UCLA
UCLA Electronic Theses and Dissertations

Title
Design and Optimization of Stacked Nanosheet FET and FinHBT for Ultra-Scaled SoC

Permalink
https://escholarship.org/uc/item/5h78t3h2

Author
DUAN, Xicheng

Publication Date
2022

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Design and Optimization of Stacked Nanosheet FET and FinHBT for Ultra-Scaled SoC

A thesis submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Electrical Engineering

by

Xicheng Duan

2022

© Copyright by

Xicheng Duan

2022

ABSTRACT OF THE DISSERTATION

Design and Optimization of Stacked Nanosheet FET and FinHBT for Ultra-Scaled SoC

by

Xicheng Duan

Doctor of Philosophy in Electrical Engineering

University of California, Los Angeles, 2022

Professor Jason C. S. Woo, Co-Chair

Professor Mau-Chung Frank Chang, Co-Chair

Combining digital computation, analog, and radio frequency (RF) circuitry, a highly functional systems-on-chip (SoC) design is a favorable choice for various applications like mobile systems, embedded systems, and space applications. While the aggressive scaling continues for digital applications, the analog/RF performance of these highly scaled devices is lagging. To achieve an SoC technology with advanced ultra-scaled digital transistors and compatible RF/Mixed-Mode devices, this work is divided into two major sections. In the first section, TCAD simulation of Stacked Nanosheet FETs (NSFETs) based on quantum physics has been performed. The study focuses on critical device parameters including L_G, T_{NS}, parasitic resistance, and EOT. The TDDB behavior of NSFETs with different corner rounding radii is studied closely with a discrete trap based TCAD simulation. In the second section, an innovative lateral SiGe FinHBT is proposed to leverage the lateral scaling capability of the FinFET CMOS fabrication platform to establish high-speed BiCMOS VLSI SoCs for THz/mixed-mode applications. A complete CMOS compatible process flow of fabricating a nanoscale lateral SiGe FinHBT on an SOI wafer has been developed.

A prototype of the lateral SiGe FinHBT has been fabricated and characterized. Based on the results, it is projected that the lateral SiGe FinHBT can potentially reach f_T/f_{MAX} > 750 GHz with further process optimization and scaling.

The dissertation of Xicheng Duan is approved.

Mark S. Goorsky

Chin-Kong Ken Yang

Robert Candler

Jason C S Woo, Committee Co-Chair

Mau-Chung Frank Chang, Co-Chair

University of California, Los Angeles

2022

Table of Contents

Chapter 1 Introduction......1
1.1 Background and motivation 1
1.2 Outline 3
Chapter 2 TCAD modeling 5
2.1 Transport model 5
2.2 Quantum model 7
2.3 Trap-assisted tunneling models 9
2.4 Conclusion 11
Chapter 3 FinFET and NSFET simulation 13
3.1 10nm FinFET simulation 13
3.2 Sub-5nm NSFET Simulation 15
3.3 Parasitic resistance analysis 18
3.4 Channel length scaling 20
3.5 Nanosheet thickness scaling 23
3.6 EOT scaling 26
3.7 Conclusion......28
Chapter 4 NSFET reliability study 29
4.1 Literature review 29
4.2 Discrete trap TDDB model 30
4.3 Breakdown process and calibration 34
4.4 NSFET TDDB Simulation 38
4.5 Conclusion 44
Chapter 5 FinHBT process development 45
5.1 Device Architecture 45
5.2 FinHBT Process Flow 46
5.3 Base epitaxial growth 48
5.4 Fin Patterning and contact opening 54
5.5 Silicidation and Metallization 55
5.6 Conclusion 58

Chapter 6 FinHBT characterization and simulation prediction 59
6.1 FinHBT DC Characteristics 59
6.2 FinHBT RF performance simulation 61
6.3 Conclusion 66
Chapter 7 Conclusion 67
7.1 Summary 67
7.2 Suggestions for future research 69

Reference 71

List of Figures

Fig. 2.1 Modified Vsat of electrons vs channel length. The velocity profile is fitted to the Monte Carlo simulation 6

Fig.2.2 Simulated vs experiment electron mobility. Mobility extracted as the weighted average over the charge in the channel 7

Fig. 2.3 Simulation procedure of 2D Schrodinger solver 8

Fig. 2.4 The calibrated capacitance vs silicon thickness 9

Fig 2.5. Capture involving direct tunneling at (a) low gate bias and (b) high gate bias................10

Fig 2.6. Inelastic tunneling assisted by phonon emission with energy $\hbar\omega$ 11

Fig 3.1. (a) Simulated 3D structure of the 10nm FinFET (b) cross-section along the channel direction of the FinFET 14

Fig. 3.2 Fin cross-section of the simulated FinFET and TEM 15

Fig 3.3 Simulated vs experiment results of 10nm FinFET 15

Fig. 3.4 Fig. 3.4 (a) Simulated structure of sub-5nm NSFET (b) cross-section along the channel direction of the NSFET..16

Fig. 3.5. Fin cross-section of the simulated FinFET and TEM...17

Fig. 3.6 The simulated transfer characteristics of 5nm FinFET and sub-5nm FinHBT 18

Fig. 3.7 (a) Parasitic resistance components in NSFET (b) electron current density and direction in NSFET 19

Fig. 3.8. Comparing parasitic components of 10nm FinFET, 5nm FinFET, and sub-5nm NSFET 20

Fig. 3.9 Ion and Ioff of NSFET versus L_G. The gate work function is fixed at 4.5eV. 21

Fig. 3.10 Ion at fixed Ioff=10nA/um versus L_G. 22

Fig. 3.11 Subthreshold swing (SS) versus L_G 22

Fig. 3.12 Intrinsic delay (CV/I) versus L_G 23

Fig. 3.13 Carrier density distribution at nanosheet cross-section at V_{GS}=0.7V. The carriers move closer to the channel/oxide interface as TNS decreases., leading to improved electrostatic control 24

Fig. 3.14 Ion and Ioff of NSFET versus T_{NS}. The gate work function is fixed at 4.5eV 24

Fig. 3.15 Ion at fixed Ioff=10nA/um versus T_{NS}. 25

Fig. 3.16 Subthreshold swing (SS) versus T 25

Fig. 3.17 Charge distribution in the channel. Charge peak moves closer to the channel/gate dielectric interface as EOT scales down 27

Fig. 3.18 EOT_{total} and EOT_{si} versus EOT of NSFET. 27

Fig. 3.19 Ion and intrinsic delay versus EOT 28

Fig. 4.1 Simulated and experiment MOSCAP gate current vs stress voltage 31

Fig. 4.2 The procedure of TDDB simulation with discrete traps 33

Fig. 4.3 Trap locations and electric field of the initial state (t=0) 35

Fig. 4.4 The traps and electric field during the breakdown process 35

Fig. 4.5 Trap number in IL and HK versus time. 36

Fig. 4.6 Gate current versus time 36

Fig. 4.7 Experiment current vs time at V_G=2.2 – 2.5V...37

Fig. 4.8 Calibrated I_G versus time simulation. The I_G falls between the lowest and highest current of the experiment data...37

Fig. 4.9 3D simulated structure and cross-section of NSFET TDDB..38

Fig. 4.10 Simulated nanosheets with different corner rounding radius 38

Fig. 4.11. The electric field near the corners of NSFET with different corner rounding radius ...39

Fig. 4.12 (a) The cutline position in <100> direction (b) The electric field along the cutline (c) trap density in the interfacial layer (d) trap density in the high-k layer.......................................40

Fig. 4.13 (a) The cutline position in <110> direction (b) The electric field along the cutline (c) trap density in the interfacial layer (d) trap density in the high-k layer.......................................41

Fig. 4.14 (a) The cutline position in corner direction (b) The electric field along the cutline. The electric field decreases quickly towards the gate 41

Fig. 4.15 Number of traps versus time in devices with different rounding. The traps gradually build up in IL and increase abruptly in HK. 42

Fig. 4.16 Gate current versus stress time. 2.5nm corner rounding improves TDDB time by 10% 43

Fig. 4.17 3D electric field distribution at the top of HK layer 43

Fig. 4.18 3D current distribution at the silicon/IL interface 44

Fig. 5.1. 3D structure of FinHBT with T-shaped extrinsic base..46

Fig. 5.2. Key process steps of T-base FinHBT..47

Fig. 5.3. Flow chart of T-base FinHBT process 48

Fig. 5.4 (a) SiGe blanket growth rate and Ge composition vs Germane/DCS mass flow rate. (b) AFM imaging of SiGe grown at 650°C 50

Fig. 5.5. SiGe SEG growth rate and Ge composition vs HCl flow..50

Fig. 5.6 Borane doping vs Borane flow rate 51

Fig. 5.7 The process flow of SiGe SEG on SOI substrate including the extrinsic base growth....52

Fig. 5.8 Cross-section view of 100nm base after T-base growth...53

Fig. 5.9 SEM of a 300nm T-base..53

Fig. 5.10. cross-section SEM of the fin 54

Fig. 5.11 FinHBT after contact opening 55

Fig. 5.12 Silicide film resistivity versus annealing temperature... 56

Fig. 5.13 Cross-section SEM of the contact after the silicidation process 56

Fig. 5.14 Contact resistivity versus RTA temperature... 57

Fig. 5.15 The device after contact opening formation. The alignment error is negligible. 58

Fig. 6.1 Measured Gummel plot of the FinHBT 60

Fig. 6.2 Measured Gain versus I_B of FinHBT...60

Fig. 6.3 Measured common-emitter characteristics of FinHBT 61

Fig. 6.4 Simulated and measured Gummel plot of FinHBT 62

Fig. 6.5. Simulated T-base FinHBT..63

Fig. 6.6 Doping and Ge profile in the simulated FinHBT with a base width of 20nm.................63

Fig. 6.7. Simulate Gummel plot and f_T/f_{MAX} of test device 1 64

Fig. 6.8. Simulate Gummel plot and f_T/f_{MAX} of test device 2 65

Fig. 6.9. Carrier velocity profile and delay components extraction..66

List of Tables

Table 3.1 The key parameters of the sub-5nm and 5nm FinFET 17

Table 4.1 Parameters for trap-assisted tunneling 31

Table 4.2 Parameters of the 2D TDDB simulation 34

Table 5.1 Comparison between FinFET and FinHBT process 48

Table 6.1 Parameters of the fabricated FinHBT 59

Table 6.2. Simulated characteristics of T-base FinHBT 65

Table 6.3. Delay components of device 1 and device 2 66

Acknowledgements

Firstly, I would like to express my sincere thanks to my advisor Prof. Jason Woo for his guidance and support during through my research journey. I would like to thank Prof. Frank Chang for his stewardship of the FinHBT project. I would also like to thank Prof. Mark Goorsky, and Prof. Rob Candler and Prof. Ken Yang for sparing the time to be in my committee.

The FinFET and NSFET simulation portion of the dissertation is based on cooperated team project with Applied Material. I would like to thank my colleague Peng Lu and Yifei Li for the great teamwork during the project. Also, special thanks to Dr. Steven Hung and Dr. Benjamin Colombeau from Applied Materials.

The experimental work of the FinHBT project is based on research sponsored by Air Force Research Laboratory (AFRL) and Defense Advanced Research Projects Agency (DARPA). The experiments are mainly carried out at Stanford Nanofabrication Facility (SNF). Many thanks to all the staff members at SNF for their technical supports.

I would like to thank all my lab mates, Wilson, Peng, Allen, Scott, and Yifei for the great teamwork and inspiring discussions.

In the end, special gratitude to my parents for their unconditional support even during the toughest part of my Ph. D. career.

VITA

2014	Bachelor of Engineering in Electrical and Computer Engineering, Hong Kong University of Science and Technology
2016	Master of Science, Electrical Engineering University of California, Los Angeles
2018	Ph. D. Candidate, Electrical Engineering, University of California, Los Angeles

Publications

X. Duan, B. Huang, J. C. S. Woo, "Device and Process Design of FinHBT for THz/Mixed-Mode Applications", 2021 GOMACTech Conference

Lu, Peng and Colombeau, Benjamin and Hung, Steven and Li, Weicong and Duan, Xicheng and Li, Yifei and Bazizi, El Mehdi and Natarajan, Sanjay and Woo, Jason C. S"Source/Drain Extension Doping Engineering for Variability Suppression and Performance Enhancement in 3-nm Node FinFETs," in *IEEE Transactions on Electron Devices*, vol. 68, no. 3, pp. 1352-1357, March 2021, doi: 10.1109/TED.2021.3052432.

Xicheng Duan, Peng Lu, Weicong Li, and Jason C. S. Woo, "Parasitic resistance modeling and optimization for 10nm-node FinFET," 2018 18th International Workshop on Junction Technology (IWJT), Shanghai, 2018, pp. 1-4, DOI: 10.1109/IWJT.2018.8330306.

Peng Lu, Po-Yen Chien, Xicheng Duan, Jason C. S. Woo, "Deeply scaled VLSI analog transistor design and optimization," *2017 IEEE 12th International Conference on ASIC (ASICON)*, Guiyang, 2017, pp. 1061-1064, DOI: 10.1109/ASICON.2017.8252662

CHAPTER 1
Introduction

1.1 Background and motivation

Over the last 50 years, Moore's law has provided continuous improvement in semiconductor device/circuit technology and has resulted in unprecedented advancement in electronic systems. Highly functional Systems-on-Chip (SoC) has been developed, combining RF/mixed-mode with signal conversion and enormous amounts of digital computation and signal processing. Due to its excellent electrostatic control and capability to deliver high current density per footprint, the FinFET structure has been adopted for sub-22nm/14nm node. While the FinFET technology has granted successful technology node scaling for a decade, it will eventually hit a bottleneck as scaling proceeds to the sub-5nm era. The aggressively scaling channel length, EOT, and gate/fin pitch have brought up numerous issues like short channel effect, parasitic resistance/capacitance, and reliability problems. To enable further scaling, gate-all-around (GAA) FETs have been proposed as the next mainstream technology for digital applications thanks to their excellent electrostatic control. Many different GAAFET designs have been reported, including lateral stacked nanowire FET [1], vertical nanowire FET [2], and stacked nanosheet FET (NSFET) [3]. Among all the different GAAFET designs, NSFET stands out as the most promising solution for sub-5nm VLSI due to its enhanced electrostatic control, large effective channel width per footprint, and minimized deviation from the standard FinFET process. In [3], NSFET with aggressively scaled 12nm Lg can still maintain a subthreshold swing (SS) of 75mV/dec. A 3-nanosheet stack can also reach comparable effective channel width per footprint with aggressively scaled FinFET

while being able to scale the contact poly pitch to 44nm (7nm node standard). This enables further scaling to sub-5nm nodes.

With all the benefits of adapting to NSFET, it is most likely to become the technology for ultra-scaled VLSI and SoC technology. However, as an emerging device, the relationship between the device's performance and key design parameters like L_G, EOT, and T_{NS} hasn't been fully understood. Its unique gate-all-around structure is accompanied by a severe quantum effect, which will significantly affect the carrier distribution in the channel region. Moreover, as the scaling continues, aggressive EOT scaling is needed to compensate for the increased parasitic capacitance. This will lead to significant reliability concerns, including gate leakage, TDDB, and BTI. Therefore, it is crucial to obtain a thorough understanding of the performance constraints and reliability issues of sub-5nm NSFET. In this work, a comprehensive TCAD platform is constructed for sub-5nm NSFET. Quasi-ballistic transport, 2D Schrodinger equation, and trap-assisted tunneling have been solved self-consistently. The device performance of sub-5nm NSFET is simulated and compared to FinFETs. Based on the TCAD platform, a discrete trap TDDB model has been constructed. A detailed breakdown analysis is done to guide the bottleneck of the TDDB in NSFET.

While the FinFET and NSFET keep scaling to sub-5nm nodes, it is very hard to utilize these devices for very high-frequency mixed-signal SoC applications because of their large parasitic resistance/capacitance and degraded mobility. An f_T/f_{MAX} of 300GHz/450GHz has been demonstrated on Intel 22nm Si FinFET platform [4]. However, it is still much too insufficient for terahertz sensing and high bandwidth communications. While III-V-based devices can approach THz [5] [6] [7], the lack of integration density constrains their applications in a highly scaled System-on-Chip system. In the past, BiCMOS technology has been the go-to solution for

RF/mixed-mode applications. Vertical SiGe HBT on 130nm node has demonstrated an f_T/f_{MAX} up to 505GHz/720GHz [8], which shows potential for THz operations. However, implementing the conventional vertical HBTs on the FinFET platform is extremely challenging due to the difference in layout dimensions and heights.

To solve this issue, the lateral SiGe FinHBT is proposed as a promising solution to realize the FinFET-compatible BiCMOS for THz/mixed-mode applications. In this project, a T-base FinHBT design capable of reaching $f_T/f_{MAX} > 750$GHz is presented through simulation. A process flow featuring lateral SiGe epitaxy growth is proposed to create the lateral SiGe profile to improve the carrier transfer in the base region. Key process steps, including base definition, SiGe SEG, fin patterning, and contact silicide formation are developed and demonstrated. An improved T-shaped base (T-base) structure is proposed to resolve the base resistance issue and improve the DC and RF performance of FinHBT. The ultimate goal is to develop ultrahigh-performance THz/mixed-mode devices that can be co-integrated with current FinFET technology.

1.2 Outline

Stacked Nanosheet FET (NSFET) technology is going to replace FinFET in sub-3nm nodes. With aggressive scaling of both device geometry and effective oxide thickness (EOT), it is crucial to understand how the change of important parameters affects the performance and reliability of this novel device. In this study, a physics based TCAD platform is set up for a comprehensive understanding of the effect of the performance and reliability of NSFET.

Key messages:

- TCAD modeling setup: 2D Schrodinger equation, mobility model, carrier tunneling
- FinFET and NSFET comparison
- Parasitic resistance analysis

- NSFET performance versus L_G, T_{NS}, EOT
- TDDB simulation with discrete traps in NSFET

A lateral SiGe FinHBT is proposed as a promising solution to realize the FinFET-compatible BiCMOS for THz/mixed-mode applications. A process flow featuring lateral SiGe selective epitaxy growth (SEG) is proposed to create the lateral SiGe profile and the T-shaped extrinsic base to improve the carrier transfer in the base region. Key process steps, including base definition, SiGe SEG, T-base patterning, fin patterning, and contact silicide formation are developed and demonstrated. Measurement of the fabricated DC test devices shows an HBT device with a peak current gain of 6.5. TCAD Simulation shows that with further base width scaling and doping/Ge profile optimization, the T-shaped base (T-base) FinHBT is capable of reaching $f_T/f_{MAX} > 750$GHz.

Key messages:

- FinHBT concept and structure explanation
- FinHBT process flow: SiGe epitaxial growth, dry etch, contact formation
- FinHBT DC measurement results and analysis
- TCAD simulation of FinHBT with scaled base width, DC and RF characteristics

.

CHAPTER 2
TCAD modeling

In this work, Sentaurus TCAD platform [9] is used to study the performance of transistors. In ultra-scaled FinFET and NSFET, quasi-ballistic transport (described in [10]) needs to be considered due to the sub-20nm channel length. A carrier transport model that is capable of capturing quasi-ballistic transport is necessary to correctly model the performance of FinFET and NSFET. Apart from that, the ultra-thin silicon body of FinFET and NSFET leads to severe quantum effects. Therefore, it is crucial that the quantum effect can be accurately modeled in the simulation. In the reliability study, the trap-assisted tunneling effect needs to be captured to model the gate leakage current under gate stress. In this chapter, the detailed model setup and calibration are demonstrated as the foundation of the TCAD analysis in the later chapters.

2.1 Transport model

In traditional device simulation, a macroscopic approach like the drift-diffusion (DD) model is used to simulate carrier transport in semiconductors. With the scaling of the device dimension, the traditional DD model with fixed carrier saturation velocity cannot capture the physics of microscopic effects like quasi-ballistic transport. Microscopic models based on the Boltzmann transport equation (BTE) can capture the physics of quasi-ballistic transport. However, it can be challenging to utilize this approach in a complex 3D device architecture like NSFET. Therefore, a modified DD model is adapted in this study.

Although the DD model fails to capture the physics of velocity overshoot, it provides a reasonably accurate electron velocity profile at a low electric field. At a high electric field, the carrier transport velocity is limited by the saturation velocity (v_{sat}). To account for the carrier velocity overshoot in short channel transistors, one practical method is to establish a field-dependent v_{sat} as described in [11]. With modified v_{sat}, the carrier velocity profile in the DD simulation can match the results in the Monte Carlo simulation for L_g from 10nm to 50nm. Therefore, this method provides an efficient approach to simulate the carrier transport in short channel FinFET and NSFET.

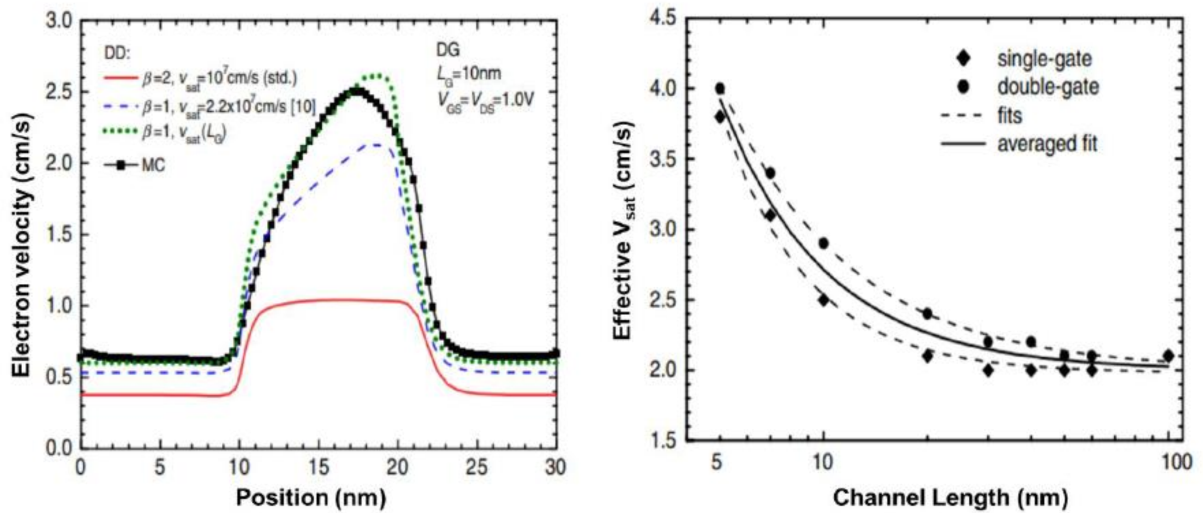


Fig. 2.1 Modified Vsat of electrons vs channel length. The velocity profile is fitted to the Monte Carlo simulation [11]

The carrier mobility degradation is modeled by a surface roughness scattering model. The model is described in [9]:

$$\frac{1}{\mu(x)} = \frac{1}{\mu_0} + \frac{exp(-x/l_{crit})}{\mu_S(E_{\perp}(x))} \qquad (2.1)$$

Where $\mu 0$ is the bulk semiconductor mobility, x is the distance from the silicon/gate dielectric interface, l_{crit} is a fitting parameter, and μ_s is the surface mobility component as a function of the electric field normal to the silicon/gate dielectric interface. The simulated mobility versus silicon body thickness is shown in Fig.2.2. The simulated mobility is compared to the reported results in [12] [13].

