



UNIVERSITÀ DEGLI STUDI DI BRESCIA
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
DOTTORATO DI RICERCA IN INGEGNERIA DELL'INFORMAZIONE
ING-INF/01 - ELETTRONICA

STRESS-RELATED LOCAL LAYOUT EFFECTS IN FINFET DEVICES

Candidato

Angelo Rossoni

Supervisori

Dr. Tomasz Brożek

Prof. Zsolt Miklós Kovács-Vajna

Prof. Luigi Colalongo

CICLO XXXVIII, 2022-2025

Università degli Studi di Brescia, Dipartimento di Ingegneria
dell'Informazione.

Thesis submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Information Engineering.
Copyright © 2026 by Angelo Rossoni.

To my parents,
*who taught me the value of curiosity and kindness
your love was my compass.*

To my mentors,
for their guidance, patience, and belief in my potential.

To my wife,
*for your endless patience, encouragement,
and belief in me even when I doubted myself.*

And to my younger self,
this is the proof that you were right to dream big.

Acknowledgments

Let's be honest, no one finishes a PhD alone. If I had, this would be a very different document, probably written in crayon and submitted via smoke signal.

First, a massive thank-you to my supervisors, Dr. Tomasz Brozek and Prof. Zsolt Miklos Kovacs Vajna. Your wisdom, patience, and ability to decipher my ramblings have been nothing short of heroic. You've guided me through academic chaos with the calm of Jedi Masters, and for that, I am eternally grateful.

To Dr. Sharad Saxena, thank you for your technical wizardry and for answering my questions without ever sighing audibly (at least not over Teams).

To my parents, my wife, and my kids: thank you for your love, support, and for pretending to understand what I was working on. You've endured, my thesis-induced mood swings with grace, snacks, and the occasional eye-roll. I owe you everything.

A big shout-out to Dr. Kimon Michaels, Dr. Michael Yu, Dr. P.K. Mozumder, and Dr. Howard Read for letting me chase academic dreams while juggling a full-time job. You gave me the green light to be both a researcher and a professional, and somehow, I didn't explode.

To Dr. Dong-Won Kim, whom I met at a conference and somehow ended up discussing FinFETs, stress modeling, and Italian travel itineraries. Thank you for the insightful discussions and for bringing laughs to our chats.

To my family: thank you for surviving three years of distance learning with me. I know I was grumpier than a cat in a bathtub during thesis season. You deserve medals. Or at least a vacation from me.

And to my wife, Leslie Ann, thank you for being the spark that lit this fire. I still remember your ultimatum: no ring until I hit the books. Well, here we are. You've been, my motivator, my editor, my therapist, and my biggest fan. I couldn't have done this without you. Also, I now fully support your right to issue ultimatums.

Finally, to the amazing crew of friends and colleagues who kept me sane: Alberto, Alfredo, Davide, Elisa, Fabrizio, Giorgio, Giulia, Lara, Luca, Marco, Massimo, Michele, Paolo, Samuele, Simone and Vito, you're the real MVPs. Thanks for the laughs, the pep talks, and the occasional reality checks.

This thesis is the result of science, sweat, and a whole lot of coffee. Thank you all for being part of the madness.

Abstract

Semiconductor devices are scaling into the nanometer regime. Managing variability in this context is now a critical challenge for ensuring performance, reliability, and yield. One of the most significant, and often underappreciated, sources of variability in FinFET technologies is mechanical stress induced by Local Layout Effects (LLEs). These effects arise from the physical configuration of nearby structures and can significantly alter carrier mobility and transistor behavior.

This thesis presents a novel and comprehensive methodology to isolate, characterize, and model stress-related LLEs in advanced FinFET nodes. A high-density test chip, fabricated using a commercial 7nm FinFET process, integrates over 30,000 devices across a wide range of layout configurations. Electrical measurements from this chip are used to calibrate a 3D Technology Computer-Aided Design (TCAD) model that accurately predicts stress-induced performance variations.

The model incorporates detailed process assumptions, including epitaxial growth, gate stack deposition, and thermal steps, and accounts for the anisotropic stress response of silicon and silicon-germanium materials. Key findings reveal that p-type devices are particularly sensitive to stress perturbations, with performance variations exceeding 10% near certain layout features. N-type devices, while generally less sensitive, may exhibit complex behavior due to the interplay of all stress orthogonal components.

Beyond characterization, the model enables predictive evaluation of process innovations, including gate cut strategies, spacer deposition methods, and thermal effects. While it does not simulate full wafer-scale variability, it can incorporate localized process variability due to wafer gradients (e.g., deposition thickness, lithographic Critical Dimension variation, mask misalignment) to assess their impact on device behavior.

Overall, this work delivers a predictive, scalable, and experimentally validated framework for analyzing stress-related LLEs in FinFET technologies. It provides actionable insights for Design-Technology Co-Optimization (DTCO), Process Design Kit (PDK) development, and process optimization in advanced CMOS technologies.

These contributions are directly applicable to industrial semiconductor design workflows, enabling early-stage layout decisions that mitigate stress-induced variability. The methodology supports both product-level DTCO

and foundry-level process tuning, particularly in the context of 7nm and more advanced FinFET nodes.

Abstract (Italian)

I dispositivi a semiconduttore stanno raggiungendo dimensioni dell'ordine dei nanometri, rendendo la gestione della variabilit  un aspetto cruciale per garantire prestazioni, affidabilit  e resa produttiva. Tra le principali fonti di variabilit  nelle tecnologie FinFET, spesso sottovalutate, vi sono gli effetti meccanici indotti dalla configurazione locale del layout, noti come Local Layout Effects (LLEs).

Il lavoro sviluppa una metodologia sperimentale e un modello di simulazione per analizzare gli effetti di layout locale indotti da stress meccanico in dispositivi FinFET in nodi avanzati. E' stato progettato e fabbricato un chip di test ad alta densit , realizzato con un processo commerciale FinFET a 7nm, che integra oltre 30.000 dispositivi distribuiti su un'ampia variet  di configurazioni di layout. Le misure elettriche ottenute da questo chip sono state utilizzate per calibrare un modello tridimensionale TCAD in grado di prevedere con precisione le variazioni prestazionali indotte dallo stress.

Il modello incorpora assunzioni dettagliate sul processo, tra cui crescita epitassiale, deposizione dello stack di gate e trattamenti termici, tenendo conto della risposta anisotropa allo stress dei materiali in silicio e silicio-germanio. I risultati evidenziano una particolare sensibilit  dei dispositivi di tipo P alle perturbazioni da stress, con variazioni prestazionali superiori al 10% in prossimit  di specifici elementi di layout. I dispositivi di tipo N, sebbene generalmente meno sensibili, mostrano comportamenti complessi dovuti all'interazione tra le componenti ortogonali dello stress.

Oltre alla caratterizzazione, il modello consente valutazioni predittive di innovazioni tecnologiche, incluse strategie di gate cut, metodi di deposizione degli spacer di gate e impatti termici. Sebbene non simuli la variabilit  a livello di intero wafer, il modello integra la possibilit  di simulare variabilit  locali dovute a gradienti di processo (es. spessore di deposizione, variazioni litografiche di CD, disallineamenti di maschera), permettendo di valutarne l'impatto sul comportamento dei dispositivi.

Nel complesso, questo lavoro fornisce un framework predittivo, scalabile e validato sperimentalmente per l'analisi degli LLEs legati allo stress nelle tecnologie FinFET. I risultati ottenuti offrono spunti concreti per l'ottimizzazione

del DTCO (Design-Technology Co-Optimization), lo sviluppo dei PDK (Process Design Kit) e il miglioramento dei processi nelle tecnologie CMOS avanzate.

Il contributo di questo studio e' direttamente applicabile ai sistemi industriali per la progettazione di semiconduttori e circuiti integrati, consentendo sin dalle fasi iniziali di ottimizzare il design per mitigare la variabilita' indotta dallo stress. La metodologia supporta sia la co-ottimizzazione tra progettazione e tecnologia a livello di prodotto (DTCO), sia la messa a punto dei processi a livello di fonderia, in particolare nel contesto dei nodi FinFET a 7nm e piu' avanzati.

Confidentiality Disclaimer

This thesis does not disclose proprietary or confidential information belonging to PDF Solutions, Inc. or any semiconductor foundry. All process assumptions, material properties, and simulation parameters are based on publicly available sources or predictive models and are intended solely for academic research purposes.

Trademark Acknowledgment

The author gratefully acknowledges the use of proprietary tools and methodologies provided by PDF Solutions, Inc. These technologies were instrumental in the design, testing, and analysis phases of this research.

Characterization Vehicle®(CV®), pdFasTest®, and Exensio®Analytic Platform are registered trademarks of PDF Solutions, Inc. All references to these technologies in this thesis have been made in accordance with corporate trademark usage guidelines and with explicit authorization for disclosure.

This acknowledgment is made in recognition of the technical support and resources provided by PDF Solutions, Inc., which enabled the successful completion of the experimental and modeling work presented herein.

Acknowledgment of Co-authors

Several chapters incorporate material from peer-reviewed publications and submissions co-authored with colleagues. Their contributions are acknowledged as follows:

- **Dr. Tomasz Brożek, Prof. Zsolt Miklós Kovács-Vajna:** Provided scientific guidance and critical review of methodology and results
- **Dr. Tomasz Brożek, Dr. Sharad Saxena:** Offered technical insights on process assumptions and supported access to proprietary tools during experimental phases
- **Dr. Christopher Hess:** Led the collaborative development of Passive Device Array and Passive Matching Array concepts for test infrastructure optimization

All co-authors participated in discussions and manuscript reviews but did not influence the interpretation of results or the conclusions presented herein. The candidate retains full responsibility for the originality, integrity, and completeness of this dissertation.

Contents

Acknowledgments	iv
Abstract	v
Confidentiality Disclaimer and Acknowledgments	vii
Contents	ix
List of Figures	xiii
List of Tables	xix
1 Introduction and Motivation	1
1.1 Power, Performance, and Area Metrics	2
1.2 Scaling Principles and Limitations	4
1.3 Mechanical Stress in Silicon Devices	6
1.4 Carrier Mobility and Stress Sensitivity	7
1.5 Summary of Stress Effects on Silicon	9
1.6 Strain Engineering: Global and Local	10
1.6.1 Global Strain Engineering	12
1.6.2 Local Strain Engineering	13
1.6.3 Unwanted Sources of Stress	14
1.7 Scaling Challenges and Process Innovations	15
1.7.1 Process Innovations	15
1.7.2 Architecture Innovations	17
1.7.3 Intel's Process Technology Innovations	20
1.7.4 Design-Technology Co-Optimization (DTCO)	22
1.7.5 Sources of Device Performance Variability	23
1.8 Conclusions	26

2	Transistor Fundamentals and Parameters	27
2.1	MOSFET Fundamentals	28
2.1.1	Structure	28
2.1.2	Operating Regimes	28
2.1.3	Key Performance Metrics	31
2.2	Mobility as Stress Monitor	33
2.3	Conclusions	34
3	Characterization Methodology and Test Structures	35
3.1	Introduction to Local Layout Effects	37
3.2	Design of Experiment and Layout Strategy	38
3.3	Characterization Modules	48
3.3.1	Active Extension Module	49
3.3.2	Active Proximity Module	50
3.3.3	Gate Cut Proximity Module	51
3.3.4	V_T /NP Boundary Proximity Module	52
3.3.5	Fin Pitch Module	53
3.3.6	Poly Pitch Module	55
3.4	Device Testing and Data Acquisition	56
3.5	Results and Analysis	57
3.5.1	Die Level Analysis	57
3.5.2	Wafer Level LLE Analysis	57
3.6	Conclusions	61
4	Modeling Framework and Assumptions	63
4.1	FinFET Architecture and Process Flow	65
4.2	Simulation Workflow	67
4.3	Process Assumptions: Geometry and Materials	71
4.3.1	Device Geometries	72
4.3.2	High-K Gate Dielectric	72
4.3.3	Gate Metal Stack Deposition	74
4.3.4	Gate Cut Isolation	76
4.3.5	Epitaxial Growth in Source and Drain Module	77
4.4	Conclusions	79
5	Simulation Configuration and Model Calibration	81
5.1	TCAD Process Simulator Setup	82
5.1.1	General Setup	82

5.1.2	Tensor Field Simulation	83
5.1.3	Mesh Definition	84
5.1.4	Mesh Independence and Solver Convergence	85
5.2	Implemented Process Steps	86
5.2.1	Integration with Device Simulation	87
5.3	TCAD Device Simulator Setup	88
5.3.1	Mesh re-Definition	88
5.3.2	Physical Models	89
5.4	Calibration Strategy and Results	100
5.4.1	Gate Capacitance and EOT	100
5.4.2	Subthreshold Swing	101
5.4.3	Drain Current in Linear Region	101
5.4.4	Threshold Voltage	102
5.4.5	Stress Effects on Saturation Velocity in High Electric Fields	102
5.4.6	Calibration Results	103
5.5	Conclusions	104
6	Experimental Results and Simulation Validation	105
6.1	Performance Variability Summary and Methodology	107
6.2	Impact of Diffusion Break Proximity	108
6.3	Gate Cut Proximity Effects	115
6.4	Poly Pitch and Stress Modulation	121
6.5	Fin Pitch and Patterning Variability	127
6.5.1	Spacer-Assisted Quadrupole Patterning	132
6.6	Fin Shielding Effect	135
6.7	Conclusions	137
7	Process Innovations and Impact on Performance	139
7.1	Impact of Gate Cut Strategies	140
7.2	Gate Spacer Deposition Techniques	143
7.3	Impact of Dielectric Isolation Materials on Performance	146
7.4	SDB Etching Process Window	147
7.4.1	Process Window Optimization	149
7.5	Impact of Spacer Thickness Variability	150
7.5.1	Importance of Process Control	152
7.6	Temperature Effects on Stress Sensitivity	153

7.7	Conclusions	155
8	Applications, Limitations, and Future Extensions	157
8.1	Applications	158
8.1.1	Design-Technology Co-Optimization	158
8.1.2	Process Design Kit Development	159
8.1.3	Process Optimization	161
8.1.4	Sensitivity to Process Variability	161
8.1.5	Impact on Device and Circuit Reliability	162
8.2	Limitations of the Current Framework	163
8.2.1	Assumptions in Process Parameters	163
8.2.2	Applicability to Other Nodes or Foundries	163
8.2.3	Impact of Measurement Noise and Tester Limitations	164
8.2.4	Full Wafer-Scale Variability Modeling	164
8.2.5	Simplified Stress Tensor Analysis	165
8.2.6	Static Process Flow and Geometry	165
8.2.7	Computational Cost and Scalability	165
8.3	Future Extensions	166
8.3.1	Integration of AI/ML	166
8.3.2	Modeling Compound LLE Interactions	169
8.3.3	Dynamic Process Modeling and Geometry Adaptation	170
8.3.4	Extension to Gate-All-Around	170
8.4	Conclusions	172
A	Summary of Physical Models Used in TCAD Simulation	175
B	Mechanical Properties of Dielectric Materials	177
C	Compact-Model Integration (BSIM-CMG v111.x)	179
D	Reproducibility Pack	183
E	Publications	189
	Glossary	191
	Bibliography	195

List of Figures

1.1	Typical standard cell layout	3
1.2	Cost trends and scaling projections according to Moore's Law	4
1.3	Dennard's scaling	5
1.4	Definitions of stress and strain	6
1.5	Strain vs stress relation	7
1.6	Mobility variation as a function of stress	8
1.7	Impact of stress components on transistor drain current . . .	9
1.8	Lattice mismatch technique applied to a bulk device	12
1.9	Epitaxial growth concept in p-type FinFET structure	14
1.10	Cross-sectional SEM image of a notched-gate MOSFET . . .	16
1.11	Energy Band Diagrams	17
1.12	Architecture evolution: Planar, FinFET, GAA	18
1.13	Cross-sectional TEMs of several architectures	18
1.14	Example of technology node evolution (from Intel TM)	20
1.15	Standard cell scaling from Intel 7 to Intel 4	22
1.16	Electron microscope photograph of FinFET	25
2.1	Typical MOSFET cross-section and layout	29
2.2	Band diagram of the MOSFET structure	29
2.3	MOSFET $I_D(V_{GS})$ curves in linear and saturation regimes .	31
3.1	Overview of the testing workflow	39
3.2	Illustration of the transistor layout in standard cell environment	41
3.3	Local environment of the Device Under Test	42
3.4	Schematic of Single Padded infrastructure	43
3.5	Schematic of a typical Passive Device Array infrastructure .	44
3.6	Schematic of a typical Active Device Array infrastructure . .	45
3.7	Modified schematic of the DUT of the Passive Matching Array	46

3.8	Schematic of Passive Matching Array infrastructure	47
3.9	Comparison of Area Efficiency, Design Complexity, Reliability, and Test Complexity for different test infrastructure architectures	47
3.10	Test structure to evaluate the Diffusion Break proximity impact	49
3.11	Test structure to evaluate the Active proximity impact . . .	50
3.12	Test structure to evaluate the Gate Cut proximity impact . .	51
3.13	Test structure to evaluate the Vt proximity impact	52
3.14	Test structure to evaluate fin-to-fin space variation impact .	53
3.15	Test structure to evaluate the Poly Pitch impact	56
3.16	Raw I_d - V_{gs} curves measured on Reference and SDB-terminated devices	58
3.17	Raw I_d - V_{gs} logarithmic curves measured on Reference and SDB-terminated devices	58
3.18	Measured results of the Drain current degradation in p-type transistors neighboring a SDB or DDB	59
3.19	Gate Cut distance dependence of the degradation of the p-type device drain current	59
3.20	Cumulative plots of measured drive currents for Reference and SDB transistors	60
3.21	Wafer maps of I_{dLIN} for Reference and SDB terminated transistors	60
4.1	Intel™ 22nm FinFET technology node	66
4.2	Simplified process steps for a 7nm FinFET technology node	67
4.3	Typical layout of a FinFET transistor surrounded by different layout features	68
4.4	Simulated cross-section transversal to the Fin	68
4.5	Workflow developed in this study to predict Local Layout Effects	69
4.6	Sample of the process simulator output	70
4.7	3D structure of a double Fin p-type device and its cross-section, showing geometric parameters	73
4.8	Equivalent Oxide Thickness schematic concept	74
4.9	Gate metal stack deposited over silicon fin FET devices . . .	75
4.10	Different implementations of the Gate Cut process step order	77
4.11	Cross-sectional SEM transversal to the fin	77
4.12	Silicon epitaxial growth simulated over time by LKMC module	78

4.13	Typical diamond shapes in single and double fin devices . . .	79
5.1	Convention of the 3D orthogonal directions	83
5.2	Mesh examples in process simulator for a double fin structure	85
5.3	Process flow described and modelled in this study	87
5.4	Device channel mesh used as input for the device simulator .	88
5.5	Impact of the calibration steps on the transfer function . . .	103
5.6	Experimental and simulated transfer function comparison for reference devices	103
5.7	Experimental and simulated transfer function difference for reference devices	104
6.1	Test structure to evaluate the Diffusion Break proximity effect	109
6.2	3D structure used to evaluate the Diffusion Break proximity effect	109
6.3	Changes in p-type drain current as function of DB proximity	110
6.4	Variation of the 3D components of the channel stress as a function of DB proximity in p-type devices	111
6.5	Longitudinal compressive stress of the silicon as a function of the distance from the Diffusion Break	111
6.6	Cross-section longitudinally to the fin and the corresponding compressive stress	112
6.7	Changes in n-type drain current as function of DB proximity	113
6.8	Analysis of the longitudinal stress variation across several de- vice channels as function of SDB proximity	114
6.9	Test structure layout used to evaluate the impact of Gate Cut proximity on device performance	116
6.10	3D structure created by the process simulator to study GC proximity	116
6.11	Changes in p-type drain current as function of GC proximity	117
6.12	Variation of the 3D components of the channel stress as a function of GC proximity in p-type devices	118
6.13	Representation of the vertical stress in the nitride walls with different GC distances	118
6.14	Representation of the longitudinal stress in the nitride walls with different GC distances	119
6.15	Distribution of the vertical component of the stress with dif- ferent GC proximity	119

6.16	Changes in n-type drain current as function of GC proximity	120
6.17	Variation of the 3D components of the channel stress as a function of GC proximity in n-type devices	121
6.18	Detail of the 3D process simulation of different poly pitches with relative source epitaxial growth	123
6.19	Changes in p-type drain current as function of poly space variation	124
6.20	Distribution of the stress variation in p-type devices as function of poly space variation	124
6.21	Changes in n-type drain current as function of poly space variation	125
6.22	Simulated stress profiles at the channel center in n-type devices	126
6.23	Distribution of the stress variation in n-type devices as function of poly space variation	126
6.24	3D process simulation of different fin pitches showing the different stressor shapes	128
6.25	Changes in n-type drain current as function of fin pitch variation	130
6.26	Distribution of the stress variation in n-type devices as function of fin pitch variation	130
6.27	Changes in p-type drain current as function of fin pitch variation	131
6.28	Distribution of the stress variation in p-type devices as function of fin pitch variation	132
6.29	SAQP pattern propagation sequence	133
6.30	Positions in the sea of fins of a 4-fins device relative to the mandrel of a SAQP patterning technique	134
6.31	Simulated cross section across the channel of 4-fins p-type	134
6.32	Changes in p-type drain current as function of fin pitch variation for different spatial positions	135
6.33	Profiles of vertical and transversal stress variation in both channels of a n-type devices as function of GC proximity . .	136
7.1	Different implementations of the Gate Cut process step order	141
7.2	3D process simulation results for different GC processes . . .	142
7.3	Comparison of the drain current variation in the linear region for p-type devices as a function of the distance of the Gate Cut for different process techniques	143

7.4	Performance change in n-type and p-type devices depending on the gate nitride spacers deposition methods	145
7.5	Cross-sectional view along the fin line showing longitudinal stress variation for two SDB trench widths for p-type devices: (left) small SDB size and (right) large SDB size. The color scale represents longitudinal stress in Pascals	148
7.6	3D stress components as a function of SDB trench width for (a) n-type and (b) p-mos devices. Components follow the coordinate convention of Fig. 5.1: x vertical (normal to wafer), y lateral (across fin), z transport (along fin)	149
7.7	Impact of SDB trench width on I_{Dlin} for (a) n-type and (b) p-type devices	150
7.8	Analysis of stress and performance variations induced by a change in the silicon nitride gate spacer thickness	152
7.9	Simulated I-V curves for p-type devices at 300K and 375K	153
7.10	Simulated linear-region drain current for (a) n-type and (b) p-type devices as a function of distance from an SDB at 300K and 375K	154
7.11	Simulated variation in linear-region drain current versus SDB proximity at 300K and 375K. Mobility sensitivity to stress decreases significantly with temperature	155
8.1	Integration of AI/ML into TCAD-PDK-DTCO workflow for stress-aware modeling and optimization	166
8.2	Process flow modeled in this work, with the additional the process steps dedicated to GAA manufacturing (left); a detail of the 3D GAA structure simulated by the updated process model (right)	172

List of Tables

1.1	Summary of the channel mobility sensitivity to stress	11
2.1	Transistor main metrics	32
3.1	Most known Local Layout Effects impacting transistor variability in advanced FinFET technologies	38
3.2	Layout experiments implemented on the test chip to study and monitor stress-related LLEs	42
5.1	Masetti model: default coefficients	90
5.2	RPS model: default coefficients	91
5.3	Canali model: default coefficients	91
5.4	Doping-dependent SRH lifetime: default coefficients	93
5.5	Auger recombination model: default coefficients	94
6.1	Performance Variability Caused by Stress-LLEs	108
7.1	Predicted ΔI_{Dlin} (%) as a function of material in SDB trench	147
7.2	Summary of DTCO and PDK Actions for Each Process Innovation	156
8.1	Mapping of LLEs to BSIM Parameters and Physical Impact	160
A.1	Summary of Physical Models	176
B.1	Estimated Thermal and Mechanical Properties of Alternative Dielectric Materials	178
D.1	Geometry & orientation assumptions (FinFET for 7 nm node)	184
D.2	High- k metal gate stack and EOT target	184
D.3	Doping/epi assumptions for front end-of-line process layers .	184

D.4	Key thermal steps assumptions (nominal setpoints)	185
D.5	Baseline meshing and solver acceptance criteria (Sections 5.1.3- 5.1.4)	185
D.6	Parameters adjusted during calibration on reference struc- tures (Section 5.4)	185

Chapter 1

Introduction and Motivation

As semiconductor devices scale into the nanometer regime, variability induced by mechanical stress and layout interactions has become a major challenge in maintaining performance and reliability.

This chapter introduces the technological context of the research and analyzes the evolution of FinFET nodes toward sub-10nm geometries. It discusses how Design-Technology Co-Optimization directly influences Power, Performance, and Area metrics, with specific reference to strain engineering techniques. Particular attention is given to stress-related Local Layout Effects, which are increasingly significant in FinFET and GAA technologies. These effects, often overlooked in early design stages, can critically impact carrier mobility and device behavior. This chapter sets the stage for the detailed investigation of stress-induced variability presented in the following sections of this dissertation.

Overview

In contemporary life, the use of Integrated Circuits (ICs) is pervasive.

ICs are embedded in a multitude of everyday devices, including but not limited to cellphones, laptops, personal computers, dishwashers, automobiles, LED lamps, and credit cards.

The impetus for upgrading our devices often stems from the desire for extended battery life (reduced power consumption), enhanced performance (increased speed), and/or additional integrated features (miniaturized circuits).

This chapter introduces the technological context and motivation for this research. It begins by reviewing the evolution of CMOS scaling and the associated challenges in Power, Performance, Area (PPA). It then explores the role of mechanical stress in modern FinFET devices, culminating in a discussion of Local Layout Effects (LLEs) and their impact on device variability. The chapter concludes by motivating the need for a predictive and experimentally validated modeling framework.

1.1 Power, Performance, and Area Metrics

This section delves into the PPA metrics, which are crucial for characterizing CMOS technology. These metrics are considered alongside other important factors such as cost, yield, and reliability.

Power and Performance

Power and performance are closely related in the characterization of devices. In terms of power of a specific circuit, the dynamic and static powers are deduced from its currents as following:

$$P_{TOT} = \alpha P_{DYN} + P_{STAT} \quad (1.1)$$

where α is the activity factor, which depends on the application.

The dynamic power of a circuit is given by[1]:

$$P_{DYN} = I_{DYN} * V_{DD} \quad (1.2)$$

where the dynamic current is the current flow that charges and discharges all the parasitic capacitances and the output load when a switch on the input

of the circuit occurs. The static power by:

$$P_{STAT} = I_{DDQ} * V_{DD} \quad (1.3)$$

where I_{DDQ} is the current flowing through the circuit when it is not switching, and is also known as standby current or leakage current. So, the total power is:

$$P_{TOT} = V_{DD}(\alpha I_{DYN} + I_{DDQ}) \quad (1.4)$$

And it is directly proportional to the supply voltage and to the currents flowing through the circuits. Therefore, as it is seen in Figure 1.3a, the power metric is improved by scaling.

Area

The final component of PPA analysis is Area, which is evaluated here with reference to a typical standard cell layout (see Figure 1.1). A standard cell represents the physical implementation of a logic function using transistors. These cells include combinational logic elements (e.g., inverters, NAND, NOR gates), sequential elements (e.g., flip-flops), and other functional blocks.

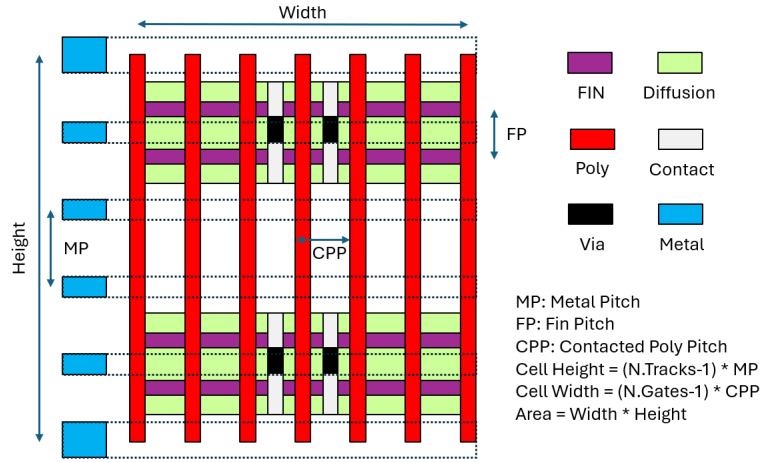


Figure 1.1: Typical standard cell layout: active (green), fin (purple), gate (red), active contact (white), gate contact (black), first level of metal (blue)

The area of a standard cell is defined by its height and width. The height is typically determined by the minimum metal pitch (MP), which dictates

the maximum number of vertical routing tracks. The width is defined in terms of the number of Contacted Poly Pitch (CPP), which corresponds to the horizontal pitch of transistor gates. The product $MP \times CPP$ is therefore a robust FOM for assessing technology node density.

Another reliable indicator of technology node density is the area of an SRAM bitcell, which does not conform to the standard cell footprint but employs a full custom layout designed to minimize area. By definition, the area metric can also be enhanced through scaling.

1.2 Scaling Principles and Limitations

Moore observed that there exists an optimal transistor count that minimizes the cost per component (Figure 1.2a) [2]. This insight stems from the interplay between two opposing trends: on one hand, miniaturization enhances manufacturability and reduces cost; on the other, increased circuit complexity leads to yield degradation, thereby raising costs.

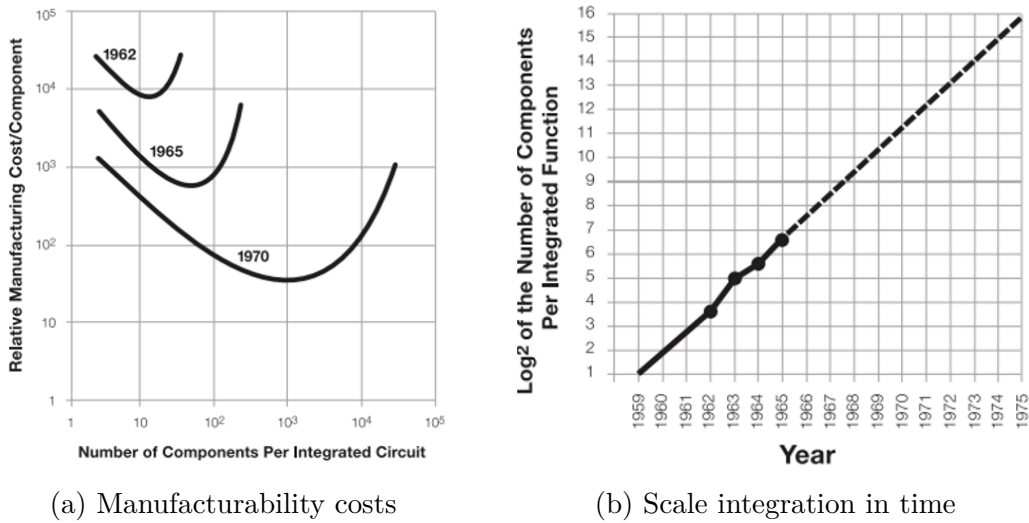


Figure 1.2: Cost trends and scaling projections according to Moore's Law [2]

These considerations led Moore to formulate his well-known projection that the number of transistors on a chip would approximately double every two years [2] (Figure 1.2b). For several decades following the invention of the transistor, empirical data consistently validated this trend. Transistor dimensions and overall chip area have been continuously reduced. This

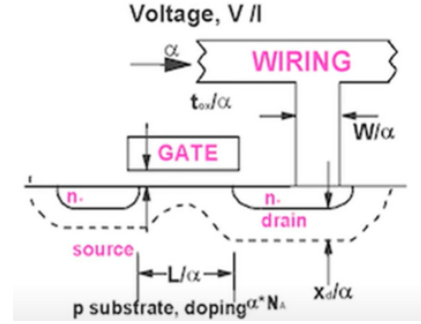
has enabled cost reduction and established scaling as the primary driver of CMOS technology advancement.

It soon became evident that device scaling not only reduced costs but also improved performance. This phenomenon was rigorously analyzed by Dennard et al. [3], who demonstrated that uniform scaling of all three spatial dimensions of the transistor, including material thicknesses, by a factor of k , results in reduced delay while maintaining constant power density.

The simultaneous benefits in cost, performance, and power efficiency resulting from dimensional scaling are collectively referred to as happy scaling, a term that encapsulates the synergistic effects of Moore's and Dennard's scaling principles. As shown in Table 1.3a, improvements in PPA metrics are predominantly attributed to aggressive scaling strategies.

SCALING RESULTS FOR CIRCUIT PERFORMANCE	
Device or Circuit Parameter	Scaling Factor
Device dimension t_{ox} , L , W	$1/\kappa$
Doping concentration N_a	κ
Voltage V	$1/\kappa$
Current I	$1/\kappa$
Capacitance $\epsilon A/t$	$1/\kappa$
Delay time/circuit VC/I	$1/\kappa$
Power dissipation/circuit VI	$1/\kappa^2$
Power density VI/A	1

(a) Scale parameters



(b) Parameters representation

Figure 1.3: Dennard's scaling[3]. Reducing transistor dimensions (gate length, width, and oxide thickness) by a constant factor k improves delay while maintaining constant power density. This approach, known as "happy scaling," benefits cost, functionality, and performance

In the context of ideal scaling, all physical dimensions of the transistor, including channel length, oxide thickness, and junction depth, as well as the supply and threshold voltages, are scaled down simultaneously (Figure 1.3b).

Applying a linear scaling factor of approximately 0.7x leads to significant improvements in key performance metrics. Specifically, power consumption is reduced by approximately 50%, while circuit delay improves by a factor of 0.7, thereby enhancing overall performance. Moore's projection of doubling the transistor count per unit area is realized, as the area occupied by each transistor scales by a factor of *area scaling factor* = $0.7^2 \simeq 0.5$. This quadratic reduction in area per device underpins the exponential growth in

integration density that has characterized CMOS technology scaling.

1.3 Mechanical Stress in Silicon Devices

Applying a force P , called stress, to a 1-dimensional material of length L , the Strain ε is defined as:

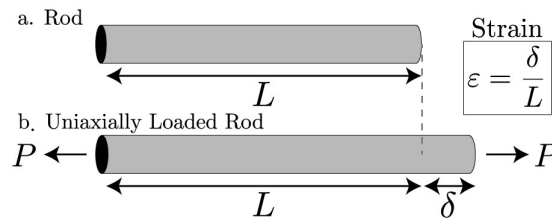


Figure 1.4: Definitions of stress and strain

Each material exhibits a distinct response when subjected to mechanical stress. By plotting the stress-strain curve $P(\varepsilon)$ for a generic material, as illustrated in Figure 1.5, we can typically identify four characteristic regions:

Hooke's Law In this initial phase, the stress-strain relationship is linear, and the slope $\frac{\Delta P}{\Delta \varepsilon}$ corresponds to Young's Modulus. The material behaves elastically, meaning the deformation is fully reversible upon removal of the applied stress.

Strain hardening Beyond a certain strain threshold, the material begins to exhibit plastic behavior. The deformation becomes partially irreversible, as the crystal lattice starts to accumulate permanent changes.

Necking After reaching the ultimate tensile strength, the material enters a fully plastic regime. The lattice structure is significantly damaged, and the material can no longer recover its original shape.

Fracture At this final stage, the material can no longer withstand the applied stress and ultimately breaks.

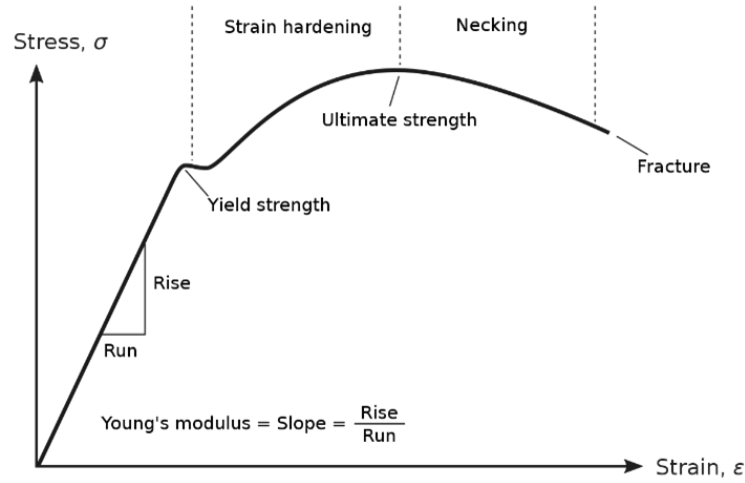


Figure 1.5: Strain vs stress relation

The Hooke's Law region is extensively studied in silicon engineering because it represents the elastic regime of mechanical deformation, where the material returns to its original state without introducing defects into the crystal lattice. Operating within this region allows for modulation of the semiconductor's electrical properties, such as carrier mobility and band structure, without degrading the material's structural integrity [4].

1.4 Carrier Mobility and Stress Sensitivity

Carrier mobility (μ) is a fundamental parameter in semiconductor physics, quantifying how quickly charge carriers (electrons or holes) can move through a material under an electric field [5]. Understanding how mechanical stress modulates carrier mobility is essential for quantifying the impact of layout-induced stress variations in FinFET devices.

In silicon, electrons exhibit significantly higher intrinsic mobility than holes due to their lower effective mass and reduced scattering rates. This asymmetry is well-documented in the literature [6], and it underpins the inherent performance advantage of n-type MOSFETs over their p-type counterparts.

This type of imbalance presents challenges in CMOS design, particularly in achieving symmetric rise and fall times, maintaining robust noise margins, and minimizing leakage energy. Historically, designers compensated for this

by increasing the width of p-type transistors. However, as scaling constraints tighten, this approach becomes less viable, necessitating alternative strategies such as strain engineering to enhance hole mobility.

Building on the intrinsic mobility asymmetry, Esseni and Conzatti [7, 8] conducted a seminal study quantifying the impact of mechanical stress on carrier mobility in silicon. In the literature, hole mobility enhancement in p-type devices under uniaxial compressive stress along $\langle 110 \rangle$ can exceed +100–140% at ~ 1 –1.5 GPa, with a non-linear trend above ~ 1 GPa and no saturation up to ~ 3 GPa in some configurations. Electron response is more modest and strongly direction-dependent. These sensitivities depend on wafer orientation and stress components (longitudinal, vertical, transverse), motivating the full-tensor approach adopted in this work. This differential sensitivity makes stress engineering particularly effective for enhancing p-type device performance, thereby narrowing the performance gap with n-type devices.

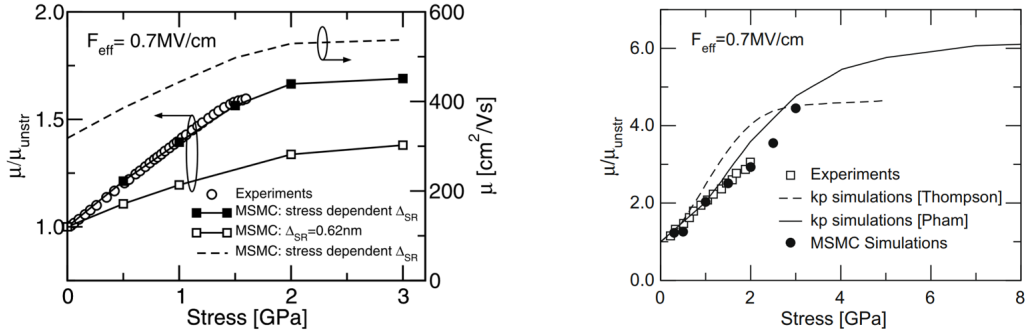


Figure 1.6: Mobility variation of electrons (left) [7] and holes (right) [8] as a function of longitudinal stress

Further refinement of stress engineering strategies requires understanding the directional dependence of stress effects. Chiang et al. [9] investigated the influence of stress orientation on planar MOSFETs fabricated on both Si and SiGe materials with different wafer surface orientations (see Figures 1.7). Their results showed that longitudinal compressive stress enhances hole mobility but degrades electron mobility, while vertical and transverse stress components exhibit more complex, orientation-dependent effects. These insights are critical for optimizing stress profiles in FinFET and Gate-All-Around Field Effect Transistor architectures, where three-dimensional geometries introduce additional complexity in stress distribution.

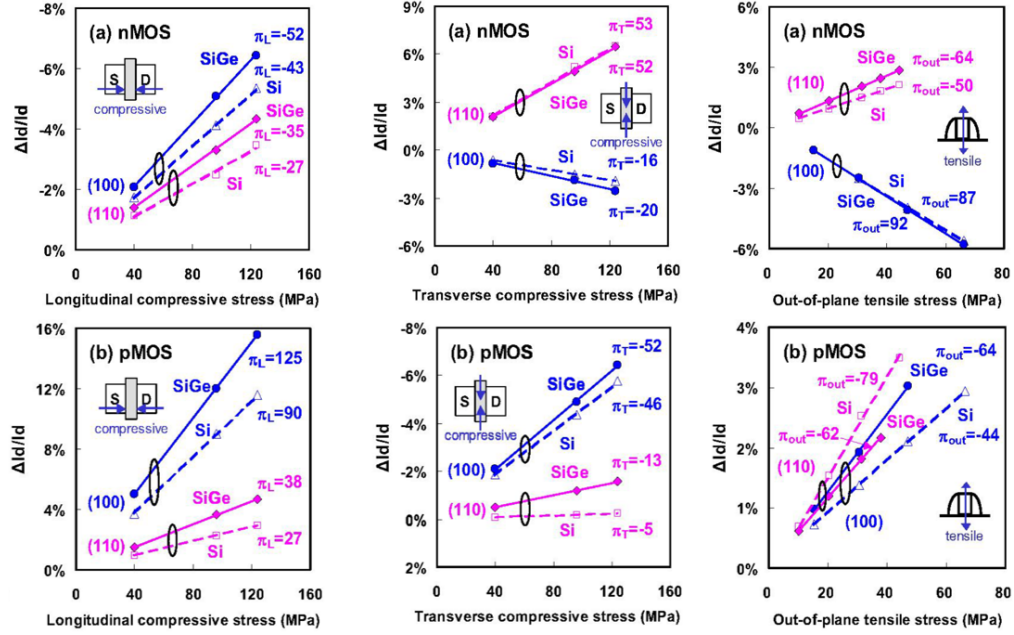


Figure 1.7: Impact of different stress components on the drain current of a MOSFET fabricated on (100) and (110) silicon wafer orientations [9]

1.5 Summary of Stress Effects on Silicon

Silicon wafers employed in the fabrication of advanced FinFET logic devices, including the technology characterized in this work, are predominantly oriented along the (100) crystallographic plane. This orientation is favored due to several critical advantages that enhance device performance and manufacturability [10, 11, 12].

Patterning of Smaller, Higher Density Transistors

The (100) orientation offers a smoother surface morphology, which is instrumental in enabling the patterning of transistors with smaller feature sizes and higher packing densities. The reduced surface roughness minimizes defect formation and enhances lithographic precision, thereby supporting the continued scaling of device dimensions.

Superior Thermal Oxide Growth Quality

Thermal oxidation on (100) silicon yields higher-quality gate oxides compared to other orientations. This results in improved dielectric integrity, which is essential for ensuring low leakage currents and robust electrical performance in gate stacks. The superior oxide quality directly contributes to the reliability and efficiency of modern CMOS devices.

Ease of Ion Implantation

The (100) orientation facilitates more uniform and controllable ion implantation through the gate oxide. This leads to well-defined doping profiles and enhanced reproducibility of electrical characteristics across devices. Such precision is vital for achieving the stringent variability and performance targets of advanced CMOS technologies.

Standardization in Advanced CMOS Logic Devices

Owing to these cumulative benefits, the (100) orientation has become the *de facto* standard for silicon wafers in advanced CMOS logic fabrication. Its optimal combination of surface smoothness, oxide quality, and implantation behavior makes it the preferred substrate for high-performance semiconductor manufacturing.

The qualitative impact of mechanical stress on carrier mobility in (100)-oriented silicon is summarized in Table 1.1. These dependencies serve as a foundational reference for the analysis of Local Layout Effects presented in subsequent sections of this work.

1.6 Strain Engineering: Global and Local

Over the years, various techniques have been developed to apply controlled stress to transistors, both to enhance overall performance and to meet specific design targets.

When stress is applied uniformly across the entire wafer, the approach is referred to as Global Strain Engineering.

As discussed in Section 1.4, the intrinsic mobility of electrons is higher than that of holes, making n-type MOSFETs inherently more efficient than p-type MOSFETs. This imbalance presents several challenges in CMOS

Wafer Surface (100)		
Stress Direction	Electron mobility	Hole mobility
Vertical <100>	<u>Strong dependency</u> Increases with Compressive stress	Medium dependency Increases with Tensile Stress
Transverse <110>	Weak dependency Increases with Tensile Stress	Medium dependency Increases with Tensile Stress
Longitudinal <110>	Medium dependency Increases with Tensile Stress	<u>Strong dependency</u> Increases with Compressive stress

Table 1.1: Qualitative summary of the channel mobility sensitivity in Fin-FET devices fabricated on a (100)-oriented silicon wafer

technology, which dominates Integrated Circuit design. Specifically, it contributes to major issues: noise margin and V_{DDmin} degradation [13], and increased system leakage energy [14].

In earlier technology nodes, this disparity was partially mitigated by designing p-type MOSFETs with larger dimensions than their n-type counterparts. However, as scaling progressed, area constraints and architectural demands necessitated new process innovations to boost p-type device performance. This led to the development of Local Strain Engineering, where compressive or tensile stress is selectively applied to individual transistors.

The most predictable sources of stress are those intentionally introduced during the manufacturing process. However, as scaling continues, new manufacturability solutions often introduce unintended stress factors that must be accurately modeled to predict device behavior.

Additionally, layout dependent stress variability arising from the physical surroundings of a device poses a significant modeling challenge. These effects are highly sensitive to the manufacturing process and layout context, making them difficult to predict and often undesirable, as they contribute to increased variability in device performance.

1.6.1 Global Strain Engineering

Global Strain Engineering involves applying specific and controlled stress to the entire wafer.

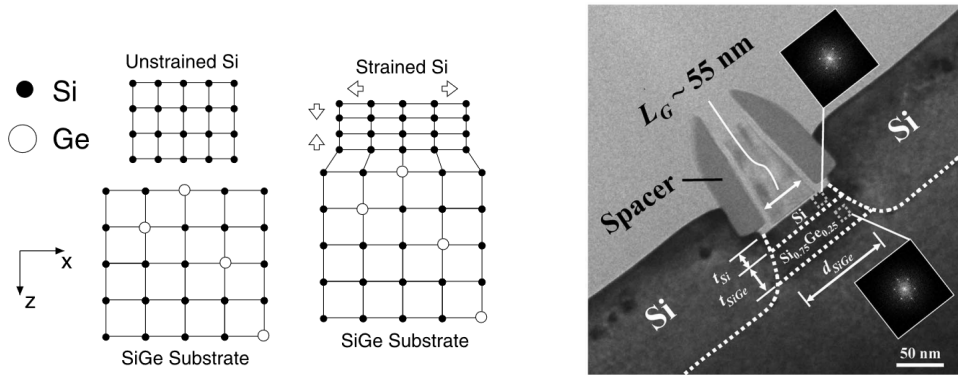
Strain Relaxed Buffer (SRB)

Lattice constant (or lattice parameter) is one of the fundamental physical dimensions that define the geometry of a unit cell in a crystal lattice. For unstrained silicon, the lattice constant is approximately 5.431 Å. In the case of silicon-germanium ($Si_{1-x}Ge_x$) alloys, the lattice constant varies with the germanium concentration x , and is expressed as:

$$Si_{1-x}Ge_x \rightarrow (5.431 + 0.20x + 0.027x^2) \text{ Å} \quad (1.5)$$

This lattice mismatch between materials is exploited to introduce mechanical stress into the crystal structure.

For example, growing a silicon layer on a silicon-germanium substrate creates a strained silicon region, as illustrated in Figure 1.8a. This technique is widely used in planar technologies to apply tensile stress to the channel region. It was successfully implemented in standard bulk devices (see Figure 1.8b), Partially Depleted Silicon on Insulator (PD-SOI), and Fully Depleted Silicon on Insulator (FD-SOI) technologies.



(a) Representation of Si growth over a substrate of SiGe, showing the lattice constant change[7] (b) Illustration of the lattice interactions at the vertical heterojunction between the Si S/D region and the $Si_{1-x} - Ge_x$.

Figure 1.8: Lattice mismatch technique applied to a bulk device

This method is particularly effective for planar n-type MOSFETs, as the induced tensile stress in both the longitudinal and transverse directions increases carrier mobility and drive current (see Table 1.1). However, the same stress conditions tend to degrade the performance of p-type MOSFETs due to the longitudinal tensile stress applied.

While global strain engineering has proven effective in earlier planar technologies, its applicability diminishes in modern 3D architectures such as FinFETs and GAA. To address the limitations of global approaches in these advanced nodes, local strain engineering techniques have been developed to target individual devices more precisely.

1.6.2 Local Strain Engineering

To address the limitations of traditional stress engineering methods in FinFET and GAA technologies, local strain engineering was developed. This technique involves the direct application of strain to the channel region to enhance carrier mobility and improve device performance. Local strain engineering is implemented through several methods, including:

- **Epitaxial Source/Drain Stressors:** Selective epitaxial growth of materials with different lattice constants induces strain in the channel region. In Figure 1.9(a), the lattice mismatch between the silicon channel and the silicon-germanium source/drain regions (see Figure 1.9) generates longitudinal compressive stress, which enhances hole mobility in p-type devices. This technique remains highly relevant even in the most advanced nodes, including FinFET and GAA architectures such as nanowires and nanosheets [15], due to their ability to uniformly distribute stress across the channel(s).
- **Contact Etch Stop Layer (CESL):** A stressed dielectric layer is deposited over the channel to stop the etch gate during contact formation [16] (see Figure 1.10). As a side effect, this layer can be engineered to induce strain in the underlying channel. Originally optimized for planar devices, CESL can also be applied to FinFET and GAA technologies, although its effectiveness is reduced due to the three-dimensional nature of these architectures.
- **Stress Memorization Technique (SMT):** This method involves applying a temporary stress layer during fabrication to induce strain in

the silicon channel [17]. Once the desired stress is transferred, the layer is removed, but the strain remains *memorized* in the silicon lattice.

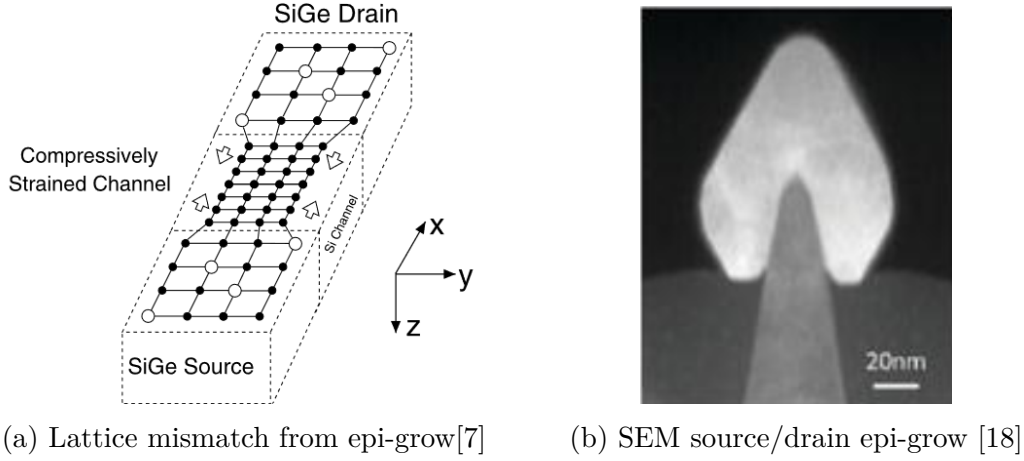


Figure 1.9: Epitaxial growth concept in p-type FinFET structure

These techniques are tailored to the specific requirements of FinFET and GAA architectures, ensuring that the benefits of strain engineering are realized in these advanced devices.

Local strain engineering enables targeted performance enhancement by applying stress directly to individual transistors, making it indispensable for modern FinFET and GAA technologies.

1.6.3 Unwanted Sources of Stress

While intentional strain engineering was widely adopted to enhance device performance, unintentional stress sources, particularly those arising from layout context, introduce variability that is more difficult to control. A stress source is generally considered undesirable when it is poorly controlled, inadequately modeled, or introduces unwanted variability in device performance. Understanding and mitigating unintended stress sources is critical, as they introduce variability that can undermine the benefits of intentional strain engineering.

These effects are collectively referred to as Local Layout Effects, which are the focus of the following sections.

1.7 Scaling Challenges and Process Innovations

”For a long time, we have lived in the happy scaling era, where every technology node reshinks and redoubles the number of transistors per area, for the same cost, but the last 10-12 years, we have not been following that happy scaling path. The number of transistors still doubles, but device scaling provides us with diminishing returns. We’ve seen these dark periods of *dark silicon* before, but, fortunately, we’ve always managed to get out of these periods. Again, the technology box will provide new features to help manage power, performance and area node by node as we move to the next generation”

An Steegen, Imec Vice President, 2017

The current challenges in semiconductor scaling arise from the physical limitations of linear scaling. Reducing the silicon oxide thickness to the angstrom scale has led to significant gate leakage and manufacturability challenges. The pitches of metals, contacts, and vias have become so small that standard lithography techniques are unable to print without introducing defects. Additionally, the length of MOSFET channels has decreased to the point where Short Channel Effect (SCE) dominates, significantly increasing leakage current.

To address these challenges and enhance scalability, new solutions have been adopted. These include advanced process techniques, innovative design approaches, and modifications to transistor architecture.

1.7.1 Process Innovations

Process innovations are pivotal not only for enabling novel device architectures but also for improving yield and scalability within a given technology node.

An example is the implementation of the Contact Etch-Stop Layer [19] in early planar CMOS technologies, which was introduced to enhance the lithographic printability of source and drain contacts and to mitigate the risk of gate-to-contact shorts. CESL is a conformal thin film deposited atop the device structure (see Figure 1.10), serving as a protective barrier for the gate

stack and underlying layers during subsequent etching processes [20]. While silicon nitride is the most commonly employed material, alternative compositions have also been explored. Deposition techniques include Low Pressure Chemical Vapor Deposition, Plasma-Enhanced Chemical Vapor Deposition, and Atomic Layer Deposition, each offering distinct advantages in terms of film quality, conformality, and stress characteristics.

