. The parameters for the TCAD models are listed in Table 5.1. In addition, the drift-diffusion model is used in the simulation. Consequently, a comparison of the transistor output characteristics between the TCAD and PDK models is shown in Fig. 5.2(a), while the structure of the FDSOI is depicted in Fig. 5.2(b).

5.2.2. Modelling NBTI Effect

The NBTI effect is modeled using the Sentaurus device simulator, integrated with the RD model to calculate the interface trap generation using the Capture Emission Depassivation model referenced in [49]. The RD model concept is depicted in Fig. 2.5. Moreover, stress duration in the simulation ranges from 0 to 3 years, under conditions of nominal stress ($V_{DD} = 0.8V$, $T = 298.15K$) and high stress ($V_{DD} = 0.96V$,

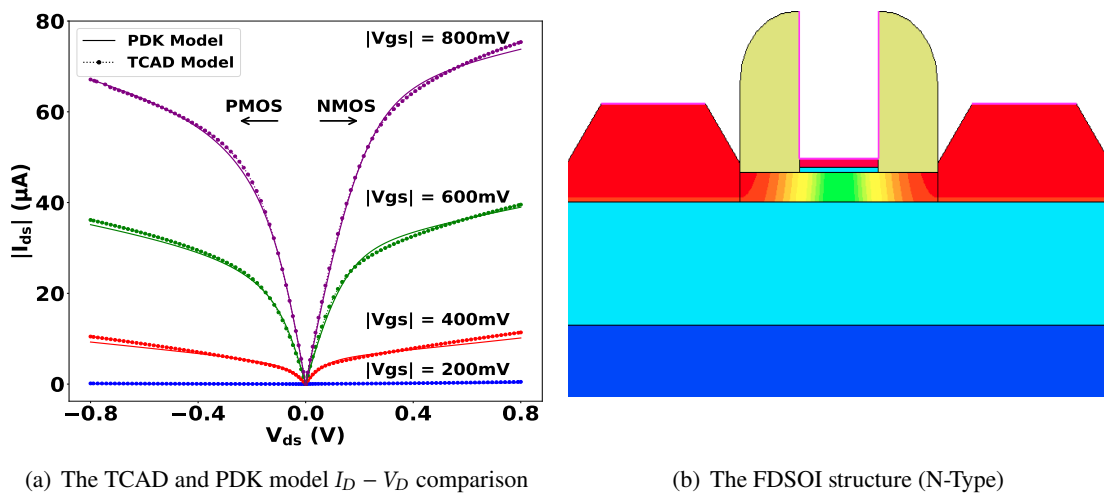


Figure 5.2.: Output characteristics of the TCAD and PDK MOSFETs and the proposed raised source/drain FDSOI structure.

Table 5.1.: FDSOI parameters

Parameters	NMOS	PMOS
Gate Width / Length	80nm / 20nm	
Effective Gate Thickness	1.25nm	
Channel Thickness	6nm	
Box Thickness	25nm	
Body Material	Silicon	$Si_{0.75}Ge_{0.25}$
S/D Doping	$4e+20 \text{ cm}^{-3}$	
Substrate Doping	$1e+18 \text{ cm}^{-3}$	
Channel Doping	Undoped	
Work Function	4.45eV	4.62eV

$T = 398.15K$), as illustrated in Fig. 5.3(a). The simulation initiates with a specified number of intact Si–H bonds at the channel-gate oxide interface and employs the charge trapping model in both SiO₂ and HfO₂ layers, following the setup in [49].

In addition, ΔV_{th} is quantified by the I_D - V_G test, where V_{DS} is set to half the value of V_{DD} , i.e. $V_{DS} = 0.5 * V_{DD}$. A Verilog-A model is used to represent ΔV_{th} as a stress time-dependent voltage source. This model not only facilitates circuit-level simulations but also interpolates ΔV_{th} values for cases where the corresponding stress time is not available. To assess the worst-case degradation of a PMOS device, a high-stress scenario with no recovery is considered and then implemented in the circuit-level simulations.

5.2.3. Modelling Radiation Current

As mentioned in the previous section, our first investigation focuses on the sensitivity of the transistor to radiation. This analysis employs a heavy ion model to emulate radiation effects, characterized by a track length of $1 \mu\text{m}$ and a Gaussian spatial distribution with a characteristic width of 10 nm. Two different particle energies are selected for this study: $10 \text{ MeV cm}^2\text{mg}^{-1}$ and $40 \text{ MeV cm}^2\text{mg}^{-1}$, which are referred

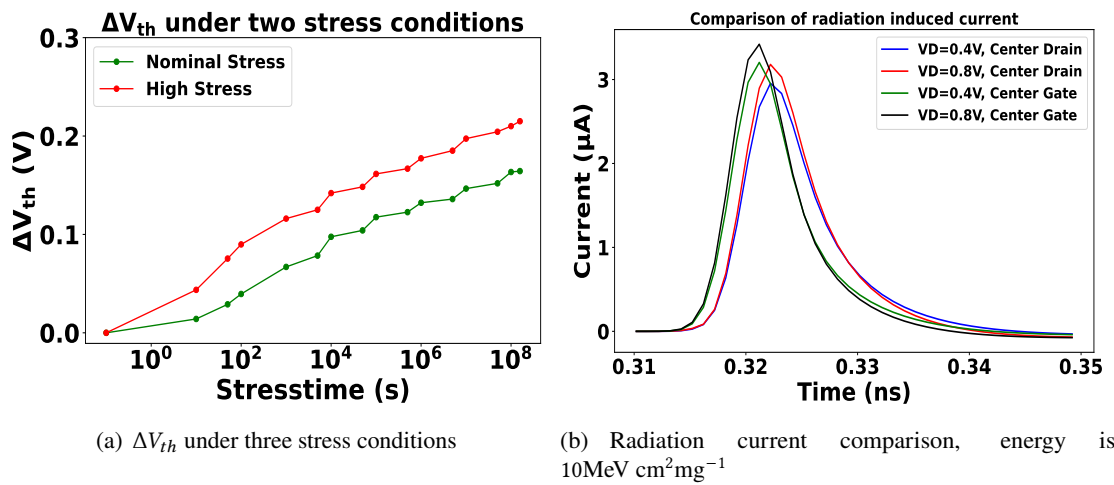


Figure 5.3.: Analysis of ΔV_{th} under three stress conditions and variation of radiation-induced current across diverse locations and operating conditions.

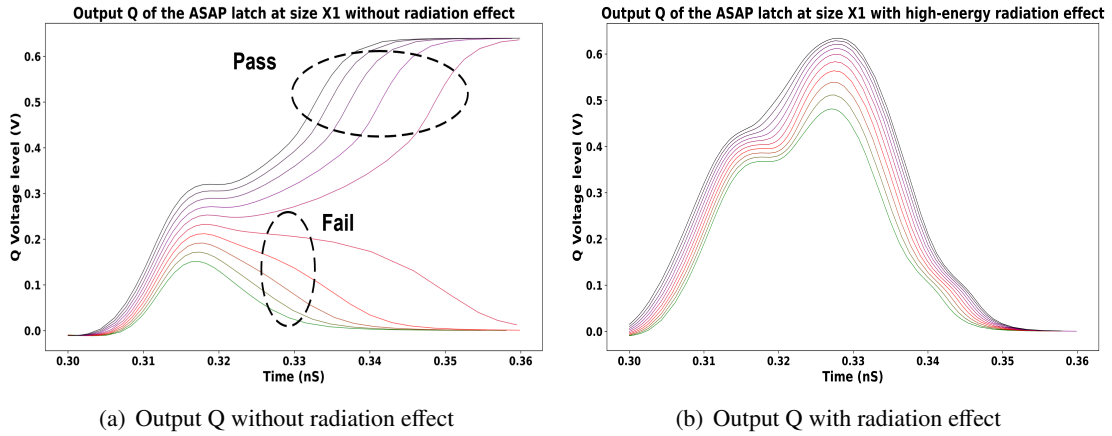


Figure 5.4.: Analysis of T_{setup} violation for the ASAP latch at size X1, studied under conditions without and with high-energy ion injection at N7. Initial delays for CLK and DAT are 300ps, with DAT shifted to the CLK edge in 0.4ps increments.

to as low-energy and high-energy particles, respectively. These energies are representative of critical particles, such as alpha particles, which are important for FDSOI technologies [92]. Furthermore, the investigation includes an analysis of different strike locations within the device, as well as a variation of operating conditions. The results show that the radiation current varies only slightly in the different scenarios, as shown in Fig. 5.3(b). Consequently, the largest observed current is selected and converted into a current source specifically for circuit-level simulations, enabling a precise evaluation of the circuits under worst-case radiation scenarios.

5.2.4. Setup Time Simulation

The T_{setup} , defined as the interval during which data (DAT) must remain stable before the clock (CLK) arrives at a latch, is challenging to model due to the latch's near-metastable behavior. We use circuit-level simulations to assess T_{setup} and its sensitivity to size, V_{dd} , NBTI, and radiation. As shown in Fig. 5.4(a), as DAT shifts towards the CLK edge, the latch fails to maintain a switch to logic '1' and reverts to '0', as prior data at Q prevails. Thus, T_{setup} is the critical interval between CLK and DAT where Q degradation begins, causing functional failure.

5.3. Model Implementation and Analysis

In this study, we seamlessly integrate the proposed NBTI and radiation-induced current models into the evaluation framework for digital latch circuits. Our investigation primarily focuses on the latch t_{setup} to evaluate the accuracy and operation efficiency of the digital circuit. The evaluated latches are shown in Fig. 5.1, and the individual load is selected based on the fan-out-of-four sizing method to ensure performance. The objectives of our comprehensive test include:

- Adjusting the transistor size within each latch to ensure correct operation of the latch.
- Evaluating the performance of the latches with particular attention to t_{setup} under nominal V_{DD} as well as under variations of $\pm 20\%$.
- Investigating the influence of the transistor size on the latch performance by systematically doubling the transistor size in each latch and the resulting influence on t_{setup} .

- Performing separate NBTI and radiation tests to evaluate the impact of both low- and high-energy radiation, in addition to varying stress duration, on the devices.
- Performing analysis of NBTI and radiation testing to determine the combined effects of aging on the heavy ion resistance of the selected latches. This integrated evaluation also examines the effects of V_{DD} and the size of the latches on the resistance of the latches to radiation.

We present the findings of our investigation into latch performance and resilience. Initially, we elucidate the T_{setup} for fresh latches, with a focus on analyzing the impact of V_{DD} and transistor size on T_{setup} . Subsequently, we assess the radiation resilience of each fresh latch variant by comparing T_{setup} under conditions of ion strikes at the most susceptible nodes in each latch. Additionally, the investigation encompasses the effects of aging on latch performance. This includes an analysis of the impact of two aging mechanisms: the aging associated with storing a logic ‘0’ and the aging associated with storing a logic ‘1’. To offer comprehensive insights into how aging influences a latch’s resilience to radiation, we also conduct radiation tests on aged latches. Through this systematic approach, we aim to provide a detailed understanding of the variables affecting latch performance and durability, particularly concerning T_{setup} .

5.3.1. Simulation Setup

The circuit-level simulation is conducted at a temperature of 300 K under the typical-typical corner. We choose a transistor W/L ratio of 160 nm/20 nm as the minimum size (X1) for transistors to ensure a higher noise margin compared to the PDK’s minimum ratio of 80 nm/20 nm. To determine the size of the transistors in a latch, we apply the fan-out-of-4 method and also improve the driving ability by increasing the size of the output transistors. For example, as shown in Fig. 5.1(d), the transistors N4 and P4 are sized as 320 nm/20 nm, while other transistors maintain a size of 160 nm/20 nm. The size of the load inverter is increased to 640 nm/20 nm. In addition, the rise time for *DAT* and *CLK* signals is set to 15 ps, and the simulation time is 1 ns. The T_{setup} is subsequently assessed using the methodology described in Section 5.2.4.

5.3.2. Fresh Latch Analysis Excluding Radiation Impact

We investigate the impact of transistor size and V_{DD} on the T_{setup} of fresh latches, without taking radiation into account. Fig. 5.5 shows that increasing the transistor size can significantly reduce the T_{setup} of the WPMS latch, especially at lower V_{DD} . For example, doubling the transistor size shortens the T_{setup} of the WPMS latch by 31% (from 74.2 ps to 51.2 ps) at $V_{DD} = 0.64$ V, and by 15.6% (from 40.5 ps to 35.4 ps) at $V_{DD} = 0.96$ V. However, this effect is less pronounced in other latches. For instance, the ASAP latch’s T_{setup} decreases by 7.7% at $V_{DD} = 0.64$ V and 6% at $V_{DD} = 0.96$ V. Additionally, an increase in V_{DD} reduces T_{setup} , especially for smaller transistors, with the WPMS latch having the highest sensitivity to V_{DD} changes. At X1 size, its T_{setup} is 45% lower at $V_{DD} = 0.96$ V than at 0.64 V.

5.3.3. Fresh Latch Analysis Including Radiation Impact

Further, we explore the radiation effects on fresh latches, focusing on the influence of transistor size and V_{DD} on their radiation resilience. We inject a radiation-induced current 1 ps after the *CLK* has stabilized after the transition. Notably, the SSTC latch exhibits the highest sensitivity to radiation, which can only recover from low-energy radiation at $V_{DD} = 0.96$ V at size X1, with T_{setup} increasing by 22.4%. It fails to recover from high-energy radiation at both X1 and X2 sizes. In contrast, at size X1, the other three latches can recover from low-energy radiation across all V_{DD} with a negligible increase in T_{setup} , and can only recover from high-energy radiation at high V_{DD} .

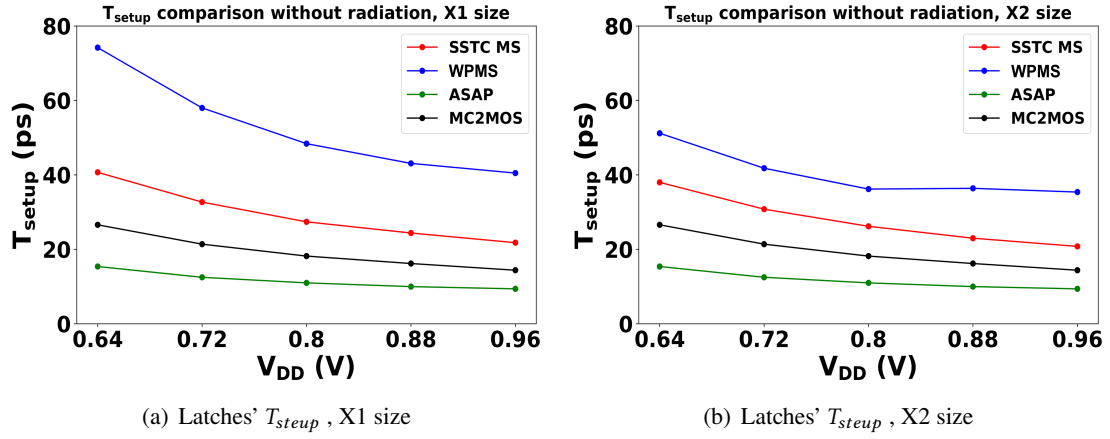


Figure 5.5.: Comparison of fresh latches' T_{setup} without radiation.

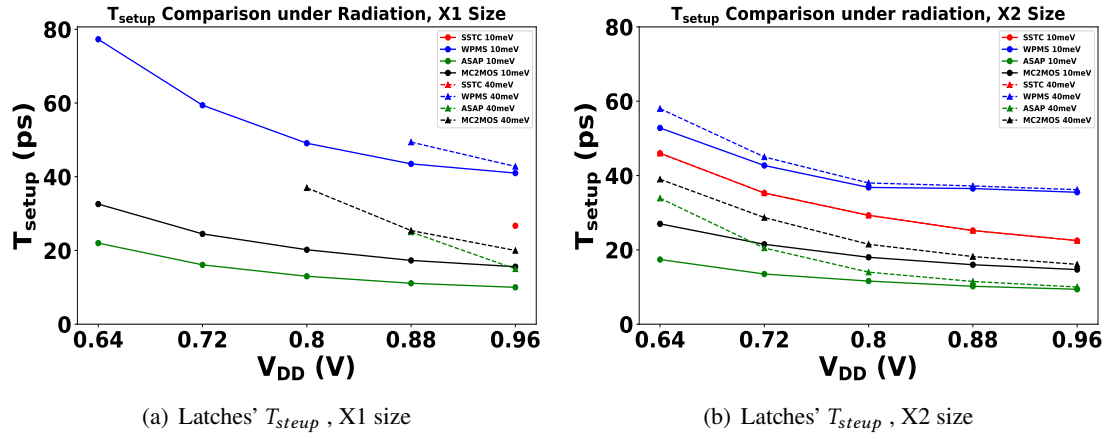


Figure 5.6.: Comparison of fresh latches' T_{setup} under radiation. Missing data points indicate the latch fails to retain data.

At size X2, the WPMS latch shows a slight variation in T_{setup} under both low- and high-energy radiation, in contrast to the ASAP and MC2MOS latches, which are more susceptible to high-energy radiation at lower V_{DD} , as illustrated in Fig. 5.6(b). For example, with high-energy radiation at size X2, the T_{setup} for the ASAP latch increases to 33.9 ps, which is a significant increase of 94.8% and 120.1% compared to its low-energy radiation and without radiation conditions, respectively.

5.3.4. Aged Latch Analysis Excluding Radiation Impact

In our subsequent study, we evaluate the influence of NBTI on the T_{setup} for latches, excluding the radiation effect. Two distinct aging mechanisms are considered, one associated with the storage of logic '0' and one associated with the storage of logic '1'. First, we identify the transistors in each latch that are affected by each aging phenomenon. For instance, in the ASAP latch as shown in Fig. 5.1(d), transistors P1, P2, P5, and P6 are susceptible to NBTI when storing logic '0', while transistors P1, P4, P6, and P7 are sensitive when storing logic '1'. Subsequently, the V_{th} of the affected transistors is modified using the previously described NBTI model, in which two predetermined periods of constant stress are considered: short-term stress (1e4 seconds) and long-term stress (1e8 seconds).

The latches demonstrate different responses to the two aging mechanisms. In particular, the transistor size shows a smaller influence on T_{setup} compared to V_{DD} under the stress of storing '0', as shown in

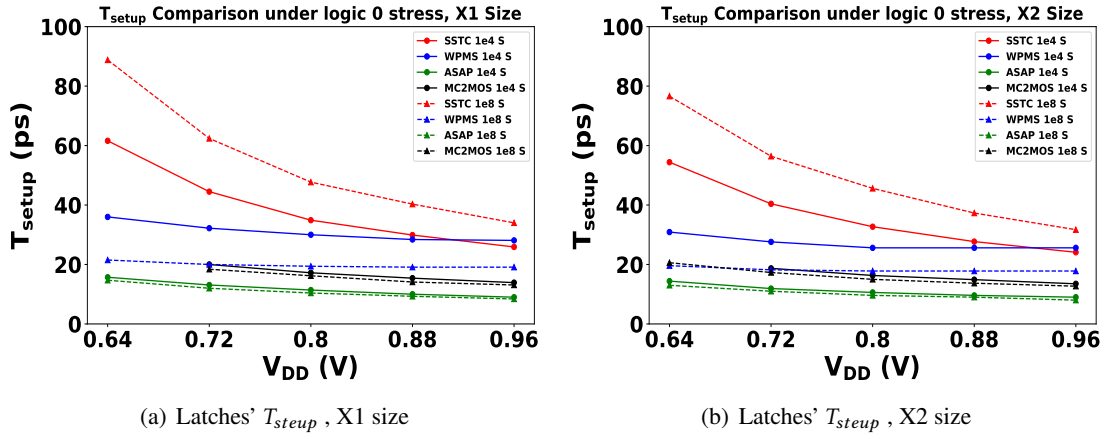


Figure 5.7.: Latches' T_{setup} under logic '0' stress without radiation.

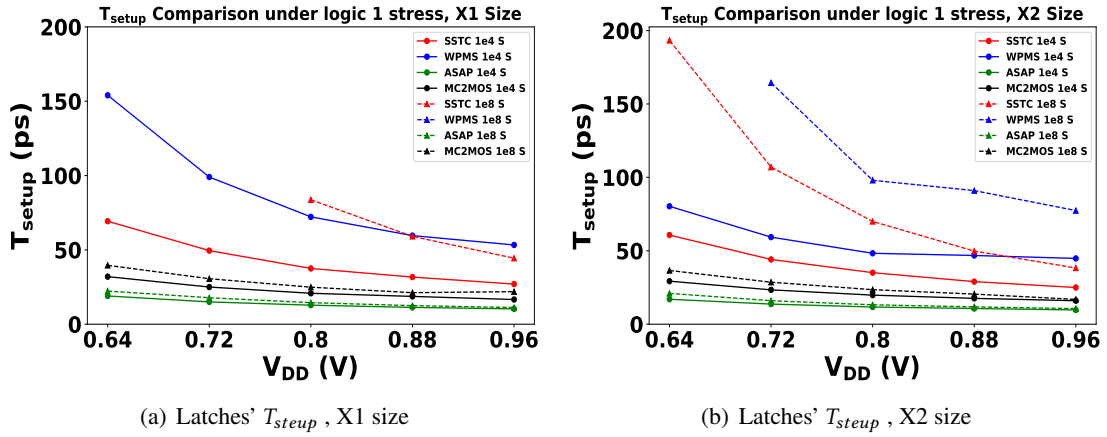
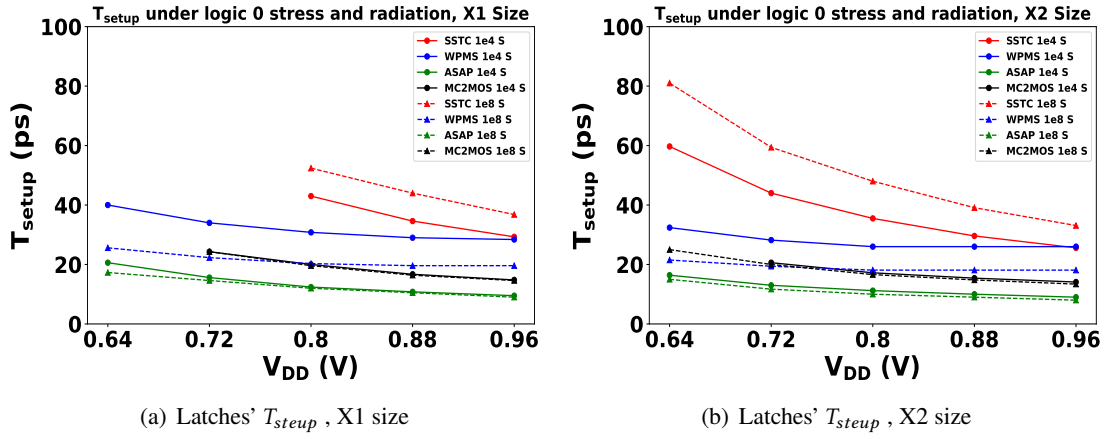
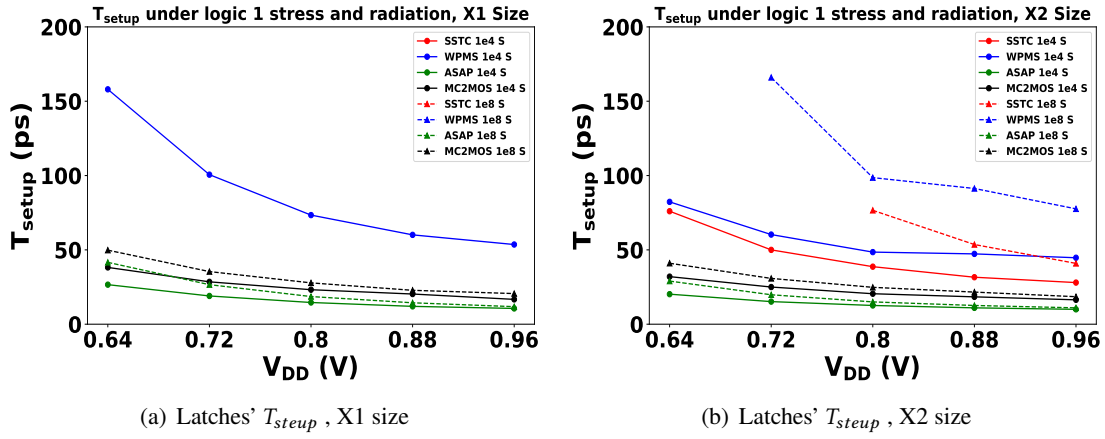


Figure 5.8.: Latches' T_{setup} under logic '1' stress without radiation.

Fig. 5.7(a) and 5.7(b). Storing '0' reduces the switching speed of the SSTC latch, but accelerates that of the other three latches. For the ASAP and MC2MOS latches, however, the effect of storing '0' aging is negligible. For both transistor sizes, the MC2MOS latch cannot retain the data at lower V_{DD} levels. Moreover, at $V_{DD} = 0.64$ V, the SSTC latch shows an increase in T_{setup} of 51.4% under short-term stress and 118.1% under long-term stress at size X1, with corresponding increases of 43.2% and 101.6% at size X2 compared to the fresh device. Conversely, storing '0' significantly reduces the T_{setup} of the WPMS latch, especially at lower V_{DD} levels. For example, its T_{setup} decreases by 71.0% and 52.8% at $V_{DD} = 0.64$ V and 0.96 V at size X1 respectively.

In contrast to storing '0' stress, the transistor size can significantly influence the T_{setup} under storing '1' stress, as shown in Fig. 5.8(a) and 5.8(b). While storing '1' slows down all latches, it has a negligible effect on the ASAP and the MC2MOS latch. However, the WPMS latch is highly sensitive to storing '1' and fails at all V_{DD} levels for long-term stress at size X1. In particular, at lower V_{DD} , its T_{setup} increases significantly (107.5% at size X1 and 56.8% at size X2) compared to fresh conditions, suggesting that larger transistor sizes can mitigate the aging effects of storing '1'. In addition, the T_{setup} of the SSTC latch increases more significantly under long-term storing '1' stress, especially at lower V_{DD} , with an increase of 152% at size X2 and $V_{DD} = 0.64$ V to storing '0' stress, highlighting the pronounced effect of storing '1' stress.

Figure 5.9.: Latches' T_{setup} under logic '0' stress with radiation.Figure 5.10.: Latches' T_{setup} under logic '1' stress with radiation.

5.3.5. Aged Latch Analysis Including Radiation Impact

In further tests, we evaluate the effects of aging on the radiation resilience of all tested latches. Our results show that aging can significantly increase the sensitivity of all latches to the effects of radiation. In particular, only the MC2MOS latch shows the ability to recover from high-energy radiation across all V_{DD} levels after being stressed for a long time at size X2. Therefore, we present the results of tests with low-energy radiation.

As shown in Fig. 5.9 and Fig. 5.10, the SSTC latch is most sensitive to low-energy radiation under storing '0' stress and storing '1' stress and cannot recover at size X1 in the scenario of storing '1' across all V_{DD} levels. Compared to its fresh state (Fig. 5.6(a)), it is noteworthy that the stress of storing '0' allows recovery from radiation at higher V_{DD} levels. At size X2, storing '1' stress worsens the radiation sensitivity of the SSTC latch, especially at lower V_{DD} levels. For example, for short-term stress, its T_{setup} increases by 65.2% (from 46 ps to 76 ps) with storing '1' stress and by 23.9% (from 46 ps to 59.7 ps) with storing '0' stress, relative to its fresh radiation response.

The MC2MOS and WPMS latches display only minor sensitivity to the combined effects of aging and radiation. Comparing Fig. 5.7 to Fig. 5.10, most of the changes in T_{setup} are due to aging rather than radiation. In addition, the ASAP latch exhibits significant radiation sensitivity only with storing '1' stress at low V_{DD} levels. At size X1 and $V_{DD} = 0.64$ V under long-term stress, its T_{setup} increases by 89.1%

(from 22 ps to 41.6 ps), while storing ‘0’ stress leads to a T_{setup} reduction of 21.4% (from 22 ps to 17.3 ps), compared to its fresh state with radiation.