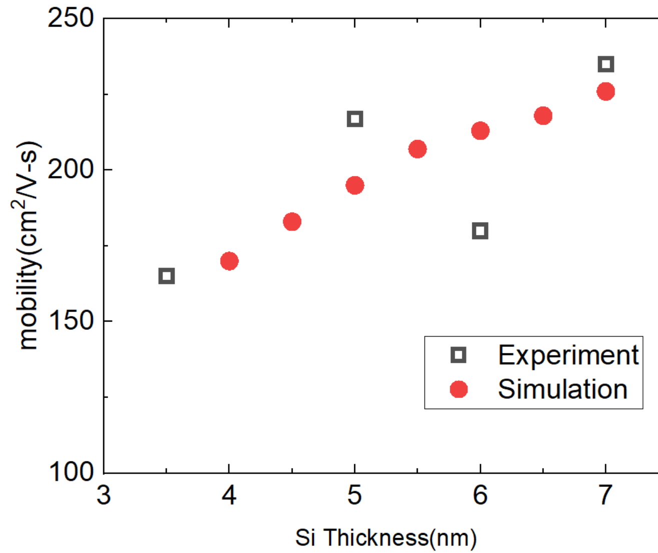


Fig.2.2 Simulated vs experiment electron mobility. Mobility extracted as the weighted average over the charge in the channel

2.2 Quantum model

The ultra-thin silicon body of FinFET and NSFET leads to severe quantum effects. In this study, a 2D Schrodinger solver is set up to capture the quantum confinement effect in the channel region of ultra-scaled FinFET and NSFET. The 2D Schrodinger solver is described in [9]. The 3D device structure is divided into 2D cross-sections. From the band profile, quasi-Fermi potential, and lattice temperature, the 2D Schrodinger solver computes the quantum mechanical carrier density in the

2D cross-section. The result is then passed back to the 3D device simulator. The Poisson equation and quantum-mechanical potential correction are calculated based on the carrier density. The procedure of the quantum aware TCAD simulation framework is shown in Fig. 2.3.

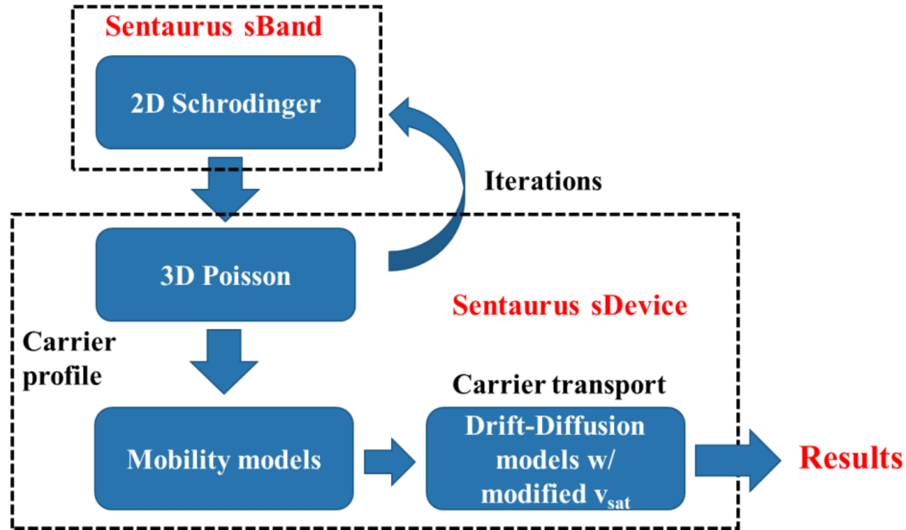


Fig. 2.3 Simulation procedure of 2D Schrodinger solver.

The quantum model is verified with the Nanosheet FET results in [14]. The calibrated capacitance vs silicon thickness is shown in Fig. 2.4. The calibrated parameters are taken from [36] The simulated gate capacitance (C_G) can fit the experiment results. This indicated that the carrier density and quantum capacitance can be accurately modeled by the Schrodinger solver.

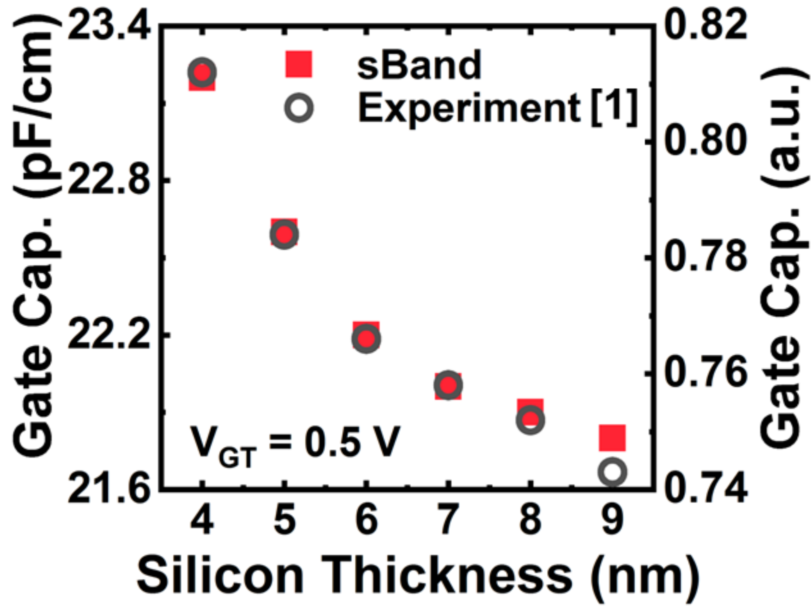


Fig. 2.4 The calibrated capacitance vs silicon thickness [36]

2.3 Traps-assisted tunneling model

The reliability issues of the scaled gate dielectric in NSFET are highly linked to traps in the gate dielectric. The trap-assisted tunneling (TAT) is one of the main mechanisms of carrier transport across the gate stack. The TAT model is explained and calibrated in this section.

In general, the trap occupation rate (f) is determined by the capture and emission rate:

$$\frac{df}{dt} = (1 - f)c - fe \qquad (2.2)$$

where c is the capture rate and e is the emission rate. The capture rate is determined by the tunneling process and the emission rate can be obtained from the capture rate using the principle of detailed balance. With a higher capture rate, the trapping time constant will be smaller, and the traps can be more easily filled.

For elastic tunneling, the initial and final states have the same energy level. The tunneling process is demonstrated in Fig 2.5. As the gate bias increases, the crossing point between the trap level

and the Fermi level in silicon shifts towards the silicon/dielectric interface. This indicated that the trapping occurs closer to the interface due to direct tunneling with a shorter tunneling distance, smaller barrier, and higher capture rate. According to [15], the capture rate can be modeled as:

$$C_{el}(E,z) = D(E)f(E)W_{el}(E,z) \qquad (2.3)$$

$$W_{el} = \frac{2\pi}{\hbar}T(E,z)^2\delta(E_T - E) \qquad (2.4)$$

Where D(E) is the density of states, f(E) is the Fermi function and T(E,z) is the tunneling rate at location z, which can be calculated by Wentzel–Kramers–Brillouin (WKB) approximation:

$$\mathrm{T(E,x)} = \exp\left(-2\int_0^x dx\sqrt{\frac{2m}{\hbar^2}(V(x)-E)}\right) \qquad (2.5)$$

where x is the distance between channel and trap V(x)-E indicates the barrier height along the tunneling path.

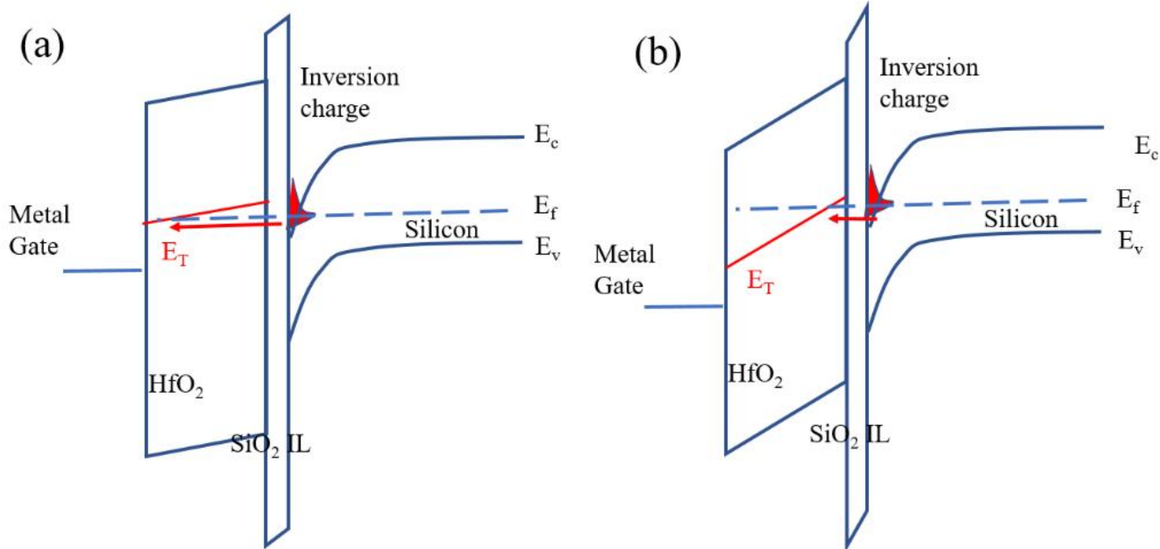


Fig 2.5. Capture involving direct tunneling at (a) low gate bias and (b) high gate bias

The process of inelastic tunneling is shown in Fig. 2.6. The electron in the channel can tunnel to a trap at a lower energy level by emitting one of more phonons with the energy $\hbar\omega$. Therefore, the capture rate is determined by phonon energy, the number of an emitted phonon (or energy

difference between electron and trap) and tunneling probability. Due to phonon emission, a larger range of trap energy can be involved in the tunneling process, and the thermal dependence of this process can be important. In [9], with single energy phonon approximation, the transition rate between electron and trap state can be modeled as:

$$W_{ph} = \frac{\pi}{\hbar} S \left(1 - \frac{p}{S}\right)^2 G(E_T, \hbar\omega) T(E)^2 \qquad (2.6)$$

Where $\hbar\omega$ is the phonon energy, p is the number of emitted phonon numbers, S is the Huang-Rhys factor and $G(E_T)$ is a function of trap energy, temperature, and phonon energy as described in [16].

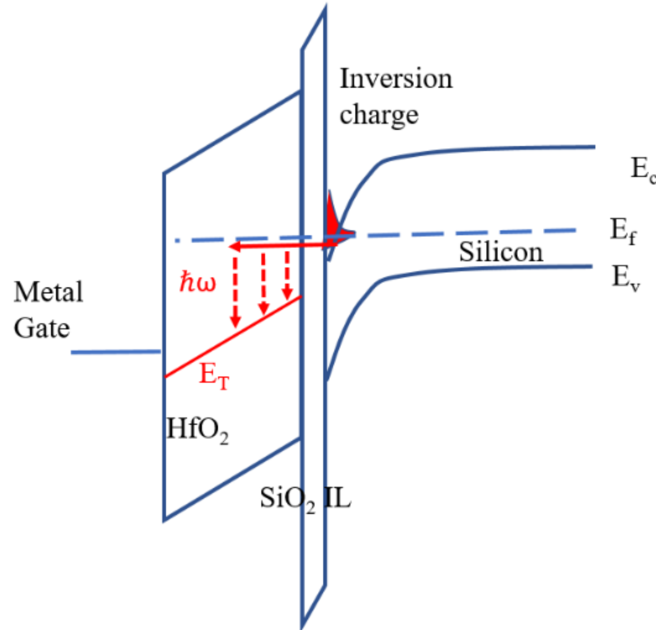


Fig 2.6. Inelastic tunneling assisted by phonon emission with energy $\hbar\omega$

2.4 Conclusion

In this chapter, the detailed model setup and calibration are demonstrated as the foundation of the TCAD analysis in the later chapters. A DD model with modified saturation velocity is used to

simulate the velocity profile in the ultra-scaled channel in FinFET and NSFET. A 2D Schrodinger equation solver is used to self-consistently solve the carrier density and potential in the thin silicon body. The trap-assisted tunneling model considering inelastic phonon emission is used to simulate the gate leakage current. The models are calibrated to experimental data so reliable simulation results can be obtained.

CHAPTER 3
FinFET and NSFET simulation

In this section, the device performance of sub-5nm n-type NSFET is simulated and compared to n-type FinFETs. The simulation platform is first fitted to 10nm FinFET. Based on the calibrated parameters, the performance of 5nm FinFET is projected. Stacked Nanosheet FET (NSFET) technology is going to replace FinFET in sub-5nm nodes. With aggressive scaling of both device geometry and effective oxide thickness (EOT), it is crucial to understand how the change of important parameters affects the performance and reliability of this novel device. In this section, the sub-5nm NSFET structure is set up, simulated and compared to FinFET. The performance of NSFET vs important design parameters is studied based on the simulation results.

3.1 10nm FinFET simulation

The simulated 3D structure of 10nm n-type FinFET is shown in Fig. 3.1. The device structure and design parameters are determined based on the Intel 10nm technology reported in [17]. The cross-section of the fin is matched to the SEM picture in [17]. The sidewall taper angle is 2°. The gate dielectric is a dual-layer stack with a SiO_2 interfacial layer and HfO_2 high-k layer. Silicon nitride is used as the spacer dielectric between gate and source/drain.

As mentioned in Chapter 2, a drift-diffusion model with modified saturation velocity is used to capture the carrier transport in short channel devices. A 2D Schrodinger equation solver is used to account for the quantum confinement of carriers in the narrow fin structure. Source/drain doping

and contact resistivity (ρ_c) were used as fitting parameters to calibrate the simulation results to the experimental data.

The simulated device transfer and output characteristics are shown in Fig. 3.3. The simulated I-V can fit the reported data with an error smaller than 5% for both on and off regions. This indicates that a combination of modified DD mode and 2D Schrodinger solver can correctly model the device performance of scaled FinFET. The calibrated contact resistance (ρ_c) is $1x10^{-8}\Omega$-cm. The source/drain doping concentration is 3×10^{20} cm^{-3}.

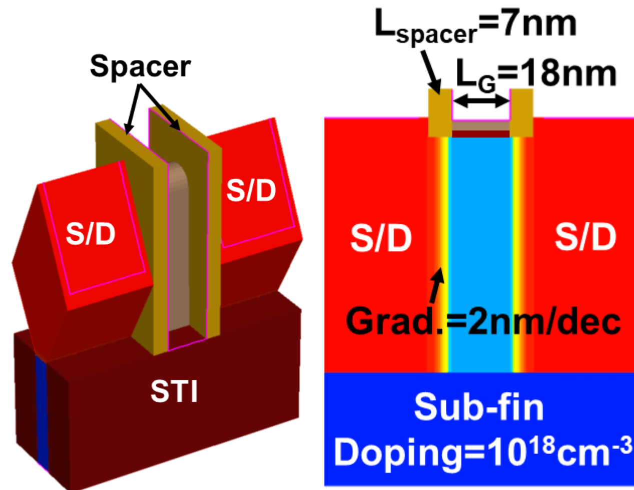


Fig. 3.1 (a) Simulated 3D structure of 10nm FinFET (b) cross-section along the channel direction

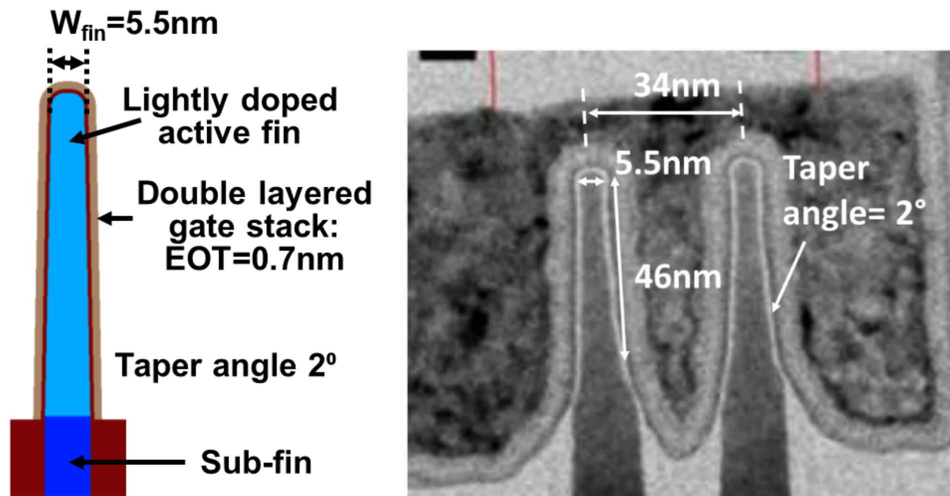


Fig. 3.2 Fin cross-section of the simulated FinFET and TEM in [17]

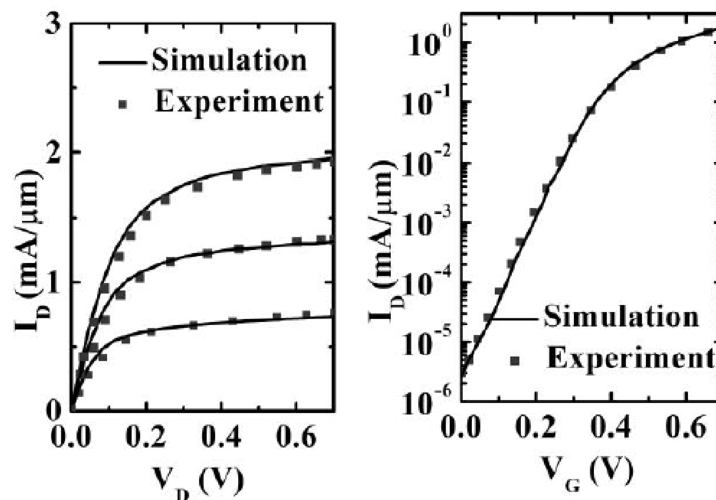


Fig 3.3 Simulated vs experiment results of 10nm FinFET

3.2 Sub-5nm NSFET simulation

Based on the results of the 10nm FinFET simulation, 5nm n-type FinFET and sub-5nm n-type NSFET are simulated. The structure of an NSFET for sub-5nm technology is shown in Fig. 3.4. The structure is based on the reported data in [3]. A 3-nanosheet stack is used to improve the

effective channel width per footprint. The thickness of each nanosheet is set to be 5nm, which is the same as the fin width of a 5-nm node FinFET. With the ultra-thin nanosheet, the short channel effect can be greatly reduced, allowing further scaling of the channel length. The width of the silicon nanosheet is set to 30nm. With the gate-all-around silicon body, the short channel effect can be further reduced. The total height of the 3-nanosheet stack is 50nm, which is the same as a typical FinFET technology. High-k metal gate technology is used for the gate stack. The gate dielectric consists of SiO_2 interfacial layer and HfO_2 high-k layer. The key parameters of the sub-5nm stacked NSFET are summarized in Table 3.1. For comparison, a 5nm FinFET is also simulated. The projected 5nm FinFET parameters ([18]) were also added as a comparison.

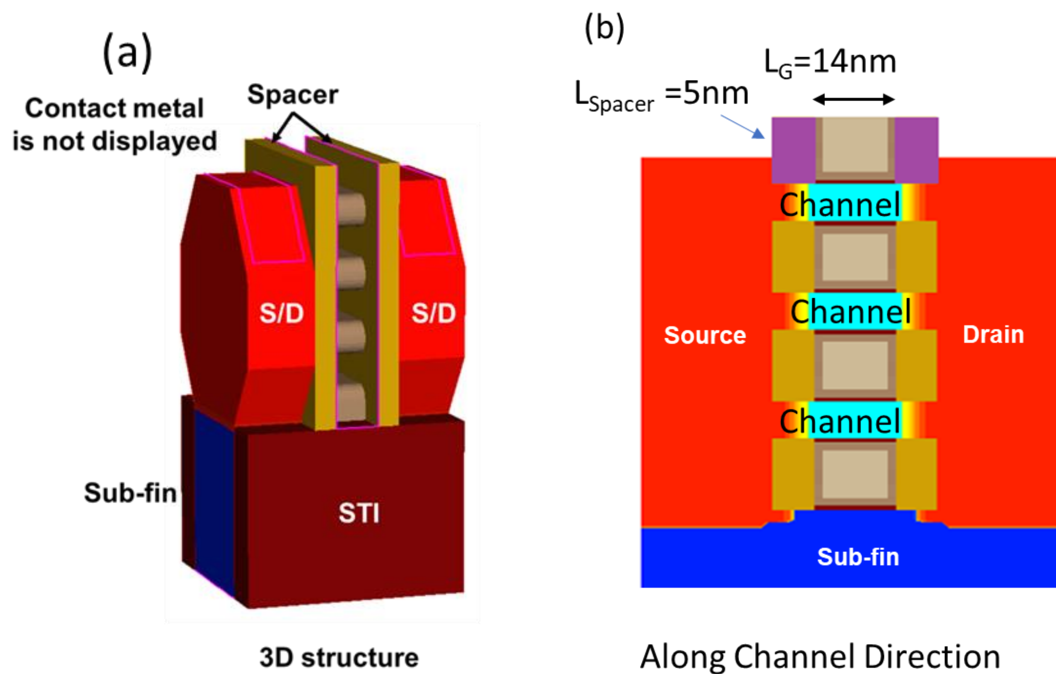


Fig. 3.4 (a) Simulated structure of sub-5nm NSFET (b) cross-section along the channel direction of the NSFET

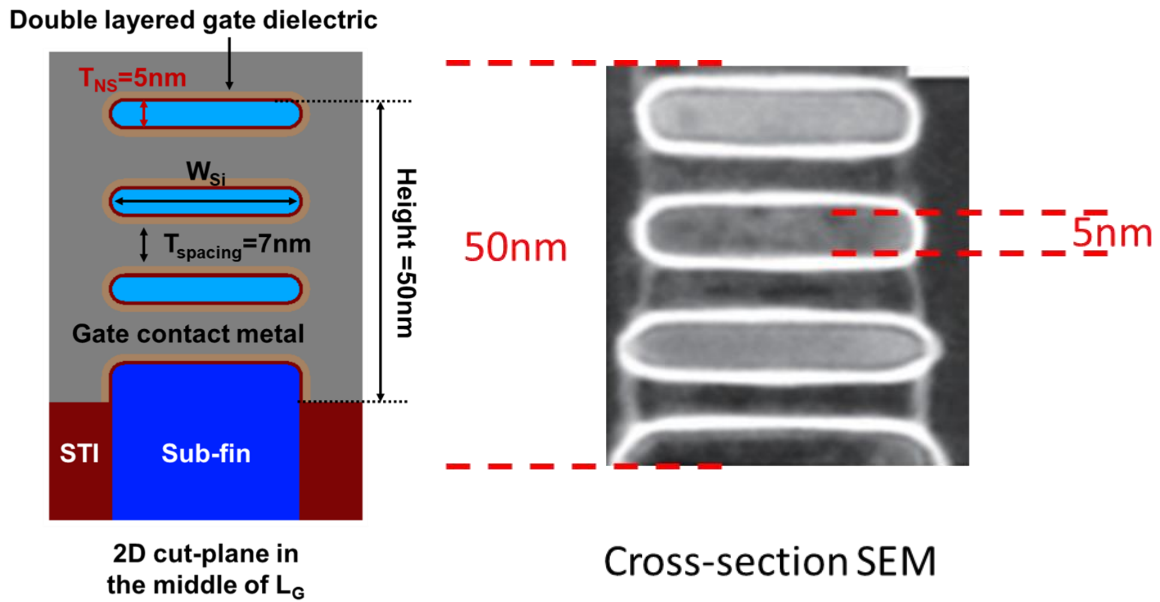


Fig. 3.5. Fin cross-section of the simulated FinFET and TEM in [3]

Table 3.1 The key parameters of the sub-5nm and 5nm FinFET

Sub-5nm NSFET		5nm FinFET	
H_{NS}	50nm	H_{fin}	50nm
L_G	14nm	L_G	14nm
L_{spacer}	5nm	L_{spacer}	5nm
EOT	0.6nm	EOT	0.6nm
T_{NS}	5nm	T_{fin}	5nm
NS pitch	52nm	Fin pitch	26nm
W_{NS}	30nm		

The simulated transfer characteristics of sub-5nm NSFET and 5nm FinFET are shown in Fig. 3.6. The Ion is normalized to the device footprint (fin pitch) with fixed Ioff. As shown in Fig. 3.6, the normalized Ion of NSFET is 20% higher than FinFET. This is due to the better SS in the sub-threshold region. With the gate-all-around structure, the short channel effect is suppressed in

NSFET. The SS of the NSFET is 68.5 mV/dec whereas the SS of FinFET is 74.6 mV/dec. Therefore, the Ion/Ioff is better for NSFET than FinFET with the same L_G.