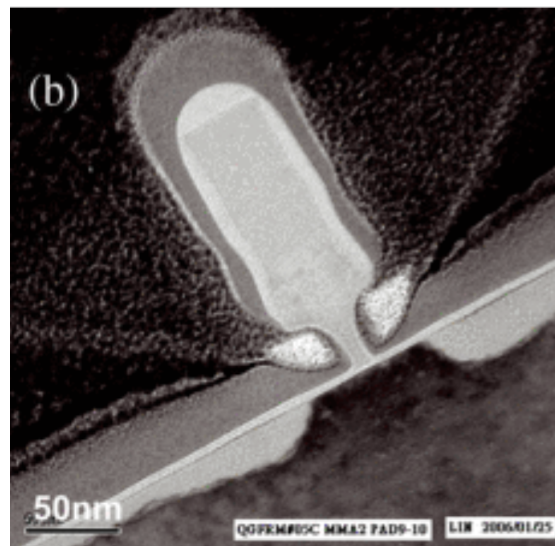


Figure 1.10: Cross-sectional SEM image of a notched-gate MOSFET. CESL induces tensile stress in the channel region via the notched-gate edge, thereby enhancing carrier mobility [19]

As the role of strain engineering in enhancing carrier transport became increasingly evident, extensive research efforts were directed toward the deliberate application of stress to transistor channels. One such technique involves source/drain engineering through SiGe epitaxial growth, which introduces longitudinal compressive stress in p-type transistors, thereby improving hole mobility.

In early VLSI fabrication, pattern fidelity was achieved using optical lithography systems operating at a wavelength of 193nm [21]. However, as technology nodes scale below 30nm, fundamental optical limitations, such as diffraction and aberration, compromise pattern resolution, leading to edge roughness, line bridging, and open defects. These challenges necessitated the development of advanced patterning techniques, most notably the Double Patterning Lithography (DPL).

DPL enables the decomposition of dense layout features into two separate masks, which are sequentially exposed and interleaved to form the target pattern. This approach significantly enhances resolution and pattern fidelity, thereby supporting higher layout densities and improved device performance.

A multitude of additional process innovations have been introduced to sustain Moore's Law by improving scalability and yield. This dissertation focuses specifically on those innovations that inadvertently introduced mechanical stress into the silicon lattice, resulting in performance variability and reliability concerns in advanced CMOS devices.

1.7.2 Architecture Innovations

Over time, scaling limitations have been addressed by improving the architecture of transistors themselves. In a planar MOSFET, as dimensions decrease, the proximity between the drain and the source reduces the gate electrode's ability to control the current flow in the channel region. The shorter the gate length, the lower the barrier from the source to the channel. This is due to the band curvature induced by the source-channel and drain-channel PN junctions, resulting in a lower threshold voltage caused by Short Channel Effect (SCE) (see Figure 1.11).

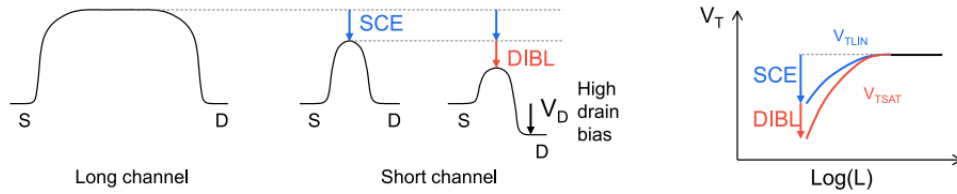


Figure 1.11: Energy Band Diagram in a long channel device (left) and in a short channel device (center) showing the Short Channel Effect (SCE) and Drain-Induced Barrier Lowering (DIBL); Effect of SCE and DIBL to the threshold voltage depending on the length of the device in logarithmic scale (right)

Similarly, due to DIBL, shrinking the gate length below 90 nm produces significant leakage current, and below 28 nm, the leakage becomes excessive, rendering the transistor less effective. In other words, decreasing the width of the devices to limit the leakage current reduces the drive current and slows down the switching.

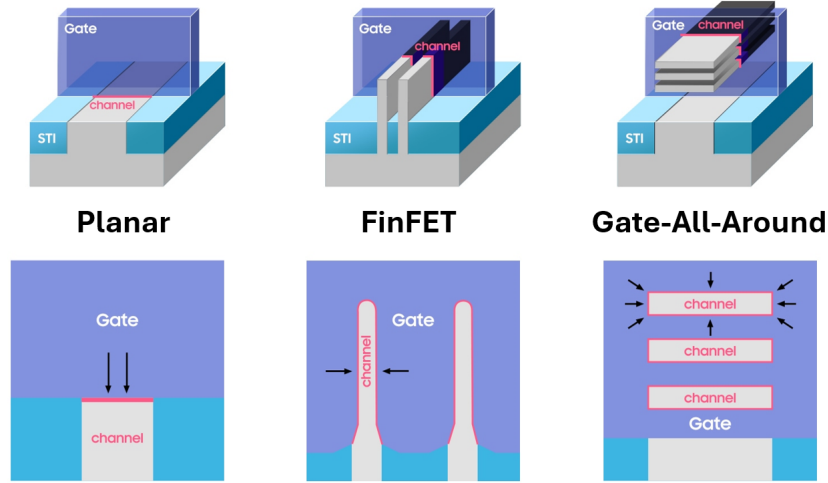


Figure 1.12: Architecture evolution: Planar device, FinFET, and GAA

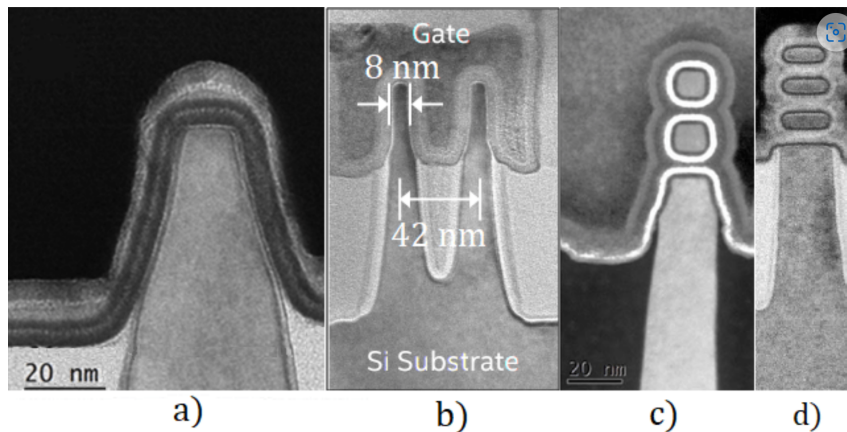


Figure 1.13: Cross-sectional TEM¹ pictures of novel 3D structures[22]: ordinary bulk sub-100 nm FinFET structure (a), 14-nm FinFET by Intel (b), 15-nm Ge NWFET² (c), and 5-nm NWFET by IBM (d)

To overcome the limitations of planar MOSFET architectures, novel device structures have been developed that offer superior electrostatic control [23], reduced switching delay, and enhanced current drive capabilities. Among these, the FinFET and the GAA transistor architectures (illustrated

¹Cross-sectional Transmission Electron Microscopy (TEM) is a high-resolution imaging technique used to analyze the internal structure of semiconductor devices.

²Nanowire Field-Effect Transistor (NWFET) and Nanosheet transistor (NSFET) are all using GAA architecture.

in Figure 1.12) represent significant milestones in transistor scaling. Representative cross-sectional images of fabricated FinFET and GAA devices are shown in Figure 1.13.

Despite their advantages, continued scaling of FinFET technology introduces challenges related to drive current and electrostatic integrity. The transition of CMOS standard cell design toward reduced track heights imposes stringent constraints on fin geometry, limiting the number of fins that is accommodated per device. Consequently, single fin devices at sub-5nm nodes are projected to fall short of the drive current requirements for high-performance applications.

Furthermore, the FinFET architecture inherently controls only three sides of the fin with the gate electrode, leaving the bottom surface electrostatically ungoverned. As gate lengths shrink, this partial gate control exacerbates short-channel effects (SCE) and increases subthreshold leakage, thereby undermining power efficiency and performance targets.

To address these limitations, the GAA transistor architecture has emerged as a promising successor. In this configuration, the gate electrode envelops the channel from all sides (see Figure 1.12), thereby maximizing electrostatic control and enabling further scaling. Various GAA implementations have been proposed, with early prototypes employing vertically stacked nanosheets as the channel medium. These nanosheets consist of discrete horizontal layers, each fully surrounded by gate material, offering superior gate-channel coupling compared to FinFET devices.

Unlike FinFET architectures, where increased drive current necessitates the use of multiple parallel fins, GAA devices achieve higher current densities by vertically stacking multiple nanosheets, effectively forming parallel conduction paths within a compact footprint. The dimensions of these nanosheets, particularly their width and thickness, is independently tuned to optimize device performance.

However, as lithographic resolution continues to improve and feature sizes shrink, the lateral dimensions of the nanosheets will inevitably approach their vertical thickness. At this scaling limit, the nanosheets effectively transition into nanowires, giving rise to the nanowire FET (NWFET) architecture, which represents the ultimate form of GAA scaling.

1.7.3 Intel's Process Technology Innovations

As a practical example of process innovations from a leading semiconductor company, this section presents Intel's roadmap and recent advancements in VLSI technology as of 2025. Intel continues to push the boundaries of CMOS scaling through a series of groundbreaking developments aimed at improving device performance, energy efficiency, and integration density. These innovations [24] include novel transistor architectures such as RibbonFET, advanced power delivery schemes like PowerVia, and next-generation lithographic techniques. Together, they represent a strategic response to the challenges of sub-7nm nodes and serve as a benchmark for industry-wide process evolution. Figure 1.14 illustrates the process innovations for the most recent technology nodes, such as FinFET and GAA, including some future projections.

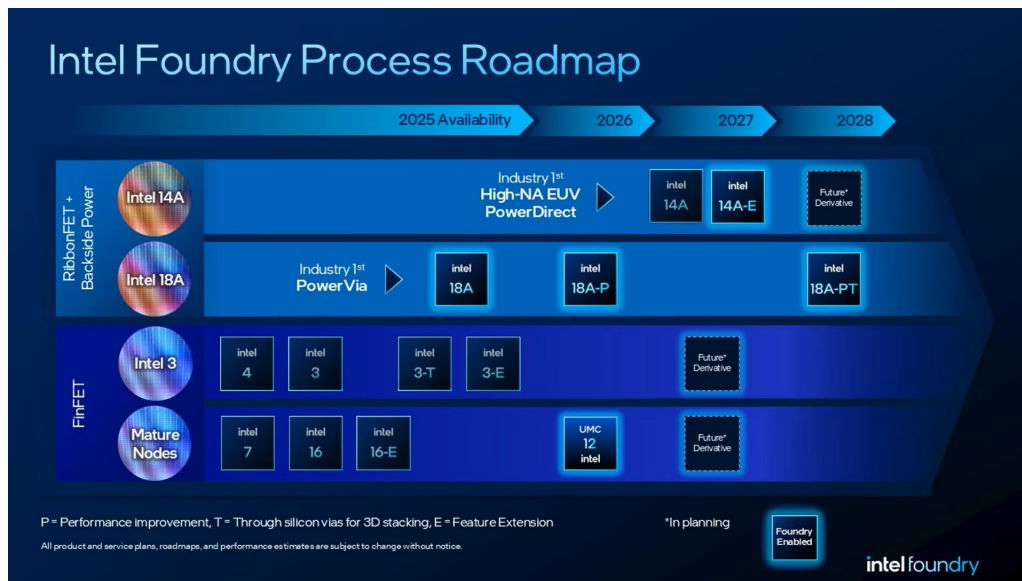


Figure 1.14: Example of technology node evolution (from IntelTM)

Joint Intel-UMC 12-nm Initiative

In a strategic move to address high-growth markets such as mobile, communication infrastructure, and networking, Intel and UMC have announced a joint development of a 12-nm photolithography process. Production at this node is scheduled to begin in 2027 at Intel's Fabs located in Arizona.

This collaboration underscores the continued evolution and relevance of FinFET technologies, demonstrating that FinFET nodes remain under active development and optimization. The 12-nm process aims to deliver enhanced performance and energy efficiency, leveraging the maturity of FinFET architecture while addressing the demands of emerging applications.

RibbonFET Technology

RibbonFET represents Intel's transition from traditional FinFET architectures to a Gate-All-Around Field Effect Transistor (GAA) transistor design. This architecture employs stacked nanosheet channels fully surrounded by the gate, thereby offering superior electrostatic control, reduced leakage, and enhanced drive current. The RibbonFET structure enables increased transistor density and improved performance-per-watt, making it a critical enabler for continued scaling beyond the 7nm node.

PowerVia Architecture

PowerVia is Intel's industry-first implementation of a Backside Power Delivery Network. By relocating power rails to the backside of the wafer, PowerVia decouples power and signal routing layers, thereby reducing IR drop, improving power delivery efficiency, and freeing up front-side routing resources. This innovation is particularly beneficial for High-Performance Computing and Artificial Intelligence workloads, where power integrity is paramount.

14A and 18A Process Nodes

Intel's 18A process node is planned for high-volume manufacturing in 2025. It incorporates both RibbonFET and PowerVia technologies, delivering substantial improvements in performance and energy efficiency. The subsequent 14A node is projected to offer a 15%-20% increase in performance and a 25%-35% reduction in power consumption relative to 18A, positioning it as a key milestone in Intel's roadmap.

Extreme Ultraviolet Lithography (EUV)

To support these advanced nodes, Intel is adopting next-generation Extreme Ultra-Violet (EUV) lithography. EUV enables the patterning of sub-20nm

features with higher precision and fewer process steps compared to traditional 193nm immersion lithography. This transition is essential for maintaining pattern fidelity and yield at advanced nodes.

1.7.4 Design-Technology Co-Optimization (DTCO)

While significant advancements have been made in transistor architecture and manufacturability, recent efforts have increasingly focused on optimizing the standard cell design itself. This holistic approach, known as Design-Technology Co-Optimization, integrates design and process technology considerations to maximize performance, power efficiency, and area utilization.

An illustrative example of the impact of DTCO is presented in Figure 1.15 [25], which compares the layout of an inverter standard cell implemented in the *Intel 7* and *Intel 4* technology nodes. Despite a modest $0.9\times$ reduction in CPP between the two nodes, aggressive DTCO strategies enable a $0.5\times$ reduction in cell area. This demonstrates the critical role of co-optimization in sustaining scaling trends even when traditional pitch scaling slows.

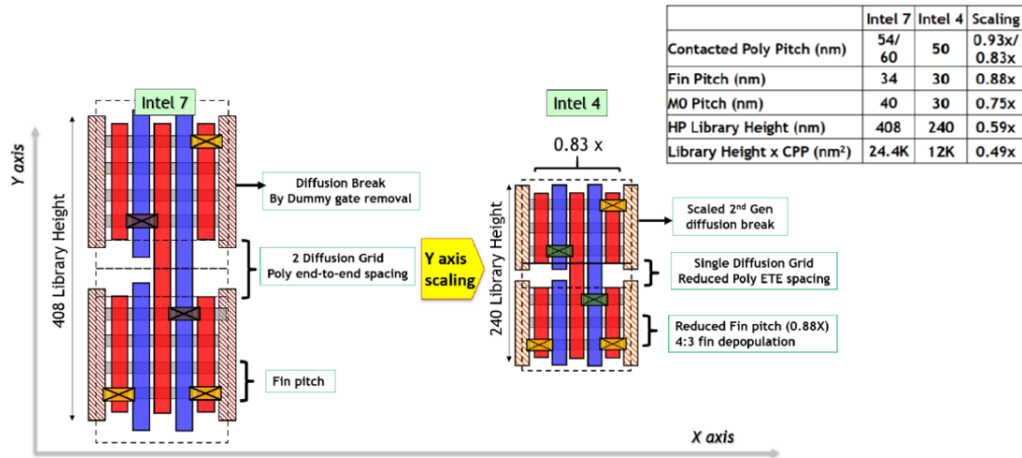


Figure 1.15: Illustration of DTCO-driven standard cell scaling from Intel 7 to Intel 4 nodes [25]

This dissertation focuses on the electrical characterization of transistors fabricated using the FinFET architecture, with particular emphasis on their sensitivity to surrounding layout features, such as gate cut and diffusion break proximity, fin pitch and poly pitch. As DTCO continues to drive the scaling of these layout features and their proximity to active devices, un-

derstanding their impact becomes increasingly critical. This understanding is essential not only for academic modeling but also for industrial applications, where layout-aware variability control directly affects yield, reliability, and time-to-market in advanced CMOS manufacturing. As will be detailed in the following chapters, variations in the local layout environment introduce mechanical stress fluctuations within the FinFET channel. These stress variations modulate carrier mobility, thereby influencing overall device performance.

1.7.5 Sources of Device Performance Variability

Device performance variability is a paramount challenge in nanoscale circuit design. Transistors intended to be identical (i.e., drawn with the same layout dimensions) exhibit discrepancies in their electrical characteristics, such as threshold voltage (V_t) or drive current (I_D). This variability stems from the inherent and unavoidable tolerances and fluctuations within the complex fabrication process.

These variations can be broadly classified into fundamental process variations, which introduce both random and local deviations, and systematic variations, which are pattern-dependent.

The electrical behavior of a transistor is extremely sensitive to its physical dimensions and composition. Minor fluctuations in these parameters, often at the nanometer or even atomic scale, translate directly into significant performance shifts. The most impactful sources include:

1. **Channel Dimensions (L, W):** The gate length (L_g) is the dominant parameter controlling Short-Channel Effects (SCEs). Even minimal variations in the effective channel length, often manifested as Line-Edge Roughness (LER), cause significant V_t fluctuations and impact subthreshold leakage. Similarly, the effective channel width (W), or in FinFETs, the 3D fin geometry (e.g., fin height H_{fin} and fin width W_{fin}), directly dictates the drive current I_D .
2. **Gate Dielectric Thickness (t_{ox}):** The thickness of the gate oxide (or high-k dielectric) directly influences the gate capacitance (C_{ox}) and, consequently, the gate's electrostatic control over the channel. Atomic-level variations in t_{ox} can alter V_t and gate leakage (I_G).

3. **Parasitic Resistance (R_{sd}, R_c):** In highly scaled nodes, parasitic resistances in the Source/Drain regions (R_{sd}) and, most critically, the contact resistance (R_c) between the metallization and the silicon, have become dominant components of the total device resistance. Variability in the contact area or the quality of the silicide-silicon interface creates strong variations in the extrinsic device performance.

Beyond these dimensional variations, a fundamental source of purely random variability is Random Dopant Fluctuation (RDF). In ultra-small channels, the discrete number and precise location of dopant atoms vary statistically from one device to another. This discrete atomistic effect causes significant, inherent device-to-device mismatch.

While the sources described above introduce a random component to variability, another critical category consists of systematic variations known as Local Layout Effects (LLEs). The electrical characteristics of the transistors depend not only on its own drawn dimensions but also on what is patterned nearby.

Unlike random variations, LLEs are systematic and, in principle, predictable and modelable.

Their impact is a primary concern during physical design.

Among the most well-known LLEs are:

- **Active area shape and distance to neighboring active regions**[26]: This includes the dimensions and spaces along the transistor channel and perpendicular to that direction. The introduction of the RMG process has restricted gate pattern design to unidirectional orientations, simplifying design and facilitating the introduction of stress as mobility boosters.
- **Gate extension beyond the edge of the channel**[27]: This has a strong impact on off-state leakage control in traditional planar MOS transistors. Additional sensitivity is introduced by the PN boundary of polysilicon gate doping or the boundary between work function metals of p-type and n-type in the case of metal gates. For RMG integration, the gate metal stack also modulates mechanical stress in the transistor channel, and the placement of the gate line end may impact device performance.

- **Well proximity and P-N boundary effects**[17]: These effects become more pronounced when the border between regions of opposite doping types is very close to the edge of the transistor, potentially leading to contamination by opposite-type dopants through implant scattering or out-diffusion, affecting either the channel or the gate of the device.
- **Device type neighborhood**: The proximity of doping boundaries in the case of threshold voltage types and boundaries between various regions can introduce dopant gradients, cross-diffusion, and channel doping variability.
- **Contact placement sensitivity**: This is a new layout effect, particularly for technologies with high contact resistance, due to current crowding effects and possible modulation of stress effects in strained dielectric layers surrounding the transistor.

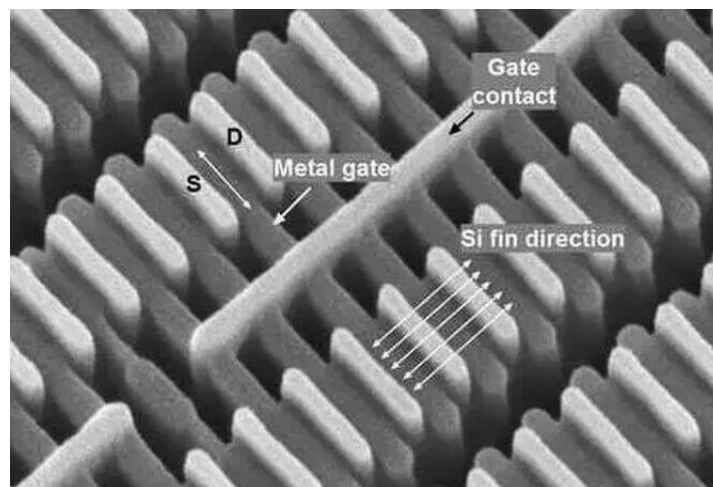


Figure 1.16: The electron microscope photograph of FinFET³

These effects become more pronounced in three-dimensional structures like FinFETs, where the fin-shaped active region is surrounded by a 3D gate structure composed of multiple metal layers, as shown in Figure 1.16. Many of these effects can occur simultaneously, triggering complex interactions

³*Introduction to FinFET* from Utmel Electronic (<https://www.utmel.com/>)

that affect transistors in multifaceted ways. For instance, an isolated p-type transistor is impacted by a P-N boundary surrounding it from all sides, with additional effects arising from density-driven volume changes in Silicon-Germanium epitaxially grown, which acts as a stressor and mobility booster in p-type devices.

The management of LLEs is a critical task for process modeling and physical design, where Design-for-Manufacturing rules and compensation techniques are employed to mitigate these systematic variations and ensure circuit yield and robustness.

1.8 Conclusions

As semiconductor devices continue to scale, the control of variability induced by Local Layout Effects has become essential to maintaining device performance, reliability, and manufacturing yield.

This chapter has traced the evolution from planar to 3D architectures, emphasizing the role of both global and local strain engineering in modulating carrier mobility. It further introduced emerging innovations such as RibbonFET, PowerVia, and DTCCO, which redefine the interaction between layout and process.

The research presented here focuses on stress-induced LLEs arising from both intentional design choices and unintended process interactions, offering a predictive framework to mitigate their impact through calibrated modeling and targeted design strategies.

Chapter 2

Transistor Fundamentals and Parameters

This chapter provides a detailed theoretical framework for evaluating MOSFET performance, with emphasis on how stress-induced variability affects transistor behavior. It links the theory of carrier mobility to TCAD modeling and experimental calibration, highlighting the role of linear drain current as a reliable indicator of stress effects in FinFET devices.

Overview

This chapter establishes the theoretical foundation for evaluating MOSFET device performance, with particular emphasis on stress-induced variability. It introduces the physical structure and operating principles of MOSFETs, defines key performance metrics, and discusses the role of carrier mobility as a stress monitor. These concepts underpin the modeling and characterization methodologies presented in subsequent chapters.

2.1 MOSFET Fundamentals

To understand how mechanical stress affects transistor behavior, it is first necessary to review the structural and operational principles of MOSFETs. This section introduces the physical layout and key functional regimes of the device.

2.1.1 Structure

The MOSFET structure determines the electrostatic control and current flow characteristics of the device. Figure 2.1(a) illustrates the cross-sectional and layout views, highlighting the gate, source, drain, and oxide regions. The control electrode, known as the *gate*, forms a MOSFET capacitance. The geometry of this capacitance is defined by the gate length (L) and the transistor width (W). The gate oxide, characterized by its thickness t_{OX} , plays a critical role in the device's performance. The *source* and *drain*, which serve as the two carrier reservoirs, are positioned on either side of the gate. Figure 2.1(b) presents a typical layout of a transistor, where the active area is illustrated in green, and the gate is shown in red [28].

2.1.2 Operating Regimes

MOSFET operation is governed by the gate voltage and its effect on the semiconductor surface potential. The MOS capacitance can operate in different regimes, determined by the potential at the semiconductor/oxide interface (ψ_s), which is influenced by the gate voltage (V_{GS}). The band diagram of the MOS structure under flat-band condition is illustrated in Figure 2.2(A).

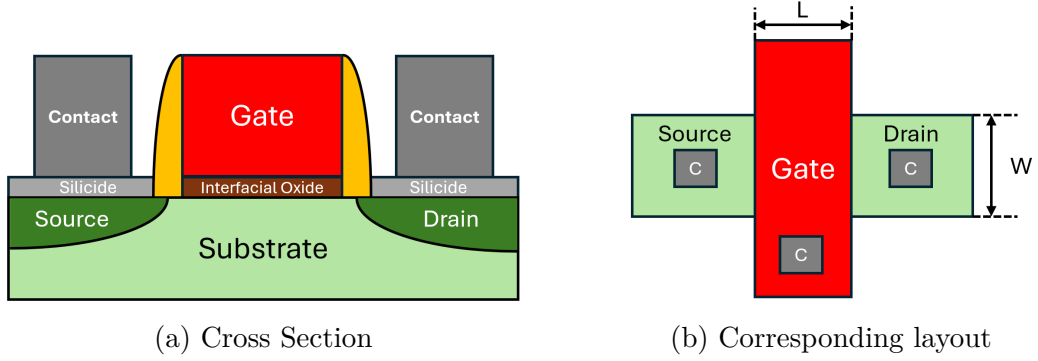


Figure 2.1: A Typical MOSFET (a) x-Section and (b) layout

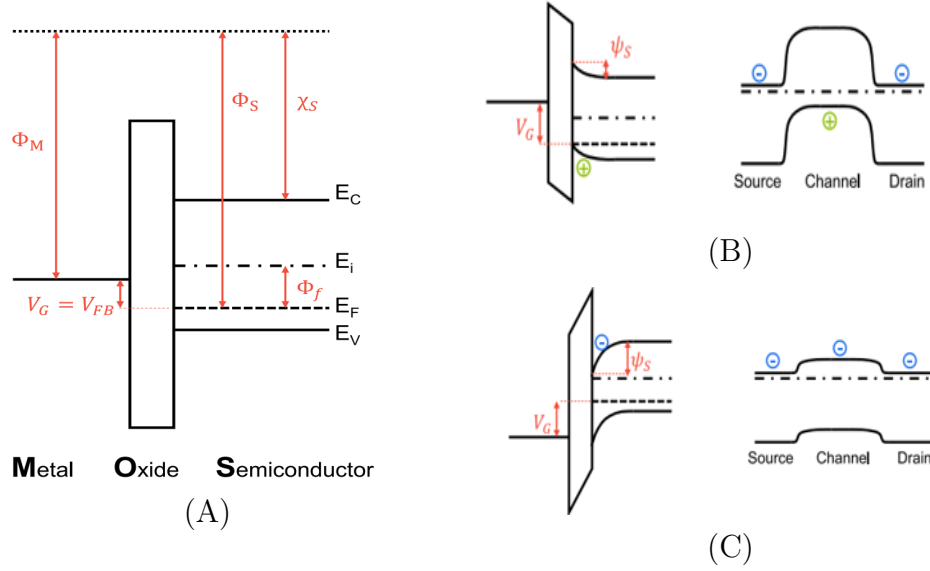


Figure 2.2: Band diagram of the MOSFET structure: Flat Band (A), Accumulation (B), Inversion (C)

Here, ϕ_M and ϕ_S represent the metal and semiconductor work functions, respectively. χ_S denotes the electron affinity, while ϕ_f is the Fermi potential. E_F is the Fermi level, and E_i is the intrinsic Fermi level. E_C represents the bottom of the conduction band, and E_V indicates the top of the valence band [5].

For a MOSFET with a p-type semiconductor, such as an n-type MOSFET, there are three distinct regimes:

Accumulation ($\psi_s < 0; V_{GS} < V_{FB}$): In this regime, the majority carriers (holes in a p-type substrate) are attracted towards the gate by the electric field. The flat-band voltage (V_{FB}) is the gate voltage required to achieve a flat energy band in the semiconductor, indicating no charge in the capacitor. Figure 2.2 (A) shows the band diagram of the MOSFET structure under flat-band condition and accumulation (B). The band diagram from the source to the drain is also depicted. A potential barrier prevents electrons from flowing from the source to the drain, placing the MOSFET in the OFF state.

Depletion ($0 < \psi_s < \psi_f; V_{GS} > V_{FB}$): The Fermi potential (ψ_f) is given by $\psi_f = \frac{kT}{q} \ln(\frac{N_A}{n_i})$, where N_A is the dopant concentration in the substrate, and n_i is the intrinsic carrier concentration. In the depletion regime, most carriers are repelled from the oxide/semiconductor interface. The charge in the semiconductor is due to ionized dopants (negative charge in the case of an n-type MOSFET).

Inversion ($\psi_s > \psi_f; V_{GS} \gg V_{FB}$) **or strong inversion** ($\psi_s > 2\psi_f$): In this regime, the previously minority carriers become the majority. In strong inversion, there are more minority carriers (electrons) than majority carriers (holes) in the bulk. A conduction path, known as the channel, forms, allowing electrons to flow between the source and drain, as illustrated in Figure 2.2 (C). The MOSFET is in the ON state.

The threshold voltage is defined as the criterion for strong inversion. In the case of the historical planar bulk technology, the threshold voltage is defined by:

$$V_T = V_{FB} + 2\psi_f + \frac{Q_{dep}}{C_{ox}} \quad (2.1)$$

where

$$V_{FB} = \psi_M - (\chi_S + \frac{E_G}{2} + \psi_f), \quad (2.2)$$

with Q_{dep} being the depletion charge and C_{ox} the gate oxide capacitance [29].

2.1.3 Key Performance Metrics

Quantitative evaluation of MOSFET performance relies on several electrical parameters, including threshold voltage, drain current, subthreshold swing, and on/off ratios. These metrics are sensitive to process variations and mechanical stress, making them essential for variability analysis.

A MOSFET is a four-terminal device, comprising the Source (S), Drain (D), Gate (G), and Bulk (B). The gate voltage (V_{GS}) serves as the input signal, while the drain current (I_D) is the output. The transfer function can be defined as $I_D = f(V_{GS})$, as shown in Figure 2.3 (logarithmic scale).

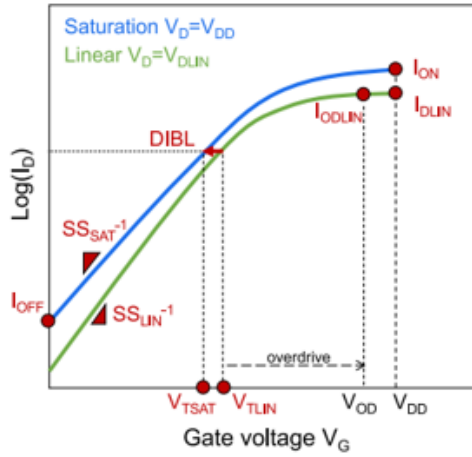


Figure 2.3: MOSFET $I_D(V_{GS})$ curves in linear and saturation regimes

Two distinct regimes are identified:

- The *linear regime*, where the voltage applied to the drain is small (typically 50 mV)
- The *saturation regime*, where the voltage applied to the drain is V_{DD}

The device's electrical performance is characterized by several parameters, as detailed in Table 2.1.

This work consistently refers to the one-point V_T .

In the subthreshold regime, the drain current increases exponentially with the gate voltage:

$$I_D = \frac{W}{L} \mu_0 C_{\text{dep}} \frac{kT}{q} \left(1 - e^{-qV_{DS}/kT}\right) \exp\left[\frac{q(V_{GS} - V_T)}{(1 + C_{\text{dep}}/C_{\text{ox}}) kT}\right] \quad (2.3)$$

Parameter	Definition	V_{GS}	V_{DS}	Unit
I_{OFF}	Leakage current	0	V_{DD}	A/um
I_{ON}	Saturation current	V_{DD}	V_{DD}	uA/um
V_{Tsat}	Saturation threshold voltage		V_{DD}	V
V_{Tlin}	Linear threshold voltage		V_{Dlin}	V
SS_{sat}	Saturation Subthreshold Swing		V_{DD}	mV/dec
SS_{lin}	Linear Subthreshold Swing		V_{DD}	mV/dec
I_{Dlin}	Linear drain current	V_{DD}	V_{Dlin}	uA/um
I_{ODlin}	Overdrive drain current	$V_{Dlin} + V_{DD}$	V_{Dlin}	uA/um
R_{ON}	ON resistance	$V_{Tlin} + V_{OD}$	V_{Dlin}	$\Omega/\mu m$

Table 2.1: Transistor main metrics

where μ_0 represents the low-field mobility and C_{dep} denotes the depletion capacitance [30].

In strong inversion, the drain current in linear region is:

$$I_D = \frac{W}{L} \mu_{eff} C_{ox} (V_{GS} - V_T - \frac{V_{DS}}{2}) V_{DS} \quad (2.4)$$

where μ_{eff} is the effective mobility of the carriers in the channel.

The drain current in saturation region is described by:

$$I_D = \frac{1}{2} \frac{W}{L} \mu_{eff} C_{ox} (V_{GS} - V_T)^2 (1 + \lambda V_{DS}) \quad (2.5)$$

where λ is the channel length modulation coefficient, which incorporates the short channel effects [30].

A key indicator of electrostatic control is the subthreshold swing (SS):

$$SS = \frac{dV_{GS}}{d \log(I_D)}$$

This is expressed according to equation 2.3 as:

$$SS = \frac{kT}{q} \ln(10) \left(1 + \frac{C_{dep} + C_{SS}}{C_{ox}} \right) \quad (2.6)$$

where C_{SS} is the capacitance associated with surface states. Neglecting C_{SS} and C_{dep} , the theoretical subthreshold swing is $SS = 60\text{mV/dec}$, which is a lower limit that cannot be crossed in silicon semiconductor devices. In practice, measured SS is often higher than this theoretical limit due to

interface trap capacitance and short-channel effects such as Drain-Induced-Barrier Lowering (DIBL) which degrade electrostatic control and will be analyzed in later sections.

The current in the subthreshold regime is expressed in terms of the subthreshold swing as:

$$I_D = I_{th} \exp\left(\frac{V_{GS} - V_T}{SS/\ln(10)}\right)$$

This leads to the expression for the leakage current (assuming an OFF state in the subthreshold regime):

$$I_{OFF} = I_{th} \exp\left(-\frac{V_T}{SS/\ln(10)}\right)$$

2.2 Mobility as Stress Monitor

Carrier mobility μ is a fundamental parameter that quantifies how easily charge carriers (electrons or holes) move through a semiconductor under the influence of an electric field E .

The drift velocity is given by $v_{DS} = \mu E$. As discussed in Section 1.4, mobility is highly sensitive to mechanical stress in the transistor channel. Since stress alters the band structure and scattering mechanisms, it directly impacts carrier mobility. Therefore, monitoring mobility provides a reliable means of assessing stress-induced variability in device performance.

Among the various electrical parameters influenced by mobility, the drain current is particularly informative. Both the linear and saturation regimes of MOSFET operation exhibit mobility dependence, as shown in expressions (2.4) and (2.5) respectively.

While both expressions depend on the effective mobility μ_{eff} , the linear region current is the preferred monitor for stress variability for several reasons:

- **Analytical Accessibility:** In the linear regime, the relationship between I_D and μ_{eff} (Equation 2.4) is more direct and linear. This simplifies the extraction of mobility and reduces the influence of second-order effects such as saturation velocity and channel length modulation.
- **Reduced Impact of Parasitics:** At low drain voltages (typically 50 mV), the influence of parasitic resistances (e.g., contact and interconnect resistance) is minimized. In contrast, in the saturation regime,

higher drain voltages amplify the impact of these parasitics, complicating the interpretation of measured current.

- **Avoidance of Velocity Saturation:** In the saturation regime, carriers may reach their saturation velocity v_{sat} , especially under high longitudinal electric fields. This introduces a non-linear dependence on μ_{eff} , making it less sensitive to stress-induced changes in mobility. In contrast, the linear regime operates well below this limit, preserving the proportionality between current and mobility.
- **Measurement Stability:** The linear region current is less sensitive to short-channel effects and variability in threshold voltage extraction. This makes it a more stable and reproducible metric for comparative analysis across devices or process conditions.

To study the impact of the Design of Experiment factors, relative changes in drive current are reported, measured at a fixed gate overdrive voltage ($V_{GS} - V_T$) and a drain voltage of $V_{DS} = 0.05V$. This approach allows to decouple threshold voltage shifts and subthreshold swing variations from stress-induced mobility changes, providing a more accurate assessment of performance sensitivity.

In summary, although both regimes reflect mobility variations, the drain current in the linear region offers a more accurate, stable, and interpretable indicator of stress-induced variability. It is therefore the preferred metric for performance monitoring in the context of this work.

2.3 Conclusions

This chapter has provided a comprehensive overview of MOSFET operation, including its physical structure, operating regimes, and key performance metrics. A central theme was the role of carrier mobility as a sensitive indicator of stress-induced variability. Through analytical modeling, it was demonstrated that while both linear and saturation drain currents depend on mobility, the linear region current offers superior sensitivity, stability, and interpretability. Its reduced susceptibility to parasitic effects and saturation velocity makes it the preferred metric for performance monitoring in advanced CMOS technologies. These findings establish a robust framework for the stress-aware modeling and characterization techniques that follow in this dissertation.

Chapter 3

Characterization Methodology and Test Structures

This chapter presents a structured methodology for characterizing stress-related Local Layout Effects in a commercial 7nm Fin-FET technology. It introduces a high-density test chip integrating over 30,000 devices, and details the Design of Experiment, test infrastructure, and data analysis workflow used to isolate and quantify layout-induced stress variability.

Overview

This chapter presents the experimental methodology developed to characterize stress-related Local Layout Effect in FinFET devices fabricated in a commercial 7nm technology node. The approach is based on a structured Design of Experiment (DOE) implemented through a high-density Characterization Vehicle[®] (CV[®]) test chip, which integrates over 30,000 Device Under Test (DUT) across multiple layout configurations.

The methodology isolates the electrical impact of individual layout features, such as Diffusion Breaks (DB), Gate Cuts (GC), Fin Pitch, and Poly Pitch. Each DUT is embedded in a controlled environment, and this setup ensures that any performance variation measured is due to the specific LLE being studied, reducing interference from other factors.

The chapter details the layout design strategy, the test infrastructure selection, the automated data acquisition and analysis flow. The test structures are grouped into modular blocks, each targeting a specific LLE. Electrical measurements are performed using a high-speed parametric tester, and the resulting data are used to extract key device parameters such as threshold voltage, drive current, and subthreshold swing.

The methodology described in this chapter is validated through silicon measurements and enables the calibration of a predictive 3D TCAD model, as discussed in Chapter 4. The experimental framework and test chip architecture have been presented in the following peer-reviewed publications:

- A. Rossoni et al., “Methodology and test structures to characterize Local Layout Effects in FinFET transistors,” submitted to *IEEE Transactions on Semiconductor Manufacturing*, 2025.
- A. Rossoni et al., “Layout experiments and test structures to characterize Local Layout Effects due to mechanical stress in FinFET transistors,” *IEEE ICMTS*, 24-27 March 2025. [31]

This methodology provides a scalable and reproducible platform for LLE characterization, supporting both Process Design Kit¹ development and Design-Technology Co-Optimization in advanced CMOS nodes.

¹Process Design Kit (PDK): set of models and rules enabling circuit design for a given technology node.

3.1 Introduction to Local Layout Effects

The continuous scaling of semiconductor devices has led to significant advancements in microelectronic engineering. As feature sizes decrease, the interaction of the transistors with their design-dependent layout neighborhood introduces significant and often unpredictable variability. Also, as the distances between neighboring features decrease, the sensitivity of these effects increases [32].

These variations, known as Local Layout Effect (LLE), can have a profound impact on device performance and reliability, and they are difficult to mitigate, even with very restrictive design rules [33].

LLEs are critical in semiconductor devices, driven by mechanisms such as mechanical stress perturbations, pattern-induced loading factors, lithography interactions, and the interplay of dopants or work function metal materials between neighboring transistors. Mechanical stress arises from mismatches in thermal expansion and deposition of heterogeneous materials, impacting the carrier mobility in the semiconductor (see Section 1.4). Pattern-induced loading factors, such as variations in profile and feature height, result from differences in pattern density, affecting surface topology and device performance. Light interferences during lithography illumination are also a source of performance variability, which is not directly related to process steps but is pattern dependent. Additionally, the distribution of dopants and work function metals can vary due to local layout, influencing electrical behavior and threshold voltage.

In this research, we examine the LLEs caused by the modulation of mechanical stress in the transistor channel. Various structures have been developed to characterize the prevalent LLEs in VLSI FinFET technology nodes, and the established workflow is applicable to all types of LLEs.

This chapter aims to establish a comprehensive set of test structures that are sensitive to specific LLEs in 7nm FinFET technology. Each set is designed to isolate the performance variation due to individual layout features. By avoiding the convolution of multiple effects within a single set, the comparison among DUTs is simplified, enabling a swift classification of the impact.

Table 3.1 provides examples of detectable LLEs, highlighting their influence on device performance in advanced FinFET technologies.

Local Layout Effect	Impact
P-N Well Boundary	Dopant cross-diffusion
Isolation proximity	Stress modulation
Diffusion Break type	Stress modulation
Gate Cap extension	Stress modulation
Gate Contact Location	Gate Dielectric Leakage
Active-Active Space	Etch Profiles, Stress modulation
Vt flavor transition	Implant interaction
Number of Fins	LLE screening

Table 3.1: Most known Local Layout Effects impacting transistor variability in advanced FinFET technologies

3.2 Design of Experiment and Layout Strategy

To systematically characterize stress-related LLEs, a structured Design of Experiment (DOE) was implemented using a high-density Characterization Vehicle[®] test chip.

The design of the test structures employed in this study leverages the candidate's expertise in semiconductor device characterization, notably as an inventor of structures specifically designed to isolate LLEs due to stress, which are central to this work. This research was conducted in close collaboration with PDF Solutions, Inc., the industrial partner and the candidate's employer, where the candidate also contributed as a co-inventor to Passive Device Array testing methodologies. The realization of the test chip and subsequent data acquisition utilized an existing PDF Solutions technology characterization project. PDF Solutions' Characterization Vehicle[®] (CV[®]) team provided the essential infrastructure, encompassing the test chip design platform, the measurement platform (pdFasTest[®] tester), test programs, and the final experimental results.

The DOE isolates the electrical impact of individual layout features by embedding Device Under Test in controlled environments. This section describes the layout strategy, the infrastructure selection, and the rationale

behind the modular test architecture.

Figure 3.1 illustrates the workflow utilized in this study for characterizing LLEs. The layout was fabricated in a commercial 7nm FinFET technology node. The structures have been tested, and I-V curves and key device parameters were extracted. Subsequently, the results were analyzed using a dedicated Data Analysis Software.

Each step of the workflow is detailed below.

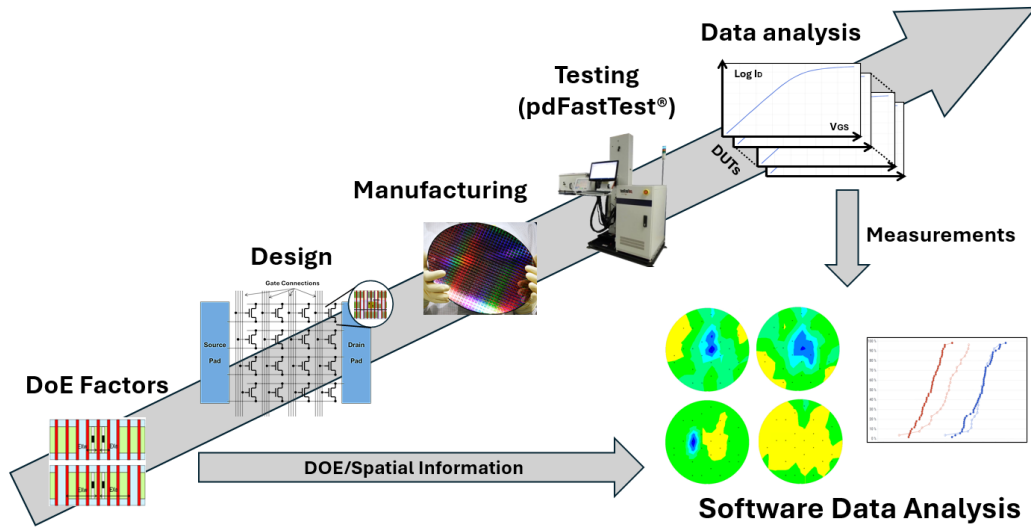


Figure 3.1: Overview of the testing workflow. Designed test structures are fabricated on silicon and evaluated using an automated tester to extract I-V curves and key device parameters. The data are then processed and analyzed using dedicated software templates

Design of Experiment

First, a Design of Experiment is developed to describe each effect. In VLSI testing, the DOE is a methodology used to systematically plan and analyze tests to enhance the performance and reliability of integrated circuits. This method involves creating experiments that vary different factors to pinpoint the conditions that cause defects in VLSI chips, thereby improving the yield and reliability of the manufacturing process.

Furthermore, DOE approach is utilized to assess the performance of VLSI devices and circuits under various conditions. The data gathered from these

experiments are analyzed statistically to determine the impact of different variables on the performance and reliability.

In this work, the developed DOE is partitioned into several sets, each detailing the layout of a specific LLE. Within each set, all factors describing the layout are kept constant except those dedicated to the LLE being characterized.

The DOE was designed to:

- Vary one layout parameter at a time while keeping others constant
- Enable direct comparison between DUTs affected by specific LLEs
- Ensure statistical robustness through replication across device types and threshold voltage flavors

This strategy enables the creation of distinct groups of test structures, each sensitive to a single LLE, simplifying the analysis of its impact on the performance of the DUT.

Layout Design

As part of the Characterization Vehicle[®] (CV[®]) test chip for 7nm FinFET technology development, an extensive set of test structures described by the DOE was designed to explore and characterize several LLEs in standard cell layouts. Figure 3.2 illustrates an example of a transistor layout neighborhood and the layout attributes that need to be considered to characterize LLE sensitivities[34].

Each structure is specifically designed in a controlled environment to shield it from unwanted sources of performance variability, as depicted in Figure 3.3. The scope is to ensure constant density of the main front end-of-line layers (typically all the manufactured layers up to the active and gate contacts), constant distance from the DUTs to the environment, and constant connection resistance to the testing infrastructure. This method simplifies the process of targeting each source of stress [34]. With these measures, the comparison process between the DUTs is facilitated, and the differences measured are directly attributable to the LLE under investigation.

The characterization of each LLE is performed using a dedicated set of structures where only the parameters strictly related to it can vary. Many effects have been characterized in the CV[®], but in this study we focus on the

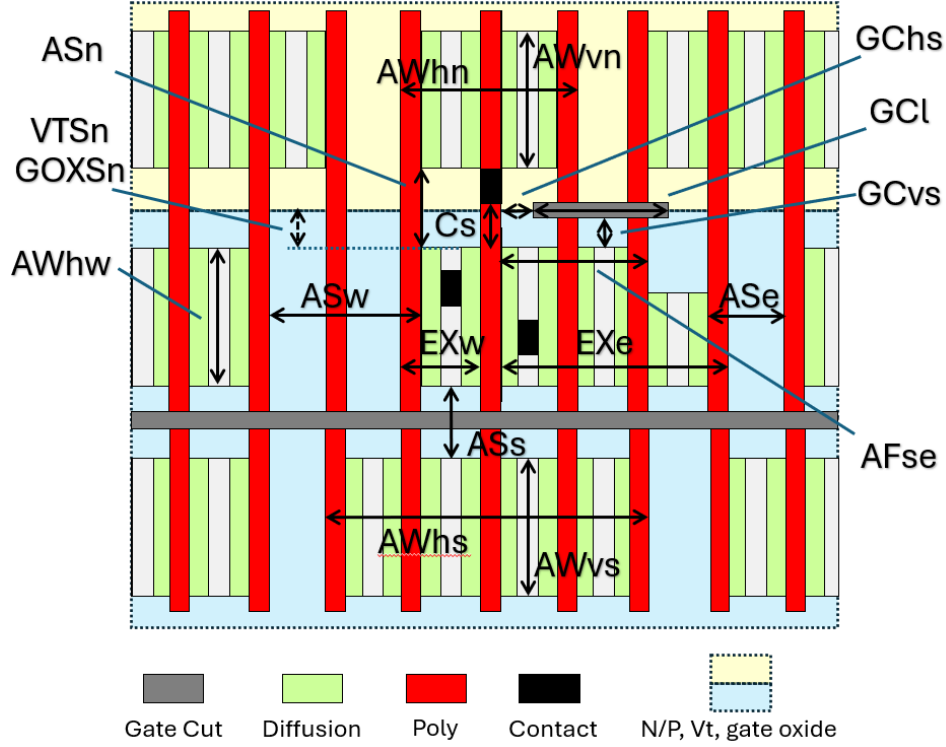


Figure 3.2: Illustration of the transistor layout in a standard cell environment and the attributes defining the LLEs

LLE strictly related to the stress variations. The layout DOE factors and parameter levels shown in Table 3.2 ensure comprehensive coverage with significant resolution and range of the DOE parameters, enabling the characterization of all stress-related LLEs listed in Table 3.1.

All the designed structures have been replicated across transistor types and threshold voltage flavors, and they cover approximately 30% of the entire CV[®].

Testing Design Infrastructure

The result is a set of DUTs designed to characterize LLEs, comprising over 30,000 cells. Given the large number of cells, multiple test infrastructures were evaluated. The analysis considered area efficiency, design complexity, reliability, and measurement accuracy.

In *Single Padded* design, DUT terminals are directly connected to the

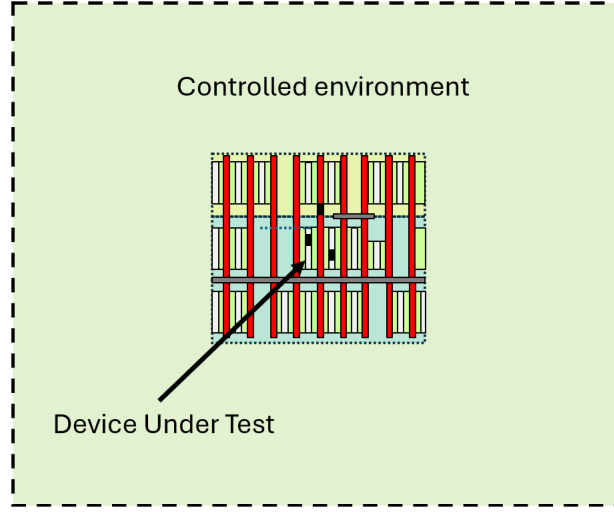


Figure 3.3: The Devices Under Test are surrounded by a controlled environment which allows a direct comparison

LLE Experiment	DoE factors	Parameters	Levels and Value Range in DoE
Diffusion Break (isolations)	Distance between Active Areas ("West" and "East"); Single/Double Diffusion Break	ASw ASe	5 DUTs in a range of 1 $\leftrightarrow$ 20 CPP 5 DUTs in a range of 1 $\leftrightarrow$ 20 CPP
Active Area extension	Active Area extension from DUT to the edge of Active ("West" and "East")	EXw EXe	4 DUTs in a range of 1 $\leftrightarrow$ 10 CPP 4 DUTs in a range of 1 $\leftrightarrow$ 10 CPP
Gate Cut	Distance from DUT's Gate to next Gate Cut Gate Cut length Distance from Active edge to Gate Cut	GChs GCl GXvs	5 DUTs in a range of 0 $\leftrightarrow$ 5 CPP 6 DUTs in a range of 1 $\leftrightarrow$ 8 CPP 7 DUTs in a range of min $\leftrightarrow$ 8 Fin pitches
Neighborhood Diffusion Area	Distance between Active Areas ("North" and "South") Width of Neighbor Active	AS[n,s] AW[n,s]	7 DUTs in a range of 2 $\leftrightarrow$ 8 Fin pitches 2 DUTs in a range of 2 $\leftrightarrow$ 4 Fin pitches

Table 3.2: Layout experiments implemented on the test chip to study and monitor stress-related LLEs. All DoE factors and levels are replicated across device types (N/P) and two Vt flavors

PADs (see Figure 3.4). Due to their straightforward structure, they typically involve simple testing methodologies with high accuracy, which allows parallel testing and ensures high reliability.

This makes them particularly useful during the Technology Development (TD) phase, where yield is still very low. However, protection diodes connected to the gates of the devices are still needed to protect the channels from plasma-induced gate oxide damage (aka Antenna effect [35][36][37]), which prevents testing for gate leakage current. The main limitation of this type of infrastructure is the low area efficiency, which limits the number of

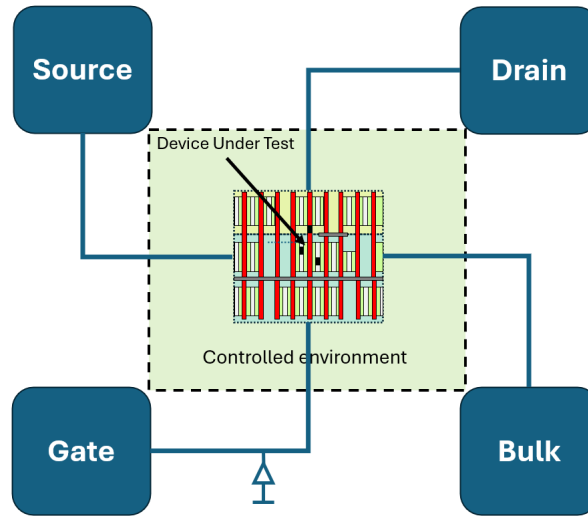


Figure 3.4: Schematic of Single Padded infrastructure, with the transistor terminals directly connected to the PADS, and a protection diode connected to the gate routing

DUTs that are measured on a wafer.

Passive Arrays, which do not contain active components like transistors, rely heavily on the integrity of interconnections (see Figure 3.5). These structures are well-suited for parallel testing, significantly increasing area efficiency using an addressable device matrix, and their high reliability makes them an excellent choice throughout all phases of the TD process. However, the complexity of the interconnections and their intrinsic resistance often necessitates the use of additional PADS dedicated to sensing critical device voltages to increase the accuracy in the measurements.

Active Arrays (Figure 3.6) are characterized by the presence of a selection logic with active components, which significantly reduces yield, especially during the TD phase. Although they offer the highest area efficiency, this is counterbalanced by increased complexity and time during testing. Additionally, the bias voltages applied to the DUTs through the logic typically have a narrower range, effectively preventing DUT bias outside the GND-VDD range. The testing of specific device parameters, such as off currents, is also limited due to the presence of selection logic and due to the parallel device architecture.