5.4. Conclusion

Radiation has a significant impact on latch timing and affects dynamic voltage and frequency scaling, especially for latches that are smaller and operate at a lower supply voltage. Among the tested latches, the MC2MOS latch shows the most consistent performance and proves resistant to the effects of Negative Bias Temperature Instability, radiation, and their combined effects. The radiation resistance of the ASAP latch improves when subjected to ‘0’ storing stress. However, it shows lower reliability under ‘1’ storing stress. Furthermore, improving the transistor size and increasing the supply voltage of both the SSTC and the WPMS latch can significantly increase their resistance to both NBTI and radiation effects. This observation underscores a crucial tactic for designers seeking to improve the reliability of digital circuits with these latch designs.

6. Addressing the Combined Effect of Transistor and Interconnect Aging in SRAM towards Silicon Lifecycle Management

The long-term reliability of SRAM, as a key component in modern computing architectures, is becoming increasingly important for safety-critical applications such as automotive systems. In advanced technology nodes, both FEOL transistors and BEOL interconnects are affected by aging mechanisms that degrade circuit reliability over time. In the FEOL, NBTI causes threshold voltage (V_{th}) degradation in PMOS transistors, leading to increased delay and reduced switching performance. In the BEOL, EM gradually increases interconnect resistance due to metal atom migration caused by high current density and temperature. As device dimensions continue to scale and integration density increases, the combined impact of NBTI and EM becomes increasingly critical for SRAM reliability and writability.

Existing SRAM reliability studies mainly focus on individual degradation mechanisms in isolation. However, in advanced scaled technologies, SRAM reliability increasingly depends on the interaction between transistor aging and interconnect degradation. NBTI-induced threshold voltage shifts degrade write-driver strength and increase switching delay, while EM-induced resistance degradation further weakens voltage delivery through scaled bit-line and word-line interconnects. As a result, the combined effects of FEOL and BEOL aging mechanisms can considerably aggravate SRAM writability degradation and timing failures. For safety-critical applications, reliable aging analysis is an essential part of SLM infrastructures and requires analysis of the combined impact of FEOL transistor aging and BEOL interconnect degradation on SRAM operation, together with effective aging-aware testing, monitoring, and mitigation strategies.

This chapter is based on the work published in [4]. It investigates the combined impact of NBTI and EM degradation on SRAM writability in 22 nm FDSOI. By integrating transistor aging and interconnect degradation models into circuit-level simulations, the proposed framework enables cross-layer reliability analysis of SRAM write behavior under combined FEOL and BEOL aging stress. The main contributions of this chapter are summarized as follows:

1. Development of a cross-layer reliability analysis framework considering both FEOL transistor aging and BEOL interconnect degradation in advanced SRAM structures.
2. Modeling the impact of NBTI-induced threshold voltage degradation and EM-induced interconnect resistance increase on SRAM writability and switching behavior.
3. Identification of possible failure mechanisms during SRAM write operation under FEOL and BEOL degradation.
4. Investigation of the individual and combined effects of FEOL and BEOL degradation on SRAM switching delay and write reliability.
5. Providing reliability insights toward aging-aware SRAM design and SLM methodologies.

The structure of this chapter is as follows: Section 6.1 introduces the impact of NBTI and EM degradation on SRAM operation. Section 6.2 describes the proposed methodology, including TCAD

calibration, transistor aging modeling, and EM-induced interconnect degradation analysis. Section 6.3 discusses the evaluation of SRAM writability under NBTI, EM, and their combined effects. Finally, Section 6.4 concludes the chapter with a summary.

6.1. Related work

EM-induced failures in power interconnects have attracted significant attention because power lines typically carry unidirectional current with relatively high current density. For example, the work in [93] employed a tree-based interconnect structure together with physics-based EM modeling. The proposed methodology considers both void nucleation and void growth to evaluate resistance degradation while also accounting for the statistical nature of EM caused by random grain-size distribution. Furthermore, [94] proposed an on-chip power-grid optimization framework considering EM void saturation volume in multi-segment interconnects to improve long-term interconnect reliability.

In SRAM structures, [95] investigated the EM susceptibility of SRAM BLs under uniform workload conditions using the empirical Black's equation for EM modeling. Under uniform workloads, where the number of Write '1' and Write '0' operations are assumed to be equal, the EM-induced degradation is dominated by read operations. In addition, [96] proposed a workload-aware EM analysis methodology for STT-MRAM bit-lines. However, the modeling of bit-line current behavior remains insufficiently realistic, resulting in inaccurate EM lifetime prediction.

Although previous studies have investigated EM degradation in power grids and memory interconnects, comprehensive analysis of workload-dependent EM degradation together with FEOL transistor aging effects in advanced SRAM structures remains insufficiently explored.

6.2. Modeling Methodology

This section introduces our proposed method for TCAD-based NBTI and EM analysis. The major steps include:

- *TCAD Calibration*: The FDSOI devices are constructed using TCAD and calibrated against the PDK models by conducting a detailed comparison of the drain current versus drain voltage ($I_D - V_D$) and drain current versus gate voltage ($I_D - V_G$) curves to ensure accuracy and consistency with the PDK's specifications.
- *Building the NBTI Model*: The ΔV_{th} due to NBTI is modeled under constant stress, assessing the impact of V_{dd} and temperature on ΔV_{th} .
- *Injecting the EM fault*: Considering the impact of the EM in the resistance-increase phase by increasing the per-SRAM parasitic resistance of the bit-line interconnect in small steps.
- *Model Exploitation*: Evaluating the individual and combined effects of NBTI and EM on SRAM cell's writability.

6.2.1. TCAD Model Calibration

In this chapter, both N-type and P-type transistors are simulated using the Sentaurus device simulator with the drift-diffusion model and calibrated to GlobalFoundries' SLVT FDSOI PDK. The parameters of the proposed TCAD models are listed in Table 5.1. The calibration process involved the adjustment of various geometric and doping parameters to achieve a good qualitative fit for digital logic applications. Consequently, a comparison of the transistor output characteristics between the TCAD and PDK models is shown in Fig. 5.2(a), while the structure of the FDSOI is depicted in Fig. 5.2(b).

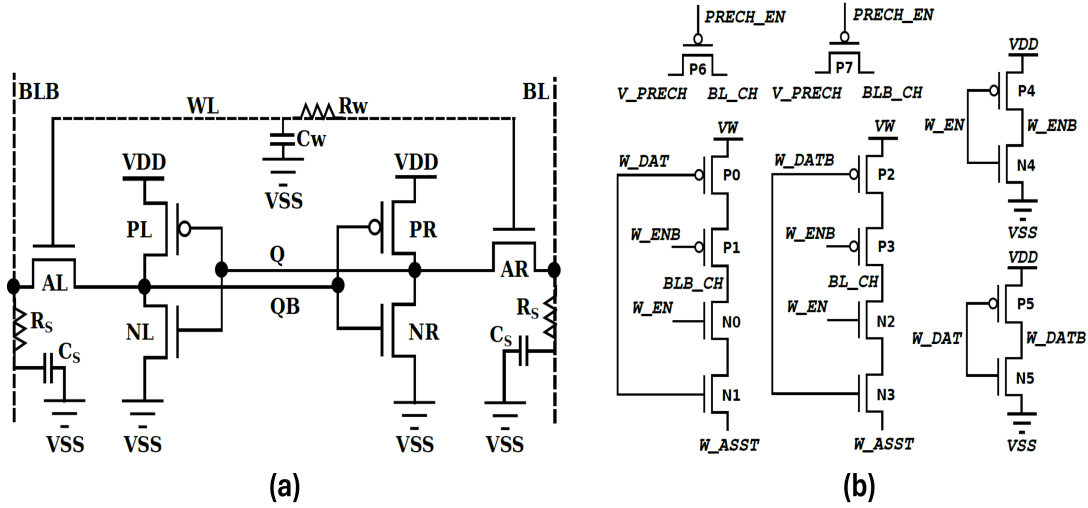


Figure 6.1.: SRAM cell schematics (a) and peripheral circuit (b), with per-cell parasitic resistance on BL and word line (WL).

6.2.2. Building the NBTI Model

The NBTI effect is modeled using the Sentaurus device simulator, integrated with the MSC model mentioned in Section 5.2.2. This integration facilitates the estimation of the ΔV_{th} under different supply voltage (V_{DD}) and temperature (T) conditions, covering a stress duration from 0 to 5 years. Fig. 5.3(a) shows ΔV_{th} for two specific stress scenarios: a nominal stress scenario ($V_{DD} = 0.8\text{ V}$, $T = 100^\circ\text{C}$) and a high stress scenario ($V_{DD} = 0.96\text{ V}$, $T = 125^\circ\text{C}$), indicating the different effects on ΔV_{th} . As described in Table 6.1, it is assumed in this work that the SRAM operates under nominal stress conditions.

6.2.3. NBTI Fault Injection

The six-transistor SRAM cell shown in Fig. 6.1(a) combines two access transistors (AL and AR) for external connections via bit lines (BL and BLB) and two cross-coupled inverters for data storage. Since an SRAM cell can store identical data over prolonged periods, asymmetric degradation of the transistors may occur [97]. Therefore, we investigate the consequences of NBTI and EM under the assumption that the SRAM cell permanently stores a ‘0’ and thus forgoes any recovery mechanism, which particularly affects the V_{th} of PL in Fig. 6.1(a).

6.2.4. EM Fault Injection

In our EM fault injection, we assume that the resistance increase due to the void will affect all the per-cell parasitic resistances of BL equally. However, the effects of the nucleation and incubation phases cannot be analyzed at the circuit level because their impacts are not propagative to this level. The phase of resistance increase, characterized by its rapid progression, is of greater significance. Therefore, this work focuses exclusively on the resistance-increase phase for functional failures.

6.2.5. Model Exploitation

We investigate the writability of an SRAM array featuring dimensions of 128 columns and 40 rows, considering both the individual and combined impacts of NBTI and EM. The analysis focuses on the SRAM cell located farthest from the driver, performing a write ‘1’ $\rightarrow$ ‘0’ operation, representing the most critical scenario. In the circuit-level simulation, we fully simulate the peripheral circuitry, including

the WL driver, read and write circuitry, and pre-charger. For the SRAM array, however, we employ a critical path simulation approach, i.e., we simulate the entire 128 rows and represent the columns with an equivalent resistor-capacitor circuit and an SRAM cell.

6.3. Results and Discussion

In this section, the simulation results for the writability of the SRAM under the individual and combined influence of NBTI and EM are presented. Our main focus in this investigation is the switching time (T_{switch}) violation. The SRAM write operation is a sequence of the following operations:

- Activation of the corresponding WL.
- Activation of the write-enable signal.
- Switching the SRAM content.
- Pre-charging the BL and BLB.

The individual and combined effects of NBTI and EM can lead to significant challenges for SRAM cells. First, it can prolong the T_{switch} , resulting in delay faults where the SRAM cell cannot complete a write '1' → '0' within the critical T_{switch} . Second, hard stuck-at faults can occur, rendering the SRAM cell completely inoperable and unable to change its state. Our analysis distinguishes between successful operations, where the cell can successfully perform a write '1' → '0' operation within the critical T_{switch} , and these faults, examining the impact of NBTI and EM effects on SRAM writability.

6.3.1. Simulation Setup

The circuit-level simulations were performed using the typical-typical corner setting with a BL as a critical path containing a 128-cell SRAM. The simulation setup is listed in Table 6.1. A transistor W/L ratio of 100 nm/20 nm was assumed for the peripheral circuit, while a ratio of 80 nm/20 nm was chosen for the cell and sense amplifier. To focus on evaluating SRAM performance under extreme conditions, write-boost was excluded from our analysis to examine the inherent vulnerabilities of the SRAM array.

6.3.2. Individual Effect of NBTI and EM on Writability

To model the individual aging effects of NBTI and EM, the V_{th} of PL and the value of the R_s are increased separately and gradually, as shown in Fig. 6.1(a). The T_{switch} is defined as the duration required for the

Table 6.1.: Nominal setup parameters

SRAM parameters (FEoL and BEoL)	
SRAM array configuration	128 rows, 128 columns
Vdd (SRAM array and periphery circuit)	800 mV
Pre-charge level (BL and BL-bar)	V_{DD}
Write voltage	High: V_{DD} , Low: 0 V
Operating temperature	100 °C
Operating frequency	1 GHz
Maximum switching time	570 ps
BL width (W) and height (H)	W = 22 nm, H = 41.8 nm
Wire (R(Ω), C(aF)) per cell [98], [99]	BL: (35.18, 89.92) WL: (24.09, 61.59)
BL average current	1.05 μ A

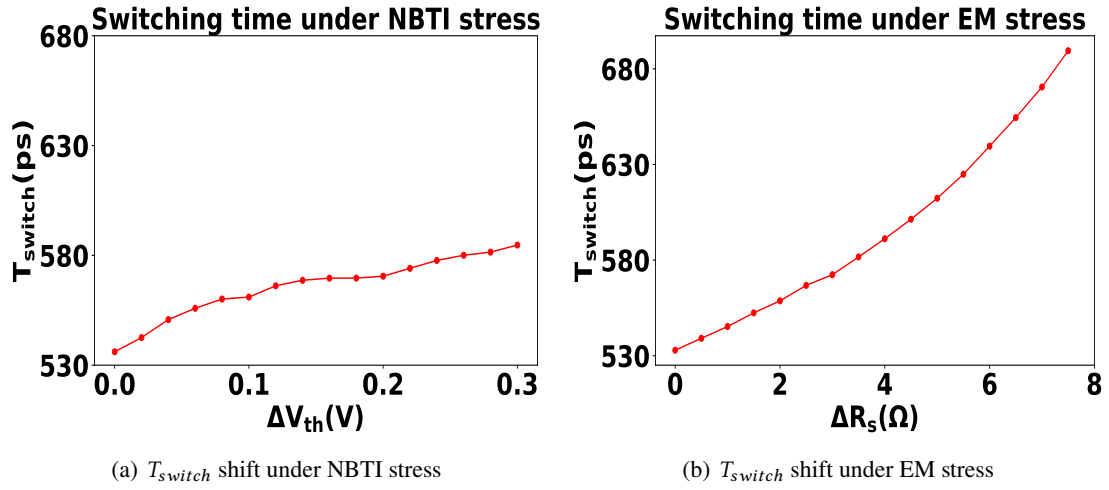


Figure 6.2.: Individual effects of NBTI and EM on T_{switch} of the write '0' operation.

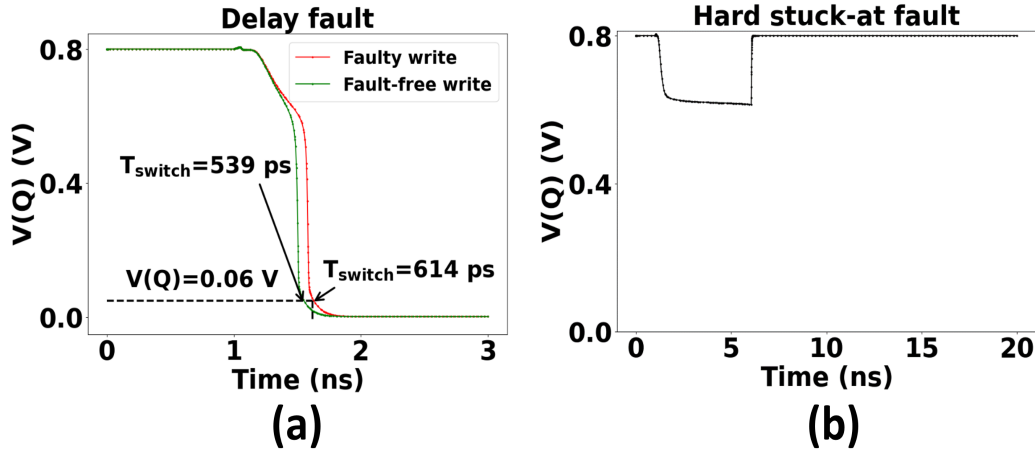


Figure 6.3.: Examples of faults in the FDSOI SRAM: (a) delay fault with a critical T_{switch} of 600 ps, and (b) stuck-at fault.

Q of the SRAM cell to drop to 0.05 V during a write '1' $\rightarrow$ '0' operation. Fig. 6.2 illustrates that as V_{th} and R_s change (ΔV_{th} and ΔR_s), the T_{switch} increases accordingly for both NBTI and EM scenarios. However, the increase in T_{switch} attributable to NBTI begins to saturate at larger ΔV_{th} values, indicating a decreasing effect over a longer stress duration. In contrast, the increase in T_{switch} due to EM, influenced by ΔR_s , shows an almost linear trend, indicating a consistent influence of EM on the T_{switch} .

6.3.3. Combined Effect of NBTI and EM on Writability

To capture the combined aging effects of NBTI and EM, the V_{th} of PL and the value of R_s are increased simultaneously. The simulation results shown in Fig. 6.2 illustrate that the time to failure of the SRAM is significantly affected by both EM and NBTI. Our analysis shows that NBTI and EM degradation can lead to either delay faults or stuck-at faults, as illustrated in Fig. 6.3. It is important to note that NBTI alone does not lead to stuck-at faults within the evaluated stress duration. Conversely, Fig. 6.4 shows that EM can induce stuck-at faults independent of the effects of NBTI. Furthermore, it can be observed that NBTI

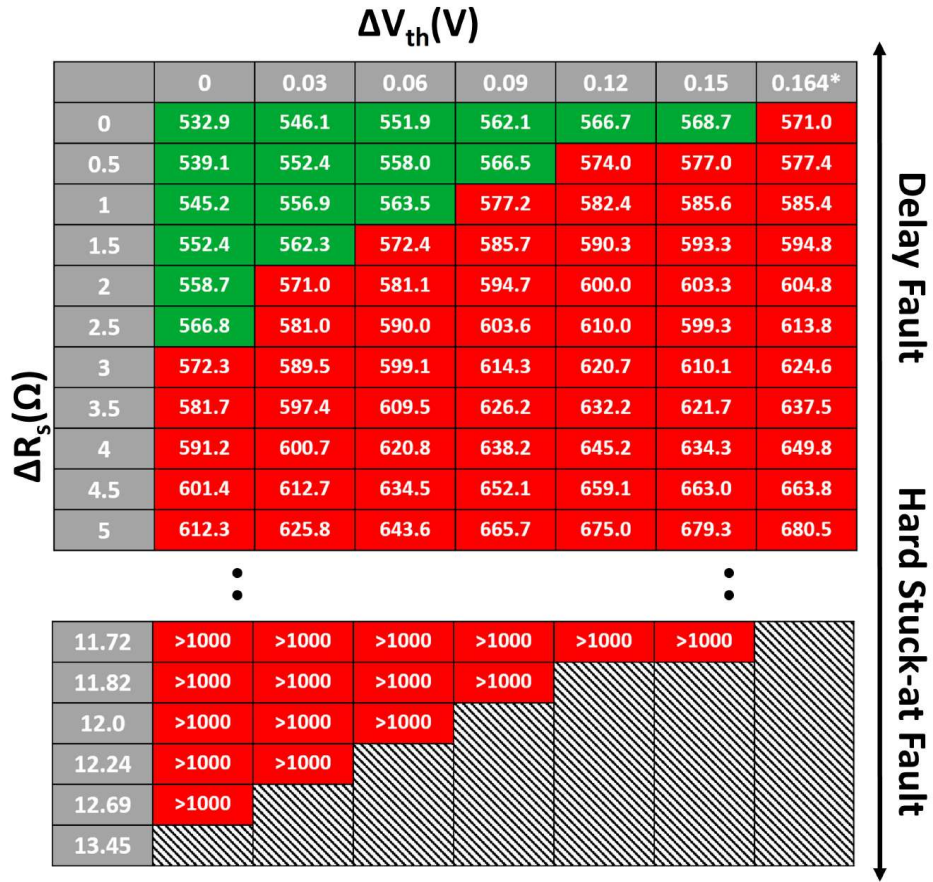


Figure 6.4.: Delay and stuck-at faults caused by NBTI and EM, categorized by cell color. Green cells denote successful write operations within the critical switching time (T_{switch}), while red cells indicate delay faults, as illustrated in Fig. 6.3(a). Pattern cells represent stuck-at faults, as shown in Fig. 6.3(b). *: Delay faults start to occur at a ΔV_{th} of 0.164 V.

can amplify the effects of EM and vice versa, exacerbating both delay and stuck-at faults. Therefore, analyzing the effects of NBTI or EM alone leads to a significant overestimation of SRAM lifetime. In particular, under the sole influence of NBTI, the threshold stress duration for faultless writing ‘1’→‘0’ operation is about 4.3 years. However, by including EM effects, the faultless writing ‘1’→‘0’ operation is drastically reduced to only 1.13 years of stress, with the resistance increasing by a small value of 0.5 Ω . This reduction illustrates the accelerated fault manifestation due to the EM-induced increase in resistance. These findings emphasize the need to consider both NBTI and EM when evaluating and predicting the time to failure of SRAM under aging conditions.

6.4. Conclusion

In this study, we have thoroughly investigated the impact of NBTI and EM on the writability of SRAM, focusing on worst-case operating scenarios to avoid overestimating the lifetime of SRAM. Our results emphasize the mutual acceleration of delay faults and hard stuck-at faults caused by NBTI and EM and highlight a synergistic exacerbation of the fault mechanisms. Significantly, our results demonstrate that while NBTI’s impact on the SRAM content switching tends to reach saturation, the influence of EM

becomes increasingly dominant with prolonged stress time. This analysis helps to devise more effective test and mitigation strategies during device lifetime in the context of SLM.

7. Testing for Aging in Advanced SRAM: From Front-End-of-The-Line Transistors to Back-End-of-The-Line Interconnects

Chapter 6 demonstrated that the combined effects of FEOL transistor aging and BEOL interconnect degradation can considerably degrade SRAM writability and lead to delay faults and hard stuck-at faults, potentially resulting in overestimation of the fault-free operational lifetime of SRAM arrays. Therefore, effective aging-aware testing and monitoring methodologies are required to enable earlier prediction and detection of aging-induced SRAM failures, allowing system-level mitigation techniques to be applied before catastrophic functional failures occur.

In advanced FinFET technologies, reduced voltage margins, increasing integration density, and aggressive interconnect scaling further increase the sensitivity of SRAM write operations to aging-induced degradation. In particular, SRAM write stability strongly depends on write operation conditions, especially pre-write operations and reduced supply voltage conditions, making aging-aware testing increasingly challenging in advanced FinFET SRAM structures. Furthermore, the aging behavior and write sensitivity of FinFET SRAM structures can differ considerably from previously investigated FDSOI-based SRAM designs due to technology-dependent device and interconnect characteristics.

To address these challenges, this chapter presents an aging-aware SRAM testing framework under combined NBTI and EM degradation in 12 nm SLVT FinFET technology, based on the work published in [5]. TCAD-based transistor models are extended with detailed NBTI degradation modeling, while a physically based and measurement-calibrated EM model is employed to capture BEOL interconnect degradation. By integrating both models into circuit-level simulations, the combined effects of transistor and interconnect aging on SRAM writability are investigated for a complete SRAM subarray, including both bit-cell arrays and peripheral circuits. Furthermore, a voltage-scaling-based sensitization and testing methodology is proposed to detect aging-induced SRAM failures at earlier aging stages, enabling earlier prediction of SRAM failures and estimation of the remaining operational lifetime.

The rest of this chapter is structured as follows: Section 7.1 introduces the NBTI and EM models together with FinFET SRAM reliability considerations. Section 7.2 outlines the aging modeling methodology, including TCAD calibration, NBTI simulation, and EM model generation. Section 7.3 presents the proposed aging-aware testing methodology, including NBTI and EM fault injection, sensitization techniques, and the voltage-scaling-based early detection method. Section 7.4 discusses the SRAM writability analysis and early failure detection results under combined aging effects. Finally, Section 7.5 concludes the chapter.

7.1. Related Work

In [100], the authors compare SOI and bulk supersteep retrograde (SSR) FinFET technologies for 6T SRAMs at the 7/8 nm node in terms of static noise margin (SNM), write ability, variability, and minimum operating voltage (V_{DD}). The results show that SOI FinFET SRAMs achieve slightly improved read stability and lower $V_{DD,min}$ due to enhanced electrostatic integrity, although the overall improvement over optimized bulk SSR FinFET SRAMs remains limited.

To improve long-term SRAM reliability, [101] proposes predictive testing methodologies considering the combined effects of process variation and BTI aging, with particular focus on read delay and read-destructive faults. The work introduces wordline- and supply voltage-based predictive testing together with adaptive DVFS mitigation techniques, demonstrating that periodic predictive testing can considerably reduce aging-induced failures and DPPM with relatively small energy overhead.

In addition, [98] investigates the impact of technology scaling on electromigration reliability in SRAM WLs and BLs during DTCO. By comparing SRAM interconnects with 22 nm and 12 nm critical dimensions, the authors demonstrate that tighter-pitch BEOL interconnects significantly increase current density and EM-induced stress, resulting in a $2.53\times$ increase in EM degradation despite achieving $2.68\times$ improved energy efficiency. The work further proposes an EM-aware DTCO methodology balancing SRAM performance, energy efficiency, and EM reliability through workload-aware current analysis and operating-condition optimization.

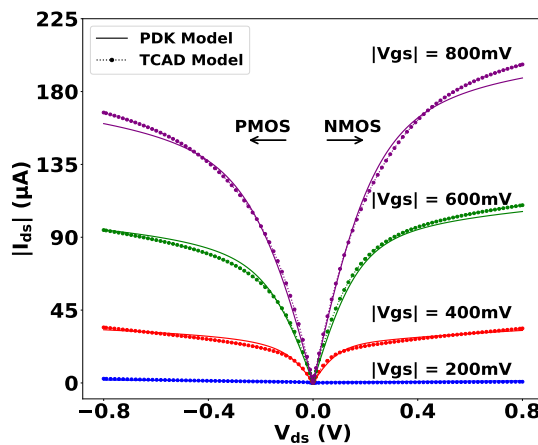
Although previous studies have investigated FinFET SRAM reliability, predictive aging-aware testing, and interconnect scaling effects independently, comprehensive aging-aware SRAM testing methodologies considering the combined effects of FEOL transistor aging and BEOL interconnect degradation in advanced FinFET SRAM structures remain insufficiently explored.

7.2. Aging Modeling

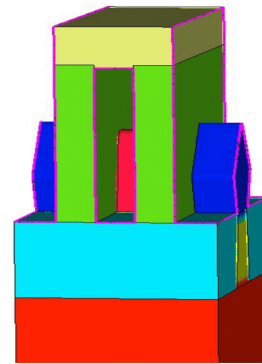
In this section, we present our methodology for modeling NBTI and EM, aimed to accurately predict their combined effects over time and improve design reliability. The major steps are outlined as follows.

7.2.1. TCAD Model Calibration

In this study, both N-type and P-type transistors are simulated using the Sentaurus device simulator, employing the drift-diffusion model. Each model is calibrated against PDK models [102] by conducting a detailed comparison of the drain current versus drain voltage ($I_{DS} - V_{DS}$) and drain current versus gate voltage ($I_{DS} - V_{GS}$) curves to ensure accuracy and consistency with the PDK's specifications, as shown in Fig .7.1(a). The structure of the FinFET is illustrated in 7.1(b). Moreover, the parameters of the proposed TCAD models for the FinFET devices are listed in Table 7.1.



(a) The TCAD and PDK model $I_D - V_D$ comparison



(b) The FinFET structure (P-Type)

Figure 7.1.: Output characteristics of the TCAD and PDK MOSFETs and the proposed FinFET structure with diamond-shaped epitaxy.

