Compact Modeling of Advanced CMOS and Emerging Devices for Circuit Simulation

Yen-Kai Lin

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2019-28
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2019/EECS-2019-28.html

May 9, 2019

Compact Modeling of Advanced CMOS and Emerging Devices for Circuit Simulation

By

Yen-Kai Lin

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering − Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Chenming Hu, Chair
Professor Sayeef Salahuddin
Professor Junqiao Wu

Spring 2019

Compact Modeling of Advanced CMOS and Emerging Devices for Circuit Simulation

Abstract

Compact Modeling of Advanced CMOS and Emerging Devices for Circuit Simulation

by

Yen-Kai Lin

Doctor of Philosophy in Engineering − Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Chenming Hu, Chair

Compact model plays an important role in designing integrated circuits and serves as a bridge to share the information between foundries and circuit designers. Since various flavors of transistor architectures like FDSOIs and FinFETs are proposed to improve device performances, the accurate, fast, and robust compact models, which are capable of reproducing the very complicated transistor characteristics like transconductance, are urgently required. Novel device concept, such as tunnel FETs (TFETs) and negative capacitance FETs (NCFETs), needs new device modeling methodology and understanding of device physics. In addition to transistors, memory device like magnetic tunnel junction (MTJ) compact model is also crucial for circuit designs. This dissertation presented the advanced research on compact models for the state-of-the art transistor and memory technologies: FDSOIs, FinFETs, TFETs, NCFETs, and MTJs.

Due to the limitations in the aggressively scaled planar transistors, the devices with good electrostatic control are discussed and modeled into the industry standard model − BSIM-IMG for FDSOIs and BSIM-CMG for multi-gate FETs. Although the dynamic back-gate bias change help reduce the static power in FDSOIs, the leakages, overlap capacitance, and carrier transport are thus showing back-gate bias-dependence. The enhanced gate-related leakage, overlap capacitance, and mobility compact models are validated against the silicon data and incorporated into BSIM-IMG. The leakages through subsurface path and source-to-drain direct tunneling due to extremely short channel are also included in this work, which are in excellent agreement with the technology computer-aided design (TCAD) and atomistic simulations. The computationally efficiency of these models are the key solutions for evaluating the circuit performance of future technology nodes.

Two paradigms of steep subthreshold slope transistors − TFETs and NCFETs as the promising candidates for future Internet of Things (IoT) and logic/analog applications are also presented in this thesis. TFET has a gated p-i-n diode structure, where the current relies on direct band-to-band tunneling in source/channel junction. Such tunneling mechanism breaks the tradition limitation of MOSFET turn-ON characteristics called the Boltzmann tyranny. The improvements in power consumption and delay of circuits are thus the emphasis and attention of device community, where the need of TFET compact model is fulfilled with the developed model in this work. NCFET is rapidly emerging as a preferred replacement for traditional MOSFET since the recent discovery of ferroelectric (FE) materials to amplify the voltage suggests

that further scaling supply voltage is possible with the CMOS-compatible fabrication process of NCFET. The short channel effect, ferroelectric variability, and spacer optimization design are the focus in this thesis. The compact model of NCFET is improved to be more predictive for ferroelectric properties with verification against TCAD simulations. Monte-Carlo method is carried out in FE variability study, where the main finding is that the dielectric phase is critical but fortunately is theoretically possible to be absent. The spacer design reveals that further engineering the capacitance matching via parasitic capacitance is the key solution for future technology nodes.

In addition to transistor compact models and physics, the memory device − spin-transfer-torque magnetic tunnel junction (STT-MTJ) is also presented. The resistances and critical currents are derived from the Landau-Lifshitz-Gilbert (LLG) equation and modeled analytically. The RC sub-circuit is found to describe the dynamic switching behavior of MTJ due to the precession and thermal fluctuation. The proposed MTJ compact model has been validated with silicon data from the industry and is capable of simulating a memory circuit with previously mentioned BSIM models.

Dedicated to my family

Contents

List of Figures

List of Tables

Acknowledgements

First of all, I would like to express, with all my heart and my sincere gratitude, to my research advisor, Professor Chenming Hu. He has supported me in my professional and personal life for four years of guidance and mentorship and maintained an exciting research environment within the BSIM group. His big enthusiasm and wealth of technical knowledge about research motivated me to keep forward to overcome the challenges and to generate new ideas. He always encouraged me to creatively understand, analyze and express engineering problems. In the personal aspect, he always accommodated to my personal needs and gave me a lot of freedom. It is my great pleasure and honor to have known him.

I am also greatly indebted to thank my co-advisor, Professor Sayeef Salahuddin, for his guidance in negative capacitance transistor research and for being the chair of my qualifying exam committee and member of my dissertation committee. His insights on solid state physics and device physics have inspired me a lot during the past four years. Serving as a teaching assistant of his course *Integrated Circuit Devices* and *Solid State Electronics* in 2018 improved my teaching and presentation skills, and most importantly Professor Salahuddin really helps me a lot fulfill the teaching requirement.

I also would like to thank Professor Junqiao Wu in the materials science and engineering department for serving on my qualifying exam and dissertation committees, as well as Professor Ali Javey for serving as the member of my qualifying exam committee. The fruitful discussions with them regarding my research were beneficial to shape my dissertation. Also thank Professor Jaeseok Jeon for hiring me as a teaching assistant in his course *Integrated Circuit Devices* in Fall 2018 for meeting the teaching requirement.

Life is hard as a PhD student. I am lucky and privileged to be able to work with talented fellow graduate students, postdocs, and industrial sponsors. I would like to thank my colleagues in BSIM group and Sayeef's group: Sourabh, Yogesh, Juan, Angada, Aditya, Huan-Lin, Harshit, Pragya, Ming-Yen, Avirup, Yu-Hung, Korok, Daewoong, Jiuren, and Shehrin, for their valuable discussion about research works with me, and all the companies who financially supported our works in the past four years: Semiconductor Research Corporation (SRC), Compact Modeling Council (CMC), Berkeley Device Modeling Center (BDMC) with its sponsors, Berkeley Center for Negative Capacitance Transistors (BCNCT) with its sponsors, and Ministry of Education in Taiwan. I also would like to thank my friends: Li-Chen, Chen-Yu, Ping-Hung, Wen-Chin, Jerry, Yu-Hung, Hsin-Yi, Yow-Gwo, Hsiou-Yuan, Howard, Fang-Chieh, Chi-Heng, Tim, Angela, Jeffery, Michael, Yi-An, Huanting, Ching-Fu, Jason, Alex, Rosa, Fu-Chi, Jong-Kai, Jesse, and all my BATS friends, for regularly working out with (spotting) me and having fun with me.

Finally, I would like to thank my girlfriend, Li-Chen, for being so supportive during the last year of my PhD journey, and also would like to express my gratitude to my family, especially my parents, for their lasting love, endless support, and encouragement in the past years. Without them, I cannot dedicate to my research, and this work would have not been possible.

Chapter 1

Introduction

Since early 1960s, the semiconductor industry has grown up significantly, which also leads the growth in other related fields and changes our live styles. The semiconductor industry consists of the fabrication of semiconductor devices and circuit design companies where they communicate with each other often to exchange the information from product definition to final product validation through a set of technology definitions called a process design kit (PDK). The compact SPICE model serves as a bridge to connect the foundry and circuit designers. This dissertation will present some key enhancements in the existent Berkeley Short-channel IGFET Model (BSIM) family SPICE models [1] and derivations for the emerging devices, and discuss their applications to the integrated circuit (IC) design.

1.1 Why Compact Model Is Needed?

1.1.1 TCAD and Compact Model

Semiconductor devices consist of multiple materials in various structures on wafers. To predict the device performances, the underlying physics in such material properties is required. Usually, the technology computer-aided design (TCAD) tools like Sentaurus TCAD [2], are run by device engineers/researchers to look deeply in the potential, electric field, and so on of devices. Such TCAD tools generally solve the physical equations like Poisson's equation and Schrodinger's equation numerically based on the finite-element method. Although TCAD enables researchers to evaluate and optimize devices, the complexity of solving differential equations self-consistently results in longer simulation time (from few minutes to few days) which may not be satisfying with consideration of the product design cycle. Therefore, for circuit design perspective, a speedy way to evaluate circuit performances based on silicon data is required.

Compact models for semiconductor devices, including transistors and memories, are the concise mathematical descriptions of their complicated behaviors, and they are usually implemented in a computer programming language like C or Verilog-A. Compact models enable accurate IC design simulations and are an integral part of the PDK. There are four benefits of compact model against TCAD simulation: fast, robust, accurate, and flexible. The structure of an BSIM family model [3] is shown in Fig. 1.1. The core model consists of current-voltage ($I-V$) and capacitance-voltage ($C-V$) models for long channel devices since the equation can be further simplified to

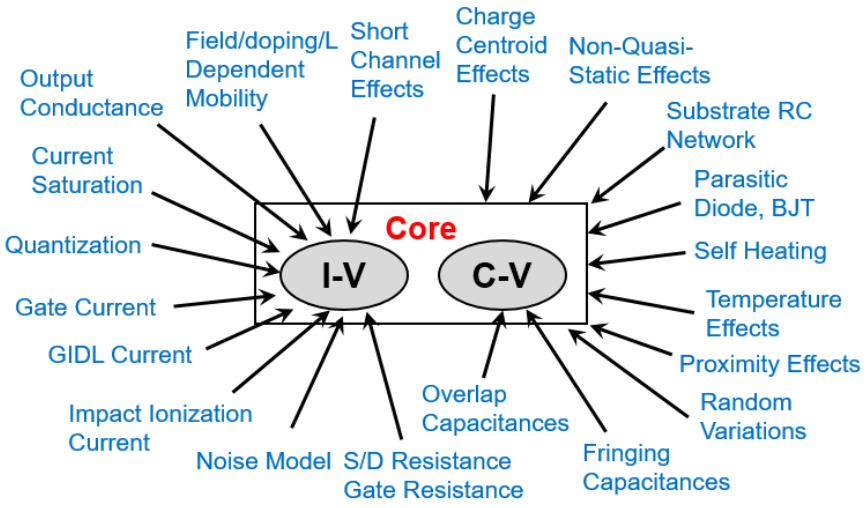


Figure 1.1: Structure of BSIM compact model, including core model for current and capacitance calculations and real device (non-ideal) effects.

be analytical. The real device effects, such as the mobility degradation and the short channel effects, are added into the core model as the modifications, which ensures that the models are written in analytical manner to achieve good simulation efficiency. Normally, a bias point takes about 10 μs in compact model. Furthermore, the robustness over wide range of parameter, bias, and temperature, is also achieved since the singularities can be avoided by clamping or smoothing the analytical equations. With a good parameter extraction procedure, the model can reproduce the silicon data very accurately with RMS error less than 1% [4]. Finally, the compact models are also flexible with model parameters to accommodate technologies from multiple foundries and to tolerate process variations.

As the advanced device structures like FinFET [5] and gate-all-around (GGA) FET [6] were proposed, some more new device features should be added into the compact model for circuit simulations, which can be done by building up the new models on the top of existent model like BSIM. Furthermore, the new device prototype also brings new physical concepts which requires a distinct compact model. For example, negative capacitance field-effect transistor (NCFET) involves solving Landau-Khalatnikov equation and the underlying transistor physics [7], so the core model needs to be refined which will be discussed in Chapter 5. Therefore, the compact model is expandable to capture future technologies.

1.1.2 Overview of BSIM

Berkeley Short-channel IGFET Model (BSIM) is an industry standard compact SPICE model. The research on the complexity of the transistor model initiated in early 1980s at Berkeley [8, 9]. As device structure evolves from planar MOSFET, SOI MOSFET, to multi-gate MOSFET, BSIM-BULK, BSIM-IMG, and BSIM-IMG have been developed to address the needs for IC design. The timeline of BSIM family compact model is shown in Fig. 1.2. All BSIM models have the inputs of the terminal voltages and outputs of currents and charges, and they share similar model architecture shown in Fig. 1.1. Currently, BSIM1, 2, 3, and 5 for the planar MOSFET are

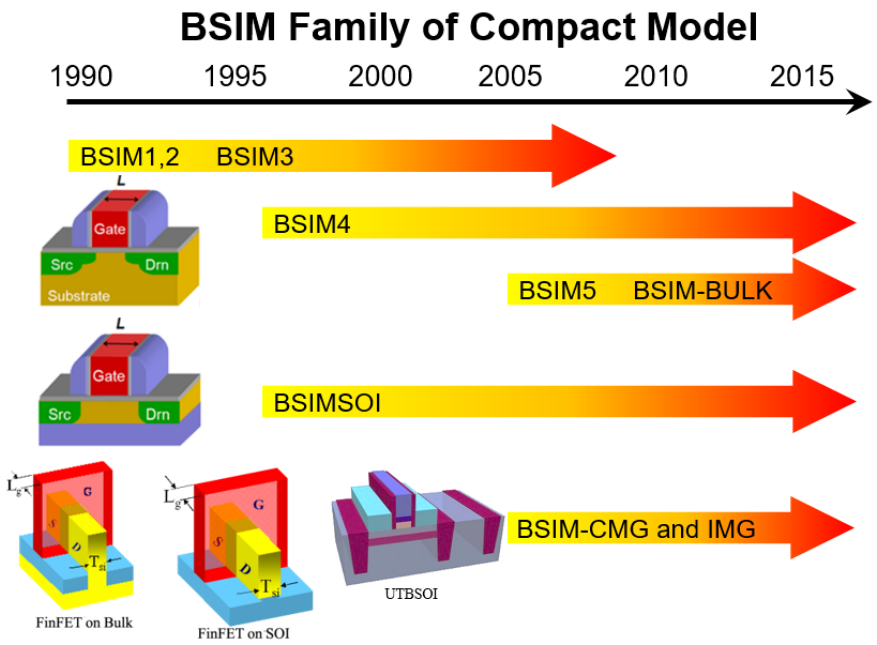


Figure 1.2: Evolution and timeline of BSIM family of compact models [10]. BSIM1,2,3,5 now are not supported by BSIM group, while BSIM4, BSIM-BULK, -SOI, -CMG, and -IMG are maintained and updated by BSIM group.

no longer supported, but BSIM4 [11] is still under maintenance. BSIM1 to BSIM5 are implemented by C, while BSIM-BULK, -CMG, and -IMG are implemeneted by Verilog-A. BSIM-BULK is a charge-based model for planar MOSFETs with an emphasis on satisfaction of RF applications against the threshold voltage-based BSIM4 due to continuity nature of the charge-based model [3]. A sub-surface leakage in an ultra-scaled planar MOSFET as discussed in Chapter 3 is built on top of BSIM-BULK and is validated with TCAD simulated data. BSIM-IMG (Independent Multi-Gate) is a surface potential-based (continuous) model for ultra-thin body SOI MOSFETs (UTBSOI) or fully-depleted SOI MOSFET (FDSOI). The accurate surface potentials on front and back sides are derived simultaneously by solving Poisson's equation, enabling multiple combinations of biases [12, 13]. The back-gate bias effects are the distinct features in BSIM-IMG and will be further discussed in Chapter 2. BSIM-CMG (Common Multi-Gate) is a surface potential-based model for multi-gate MOSFETs like FinFETs (triple gates) [14], which is an unified model for arbitrary channel cross-sectino shape [15]. Due to the superior scalability of multi-gate devices, the channel length can be shrunk down to sub-10 nm regime which leads to fundamental limit of source-to-drain tunneling as being described in Chapter 3.

1.2 Advanced CMOS and Emerging Devices

1.2.1 Challenges in CMOS Scaling

Over the past few decades, the dimension of complementary MOSFET (CMOS) has been aggressively reduced with the guidance of Moore's law. The scaling enables much denser circuit and more functional chip in a given area and decreases the man-

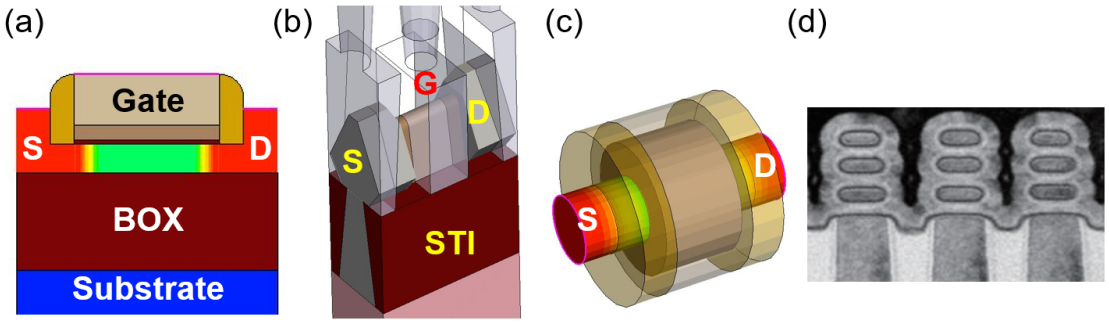


Figure 1.3: Illustrations of various flavors of multi-gate transistors: (a) UTBSOI (FDSOI), (b) FinFET, (c) Gate-all-around (GAA) nanowire FET, (d) Stacked nanosheet GAA FET [22].

ufacturing cost. However, simple scaling has encountered more and more challenges due to fundamental device physics limitations. Due to bad electrostatic gate control, short channel effects (SCEs) lead to threshold voltage (V_{TH}) roll-off and drain-induced barrier lowering (DIBL) [16]. In addition, the velocity saturation [16] and ballistic transport [17] further limit the improvement in device performances as the gate length is shrunk down below 100 nm or even smaller. Note that the total capacitance can be also scaled, and thus the speed of circuit is improved even without enhancement in the drive current. Gate tunneling leakage also becomes an issue in power consumption when the gate oxide thickness is down below 3 nm [18]. The high-κ dielectric was introduced to suppress the gate leakage by keeping the effective oxide thickness (EOT) but it could lead to reliability issues [19]. Metal gate electrode is also adopted to eliminate the undesired poly-depletion effect [16] and improve the mobility by screening the remote phonon scattering caused in high-κ dielectrics [20]. Even with these advances, MOSFET is still unable to be further aggressively scaled since the limitation of thinning EOT is inevitable. Therefore, to provide stronger gate control to resolve the short channel effects, a new approach is required to allow future reduction in channel length. The multi-gate transistors, including FDSOI and FinFET, are the promising approaches. Furthermore, applying novel device operation concepts, such as negative capacitance field-effect transistor (NCFET) and tunnel field-effect transistor (TFET) [21], has been also proposed.