The choice of passive arrays over addressable active arrays[38] was strategically motivated by the need to support PDK development throughout the

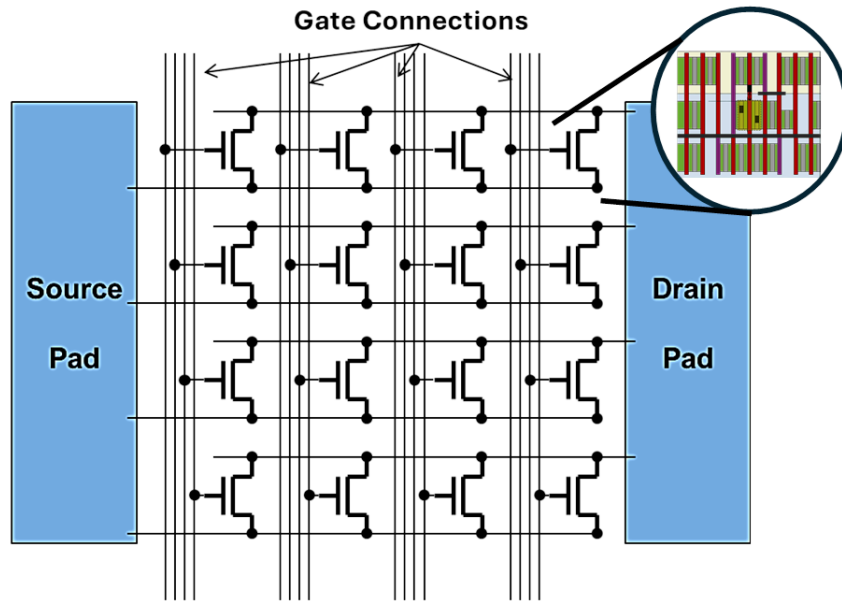


Figure 3.5: Schematic of a typical Passive Device Array infrastructure, with the shared Source and shared Drain with metal connection bus to select the transistor under test

entire TD cycle and the need of high area efficiency. Passive arrays offer superior reliability due to their reduced complexity and fewer potential failure points, critical especially during the stages of TD. Consequently, passive arrays provide a stable and consistent platform for the rigorous testing and validation of PDK, which is essential for achieving high-quality, reliable technological outcomes.

Proposed Upgrade: Passive Matching Array² (future work)

Further studies were dedicated to optimize the test infrastructure. A traditional passive array approach (Figure 3.5) was modified to enhance reliability, area efficiency, and measurement accuracy, albeit with a slight increase in design cost and test complexity.

The standard passive array DUT is divided into two identical devices connected in differential pair configuration, as depicted in Figure 3.7a. The drain routing connections of the two devices to the PADs are kept symmet-

²The concept of the Passive Matching Array (PMA) was developed under the leadership of Dr. Christopher Hess, in collaboration with the PDF Solutions Design Team.

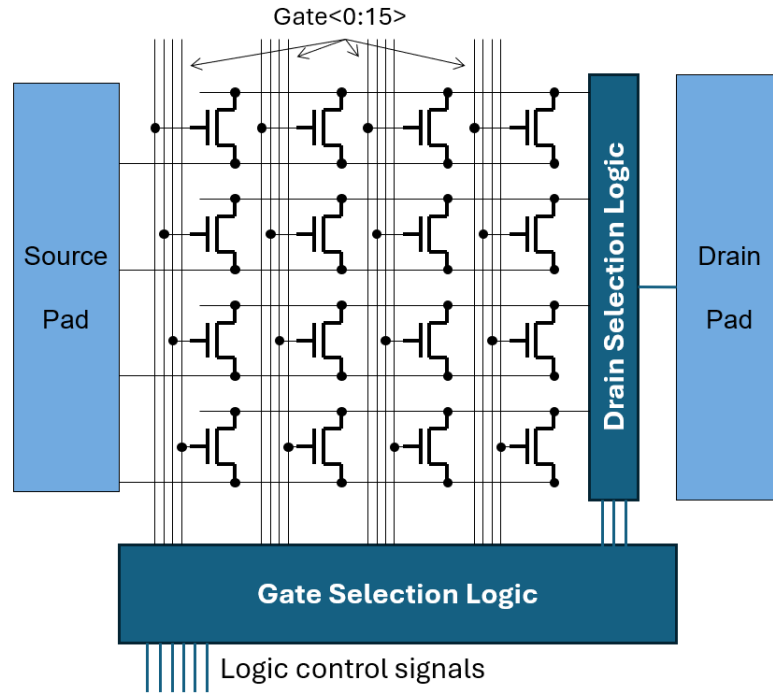


Figure 3.6: Schematic of a typical Active Device Array infrastructure, with the shared Source and selection logic for Gate and Drain which increases the area efficiency

rical to ensure constant resistance. The resulting architecture is shown in Figure 3.8.

Three test modes are enabled, as shown in Figure 3.7b:

- Parallel Test Mode facilitates a rapid test of the differential pair by directly measuring the drain current difference.
- T1 Test Mode measures the first device using a voltage source sense with Kelvin measurement through the second device.
- T2 Test Mode measures the second device, also utilizing a voltage source sense.

These modes are used either to extract the actual routing resistance from the source PAD to the DUT and include it in the testing procedures or to back bias the source PAD to provide $V_s = 0V$ at the DUT source contact, allowing an exact bias of the device. The area efficiency is doubled compared to a standard passive array configuration, utilizing a single gate line to select two

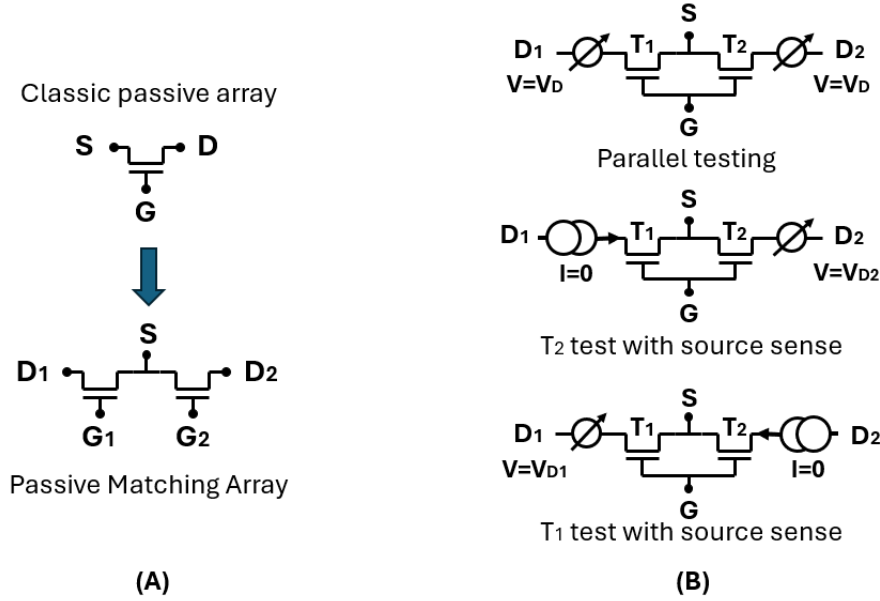


Figure 3.7: Modified schematic of the DUT of the Passive Matching Array (PMA) scheme (A); PMA test modes which allow close mismatch measurement with voltage source sense and no area penalty

devices. The PMA does not require additional PADs for source voltage sensing, as is necessary in standard passive architectures, further enhancing the area efficiency. The close proximity of the two coupled devices in the DUT allows for sensitivity to very close mismatch. The distance between different diff-pairs can vary between $10 - 20\mu m$, enabling sensitivity to medium-range mismatch. Long-range mismatch sensitivity is also ensured by measuring different dies in the wafer.

Figure 3.9 highlights the main factors that influenced the choice of the most suitable test architecture for our application. The decision favored the PMA as reliability was considered the dominant factor.

Testing Equipment

The wafers integrating the LLE test structures were tested using a highly parallel high speed parametric tester³ optimized for the CV[®]. The PAD layout (PAD footprint), the PAD signal assignment (PAD mapping), and the

³PDF Solutions, pdFasTest[®] Tester[39]

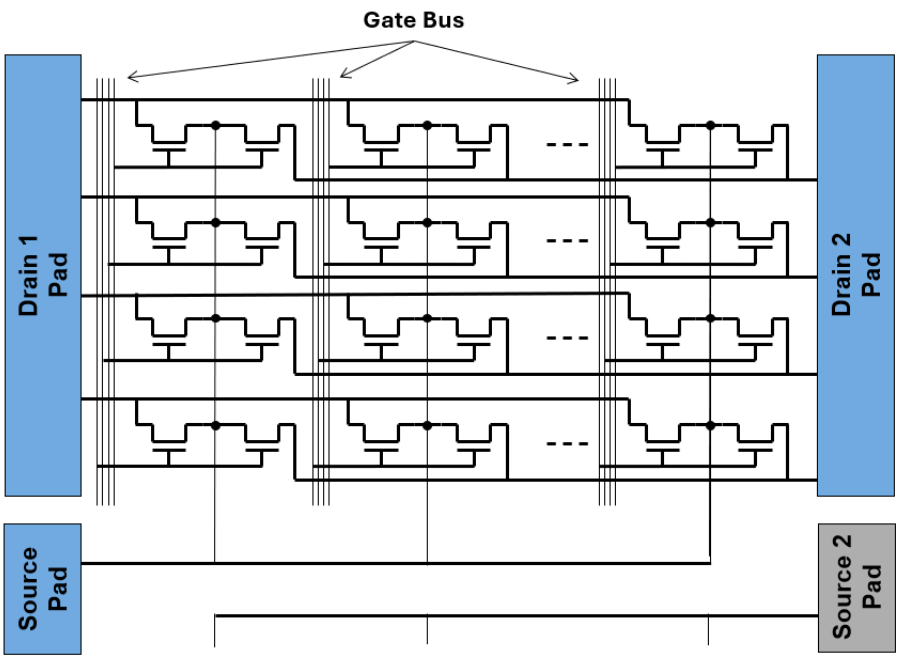


Figure 3.8: Passive Matching Array (PMA) for high density arrangement of DUTs in LLE DoE. Each DUT consists in a differential pair of two identical devices



Figure 3.9: Comparison of Area Efficiency, Design Complexity, Reliability, and Test Complexity for different test infrastructure architectures: Single Padded, Passive Arrays, Passive Matching Arrays, and Active Arrays

global structure floorplan within dies are all optimized to expedite testing procedures and enhance the throughput of the entire system, thereby achiev-

ing rapid feedback from silicon data, which is essential especially during the Technology Development phase.

The I_D - V_{GS} curves were measured for both linear and saturation regimes, and key device parameters such as Threshold Voltage (V_T), Drive Current (I_d), and Subthreshold Swing (SS) were extracted for each transistor. Median values are used to represent the electrical measurements of device parameters.

This approach minimizes sensitivity to local and wafer spatial variations and generates a reliable and consistent data. The calculated standard deviation of measured parameters for the experiments under investigation in this study is less than 1.3% for the drain current in both linear and saturation regions, and less than 1.6% for the threshold voltage and the subthreshold swing.

This data set played a crucial role in both calibrating and validating the 3D TCAD model.

Data Collection and Software Data Analysis

The so obtained measurements, statistical data, layout characteristics (DoE), and spatial information (X-Y position in the wafer of each DUT) were automatically ingested by the data analysis software⁴, which is able to perform automated analysis, producing Cumulative Distribution Functions (CDF), Wafer Maps (WM), and reports of the key device parameters of each DUT. Design generation, testing, and analysis report generation are integrated in a Characterization Vehicle System[®] ⁵.

3.3 Characterization Modules

To facilitate targeted analysis of individual LLE, the test structures were partitioned into dedicated modules. Each module varies only the layout parameter associated with a specific LLE, while all other design attributes

⁴PDF Solutions, Exensio[®] Analytic Platform[39]

⁵Characterization Vehicle System[®]. The CV System is the combination of high-density CV test chip layouts targeted to a proprietary pdFasTest[®] parallel parametric tester[39], producing data with automatic ingestion into the Exensio[®] Analytics Platform. The entire stack is optimized to generate clean data labeled with all key design attributes for efficient summary and drilldown analysis.

remain constant. This approach enables high-resolution characterization of stress-induced performance variability.

There are some cases where layout feature variations between DUTs lead to multiple LLEs simultaneously. An example is the study of the Contacted Poly Pitch, which intrinsically induces a variation in the distance between the measured device and the Diffusion Break. To deconvolve the effect of the two LLEs, a very high distance with the DB is imposed for all DUTs in the module where the CPP is varied, making the measurements insensitive to the DB proximity.

This approach ensures that the signal measured from each structure is independent of all other LLEs considered.

The partitioning of the modules allows for simpler and more accurate analysis of the measurements, as it eliminates the need to decompose various effects from one another.

Here's an example of modular partitioning.

3.3.1 Active Extension Module

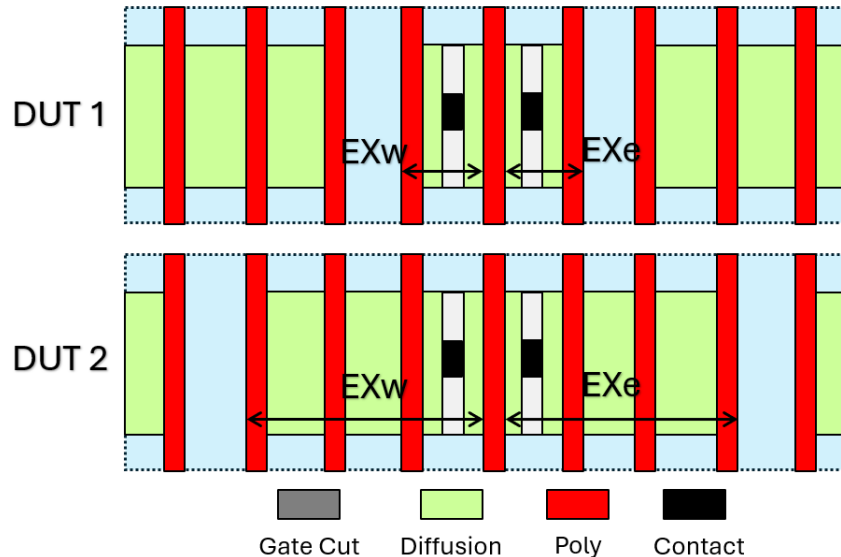


Figure 3.10: Test structure to evaluate the Diffusion Break proximity impact

One of the strongest LLEs affecting device performance is the proximity of a DB. In recent FinFET technologies, as well as the more recent Gate-All-Around Field Effect Transistor technologies, strain is introduced in the

source and drain regions to enhance carrier mobility and improve device performance. However, the presence of fin isolations disrupts the periodicity of the engineered stress and introduces perturbations.

The engineered stress is designed to be uniform across the device to maximize performance benefits. However, fin isolations, which are necessary for device isolation, break this uniformity and create localized stress variations, which can significantly impact the electrical characteristics of the device, leading to performance changes.

To study and quantify this performance variability, the Active Extension Module is employed (see Figure 3.10). This test structure is designed to capture the effects of the DB on device performance. By placing devices in a variable proximity to the DB, the module can measure the impact on key device parameters.

3.3.2 Active Proximity Module

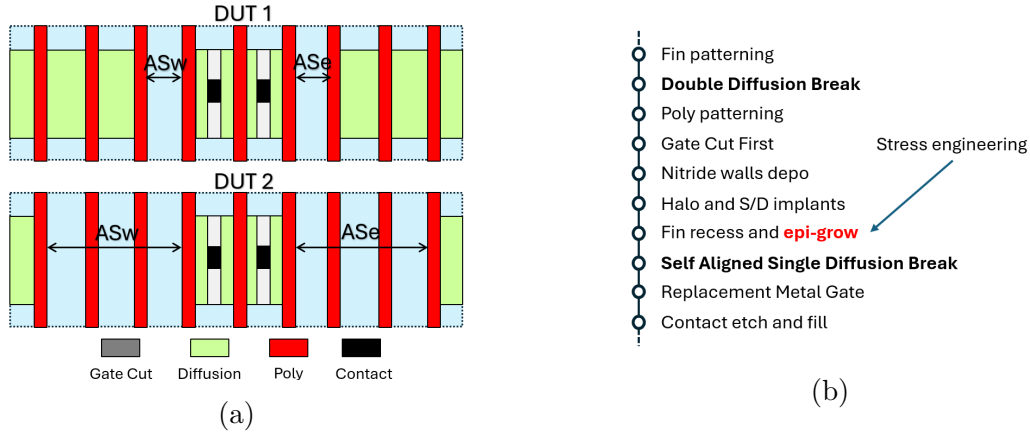


Figure 3.11: (A) Test structure designed to assess the impact of horizontal size variation in DDB; (B) Process flow schematic illustrating the module sequence: the DDB is preformed prior to stress engineering during epitaxial growth, while the SDB is implemented afterward

The Active Extension Module was used in combination with the Active Proximity Module (see Figure 3.11a), which characterizes the type of DB surrounding the DUT. Single Diffusion Break (SDB) and Double Diffusion Break (DDB), or larger, are manufactured at different stages in the process, as shown in Figure 3.11b.

SDB are typically introduced late in the manufacturing process, after the stress engineering of the source and drain regions. This timing makes the device more sensitive to the stress perturbation introduced by the DB because it is performed on a structure where the stress is already calibrated. The introduction of SDBs after the stress engineering phase means that any perturbations caused by the DB can significantly impact the already established stress profile. In p-type devices, the high compressive stress from the Silicon-Germanium epitaxial layer is crucial for enhancing hole mobility and overall device performance. However, the presence of a diffusion break can disrupt this stress, leading to variations in device characteristics.

In contrast, DDB or larger diffusion breaks are introduced earlier in the process, just after fin patterning. This earlier introduction can lead to different stress profiles and, consequently, different impacts on device performance.

Due to these differences in the manufacturing stages, the device performance is affected differently by SDB and DDB. The Active Proximity Module, in conjunction with the Active Extension Module, allows for a detailed characterization of these effects.

3.3.3 Gate Cut Proximity Module

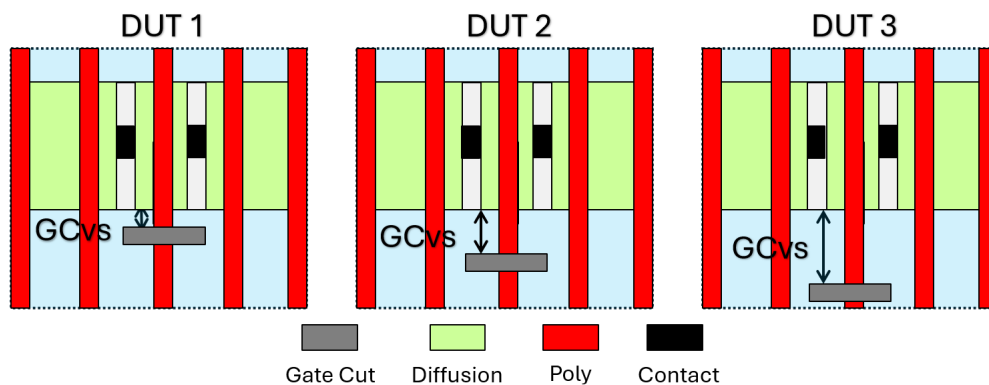


Figure 3.12: Test structure to evaluate the Gate Cut proximity impact

In advanced semiconductor manufacturing, the gate terminal of a transistor is isolated through a series of specific process steps. This involves using a dedicated GC mask to create a cut in the gate line, which is subsequently filled with an isolation dielectric material (more details about Gate Cut process steps will be described in Section 7.1). This process is crucial for ensuring

that the gate terminal is properly insulated from other components, thereby preventing electrical interference and ensuring reliable device operation.

However, the proximity of this isolation cut to the transistor's channel region can introduce stress perturbations. These perturbations arise because the isolation dielectric material, the gate spacers and the surrounding semiconductor material have different mechanical properties. When the dielectric is inserted, it can induce mechanical stress in the nearby semiconductor regions.

Also, the presence of the GC modifies the complexity of the gate structure (oxide spacers, silicon nitride spacers, work-metal functions, etc.) in the close proximity of the channel. The stress induced by the epitaxial growth interacts with these structures differently, resulting in variations in transistor performance.

To address this issue, the Gate Cut Proximity Module was developed (see Figure 3.12). This module is designed to capture the LLE associated with the (GC) without the risk of convolving other effects. By isolating the impact of GC, this test structure allows for a more accurate assessment of how the gate cut isolation affects device performance.

3.3.4 V_T /NP Boundary Proximity Module

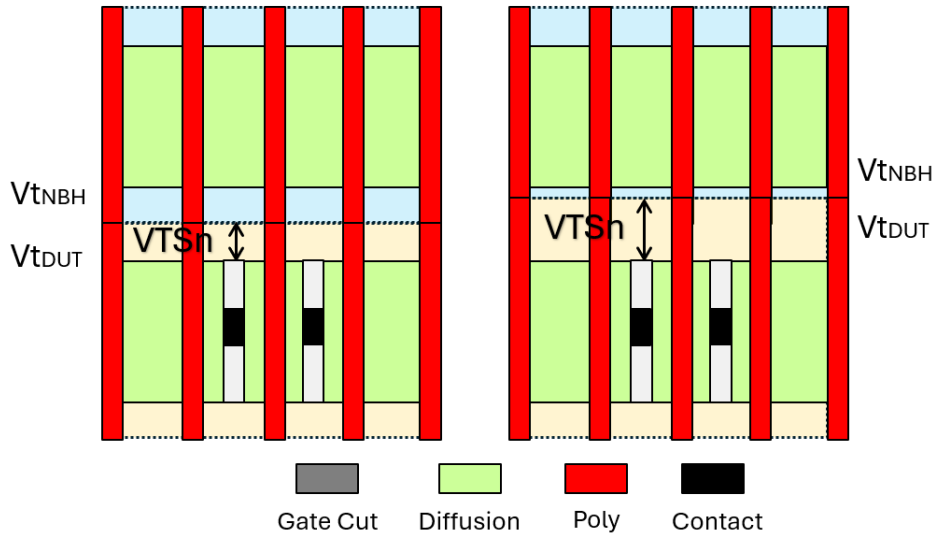


Figure 3.13: Test structure to evaluate the V_t transition or the P to N transition proximity impact

In recent technologies, the channel of the transistors is composed of undoped silicon. This approach drastically reduces the doping interactions between n-type and p-type regions and different threshold voltages. The definition of these parameters is primarily delegated to process steps that define the gate metal stack.

Despite this, a strong interaction remains between devices with different materials and/or thicknesses in the layers that compose the gate metal stack. This interaction is due to the modulation of stress and potential cross-diffusion of metallic elements within the gate, which places this LLE in a category that is not strictly stress-related.

These interactions create the need for a dedicated module to characterize these effects. The V_T /NP Boundary Proximity Module is designed to fulfill this role by modulating the relative position of the transition of the n-type/p-type regions and the V_T transitions with respect to the device under test (DUT).

3.3.5 Fin Pitch Module

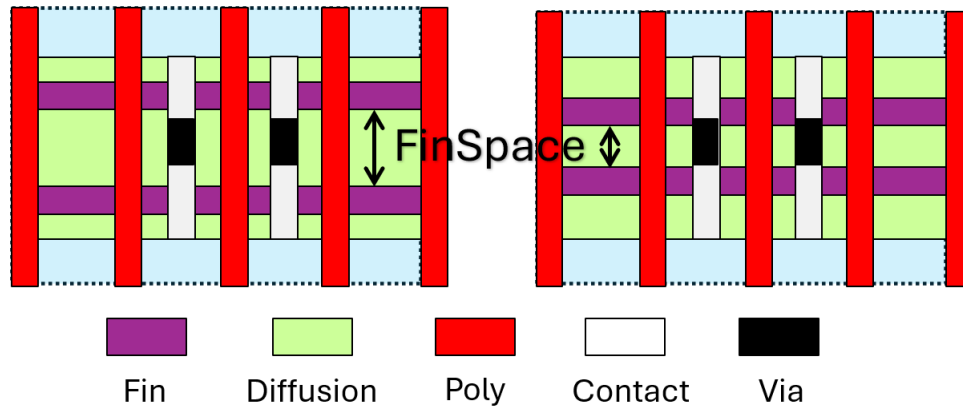


Figure 3.14: Test structure to evaluate fin-to-fin space variation impact

FinFET technology emerged in the late 1990s to address the limitations of planar transistors. Initially, challenges such as fin patterning and uniformity were prominent, but fin pitch walking was not a significant concern due to the larger feature sizes. As technology nodes have scaled down, the demand for more precise patterning techniques has become evident. Advanced lithog-

raphy methods, such as 193nm immersion lithography, have enhanced fin patterning but also exposed limitations in resolution and overlay accuracy.

Scaling involves reducing the dimensions of various components, as well as the spacing between them. This process raises critical questions about the behavior of stress within the devices. Engineers must consider whether stress will scale proportionally and how this will impact device performance, particularly in relation to device patterning.

For FinFET transistors the scaling also involves the change in the fin size and in the fin pitch (fin width + inter-fin space). Variations in Critical Dimension during patterning can modulate the fins, leading to two primary issues: changes in fin width and fin spacing. With currently dominating patterning method based on Spacer-Assisted Patterning, fin width has much lower variability than the space between fins. Since the space between the fins determines the volume of epitaxially grown Source/Drain, these variations can have an impact on transistor performance.

As semiconductor technology progresses into the nanometer regime, maintaining uniform fin pitch in FinFET devices becomes increasingly challenging. Fin pitch walking can result from variations in pattern density, etching processes, and sidewall angle inconsistencies during even more advanced patterning techniques like SAQP⁶ [40].

Silicon epitaxy is a complex process in which an additional layer of monocrystalline silicon is deposited onto the polished surface of a silicon wafer. Unlike simple epitaxy, silicon epitaxy involves multiple simultaneous chemical reactions. During epitaxial layer growth, chemical substances are deposited on top of an existing crystal layer, resulting in layered growth, also known as "stacking". This process allows for the fabrication of silicon-based devices with layers formed in two different directions.

The epitaxial growth performed on the source and drain regions of a device is an essential process step that defines the performance of the device, because it generates the main stress components affecting the channel of the transistor, especially in p-type devices where Silicon-Germanium is used to increase the stress.

The final shape of the epitaxial layer is influenced by several process parameters, as well as the geometry of the fins from which the growth originates. Parameters such as height, width, volume, and the merging between

⁶Spacer-Assisted Quadruple Patterning (SAQP): advanced lithography technique for sub-10nm pitch definition.

epi-grow shapes vary depending on the fin width, height and spacing. These variations can lead to differences in the stress profile and, consequently, the electrical characteristics of the device.

While fin pitch is less critical in mature technologies due to the precise calibration of lithography and etching tools in the production chain, it remains a significant source of variability for the epitaxial growth stressors in source and drain regions during technology development stage.

This behavior is characteristic of merged-fin technology, where the epitaxial growths of adjacent devices merge due to the high fin height-to-pitch ratio. In contrast, in non-merged-fin technology, the stressor volume remains unchanged, and the stress-related impact of fin pitch walking on performance tends to be negligible.

Understanding the impact of fin pitch on stress modulation is crucial for optimizing device performance and achieving the desired electrical characteristics of transistors in advanced VLSI technology.

To address the performance variability caused by epitaxial growth shape variability due to fin pitch, the Fin Pitch Module was developed. The realization of this particular structure is typically not feasible with a standard PDK. It was necessary to use reserved design layers, specially enabled by the foundry, to modify the relative position of the fins, emulating the fin pitch variability.

This module characterizes the performance variability by analyzing the impact of different epitaxial growth shapes on device parameters. By understanding how the epitaxial growth shape affects device performance, it is possible to optimize the manufacturing process and achieve more consistent and reliable device behavior.

3.3.6 Poly Pitch Module

In the same way that fin spacing affects epitaxial growth, the gate-to-gate space also plays a crucial role. The volume available for silicon to grow is modulated by the poly space, which changes the source and drain volume and, consequently, modifies the stress generated by the stressor.

To address this LLE, the Poly Pitch Walking Module was designed (see Figure 3.15).

Similarly the previous module, achieving this specific structure is generally impractical with a conventional PDK. It required the utilization of

reserved design layers, which were specially activated by the foundry, to adjust the relative positioning of the gates keeping the same gate length, thereby simulating the poly pitch walking phenomenon.

This module is specifically aimed at characterizing the performance variability induced by variations in the gate-to-gate spacing.

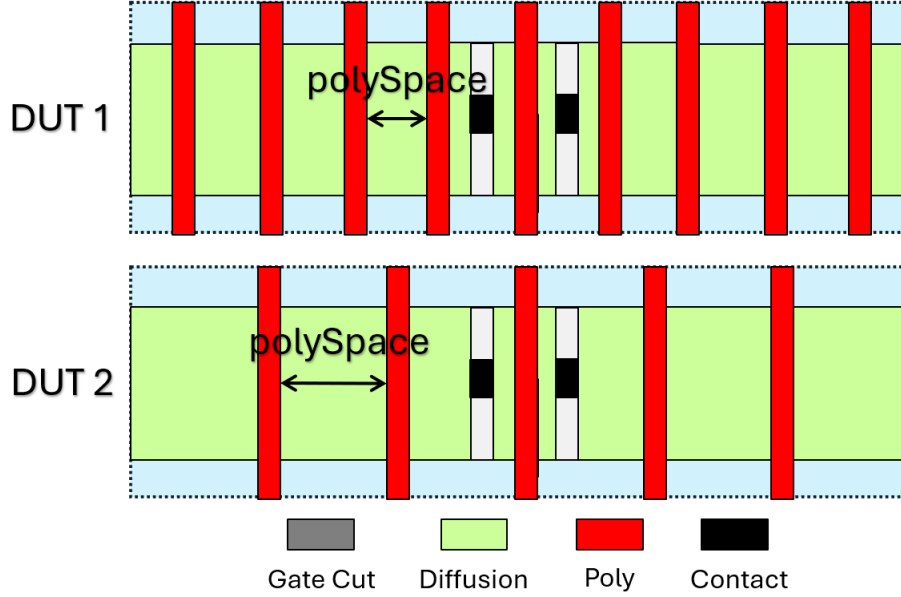


Figure 3.15: Test structure to evaluate the impact of gate-to-gate spacing on epitaxial growth and device performance

3.4 Device Testing and Data Acquisition

Each module was replicated across multiple device types and threshold voltage flavors to ensure comprehensive coverage. The modular design simplifies analysis and supports direct correlation between layout features and electrical performance metrics such as threshold voltage, drive current, and sub-threshold swing.

Electrical characterization was performed using a highly parallel high speed parametric tester (PDF Solutions' pdFasTest®), optimized for rapid data acquisition and automated analysis. The entire stack is optimized to generate clean data labeled with all key design attributes for efficient summary and drilldown analysis.

I-V curves were measured for each DUT at room temperature (25°C)

for linear and saturation regimes and key device parameters were extracted. The software platform offers comprehensive wafer coverage and automated analysis, supporting all DOE factors, including details about packing and DUT location on the wafers. These factors are directly ingested from the design engine into the analysis template configuration, ensuring seamless integration and efficient data processing.

3.5 Results and Analysis

3.5.1 Die Level Analysis

The impact of the LLE is evident in the raw I-V curves. Figure 3.16 illustrates how a SDB near the DUT can affect the drive current and the threshold voltage compared to a Reference DUT, which has continuous active and gate regions across several neighboring devices.

P-type transistors generally exhibit greater sensitivity to stress modulation compared to their n-type counterparts. This is due to the use of Silicon-Germanium epitaxial source/drain regions in p-type, which induces high compressive stress in the device channel.

The results and the detailed analysis for several LLEs are described in Chapter 6.

Examples of key findings are showed in Figures 3.18 and 3.19, which illustrate the degradation of the drive current performance in p-type devices due to the proximity of SDB or DDB and GC. The results indicate a clear dependence of performance degradation on the distance to these layout features, with drive loss reaching up to 10% compared to reference DUTs with continuous active and gate regions.

3.5.2 Wafer Level LLE Analysis

Device array design combined with high-speed testing enable collection of large statistics of data from each wafer, which is critical for proper evaluation of sources of variability, especially at early stages of process development. Each DUT is replicated eight times, and each die is placed 65 times across the wafer.

Figure 3.20 illustrates the CDF plots for the drive currents of n-type and p-type devices fabricated on the same wafer. Cumulative plots are essential

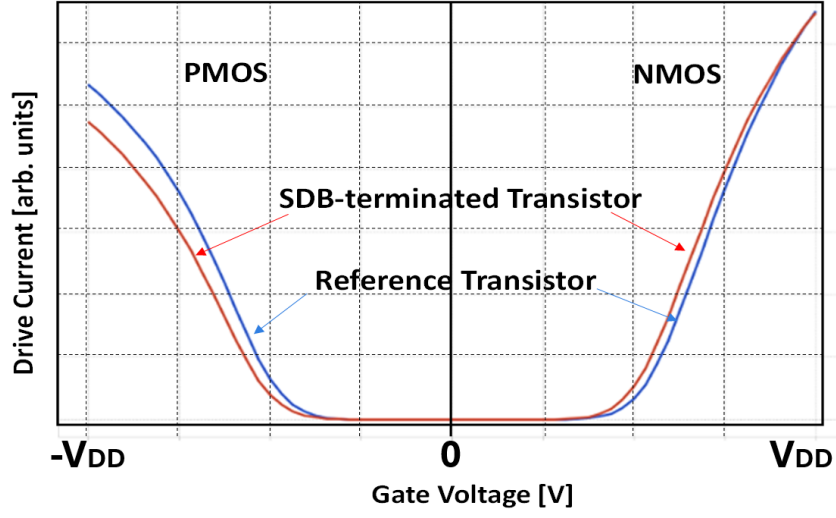


Figure 3.16: Raw I_d - V_{gs} curves measured on Reference and SDB-terminated devices from LLE DoE, showing the impact of the isolation edge proximity. P-type devices show significant degradation

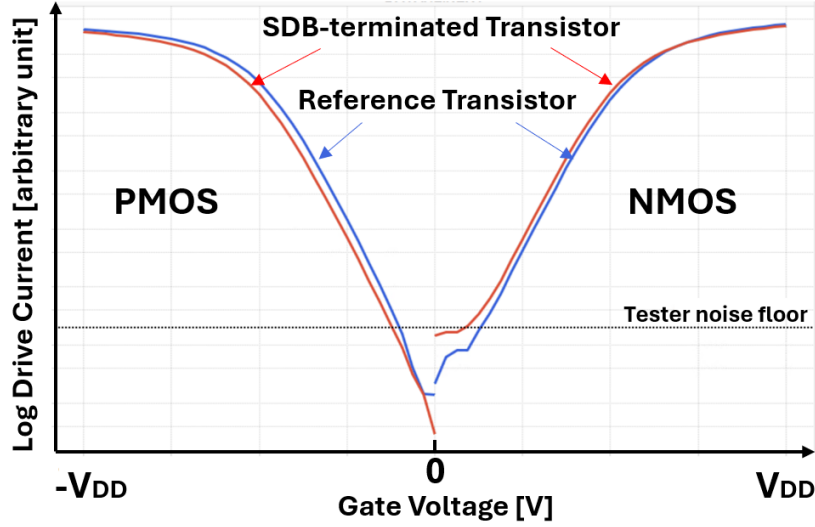


Figure 3.17: Raw $\text{Log} I_d$ - V_{gs} curves measured on Reference and SDB-terminated devices from LLE DoE, showing the impact of the isolation edge proximity. P-type devices show significant degradation

for assessing the uniformity of the effects across the wafer - parallel shift indicates good wafer process uniformity. The observed variations in the distribution shapes are attributed to process non-uniformity and differential

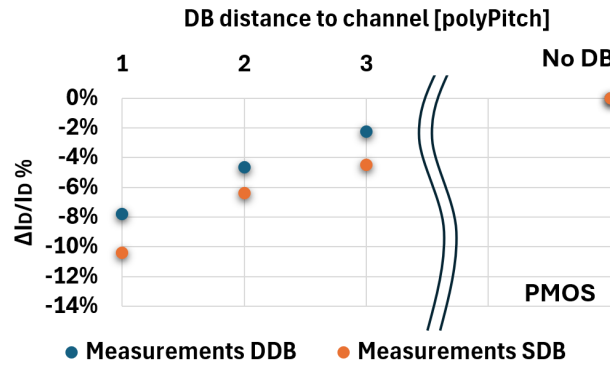


Figure 3.18: Measured results of the Drain current degradation in p-type transistors neighboring a SDB or DDB; stress relaxation due to discontinuity of the active region results in lower performance of p-type FinFETs

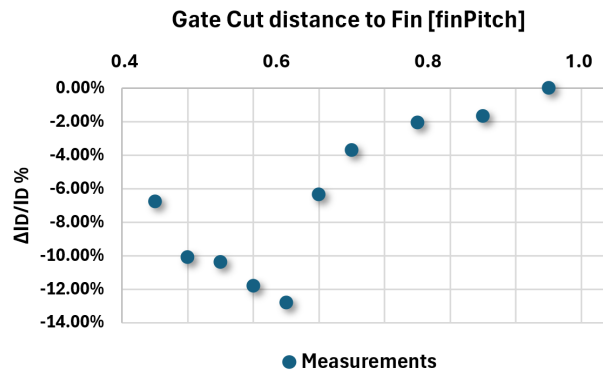


Figure 3.19: Gate Cut distance dependence of the degradation of the p-type device drain current measured on the test structures from the DOE DUTs. Gate Cut presence causes a decrease of the device performance. The results will be discussed in detail in Chapter 6

stress changes across various regions of the wafer.

This phenomenon is further supported by the wafer maps of drive currents presented in Figure 3.21, which highlight the spatial disparities in device performance. Development engineers are concerned not only with the magnitude of LLEs at the median level but also with the additional variability introduced by process inconsistencies across the wafer. The edge regions, in particular, can significantly impact overall wafer uniformity, making thorough testing and analysis indispensable. By examining the full wafer and focusing on edge effects, engineers can better anticipate and mitigate the im-

part of process-induced stress variations, leading to more robust and reliable devices.

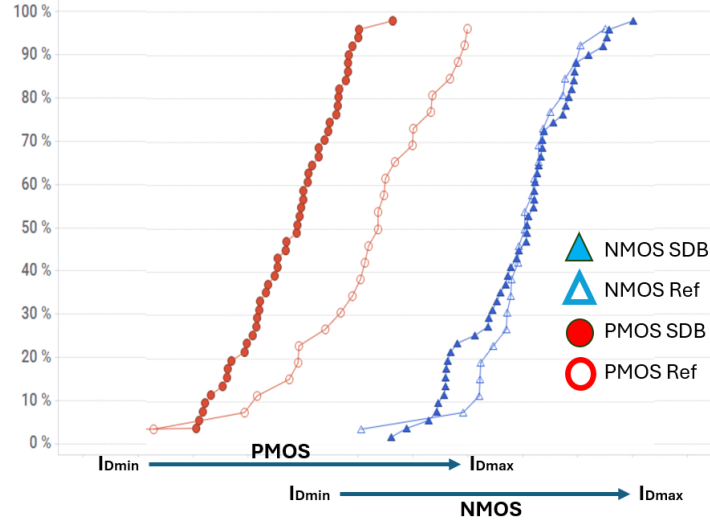


Figure 3.20: Cumulative plots of measured drive currents for Reference and SDB transistors. There is no impact in n-type devices while p-type devices show 10% degradation

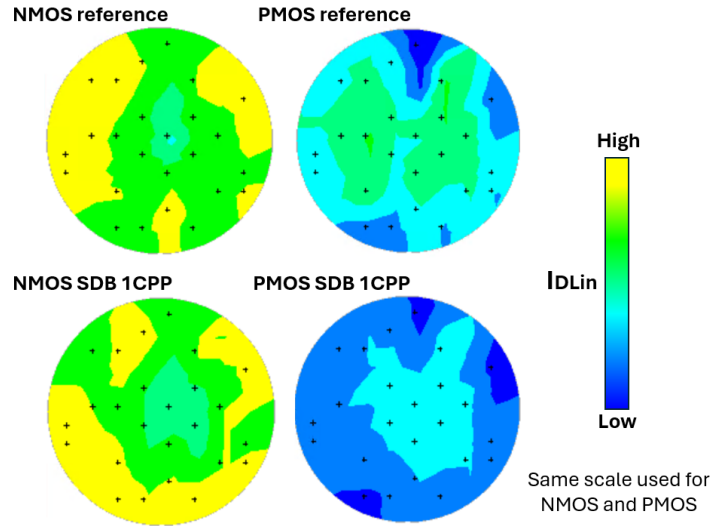


Figure 3.21: Wafer maps of I_{dLIN} for Reference and SDB terminated transistors - n-type devices show similar spatial patterns while p-type devices show overall degradation with some deviations in wafer center

3.6 Conclusions

This chapter introduced a rigorous and scalable methodology for the characterization of Local Layout Effects (LLEs) in advanced 7nm FinFET technologies. Recognizing the critical role of LLEs in influencing device variability and performance, a comprehensive Design of Experiment (DOE) was developed to isolate and quantify the impact of individual layout-dependent stress mechanisms. The methodology was implemented through a suite of custom-designed test structures embedded in a high-density Characterization Vehicle[®], enabling precise and reproducible measurements.

The modular design of the test structures, each targeting a specific LLE, ensures that the extracted electrical parameters are directly attributable to the effect under investigation, thereby eliminating confounding variables and simplifying data interpretation.

Automated data collection and analysis tools streamlined the workflow, allowing fast extraction of key metrics like threshold voltage, drive current, and subthreshold swing. The resulting dataset, characterized by low variability and high statistical confidence, serves as a foundation for the calibration and validation of predictive TCAD models.

Overall, the methodology developed in this chapter provides a powerful framework for understanding and mitigating LLE-induced variability in nanoscale CMOS technologies. It not only supports the refinement of process design kits (PDKs) but also informs layout optimization strategies, ultimately contributing to improved device performance, yield, and manufacturability in future technology nodes.

Chapter 4

Modeling Framework and Assumptions

This chapter briefly describes the FinFET architecture and a schematized manufacturing process. The modeling approach used in this work is also presented. The details of the VLSI FinFET technology node used to implement the testing structures analyzed in this study are not fully disclosed. Therefore, this chapter also describes the assumptions made to develop the process model.

Overview

This chapter introduces the modeling methodology adopted to simulate stress-related Local Layout Effect in FinFET devices. The approach is based on a calibrated 3D Technology Computer-Aided Design (TCAD) framework, developed to replicate the physical and electrical behavior of devices fabricated in a commercial 7nm technology node.

The chapter starts with a description of the FinFET architecture and the rationale behind the modeling assumptions, including device geometries, material stacks, and process integration steps. Due to the proprietary nature of the foundry process, the model relies on informed assumptions derived from literature and predictive PDKs. These assumptions are validated through comparison with silicon measurements.

The modeling flow integrates process and device simulation stages. The process simulation captures the mechanical stress distribution induced by key fabrication steps, such as epitaxial growth¹, gate cut formation, and metal stack deposition, while the device simulation evaluates the electrical response under stress conditions. The model integrates several advanced physical models to capture stress-induced performance variations:

- **Carrier mobility:** including doping, phonon, and surface scattering.
- **Recombination mechanisms:** Shockley-Read-Hall (SRH), Auger, and band-to-band tunneling.
- **Piezoresistive effects:** modeled using deformation potential theory².

The simulated structures include layout features such as Single and Double Diffusion Breaks (SDB/DDB) and Gate Cuts (GC), which are known to modulate stress in the channel region. The model was calibrated using reference structures with continuous active and gate regions. It was then validated against experimental data.

The calibrated model described in this chapter forms the foundation for the predictive simulations presented in Chapters 6 and 7. The modeling

¹A process where a crystalline layer is grown on a substrate, often used to form source/drain regions with specific stress properties.

²A model that describes how mechanical strain affects the energy levels (band structure) in a semiconductor.

framework and its validation have been published in the following peer-reviewed works:

- A. Rossoni et al., “Stress-Related Local Layout Effects in FinFET Technology and Device Design Sensitivity,” *IEEE Trans. on Semiconductor Manufacturing*, vol. 38, no. 2, 2025. [41]
- A. Rossoni et al., “Layout experiments and test structures to characterize Local Layout Effects due to mechanical stress in FinFET transistors,” *IEEE ICMTS*, 24-27 March 2025. [31]
- A. Rossoni et al., “Characterization and Modeling of FinFET Performance Variability from Layout-Induced and Process-Induced Stress Modulation,” submitted to *IEEE Trans. on Semiconductor Manufacturing*, 2025.

This modeling strategy enables a robust and scalable approach to analyze layout-dependent variability in advanced CMOS nodes, supporting both design and process optimization.

4.1 FinFET Architecture and Process Flow

The concept of a double-gate thin-film transistor was first proposed by H. R. Farrah (Bendix Corporation) and R. F. Steinberg in 1967. The term *FinFET* (fin field-effect transistor) was later introduced in a seminal paper by Hisamoto *et al.* in December 2000 to describe a non-planar, double-gate transistor fabricated on a Silicon on Insulator (SOI) substrate [23].

Commercial adoption of FinFET technology began at the 22 nm node, with Intel introducing its tri-gate FinFET variant in 2011 [42] as part of the Ivy Bridge microarchitecture, as shown in Figure 4.1 [42]. These devices began shipping in 2012. From 2014 onward, major foundries such as TSMC, Samsung, and GlobalFoundries adopted FinFET designs at the 14 nm (or 16 nm) technology node, marking a significant industry-wide transition from planar to 3D transistor architectures [43].

A Fin Field Effect Transistor (FinFET) is a modern type of transistor with a three-dimensional structure that offers superior performance compared to traditional planar Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET). It features a thin silicon fin that rises above the substrate,

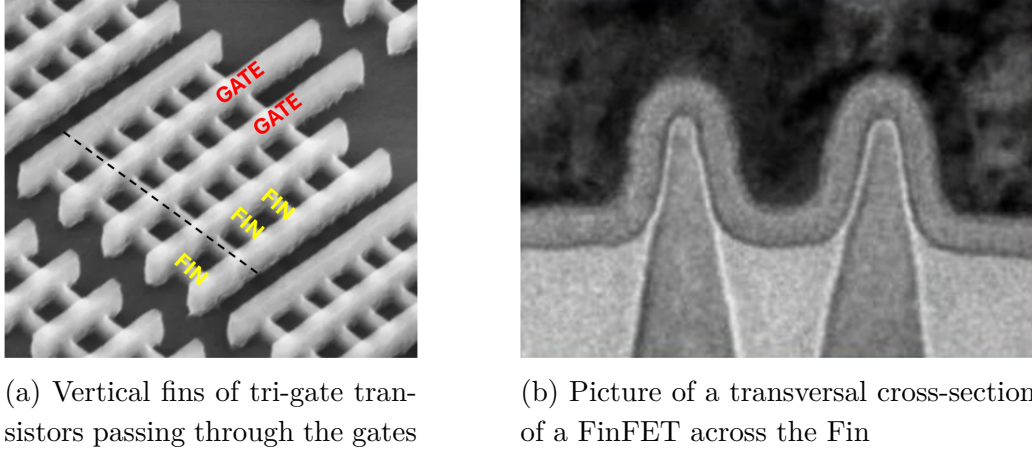


Figure 4.1: Intel™ 22nm FinFET technology node [42]

allowing the gate to control the channel from three sides instead of just one, and its name derives from the fact that the source/drain regions form fins on the silicon surface. The improved electrostatic control reduces short-channel effects and leakage currents [44] makes FinFETs highly efficient compared to planar devices.

Because of their higher drive current and lower power consumption, FinFETs are widely used in advanced semiconductor fabrication, particularly in 14nm, 10nm, 7nm, and 5nm process nodes. They are commonly found in CPUs, GPUs, and mobile processors, providing better speed, efficiency, and scalability for modern electronics.

The fabrication of FinFET consists of several steps, schematized in Figure 4.2a. Typical geometric parameters of a 7nm FinFET technology are summarized in Table 4.2b.

A typical FinFET layout is illustrated in Figure 4.3(a), with its simulated longitudinal cross-section shown in Figure 4.3(b), and the transversal cross-section in Figure 4.4. The Device Under Test (DUT) is centrally located (dashed area) and surrounded by various isolation regions.

The fin isolation on the left is known as Double Diffusion Break (DDB), occupying two Contacted Poly Pitches (CPP). This isolation is implemented early in the process, immediately following the fin patterning, as depicted in Figure 4.2a.

Conversely, the isolation on the right is called Single Diffusion Break (SDB), requiring only a single CPP. In order to increase device density, SDB

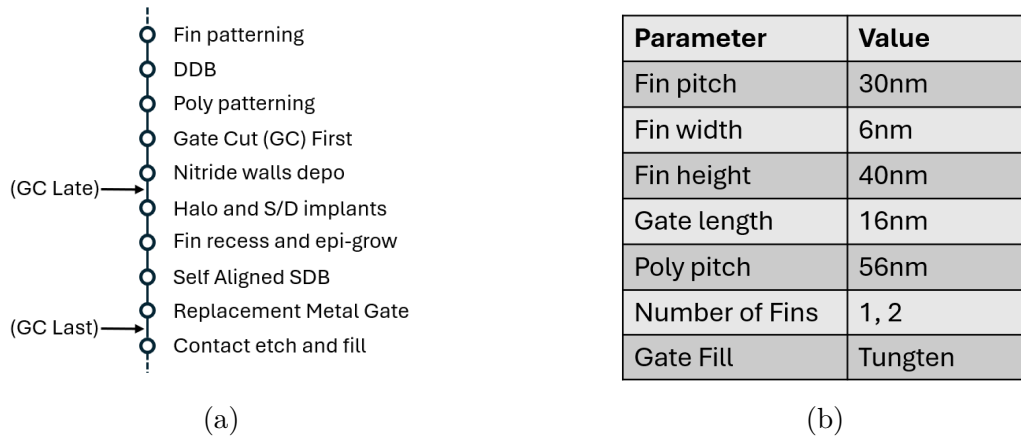


Figure 4.2: Simplified process steps for a 7nm FinFET technology node (a) and the parameters used in the process simulation (b)

was developed to conserve area compared to DDB, achieving an area saving of approximately 30% statistically. The fabrication of SDB is processed following the stress engineering applied to the source and drain regions (as shown in Figure 4.2a). This sequence induces more significant perturbations to the overall stress equilibrium compared to DDB, which is implemented prior to the polysilicon patterning. Consequently, it is expected that SDB induces more transistor performance variability relative to its DDB counterpart.

The simulated cross-section of the GC that isolates a device gate from the adjacent gate is shown in Figure 4.4.

4.2 Simulation Workflow

The workflow developed in this study is illustrated in Figure 4.5, and starts with the definition of the device layout, which includes nearby isolation structures, as specified by the Design of Experiment (DOE).

The DOE refers to a systematic methodology used to plan, design, and analyze experiments aimed at evaluating the performance, reliability, and manufacturability of various components and process parameters within an integrated circuit. In this work, the DOE approach is applied to define a set of test structures specifically designed to be sensitive to Local Layout Effects, enabling targeted analysis of layout-dependent effects.

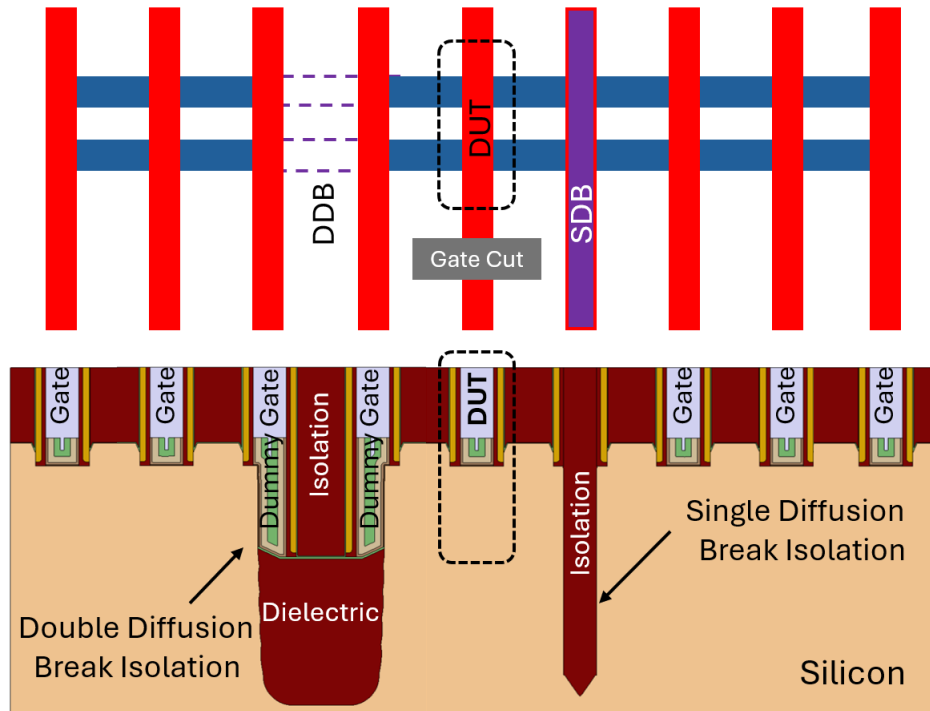


Figure 4.3: Typical layout of a FinFET transistor (a) surrounded by Fin isolations (DDB on the left and SDB on the right), and a Gate Cut on the bottom. (b) is the corresponding simulated cross-section representing the DDB on the left and the SDB on the right of the DUT

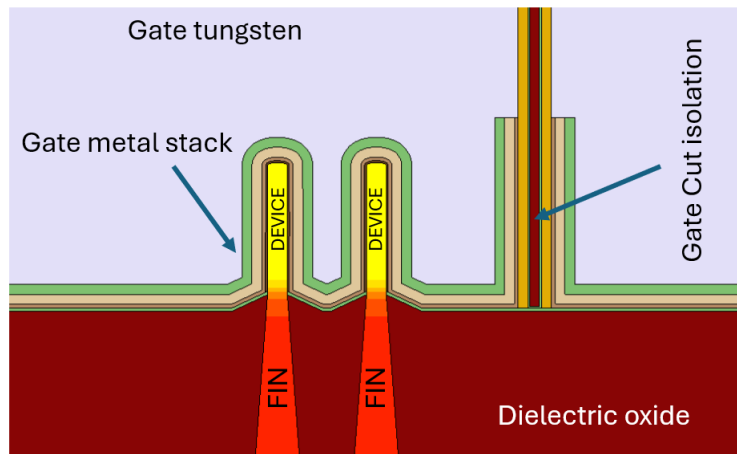


Figure 4.4: Simulated cross-section transversal to the Fin representing the Gate Cut isolation when the gate metal stack is completely formed

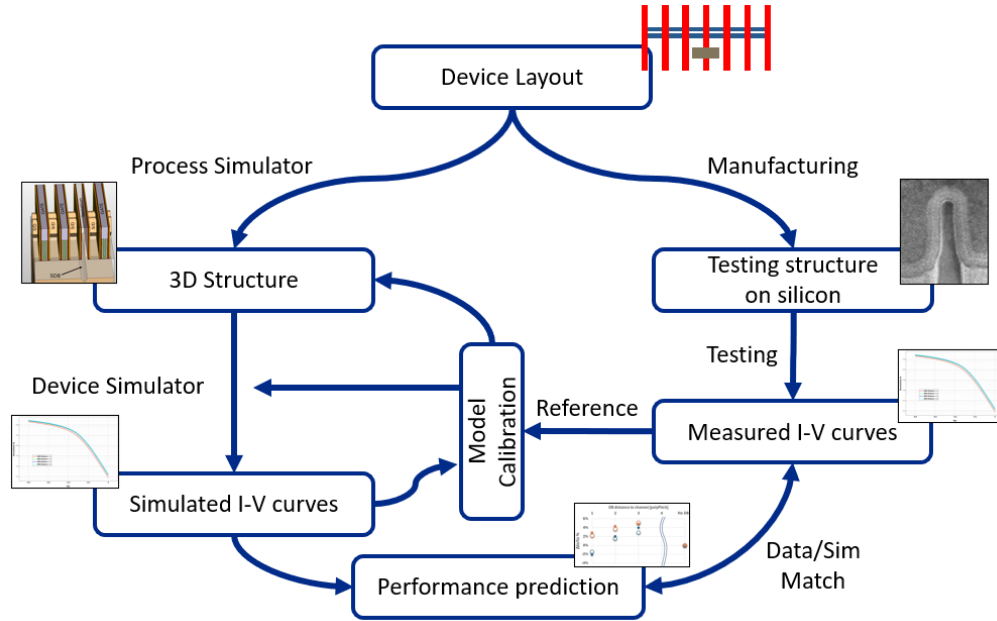


Figure 4.5: Workflow developed in this study to predict Local Layout Effects. It begins with the definition of the device layout, followed by two parallel branches: the left branch simulates the electrical behavior, while the right branch involves manufacturing and testing. Simulation results and silicon measurements are then compared and analyzed to validate the model

The test structures were fabricated using a commercial advanced 7nm FinFET technology node and subsequently characterized using a high-speed tester. Transfer I-V curves were measured, and key electrical parameters were extracted for all transistors (Figure 4.5, right branch).