Table 7.1.: FinFET device parameters

Parameters	N-type	P-type
Fin Width / Height	16 nm / 40 nm	
Gate Length	14 nm	
Effective Gate Thickness	1.35 nm	1.3 nm
Body Thickness	40 nm	
Material	Silicon	
S/D Doping Peak	$3.8\text{e}+20 \text{ cm}^{-3}$	$5\text{e}+20 \text{ cm}^{-3}$
Channel / Body Doping	$1\text{e}+15 \text{ cm}^{-3}$	
Work Function	4.5 eV	4.82 eV

7.2.2. Building the NBTI and EM Model

The aging modeling methodology employed in this chapter generally follows the FEOL and BEOL degradation framework presented in Chapter 6. Specifically, the MSC framework is used to model the threshold voltage degradation (ΔV_{th}) caused by NBTI, while Equation 2.2 is employed to model the EM-induced interconnect resistance increase (ΔR_s) of SRAM BLs. However, the TCAD device structures, transistor parameters, and interconnect dimensions are recalibrated for the 12 nm SLVT FinFET technology node, as summarized in Table 7.1 and Table 7.2. Consequently, the technology-dependent aging behavior and SRAM write sensitivity of advanced FinFET SRAM structures differ from those of the previously investigated 22 nm SLVT FDSOI SRAM designs.

7.3. Aging Testing

This study investigates the writability of an SRAM array with 128 rows and 40 columns (32 data bits plus 8 parity bits needed for single error correction and double error detection), considering both the

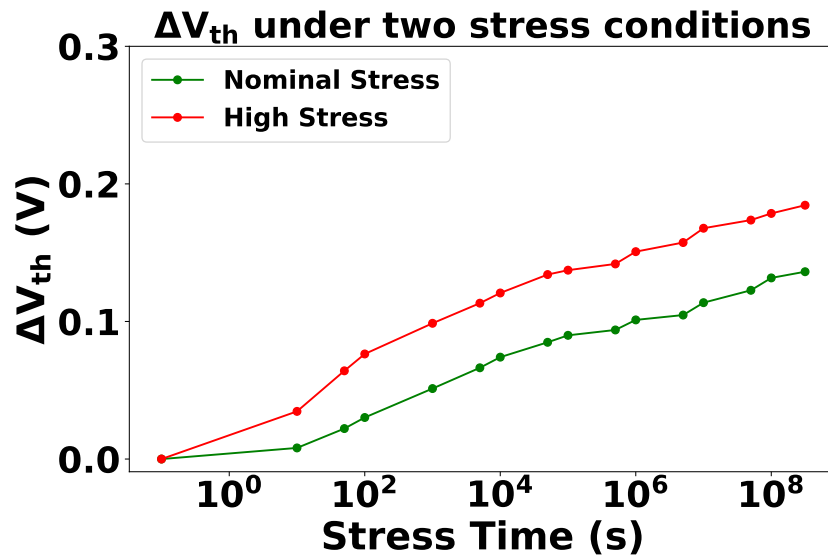
**Figure 7.2.:** The V_{th} increase due to NBTI in a sole P-type transistor.

Table 7.2.: Simulation setup parameters

SRAM general parameters	
SRAM array configuration	128 rows, 40 columns
V_{dd} (SRAM array and periphery circuit)	800 mV
Pre-charge level	V_{DD}
Write voltage	High: V_{DD} , Low: 0 V
Operating temperature	100 °C
Maximum switching time	220 ps
Subarray level latency	700 ps
Negative write driver voltage	-0.1 V
BL width (W) and height (H)	W = 17 nm, H = 32.3 nm
Cell transistors size	2 Fins, 1 Finger
Wire (R(Ω), C(aF)) per cell [98], [99]	BL: (54.59, 63.81)WL: (18.58, 84.295)
BL average current	300 nA

individual and combined impacts of NBTI and EM. The read operation is more resilient to the aging effects; therefore, we focus on the write operation for earlier detection. Our analysis centers on the SRAM cell with the weakest writability, located farthest from the driver, i.e., located in row 128, column 40 [103]. The NBTI and EM models from Section 7.2 were used in the testing of the SRAM subarray, the major steps are outlined as follows:

7.3.1. NBTI Fault Injection

The six-transistor SRAM cell shown in Fig. 6.1(a) includes two access transistors (AL and AR, connected with the word line (WL)) for external connections through bit-lines (BL and BLB) and two cross-coupled inverters for data storage. Notably, storing identical data for long periods in SRAM can cause asymmetric aging of the transistors. Therefore, we investigate NBTI and EM effects, assuming the SRAM cell permanently stores a ‘0’, affecting the V_{th} of PL in Fig. 6.1(a). Results for storing a ‘1’ are similar. We also consider the threshold voltage shift (ΔV_{th}) in peripheral circuit transistors in Fig. 6.1(b) and the sense amplifier to provide a comprehensive analysis of potential degradation across all SRAM components.

7.3.2. EM Fault Injection

For EM fault injection, we assume that the resistance increases due to the sufficiently large void, which equally affects all the per-cell parasitic resistances of BL. However, the effects of the nucleation and incubation phases cannot be observed and therefore analyzed at the electrical level. The phase of resistance increase, characterized by its rapid escalation, is the only phase that has an observable effect at the electrical level and can impact the SRAM functionality. Therefore, this study focuses exclusively on the resistance-increase phase to investigate in-field failures.

We investigated the effects of NBTI and EM on the write operations of SRAM cells by varying the ΔV_{th} and ΔR_s . Our main focus was on switching time (T_{switch}) violation. The individual and combined effects of NBTI and EM can severely impact the functionality of SRAM subarray in multiple ways. First, these effects can prolong the T_{switch} , resulting in delay faults where the SRAM fails to complete a write operation within the expected switching time. These faults can be mitigated by increasing the operating latency. Second, they can lead to hard stuck-at faults, which make the SRAM cell unable to switch. These faults cannot be corrected by increasing the operational cycle. Our analysis distinguishes between successful operations, where the cell can complete a write operation within the assumed latency, and these faults, and thus investigates the impact of NBTI and EM on the writability of SRAM.

In order to prevent overestimation of the fault-free lifetime and realize early detection of write failures in SRAM, our test strategies are as follows:

- Investigation of T_{switch} at the nominal operation corner: We investigated the T_{switch} under the condition that both the supply voltage of the SRAM cell (V_{cell}) and the supply voltage of the peripheral circuitry (V_{peri}) were set to 0.8 V. First, we compared the T_{switch} in the write '1'→'0' operation ($W0\downarrow$) with that of the '0'→'1' operation ($W1\uparrow$) to identify the critical write operation.
- T_{switch} analysis at the test operation corner: At this test operation corner, V_{peri} was slightly reduced. In contrast to the adjustments of V_{peri} , we did not increase the V_{cell} , as we found that a slight increase could cause write failures in the SRAM.
- Earlier detection of write failures: Test voltage corner analysis aims to sensitize the write operation of SRAM, thereby enabling earlier detection of aging-related failures.

7.4. Results and Discussion

In this section, we present the key results of our FinFET SRAM testing, including identification of the critical write operation, estimation of the fault-free lifetime under NBTI alone or in combination with EM, and sensitization of the write operation with voltage test corners for early fault detection.

7.4.1. Critical Write Operation

The switching write operations, i.e., $W0\downarrow$ and $W1\uparrow$, are asymmetrical, and one of them typically requires more energy to be performed and is therefore more critical. To achieve earlier detection of aging effects, identifying the most critical writing direction is a prerequisite. Transistor leakage, along with the slight difference in the biasing of the access transistors (as indicated by AL and AR in Fig. 6.1(a)) is the main reason for such an asymmetric behavior. Based on our simulation setup outlined in Table 7.2, $W1\uparrow$ was observed to be more critical in FinFET SRAM.

7.4.2. T_{switch} at the Nominal Operation Corner

During testing, both delay and stuck-at faults occurred, as illustrated in Fig. 7.3(a) and (b), respectively. To model the aging effects of NBTI and EM, we increased the V_{th} of the corresponding transistors under NBTI stress and the value of R_s in small steps, as shown in Fig. 7.4. The T_{switch} is defined as the duration required for the state of the SRAM (Q in Fig. 6.1(a)) to exceed 90% of the final value. Furthermore, in Fig. 7.4, V_{th} is increased in increments of 0.05 V and R_s in increments of 1 Ω . The intermediate results marked on the ticks indicate the starting points at which either a delay fault or a stuck-at fault was observed. The results in Fig. 7.4 indicate several important findings related to SRAM cell write operations, categorized as follows:

- Successful operations (solid green entries, with no need for frequency downscaling): These cells represent scenarios where SRAM cell write operations complete within the prescribed subarray-level latency.
- Immediate delay faults (solid red entries, recoverable with a frequency downscaling of less than 50%): These cells mark the first occurrence of a delay fault at certain ΔV_{th} and ΔR_s values. For example, a red cell at $\Delta V_{th} = 0.05$ V and $\Delta R_s = 2.46$ Ω in Fig. 7.4 marks the first occurrence of a write delay fault under these specific conditions.

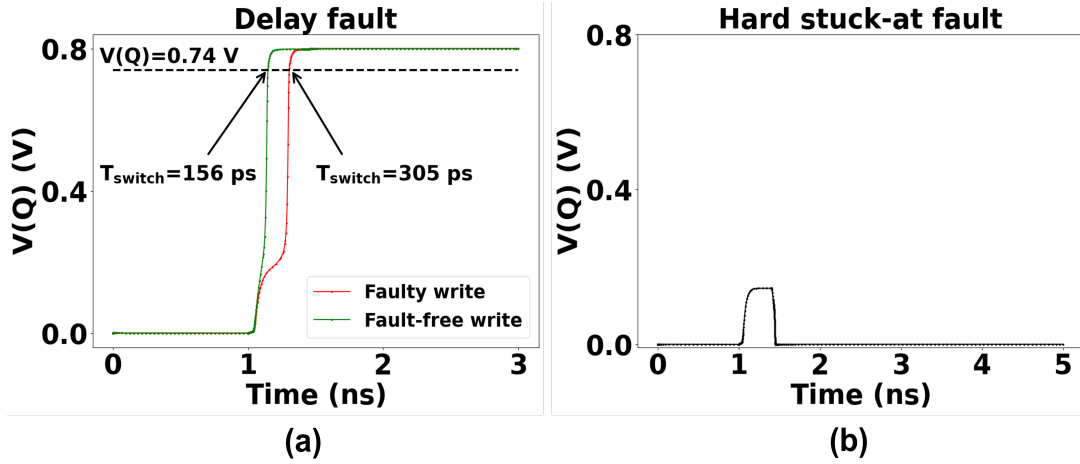


Figure 7.3.: Example of a delay fault in (a), where the critical switching time is 220 ps, and a hard stuck-at fault in (b).

- Extended delay faults (red dashed entries, recoverable with a frequency downscaling of more than 50%): These cells represent conditions where the write delay exceeded the prescribed latency at the subarray level and did not meet the predefined latency requirements, but did not result in a stuck-at fault. An example of such a fault is shown at $\Delta V_{th} = 0.15$ V and $\Delta R_s = 4.09$ Ω in Fig. 7.4.
- Hard stuck-at faults (black dashed entries, not recoverable through frequency downscaling): These cells indicate scenarios in which a stuck-at fault occurred, i.e., the write operation has not successfully changed the state of the SRAM cell within an extremely prolonged period of 5 ns. An example of such a fault is recorded at $\Delta V_{th} = 0$ V and $\Delta R_s = 7.2$ Ω in Fig. 7.4.

In Fig. 7.4, important observations can be drawn from the effects of NBTI and EM on the writability of SRAM. In particular, NBTI alone does not cause stuck-at faults within the evaluated stress duration. However, EM can cause stuck-at faults regardless of the influence of NBTI. Delay faults are only observed when the ΔV_{th} is 0.23 V. According to the TCAD results shown in Fig. 7.2, it is unlikely that such threshold voltage shifts occur within the considered stress duration and conditions for the assumed FinFET technology. Therefore, a write delay fault at the nominal operation corner is unlikely if only the NBTI effect is considered. However, our studies suggest that the co-existence and interaction between NBTI and EM can exacerbate both delay and stuck-at faults in FinFET SRAM. Specifically, the combined effects result in a significantly shorter fault-free lifetime for SRAM. Write delay faults occur at a minimum ΔV_{th} of 0.05 V with a small ΔR_s of 2.46 Ω , much less than the critical ΔV_{th} (0.23 V) when only NBTI is considered.

7.4.3. T_{switch} at the Test Operation Corner

As shown in Fig. 7.5, decreasing V_{peri} significantly increases the T_{switch} of fresh devices and further increases the sensitivity of the SRAM writability to the combined effect of NBTI and EM, reducing the write operation's time-to-failure. In tests with fresh SRAM, the T_{switch} at the proposed test operation corner increases from 155.7 ps to 170.9 ps compared to the nominal operation corner, i.e., an increase of 9.8%. The time-to-failure due to delay faults in the SRAM is very sensitive to reductions in V_{peri} . For example, a stuck-at fault occurs at a resistance increase of 2.34 Ω with a ΔV_{th} of 0.15 V, which is 46.0% less than the resistances observed at the nominal operation corner ($\Delta R_s = 4.33$ Ω), emphasizing heightened sensitivity of the SRAM's writability to peripheral voltage changes.

		$\Delta V_{th} \text{ (V)}$									
		0	0.05	0.1	0.15	0.2	0.23	0.25	0.3	0.35	0.4
$\Delta R_s \text{ (}\Omega\text{)}$	0	155.7	173.5	190.2	203.9	214.2	220.1	222.0	232.9	248.0	274.6
	0.28	158.3	178.0	196.4	210.9	220.9	226.8	230.1	241.1	258.0	286.1
	0.68	162.2	183.8	202.5	220.5	230.3	237.3	241.5	254.9	274.5	307.3
	1	165.5	187.7	210.8	228.8	241.0	248.3	252.8	265.9	285.9	324.4
	1.36	170.3	195.2	220.3	240.4	255.2	263.3	268.2	283.4	308.1	355.3
	2	178.0	209.7	239.4	265.6	284.5	296.7	303.8	323.8	356.6	438.1
	2.46	185.4	220.3	260.0	288.7	313.3	330.5	339.8	366.4	426.9	574.0
	2.93	193.5	236.6	284.3	330.3	364.4	384.9	410.0	454.4	568.8	
	3	194.8	235.5	289.7	337.7	374.9	396.8	414.0	474.0	599.9	
	3.4	201.1	254.4	319.7	390.6	451.8	492.5	532.7			
	3.7	209.2	271.0	289.7	355.2	583.4	682.4				
	3.91	214.0	280.4	391.6	538.2						
	3.98	216.8	286.1	406.8	599.6						
	4	216.8	286.8	412.1	617.3						
	4.09	219.5	290.7	421.7							
	4.11	220.1	292.9	430.2							
	4.33	227.8	315.0	522.5							
	4.77	241.1	372.4								
	5	252.5	425.7								
	5.66	291.4									
	6	317.9									
	7	693.9									
	7.2										
	8										

Figure 7.4.: Delay and stuck-at faults due to NBTI and EM under nominal operating conditions in FinFET SRAM. Delays are shown in picoseconds. Green entries indicate successful write operations within the critical T_{switch} . Solid red entries represent delay faults that a 50% frequency downscaling can mitigate, while patterned red entries indicate delay faults that this downscaling cannot mitigate. Patterned black entries represent stuck-at faults.

7.4.4. Earlier Detection of Write Failures

We demonstrate the effectiveness of our proposed voltage scaling-based earlier detection method by focusing on the stuck-at faults as a case study. Early stuck-at faults occur after a ΔV_{th} of approximately 0.13 V and a parasitic resistance increase (ΔR_s) of about 4.51 Ω . These values are interpolated from the data presented in Fig. 7.4, specifically between the points at $\Delta V_{th} = 0.1$ V, $\Delta R_s = 4.77$ Ω , and $\Delta V_{th} = 0.15$ V, $\Delta R_s = 4.33$ Ω . Notably, this ΔV_{th} , occurring in the stagnation region, typically appears after approximately 2.79 years. At this point, EM becomes the determining factor in in-field failures, reaching a critical resistance of about 4.51 Ω after approximately 7.13 years, marking the time-to-failure. By utilizing the test corner, the combined effect of NBTI and EM can be detected 44% earlier (3.15 years earlier) than in the nominal operation corner.

$\Delta V_{th}(V)$

	0	0.05	0.087	0.1	0.15	0.2	0.25	0.3	0.35	0.4
0	170.9	199.1	220.3	227.4	253.5	271.3	287.8	314.2	354.3	462.6
0.75	181.8	217.9	247.6	257.9	297.9	327.5	352.7	397.3	491.8	
0.82	183.0	220.2	251.2	262.0	297.1	326.1	362.9	400.9	525.6	
1	186.6	226.5	261.9	274.3	316.1	351.6	393.9	456.7	638.8	
1.32	192.8	235.3	277.9	293.8	352.2	410.4	473.4	600.0		
1.67	201.0	254.9	311.1	332.5	424.7	528.5				
1.91	204.3	269.8	332.6	363.1	510.4					
2	207.3	276.5	347.8	382.1	569.0					
2.1	210.5	278.6	360.1	406.3	652.3					
2.34	218.2	301.0	412.8	484.9						
2.4	220.2	308.4	436.5	520.1						
2.78	233.0	356.0								
2.95	241.0	391.6								
3	243.8	395.6								
3.64	243.8									
4	313.5									
5										
5.14										

$\Delta R_s(\Omega)$

Figure 7.5.: Delay and stuck-at faults caused by NBTI and EM at the test operation corner in the FinFET SRAM. Delays are in picoseconds. Colors and patterns of entries (green, solid red, patterned red, and patterned black) indicate the same operation statuses as in Fig. 7.4.

7.5. Conclusion

In this chapter, we thoroughly investigated the impact of NBTI and EM on SRAM's writability, focusing on worst-case operating scenarios. Our approach avoids the overestimation of SRAM write operation lifetime and facilitates early detection of aging-induced write faults using appropriate voltage test corners. Our results show a significant acceleration of delay and hard-stuck-at faults in write operations due to NBTI and EM, highlighting their exacerbated combined impact in FinFET SRAM. This analysis helps to develop effective strategies for early detection and mitigation strategies throughout the device lifecycle, especially in the context of SLM for safety-critical applications such as automotive.

8. Effect of Transistor and Interconnect Scaling on Aging in Advanced SRAM

Chapter 6 investigated the combined effects of NBTI and EM on SRAM writability in 22 nm FDSOI technology, while Chapter 7 extended the analysis toward 12 nm FinFET SRAM and introduced a voltage-scaling-based earlier detection methodology. However, several important challenges remain insufficiently explored.

First of all, a comprehensive comparison between FDSOI and FinFET SRAM under identical aging conditions is still missing. Due to differences in transistor architecture, device electrostatics, and interconnect scaling, the sensitivity of SRAM to NBTI- and EM-induced degradation can vary significantly across technologies. Therefore, a unified cross-technology analysis is required to understand the impact of technology scaling on SRAM aging behavior.

Furthermore, although a voltage-scaling-based earlier detection methodology was introduced for FinFET SRAM in Chapter 7, its applicability to FDSOI SRAM has not been investigated. In addition, the influence of operation-dependent sensitization techniques, such as applying specific pre-operations before write operations, remains insufficiently explored.

Moreover, the impact of process variation and cell location on aging-induced SRAM failures was not considered in the previous chapters. In advanced technology nodes, process variation can strongly influence SRAM writability and timing robustness, while its interaction with NBTI- and EM-induced degradation may further aggravate aging-induced failures. In addition, SRAM cell location within the SRAM array also affects aging behavior and write vulnerability, since cells located farther from the BL driver experience larger interconnect delay and voltage degradation.

In addition, in the previous chapters, the aging effects of SRAM bit-cells and peripheral circuits were evaluated together. However, the individual contributions of cell degradation and peripheral-circuit degradation to SRAM writability failures were not separately analyzed. Since write failures can originate either from the degraded SRAM cell or from the degradation in peripheral circuitry, separating these effects is necessary for identifying the dominant aging bottleneck in advanced SRAM architectures.

Finally, although the previous chapters focused on aging analysis and fault detection, mitigation strategies for extending SRAM lifetime were not investigated. Therefore, this chapter additionally introduces an aging-aware DVFS methodology and analyzes the associated trade-offs between reliability, performance, and energy consumption.

To address these challenges, this chapter presents the following contributions, which are based on the work published in [6]:

- A unified simulation framework combining TCAD-based NBTI models with a physics-based EM model for consistent cross-technology aging analysis in both FDSOI and FinFET SRAM technologies. The framework considers both SRAM bit-cell arrays and peripheral circuitry, enabling separate analysis of their contributions to SRAM writability degradation.
- An earlier failure-detection methodology that sensitizes aging-induced SRAM failures through operation-sequence control and supply-voltage scaling, enabling significantly earlier manifestation of delay faults and hard stuck-at faults compared with nominal operating conditions.

- A comprehensive investigation of process-variation and cell-location effects on SRAM aging behavior and failure mechanisms under combined NBTI and EM degradation, together with detailed analysis of both SRAM cell aging and peripheral-circuit aging.
- An aging-aware DVFS methodology as a system-level mitigation strategy, including analysis of the trade-offs among SRAM reliability, performance, and energy efficiency under long-term aging conditions.

The remainder of this chapter is organized as follows: Section 8.1 describes our methodology, which includes NBTI and EM fault injection. Section 8.2 discusses the evaluation of SRAM's writability under the combined influence of NBTI and EM, detailing the methods for earlier detection and the aging-aware DVFS strategy. Finally, Section 8.3 concludes with a summary.

8.1. Cross-Technology Aging Analysis and Mitigation Methodology

This chapter evaluates the writability of an SRAM subarray consisting of 128 rows and 40 columns. The investigation considers both the individual and combined FEOL and BEOL aging by varying ΔV_{th} and ΔR_s , following the methodology presented in Chapters 6 and 7.

The primary focus of the analysis is the identification of violations in the SRAM T_{switch} . Similar to the previous chapters, the analysis mainly targets the SRAM cell with the weakest writability, i.e., the cell located farthest from the driver. Furthermore, to investigate the impact of cell location, the FDSOI SRAM design is selected as a case study in which the first SRAM cell, located closest to the driver, is additionally analyzed.

The NBTI and EM aging models presented in Section 6.2 and Section 7.2 are integrated into the SRAM subarray evaluation flow. The key components of this analysis are outlined as follows:

8.1.1. NBTI and EM Fault Injection

We evaluate SRAM write '0' delay under three aging scenarios, corresponding to different dominance regimes of BL forcing and cell-internal regeneration:

1. Driver aged, cell aged (write '1' stress, cell stores '0'): Both BL forcing and regeneration are degraded, resulting in the worst-case write delay.
2. Driver aged, cell moderately aged (write '1' stress, cell switching): Both mechanisms are degraded in a more balanced manner, leading to moderate delay degradation.
3. Driver weakly aged, cell aged (write '0' stress, cell stores '0'): BL degradation is limited due to weak driver aging, and the delay is dominated by the weakened regeneration process.

In the following, Scenario (3) is used as the primary focus to isolate regeneration-limited behavior. Under this condition, the SRAM cell is assumed to continuously store '0', which induces NBTI degradation in the pull-up transistor PL in Fig. 6.1(a). Similar trends are observed when storing '1'. We further analyze the ΔV_{th} of the peripheral circuit transistors shown in Fig. 6.1(b), to analyze degradation across the SRAM array.

For EM fault injection, similar to the previous chapters, we assume that EM-induced degradation results in the formation of a sufficiently large void, leading to a uniform increase in the per-cell parasitic resistances along the BL.

8.1.2. Sensitization and Voltage-Scaling Test Strategies

In order to prevent overestimation of the fault-free lifetime and realize early detection of write faults in SRAM, our test strategies are as follows:

- Investigation of T_{switch} at the nominal operation corner: The T_{switch} was analyzed with the supply voltage of the SRAM cell (V_{cell}) and the supply voltage of the periphery circuitry (V_{peri}) both set to 0.8 V. Initially, the T_{switch} for the write '1'→'0' operation ($W0\downarrow$) was compared with that of the '0'→'1' operation ($W1\uparrow$) across the two technologies to identify the critical write operation.
- Sensitization of critical write operations: A test vector was employed to sensitize the critical write operation. Prior to each write operation, various pre-operations were performed to examine their influence on T_{switch} . This step is essential for accurately estimating the fault-free lifetime of the SRAM. To achieve this, simulations were conducted on the entire SRAM macro, encompassing both the bit-cell array and periphery circuit.
- T_{switch} analysis at the test operation corner: At this test operation corner, the V_{peri} was slightly reduced. Unlike the adjustments of V_{peri} , the V_{cell} was not increased, as even a slight decrease was found to induce functional faults in the SRAM for both technologies under investigation.
- Earlier detection of write faults: The analysis of the test corner and the test vector is designed to sensitize the write operation of SRAM, facilitating earlier detection of aging-induced faults in SRAM at the nominal operation corner.

8.1.3. Aging-Aware DVFS Strategy

As a part of the SLM infrastructure, we have implemented a DVFS strategy to ensure a prolonged lifetime of the SRAM module, mitigating the combined effects of NBTI and EM. This approach involves decreasing the V_{cell} and increasing the V_{peri} by 5%, while also considering decreasing the operating cycle of SRAM by 50%. These adjustments expand the lifetime of aged devices through graceful degradation, ensuring fault-free operation over time after the in-field aging testing confirms degradation in the SRAM module.

Estimating the actual lifetime of SRAM under the combined influence of NBTI and EM presents significant challenges due to the strong interdependence of these mechanisms. The key challenges are:

- **Mutual dependence of degradation:** NBTI-induced threshold voltage shift ΔV_{th} reduces the drive current of transistors, which lowers the current density in the bit-line metal and directly affects the rate of EM-induced resistance increase ΔR_s .
- **Different time evolutions:** NBTI follows an exponential time dependence, rising quickly in the early aging stage and then saturating, whereas EM exhibits an approximately linear increase in ΔR_s over time (Equation. 2.2).
- **Shift in dominant mechanism over time:** When ΔV_{th} is small, NBTI dominates the degradation behavior. As ΔV_{th} approaches its saturation region, EM-induced resistance increase becomes the dominant contributor to degradation.
- **Difficulty in unified lifetime prediction:** Because the dominant aging mechanism changes during the device lifetime, deriving a single closed-form model for combined NBTI–EM lifetime estimation is nontrivial.

8.2. Results and Discussion

We investigated the effects of NBTI and EM on SRAM write operations by varying ΔV_{th} and ΔR_s . The primary focus of the analysis is the identification of SRAM T_{switch} violations under nominal stress conditions. The analyses described in Section 8.1 were performed to evaluate SRAM writability under different aging conditions. The simulation setup follows the methodology presented in the previous chapters.

8.2.1. T_{switch} at the Nominal Operation Corner

To model the aging effects of NBTI and EM, we increased the V_{th} of the corresponding transistors under NBTI stress and the value of R_s in small steps, as shown in the diagrams in Fig. 8.1 (FDSOI) and Fig. 8.2 (FinFET). The T_{switch} is defined as the duration required for the state Q of SRAM to reach over 90% of the final value. Furthermore, in this diagram V_{th} is ramped with an increment of 0.05 V and R_s with an increment of 1 Ω . These increments represent routine steps along the axes that do not convey any information about faults. The intermediate results marked on the ticks indicate the starting points at which either a delay fault or a stuck-at fault was observed. The results indicate several important findings related to SRAM cell write operations, which are categorized as follows:

- **Successful operations** (solid green entries): write operations that complete within the prescribed subarray-level latency, without requiring any frequency downscaling.
- **Immediate delay faults** (solid red entries): write operations whose switching time exceeds the latency limit, but can be recovered by a frequency downscaling of less than 50%.
- **Extended delay faults** (red dashed entries): write operations that violate the latency constraint and require more than 50% frequency downscaling to be recovered, but do not result in stuck-at faults.
- **Hard stuck-at faults** (black dashed entries): cases where the write operation fails to change the cell state even within an extended time window of 5 ns and is therefore not recoverable by frequency downscaling.

This structured approach allows a detailed analysis of the effects of NBTI and EM on SRAM writability under different conditions and provides important insights to improve device reliability and performance.