1.2.2 State-of-the-Art Devices

Multi-Gate Transistors

The key advantage of multi-gate transistors, including FDSOI, FinFET, and GAA FET shown in Fig. 1.3, is their strong gate control to suppress SCEs since thin channel is wrapped by gates from multiple sides. The thin channel also allows lightly doped or even undoped channel which improves mobility by reducing Coulombic scattering and surface roughness scattering [16], and it cuts the most leaky path underneath the channel. Low doping concentration minimizes the random dopant fluctuation (RDF) [23]. The planar FDSOI transistor [Fig. 1.3 (a)] offers the dynamic threshold voltage control through the back-gate bias, which is favorable for low-power applications without relying on multiple channel doping concentrations [24]. The three-dimensional

multi-gate transistor like FinFET [Fig. 1.3 (b)] provides a wider channel width with a small footprint in area, which raises the drive current and is crucial for driving larger load capacitance [5, 25]. These benefits incentivize people to further work on multi-gate transistors. Recently, the stacked nanosheet GAA transistor [Fig. 1.3 (d)] [22] has been proposed to increase the effective channel width at the same footprint as FinFET, although the fabrication is more sophisticated. With such continuously aggressive scaling, the quantum mechanical confinement should be taken into account, and the sub-band effect may appear in the transconductance and capacitance [26]. Therefore, there is still room left for investigating multi-gate transistor physics, fabrication, and modeling.

Steep Subthreshold Transistors

Over past few decades, the transistor dimension has been reduced as mentioned earlier, whereas the supply voltage (V_{DD}) can not be scaled accordingly [Fig. 1.4 (a)] due to the limitation of traditional MOSFET turn-on mechanism [16] which leads to power consumption problem. The gate voltage is capacitively coupled to the channel and changes the surface potential of the channel. The carriers in the source allowed to surmount the potential barrier statistically follow the Boltzmann distribution, which is exponentially proportional to the surface potential. The drain current in the subthreshold region thus is an exponential function of the surface potential (ϕ_S).

$$I_{DS} \propto \exp\left(\frac{q\phi_S}{k_B T}\right) \tag{1.1}$$

Even the gate voltage is able to perfectly couple to the surface potential, one order magnitude of drain current only can be obtained by 60 mV or more at room temperature. In other words, the subthreshold slope (SS) cannot exceed 60 mV/dec. This is called the Boltzmann Tyranny. In real devices, the gate voltage cannot be perfectly coupled to the channel due to the additional interfacial capacitance and drain-coupled capacitance (SCEs), which is also true for the state-of-the-art multi-gate transistor [27] as shown in Fig. 1.4 (b). Therefore, the impetus to invent new prototype of device operation mechanisms was born to reduce V_{DD}.

Nano-electromechanical (NEM) switches [29], impact-ionization MOSFETs (i-MOS) [30], tunnel field-effect transistor (TFET) [31], and negative capacitance field-effect transistor (NCFET) [32], were being developed to overcome the Boltzmann tyranny and replace the traditional MOSFETs. However, NEM and i-MOS switches suffer from reliability and scalability issues [29]. TFET and NCFET are promising candidates for future technologies, such as low power Internet of Things (IoT), since they are fully compatible withe the modern semiconductor processes. The detailed operation mechanisms and compact models of TFET and NCFET will be discussed in Chapter 4 and 5.

Memory Technologies

The aggressively scaled transistors with small footprint enable more functionalities in a chip. Another important aspect of semiconductor chips is the memory. There is a wide variety of memories, such as dynamic random-access memory (DRAM), static

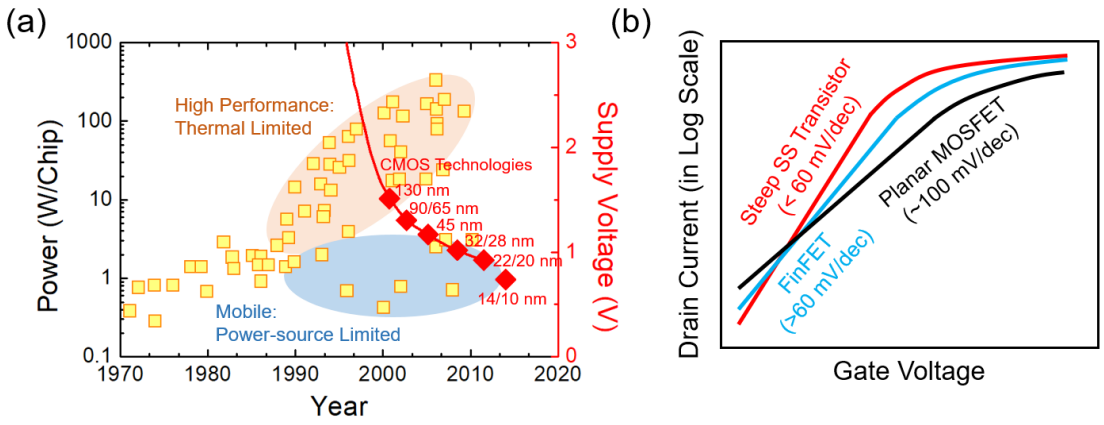


Figure 1.4: (a) Power density on chip over time for high performance and mobile CPUs and supply voltage scaling for CMOS technologies (reproduce from [28]). (b) Turn-on characteristics of steep SS transistor, FinFET, and planar MOSFET.

Table 1.1: Comparison of memory technologies (reproduce from [34]).

Technology	**DRAM**	**SRAM**	**NOR Flash**	**STT-MRAM**
Energy/bit (fJ)	1000	100	10^6	100
Write speed (ns)	20	1	1000	1−10
Read speed (ns)	30	1	10	1−10
Density (area in F^2)	6−10	>30	4−8	10−30
Endurance (cycles)	Very high	Very high	Low	Very high
Non-volatile	No	No	Yes	Yes
Standby power	Refresh current	Leakage current	None	None
Cost overhead versus CMOS	Separate process	Large area (6T)	Separate process	Back-end (BEOL) process
Non-volatile logic capability	No	No	No	Very limited due to power

random-access memory (SRAM), flash memory, and magnetoresistive random-access memory (MRAM), satisfying various needs of circuit applications. To further achieve high density and performance in semiconductor memories at low power consumption, new materials and structures were explored. One of promising memory devices is the spin-transfer torque magnetic tunnel junction (STT-MTJ), a paradigm of MRAM, since it serves as an universal memory with the speed of SRAM, the density of DRAM, and non-volatility of flash memories but simultaneously with power efficiency and higher endurance [33]. Table 1.1 shows the comparison of various memory technologies.

A STT-MTJ basically consists of two layers of ferromagnetic materials (fixed and free layers) and tunnel oxide. A typical MTJ resistance hysteresis loop and 1Transistor-1MTJ MRAM cell are shown in Fig. 1.5 (a) and (b), respectively. By flipping the magnetization in ferromagnetic material using the spin-polarized currents, the tunnel resis-

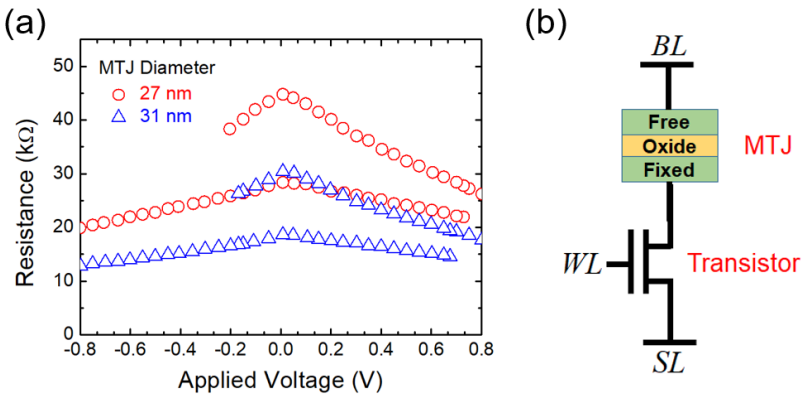


Figure 1.5: (a) A typical resistance hysteresis loop of MTJ (reproduce from [35]). (b) 1Transistor-1MTJ MRAM cell.

tance changes accordingly [36]. Because of the mentioned advantages of STT-MRAM, many companies, including IBM, Samsung, GlobalFoundries, Qualcomm, Applied Materials, TSMC, have developed the STT-MRAM technologies [37, 38, 39, 40, 41]. The compact model of STT-MRAM thus becomes increasingly attractive. The detailed operation mechanism of STT-MTJ and compact model methodology will be discussed in Chapter 6.

Chapter 2

Modeling Back-Gate Bias Effects in FDSOIs

2.1 Introduction

As mentioned in Chapter 1, FDSOIs offer dynamic threshold voltage control by varying back-gate bias (V_{BG}) without relying on multiple channel doping for devices to suppress the off-state leakage, which is desirable for low power logic applications. However, such applied V_{BG} alters the electrostatics in the channel, leading to back-gate bias-dependent physical phenomena in FDSOIs [24, 42]. In addition to logic circuits, in order to fulfill the need for image sensors, the thick front gate oxide is adopted to reduce the noises [43, 44, 45]. Furthermore, the thick front oxide is also helpful to sustain high electric field in high voltage devices [46]. However, the charge and current contributions from the back side channel is non-negligible if the front gate oxide is thick compared to the BOX thickness. The back side inversion effect [47, 48] now becomes important and affects the transconductance of FDSOIs, where an accurate transconductance model is required to describe the analog gain in analog signal processor and then output to the analog-to-digital converter in CMOS image sensor architecture design [49]. Therefore, the aim of this chapter is to analyze and develop compact models for the back-gate bias-dependent effects with validation of silicon data from industry and TCAD simulations.

2.2 Back-Gate Bias Dependent Leakages Models

2.2.1 Gate-Induced Drain Leakage (GIDL)

Gate-induced drain leakage (GIDL) is the band-to-band tunneling leakage at OFF-state, which strongly depends on the drain voltage [50]. Due to high drain-to-gate voltage ($V_{DG} = V_{DS} - V_{GS}$), an inversion layer is formed in source/drain extension (overlap) region and high electric field leads to band-to-band tunneling, which is typically modeled by [50]

$$I_{\mathrm{GIDL}} = \mathrm{AGIDL} \cdot W \cdot E^{\mathrm{PGIDL}} \cdot \exp\left(-\frac{\mathrm{BGIDL}}{E}\right) \tag{2.1}$$

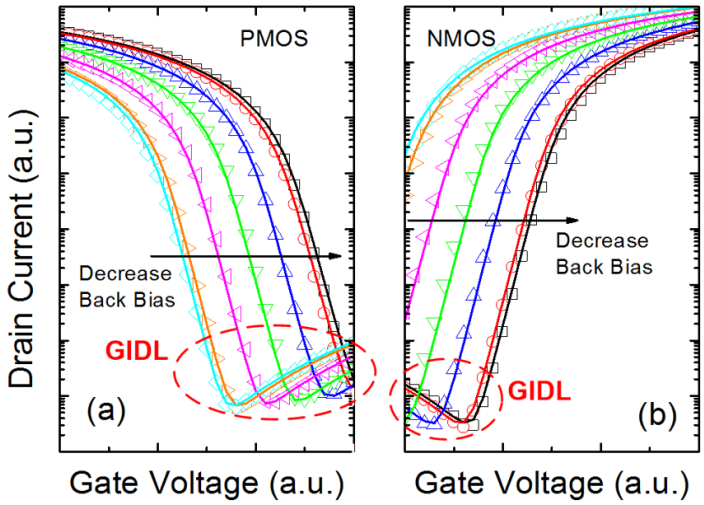


Figure 2.1: $I_{DS} - V_{GS}$ characteristics of FDSOI (a) P-MOSFET and (b) N-MOSFET. The GIDL currents show back-gate dependent behaviors. The lines and symbols represent models with BSIM-IMG and measured data, respectively. The drain is biased at saturation region. Back-gate bias is from positive (left) to negative (right) values.

where W is the device width, AGIDL, BGIDL, and PGIDL are the model parameters which are associate with source/drain material, and E is the electric field in the source/drain overlap region and is given by

$$E = \frac{V_{DS} - V_{GS} + V_{FBSD} - \text{EGIDL}}{\epsilon_{\text{ratio}} \cdot \text{EOT}} \tag{2.2}$$

where V_{DS}, V_{GS}, and V_{FBSD} are the drain voltage, front gate voltage, and flat band voltage of the overlap region, ϵ_{ratio} is the permittivity ratio of the semiconductor to the front gate oxide, EOT is the effective front gate oxide thickness, and EGIDL is the model parameter.

Fig. 2.1 (a) and (b) shows the back-gate bias-dependent GIDL currents of FDSOI P-MOSFET and N-MOSFET which are fabricated by an industry lab. By Gausss law, the effects of back-gate bias can be viewed as the threshold voltage shift in the overlap region, which makes V_{GS} in (2.2) becomes an effective front gate voltage V_{GSeff}

$$V_{GS} = V_{GSeff} - \text{VBGIDL} \cdot \gamma_0 \cdot (V_{BGS} - V_{FBSDBG} - \text{VBEGIDL}) \tag{2.3}$$

where VBGIDL and VBEGIDL are the model parameters for nonuniform doping in the overlap region, V_{BGS} is the back-gate bias with respect to the source, V_{FBSDBG} is the flat band voltage of the overlap region with respect to the back gate, and γ_0 is the capacitive coupling ratio between the body and the back-gate capacitance with the front gate capacitance [51]

$$\gamma_0 = -\frac{C_{\text{Si}} \cdot C_{\text{OX2}}}{(C_{\text{Si}} + C_{\text{OX2}}) \cdot C_{\text{OX1}}} \tag{2.4}$$

where C_{Si}, C_{OX2}, and C_{OX1} are the capacitances of the thin body, back gate, and front gate, respectively. The model parameter VBGIDL is around 1, while VBEGIDL is on the order of 100 mV. In (4), a fully depleted thin body (T_{Si}) is assumed so that C_{Si} is

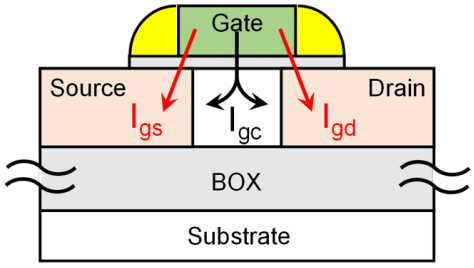


Figure 2.2: Schematic of leakage current components. $I_{\rm gs}$ and $I_{\rm gd}$ are the leakage currents at the overlap region. $I_{\rm gc}$ is flowing between the gate and the channel.

a constant ($= \epsilon_{\rm Si}/T_{\rm Si}$). From (2.3) and (2.4), at OFF-state ($V_{GS} < 0$), the magnitude of the effective front gate voltage increases with decreasing (increasing) the back-gate bias for N-MOSFET (P-MOSFET) so that the electric field in (2.2) becomes higher, giving rise to higher tunneling current. This model is implemented into BSIM-IMG, and the model shows good agreement with the experimental data, as shown in Fig. 2.1.