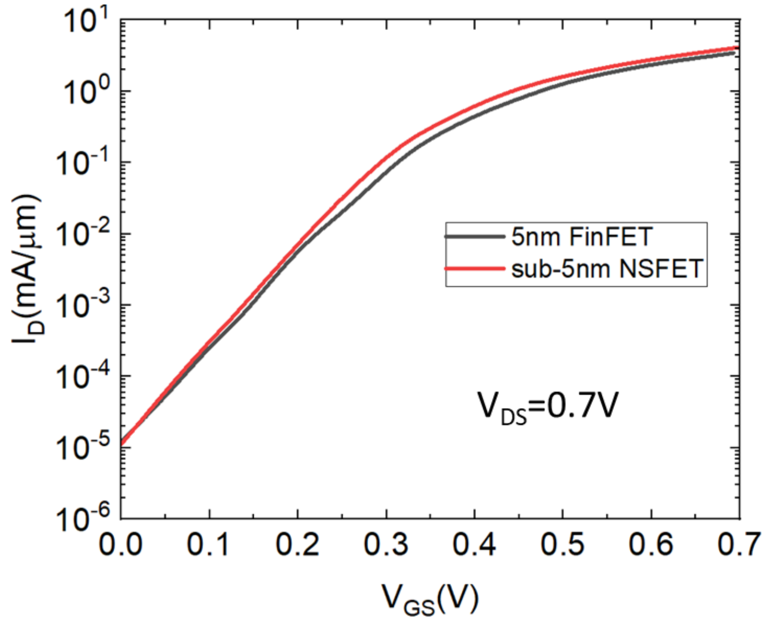


Fig. 3.6 The simulated transfer characteristics of 5nm FinFET and sub-5nm FinHBT

3.3 Parasitic resistance analysis

As shown in Fig. 3.7, the parasitic resistance can be divided into 3 in-series components. Each component of the $R_{parasitics}$ is subsequently extracted with the total $R_{parasitics}$ expressed as:

$$R_{parasitic} = R_{contact} + R_{SD} + R_{spacer} \qquad (3.1)$$

where $R_{contact}$ is the contact resistance, R_{spacer} is the resistance of the spacer region (also known as the gate underlap region), and R_{SD} is the total resistance of the heavily doped source/drain.

The extracted parasitic components of 10nm FinFET, 5nm FinFET and sub-5nm NSFET are shown in Fig. 3.8. The contact resistivity (ρ_c) is set to be $1x10^{-9}\Omega cm$, assuming the laser annealing technique ([19]) is used. With the small ρ_c, the contact resistivity is greatly suppressed. For all 3

devices, the biggest contributor of parasitic resistance is R_{spacer}. As FinFET scales from 10nm node to 5nm node, the $R_{contact}$ will increase by 100% due to the decrease of the source/drain size as contact poly pitch (CPP) scales down. Meanwhile, R_{spacer} decreases slightly because of the shorter spacer length. As a result, contact resistance is more important in scaled devices. For sub-5nm NSFET, the resistance per device is smaller than 5nm FinFET. This is due to the larger device size with the wide nanosheet structure. The R_{spacer} is about 50% smaller than the 5nm FinFET. This is due to the increased conductive area of the spacer region in NSFET. However, the $R_{contact}$ and R_{SD} were only reduced by about 20% as the resistance of these regions are limited to the size of the regrown source/drain. Moreover, carriers in source/drain of NSFET funnel into the 3 individual nanosheets when entering the channel. Therefore, the effective resistance of the source/drain region is higher than FinFET. As a result, the R_{SD} and $R_{contact}$ occupy a larger portion of the total parasitic resistance. However, R_{spacer} is still the biggest contributor to $R_{parasitic}$, as in 10nm and 5nm FinFET. Advanced spacer region doping technique like described in [35] can potentially improve the performance by reducing R_{spacer}.

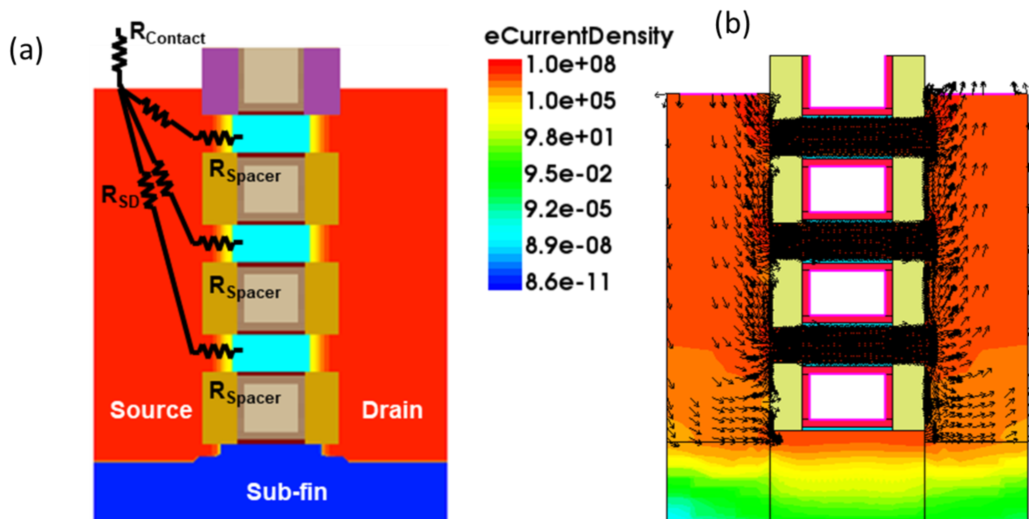


Fig. 3.7 (a) Parasitic resistance components (b) electron current density and direction in NSFET

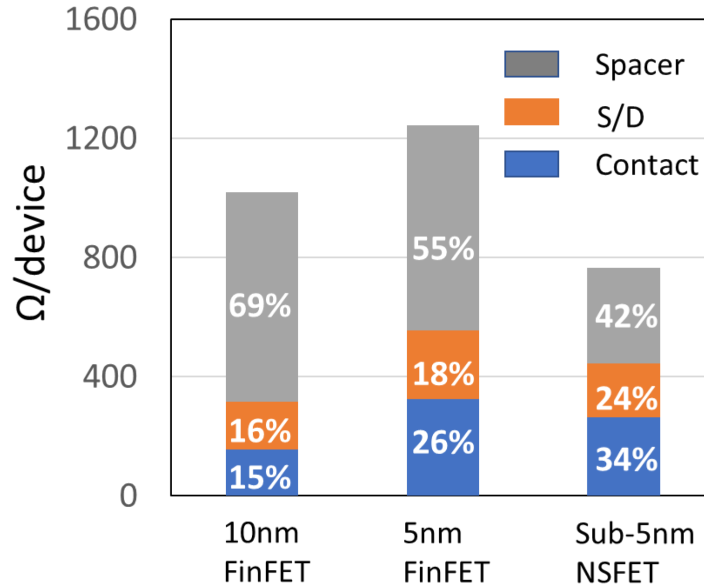


Fig. 3.8. Comparing parasitic components of 10nm FinFET, 5nm FinFET, and sub-5nm NSFET

3.4 Channel length scaling

Channel length (L_G) is a crucial design parameter of NSFET. In this work, the effect of scaling L_G from 14nm to 10nm is analyzed through TCAD. The results are shown in Fig. 3.9 - Fig. 3.12. T_{NS} is fixed to be 5nm in this simulation. Key device characteristics like Ion, Ioff, SS are compared. As the channel length decreases from 14nm to 10nm, the $_{Ion}$ increases from 3.12 mA/um to 3.49 mA/um with fixed work function (WF). This is due to a combination of the reduction of channel resistance and the V_{th} roll-off. When L_G reduces, there is a higher electric field in the channel region and thus higher carrier transport velocity and smaller channel resistance. On the other hand, the SS also decrease from 68.6 mV/dec to 73.4 mV/dec due to the short channel effect. This increases I_{off} by about 10 times. Moreover, the device is more subjected to DIBL at smaller L_G. Therefore, there is a larger Vth roll-off when L_G reduces, increasing both I_{on} and I_{off}. To further compare the effect of L_G scaling, the I_{on} at a fixed I_{off} is shown in Fig. 10. It can be observed that I_{on}/I_{off} decreases as L_G scales down due to more severe short channel effect.

The total capacitance and intrinsic delay (CV/I) vs L_G are shown in Fig. 3.12. As L_G decreases the intrinsic gate capacitance also decreases. This results in a smaller intrinsic delay even with reduced Ion at smaller L_G. Therefore, while further scaling L_G cannot improve I_{on}/I_{off}, it still benefits the digital performance of NSFET by reducing the intrinsic delay.

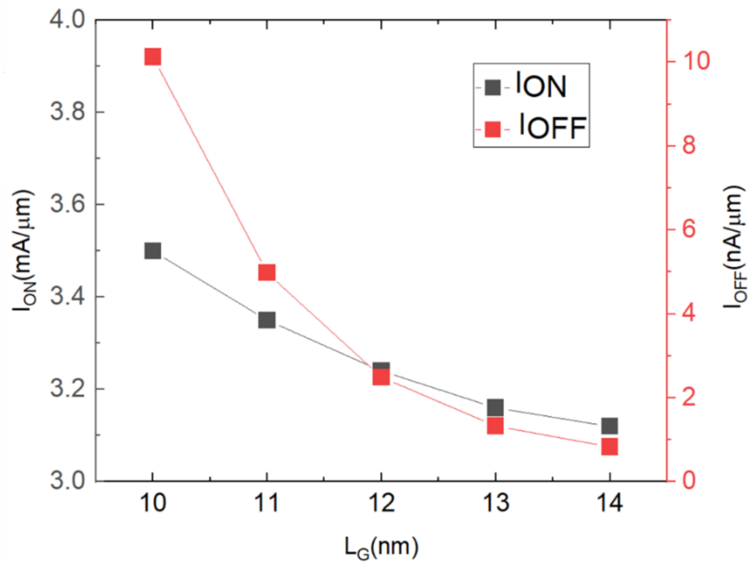


Fig. 3.9 Ion and Ioff of n-type NSFET versus L_G. The gate work function is fixed at 4.5eV.

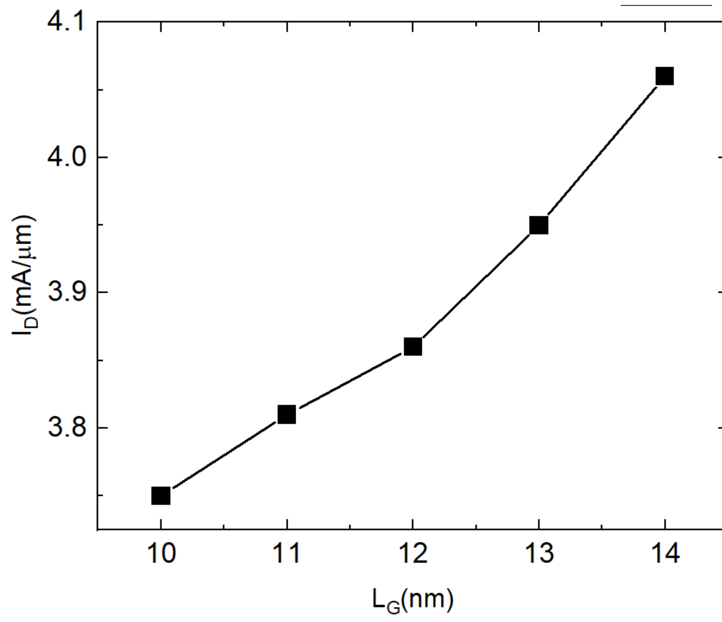


Fig. 3.10 Ion at fixed Ioff=10nA/um versus L_G.

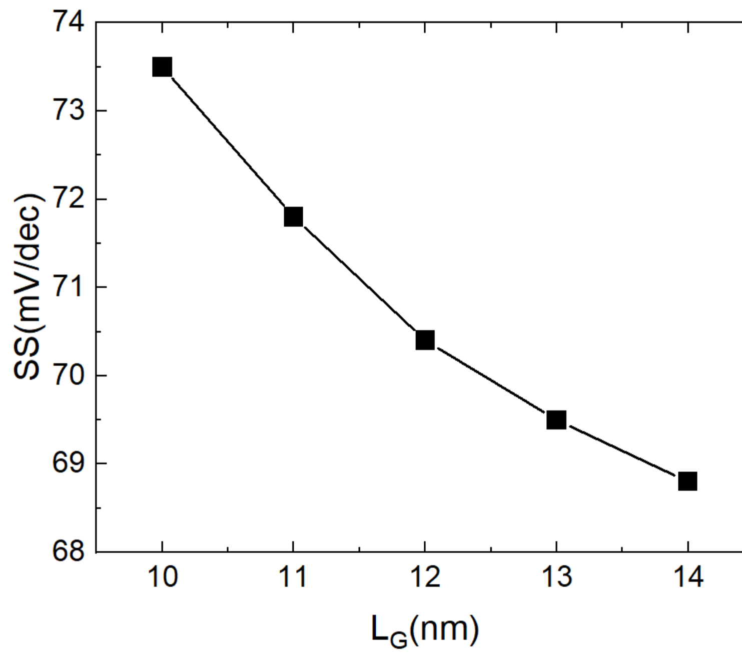


Fig. 3.11 Subthreshold swing (SS) versus L_G

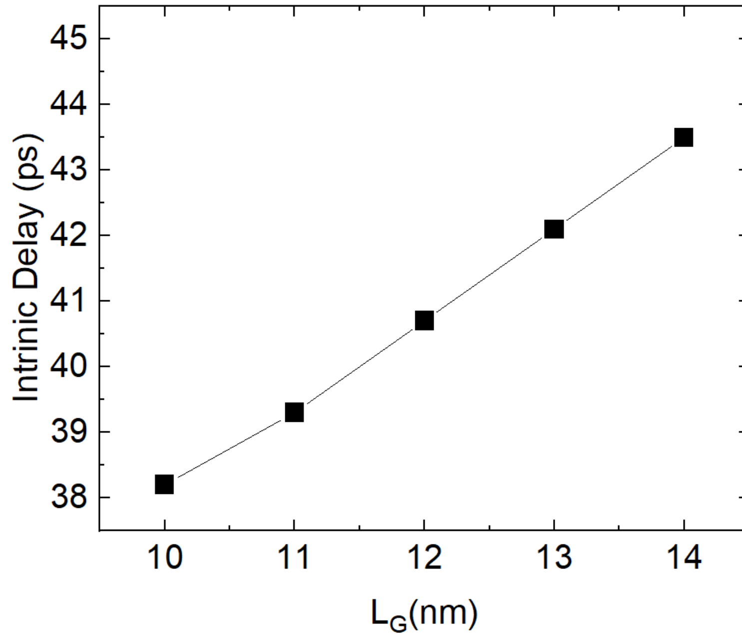


Fig. 3.12 Intrinsic delay (CV/I) versus L_G

3.5 T_{NS} scaling

Nanosheet thickness (T_{NS}) is another key design parameter of NSFET. The effect of scaling T_{NS} is shown in Fig. 3.14 -Fig. 3.16. When TNS reduces, the conductive area per nanosheet decreases, leading to larger series resistance. This results in the smaller I_{on}. On the other hand, the carrier density in the NS is higher when T_{NS} reduces. As the silicon volume reduces, the carrier moves closer to the Si/gate dielectric interface. There is a stronger coupling between the gate and the carriers in the channel. The short channel control of NSFETs improves as T_{NS} reduces.

The I_{on} & I_{off} of n-type NSFET with different T_{NS} is shown in Fig. 14. As T_{NS} increase from 4nm to 6nm, the Io_n increases by 28% with fixed WF. This is due to the increased conduction area and reduced parasitic resistance in the spacer region. On the other hand, I_{off} increases by about 10 times

due to the degraded short-channel control. The SS and Ion at fixed Ioff are shown in Fig. 3.15 and Fig 3.16. I_{on}/I_{off} increases, and SS decreases as T_{NS} decreases due to better short channel control.

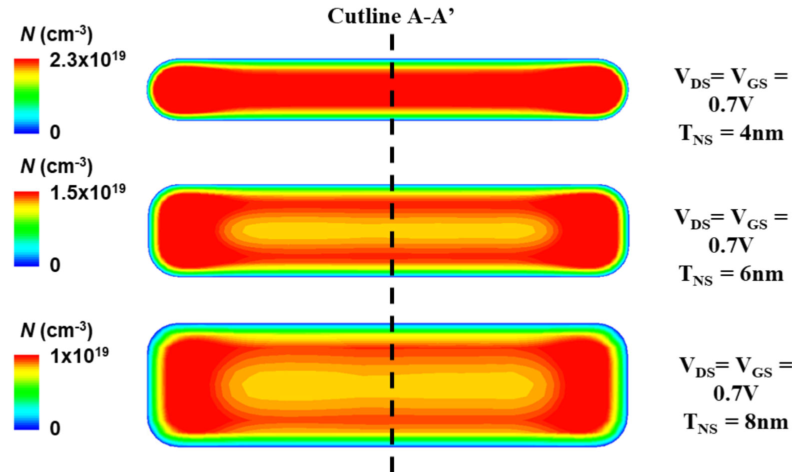


Fig. 3.13 Carrier density distribution at nanosheet cross-section at V_{GS}=0.7V. The carriers move closer to the channel/oxide interface as TNS decreases., leading to improved electrostatic control

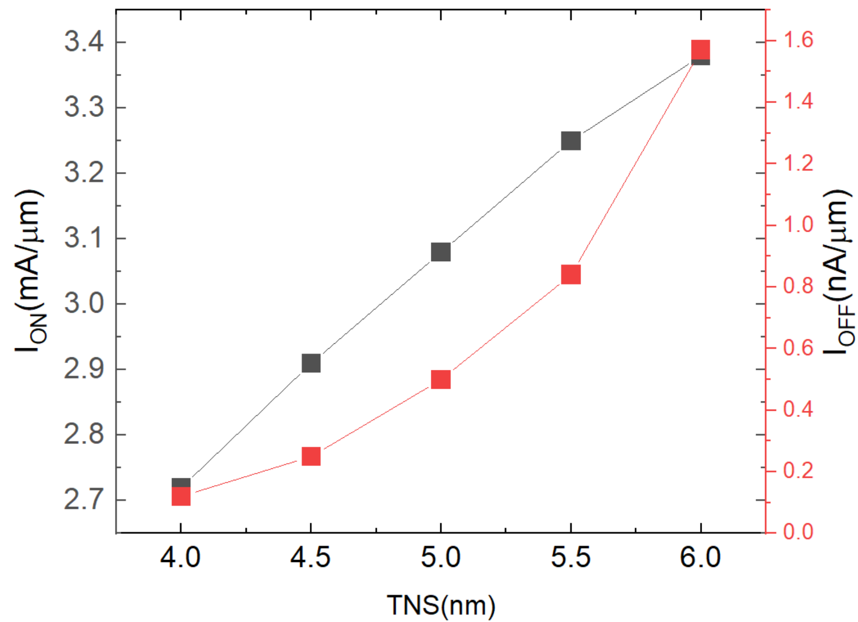


Fig. 3.14 Ion and Ioff of NSFET versus T_{NS}. The gate work function is fixed at 4.5eV.

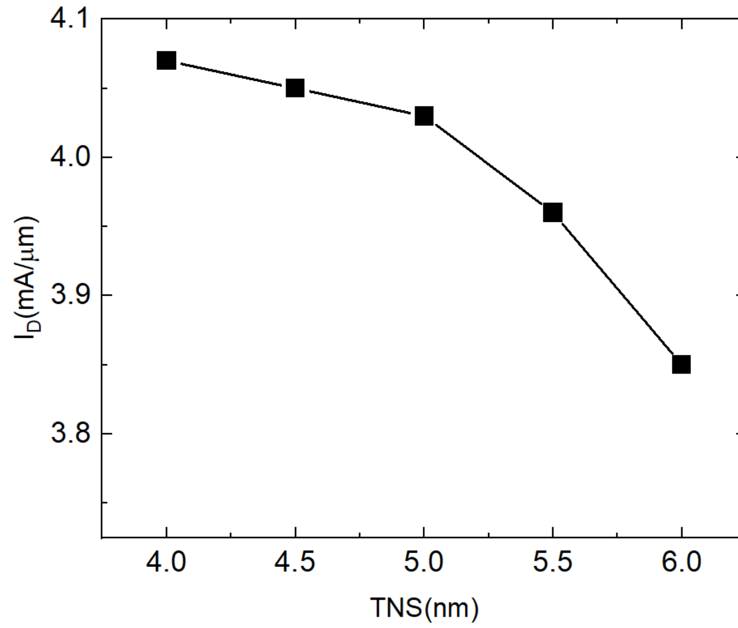


Fig. 3.15 Ion at fixed Ioff=10nA/um versus T_{NS}.

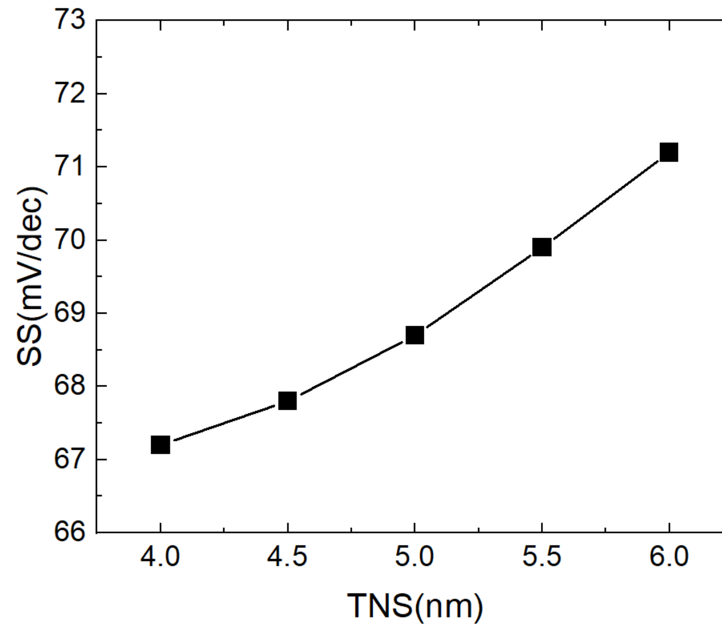


Fig. 3.16 Subthreshold swing (SS) versus T_{NS}

3.6 EOT scaling

Scaling down EOT can both affect I_{on} by increasing the gate capacitance (C_{gate}) and reduce I_{off} by having a stronger gate-channel coupling. NSFETs (L_G=14nm) with EOT from 0.8nm to 0.5nm have been simulated. The charge distribution in the nanosheet at V_{GT}=0.5V is shown in Fig. 3.17. As EOT scales down, the electric field in the gate dielectric increases. As a result, the charge peak moves closer to the Si/gate dielectric interface, enhancing the gate-channel coupling. Due to quantum effect, the intrinsic gate capacitance (C_{Gint}) can be described as the capacitance of the gate stack (C_{gate}) and silicon quantum capacitance (C_{si}) in series. The total EOT (EOT_{total}), which corresponds to the C_{Gint}, can be described as:

$$EOT_{total} = EOT + EOT_{Si} \qquad (3.2)$$

where EOT_{Si} is the EOT corresponding to C_{Si}. As shown in Fig. 3.18, EOT_{si} reduces from 0.7nm to 0.6nm as EOT reduces from 0.8nm to 0.5nm. This indicates that as EOT scales down, the NSFET is less susceptible to quantum confinement due to the stronger gate-channel coupling. EOT_{total} reduces from 1.5nm to 1.1nm, indicating a 35% increase in the intrinsic gate capacitance. Apart from that, the enhanced gate-channel coupling leads to an improved SS from 71.3 mV/dec to 67.1 mV/dec as EOT scale from 0.8nm to 0.5nm. A combination of higher C_{gate} and lower SS leads to a higher I_{on}/I_{off}. The I_{on} at fixed I_{off} and intrinsic delay (CV/I) vs EOT is shown in Fig. 3.19. The I_{on} at fixed I_{off} increased by 28%. The intrinsic delay, calculated by CV/I at V_G=V_D=0.7V, decreased by 3.5% as EOT scales down.