Important insights can be drawn regarding the impact of NBTI and EM on SRAM writability. According to the TCAD-predicted results in Fig. 5.3(a) and Fig. 7.2, the NBTI-induced ΔV_{th} alone is unlikely to cause any write faults at the nominal operation corner. T_{switch} exhibits a slight non-monotonic variation with increasing ΔV_{th} (Fig. 8.1) due to the competition between BL forcing and cell-internal regeneration. At moderate ΔV_{th} , degradation of the pull-up transistor weakens the feedback that maintains the previous cell state, allowing the internal node to reach the switching point more easily, which can slightly reduce the switching delay. At higher ΔV_{th} , however, the same degradation slows down the regeneration of the cross-coupled inverter, leading to an increase in T_{switch} . Additionally, FinFET SRAM exhibits greater sensitivity to NBTI-induced delay degradation, indicating a tighter write margin compared to FDSOI SRAM. Moreover, our studies suggest that the co-existence and interaction between NBTI and EM can exacerbate both delay and stuck-at faults in both technologies. Specifically, FDSOI SRAM shows greater resilience to combined aging effects, with the corresponding delay and stuck-at fault thresholds summarized in Table 8.1. In contrast, FinFET SRAM shows a much shorter fault-free lifetime, entering delay fault regimes earlier under the same assumptions, as also shown in Table 8.1. In addition, FinFET technology is significantly more susceptible to EM, showing a reduced ΔR_s margin against stuck-at faults. These results emphasize that FDSOI SRAM has higher resilience to aging effects at the nominal operation corner, highlighting the critical role of such aging mechanisms in advanced nodes.

Table 8.1.: Fault thresholds under EM-only and combined NBTI+EM stress at the nominal corner.

Technology	Long-term stuck-at faults	Early-stage delay faults	
	ΔR_s	ΔR_s	ΔV_{th}
FDSOI	14.39 Ω	2.96 Ω	0.15 V
FinFET	7.2 Ω	2.46 Ω	0.05 V

To complement the analysis, we also analyze representative cases from Scenarios (1) and (2) mentioned in Section 8.1.1 for the FDSOI SRAM under early-stage aging. The results are summarized in Fig. 8.3. Scenarios (1) and (2) exhibit a larger delay due to the additional impact of peripheral circuit aging, which reduces the bitline driving strength. However, these scenarios combine multiple degradation sources. In contrast, Scenario (3) isolates cell-internal degradation, where the delay is mainly dominated by the regeneration process. Therefore, Scenario (3) is used as the primary focus to enable a clearer analysis of regeneration-limited behavior.

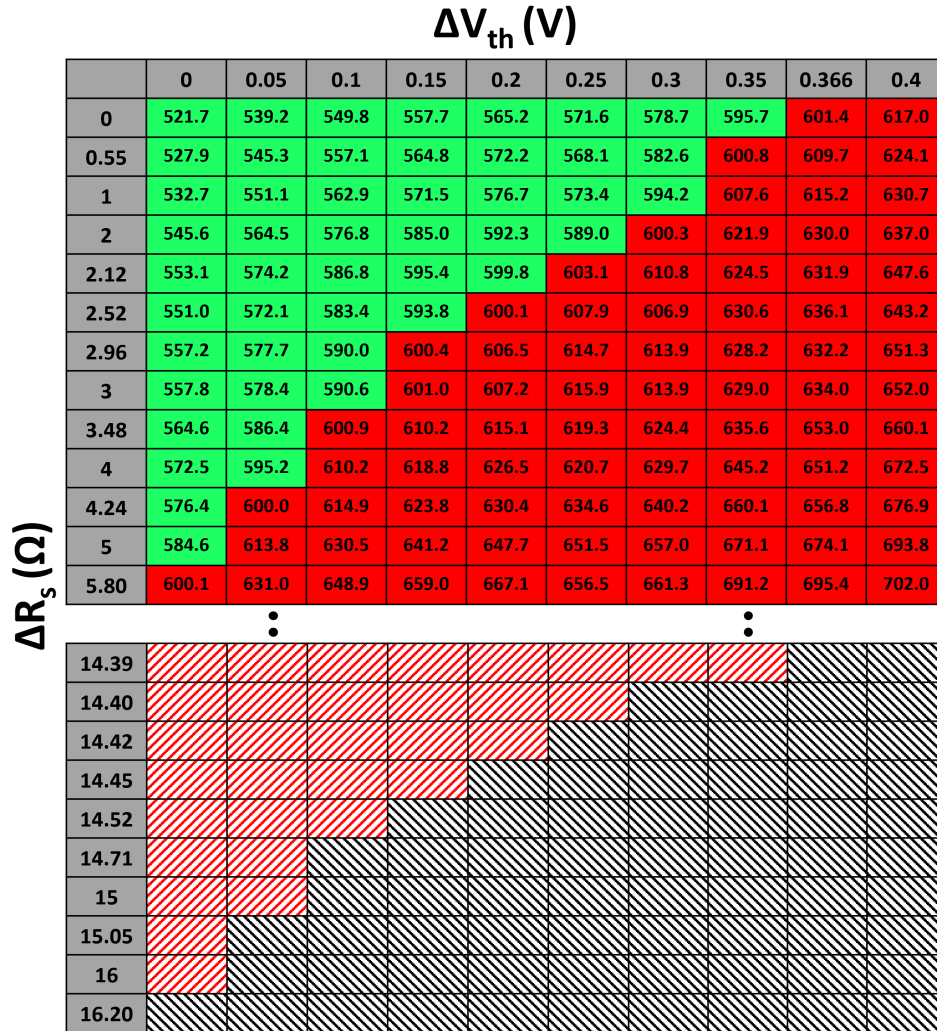


Figure 8.1.: Delay and stuck-at faults due to NBTI and EM under nominal operating conditions in the FDSOI SRAM. Delays are shown in picoseconds. Green entries indicate successful write operations within the critical T_{switch} . Solid red entries represent delay faults that a 50% frequency downscaling can mitigate, while patterned red entries indicate delay faults that this downscaling cannot mitigate. Patterned black entries denote stuck-at faults.

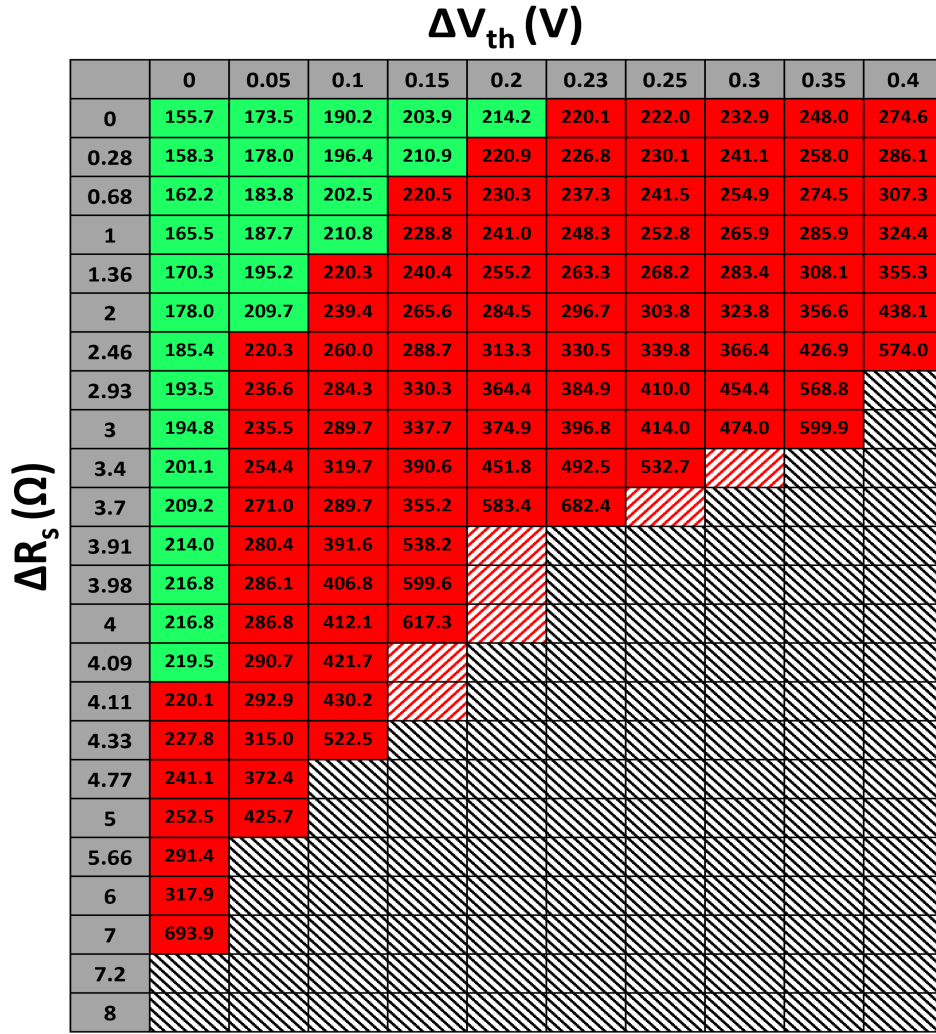


Figure 8.2.: Delay and stuck-at faults due to NBTI and EM under nominal operating conditions in the FinFET SRAM. The color coding is the same as in Fig. 8.1.

8.2.2. Impact of Process Variation on T_{switch}

We evaluate the impact of process variation and cell location on T_{switch} in the FDSOI SRAM using MC simulations across all process corners. As shown in Fig. 8.4 and summarized in Table 8.2, the last cell, which is furthest from the write driver, is significantly more sensitive than the first cell. For fresh devices, the last cell that is farthest from the write driver exhibits a 37.4% higher mean T_{switch} compared

	0	0.02	0.05		0	0.02	0.05
0	521.7	563	642.3	0	521.7	552.0	611.1
2	545.6	590.3		2	545.6	578.2	644.9
4	572.5	623.8		4	572.5	609.5	

(1) (2)

Figure 8.3.: Impact of aging scenarios on SRAM write '0' delay. Scenario (1): combined peripheral and cell aging. Scenario (2): peripheral-dominant aging. The color coding is the same as in Fig. 8.1.

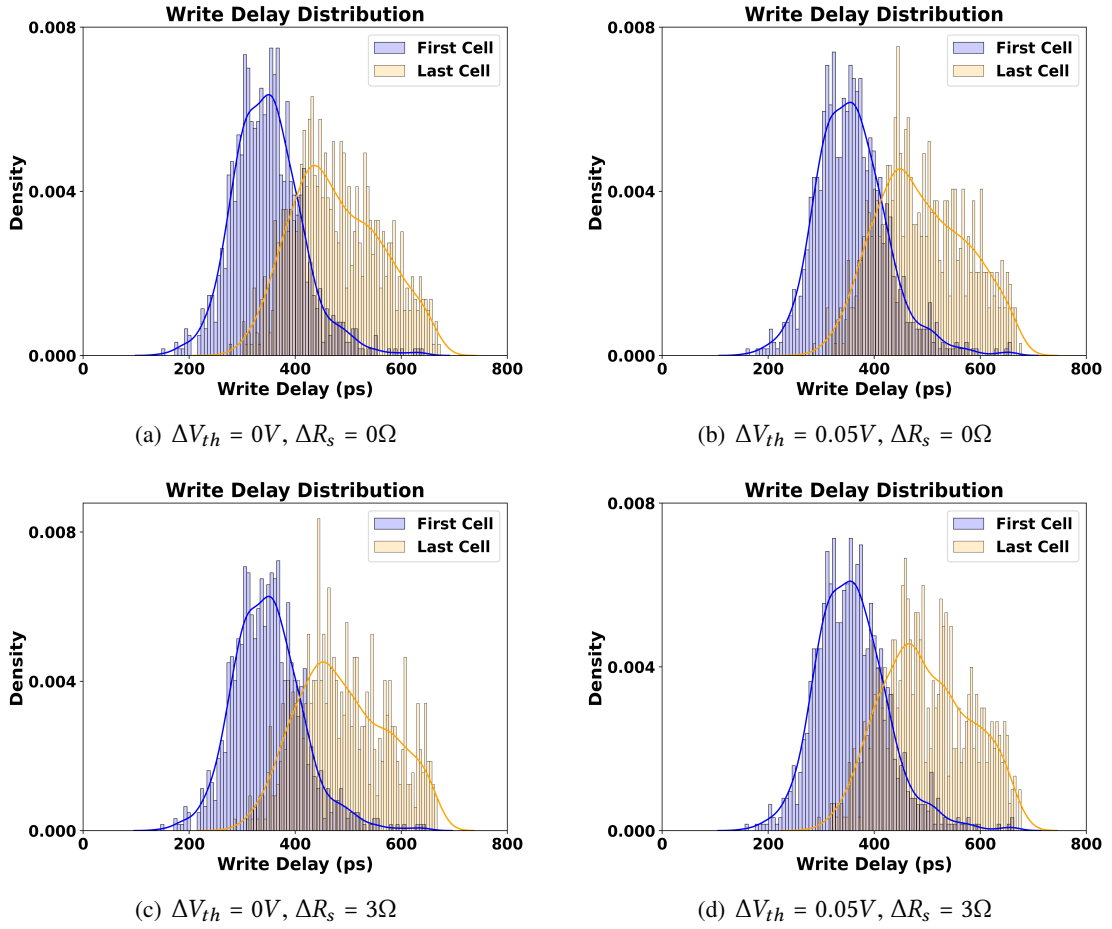


Figure 8.4.: Impact of process variation and cell location on T_{switch} and write failure probability (First vs. Last cell), considering early-stage of NBTI and EM.

Table 8.2.: Impact of process variation and cell location on T_{switch} and write failure probability (First vs. Last cell). μ : mean of T_{switch} , P_{delay} : probability of delay faults, P_{stuck} : probability of stuck-at faults.

Case	μ (ps)	P_{delay} (%)	P_{stuck} (%)
(a)	348.1/478.4	0.4/9.3	0.4/26.7
(b)	357.7/489.6	0.4/12.5	0.5/29.8
(c)	348.6/490.5	0.4/13.4	0.4/32.7
(d)	358.2/498.1	0.4/14.5	0.5/37.4

to the first cell. Moreover, the corresponding stuck-at failure probability is up to two orders of magnitude higher. This demonstrates a strong spatial dependence of write reliability within the SRAM array. This location dependence becomes more pronounced in process variation under combined NBTI and EM aging, highlighting the vulnerability of cells located farther from the write driver.

8.2.3. Sensitization of Critical Write Operation with Test Vector

To explore how a pre-operation influences the sensitization of the critical write operation as the test vector, potentially leading to a prolonged T_{switch} and thereby enabling earlier detectability, we devised a sequence of operations involving two specifically chosen SRAM cells. The first operation was performed

in $Cell_0$, the cell nearest to the driver, followed by a write operation in $Cell_{128}$, the cell farthest from the driver, to establish a worst-case scenario. The series of operations began with various pre-operations at $Cell_0$, including writing '1' $\rightarrow$ '1' (W1), '0' $\rightarrow$ '0' (W0), W1 $\uparrow$, W0 $\downarrow$, and reading '0' (R0) or '1' (R1). These pre-operations were designed to evaluate their respective impacts on the subsequent critical write operation at $Cell_{128}$.

The detailed sequence of operations is as follows:

1. Activation of the WL for $Cell_0$.
2. Activation of the write-enable signal for writing or the sense amplifier-enable signal for reading.
3. Execution of one of the pre-operations in $Cell_0$, which could be either writing or reading.
4. Deactivation of the WL for $Cell_0$ and pre-charging the BL and BLB.
5. Deactivation of the pre-charger and activation of the WL for $Cell_{128}$ as well as the write-enable signal.
6. Writing to $Cell_{128}$.

In the FDSOI SRAM tests, among all the pre-operations evaluated, an R0 operation before a W1 $\uparrow$ operation was found to significantly slow down the switching process in $Cell_{128}$. Specifically, for a fresh device, the switching time (T_{switch}) increases from 521.7 ps to 535.4 ps, representing a 2.6% increase. In contrast, no single pre-operation substantially influenced the switching speed in FinFET SRAM. Generally, the dependency between the consequent SRAM operations can be explained by the leakage profile. So, this discrepancy observed in FinFET can be attributed to the generally lower leakage properties compared to FDSOI.

As shown in Fig. 8.1 and Fig. 8.5, applying a suitable pre-operation test vector before a W1 $\uparrow$ operation shortens the time-to-failure in FDSOI SRAM compared to the nominal operation corner, meaning that the write operation becomes more vulnerable to failure under aging. Assuming a ΔV_{th} of 0.15 V, which corresponds to about 1.13 years of NBTI-only degradation, the T_{switch} remains almost the same at both the nominal and test operation corners, allowing successful writes. However, delay and stuck-at faults can occur when the combined effects of NBTI and EM are considered, as summarized in Table 8.3. For this ΔV_{th} , the ΔR_s threshold required to induce a delay fault or a stuck-at fault during a W1 $\uparrow$ write becomes significantly lower when a read '0' is used as a pre-operation test vector.

8.2.4. Sensitization of Critical Write Operation with Test Corner

To investigate how supply voltage scaling aids earlier detection of write faults, we slightly reduced the V_{peri} . In contrast, we refrained from decreasing V_{cell} as our tests indicated that even a slight decrease in V_{cell} could introduce functional faults in SRAM for both technologies investigated. As illustrated in Fig. 8.6 (FDSOI) and Fig. 8.7 (FinFET), lowering the V_{peri} increases the T_{switch} of fresh devices and enhances the sensitivity of SRAM to the combined impact of NBTI and EM, resulting in a significantly reduced time-to-failure for the write operation. These switching time degradations for fresh devices at

Table 8.3.: Sensitization-induced changes in fault-threshold under a short-term ΔV_{th} stress.

Technology	Nominal Corner	Test Vector
FDSOI (Delay fault)	2.96 Ω	1.42 Ω (-52.0%)
FDSOI (Stuck-at fault)	14.52 Ω	13.56 Ω (-6.6%)
FinFET (Delay fault)	No noticeable change	
FinFET (Stuck-at fault)	No noticeable change	

		$\Delta V_{th} (V)$									
		0	0.05	0.1	0.15	0.2	0.25	0.3	0.309	0.35	0.4
$\Delta R_s (\Omega)$	0	535.4	553.2	569.1	577.9	583.5	587.5	596.2	603.9	608.7	633.2
	0.14	537.8	554.8	570.6	580.0	585.8	598.0	600.1	602.3	610.3	636.0
	0.81	544.8	567.5	581.0	589.9	596.5	600.8	612.9	615.6	621.4	645.2
	1	549.1	567.9	584.0	593.1	599.7	604.0	616.4	617.6	631.3	645.6
	1.02	548.8	568.2	584.3	593.4	600.2	604.3	616.8	617.9	624.6	645.8
	1.42	554.7	574.3	590.9	600.3	607.0	611.7	622.5	624.4	630.9	653.2
	1.96	562.5	582.3	600.4	610.0	616.5	621.6	632.2	634.7	642.7	668.7
	2	563.8	583.0	601.3	610.8	617.9	622.3	633.3	635.5	608.7	670.2
	3	579.0	599.8	620.8	632.0	637.5	643.0	653.7	656.2	658.0	677.1
	3.03	579.0	600.4	619.3	632.7	638.2	643.8	654.9	656.7	655.9	678.0
	4	593.1	618.3	643.2	655.4	662.5	656.0	666.2	669.0	690.0	704.4
	4.42	601.8	633.3	654.2	666.3	674.7	667.7	678.7	680.8	701.3	645.6
	:	:	:	:	:	:	:	:	:	:	:
	12.39										
	12.40										
	12.49										
	12.56										
	12.73										
	13										
	13.14										
	14										
	14.17										

Figure 8.5.: Delay and stuck-at faults caused by NBTI and EM with the pre-operation test vector in the FDSOI SRAM. The color coding is the same as in Fig. 8.1.

the proposed voltage test corner are summarized in Table 8.4, demonstrating that fresh FinFET SRAM exhibits stronger sensitivity to a reduction in V_{peri} than the FDSOI SRAM.

Considering the aging conditions consistent with the pre-operation test vector analysis in the previous subsection, FDSOI SRAM can still successfully perform write operations when only the NBTI effect is present. However, when EM is included, the resistance degradation substantially accelerates write faults. As summarized in Table 8.4, the ΔR_s threshold required to induce delay faults during a $W\uparrow$ write is significantly reduced at the voltage test corner compared to the nominal operation corner. Although stuck-at faults also occur earlier under the reduced V_{peri} condition, their sensitivity to voltage scaling in FDSOI SRAM remains less pronounced. In contrast, FinFET SRAM exhibits a smaller reduction in the failure threshold, indicating a relatively lower sensitivity to V_{peri} scaling for both delay and stuck-at faults.

8.2.5. Earlier Detection of Aging-Induced Write Failures

We demonstrate the effectiveness of our proposed earlier detection method by focusing on stuck-at write faults in FDSOI SRAM as a case study. The performance of the proposed earlier detection method for stuck-at write fault detection in both FDSOI and FinFET SRAM is shown in Table 8.5. Early stuck-at write faults occur after a ΔV_{th} due to NBTI in the stagnation region (approximately 0.165 V, 6.25 years) and a

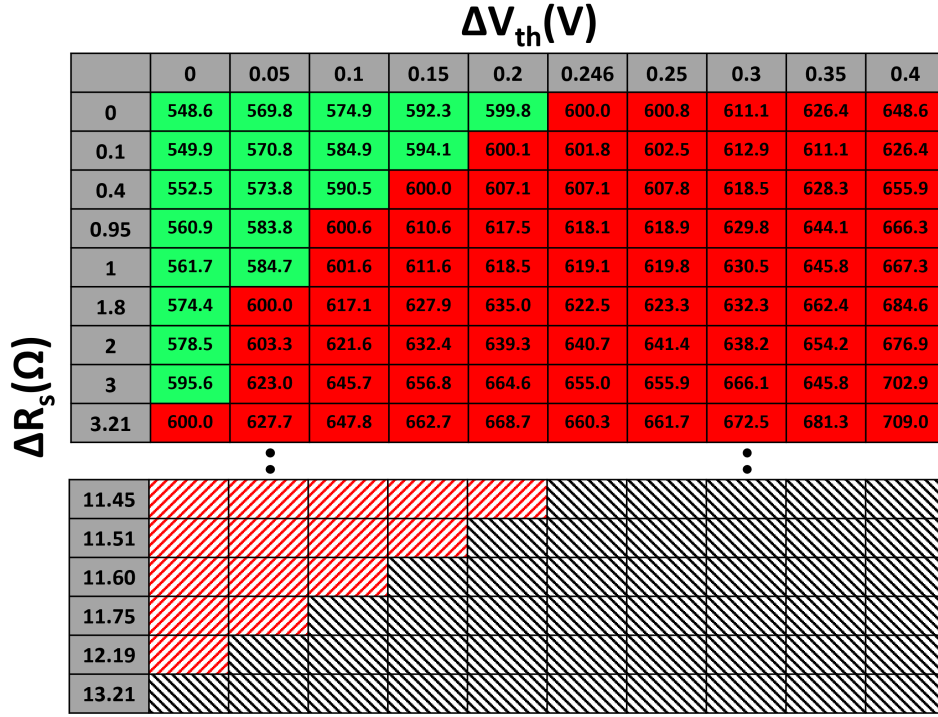


Figure 8.6.: Delay and stuck-at faults caused by NBTI and EM at the voltage test corner in the FDSOI SRAM. The color coding is the same as in Fig. 8.1.

Table 8.4.: Switching-time degradation of fresh devices and fault-threshold resistance shifts under the same NBTI stress as Table 8.3

Technology	Metric	Nominal Corner	Voltage Corner
Fresh devices: switching time T_{switch}			
FDSOI	T_{switch}	521.7 ps	548.6 ps (+5.2%)
FinFET	T_{switch}	155.7 ps	170.9 ps (+9.8%)
Aged devices: fault-threshold ΔR_s			
FDSOI	Delay fault	2.96 Ω	0.40 Ω (-86.5%)
	Stuck-at fault	14.52 Ω	11.60 Ω (-20.1%)
FinFET	Delay fault	4.33 Ω	2.34 Ω (-46.0%)
	Stuck-at fault	7.20 Ω	6.0 Ω (-16.7%)

ΔR_s of approximately 14.62 Ω (8.3 years) at the nominal operation corner. At this point, EM becomes the dominant failure mechanism. By applying the proposed pre-operation test vector, the combined effects of NBTI and EM are detected at an earlier stage compared to the nominal operation corner, as summarized in Table 8.5. Furthermore, the voltage test corner accelerates the detection even further compared to the test vector. Notably, earlier detection in FinFET SRAM can only be achieved through the voltage test corner, indicating a stronger sensitivity to supply-voltage variations compared to FDSOI SRAM.

8.2.6. Aging-Aware DVFS for Device Lifetime Extension

As a part of the SLM infrastructure, we have developed DVFS strategies to mitigate the combined effects of NBTI and EM and be able to prolong device lifetime using graceful degradation. Our approach involves identifying a region of interest where voltage scaling alone or both frequency and voltage scaling are required. We also define a metric to quantify the cost of frequency downscaling.

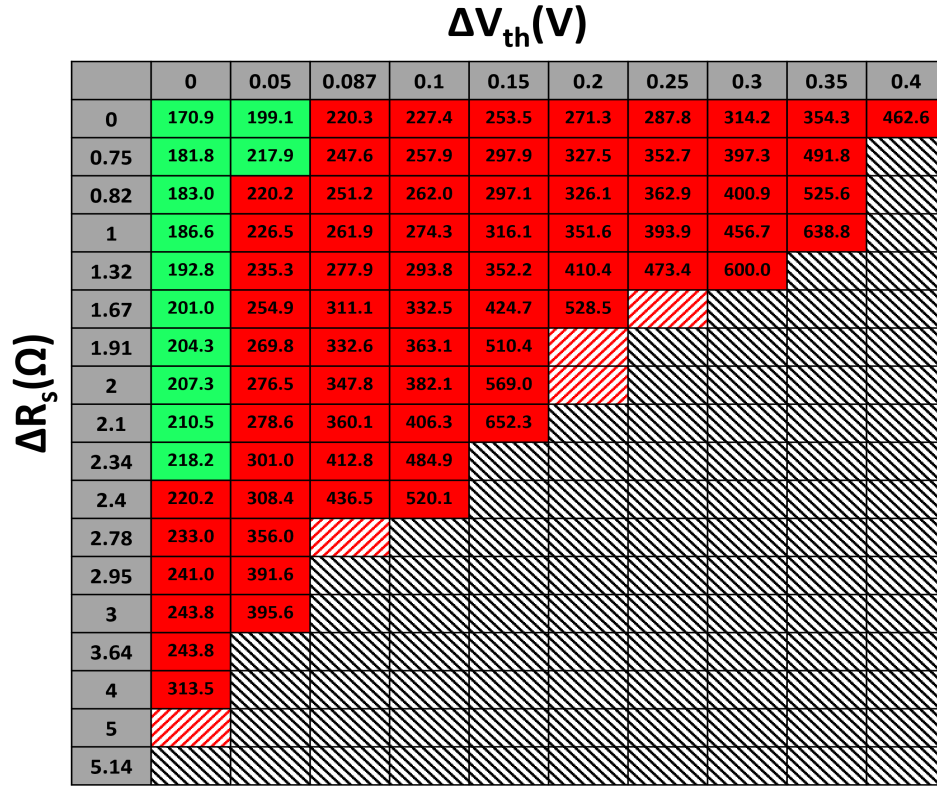


Figure 8.7.: Delay and stuck-at faults caused by NBTI and EM at the voltage test corner in the FinFET SRAM. The color coding is the same as in Fig. 8.1.

Table 8.5.: Earlier detection performance

Technology	ΔV_{th} stagnation	Mechanism	Earlier Detection
FDSOI	0.165 V	Test vector	1.15 years
		Voltage corner	1.73 years
FinFET	0.13 V	Voltage corner	3.10 years

As previously mentioned, the simulations were performed with a reduced V_{cell} and an increased V_{peri} . We then defined the new critical T_{switch} by reducing the latency at the subarray level by a maximum of 50%. For FDSOI SRAM, a T_{switch} below 1.6 ns enables successful writes with frequency downscaling, while for FinFET SRAM, the threshold is 920 ps. The results in Fig. 8.8 (FDSOI) and Fig. 8.9 (FinFET) highlight several important findings related to aging-aware DVFS. We identify regions that are exclusively suitable for voltage scaling with the following method:

1. Finding the non-stagnation region of ΔV_{th} due to NBTI.
2. Considering only EM to estimate the critical ΔR_s needed for a successful write operation.
3. Identify the range with 50% of the critical ΔR_s and the non-stagnation ΔV_{th} .