2.2.2 Gate Leakage

As the oxide continuously scales, the gate tunneling gets severe. The gate tunneling currents, including gate-to-channel ($I_{\rm gc}$) and gate-to-source/drain ($I_{\rm gs}$, $I_{\rm gd}$) as shown in Fig. 2.2, have been modeled [18, 52]. In general, the gate direct tunneling leakage current can be expressed as [52]

$$\begin{aligned} I_G =& W \cdot L \cdot A \cdot \left(\frac{T_{\rm oxref}}{t_{\rm ox}}\right)^{\rm NTOX} \cdot \frac{V_{GS(D)} \cdot V_{\rm aux}}{t_{\rm ox}^2} \\ & \cdot \exp\left[-B \cdot (\alpha - \beta|V_{\rm ox}|) \cdot (1 + \gamma|V_{\rm ox}|) \cdot t_{\rm ox}\right] \end{aligned} \tag{2.5}$$

where L is the length of tunneling region, A ($= 4.97232 \times 10^{-7}$ for N-MOSFET, 3.42537×10^{-7} for P-MOSFET using silicon and silicon dioxide parameters) and B ($= 7.45669 \times 10^{11}$ for N-MOSFET, 1.16645×10^{12} for P-MOSFET using silicon and silicon dioxide parameters) are the material-related constants, NTOX, α, β, and γ are the model parameters, $V_{\rm ox}$ is the oxide voltage, $t_{\rm ox}$ is the physical thickness of the oxide, $T_{\rm oxref}$ is the reference oxide thickness at which all the parameters are extracted, and $V_{\rm aux}$ (in unit of volt) is an auxiliary function which represents the density of tunneling carriers as well as available states. Equation (2.5) has been widely used in gate currents of planar MOSFETs [52]. However, due to lightly doped thin film used in the state-of-the-art devices, the body effect or back bias effect on gate current characteristic has not received much attention and physics remains unclear [53].

Fig. 2.3 shows the gate leakage current characteristics of long- and short-channel FDSOI N-MOSFETs fabricated by an industry lab. The source and drain are shorted when measuring gate current. Note that in Fig. 2.3 the gate length of long-channel device is 100 times longer than that of short-channel device, while their gate widths are the same. The gate leakage current exhibits V_{BG}-dependence, indicating that the back-gate bias affects the electrostatics in the channel and source/drain overlap regions. Interestingly, the gate leakage current of long-channel device at accumulation

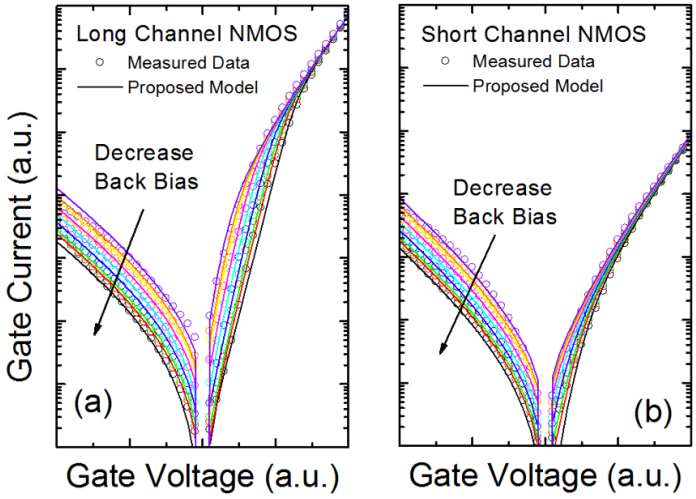


Figure 2.3: I_{GS}−V_{GS} characteristics of FDSOI (a) long- and (b) short-channel N-MOSFETs. The source and drain are shorted when measuring. The ratio of long and short gate length is 100 with the same gate width.

regime ($V_{GS} < 0$, left side) is approximately the same as that of short-channel device, implying that the current is flowing through the overlap region whose dimension does not scale with channel dimension [54]. Note that the term accumulation used here is for convenience. There is no accumulation layer in FDSOI device because of insufficient supply of majority carriers from the thin body [55]. Thus, the gate-to-source ($I_{\rm gs}$) and gate-to-drain ($I_{\rm gd}$) currents dominate the gate leakage at accumulation region. Because part of the overlap region is heavily doped, the gate-to-overlap region is assumed to be an metal-insulator-metal capacitor, although some lightly doped region is present so that the oxide electric field in the overlap region is affected by the back-gate bias due to the vertical and lateral electric field profile. This assumption means that the gate voltage mostly drops on the oxide and simplifies the auxiliary function. Therefore, based on (2.5), $|V_{\rm ox}|$ and $V_{\rm aux}$ of gate-to-source (drain) currents can be written as

$$\begin{aligned} |V_{\rm ox}| &= |V_{GS(D)} - V_{FBSD} - [\phi_{s\rm SD} - \eta \cdot \gamma_0 \cdot (V_{BGS} - V_{FBSDBG})]\,| \\ &\approx |V_{GS(D)} - V_{FBSD} + \eta \cdot \gamma_0 \cdot (V_{BGS} - V_{FBSDBG})\,| \end{aligned} \tag{2.6}$$

and

$$V_{\rm aux} = V_{\rm ox} \tag{2.7}$$

where η is the model parameters for nonuniform doping in the overlap region. Note that, in (2.6), the absolute value is taken using the smoothing function in order to avoid the possible discontinuity in the derivative of the current. In (2.6), $\phi_{s\rm SD}$ is the band bending in the overlap region without back-gate effect, which is negligible. A negative back-gate bias can raise the band in the overlap region and thus reduce the oxide electric field which is demonstrated by Sentaurus TCAD simulation [2] in Fig. 2.4. Thus, the gate-to-source (drain) leakages increase with the back-gate bias.

In Fig. 2.3, at inversion bias region ($V_{GS} > 0$, right side), it is observed that the gate leakage current of long-channel device is 100 times larger than that of short-channel device, which is exactly the same as the ratio of gate lengths of these two devices. This fact implies that the gate-to-channel current ($I_{\rm gc}$) dominates because

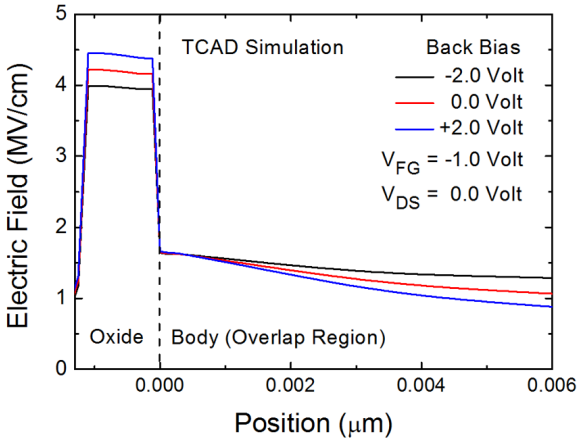


Figure 2.4: Simulated electric field at $V_{GS} < 0$ with various back biases in the overlap region. The front gate oxide electric filed increases with back bias.

it is proportional to the gate length. Note that the gate-to-channel current partition due to the drain voltage has been modeled in BSIM4 [52] and BSIM-IMG [12], which is important when the drain bias is nonzero. At inversion, V_{aux} can be represented as the average charge q_{ia} (in unit of volt) in the channel, which is directly calculated by the core model of BSIM-IMG model [12]. V_{ox} can be written as

$$V_{\text{ox}} = V_{GS} - V_{FB} - \zeta \cdot \phi_s \tag{2.8}$$

where ζ is the model parameter to capture nonuniform electric potential along the channel due to drain voltage, V_{FB} is the flat band voltage, and ϕ_s is the surface potential of the front side of body at the source which is V_{BGS}-dependent and is determined by the core model of BSIM-IMG model [12]. From (2.8), ϕ_s decreases with decreasing V_{BGS} [12, 56] so that V_{ox} increases with decreasing V_{BGS}. However, the available charges for tunneling (q_{ia}) exponentially increase with V_{BGS} before the threshold. Thus, the net result is that the gate-to-channel leakage current increases with V_{BGS}. Furthermore, at weak inversion the inversion charges are not abundant, so the back-gate effect is still dominant, giving rise to V_{BGS}-dependence. Nevertheless, at strong inversion the inversion charges are able to screen out the back-gate electric field, which means that the back-gate effect becomes less significant. This fact leads that the gate leakage currents at high V_{GS} for various V_{BGS} start to merge [13].

Equations (2.5)−(2.8) are implemented into BSIM-IMG model. Fig. 2.3 shows the model results are in good agreement with the measured data of both long- and short-channel devices. To examine the model, different drain biases are applied for gate currents. Fig. 2.5 shows the long-channel gate and drain currents versus gate voltage at various back-gate biases at linear and saturation regions. After fitting the drain current to get accurate q_{ia}, the gate current is fitted. The drain currents of both linear and saturation bias regions at OFF-state ($V_{GS} < 0$, left side) are dominated by the overlap gate current (I_{gs} and I_{gd}) in this device. This leakage current shows opposite V_{BGS}-dependence to the GIDL mentioned in Section 2.2.1. This distinct characteristic is helpful for distinguishing the leakage components and is a guideline to reduce leakages. Note that the gate current at saturation bias region shows V_{BGS}-dependence crisscrossing at high V_{GS} because the V_{ox} effect overwhelms q_{ia} effect. The

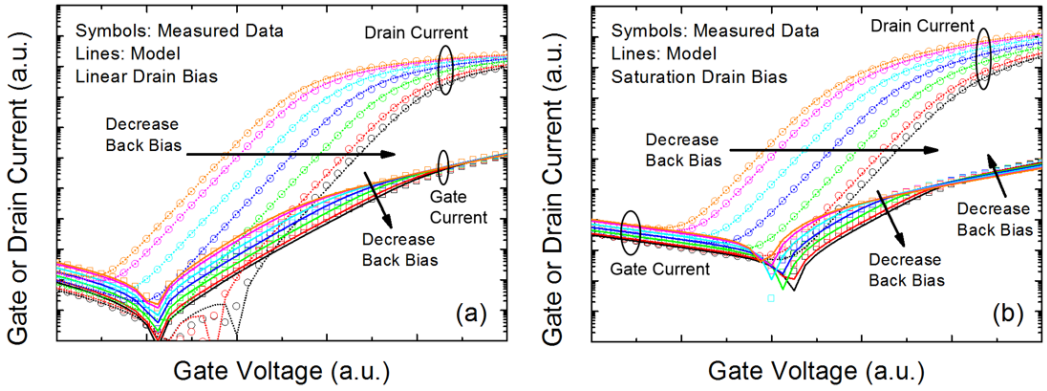


Figure 2.5: Gate and drain currents versus gate voltage at various back-gate biases and at (a) linear and (b) saturation drain biases in a long-channel N-MOSFET. The back-gate bias varies from positive to negative values.

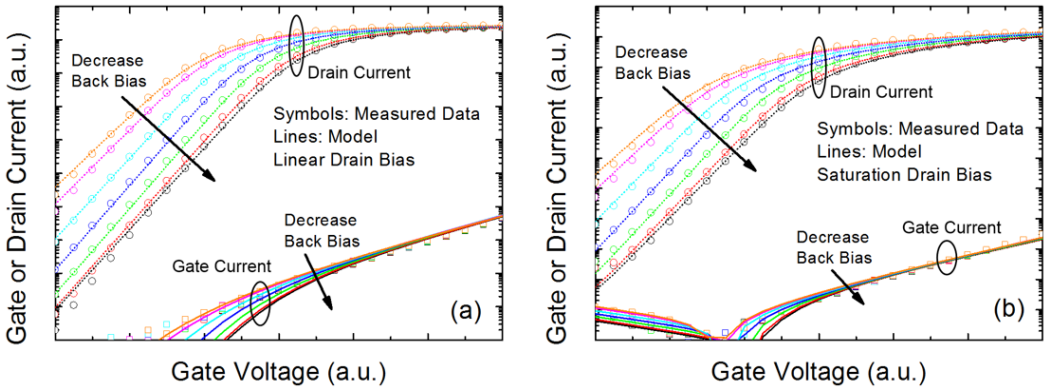


Figure 2.6: Gate and drain currents versus gate voltage at various back-gate biases and at (a) linear and (b) saturation drain biases in a short-channel N-MOSFET. The back-gate bias varies from positive to negative values.

drain voltage would reduce the average charges in the channel as well as the screening effect for the back electric field to the front electric field. Fig. 2.6 shows the short-channel gate and drain currents. The proposed model also matches the experimental data well. Interestingly, the short-channel gate currents do not show V_{BGS}-dependence crisscrossing at high V_{GS} because the charges are more abundant than that in long channel due to lower threshold voltage. Also, the charge distribution is also more uniform compared to that of long channel, as shown in Fig. 2.7. Thus, the electric field from the back gate is screened out, leading to less V_{BGS}-dependence but not crisscrossing.

2.3 Dual Mobility Model

2.3.1 Transconductance Behavior

One of the advantages of the FDSOI MOSFET is the changeable threshold voltage via applying the back-gate bias. If the back-side inversion is absent (only depletion at the back interface), the threshold and thus the transconductance (g_m) are only

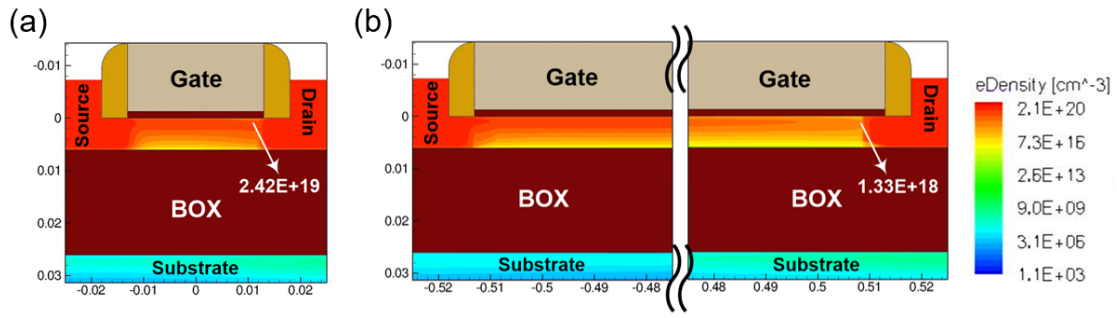


Figure 2.7: TCAD simulated electron density in (a) short-channel ($\approx$ 20 nm) and (b) long-channel ($\approx$ 1 μm) devices at $V_{GS} = 1V$, $V_{DS} = 0.8V$, and $V_{BGS} = -2V$. The electron distribution is uniform in short-channel device.

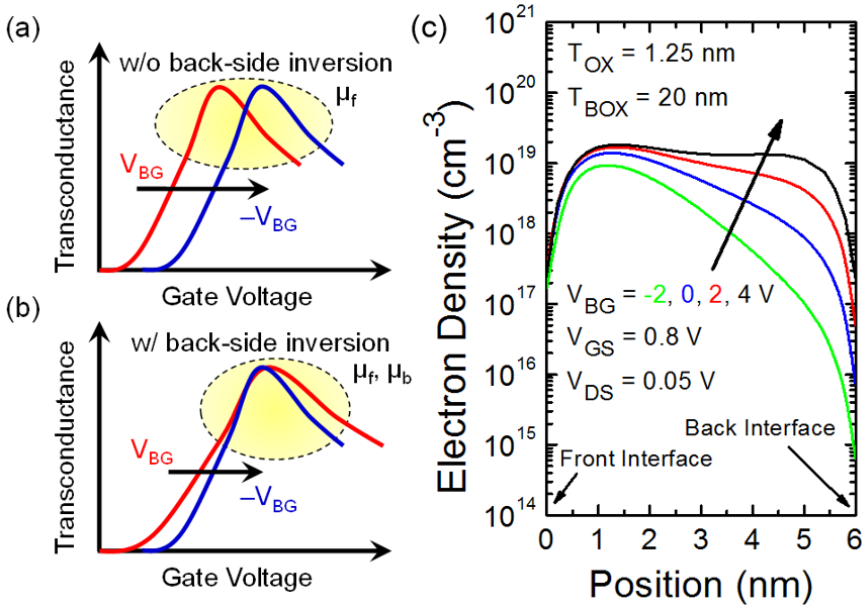


Figure 2.8: (a) Typical and (b) anomalous transconductances of the FDSOI nMOSFETs. Due to the back-side inversion, both the front- and backside mobilities affect the transconductance behavior. (c) TCAD simulated electron density with various V_{BG}.

shifted by the back-gate bias via capacitive coupling, indicating that only the front mobility is important, as shown in Fig. 2.8 (a). In contrast, if there is an inversion at the back interface, the contribution of the back channel is not negligible and the back-side mobility should be considered, as shown in Fig. 2.8 (b). In this case, the transconductance behavior with various V_{BG} is not just threshold shift. The peak and the slope of transconductance may change due to the combination of the front and back mobilities, depending on the interfacial qualities of the front and back sides. This anomalous transconductance behavior can be observed when: 1) applying high V_{BG} [positive (negative) for n(p)MOSFET] to invert the back-side channel and 2) the front- (back)- gate oxide is thick (thin). Sentaurus technology computer-aided design (TCAD), which includes the density gradient quantization model and thin-layer mobility model with Philips unified mobility model [2], is performed for the charge distribution. As shown in Fig. 2.8 (c), a high V_{BG} leads to the back-side inversion,

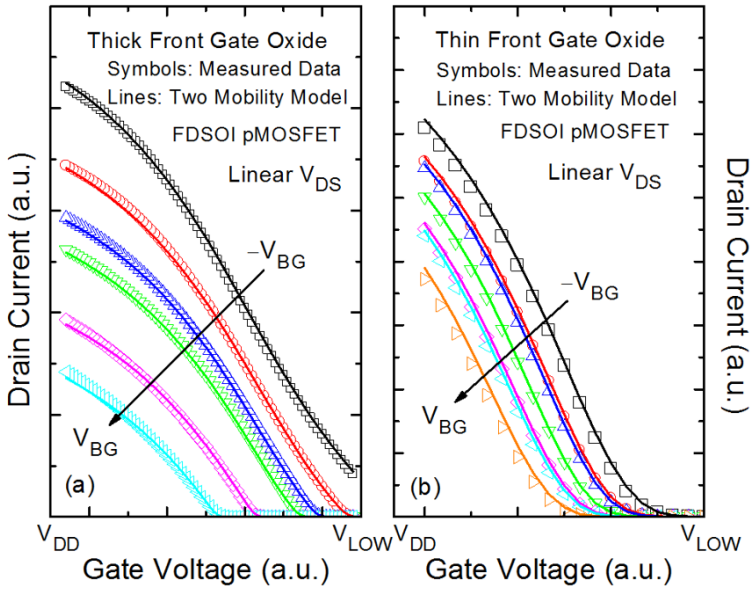


Figure 2.9: $I_{DS} - V_{GS}$ curves of FDSOI pMOSFETs with (a) thick and (b) thin front-gate oxides at linear drain bias region. The same BOX and body thicknesses are used in both devices. The back-gate bias is applied from negative to positive.

so the total current is affected by the front and back mobilities. Furthermore, the potential in the channel is determined by the frontand back-gate capacitances. A thick front-gate oxide would lead to less charges in the front channel and thus higher impact of the back-side charges and mobility. Therefore, both considering the front and back mobilities are crucial for the FDSOI MOSFET with wide range of biases and various device structures.