In parallel, a process model of a FinFET device was developed using a process simulator³, based on the same 7nm technology. The output of the process simulation is a three-dimensional device structure corresponding to the input layout, as shown in Figure 4.6.

The resulting structure was then used as input for the device simulator⁴ to extract the simulated I-V characteristics (Figure 4.5, left branch).

The model was calibrated using measurements from reference structures, based on devices with continuous active and gate regions extending across multiple neighboring devices, fabricated and tested on silicon. Model param-

³Synopsys Sentaurus TCAD Process simulator (version U-2022.12)

⁴Synopsys Sentaurus TCAD Device simulator (version U-2022.12)

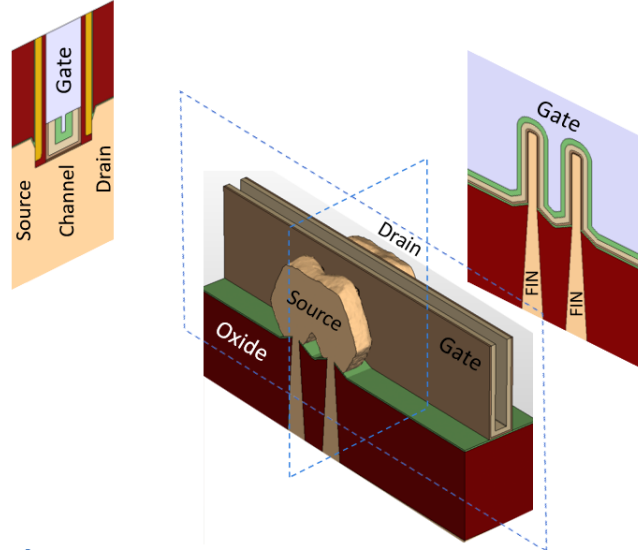


Figure 4.6: A sample of the process simulator output (center), consisting in a double FinFET device, and its transversal (right) and longitudinal (left) cross-sections

eters were tuned to ensure close agreement with the experimentally extracted I-V curves of the reference devices.

Once calibrated, the model was employed to predict the behavior of various LLEs, such as Diffusion Break and Gate Cut proximity effects. The model was validated against silicon data obtained from the corresponding structures.

The calibrated model predicts performance variations in LLEs caused by stress modulation. It helps identify stress-related performance issues to assess how much a specific LLE is influenced by stress-induced effects. This functionality serves as a powerful diagnostic tool for discerning the degree to which a specific LLE is influenced by stress-induced effects.

Moreover, the developed workflow provides a flexible platform for assessing the impact of potential changes in the manufacturing process, material choices, or annealing temperatures on device performance. As such, it constitutes a robust framework for process optimization and advanced technology development. Representative examples of these applications will be presented in Chapter 7.

A detailed mesh-independence⁵ study and solver convergence criteria are discussed in Section 5.1.4 to ensure numerical stability and accuracy.

Recent literature has increasingly emphasized the importance of capturing all components of the stress tensor when modeling variability in advanced CMOS devices. While longitudinal compressive stress is well known to enhance hole mobility in p-type FinFETs, vertical and transverse stress components, originating from gate stack deposition and isolation structures, can significantly degrade electron mobility in n-type devices. However, these effects are often neglected in simplified modeling approaches, leading to incomplete or misleading conclusions about device behavior. This gap in the literature underscores the need for a comprehensive simulation framework that retains the full three-dimensional stress tensor and propagates it through all relevant physical models. The methodology adopted in this work addresses this need by explicitly modeling longitudinal, vertical, and transverse stress components, thereby enabling accurate prediction of layout-induced performance variability across both device polarities.

4.3 Process Assumptions: Geometry and Materials

The fabrication of the test structures utilized in this study was carried out using a 7nm bulk FinFET technology node. However, the complete set of process details and proprietary parameters associated with this advanced technology node have not been publicly disclosed by the foundry. This lack of transparency presents a challenge in accurately replicating the manufacturing conditions and device characteristics in a research setting.

To address this limitation, it becomes necessary to adopt a modeling approach that incorporates informed assumptions based on industry trends, published literature, and known characteristics of similar technology nodes. These assumptions pertain to critical aspects such as device geometry, doping profiles, gate stack composition, and process integration schemes. While this introduces a degree of uncertainty, the process model developed herein aims to approximate the actual fabrication process with a high degree of fidelity, enabling meaningful analysis and interpretation of the experimental results.

⁵Mesh-independence: verification that simulation results do not change significantly with finer mesh discretization.

This approach is consistent with standard practices in semiconductor research, where access to proprietary process flows is often restricted. For instance, the ASAP7 predictive Process Design Kit (PDK) was developed to emulate 7nm FinFET behavior in the absence of foundry-disclosed data, and was widely adopted in academic research for this purpose [45]. By carefully calibrating the model against available empirical data and benchmarking it against known performance metrics, the study ensures that the derived insights remain relevant and applicable to the target technology.

The following process details are hypothetical and derived from literature and predictive PDKs. They do not represent actual foundry data.

4.3.1 Device Geometries

Devices with a physical channel length of 16 nm, a fin height of 40 nm, and a fin width of 6 nm are considered. Figure 4.7 depicts the transversal and the longitudinal cross-sections showing the main geometrical parameters and the net active charge distribution in the silicon simulated for a p-type device, while additional technology parameters are listed in Table 4.2b.

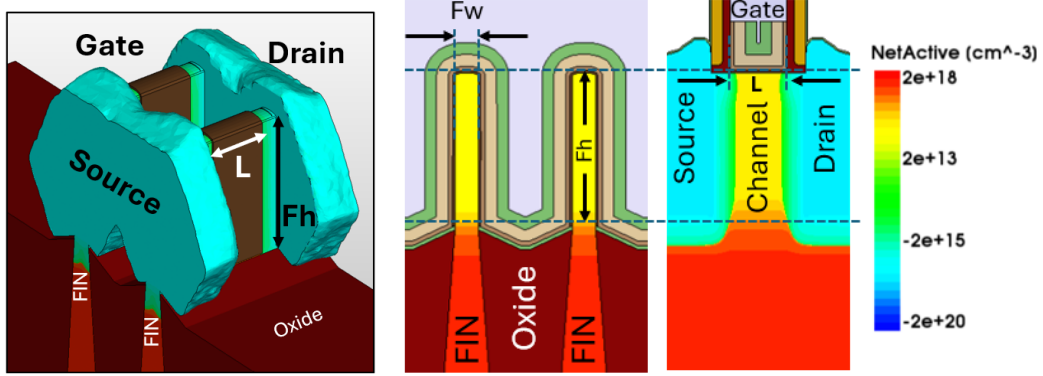
In planar transistor technologies, the device width is defined by the overlap between gate and active regions. In contrast, FinFET devices feature a channel that extends vertically, orthogonal to the wafer surface, and the transistor width is given by the following equation:

$$W = 2 \cdot Fh + Fw \approx 2 \cdot Fh \quad (4.1)$$

4.3.2 High-K Gate Dielectric

In the realm of VLSI technology, the continuous scaling down of transistor dimensions has necessitated the development of advanced materials and structures to maintain device performance and reliability. One such innovation is the high- k metal gate technology [46], which addresses the limitations of traditional silicon dioxide (SiO_2) gate dielectrics.

The term *high- k* refers to materials with a high dielectric constant (k), which is a measure of a material's ability to store electrical energy in an electric field. Traditional SiO_2 gate dielectrics have a dielectric constant of approximately 3.9. As transistor dimensions shrink, the gate dielectric thickness must also decrease to maintain adequate gate control over the



(a) 3D structure of a double Fin p-type device (b) The transversal (left) and longitudinal (right) cross-section of a double Fin p-type device

Figure 4.7: Parameter specification: Fin width (Fw), Fin height (Fh), device length (L), and the net active charge distribution in the silicon

channel [3]. However, this thinning leads to increased gate leakage currents due to direct tunneling, adversely affecting device performance and power consumption.

High- k materials, such as hafnium dioxide (HfO_2) with a dielectric constant ranging from 20 to 25, allow for a physically thicker dielectric layer while maintaining a high capacitance. This thicker layer significantly reduces gate leakage currents, thereby enhancing device performance and reliability.

The integration of high- k dielectrics necessitates the use of metal gates to replace the traditional polysilicon gates[47]. Metal gates provide several advantages, including improved work function tuning, reduced gate depletion effects, and enhanced thermal stability. The combination of high- k dielectrics and metal gates (HKMG) has become a standard in advanced CMOS technology nodes.

The concept of Equivalent Oxide Thickness is crucial in evaluating the performance of high- k dielectrics. EOT is defined as the thickness of a hypothetical SiO_2 layer that would provide the same capacitance per unit area as the high- k dielectric layer stack, as shown in Figure 4.8.

Mathematically, EOT is expressed as:

$$\frac{1}{C_{EOT}} = \frac{1}{C_{SiO_2}} + \frac{1}{C_{HfO_2}} \quad (4.2)$$

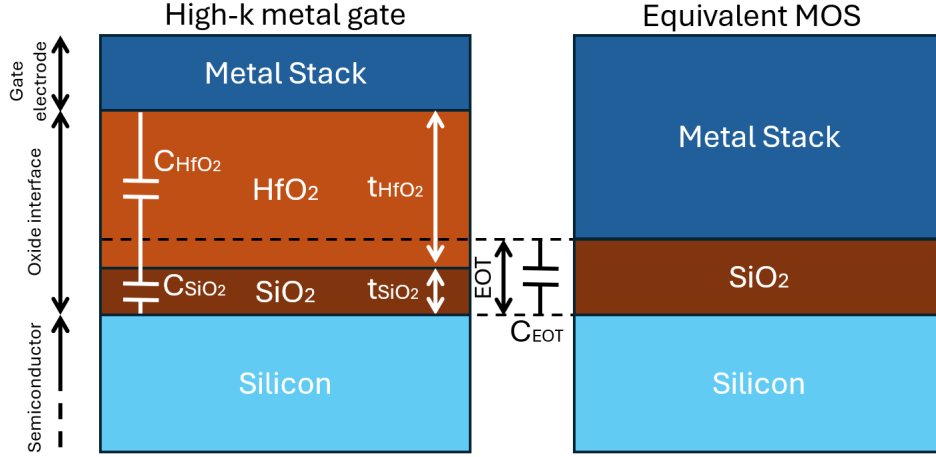


Figure 4.8: Equivalent Oxide Thickness schematic concept

$$\frac{EOT}{\epsilon_{SiO_2} \cdot A} = \frac{t_{SiO_2}}{\epsilon_{SiO_2} \cdot A} + \frac{t_{HfO_2}}{\epsilon_{HfO_2} \cdot A} \quad (4.3)$$

where ϵ_{SiO_2} and ϵ_{HfO_2} are the dielectric constants of the Silicon Oxide and the Hafnium Oxide respectively, t_{SiO_2} and t_{HfO_2} are the thickness of Silicon Oxide and the Hafnium Oxide layers respectively, A is the gate interface area.

Isolating EOT:

$$EOT = t_{SiO_2} + \frac{t_{HfO_2} \cdot \epsilon_{SiO_2}}{\epsilon_{HfO_2}} \quad (4.4)$$

EOT provides a standardized metric to compare the electrical performance of different gate dielectrics, regardless of their physical thickness. A lower EOT indicates a higher capacitance, which translates to better gate control and improved transistor performance.

This concept will be applied in this work during the process modeling and calibration phases, serving as a foundational element for accurately capturing layout-dependent effects (Section 5.4).

4.3.3 Gate Metal Stack Deposition

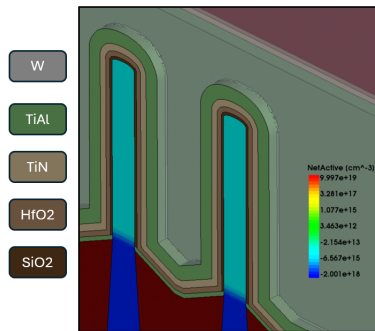
A HKMG is envisioned using hafnium dioxide (HfO_2), and a metal stack is constructed with layers of titanium nitride (TiN) with different thickness for n-type and p-type devices [48] and for threshold voltage tuning, titanium

aluminum ($TiAl$) as a barrier against tungsten diffusion towards the semiconductor and to reduce the gate resistivity[49, 48], and tungsten (W) to fill the top gate volume. The deposition temperatures of the materials are very important from stress management perspective. As the materials cool down from deposition to room temperature, the mismatch in thermal expansion rates can induce residual stresses, thereby impacting the performance of the devices.

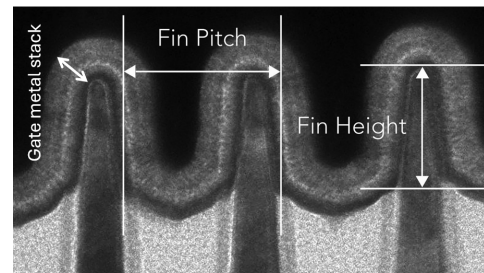
Deposition temperatures and material stacks are based on generic literature values and do not reflect proprietary process conditions. The deposition process is assumed as following[50, 51, 52]:

- **Silicon dioxide:** 900°C thermal oxidation of Fins
- **Hafnium dioxide:** 250°C ALD
- **Titanium Nitride:** 200°C ALD
- **Titanium Aluminum:** 350°C CVD
- **Tungsten fill:** 450°C CVD

The structure generated by the process simulator is shown in Figure 4.9a, and a cross-sectional TEM of a 10nm FinFET structure as an example where the gate metal stack is visible in Figure 4.9b.



(a) Result of the process simulation after the gate metal stack deposition in a n-type device



(b) Cross-sectional TEM transversal to the fin, showing the gate metal stack deposition over the fins[53]

Figure 4.9: Gate metal stack deposited over silicon fin FET devices

4.3.4 Gate Cut Isolation

In the early stages of VLSI, devices were relatively large. The precision control of the gate line ends was primarily achieved through lithography and patterning techniques. The focus was on integrating more transistors onto a single chip, moving towards higher density designs. As device dimensions shrank, the need for precise gate line end definition became more pronounced. The gate line end control is important in CMOS design, when the gates of neighbor p-type and n-type transistors need to be separated, and in order to reduce the space between the devices and avoid gate tip-to-tip shorts, a dedicated Gate Cut (GC) patterning step was introduced.

The transition from planar transistors to FinFETs brought about significant changes. FinFETs, with their three-dimensional structure, required more sophisticated GC techniques to ensure proper isolation and performance. The GC process became crucial for defining the fin structure and maintaining electrostatic control[54].

The GC process involves etching to isolate individual transistor gates from continuous gate strips. Breaking the continuity of a metal gate (which in fact is a strained metal bar) results in modification of the structure which can alter the mechanical stress profiles within the FinFET device, affecting the mechanical state of the device.

At advanced technology nodes, such as 7nm and 5nm, the GC process has become even more critical, and can be performed at various stages of fabrication, as shown in Figure 4.10. Some technologies implement it immediately after sacrificial polysilicon gate patterning (*GC first*), others after the deposition of the nitride wall (*GC late*), and some delay it until after the deposition of the gate metal stack (*GC last*).

Examples of GC First and GC Last cross-sections are shown in Figure 4.11. It is notable that the two implementations leads to a different deposition of the gate metal stack, which leads to different mechanical characteristics.

These approaches impact the mechanical stress of the devices in different ways. Managing the stress introduced by the multi-metal stack is essential to maintain device reliability and performance.

The proposed FinFET process model can effectively simulate all four types of Gate Cut procedures discussed earlier. 7nm FinFET technologies use preferably the *Gate Cut First* approach and it is also adopted it in this

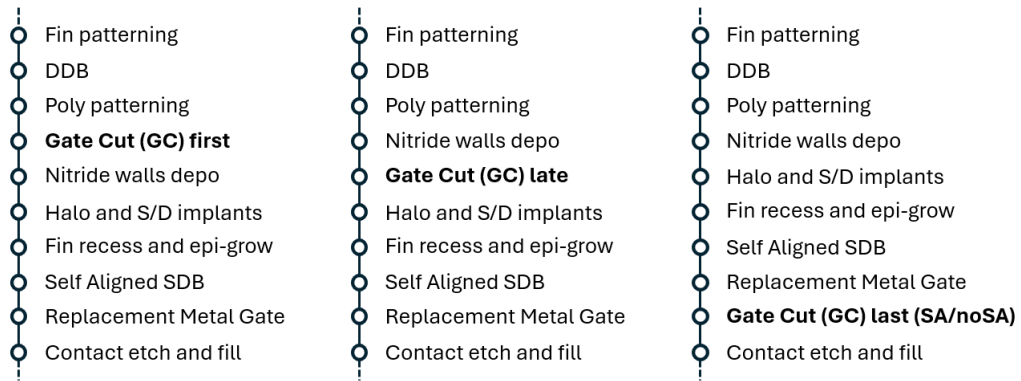
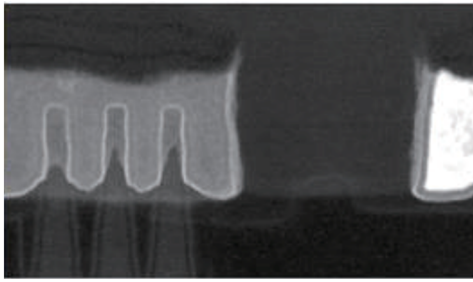
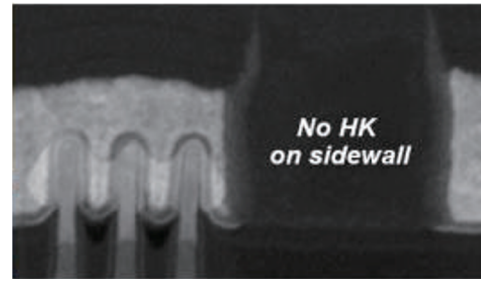


Figure 4.10: Different implementations of the Gate Cut process step order: *GC First*, during polysilicon patterning (left), *GC Late*, after the gate nitride spacers (center), and *GC Last*, after the Replace Metal Gate steps (right)



(a) GC First, performed before the gate spacers are deposited



(b) GC Last: Etch is performed after the metal gate stack is deposited

Figure 4.11: Cross-sectional SEM transversal to the fin: the gate metal stack climbs upward on the gate cut in case of GC First (left), while it is etched away in case of GC Last (right)[54].

study. A comprehensive description of this process module is provided in[55].

4.3.5 Epitaxial Growth in Source and Drain Module

In advanced FinFET CMOS technologies, the source and drain regions of p-type and n-type transistors are grown epitaxially on top of the fin, between gate poly-Si lines. P-type devices uses boron-doped SiGe while n-type gets phosphorus doped Si. Since those process operations are performed in sequence, they can be optimized independently to meet distinct performance targets.

Before the epitaxial growth, the device well is doped with Boron for n-type devices and Arsenic for p-type devices, with a concentration of approximately $1 \times 10^{18} \text{ cm}^{-3}$. The well doping defines the background doping profile and sets the initial electrical characteristics of the channel region. Subsequently, a Lightly Doped Drain (LDD) implantation is performed in the source and drain extensions, using Arsenic for n-type and Boron for p-type devices, to reduce electric field peaks and improve short-channel behavior. The LDD step is modeled using the Sentaurus Kinetic Monte-Carlo module[56], executed with the atomistic solver to accurately capture the stochastic nature of dopant distribution in the complex 3D fin architecture.

An approach with silicon epitaxy is used for the source and drain regions of n-type transistors, while an epitaxial growth enriched with 50% of Germanium concentration is used for p-type transistors to induce compressive stress into the channel. The epitaxial growth is performed with in-situ doping, reaching concentrations on the order of $1 \times 10^{20} \text{ cm}^{-3}$, to form highly conductive source and drain regions. This doping profile ensures low series resistance and contributes to the stress engineering required for mobility enhancement, particularly in p-type devices where SiGe is used.

An atomistic Lattice Kinetic Monte-Carlo module was employed to simulate the epitaxial growth [57]. Figure 4.12 depicts an example of the epitaxial growth simulated over time by the module integrated in the model.

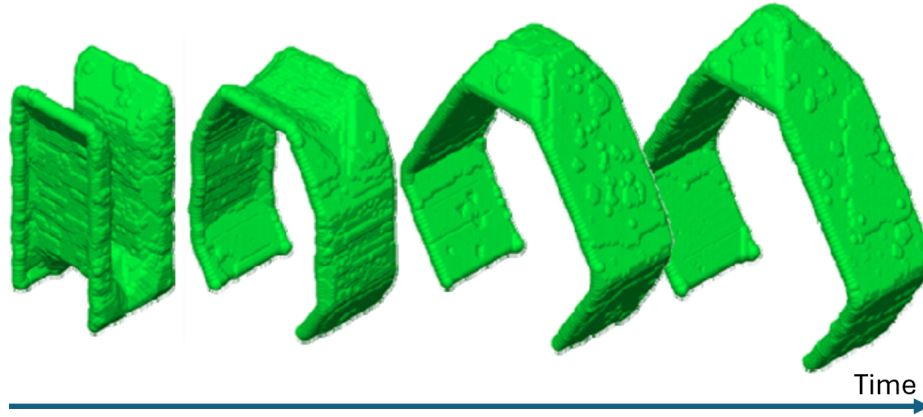


Figure 4.12: Silicon epitaxial growth simulated over time by LKMC module

Figure 4.13 illustrates the results of the module applied to a single Fin device (left) and a double Fin device (right).

This approach enables the prediction of interactions between the source

and drain stressors and the device architecture, which is fundamental for forecasting the actual stress induced in the device channel and its effects on mobility.

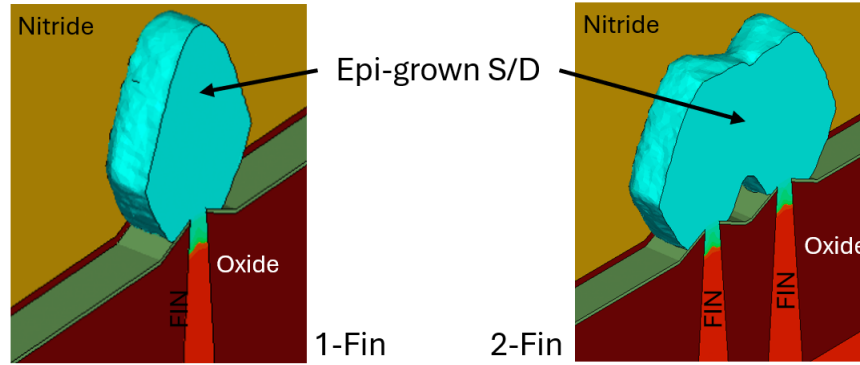


Figure 4.13: A typical diamond shape epitaxial growth simulation of a single fin device (left) and two merged growths of a double fin device (right)

4.4 Conclusions

This chapter introduces the FinFET architecture and outlines the TCAD-based modeling methodology used to simulate stress-related layout-dependent effects.

Given the proprietary nature of the 7nm FinFET process, the study adopted a calibrated TCAD-based approach, combining empirical measurements with informed assumptions to replicate real-world device behavior.

Key process elements, such as 3D geometries, high- k Metal Gate stacks, Gate Cut isolation strategies, and epitaxial growth, were modeled to capture their influence on mechanical stress and electrical variability. The workflow, grounded in a Design of Experiment framework, enabled the extraction and simulation of I-V characteristics, allowing for accurate prediction of Local Layout Effects such as Diffusion Break and Gate Cut proximity.

By aligning simulation outputs with silicon data, the model serves as a reliable tool for isolating stress-induced performance variations. This modeling foundation supports the analyses in the following chapters, where the impact of specific LLEs on device behavior will be explored in greater detail.

Chapter 5

Simulation Configuration and Model Calibration

This chapter outlines the setup and calibration of the TCAD environment for simulating 7nm FinFET devices. The simulation workflow is structured into process and device stages, enabling accurate modeling of mechanical stress and its impact on electrical behavior. By integrating realistic fabrication steps and advanced transport models, the framework supports predictive analysis of FinFET performance under varying stress and doping conditions, laying the groundwork for variability studies and design optimization in advanced CMOS nodes.

Overview

This chapter presents the configuration of the Technology Computer-Aided Design (TCAD) environment used to simulate 7nm FinFET devices. It is divided into two main sections: the process simulator¹ setup and the device simulator² setup. The process simulation models how mechanical stress is distributed inside the transistor, using detailed meshing and crystal orientation settings. The device simulation builds upon the process results, incorporating quantum and classical transport models to evaluate electrical behavior under various stress and doping conditions. Key aspects include the generation of high-resolution, symmetry-preserving meshes, the simulation of tensor stress fields, and the calibration of physical models for carrier mobility and recombination. The chapter also details the implementation of mobility degradation models and recombination mechanisms, ensuring a comprehensive and realistic simulation framework. This setup forms the foundation for subsequent analysis of layout-induced variability and performance optimization in advanced CMOS technologies.

5.1 TCAD Process Simulator Setup

5.1.1 General Setup

The process simulations were conducted using Synopsys Sentaurus TCAD 3D (version U-2022.12) to model a 7nm FinFET structure. Standard material parameters, including stiffness constants and stress-strain conversion coefficients, were sourced from the simulator's internal libraries and integrated into the simulation environment.

The thermal budget was carefully defined to reflect realistic fabrication conditions and to accurately model stress generation, dopant diffusion, and relaxation mechanisms. The process flow included several key thermal steps, such as annealings, material deposition temperatures, epitaxial growth.

Crystallographic orientations were selected to match standard CMOS manufacturing practices: wafer surface (100), fin sidewalls (110), and channel

¹A virtual reproduction of the fabrication steps (e.g., doping, etching, deposition) to generate a 3D structure of the device.

²A simulation that evaluates the electrical behavior (e.g., current-voltage characteristics) of the device based on the structure from the process simulation.

direction $\langle 110 \rangle$. These orientations are illustrated in Figure 5.1, which defines the 3D orthogonal directions relevant to the FinFET structure.

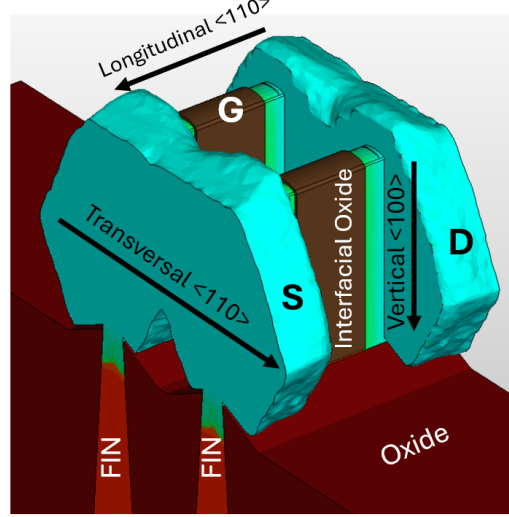


Figure 5.1: Convention of the 3D orthogonal directions

5.1.2 Tensor Field Simulation

In all numerical simulations, we retain the *full three-dimensional* (3D) stress tensor $\boldsymbol{\sigma}(\mathbf{r})$ at every mesh location $\mathbf{r}$. The tensor includes all six independent components (three normal and three shear) and is treated as symmetric ($\sigma_{ij} = \sigma_{ji}$):

$$\boldsymbol{\sigma}(\mathbf{r}) = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix}, \quad \sigma_{ij} = \sigma_{ji}. \quad (5.1)$$

As shown in Figure 5.1, the coordinate frame is chosen as follows: x is the vertical direction (normal to the wafer), y is the lateral direction (across the fin), and z is the transport direction (along the fin).

The *simulation framework always uses* the full 3D stress tensors of Eq. (5.1) as provided by the TCAD solver, without suppressing diagonal components. The complete tensor field is propagated to all physical models that require stress input (e.g., strain-dependent mobility or band-structure corrections), ensuring maximum physical fidelity.

For interpretability, sensitivity studies, and visualization, we compute *simplified 3D* descriptors *in post-processing only*. These quantities do not

feed back into the simulation kernel and therefore do not affect the computed device behavior. Specifically, we use the principal stress directions: the eigenvalues $\{\lambda_1, \lambda_2, \lambda_3\}$ and corresponding eigenvectors of $\boldsymbol{\sigma}$, used to summarize dominant tensile/compressive states along physically axes.

This separation between full-tensor simulation and simplified-tensor analysis maintains *full physical accuracy* where it matters (the numerical solution and device model evaluation), while providing *tractable, human-readable* summaries for trend analysis, visual inspection, and DTCO rule extraction. In short, *full 3D tensors* are used in simulations, while *simplified 3D tensors* are used for analysis only.

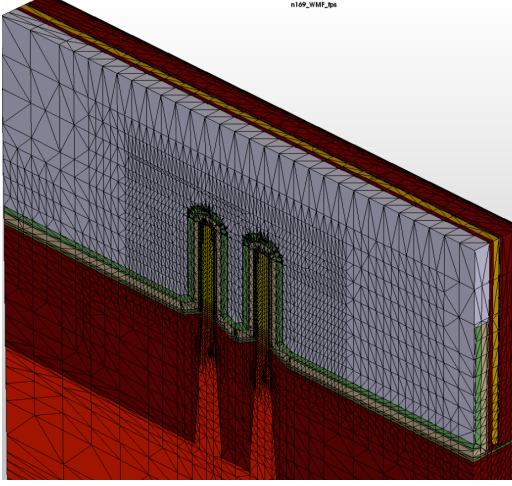
5.1.3 Mesh Definition

To ensure accurate resolution of physical phenomena in critical regions of the FinFET structure, an adaptive meshing³ strategy was employed throughout the process simulation. Particular attention was given to the areas surrounding the channel, the source/drain junctions, and the gate dielectric interfaces, where abrupt dopant gradients and high stress concentrations are typically observed. These regions demand high spatial resolution to capture the intricate variations in material properties and mechanical interactions.

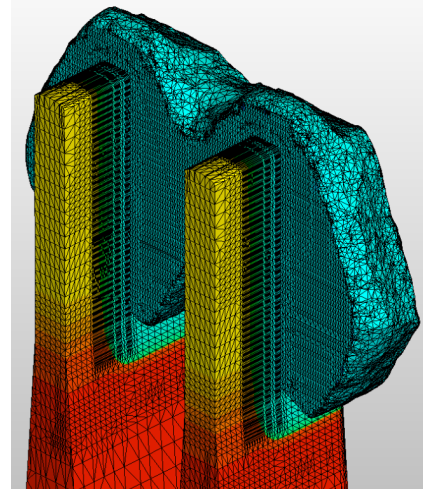
The mesh was automatically refined in areas with high stress or dopant gradients to improve accuracy without increasing simulation time unnecessarily. During this phase, symmetry planes were introduced to reduce computational complexity and to minimize numerical artifacts. The mesh structures were carefully designed to be geometrically balanced across the device, ensuring consistency in simulation results across different instances of the Device Under Test.

Examples of the generated mesh are shown in Figure 5.2. The global mesh demonstrates localized refinement near the channel, while the detailed view of the double-fin structure highlights the high-resolution meshing applied around the junctions. This meshing approach was essential for achieving convergence in subsequent simulations and for preserving the fidelity of the physical models applied in later stages.

³A technique where the mesh is automatically refined in regions with high gradients (e.g., near junctions or interfaces).



(a) Global mesh with higher resolution in the channel's surroundings



(b) Channels' mesh with high resolution around the junctions

Figure 5.2: Mesh examples in process simulator for a double fin structure

5.1.4 Mesh Independence and Solver Convergence

A mesh-independence study was performed using three discretization levels, minimum spacing near the oxide interface: *coarse* ($\approx 2.5 \text{ \AA}$), *medium* ($1.5 \text{ \AA}$), and *fine* ($1.0 \text{ \AA}$).

The minimum mesh spacing near the channel was set to $\sim 1.0 \text{ \AA}$ in the finest discretization level. Although this resolution is unusually fine for conventional TCAD, it is required for two reasons:

1. **Density Gradient (DG) Quantum Correction:** The DG model introduces a quantum potential term proportional to $\nabla^2 \sqrt{n}$, which is highly sensitive to local curvature of the carrier density near the oxide/semiconductor interface. A coarse mesh would distort this term, leading to inaccurate inversion charge and threshold voltage predictions.
2. **Stress Tensor Accuracy:** Capturing steep stress gradients at the fin sidewalls and gate dielectric interfaces demands sub-nanometer resolution. These regions exhibit abrupt material transitions and anisotropic elastic constants, which strongly influence the propagation of σ_{ij} components into the channel.

A mesh-independence study confirmed that reducing the minimum spacing from 2.5 Å to 1.0 Å changes the integrated stress and doping values by less than 1%, satisfying the acceptance criterion. This fine mesh ensures convergence of Newton iterations (residual norm⁴ $\leq 10^{-6}$) and numerical stability for DG-enabled simulations without introducing discretization artifacts. This condition was satisfied for all reference and representative LLE configurations (Diffusion Break and Gate Cut).

The most challenging convergence occurred near the subthreshold region ($V_{GS} \approx V_T$), where strong nonlinearity and stress-dependent mobility effects required fine bias stepping. Above-threshold and saturation regions converged within 5-10 iterations on average.

5.2 Implemented Process Steps

Figure 5.3 illustrates the main process steps modeled in this work. The nodes highlighted in red correspond to stages that are particularly sensitive to stress modulation, and thus have a significant impact on device performance. The modeling framework places special emphasis on the thermal budget, with detailed simulation of the thermal variations characterizing several critical steps.

An oxidation step at 900°C was employed to form the interfacial gate oxide layer during the Replace Metal Gate step. Subsequently, source and drain implantation was simulated using a Kinetic Monte-Carlo (KMC) solver, capturing the stochastic nature of dopant distribution and damage generation. This was followed by a laser annealing step at 1000°C, designed to activate dopants and repair implantation-induced defects. This annealing process was modeled as a short, high-intensity thermal pulse to accurately reflect its localized and transient thermal profile.

Epitaxial growth of silicon or silicon-germanium was performed at 450°C to form raised source/drain regions, engineered to introduce controlled stress for enhanced carrier mobility. Finally, the deposition of the gate metal stack was included to account for front-end-of-line (FEOL) stress contributions arising from the mechanical properties of the gate electrode materials.

⁴Residual norm refers to the magnitude of the error in the system of equations during iterative solution. A threshold of 10^{-6} ensures numerical convergence and physical accuracy.

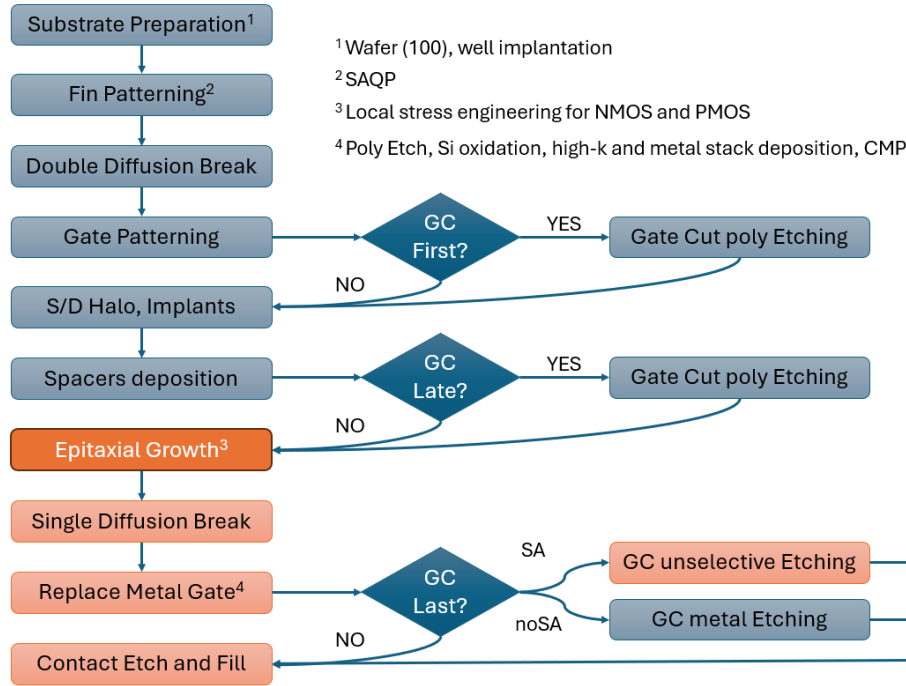


Figure 5.3: Process flow modeled in this work, up to the contact formation step. Particular attention is given to the stages where stress modulation is most pronounced (highlighted in red), as these are critical to the electrical and mechanical behavior of the final device

5.2.1 Integration with Device Simulation

The results obtained from the process simulation stage were seamlessly integrated into the device simulation environment to ensure continuity and physical consistency across the simulation workflow. The final 3D structures, including the detailed mesh and spatially resolved stress profiles, served as the input for the electrical simulation phase conducted using Sentaurus Device.

This integration allowed the transfer of critical information such as dopant distributions, mechanical stress tensors, and material interfaces, which are essential for accurately modeling carrier transport phenomena. The stress fields computed during the process simulation were crucial for evaluating their impact on carrier mobility. They directly influence the band structure and scattering mechanisms within the channel region.

The mobility and recombination models were adjusted using the stress

and doping data from the process simulation. This step was crucial for achieving accurate predictions of device characteristics, including threshold voltage, subthreshold swing, and drive current. By maintaining a consistent simulation framework from process to device level, the methodology enabled a comprehensive and predictive analysis of FinFET performance under realistic fabrication conditions.

5.3 TCAD Device Simulator Setup

5.3.1 Mesh re-Definition

After the process simulation, the mesh was rebuilt to support advanced models, especially those that include quantum effects. High-resolution meshing was applied orthogonally to the oxide-semiconductor interface and across the source/drain-channel junctions to accurately capture critical phenomena such as quantum confinement and tunneling, as shown in Figure 5.4.

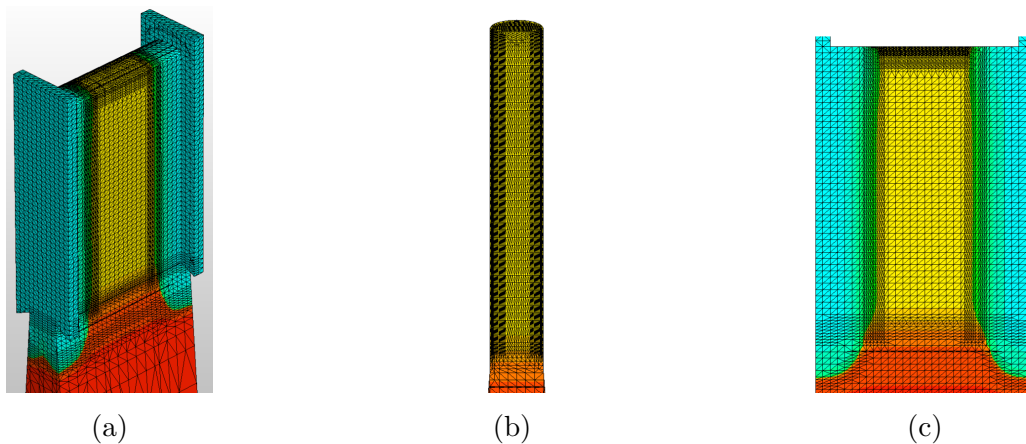


Figure 5.4: Device channel mesh used as input for the device simulator: 3D structure (a), transversal (b) and longitudinal (c) sections

To support these models, nonlocal meshing features were enabled⁵ [58].

⁵Nonlocal meshes are one-dimensional, special-purpose meshes that Sentaurus Device needs to implement one-dimensional, nonlocal physical models. A nonlocal mesh consists of nonlocal lines. Each nonlocal line is subdivided by nonlocal mesh points, to allow for the discretization of the equations that constitute the physical models. Sentaurus Device performs this subdivision automatically to obtain optimal interpolation between the nonlocal mesh and the normal mesh

These one-dimensional, special-purpose meshes are essential for implementing nonlocal physical models, such as the 1D Schrödinger equation and trap-assisted tunneling. Sentaurus Device automatically subdivides nonlocal lines into mesh points to ensure optimal interpolation and numerical stability.

5.3.2 Physical Models

Electrical device simulation was performed on the 3D structures generated by the process simulator. Two types of setup were required, for n-type and p-type devices, to properly configure the Synopsys Sentaurus Device Simulator. The simulation incorporated models for carrier mobility, recombination, and mechanical stress effects, with particular attention to electron/hole mobility, device stress, and Germanium concentration. The following models are described in details in Sentaurus Device User Guide [58].

The analysis focused on how modifications in the surrounding layout influence the stress profiles within the device channel, and how these stress variations, in turn, affect transistor performance. The trans-characteristic curve ($I_D - V_{GS}$) was derived by combining the outputs of the employed physical models, enabling a comprehensive evaluation of the electrical behavior of both FinFET variants.

Mobility Models

Several phenomena affecting carrier mobility were considered. Normally, mobility is a function of lattice temperature. In doped materials, carriers scatter with impurities, leading to mobility degradation.

The enabled models account for mobility degradation mechanisms including doping concentration, surface phonon scattering, surface roughness, and high electric field effects. All simulations were conducted at a constant room temperature of 25°C, consistent with the conditions under which the test structure measurements were performed.

Masetti Model⁶, a doping-dependent mobility model, is adopted and follows the equation:

⁶Described in [58] pages 399-400

$$\mu_{\text{dop}} = \mu_{\text{min1}} e^{-\frac{P_c}{N_{A,0} + N_{D,0}}} + \frac{\mu_{\text{cost}} - \mu_{\text{min2}}}{1 + \left(\frac{N_{A,0} + N_{D,0}}{C_r}\right)^\alpha} - \frac{\mu_1}{1 + \left(\frac{C_s}{N_{A,0} + N_{D,0}}\right)^\beta} \quad (5.2)$$

The default values for each coefficient for silicon are reported in Table 5.1.

Symbol	Electrons	Holes	Unit
μ_{min1}	52.2	44.9	$\text{cm}^2/\text{V}\cdot\text{s}$
μ_{min2}	52.2	0	$\text{cm}^2/\text{V}\cdot\text{s}$
μ_1	43.4	29.0	$\text{cm}^2/\text{V}\cdot\text{s}$
P_c	0	9.23×10^{16}	cm^{-3}
C_r	9.68×10^{16}	2.23×10^{17}	cm^{-3}
C_s	3.43×10^{20}	6.10×10^{20}	cm^{-3}
α	0.680	0.719	1
β	2.0	2.0	1

Table 5.1: Masetti model: default coefficients

The low-doping reference mobility μ_{cost} is determined by the constant mobility model, which accounts only for phonon scattering and, therefore, it is dependent only on the lattice temperature⁷.

Remote Phonon Scattering Model⁸ describes mobility degradation caused by scattering with surface phonons and surface roughness. The RPS degradation mobility is obtained by:

$$\mu_{\text{rps}} = \mu_{\text{rps0}} \left(\frac{F_\perp}{10^6 \text{V/cm}} \right)^{\gamma_1} \left(\frac{T}{300\text{K}} \right)^{\gamma_2} \quad (5.3)$$

The default values for each coefficient for silicon are reported in Table 5.2.

In high electric fields, the carrier drift velocity is no longer proportional to the electric field; instead, the velocity saturates to a finite speed v_{sat} . The High-Field Saturation model used is the Canali model, which derives from the Caughey-Thomas formula but includes temperature-dependent parameters.

⁷Described in [58] pages 397

⁸Described in [58] pages 433-435

Symbol	Electrons	Holes	Unit
μ_{rps0}	496.7	496.7	$\text{cm}^2/\text{V}\cdot\text{s}$
γ_1	0.68	0.68	1
γ_2	0.34	0.34	1

Table 5.2: RPS model: default coefficients

Canali Model⁹ considers the mobility as a function of the electric field:

$$\mu(F) = \frac{(\alpha + 1)\mu_{\text{low}}}{\left[1 + \left(\frac{(\alpha+1)\mu_{\text{low}}F_{\text{hfs}}}{v_{\text{sat}}}\right)^\beta\right]^{1/\beta}} \quad (5.4)$$

where μ_{low} is the low-field mobility. The exponent β is temperature-dependent according to:

$$\beta = \beta_0 \left(\frac{T}{300\text{K}}\right)^{\beta_{\text{exp}}} \quad (5.5)$$

The velocity saturation v_{sat} is given by:

$$v_{\text{sat}} = v_{\text{sat},0} \left(\frac{300\text{K}}{T}\right)^{v_{\text{sat},\text{exp}}} \quad (5.6)$$

The driving force F_{hfs} is calculated as:

$$F_{\text{hfs}} = |\nabla\phi_n| \quad (5.7)$$

The default values for each coefficient for silicon are reported in Table 5.3.

Symbol	Electrons	Holes	Unit
β_0	1.109	1.213	1
β_{exp}	0.66	0.17	1
α	0	0	1
$v_{\text{sat},0}$	1.07×10^7	8.37×10^6	cm/s
$v_{\text{sat},\text{exp}}$	0.87	0.52	1

Table 5.3: Canali model: default coefficients

⁹Described in [58] pages 435-436

Inversion and Accumulation Layer Mobility Model¹⁰ is based on the unified portion of the model described in [59]. This model includes doping and transverse-field dependencies and is similar to the Lucent (Darwish) model [60], but it contains additional terms that account for two-dimensional Coulomb impurity scattering. The implementation of the IALMob model is based on modified versions of the Philips Unified Mobility Model and the Enhanced Lombardi Model, but it is completely self-contained, and all parameters associated with this model are specified independently of those models.

The model has contributions from Coulomb impurity scattering, phonon scattering, and surface roughness scattering:

$$\mu_{\text{IALMob}} = \left(\frac{1}{\mu_{\text{Coulomb}}} + \frac{1}{\mu_{\text{phonon}}} + \frac{1}{\mu_{\text{surface}}} \right)^{-1} \quad (5.8)$$

where:

- μ_{Coulomb} is the mobility due to Coulomb impurity scattering.
- μ_{phonon} is the mobility due to phonon scattering.
- μ_{surface} is the mobility due to surface roughness scattering.

The distance from the interface plays a significant role in determining the mobility contributions from these scattering mechanisms.

Recombination Models

Recombination processes exchange carriers between the conduction band and valence band. In this analysis of FinFETs, three types of recombination are considered: Shockley-Read-Hall (SRH), Auger, and Band-to-Band tunneling.

Shockley-Read-Hall Recombination¹¹ is a non-radiative process occurring when a trap level is created within the energy gap due to a defect in the silicon lattice. The SRH recombination follows the model:

$$R_{\text{SRH,net}} = \frac{np - n_i^2}{\tau_p(n + n_1) + \tau_n(p + p_1)} \quad (5.9)$$

¹⁰Described in [58] pages 418-426

¹¹Described in [58] pages 484-497

where:

$$n_1 = n_i e^{\frac{E_{\text{trap}}}{kT}}, \quad p_1 = n_i e^{-\frac{E_{\text{trap}}}{kT}} \quad (5.10)$$

The lifetimes τ_n and τ_p are derived from the doping-dependent factor modeled with the Scharfetter relation:

$$\tau_{\text{dop}}(N_{A,0} + N_{D,0}) = \tau_{\text{min}} + \frac{\tau_{\text{max}} - \tau_{\text{min}}}{1 + \left(\frac{N_{A,0} + N_{D,0}}{N_{\text{ref}}}\right)^\gamma} \quad (5.11)$$

The default parameters for doping-dependent SRH lifetime are summarized in Table 5.4.

Symbol	Electrons	Holes	Unit
τ_{min}	0	0	s
τ_{max}	1×10^{-5}	3×10^{-6}	s
N_{ref}	1×10^{16}	1×10^{16}	cm^{-3}
γ	1	1	1

Table 5.4: Doping-dependent SRH lifetime: default coefficients

Auger Recombination Model¹² is a non-radiative process that occurs when two electrons collide in the vicinity of a hole, leading to a recombination event where the energy and momentum are absorbed by the second electron. The rate of band-to-band Auger recombination $R_{A,\text{net}}$ is given by:

$$R_{A,\text{net}} = (C_n n + C_p p) np \quad (5.12)$$

The Auger recombination rate $R_{A,\text{net}}$ is influenced by temperature-dependent Auger coefficients. The coefficient $C_p(T)$ for holes is given by:

$$C_p(T) = \left(A_{A,p} + B_{A,p} \left(\frac{T}{T_0} \right) + C_{A,p} \left(\frac{T}{T_0} \right)^2 \right) \left(1 + H_p e^{-\frac{p}{N_{0,p}}} \right) \quad (5.13)$$

where $T_0 = 300 \text{ K}$. There is experimental evidence for a decrease in the Auger coefficients at high injection levels. This effect is explained as resulting from exciton decay: at lower carrier densities, excitons, which are

¹²Described in [58] pages 500-501

loosely bound electron-hole pairs, increase the probability for Auger recombination. Excitons decay at high carrier densities, resulting in a decrease in recombination. This effect is modeled by the terms $1 + He^{-n/N_0}$.

The default values of the coefficients for silicon are listed in Table 5.5.

Symbol	Electrons	Holes	Unit
A_A	6.7×10^{-32}	7.2×10^{-32}	cm^6/s
B_A	2.45×10^{-31}	4.5×10^{-33}	cm^6/s
C_A	-2.2×10^{-32}	2.63×10^{-32}	cm^6/s
H	3.46667	8.25688	1
N_0	1×10^{18}	1×10^{18}	cm^{-3}

Table 5.5: Auger recombination model: default coefficients

Band-to-Band Tunneling¹³ is a generation-recombination process that occurs when an electron passes through the band gap from the valence band to the conduction band without the assistance of traps. Several models represent this phenomenon, and in this case, the Hurkx model is used. The equation for the net recombination rate $R_{\text{BB,net}}$ is:

$$R_{\text{BB,net}} = A \cdot D \cdot \left(\frac{F}{1 \text{ V/cm}} \right)^P e^{-\frac{BE_g(T)^{3/2}}{E_g(300 \text{ K})^{3/2} F}} \quad (5.14)$$

where:

$$D = \frac{np - n_i^2}{(n + n_i)(p + n_i)} ((1 - |\alpha|) + \alpha) \quad (5.15)$$

In this model:

- A and B are coefficients with units of $\text{cm}^{-3}\text{s}^{-1}$ and V/cm , respectively.
- P and α are dimensionless parameters.
- F is the electric field.
- $E_g(T)$ is the bandgap energy at temperature T .
- n and p are the electron and hole concentrations, respectively.
- n_i is the intrinsic carrier concentration.

¹³Described in [58] pages 533-548

The parameter α determines the nature of the process:

- When $\alpha = -1$, D is negative, indicating only generation.
- When $\alpha = 1$, D is positive, indicating only recombination.
- When $\alpha = 0$, there is a generation-recombination process. By default, $\alpha = 0$.

Mechanical Stress Models

Mechanical distortion of semiconductor microstructures leads to significant changes in the electronic band structure and carrier mobility. These effects are well documented [61] and are typically analyzed using deformation potential theory, which provides a framework for computing strain-induced modifications in the band structure. In this study, the calculation of mechanical stress and its impact on carrier transport was based on the combined use of three models: the *Deformation Potential Model*, which captures the shift in energy bands due to strain; the *Density of States Model*, which accounts for changes in the effective density of states under stress; and the *Piezoresistance Mobility Model*, which describes the variation in carrier mobility as a function of mechanical stress.

Deformation Potential Model¹⁴

For silicon, Bir and Pikus [62] propose the following model in which three Δ_2 electron valleys, heavy-hole, and light-hole are considered:

$$\Delta E_{C,i} = \Xi_d(\epsilon_{11}^0 + \epsilon_{22}^0 + \epsilon_{33}^0) + \Xi_u\epsilon_{ii}^0 \quad (5.16)$$

$$\Delta E_{V,i} = a(\epsilon_{11}^0 + \epsilon_{22}^0 + \epsilon_{33}^0) \pm \delta E \quad (5.17)$$

$$\delta E = \sqrt{\frac{b^2}{2} + \eta_i} \quad (5.18)$$

$$\eta_i = 4\Xi_u^0\epsilon_{jk}^0 \quad (5.19)$$

where:

¹⁴Described in [58] pages 952-963

- ϵ_{jk}^0 is a shear strain component.
- Δ is the band separation between the two lowest conduction bands.
- Ξ_u^0 is the deformation potential responsible for the band-splitting of the two lowest conduction bands.
- $j \neq k \neq i$.

Density of States Model¹⁵

The effect of strain leads to a change in the effective mass for both electrons and holes, and consequently, an alteration of the conduction band and valence band effective density-of-states (DOS).