Voltage scaling has proven effective at extending the time to failure of write operations, particularly in preventing stuck-at faults. Assuming an EM-only scenario, DVFS can extend the critical ΔR_s to around 10 times its nominal value for FDSOI SRAM and 5 times for FinFET SRAM. Moreover, when accounting for the combined effects of NBTI and EM, adjusting the frequency and supply voltage significantly enhances the resilience of SRAM. As summarized in Table 8.6, a DVFS strategy enables both FDSOI

		$\Delta V_{th}(V)$								
		0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4
$\Delta R_s(\Omega)$	0	388.7	394.5	399.7	404.7	418.4	426.7	438.1	453.7	479.8
	10	417.8	424.6	430.0	435.9	447.0	457.4	468.1	484.8	509.1
	20	455.0	463.0	469.8	476.3	487.0	496.0	505.3	520.9	545.8
	29.9	507.9	517.2	526.9	531.6	541.4	543.7	559.2	576.5	600.0
	30	508.5	517.4	527.6	532.3	542.1	544.5	560.0	577.9	600.8
	32.8	527.7	538.4	549.1	553.2	556.3	568.2	582.5	600.5	624.7
	35.48	549.6	560.8	572.4	577.6	583.0	594.3	600.1	617.6	641.6
	36.95	561.9	574.7	587.8	595.7	598.2	600.2	616.6	634.2	659.0
	37.12	564.1	576.4	590.0	597.2	600.0	602.1	618.7	636.2	661.0
	37.39	566.8	579.2	588.5	600.1	603.5	609.6	621.9	639.8	664.4
	38.44	577.4	590.0	600.0	612.5	604.9	622.6	634.9	652.0	676.0
	39.24	584.6	600.1	607.3	622.5	615.8	633.2	645.6	662.4	687.2
	40	593.3	607.5	616.8	631.7	626.1	630.4	645.6	669.0	687.0
	40.52	600.1	614.3	623.5	639.3	635.1	639.6	652.3	676.7	696.9
	50	808.3	836.1	873.8	891.0	872.0	879.6	892.4	911.5	937.0
	54.72	1182.3	1354.6	1471.5	1531.1	1526.9	1541.0	1557.4	1542.9	1599.9
	54.82	1200.1	1388.5	1516.2	1571.4	1565.8	1581.1	1597.8	1594.8	
	54.85	1205.9	1397.5	1555.5	1590.7	1581.6	1597.1			
	54.86	1207.8	1410.0	1550.7	1596.5	1594.2				
	54.96	1226.7	1430.0	1599.3						
	55.34	1311.7	1591.4							
	56.21	1598.5								
	60									

Figure 8.8.: DVFS strategy to mitigate delay faults caused by NBTI and EM at the nominal operation corner in the FDSOI SRAM. Delays are shown in picoseconds. Green entries indicate successful write operations within the critical T_{switch} . Solid red entries indicate unrecoverable delay or stuck-at faults. Solid red entries with green triangles denote faults that can be recovered by 50% frequency downscaling. White entries represent regions where frequency scaling is necessary.

and FinFET SRAM to withstand substantially larger ΔR_s due to EM compared to the nominal operation corner. This improvement is especially notable for FinFET SRAM. These results indicate that properly selected voltage and frequency settings can effectively extend the operational lifetime of SRAM under combined aging conditions and prevent functional failures.

Furthermore, evaluating the trade-off associated with extending the tolerable ΔR_s through frequency downscaling is essential, since excessive reductions in operating frequency can introduce unacceptable performance degradation in high-performance applications. As summarized in Table 8.7, increasing the ΔR_s margin via a DVFS strategy results in a higher T_{switch} , and the incurred performance penalty grows nonlinearly as the reliability margin increases. This trend is observed in both FDSOI- and FinFET SRAM devices, although with different magnitudes. FDSOI SRAM exhibits a stronger increase in T_{switch} for a given resistance margin extension, indicating a higher sensitivity to EM in this DVFS trade-off. In contrast, FinFET SRAM shows a comparatively lower delay degradation. Therefore, a carefully balanced DVFS configuration is required to improve write-margin robustness without incurring excessive performance loss under combined NBTI and EM aging. Notably, DVFS increases energy consumption. For the FDSOI SRAM, the total energy rises from 5.87 μJ to 6.34 μJ for an operation time of 2 ns, corresponding to an increase of approximately 8%.

		$\Delta V_{th} \text{ (V)}$								
		0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4
$\Delta R_s \text{ (}\Omega\text{)}$	0	122.7	127.5	131.7	135.0	138.1	141.3	144.3	147.6	152.3
	5	138.7	146.7	152.6	157.2	160.1	163.8	167.5	171.9	177.1
	10	163.3	175.1	184.4	191.3	195.5	200.2	205.1	210.5	218.0
	10.2	164.4	186.0	186.0	192.8	197.3	202.1	207.1	212.7	220.1
	10.88	168.4	181.5	191.8	199.0	204.2	209.0	212.9	220.1	228.1
	11.42	172.1	185.9	196.6	204.0	209.5	215.0	220.1	226.3	234.7
	11.9	175.4	190.9	201.6	209.7	214.8	220.4	226.1	232.7	241.5
	12.32	179.3	194.4	206.7	214.5	220.1	226.3	231.9	238.4	248.5
	12.76	182.7	199.3	212.4	220.1	226.8	232.6	238.6	246.1	255.7
	13.42	187.5	206.2	220.1	225.2	236.5	241.3	247.9	257.0	266.5
	14.6	199.1	220.2	238.9	248.5	258.3	265.5	273.2	281.5	295.4
	15	203.3	226.8	244.0	257.4	267.7	274.7	279.2	292.0	304.1
	16.39	220.2	251.3	275.1	292.7	307.9	317.8	329.0	341.8	358.5
	18.92	267.0	330.1	405.3	453.7	486.3	530.8	587.3	676.1	909.6
	19.25	277.5	352.4	445.6	521.2	574.4	622.0	718.3	908.1	
	19.49	284.5	372.6	482.2	580.4	654.7	755.7	908.9		
	19.67	289.6	378.5	514.8	644.1	754.4	911.8			
	19.83	295.7	398.2	554.7	734.3	917.1				
	20	302.7	419.0	591.4	885.8					
	20.03	318.5	422.8	619.9	910.9					
	20.43	295.7	479.8	906.4						
	21.32	374.9	917.1							
	23.05	907.7								
	25									

Figure 8.9.: DVFS strategy to mitigate delay faults caused by NBTI and EM at the nominal operation corner in the FinFET SRAM. The color coding is the same as in Fig. 8.8.

Table 8.6.: Improvement in tolerable ΔR_s due to DVFS under the same ΔV_{th} aging condition as Table 8.3

ΔR_s tolerance margin improvement		
Technology	Nominal Corner	DVFS
FDSOI	2.96 Ω	54.96 Ω (18.6X)
FinFET	0.68 Ω	20.03 Ω (29.46X)

In summary, although DVFS improves aging tolerance, it also leads to reduced operating frequency and increased energy consumption, highlighting the inherent trade-off between reliability and performance.

8.3. Conclusion

We investigate the combined impact of negative bias temperature instability (NBTI) and electromigration (EM) on the writability of SRAM under worst-case conditions. Our methodology prevents lifetime overestimation and enables earlier detection of write faults using tailored test vectors and voltage test corners. We also propose an aging-aware Dynamic Voltage and Frequency Scaling (DVFS) scheme that mitigates degradation and extends operational lifetime. Our results show that NBTI and EM jointly

Table 8.7.: T_{switch} degradation under DVFS with ΔR_s scaled in 1.5 $\times$ increments

T_{switch} degradation		
Technology	ΔR_s margin	T_{switch}
FDSOI	30 Ω	532.3 ps (Baseline)
	45 Ω	749.3 ps (1.41X)
	67.5 Ω	Unrecoverable
FinFET	10 Ω	191.3 ps (Baseline)
	15 Ω	257.4 ps (1.35X)
	22.5 Ω	Unrecoverable

accelerate delay and stuck-at faults, with FinFET SRAM exhibiting higher sensitivity than FDSOI. Furthermore, a strong spatial dependence of write reliability is observed, where cells farther from the write driver exhibit significantly higher delay and failure probability under aging. These findings support the development of methods for early detection and mitigation of aging across the device lifecycle, particularly within Silicon Lifecycle Management (SLM). Future work will extend this framework to include process, voltage, temperature (PVT) variations and Monte-Carlo analysis to evaluate the statistical and spatial robustness of SRAM under combined NBTI and EM aging, as well as extend the framework to sub-10 nm technologies and integrate in-field monitors for predictive SLM in safety-critical domains such as automotive and aerospace.

9. Analysis and Mitigation of Radiation Effects in SRAM-based Register Files

Continuous transistor scaling in nanoscale technology nodes has enhanced the performance and energy efficiency of VLSI systems, driving advancements in high-performance computing (HPC). To meet stringent PPA requirements, HPC architectures, especially GPUs, increasingly employ SRAM-based register files in place of traditional latches or flip-flops, thereby improving efficiency in compute-intensive tasks such as AI training and inference.

While advanced technology nodes offer significant performance improvements, they also introduce critical reliability challenges. In particular, radiation-induced soft errors remain a major concern for SRAM-based memory structures in nanoscale technologies, especially in reliability-critical HPC applications. The reliability of SRAM-based register files is therefore crucial for robust system operation.

Conventional radiation studies on SRAM mainly focus on the hold state, since the hold duration dominates the operating time of SRAM arrays. Consequently, most SRAM soft-error analyses primarily investigate radiation-induced failures during data retention. However, SRAM-based register files exhibit substantially different access behavior compared with conventional SRAM caches. In register files, read operations occur very frequently, and the read duration can become comparable to the hold duration during normal operation. As a result, not only the hold state but also read operations can become susceptible to radiation-induced soft errors.

While the vulnerability of SRAM caches to radiation-induced errors during hold operation has been extensively studied, research on SRAM-based register files has primarily focused on design and optimization, with limited attention to their behavior under radiation during read operations. Therefore, the impact of radiation-induced soft errors during SRAM register-file read operations remains largely unexplored.

Therefore, this chapter addresses these gaps by evaluating the resilience of SRAM-based register files to radiation during access time, analyzing the impact of various factors on soft error probability, and proposing architectural- and circuit-level mitigation strategies, based on the work published in [7]. The rest of this chapter is structured as follows: Section 9.1 introduces radiation current models and SRAM-based register files. Section 9.2 describes the methodology, including TCAD calibration and radiation simulation. Section 9.3 evaluates SRAM-based register file readability under the influence of radiation. Section 9.4 details the proposed technology-dependent mitigation methods, and Section 9.5 concludes the work.

9.1. Related Work

HPC systems, for example, those that incorporate GPUs, utilize distributed processing units known as streaming multiprocessors (SMs) for parallel processing. Workloads are partitioned into thread arrays (e.g., cooperative thread arrays, or CTAs) dynamically scheduled across SMs. In this structure, the performance highly depends on efficient hardware-managed context switching between groups of active threads. This requires rapid state retention via SRAM-based register files, which are preferred over latches/flip-flops for PPA efficiency. Parallelism is fundamentally constrained by register file capacity, driving the progressive scaling of resources [104]. However, increased parallelism elevates access frequency, exposing reliability vulnerabilities: unlike RAM structures, for which data hold is predominant,

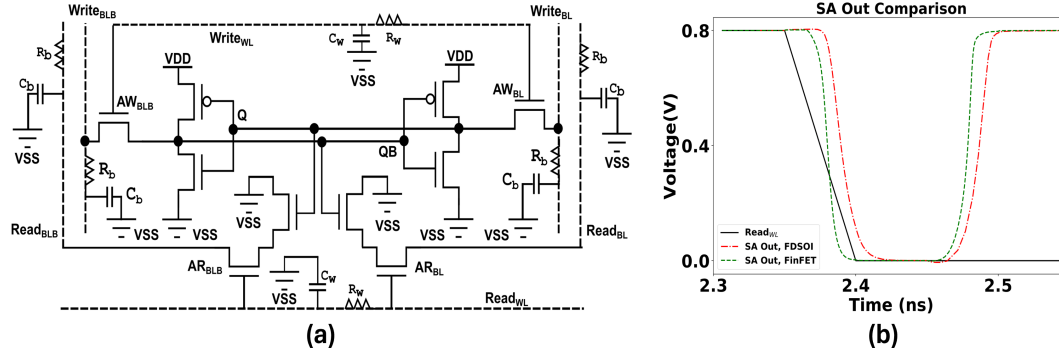


Figure 9.1.: (a): The one-write-one-read (1W1R) register file bit-cell includes labeled driver locations (e.g., write bit line, $Write_{BL}$), access transistors (e.g., AW_{BL}) connected to bit lines (BL and BLB), and parasitic elements (e.g., R_b , C_b) on BL and BLB. (b): Comparison of SA switching speed for FDSOI and FinFET SRAM-based register files at $WL = 300$ ps for the nearest cell to the SA.

register files operate in active access modes during execution and their radiation susceptibility remains underexplored.

9.2. Radiation Modeling and Testing

In this section, we present our TCAD-based radiation modeling and testing method to evaluate the reliability of an SRAM-based register file (128 rows, 40 columns), focusing on radiation effects during read access time. Since write operations are inherently more resilient to radiation-induced soft errors, our analysis primarily focuses on read operations. In addition, error probability is evaluated based on various key factors such as cell location. The key steps are shown below.

9.2.1. TCAD Model Calibration

The Sentaurus Device Simulator is used to simulate N- and P-type transistors using the drift-diffusion model, calibrated to process design kit (PDK) specifications. TCAD parameters for FDSOI and FinFET are detailed in previous chapters.

9.2.2. Modeling Radiation-Induced Current

We use TCAD to simulate radiation effects with a track length of $1 \mu\text{m}$ and a Gaussian width of 10 nm . For the FDSOI register file, the worst-case error probability stabilized at approximately 30% for linear energy transfer (LET) exceeding $60 \text{ MeV} \cdot \text{cm}^2/\text{mg}$, which we set as the baseline LET for testing with a range of $\pm 20\%$ to evaluate the effect of particle energy variation on error probability. However, for the FinFET register file, these energy levels lead to nearly 100% error probability, as FinFETs require a much lower LET to match the error probability of 22nm FDSOI [105]. To ensure comparability, a LET of $1.82 \text{ MeV} \cdot \text{cm}^2/\text{mg}$ is set as the baseline. Although LET ranges differ across technologies, the trends in error probability remain consistent, enabling comparative analysis of technology scaling effects on radiation-induced failures.

9.2.3. Identifying Impactful Factors

We investigate the effects of radiation on the SRAM-based register file by exploring the effect of several factors. The analysis considers multiple dependencies affecting error probability. Since charge collection

dynamics vary depending on stored bit value, we evaluate error susceptibility separately for reading a cell storing logic ‘0’ ($R0$) and logic ‘1’ ($R1$). Location dependency is examined using three test cells in column 39, the farthest from the WL drivers, based on their distance from the sense amplifier (SA): the nearest cell (row 127), the farthest cell (row 0), and the middle cell (row 63). Moreover, radiation energy dependency is assessed using different LETs for the FDSOI and FinFET register files, as described in Section 9.2.2, to analyze the effects of LETs and technology scaling on error probabilities. In addition, read word-line activation (WL) width dependency is investigated by varying the WL width between 200 ps and 300 ps in 25 ps steps while maintaining an operating frequency of 1.3 GHz.

9.2.4. Calculating Error Probability

Error probabilities are determined by monitoring the SA output during read operations. The output is measured at the midpoint of the SA enable signal before pre-charging read bit lines ($Read_{BL}$ and $Read_{BLB}$) for the subsequent read operation. An error is recorded if the outputs of the SA fail to reach the 90% proximity of the expected value.

9.3. Results and Discussion

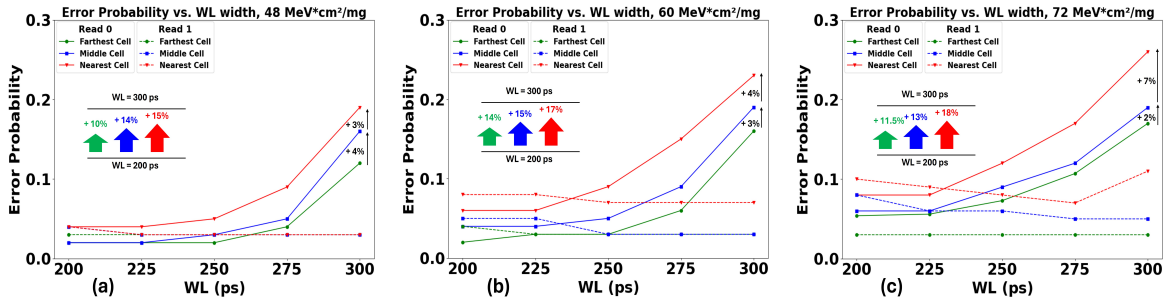


Figure 9.2.: Error probability of the FDSOI SRAM-based register file under various linear energy transfer (LETs), with arrows representing the increase in error probability for $R0$ operations. LET: (a) 48 MeV*cm²/mg, (b) 60 MeV*cm²/mg, (c) 72 MeV*cm²/mg.

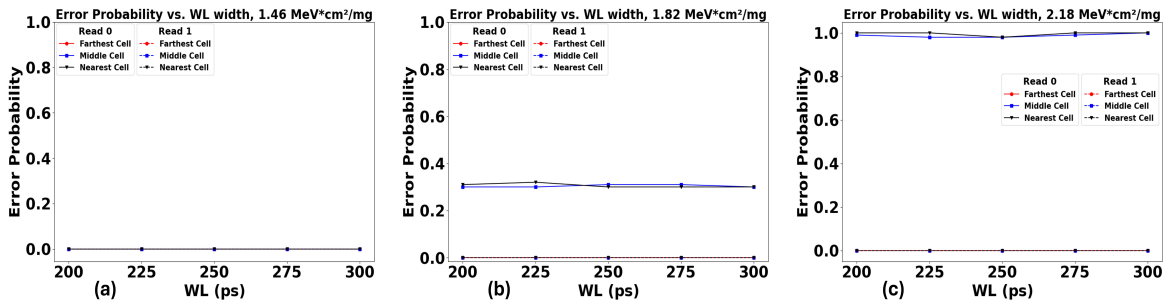


Figure 9.3.: Error probability of the FinFET SRAM-based register file under various LETs. LET: (a) 1.46 MeV*cm²/mg, (b) 1.82 MeV*cm²/mg, (c) 2.18 MeV*cm²/mg.

This section presents the key findings from the radiation testing of SRAM-based register files, focusing on factors affecting read error probability, including data type, cell location, strike energy, technology, and their dependencies on RC load and WL width.

9.3.1. Simulation Setup

The one-write-one-read (1W1R) register file bit-cell (Fig. 9.1(a)) comprises two cross-coupled inverters for data storage and four access transistors, enabling independent read and write operations. The writing and reading processes are controlled by dedicated WL (WLB) and BL (BLB). The parasitic RCs per cell are summarized in Chapter 6 and Chapter 7. Electrical-level simulations are performed with 1000 Monte Carlo iterations to account for the impact of process variation.

9.3.2. Impact of RC Load

During a read operation, either $Read_{BL}$ or $Read_{BLB}$ discharges while the other remains at a high pre-charged level. The transient decay rate of the discharging BL determines the register file's susceptibility to radiation-induced errors. The voltage decay, modeled by Equation (9.1), assumes an initial pre-charge of V_{dd} , where R and C represent the total resistive and capacitive load of the discharge path:

$$V(t) = V_{dd} \cdot e^{-\frac{t}{\tau}}, \tau = R \cdot C \quad (9.1)$$

The transient discharge rate, given by $\frac{dV}{dt} \propto -\frac{1}{RC}$, influences register file stability during read operations. A lower RC time constant (τ) accelerates BL discharge, reducing the SA's noise margin and increasing vulnerability to radiation-induced transients. This model characterizes SRAM-based register file behavior under radiation during read operations.

9.3.3. Impact of Read Word-Line Activation Pulse-Width

As shown in Fig. 9.1(b), the FinFET register file has a much smaller sensitive time window to radiation due to the fast charging and discharging speed of read bit-lines. As a result, as seen in Fig. 9.2(a) to Fig. 9.2(c), the FDSOI register file exhibits a strong dependency on WL width, as a longer WL prolongs data retention during read operations, increasing susceptibility to radiation-induced errors. Expanding WL from 200 ps to 300 ps can raise error probability by up to 18%. In contrast, the FinFET register file is more sensitive to $R0$ operations, but WL width does not affect its error probability (Fig. 9.3).

9.3.4. Impact of Read Operation

Fig. 9.2 and 9.3 show radiation-induced error probabilities during read operations. In both FDSOI and FinFET SRAM-based register files, $R0$ operations exhibit significantly higher error rates than $R1$. In the FDSOI register file, the error probability for $R0$ operations increases with WL width, while $R1$ operations are minimally affected by WL width. In contrast, for the FinFET register file, WL width has no effect on error probabilities for $R0$ operations. The difference in radiation susceptibility between $R0$ and $R1$ operations stems from the discharge dynamics during read operations. In an $R1$ operation, AR_{BL} discharges $Read_{BL}$, but parasitic RC on $Read_{WL}$ slightly lowers its gate voltage. In contrast, during $R0$, AR_{BLB} receives a higher gate voltage, resulting in lower equivalent resistance and faster discharge, as discussed in Section 9.3.2. This faster discharge increases the susceptibility of $R0$ operations to radiation-induced errors compared to $R1$.

9.3.5. Impact of Cell Location

Error probability analysis reveals a strong dependence on the radiation-affected cell's location relative to the SA. In the FinFET register file, the farthest cell exhibits no errors across 1000 Monte Carlo simulations for both $R0$ and $R1$ operations, while the middle and nearest cells show similar error probabilities. In contrast, for the FDSOI register file, the nearest cell consistently has the highest error probability, with the gap widening at higher radiation energy levels. Additionally, WL width significantly affects $R0$ errors

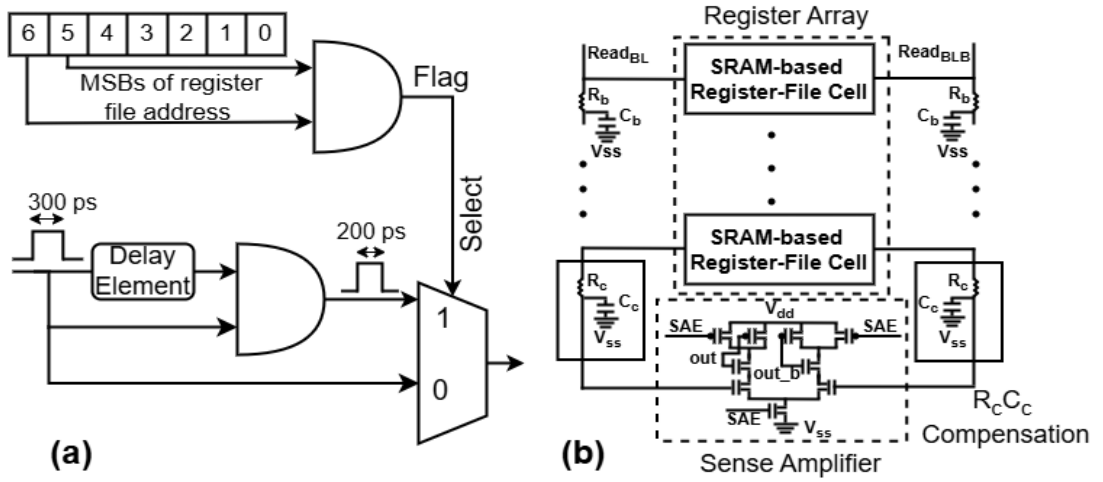


Figure 9.4.: Radiation-aware mitigation strategies. (a) FDSOI: Adaptive location-dependent WL width modulation, (b) FinFET: Compensating RC parasitic structure.

across different cell locations. As shown in Fig. 9.2(c), reducing the WL width to 200 ps significantly reduces the error probability gap between the nearest and farthest cells. However, R1 errors remain insensitive to WL width but still vary with cell location. These results highlight the critical impact of cell location, as the increased distance from the SA introduces greater RC delay, slows BL discharge, and reduces radiation susceptibility.

9.3.6. Impact of Radiation Energy

We analyze the impact of radiation energy and WL width on error probabilities in FinFET and FDSOI SRAM-based register files. As shown in Fig. 9.3, increasing LET causes a sharp rise in R0 error probability in the FinFET register file, reaching nearly 100% at high LET values. However, WL width has no observable impact in this case. Sensitivity also varies with cell location, with those closest to the SA being the most vulnerable, as discussed in Equation (9.1). In contrast, the FDSOI register file shows increased radiation sensitivity with wider WL pulses, although LET remains the dominant factor. In summary, the FinFET register file is robust at low LET but becomes highly error-prone at elevated energies, while the FDSOI register file exhibits stronger WL width dependence at low LET and error saturation at higher values.

9.4. Technology-Dependent Mitigation Strategies

For FDSOI technology, the radiation-induced errors are not only location-dependent but also susceptible to the WL width. We propose an adaptive, location-dependent WL width for the SRAM-based register files as shown in Fig. 9.4(a), with a longer WL width applied to the less susceptible farthest cells and a shorter WL width assigned to the more vulnerable nearest cells to reduce radiation impact. The flag signal for adaptive WL width can be generated using a few of the most significant bits (MSBs) of the register files' row address. The number of selected MSBs defines the group of rows in the register file that operate under the shorter WL width. For instance, selecting two MSBs will enable the 32 rows (nearest to the SA) under the effect of shorter WL. As shown in Fig. 9.2(c), decreasing the WL width (e.g., from 300 ps to 200 ps) for the nearest cell can reduce the error probability by 18%. However, while increasing WL width improves the noise margin, it also exacerbates radiation-induced errors. This trade-off must be carefully balanced when selecting adaptive WL width.

Table 9.1.: Results of error probability reduction due to mitigation for the worst-case scenario (Fig. 9.3(c)) in the nearest cell (baseline: no compensating RC) and the SRAM-based register file's read energy.

Error probability (%)	$R_c(\Omega)$	$C_c(f)$	Energy (pJ)
100%	No compensating RC		1.37
43%	10k	8f	1.55
0	10k	10f	1.58

For the FinFET SRAM-based register file, whose radiation susceptibility is relatively insensitive to WL width, we propose enhancing radiation resilience by adding a compensating RC parasitic structure at the SA inputs (Fig. 9.4(b)). Inspired by location-dependent radiation susceptibility during read operations, this approach leverages the differential nature of the sense amplifier by adding identical dummy RC components without disturbing the functionality while improving radiation resilience. Table 9.1 summarizes the results. However, this compensation method increases energy consumption by up to 15.3% during read operations under worst-case conditions.

9.5. Conclusion

This study investigates radiation-induced soft errors in SRAM-based register files implemented in 22 nm FDSOI and 12 nm FinFET technologies, with a focus on read operations rather than data hold time. Error probability is shown to depend on several factors, such as word-line width, with reads of logic '0' being more vulnerable than logic '1'. To mitigate these effects, we propose technology-specific solutions: an architecture-level adaptive location-dependent word-line width modulation for FDSOI, and a circuit-level compensating resistor-capacitor structure for FinFET. Our study highlights the need for technology-specific optimizations to enhance the resilience of SRAM-based register files in high-performance computing systems operating in radiation-prone environments.

10. ChatTCAD: Leveraging Large Language Models for Automated TCAD Simulation File Generation

Continued progress in semiconductor technology increasingly depends on innovations in both device architectures and materials to maintain improvements in performance, power efficiency, and cost. As conventional geometric scaling approaches physical limits, advanced device architectures and emerging materials are being introduced to extend CMOS scaling and address new application requirements.

However, these advances are increasingly accompanied by reliability challenges. Aging mechanisms such as NBTI, Hot Carrier Injection (HCI), and Time-Dependent Dielectric Breakdown (TDDB) contribute to long-term circuit degradation. Accurate analysis and mitigation of these effects require detailed wafer-level testing and physics-based simulation, which are both time-consuming and costly.