The linear region currents of FDSOI pMOSFETs with thick- and thin-front-gate oxide devices fabricated by a commercial foundry are measured, as shown in Fig. 2.9. Note that the device dimensions and bias conditions are similar to that in [43]. These two devices show distinct transconductance behaviors as shown in Figs. 2.10 and 2.11. The peak of transconductance of the thick front-gate oxide device (Fig. 2.10) non-monotonically shifts (to the right at low $|V_{GS}|$ and then to the left at high $|V_{GS}|$) with increasing back-gate bias voltage (denote as anomalous transconductance), while that of the thin frontgate oxide device (Fig. 2.11) exhibits a monotonic characteristic with the back-gate bias voltage (to the left). It is observed that the single mobility model (dashed lines in Fig. 2.10) cannot capture the non-monotonic shift of the transconductance peak. The transconductance peak is directly related to the turn-on phenomenon of the multiple channels as well as the mobility [57] and a new model is required to capture this non-monotonic effect.

2.3.2 Modeling and Parameter Extraction

Due to the back-gate bias, the charge centroid moves, resulting in back bias-dependent degradation and ambiguous effective electric field [48]. Furthermore, the back-side inversion or back-channel also affects the current and its derivative. That is, the device with thicker front-gate oxide sees more influence of the back gate in its behavior due to the weaker front-gate control. For this device, the front- and back-

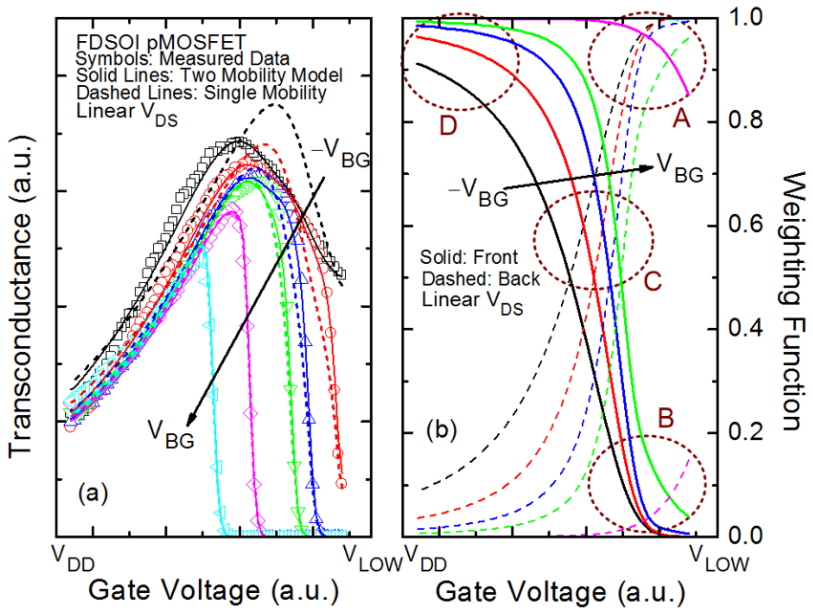


Figure 2.10: (a) Transconductance and (b) extracted weighting function of an FDSOI pMOSFET with thick front-gate oxide at linear drain bias region. The back-gate bias is applied from negative to positive.

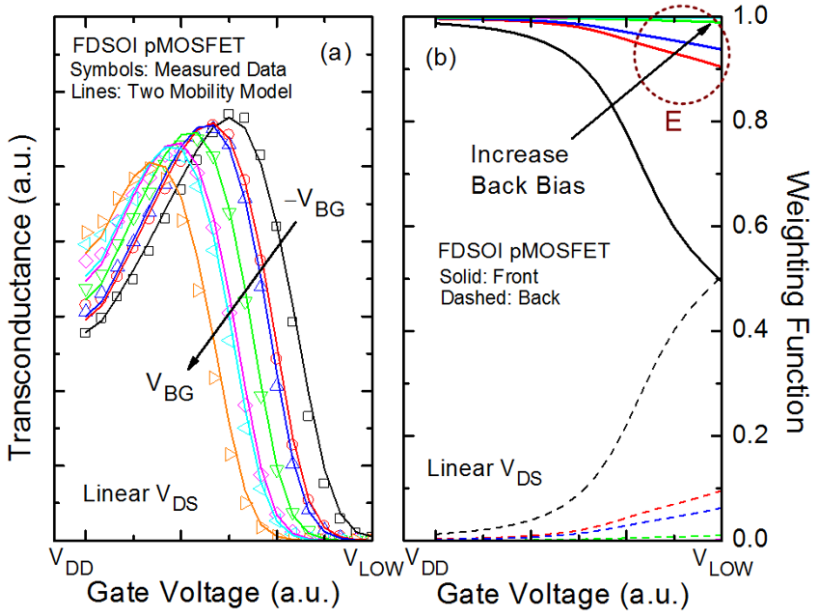


Figure 2.11: (a) Transconductance and (b) extracted weighting function of an FDSOI pMOSFET with thin front-gate oxide at linear drain bias region. The back-gate bias is applied from negative to positive.

side channels co-exist and show different effective mobilities due to distinct qualities of front and back interfaces [58]. Moreover, the inversion charges of the front and back sides experience different electric fields because of the various applied biases scenario. According to above reasons, the mobility model based only on the front surface electrostatics may not be sufficient and a two-mobility model is required. For planar MOSFETs and FinFETs, the mobility model, which includes the surface roughness scattering and Coulombic scattering, has been widely used and is accurate [3, 14]. Thus, the mobility formulas for front- and back-side channels of FDSOI MOSFETs are similar but the parameters are separated.

$$\mu_{\text{eff1(2)}} = \frac{\mu_{1(2)}}{1 + [\text{UA1(2)} + \text{UC1(2)} \cdot V_{BG}] \cdot E_{\text{eff1(2)}}^{\text{EU1(2)}+\text{EUB1(2)}\cdot V_{BG}} + \frac{\text{UD1(2)}+\text{UDB1(2)}\cdot V_{BG}}{(0.5+0.5 q_{ia} C_{\text{ox}})^{\text{UCS1(2)}}}} \tag{2.9}$$

In (2.9), as shown at the top of this page, $\mu_{1(2)}$ is the low-field carrier mobility, UA1(2), UC1(2), EU1(2), EUB1(2), UD1(2), UDB1(2), and UCS1(2) are the model parameters extracted from the experimental data, q_{ia} is the average charge in the channel in unit of volt, and $E_{\text{eff1(2)}}$ is the effective electric field [59].

$$E_{\text{eff1(2)}} = \frac{Q_{\text{dep}} + \eta Q_{\text{inv1(2)}}}{\epsilon_{\text{Si}}} \tag{2.10}$$

where Q_{dep} and $Q_{\text{inv1(2)}}$ are the depletion and inversion charges, ϵ_{Si} is the permittivity of silicon, and η is 1/2 for nMOSFET and 1/3 for pMOSFET. The parameters for the front- and back-side channels are denoted by 1 and 2, respectively. The second term in the denominator of (2.9) represents the surface roughness scattering, and the third term stands for the Coulombic scattering [12, 60].

The total effective mobility is calculated considering the front- and back-side mobilities in a weighted manner as [61, 62, 63]

$$\mu_{\text{total}} = w \cdot \mu_{\text{eff1}} + (1 - w) \cdot \mu_{\text{eff2}} \tag{2.11}$$

The weights w in (2.11) is a dimensionless function of the ratio of charges. The more the amount of charge at an interface, the larger contribution in the effective mobility that interface makes, as shown in Fig. 2.10 (b). The weighting functions in [61] are only valid for the strong inversion, and they are improved in this paper to capture both subthreshold and inversion regions. The amount of inversion charges is an exponential function of surface potential [13] derived from Poissons equation assuming Boltzmann statistics (approximation of FermiDirac statistics). In the subthreshold region, the surface potential varies linearly with the gate bias, so the inversion charge is an exponential function of the gate voltage. In the strong inversion, since the surface potential has weak dependence on the gate bias, the inversion charge shows linear dependence on the gate voltage [64]. Hence, the weighting functions for the front and back sides are

$$w = \frac{e^{(\phi_f - V_{ch})/V_t}}{e^{(\phi_f - V_{ch})/V_t} + e^{(\phi_b - V_{ch})/V_t}} = \frac{e^{\phi_f/V_t}}{e^{\phi_f/V_t} + e^{\phi_b/V_t}} \tag{2.12}$$

$$1 - w = \frac{e^{(\phi_b - V_{ch})/V_t}}{e^{(\phi_f - V_{ch})/V_t} + e^{(\phi_b - V_{ch})/V_t}} = \frac{e^{\phi_b/V_t}}{e^{\phi_f/V_t} + e^{\phi_b/V_t}} \tag{2.13}$$

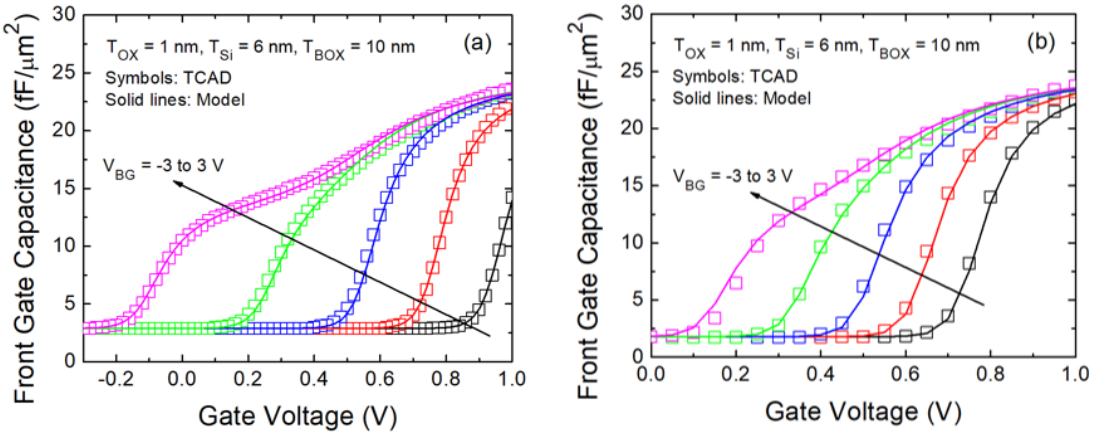


Figure 2.12: The TCAD simulated front-gate capacitances [66] with the BOX thicknesses of (a) 10 and (b) 20 nm. The core model described in [13] for the charges and surface potentials accurately captures the back-side inversion effect.

where ϕ_f and ϕ_b are the surface potentials of the front and back channels, V_{ch} is the quasi-Fermi potential, and $V_t(= k_BT/q)$ is the thermal voltage. A thin front (back) gate oxide increases the surface potential of the front (back) channel, leading to higher weighting factors. If the front-gate oxide is thick enough (about one fifth of the back oxide), the back channel would be apparent in the g_m due to large weighting factor of the back-side as shown in Fig. 10 (b). Finally, the total effective mobility is used into the drain current model [12, 13, 65].

To evaluate the proposed model, the model parameters in (2.9) should be properly extracted. The general guideline of parameters extraction is available in [1]. After extracting the model parameters for charges, threshold voltage, and subthreshold slope from the device $I_{DS}-V_{GS}$ plot, the mobility model parameters at strong inversion bias regime are then extracted as follows. First, the mobility parameters such as EU, UA, UD, and UCS for the front side are extracted for $V_{BG} = 0$. At $V_{BG} = 0$, these parameters do not depend on the back-gate bias. Then, only the back-gate bias-related parameters such as UC, EUB, and UDB for the front side at the back-side accumulation bias condition [negative (positive) V_{BG} for n(p)MOSFET] are extracted. Finally, all the parameters related to the back-side mobility are tuned at the back-side inversion bias condition [large positive (negative) V_{BG} for n(p)MOSFET].

2.3.3 Results and Discussion

The two-mobility model has been incorporated into BSIM-IMG models. Fig. 2.12 shows the comparison of the model and the TCAD simulation data from [66] of the front gate capacitances with the buried oxide (BOX) thicknesses of 10 and 20 nm. The core model described in [13] accurately captures the back-side inversion effect, which causes the first plateau of the gate capacitance at $V_{BG} = 3V$. Based on the accurate charge and surface potential calculations from the core model, weighting functions (2.12) and (2.13) are robust. Fig. 2.9 shows that the proposed model matches well the experimental data of both thick and thin front-gate oxides FDSOI pMOSFETs. Fig. 2.10 shows the transconductance and extracted weighting function from experimental

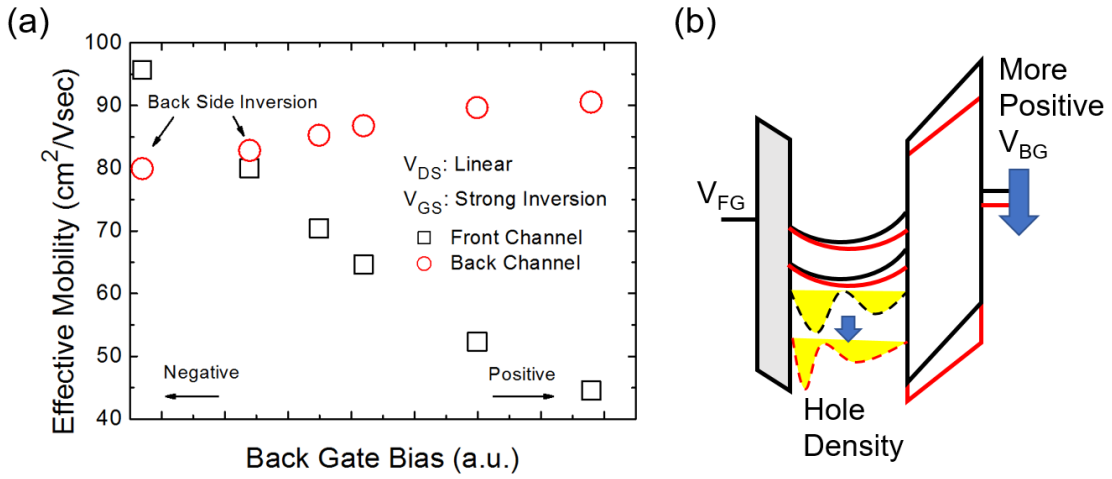


Figure 2.13: (a) Extracted effective mobility of the front and back channels from Fig. 2.9 (a). The back-side inversion occurs when the back-gate bias is negative in FDSOI pMOSFET. (b) Illustration of hole density distribution due to applied V_{BG}.

data for an FDSOI pMOSFET with thick front-gate oxide. At high $|V_{BG}|$ and low $|V_{GS}|$ (subthreshold region), i.e., the bias condition in which the back-side inversion occurs, the channel charge (hole) is dominated by the back-side channel charges and the weighting function of the back side is larger than that of the front side, as shown by circles A and B in Fig. 2.10 (b). As the front-gate voltage increases, the weighting factor for the front side increases and the two factors cross, as shown by circle C in Fig. 2.10 (b). Ultimately, the front weighting factor becomes much larger than the back one, indicating the domination of the front charge and mobility, as shown by circle D in Fig. 2.10 (b). The crossing point happens near the onset of the surface roughness scattering. In other words, the front-channel charge concentration is high enough and the front-gate voltage attracts the inversion holes to the body/oxide interface, leading to scattering. For a thin front-gate oxide devices shown in Fig. 2.11, a different behavior of weighting function is found. The transconductance shows monotonic behavior with the back-gate bias, just like threshold voltage shift. This is due to the larger front-gate capacitance and charge in thin front-gate oxide device. Even at high $|V_{BG}|$, the front-channel charge dominates the effective mobility for this device, as shown by circle E in Fig. 2.11 (b). It is worth noting that the different device behavior due to the front-gate oxide thickness change is captured physically in our model.