For the conduction band in silicon, it can be approximated as three Δ_2 valleys. Without stress, the DOS of each valley is:

$$N_{C,i} = \frac{N_C}{3}, \quad i = 1, 2, 3 \quad (5.20)$$

where N_C is given by:

$$N_C(m_n, T_n) = 2.5094 \times 10^{19} \left(\frac{m_n}{m_0} \right)^{3/2} \left(\frac{T_n}{300 \text{ K}} \right)^{3/2} \text{ cm}^{-3} \quad (5.21)$$

Applying stress, there is a relative shift of the energy $\Delta E_{C,i}$ that is different for each Δ_2 valley. Moreover, the presence of shear stress leads to a significant change in the effective mass for electrons. The stress-induced change of the effective DOS for each Δ_2 valley can be written as:

$$N_{C,i} = \sqrt{\frac{m_{t1,i}}{m_t}} \sqrt{\frac{m_{t2,i}}{m_t}} \sqrt{\frac{m_{1,i}}{m_1}} \cdot \frac{N_C}{3} \quad (5.22)$$

Accounting for the change in the stress-induced valley energy $\Delta E_{C,i}$ and the carrier redistribution between valleys, the strain-dependent conduction-band effective DOS can be derived for Boltzmann statistics:

$$N'_C = \gamma \cdot N_C \quad (5.23)$$

where:

¹⁵Described in [58] pages 328-330

$$\gamma = \frac{1}{N_C} \sum_{i=1}^3 N_{C,i} e^{\frac{\Delta E_{C,\min} - \Delta E_{C,i}}{kT_n}} \quad (5.24)$$

and:

$$\Delta E_{C,\min} = \min(\Delta E_{C,i}) \quad (5.25)$$

The total hole effective DOS mass for arbitrary strain in silicon is:

$$m'_p(T) = \left(m_{cc,1}^{3/2} e^{-\frac{E_3-E_1}{kT}} + m_{cc,2}^{3/2} e^{-\frac{E_3-E_2}{kT}} + m_{cc,3}^{3/2} \right)^{2/3} \quad (5.26)$$

where E_1 , E_2 , and E_3 are the ordered band edges for each of the three valence valleys, and $m_{cc,1}$, $m_{cc,2}$, and $m_{cc,3}$ are the carrier-concentration masses for each of the valleys given by:

$$m_{cc}^{3/2}(T) = 2\sqrt{\pi}(kT)^{-3/2} \int_0^\infty dE m_{\text{DOS}}^{3/2}(E) \sqrt{E} e^{-E/(kT)} \quad (5.27)$$

The energy-dependent DOS mass of each valley m_{DOS} is equal to:

$$m_{\text{DOS}}^{3/2}(E) = \sqrt{2\pi} \left(\frac{2\hbar^2}{E} \right)^{3/2} \frac{1}{(2\pi)^3} \int_0^{2\pi} d\phi \int_0^\pi d\Theta \sin(\Theta) \left| \frac{\partial k}{\partial E} \right|_{k,\phi,\Theta,E} k^2 \quad (5.28)$$

The strain-dependent valence-band effective DOS is then calculated from:

$$N'_V = \left(\frac{m'_p}{m_p} \right)^{3/2} N_V \quad (5.29)$$

Piezoresistance Mobility Model¹⁶ is based on an expansion of the mobility enhancement tensor in terms of stress. The electron or hole mobility enhancement tensor follows the equation:

$$\frac{\mu_{ij}}{\mu_0} = \delta_{ij} + \sum_{k=1}^3 \sum_{l=1}^3 \Pi_{ijkl} \sigma_{kl} + \sum_{k=1}^3 \sum_{l=1}^3 \sum_{m=1}^3 \sum_{n=1}^3 \Pi_{ijklmn} \sigma_{kl} \sigma_{mn} \quad (5.30)$$

where:

¹⁶Described in [58] pages 986-998

- μ_{ij} is a component of the electron or hole stress-dependent mobility tensor.
- μ_0 is the isotropic mobility without stress.
- δ_{ij} is the Kronecker delta function.
- σ_{kl} is a component of the stress tensor.
- Π_{ijkl} is a component of the first-order electron or hole piezoconductance tensor.
- Π_{ijklmn} is a component of the second-order electron or hole piezoconductance tensor.

The temperature and doping dependency of the piezoresistance coefficients is obtained by combining the following two equations:

$$\pi_{ij} = -\Pi_{ij} \quad \leftrightarrow \quad \Pi_{ij} = -\pi_{ij} \quad (5.31)$$

$$\pi_{ijk} = -\Pi_{ijk} + \Pi_{ij}\Pi_{ik} \quad \leftrightarrow \quad \Pi_{ijk} = -\pi_{ijk} + \pi_{ij}\pi_{ik} \quad (5.32)$$

with these two equations:

$$\Pi_{ij}(N, T) = P_1(N, T)\Pi_{ij}(0, 300 \text{ K}) \quad (5.33)$$

$$\Pi_{ijk}(N, T) = P_2(N, T)\Pi_{ijk}(0, 300 \text{ K}) \quad (5.34)$$

where:

- π_{ij} and π_{ijk} are components of the first-order and second-order piezoresistance tensors, respectively.
- $\Pi_{ij}(0, 300 \text{ K})$ and $\Pi_{ijk}(0, 300 \text{ K})$ are piezoconductance coefficients for low-doped silicon at 300 K.
- $P_1(N, T)$ and $P_2(N, T)$ are doping factors given by:

$$P_1(N, T) = \left(\frac{300 \text{ K}}{T} \right) \frac{F_{-1}(\eta)}{F_0(\eta)} \quad (5.35)$$

$$P_2(N, T) = \left(\frac{300 \text{ K}}{T} \right)^2 \frac{F_{-2}(\eta)}{F_0(\eta)} \quad (5.36)$$

$F_r(\eta)$ is a Fermi-Dirac integral of order r and η is the Fermi energy measured from the band edge in units of kT ($\eta_n = (E_{F,n} - E_C)/kT$ and $\eta_p = (E_V - E_{F,p})/kT$).

The mobility enhancement factor γ is applied to each individual mobility component. This factor is equal to:

$$\gamma = 1 + P_1(N, T) \frac{\Delta\mu_{\text{low}}}{\mu_{\text{low},0}} \quad (5.37)$$

In high electric fields, the mobility along the main directions is proportional to the saturation velocity. The stress-dependent saturation velocity is modeled as:

$$v_{\text{sat},i} = v_{\text{sat},0} \cdot \frac{1 + \alpha \cdot t_i}{1 + t_i} \quad (5.38)$$

where $v_{\text{sat},0}$ is the stress-independent saturation velocity and α is the saturation factor.

By default, $\alpha = 1$ for planar devices, $\alpha = 0.2$ for n-type FinFETs, and $\alpha = 0.27$ for p-type FinFETs, which are kept unchanged in this study.

Density Gradient Quantum Correction¹⁷

Quantum mechanical effects become increasingly relevant in nanoscale semiconductor devices, particularly in the subthreshold and inversion regimes. TCAD Sentaurus incorporates these effects using the *Density Gradient* (DG) model, which offers a computationally efficient alternative to the full Schrödinger-Poisson approach.

The DG model modifies the classical drift-diffusion framework by introducing a quantum correction term derived from the gradient of the carrier density. The corrected electron and hole densities are expressed as:

$$n = n_{\text{cl}} \cdot \exp \left(\frac{\hbar^2}{12m_n^*q} \cdot \frac{\nabla^2 \sqrt{n_{\text{cl}}}}{\sqrt{n_{\text{cl}}}} \right) \quad (5.39)$$

$$p = p_{\text{cl}} \cdot \exp \left(-\frac{\hbar^2}{12m_p^*q} \cdot \frac{\nabla^2 \sqrt{p_{\text{cl}}}}{\sqrt{p_{\text{cl}}}} \right) \quad (5.40)$$

¹⁷Described in [58] pages 986-998

where n_{cl} and p_{cl} are the classical electron and hole densities, $\hbar$ is the reduced Planck constant, m_n^* and m_p^* are the effective masses of electrons and holes, and q is the elementary charge.

This correction captures key quantum phenomena such as carrier confinement and threshold voltage shifts, while avoiding the computational cost of solving the Schrödinger equation. The DG model is typically calibrated to reproduce experimental subthreshold swing and inversion charge in reference devices.

5.4 Calibration Strategy and Results

The device simulation was calibrated using a reference test structure, which has continuous active and gate regions across several neighboring devices, whose measured electrical characteristics were instrumental in extracting specific parameters for the physical models described earlier. This calibration ensured that the simulation accurately reflected real-world device behavior.

5.4.1 Gate Capacitance and EOT

The approach described in [63] was applied to a FinFET structure generated by the process simulator, based on the assumptions outlined in Section 4.3. Analysis of the measured off-state current (I_{off}) and the $I_{\text{off}}/I_{\text{on}}$ ratio from reference silicon structures indicated an equivalent oxide thickness (EOT) of approximately $7.7\text{\AA}$. However, a single SiO_2 layer of this thickness does not provide sufficient gate isolation, resulting in excessive gate leakage current due to quantum tunneling.

To ensure proper gate insulation, a high- k dielectric stack composed of silicon dioxide (SiO_2) and hafnium dioxide (HfO_2) was introduced. Considering a minimum SiO_2 thickness of $6\text{\AA}$ to maintain manufacturability and reliability, the required HfO_2 thickness to achieve an EOT of approximately $7.7\text{\AA}$ was calculated using the standard EOT formula (Equation 4.4):

$$t_{\text{HfO}_2} = (\text{EOT} - t_{\text{SiO}_2}) \cdot \frac{\epsilon_{\text{HfO}_2}}{\epsilon_{\text{SiO}_2}} \approx 11\text{\AA} \quad (5.41)$$

with $\text{EOT} \approx 7.7\text{\AA}$, $t_{\text{SiO}_2} = 6\text{\AA}$, $\epsilon_{\text{SiO}_2} = 3.9$, $\epsilon_{\text{HfO}_2} = 25$,

Based on this calculation, the high- k gate stack in the process model was defined with $t_{\text{SiO}_2} = 6\text{\AA}$ and $t_{\text{HfO}_2} = 11\text{\AA}$.

This stack reduces direct tunneling relative to an all- SiO_2 film of equal EOT, consistent with HKMG practice; gate leakage met measurement constraints.

5.4.2 Subthreshold Swing

Achieving the desired Subthreshold Swing (SS) requires precise control over the density of traps at the oxide-semiconductor interface. These traps, originating from defects or impurities, can capture and release charge carriers, thereby influencing the subthreshold behavior of the device. The interface trap density directly affects both the SS and the Threshold Voltage (V_T), with higher trap densities typically leading to increased SS and a shift in V_T .

The Sentaurus Device simulator provides the capability to define trap concentrations across various interfaces, including those within the High- k Metal Gate (HKMG) stack. This allows for accurate modeling of interface-related effects on device performance.

5.4.3 Drain Current in Linear Region

The calibration of the drain current in the linear region, characterized by $V_S = 0V$, $V_{DS} = 0.05V$, and $V_{GS} = V_{DD}$, is executed in two distinct phases to ensure high accuracy and alignment with measured values.

Initially, a coarse calibration is performed at the process simulation level. This involves adjusting the fin height parameter, which is a critical dimension in FinFET devices, to achieve a drain current that is reasonably close to the experimentally measured values. The fin height directly influences the channel dimensions, more precisely the channel width, and consequently the current flow through the device. By tuning this parameter, a preliminary alignment with the target drain current is achieved.

Following the coarse calibration, a fine-tuning phase is conducted at the device simulation level. In this phase, the gain factor, which represents the transconductance efficiency of the device, is meticulously adjusted. The gain factor is a crucial parameter that affects the amplification of the input signal and the overall current flow. By modifying the gain factor, the simulation is refined to minimize the discrepancy between the simulated drain current and the reference structure current. This ensures that the simulated device behavior closely matches the actual measured performance, resulting in a

negligible difference between the two.

Through this two-phase calibration process, the drain current in the linear region is accurately calibrated, ensuring that the simulation results are reliable and reflective of the real device characteristics.

5.4.4 Threshold Voltage (V_T)

The threshold voltage V_T is a crucial parameter that determines the voltage at which the device begins to conduct significantly. While the interface trap density plays a significant role in determining V_T , it can also be finely calibrated through the engineering of the gate metal stack.

The final threshold voltage is defined setting the voltage drop across the gate contact and the Hafnium oxide in the device gate metal stack for both n-type and p-type devices.

5.4.5 Stress Effects on Saturation Velocity in High Electric Fields

Measured data indicates that the stress dependency of minority carrier mobility, such as the mobility of carriers in the channel of a MOSFET, differs from the stress dependency of majority carrier mobility. Specifically, the stress effect on minority carriers may depend on the electric field and potentially exhibit a different doping dependency. To address this, Sentaurus Device provides an additional calibration factor for the stress effect applied to minority carrier mobility.

In high electric fields, the mobility (μ) along the main crystallographic directions is proportional to the saturation velocity (v_{sat}). Consequently, the dependence of the saturation velocity on stress is analogous to that of mobility, typically expressed with a proportionality factor. However, experimental data suggests that the saturation velocity can exhibit varying stress effects or, in some instances, no stress effect at all.

Sentaurus Device offers a calibration option to adjust the dependency of saturation velocity on stress. In this study, the primary mobility monitor is the drain current in the linear region; therefore, this parameter plays only a secondary role in model calibration. Nevertheless, the option was tuned to align with experimental data, ensuring that the simulated device behavior under stress conditions accurately reflects measured performance.

5.4.6 Calibration Results

The calibration steps have been iterated several times for both n-type and p-type devices to reach a good match with the measured data.

Figure 5.5 depicts the impact that the more sensitive calibration steps have to the transfer curve I_D - V_{GS} .

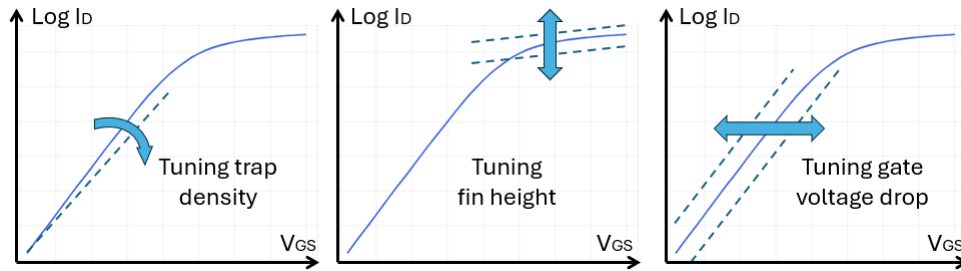


Figure 5.5: Impact of the calibration steps: (left) Channel width, (center) Interface traps, (right) Gate metal stack

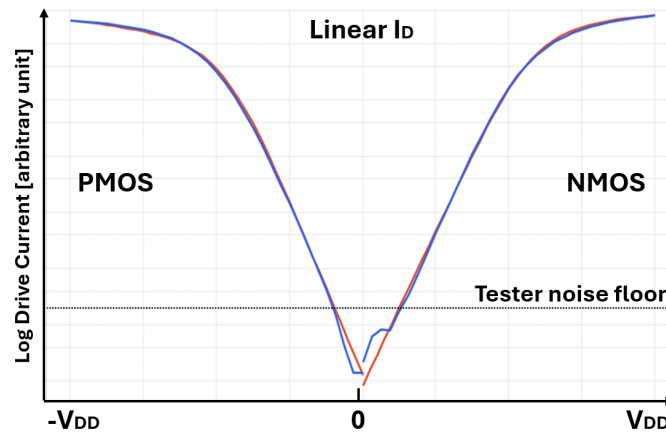


Figure 5.6: Overlapping between I_D - V_{GS} curves in linear region for measured data of reference structures (blue) and simulations (red).

Figure 5.6 depicts the I_D - V_{GS} curves in linear region extracted from n-type and p-type reference devices on silicon in blue and the simulated devices in red after the model calibration (y-axis scale is not reported to protect the confidentiality of the particular silicon process and the raw measurement data). The comparison is meaningful when considered above the tester noise floor (horizontal dashed line in the graph).

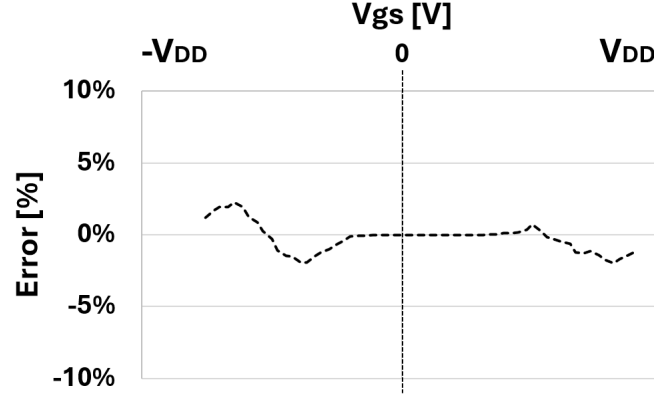


Figure 5.7: Difference between simulation and average value of measured Drain current (in linear region $V_{ds}=0.05\text{V}$) as a function of the gate bias V_{gs} .

The difference of the final result and the measured parameters are comparable to the standard error of the available data set, reaching a maximum and an average deviation of 2.2% and 0.2% respectively, as reported in Figure 5.7.

5.5 Conclusions

In this chapter, a comprehensive setup of the TCAD simulation environment for 7nm FinFET devices was established. The process simulation accurately captured the mechanical stress distribution using a finely tuned mesh and crystallographic orientation settings. These stress profiles were critical for understanding mobility variations and their impact on device performance.

The device simulation phase incorporated advanced physical models, including quantum effects and high-field transport phenomena. The use of detailed mobility and recombination models enabled a realistic representation of carrier dynamics under various operating conditions.

Overall, the simulation framework developed in this chapter provides a robust foundation for analyzing the influence of process-induced stress and layout variations on FinFET behavior. This setup ensures high fidelity in predicting electrical characteristics and supports further exploration of performance optimization strategies in advanced CMOS technologies.

Chapter 6

Experimental Results and Simulation Validation

This chapter presents a validated correlation between electrical measurements and 3D TCAD simulations to assess stress-induced Local Layout Effects in 7nm FinFETs. Using a high-density test chip, the impact of Diffusion Breaks, Gate Cuts, and Gate and Fin pitch variations on device performance is quantified. A vectorial decomposition of the stress tensor confirms the dominant role of longitudinal and vertical stress in mobility modulation, reinforcing the predictive capability of the proposed modeling framework.

Overview

Having calibrated the 3D model on silicon data, we now quantify LLE trends across Diffusion Break, Gate Cut, Contacted Poly Pitch, and .

This chapter presents a detailed correlation between experimental data and TCAD simulation results, focusing on the impact of stress-induced Local Layout Effect in FinFET devices fabricated in a commercial 7nm technology node. The analysis is based on a comprehensive set of test structures specifically designed to isolate the influence of individual layout features, such as Diffusion Break (DB), Gate Cut (GC), Contacted Poly Pitch (CPP), and Fin Pitch, on device performance.

The electrical measurements, performed on over 30,000 devices, are compared against a calibrated 3D TCAD model. The results confirm that p-type devices exhibit higher sensitivity to stress perturbations, particularly in proximity to SiGe-induced compressive stress regions, while n-type devices show more complex, non-monotonic behavior due to the interplay between stress relaxation and dielectric fill effects.

The chapter also includes a vectorial decomposition of the stress tensor components (longitudinal, vertical, and transversal) and their respective contributions to carrier mobility modulation. The analysis highlights the importance of layout symmetry, the shielding effect in multi-fin devices, and the impact of patterning variability (e.g., Fin Pitch Walking) on stress distribution.

The findings presented in this chapter have been validated through silicon measurements and have been disseminated in the following peer-reviewed publications:

- A. Rossoni et al., “Stress-Related Local Layout Effects in FinFET Technology and Device Design Sensitivity,” *IEEE Trans. on Semiconductor Manufacturing*, vol. 38, no. 2, 2025. [41]
- A. Rossoni et al., “Modeling Stress Effects from Fin Isolation in 7nm FinFET Transistors,” *IEEE CAS*, 09-11 October 2024. [64]
- A. Rossoni et al., “Impact of the Inter-Fin Space on Stress Modulation and FinFET Transistor Performance,” *IEEE ICICDT*, 23-25 June 2025. [65]
- A. Rossoni et al., “Impact of Gate and Inter-Fin Space on Stress Modulation and FinFET Transistor Performance,” submitted to *IEEE*

Trans. on Electron Devices, 2025.

These results provide a robust validation of the simulation methodology and support its applicability in predictive modeling of LLEs for advanced CMOS nodes.

6.1 Performance Variability Summary and Methodology

To provide a quick reference for the relative impact of each layout feature, Table 6.1 summarizes the observed performance variability for both n-type and p-type devices across the key LLE categories analyzed in this chapter.

To isolate stress-induced mobility effects from other contributors such as V_t shifts and external resistance, a constant-overdrive extraction method has been adopted. Drive current I_d was measured at a fixed gate overdrive voltage of $V_{ov} = (V_{gs} - V_t) = 0.3V$ and low drain bias ($V_{ds} = 0.05V$), ensuring that variations in V_t do not influence the reported performance changes. This approach is widely used in advanced CMOS characterization and variability studies [66].

In addition, the impact of external resistance R_{ext} , defined as the combined source and drain series resistance (outside of the gate-controlled channel region) has been evaluated. To accurately separate intrinsic channel effects (mobility) from parasitic series resistance (R_{ext}) and V_t shifts, a robust de-embedding methodology is required. The measured drive current (I_d) and peak transconductance (g_m) are both inherently degraded by R_{ext} . The measured g_m is related to the intrinsic transconductance (g_{mi}) by $g_m = g_{mi}/(1 + g_{mi} \cdot R_{ext})$ [67]. Therefore, to de-embed these effects, we employed the Y-Function (Ghibaudo) method ($Y = I_d/\sqrt{g_m}$) [68]. For each device, $I - V$ characteristics were measured in linear region. The I_d data was smoothed using a Savitzky-Golay filter to extract a clean g_m curve.

By performing a linear fit on the $Y - V_{gs}$ plot, we extracted the intrinsic threshold voltage (V_t) from the x-intercept and the intrinsic mobility parameter ($m^2 \propto \mu_0$) from the slope m . This V_t was then used to extract R_{ext} by plotting the total resistance ($R_{total} = V_{ds}/I_d$) against $1/(V_{gs} - V_t)$ and extrapolating the linear fit to its y-intercept.

This methodology allows for the independent tracking of intrinsic mobility and parasitic resistance. Our analysis confirms that R_{ext} is a significant

component, varying from 35% to 41% of R_{total} depending on device type and layout.

Stress values reported in this work are averaged stress integrated over the entire channel region. This approach captures the overall stress state experienced by carriers rather than localized peaks.

Unless otherwise noted, stress maps use the sign and color convention *compressive* < 0 and *tensile* > 0 .

This overview helps contextualize the detailed discussions that follow.

LLE	N-type	P-type
Single Diffusion Break	+5%	-10%
Double Diffusion Break	-2% to +4%	-8%
Gate Cut	-5% to +2%	-13%
Poly Pitch	-11% to +3%	-12% to +8%
Fin Pitch	-2% to +2%	-1% to +1%

Table 6.1: Performance Variability Caused by Stress-LLEs

6.2 Impact of Diffusion Break Proximity

With the overall methodology established, we begin by analyzing the impact of Diffusion Break (DB), a key isolation feature, on device performance. Figure 6.1 illustrates the layout used to predict device performance variability resulting from the proximity of a Diffusion Break (DB) to the Device Under Test (DUT). The layout is then processed by the process simulator, producing the 3D structure shown in Figure 6.2. The model captures key features of the device at the end of the gate metal stack deposition step, including the gate metal layers, gate spacers, and the underlying silicon. This representation enables accurate analysis of stress distribution and its impact on the electrical behavior of the DUT in the presence of nearby isolation structures.

The 3D structures in Figure 6.2 are used as input to extract device performance and key parameters. The measurements of the devices manufactured on silicon are then compared with the simulation results to validate the model and the proposed workflow.

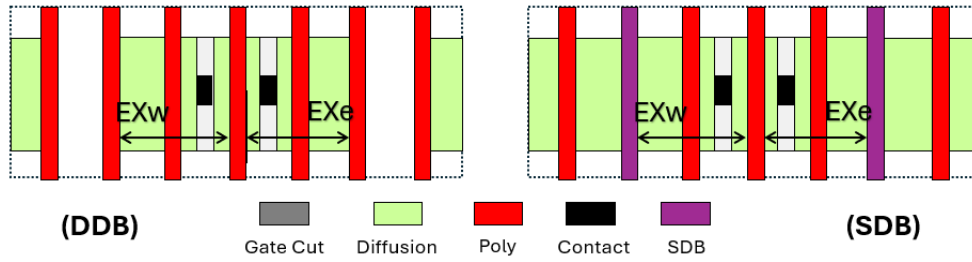


Figure 6.1: Test structure to evaluate the Double Diffusion Break (DDB) and the Single Diffusion Break (SDB) proximity impact on device performance. The layout is used for both device manufacturing and device simulation

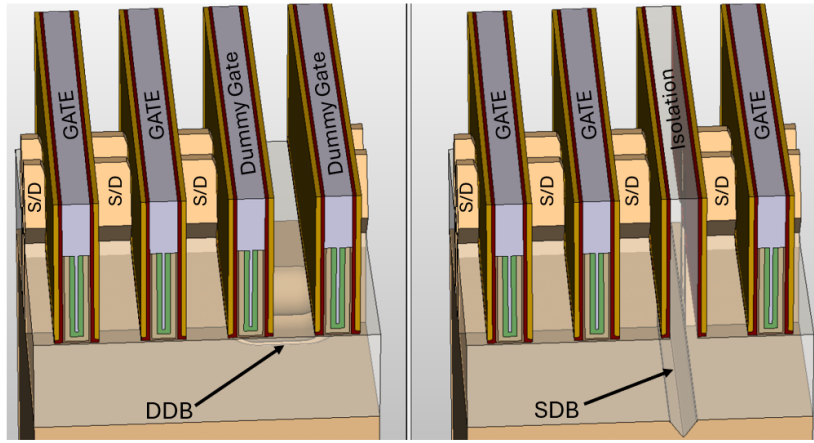


Figure 6.2: 3D structure used to characterize the DDB (left) and the SDB (right) proximity effect to the transistor performance

(DB) P-type Device Analysis

Simulation and Data

The graph in Figure 6.3 represents the change in the p-type drain current in the linear region, referred to a reference device with a continuous fin line, as a function of the distance from the DB boundary.

The plot shows both the simulated electrical performance (open circles marked "Sim") and the electrical results measured on the silicon (closed circles marked "Data") in both DDB and SDB cases. Double diffusion breaks are performed prior to the formation of the source and drain regions (as described in Section 4.3.4), which, in the case of p-type, are rich in Germanium and generate the majority of the stress within the device. Conversely,

the process steps that form single diffusion break occur after the epitaxial growth and are thus carried out under conditions of high structural stress. It is inevitable that the fin cut in the SDB causes substantial relaxation, which is only partially contained by the nitride that constitutes the gate spacers. For this reason, the performance in the case of SDB is more sensitive to the distance of the cut compared to the DDB case.

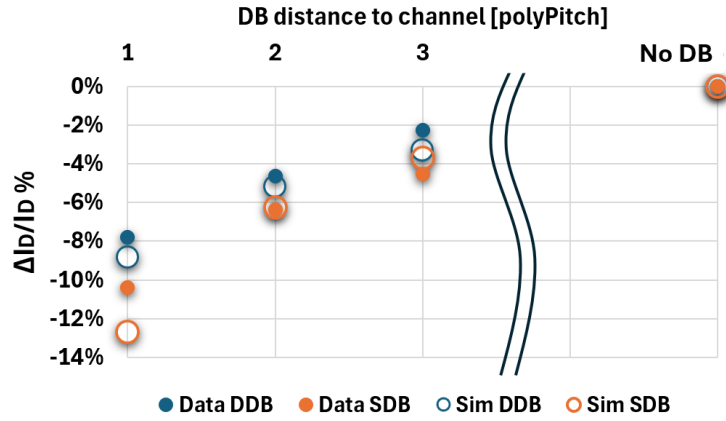


Figure 6.3: Changes in p-type drain current (in linear region @ $V_{sd} = 0.05V$) as a function of the distance from the Diffusion Break boundary

As expected, the reduction in performance due to the DB increases when the DB is closer to the device. The simulation may slightly overestimate the effect near the DB due to uncertainty in the etch profile and depth.

Stress Analysis

Given the good correspondence between simulations and data, this aspect can be analyzed from the stress perspective using the process model.

Figure 6.4 represents the variation of the three orthogonal stress components to which the channel is subjected relative to the distance of the SDB. It is evident that the stress in the fin direction (longitudinal) is the main cause of performance variability, as the variation of the other components is negligible.

Figure 6.5 visualizes the longitudinal compressive stress of the silicon in blue and shows that this stress increases with the distance from the DB in both cases, but is less intense in the case of SDB. This fact confirms the hypothesis previously deduced about the different sensitivity of the two types

of DB.

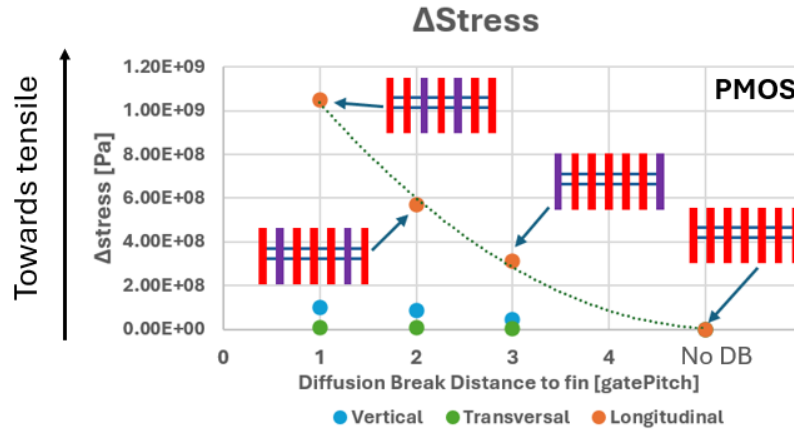


Figure 6.4: Variation of the 3D components of the channel stress as a function of the distance to the Diffusion Break in p-type devices

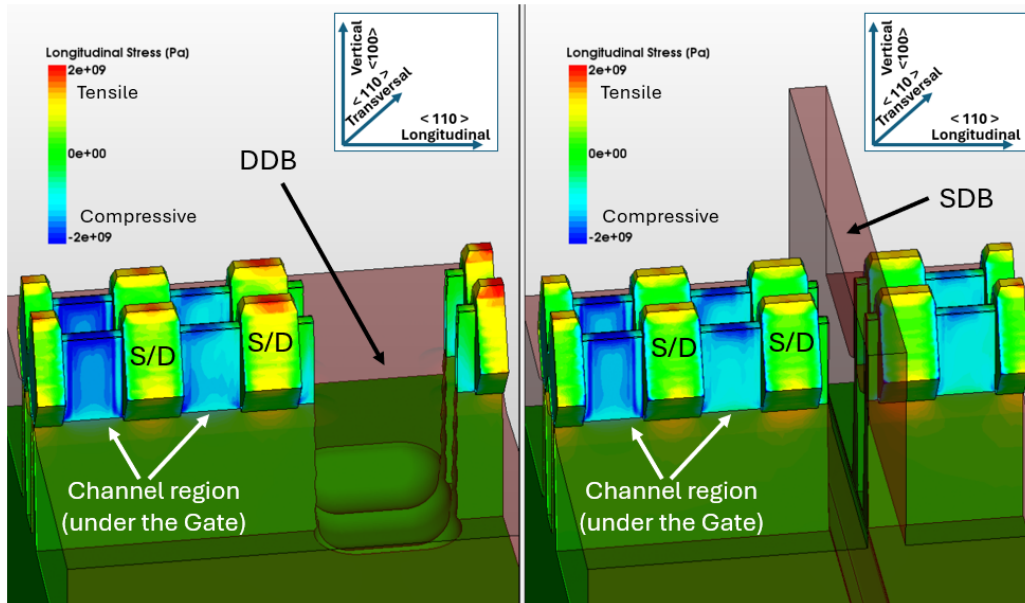


Figure 6.5: Longitudinal compressive stress of the silicon as a function of the distance from the Diffusion Break

Figure 6.6 represents the cross-section longitudinally to the fin and the corresponding compressive stress in blue, in the case of a structure without DB (top) and with SDB near the DUT (bottom). From this, it can be

deduced that the stress generated by the SiGe, which in the absence of DB is distributed uniformly across all device channels, partially dissipates towards the trench of the SDB created by etching.

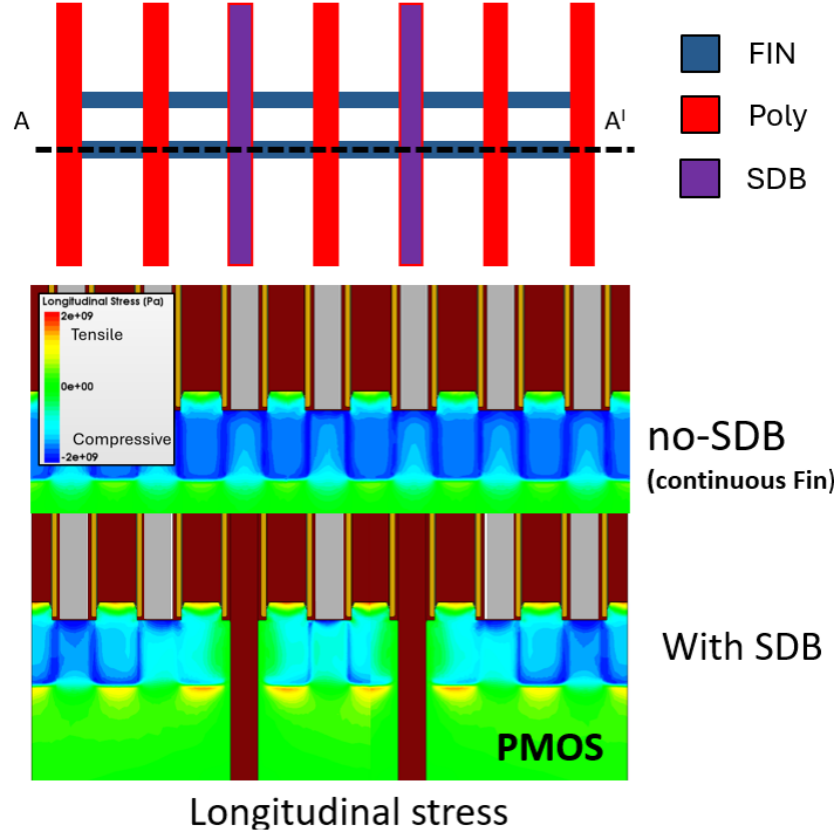


Figure 6.6: Cross-section longitudinally to the fin and the corresponding compressive stress in blue, for a structure without DB (top) and with SDB near the DUT

(DB) N-type Device Analysis

Simulation and Data

We now turn to n-type devices, which respond differently to stress compared to their p-type counterparts due to their distinct channel properties and stress sensitivities.

The graph in Figure 6.7 represents the change in the n-type drain current in the linear region, referred to a device without DB, as a function of the

distance from the Diffusion Break boundary.

The plot shows both the simulated electrical performance (open circles marked "Sim") and the electrical results measured on the silicon (closed circles marked "Data") in DDB and SDB cases. The results are similar to the p-type case, except for the fact that there is a non-monotonic behavior when the DB is far away from the DUT.

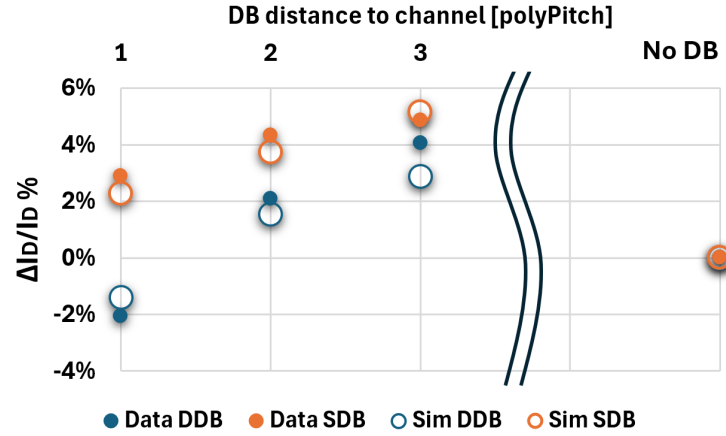


Figure 6.7: Changes in n-type drain current (in linear region @ $V_{ds} = 0.05V$) as a function of the distance from the Diffusion Break boundary

Stress Analysis

The non-monotonic electrical characteristics of the n-type device shown in Figure 6.7 warrant a detailed discussion. Figure 6.8 (top) illustrates a longitudinal section of the fin, depicting the distribution of longitudinal stress in silicon. Figure 6.8 (bottom) presents the graph of this stress across three consecutive process steps: before the SDB etching (blue line), after performing the cut (green line), and after filling the cut with the insulating dielectric (red line).

Compressive stress is not beneficial for n-type device performance; rather, it is detrimental. Therefore, a traditional approach with silicon epitaxy was used for the source and drain regions of n-type transistors. As a result, the overall stress levels in the n-type channel are lower (on an absolute scale) than in p-type engineered FinFETs. Simulation shows that in a continuous fin (no SDB in close proximity) of n-type transistor, the stress is compressive, and its value is relatively small. The DB performed through the etch of a small fin

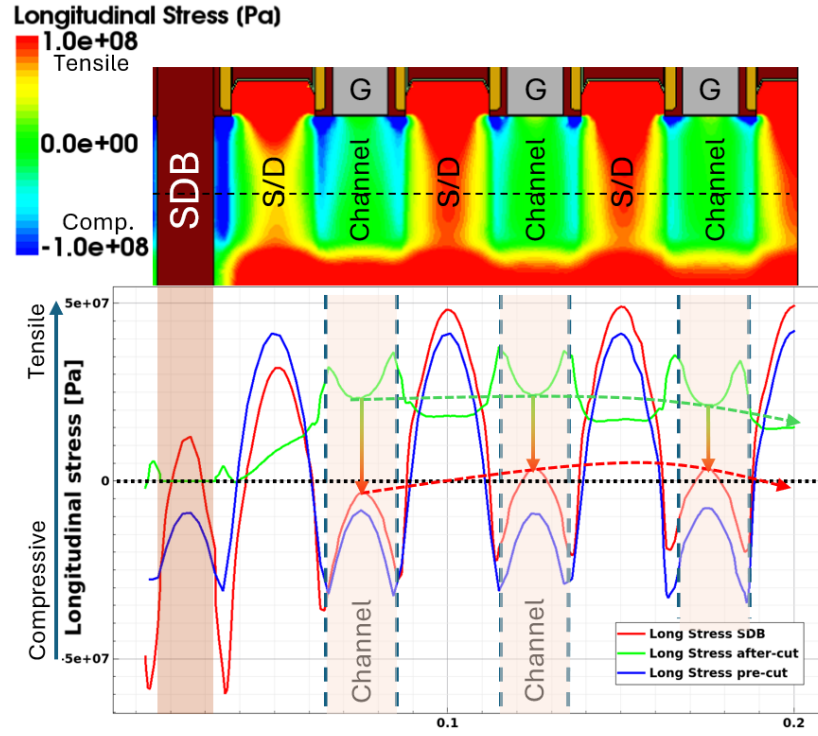


Figure 6.8: The top shows the cross-section of the testing structure with the longitudinal stress at the end of the process in blue. The bottom plot the longitudinal stress distribution in silicon across three process steps: before SDB etching (blue), after the cut (green), and after filling with dielectric (red)

segment causes a relaxation of the compressive stress and a transition into a tensile state (Figure 6.8 blue to green). The stress profile in the source/drain region becomes flat, lowering the central tensile peak and raising the lateral wings to the same level. The channels, which have a smaller silicon volume compared to the source/drain regions, absorb this change from both sides, transitioning from a compressive to a U-shaped tensile stress.

Following the SDB etching, the dielectric material fills the created gap. This dielectric fill exerts a compressive force on the adjacent fins, pushing back the stress towards its original state before the cut (Figure 6.8 green to red). Notably, even the maximum compressive stress induced by the dielectric fill is not able to completely revert the stress change induced by the cut.

The impact of both aforementioned process steps on stress gradually di-

minishes with increasing distance from the DB (green and vertical arrows). Consequently, near the SDB, the compressive stress decreases as the distance from the SDB cut increases, due to the dominating stress gradient generated by the filling dielectric. This stress gradient positively impacts the performance of n-type transistors, as reduced compressive stress enhances electron mobility.

For distances greater than four poly pitches, the relaxation due to the SDB cut starts to diminish, pushing back the longitudinal stress towards the original state before the cut. This stress gradient from tensile to compressive becomes the dominant component, thereby reversing its impact on mobility. Fins that remain in an uncut line exhibit higher compressive stress compared to those that have undergone the SDB process. This elevated stress correlates with inferior performance metrics, underscoring the benefits of the SDB architecture in optimizing n-type device functionality.

Stress variations from process steps and material properties cause non-monotonic behavior in n-type devices. These results confirm the expected mechanism: etch-driven relaxation near the cut followed by dielectric-fill recompression, which dominates in SDB and is milder in DDB. The simulation achieves excellent agreement with silicon, with RMSE ~ 0.056 (DDB) and ~ 0.036 (SDB) and negligible bias, validating the calibration strategy for stress-sensitive n-type behavior.

In contrast, in p-type devices, the longitudinal stress variation generated by SiGe dominates all other components, maintaining a monotonic performance curve and making the devices less sensitive to the material properties used in the DB process. The sensitivities (linear fit) are approximately $+2.23\%/CPP$ (DDB) and $+3.30\%/CPP$ (SDB) for simulation, in good agreement with the measured data ($+1.82$ and $+2.93\%/CPP$, respectively). The quantitative match is strong (RMSE ~ 0.055 for DDB and 0.085 for SDB, with negligible bias), confirming that the model captures the dominant physics: stress relaxation near the trench (more pronounced for SDB due to its later process insertion) and progressive recovery as the distance increases.

6.3 Gate Cut Proximity Effects

After examining the effects of fin isolation, we next explore gate isolation through Gate Cut (GC), which introduce stress perturbations via a different

mechanism.

Figure 6.9 represents the layout used as input to the process simulator, and Figure 6.10 depicts the generated 3D structure used to predict the electric characteristics resulting from the proximity of a GC to the DUT. It shows the gate metal stack deposited over the silicon fins, the Gate Cut, and the filling dielectric isolation.

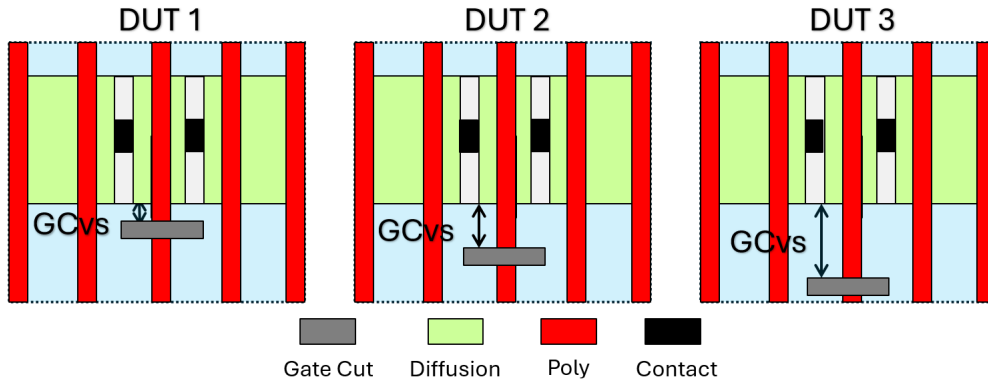


Figure 6.9: Test structure layout used to evaluate the impact of Gate Cut proximity on device performance, applied in both fabrication and simulation

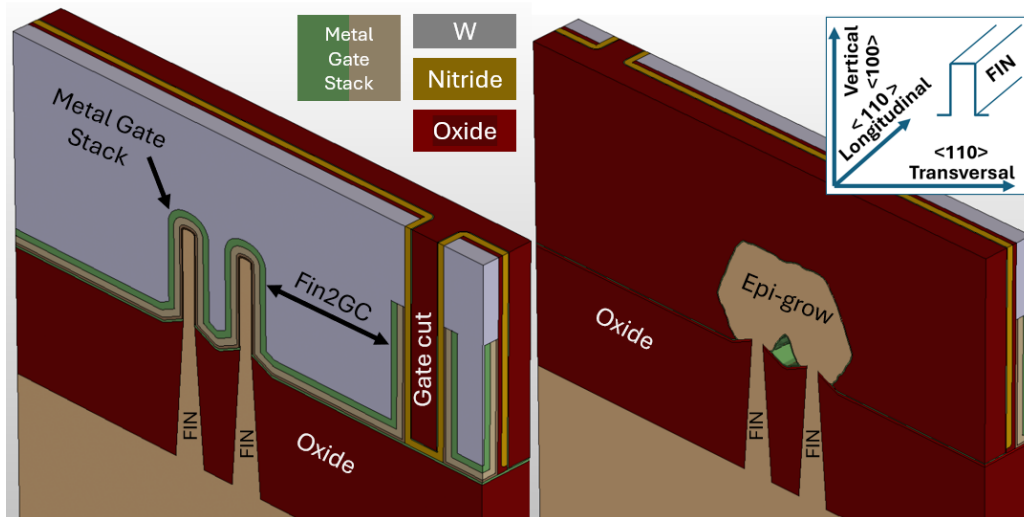


Figure 6.10: Front (left) and back (right) views of the 3D structure generated by the process simulator, illustrating the Gate Cut and its proximity-dependent impact on performance

P-type Device Analysis

Simulation and Data

Figure 6.11 illustrates the variation in p-type drain current within the linear regime as a function of the GC proximity. Both simulated results (open circles labeled "Sim") and silicon measurements (closed circles labeled "Data") are presented. The data reveal a non-linear trend, suggesting the presence of two overlapping stress mechanisms: vertical stress enhancement near the GC and longitudinal stress recovery at greater distances. These competing effects contribute to the observed performance modulation, which is further analyzed in Figure 6.16.

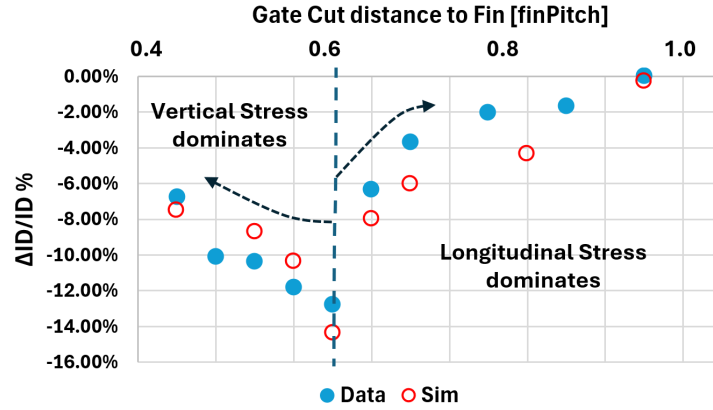


Figure 6.11: Changes in p-type drain current (in linear region @ $V_{sd} = 0.05V$) as a function of the distance from the Gate Cut boundary

Stress Analysis

Figure 6.12 illustrates the simulated 3D stress component distributions within the device channel region. At shorter GC distances from the fin, the vertical stress enhancement on the mobility dominates[69], while at longer distances, the longitudinal stress effect prevails (see Table 1.1).

Figures 6.13 and 6.14 illustrates how the GC and the gate spacers contribute to the stress balance, affecting hole mobility and, consequently, p-type transistor performance.

Figure 6.15 presents the vertical stress in the nitride walls for devices at varying distances from the GC. The figure shows a simplified view of the

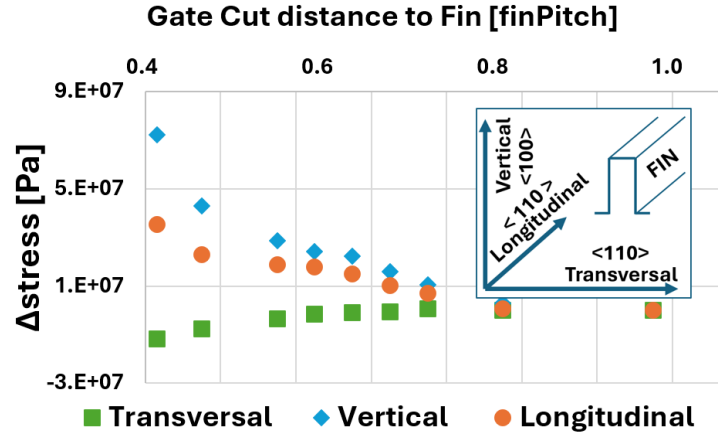


Figure 6.12: Variation of the 3D components of the channel stress as a function of the distance to the Gate Cut in p-type devices

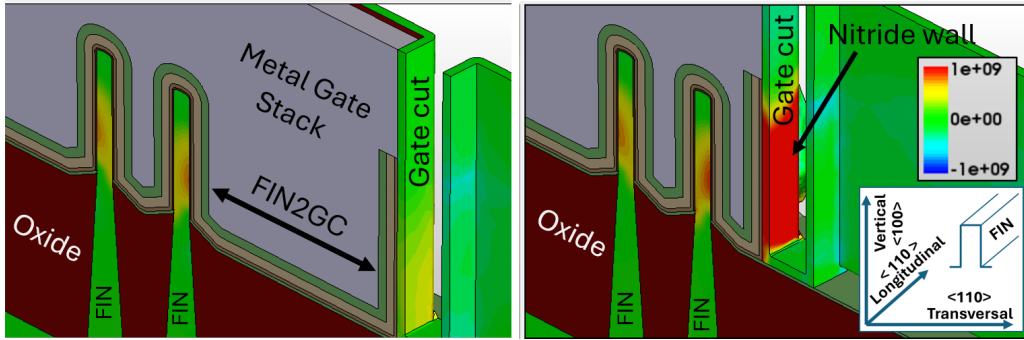


Figure 6.13: Representation of the vertical stress in the nitride walls with different GC distances: GC far from the device (left) and very close to the device (right)

epitaxial source/drain grown next to the gate spacer. The stress developed along their interface contributes to the mobility modulation in the channel.

However, the behavior of the longitudinal stress is different from the vertical stress. The nitride spacer layer around the GC region creates a fulcrum that absorbs part of the compressive longitudinal stress generated by the source/drain SiGe, as shown in Figure 6.14. The reduction of compressive longitudinal stress in the channel results in a reduction in hole mobility, but this effect is less pronounced compared to the impact of vertical stress in the case of fin architecture with (110) channel orientation.

At a larger distance from the GC, the longitudinal compressive stress

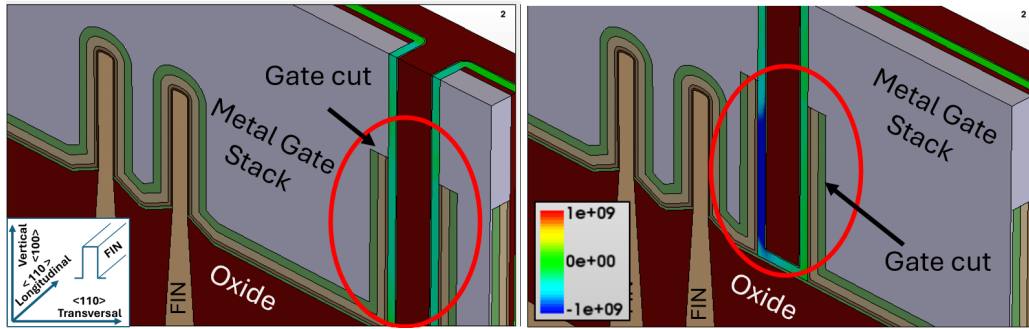


Figure 6.14: Representation of the longitudinal stress in the nitride walls with different GC distances: GC far from the device (left) and very close to the device (right)

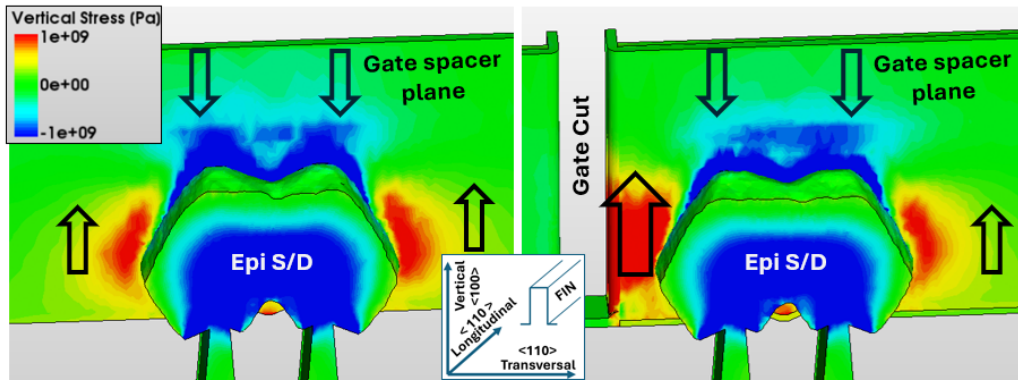


Figure 6.15: Distribution of the vertical component of the stress arising from the interaction between epitaxial growth of source and drain and the surface of the gate spacer; (left) continuous gate without cut; (right) Gate Cut close to the fin (compressive stress is shown in blue and tensile in red)

increases, while the vertical component of the stress becomes constant. The new stress state leads to an overall increment in hole mobility as the GC distance increases. Transversal stress appears to have a negligible effect on hole mobility.

Data and simulation show the same non-linear response: the drive current drops to a minimum of -13% , then recovers steadily with distance, reaching the equilibrium around fin pitch value. The simulation matches the magnitude and curvature of the trough (steep near-field slope, weaker sensitivity before fin pitch value). This confirms the mechanism discussed in the thesis: near-cut vertical stress from the gate stack depresses hole mobil-

ity, while longitudinal compressive stress from SiGe S/D re-establishes with distance, restoring p-type drive.

N-type Device Analysis

Simulation and Data

Unlike p-type, n-type transistors exhibit a monotonic dependence of the device performance on the distance to the GC (see Figure 6.16).

Stress Analysis

This is a consequence of the transversal and vertical stress perturbations that impact the channel mobility in a synergistic way (see Table 1.1). The impact of GC on the longitudinal component is negligible, as shown in Figure 6.17.

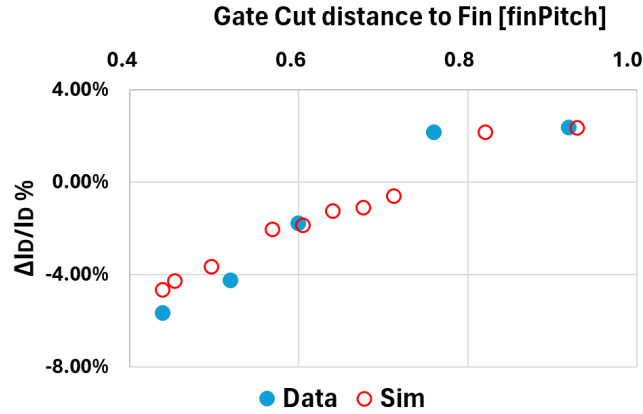


Figure 6.16: Changes in n-type drain current (in linear region @ $V_{ds} = 0.05V$) as a function of the distance from the Gate Cut boundary

The variation of the transversal and vertical stress is driven by the stress from the gate metal stack. The metal stack is fabricated through multiple steps, with each layer deposited at different temperatures ranging from $200^{\circ}C$ to $450^{\circ}C$, followed by cooling. Each of these thermal cycles generates a distinct stress profile within the device. Ultimately, the device will experience vertical compressive stress along the channel surface and transverse tensile stress perpendicular to the fin sidewall. These stress degrade electron mobility, thereby reducing n-type device performance.