TCAD has become an important tool for physics-based modeling of semiconductor device behavior and reliability. However, the complexity of TCAD environments limits their accessibility for many circuit designers, especially in the context of DTCO and STCO, where efficient interaction between device, circuit, and system levels is increasingly required.

In recent years, LLMs have demonstrated strong capabilities in EDA tasks such as code generation, circuit optimization, and verification. However, the application of LLMs to TCAD automation and semiconductor reliability analysis remains largely unexplored.

This chapter is based on the work published in [8]. It introduces ChatTCAD, the first framework that leverages fine-tuned LLMs for automated TCAD device file generation through natural language interaction. ChatTCAD transforms high-level user specifications into executable and physically consistent simulation scripts for a wide range of two-dimensional devices. By automating the code generation process, ChatTCAD significantly reduces the learning curve for TCAD users and accelerates the simulation setup. The main contributions of this chapter are as follows:

- A TCL-to-Python translation approach that transforms complex TCL-based TCAD simulation commands into Python functions, improving both model training efficiency and extensibility.
- A dataset generation pipeline for the construction of synthetic datasets using GPT-4o [106], specifically for fine-tuning LLMs on TCAD device files.
- A single-shot prompting strategy for generating and debugging individual simulation functions (e.g., region generation, doping definition).
- An enhanced table-based prompting strategy that decomposes complex devices into modular subcomponents. This enables the generation of both single-device structures (e.g., transistors) and multi-device systems (e.g., inverters), expanding the versatility of ChatTCAD.

The structure of this chapter is as follows: Section 10.1 provides background on the use of LLMs in EDA and machine learning applications in TCAD simulations. Section 10.2 outlines our methodology, including the proposed workflow, prompt strategies, and our approach for generating synthetic datasets used to fine-tune the Llama-3-8B-Instruct model. Section 10.3 presents an evaluation of the fine-tuned model's performance. Finally, Section 10.4 concludes with a summary of the work.

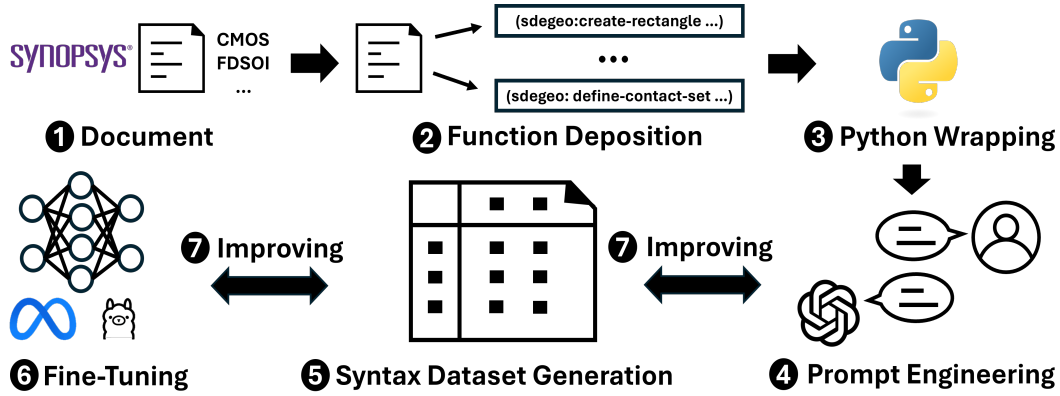


Figure 10.1.: Overall workflow of the ChatTCAD framework.

10.1. Preliminaries

10.1.1. Large Language Model in EDA

LLMs have gained prominence in the ML community for their ability to interpret both natural language and complex structured data. LLMs such as Llama-3-8B-Instruct [107] have demonstrated strong performance across a wide range of tasks, and their potential within EDA is gaining interest due to their ability to process and generate technical content. For example, Auto-SPICE is proposed in [108] for extracting SPICE netlists from analog schematics, using You Only Look Once-based component detection and Deep Hough Transform-based net recognition. In addition, LayoutCopilot [109] leverages LLM-driven design suggestions in combination with human feedback to enhance analog layout productivity.

10.1.2. Machine Learning for TCAD Simulation

ML techniques have proven effective in accelerating simulations and simplifying calibration tasks in TCAD. For example, [110] proposed a TCAD-ML framework combining model calibration with principal component analysis and polynomial regression to enable device variation analysis from current–voltage (I–V) characteristics. To date, only one study [111] has applied LLMs in TCAD, generating FinFET device files by tuning geometrical parameters. However, its focus on similar FinFET structures limits generalization to broader TCAD use cases, which has been addressed in our work.

10.2. Methodology

In this section, we introduce our proposed LLM-based framework for generating TCAD device files from natural language descriptions. The goal is to bridge the gap between circuit designers who may lack detailed knowledge of device construction and the complexity involved in writing valid TCAD device files. As shown in Fig. 10.2, the user provides a high-level natural language description of the semiconductor device, such as its doping characteristics. The framework then generates a complete, executable TCAD device file that includes all necessary components (e.g., doping and meshing). To build this system, we first review a set of representative TCAD device files and extract commonly used TCL functional blocks, which are then abstracted into structured Python functions. We use GPT-4o to generate synthetic datasets for fine-tuning the Llama-3-8B-Instruct model, selected for its strong performance and computational efficiency. After the initial fine-tuning, we conduct iterative prompt refinement to enhance dataset quality and model robustness.

10.2.1. Function Block Decomposition

Training LLMs on full device simulations is challenging due to limited data diversity and accuracy degradation with long input token sequences [112]. These issues make it particularly difficult to fine-tune lightweight LLMs on complete simulation files. To address both problems, we decompose full TCAD device files into independent functional blocks. These blocks represent key components in the device files and can be reused across different two-dimensional device types. Each block is used to fine-tune the Llama-3-8B-Instruct model individually, effectively mitigating both data scarcity and token-length limitations. Specifically, our workflow is based on five fundamental TCL code blocks:

- **Region generation:** Defines the geometric parameters, material types, and names of individual device regions. This includes attributes such as region dimensions (e.g., substrate thickness) and material assignments (e.g., silicon or oxide).
- **Doping definition:** Specifies dopant species (e.g., arsenic or boron) along with parameters such as peak concentration, expansion coefficient, and doping depth, which collectively determine the impurity distribution profile.
- **Contact definition:** Identifies the physical locations of electrical contacts. Material properties like work function are excluded at this stage, as they are typically specified later during device-level simulation.
- **Global meshing:** Applies coarse mesh settings to non-critical regions such as deep substrate layers to accelerate simulation while maintaining adequate resolution where precision is less critical.
- **Refined meshing:** Introduces fine mesh refinement in regions requiring higher accuracy, such as material interfaces (e.g., silicon/oxide) or areas with steep doping gradients. This improves numerical convergence and overall simulation precision.

10.2.2. Python Function Wrapper

TCL is not widely used in modern software development, making it a challenging target language for LLMs. Compared to more commonly supported languages such as Python, LLMs often struggle to produce accurate and reliable TCL code. Additionally, device files follow strict syntactic rules, where even minor punctuation errors can render an entire file invalid. To mitigate these challenges, each block is wrapped as a Python function. This abstraction allows the Llama-3-8B-Instruct model to generate only

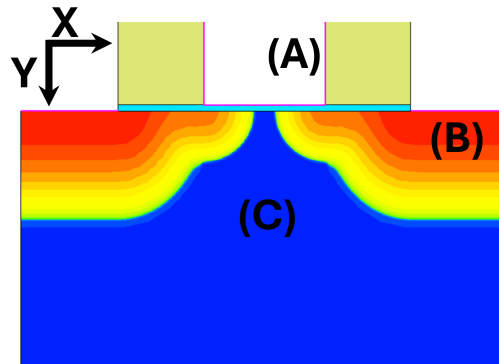


Figure 10.2.: Example CMOS device generated using the proposed ChatTCAD framework. (A) Gate Contact, (B) Source/Drain Doping and Lightly Doped Drain (LDD), (C) P-Well region.

the function arguments, rather than full TCL code. By simplifying the generation task and enforcing a structured output format, this approach significantly reduces the likelihood of syntax-related errors.

10.2.3. Prompt Engineering for Single-Function Mode

Prompt design is essential for guiding LLMs to generate accurate and structured outputs [113], particularly during fine-tuning. Our system supports two generation modes for different use cases: Single-Function Mode (SFM) and Complete-Device Mode (CDM), as shown in Fig. 10.3. SFM is intended for generating individual functional blocks, enabling focused debugging and learning. This mode is particularly useful during the early stages of development, where full-device simulation is not required. CDM, in contrast, supports the generation of complete device files from a single prompt. For SFM, we use one-shot textual prompts based on synthetic data generated by GPT-4o. For CDM, we employ a tabular Chain-of-Thought (CoT) prompting strategy, which helps guide the model through step-by-step reasoning for complete device construction. As highlighted in prior work [114], synthetic datasets often suffer from issues such as low diversity, limited robustness, and noisy inputs. We address the most relevant challenges as follows.

10.2.3.1. Lack of diversity

Synthetic datasets often overrepresent certain keywords while underrepresenting others, limiting their linguistic diversity. To better simulate the variability of real user inputs, we prompt GPT-4o to generate more natural phrasings using passive constructions, imperative forms, and paraphrasing. Building on the approach from [115], we adopt a hierarchical *trigger-root-branch-leaf* keyword strategy, as illustrated in Fig. 10.3 ①:

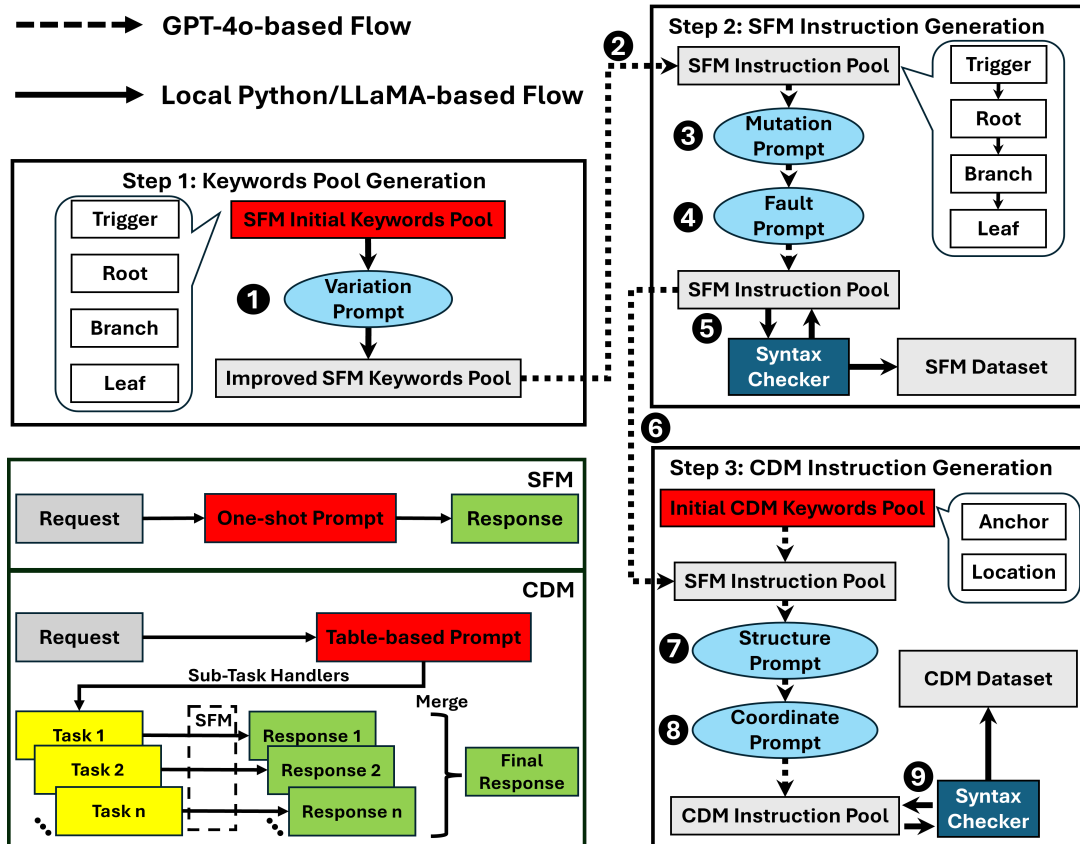


Figure 10.3.: Detailed prompt engineering and synthetic dataset generation steps for SFM and CDM.

- **Trigger:** Verbs or phrases that initiate user intent (e.g., “define”, “I want to create”).
- **Root:** The object being created (e.g., "region").
- **Branch:** Attributes or descriptors (e.g., “length”).
- **Leaf:** Specific instance names or materials (e.g., "contact", "Silicon").

We generate user-like instructions by randomly sampling keywords from each pool and combining them into a complete SFM prompt instruction, as shown in Fig. 10.3 ②.

10.2.3.2. Insufficient Mutation

Synthetic datasets may not capture the complexity and imperfections found in real-world user inputs. To address this, we introduce realistic user-like variation, including typographical errors and phrasing diversity by introducing mutations into correct samples. These include typos, sentence reordering, and variations in command length, all applied within the trigger–root–branch–leaf framework. This strategy improves model robustness and enhances generalization to noisy inputs, as illustrated in Fig. 10.3 ③.

10.2.3.3. Faulty Data

Synthetic data may lack realistic errors, leading to overfitting on perfect examples. To address this, we introduce incomplete commands and incorrect function arguments in our training data. These are paired with clear error type annotations to help the model learn to recognize common user input mistakes (Fig. 10.3 ④).

10.2.3.4. Duplicated or Incorrect Data

Since LLM-generated synthetic data may include duplicated or syntactically incorrect instructions that do not conform to the defined Python functions, we apply a syntax checker to filter out such cases. As illustrated in Fig. 10.3 ⑤, this step eliminates both redundant entries and invalid attribute combinations, resulting in a high-quality, robust pool of human-like SFM instructions.

An example of a complete SFM prompt instruction is shown in Fig. 10.4. The system prompt defines the model’s expected behavior and output format, with key instructional keywords highlighted in green. The output is required in JavaScript Object Notation (JSON) format, which has been shown to improve LLM precision compared to plain text [116]. In the user prompts, red-highlighted keywords illustrate natural variation and minor spelling errors, simulating the types of imperfect input that may occur in real-world usage. Prompt A includes all required attributes, allowing the model to generate a valid Python function. Prompt B, in contrast, serves as an example of incomplete input. It omits the material field and uses a more formal phrasing style. The model correctly identifies the missing information and returns an appropriate error. By fine-tuning the LLM on this dataset, generated using our prompt strategy, we improve its ability to handle phrasing variability and to detect incomplete or faulty user inputs.

10.2.4. Prompt Engineering for Complete-Device Mode

Prompt engineering for CDM presents additional challenges compared to SFM. The key challenges are:

- **Multi-step generation:** Constructing a complete device requires multiple interdependent steps. Simple prompting strategies, such as one-shot or few-shot learning, often fail to produce satisfactory results [117].

- **High inter-step dependency:** The generation process involves strong dependencies between steps, particularly in coordinate assignments and region sequencing, making it a complex reasoning task for the model.
- **Long sequences and limited debuggability:** Complete device files are token-heavy, which can hinder the performance of lightweight LLMs. Furthermore, one-shot or few-shot generation lacks transparency and makes debugging more difficult.

To address these challenges, we adopt CoT prompting [118] as a strategy for CDM. CoT enhances the reasoning capabilities of LLMs by encouraging them to work through intermediate steps, improving performance and robustness on complex tasks compared to conventional prompting methods. To structure this reasoning more effectively, we employ tabular CoT, which organizes information in a compact and interpretable format. This approach reduces verbosity and enables structured updates across both rows and columns [119]. It is particularly well-suited to our task, where the coordinates of regions, contacts, and doping steps are interdependent and must be computed incrementally. Tabular CoT allows the model to focus on spatial relationships with minimal distraction from irrelevant text.

10.2.4.1. Prompt Strategies for CDM

To enable text-to-simulation file translation in CDM, we define a system prompt that guides the LLM to act as a structural reasoning assistant for complete semiconductor device modeling. The prompt instructs the model to convert a natural language structure description into a table-based reasoning format, where

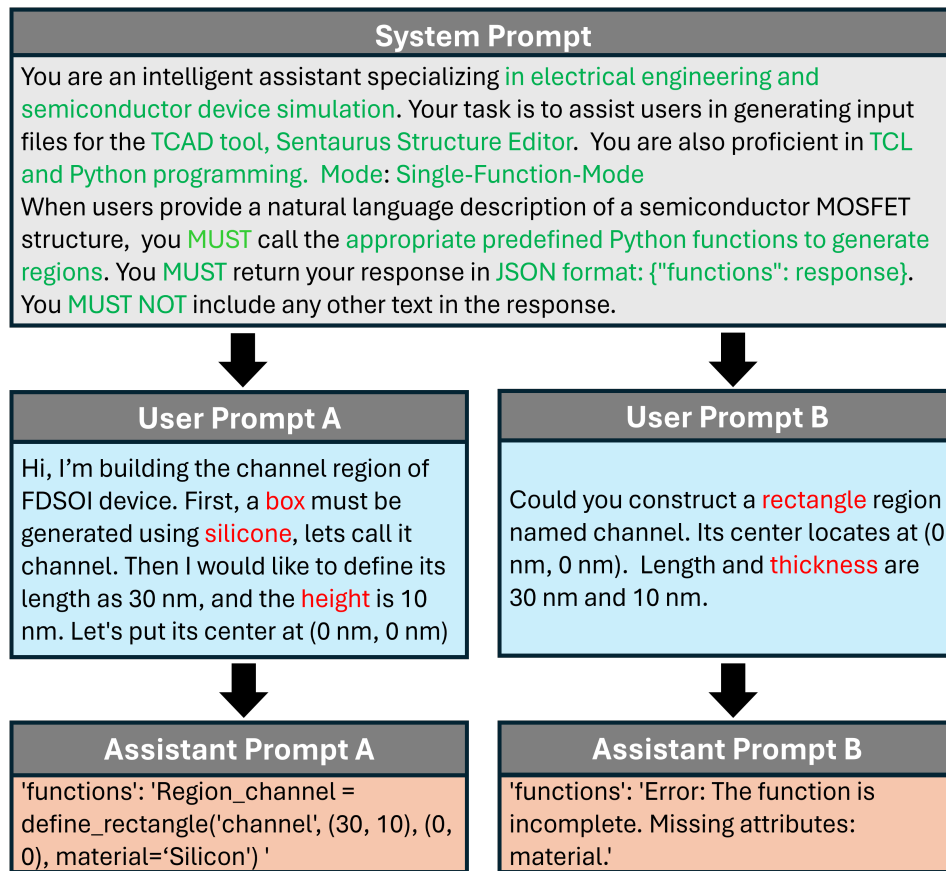


Figure 10.4.: Example of a complete SFM prompt for region block generation, with key keywords highlighted.

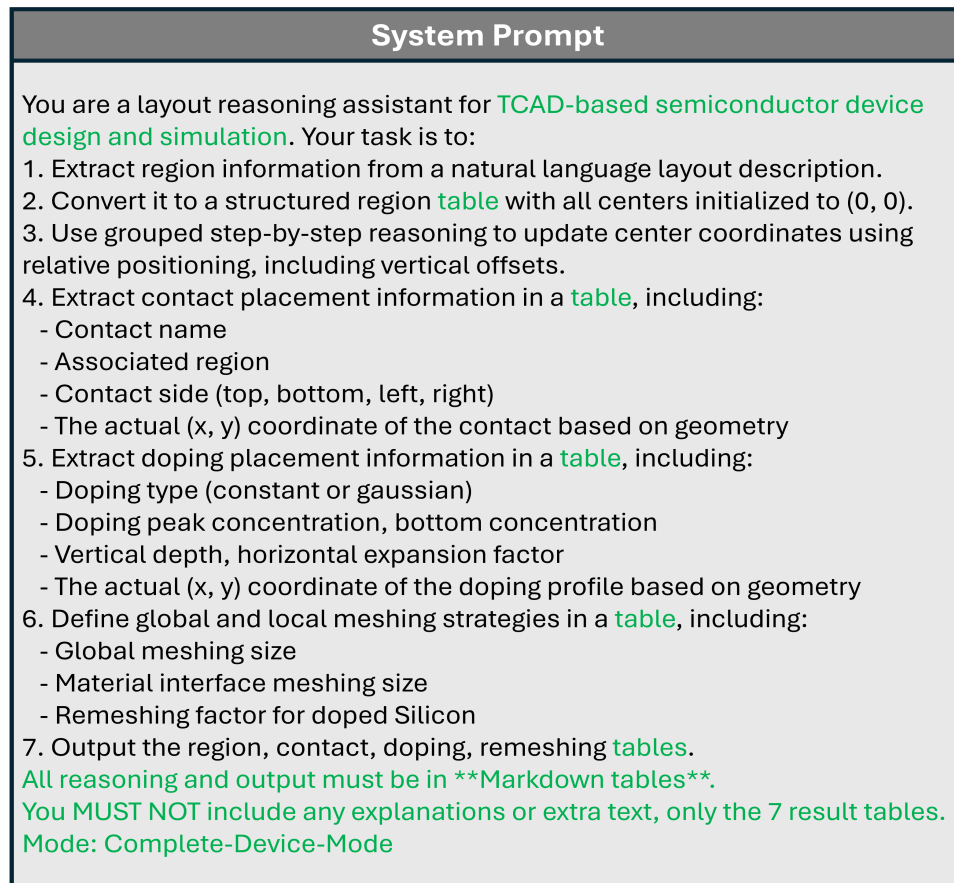


Figure 10.5.: Example of a CDM system prompt, with key instructional keywords highlighted in green.

the coordinates and attributes of each functional block (e.g., regions, contacts) are computed step by step. These resolved blocks are then represented as Python function calls and ultimately merged into a complete simulation file, as illustrated in Fig. 10.3. The system prompt in Fig. 10.5 defines the following core components, each executed through a structured, table-driven process within the user and assistant prompts in Fig. 10.6:

- **Region initialization:** Extract device regions from the input and initialize their center coordinates to (0, 0) in a region table.
- **Coordinate reasoning:** Update region coordinates incrementally using spatial relationships (e.g., ‘above’, ‘left of’), with each step recorded in a reasoning table.
- **Contact extraction:** Identify contacts, associated regions, and placement sides. Compute their coordinates and store them in a contact table.
- **Doping extraction:** Extract doping types, concentration parameters, spatial extents, and placement coordinates, organized in doping tables.
- **Meshing definition:** Define global and local mesh sizes, material interface refinement, and remeshing factors in a meshing table.
- **Output enforcement:** The model is instructed to return only Markdown tables. A Python-based function mapper then converts these structured tables into predefined function calls to generate a complete, executable device file.

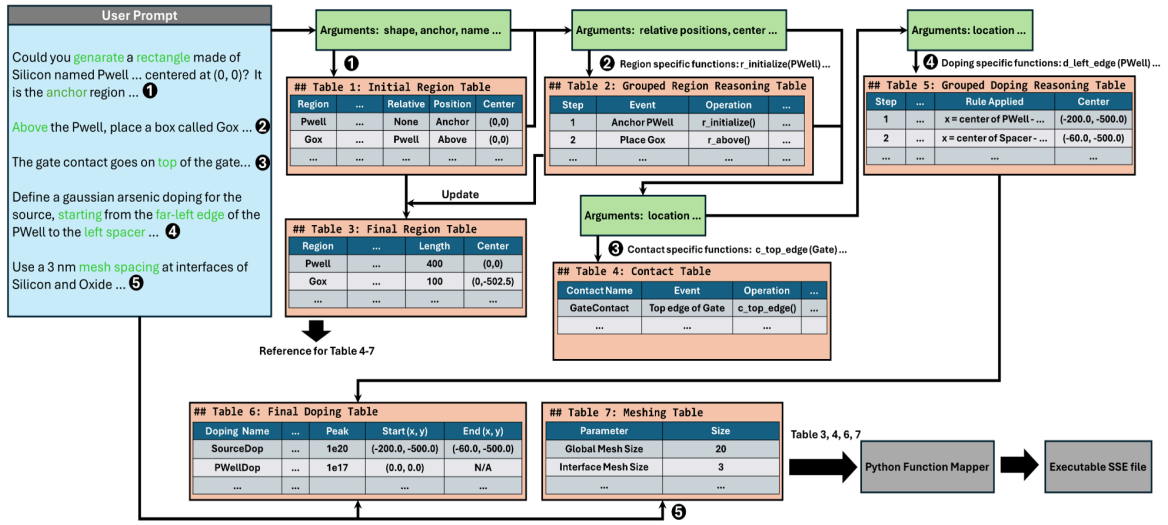


Figure 10.6.: Step-by-step tabular CoT reasoning process for CDM. Tables 1–7 show incremental spatial and physical feature resolution.

10.2.4.2. The proposed Tabular COT

At the core of our framework is a tabular CoT reasoning process that computes structural coordinates and attributes step by step, as shown in Fig. 10.6. All information in the tables is derived from the user’s natural language description of the device structure.

The process begins with region initialization (Table 1), where structural blocks such as PWell are defined using geometric attributes, material information, and relative spatial constraints such as above PWell. These relations are progressively resolved in the Grouped Region Reasoning Table (Table 2) through symbolic spatial operators such as r_{above} , each associated with a corresponding Rule Applied field. The applied rule determines the coordinate transformation required to update the center position of a region relative to a reference structure, enabling consistent spatial alignment. The resolved coordinates and dimensions are then consolidated in the Region Table (Table 3), which serves as the geometric foundation for subsequent placement of functional structures, including contacts and doping regions.

Based on Table 3, the model infers additional physical features: contacts are placed relative to region boundaries (Table 4), and doping profiles are determined through edge-based reasoning steps (Tables 5–6), each accompanied by its own update rule. Finally, global and local meshing settings are defined in the Meshing Table (Table 7). Once all tables are completed, they are passed to a Python-based function mapper, which invokes the fine-tuned SFM model to generate the corresponding function calls and assemble a complete, simulation-ready device file.

10.3. Results and Discussion

The fine-tuning is conducted using two NVIDIA A100 GPUs. To reduce memory overhead, the Llama-3-8B-Instruct model is loaded with 4-bit quantization. For supervised fine-tuning, we use a learning rate of 2×10^{-4} , a context length of 1024 tokens, and FP16 precision to balance training speed and numerical accuracy. The Adam optimizer with weight decay is employed for optimization. Each dataset generated using the proposed synthetic data generation method contains approximately 10,000 samples. For evaluation, we monitor the step-wise cross-entropy loss and record the value every 10 steps to track training progress. As we observe that training consistently converges within 300 steps

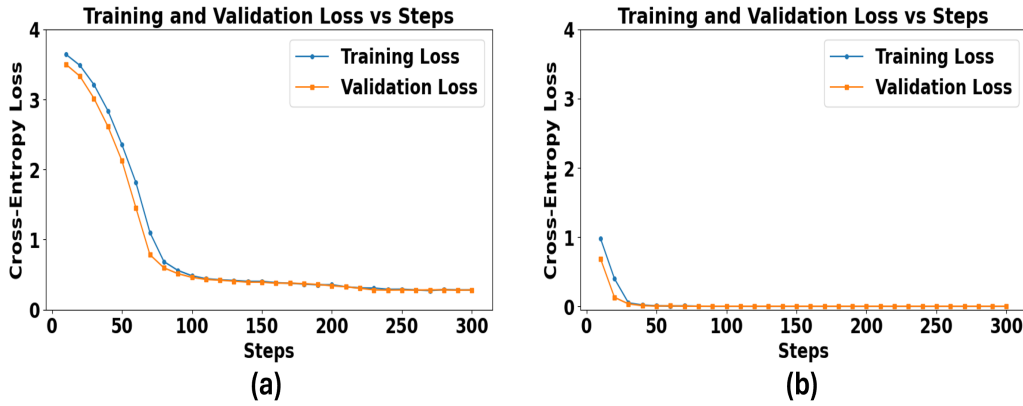


Figure 10.7: Cross-entropy loss comparison during supervised fine-tuning for region generation in SFM. (a) Training with a baseline GPT-4o synthetic dataset. (b) Training with the proposed pool sampling-based dataset, showing faster convergence and lower final loss.

across all experiments, we report the results of supervised fine-tuning within this 300-step window. The corresponding training curves are presented in the following subsections.