Fig. 2.13 (a) shows the extracted effective mobility of the front and back channels as functions of the back-gate bias. The back-channel mobility always exceeds the front-channel mobility [58] except at high $|V_{BG}|$. In addition, the dependence of the front-channel mobility on V_{BG} is opposite to that of the back-channel mobility, i.e., opposite sign of slope of mobilityV_{BG} plot. This is because more $|V_{BG}|$ attracts the front-channel holes toward the body but moves the back-channel holes closer to body/BOX interface, resulting in reduction of scattering for the front side but mobility degradation for the back-side [see Fig. 2.13 (b)]. This effect is captured by model parameters UC and EUB. Furthermore, the front and back mobilities in Fig. 2.13 (a) show different magnitude of slopes, indicating that the interfacial quality of the front and back surfaces is different for the surface roughness scattering [58]. In Fig. 2.14, the scalability of the BSIM-

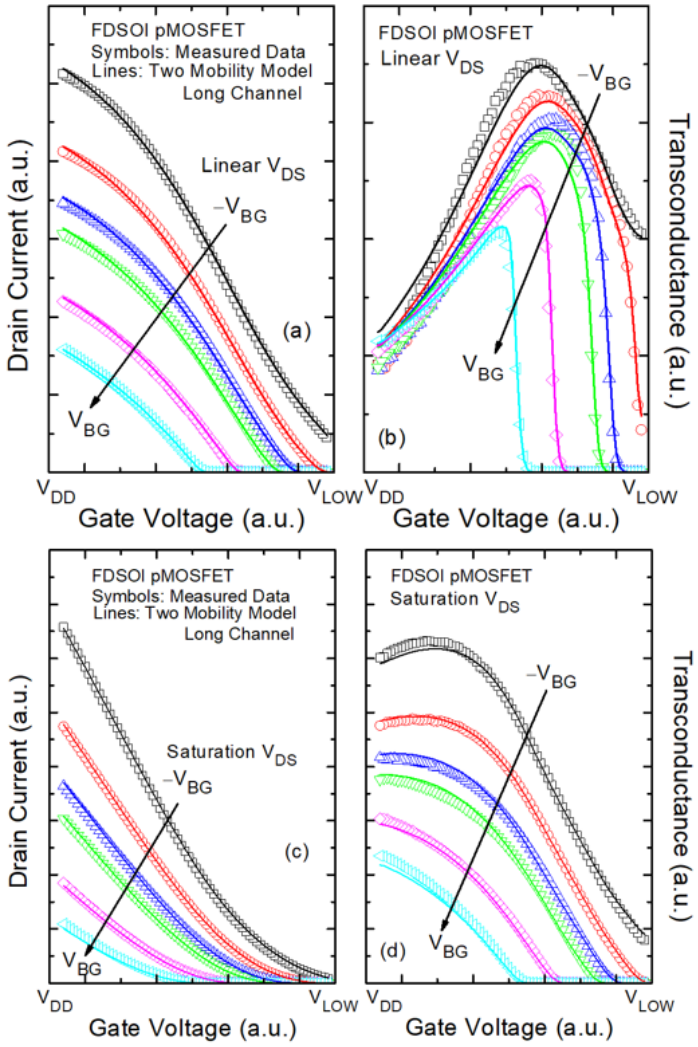


Figure 2.14: $I_{DS} - V_{GS}$ at (a) linear and (c) saturation drain biases, and transconductance at (b) linear and (d) saturation drain biases of a long channel FDSOI pMOSFET with the same body and BOX thicknesses as the device shown in Fig. 2.10.

IMG model is demonstrated with another longer channel FDSOI pMOSFET with thick front-gate oxide using similar mobility parameters as the shorter channel device in Fig. 2.10. The nonmonotonic V_{BG}-dependence on the transconductance at the linear drain bias is observed [see Fig. 2.14 (a) and (b)], indicating that the impact of the longitudinal electric field is less crucial to the mobility than the transverse one. The mobility is mainly influenced by the electric field from the front-gate and the back-gate. Moreover, the transconductance at the saturation drain bias is also accurately captured by the proposed model [see Fig. 2.14 (c) and (d)], showing the model capability for the pinch off and the channel length modulation effects.

To examine the model for a different doping polarity, a long-channel FDSOI nMOSFET with the thick front-gate oxide is well fit with the developed model as shown in Fig. 2.15. Due to the smaller effective mass of electron, the charge centroid is closer to the back surface and thus the surface roughness scattering becomes more significant than that of hole [67]. Hence, the peak of transconductance of FDSOI nMOSFET with the thick front-gate oxide at the back-side inversion bias is not as

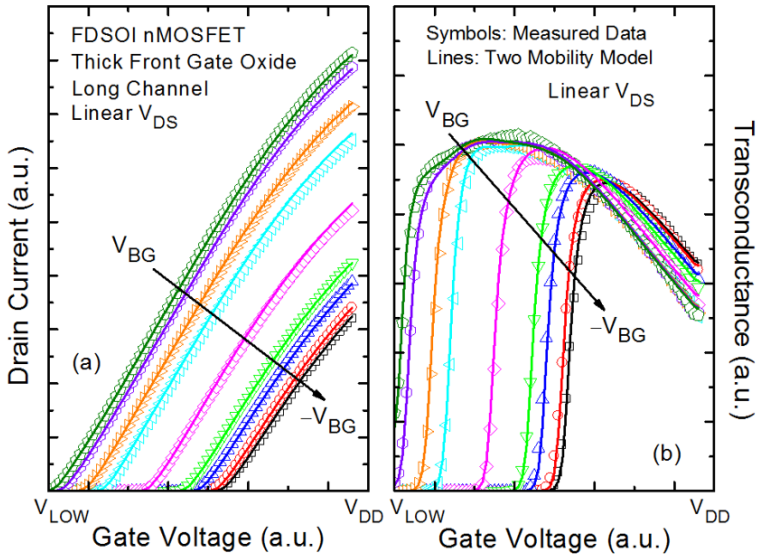


Figure 2.15: (a) $I_{DS} - V_{GS}$ and (b) transconductance of a long-channel FDSOI nMOSFET with thick front-gate oxide and similar device structure as the pMOSFET shown in Fig. 2.14.

apparent as FDSOI pMOSFET, as shown in Fig. 2.15 (b). Fig. 2.16 shows measured data from the Laboratoire dlectronique et de technologie de linformation (LETI) device [59] and model results of an FDSOI nMOSFET. Different technologies from the other previously shown devices are adopted in Fig. 2.16, indicating that the proposed model is not only physical but also flexible for the technology variation. To further validate the developed model for the device scaling, a short-channel ($L = 20$ nm) FDSOI nMOSFET with $T_{\mathrm{BODY}} = 4$ nm is simulated using Sentaurus TCAD with the density gradient quantization model and thin-layer mobility model with the Philips unified mobility model [2], as shown in Fig. 17. The proposed model is in good agreement with the simulated data, showing great model capability of device scaling with even thinner body where the charge ratio of front and back sides are correctly captured by the core model.

2.4 Non-Monotonic DIBL

2.4.1 Device Characterization

Non-Monotonic Threshold Voltage Degradation

Sentaurus technology computer-aided design (TCAD) is utilized to study the FDSOI MOSFETs in this section, where the physical model includes the density gradient quantization model and thin-layer mobility model with Philips unified mobility model [2]. Since the TCAD is able to accurately solve the electrostatics in the subthreshold region where basically only involves the Poissons equation, the threshold voltage of the experimental FDSOI MOSFET can be reproduced with similar structure adopted in our simulation. The non-monotonic threshold voltage degradation is first found in FDSOI devices with thick front gate oxide (6 nm SiO_2) as shown in Fig. 2.18. The threshold voltage shift is defined as $V_{Tsat} - V_{Tlin}$, where V_{Tsat} and V_{Tlin} are the thresh-

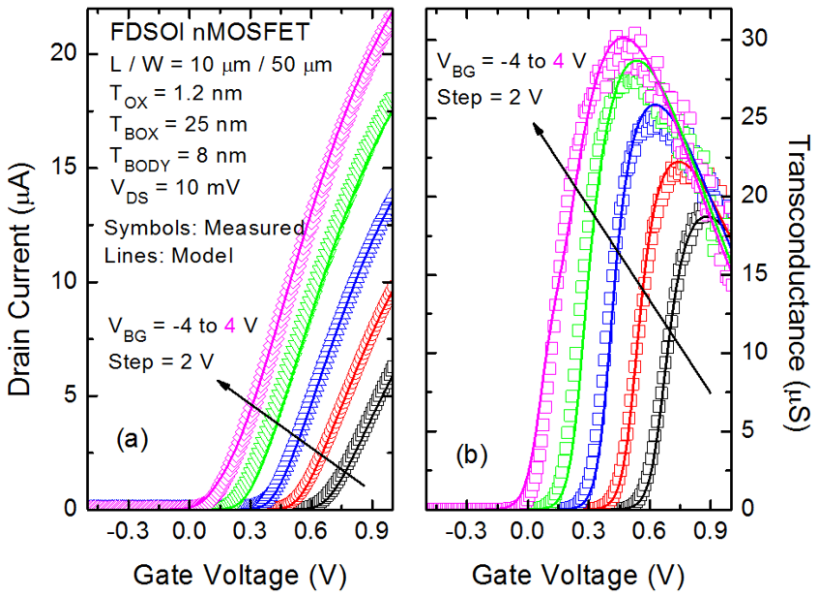


Figure 2.16: (a) $I_{DS}-V_{GS}$ and (b) transconductance of an FDSOI nMOSFET with $L = 10$ μm, $W = 50$ μm, $T_{\mathrm{OX}} = 1.2$ nm, $T_{\mathrm{BOX}} = 25$ nm, and $T_{\mathrm{BODY}} = 8$ nm. Measurement data are from the LETI device [59].

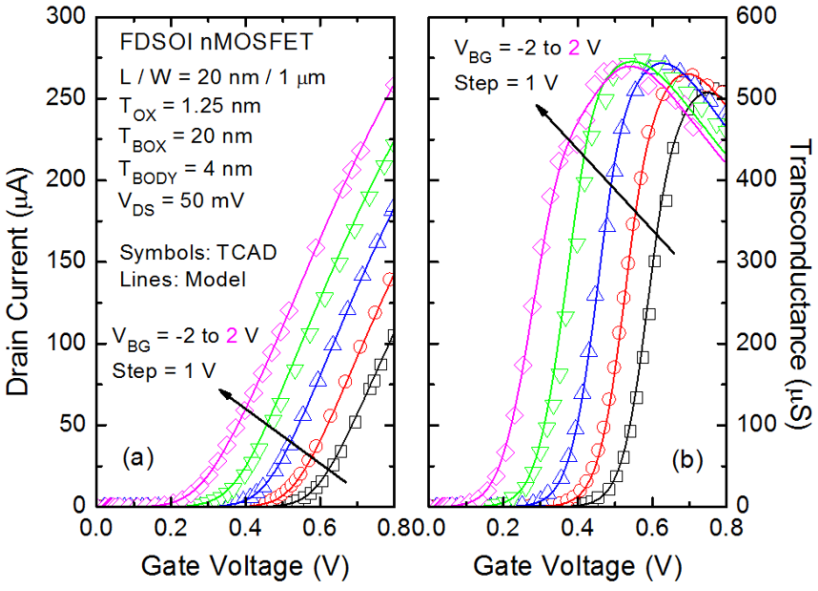


Figure 2.17: TCAD simulated (a) $I_{DS}-V_{GS}$ and (b) transconductance of an FDSOI nMOSFET with $L = 20$ nm, $W = 1$ μm, $T_{\mathrm{OX}} = 1.25$ nm, $T_{\mathrm{BOX}} = 20$ nm, and $T_{\mathrm{BODY}} = 4$ nm at various V_{BG}.

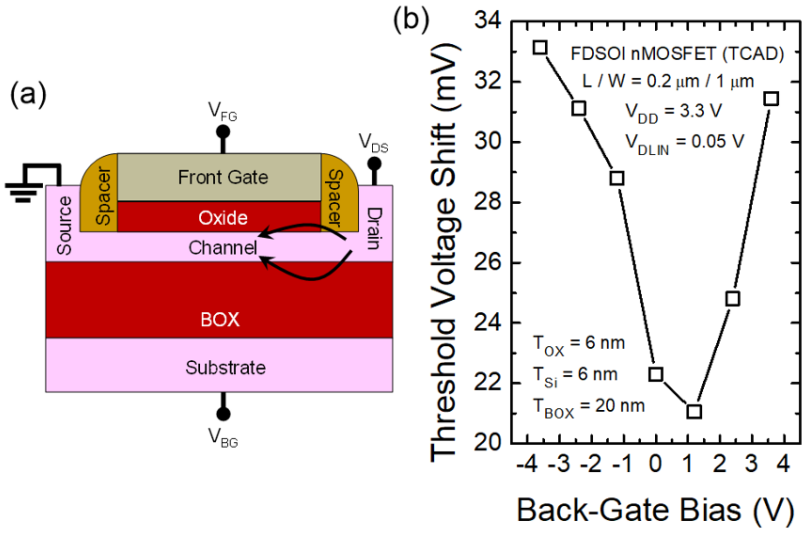


Figure 2.18: (a) TCAD simulated structure of a FDSOI MOSFET with thick front gate oxide. The arrows represent the fringing fields through the BOX and front gate oxide. (b) The threshold voltage shift versus back-gate bias plot.

old voltage at $V_{DS} = V_{DD}(= 3.3V)$ and $V_{Dlin}(= 50mV)$ respectively. The threshold voltage (V_{TH}) is determined by the constant current method (300nA $\times W/L$ for nMOSFET; 70nA $\times$ W/L for pMOSFET) [68]. However, the monotonic threshold voltage shift (here defined as DIBL) has been reported in literature [69, 70, 71]. The monotonic DIBL can be understood by the concept of the charge centroid in the channel. For nMOSFET, the charge centroid moves toward the front surface as decreasing the back-gate bias (V_{BG}), so the front gate control becomes better and thus DIBL is reduced. Nevertheless, the non-monotonic threshold voltage shift (Fig. 2.18) appears at devices with few hundred nanometers gate lengths. It is expected that the DIBL will not happen in such long channel devices. Fig. 2.19 shows the conduction band energy of the same device as shown in Fig. 2.18 at various drain biases. The barrier height of the channel is not lowered much (less than $1meV$) by the drain biases (inset of Fig. 2.19), implying that the non-monotonic threshold voltage shift (degradation) is not caused by the DIBL. This is also the reason why the term threshold voltage shift is used instead of DIBL. Therefore, more detailed physics should be taken into account to explain the non-monotonic threshold voltage shift.

Channel Length Modulation

One of the benefits of the FDSOI device is the mitigated random dopant fluctuation due to the lightly doped channel, which further boosts the carrier mobility. However, although the DIBL is negligible because of good gate control, the channel length modulation (CLM) which even can be observed in long channel device is important since the channel doping is quite low (nearly undoped). As shown Fig. 2.20, it is observed that the modulated channel length (ΔL) is $\approx$ 65 nm at $V_{DS} = 3.3V$, which is defined at the point where the conduction band energies of $V_{DS} = 50mV$ and $3.3V$ cross. Due to shorter channel length, the drain current is enhanced, which can be manifested in terms of the threshold voltage shift. The amount of the threshold

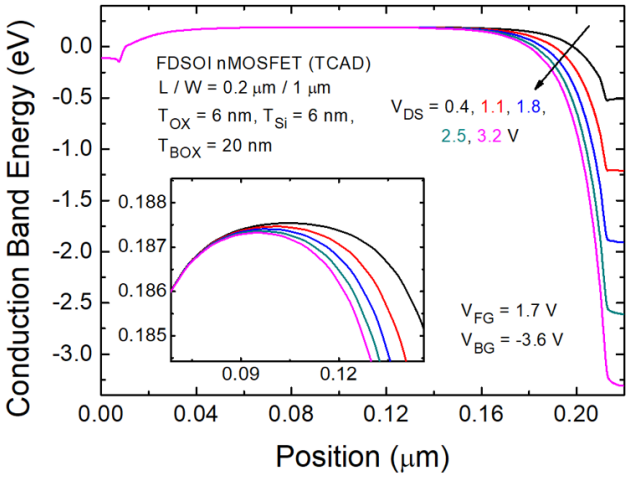


Figure 2.19: Simulated conduction band energy profile along source-channel-drain direction, which is cut at 0.5 nm below the front surface. The front gate is biased at $\approx 0.25V$ below the threshold voltage. The inset shows barrier height difference for various V_{DS} is less than $1meV$.