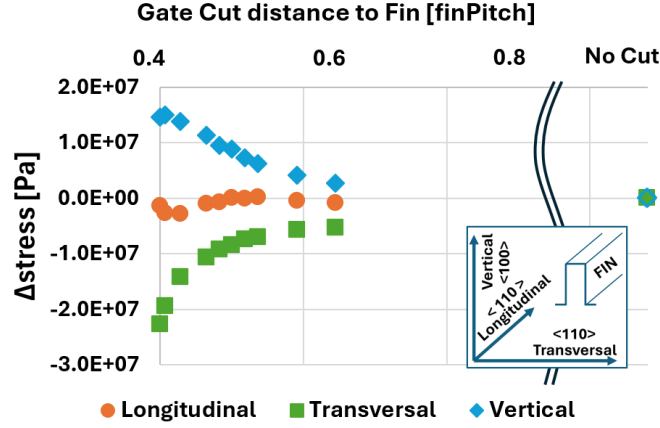


Figure 6.17: Variation of the 3D components of the channel stress as a function of the distance to the Gate Cut in n-type devices

Figure 6.17 shows that both vertical and transversal stress components intensify as the GC distance to the fin is reduced, degrading the performance.

The simulation and measured data show the same trend: drive current decreases sharply as the GC approaches the fin, with the strongest penalty within fin pitch range. The effect then saturates and additional spacing offers minimal benefit. Simulation aligns well with silicon, capturing both the magnitude and the non-linear recovery, with only minor residual differences. This confirms that vertical and transverse stress from the gate stack dominates the LLE, and proper GC spacing is critical for n-type device performance.

6.4 Poly Pitch and Stress Modulation

In FinFET technology, scaling involves modifications to both fin height and fin pitch (defined as the sum of fin width and inter-fin spacing), as well as gate length and gate pitch. Variations in Critical Dimension during the patterning process can lead to changes in width and spacing. With spacer-assisted patterning, currently the dominant technique, fin width and gate length exhibit significantly lower variability compared to their corresponding spacing.

Modern FinFET devices employ multiple performance boosters, among them the *stressors*, architectural elements which can change the stress in the

transistor channel. The most common approach is to use stress-introducing material in Source/Drain regions. In such cases, variations in fin and gate spacing alter the stress distribution in the channel region, which in turn modulates carrier mobility and leads to shifts in drive current[9].

While our analysis focuses on stress modulation as a primary contributor to performance variability, we acknowledge that in this experiment other factors, such as fin profile or contact area and resistance, can also influence device behavior. These effects are either minimized by our test methodology or considered to have a much smaller impact.

In fact, the methodology compensates for threshold voltage shifts by computing the drain current at fixed overdrive, ensuring that the extracted trends primarily reflect stress-related mobility changes. Similarly, the contribution of parasitic series resistance was quantified and compensated by the applied methodology described in Section 6.1.

We now investigate the impact of poly pitch variations on device behavior. The Poly Pitch Module (see Section 3.3.6) modulates the distance between the gates of the devices, while maintaining the same transistor gate length. To quantify the impact of gate patterning on device performance, a special set of dedicated test structures were designed. The layout Design of Experiment was built for 2-fin n-type and p-type devices, with polySpace as major factor which was systematically varied by up to $\pm 7\%$.

This variation immediately affects the volume of the epitaxial growth in the source and drain regions, maintaining the transversal cross-section of the epitaxial growth almost unchanged.

Altering the stressor volume leads to a different amount of stress induced into the device channel, and consequently, it affects performance levels. Figure 6.18 shows two devices with different gate space, where the size of the epitaxial growth increases with it.

In this technology, Spacer-Assisted Double Patterning (SADP) is employed for gate patterning [70, 71]. This is a standard technique widely adopted by leading FinFET manufacturers due to their superior Critical Dimension (CD) uniformity and low line edge roughness (LER). However, while the feature CD remains tightly controlled, the spacing between features is more prone to variability, primarily due to process-induced deviations in mandrel lithographic CD control. This variability in spacing is a key contributor to stress modulation and, consequently, to performance variation in scaled devices.

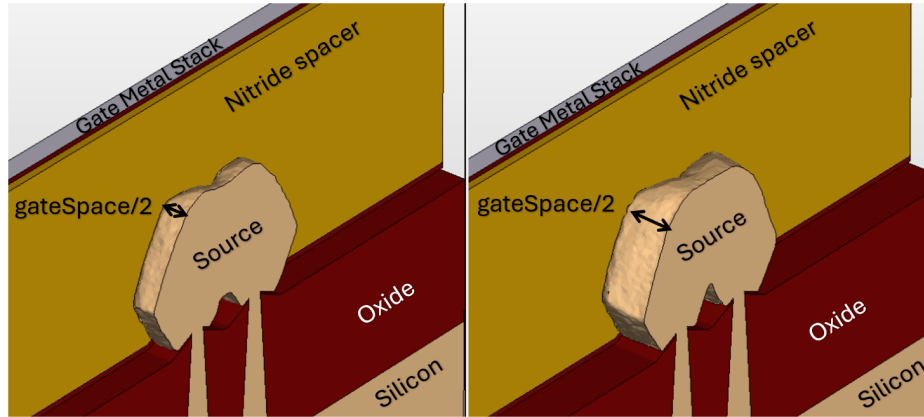


Figure 6.18: Detail of the 3D process simulation of different poly pitches with relative source epitaxial growth. The gate length of the device is fixed by process construction, while the gate space is the Design of Experiment main factor

P-type Device Analysis

Simulation and Data

Figure 6.19 presents the variation in p-type drain current (linear regime) as a function of gate-to-gate spacing (poly space). Both simulation results (open circles labeled "Sim") and experimental silicon data (closed circles labeled "Data") are shown. The results indicate a clear linear improvement in drive current with increasing gate spacing, with a total variation of approximately -12% to $+8\%$ for a $\pm 7\%$ change in poly pitch. This behavior is consistent with expectations: increasing the gate spacing enlarges the volume of the Silicon-Germanium source/drain stressors relative to the channel, thereby enhancing longitudinal compressive stress and improving hole mobility.

Stress Analysis

This is further confirmed with simulation. Figure 6.20 shows the simulated variation of the three components of stress on the p-type channel. The transversal and vertical components do not significantly change with the gate space, because the transversal cross-section of the stressors does not undergo substantial deformations. However, while Ge concentration remains constant, the volume of the SiGe Source and Drain stressor increases with poly spacing, inducing an increase in longitudinal stress and a resulting per-

formance improvement. Modulation of the S/D volume also impacts the resistance of those regions (and the interfacial area of the contacts) - the impact on R_{ext} was discussed already in Section 6.1.

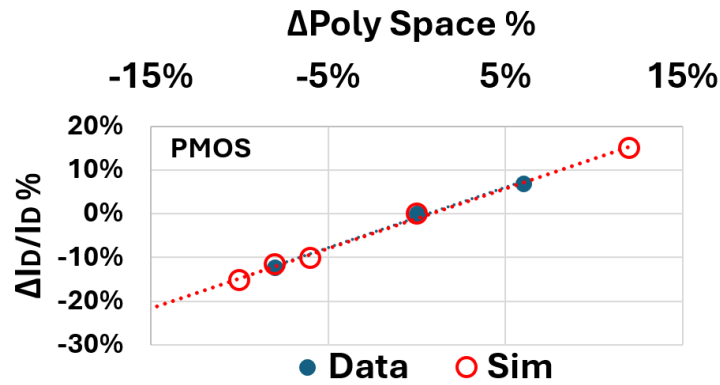


Figure 6.19: Changes in p-type drain current (in linear region @ $V_{sd} = 0.05V$) as a function of the gate space variation from the reference value of the process

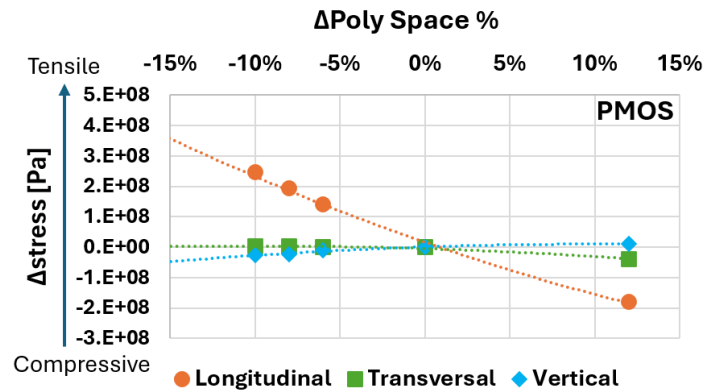


Figure 6.20: Distribution of the stress variation (referenced to the process standard) as a function of variation of the gate space for p-type FinFET; three vector components of the stress are shown: longitudinal (circle), transversal (square) and vertical (diamond)

N-type Device Analysis

Simulation and Data

Figure 6.21 illustrates the change in n-type drain current in the linear region as a function of the poly space. The performance gradually increases with larger gate space, but the sublinear behavior suggests a more complex stress interaction when compared to the p-type case. The performance change is between -11% and +3% for a $\pm 7\%$ change in gate space.

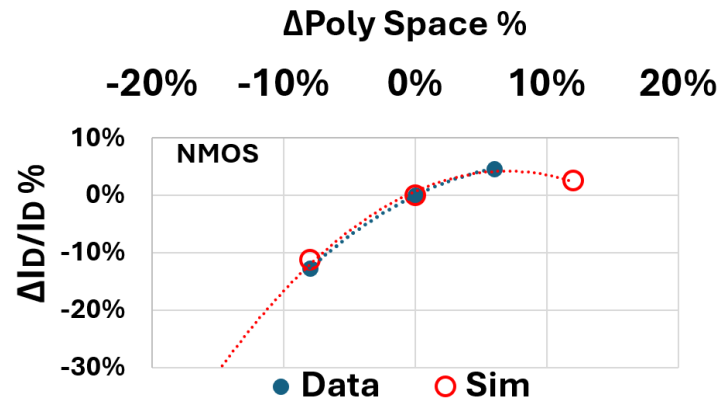


Figure 6.21: Changes in n-type drain current (in linear region @ $V_{ds} = 0.05V$) as a function of the gate space variation from the reference value of the process

Stress analysis

Figure 6.22 shows the simulated stress distribution at the channel center. The left panel presents the longitudinal stress profile across the fin cross-section, while the right panel plots the vertical and longitudinal stress components along the indicated section for different poly pitch values. Because stress along the fin can exhibit complex local variations, as evident in Figure 6.22, this work reports stress values as averages integrated over the entire channel region.

The analysis of the stress components is presented in Figure 6.23. Both the longitudinal component, and, to a lesser extent, the vertical component, increase approximately linearly with poly space. The overall impact on mobility is primarily driven by the longitudinal component, which enhances

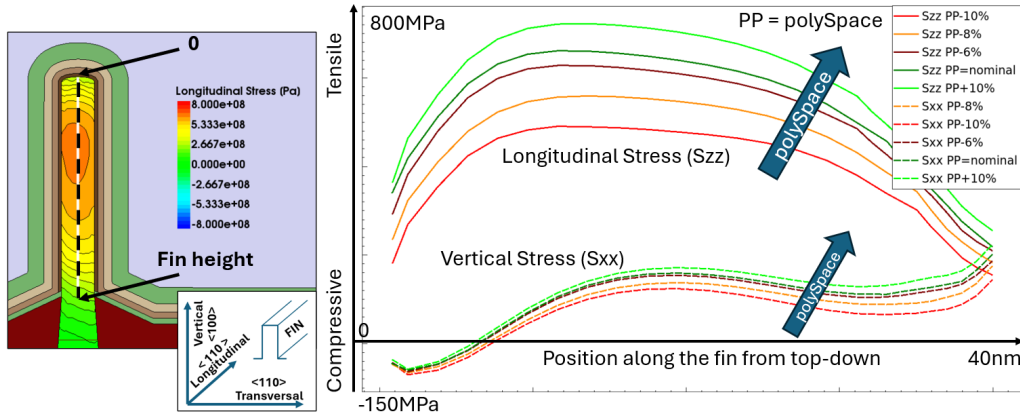


Figure 6.22: Simulated stress profiles at the channel center in n-type devices. Left: fin cross-section showing longitudinal stress distribution. Right: plot of vertical and longitudinal stress along the dashed section on the left for different poly pitches

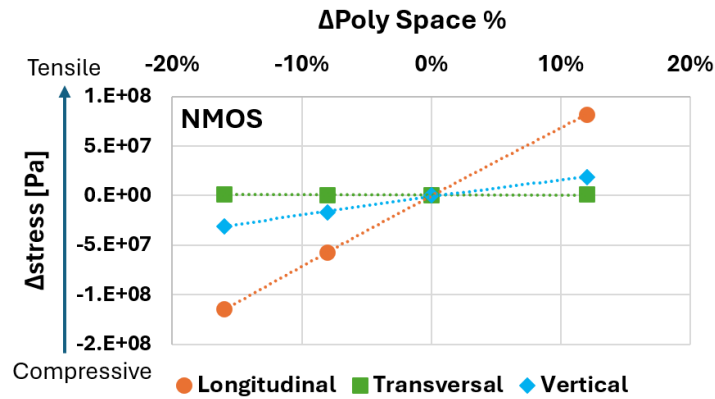


Figure 6.23: Distribution of the stress variation (referenced to the process standard) as a function of variation of the gate space for n-type Fin-FET; three vector components of the stress are shown: longitudinal (circle), transversal (square) and vertical (diamond)

device performance in a linear fashion. In contrast, although the vertical component exhibits a smaller variation, it has a stronger influence on electron mobility [9], leading to a sublinear trend in the current, as illustrated in Figure 6.21.

Since compressive stress degrades electron mobility, SiGe is not used for n-type Source and Drain regions. Instead, n-type devices typically utilize sil-

icon epitaxy, which introduces minimal stress and maintains favorable electron transport characteristics. Consequently, the causes of stress variations must be sought in a different process step other than epitaxial S/D growth.

In n-type devices, stress modulation is primarily driven by tungsten contact deposition. To capture this effect, the TCAD model includes a thermo-mechanical simulation step after tungsten fill, assuming deposition at 450°C followed by cooling to 25°C. The stress tensor is computed based on the thermal expansion mismatch between tungsten and silicon, using linear elastic constants for both materials. Poly pitch variation alters the contact area between tungsten and silicon, which changes the magnitude of tensile stress transferred to the channel.

To validate this hypothesis, a TCAD ablation simulation was performed by disabling tungsten-induced stress. This was achieved by setting the Coefficient of Thermal Expansion of tungsten equal to that of silicon in the process simulation, effectively nullifying the stress contribution during cooldown. When Tungsten stress was removed, the CPP trend became a monotonic degradation of performance, as expected from pure spacer and epitaxial geometry effects. This confirms that tungsten cooling-induced stress is the dominant source of longitudinal tensile stress in the channel, which counteracts the expected stress benefits from layout scaling.

Residual longitudinal tensile stress from tungsten fill (typically 0.8-1.5 GPa after cool-down) was also reported to modulate channel strain in Fin-FETs [72, 73].

This stress effect is more pronounced in n-type devices, where it becomes a significant factor in performance modulation. In contrast, for p-type devices, the stress impact from tungsten is relatively minor compared to the dominant influence of SiGe-induced stress.

The observed changes in drive current are very significant for both n-type and p-type transistors. In order to avoid high variability of devices and negative impact on product performance, the process control needs to ensure low variability of the gate pitch.

6.5 Fin Pitch and Patterning Variability

Similar to gate spacing, the spacing between fins, Fin Pitch (Fp), affects the shape and volume of stressors. This section analyzes how fin pitch variations influence stress and performance.

The Fin Pitch Module, described in Section 3.3.5, modulates the distance between the fins of the device.

Figure 6.24 displays examples of 3D structure generated by the model, featuring devices with F_p , and illustrates how the epitaxial growth varies in a two-fin device as a function of F_p . As the fin pitch (F_p) increases, the overall volume of the stressor region also increases, and its shape evolves, tending toward the separation of the epitaxial diamond-like shapes associated with each fin.

Intuitively, one might expect that this volume change directly correlates with the magnitude of stress variation in the channel. However, it is not only the volume but also the morphological transformation of the stressor that plays a critical role. Simulation shows that changes in the stressor shape can significantly alter the spatial distribution of stress within the device channel, potentially leading to non-linear or unexpected effects on carrier mobility and device performance.

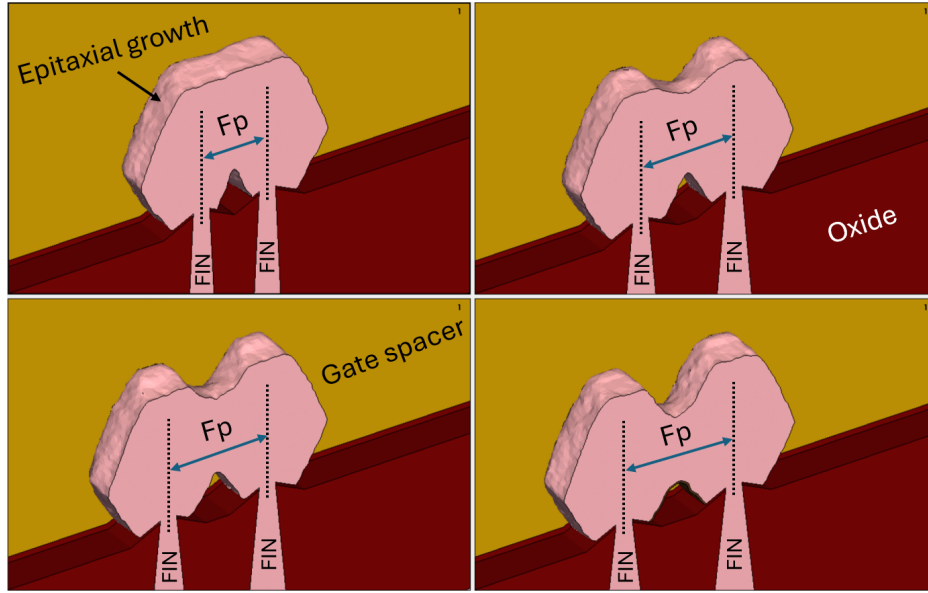


Figure 6.24: 3D process simulation of different fin pitches ($\pm 20\%$ from the nominal process) with relative source epitaxial growth

To quantify the impact of the fin patterning to the performance, structures were designed specifically for this purpose. Test structures for 2-fin and 4-fin NFETs and PFETs with fin pitches varying by up to $\pm 10\%$ were manufactured using a 7nm foundry FinFET technology, using the layout

illustrated in Section 3.3.5.

In this study, we simulated the mechanical stress in the channel of the devices as a function of the fin pitch. The results of the simulation were then used to model the impact on transistor performance due to the modulation of carrier mobility.

The realization of this particular structure is typically not feasible with a standard Process Design Kit. It was necessary to use reserved design layers, specially enabled by the foundry, to modify the relative position of the fins.

N-type Device Analysis

In advanced CMOS technologies, the source and drain regions of p-type and n-type transistors are customized individually to achieve specific performance objectives. Variations in the shape or volume of these stressors result in changes in the stress applied to the channel, thereby affecting the overall device performance.

Simulation and Data

Figure 6.25 illustrates the change in n-type drain current in the linear region as a function of F_p . The plot displays both the simulated electrical performance (open circles marked "Sim") and the measured electrical results on silicon (closed circles marked "Data").

The data shows a linear dependence of the n-type transistor drive current within the $\pm 10\%$ of nominal fin space change ($\sim 2\%$ current change), and simulation indicating the same linear dependence within the range extended to $\pm 20\%$.

Stress Analysis

Figure 6.26 presents the simulation results for n-type FinFET, illustrating the variation of the three stress components in the channel as a function of fin pitch. As F_p increases, the volume of the stressor also increases, leading to gradual longitudinal and vertical compression, along with slightly greater transversal tension.

The results indicate that the change in the transversal stress component is very small, making its impact negligible. On the other hand, the longitudinal and vertical components exhibit a counterbalancing effect on mobility,

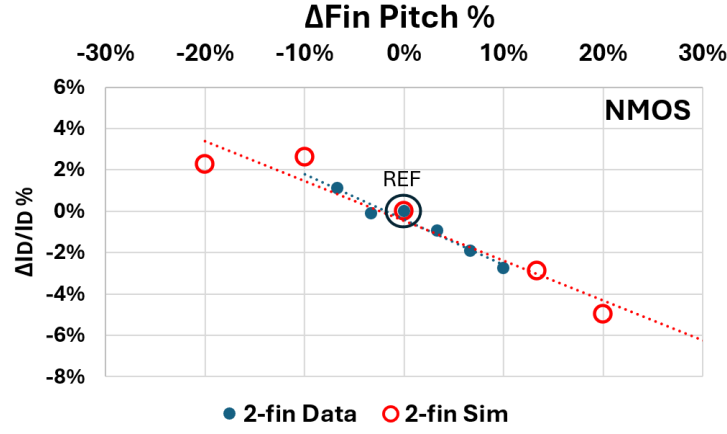


Figure 6.25: Changes in n-type drain current (in linear region @ $V_{ds} = 0.05V$) as a function of the fin pitch variation from the reference value of the process

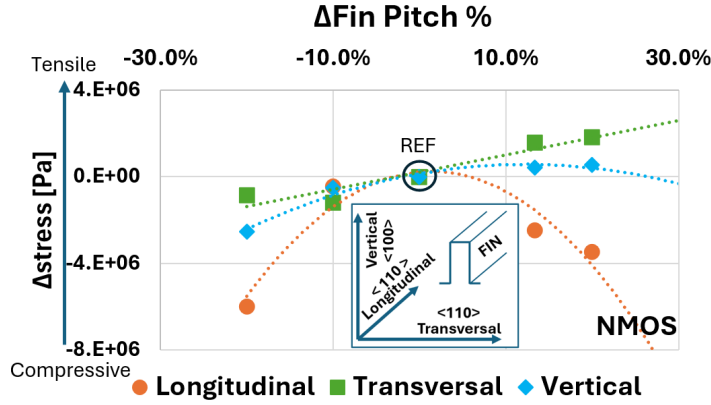


Figure 6.26: Distribution of the stress variation (referenced to the process standard) as a function of variation of the fin pitch for n-type FinFET; three vector components of the stress are shown: longitudinal (circle), transversal (square) and vertical (diamond)

as summarized in Table 1.1. The overall impact of the stress on n-type performance is a gradual degradation with increasing F_p .

Indeed, while electron mobility is sensitive to longitudinal stress, it is even more sensitive to vertical stress [9]. Therefore, device optimization should take into account all stress components.

The process appears to have been optimized in terms of fin pitch to maximize the longitudinal tensile stress in n-type devices. However, as can

be deduced from the data and simulations, this may not represent the peak performance point.

The agreement between simulation and experimental data confirms that the transistor performance degradation can be attributed to the stress change caused by inter-fin space modulation.

P-type Device Analysis

P-type devices gain performance advantages from longitudinal compressive stress induced by Silicon-Germanium stressors. As a result, variations in these stressors due to changes in fin pitch can cause significant stress perturbations, suggesting that p-type devices might be more sensitive to this effect compared to n-type devices.

Simulation and Data

Figure 6.27 illustrates both the data and simulation results for the change in p-type drain current in the linear region as a function of Fp. The performance gradually degrades with larger Fp, with the total change of $\pm 1\%$ for a $\pm 10\%$ variation in fin pitch process standards.

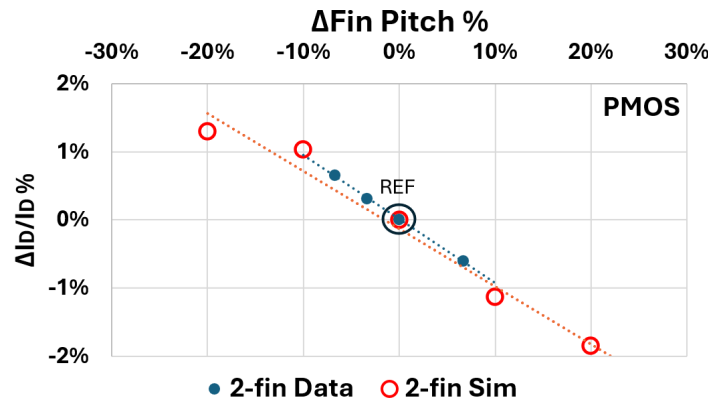


Figure 6.27: Changes in p-type drain current (in linear region @ $V_{sd} = 0.05V$) as a function of the fin pitch variation from the reference value of the process

Stress Analysis

Figure 6.28 presents the simulation results for a p-type FinFET, illustrating changes of the three stress components in the channel as a function of the

fin pitch. The results indicate that for a positive change in F_p , the increase in mobility due to the vertical component is counterbalanced by the longitudinal component. For a negative change in F_p , the the vertical component change becomes stronger, improving the overall performance.

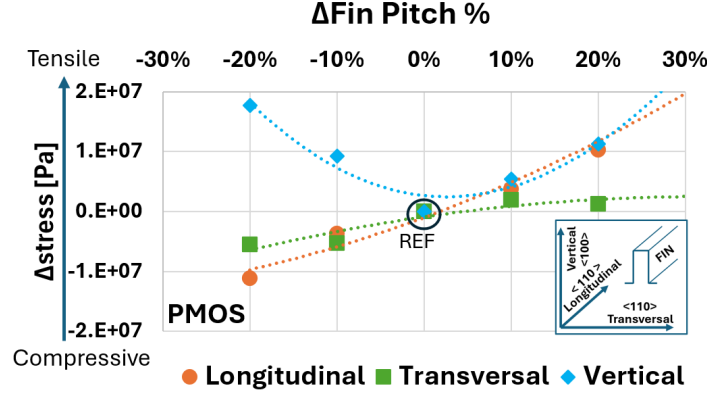


Figure 6.28: Distribution of the stress variation (referenced to the process standard) as a function of variation of the fin pitch for p-type FinFET; three vector components of the stress are shown: longitudinal (circle), transversal (square) and vertical (diamond)

The use of Silicon-Germanium as a stressor for p-type transistors amplifies stress variations by an order of magnitude compared to n-type transistors. Notably, the vertical component varies more significantly than the longitudinal component, but not enough to dominate it. This overall effect reduces the sensitivity to inter-fin space in p-type devices compared to n-type devices.

The strong data-simulation agreement confirms that stress modulation explains the behavior.

6.5.1 Spacer-Assisted Quadrupole Patterning

The standard fin patterning technique in FinFET technology nodes is the SAQP [40, 74, 75]. Figure 6.29a illustrates the principle of SAQP and the impact of the mandrel CD variability (marked by x, and exaggerated for illustration purposes). The cross-sectional micro-image in Figure 6.29b shows variation in distances between the fins resulting from the mandrel CD patterning errors.

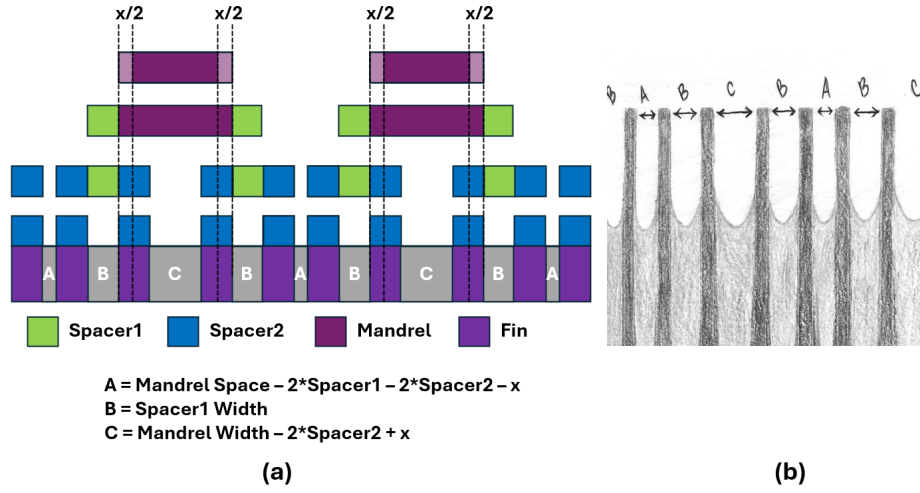


Figure 6.29: (a) SAQP pattern propagation sequence, from the top: Mandrel lithography (purple) with CD variability (x), first spacer deposition (green), second spacer deposition (blue), final Fin placement (violet); (b) micro-image of fins demonstrating variability of inter-fin space (fin "pitch walking")

The variability in the mandrel width leads to a phenomenon known as fin pitch walking, where the space between fins changes, but the sum of four subsequent pitches remains constant, as schematized in Figure 6.29a. Analyzing the effects of pitch walking on device performance is essential to understanding the overall impact of scaling on stress modulation.

The sensitivity of 2-fin and 4 fin devices (typical device sizes in standard cell libraries of 7nm node) to the pitch walking effect can be quite different. 2-fin devices modulation is always related to the model described with a single inter-fin space, which we already discussed in previous section for variable Fin pitches. Below we discuss 4-fin device, which may be impacted in different way, depending on its layout and placement.

Figure 6.29 shows three pitch-walk parameters of interest (spaces A, B, and C), linked to the lithography precision of the mandrels' width. x is the error in defining the mandrel width (exaggerated for illustration purposes). As can be deduced from the formulas shown in Figure 6.29a, we can observe that a poor mandrel CD control generates a direct mismatch between A and C, with evident effects in fin patterning on silicon (Figure 6.29b).

Single fin devices are not affected by this phenomenon by construction. In 3-fin devices, only 1 fin space is affected, rendering the device relatively

insensitive.

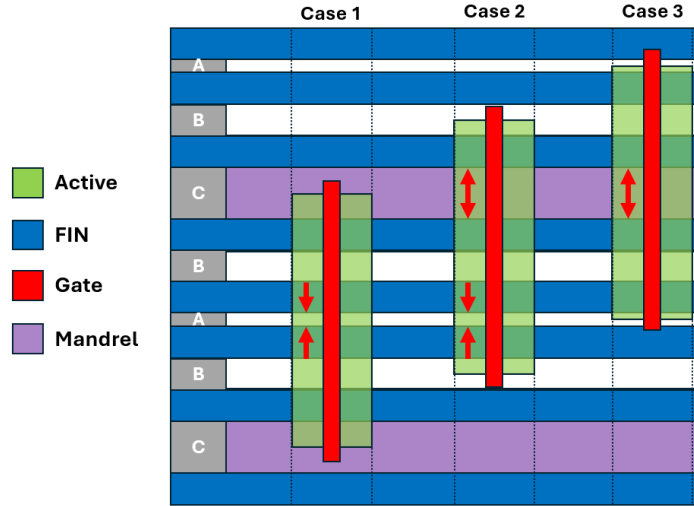


Figure 6.30: Positions in the sea of fins of a 4-fins device relative to the mandrel of a SAQP patterning technique

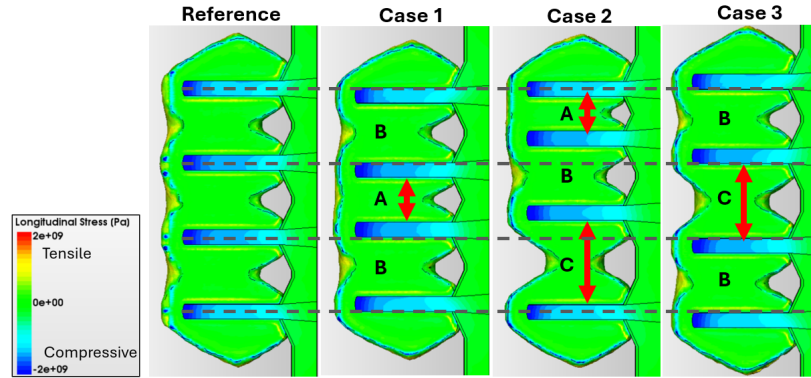


Figure 6.31: Simulated cross section across the channel of 4-fins p-type with in blue the longitudinal compressive stress, for the three possible configurations 1, 2, and 3

The impact of fin pitch walking on 4-fin transistors depends on device placement and configuration of fins used to define its location within the uniform array of fins. Figure 6.30 defines three possible positions of a 4-fin device relative to the SAQP mandrel (the reference nominal case would have all fin spaces equal). Figure 6.31 shows the simulated cross-section across

the channel with the longitudinal stress in blue and the epitaxial growth formation on source and drain which varies depending on the relative fin space.

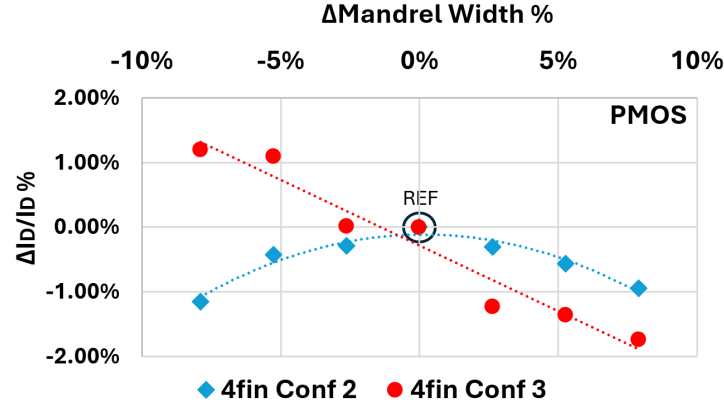


Figure 6.32: Experimental 4-fin data of the drain current change (in linear region @ $V_{sd} = 0.05V$) as a function of the variation of the mandrel width for 4-fin p-type positioned in configuration 2 and 3

From the 2-fin model, it is demonstrated that the close proximity of the fins enhances the efficiency of epitaxial growth, thereby improving performance. Consequently, Cases 1 and 3 exhibit a monotonic but opposite effect on performance due to mandrel variations influencing the fin spacing centered on the device in opposite directions (see Figure 6.29 formulas). In contrast, Case 2 exhibits parabolic behavior because the lithography error x influences two fin spaces in opposite directions within the device, creating a geometrically symmetrical structure.

Figure 6.32 illustrates the performance of a 4-fin p-type as a function of variations in mandrel width for Cases 2 and 3 as an example. This confirms the earlier analysis and demonstrates that Case 3 is less sensitive to the fin pitch walking effect by construction.

6.6 Fin Shielding Effect

Most of the FinFET transistors use at least two fins to provide the required performance and reduce variability of a single-fin channel.

The study in this work is focused on two-fin transistors because the technology does not allow single-fin devices. For Diffusion Break, Fin Pitch and

Poly Pitch effects, the stress variations are identical for both fins, since the isolation impact is the same for both by construction. This gives the same stress and performance modulation regardless of the number of fins in the device.

However, in the case of the single-sided GC, the structure is asymmetrical, and the two fins of the transistor have different sensitivities to the GC [76].

For n-type devices, the primary stress components affecting their performance are vertical and transverse, both of which are mainly influenced by the morphology of the gate metal stack. The fin closest to the Gate Cut is more affected due to the significant reduction in the volume filled by the gate metal stack between the fin and the cut. Conversely, the second fin does not experience significant changes.

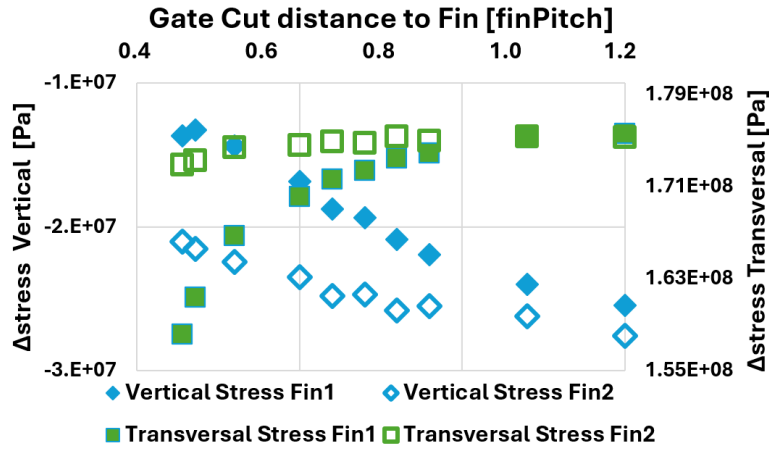


Figure 6.33: Profiles of vertical and transversal stress in n-type devices showing the change between stress inside the outer fin for the case with GC compared to continuous gate

Figure 6.33 illustrates the variation of vertical and transverse stress in n-type devices for both fins separately. It is observed that the fin farther from the Gate Cut is minimally affected by stress variations.

Due to this shielding effect, the impact on the device performance decreases for the devices with a larger number of fins.

6.7 Conclusions

Strain engineering plays a pivotal role in enhancing the performance of modern FinFET devices. This chapter has explored how localized layout effects, specifically diffusion and gate isolations, and variations in poly and fin pitch, modulate mechanical stress within the transistor channel, thereby influencing electrical behavior. Through the use of a calibrated 3D TCAD model applied to a 7nm FinFET technology, the study achieved strong agreement between simulated stress distributions and experimental silicon data. This validation confirms the model's reliability in predicting the impact of layout-induced stress on both p-type and n-type devices.

The analysis revealed that p-type transistors are particularly sensitive to stress variations introduced by single and double diffusion breaks, with performance changes reaching up to 12% and 8%, respectively. These effects are primarily driven by the compressive stress induced by silicon-germanium in the source and drain regions. In contrast, n-type devices exhibited smaller performance shifts, typically under 5%, but displayed a non-monotonic response due to the interplay of stress relaxation and dielectric fill effects during fabrication.

Gate cuts were shown to significantly affect device behavior as well. In p-type devices, the interaction between vertical and longitudinal stress components led to complex, non-linear performance trends, especially when the Gate Cut was in close proximity to the channel. For n-type devices, the degradation in performance followed a more predictable, monotonic trend, largely influenced by vertical and transverse stress originating from the gate metal stack.

Variations in poly pitch and fin pitch further demonstrated the importance of stressor volume and geometry. Increasing the gate space enhanced the stressor volume, resulting in improved performance for both device types, though the trend was linear for p-type and sublinear for n-type devices due to opposing stress effects. Fin pitch changes, on the other hand, altered the shape and distribution of stressors, leading to performance degradation, particularly in n-type devices where vertical stress had a more pronounced impact.

The study also highlighted the fin shielding effect, where asymmetrical structures such as single-sided Gate Cuts caused uneven stress distribution across multiple fins. This effect was more pronounced in devices with fewer

fins and diminished as the number of fins increased, thereby reducing overall sensitivity to stress variations.

In summary, this work underscores the critical importance of accurately modeling and managing stress in FinFET design. The insights gained provide a solid foundation for optimizing device architecture, improving layout strategies, and enhancing manufacturing yield. The validated simulation framework offers a powerful tool for design-for-manufacturing applications, enabling more robust and predictable performance in advanced CMOS technologies.

Chapter 7

Process Innovations and Impact on Performance

As semiconductor devices scale, process-induced variability has become a critical challenge to performance and reliability. This chapter investigates how both deterministic process innovations and stochastic manufacturing variations impact FinFET behavior. Using the calibrated TCAD framework developed in earlier chapters, the analysis explores the effects of Gate Cut strategies, gate spacer deposition techniques and thickness variability, the selection of dielectric isolation materials, and the SDB etching process window. The study also examines how operating temperature modulates stress sensitivity. This analysis bridges process integration with device performance, offering insights to guide the optimization of advanced CMOS manufacturing

Overview

This chapter investigates how process innovations, material choices, and manufacturing variability influence FinFET device behavior. Using the calibrated TCAD framework, we analyze deterministic process knobs, such as Gate Cut strategies, spacer deposition techniques, and dielectric selection, as well as stochastic variations in Single Diffusion Break trench width. These studies provide actionable insights for Design-Technology Co-Optimization (DTCO), Process Design Kit (PDK) development, and yield-aware design, bridging process integration with variability modeling.

Some of the findings presented in this chapter have been presented in the following peer-reviewed publications:

- A. Rossoni et al., “Characterization and Modeling of FinFET Performance Variability from Layout-Induced and Process-Induced Stress Modulation,” submitted to *IEEE Transactions on Semiconductor Manufacturing*, 2025.
- T. Brozek, A. Rossoni, Z. Kovacs, “Impact of Process Variation on Parametric Performance of FinFET Transistors - Stress-aware Variability Control,” submitted to *IEEE Electron Devices Technology and Manufacturing*, 2026.

7.1 Impact of Gate Cut Strategies

The fabrication of FinFETs involves several critical steps to ensure precise control over the device’s electrical characteristics. Among these steps, the poly cut process plays a pivotal role in defining the gate structure. This section delves into different implementations, discussing their impact on stress and performance variability. Each implementation is executed at different stages in the manufacturing process [55] (see Figure 7.1).

Gate Cut First

The *Gate Cut First* approach is implemented at an early stage of the fabrication flow, immediately following gate patterning and preceding the deposition of gate spacers. While this technique does not inherently introduce stress perturbations, the structural modifications it induces in the gate

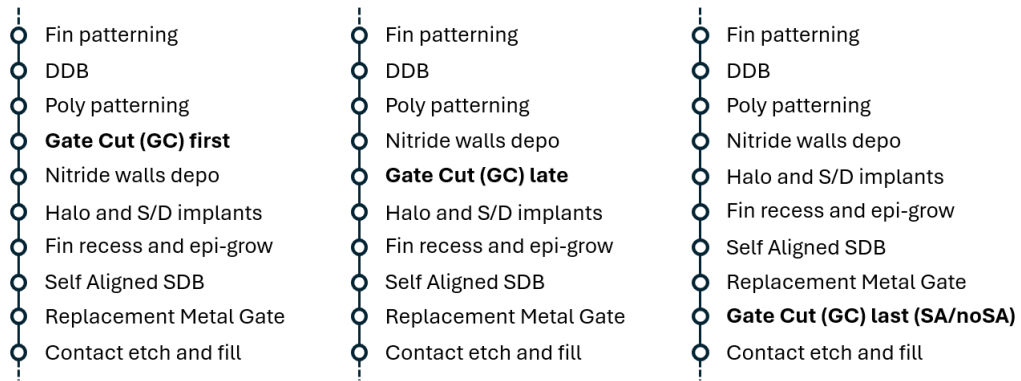


Figure 7.1: Different implementations of the Gate Cut process step order: *GC First*, during polysilicon patterning (left), *GC Late*, after the gate nitride spacers (center), and *GC Last*, after the Replace Metal Gate steps (right)

spacer geometry render the FinFET device more susceptible to subsequent stress-inducing process steps, such as the epitaxial growth of source/drain stressors.

Gate Cut Late

Conversely, the GC Late process is performed immediately after the deposition of the gate spacers in Silicon Nitride, a material harder than Silicon or Silicon Oxide. This ensures that the device structure is not altered by the proximity of the GC, resulting in performance that is less sensitive compared to the GC First process. This reflects the simulation results reported in Figure 7.3.

Gate Cut Last

In the *Gate Cut Last* approach, GC is performed after the final metal gate formation [54]. This method introduces additional complexity due to the need to remove multiple gate materials with differing etch rates.

There are two main implementations of the GC Last scheme: *Self-Aligned* and *Non-Selective*.

The first approach, referred to as the "Self-Aligned" scheme, involves removing only the gate materials while preserving the silicon nitride spacer. Although this process step is performed after the formation of the source and drain stressors, the structural framework of the transistor, defined by

the silicon nitride, remains unaffected. As a result, the device performance is expected to be only marginally influenced by the presence of the GC.

The second approach, known as the "Non-Selective" scheme, entails the removal of both the gate materials and the silicon nitride spacer. In this case, the FinFET structure is significantly altered, leading to a substantial perturbation of the previously engineered stress profile.

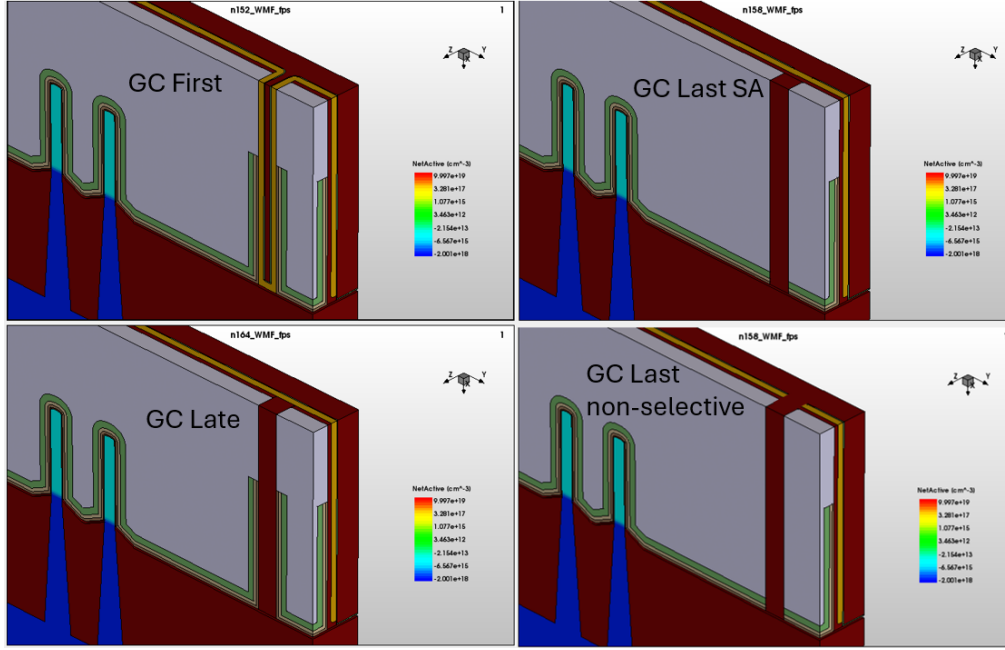


Figure 7.2: 3D process simulation results for different GC processes

Effects on Performance

The process steps required for each type of Gate Cut are well described in literature [55], along with a detailed comparison of the advantages and disadvantages of the different approaches. However, an analysis from the perspective of induced stress and, consequently, the variability in performance of the various methods is still lacking.

Using the workflow developed in this study, it is possible to compare these technologies from this perspective. Figure 7.3 depicts the comparison of the drain current variation sensitivity to its proximity and its specific implementation methodology.

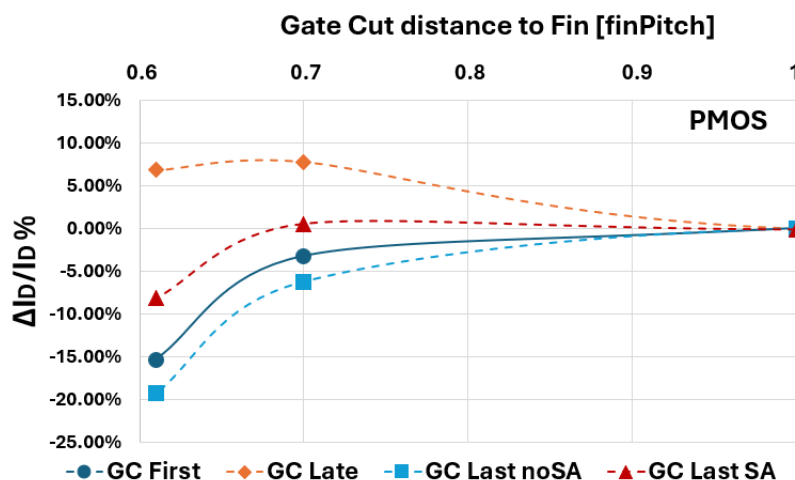


Figure 7.3: Comparison of the drain current variation in the linear region for p-type devices as a function of the distance of the Gate Cut for different process techniques

While GC strategies influence stress through structural isolation, another critical contributor to stress modulation is the integrity of the gate spacers. These spacers, typically composed of silicon nitride, not only define the gate profile but also modulate the mechanical stress that impacts carrier mobility.

The following section examines various spacer deposition methods and evaluates their implications for overall device performance.

7.2 Gate Spacer Deposition Techniques

The continuous evolution of semiconductor manufacturing has driven interest in silicon dioxide and nitride films that can be deposited at reduced temperatures, particularly for BEOL and 3D integration schemes. Lower wafer temperatures can minimize thermal ramp-up times and reduce energy consumption. However, deposition rates in most Chemical Vapor Deposition (CVD) techniques are exponentially dependent on temperature, and higher temperatures typically yield superior film quality and stress control. Therefore, process engineers must carefully balance thermal budget constraints with the need for high-performance dielectric films, especially in stress-sensitive FinFET architectures. To address these challenges, various CVD techniques have been investigated to enable high-quality film deposi-

tion at reduced temperatures, while preserving the electrical and structural integrity essential for advanced device performance [77].

The proposed model can also be employed to predict performance variations as a function of the silicon nitride deposition process used for gate spacers in transistors. Using this framework, we investigated the impact of deposition temperature across several CVD techniques, including Inductively Coupled Plasma CVD (ICPCVD), Plasma Enhanced CVD (PECVD), Low-Temperature Low-Pressure CVD (LT-LPCVD), and High-Temperature LPCVD (HT-LPCVD). Each of these methods exhibits distinct thermal budgets [78] and stress profiles, which in turn influence the mechanical stress distribution in the spacer region and ultimately affect device performance. The model enables a comparative analysis of these deposition schemes, providing predictive insights into their impact on transistor characteristics.

Low-Pressure Chemical Vapor Deposition

LPCVD is a process used to deposit thin films on semiconductor substrates at subatmospheric pressures. LPCVD is favored for its ability to produce uniform films over large substrates, making it ideal for applications in the semiconductor industry. By operating at lower pressures, LPCVD can achieve high-quality film deposition at typical temperatures between 400°C to 900°C .

Plasma-Enhanced Chemical Vapor Deposition

PECVD uses plasma to activate gas-phase reactions at lower temperatures. This technique allows for faster deposition rates and greater process flexibility, making it suitable for a wide range of materials, including silicon nitride and silicon oxide. PECVD operates at lower temperatures compared to conventional CVD (between 200°C to 350°C), which enables successful applications in situations where higher temperatures could potentially damage the device or substrate being coated.

Inductively Coupled Plasma Chemical Vapor Deposition

ICPCVD is an advanced form of PECVD characterized by higher plasma density and more uniform energy distribution. This results in superior film quality and uniformity, even at lower pressures and temperatures (less than 200°C). ICPCVD is particularly advantageous for depositing films on

complex-shaped surfaces and for applications requiring high-quality, low-temperature films.

Effects on Performance

Leveraging the model developed and calibrated, it was possible to predict the performance variability caused by the different gate spacer deposition processes. Figure 7.4 shows that the devices are affected by the performance variability up to 0.65% for n-type and 2.30% for p-type devices.

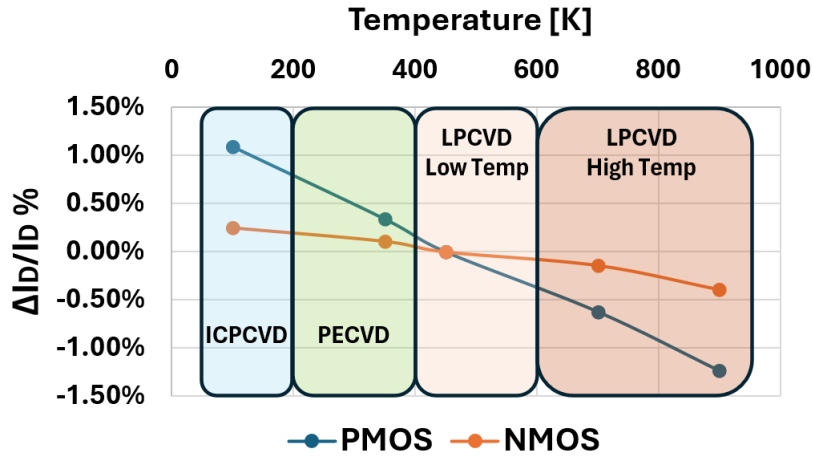


Figure 7.4: Performance change in n-type and p-type devices depending on the gate nitride spacers deposition and temperatures method

The observed performance variation across spacer deposition methods ($\leq 0.65\%$ for n-type and $\leq 2.3\%$ for p-type devices) falls within a narrow range, indicating that the importance of mechanical property of the spacer is relatively minor compared to other process knobs. However, these differences become relevant when combined with thermal budget constraints. The choices of the spacer deposition process and temperature may be dictated by manufacturing flow. However, the film stress and uniformity need to be taken into account for process window optimization to balance performance, reliability, and manufacturing efficiency.

Simulations on the sensitivity of GC proximity shows that the difference in sensitivity to the LLE depending on the deposition method is negligible (less than 0.3%).

7.3 Impact of Dielectric Isolation Materials on Performance

Following the analysis of gate spacer deposition techniques, this section extends the investigation to the role of dielectric isolation materials in modulating mechanical stress and, consequently, device performance in advanced FinFET technologies. The dielectric used to fill isolation regions, such as SDB trenches, interacts with the surrounding silicon and metal structures through thermal expansion and elastic properties, thereby altering the stress distribution in the transistor channel. These stress variations directly affect carrier mobility, making dielectric selection a critical factor in Design-Technology Co-Optimization.

In this study, silicon dioxide (SiO_2) is adopted as the reference material, given its widespread use and well-characterized mechanical properties. Multiple variants of silicon nitride (Si_3N_4)—including undoped, silicon-rich, and doped formulations (e.g., Y, Al, H)—as well as tailored stacks incorporating ZrO_2 and TiN were evaluated. Alternative dielectrics such as SiCOH, hafnium oxide (HfO_2), and zirconium oxide (ZrO_2) were also considered. Each material was characterized by its Coefficient of Thermal Expansion (CTE), Young’s modulus, and shear modulus, as summarized in Appendix B.

Table 7.1 summarizes the predicted sensitivity of I_{Dlin} for p-type and n-type devices when varying the dielectric material used to fill the SDB trench. These results are obtained from a calibrated TCAD process/device model and should be interpreted as simulation-based predictions, not yet confirmed on silicon wafers in this study.

- **N-type devices** exhibit moderate sensitivity, with performance variations typically within -10% relative to the SiO_2 baseline. Among the simulated materials, optimized Si_3N_4 formulations minimize performance degradation, whereas low- k dielectrics such as SiCOH induce the most detrimental stress profiles.
- **P-type devices** show far greater sensitivity, with performance shifts ranging from $+23\%$ to -21% compared to SiO_2 . Materials that enhance longitudinal compressive stress—such as doped or tailored Si_3N_4 stacks—deliver substantial mobility gains, while high- k oxides like HfO_2 and ZrO_2 can introduce unfavorable stress states, leading to severe performance penalties.