10.3.1. Evaluation of Dataset Generated for SFM

We evaluate the effectiveness of the proposed pool sampling-based synthetic dataset generation method using the region generation task in SFM as an example. As shown in Fig. 10.7, the method significantly improves training efficiency by reducing the cross-entropy loss during fine-tuning of the Llama-3-8B-Instruct model.

Without the proposed method, the loss plateaus around 0.27 and converges slowly, stabilizing only after approximately 200 steps (Fig. 10.7(a)). In contrast, the dataset generated with our sampling strategy demonstrates faster convergence and achieves a substantially lower final loss, dropping below 0.001 within the first few steps (Fig. 10.7(b)). Evaluation results for other functional blocks, including contact, doping, and meshing, are shown in Fig. 10.8(a)–(c), where the final training loss remains consistently below 0.001, confirming the generalizability and robustness of the synthetic dataset generation method across tasks.

10.3.2. Similarity Evaluation

As shown in Fig. 10.8(d), the proposed tabular CoT prompt strategy also demonstrates strong performance in CDM. The final training loss for CDM converges around 0.01, which is slightly higher than that for SFM tasks. This is expected due to the increased input complexity and longer token sequences in CDM prompts. Nevertheless, the results confirm that the model can learn structured multi-step reasoning effectively, validating the efficiency of our proposed tabular CoT strategy for complete device simulation file generation.

To assess the correctness of LLM-generated device files, we compare them against expert-designed reference device files. The goal is to determine whether the generated file accurately represents the intended device structure and is functionally usable for simulation. While text-based metrics such as Bilingual Evaluation Understudy (BLEU) are commonly used in language generation tasks, they are not well-suited for TCAD device files. BLEU is sensitive to token order and does not account for structural meaning. For example, two files may describe the same physical device using different block orders, which would lead BLEU to assign a low score.

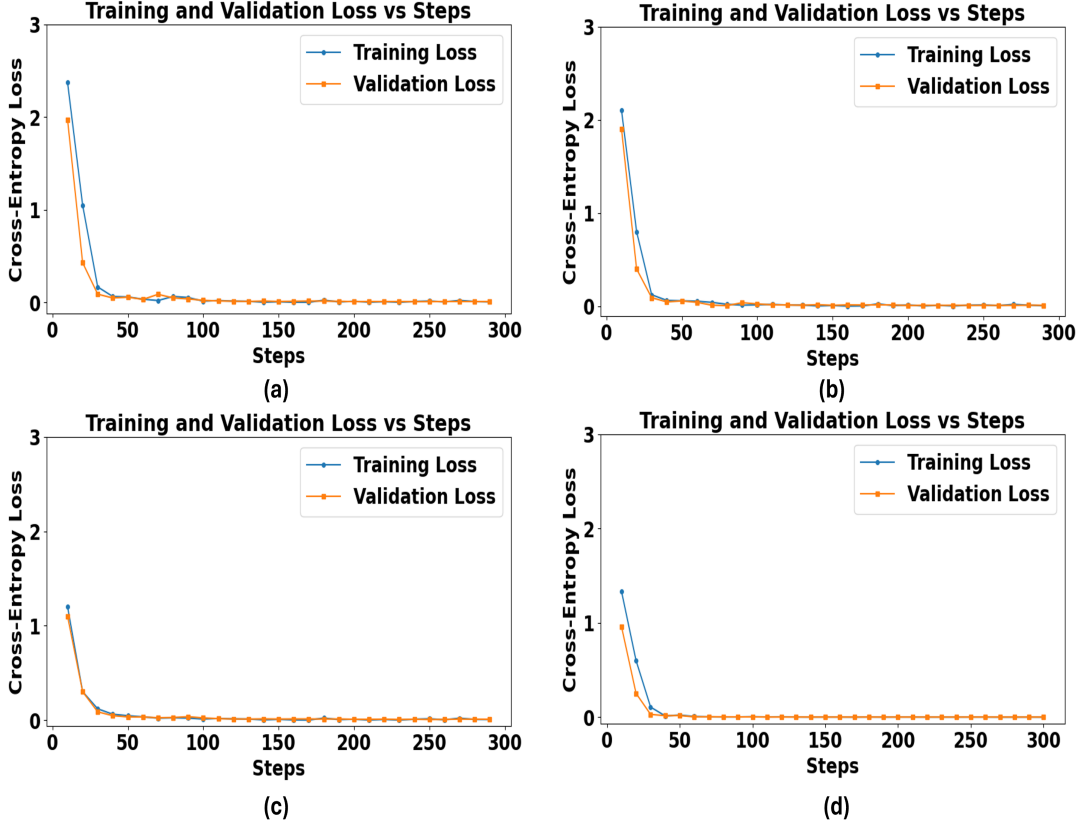


Figure 10.8.: Cross-entropy loss across various functional blocks. (a) Contact, (b) Doping, (c) Meshing, (d) CDM with Tabular CoT prompting.

To address this limitation, we adopt a structure-aware, block-level comparison strategy. Both the generated file G and the reference file R are parsed into sets of functional blocks—regions, contacts, doping, and meshing. Each block is represented as a structured set of attributes such as material, dimensions, and coordinates. If a one-to-one alignment between blocks is known, we compute block-level BLEU scores for each aligned block pair and report the average scores. To quantify similarity, we define the score as follows. $\mathcal{T}$ denotes the set of block types. For each type $t \in \mathcal{T}$, $G_t = \{g_1, \dots, g_{n_t}\}$ and $R_t = \{r_1, \dots, r_{n_t}\}$ are the aligned blocks from the generated and reference files, respectively.

$$S = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \frac{1}{n_t} \sum_{i=1}^{n_t} \text{BLEU}(G_i, R_i)$$

We applied this evaluation to the BJT and SiGeHBT example codes provided in the Sentaurus TCAD. Both generated files achieved an average block-level BLEU score above 0.96, demonstrating the high fidelity of our framework in replicating expert-designed device files.

10.4. Conclusion

In this work, we presented ChatTCAD, a framework that leverages fine-tuned large language models to generate structural Technology computer-aided design (TCAD) device files for two-dimensional devices from natural language inputs. By modularizing TCAD scripts and fine-tuning the Llama-3-8B-Instruct model with our proposed synthetic dataset and prompt strategies, ChatTCAD can reliably generate

simulation files that closely match those written by experts. As Design-Technology Co-Optimization and System-Technology Co-Optimization require tighter links between device-level modeling and circuit/system design, ChatTCAD enables circuit designers to explore and evaluate technology options without deep TCAD expertise.

11. Conclusion and Perspectives

11.1. Conclusion

In this thesis, reliability challenges in advanced semiconductor circuits were investigated from a multi-scale perspective, with the main focus on aging effects, radiation-induced soft errors, and AI-assisted EDA automation. The presented work demonstrated that technology scaling not only improves performance and integration density but also considerably increases the vulnerability of modern integrated circuits to device degradation, interconnect reliability issues, and transient fault mechanisms. Therefore, reliable operation in advanced technology nodes requires comprehensive device-to-circuit co-analysis methodologies capable of capturing both FEOL and BEOL interactions together with workload-dependent effects.

The main contributions of this thesis can be summarized as follows. In Chapter 3 and Chapter 4, machine-learning-based methodologies for efficient SER estimation and structural SET propagation analysis in PTL circuits were proposed to accelerate radiation reliability evaluation. Furthermore, Chapter 5 investigated the combined impact of radiation-induced transient effects and NBTI aging on latch timing behavior through multi-scale TCAD-to-circuit-level analysis.

This thesis also presented a unified FEOL/BEOL aging analysis framework for evaluating the combined effects of transistor aging and interconnect degradation on SRAM writability in advanced FDSOI and FinFET technologies. In particular, Chapter 6 analyzed the interaction between transistor aging and interconnect degradation in SRAM structures, while Chapter 7 proposed aging-aware testing and sensitization methodologies, including voltage-scaling-based earlier detection techniques, for predicting aging-induced SRAM failures and estimating the remaining operational lifetime. Moreover, Chapter 8 comprehensively investigated the impact of transistor and interconnect scaling, process variation, operating conditions, and cell location on SRAM aging behavior, together with DVFS-based mitigation methodologies for improving SRAM reliability and operational robustness.

In addition, Chapter 9 investigated radiation-induced soft errors in SRAM-based register files and evaluated technology-dependent mitigation strategies. Finally, Chapter 10 presented ChatTCAD, an LLM-assisted framework for automated TCAD simulation file generation and reliability-oriented EDA workflow automation, demonstrating the potential of AI-assisted methodologies for improving the efficiency and automation of reliability-oriented semiconductor design flows.

Overall, the results presented in this thesis demonstrate that future semiconductor reliability analysis increasingly requires interdisciplinary methodologies combining device physics, circuit-level analysis, reliability-aware testing, and AI-assisted automation. The presented methodologies provide important insights toward reliability-aware design, testing, and evaluation of future advanced semiconductor technologies and systems under continued technology scaling.

11.2. Future Perspectives

Several directions for future work can be identified based on the results presented in this thesis. First, reliability-driven DTCO methodologies could be further explored to jointly optimize PPA and long-term reliability under continued technology scaling. In addition to NBTI and EM, other important degradation mechanisms, such as TDDB and HCI, should also be incorporated into future multi-scale reliability analysis frameworks.

Furthermore, future studies could investigate system-level reliability-aware workload management and adaptive operating-condition adjustment methodologies to dynamically mitigate aging-induced degradation during runtime. In particular, workload-aware reliability optimization combined with adaptive voltage and frequency control may provide promising approaches for extending the operational lifetime of advanced semiconductor systems.

Finally, the rapid development of LLMs and AI-assisted methodologies opens new opportunities for reliability-oriented semiconductor design automation. Future work may explore LLM-assisted reliability-aware circuit optimization, automated reliability-driven design guidance, and intelligent TCAD/EDA workflow automation for accelerating reliability analysis and improving design productivity in advanced semiconductor technologies.

Part III.

Appendix

Bibliography

- [1] Z. Zhang, J. Lappas, A. Chinazzo, C. Weis, Z. Wu, L. Ni, N. Wehn, and M. Tahoori, “Machine learning based soft error rate estimation of pass transistor logic in high-speed communication”, in *2022 IEEE European Test Symposium (ETS)*, IEEE, 2022, pp. 1–4.
- [2] Z. Zhang, Z. Wu, C. Weis, N. Wehn, and M. Tahoori, “A learning-based approach for single event transient analysis in pass transistor logic”, in *2023 IEEE 29th International Symposium on On-Line Testing and Robust System Design (IOLTS)*, IEEE, 2023, pp. 1–7.
- [3] Z. Zhang, C. Weis, N. Wehn, M. Tahoori, and S. Nassif, “Do radiation and aging impact dvfs? tcad-based analysis on 22 nm fdsoi latches”, in *2024 IEEE 30th International Symposium on On-Line Testing and Robust System Design (IOLTS)*, IEEE, 2024, pp. 1–6.
- [4] Z. Zhang, M. Mayahinia, C. Weis, N. Wehn, M. Tahoori, S. Nassif, G. Tshagharyan, G. Harutyunyan, and Y. Zorian, “Addressing the combined effect of transistor and interconnect aging in sram towards silicon lifecycle management”, in *2024 IEEE 42nd VLSI Test Symposium (VTS)*, 2024, pp. 1–5.
- [5] Z. Zhang, M. Mayahinia, C. Weis, N. Wehn, M. Tahoori, S. Nassif, G. Tshagharyan, G. Harutyunyan, and Y. Zorian, “Testing for aging in advanced sram: From front end of the line transistors to back end of the line interconnects”, in *2024 IEEE International Test Conference (ITC)*, IEEE, 2024, pp. 46–50.
- [6] Z. Zhang, M. Mayahinia, C. Weis, N. Wehn, M. Tahoori, S. Nassif, G. Tshagharyan, G. Harutyunyan, and Y. Zorian, “Effect of transistor and interconnect scaling on aging in advanced sram”, *IEEE Transactions on Device and Materials Reliability*, 2026.
- [7] Z. Zhang, S. Hemaram, M. Mayahinia, C. Weis, N. Wehn, M. Tahoori, S. Nassif, G. Tshagharyan, G. Harutyunyan, and Y. Zorian, “Analysis and mitigation of radiation effects in sram-based register files”, in *2025 IEEE European Test Symposium (ETS)*, IEEE, 2025, pp. 1–4.
- [8] Z. Zhang, S. Nassif, and M. Tahoori, “Chattcad: Leveraging large language models for automated tcad simulation file generation”, in *2025 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, IEEE, vol. 1, 2025, pp. 1–6.
- [9] G. E. Moore, “Cramming more components onto integrated circuits, reprinted from electronics, volume 38, number 8, april 19, 1965, pp. 114 ff.”, *IEEE solid-state circuits society newsletter*, vol. 11, no. 3, pp. 33–35, 2009.
- [10] IEEE International Roadmap for Devices and Systems, *2024 update: More moore*, 2024.
- [11] N. Loubet, T. Hook, P. Montanini, C.-W. Yeung, S. Kanakasabapathy, M. Guillom, T. Yamashita, J. Zhang, X. Miao, J. Wang, et al., “Stacked nanosheet gate-all-around transistor to enable scaling beyond finfet”, in *2017 symposium on VLSI technology*, IEEE, 2017, T230–T231.
- [12] N. Horiguchi, H. Mertens, T. Chiarella, S. Demuynck, V. Vega-Gonzalez, A. Vandooren, A. Veloso, M. G. Bardon, G. Sisto, A. Gupta, et al., “3d stacked devices and mol innovations for post-nanosheet cmos scaling”, in *2023 International Electron Devices Meeting (IEDM)*, IEEE, 2023, pp. 1–4.

- [13] Z. Wang, J. Sun, A. Goksoy, S. K. Mandal, Y. Liu, J.-S. Seo, C. Chakrabarti, U. Y. Ogras, V. Chhabria, J. Zhang, et al., “Exploiting 2.5 d/3d heterogeneous integration for ai computing”, in *2024 29th Asia and South Pacific Design Automation Conference (ASP-DAC)*, IEEE, 2024, pp. 758–764.
- [14] S. Naffziger, K. Lepak, M. Paraschou, and M. Subramony, “2.2 amd chiplet architecture for high-performance server and desktop products”, in *2020 IEEE International Solid-State Circuits Conference-(ISSCC)*, IEEE, 2020, pp. 44–45.
- [15] D. Lembke, S. Bertolazzi, and A. Kis, “Single-layer mos2 electronics”, *Accounts of chemical research*, vol. 48, no. 1, pp. 100–110, 2015.
- [16] Q. Zhang, Y. Zhang, Y. Luo, and H. Yin, “New structure transistors for advanced technology node cmos ics”, *National Science Review*, vol. 11, no. 3, nwae008, 2024.
- [17] R. H. Dennard, F. H. Gaensslen, H.-N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, “Design of ion-implanted mosfet’s with very small physical dimensions”, *IEEE Journal of solid-state circuits*, vol. 9, no. 5, pp. 256–268, 2003.
- [18] International Technology Roadmap for Semiconductors 2.0 (ITRS 2.0), “Interconnect: 2015 edition”, Semiconductor Industry Association, Roadmap Report, 2015. [Online]. Available: <https://www.semiconductors.org>.
- [19] A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva, “Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale mosfets”, *IEEE transactions on electron devices*, vol. 50, no. 9, pp. 1837–1852, 2003.
- [20] V. Huard, M. Denais, and C. Parthasarathy, “Nbti degradation: From physical mechanisms to modelling”, *Microelectronics Reliability*, vol. 46, no. 1, pp. 1–23, 2006.
- [21] R. Baumann, “Radiation-induced soft errors in advanced semiconductor technologies”, *IEEE Transactions on Device and materials reliability*, vol. 5, no. 3, pp. 305–316, 2005. doi: 10.1109/TDMR.2005.853449.
- [22] J. R. Black, “Electromigration—a brief survey and some recent results”, *IEEE Transactions on Electron Devices*, vol. 16, no. 4, pp. 338–347, 2005.
- [23] A. P. Chandrakasan, S. Sheng, and R. W. Brodersen, “Low-power cmos digital design”, *IEICE Transactions on Electronics*, vol. 75, no. 4, pp. 371–382, 1992.
- [24] A. Tajalli and Y. Leblebici, “Design trade-offs in ultra-low-power digital nanoscale cmos”, *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 58, no. 9, pp. 2189–2200, 2011.
- [25] J. A. Davis, R. Venkatesan, A. Kaloyeros, M. Beylansky, S. J. Souri, K. Banerjee, K. C. Saraswat, A. Rahman, R. Reif, and J. D. Meindl, “Interconnect limits on gigascale integration (gsi) in the 21st century”, *Proceedings of the IEEE*, vol. 89, no. 3, pp. 305–324, 2001.
- [26] R. Zimmermann and W. Fichtner, “Low-power logic styles: Cmos versus pass-transistor logic”, *IEEE journal of solid-state circuits*, vol. 32, no. 7, pp. 1079–1090, 1997.
- [27] K. Yano, T. Yamanaka, T. Nishida, M. Saito, K. Shimohigashi, and A. Shimizu, “A 3.8-ns cmos 16* 16-b multiplier using complementary pass-transistor logic”, *IEEE journal of solid-state circuits*, vol. 25, no. 2, pp. 388–395, 2002.
- [28] K. Yano, Y. Sasaki, K. Rikino, and K. Seki, “Top-down pass-transistor logic design”, *IEEE journal of solid-state circuits*, vol. 31, no. 6, pp. 792–803, 2002.
- [29] S. Nishizawa, T. Ishihara, and H. Onodera, “Analysis and comparison of xor cell structures for low voltage circuit design”, in *International Symposium on Quality Electronic Design (ISQED)*, IEEE, 2013, pp. 703–708.

- [30] J.-M. Wang, S.-C. Fang, and W.-S. Feng, “New efficient designs for xor and xnor functions on the transistor level”, *IEEE Journal of solid-state Circuits*, vol. 29, no. 7, pp. 780–786, 1994.
- [31] R. N. Oliveira, A. D. Lüdke, and C. Meinhardt, “Radiation effects in xor logic gates at 16nm cmos and finfet technology”, in *2019 26th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, IEEE, 2019, pp. 590–593.
- [32] N. H. Weste and D. Harris, *CMOS VLSI design: a circuits and systems perspective*. Pearson Education India, 2015.
- [33] S.-F. Hsiao and P.-C. Wu, “Design of low-leakage multi-port sram for register file in graphics processing unit”, in *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, IEEE, 2014, pp. 2181–2184.
- [34] H. Jeon, G. S. Ravi, N. S. Kim, and M. Annavaram, “Gpu register file virtualization”, in *Proceedings of the 48th International Symposium on Microarchitecture*, 2015, pp. 420–432.
- [35] M. Omana, D. Rossi, and C. Metra, “Latch susceptibility to transient faults and new hardening approach”, *IEEE Transactions on Computers*, vol. 56, no. 9, pp. 1255–1268, 2007.
- [36] V. Chandra and R. Aitken, “Impact of voltage scaling on nanoscale sram reliability”, in *2009 Design, Automation & Test in Europe Conference & Exhibition*, IEEE, 2009, pp. 387–392.
- [37] H.-H. Liu, S. M. Salahuddin, D. Abdi, R. Chen, P. Weckx, P. Matagne, and F. Catthoor, “Extended methodology to determine sram write margin in resistance-dominated technology node”, *IEEE Transactions on Electron Devices*, vol. 69, no. 6, pp. 3113–3117, 2022.
- [38] R. Baumann, “Soft errors in advanced computer systems”, *IEEE design & test of computers*, vol. 22, no. 3, pp. 258–266, 2005.
- [39] N. Mahatme, N. Gaspard, T. Assis, S. Jagannathan, I. Chatterjee, T. Loveless, B. Bhuvu, L. W. Massengill, S. Wen, and R. Wong, “Impact of technology scaling on the combinational logic soft error rate”, in *2014 IEEE international reliability physics symposium*, IEEE, 2014, 5F–2.
- [40] T. Karnik and P. Hazucha, “Characterization of soft errors caused by single event upsets in cmos processes”, *IEEE Transactions on Dependable and secure Computing*, vol. 1, no. 2, pp. 128–143, 2004.
- [41] L. W. Massengill, B. L. Bhuvu, W. T. Holman, M. L. Alles, and T. D. Loveless, “Technology scaling and soft error reliability”, in *2012 IEEE International Reliability Physics Symposium (IRPS)*, IEEE, 2012, pp. 3C–1.
- [42] D. A. Black, W. H. Robinson, I. Z. Wilcox, D. B. Limbrick, and J. D. Black, “Modeling of single event transients with dual double-exponential current sources: Implications for logic cell characterization”, *IEEE Transactions on Nuclear Science*, vol. 62, no. 4, pp. 1540–1549, 2015.
- [43] F. Wrobel, L. Dilillo, A. Touboul, V. Pouget, and F. Saigné, “Determining realistic parameters for the double exponential law that models transient current pulses”, *IEEE Transactions on Nuclear Science*, vol. 61, no. 4, pp. 1813–1818, 2014.
- [44] M. Zerarka, P. Austin, A. Bensoussan, F. Morancho, and A. Durier, “Tcad simulation of the single event effects in normally-off gan transistors after heavy ion radiation”, *IEEE Transactions on Nuclear Science*, vol. 64, no. 8, pp. 2242–2249, 2017.
- [45] R. R. Rao, K. Chopra, D. T. Blaauw, and D. M. Sylvester, “Computing the soft error rate of a combinational logic circuit using parameterized descriptors”, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 26, no. 3, pp. 468–479, 2007.

- [46] J. R. Cho, S. Go, J. Park, T. J. Yang, S. Lee, N. Lee, D. K. Lee, Y. Kim, M. Kang, R.-H. Baek, et al., “Analysis of bias temperature instability in peripheral cmos devices for low-temperature memory applications”, *IEEE Access*, 2025.
- [47] S. Ahmad, B. Iqbal, N. Alam, and M. Hasan, “Low leakage fully half-select-free robust sram cells with bti reliability analysis”, *IEEE Transactions on Device and Materials Reliability*, vol. 18, no. 3, pp. 337–349, 2018.
- [48] M. A. Alam and S. Mahapatra, “A comprehensive model of pmos nbti degradation”, *Microelectronics Reliability*, vol. 45, no. 1, pp. 71–81, 2005.
- [49] R. Tiwari, M. Duan, M. Bajaj, D. Dolgos, L. Smith, H. Y. Wong, and S. Mahapatra, “A physics-based tcad framework for nbti”, *Solid-State Electronics*, vol. 202, p. 108 573, 2023.
- [50] T. Grasser, B. Kaczer, W. Goes, T. Aichinger, P. Hehenberger, and M. Nelhiebel, “A two-stage model for negative bias temperature instability”, in *2009 IEEE international reliability physics symposium*, IEEE, 2009, pp. 33–44.
- [51] R. De Orío, H. Cerić, and S. Selberherr, “Physically based models of electromigration: From black’s equation to modern tcad models”, *Microelectronics Reliability*, vol. 50, no. 6, pp. 775–789, 2010.
- [52] S. M. Nair, R. Bishnoi, M. B. Tahoori, H. Zahedmanesh, K. Croes, K. Garelo, G. S. Kar, and F. Catthoor, “Physics based modeling of bimodal electromigration failure distributions and variation analysis for vlsi interconnects”, in *2020 IEEE International Reliability Physics Symposium (IRPS)*, 2020, pp. 1–5. doi: 10.1109/IRPS45951.2020.9128313.
- [53] K. Settaluri, A. Haj-Ali, Q. Huang, K. Hakhamaneshi, and B. Nikolic, “Autockt: Deep reinforcement learning of analog circuit designs”, *arXiv preprint arXiv:2001.01808*, 2020.
- [54] Y.-S. Yang, Y. Li, and S. R. R. Kola, “A physical-based artificial neural networks compact modeling framework for emerging fets”, *IEEE Transactions on Electron Devices*, vol. 71, no. 1, pp. 223–230, 2023.
- [55] L. Alrahis, J. Knechtel, F. Klemme, H. Amrouch, and O. Sinanoglu, “Gnn4rel: Graph neural networks for predicting circuit reliability degradation”, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 41, no. 11, pp. 3826–3837, 2022.
- [56] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, et al., “Gpt-4 technical report”, *arXiv preprint arXiv:2303.08774*, 2023.
- [57] A. Liu, B. Feng, B. Xue, B. Wang, B. Wu, C. Lu, C. Zhao, C. Deng, C. Zhang, C. Ruan, et al., “Deepseek-v3 technical report”, *arXiv preprint:2412.19437*, 2024.
- [58] A. Wei, H. Tan, T. Suresh, D. Mendoza, T. S. Teixeira, K. Wang, C. Trippel, and A. Aiken, “Vericoder: Enhancing llm-based rtl code generation through functional correctness validation”, *arXiv preprint arXiv:2504.15659*, 2025.
- [59] H. Liu, Y. Lu, M. Wang, X. Yao, and B. Yu, “Llm-assisted circuit verification: A comprehensive survey”, in *2026 31st Asia and South Pacific Design Automation Conference (ASP-DAC)*, IEEE, 2026, pp. 439–446.
- [60] K. Qayyum, C. K. Jha, S. Ahmadi-Pour, M. Hassan, and R. Drechsler, “Llm-assisted bug identification and correction for verilog hdl”, *ACM Transactions on Design Automation of Electronic Systems*, vol. 30, no. 6, pp. 1–28, 2025.
- [61] Y. Shi, Z. Tao, Y. Gao, T. Zhou, C. Chang, Y. Wang, B. Chen, G. Zhang, A. Liu, Z. Yu, et al., “Amsnet-kg: A netlist dataset for llm-based ams circuit auto-design using knowledge graph rag”, *ACM Transactions on Design Automation of Electronic Systems*, vol. 30, no. 6, pp. 1–37, 2025.

- [62] V. V. Shubin, "Analysis and comparison of ripple carry full adders by speed", in *2010 11th International Conference and Seminar on Micro/Nanotechnologies and Electron Devices*, IEEE, 2010, pp. 132–135.
- [63] H. Naseri and S. Timarchi, "Low-power and fast full adder by exploring new xor and xnor gates", *IEEE transactions on very large scale integration (VLSI) systems*, vol. 26, no. 8, pp. 1481–1493, 2018.
- [64] H. T. Bui, Y. Wang, and Y. Jiang, "Design and analysis of low-power 10-transistor full adders using novel xor-xnor gates", *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 49, no. 1, pp. 25–30, 2002.
- [65] Y. Chich et al., "An 89 tops/w and 16.3 tops/mm² all-digital sram-based full-precision compute-in-memory macro in 22nm for machine-learning edge applications", in *ISSCC*, 2021, pp. 252–253.
- [66] J. Lappas, A. Chinazzo, C. Weis, C. Xia, Z. Wu, L. Ni, and N. Wehn, "Revisiting pass-transistor logic styles in a 12nm finfet technology node", in *2022 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, IEEE, 2022, pp. 1083–1084.
- [67] H. Cha, E. Rudnick, J. Patel, R. Iyer, and G. Choi, "A gate-level simulation environment for alpha-particle-induced transient faults", *IEEE Transactions on Computers*, vol. 45, no. 11, pp. 1248–1256, 1996. DOI: 10.1109/12.544481.
- [68] P. Shivakumar, M. Kistler, S. Keckler, D. Burger, and L. Alvisi, "Modeling the effect of technology trends on the soft error rate of combinational logic", in *Proceedings International Conference on Dependable Systems and Networks*, 2002, pp. 389–398. DOI: 10.1109/DSN.2002.1028924.
- [69] J. S. Kim, C. Nicopoulos, N. Vijaykrishnan, Y. Xie, and E. Lattanzi, "A probabilistic model for soft-error rate estimation in combinational logic", in *Proc. of the Int'l Workshop on Probabilistic Analysis Techniques for Real-time and Embedded Systems*, 2004.
- [70] B. Zhang, W.-S. Wang, and M. Orshansky, "Faser: Fast analysis of soft error susceptibility for cell-based designs", in *7th International Symposium on Quality Electronic Design (ISQED'06)*, IEEE, 2006, 6–pp.
- [71] M. Hashimoto, W. Liao, and S. Hirokawa, "Soft error rate estimation with tcad and machine learning", in *2017 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, IEEE, 2017, pp. 129–132.
- [72] M. Omana, G. Papasso, D. Rossi, and C. Metra, "A model for transient fault propagation in combinatorial logic", in *9th IEEE On-Line Testing Symposium, 2003. IOLTS 2003.*, IEEE, 2003, pp. 111–115.
- [73] M. Song, G. Rang, S. Kim, and B. Kang, "Design methodology for high speed and low power digital circuits with energy economized pass-transistor logic (eepl)", in *ESSCIRC'96: Proceedings of the 22nd European Solid-State Circuits Conference*, IEEE, 1996, pp. 120–123.
- [74] M. Pelikan, "Bayesian optimization algorithm", in *Hierarchical Bayesian optimization algorithm: toward a new generation of evolutionary algorithms*, Springer, 2005, pp. 31–48.
- [75] L. Chang, K. Vo, and J. Berg, "A simplified model to predict the linear temperature coefficient of a cmos inverter's delay time", *IEEE transactions on electron devices*, vol. 34, no. 8, pp. 1834–1837, 1987.
- [76] W. Schneider, "Analytical inverter delay modeling using matlab's curve fitting toolbox", *arXiv preprint arXiv:1801.00005*, 2017.
- [77] D. Kobayashi, H. Saito, and K. Hirose, "Estimation of single event transient voltage pulses in vlsi circuits from heavy-ion-induced transient currents measured in a single mosfet", *IEEE Transactions on Nuclear Science*, vol. 54, no. 4, pp. 1037–1041, 2007. DOI: 10.1109/TNS.2007.891397.