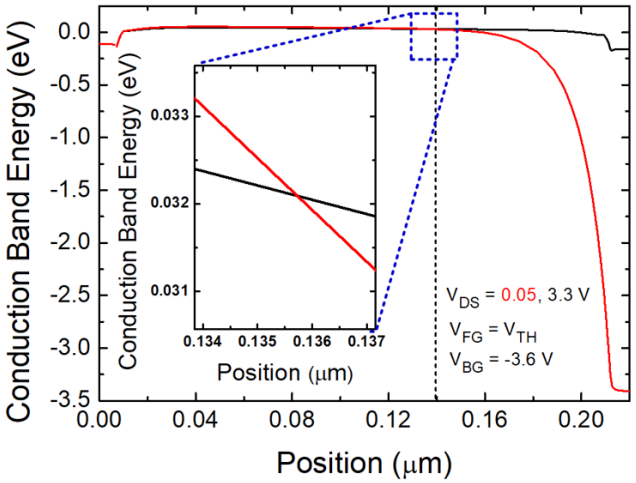


Figure 2.20: Simulated conduction band energy profile in linear and saturation drain bias regions at threshold voltage, where is at the maximum current flowing path along the channel. The dashed line is the boundary of the ΔL. The inset shows the crossing of the conduction band. The device dimensions are the same as shown in Fig. 2.19.

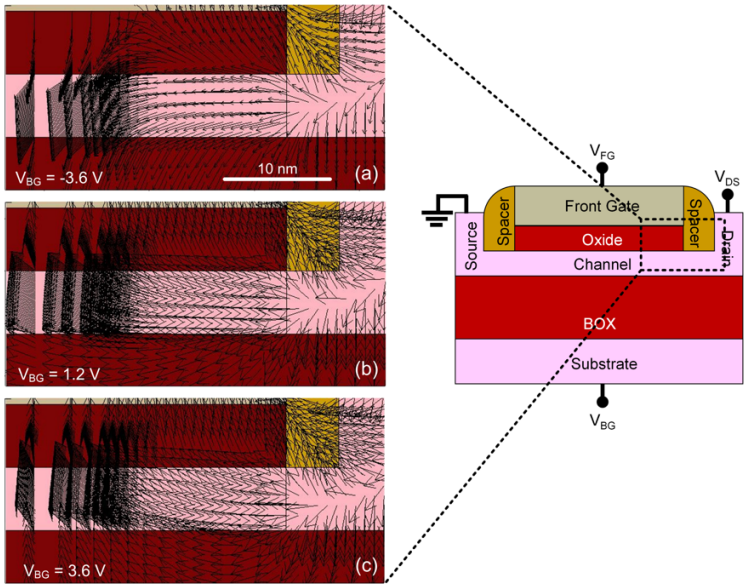


Figure 2.21: Electric field lines at $V_{DS} = 3.3V$, $V_{FG} = V_{TH}$, and V_{BG} = (a) -3.6, (b) 1.2, and (c) $3.6V$. The device dimensions are the same as shown in Fig. 2.19.

voltage shift can be calculated as follows. The subthreshold slope (SS) at the threshold voltage is extracted from the drain current versus front gate voltage ($I_{DS} - V_{FG}$) plot (see Fig. 2.22). At $V_{BG} = -3.6V$, the SS is about 183.8 mV/dec at $V_{FG} = V_{TH}$, which leads to the SS degradation factor (n) of 3.064. From CLM, the drain current is boosted due to the drain bias by $L/(L - \Delta L) = 1.481$. Because the current is proportional to the charge density, one can write

$$\frac{I_{DSsat}}{I_{DSlin}} = \exp\left(\frac{\Delta V_{TH}}{nV_t}\right) = \frac{L}{L - \Delta L} \tag{2.14}$$

where V_t is the thermal voltage ($= 26mV$ at room temperature). From equation (2.14), ΔV_{TH} is $31.31mV$ at $V_{BG} = -3.6V$, which is reasonable as compared to the value shown in Fig. 2.18(b). The same analysis is applied for the cases of $V_{BG} = 1.2V$ and $3.6V$ (SS = 139.5 mV/dec, 175 mV/dec), which gives the threshold voltage shifts of $23.76mV$ and $29.81mV$, respectively. Therefore, the threshold voltage degradation in such thick oxide device can be attributed to the CLM.

To explain the root cause that the SS is back-gate bias-dependent, the fringing field through the front gate oxide should be taken into account. Due to thick oxide, the fringing field is enhanced, which is similar to the fringing field through the thick BOX [72, 73, 74]. However, the back-gate bias dependences on the fringing fields through the front gate oxide and BOX are opposite. Fig. 2.21 shows the electric field lines distribution in the device at various V_{BG}. For the front side fringing field, more negative V_{BG} attracts the part of the field (originally from the drain to the gate) to point to the back side, giving rise to the front side fringing field. In this scenario, the part of the field from the drain can directly terminate in the substrate, so the fringing field through the BOX is minimized. As increasing V_{BG}, the field to the substrate is repelled by V_{BG}, causing the fringing field through the BOX and field directly from the drain to the gate. With these fringing field (capacitance), the SS is degraded, and thus is back-gate bias-dependent.

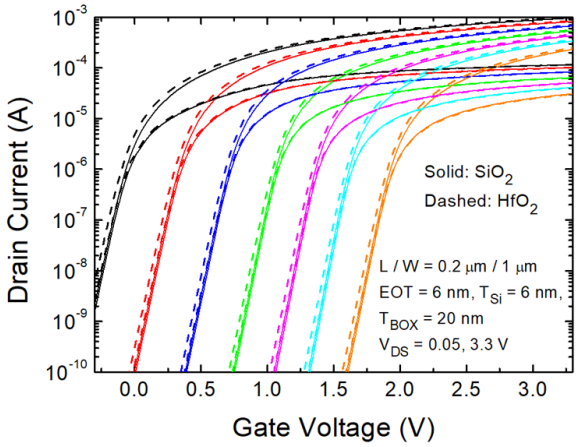


Figure 2.22: $I_{DS} - V_{FG}$ characteristics of the devices with SiO_2 and HfO_2 as the front gate oxide of EOT = 6 nm. The V_{BG} is biased from $-3.6V$ (right) to $3.6V$ (left) with the step of $1.2V$.

2.4.2 Discussions

Although it has been proved that the CLM could cause a threshold voltage shift of few tens mV, how it varies with the device structure should be carefully analyzed to generalize the developed theory.

Front Gate Oxide Configurations

For diverse applications, the gate oxide configurations would be adopted to achieve the device performance requirement. As mentioned earlier, for high voltage [46] or image sensor applications [43, 44, 45], a thick EOT is used. Furthermore, the high-κ gate stack has been widely used in low power applications because thicker physical oxide thickness (but EOT keeps the same) dramatically reduces the gate tunneling current [75]. Therefore, revisiting the impact of the CLM on various front gate oxide scenarios is important.

Fig. 2.22 and Fig. 2.23 show the $I_{DS} - V_{FG}$ plot and the threshold voltage shift [extracted from TCAD and calculated from (2.14) of devices with the same EOT (= 6 nm) of SiO_2 and HfO_2. The same EOT means that the vertical electric field is the same for two devices. However, it can be observed that the device with HfO_2 has amplified threshold voltage shift because the greater physical oxide thickness of HfO_2 leads to more fringing field and thus changes the modulated channel length. From TCAD simulation, at $V_{FG} = V_{TH}$ the modulated channel length is about 95 nm and the SS are 193 mV/dec and 138.8 mV/dec at $V_{BG} = -3.6V$ and $1.2V$, which gives $\Delta V_{TH} = 53.89mV$ and $38.76mV$ using (2.14). The obtained results are reasonable for the threshold voltage shift shown in Fig. 2.23. In Fig. 2.22, the linear drain currents of two devices are approximately on the top of each other, while the saturation currents are clearly separated. This implies that the drain coupling through the front side fringing capacitance is greater in the HfO_2 device.

Fig. 2.23 also shows the threshold voltage shift of devices with various SiO_2 thicknesses of 6 and 10 nm and a high-κ gate stack (HfO_2 on the top of SiO_2) of EOT = 1.3 nm. As mentioned earlier, the thicker gate oxide (EOT = 10 nm) induces more

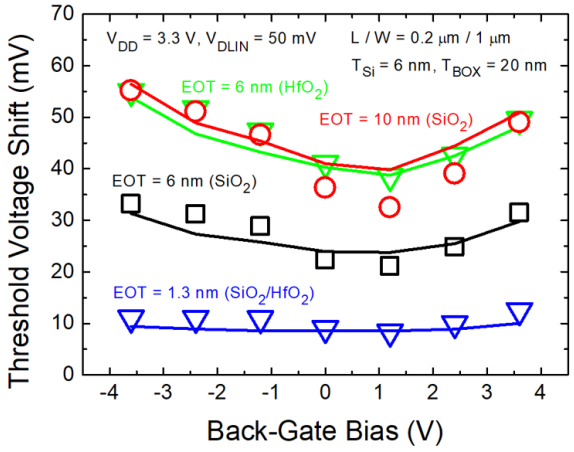


Figure 2.23: Threshold voltage shift of the devices with various front gate oxide configurations. [Symbols: extracted from TCAD; lines: calculated from (2.14)]

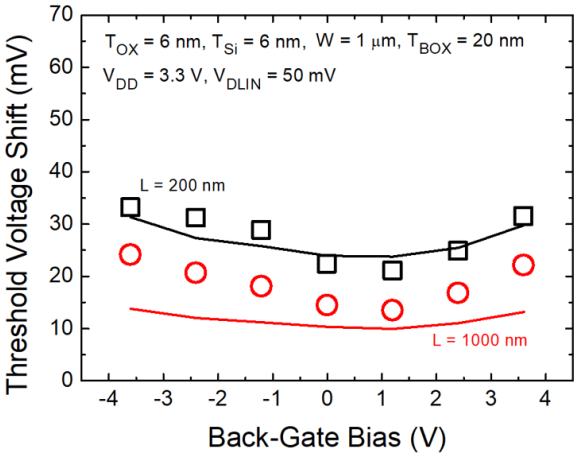


Figure 2.24: Threshold voltage shift of the devices with $L = 200$ and 1000 nm. The front gate oxide is 6 nm SiO_2.

front side fringing field, which degrades the threshold voltage and cause the back-gate bias-dependence. As reducing the EOT, the vertical electric field becomes stronger (better gate control), the modulated channel length is reduced to $\approx$ 40 nm, which gives $9.45mV$ threshold voltage shift at $V_{BG} = -3.6V$ according to (2.14). The back-gate bias-dependence is still present in Fig. 2.23 due to the fringing field, but the magnitude is significantly suppressed which may not be seen in the experimental data because of the thermal variation.

Gate Length Scaling

Although the proposed model can quantitatively explain the back-gate bias dependent threshold voltage shift of various oxide configurations based on the CLM, the gate length effect on the threshold should be explored because of the short channel effect on DIBL reported in [71]. As shown in Fig. 2.24, the threshold voltage shift is mitigated by enlarging the gate length. The same calculation as described in previous section can be applied. The modulated channel length ΔL is 160 nm extracted from

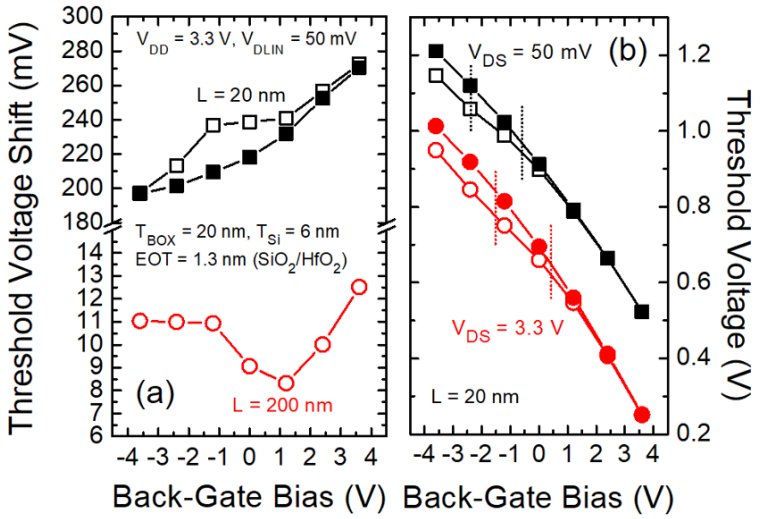


Figure 2.25: (a) Threshold voltage shift of devices with EOT of 1.3 nm and $L = 20$ and 200 nm. (b) Threshold voltage of $L = 20$ nm device at $V_{DS} = 50mV$ and $3.3V$. The dashed lines represent the regions where the substrate depletion happens in moderately doped substrate. (Solid symbols: heavily doped substrate 6.5×10^{18} cm^{-3}; Open symbols: moderately doped substrate 6.5×10^{17} cm^{-3})

the simulated results. The ΔL is longer in 1 μm device than in 200 nm device since 200 nm channel and heavily doped drain composes a short diode. The ΔL in 200 nm device is limited by the source potential. The SS of 1 μm device at $V_{FG} = V_{TH}$ and $V_{BG} = -3.6$, 1.2, and $3.6V$ are given by 182.77, 131.5, 174.31 mV/dec and from $I_{DS} - V_{FG}$ plot. The corresponding threshold voltage shifts are 13.8, 9.94, $13.17mV$. The discrepancy between the proposed model and the TCAD results may come from the non-uniform spatial charge (current) distribution. Notice that the ΔL is extracted along the line cut with the highest current. The other current paths are also contributing to the total current and thus complicates the ΔL. Therefore, the CLM results in the reasonable threshold voltage shift and is considered as the root cause of the threshold voltage degradation from the gate length of few hundreds nm to μm.

Since the short channel effect and DIBL become stronger as reducing the gate length, the thin EOT (high-κ gate stack) should be adopted to boost gate control. Fig. 2.25 (a) shows the threshold voltage shift of the devices with 1.3 nm EOT at $L = 20$ and 200 nm. As shortening the gate length, the threshold voltage becomes monotonic with V_{BG}. The electric field from the drain to the source in the short channel device can directly penetrate the channel. The threshold voltage shift now is dominated by the DIBL [71]. The back-gate bias effect comes into picture through the charge centroid. That is, for nMOSFET more positive (negative) V_{BG} move the charge centroid toward the back (front) side, degrading (improving) the front gate control. Interestingly, at $-1V < V_{BG} < 1V$ the threshold voltage shift tends to be flat in $L = 20$ nm device. This can be understood by the substrate depletion effect [51]. At linear drain bias ($V_{DS} = 50mV$), the substrate depletion is uniformly present in the substrate at around $-2.5V < V_{BG} < -0.5V$ as shown in in Fig. 2.25 (b) and Fig. 2.26. Nevertheless, at saturation drain bias ($V_{DS} = 3.3V$), the substrate depletion happens at $-1.5V < V_{BG} < 0.5V$ because the high drain bias induces more depletion

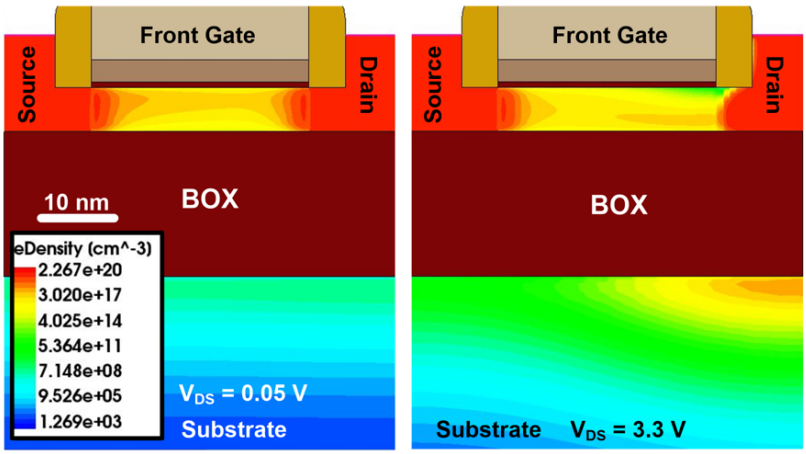


Figure 2.26: Electron density distribution for $L = 20$ nm with EOT $= 1.3$ nm device. The back-gate bias for both cases is $-1.2V$. At $V_{DS} = 3.3V$, the region under the drain is inverted and affects the center of channel, suggesting that the substrate depletion happens at more positive V_{BG} compared to that at $V_{DS} = 50mV$.

charges and inversion charges which help the back-gate to invert the electrical polarity at the substrate/BOX interface near the drain side. This effect is only obvious in short channel device since the drain bias is able to push the inversion layer in the substrate under the drain side toward the center of channel (see Fig. 2.26). Once the substrate depletion emerges, the threshold voltage becomes less sensitive to the V_{BG}. When decreasing V_{BG}, the substrate depletion at saturation region happens earlier (the DIBL would be degraded). However, due to the charge centroid effect, the DIBL is eased by the more negative V_{BG}, which compensates the substrate depletion effect and thus the DIBL becomes flat. In contrast, the substrate depletion at linear drain bias happens later but the DIBL gets better because the threshold voltage does not increase much with V_{BG}. The substrate depletion effect is reduced by the heavily doped substrate [51]. It is evident that in Fig. 2.25 the DIBL does not have a flat region with V_{BG} when the substrate doping is 6.5×10^{18} cm^{-3}. The reported DIBL data from [76] thus can be explained by the mentioned V_{DS}-induced non-uniform substrate depletion effect.