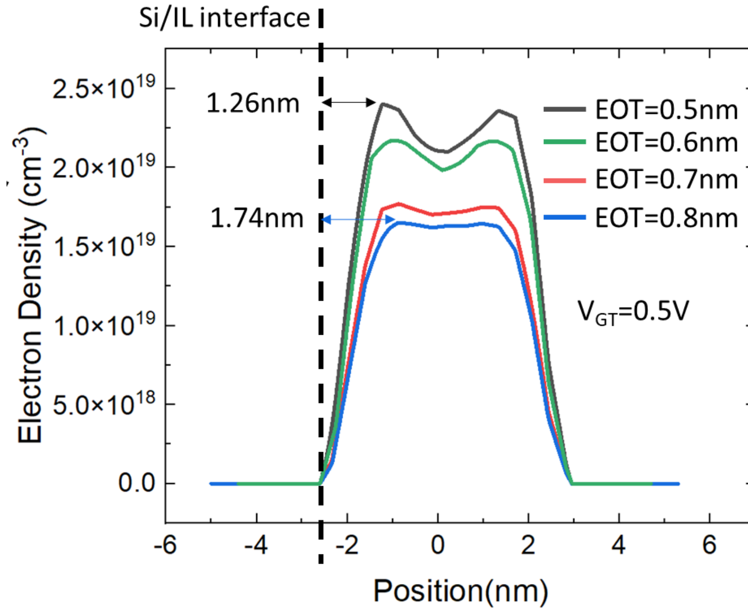


Fig. 3.17 Charge distribution in the channel. Charge peak moves closer to the channel/gate dielectric interface as EOT scales down

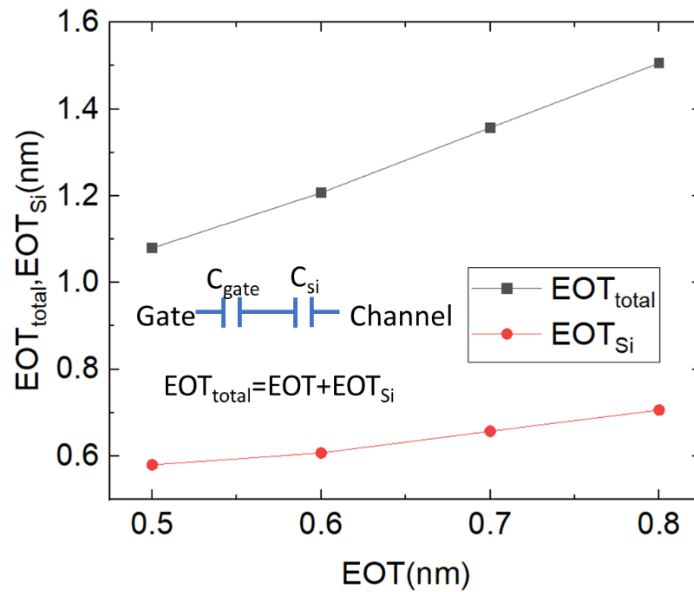


Fig. 3.18 EOT_{total} and EOT_{si} versus EOT of n-type NSFET.

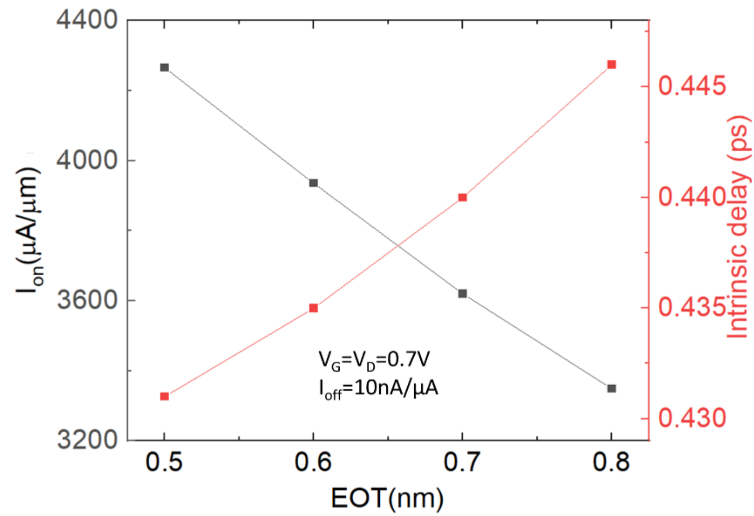


Fig. 3.19 Ion and intrinsic delay versus EOT

3.7 Conclusion

In this section, the device performance of sub-5nm NSFET is simulated and compared to FinFETs. Simulation shows that NSFET has a better I_{on}/I_{off} and SS over FinFET with the same channel length and body thickness due to the superior electrostatic control of the gate-all-around structure. Parasitic resistance analysis shows that R_{spacer} is the main contributor to parasitic resistance. Design parameter analysis shows that Ion/Ioff of NSFET can be improved by enhancing electrostatic control, like increasing L_G, reducing TNS, and scaling down EOT. Scaling EOT is considered a major pathway to further scaling, as it increases the carrier density in the channel, enhances electrostatic control, and suppresses the effect of parasitic capacitance.

CHAPTER 4
NSFET reliability study

As discussed in Chapter 3, scaling EOT can both improve on-state current and off-state electrostatic control of NSFET. Aggressive scaling will increase the parasitic capacitance. To alleviate this effect, a smaller EOT is needed to increase intrinsic gate capacitance. However, ultra-scaled EOT can also lead to reliability issues like gate leakage current, PBTI (NBTI), and TDDB. In the early 2000s, high k/metal gate (HK/MG) technology has been used to improve reliability and enable EOT scaling to sub-1nm. However, as the EOT scaling continues, reliability has become an increasingly important issue even with HK/MG. In this work, a TDDB model based on the Sentaurus TCAD platform is developed. Thermal-chemical stress-induced defect generation model is used to capture the trap generation under continuous voltage stress. Trap-to-trap tunneling and trap-assisted tunneling models are used to simulate the carrier transport through dielectric layers with defects. The model is calibrated and applied to 3D NSFETs. The TDDB behavior of NSFETs with different corner rounding radii is studied closely. This study aims to provide a guidance to the geometry related TDDB behavior in NSFET.

4.1 Literature review

TDDB effect in thin gate dielectric has been studied by multiple different methods. One common practice is the percolation method, like the example in [20]. The gate dielectric is represented by cubic grids with cell size equal to the lattice constant of the dielectric material. Defects are then generated randomly in the grids. When there is a pathway between gate and semiconductor where every cell on the pathway contains a defect, a dielectric breakdown happens. This method has been

proven to predict the Weibull TDDB behavior of gate stacks with different thicknesses. However, as a statistical model, it does not provide detailed physical information like the electric field, gate current, and device geometry-related effects.

Another method is full atomic level kinetic Monte Carlo simulation. As discussed in [21], charge trapping, charge transport, and trap generation can all be captured based on ab-initio calculation. The generation, diffusion, and recombination of all the atomic species are calculated individually through a kinetic Monte Carlo framework. While this method can most accurately describe all the important physics in TDDB, it requires excessive computational power and works best with a simple structure, like MIM stack. Therefore, it is very hard to utilize this method on a complicated 3D structure like a NSFET.

In this work, a discrete trap method is based on TCAD and external code. This method is inspired by the approach in [22]. Sentaurus TCAD is used to solve the Poisson equation, current continuity equation, and trap tunneling. The electric field in the gate stack is then extracted for trap generation rate calculations with a thermal-chemical stress-induced defect generation equation. The model allows for detailed physical analysis of complicated 3D device structures.

4.2 Discrete trap TDDB model

Two models are used to calculate the gate leakage current. The trap-assisted tunneling, as described in Chapter 2, is used to simulate the gate current at low trap density. The model is calibrated to MOSCAP experiment data with different trap densities, as shown in Fig. 4.1. The traps are assumed to concentrate near the interface between interfacial layer and High-k layer (IL/HK). The calibrated parameters of the traps are shown in Table. 4.1. The trap species in HfO_2 and SiO_2 are assumed to be oxygen vacancies and the energy levels are reported in [25]. By modifying trap

density in the gate stack, the simulated I_G can fit experiment results. The calibrated TAT model is used to calculate the gate leakage current at the initial state.

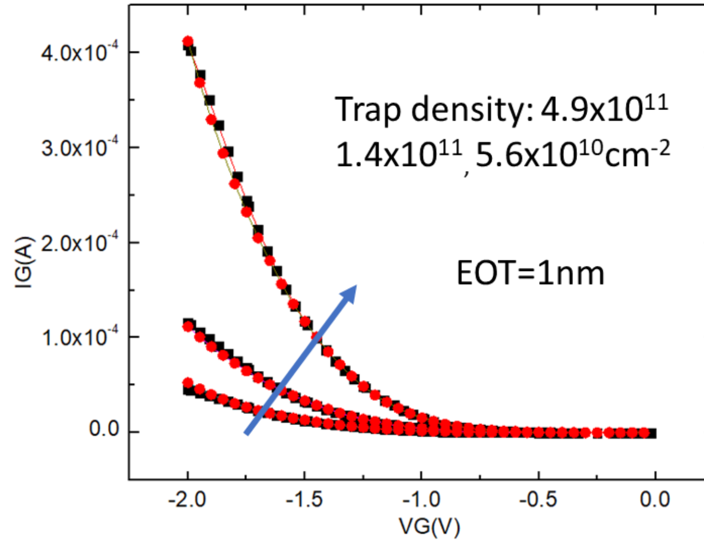


Fig. 4.1 Simulated and experiment MOSCAP gate current vs stress voltage

Table 4.1 Parameters for trap-assisted tunneling

E_T(HfO2)	2.5 eV from E_C
E_T(SiO2)	3.3 eV from E_C
m_e/m_h (HfO_2)	0.42/0.58 [23]
m_e/m_h (SiO_2)	0.3/0.42 [24]
Phonon Energy	0.07 eV
Huang-Rhy Factor	17

As the trap density increase during the breakdown process, the trap-to-trap transport becomes more significant. The trap-to-trap tunneling is described by the inelastic phonon model [9]. The capture rate between two individual traps is described by:

$$c_{j,i}(\vec{x}_i) = \frac{1}{\tau_0} T_{j,i}((\vec{x}_j | E), \vec{x}_i) M_{i,j}(l) \quad (4.1)$$

Where T is the tunneling rate and M is the phonon transition probability. T is calculated using a WKB method in equation (2.5). M is computed by the phonon-assisted tunneling model in equation (2.6). The carrier transport through the gate stack is calculated by direct tunneling, trap-assist tunneling, and trap-to-trap tunneling. This enables gate current simulation with a given trap distribution.

As described in [26], trap generation in the TDDB process is due to the strong dipolar coupling of intrinsic defect states with the local electric field in the dielectric. In [27], the probability of bond breakage can be described by:

$$G(x, y, z) = v \cdot \exp\left(-\frac{E_A - p_0 \frac{2+\kappa}{3} \cdot F(x, y, z)}{k_B T(x, y, z)} \right) \quad (4.2)$$

Where E_A is the activation energy with zero external electric fields, p_0 is the dipole moment and v is the bond vibration frequency. F(x,y,z) and T(x,y,z) are the local electric field and lattice temperature at points (x,y,z).

The trap generation in the gate dielectric is calculated from the generation probability. The gate stack is divided into numerous small regions. The average electric field in the regions is then extracted from the TCAD result. Assume the number of individual cells in each region is N. Probability of generating x traps after time t can be described as Poisson function:

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad (4.3)$$

$$\lambda = NGt \quad (4.4)$$

A random number following the Poisson function is generated from external code as the number of traps generated in the region. The traps are then randomly placed into each region to update the trap profile in the gate stack.

The schematic of the TDDB simulation framework is shown in Fig. 4.2. The simulation starts with the initial trap profile at T=0. The electric field profile in the gate stack and current is calculated by TCAD. The electric field profile is then extracted. External code is then used to calculate the generation rate based on the electric field. With the generation rate, the new trap profile after Δt can be calculated. The updated trap profile is then imported into TCAD, starting the electric field and trap generation calculation of the next Δt cycle. The cycle repeats until the simulated gate current reaches a certain threshold where a dielectric breakdown is considered to occur.

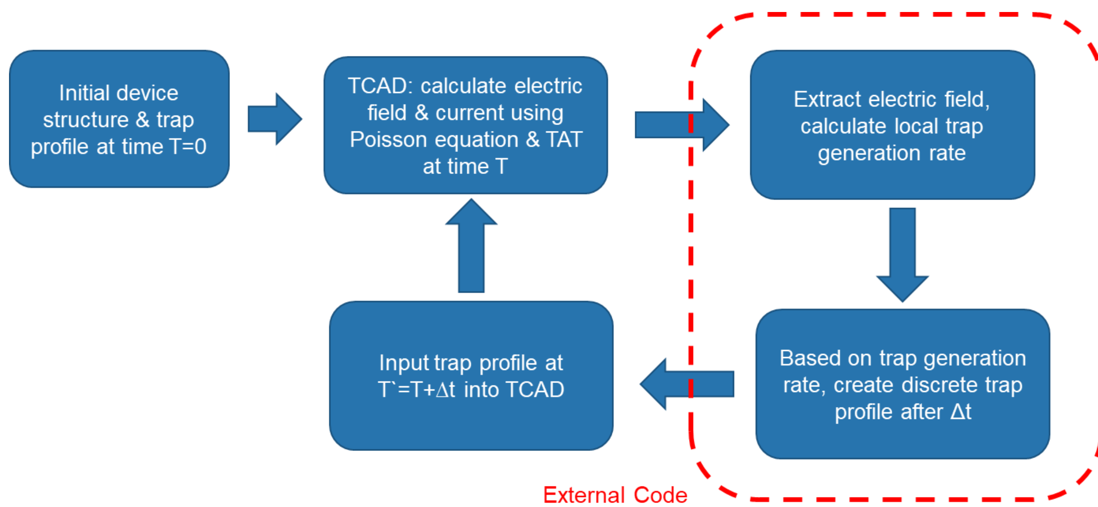


Fig. 4.2 The procedure of TDDB simulation with discrete traps

4.3 Breakdown process and calibration

The TDDB model has been demonstrated in a 2D simulation. A metal-insulator-semiconductor structure is used. A SiO_2-HfO_2 stack is used as the gate insulator. The detailed parameters of the simulation are shown in Table. 4.2.

Table 4.2 Parameters of the 2D TDDB simulation

IL thickness	1nm	Stress voltage	2V
IL dielectric constant	4	Temperature	125°C
HK thickness	2nm		
HK dielectric constant	26		

The simulated process of dielectric breakdown is shown in Fig. 4.3 and Fig. 4.4. In the initial state, there are only initial traps located at the IL/HK interface. In this state, the electric field is uniform across the simulated structure. The next state is the "build-up" state. Traps are first generated in the IL. As the number of random traps in the dielectric increases, trapped charges lead to a non-uniform electric field. Apart from that, the area with more traps will have a larger trap-assisted tunneling current which will cause electric field redistribution. The non-uniform electric field will eventually cause trap generation to be localized, as shown in Fig. 4.4. In this state, the regions with more traps in IL will have a larger electric field in HK. Therefore, the trap generation rate in these regions is higher than in the surrounding regions, forming "weak spots". In the final state, the trap density in the "weak spots" increases rapidly. The size of the "weak spot" also increases, leading to a larger gate leakage current. The "weak spots" will eventually become the breakdown sites in TDDB.

The simulated trap density and gate current vs stress time is shown in Fig. 4.5. The trap density in IL increases gradually as the stress time increases. When the trap density in IL reaches a certain level, the trap density in HK starts to increase rapidly. The increase in HK traps coincides with the increase in gate leakage current, as shown in Fig. 4.6.

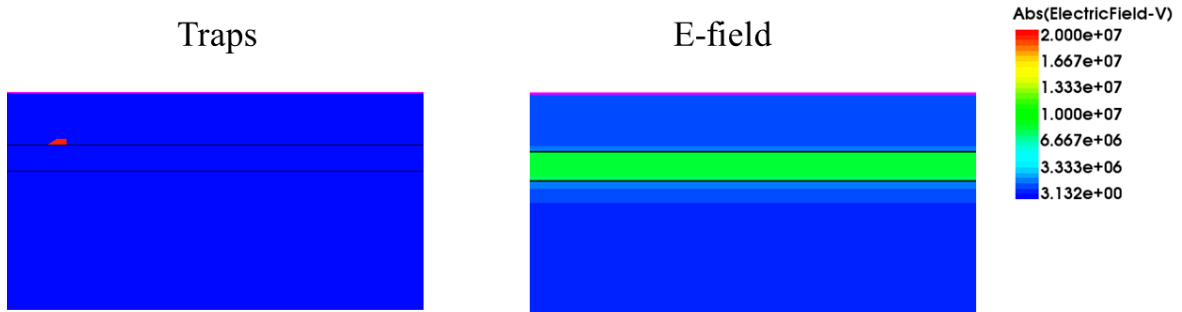


Fig. 4.3 Trap locations and electric field of the initial state (t=0)

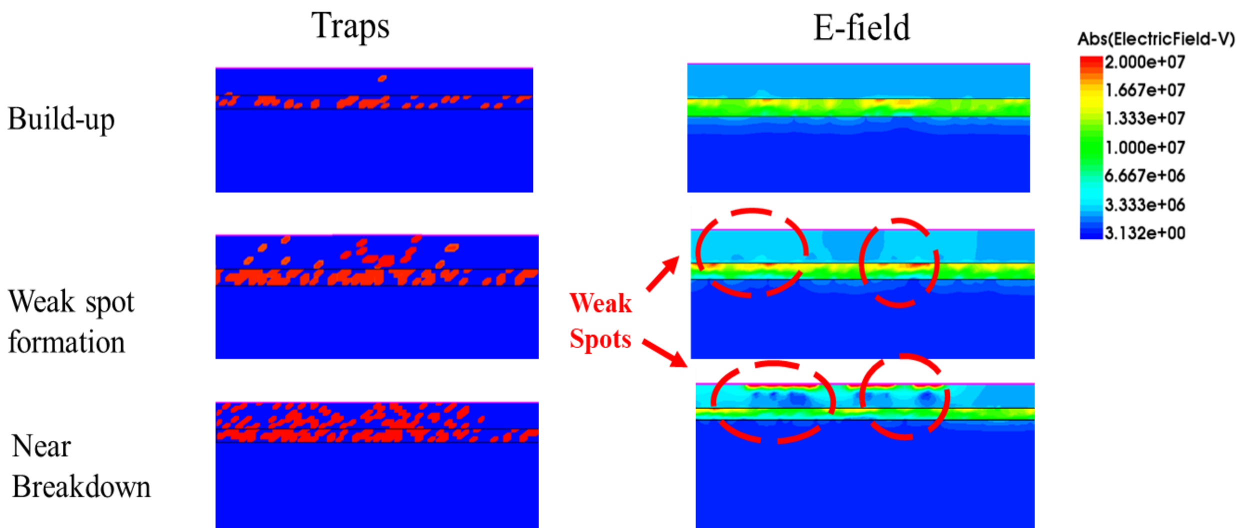


Fig. 4.4 The traps and electric field during the breakdown process

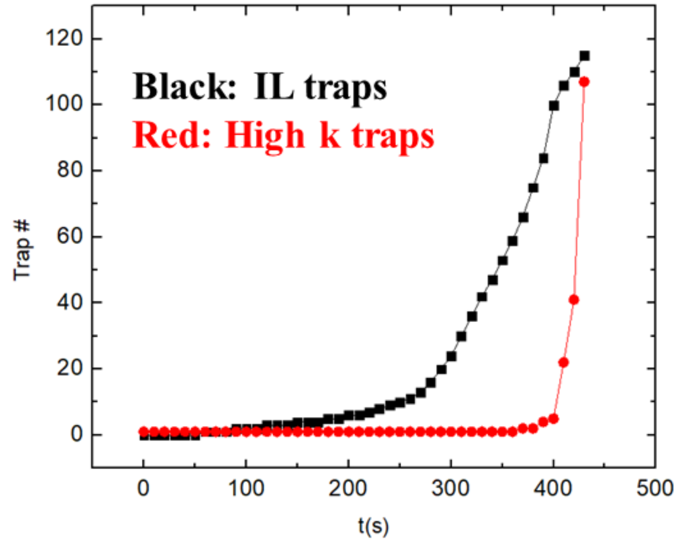


Fig. 4.5 Trap number in IL and HK versus time.

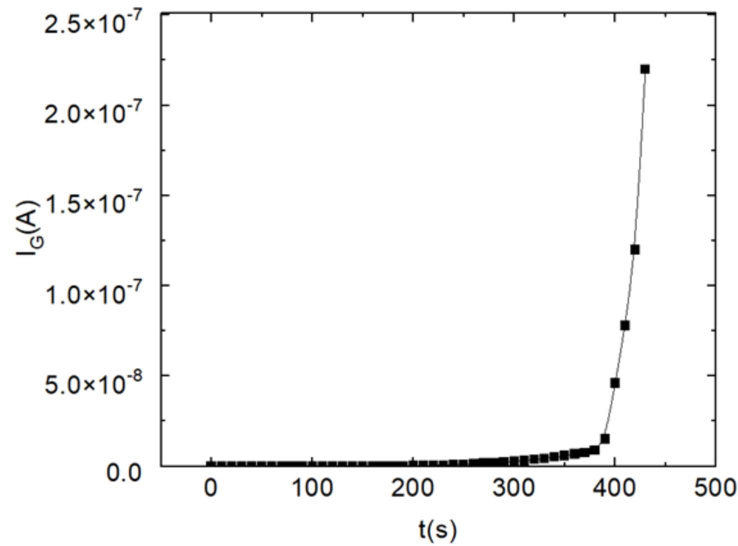


Fig. 4.6 Gate current versus time

The model is calibrated to the experimental data in [28]. The activation energy and trap capture volume are adjusted to fit the simulation results into the range of the experiment data. The results are shown in Fig. 4.8. The simulated gate current vs time fit the range of the experiment data from

V_G=2.2V to 2.5V. The calibrated activation energy of SiO_2 and HfO_2 in equation (4.4) are 2.38eV and 2.8eV, respectively.