- [78] F. Wang and Y. Xie, “An accurate and efficient model of electrical masking effect for soft errors in combinational logic”, in *2nd Workshop of System Effects of Logic Soft Error*, 2006.
- [79] C. Xu, Y. Liu, X. Liao, J. Cheng, and Y. Yang, “Machine learning regression-based single-event transient modeling method for circuit-level simulation”, *IEEE Transactions on Electron Devices*, vol. 68, no. 11, pp. 5758–5764, 2021.
- [80] R. Rajaraman, J. Kim, N. Vijaykrishnan, Y. Xie, and M. Irwin, “Seat-la: A soft error analysis tool for combinational logic”, in *2006 19th International Conference on VLSI Design*, 2006, 4 pp.-. doi: 10.1109/VLSID.2006.143.
- [81] H. Ren, G. F. Kokai, W. J. Turner, and T.-S. Ku, “Paragraph: Layout parasitics and device parameter prediction using graph neural networks”, in *2020 57th ACM/IEEE Design Automation Conference*, 2020, pp. 1–6. doi: 10.1109/DAC18072.2020.9218515.
- [82] S. Cai, W. Wang, F. Yu, and B. He, “Single event transient propagation probabilities analysis for nanometer cmos circuits”, *Journal of Electronic Testing*, vol. 35, no. 2, pp. 163–172, 2019.
- [83] R. Palin, D. Ward, I. Habli, and R. Rivett, “Iso 26262 safety cases: Compliance and assurance”, 2011.
- [84] S. Mahapatra and N. Parihar, “A review of nbti mechanisms and models”, *Microelectronics Reliability*, vol. 81, pp. 127–135, 2018, issn: 0026-2714. doi: <https://doi.org/10.1016/j.microrel.2017.12.027>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0026271417305851>.
- [85] S. Mahapatra, N. Goel, S. Desai, S. Gupta, B. Jose, S. Mukhopadhyay, K. Joshi, A. Jain, A. E. Islam, and M. A. Alam, “A comparative study of different physics-based nbti models”, *IEEE Transactions on Electron Devices*, vol. 60, no. 3, pp. 901–916, 2013. doi: 10.1109/TED.2013.2238237.
- [86] N. Rostand, S. Martinie, J. Lacord, O. Rozeau, T. Poiroux, and G. Hubert, “Single event transient compact model for fdsoi mosfets taking bipolar amplification and circuit level arbitrary generation into account”, in *2019 International Conference on Simulation of Semiconductor Processes and Devices*, 2019, pp. 1–4. doi: 10.1109/SISPAD.2019.8870520.
- [87] V. Stojanovic and V. Oklobdzija, “Comparative analysis of master-slave latches and flip-flops for high-performance and low-power systems”, *IEEE Journal of Solid-State Circuits*, vol. 34, no. 4, pp. 536–548, 1999. doi: 10.1109/4.753687.
- [88] D. Markovic, J. W. Tschanz, and V. K. De, *Transmission-gate based flip-flop*, US Patent 6,642,765, Nov. 2003.
- [89] L. T. Clark, V. Vashishtha, L. Shifren, A. Gujja, S. Sinha, B. Cline, C. Ramamurthy, and G. Yeric, “Asap7: A 7-nm finfet predictive process design kit”, *Microelectronics Journal*, vol. 53, pp. 105–115, 2016.
- [90] www.synopsys.com/manufacturing/tcad.html.
- [91] J. Hoentschel, L. Pirro, R. Carter, and M. Horstmann, “22fdx® technologies for ultra-low power iot, rf and mmwave applications”, *Nanoelectronic Devices*, vol. 2, 2019.
- [92] G. Hubert, L. Artola, and D. Regis, “Impact of scaling on the soft error sensitivity of bulk, fdsoi and finfet technologies due to atmospheric radiation”, *Integration*, vol. 50, pp. 39–47, 2015.
- [93] X. Huang, T. Yu, V. Sukharev, and S. X.-D. Tan, “Physics-based electromigration assessment for power grid networks”, in *Proceedings of the 51st Annual Design Automation Conference*, 2014, pp. 1–6.

- [94] H. Zhou, S. Yu, Z. Sun, and S. X.-D. Tan, "Reliable power grid network design framework considering em immortalities for multi-segment wires", in *2020 25th Asia and South Pacific Design Automation Conference (ASP-DAC)*, IEEE, 2020, pp. 74–79.
- [95] Z. Guan, M. Marek-Sadowska, and S. Nassif, "Sram bit-line electromigration mechanism and its prevention scheme", in *International Symposium on Quality Electronic Design (ISQED)*, IEEE, 2013, pp. 286–293.
- [96] S. M. Nair, M. Mayahinia, M. B. Tahoori, M. Perumkunnal, H. Zahedmanesh, K. Croes, K. Garelo, T. Marinelli, T. Evenblij, G. S. Kar, et al., "Workload-aware electromigration analysis in emerging spintronic memory arrays", *IEEE Transactions on Device and Materials Reliability*, vol. 21, no. 2, pp. 258–266, 2021.
- [97] A. Bansal, R. Rao, J.-J. Kim, S. Zafar, J. H. Stathis, and C.-T. Chuang, "Impacts of nbtj and pbtj on sram static/dynamic noise margins and cell failure probability", *Microelectronics Reliability*, vol. 49, no. 6, pp. 642–649, 2009.
- [98] M. Mayahinia, H.-H. Liu, S. Mishra, Z. Tokei, F. Catthoor, and M. Tahoori, "Electromigration-aware design technology co-optimization for sram in advanced technology nodes", in *2023 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, IEEE, 2023, pp. 1–6.
- [99] X. Dong, C. Xu, Y. Xie, and N. P. Jouppi, "Nvsim: A circuit-level performance, energy, and area model for emerging nonvolatile memory", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 31, no. 7, pp. 994–1007, 2012.
- [100] X. Zhang, D. Connelly, H. Takeuchi, M. Hytha, R. J. Mears, and T.-J. K. Liu, "Comparison of soi versus bulk finfet technologies for 6t-sram voltage scaling at the 7-/8-nm node", *IEEE Transactions on Electron Devices*, vol. 64, no. 1, pp. 329–332, 2016.
- [101] Y. Lin, M. Li, and S. Gupta, "Predictive testing for aging in srams and mitigation", in *2024 IEEE International Test Conference (ITC)*, IEEE, 2024, pp. 293–302.
- [102] H. Lo, D. Choi, Y. Hu, Y. Shen, Y. Qi, J. Peng, D. Zhou, M. Mohan, C. Yong, H. Zhan, et al., "A 12nm finfet technology featuring 2nd generation finfet for low power and high performance applications", in *2018 IEEE Symposium on VLSI Technology*, IEEE, 2018, pp. 215–216.
- [103] M. Mayahinia, M. Tahoori, G. Harutyunyan, G. Tshagharyan, and K. Amirkhanyan, "An efficient test strategy for detection of electromigration impact in advanced finfet memories", in *2022 IEEE International Test Conference (ITC)*, IEEE, 2022, pp. 650–655.
- [104] M. Sadrosadati, A. Mirhosseini, A. Hajiabadi, S. B. Ehsani, H. Falahati, H. Sarbazi-Azad, M. Drumond, B. Falsafi, R. Ausavarungnirun, and O. Mutlu, "Highly concurrent latency-tolerant register files for gpus", *ACM Transactions on Computer Systems (TOCS)*, vol. 37, no. 1-4, pp. 1–36, 2021.
- [105] J. Neuendank, M. Spear, T. Wallace, D. Wilson, J. Solano, G. Irumva, I. S. Esqueda, H. J. Barnaby, L. T. Clark, J. Brunhaver, M. Turowski, E. Mikkola, D. Hughart, J. Young, J. Manuel, S. Agarwal, B. Vaandrager, G. Vizkelethy, A. Gutierrez, J. Trippe, M. King, E. Bielejec, and M. Marinella, "Single event upset and total ionizing dose response of 12lp finfet digital circuits", in *2022 IEEE Radiation Effects Data Workshop (REDW)(in conjunction with 2022 NSREC)*, 2022, pp. 1–9. doi: 10.1109/REDW56037.2022.9921478.
- [106] A. Hurst, A. Lerer, A. P. Goucher, A. Perelman, A. Ramesh, A. Clark, A. Ostrow, A. Welihinda, A. Hayes, A. Radford, et al., "Gpt-4o system card", *arXiv preprint:2410.21276*, 2024.
- [107] A. Grattafiori, A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Vaughan, et al., "The llama 3 herd of models", *arXiv preprint:2407.21783*, 2024.

- [108] J. Bhandari, V. Bhat, Y. He, S. Garg, H. Rahmani, and R. Karri, “Auto-spice: Leveraging llms for dataset creation via automated spice netlist extraction from analog circuit diagrams”, *arXiv preprint:2411.14299*, 2024.
- [109] B. Liu, H. Zhang, X. Gao, Z. Kong, X. Tang, Y. Lin, R. Wang, and R. Huang, “Layoutcopilot: An llm-powered multi-agent collaborative framework for interactive analog layout design”, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, pp. 1–1, 2025. doi: 10.1109/TCAD.2025.3529805.
- [110] H. Y. Wong, M. Xiao, B. Wang, Y. K. Chiu, X. Yan, J. Ma, K. Sasaki, H. Wang, and Y. Zhang, “Tcad-machine learning framework for device variation and operating temperature analysis with experimental demonstration”, *IEEE Journal of the Electron Devices Society*, vol. 8, pp. 992–1000, 2020.
- [111] L. M. Long Nguyen, A. Lu, and H. Y. Wong, “Tcad structure input file generation using large language model”, in *2024 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2024, pp. 1–4. doi: 10.1109/SISPAD62626.2024.10733015.
- [112] M. Levy, A. Jacoby, and Y. Goldberg, “Same task, more tokens: The impact of input length on the reasoning performance of large language models”, *arXiv preprint:2402.14848*, 2024.
- [113] H. Dang, L. Mecke, F. Lehmann, S. Goller, and D. Buschek, “How to prompt? opportunities and challenges of zero-and few-shot learning for human-ai interaction in creative applications of generative models”, *arXiv preprint:2209.01390*, 2022.
- [114] S. Hao, W. Han, T. Jiang, Y. Li, H. Wu, C. Zhong, Z. Zhou, and H. Tang, “Synthetic data in ai: Challenges, applications, and ethical implications”, *arXiv preprint:2401.01629*, 2024.
- [115] S. Liu, W. Fang, Y. Lu, Q. Zhang, H. Zhang, and Z. Xie, “Rtlcoder: Outperforming gpt-3.5 in design rtl generation with our open-source dataset and lightweight solution”, in *2024 IEEE LLM Aided Design Workshop (LAD)*, 2024, pp. 1–5. doi: 10.1109/LAD62341.2024.10691788.
- [116] J. He, M. Rungta, D. Koleczek, A. Sekhon, F. X. Wang, and S. Hasan, “Does prompt formatting have any impact on llm performance?”, *arXiv preprint:2411.10541*, 2024.
- [117] M. Suzgun, N. Scales, N. Schärli, S. Gehrmann, Y. Tay, H. W. Chung, A. Chowdhery, Q. V. Le, E. H. Chi, D. Zhou, et al., “Challenging big-bench tasks and whether chain-of-thought can solve them”, *arXiv preprint:2210.09261*, 2022.
- [118] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou, et al., “Chain-of-thought prompting elicits reasoning in large language models”, *Advances in neural information processing systems*, vol. 35, pp. 24 824–24 837, 2022.
- [119] Z. Wang, H. Zhang, C.-L. Li, J. M. Eisenschlos, V. Perot, Z. Wang, L. Miculicich, Y. Fujii, J. Shang, C.-Y. Lee, et al., “Chain-of-table: Evolving tables in the reasoning chain for table understanding”, *arXiv preprint:2401.04398*, 2024.

List of Figures

2.1. Development history of and trends in MOSFETs with differing structures and scaling stages [16].	7
2.2. Simplified organization of modern computing systems, including processing logic and memory hierarchy.	8
2.3. Examples of DPL-based (b) and inverter-buffered PTL XOR circuits (c) adapted from [29], [30]. [29] reported that DPL-based XOR designs achieved up to 24% area reduction and 35% lower power consumption in synthesized multiplier circuits under low-voltage operation compared with complementary CMOS XOR implementations (a), whereas inverter-buffered single-pass-transistor structures (c) suffered from degraded voltage swing and failed to satisfy timing constraints at low supply voltages due to weak drivability.	10
2.4. (a) Conventional six-transistor (6T) SRAM cell with associated parasitic resistance and capacitance on WL and BL interconnects, (b) Simplified SRAM array architecture with the required peripheral circuitry.	11
2.5. Conceptual illustration of the reaction-diffusion (R/D) model for NBTI degradation, including Si-H bond dissociation, hydrogen diffusion, and recovery mechanisms.	13
3.1. Two of the four PTL-based full adders investigated in this chapter: (a) NASERI 20T full adder [63], (b) BUI02A9 (BA) full adder [64].	20
3.2. SET propagation behavior in the BA PTL full adder under different injected pulse conditions. Long-tail transient responses are observed due to incomplete voltage recovery at the output node.	21
3.3. Proposed method	22
3.4. Structure of the evaluated 8-bit RCA. Input and output ports are connected to X1 load buffers, while SET injection is performed at the first full-adder stage (LSB).	25
3.5. SET injection in the first PTL stage, SER comparison over all output bits (all sum bits and final carry out).	26
3.6. Comparison between simulated and predicted SERs at the carry-output ports of different RCA implementations.	26
3.7. Comparison between simulated and predicted SERs at the sum-output ports of different RCA implementations.	27
4.1. (a) BUI02 A9 PTL full adder [64], and (b) long-tail transient behavior observed during SET propagation in PTL circuits.	30
4.2. SET generation and propagation through a combinational logic chain.	30
4.3. Overall workflow of the proposed learning-based SET propagation modeling framework. .	32
4.4. Common PTL structures considered in the proposed SET propagation modeling framework.	32
4.5. Regression cases and their ground truth.	33
4.6. Curve-fitting results for irregular SET propagation modeling as well as an example of irregular SETs.	36
4.7. Predicted versus actual fitting parameters in the falling subthreshold region. (a)-(b) correspond to regular-to-irregular propagation cases, while (c)-(d) correspond to irregular-to-irregular propagation through a single NMOS structure.	38
4.8. 4-bit RCA used as the DUT for validation.	39

4.9.	Comparison between predicted and simulated waveform and delay errors.	39
5.1.	Schematics of the considered latches under test. SSTC MS: static single-transistor-clocked master-slave latch, WPMS: write-port master-slave latch, ASAP7: latch in Arizona State University's 7-nm FinFET predictive PDK, MC2MOS: modified clocked CMOS logic latch.	43
5.2.	Output characteristics of the TCAD and PDK MOSFETs and the proposed raised source/drain FDSOI structure.	44
5.3.	Analysis of ΔV_{th} under three stress conditions and variation of radiation-induced current across diverse locations and operating conditions.	45
5.4.	Analysis of T_{setup} violation for the ASAP latch at size X1, studied under conditions without and with high-energy ion injection at N7. Initial delays for CLK and DAT are 300ps, with DAT shifted to the CLK edge in 0.4ps increments.	46
5.5.	Comparison of fresh latches' T_{steup} without radiation.	48
5.6.	Comparison of fresh latches' T_{steup} under radiation. Missing data points indicate the latch fails to retain data.	48
5.7.	Latches' T_{steup} under logic '0' stress without radiation.	49
5.8.	Latches' T_{steup} under logic '1' stress without radiation.	49
5.9.	Latches' T_{steup} under logic '0' stress with radiation.	50
5.10.	Latches' T_{steup} under logic '1' stress with radiation.	50
6.1.	SRAM cell schematics (a) and peripheral circuit (b), with per-cell parasitic resistance on BL and word line (WL).	55
6.2.	Individual effects of NBTI and EM on T_{switch} of the write '0' operation.	57
6.3.	Examples of faults in the FDSOI SRAM: (a) delay fault with a critical T_{switch} of 600 ps, and (b) stuck-at fault.	57
6.4.	Delay and stuck-at faults caused by NBTI and EM, categorized by cell color. Green cells denote successful write operations within the critical switching time (T_{switch}), while red cells indicate delay faults, as illustrated in Fig. 6.3(a). Pattern cells represent stuck-at faults, as shown in Fig. 6.3(b). *: Delay faults start to occur at a ΔV_{th} of 0.164 V.	58
7.1.	Output characteristics of the TCAD and PDK MOSFETs and the proposed FinFET structure with diamond-shaped epitaxy.	62
7.2.	The V_{th} increase due to NBTI in a sole P-type transistor.	63
7.3.	Example of a delay fault in (a), where the critical switching time is 220 ps, and a hard stuck-at fault in (b).	66
7.4.	Delay and stuck-at faults due to NBTI and EM under nominal operating conditions in FinFET SRAM. Delays are shown in picoseconds. Green entries indicate successful write operations within the critical T_{switch} . Solid red entries represent delay faults that a 50% frequency downscaling can mitigate, while patterned red entries indicate delay faults that this downscaling cannot mitigate. Patterned black entries represent stuck-at faults.	67
7.5.	Delay and stuck-at faults caused by NBTI and EM at the test operation corner in the FinFET SRAM. Delays are in picoseconds. Colors and patterns of entries (green, solid red, patterned red, and patterned black) indicate the same operation statuses as in Fig. 7.4.	68
8.1.	Delay and stuck-at faults due to NBTI and EM under nominal operating conditions in the FDSOI SRAM. Delays are shown in picoseconds. Green entries indicate successful write operations within the critical T_{switch} . Solid red entries represent delay faults that a 50% frequency downscaling can mitigate, while patterned red entries indicate delay faults that this downscaling cannot mitigate. Patterned black entries denote stuck-at faults.	73

8.2.	Delay and stuck-at faults due to NBTI and EM under nominal operating conditions in the FinFET SRAM. The color coding is the same as in Fig. 8.1.	74
8.3.	Impact of aging scenarios on SRAM write ‘0’ delay. Scenario (1): combined peripheral and cell aging. Scenario (2): peripheral-dominant aging. The color coding is the same as in Fig. 8.1.	74
8.4.	Impact of process variation and cell location on T_{switch} and write failure probability (First vs. Last cell), considering early-stage of NBTI and EM.	75
8.5.	Delay and stuck-at faults caused by NBTI and EM with the pre-operation test vector in the FDSOI SRAM. The color coding is the same as in Fig. 8.1.	77
8.6.	Delay and stuck-at faults caused by NBTI and EM at the voltage test corner in the FDSOI SRAM. The color coding is the same as in Fig. 8.1.	78
8.7.	Delay and stuck-at faults caused by NBTI and EM at the voltage test corner in the FinFET SRAM. The color coding is the same as in Fig. 8.1.	79
8.8.	DVFS strategy to mitigate delay faults caused by NBTI and EM at the nominal operation corner in the FDSOI SRAM. Delays are shown in picoseconds. Green entries indicate successful write operations within the critical T_{switch} . Solid red entries indicate unrecoverable delay or stuck-at faults. Solid red entries with green triangles denote faults that can be recovered by 50% frequency downscaling. White entries represent regions where frequency scaling is necessary.	80
8.9.	DVFS strategy to mitigate delay faults caused by NBTI and EM at the nominal operation corner in the FinFET SRAM. The color coding is the same as in Fig. 8.8.	81
9.1.	(a): The one-write-one-read (1W1R) register file bit-cell includes labeled driver locations (e.g., write bit line, $Write_{BL}$), access transistors (e.g., AW_{BL}) connected to bit lines (BL and BLB), and parasitic elements (e.g., R_b , C_b) on BL and BLB. (b): Comparison of SA switching speed for FDSOI and FinFET SRAM-based register files at $WL = 300$ ps for the nearest cell to the SA.	84
9.2.	Error probability of the FDSOI SRAM-based register file under various linear energy transfer (LETs), with arrows representing the increase in error probability for $R0$ operations. LET: (a) $48 \text{ MeV} \cdot \text{cm}^2/\text{mg}$, (b) $60 \text{ MeV} \cdot \text{cm}^2/\text{mg}$, (c) $72 \text{ MeV} \cdot \text{cm}^2/\text{mg}$	85
9.3.	Error probability of the FinFET SRAM-based register file under various LETs. LET: (a) $1.46 \text{ MeV} \cdot \text{cm}^2/\text{mg}$, (b) $1.82 \text{ MeV} \cdot \text{cm}^2/\text{mg}$, (c) $2.18 \text{ MeV} \cdot \text{cm}^2/\text{mg}$	85
9.4.	Radiation-aware mitigation strategies. (a) FDSOI: Adaptive location-dependent WL width modulation, (b) FinFET: Compensating RC parasitic structure.	87
10.1.	Overall workflow of the ChatTCAD framework.	90
10.2.	Example CMOS device generated using the proposed ChatTCAD framework. (A) Gate Contact, (B) Source/Drain Doping and Lightly Doped Drain (LDD), (C) P-Well region. . .	91
10.3.	Detailed prompt engineering and synthetic dataset generation steps for SFM and CDM. . .	92
10.4.	Example of a complete SFM prompt for region block generation, with key keywords highlighted. .	94
10.5.	Example of a CDM system prompt, with key instructional keywords highlighted in green. .	95
10.6.	Step-by-step tabular CoT reasoning process for CDM. Tables 1–7 show incremental spatial and physical feature resolution.	96
10.7.	Cross-entropy loss comparison during supervised fine-tuning for region generation in SFM. (a) Training with a baseline GPT-4o synthetic dataset. (b) Training with the proposed pool sampling-based dataset, showing faster convergence and lower final loss.	97
10.8.	Cross-entropy loss across various functional blocks. (a) Contact, (b) Doping, (c) Meshing, (d) CDM with Tabular CoT prompting.	98

List of Tables

3.1. Hyperparameters	23
3.2. Simulation setup and adder types	24
3.3. Performance of SVC and NN models	24
3.4. Runtime comparison of ML and MC methods. ML: proposed Machine Learning approach, MC: Monte Carlo simulation, ASU: Average Speed-up	27
4.1. Simulation setup	33
4.2. Hyperparameters	34
4.3. Performance of regression models for SET propagation through NMOS structures evaluated on test datasets using RMSE and R^2 metrics	37
4.4. Performance of regression models for SET propagation through inverter structures evaluated on test datasets using RMSE and R^2 metrics	38
5.1. FDSOI parameters	45
6.1. Nominal setup parameters	56
7.1. FinFET device parameters	63
7.2. Simulation setup parameters	64
8.1. Fault thresholds under EM-only and combined NBTI+EM stress at the nominal corner. . .	73
8.2. Impact of process variation and cell location on T_{switch} and write failure probability (First vs. Last cell). μ : mean of T_{switch} , P_{delay} : probability of delay faults, P_{stuck} : probability of stuck-at faults.	75
8.3. Sensitization-induced changes in fault-threshold under a short-term ΔV_{th} stress.	76
8.4. Switching-time degradation of fresh devices and fault-threshold resistance shifts under the same NBTI stress as Table 8.3	78
8.5. Earlier detection performance	79
8.6. Improvement in tolerable ΔR_s due to DVFS under the same ΔV_{th} aging condition as Table 8.3	81
8.7. T_{switch} degradation under DVFS with ΔR_s scaled in 1.5 $\times$ increments	82
9.1. Results of error probability reduction due to mitigation for the worst-case scenario (Fig. 9.3(c)) in the nearest cell (baseline: no compensating RC) and the SRAM-based register file's read energy.	88

Acronyms

AI	Artificial Intelligence.
ALU	Arithmetic Logic Unit.
ANN	Artificial Neural Network.
ASU	Average Speed Up.
BEOL	Back-End-of-the-Line.
BL	Bit-Line.
BLB	Bit-Line Bar.
BTI	Bias Temperature Instability.
CDM	Complete-Device Mode.
CFET	Complementary FET.
CL	Capacitive Load.
CLK	Clock.
CoT	Chain-of-Thought.
CPL	Complementary Pass-Transistor Logic.
CTA	Cooperative Thread Array.
Cu	Copper.
DAT	Data.
DPL	Double Pass-Transistor Logic.
DPPM	Defective Parts Per Million.
DTCO	Design–Technology Co-Optimization.
DUT	Device Under Test.
DVL	Dual-Value Logic.
EDA	Electronic design automation.
EDP	Energy-Delay Product.
EM	Electromigration.
FA	Full Adder.

FDSOI Fully Depleted Silicon-on-Insulator.

FEOL Front-End-of-the-Line.

FinFET Fin Field-Effect Transistor.

FT Fall Time.

GAA Gate-All-Around.

GNN Graph Neural Network.

GPU Graphics Processing Unit.

HCI Hot Carrier Injection.

HPC High Performance Computing.

I-V Current-Voltage.

JSON JavaScript Object Notation.

LDPC Low Density Parity Check.

LET Linear Energy Transfer.

LLM Large Language Model.

LUT Lookup Table.

MAE Mean Absolute Error.

MC Monte Carlo.

ML Machine Learning.

MSB Most Significant Bit.

MSC Multi-State Configuration.

MSE Mean Squared Error.

NBTI Negative Bias Temperature Instability.

PBTI Positive Bias Temperature Instability.

PDK Process Design Kit.

PDP Power-Delay Product.

PH Pulse Height.

PPA Power, Performance, and Area.

PTL Pass-Transistor Logic.

PVT Process, Voltage, and Temperature.

PW Pulse Width.

RC Resistive-capacitive.

RCA Ripple-Carry Adder.

RD Reaction Diffusion.

RMSE Root Mean Squared Error.

RT Rise Time.

SA Sense Amplifier.

SER Soft Error Rate.

SET Single Event Transient.

SEU Single Event Upsets.

SFM Single-Function-Mode.

SLM Silicon Lifecycle Management.

SM Streaming Multiprocessor.

SNM Static Noise Margin.

SoC System-on-Chip.

SRAM Static Random Access Memory.

SSR Supersteep Retrograde.

STA Static Timing Analysis.

STCO System–Technology Co-Optimization.

SVC Support Vector Classifier.

Ta/Ta-N Tantalum.

TCAD Technology Computer-Aided Design.

Tddb Time-Dependent Dielectric Breakdown.

VLSI Very Large Scale Integration.

WL word-Line.