2.5 Back Gate Bias Dependent Overlap Capacitance Model

2.5.1 Model Description

The space charge distribution of FDSOI MOSFET is firstly studied using TCAD simulation [2], and then a simple analytical model is developed. The model is verified by comparing it against TCAD simulations and experimental data. The structure and doping profile of FDSOI nMOSFET simulated in TCAD are shown in Fig. 2.27 (a). The gate length (L_G), front gate oxide thickness (T_{OX}), back gate oxide thickness (T_{BOX}), body thickness (T_{Si}), and source/drain extension (SDE) length, are 26, 1.25, 20, 6, and 3 nm, respectively. The physical models used in the TCAD simulations

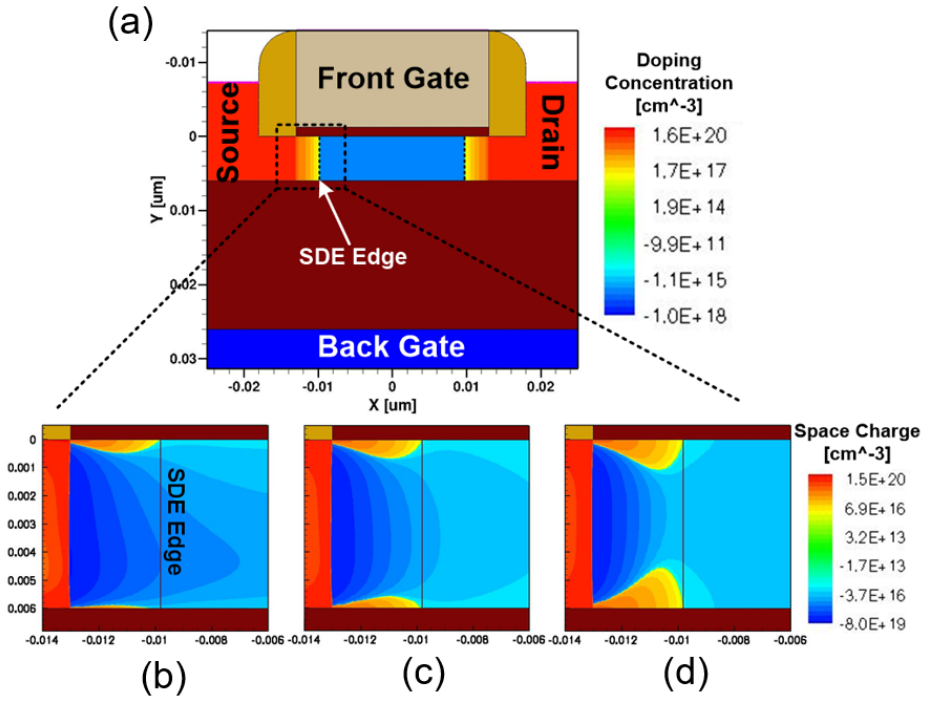


Figure 2.27: (a) Doping concentration profile in the simulated FDSOI MOSFET with front and back gate oxide thicknesses of 1.25 and 20 nm, respectively. The space charges in the source/drain extension (SDE) regions at (b) $V_{BG} = 2V$, (c) $V_{BG} = 0V$, and (d) $V_{BG} = -2V$, are with $V_{FG} = -0.1V$ (OFF state) and $V_{DS} = 0V$.

include doping-dependent mobility with high-field saturation and degradation, SRH, and density gradient quantization model for electrons and holes (quantum potential) [2]. The default values parameters in these models are used for simulations. Fig. 2.27 (b)−(d) show the space charges in the source extension overlap region at three different back gate biases (V_{BG}) of 2, 0, and $-2V$ at the OFF state ($V_{FG} = -0.1V$). Due to the electric field from the back gate, the region of space charge of front gate (namely, depletion region) spreads further by increasing negative back gate voltage even with the same V_{FG}, which means that the overlap capacitance will decrease with reducing V_{BG}. Note that the overlap region is similar to a pMOSFET due to source/drains opposite doping polarity to bodys, so a negative applied back gate voltage can deplete the overlap region while forms an accumulation layer in the body region. At the OFF state bias condition, the front gate also depletes the overlap region. Thus, the depletion charge in the overlap region is dependent on both front and back gate biases. An exact solution requires two-dimensional device electrostatics to be solved. The complex solution of 2-D electrostatics is less desirable for compact modeling purpose where a simplified solution is preferred. From BSIM4 [11], the amount of charges in the overlap region on the front gate Q_{OV} is written as

$$\begin{aligned} Q_{OV} &= W \cdot L_{OV} \cdot C_{\mathrm{OX1}} \cdot V_{FG} + W \cdot C_{OV} \cdot q \cdot N_{SDE} \cdot W_d / C_{\mathrm{OX1}} \\ &= W \cdot L_{OV} \cdot C_{\mathrm{OX1}} \cdot V_{FG} + W \cdot C_{OX} \\ &\times \left[V_{FG} - V_{FBSD} - V_{OV} - \frac{1}{2} C_K \left(\sqrt{1 - \frac{4V_{OV}}{C_K}} - 1 \right) \right] \end{aligned} \tag{2.15}$$

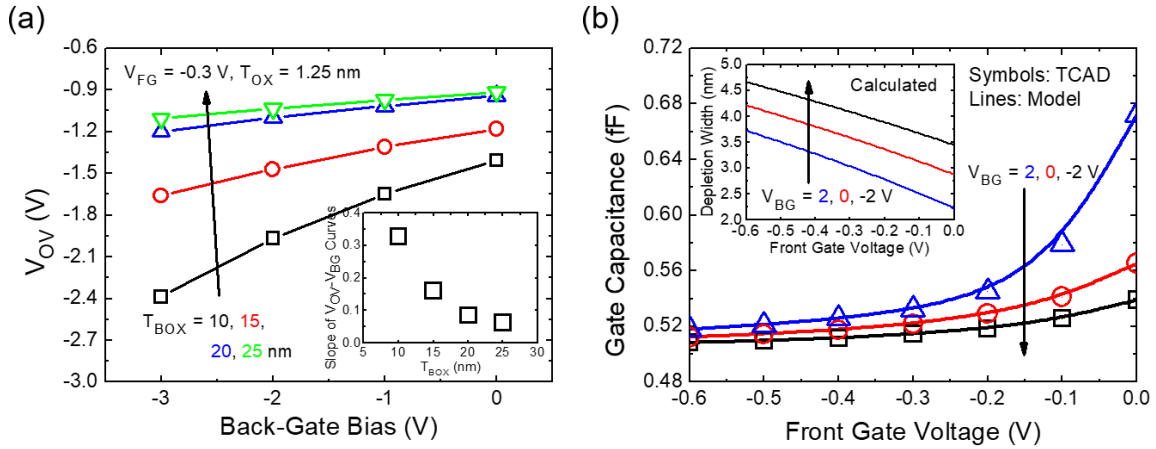


Figure 2.28: (a) Extracted V_{OV} with various T_{BOX} from the simulated capacitances. The inset shows the slope of $V_{OV} - V_{BG}$ curves. (b) Comparison of TCAD data (symbols) and BSIM-IMG model (lines) with equation (2.16). The inset shows the calculated depletion width.

where W is the device width, L_{OV} is the overlap length for the highly doped SDE region, C_{OX1} is the front gate oxide capacitance per unit area, C_{OV} is the model parameter for total bias-independent overlap capacitance per unit width, W_d is the depletion width, C_K is the model parameter of overlap capacitance in unit of volt, and V_{OV} is the effective potential in the overlap region. Hence, the (front) gate capacitance at the OFF state C_{OFF} can be approximately calculated by taking derivative of Q_{OV} with respect to V_{FG}

$$\begin{aligned} C_{OFF} &\approx \frac{\partial Q_{Fringing} + \partial Q_{OV}}{\partial V_{FG}} \\ &= W \cdot C_F + W \cdot L_{OV} \cdot C_{\mathrm{OX1}} + \frac{W \cdot C_{OV}}{\sqrt{1 - \frac{4V_{OV}}{C_K}}} \end{aligned} \tag{2.16}$$

The first term of equation (2.16) is the bias-independent fringing capacitance with the model parameter C_F in unit of F/m. The second term results from the highly doped SDE region, which is in parallel with the capacitance of the lightly doped region (the third term). Fig. 2.28 (a) shows the extracted V_{OV} versus V_{BG} curves from the simulated capacitances for different T_{BOX} by using equation (2.16) and the parameters used in Fig. 2.28 (b). Fig. 2.28 (a) demonstrates that V_{OV} is approximately linearly dependent on V_{BG} and can be written as

$$V_{OV} = (V_{FG} - V_{FBSD} - V_{\mathrm{ox}}) - \mathrm{PCOVBS1} \cdot \gamma_0 \cdot (V_{BG} - V_{FBSDBG} - \mathrm{PCOVBS0}) \tag{2.17}$$

where PCOVBS0 and PCOVBS1 are the model parameters considering the non-uniformity of the doping profile and the electric field in the overlap region, and γ_0 is defined in (2.4). The first term of equation (2.17) is already used for V_{FG}-dependency in BSIM4 [11]. The second term of equation (2.17) stands for the coupling of the back gate to the front gate. As shown in the inset of Fig. 2.28 (a), the slope of $V_{OV} - V_{BG}$ curve changes with T_{BOX}. For thicker T_{BOX}, the back gate control is much smaller against the front gate control, resulting in smaller V_{BG}-dependency. Note that V_{OV}

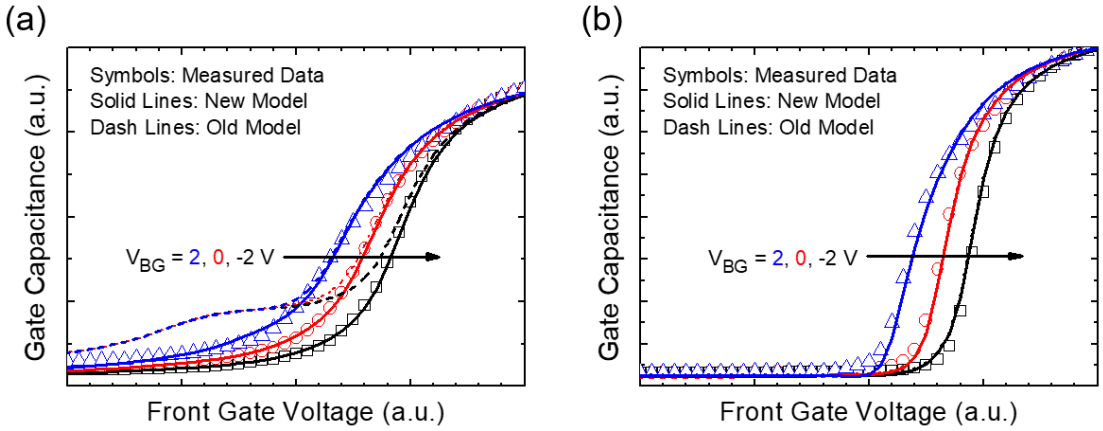


Figure 2.29: Comparison of measured data and BSIM-IMG model for FDSOI MOSFETs with (a) short and (b) long channels. The solid and dash lines are for new and old models. There is no V_{BG}-dependency in gate capacitance in the old model.

will be clamped by

$$V_{OV} = \frac{1}{2}\left[(V_{OV} + \delta) - \sqrt{(V_{OV} + \delta)^2 + 4\delta}\right] \tag{2.18}$$

where δ is 0.02 V. Equation (2.18) ensures the function continuity when V_{OV} goes from negative to positive values, since a positive V_{OV} leads to the accumulation in the overlap region instead of the depletion. Fig. 2.28 (b) shows a good fitting result with the TCAD data by incorporating equation (2.17) into BSIM-IMG model. The extracted parameters of C_F, L_{OV}, C_{OV}, C_K, V_{ox}, PCOVBS0, and PCOVBS1, are 200 pF/m, 8.8 nm, 75 pF/m, 0.488 V, 0.1 V, 1.379 V, and 2.091, respectively. W_d is calculated by using equation (2.15), where N_{SDE} is estimated to be 2.5×10^{19} cm^{-3}, as shown in the inset of Fig. 2.28 (b). W_d and L_{OV} are around few nm, which are expected from the simulated device structure.

2.5.2 Results and Discussion

Fig. 2.29 (a) and (b) show the measured data and BSIM-IMG model simulation results of the gate capacitances as functions of gate voltage for short and long channel FDSOI MOSFET. The measured data are from the devices fabricated by an industry lab. In Fig. 2.29 (a), the gate capacitance shows a strong function of the back gate bias at the OFF state, which was not captured by the previous model with only front gate bias dependency. With modified V_{OV} model, the back gate bias-dependent behavior of the gate capacitance at the OFF state can be described because the overlap length is a great portion of the gate length in the short channel device. In contrast, the long channel device does not exhibit the back gate bias dependency since the overlap length is negligible compared with the gate length as shown in Fig. 2.29 (b). Fig. 2.30 shows the delays per stage of an 17-stage ring oscillator (RO) at different back gate biases with and without the developed overlap capacitance model. The delay of each inverter can be obtained by using the RO frequency, which is given by $1/(2 \times 17 \times \text{inverter delay})$. The delay error for each stage could be up to 12.5 % if the back gate bias-dependent overlap capacitance is not considered. The nFET and pFET in an inverter of the RO

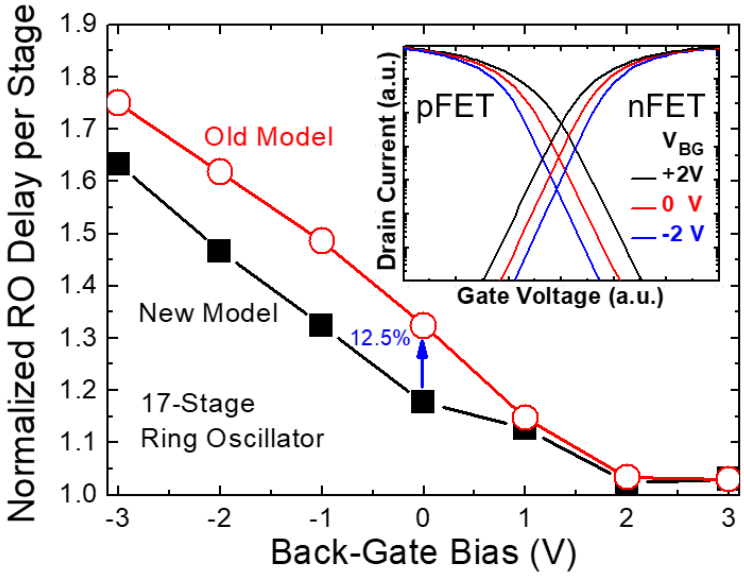


Figure 2.30: Delays per stage of an 17-stage ring oscillator (RO) at various back gate biases with new and old back gate bias-dependent overlap capacitance. The CMOS inverter consists of the short channel FDSOI MOSFETs with the same parameters used Fig. 2.29 (a).

have the symmetric drain currents by tuning the gate workfunction of pFET (see the inset of Fig. 2.30). At the same bias condition, the drain currents are the same with and without the proposed overlap capacitance. Thus, the discrepancy of the delays between new and old models results from the capacitance. Interestingly, when V_{BG} is greater than 1 V, the delays with and without the proposed model are approximately the same. This is because a positive V_{BG} reduces the threshold voltage (the curve is shifted to the left (right) for n(p)FET), resulting that the nFET and pFET are always operated in the strong inversion where the overlap capacitance is insignificant. Furthermore, the delay increases with reducing V_{BG} because of higher threshold voltage and lower drive current.

2.6 Conclusion

Back-gate bias-dependent effects, including GIDL, gate tunneling leakage, dual mobility, and non-monotonic DIBL, in FDSOIs are analyzed and modeled in this Chapter. Since the back-gate bias affects the electrostatics in the channel and front gate oxide, all the carrier transport, leakage currents, overlap capacitance, and electrical performances would exhibit V_{BG}-dependence.