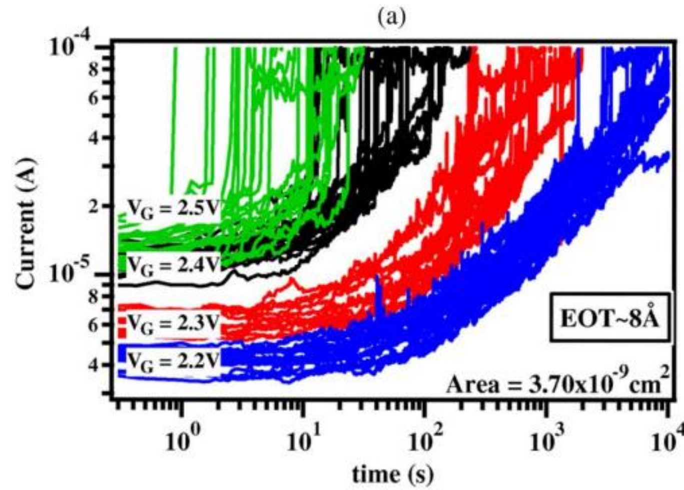


Fig. 4.7 Experiment current [28] vs time at V_G=2.2 – 2.5V

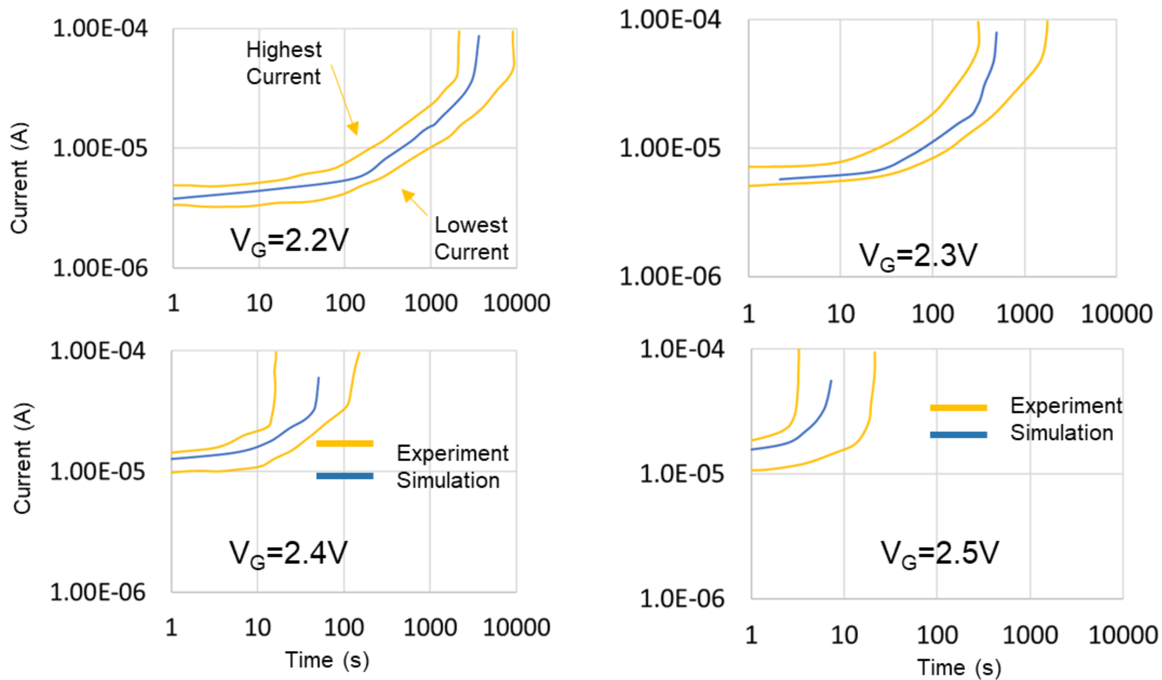


Fig. 4.8 Calibrated I_G versus time simulation. The I_G falls between the lowest and highest current of the experiment data

4.4 NSFET TDDB Simulation

After calibration, the TDDB model is adapted to 3D n-type NSFET. The 3D structure of the NSFET TDDB simulation is shown in Fig. 4.9. A single nanosheet is used to reduce the computation complexity. To understand the effect of the thin-body nanosheet geometry on TDDB, the electric field and trap generation on the top surface (100), side surfaces (110), and corners will be monitored. Moreover, different corner rounding schemes (1nm rounding radius and 2.5nm rounding radius) are simulated to examine the effect of nanosheet geometry on the breakdown process.

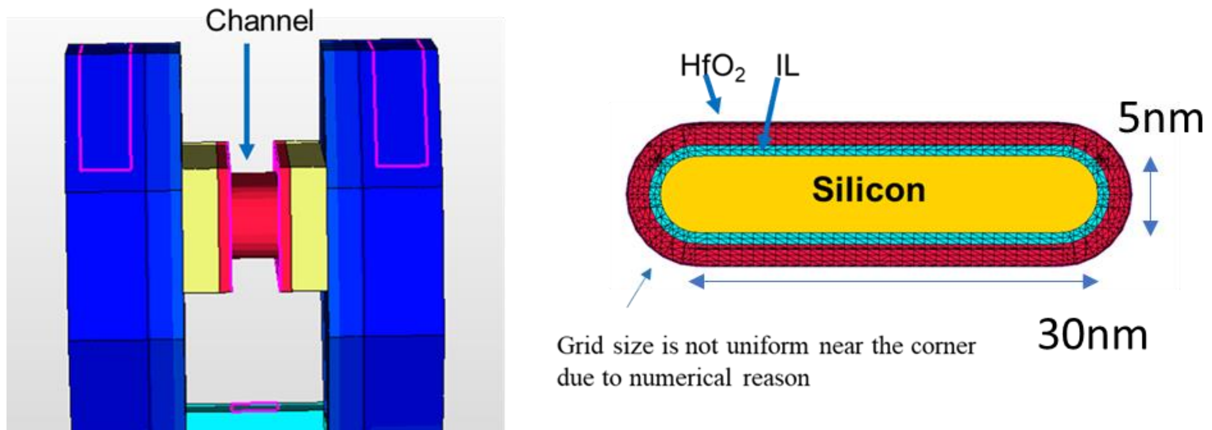


Fig. 4.9 3D simulated structure and cross-section of NSFET TDDB

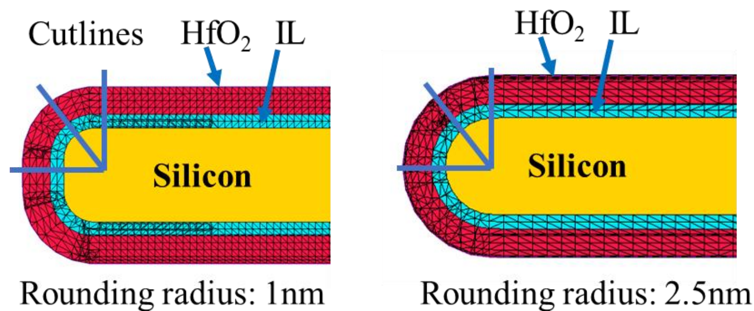


Fig. 4.10 Simulated nanosheets with different corner rounding radius

The electric field at time=0 is extracted and shown in Fig. 4.11. It can be observed that the electric field at Si/IL interface is higher near the corners than the flat surfaces. However, the electric field in the gate dielectric, especially in the HK layer, the electric field is lower near the corners. To verify this effect, the electric field profiles in the [100] direction, [110] direction, and 45° in the corner direction are extracted and shown in Fig. 4.12 -4.14. In the [100] direction, the electric field profiles in the gate stack of both corner rounding schemes are similar. The electric field in silicon increases when getting closer to the silicon/IL interface. In the IL and HK layer, the electric field remains almost as contents. The electric field in the gate stack of the 1nm rounding device is observed to be ~1% larger than that of the 2.5nm rounding device. This can be a result of the different quantum capacitance in Si ($C_{quanmtum}$) due to the different shapes of the silicon body.

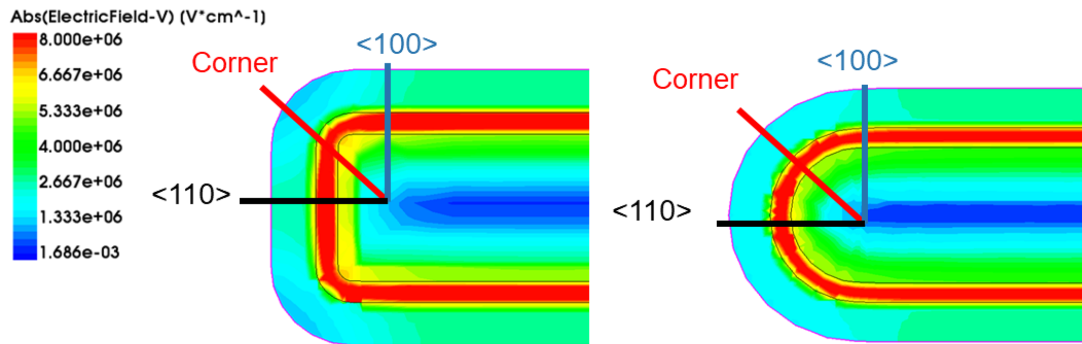


Fig. 4.11. The electric field near the corners of NSFET with different corner rounding radius

In the [110] direction, the difference between corner rounding schemes is more significant. The electric field in silicon is higher with 2.5nm rounding. This leads to a higher potential drop in silicon. As a result, the potential drop in the gate stack becomes smaller. On top of that, the electric field reduces significantly in IL and HK when getting closer to the gate. This is due to the electric field spreading out along the rounding curvature of the gate dielectric. For the NSFET with 1nm

rounding, there is less curvature on the <110> surface, meaning that the electric field drops less than the 2.5nm rounding NSFET.

In the 45-degree corner direction, the geometry effect is more significant. The electric field reduces significantly in IL and HK. Moreover, the strong quantum effect near the corners further increases the potential drop in silicon and reduces the potential drop in the gate dielectric. As a result, the electric field in the gate stack near the corner of the NSFET is smaller than the flat surfaces in <100> and <110> directions. Subsequentially, the trap generation rate near the corners will be lower, making the corners less susceptible to breakdown.

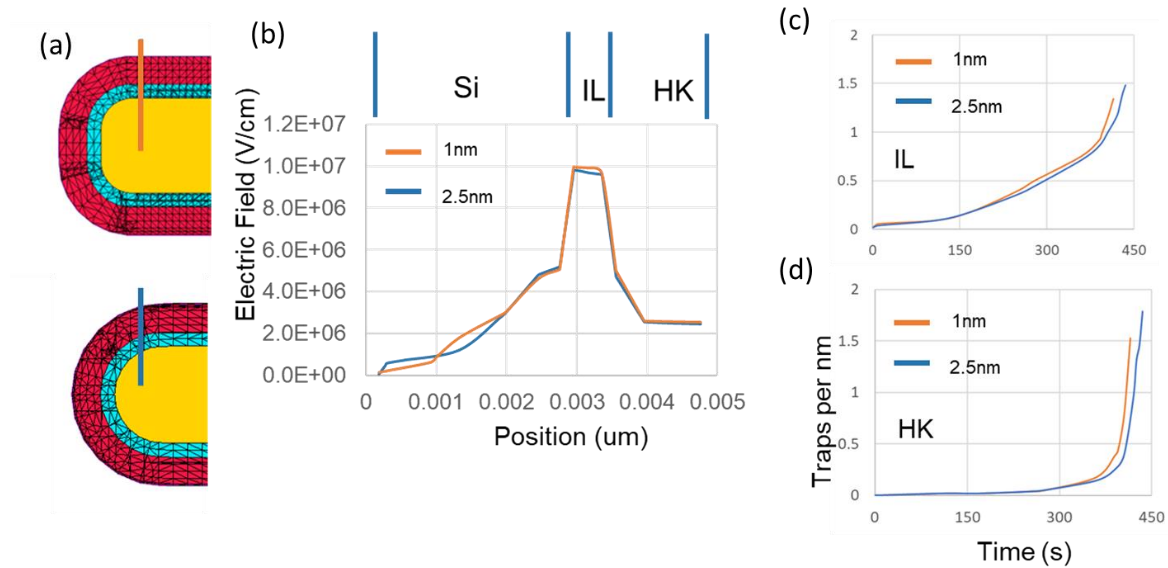


Fig. 4.12 (a) The cutline position in <100> direction (b) The electric field along the cutline (c) trap density in the interfacial layer (d) trap density in the high-k

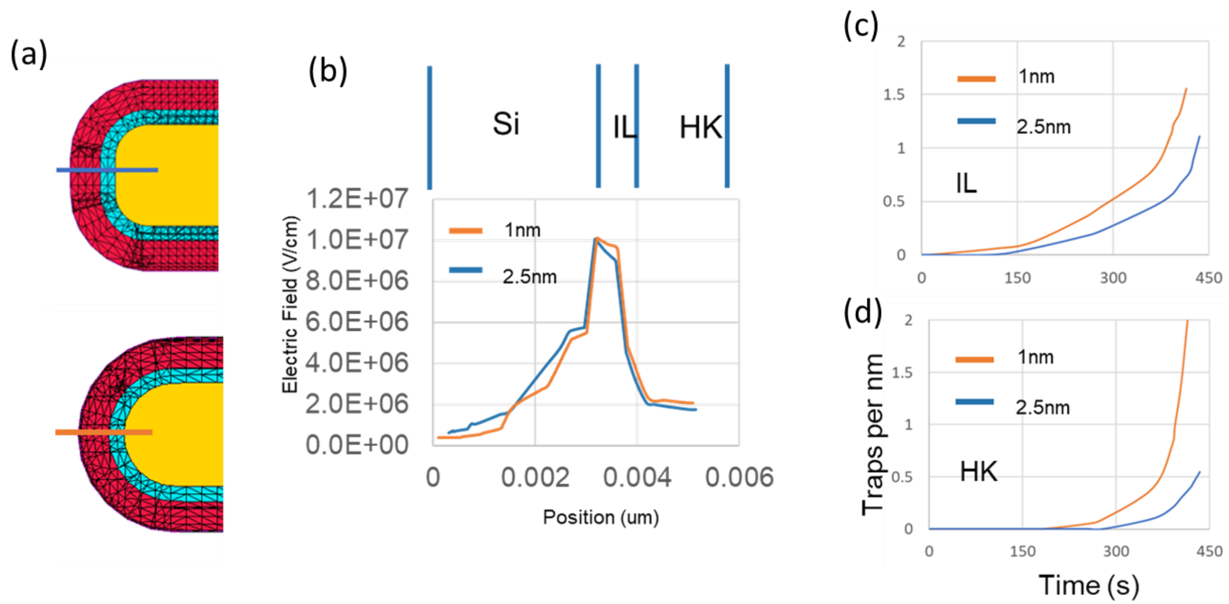


Fig. 4.13 (a) The cutline position in <110> direction (b) The electric field along the cutline (c) trap density in the interfacial layer (d) trap density in the high-k layer

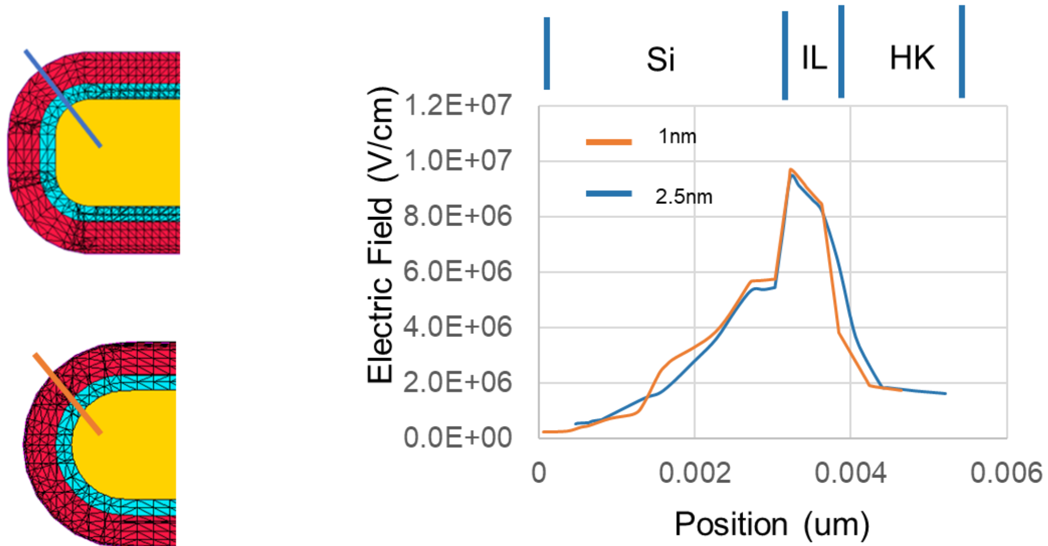


Fig. 4.14 The electric field near the corners. The electric field decreases quickly towards the gate

The normalized trap density vs stress time is plotted in Fig. 4.12 – 4.14. In the <100> direction, the trap density of the 1nm rounding device is 10% higher than the 2.5nm rounding device in the

trap build-up phase. This is due to the 1% higher electric field in the 1nm rounding device. On the <110> surface, the trap density is more than 2X higher in the 1nm rounding device than in the 2.5nm rounding device. As a result, the time of weak spot formation is shorter for a 1nm rounding device. This is due to the faster trap generation rate, especially in the [110] direction.

The simulated gate current vs time is shown in Fig. 4.16. Due to the larger trap density and shorter weak spot formation, the breakdown time of the 1nm rounding device is 35s shorter than the 2.5nm rounding device. This indicated that the NSFET with more corner rounding will have better TDDB resistance. The electric field and gate current density near breakdown are shown in Fig. 4.17 and Fig. 4.18. With the increased trap density in IL and HK layers, the electric field is increased in a few locations, forming weak spots. The gate current density at the weak spots is increased due to the increased trap-to-trap tunneling at the weak spots. For the NSFET with 1nm corner rounding, weak spots are formed on both (100) and (110) surfaces. For the NSFET with 2.5nm rounding, all the weak spots are located at the (100) surface. As discussed above, the trap generation rate is lowered at the rounded (110) surface due to geometry and quantum effect.

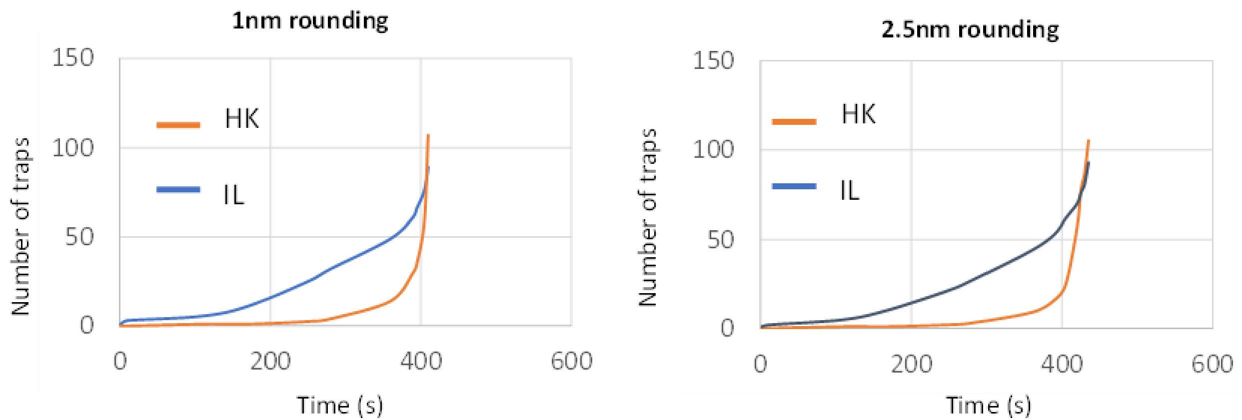


Fig. 4.15 Number of traps versus time in devices with different rounding. The traps gradually build up in IL and increase abruptly in HK.

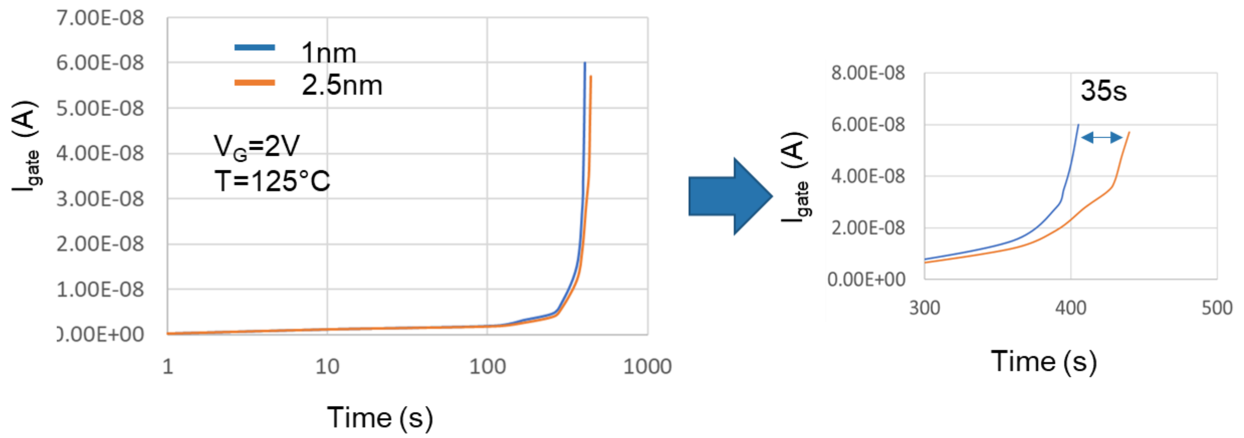


Fig. 4.16 Gate current versus stress time. 2.5nm corner rounding improves TDDB time by 10%

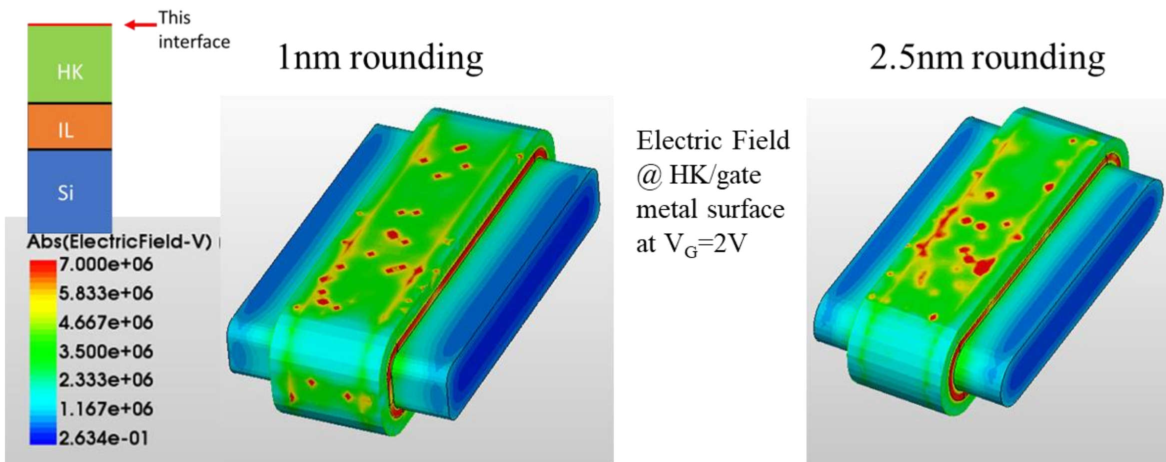


Fig. 4.17 3D electric field distribution at the top of HK layer

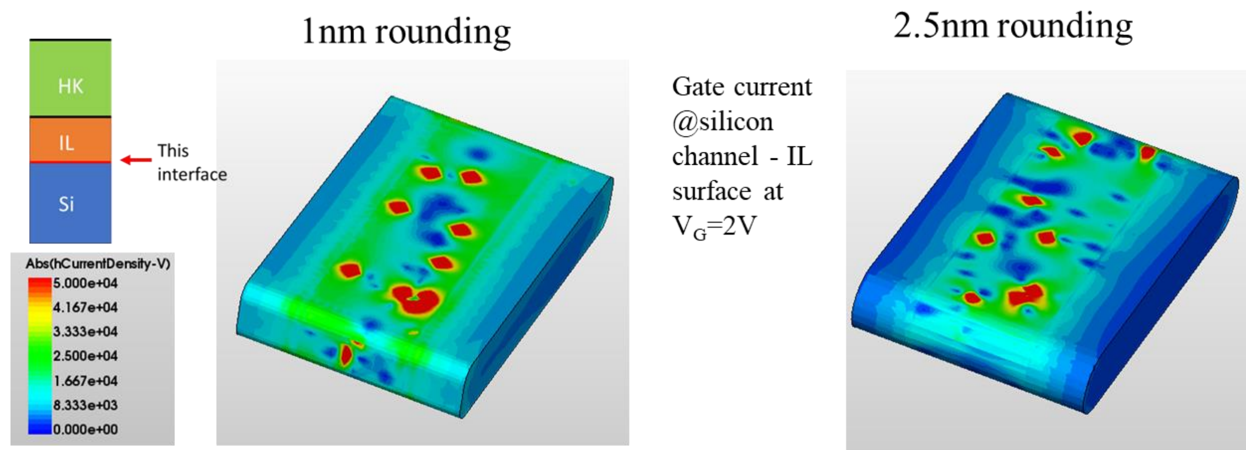


Fig. 4.18 3D current distribution at the silicon/IL interface

4.5 Conclusion

With the scaling of EOT, reliability issues like breakdown becomes increasingly important. The TDDB process in NSFET with its unique geometry has been studied. A discrete-trap-based TDDB simulation framework has been set up and calibrated. The simulation shows the trap generation process during the breakdown of the dual-layer gate dielectric in NSFET. The traps are first generated in the interfacial layer, forming weak spots. The electric field in the high-k layer is then increased near the weak spots, accelerating trap generation, and forming a breakdown path. The study reveals that the corners of the nanosheet are less susceptible to stress-induced dielectric breakdown due to quantum effect and geometry effect. By increasing the rounding radius, the NSFET can be more resilient to TDDB due to less trap generation at the rounded corners.