Type	Material	Vertical	Transversal	Longitudinal	Performance Variation
N-type	Oxide (SiO_2)	0.00E+00	0.00E+00	0.00E+00	0.0%
	Si_3N_4 (Undoped)	7.50E+06	-1.17E+06	1.22E+06	-5.09%
	Si_3N_4 (Si-rich)	8.22E+06	-8.49E+05	1.45E+07	-4.62%
	Si_3N_4 (Doped Y, Al, H)	1.60E+07	-9.57E+06	8.23E+07	-2.76%
	Si_3N_4 (Tailored ZrO_2 , TiN)	1.63E+07	-6.74E+06	6.49E+07	-3.33%
	SiCOH	-4.16E+07	3.39E+07	-2.92E+08	-11.82%
	HfO ₂	6.79E+07	-1.57E+07	1.52E+08	-3.36%
	ZrO ₂	1.12E+08	-8.90E+06	3.44E+07	-9.95%
P-type	Oxide (SiO_2)	0.00E+00	0.00E+00	0.00E+00	0.0%
	Si_3N_4 (Undoped)	4.69E+08	1.63E+08	-2.17E+08	23.40%
	Si_3N_4 (Si-rich)	4.67E+08	1.62E+08	-2.07E+08	6.08%
	Si_3N_4 (Doped Y, Al, H)	4.61E+08	1.65E+08	-2.31E+08	23.41%
	Si_3N_4 (Tailored ZrO_2 , TiN)	4.61E+08	1.63E+08	-2.22E+08	23.58%
	SiCOH	-1.70E+08	1.81E+08	-1.90E+08	-17.86%
	HfO ₂	-2.05E+08	1.79E+08	-4.24E+07	-21.18%
	ZrO ₂	-1.70E+08	1.81E+08	-1.90E+08	-17.86%

Table 7.1: Predicted ΔI_{Dlin} (%) for p-FET and n-FET as a function of dielectric material in SDB trench. Values are TCAD-based predictions

This observation aligns with prior work on CESL stress engineering, STI liner optimization, and FinFET isolation layer [79, 80, 81, 82]

These observations underscore that dielectric selection cannot be based solely on electrical properties (e.g., dielectric constant or leakage behavior). Instead, mechanical characteristics and their impact on three-dimensional stress components must be considered to achieve optimal device performance and variability control. The results clearly indicate that the optimal choice of dielectric material should differ depending on the device polarity: materials that maximize longitudinal compressive stress are preferable for p-type devices, whereas n-type devices benefit from dielectrics that minimize vertical and transversal stress components. Recent literature emphasizes that integrating material-specific stress models into TCAD and compact modeling frameworks is essential for accurate prediction of layout-dependent effects and for guiding DTCO strategies in sub-7 nm nodes [83].

7.4 SDB Etching Process Window

The SDB trench width can vary depending on the tuning of the etching process. These variations directly influence the mechanical stress distribution in adjacent devices, which in turn affects their electrical performance. Fig-

Figure 7.5 illustrates the longitudinal stress profile along the fin for two extreme trench sizes in p-type devices, highlighting how local stress states change with etch conditions. Figure 7.6 further decomposes the stress into its three principal components, longitudinal (σ_{zz}), vertical (σ_{xx}), and transversal (σ_{yy}), as a function of trench width for n-type and p-type devices.

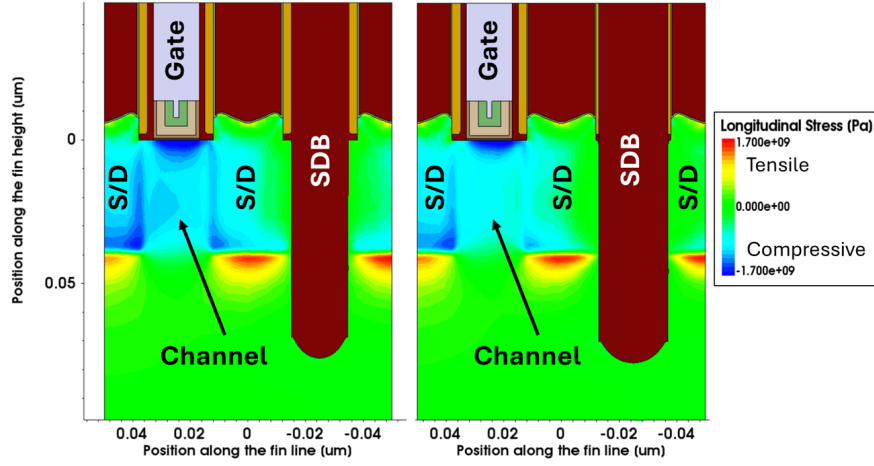
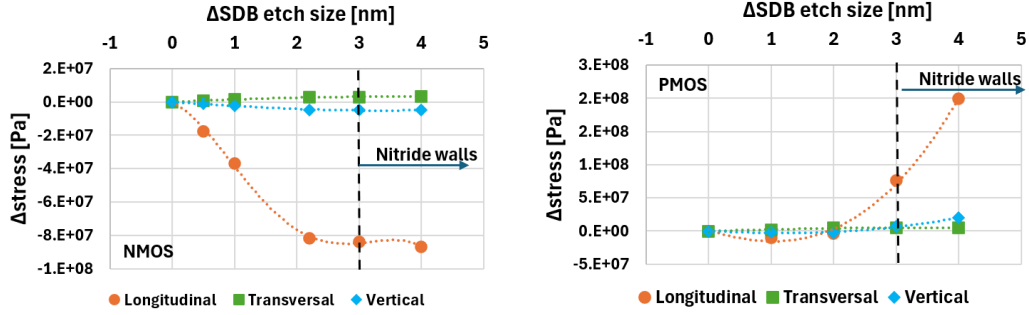


Figure 7.5: Cross-sectional view along the fin line showing longitudinal stress variation for two SDB trench widths for p-type devices: (left) small SDB size and (right) large SDB size. The color scale represents longitudinal stress in Pascals

Stress-Based Analysis

In both n-type and p-type devices, performance is primarily dominated by changes in longitudinal stress:

- **n-type devices:** The initial structural stress configuration is such that Source/Drain stressors do not significantly alter the stress state. Therefore, n-type devices are directly affected by the volume of the dielectric filling the SDB trench. Enlarging the trench increases the dielectric volume and induces additional compressive stress in the channel, which tends to degrade electron mobility
- **p-type devices:** The stress engineered by Source/Drain stressors is preserved by the structural architecture of the nitride walls, as previously analyzed in Chapter 6.3. Consequently, until the nitride gate



(a) n-type: variation of σ_{zz} (longitudinal), σ_{xx} (vertical), and σ_{yy} (transversal) vs. SDB trench width

(b) p-type: variation of σ_{zz} (longitudinal), σ_{xx} (vertical), and σ_{yy} (transversal) vs. SDB trench width

Figure 7.6: 3D stress components as a function of SDB trench width for (a) n-type and (b) p-mos devices. Components follow the coordinate convention of Fig. 5.1: x vertical (normal to wafer), y lateral (across fin), z transport (along fin)

spacers are eroded by excessive over-etch of the SDB, the overall stress remains largely unaffected. Once erosion occurs, the beneficial longitudinal compressive stress is lost, leading to performance degradation

Electrical Simulation Confirmation

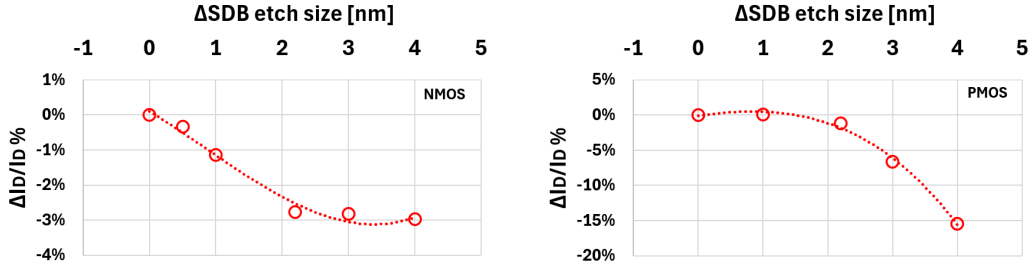
TCAD simulations confirm the mechanical analysis by quantifying the impact on drain current in the linear region (I_{Dlin}).

- **n-type devices:** Over-etching up to +3 nm reduces I_{Dlin} by approximately 3%, and the effect tends to saturate once the over-etch reaches the nitride walls, as the structural stress configuration becomes stable
- **p-type devices:** sensitivity is much higher: over-etching of +4 nm leads to a 15% degradation

Figure 7.7 compares the percentage variation of I_{Dlin} versus trench width for both device types.

7.4.1 Process Window Optimization

The combined stress and electrical analyses point to a narrow process window. Trench enlargement beyond nominal improves isolation robustness but



(a) n-type: percentage variation of I_{Dlin} vs. SDB trench width. Over-etching up to +2 nm reduces I_{Dlin} by about 3%, after which the effect saturates when nitride walls are reached

(b) p-type: percentage variation of I_{Dlin} vs. SDB trench width. Over-etching of +4 nm leads to a 15% degradation

Figure 7.7: Impact of SDB trench width on I_{Dlin} for (a) n-type and (b) p-type devices

quickly degrades p-type device performance once nitride spacers are partially eroded, due to loss of longitudinal compressive stress. Conversely, narrowing the trench improves the stress state but weakens isolation, raising leakage and reliability risk. Therefore, tight control of the SDB etch, time, bias power, plasma chemistry, and chamber pressure, is required to limit spacer erosion and trench widening while preserving isolation margins.

7.5 Impact of Spacer Thickness Variability

Although Atomic Layer Deposition (ALD) ensures exceptional precision and uniformity in spacer formation, minor thickness variations can still arise from process tolerances or equipment calibration. This section evaluates the sensitivity of FinFET device performance to small deviations in the thickness of the silicon nitride spacers surrounding the gate. Using the calibrated TCAD workflow described in Chapter 4, a parametric sweep of the spacer thickness was performed within a $\pm 10\%$ range (approximately ± 6 Å from the nominal value). The objective was to quantify the impact of these variations on the mechanical stress distribution and the resulting electrical characteristics of both n-type and p-type devices.

This analysis represents another application of the developed workflow to assess process variability, thereby supporting DTCO and PDK development.

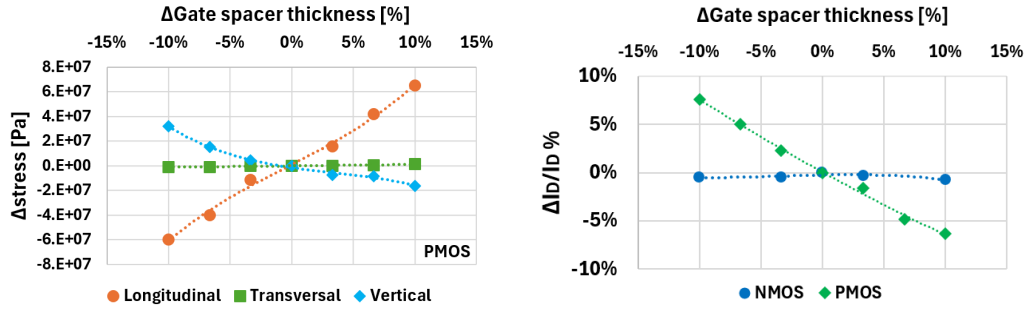
Simulation Setup

The simulation was conducted using the calibrated TCAD workflow detailed previously. The spacer material considered is silicon nitride, deposited by Atomic Layer Deposition (ALD), a technique chosen for its high precision and conformality. The nominal spacer thickness corresponds to the 6 nm baseline process flow. The analysis focused on a single-fin 7 nm bulk FinFET architecture. For each configuration, the mechanical stress distribution was examined by extracting the three principal components of the stress tensor: longitudinal, vertical, and transversal. From an electrical perspective, the primary metric was the drain current in the linear region (I_{Dlin}), measured at a constant gate overdrive voltage to isolate mobility-related effects from threshold voltage shifts.

Results and Analysis

For n-type transistors, the impact of spacer thickness variability on stress distribution and electrical performance was found to be negligible. The three principal stress components exhibited minimal changes across the variation range, and the corresponding variation in I_{Dlin} remained within the measurement noise level ($< 1\%$). This insensitivity is attributed to the relatively weak coupling between the spacer geometry and the stress propagation mechanism in n-type devices, where the Source/Drain regions are composed of silicon epitaxy without a significant lattice mismatch.

In contrast, p-type devices demonstrated a pronounced sensitivity to spacer thickness variations. Even minor deviations in spacer thickness induced structural changes in the stress propagation from the SiGe Source/Drain stressors into the channel. The simultaneous modulation of the longitudinal and vertical stress components, depicted in Figure 7.8a, resulted in coherent variations in hole mobility. This led to performance shifts from $\pm 7\%$ in I_{Dlin} for the $\pm 10\%$ thickness range, as shown in Figure 7.8b. This dependency underscores the critical role of spacer geometry in preserving the intended stress profile for p-type devices, where strain engineering is a primary performance booster.



(a) Variation of the three stress components versus spacer thickness deviation

(b) Variation of drain current as a function of spacer thickness deviation

Figure 7.8: Analysis of stress and performance variations induced by a change in the silicon nitride gate spacer thickness

7.5.1 Importance of Process Control

The observed asymmetry between n-type and p-type sensitivity highlights the importance of mechanical property of the spacer in advanced CMOS nodes. While ALD provides excellent uniformity, even small deviations can significantly impact p-type device behavior due to the complex interaction between spacer geometry, SiGe-induced stress, and channel strain. These findings suggest that:

- Spacer thickness variability should be included in DTCO analyses for p-type devices.
- Process monitoring and metrology for spacer deposition must maintain sub-angstrom precision to mitigate variability-induced performance degradation.

7.6 Temperature Effects on Stress Sensitivity

An increase in operating temperature significantly impacts the electrical behavior of silicon CMOS devices, primarily by shifting the equilibrium between carrier transport mechanisms, causing some to become more prevalent than others. As temperature rises, the intrinsic carrier concentration increases, leading to higher off-state leakage and enhanced subthreshold conduction, particularly in aggressively scaled technologies.

Elevated temperatures also intensify phonon scattering, which degrades carrier mobility. This effect is more pronounced in n-type transistors due to the higher electron effective mass and greater susceptibility to scattering.

The Subthreshold Swing (SS), with an ideal limit of 60 mV/dec at 300 K (Equation 2.6), degrades at higher temperatures as phonon scattering and interface trap dynamics increase, thereby weakening gate control. Analytical models [84, 85] and experimental studies [86] confirm this trend: between 25°C and 250°C, both n- and p-type devices exhibit significant SS degradation and mobility loss, with the n-type being more affected. These effects reduce the correlation between mobility and stress, diminishing the effectiveness of mobility-based stress monitoring at elevated temperatures.

Using the calibrated model developed in this study, the impact of Single Diffusion Break proximity was simulated at different operating temperatures.

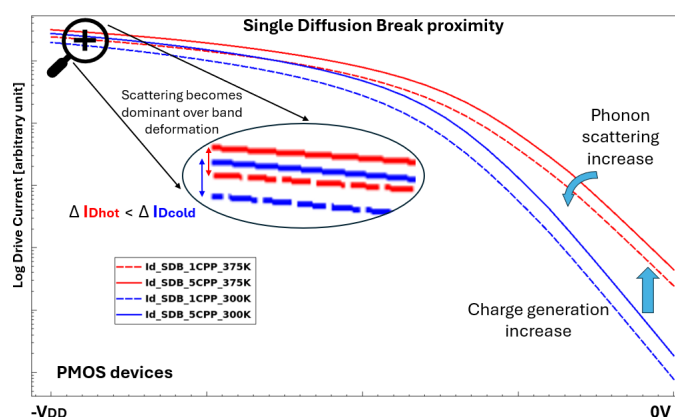


Figure 7.9: Simulated I-V curves in the linear region for p-type devices at 300K (blue) and 375K (red). The solid lines represent a device without a close SDB, while the dashed lines show a device in close proximity to SDB

Figure 7.9 presents the simulated I_D - V_G characteristics for devices with and without a nearby SDB isolation, evaluated at 300K (room temperature) and 375K (hot temperature). The simulation results align with the physical trends discussed: as temperature increases, both the drain current and the SS exhibit a noticeable rise. Conversely, the threshold voltage (V_T) decreases, and the sensitivity of the drive current to the Local Layout Effect diminishes, indicating a reduced impact of proximity-induced stress variations under thermal excitation.

Figure 7.10(a) and Figure 7.10(b) show the normalized drain current in the linear region for n-type and p-type devices, respectively, as a function of the distance from the SDB. The drain current at every isolation distance increases with temperature, which is consistent with the physical principles discussed.

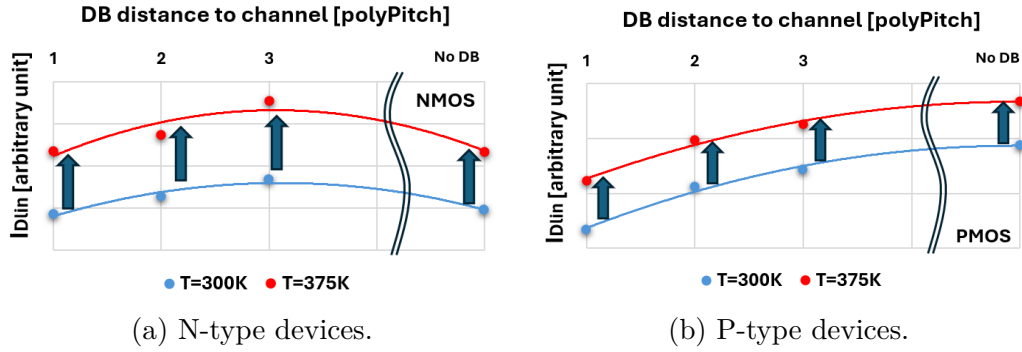


Figure 7.10: Simulated linear-region drain current for (a) n-type and (b) p-type devices as a function of distance from an SDB at 300K and 375K

Figure 7.11 depicts the variation in drain current as a function of the distance from an SDB, highlighting the temperature-dependent attenuation of LLE sensitivity. As temperature increases, the influence of the LLE on device performance diminishes. A more pronounced reduction is observed in n-type transistors, with approximately 80% attenuation, compared to about 40% in their p-type counterparts. This differential behavior underscores the greater susceptibility of electron mobility to thermally induced scattering mechanisms.

The observed attenuation in mobility sensitivity to stress variations further validates the predictive accuracy of the model. It also reinforces the hypothesis that temperature significantly weakens the correlation between mechanical stress and carrier transport, particularly in n-type devices.

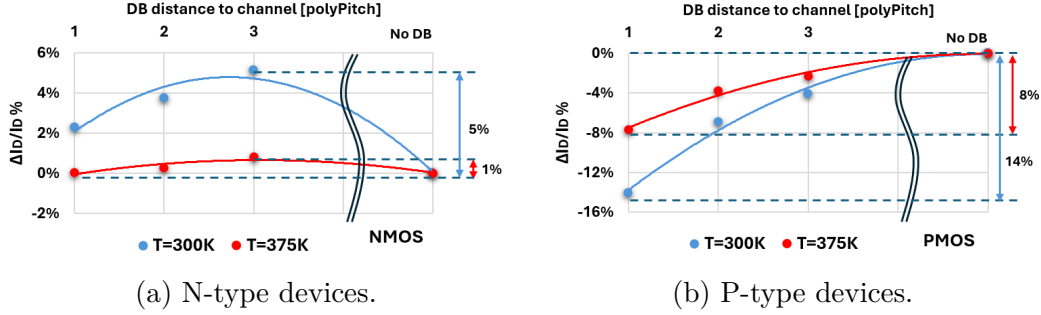


Figure 7.11: Simulated variation in linear-region drain current versus SDB proximity at 300K and 375K. Mobility sensitivity to stress decreases significantly with temperature

Summarizing, environmental conditions such as operating temperature also influence stress behavior and device variability. Understanding how temperature affects stress sensitivity is essential for robust circuit design, particularly in high-performance or thermally constrained applications.

7.7 Conclusions

This chapter demonstrated how process innovations, such as Gate Cut strategies, spacer deposition methods, dielectric material selection, and thermal conditions, affect stress distribution and FinFET performance. Using the calibrated TCAD framework, we quantified the sensitivity of n-type and p-type devices to these process variations and identified polarity-dependent trends. GC timing emerged as a dominant factor, with *GC First* showing strong performance variation, while spacer deposition and temperature effects were comparatively minor. Dielectric choice for SDB isolation proved critical, showing huge sensitivity for p-type transistor performance, depending on material properties.

Examples of these recommendations, including actionable steps for both DTCO and PDK, is provided in Table 7.2.

Process Innovation	DTCO Actions	PDK Actions
Gate Cut Strategies	Define minimum GC-to-active spacing (≥ 1 FP); prefer GC-first/late for critical paths; update cell placement guidelines.	Add LLE tables for ΔI_{dlin} vs. GC distance; include GC proximity flag in compact models; enforce DRC ¹ checks for GC spacing.
Spacer Deposition Techniques	Evaluate low-T integration impact; plan process splits for FinFET and GAA.	Provide process-specific corners; annotate stress-sensitive parameters in model cards; document spacer recipe options in PDK notes.
Dielectric Isolation Materials	Select materials with balanced stiffness/CTE for p-FET-heavy paths; combine with geometric buffers.	Create separate device models or variability factors for alternative dielectric options; update LLE coefficients in PDK.
Temperature Effects	Validate timing at hot/cold corners.	Include temperature-dependent LLE coefficients in model cards; provide thermal-aware variability tables.

Table 7.2: Summary of DTCO and PDK Actions for Each Process Innovation

Beyond deterministic process tuning, this chapter extends the framework to capture process variations, such as SDB trench width and gate spacer thickness deviations, through sensitivity analysis. This approach enables the extraction of variability bands for key electrical parameters, providing a robust foundation for yield-aware design and variability-aware PDK development. By addressing both deterministic and stochastic factors, the methodology strengthens advanced DTCO strategies for sub-7nm technologies.

¹Design Rule Check (DRC): An automated verification process that ensures an integrated circuit layout complies with all manufacturing design rules, such as minimum spacing, width, enclosure requirements and density constraints, to guarantee yield and reliability.

Chapter 8

Applications, Limitations, and Future Extensions

This chapter outlines future directions for extending the modeling framework developed in this thesis. Building on its validated application to FinFET stress-related LLEs, it proposes enhancements such as AI/ML integration, compound LLE modeling, and adaptation to Gate-All-Around architectures. These ideas aim to broaden the model's applicability in DTCO, PDK development, and variability-aware design for next-generation technologies.

Overview

This chapter outlines a strategic vision for extending the modeling framework developed in this thesis. While the current work has successfully demonstrated the predictive capabilities of a calibrated TCAD model for analyzing stress-related Local Layout Effect (LLE) in FinFET technologies, the semiconductor industry continues to evolve rapidly. Emerging device architectures, increasing design complexity, and the growing demand for variability-aware design methodologies necessitate further advancements.

To address these challenges, this chapter proposes several future directions. These include the integration of artificial intelligence and machine learning (AI/ML) to accelerate simulation workflows, the modeling of compound LLE interactions that reflect real-world layout complexity, and the adaptation of the framework to support Gate-All-Around Field Effect Transistor (GAA) architectures. Additionally, the potential to incorporate localized process variability and wafer-level gradients is explored as a mean to enhance yield-aware design and process optimization.

Together, these extensions aim to transform the current framework into a scalable, intelligent, and versatile toolset that supports Design-Technology Co-Optimization (DTCO), Process Design Kit (PDK) development, and advanced variability modeling in next-generation semiconductor technologies.

8.1 Applications

8.1.1 Design-Technology Co-Optimization

As described in Section 1.7.4, in advanced CMOS nodes, DTCO plays a critical role in sustaining performance and density improvements. The calibrated TCAD model developed in this work enables designers and technologists to evaluate how specific layout features, such as gate cuts, diffusion breaks, and fin or poly pitch variations, affect device performance through stress modulation. By simulating these effects in a controlled environment, it becomes possible to identify layout configurations that minimize performance degradation or variability.

For instance, the model can predict that placing a p-type FinFET too close to a single diffusion break may reduce drive current by over 10%, guiding layout adjustments early in the design phase. The developed model pro-

vides valuable insights into the local interactions between layout and process, which are essential for cell-level optimization.

8.1.2 Process Design Kit Development

The predictive capabilities of the model can be leveraged to enhance the accuracy and completeness of PDK. By simulating the electrical impact of various LLEs, the model supports the generation of layout-dependent device parameters that can be incorporated into compact models or lookup tables. These enriched models allow circuit designers to better anticipate performance shifts due to layout context, reducing the risk of silicon-to-model mismatches. The model enables a more granular understanding of how local layout features influence device behavior, which is particularly valuable for analog and mixed-signal design blocks where layout freedom is greater.

Compact Model Integration of Stress-Induced LLE

One of the main challenges in DTMO is ensuring that LLEs, such as gate cut proximity or poly pitch variation, are not only analyzed at the device level but also properly reflected in circuit-level simulations. While TCAD provides accurate physics-based predictions of stress-induced variability, quantifying changes in drive current (ΔI_{Dlin}) and threshold voltage (ΔV_T), these results must be translated into the language of compact models to influence timing, leakage, and variability analysis in EDA flows.

The industry-standard model for multi-gate devices, BSIM-CMG¹ (version 111.2.1), offers a robust framework for this integration. It includes dedicated parameters for threshold voltage and mobility, as well as lateral proximity effects. Specifically, threshold voltage adjustments can be captured through parameters such as DVTP0, DVTP1, and DVTP2, which scale V_T based on local geometry. Similarly, mobility variations, often the dominant contributor to ΔI_{Dlin} under stress, are controlled by U0, UA, UB, and UC, which define low-field mobility and its degradation. The parameter LPE0 acts as a length scale for lateral proximity, allowing the model to reproduce the spatial decay of these effects.

¹BSIM-CMG: industry-standard compact model for multi-gate devices; includes parameters for stress and proximity effects.

LLE Type	BSIM Parameters	Physical Effect Captured
Gate Cut (GC)	DVTP0, DVTP1 U0, UA	Threshold voltage shift due to proximity Mobility degradation (stress-induced)
Poly Pitch (CPP)	DVTP0, DVTP1 U0, UA, UB	V_T variation with poly spacing Drive current change via mobility scaling
Fin Pitch (FP)	DVTP0 U0	Minor V_T shift from fin proximity Small mobility adjustment
Diffusion Break (DB)	DVTP2 U0, UC	V_T modulation near diffusion isolation Stress relaxation effect on mobility
Proximity Scale	LPE0	Lateral decay length for all proximity-driven effects

Table 8.1: Mapping of LLEs to BSIM Parameters and Physical Impact

From the TCAD analysis performed in this work, we can establish a practical mapping. For example, a gate cut placed very close to the active region can reduce n-type drive current by about 13% and shift V_T by approximately +6 mV. In the compact model, this translates into an increase of U0 and a positive adjustment of DVTP0 to reflect the threshold shift. Similarly, a poly pitch variation of $\pm 7\%$, which induces a 10% drop in p-type current, can be modeled by scaling U0 and setting DVTP0. Even smaller effects, such as fin pitch changes, can be captured with modest adjustments (e.g., U0 +1.5%, DVTP0 +2 mV).

The integration strategy is straightforward yet powerful: first, extract global parameters (V_{TH0} , U0, UA, UB, UC) from standard $I_d - V_g$ curves; then, fit DVTP coefficients against layout metrics like gate cut distance or poly pitch variation; finally, adjust mobility terms to reproduce the TCAD-derived ΔI_{Dlin} trends. Anchoring the lateral decay to LPE0 ensures consistency across all proximity-driven effects. Once embedded in the BSIM-CMG model card, these adjustments enable stress-aware SPICE simulations, bridging the gap between physical modeling and circuit-level design.

This approach does not require introducing new parameters or breaking model compliance: it leverages documented knobs in BSIM-CMG v111.2.1, aligning with foundry practices and EDA tool support. By doing so, it transforms TCAD insights into actionable compact-model tuning, reinforcing the predictive accuracy of timing and variability analysis in advanced technology nodes.

8.1.3 Process Optimization

The TCAD-based framework provides high-resolution insights into how specific process steps interact with layout features to modulate stress and device performance. This enables process engineers to make informed decisions during technology development and integration. Examples of process optimization studies are:

Gate Cut and Spacer Strategies The model can compare different gate cut implementations, such as *GC First* versus *GC Last*, and assess their impact on stress distribution and drive current. Similarly, variations in gate spacer deposition temperature or epitaxial growth geometry can be evaluated to identify configurations that maximize beneficial stress components, such as longitudinal compressive stress in p-type devices (see Sections 7.1-7.2).

Material Selection for Isolation Dielectric choice in isolation regions, such as the Single Diffusion Break trench, significantly influences stress distribution. Materials with different mechanical properties (e.g., SiO₂, Si₃N₄, HfO₂) alter the stress tensor $\sigma(\mathbf{r})$ and, consequently, device performance. For example, replacing SiO₂ with Si₃N₄ can improve p-type drive current by more than 20% (see Section 7.3).

8.1.4 Sensitivity to Process Variability

Beyond deterministic process optimization, variability in manufacturing introduces stochastic stress perturbations that affect device performance. The framework can quantify these effects through statistical simulations.

Sources of Variability Key contributors include trench width deviations (ΔW_{SDB}), deposition thickness variations (Δt), and thermal budget fluctuations (ΔT). These factors can be modeled using Monte Carlo or DOE sweeps to capture their impact on stress and electrical metrics.

Sensitivity Analysis The sensitivity of performance to process variations can be expressed as:

$$\text{Sensitivity Index} = \frac{\partial I_{\text{Dlin}}}{\partial(\Delta W, \Delta t, \Delta T, \dots)}$$

This approach enables extraction of variability bands for critical parameters such as drain current and threshold voltage. For instance, over-etching an SDB trench by $+0.004\ \mu\text{m}$ can degrade p-type performance by 15% (see Section 7.4).

Integration into Yield-Aware Design Variability sensitivity analysis informs DTCO and PDK development by providing statistical performance envelopes and process-variation coefficients. These insights enable robust design practices by quantifying the impact of process fluctuations, such as trench width, deposition thickness, and thermal budget, on key electrical metrics. Incorporating these results into PDK supports variability-aware design rules and improves yield predictability under realistic manufacturing conditions.

8.1.5 Impact on Device and Circuit Reliability

Beyond its primary application in performance optimization, the predictive framework developed in this work holds significant potential for enhancing device and circuit reliability assessment. While the impact of mechanical stress on aging mechanisms such as Bias Temperature Instability (BTI) and Hot Carrier Injection (HCI) has been documented for planar technologies [87, 88], where lattice strain directly modulates the activation energy for bond dissociation, the available studies remain relatively limited in scope. For emerging 3D architectures, such as FinFETs and nanosheet GAA devices, these interactions require a far more in-depth and systematic investigation to fully understand their implications on reliability.

In these advanced 3D architectures, global strain models are often insufficient to capture degradation trends because reliability is increasingly dominated by localized stress gradients effects. By providing a spatially resolved description of the full stress tensor — particularly in regions characterized by steep gradients, such as near Gate Cuts and Diffusion Breaks — this model enables the identification of local *reliability hotspots* that standard layout-agnostic models overlook. Integrating these stress profiles into reliability simulation flows allows for a more physics-based prediction of time-dependent degradation (aging) and dielectric breakdown (TDDB), thereby supporting the design of robust circuits that maintain high yield and performance over their operational lifetime.

8.2 Limitations of the Current Framework

While the developed TCAD-based modeling framework demonstrates strong predictive capabilities for stress-related LLEs, several limitations must be acknowledged to contextualize its applicability and guide future improvements.

8.2.1 Assumptions in Process Parameters

Due to the proprietary nature of the commercial 7nm FinFET process used in this study, several assumptions were made regarding material properties, layer thicknesses, and thermal budgets, as described in Section 4.3. While these assumptions were informed by literature and predictive PDKs, they may not fully capture subtle process-specific effects. This introduces uncertainty in the absolute accuracy of the model, particularly in stress-sensitive regions such as gate spacers and epitaxial interfaces. A more critical limitation is the lack of access to foundry-specific calibration data, which could lead to systematic deviations in stress modeling or mobility extraction.

8.2.2 Applicability to Other Nodes or Foundries

The model presented in this work was originally calibrated using a specific 7nm FinFET process and a dedicated layout test chip. While the underlying methodology is broadly applicable, direct transferability to other technology nodes (e.g., 5nm, 3nm) or foundries (e.g., Intel, TSMC, Samsung, GlobalFoundries) may be influenced by differences in process integration, material stacks, and stress engineering strategies, such as the use of SiGe versus SiC, or variations in gate cut implementation.

Nevertheless, the modeling framework developed herein is fully parameterized and modular, enabling seamless scalability across technology nodes and fabrication environments. Device geometries, material properties, and thermal budgets can be independently configured, thereby facilitating rapid adaptation to alternative process flows.

Although this flexibility mitigates concerns regarding generalizability, accurate deployment on new platforms still requires node-specific calibration. The framework's structure supports efficient reconfiguration, making it well-suited for multi-node DTCO and layout-aware compact modeling.

8.2.3 Impact of Measurement Noise and Tester Limitations

Despite the high accuracy and repeatability of the test infrastructure, measurement noise and tester limitations can still introduce uncertainty into the calibration process. For example, while the standard deviation of extracted parameters such as threshold voltage and drive current was maintained below 1.6%, even this relatively low variability can affect the precision of model fitting.

Moreover, the resolution limits of the parametric tester, along with systematic offsets, such as contact resistance, pad parasitics, or probe-induced artifacts, may bias the extracted electrical characteristics. These effects are particularly critical when calibrating stress-sensitive metrics like subthreshold swing or carrier mobility, where small deviations can lead to significant discrepancies in the modeled behavior.

A particularly challenging region is the low-current portion of the I_D-V_{GS} curve, near the off-state. In this regime, the measured drain current approaches the noise floor of the tester, making the data increasingly unreliable. As a result, the calibration of the model in this region is inherently weak, and any conclusions drawn from it must be interpreted with caution. This limitation is especially relevant when analyzing leakage currents or subthreshold conduction, where accurate modeling is essential for low-power design.

To mitigate these issues, median values were used to represent electrical measurements, reducing the influence of outliers and local spatial variations. However, the presence of residual noise and tester-induced artifacts remains a limiting factor in achieving perfect alignment between simulation and silicon data.

8.2.4 Full Wafer-Scale Variability Modeling

The current framework is inherently local and layout-centric. It does not simulate global wafer-scale variability, such as:

- Systematic across-wafer process drifts (e.g., etch rate gradients, temperature non-uniformity, lithographic critical dimension)
- Spatially correlated lithographic distortions
- Die-to-die variation due to reticle stitching or wafer edge effects

Although the model can incorporate localized parameter emulating wafer gradients (e.g., deposition thickness, CD variation, mask misalignment), it lacks the statistical infrastructure to model spatial correlations or predict yield distributions across an entire wafer.

8.2.5 Simplified Stress Tensor Analysis

The model provides orthogonal and diagonal components of the stress tensor, but the analysis done in this work focuses on the three principal directions (longitudinal, vertical, and transversal). While this simplification has been validated against experimental data for the studied LLEs, it may not capture second-order effects or stress anisotropy in more complex 3D geometries, such as GAA or nanowire devices.

8.2.6 Static Process Flow and Geometry

The current simulation framework assumes a static process flow and fixed device geometry, which may limit its ability to capture dynamic or process-induced variations. Specifically, the model does not account for geometry distortions such as fin bending [89], gate bowing [90], or line edge roughness, which can arise from mechanical stress, etch non-uniformities, or lithographic imperfections. Additionally, time-dependent phenomena such as stress relaxation, thermal aging, and electromigration are not explicitly modeled.

Furthermore, the framework does not incorporate stochastic effects introduced by advanced patterning techniques [91], such as EUV lithography, which can lead to local variability in critical dimensions and feature placement.

Future extensions may include the integration of deformation-aware meshing, time-resolved stress modeling, and stochastic process modules to capture these effects. Coupling TCAD with finite element mechanical solvers or Monte Carlo-based patterning simulators could enhance the predictive accuracy of the framework, particularly for ultra-scaled nodes and emerging 3D architectures.

8.2.7 Computational Cost and Scalability

Although the model is efficient for single-device or small-array simulations, full-chip or library-scale analysis remains computationally prohibitive. This

limits its direct integration into standard EDA flows or large-scale DTCO pipelines without surrogate modeling or machine learning acceleration (as discussed in the next Section).

8.3 Future Extensions

8.3.1 Integration of AI/ML

Recent advances in artificial intelligence (AI) and machine learning (ML) offer promising opportunities to enhance the predictive power and efficiency of the proposed modeling framework. In particular, ML models trained on the extensive dataset generated from the Design of Experiment and TCAD simulations could be used to rapidly estimate layout-dependent stress effects without requiring full 3D simulations.

AI/ML-Enhanced TCAD-PDK-DTCO Workflow

Figure 8.1 illustrates the proposed AI/ML-enhanced workflow for integrating stress-aware modeling into TCAD, PDK development, and DTCO.

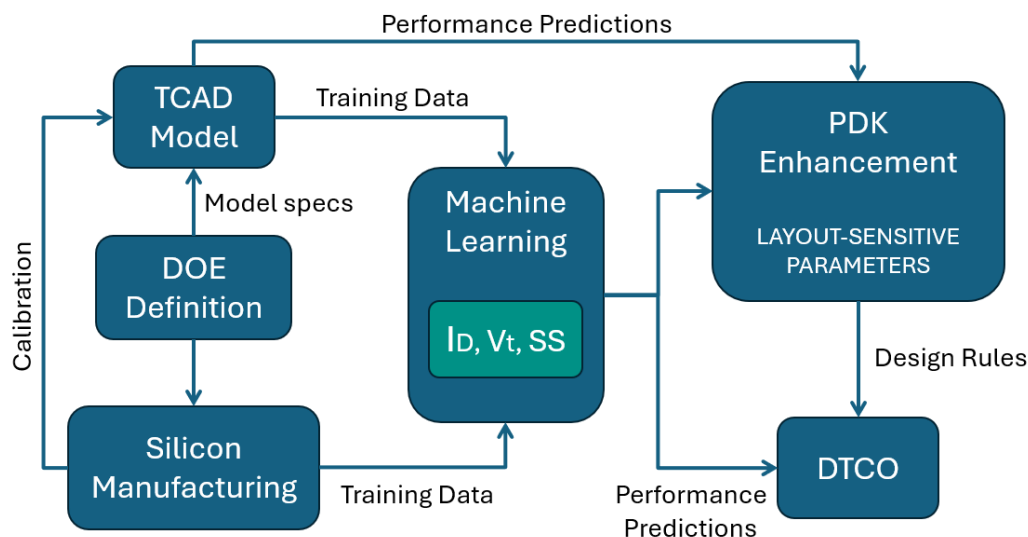


Figure 8.1: Integration of AI/ML into TCAD-PDK-DTCO workflow for stress-aware modeling and optimization

The process starts with DOE definitions, which generate structured input sets for both TCAD simulations and silicon manufacturing. The silicon

data collected from manufactured devices serves two critical roles: it helps calibrate TCAD models for improved physical accuracy and provides a rich dataset for training machine learning models.

Once trained, the ML models can predict layout-dependent performance metrics such as I_D , V_t , and SS. These predictions are then used to enhance the PDK with layout-sensitive parameters, enabling more accurate and variability-aware design rules.

The enriched PDK supports DTCO strategies aimed at minimizing variability and maximizing performance. This integrated, feedback-driven approach fosters scalable, intelligent, and variability-resilient design methodologies for advanced CMOS nodes.

This would enable faster design iterations and real-time layout optimization [92, 93].

Surrogate Modeling with Gaussian Process Regression

Another practical implementation involves training a surrogate model using Gaussian Process Regression (GPR) to approximate the output of the TCAD simulations [94]. The input features may include layout parameters such as:

- Distance to Diffusion Break (DB)
- Gate Cut (GC) proximity
- Fin pitch and poly pitch
- Device type (n/p) and V_t flavor

The output target is the simulated drain current (I_d) or threshold voltage (V_t). Once trained, the GPR model can accurately predict device performance for unseen layout configurations with minimal computational cost. This surrogate model can be embedded into EDA tools to support layout-aware PDK generation and variability-aware circuit simulation.

Reinforcement Learning for DOE Optimization

Beyond predictive modeling, AI/ML techniques can be employed to construct a consistent and optimized DOE that minimizes silicon area usage while maintaining maximum coverage of layout configuration space. Reinforcement Learning or Bayesian optimization could be used to iteratively

select the most informative layout configurations, reducing redundancy and improving test chip area efficiency [95]. This is a possible setup:

- The environment is the layout configuration space
- The agent selects layout parameters (actions) to maximize a reward function (e.g., information gain, prediction uncertainty reduction)
- The reward is computed based on how much new information a layout configuration adds to the model (e.g., via Bayesian Active Learning)

This approach can significantly reduce the number of required silicon experiments by focusing on the most informative configurations, thereby improving test chip area efficiency and accelerating technology development.

Hybrid Modeling Strategy

Integrating AI/ML into the TCAD workflow opens the door to hybrid modeling strategies. For example, Reinforcement Learning (RL) aims at inferring optimal actions in variable situation, and it could be trained to approximate TCAD outputs, enabling the prediction of a wide range of process parameters [96].

A hybrid framework could be structured as follows:

1. **Data Collection:** Use TCAD simulations and silicon measurements to build a labeled dataset.
2. **Surrogate Model:** Train a Gaussian Process Regression or neural network model to predict stress-induced performance metrics.
3. **RL Agent:** Use a policy-gradient or Q-learning agent to explore the layout space and propose new DOE points and eliminate redundancy.
4. **Feedback Loop:** Update the surrogate model with new simulation or measurement data, refining the RL policy.

This AI-enhanced workflow would enable intelligent exploration of the layout space, improving the silicon area efficiency, reducing simulation time and increasing the robustness of the predictive model.

8.3.2 Modeling Compound LLE Interactions

While the modular approach adopted in this work has proven effective in isolating and quantifying the impact of individual LLE, real-world standard cell layouts frequently include overlapping features whose combined effects, such as those from GC, DB, and fin pitch variations, may interact in non-linear and unpredictable ways. Instead, these effects can interact in complex, non-trivial ways, potentially amplifying or mitigating each other's impact on device performance.

In most cases, LLEs tend to produce additive effects on device performance. However, due to the inherently three-dimensional nature of FinFET architectures, and even more so in GAA devices, there is a strong possibility that certain LLEs may interact in a non-additive manner. For example, stress fields induced by a nearby gate cut may be modulated or shielded by the presence of a diffusion break or by variations in fin spacing, leading to unexpected compound effects that cannot be predicted by summing individual contributions.

A proposed direction for future work includes:

- Designing multi-factor DOE structures that intentionally combine two or more LLEs in controlled configurations.
- Simulating compound stress fields using the existing TCAD framework to evaluate non-linear interactions.
- Comparing compound simulations with silicon measurements to validate the presence of synergistic or antagonistic effects.
- Developing correction models or interaction terms that can be integrated into compact models or PDKs to account for these compound effects.

This extension would significantly enhance the predictive power of the model and its applicability to real-world design environments, where layout constraints often necessitate the coexistence of multiple LLE.

8.3.3 Dynamic Process Modeling and Geometry Adaptation

Future work may focus on extending the current framework to incorporate dynamic process-induced geometry variations and time-dependent effects. This includes the integration of deformation-aware meshing techniques to capture fin bending, gate bowing, and other structural distortions arising from mechanical stress or etch non-uniformities. Integrating TCAD with finite element mechanical solvers would allow for multi-physics simulations that accurately capture the interaction between mechanical stress and electrical behavior.

Additionally, the framework could be enhanced to model temporal phenomena such as stress relaxation, thermal aging, and electromigration, which are increasingly relevant in ultra-scaled nodes and reliability-critical applications. Incorporating time-resolved stress evolution would allow for predictive modeling of device degradation and performance drift over operational lifetimes.

To address variability introduced by advanced patterning techniques, such as stochastic edge placement errors in EUV lithography, Monte Carlo-based patterning simulators may be integrated. This would enable the statistical characterization of layout-dependent variability and support the development of robust, yield-aware design strategies.

These enhancements would significantly improve the fidelity and applicability of the framework, particularly for next-generation 3D architectures and sub-5nm nodes, where mechanical and stochastic effects play a dominant role in device behavior.

8.3.4 Extension to Gate-All-Around

The methodology presented in this work is expected to remain applicable to next-generation Gate-All-Around (GAA) nanosheet transistors; however, the underlying model parameters will require extensive recalibration. This is primarily due to fundamental structural and process differences between bulk-type FinFETs and GAA architectures [97]. In FinFET devices, the fin is a continuous extension of the substrate, resulting in a rigid monocrystalline structure, in which the lattice deformation in one region propagates elastically to the channel. In contrast, GAA devices employ multiple nanosheets

that become separate crystal entities during the release process, resulting in rather complex strain coupling between nanosheets, interlayered with dielectric and metal fill.

Furthermore, the amount of longitudinal stress generated in the channel can be quite different in FinFET and in GAA nanosheet devices. In FinFETs, stress modulation is achieved by fin etch-back followed by epitaxial regrowth with materials of slightly different lattice constants, preserving a monocrystalline structure and enabling effective strain engineering. In GAA devices, Source and Drain regions form as separate epitaxial nodules that merge within the cavity, creating numerous crystal plane boundaries and dislocations. This reduces the strain contribution from lattice mismatch and makes stress highly dependent on the final crystal structure, dislocation density, and thermal history. Consequently, the amount of stress experienced by each nanosheet within the same device can vary, influenced by nanosheet dimensions, spacing, inter-sheet materials, and process conditions.

Accurate modeling for GAA devices will therefore require new formulations and detailed calibration against experimental data to capture these effects. By adapting the process model to reflect the structural and material differences of GAA devices, including the vertical stacking of nanosheets or nanowires and the corresponding fabrication steps, such as the critical channel release step and inner spacer formation, the same simulation and analysis methodology can be applied. Figure 8.2 shows the proposed process flow modified to enable GAA manufacturing simulation (on the left), and a detail of the 3D GAA structure simulated by the updated process model, with nanosheets, interfacial oxide, silicon and high- k metal gate visible.

In advanced transistor architectures like FinFETs and GAA devices, the choice of wafer and channel orientation plays a crucial role in optimizing device performance. Typically, a (100) silicon wafer orientation is preferred due to its favorable surface properties and high electron mobility. For FinFETs, the channel is often aligned along the [110] direction on a (100) wafer, which provides a good balance of electron and hole mobility, making it suitable for both n-type and p-type devices. In GAA devices, which use stacked nanosheets or nanowires, the channel orientation can vary depending on the desired mobility characteristics: [100] is often chosen for n-type to maximize electron mobility, while [110] may be used for p-type to enhance hole mobility. These orientation strategies are essential for achieving high drive current, low leakage, and overall improved electrostatic control in nanoscale

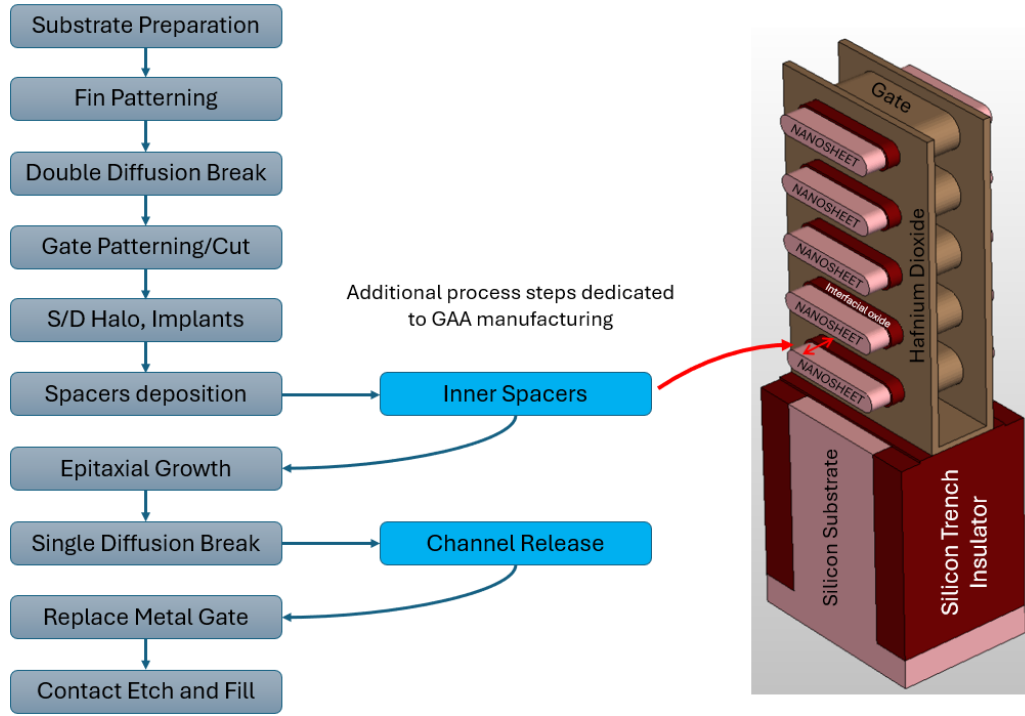


Figure 8.2: Process flow modeled in this work, with the additional the process steps dedicated to GAA manufacturing (left); a detail of the 3D GAA structure simulated by the updated process model (right)

transistors, and they need to be taken into consideration during both process and device simulation steps.

With these considerations, the developed workflow and model can be successfully adapted to the latest GAA technology nodes.

8.4 Conclusions

This section summarizes the future directions and current limitations of the proposed modeling framework. While the framework demonstrates strong predictive capability for stress-related LLE in FinFETs, it assumes static process flows, simplified stress tensor analysis, and lacks full wafer-scale variability modeling. Computational cost also limits large-scale deployment.

Future enhancements aim to overcome these constraints. AI/ML integration can accelerate simulations and enable real-time, layout-aware predictions. Modeling compound LLE interactions will capture non-linear effects

in realistic standard-cell environments, and extending the methodology to GAA architectures ensures relevance for sub-2,nm nodes. These improvements will enhance scalability, accuracy, and applicability, reinforcing the framework as a key enabler for DTCO.

By translating experimental characterization into predictive tools, the framework supports variability-aware design, process optimization, and robust manufacturing strategies for advanced FinFET and GAA technologies.

Appendix A

Summary of Physical Models Used in TCAD Simulation

This appendix provides a consolidated overview of the physical models employed in the device simulations described in Chapter 5. The models span mobility, recombination, and stress-related phenomena, and are essential for accurately capturing the behavior of FinFET devices under mechanical and electrical influences.

Category	Model Name	Purpose	Key Parameters
Mobility	Masetti Model	Models doping-dependent carrier mobility.	Doping concentration, empirical coefficients
Mobility	Canali Model	Models high-field saturation effects on mobility.	Saturation velocity, electric field, temperature
Mobility	Remote Phonon Scattering	Accounts for surface phonon and roughness scattering.	Field strength, temperature
Mobility	Inversion/Accumulation Layer (IALMob)	Models mobility in inversion/accumulation layers.	Coulomb, phonon, and surface scattering
Mobility	Piezoresistance Mobility	Models stress-induced mobility changes.	Stress tensor, piezoresistive coefficients
Recombination	Shockley-Read-Hall (SRH)	Models trap-assisted recombination.	Trap energy level, carrier lifetimes
Recombination	Auger Recombination	Models high-injection carrier recombination.	Auger coefficients, carrier concentration
Recombination	Band-to-Band Tunneling	Models direct tunneling recombination/generation.	Electric field, bandgap, tunneling coefficients
Stress	Deformation Potential Model	Models band structure changes due to strain.	Deformation potentials, strain tensor
Stress	Density of States (DOS)	Adjusts effective DOS under stress.	Effective mass, energy shifts
Stress	Stress Tensor Simulation	Captures 3D stress distribution (longitudinal, vertical, transversal).	Tensor components, spatial gradients

Table A.1: Summary of Physical Models Used in Chapter 5

Appendix B

Mechanical Properties of Dielectric Materials

Table B.1 summarizes the thermal and mechanical properties of the dielectric materials considered in the simulation study. These properties include the Coefficient of Thermal Expansion (CTE), Young’s modulus, bulk modulus, and shear modulus, which are critical for accurately modeling stress behavior in semiconductor devices. The values were obtained from a combination of literature sources and manufacturer datasheets, as indicated in the reference column.

The materials span a range of conventional and advanced dielectrics, including undoped and doped silicon nitride (Si_3N_4), hafnium oxide (HfO_2), zirconium oxide (ZrO_2), and low-k dielectrics such as SiCOH and porous SiO_2 . Tailored formulations, such as Si_3N_4 combined with ZrO_2 and TiN, were also included to explore their impact on stress distribution and device performance.

This dataset forms the basis for the mechanical simulations discussed in Section 7.3, enabling a comparative analysis of stress-induced effects across different material stacks.

Material	CTE	Young's Modulus	Bulk Modulus	Shear Modulus	Reference
SiO ₂	0.6 ppm/°C	70 GPa	35 GPa	30 GPa	[98]
Si ₃ N ₄ (Undoped)	2.3 ppm/°C	290 GPa	190 GPa	115 GPa	[99, 98, 100]
Si ₃ N ₄ (Si-rich)	4.0 ppm/°C	250 GPa	170 GPa	95 GPa	[99, 101]
Si ₃ N ₄ (Doped Y, Al, H)	4.5 ppm/°C	230 GPa	160 GPa	85 GPa	[99, 102]
Si ₃ N ₄ (Tailored ZrO ₂ , TiN)	3.5 ppm/°C	270 GPa	180 GPa	105 GPa	[99, 103]
SiCOH	1.5 ppm/°C	10 GPa	6 GPa	4 GPa	[104]
HfO ₂	5.8 ppm/°C	280 GPa	220 GPa	90 GPa	[105]
ZrO ₂	10.5 ppm/°C	200 GPa	150 GPa	80 GPa	[105]

Table B.1: Estimated Thermal and Mechanical Properties of Alternative Dielectric Materials

Appendix C

Compact-Model Integration (BSIM-CMG v111.x)

A **two-hook approach** is recommended:

- (a) **Mobility scaling:** adjust $U0, UA, UB, UC$ to reproduce ΔI_{Dlin} trends from LLE factors.
- (b) **Threshold shift (optional):** use DVTP0/1/2 for small V_T shifts when observed; keep I_{Dlin} fit dominated by mobility terms at constant overdrive.

A lateral decay length $LPE0$ may be used as a universal spatial scale for proximity terms for DB.

Recommended parameterization

- $I_{\text{Dlin}}(\text{LLE}) \approx I_{\text{Dlin}}^0 \left[1 + F_{p/n}(d_{\text{DB}}, g_{\text{GC}}, \delta_{\text{CPP}}, \delta_{\text{FP}}) \cdot S_T(T) \right]$
- $U0(\text{LLE}) \leftarrow U0^0 \left[1 + \alpha_{\text{mob}} \cdot F_{p/n}(d_{\text{DB}}, g_{\text{GC}}, \delta_{\text{CPP}}, \delta_{\text{FP}}) \right]$
- $V_T(\text{LLE}) \leftarrow V_T^0 + \Delta V_T$ (set via DVTP* only if data shows it)

Mapping summary (from dissertation, Chapter 8/Table 8.1):

- **GC proximity:** $U0, UA$ for mobility; DVTP0/1 for minor V_T if needed.
- **CPP walking:** $U0, UA, UB$ to match $\Delta I_{\text{Dlin}}(\delta_{\text{CPP}})$; DVTP0 for any systematic V_T drift.
- **FP walking:** small $U0$ tweak (linear).