CHAPTER 5
FinHBT process development

The BiCMOS technology has been the go-to solution for RF/mixed-mode applications. Vertical SiGe HBT on 130nm node has demonstrated an f_T/f_{MAX} up to 505GHz/720GHz [8]. However, implementing the conventional vertical HBTs on the FinFET platform is extremely challenging. The process flow of vertical HBT is completely different from a FinFET or NSFET. Therefore, for ultra-scaled SoC, a CMOS-compatible HBT process can be an attractive solution.

In this study, the lateral SiGe FinHBT is proposed as a FinFET-compatible solution to ultra-scaled BiCMOS. A process flow on SOI substrate featuring lateral SiGe epitaxy growth is proposed to create the lateral SiGe profile to improve the carrier transfer in the base region. A test device is fabricated and measured. Based on the measurement results, the simulation predicts that $f_T/f_{MAX} > 750GHz$ can be reached with further scaled base width.

5.1 Device Architecture

The 3D schematic of the SiGe T-base FinHBT and its Ge composition/doping profile is shown in Fig. 5.1. SOI substrate is used to eliminate the leakage path and minimize the parasitic capacitance. The device architecture resembles a FinFET, but with a wider fin because the narrow fin is not needed to suppress short channel effects. A wider fin can provide a large current conduction area and reduce the impact of surface scattering on carrier transport. The laterally graded SiGe heterostructure can be achieved by template-assisted selective epitaxy/confined lateral selective epitaxial growth. With advanced lithography technology like EUV, a very narrow base width can

be achieved, which ensures good RF performance. Unlike FinFET, the high-k metal gate process is not needed. Instead, a T-shaped extrinsic base is located on the top of the intrinsic base. The area of the extrinsic base is much larger than the intrinsic base, providing a larger contact area. Therefore, the base contact resistance can be reduced, and a larger fmax can be achieved.

Fig. 5.1. 3D structure of FinHBT with T-shaped extrinsic base

5.2 FinHBT Process Flow

In this project, a full FinHBT fabrication process is developed. The key steps of the T-base FinHBT fabrication process are illustrated in Fig. 5.2. The width of the SiGe base region is defined by the trench created by Reactive Ion Etch (RIE). A reduced-pressure chemical vapor deposition (RPCVD) system is used for base epitaxial growth, as described in [29]. This RPCVD system provides the capability of SiGe epitaxial growth, which is essential for the formation of SiGe base in lateral SiGe FinHBT.

To define the very narrow trench for the SiGe base, e-beam lithography (EBL) with the capability of defining critical dimensions down to sub-10nm is used. RIE is used to transfer the pattern to the substrate. A PECVD SiO_2 layer is used as hard mask. With the combination of SiGe lateral epitaxial growth, EBL, and RIE, a SiGe lateral pocket can be defined as the base region.

A highly doped extrinsic base is grown on top of the intrinsic SiGe base. By using a longer epitaxy

growth time, the extrinsic base extends outside the trench, forming a T-shaped extrinsic base structure. With the T-base, the base contact area is no longer limited by the base width. At the same time, the in-situ doped extrinsic base can reach a doping concentration > $1x10^{20}$ cm^{-3}, further reducing the base resistance.

After the epitaxial growth, RIE is used to etch the fin structure. ALD is then used to passivate the exposed Si surface during the fin etch. Contact openings are created by RIE. The base is contacted right on top of the T-base region for the reduction of base resistance. NiSi is used to reduce the contact resistivity with a similar process as described in [30]. For metallization, the 150nm Al metal wires and landing pads are deposited using single-layer PMMA lift-off. Fig.5.3 summarized the flow chart for the complete SiGe FinHBT process flow. The comparison between FinHBT and the standard FinFET process is shown in Table 5.1. Extra steps include base patterning, base SEG, and extrinsic base growth.

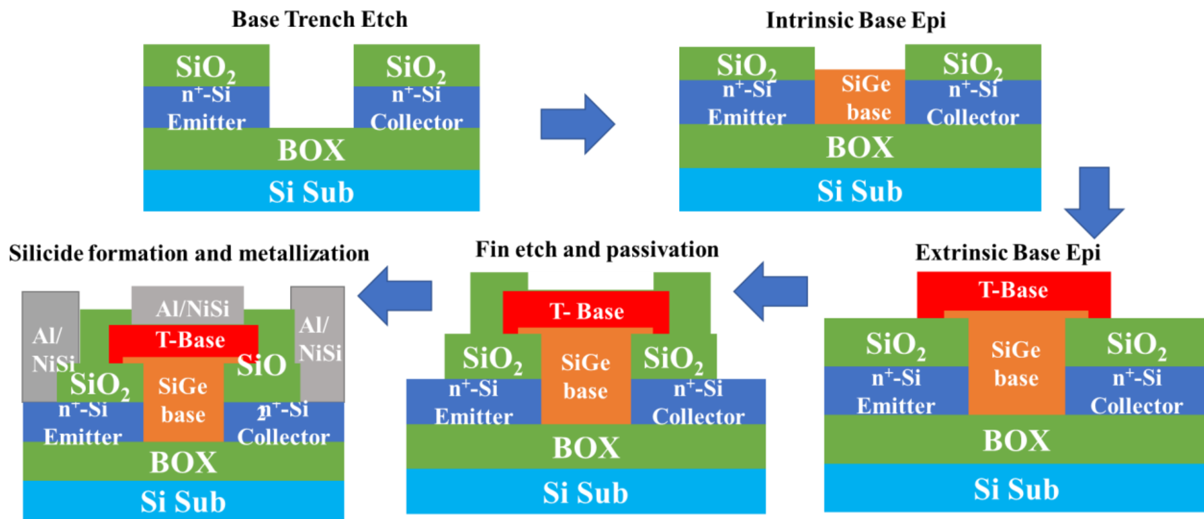


Fig. 5.2. Key process steps of T-base FinHBT

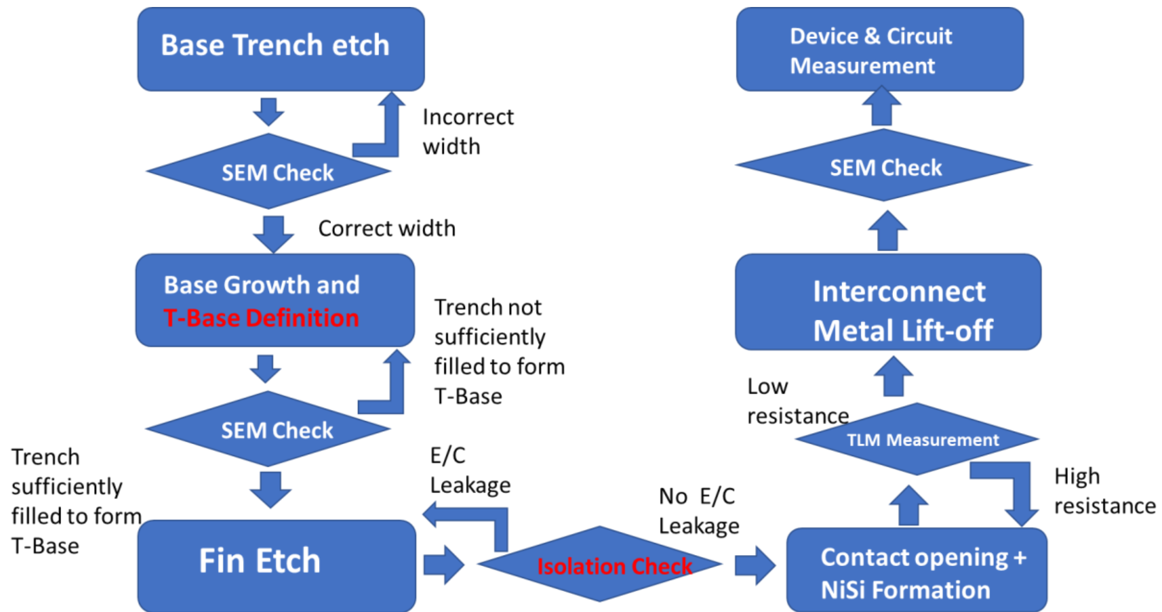


Fig. 5.3. Flow chart of T-base FinHBT process

Table 5.1 Comparison between FinFET and FinHBT process

SOI FinFET	FinHBT
SOI Substrate	SOI Substrate
	E/C doping → Base Epi → Extrinsic base Epi
Fin pattering	Fin patterning
Dummy gate	Dummy gate
Spacer	Spacer
SD regrowth	EC regrowth
Remove dummy	Remove dummy
HKMG	
Contact	Contact

5.3 Base epitaxial growth

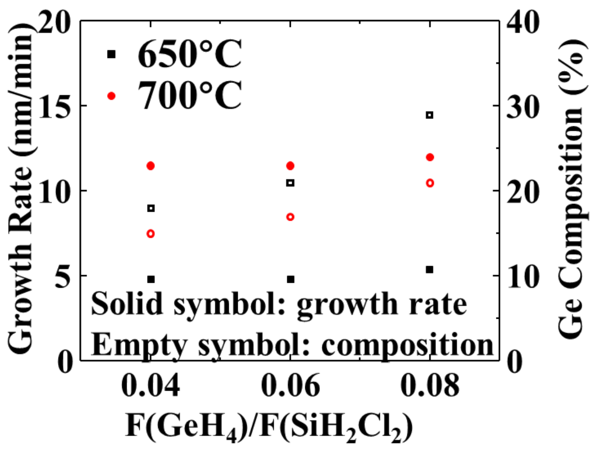
SiGe epitaxy growth is done using the Applied Materials Centurion Epitaxial System. It is an RPCVD system in which dichlorosilane (DCS) and Germane were used as a source gas in a reduced-pressure environment [29]. To find the optimal growth condition, SiGe blanket growth

was done on bare silicon wafers. The growth rate and Ge composition versus the mass flow ratio between SiDCS and Germane at 650°C and 700°C are shown in Fig. 5.4 (a). The AFM imaging (Fig. 5.4 (b)) indicates that a conformal, high-quality SiGe layer can be obtained after the growth.

To form the highly scaled SiGe base region, a SiGe selective epaxial growth (SEG) technique is used to selectively grow the SiGe pockets in the base trench. During SEG, HCl is added to the chamber to reduce the growth rate of amorphous SiGe on the SiO_2 hard mask. As shown in Fig 5.5, when the HCl flow rate increases, the selectivity of the growth improves (less growth on SiO_2). However, the SiGe growth rate in the opening regions also reduces with a high HCl flow rate. Therefore, there is an optimal HCl flow window in which both the selectivity and the growth rate are acceptable.

To reduce the number of defects in the SiGe base region, an annealing process is applied after the Epi growth. According to [31], SiGe: B/Si grown at 650°C can sustain 750°C H_2 annealing without noticeable Ge and dopant diffusion. Therefore, H_2 annealing at 750°C for 5 minutes is used to improve the crystalline quality of the SiGe base.

By introducing doping precursors in the growth process, in-situ doped SiGe can be formed. For growing the base of an NPN transistor, 1% B_2H_6 is used as the precursor. The doping concentration between $1x10^{19}$ and $1x10^{20}$ cm^{-3} can be reached by controlling the B_2H_6 mass flow. The relationship between doping concentration and borane/DCS ratio is shown in Fig. 5.6.



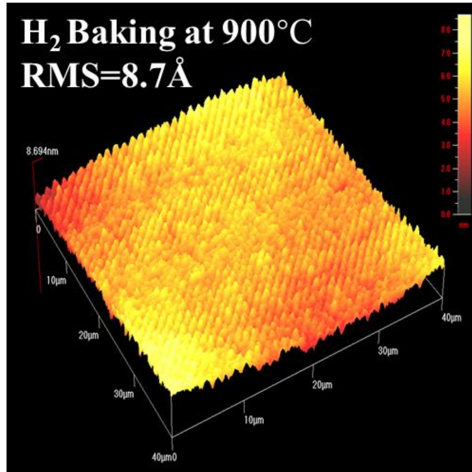


Fig. 5.4 (a) SiGe blanket growth rate and Ge composition vs Germane/DCS mass flow rate. (b) AFM imaging of SiGe grown at 650°C

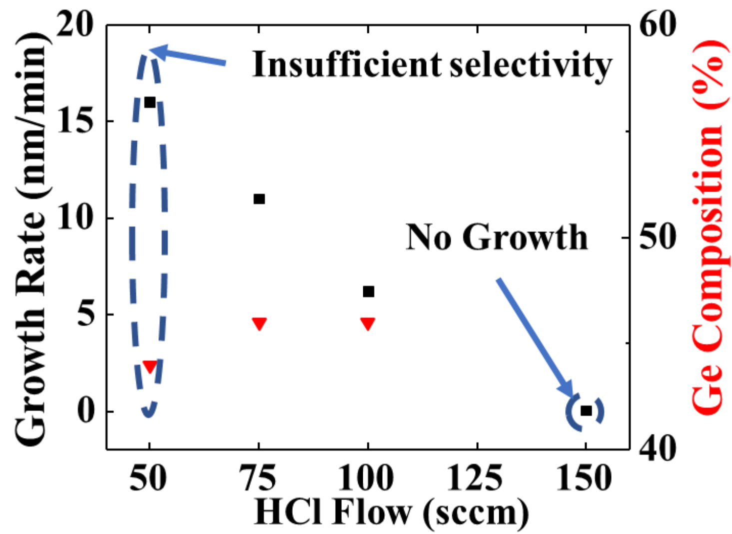


Fig. 5.5. SiGe SEG growth rate and Ge composition vs HCl flow. Temperature is 650°C. DCS flow rate is 75 sccm.

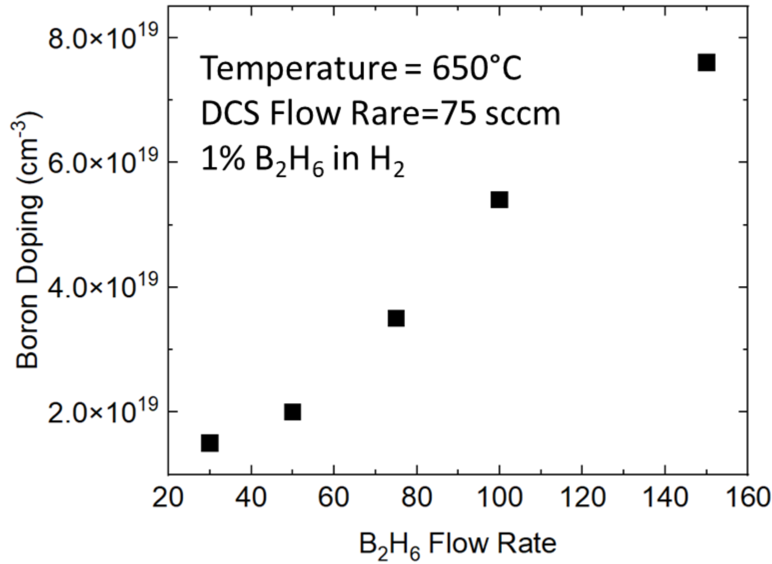


Fig. 5.6 Borane doping vs Borane flow rate

The detailed process flow of base growth with SiGe SEG is shown in Fig 5.7. Narrow trenches that are created by dry etch will be used to define the base region. SiGe SEG is then used to form the SiGe base with desired Ge composition profile inside the base. Since all the silicon in the trench is etched, there will be no seeding layer at the bottom of the trench. Therefore, SiGe will only grow from the sidewalls, enabling better control of desired Ge composition profile. After the trench is filled, an extrinsic SiGe base with a higher doping concentration is grown on the top of the intrinsic base region. The T-shaped extrinsic base will extend outside the trench, extending the contact area of the extrinsic base.

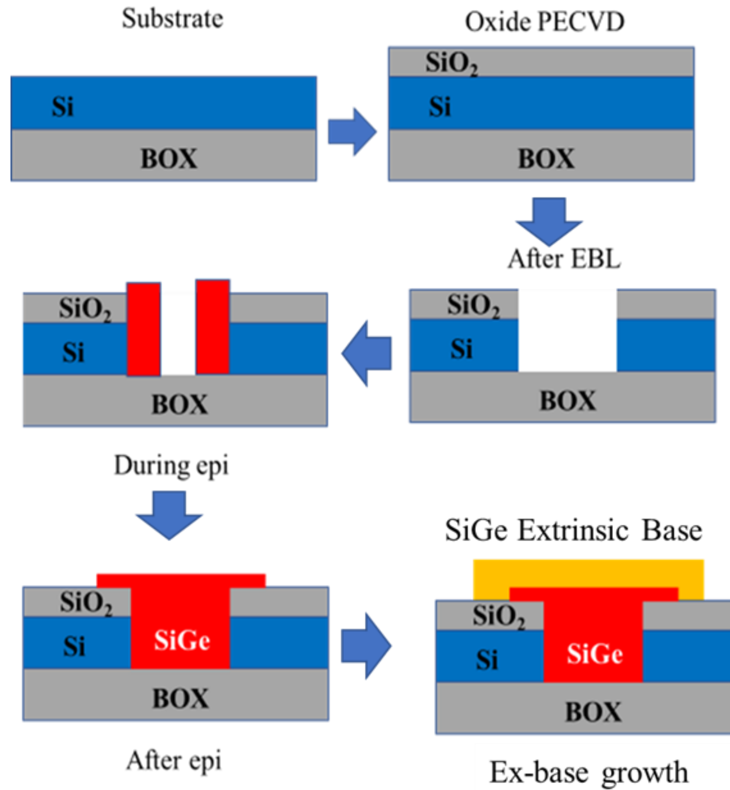


Fig. 5.7 The process flow of SiGe SEG on SOI substrate including the extrinsic base growth

The SEM picture of the base region after base SEG is shown in Fig. 5.8 – 5.9. The width of the intrinsic base and extrinsic bases are 100nm and 320nm, respectively. As shown in the cross-section SEM, the SiGe grown from the sidewalls of the base trench merges in the middle. The SiGe then grows outside the base trench, forming the extrinsic base. E-beam lithography is then used to pattern the extrinsic base to the desired width.

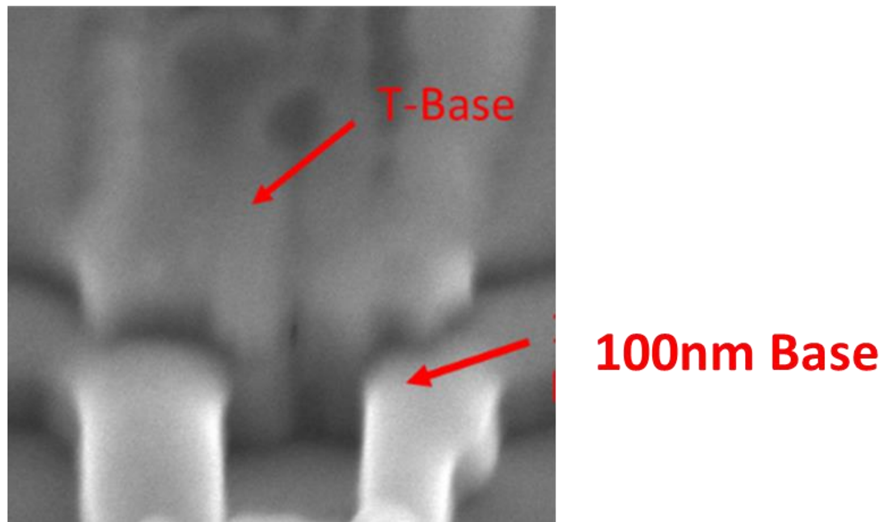


Fig. 5.8 Cross-section view of 100nm base after T-base growth

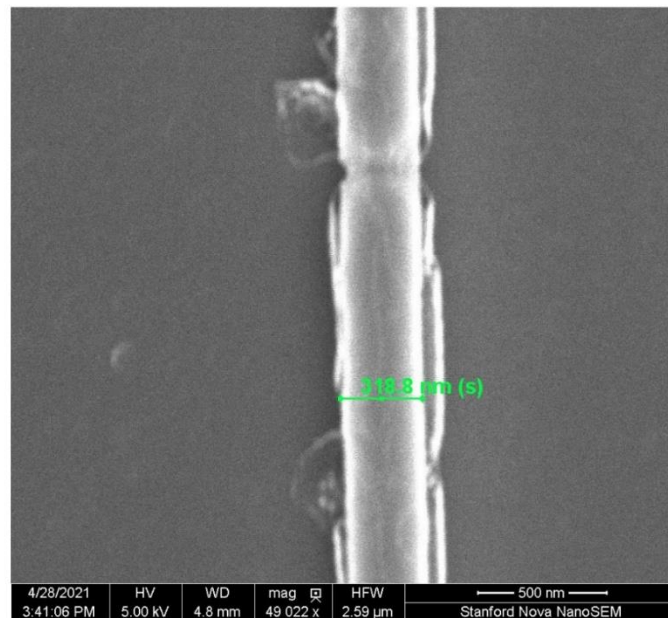


Fig. 5.9 SEM of a 300nm T-base

5.4 Fin Patterning and contact opening

Fin structure was created using EBL and Chlorine based RIE. The fin after the etch is shown in Fig. 5.10. An average fin width of 55nm is achieved, with a 77° taper angle.

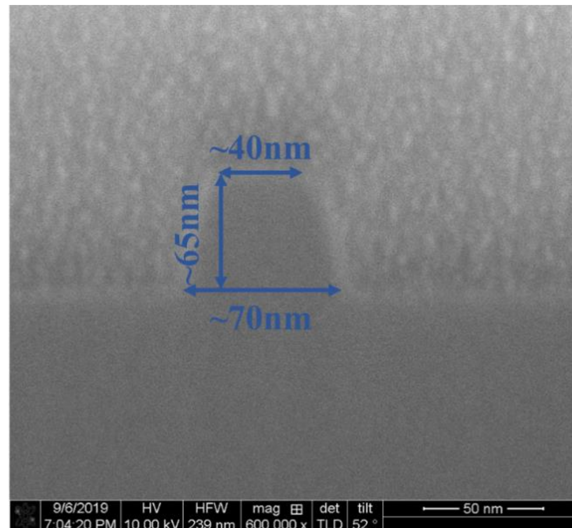


Fig. 5.10. cross-section SEM of the fin

To get desired device structure, the fin structure needs to be aligned to the base region during EBL. Two-step alignment (with global alignment and chip alignment) was used to make sure the alignment error was minimized. The alignment marks were cross-shaped, 1um deep trenches to cause good enough contrast in the back-scatter detector of the EBL system. The error of the alignment is less than 10nm, which is good enough to achieve desired device structure.

After the fin patterning, ALD is used to passivate the sidewalls exposed during the dry etch with SiO_2. The ALD also passivates the extrinsic T-base. After passivation, contact openings are created using EBL and RIE oxide etcher. The base is contacted right on top of the T-shaped extrinsic base region for the reduction of base resistance. An example of FinHBT after fin etch and contact opening is shown in Fig. 5.11.

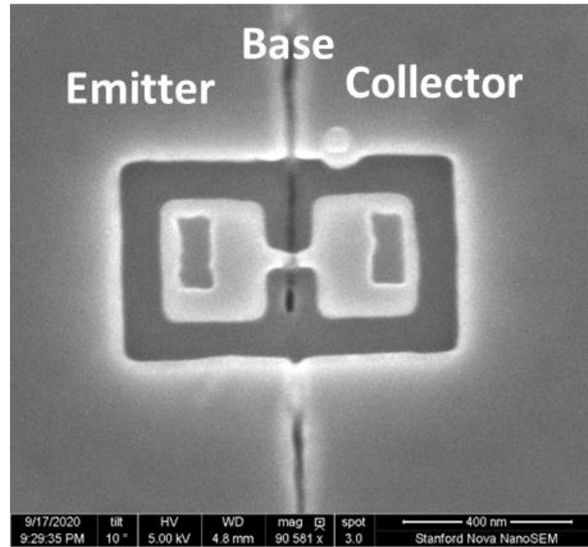


Fig. 5.11 FinHBT after contact opening

5.5 Silicidation and Metallization

Due to the small size of the FinHBT, it is crucial to have small contact resistivity for high f_T and f_{MAX}. Therefore, the self-aligned nickel silicide process is chosen to reduce the contact resistivity. The NiSi process on the SOI wafer reported in [30] is used. As described in [30], the contact resistance is minimum when 60% of the silicon body is converted to NiSi. In this study, the silicon body thickness is 50nm. Therefore, 16nm Ni is deposited on the SOI substrate to form 30nm NiSi after the annealing process. Cross-section SEM in Fig. 5.13 shows that the NiSi is 31nm after the silicidation process. Low resistance NiSi can only be formed with an annealing temperature between 450°C and 700°C. Different RTA temperatures of 450, 500, 550, 600, 650, and 700 °C also have been tested under the same Ni thickness(16nm) and base pressure (1x 10^{-6} torr). Fig. 5.14 shows the lowest contact resistivity $9.11 \times 10^{-7}\ \Omega \cdot cm^2$ is achieved at RTA= 500 °C

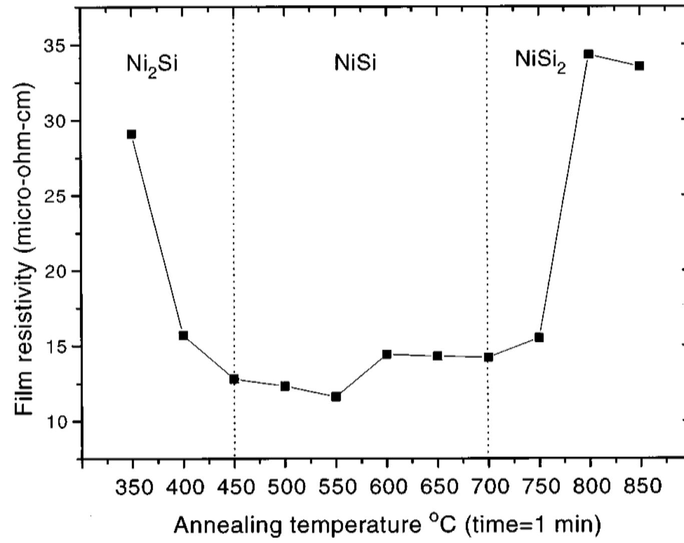


Fig. 5.12 Silicide film resistivity versus annealing temperature [30]

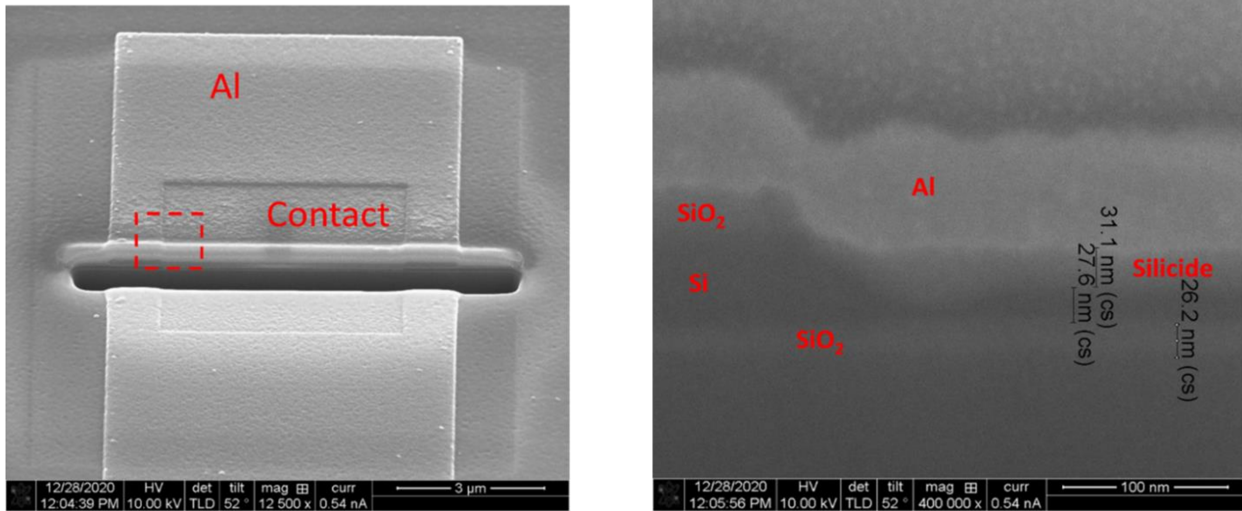


Fig. 5.13 Cross-section SEM of the contact after the silicidation process

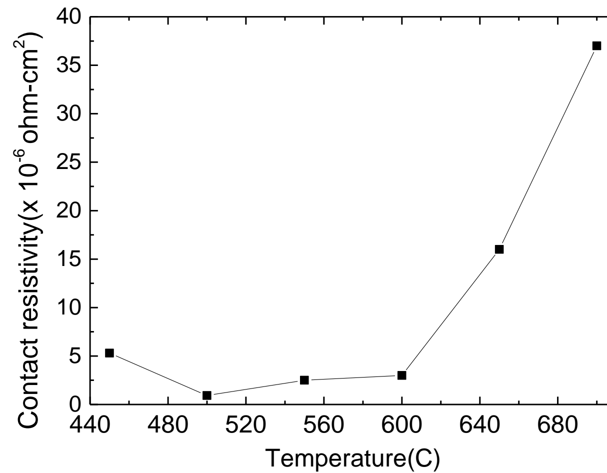


Fig. 5.14 Contact resistivity versus RTA temperature

The contact resistivity of Nickel silicide on P-type SiGe ($8x10^{19}$ cm^{-3}) has also been tested (Ni=16nm, RTA=500 C, and base pressure=1.2x 10^{-7} torr). The contact resistivity is $3.83x10^{-6}$ $\Omega \cdot cm^2$. The contact resistivity is higher on P-type SiGe than on N-type Si substrate. This is due to the rough interface and nickel germanosilicide agglomeration on the surface of P+ SiGe, as described in [32].

After the silicidation process, PMMA lift-off is used to form the metal wires and pads. The lift-off resist is a 200nm PMMA layer patterned by EBL. The 100nm Al layer used for metal pads and wires is deposited by e-beam evaporation. A completed FinHBT test device after metal deposition is shown in Fig. 5.15.

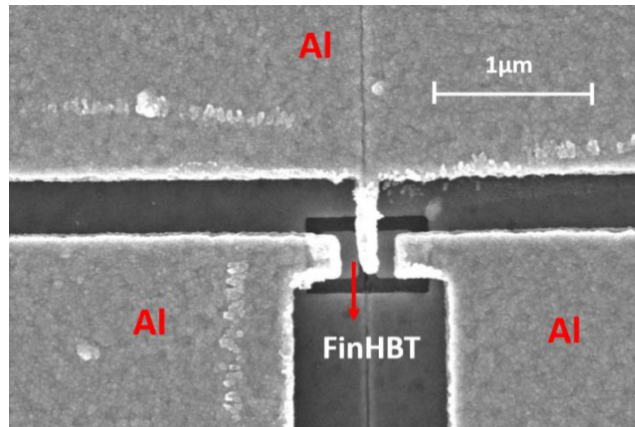


Fig. 5.15 The device after contact opening formation. The alignment error is negligible.

5.6 Conclusion

In this chapter, the lateral SiGe FinHBT is proposed as a FinFET-compatible solution to ultra-scaled BiCMOS. A process flow on SOI substrate featuring lateral SiGe epitaxy growth is developed to create the lateral SiGe profile to improve the carrier transfer in the base region. The width of the SiGe base is defined by etching a narrow trench on the SOI substrate. SiGe SEG is then performed to grow SiGe inside the narrow trench, forming a lateral SiGe pocket for the base. A T-shape extrinsic base structure is designed to reduce the base series resistance. EBL is used to define the device structure. The silicidation process and metal lift-off are used for metallization.

CHAPTER 6
FinHBT characterization and simulation prediction

6.1 FinHBT DC Characteristics

T-Base FinHBT with a base width of 100nm is successfully fabricated. The estimated parameters of the device are shown in Table. 6.1. The average doping concentration is measured by sheet-resistance calculation and the base Ge% is measured by an ellipsometer.

Table 6.1 Parameters of the fabricated FinHBT

Base Width	100nm
Fin Width	100nm
Fin Height	50nm
Emitter/Collector size	1µm/1µm
Extrinsic base width	300nm
Base Ge%	15%
Emitter/Collector Doping	$8x10^{19}cm^{-3}$
Base doping	$2x10^{19}cm^{-3}$

The measured device DC characteristics are shown in Fig. 6.1 – 6.3. The device's Gummel plot is given in Fig. 6.1. While the turn-on slope of I_C is close to the ideal 60mV/dec, the ideality factor of I_B is larger than 1, especially with a small V_{BE}. This is likely due to the existence of defects in

the base region. During the SiGe growth in the base trench, grain boundaries will form in the middle of the base when the SiGe grown from the base trench sidewalls merge in the middle. Although annealing is done after the growth to suppress the defect density, it still causes notable Shockley-Read-Hall Recombination (SRH) recombination. This leads to the <1 gain at small V_{BE}.

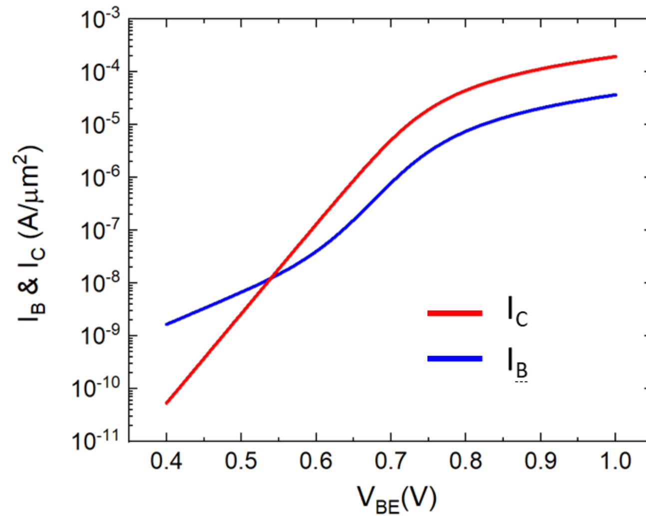


Fig. 6.1 Measured Gummel plot of the FinHBT

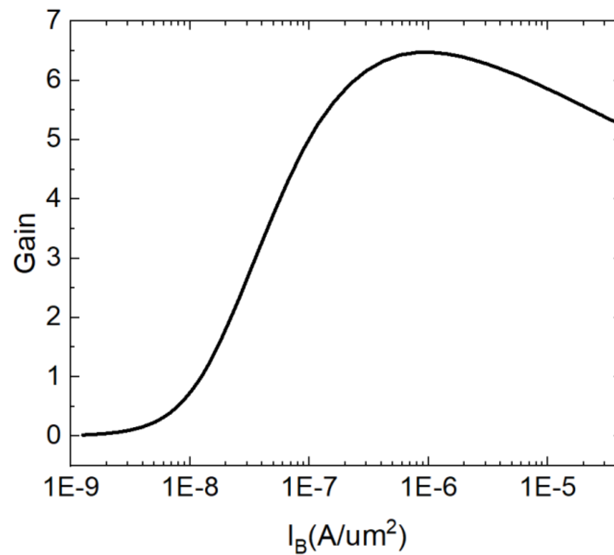


Fig. 6.2 Measured Gain versus I_B of FinHBT

The gain vs I_B is shown in Fig. 6.2. The peak DC gain is about 6.5 at I_B=2.5µA/µm^2. Measured common-emitter characteristics. Current saturation can be observed. The measured open-base breakdown voltage is 1.5V.

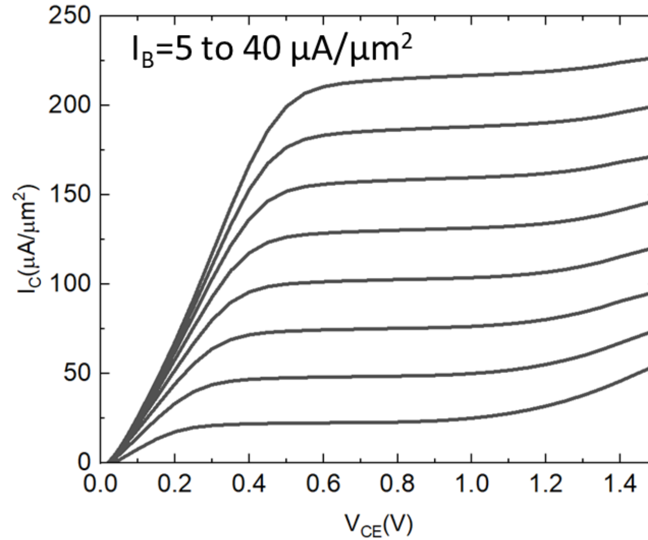


Fig. 6.3 Measured common-emitter characteristics of FinHBT

6.2 FinHBT RF performance simulation

Based on the measured data of 100nm base FinHBT, the RF performance of further scaled FinHBT can be predicted by TCAD. Like in the FinFET and FinHBT study, the Sentaurus TCAD platform is used for the FinHBT study. The calibrated hydrodynamic model [33] is adopted to capture the velocity overshoot in the deeply scaled device. In addition, high field mobility saturation, doping-dependent mobility, Auger recombination, Shockley-Read-Hall recombination, and bandgap narrowing have been enabled.

The FinHBT with parameters in Table 6.1 is simulated. The device structure is shown in Fig. 6.5. The emitter/collector and base contact resistance are used to fit the measured results. The result of

the fitting is shown in Fig. 6.4. The fitted contact resistance is 1.1 kΩ for the emitter and collector and 4.5 kΩ for the base. Although the base contact region has been enlarged by the T-base structure, the base resistance is still substantially higher than the emitter and collector. This is likely due to the relatively smaller contact area and larger silicide resistance on p-type SiGe.

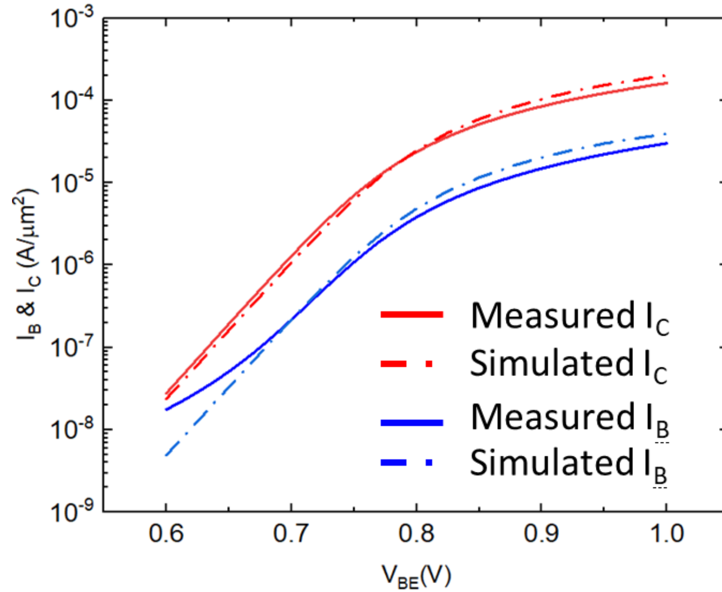


Fig. 6.4 Simulated and measured Gummel plot of FinHBT

Based on this result, the T-base FinHBT structure with scaled base width has been designed and simulated. The 3D structure and cross-section of the simulated T-base device with 20nm W_{base} are shown in Fig. 6.5. The doping and Ge profile is shown in Fig. 6.6. The parabolic-shaped Ge profile is fitted to the SIMS profile of a vertical SiGe HBT in [34]. The doping profile in the emitter/collector resembles the doping profile in the source/drain and spacer region of a FinFET ([35]). Enlarged emitter/collector regions are used to mitigate the effect of emitter/collector contact resistivity. The T-base structure is adopted to reduce base contact resistivity. The doping concentration in the T-base is $1\text{x}10^{20}$ cm^{-3}. The intrinsic SiGe base doping concentration is $8\text{x}10^{19}$ cm^{-3}.

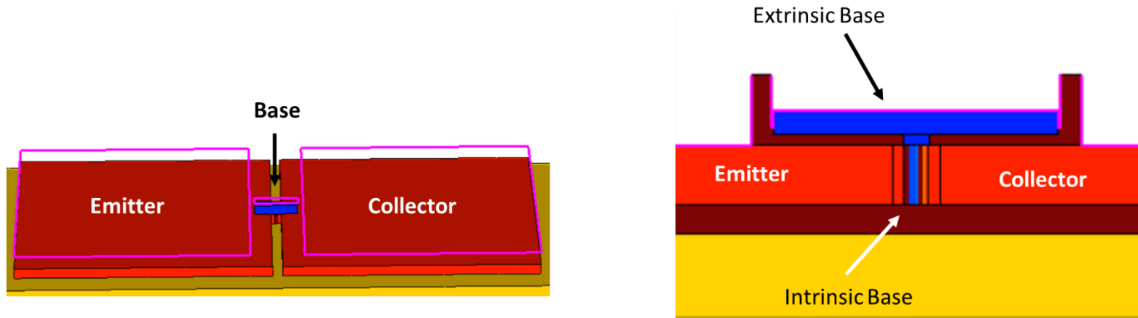


Fig. 6.5. Simulated T-base FinHBT

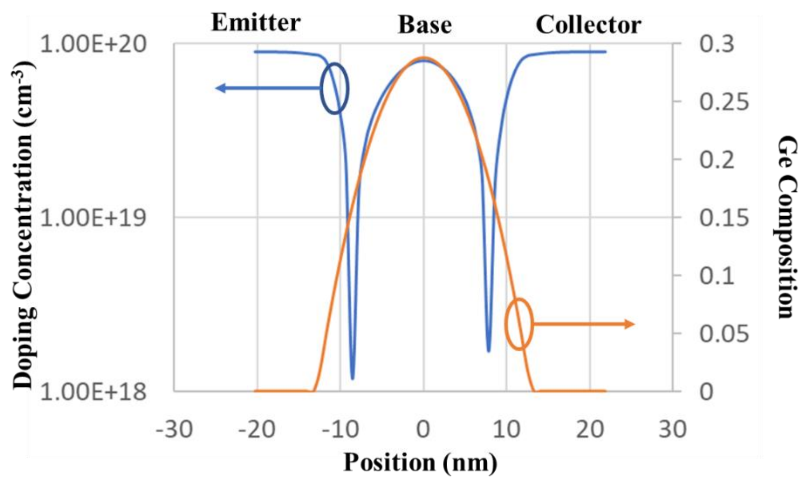


Fig. 6.6 Doping and Ge profile in the simulated FinHBT with a base width of 20nm

The detailed parameters and simulated device characteristics are shown in Table 6.2. Two types of scale FinHBT are proposed. Device 2 has a base width of 20nm, which is similar to the channel width of a 10nm FinFET. In device 1, the base width is further scaled to 10nm, which can be defined by the EUV technology in sub-5nm technologies. The contact resistance is 1×10^{-8} Ω-cm^2 for device 1 and further improved to 1×10^{-9} Ω-cm^2 in device 2. With the narrow base width, the base doping concentration needs to be increased to prevent base punch-through. The Ge composition of the base region is also increased to 30% to boost the current gain and f_T/f_{MAX} by enhancing the carrier transport in the base with the built-in electric field.

The simulated Gummel plot and f_T & f_{MAX} vs I_C for device 1 and device 2 are shown in Fig. 6.7 and Fig. 6.8. Peak f_T /f_{MAX} and I_C increase as the base width scales down. As W_B reduces from 20nm to 10nm, f_T/f_{MAX} increases from 600/620GHz to 814/751GHz. The delay components are extracted in Fig. 6.9. The forward transit time (τf) of ~156fs is extracted from Fig. 6.9(b) based on:

$$\frac{1}{2\pi f_T} = \left[\tau_f + \frac{kT}{qI_C}(C_{be} + C_{bc}) + C_{bc}(R_e + R_c)\right] \qquad (6.1)$$

where kT/q, I_C, C_{be}, C_{bc}, R_e, and R_c are thermal voltage, collector current, base-emitter capacitance, base-collector capacitance, emitter resistance, and collector resistance, respectively. The extracted $\tau_F + \tau_C$ and capacitance charging delay is shown in Table 6.3. The results show that τ_F and τ_C are the main contributors to the reduce of delay time as base width scales down. This is a result of increased carrier velocity and decreased base width. Therefore, τ_F decreases as the base width scale down, leading to the increased f_T.