- **DB proximity:** U_0, UC (stress relaxation); DVTP2 for any residual V_T proximity; $LPE0$ as decay length.
- **Temperature scaling file:** $S_T(T)$ samples for $T \in [250\text{ K}, 400\text{ K}]$ with linear interpolation.

Example model-card glue (illustrative)

Syntax may vary across simulators; adapt as needed.

Global parameters and scaling functions

Listing C.1: Global params & temperature scaling

```
* Reference temperature
.param TREF = 300

* Temperature attenuation of LLE sensitivity (linear between 300K and 375K)
.func ST_N(T) = 1 - (min(max(T,300),375)-300)/(375-300)*(1-0.2) ; ~80% reduction by 375K
.func ST_P(T) = 1 - (min(max(T,300),375)-300)/(375-300)*(1-0.5) ; ~50% reduction by 375K
```

GC proximity hooks

Listing C.2: GC proximity $\rightarrow$ mobility scaling

```
* Inputs from PEX/LDE engine (geometry extraction)
.param gGC_FP = 0.75 ; GC-to-active distance in Fin Pitches (example)

* Piecewise delta-ID/ID for p-FET (example LUT)
.func dID_REL_P_GC(x) = table(x, 0.50,-0.13, 0.75,-0.06, 1.00,0.0, 1.50,0.0)
* For n-FET (monotonic, saturating): near -5%, far +2%
.func dID_REL_N_GC(x) = table(x, 0.50,-0.05, 1.00,-0.02, 1.50,0.00, 2.00,0.02)

* Convert to mobility factor (first-order)
.param MOBSCALE_P_GC = 1 + dID_REL_P_GC(gGC_FP)*ST_P(temperature)
.param MOBSCALE_N_GC = 1 + dID_REL_N_GC(gGC_FP)*ST_N(temperature)
```

DB proximity hooks

Listing C.3: DB proximity $\rightarrow$ mobility scaling

```
* Inputs from PEX/LDE
.param dDB_CPP = 1.0
.param DBTYPE = 1 ; 1=SDB, 2=DDB

* p-FET linear slope from measurements (percent per CPP)
.param M_DB_SDB_P = 0.0293
.param M_DB_DDB_P = 0.0182
.func dID_REL_P_DB(d,t) = (t==1 ? M_DB_SDB_P : M_DB_DDB_P)*(d-1.0)

* n-FET non-monotonic: use a small LUT (example)
.func dID_REL_N_DB(d,t) = (t==1 ? table(d, 0.5,0.02, 1.0,0.05,2.0,0.01) \
: table(d, 0.5,-0.02,1.0,0.00,2.0,0.04))

.param MOBSCALE_P_DB = 1 + dID_REL_P_DB(dDB_CPP, DBTYPE)*ST_P(temperature)
.param MOBSCALE_N_DB = 1 + dID_REL_N_DB(dDB_CPP, DBTYPE)*ST_N(temperature)
```

CPP/FP walking hooks

Listing C.4: CPP & FP walking $\rightarrow$ mobility scaling

```
* Inputs from DRC/LDE
.param dCPP_pct = 0 ; in %, range [-7, +7]
.param dFP_pct = 0 ; in %, range [-10, +10]

* p-FET quasi-linear CPP fit using endpoints (-10% @ -7%, +7% @ +7%)
.func dID_REL_P_CPP(d) = table(d, -7,-0.10, 0,0.0, +7,+0.07)
* n-FET sublinear CPP (use LUT)
.func dID_REL_N_CPP(d) = table(d, -7,-0.13, 0,0.0, +7,+0.05)

* Small linear FP sensitivities
.func dID_REL_P_FP(df) = 0.001*df ; +-1% over +-10%
.func dID_REL_N_FP(df) = 0.002*df ; +-2% over +-10%

.param MOBSCALE_P_CPPFP = 1 + ( dID_REL_P_CPP(dCPP_pct) + dID_REL_P_FP(dFP_pct) ) * ST_P(temperature)
.param MOBSCALE_N_CPPFP = 1 + ( dID_REL_N_CPP(dCPP_pct) + dID_REL_N_FP(dFP_pct) ) * ST_N(temperature)
```

Applying the scale to BSIM-CMG parameters

Listing C.5: Model overrides (illustrative)

```
* Nominal
.model NCORE bsimcmg_va +U0=0.025 +UA=1e-9 +UB=1e-18 +UC=0 ... +LPE0=2e-7
.model PCORE bsimcmg_va +U0=0.010 +UA=2e-9 +UB=2e-18 +UC=0 ... +LPE0=2e-7

* Instance-level LDE wrappers (syntax varies by simulator)
* Example: scale mobility knobs multiplicatively
.param U0N_eff = 0.025 * MOBSCALE_N_GC * MOBSCALE_N_DB * MOBSCALE_N_CPPFP
.param U0P_eff = 0.010 * MOBSCALE_P_GC * MOBSCALE_P_DB * MOBSCALE_P_CPPFP

* Example of creating effective models (or use param overrides on instances)
.model NCORE_LDE bsimcmg_va +U0=@U0N_eff +UA=1e-9 +UB=1e-18 +UC=0 ... +LPE0=2e-7
.model PCORE_LDE bsimcmg_va +U0=@U0P_eff +UA=2e-9 +UB=2e-18 +UC=0 ... +LPE0=2e-7

* Optional: small VT trims if measured (DVTP*)
* .model NCORE_LDE bsimcmg_va +DVTP0=... +DVTP1=... +DVTP2=...
```


Appendix D

Reproducibility Pack

This appendix consolidates the geometry, material, and thermal assumptions used in the 3D TCAD flow together with the knobs tuned during calibration. Unless otherwise indicated, values and conventions are taken from the main text: device architecture and process assumptions (Chapter 4), simulation and calibration (Chapter 5), experimental validation (Chapter 6), process studies (Chapter 7), and model lists (Appendix A), material properties (Appendix B).

How to use this appendix

- Use Tables D.1-D.4 to construct a baseline process & device deck
- Parameters marked **Tuned=Y** were calibrated using reference (continuous active/gate) structures only (Section 5.4) and then *held* for all LLE predictions (Chapters 3-7).
- The pseudo-Sentaurus scripts in Section D provide a runnable skeleton: fill your file paths and, if needed, replace with equivalent commands for your TCAD version

Parameter	Meaning	Nominal ¹	Range used
G_l	Gate length	16 nm	Fixed
G_p	Gate pitch	56 nm	Fixed
F_h	Fin height	40 nm	$\pm 20\%$ (coarse pre-cal)
F_w	Fin width	6 nm	Fixed (SADP control)
W_{eff}	Channel width $2F_h + F_w$	$\approx 2F_h$	Derived
Wafer	Crystal plane	(100)	Fixed
Sidewall	Fin sidewall plane	(110)	Fixed
Channel	Transport direction	$\langle 110 \rangle$	Fixed

Table D.1: Geometry & orientation assumptions (FinFET for 7 nm node)

Layer	Property	Nominal	Tuned	Deposition process ²
SiO ₂ (interfacial)	Thickness t_{ox}	6 Å (Fixed)	N	Thermal ox. (900°C)
HfO ₂	Thickness	11 Å (Calculated)	N	ALD (250°C)
<i>EOT target</i>	<i>Thickness</i>	~ 7.7 Å (<i>Derived</i>)	<i>Y</i>	<i>Matches Ch. 5.4.1</i>
TiN (WF layer)	Th. & depo temp	process-spec	Y	ALD (200°C), WF trim
TiAl (barrier)	Th. & depo temp	process-spec	N	CVD (350°C)
W (gate fill)	Depo temp	process-spec	N	CVD (450°C)

Table D.2: High- k metal gate stack and EOT target

Channel Type	Module	Property	Nominal	Notes
N-type	Wells	Type / dose	B $1 \times 10^{18} \text{ cm}^{-3}$	Deep uniform implantation
	LDD	Type / dose	As process-spec	KMC ³ implant
	S/D epi	Material	Si	Minimize compressive stress
	S/D epi	Type / dose	As $1 \times 10^{20} \text{ cm}^{-3}$	In-situ LKMC ³ growth
P-type	Wells	Type / dose	As $1 \times 10^{18} \text{ cm}^{-3}$	Deep uniform implantation
	LDD	Type / dose	B process-spec	KMC ³ implant
	S/D epi	Material	Si _{0.5} Ge _{0.5}	Induce compressive stress
	S/D epi	Type / dose	B $1 \times 10^{20} \text{ cm}^{-3}$	In-situ LKMC ³ growth

Table D.3: Doping/epi assumptions for front end-of-line process layers

¹Geometries and orientation conventions as in Chapters 4.1-4.3, Figure 4.7 and Figure 5.1.

²Deposition methods, temperature, and time are assumptions derived from literature.

³See Section 4.3.5 for epitaxial, (Lattice) Kinetic Monte-Carlo usage.

Module	Step	Temp	Time	Notes
Interfacial ox	Dry oxidation	900°C	5 sec	in dry O ₂ , 0.5 atm
Epi growth	Si / SiGe	450°C	7 min	S/D stressor formation
Laser anneal	Dopant activation	1000°C	pulse	Short, localized budget

Table D.4: Key thermal steps assumptions (nominal setpoints)

Setting	Region	Notes
Min spacing	Channel edge	Coarse:2.5 Å, Medium:1.5 Å, Fine:1.0 Å
Stress tensor	σ_{ij}	Full 3D symmetric tensor propagated to device
Nonlocal mesh	Interface lines	Enabled for Schrödinger (SDevice)
Convergence	Residual norm	$\leq 10^{-6}$; max 100 Newton iterations per bias point
Bias stepping	Near V_T	Finer stepping; linear/sat regions converge in 10-20 iterations

Table D.5: Baseline meshing and solver acceptance criteria (Sections 5.1.3-5.1.4)

Parameter	Role	Prior / Range	Final (Cal.)	Tuned
Geometry (F_h)	Coarse pre-fit	From literature	Nominal	N
Effective gain factor	Align I_{Dlin}	Process-spec / $\pm 3\%$	Set per polarity	Y
Interface trap density	Match SS	Within spec window	Set per polarity	Y
EOT (High- k stack)	Gate control/leak	H_fO_2 11 Å target	Met	Y
v_{sat} stress factor	High-field fit	Default	Minor adj.	Y

Table D.6: Parameters adjusted during calibration on reference structures (Section 5.4)

Pseudo-Sentaurus decks (baseline)

The following *pseudo* decks capture the baseline process (stress) and device (electrical) stages described in Chapters 4-5. They omit foundry-specific details but preserve geometry, thermal steps, meshing, and model choices, so others can reproduce trends.

(1) Process / stress deck skeleton

```
# ----- sprocess (pseudo) -----
# Project
set ::proj "repro_baseline"
set ::tdr "$::proj/finfet7nm.tdr"

# ----- GDS import & layer mapping (pseudo) -----
# Open GDS and select top cell
gds::open file="cv7nm.gds" topcell="TOP"

# Map layers (replace <LAYER:DT> with your foundry numbers)
gds::map layer=<a:b> name=LAYER_Name
[...]

# ----- INIT SILICON -----
init wafer material=Si orientation=(100)

# Define wells from GDS
apply_mask name=NWELL op=implant species=P dose=<...> energy=<...>
apply_mask name=PWELL op=implant species=B dose=<...> energy=<...>
drive_in temperature=<...> time=<...>

# ----- FIN PATTERNING (SAQP) -----
# Mandrel pattern transfer for fins
pattern from=FIN_MANDREL type=mandrel material=SiO2 thickness=<hardmask>
deposit spacer material=Si3N4 thickness=<t1> conf=conformal
etch_back spacer anisotropic target_spacer="FIN_S1"
deposit spacer material=Si3N4 thickness=<t2> conf=conformal
etch_back spacer anisotropic target_spacer="FIN_S2"
remove mandrel method=selective

# Transfer spacers into hardmask and silicon to form fins
transfer_spacers_to_hardmask source="FIN_S2" hardmask=SiO2
etch_silicon mask=hardmask depth=40nm profile=vertical # creates fins of ~40 nm
strip_hardmask

# STI formation
form_STI depth=<...> fill=SiO2 planarity=CMP

# ----- DDB PROCESSING (EARLY) -----
# Double Diffusion Break is performed early (post-fins) to break active periodicity
apply_mask name=DDB_MASK op=etch_silicon depth=<through-active>
fill_trench material=SiO2 method=CVD densify=yes # isolation fill

# ----- GATE (POLY) PATTERNING + SPACERS -----
# Dummy gate (sacrificial poly) + SAQP for poly pitch control
deposit material=polySi thickness=<poly_thk> region=surface
pattern from=POLY_MANDREL type=mandrel material=SiO2 thickness=<hardmask>
deposit spacer material=Si3N4 thickness=<t_poly_s1>
etch_back spacer anisotropic target_spacer="POLY_S1"
deposit spacer material=Si3N4 thickness=<t_poly_s2>
etch_back spacer anisotropic target_spacer="POLY_S2"
remove mandrel method=selective
transfer_spacers_to_gate source="POLY_S2" hardmask=SiO2
etch_poly mask=hardmask stop_on=SiO2
strip_hardmask

# --- GC (configurable strategy) ---
set ::GC_STRATEGY "FIRST" # options: FIRST, LATE, LAST
if {$::GC_STRATEGY == "FIRST"} {
```

```

    gate_cut mask=GC_MASK mode=FIRST selectivity=poly_only
}

# Sidewall spacers on dummy gate
deposit spacer material=Si3N4 thickness=<t_sw>
etch_back spacer anisotropic target_spacer="GATE_SW"

# If "LATE", do GC after spacers (still before epi)
if {$::GC_STRATEGY == "LATE"} {
    gate_cut mask=GC_MASK mode=LATE selectivity=poly_only
}

# ----- LDD IMPLANTS -----
implant LDD species=As dose=<...> energy=<...> tilt=<...> for=NFET
implant LDD species=B dose=<...> energy=<...> tilt=<...> for=PFET
anneal spike temperature=<...> time=<...>

# ----- EPITAXY (IN-SITU S/D) -----
# Grow stressors and S/D, different for PFET/NFET
epi region=SD_PFET material=Si0.5Ge0.5 doping=p conc=1e20 temperature=450C
epi region=SD_NFET material=Si doping=n conc=1e20 temperature=450C
laser_anneal temperature=1000C time=<pulse> target=activation

# ----- RMG (REPLACE METAL GATE) -----
# Remove dummy poly, form HKMG
remove_dummy_gate method=selective_stop_on=Si3N4
grow_interfacial_oxide thickness=6A temperature=900C
deposit material=HfO2 method=ALD temperature=250C thickness=11A
deposit material=TiN method=ALD temperature=200C thickness=<n/p split>
deposit material=TiAl method=CVD temperature=350C thickness=<...>
fill_gate material=W method=CVD temperature=450C
CMP to planarize gate

# If "LAST", do GC now (post-RMG)
if {$::GC_STRATEGY == "LASTw/SA"} {
    gate_cut mask=GC_MASK mode=LASTsa selectivity=gate_stack_spacers
}
if {$::GC_STRATEGY == "LASTwSA"} {
    gate_cut mask=GC_MASK mode=LASTnsa selectivity=gate_stack_only
}

# ----- SDB PROCESSING (LATE) -----
# Single Diffusion Break executed late, after S/D stressors and RMG
apply_mask name=SDB_MASK op=etch_poly_silicon depth=<through-active>
fill_trench material=<SiO2|Si3N4|HfO2|SiCOH> method=CVD densify=yes

# ----- CONTACTS -----
apply_mask name=CONTACT_MASK op=open_contact dielectrics=<...> stop_on=Si/metal
deposit barrier=<Ti/TiN> seed=<...> fill=<Cu|W> method=<...>
CMP to metal level

# ----- STRESS SOLVE & EXPORT -----
mesh auto refine region=channel min_edge=1.5A
solve_stress tensor=full export_field=on
save_tdr file=$::tdr include={mesh, doping, materials, stress_tensor}
# ----- end sprocess -----

```

(2) Device / electrical deck skeleton

```

# ----- sdevice (pseudo) -----
set ::tdr "repro_baseline/finfet7nm.tdr"
set ::out "repro_baseline/ldvg_lin.log"

# Import 3D structure and stress tensor
load file=$::tdr
mesh nonlocal enable lines=interface_along_gate

# Orientation frame (x: vertical, y: lateral, z: transport)
set_axes x=vertical y=across_fin z=along_fin

# Physics models (App. A)
models \

```

```

mobility=(Masetti RPS IALMob Canali) \
recomb=(SRH Auger HurkxBTBT) \
stress=(DeformationPotential DOS Piezoresistance)

# Stress coupling
stress use_full_tensor=on
piezoresistance doping_factor=P1(N,T) second_order=P2(N,T)

# Temperature
temperature 300K

# Contacts and bias
contact gate workfunction=<n/p specific> seriesR=<...>
solve init
sweep VG from=Vmin to=Vmax step=dV at VD=0.05V VS=0V

# Convergence and stepping
newton maxiters=100 reltol=1e-6
bias_stepping region=near_VT fine_step=on

# Outputs
plot quantities=(Id Vg Vd Vs) to=$:out
save "repro_baseline/idvg_lin.tdr"
# ----- end sdevice -----

```

Run checklist

- Prepare a layer list file to import the layout into the process simulator.
- Choose one GC strategy (FIRST/LATE/LAST) per split (Section 7.1) and one SDB dielectric (Section 7.3).
- Use the **Medium** mesh (Table D.5) for production LLE curves; confirm mesh-independence on at least one point with **Fine**.
- Calibrate only on continuous active/gate references (Section 5.4), then *freeze* tuned knobs (Table D.6) for LLE predictions.

Appendix E

Publications

This research activity has led to several publications in international journals and conferences. These are summarized below.¹

International Journals

1. **A. Rossoni**, T. Brozek, S. Saxena, R. Khamankar, L. Colalongo, Z. Kovacs. “Stress-Related Local Layout Effects in FinFET Technology and Device Design Sensitivity”, *IEEE Transactions on Electron Devices*, vol. 72, Issue 5, 2025. (Special Issue: IEEE Transactions on Electron Devices) [DOI : 10.1109/TED.2025.3561974]
2. **A. Rossoni**, T. Brozek, S. Saxena, R. Khamankar, L. Colalongo, Z. Kovacs. “Stress-Related Local Layout Effects in FinFET Technology and Device Design Sensitivity”, *IEEE Transactions on Semiconductor Manufacturing*, vol. 38, Issue 2, 2025. (Special Issue: IEEE Transactions on Semiconductor Manufacturing) [DOI : 10.1109/TSM.2025.3540267]
3. **A. Rossoni**, T. Brozek, Z. Kovacs. “Impact of the Gate and Fin Space Variation on Stress Modulation and FinFET Transistor Performance”, *IEEE Transactions on Electron Devices*, [DOI : 10.1109/TED.2025.3648978]

¹The author’s bibliometric indices are the following: *H*-index = 3, *i10*-index = 3, total number of citations = 198 (source: Google Scholar on January, 2026).

International Conferences and Workshops

1. **A. Rossoni**, Z. Kovacs, L. Colalongo, R. Khamankar, S. Saxena, T. Brozek. “Modeling Stress Effects from Fin Isolation in 7nm FinFET Transistors”, *IEEE International Semiconductor Conference*, 9-11 October, 2024. [DOI : 10.1109/CAS62834.2024.10736860] - **Best Paper Award**
2. **A. Rossoni**, T. Brozek, S. Saxena, R. Khamankar, C. Hess, J. Huang, Y. Teng, Z. Kovacs, M. Quarantelli. “Layout experiments and test structures to characterize Local Layout Effects due to mechanical stress in FinFET transistors”, *IEEE International Conference on Microelectronic Test Structures*, March 24-27, 2025. [DOI : 10.1109/ICMTS63811.2025.11068908]
3. **A. Rossoni**, T. Brozek, S. Saxena, R. Khamankar, Z. Kovacs. “Impact of the Inter-Fin Space on Stress Modulation and FinFET Transistor Performance”, *IEEE International Conference on IC Design and Technology*, 23-25 June, 2025. [DOI : 10.1109/ICICDT65192.2025.11078040]
4. T. Brozek, **A. Rossoni**, Z. Kovacs. “Impact of Process Variation on Parametric Performance of FinFET Transistors - Stress-aware Variability Control”, *IEEE Electron Devices Technology and Manufacturing*, March 1-4, 2026 - **Invited Paper**

Previous related publications

1. C. Hess, S. Saxena, H. Karbasi, S. Subramanian, M. Quarantelli, **A. Rossoni**, S. Tonello, S. Zhao, D. Slisner. “Device Array Scribe Characterization Vehicle Test Chip for Ultra Fast Product Wafer Variability Monitoring”, in *IEEE International Conference on Microelectronic Test Structures*, 2007. [DOI : 10.1109/ICMTS11586.2007]
2. S. Saxena, C. Hess, H. Karbasi, **A. Rossoni**, S. Tonello, P. McNamara, S. Lucherini, S. Minehane, C. Dolainsky, M. Quarantelli. “Variation in Transistor Performance and Leakage in Nanometer-Scale Technologies”, in *IEEE Transactions on Electron Devices*, vol.55, issue 1, 2008. [DOI : 10.1109/TED.2007.911351]

US Patent

1. C. Hess, **A. Rossoni**, S. Tonello, M. Squicciarini, M. Quarantelli. “Layout for DUT Arrays Used in Semiconductor Wafer Testing”, PDF Solutions, Inc., US Patent Application US20090140762A1, filed Oct. 3, 2005, published Jun. 4, 2009. <https://patents.google.com/patent/US20090140762A1/en>

Glossary

ALD Atomic Layer Deposition

BTI Bias Temperature Instability

CDF Cumulative Distribution Function

CESL Contact Etch-Stop Layer

CMOS Complementary Metal-Oxide-Semiconductor

CPP Contacted Poly Pitch

CD Critical Dimension

CV[®] Characterization Vehicle[®]

CVD Chemical Vapor Deposition

CTE Coefficient of Thermal Expansion

DB Diffusion Break

DDB Double Diffusion Break

DFM Design for Manufacturing

DIBL Drain-Induced-Barrier Lowering

DOE Design of Experiment

DPL Double Patterning Lithography

DTCO Design-Technology Co-Optimization

DUT Device Under Test

EOT Equivalent Oxide Thickness

EUV Extreme Ultra-Violet

FinFET Fin Field Effect Transistor

FOM Figure Of Merit

GAA Gate-All-Around Field Effect Transistor

GC Gate Cut

GPR Gaussian Process Regression

HCI Hot Carrier Injection

HKMG High- k Metal Gate

IC Integrated Circuit

LDD Lightly Doped Drain

LLE Local Layout Effect

LKMC Lattice Kinetic Monte-Carlo

KMC Kinetic Monte-Carlo

LPCVD Low Pressure Chemical Vapor Deposition

MOSFET Metal-Oxide-Semiconductor Field-Effect Transistor

MP Metal Pitch

NSFET Nano-Sheet Fin Field Effect Transistor

PECVD Plasma-Enhanced Chemical Vapor Deposition

PDK Process Design Kit

PMA	Passive Matching Array
PPA	Power, Performance, Area
RL	Reinforcement Learning
RMG	Replace Metal Gate
RPS	Remote Phonon Scattering
SAP	Spacer-Assisted Patterning
SADP	Spacer-Assisted Double Patterning
SAQP	Spacer-Assisted Quadruple Patterning
SCE	Short Channel Effect
SDB	Single Diffusion Break
SEM	Scanning electron microscope
SOI	Silicon on Insulator
PD-SOI	Partially Depleted Silicon on Insulator
FD-SOI	Fully Depleted Silicon on Insulator
SRAM	Static Random Access Memory
SRB	Strained Relaxed Buffer
SS	Subthreshold Swing
STI	Silicon Trench Insulator
TCAD	Technology Computer-Aided Design
TDDb	Time-Dependent Dielectric Breakdown
TSM	Transmission Electron Microscopy
TD	Technology Development

TEM Transmission Electron Microscopy

VLSI Very Large Scale Integration

V_T Threshold Voltage

WM Wafer Map

Bibliography

- [1] S. Saini, *Low Power Interconnect Design*, 1st ed. Springer Publishing Company, Incorporated, 2016.
- [2] G. Moore, “Cramming more components onto integrated circuits,” *Proceedings of the IEEE*, vol. 86, no. 1, pp. 82–85, 1998, DOI: 10.1109/JPROC.1998.658762.
- [3] R. Dennard, F. Gaensslen, H.-N. Yu, V. Rideout, E. Bassous, and A. 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, 1974, DOI: 10.1109/JSSC.1974.1050511.
- [4] T. M. Atanackovic and A. Guran, “Hooke’s law,” in *Theory of Elasticity for Scientists and Engineers*. Birkhäuser, 2000, pp. 85–111, DOI: 10.1007/978-1-4612-1330-7.
- [5] S.M. Sze, Kwok K. Ng, *Physics of Semiconductor Devices*. John Wiley and Sons, Ltd, 2006, pp. 293–373, DOI: 10.1002/9780470068328.ch6.
- [6] S. Reggiani, M. Valdinoci, L. Colalongo, M. Rudan, G. Baccarani, A. Stricker, F. Illien, N. Felber, W. Fichtner, and L. Zullino, “Electron and hole mobility in silicon at large operating temperatures – part i: Bulk mobility,” *IEEE Transactions on Electron Devices*, vol. 49, no. 3, pp. 490–499, 2002, DOI: 10.1109/16.987121.
- [7] D. Esseni, F. Conzatti, M. De Michielis, N. Serra, P. Palestri, and L. Selmi, “Semi-classical transport modelling of CMOS transistors with arbitrary crystal orientations and strain engineering,” *Journal of Computational Electronics*, vol. 8, no. 3-4, pp. 209–224, Oct. 2009, DOI: 10.1007/s10825-009-0284-0.
- [8] F. Conzatti, M. De Michielis, D. Esseni, and P. Palestri, “Drain current improvements in uniaxially strained p-MOSFETs: A Multi-Subband monte carlo study,” *Solid State Electron.*, vol. 53, no. 7, pp. 706–711, Jul. 2009, DOI: 10.1016/j.sse.2009.02.019.

- [9] W. T. Chiang, J. W. Pan, P. W. Liu, C. H. Tsai, C. T. Tsai, and G. H. Ma, "Strain effects of si and sige channel on (100) and (110) si surfaces for advanced cmos applications," in *International Symposium on VLSI Technology, Systems and Applications (VLSI-TSA)*, 2007, pp. 1–2, DOI: 10.1109/VTSA.2007.378930.
- [10] "Understanding Silicon Wafer Orientation and Crystal Structure," <https://waferpro.com/understanding-silicon-wafer-orientation-and-crystal-structure/>, 2025.
- [11] Y. Mitani and A. Toriumi, "Re-consideration of influence of silicon wafer surface orientation on gate oxide reliability from tddb statistics point of view," in *2010 IEEE International Reliability Physics Symposium*, 2010, pp. 299–305, DOI: 10.1109/IRPS.2010.5488812.
- [12] R. Kotlyar, C. Weber, L. Shifren, S. Cea, M. D. Giles, and M. Stettler, "Effect of band warping and wafer orientation on NMOS mobility under arbitrary applied stress," *J. Comput. Electron.*, vol. 7, no. 3, pp. 95–98, Sep. 2008, DOI: 10.1007/s10825-007-0162-6.
- [13] M. Alioto, "Impact of nmos/pmos imbalance in ultra-low voltage cmos standard cells," in *20th European Conference on Circuit Theory and Design (ECCTD)*, 2011, pp. 536–539, DOI: 10.1109/ECCTD.2011.6043407.
- [14] Z. e. a. Hanson, "Exploring variability and performance in a sub-200-mv processor," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 4, pp. 881–891, 2008, DOI: 10.1109/JSSC.2008.917505.
- [15] Chen Yang et al., "Source/drain trimming process to improve gate-all-around nanosheet transistors switching performance and enable more stacks of nanosheets," *Micromachines*, vol. 13, no. 7, 2022, DOI: 10.3390/mi13071080.
- [16] M.-C. Wang, H.-C. Yang, W.-S. Liao, H.-Y. Yang, Y.-Y. Hoe, K.-H. Lin, and S.-Y. Chen, "Cesl deposition promoting n/p mosfets under 45-nm-node process fabrication," in *2010 International Symposium on Next Generation Electronics*, 2010, pp. 17–20, DOI: 10.1109/ISNE.2010.5669140.
- [17] Sato, F. et al., "Process and local layout effect interaction on a high performance planar 20nm CMOS," *Symposium on VLSI Technology Digest of Technical Papers*, 2013.
- [18] G. Wang, C. Qin, H. Yin, J. Luo, N. Duan, P. Yang, X. Gao, T. Yang, J. Li, J. Yan, H. Zhu, W. Wang, D. Chen, T. Ye, C. Zhao, and H. H. Radamson,

- “Study of SiGe selective epitaxial process integration with high-k and metal gate for 16/14nm nodes FinFET technology,” *Microelectronic Engineering*, vol. 163, pp. 49–54, Sep. 2016, DOI: 10.1016/j.mee.2016.06.002.
- [19] C.-T. Lin, Y.-K. Fang, W.-K. Yeh, C.-M. Lai, C.-H. Hsu, L.-W. Cheng, and G. H. Ma, “Impacts of Notched-Gate Structure on Contact Etch Stop Layer (CESL) Stressed 90-nm nMOSFET,” *IEEE Electron Device Letters*, vol. 28, no. 5, pp. 376–378, May 2007, DOI: 10.1109/LED.2007.895425.
- [20] Fukase, Hada, Aoki, and Kunio, “A margin-free contact process using an al/sub 2/o/sub 3/ etch-stop layer for high density devices,” in *1992 International Technical Digest on Electron Devices Meeting*, 1992, pp. 837–840, DOI: 10.1109/IEDM.1992.307487.
- [21] S. K. Selvaraja, G. Winroth, S. Locorotondo, G. Murdoch, A. Milenin, C. Delvaux, P. Ong, S. Pathak, W. Xie, G. Sterckx, G. Lepage, D. Thourhout, W. Bogaerts, J. Van Campenhout, and P. Absil, “193nm immersion lithography for high-performance silicon photonic circuits,” *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 9052, 03 2014, DOI: 10.1117/12.2049004.
- [22] K. O. Petrosyants, D. S. Silkin, and D. A. Popov, “Comparative characterization of nwfet and finfet transistor structures using tcad modeling,” *Micromachines*, vol. 13, no. 8, 2022, DOI: 10.3390/mi13081293.
- [23] D. Hisamoto, W.-C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, E. Anderson, T.-J. King, J. Bokor, and C. Hu, “Finfet-a self-aligned double-gate mosfet scalable to 20 nm,” *IEEE Transactions on Electron Devices*, vol. 47, no. 12, pp. 2320–2325, 2000, DOI: 10.1109/16.887014.
- [24] N. Shahriari and I. Corporation, “Ai era innovation matrix and intel xeon 6 soc,” <https://community.intel.com/t5/Blogs/Tech-Innovation/Edge-5G/Intel-at-ISSCC-2025-Navid-Shahriari-Invited-Talk-Eight-Papers/post/1667592>, 2025, invited Talk at ISSCC 2025.
- [25] Sell An et al., “Intel 4 cmos technology featuring advanced finfet transistors optimized for high density and high-performance computing,” in *2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits)*, 2022, pp. 282–283, DOI: 10.1109/VLSITechnologyandCir46769.2022.9830194.
- [26] M. G. Bardon, V. Moroz, G. Eneman, P. Schuddinck, M. Dehan, D. Yakimets, D. Jang, G. Van der Plas, A. Mercha, A. Thean, D. Verk-

- est, and A. Steegen, "Layout-induced stress effects in 14nm and 10nm finfets and their impact on performance," in *2013 Symposium on VLSI Technology*, 2013, pp. T114–T115.
- [27] P. Zhao, S. M. Pandey, E. Banghart, X. He, R. Asra, V. Mahajan, H. Zhang, B. Zhu, K. Yamada, L. Cao, P. Balasubramaniam, M. Joshi, M. Eller, F. Benistant, and S. Samavedam, "Influence of stress induced ct local layout effect (lle) on 14nm finfet," in *2017 Symposium on VLSI Technology*, 2017, pp. T228–T229, DOI: 10.23919/VLSIT.2017.7998182.
- [28] Y. Tsividis and C. McAndrew, *Operation and Modeling of the MOS Transistor*, ser. The Oxford Series in Electrical and Computer Engineering Series. Oxford University Press, 2011.
- [29] J. J. Liou, A. Ortiz-Conde, and F. Garcia-Sanchez, *Analysis and Design of MOSFETs: Modeling, Simulation, and Parameter Extraction*. Springer, 2005, DOI: 10.1007/978-1-4615-5415-8.
- [30] S. Takagi, M. Iwase, and A. Toriumi, "On the universality of inversion-layer mobility in n- and p-channel mosfets," in *Technical Digest., International Electron Devices Meeting*, 1988, pp. 398–401, DOI: 10.1109/IEDM.1988.32840.
- [31] A. Rossoni, T. Brozek, S. Saxena, R. Khamankar, C. Hess, J. Huang, Y. Teng, Z. Kovacs-Vajna, and M. Quarantelli, "Layout experiments and test structures to characterize local layout effects due to mechanical stress in finfet transistors," in *2025 IEEE 37th International Conference on Microelectronic Test Structures (ICMTS)*, 2025, pp. 1–4, DOI: 10.1109/ICMTS63811.2025.11068908.
- [32] K. Miyaguchi, F. M. Bufler, T. Chiarella, P. Matagne, N. Horiguchi, A. D. Keersgieter, G. Eneman, A. Spessot, B. Parvais, D. Verkest, and A. Mocuta, "Single and double diffusion breaks in 14nm finfet and beyond," 2017, DOI: 10.7567/SSDM.2017.E-2-03.
- [33] S. Saxena, C. Hess, H. Karbasi, A. Rossoni, S. Tonello, P. McNamara, S. Lucherini, S. Minehane, C. Dolainsky, and M. Quarantelli, "Variation in transistor performance and leakage in nanometer-scale technologies," *IEEE Transactions on Electron Devices*, vol. 55, no. 1, pp. 131–144, 2008, DOI: 10.1109/TED.2007.911351.
- [34] S. Saxena, C. Dolainsky, M. Lunenborg, J. Cheng, B. Yu, R. Vallishayee, and D. Ciplickas, "Impact of layout at advanced technology nodes on the

- performance and variation of digital and analog figures of merit,” in *2013 IEEE International Electron Devices Meeting*, 2013, pp. 17.2.1–17.2.4, DOI: 10.1109/IEDM.2013.6724646.
- [35] Y. Y. Watanabe T., “Dielectric breakdown of gate insulator due to reactive etching,” in *Solid State Technology*, vol. 26, no. 4, 1984, p. 263.
- [36] H. Shin, C. King, and C. Hu, “Thin oxide damage by plasma etching and ashing processes,” in *30th Annual Proceedings Reliability Physics 1992*, 1992, pp. 37–41, DOI: 10.1109/RELPHY.1992.187618.
- [37] S. Fang and J. McVittie, “Thin-oxide damage from gate charging during plasma processing,” *IEEE Electron Device Letters*, vol. 13, no. 5, pp. 288–290, 1992, DOI: 10.1109/55.145056.
- [38] C. Hess, S. Saxena, H. Karbasi, S. Subramanian, M. Quarantelli, A. Rossoni, S. Tonello, S. Zhao, and D. Slisher, “Device Array Scribe Characterization Vehicle Test Chip for Ultra Fast Product Wafer Variability Monitoring,” in *2007 IEEE International Conference on Microelectronic Test Structures*. Bunkyo-ku, Japan: IEEE, Mar. 2007, pp. 145–149, DOI: 10.1109/ICMTS.2007.374472.
- [39] PDF Solutions, Inc., “pdFasTest® , Exensio® , Characterization Vehicle® System,” <https://www.pdf.com>, 2025.
- [40] X. Ke, C. Jiang, S. Ji, Z. Zhao, B. Su, F. Li, C. Zheng, and H. Zhang, “Pitch walking simulation and process window evaluation of self aligned quadruple patterning for advanced nodes,” in *2022 China Semiconductor Technology International Conference (CSTIC)*, 2022, pp. 1–3, DOI: 10.1109/CSTIC55103.2022.9856861.
- [41] A. Rossoni, T. Brozek, S. Saxena, R. Khamankar, L. Colalongo, and Z. M. Kovacs-Vajna, “Stress-related local layout effects in finfet technology and device design sensitivity,” *IEEE Transactions on Semiconductor Manufacturing*, vol. 38, no. 2, pp. 117–125, 2025, DOI: 10.1109/TSM.2025.3540267.
- [42] M. T. Bohr and I. A. Young, “Cmos scaling trends and beyond,” *IEEE Micro*, vol. 37, no. 6, pp. 20–29, 2017, DOI: 10.1109/MM.2017.4241347.
- [43] X. Huang, W.-C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y.-K. Choi, K. Asano, V. Subramanian, T.-J. King, J. Bokor, and C. Hu, “Sub-50 nm p-channel finfet,” *IEEE Transactions on Electron Devices*, vol. 48, no. 5, pp. 880–886, 2001, DOI: 10.1109/16.918235.

- [44] T. Sekigawa and Y. Hayashi, "Calculated threshold-voltage characteristics of an XMOS transistor having an additional bottom gate," *Solid State Electron.*, vol. 27, no. 8-9, pp. 827–828, Aug. 1984, DOI: 10.1016/0038-1101(84)90036-4.
- [45] L. T. Clark, V. Vashishtha, D. M. Harris, S. Dietrich, and Z. Wang, "Design flows and collateral for the asap7 7nm finfet predictive process design kit," in *2017 IEEE International Conference on Microelectronic Systems Education (MSE)*, 2017, pp. 1–4, DOI: 10.1109/MSE.2017.7945071.
- [46] M. Khare, "High-k/metal gate technology: A new horizon," in *2007 IEEE Custom Integrated Circuits Conference*, 2007, pp. 417–420, DOI: 10.1109/CICC.2007.4405765.
- [47] Y.-C. Yeo, P. Ranade, T.-J. King, and C. Hu, "Effects of high- κ gate dielectric materials on metal and silicon gate workfunctions," *IEEE Electron Device Letters*, vol. 23, no. 6, pp. 342–344, 2002, DOI: 10.1109/LED.2002.1004229.
- [48] S. Kesapragada, R. Wang, D. Liu, G. Liu, Z. Xie, Z. Ge, H. Yang, Y. Lei, X. Lu, X. Tang, J. Lei, M. Allen, S. Gandikota, K. Moraes, S. Hung, N. Yoshida, and C.-P. Chang, "High-k/metal gate stacks in gate first and replacement gate schemes," in *2010 IEEE/SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*. San Francisco, CA, USA: IEEE, Jul. 2010, pp. 256–259, DOI: 10.1109/ASMC.2010.5551460.
- [49] B. P. Linder, V. Narayanan, and E. A. Cartier, "Interfacial layer optimization of high-k/metal gate stacks for low temperature processing," *Microelectron. Eng.*, vol. 86, no. 7-9, pp. 1632–1634, Jul. 2009, DOI: 10.1016/j.mee.2009.03.058.
- [50] J. Robertson and R. M. Wallace, "High-K materials and metal gates for CMOS applications," *Materials Science and Engineering: R: Reports*, vol. 88, pp. 1–41, Feb. 2015, DOI: 10.1016/j.mser.2014.11.001.
- [51] Y. J. Kim, D. Lim, H. H. Han, A. S. Sergeevich, Y.-R. Jeon, J. H. Lee, S. K. Son, and C. Choi, "The effects of process temperature on the work function modulation of ALD HfO₂ MOS device with plasma enhanced ALD TiN metal gate using TDMAT precursor," *Microelectronic Engineering*, vol. 178, pp. 284–288, Jun. 2017, DOI: 10.1016/j.mee.2017.05.023.

- [52] R. Saad, M. Silberg, Y. Dafne, and Z. Lando, “Optimizing the Tungsten Deposition Process,” *High Temperature Materials and Processes*, vol. 15, no. 3, pp. 217–222, Jul. 1996, DOI: 10.1515/HTMP.1996.15.3.217.
- [53] T. B. Hook, “Power and Technology Scaling into the 5 nm Node with Stacked Nanosheets,” *Joule*, vol. 2, no. 1, pp. 1–4, Jan. 2018, DOI: 10.1016/j.joule.2017.10.014.
- [54] A. Greene, H. Zhou, R. Xie, C. Park, L. Economikos, V. Chan, K. Akarvardar, R. Bao, I. Seshadri, R. Conti, M. Wang, M. Sankarapandian, J. Demarest, J. Li, L. Jiang, K. Zhao, R. Robison, T. Ando, N. Cave, A. Knorr, D. Gupta, S. Kanakasabapathy, D. Guo, B. Haran, V. Basker, and H. Bu, “Gate-Cut-Last in RMG to Enable Gate Extension Scaling and Parasitic Capacitance Reduction,” in *2019 Symposium on VLSI Technology*. Kyoto, Japan: IEEE, Jun. 2019, pp. T144–T145, DOI: 10.23919/VLSIT.2019.8776493.
- [55] S. L. Fei, W. Q. Peng, Z. J. Hong, and C. Y. Shan, “Exploring Gate-Cut Patterning Approaches Using Simulation and Defect Modelling,” in *2021 China Semiconductor Technology International Conference (CSTIC)*. Shanghai, China: IEEE, Mar. 2021, pp. 1–4, DOI: 10.1109/CSTIC52283.2021.9461509.
- [56] Synopsys, *SentaurusTM Process User Guide*, Synopsys.
- [57] R. Akis and D. Ferry, “Kinetic lattice monte carlo simulations of silicon and germanium epitaxial growth on the silicon (100) surface,” *IEEE Transactions on Nanotechnology*, vol. 4, no. 3, pp. 345–348, 2005, DOI: 10.1109/TNANO.2005.846926.
- [58] Synopsys, *SentaurusTM Device User Guide*, Synopsys.
- [59] S. A. Mujtaba, “Advanced mobility models for design and simulation of deep submicrometer mosfets,” 1995.
- [60] M. Darwish, J. Lentz, M. Pinto, P. Zeitzoff, T. Krutsick, and H. H. Vuong, “An improved electron and hole mobility model for general purpose device simulation,” *IEEE Transactions on Electron Devices*, vol. 44, no. 9, pp. 1529–1538, 1997, DOI: 10.1109/16.622611.
- [61] J. Bardeen and W. Shockley, “Deformation potentials and mobilities in non-polar crystals,” *Phys. Rev.*, vol. 80, pp. 72–80, Oct 1950, DOI: 10.1103/PhysRev.80.72.

- [62] G. Bir and G. Pikus, *Symmetry and Strain-induced Effects in Semiconductors*, ser. A Halsted Press book. Wiley, 1974.
- [63] M. Charmi, “Novel attributes and design considerations of effective oxide thickness in nano dg mosfets,” *Chinese Physics B*, vol. 24, no. 4, p. 047302, 2015, DOI: 10.1088/1674-1056/24/4/047302.
- [64] A. Rossoni, Z. M. Kovacs-Vajna, L. Colalongo, R. Khamankar, S. Saxena, and T. Brozek, “Modeling stress effects from fin isolation in 7nm finfet transistors,” in *2024 International Semiconductor Conference (CAS)*, 2024, pp. 127–130, DOI: 10.1109/CAS62834.2024.10736860.
- [65] A. Rossoni, T. Brozek, S. Saxena, R. Khamankar, and Z. M. Kovacs-Vajna, “Impact of the inter-fin space on stress modulation and finfet transistor performance,” in *2025 International Conference on IC Design and Technology (ICICDT)*, 2025, pp. 161–164, DOI: 10.1109/ICICDT65192.2025.11078040.
- [66] Y.-C. Eng, L. Hu, T.-F. Chang, C.-Y. Wang, S. Hsu, O. Cheng, C.-T. Lin, Y.-S. Lin, Z.-J. Tsai, C.-W. Yang, J. Lu, and S. Y.-W. Chen, “Significance of overdrive voltage in the analysis of short-channel behaviors of n-finet devices,” *IEEE Journal of the Electron Devices Society*, vol. 10, pp. 281–288, 2022, DOI: 10.1109/JEDS.2022.3160881.
- [67] S. Chou and D. Antoniadis, “Relationship between measured and intrinsic transconductances of fet’s,” *IEEE Transactions on Electron Devices*, vol. 34, no. 2, pp. 448–450, 1987, DOI: 10.1109/T-ED.1987.22942.
- [68] G. Ghibaudo, “New method for the extraction of mosfet parameters,” *Electronics Letters*, vol. 24, pp. 543–545, 1988, DOI: 10.1049/el:19880369.
- [69] Yang Sam, et al., “10nm high performance mobile SoC design and technology co-developed for performance, power, and area scaling,” in *2017 Symposium on VLSI Technology*. Kyoto, Japan: IEEE, Jun. 2017, pp. T70–T71, DOI: 10.23919/VLSIT.2017.7998203.
- [70] X. Guo, X. Dong, S. Yao, Z. Gan, W. Wang, Z. Yang, E. Chong, Q. Li, Z. Mao, L. Zhang, R. Li, and Y. Zhang, “Photolithography solutions for fabrication of fin and poly-gate in 14nm finfet devices,” in *2015 China Semiconductor Technology International Conference*. IEEE, Mar. 2015, pp. 1–3, DOI: 10.1109/cstic.2015.7153353.
- [71] X. Li and C. Liu, “The alignment performance study for the gate layer in finfet processes,” in *2016 China Semiconductor Technology International Conference (CSTIC)*, 2016, pp. 1–3, DOI: 10.1109/CSTIC.2016.7463978.

- [72] A. Paul, A. Bello, V. K. Kamineni, and D. Deniz, “Finfet channel stress using tungsten contacts in raised epitaxial source and drain,” Patent US8 975 142B2, March, 2015, assignee: GlobalFoundries Inc. Filed: April 25, 2013. Granted: March 10, 2015. [Online]. Available: <https://patents.google.com/patent/US8975142B2/en>
- [73] C. Auth, “45nm high-k + metal gate strain-enhanced cmos transistors,” in *2008 IEEE Custom Integrated Circuits Conference*, 2008, pp. 379–386, DOI: 10.1109/CICC.2008.4672101.
- [74] S. Baudot, S. Guissi, A. P. Milenin, J. Ervin, and T. Schram, “N7 finfet self-aligned quadruple patterning modeling,” in *2018 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2018, pp. 344–347, DOI: 10.1109/SISPAD.2018.8551646.
- [75] I. Seshadri, E. Miller, J. Church, A. Chu, J. Zhang, A. Greene, J. Frougier, T. Li, Y. Cabrera, G. Kenath, M. Burkhardt, S. Skordas, L. Meli, and N. Felix, “Scaling opportunities for gate-all-around: A patterning perspective,” in *2023 International Electron Devices Meeting (IEDM)*, 2023, pp. 1–4, DOI: 10.1109/IEDM45741.2023.10413878.
- [76] B. Ye, H. Zhu, L. Chen, J. Li, Q.-Q. Sun, and D. W. Zhang, “Suppression of stress-induced defects in finfet by implantation and sti co-optimization,” *IEEE Transactions on Electron Devices*, vol. 68, no. 5, pp. 2587–2589, 2021, DOI: 10.1109/TED.2021.3068241.
- [77] Y. Lian, *Dielectric Films Growth*, 2023, pp. 161–177, DOI: 10.1002/9781119867814.ch13.
- [78] D. G. Lishan and K. D. Mackenzie, “Comparing Chemical Vapor Deposition Systems: LPCVD vs. PECVD vs. HDPCVD (white paper),” November 2020.
- [79] Y. Xia, L. Chen, H. Zhu, Q. Sun, and D. W. Zhang, “Suppression of sti-induced asymmetric stress in finfet by cesl stressor,” *Electronics*, vol. 14, no. 11, 2025, DOI: 10.3390/electronics14112099.
- [80] T. P. Dash, S. Das, S. Dey, J. Jena, and C. K. Maiti, “Stress profile analysis in n-finfet devices,” in *Emerging Technology in Modelling and Graphics*, J. K. Mandal and D. Bhattacharya, Eds. Singapore: Springer Singapore, 2020, pp. 309–318.
- [81] G. Eneman, L. Witters, J. Mitard, G. Hellings, A. De Keersgieter, D. P. Brunco, A. Hikavyy, B. Vincent, E. Simoen, P. Favia, H. Bender, A. Veloso,

- T. Chiarella, G. Boccardi, M. Kim, M. Togo, R. Loo, K. De Meyer, N. Horiguchi, N. Collaert, and A. Thean, "Stress techniques and mobility enhancement in finfet architectures," *ECS Transactions*, vol. 50, no. 9, p. 47, mar 2013, DOI: 10.1149/05009.0047ecst.
- [82] A. Pal, E. M. Bazizi, L. Jiang, M. Saremi, B. Alexander, and B. Ayyagari-Sangamalli, "Self-aligned single diffusion break technology optimization through material engineering for advanced cmos nodes," in *2020 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2020, pp. 307–310, DOI: 10.23919/SISPAD49475.2020.9241625.
- [83] Q. Huo, Z. Wu, W. Huang, X. Wang, G. Tang, J. Yao, Y. Liu, X. Zhao, F. Zhang, L. Li, and M. Liu, "A novel general compact model approach for 7-nm technology node circuit optimization from device perspective and beyond," *IEEE Journal of the Electron Devices Society*, vol. 8, pp. 295–301, 2020, DOI: 10.1109/JEDS.2020.2980441.
- [84] A. Beckers, F. Jazaeri, and C. Enz, "Revised theoretical limit of subthreshold swing in field-effect transistors," 2018, DOI: 10.48550/arXiv.1811.09146.
- [85] G. Ghibaudo, M. Aouad, M. Casse, S. Martinie, T. Poiroux, and F. Balestra, "On the modelling of temperature dependence of subthreshold swing in mosfets down to cryogenic temperature," *Solid-State Electronics*, vol. 170, Aug. 2020, DOI: 10.1016/j.sse.2020.107820.
- [86] F. Shoucair, "Scaling, subthreshold, and leakage current matching characteristics in high-temperature (25 degrees c-250 degrees c) vlsi cmos devices," *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, vol. 12, no. 4, pp. 780–788, 1989, DOI: 10.1109/33.49047.
- [87] V. Huard, C. Ndiaye, M. Arabi, N. Parihar, X. Federspiel, S. Mhira, S. Mahapatra, and A. Bravaix, "Key parameters driving transistor degradation in advanced strained sige channels," in *2018 IEEE International Reliability Physics Symposium (IRPS)*, 2018, pp. P–TX.4–1–P–TX.4–6, DOI: 10.1109/IRPS.2018.8353699.
- [88] P. Ren, R. Wang, and R. Huang, "Layout dependent bti and hci degradation in nano cmos technology: A new time-dependent lde and impacts on circuit at end of life," in *2016 International Conference on IC Design and Technology (ICICDT)*, 2016, pp. 1–3, DOI: 10.1109/ICICDT.2016.7542063.
- [89] A. H. Gencer, D. Tsamados, and V. Moroz, "Fin bending due to stress and its simulation," in *2013 International Conference on Simulation of*

- Semiconductor Processes and Devices (SISPAD)*, 2013, pp. 109–112, DOI: 10.1109/SISPAD.2013.6650586.
- [90] L. Sun, P. Lyu, Q. Wang, Q. Zhong, K. Wang, and Y. Chi, “Investigation of fin bowing formation mechanism during sti etching by virtual fabrication,” in *2022 China Semiconductor Technology International Conference (CSTIC)*, 2022, pp. 01–03, DOI: 10.1109/CSTIC55103.2022.9856777.
- [91] P. De Bisschop and E. Hendrickx, “Stochastic effects in euv lithography,” 03 2018, p. 53, DOI: 10.1117/12.2300541.
- [92] Y. Zhang, Y. Liu, Y. Wang, and Y. Xie, “Ai-eda: Ai-enhanced electronic design automation,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 40, no. 11, pp. 2345–2358, 2021, DOI: 10.1109/TCAD.2021.3062345.
- [93] L. Xue, A. Dixit, N. Kumar, and V. Georgiev, “Generative process variation modeling and analysis for advanced technology based on variational autoencoder,” *IEEE Transactions on Electron Devices*, pp. 1–7, 2025, DOI: 10.1109/TED.2025.3570675.
- [94] S. Woo, J. Jeon, and S. Kim, “Prediction of device characteristics of feedback field-effect transistors using tcad-augmented machine learning,” *Micromachines*, vol. 14, no. 3, 2023, DOI: 10.3390/mi14030504.
- [95] H. Kuang, X. Cao, J. Li, and L. Wang, “Hgbo-dse: Hierarchical gnn and bayesian optimization based hls design space exploration,” in *2023 International Conference on Field Programmable Technology (ICFPT)*, 2023, pp. 106–114, DOI: 10.1109/ICFPT59805.2023.00017.
- [96] C. Zimmerling, C. Poppe, O. Stein, and L. Kärger, “Optimisation of manufacturing process parameters for variable component geometries using reinforcement learning,” *Mater. Des.*, vol. 214, no. 110423, p. 110423, Feb. 2022, DOI: 10.1016/j.matdes.2022.110423.
- [97] F. Teng, X. Yan, J. Jin, J. Sun, R. Wang, M. Li, H. Wu, and R. Huang, “Stress-aware performance optimization in gate-all-around nsfet and beyond,” *IEEE Transactions on Electron Devices*, vol. 72, no. 7, pp. 3379–3386, 2025, DOI: 10.1109/TED.2025.3569503.
- [98] “Material properties,” AZoM.com, 2023. [Online]. Available: <https://www.azom.com/>

- [99] *Kyocera Corporation*, “Fine ceramics (advanced ceramics) materials: Silicon nitride,” 2023. [Online]. Available: <https://global.kyocera.com/feworld/character/nitride.html>
- [100] T. Yamamoto and Y. Katayama, “Thermal expansion of α - and β - Si_3N_4 by x-ray diffraction,” *Journal of the American Ceramic Society*, vol. 84, pp. 1233–1238, 2001.
- [101] Y. Kuo and C. Lee, “Thermal and mechanical properties of si-rich silicon nitride films,” *Thin Solid Films*, vol. 478, pp. 244–249, 2005.
- [102] Y. Zhang and H. Kim, “Effect of rare earth doping on mechanical properties of Si_3N_4 ,” *Journal of Materials Science*, vol. 45, pp. 1123–1130, 2010, DOI: 10.1007/s11595-023-2841-y.
- [103] L. Chen and J. Wang, “Tailoring thermal expansion of Si_3N_4 ceramics with tin and ZrO_2 ,” *Ceramics International*, vol. 38, pp. 4567–4573, 2012.
- [104] *TCAD Central*, “Low-k dielectrics for advanced cmos,” 2022, technical report, internal simulation data.
- [105] *TCAD Central*, “High-k materials for gate and isolation layers,” 2022, technical report, internal simulation data.