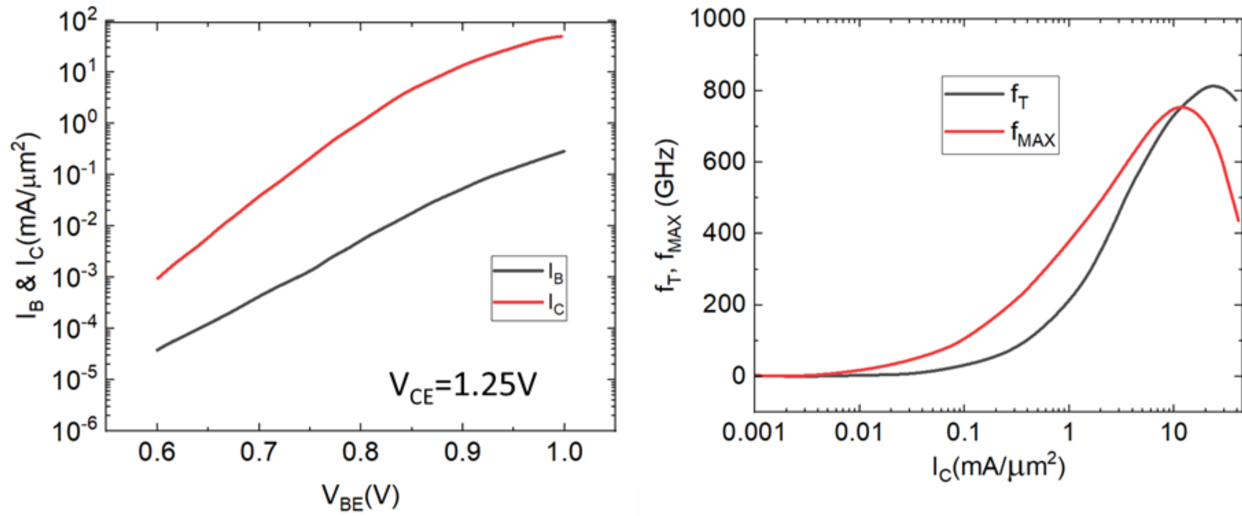


Fig. 6.7. Simulate Gummel plot and f_T/f_{MAX} of test device 1

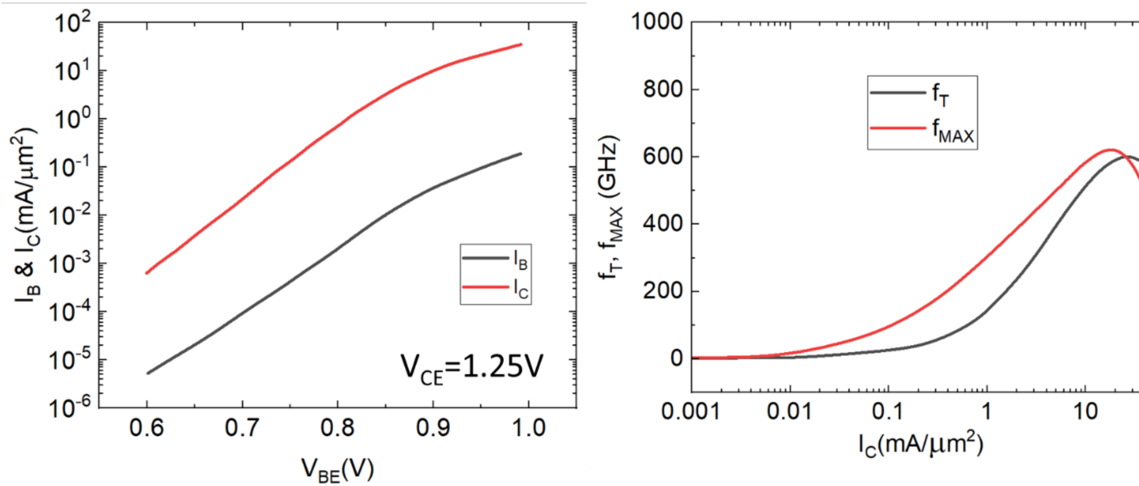


Fig. 6.8. Simulate Gummel plot and f_T/f_{MAX} of test device 2

Table 6.2. Simulated characteristics of T-base FinHBT

	Device 1	**Device 2**	**DC Test Device**
Base width	10nm	20nm	100nm
Contact resistivity	$1x10^{-9}$ Ω-cm^2	$1x10^{-8}$ Ω -cm^2	$1x10^{-6}$ Ω -cm^2
Ex-base width	100nm	100nm	300nm
Base Doping	$8x10^{19}cm^{-3}$	$8x10^{19}cm^{-3}$	$2x10^{19}cm^{-3}$
Base Ge%	30%	30%	15%
f_T	814GHz	600 GHz	
f_{MAX}	751GHz	620 GHz	
DC gain	220	169	6.5

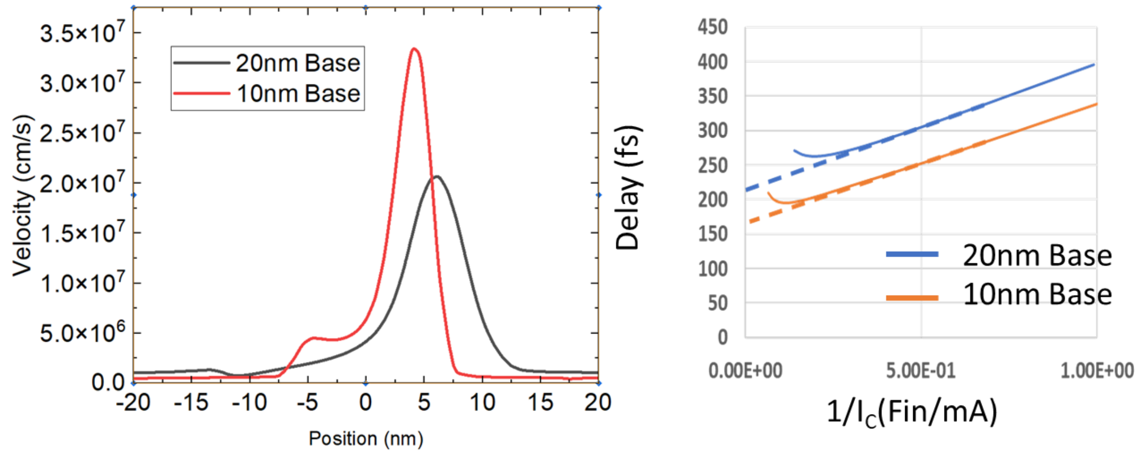


Fig. 6.9. Carrier velocity profile and delay components extraction

Table 6.3. Delay components of device 1 and device 2

	τF + τC	**Charging Delay**
Device 1	156 fs	42 fs
Device 2	203 fs	60 fs

6.3 Conclusion

A test device has been successfully fabricated and characterized. With a 100nm base width and 15% peak Ge composition, the maximum measured DC gain is 6.5. Based on this result, the T-base FinHBT structure with scaled base width has been designed and simulated. The simulation predicts that as W_B reduces to 20nm, the f_T/f_{MAX} can reach 600/620GHz. Further decreasing W_B to 10nm can further push the f_T/f_{MAX} to 814/750 GHz. This shows that with further lateral scaling, the FinHBT can potentially approach f_T/f_{MAX} close to THz.

CHAPTER 7
Conclusion

7.1 Summary

Combining digital computation, analog, and radio frequency (RF) circuitry, a highly functional systems-on-chip (SoC) design is a favorable choice for various applications like mobile systems, embedded systems, and space applications. In this work, potential devices to realize ultra-scaled SoC has been studied.

For digital computation, an extensive TCAD simulation of sub-5nm NSFET has been done. Since NSFET has been the most promising candidate for next-generation VLSI technology, it is crucial to have a thorough understanding of its performance and constraints. The study focuses on critical device parameters including L_G, T_{NS}, parasitic resistance, and EOT. Based on the Sentaurus TCAD platform, the ultra-scaled devices are simulated considering quasi-ballistic transport and quantum confinement. Simulation shows that NSFET has a better I_{on}/I_{off} and SS over FinFET with the same channel length and body thickness due to the superior electrostatic control of the gate-all-around structure. Parasitic resistance analysis shows that R_{spacer} is the main contributor to parasitic resistance. Design parameter analysis shows that the Ion/Ioff of NSFET can be improved by enhancing electrostatic control, like increasing L_G, reducing TNS, and scaling down EOT. Scaling EOT is considered a major pathway to further scaling, as it increases the carrier density in the channel, enhances electrostatic control, and suppresses the effect of parasitic capacitance.

With the scaling of EOT, reliability issues like breakdowns become increasingly important. The TDDB process in NSFET with its unique geometry has also been studied. A discrete-trap-based

TDDB simulation framework has been set up and calibrated. The simulation shows the trap generation process during the breakdown of the dual-layer gate dielectric in NSFET. The traps are first generated in the interfacial layer, forming weak spots. The electric field in the high-k layer is then increased near the weak spots, accelerating trap generation and forming a breakdown path. The study reveals that the corners of the nanosheet are less susceptible to stress-induced dielectric breakdown due to quantum effect and geometry effect. By increasing the rounding radius, the NSFET can be more resilient to TDDB due to less trap generation at the rounded corners.

For RF/Mixed-mode applications, a novel lateral SiGe FinHBT has been proposed. Utilizing the advanced lithography technology of ultra-scaled VLSI, the lateral scaling of base width can greatly enhance the RF performance of FinHBT. A FinFET compatible fabrication process has been designed and developed. The lateral base region is created by SiGe SEG. A T-shape extrinsic base structure is used to reduce the base resistance. A test device has been successfully fabricated and characterized. With a 100nm base width and 15% peak Ge composition, the maximum measured DC gain is 6.5. The simulation predicts that the $f_T/f_{MAX} > 750$GHz can be achieved with base width scaling.

Through TCAD simulations and experiments, promising devices for ultra-scaled SoC technology have been designed and studied. The FinFET and NSFET study can provide a guideline for the optimization of NSFET. The TDDB study provides an in-depth analysis of the breakdown process in NSFET and serves as guidance for gate stack reliability enhancement. The proposed lateral SiGe FinHBT introduces a FinFET-compatible RF/Mixed mode technology for further scaled BiCMOS. With further lateral scaling, the FinHBT can potentially approach 1 THz f_T/f_{MAX}, which enables high-speed BiCMOS VLSI SoCs for THz/mixed-mode applications.

7.2 Suggestions for future research

- NSFET TDDB study with further scaled EOT: In this work, a discrete trap model is used to study the breakdown process in the dual-layered gate stack in NSFET. In future work, the TDDB of NSFET with further scaled can be investigated. In ultra-scaled NSFET, the EOT can be increased by increasing the dielectric constant of the gate stack or reducing the thickness of the gate stack. Novel high-k material like $HfZrO_2$ can greatly increase the dielectric constant of the high-k layer ([37]), resulting in a smaller EOT with same thickness. Apart from that, the dielectric layer of the interfacial layer can also be increased by nitrogen treatment as discussed in [38]. While increasing dielectric constant can decrease EOT without reducing the gate stack thickness, it also results in higher dipole moment that can reduce the breakdown electric field. Therefore, an extensive study on the TDDB with higher dielectric constant material will provide guidance to the reliability of NSFET with further scaled EOT
- NSFET BTI study: BTI is another important aspect of device reliability. While the TDDB focuses on the trap generation and breakdown path formation in the gate dielectric layers, BTI is more strongly related to the increase of interfacial traps/trapped charges between gate dielectric and silicon. BTI has been studied in detail for planar MOSFET and FinFET [39], [40]. However, like in the case of TDDB, the unique geometry and quantum effect can lead to some intriguing BTI behaviors. Therefore, it is important to obtain a better understanding of BTI in NSFET and gain a more thorough understanding of the reliability constrains of the NSFET.
- Scaled FinHBT fabrication: In this study, a FinHBT with base width of 100nm has been fabricated. While simulation shows the RF performance of the FinHBT can be substantially

increased by scaling down base width, it also introduces some difficulties. One issue is the defects in the base created by the grain boundaries during the base epitaxial growth. Moreover, as the base width scales down, it can be challenging to fill the base region that has a higher aspect ratio.

- Doping/Ge profile engineering in FinHBT: As stated in this report, the lateral growth of SiGe during the base epitaxial growth provides a chance to control the lateral doping and Ge profile. With detailed engineered base doping and Ge profile, the performance of the FinHBT can be further improved. The profiles can be changed by changing the Germane/B_2H_6 flow rate with time during the SEG process. Another way is to control the thermal Ge/dopant diffusion profile by tailoring the temperature and time of the post-growth annealing process.

REFERENCE

[1] H. Mertens et al., "Vertically stacked gate-all-around Si nanowire CMOS transistors with dual work function metal gates," 2016 IEEE International Electron Devices Meeting (IEDM), 2016, pp. 19.7.1-19.7.4, doi: 10.1109/IEDM.2016.7838456.

[2] A. Veloso *et al*., "Challenges and opportunities of vertical FET devices using 3D circuit design layouts," *2016 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*, 2016, pp. 1-3, doi: 10.1109/S3S.2016.7804409

[3] N. Loubet *et al*., "Stacked nanosheet gate-all-around transistor to enable scaling beyond FinFET," *2017 Symposium on VLSI Technology*, 2017, pp. T230-T231, doi: 10.23919/VLSIT.2017.7998183.

[4] H. J. Lee, S. Rami, S. Ravikumar, V. Neeli, K. Phoa, B. Sell, and Y. Zhang, "Intel 22nm FinFET (22FFL) Process Technology for RF and mm-Wave Applications and Circuit Design Optimization for FinFET Technology," *2018 IEEE International Electron Devices Meeting*, San Francisco, CA, 2018, pp. 14.1.1-14.1.4. DOI: 10.1109/IEDM.2018.8614490

[5] D. Kim, B. Brar, and J. A. del Alamo, "f_T = 688 GHz and f_{MAX} = 800 GHz in L_g = 40 nm $In_{0.7}Ga_{0.3}As$ MHEMTs with g_{m_max} > 2.7 mS/µm," *2011 IEEE International Electron Devices Meeting*, Washington, DC, 2011, pp. 13.6.1-13.6.4. DOI: 10.1109/IEDM.2011.6131548

[6] M. Urteaga, R. Pierson, P. Rowell, V. Jain, E. Lobisser, and M. J. W. Rodwell, "130nm InP DHBTs with f_T >0.52THz and f_{MAX}>1.1THz," *69th Device Research Conference*, Santa Barbara, CA, 2011, pp. 281-282. DOI: 10.1109/DRC.2011.5994532

[7] W. Snodgrass, W. Hafez, N. Harff, and M. Feng, "Pseudomorphic InP/InGaAs Heterojunction Bipolar Transistors (PHBTs) Experimentally Demonstrating f_T = 765 GHz at 25°C Increasing to f_T = 845 GHz at -55°C," *2006 IEEE International Electron Devices Meeting*, San Francisco, CA, 2006, pp. 1-4. DOI: 10.1109/IEDM.2006.346853

[8] B. Heinemann, H. Rücker, R. Barth, F. Bärwolf, J. Drews, G. G. Fischer, A. Fox, O. Fursenko, T. Grabolla, F. Herzel, J. Katzer, J. Korn, A. Krüger, P. Kulse, T. Lenke, M. Lisker, S. Marschmeyer, A. Scheit, D. Schmidt, J. Schmidt, M. A. Schubert, A. Trusch, C. Wipf, and D. Wolansky, "SiGe HBT with f_T/f_{MAX} of 505 GHz/720 GHz," *2016 IEEE International Electron Devices Meeting*, San Francisco, CA, 2016, pp. 3.1.1-3.1.4. DOI: 10.1109/IEDM.2016.7838335

[9] Sentaurus Device User's Manual, Synopsys Inc., Mountain View, CA, USA, Jun. 2021.

[10] S. Martinie, G. Le Carval, D. Munteanu, S. Soliveres and J. Autran, "Impact of Ballistic and Quasi-Ballistic Transport on Performances of Double-Gate MOSFET-Based Circuits," in *IEEE Transactions on Electron Devices*, vol. 55, no. 9, pp. 2443-2453, Sept. 2008, doi: 10.1109/TED.2008.927656.

[11] R.Granzner, V. M. Polyakov, F. Schwierz, M. Kittler, R. J. Luyken, W. Rösner, M. Städele, "Simulation of nanoscale MOSFETs using modified drift-diffusion and hydrodynamic models and comparison with Monte Carlo results", in Microelectronic Engineering, Feb. 2006, vol. 83, issue 2, pp. 241-246, DOI: 10.1016/j.mee.2005.08.003.

[12] G. Tsutsui, M. Saitoh, T. Saraya, T. Nagumo and T. Hiramoto, "Mobility enhancement due to volume inversion in [110]-oriented ultra-thin body double-gate nMOSFETs with body thickness less than 5 nm," *IEEE InternationalElectron Devices Meeting, 2005. IEDM Technical Digest.*, 2005, pp. 729-732, doi: 10.1109/IEDM.2005.1609456.

[13] A. Khakifirooz *et al.*, "Scalability of Extremely Thin SOI (ETSOI) MOSFETs to Sub-20-nm Gate Length," in *IEEE Electron Device Letters*, vol. 33, no. 2, pp. 149-151, Feb. 2012, doi: 10.1109/LED.2011.2174411.

[14] C. W. Yeung *et al.*, "Channel Geometry Impact and Narrow Sheet Effect of Stacked Nanosheet," *2018 IEEE International Electron Devices Meeting (IEDM)*, 2018, pp. 28.6.1-28.6.4, doi: 10.1109/IEDM.2018.8614608.

[15] F. Jiménez-Molinos, F. Gámiz, A. Palma, P. Cartujo, and J. A. López-Villanueva, “Direct and trap-assisted elastic tunneling through ultrathin gate oxides”, Journal of Applied Physics 91, 5116 (2002);

[16] A. Palma, A. Godoy, J. A. Jiménez-Tejada, J. E. Carceller, and J. A. López-Villanueva, Quantum two-dimensional calculation of time constants of random telegraph signals in metal-oxide–semiconductor structures, Phys. Rev. B 56, 9565, October 1997

[17] C. Auth, A. Aliyarukunju, M. Asoro, D. Bergstrom, V. Bhagwat, J. Birdsall, N. Bisnik, M. Buehler, V. Chikarmane, G. Ding, Q. Fu, H. Gomez, W. Han, D. Hanken, M. Haran, M. Hattendorf, R. Heussner, H. Hiramatsu, B. Ho, S. Jaloviar, I. Jin, S. Joshi, S. Kirby, S. Kosaraju, H. Kothari, G. Leatherman, K. Lee, J. Leib, A. Madhavan, K. Marla, H. Meyer, T. Mule, C. Parker, S. Parthasarathy, C. Pelto, L. Pipes, I. Post, M. Prince, A. Rahman, S. Rajamani, A. Saha, J. D. Santos, M. Sharma, V. Sharma, J. Shin, P. Sinha, P. Smith, M. Sprinkle, A. St. Amour, C. Staus, R. Suri, D. Towner, A. Tripathi, A. Tura, C. Ward, A. Yeoh, " A 10 nm high performance and low-power CMOS technology featuring 3rd generation FinFET transistors self-aligned quad patterning contact over active gate and cobalt local interconnects ", *2017 IEEE International Electron Devices Meeting*, San Francisco, CA, pp. 29.1.1-29.1.4, Dec. 2017. DOI: 10.1109/IEDM.2017.8268472

[18] G. Yeap *et al*., "5nm CMOS Production Technology Platform featuring full-fledged EUV, and High Mobility Channel FinFETs with densest 0.021µm2 SRAM cells for Mobile SoC and High Performance Computing Applications," *2019 IEEE International Electron Devices Meeting (IEDM)*, 2019, pp. 36.7.1-36.7.4, doi: 10.1109/IEDM19573.2019.8993577.

[19] H. Yu, M. Schaekers, E. Rosseel, A. Peter, J.-G. Lee, W. B. Song, S. Demuynck, T. Chiarella, L. Ragnarsson, S. Kubicek, J. Everaert, N. Horiguchi, K. Barla, D. Kim, N. Collaert, A. V. Thean, and K. De Meyer, "1.5×10^{-9} Ωcm^2 Contact Resistivity on Highly Doped Si:P Using Ge Pre-amorphization and Ti Silicidation," *2015 IEEE International Electron Devices Meeting*, Washington, DC, 2015, pp. 21.7.1-21.7.4. DOI: 10.1109/IEDM.2015.7409753

[20] N. Raghavan, K. L. Pey, K. Shubhakar and M. Bosman, "Modified Percolation Model for Polycrystalline High-k Gate Stack With Grain Boundary Defects," in *IEEE Electron Device Letters*, vol. 32, no. 1, pp. 78-80, Jan. 2011, doi: 10.1109/LED.2010.2085074.

[21] A. Padovani and L. Larcher, "Time-dependent dielectric breakdown statistics in SiO2 and HfO2 dielectrics: Insights from a multi-scale modeling approach," *2018 IEEE International Reliability Physics Symposium (IRPS)*, 2018, pp. 3A.2-1-3A.2-7, doi: 10.1109/IRPS.2018.8353552.

[22] Choi, Woo Y., Gyuhan Yoon, Woo Y. Chung, Younghoon Cho, Seongun Shin, and Kwang H. Ahn. 2019. "A Technology-Computer-Aided-Design-Based Reliability Prediction Model for DRAM Storage Capacitors" *Micromachines* 10, no. 4: 256. https://doi.org/10.3390/mi10040256

[23] B. Brar, G. D. Wilk, and A. C. Seabaugh, “Direct extraction of the electron tunneling effective mass in ultrathin SiO_2”, Appl. Phys. Lett. 69, 2728 (1996);

[24] Ravi Kumar Chanana, “A new method of calculating charged deep level defects density in doped semiconductors from the band offsets of MIS device interfaces”, IOSR Journal of Applied Physics (IOSR-JAP) e-ISSN: 2278-4861.Volume 8, Issue 4 Ver. I (Jul. - Aug. 2016), PP 53-56

[25] D. Veksler, G. Bersuker, A. Koudymov and M. Liehr, "Analysis of Charge-Pumping Data for Identification of Dielectric Defects," in *IEEE Transactions on Electron Devices*, vol. 60, no. 5, pp. 1514-1522, May 2013, doi: 10.1109/TED.2013.2249070.

[26] J. McPherson, J-Y. Kim, A. Shanware, and H. Mogul, “Thermochemical description of dielectric breakdown in high dielectric constant materials”, Appl. Phys. Lett. 82, 2121 (2003)

[27] A. Padovani, L. Larcher, G. Bersuker and P. Pavan, "Charge Transport and Degradation in HfO_2 and HfO_x Dielectrics," in *IEEE Electron Device Letters*, vol. 34, no. 5, pp. 680-682, May 2013, doi: 10.1109/LED.2013.2251602.

[28] P. C. Feijoo, T. Kauerauf, M. Toledano-Luque, M. Togo, E. San Andres and G. Groeseneken, "Time-Dependent Dielectric Breakdown on Subnanometer EOT nMOS FinFETs," in *IEEE Transactions on Device and Materials Reliability*, vol. 12, no. 1, pp. 166-170, March 2012, doi: 10.1109/TDMR.2011.2180387.

[29]] Kolahdouz, Mohammadreza & Hållstedt, Julius & Östling, Mikael & Wise, R & Radamson, Henry. (2008). Selective Epitaxial Growth with Full Control of Pattern Dependency Behavior for pMOSFET Structures. ECS Transactions. 16. 10.1149/1.2986762.

[30] F Deng, RA Johnson, PM Asbeck, SS Lau , “Salicidation process using NiSi and its device application” Journal of Applied Physics 81, 8047 (1997)

[31] M. Py, J. P. Barnes, P. Rivallin, A. Pakfar, T. Denneulin1, D. Cooper, and J. M. Hartmann. Characterization and modeling of structural properties of SiGe/Si superlattices upon annealing, Journal of Applied Physics 110, 044510 (2011)

[32] Tsung-Hsi Yang, Guangli Luo, E. Y. Chang, Tsung-Yeh Yang, Hua-Chou Tseng and Chun-Yen Chang, "Study of nickel silicide contact on Si/Si1-xGex," in *IEEE Electron Device Letters*, vol. 24, no. 9, pp. 544-546, Sept. 2003

[33] G. Sasso, N. Rinaldi, G. Matz and C. Jungemann, "Accurate Mobility and Energy Relaxation Time Models for SiGe HBTs Numerical Simulation," 2009 International Conference on Simulation of Semiconductor Processes and Devices, San Diego, CA, USA, 2009, pp. 1-4, doi: 10.1109/SISPAD.2009.5290206.

[34] J. Korn, "Device Simulation of High-Performance SiGe Heterojunction Bipolar Transistors," Ph.D. dissertation, Department of Electrical Engineering and Computer Science, Technische Universität Berlin, Berlin, Germany, 2018.

[35] P. Lu *et al.*, "Source/Drain Extension Doping Engineering for Variability Suppression and Performance Enhancement in 3-nm Node FinFETs," in *IEEE Transactions on Electron Devices*, vol. 68, no. 3, pp. 1352-1357, March 2021, doi: 10.1109/TED.2021.3052432.

[36] Peng Lu, "Doping Profile Engineering for Advanced Transistors", Ph.D. dissertation, Department of Electrical Engineering, University of California, Los Angeles, United States of America, 2020.

[37] M. -C. Wang, C. -H. -T. Chang, H. -L. Hu, Z. -W. Zhong, G. -X. Qiu and W. -H. Lan, "Channel Surface Integrity with 2.4nm High-k Gate Dielectric under DPN Treatment at Different

Annealing Temperatures," *2021 9th International Symposium on Next Generation Electronics (ISNE)*, 2021, pp. 1-4, doi: 10.1109/ISNE48910.2021.9493656.

[38] J.-. Everaert, T. Conard and M. Schaekers, "SiON Gate Dielectric Formation by Rapid Thermal Oxidation of Nitrided Si.," 2005 13th International Conference on Advanced Thermal Processing of Semiconductors, 2005, pp. 135-138, doi: 10.1109/RTP.2005.1614328.

[39] T. A. Karatsori, C. G. Theodorou, S. Haendler, N. Planes, G. Ghibaudo and C. A. Dimitriadis, "Characterization and Modeling of NBTI in Nanoscale UltraThin Body UltraThin Box FD-SOI MOSFETs," in IEEE Transactions on Electron Devices, vol. 63, no. 12, pp. 4913-4918, Dec. 2016, doi: 10.1109/TED.2016.2620720.

[40] S. Mishra et al., "A Simulation Study of NBTI Impact on 14-nm Node FinFET Technology for Logic Applications: Device Degradation to Circuit-Level Interaction," in IEEE Transactions on Electron Devices, vol. 66, no. 1, pp. 271-278, Jan. 2019, doi: 10.1109/TED.2018.2875813.