The V_{BG}-dependence in leakages (GIDL and gate leakages) can be modeled using effective front gate biases based on the models implemented in BSIM-IMG, which has been validated with the silicon data shown in Section 2.2. In FDSOIs with thick front-gate oxides for high-voltage and image sensor applications, the back-side inversion effect is evident since the amount of front channel charge is not much greater than that of back channel. Hence, the drain current is constituted from the front and back channels, and both front and back mobilities are crucial for the transconductance. Dual mobilities discussed in Section 2.3 can be linked by charge-based weighting function,

and the compact model has been validated by TCAD simulations and silicon data from different technologies. Both V_{BG}-dependent leakages, and dual mobility models have been implemented into BSIM-IMG. In Section 2.4, the non-monotonic DIBL with V_{BG} is found due to the channel length modulation (CLM) enhanced by the fringing field through the front-gate oxide and BOX. The impact of different gate stack configurations on DIBL are also examined using TCAD simulations, and the proposed explanation holds well. In Section 2.5, the back-gate bias-dependent overlap capacitance model is developed based on the model from BSIM4. By introducing back-gate bias into the overlap voltage in BSIM-IMG, the model shows good agreement with TCAD simulation and experimental data.

Chapter 3

Modeling Leakages in Advanced CMOSFETs

3.1 Introduction

Continuously scaling the gate length of MOSFET over few decades significantly improves the performance of circuit applications. However, this leads to servere power consumption issues and should be avoided. For example, the gate leakages and GIDL, modeled in FDSOIs in Chapter 2, can be eliminated by introducing high-κ and engineering lightly doped drain (LDD) structure [77]. There are also various types of leakages in MOSFETs and FinFETs, such as drain-to-body/source-to-body junction leakage currents and impact ionization currents. These leakages have been modeled in BSIM-BULK, BSIM-IMG, and BSIM-CMG. However, it is important to consider the leakage path between the source and the drain beneath the channel, which has not received much attention in the literature [78, 79, 80]. Zhu *et al.* [81] have analyzed the punchthrough currents and corresponding potential distribution for long-channel MOSFET at high drain voltage but lacked a suitable model for circuit simulation. Although the heavily doped substrate and the employment of the halo doping could suppress the punchthrough effect, they may result in significant band-to-band tunneling current through drain/source-body junctions [82]. Furthermore, even though there is a parasitic bipolar-junction transistor (BJT) comprising source-body-drain (emitter-base-collector) in the MOSFET [83], it is generally impossible to attribute the OFF-state leakage current to BJT current, because body-source (base-emitter) junction is not forward biased. In addition to the MOSFETs, the leakage current underneath the fin (channel) at OFF-state is also observed in FinFETs [84]. Thus, it is crucial to have an insight into the physics of leakage current via the path beneath the channel and the corresponding compact model implemented in BSIM-BULK.

Nowadays, the state-of-the-art 7nm technology node has been proposed [85, 86]. Although multi-gate transistors improve electrostatic control of gate, ultimately the source-to-drain direct tunneling (SDT) is inevitable [87, 88, 89] even gate control is still good, which sets the limitation for device scaling. Although analytical models are given in [89, 90], the brute force integration in their models are the limitation for the compact model purpose. To evaluate the impact of SDT current on the circuit applications, a compact model of this current component is required in BSIM-CMG.

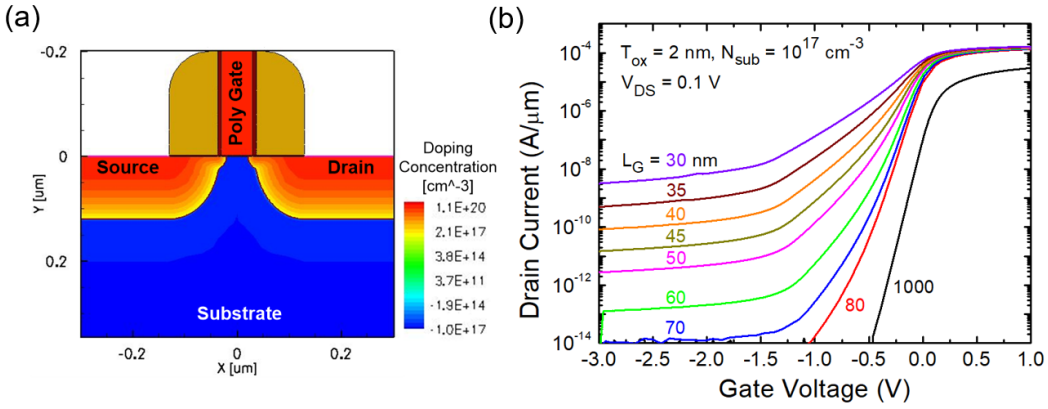


Figure 3.1: (a) Simulated structure of nMOSFET with doping concentration. (b) Simulated $I_{DS} - V_{GS}$ curves for different values of L_G at $V_{DS} = 0.1V$ and $V_S = V_B = 0V$. It shows that the subsurface leakage current at the accumulation bias region increases as the value of L_G reduces.

3.2 Sub-Surface Leakages in MOSFETs

3.2.1 Model Description

The nMOSFET structure simulated in Sentaurus TCAD [2] with gate oxide thickness $T_{\text{ox}} = 2$ nm, $N_{\text{sub}} = 10^{17}$ cm^{-3}, is shown in Fig. 3.1 (a). The peak value of Gaussian doping profile of source (drain) is 10^{20} cm^{-3}, while that of source (drain) extension is 10^{19} cm^{-3} in order to minimize the GIDL effect. The physical models used in simulations include doping-dependent mobility with high-field saturation and degradation, van Dort quantization, ShockleyReadHall, and Schenk band-to-band tunneling models [2]. Note that we choose the zero-threshold-voltage devices because they have low doping [91], and the subsurface leakage phenomenon is highly visible in these devices, which will be discussed later. The zero threshold voltage devices can be used in electrostatic discharge and I/O as well as some analog/RF applications [91, 92]. Fig. 3.1 (b) shows the simulated drain current versus gate-to-source voltage characterization for nMOSFETs for different gate lengths at $V_{DS} = 0.1V$. It can be observed that there is an approximately gate bias-independent (in a log scale) leakage current in the strong accumulation bias region and depends on the gate length. This leakage current cannot be explained by GIDL, p-n (drain-to-body) junction leakage, or parasitic BJT current. The GIDL current should increase as the gate bias becomes more negative and it can be effectively suppressed by LDD structure [77]. Furthermore, the p-n (drain-to-body) junction leakage in reverse bias is generally independent of the gate length [93]. Parasitic BJT current cannot also be considered in this case because source-body junction is not forward biased. Therefore, a new model should be developed to capture this leakage current by taking the key device parameters into account.

Drain-to-Source Voltage and Gate Length Dependence

Fig. 3.2 (a)-(c) shows the contour plots of electron currents for the devices with the gate length of 30, 50, and 70 nm at $V_{GS} = -3V$ and $V_{DS} = 0.1V$, respectively.

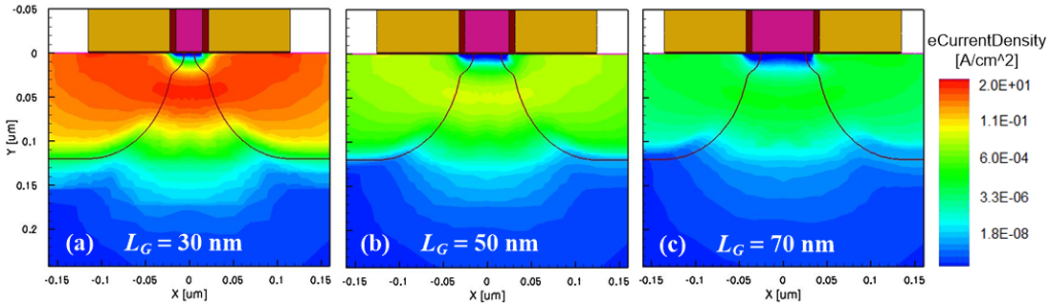


Figure 3.2: Contour plots of the electron current density for nMOSFET with the gate length of (a) 30, (b) 50, and (c) 70 nm at $V_{GS} = -3V$, $V_{DS} = 0.1V$, and $V_S = V_B = 0V$. The junction depths of source/drain and their extensions are 120 and 20 nm, respectively.

It can be observed that there are electron current paths at a distance away from the Si/SiO_2 interface which we define as subsurface leakage current. Moreover, the subsurface leakage current is gate length-dependent, as shown in Fig. 3.1 (b). This can be directly attributed to the barrier lowering induced by V_{DS}, which is similar to drain-induced barrier lowering effect [64] but at the different depths. For a shorter channel device, even a small V_{DS} can sufficiently reduce the barrier height between the source and the drain, so that the electrons are able to surmount the barrier from the source side and the corresponding current will be similar to the diode current form. Fig. 3.3 (a) shows the conduction band diagram and the extracted barrier height lowering (see the data in black in the inset) at the local minimum of conduction band due to V_{DS} of the device with $L_G = 50$ nm from the TCAD simulations. We approximate the barrier change as a linear function for simplification

$$\Delta V = \text{ASSL} \cdot V_{DS} \tag{3.1}$$

where ASSL (dimensionless) is a fitting parameter defined as barrier-change coefficient that is obtained from the TCAD data. We consider the gate length-dependence of the subsurface leakage current to be exponential relationship from the observation in Fig. 3.3 (b), which shows that the slopes in a log scale for different V_{DS} values are approximately the same. In addition, Fig. 3.3 (c) shows that the drain current exponentially depends on V_{DS}, although the curves are slightly deviating from exponential behavior at high V_{DS} due to nonlinear barrier height lowering shown in the inset of Fig. 3.3 (a) (black). The root cause of this nonlinear effect comes from the space-charge effect influenced by the high injection current, because the injected electrons can screen out the electric field provided by the drain terminal and hence decrease barrier height lowering [94]. The data in red in the inset of Fig. 3.3 (a) fortify the above hypothesis. If the mobility is low enough to reduce the drift current across the drainbody reverse-biased junction, the barrier height lowering will be much more linear, implying that the injected electrons actually are able to affect the electrostatics. Nevertheless, the linear approximation is still adopted for simplification. Thus, the subsurface leakage current can be expressed as

$$I_{DS,\text{leakage}} = \text{CSSL} \cdot e^{-\text{BSSL} \cdot L_G} \cdot \left(e^{\frac{\text{ASSL} \cdot V_{DS}}{V_t}} - 1 \right) \cdot W \tag{3.2}$$

where W is the device width in units of m, V_t is the thermal voltage ($= k_B T/q$), and BSSL and CSSL in units of m^{-1} and A/m are the variables called inverse characteristic

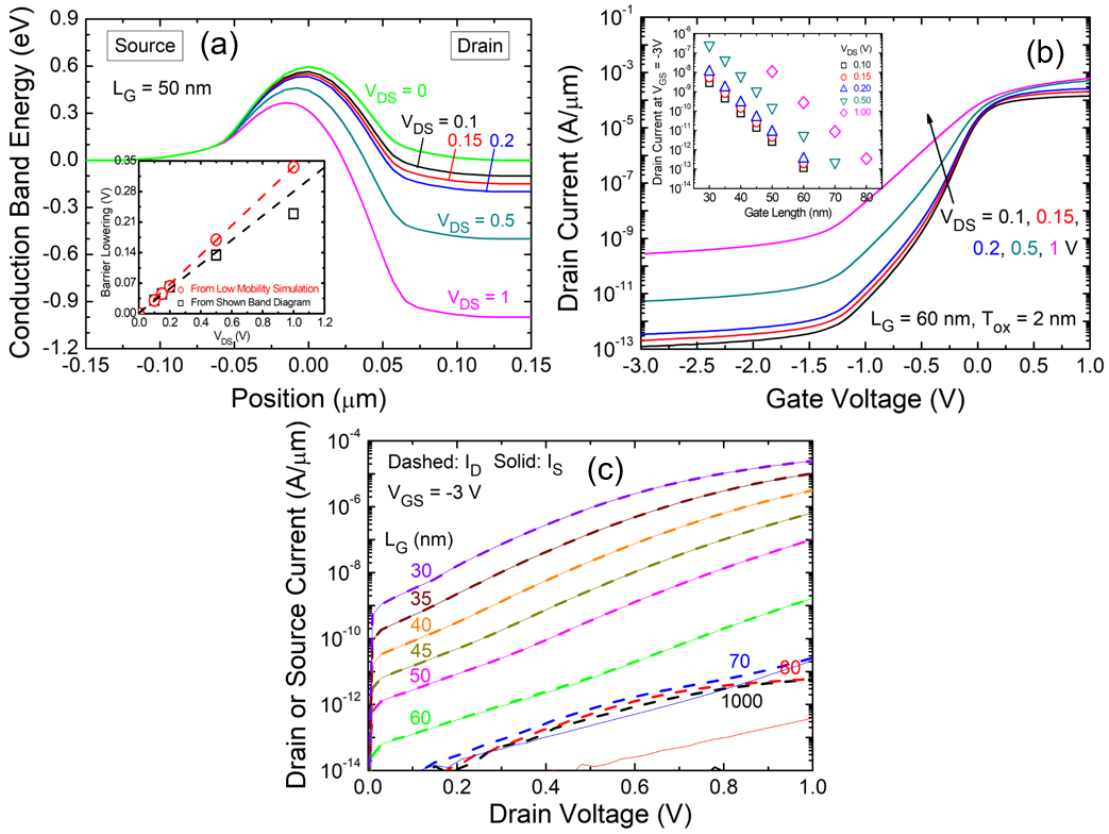


Figure 3.3: (a) Conduction band diagram of device with $L_G = 50$ nm at different values of V_{DS} but at fixed $V_{GS} = -3V$ and $V_S = V_B = 0V$ at a depth of 50 nm away from Si/SiO_2 interface where there is a local minimum of barrier height along the vertical direction. Inset: barrier lowering extracted from the band diagram (black) and from low-mobility simulation for low space-charge effect (red). (b) $I_{DS} - V_{GS}$ curves of device with $L_G = 60$ nm for different values of V_{DS}. Inset: drain current at $V_{GS} = -3V$ versus gate length for different values of V_{DS}. (c) Drain and source currents versus drain voltage curves for different values of L_G.

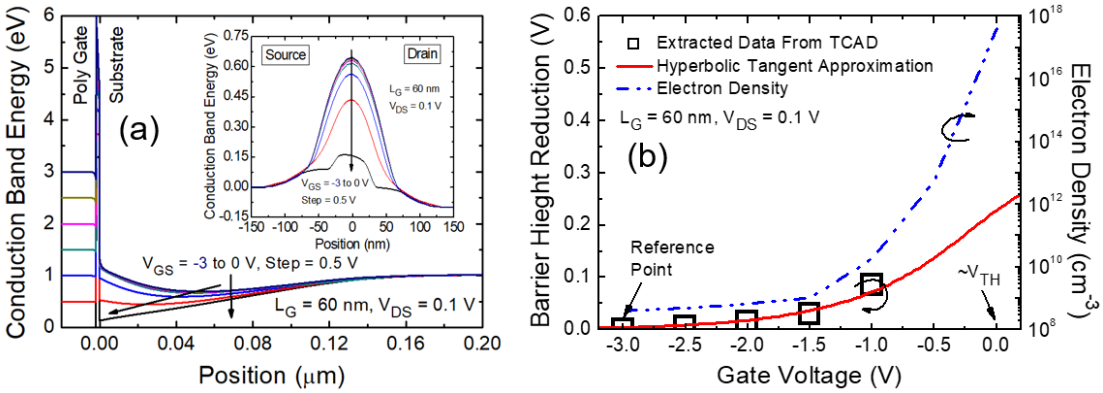


Figure 3.4: (a) Conduction band diagram in the vertical direction (from gate to substrate) and the lateral direction (from source to drain) (inset). (b) Barrier height reduction for different gate voltages extracted from (a). Hyperbolic tangent function approximation (red) is used in the model. As V_{GS} is close to V_{TH}, the electron density near the surface grows up dramatically, meaning that the barrier height reduction (band bending) leads to normal transistor current.

length and intrinsic leakage that will be discussed later. It should be noted that (3.2) guarantees that the leakage current must be zero when $V_{DS} = 0$. Furthermore, if L_G is long enough [for example, $L_G > 70$ nm, as shown in Fig. 3.3 (c)] to decouple the source and the drain, the source current is much smaller than the drain current and, the drain current is almost independent of L_G, meaning that the reverse-biased junction leakage dominates the drain current, as shown in Fig. 3.3 (c).

Impact of Gate-to-Source Voltage

As the gate voltage sweeps from the strong accumulation bias to around the flat-band voltage (V_{FB}), the local minimum of conduction band decreases further and moves toward the Si/SiO_2 interface [see Fig. 3.4 (a) (dashed arrow)], which means that the leakage current is gate voltage-dependent at the transition region from accumulation to weak inversion. Before the accumulation layer is completely formed, the gate field can penetrate into the substrate, leading to V_{GS}-dependent barrier height reduction. If the accumulation layer is built, the gate electric field will be screened out by the accumulated holes, resulting in V_{GS}-independent behavior. Nevertheless, the V_{GS}-dependent effect will no longer be prominent when V_{GS} is close to or higher than the threshold voltage (V_{TH}). In other words, if the channel is inverted, the sub-surface leakage would mix with the normal transistor current and is negligible. Fig. 3.4 (b) shows that the barrier height reduction due to V_{GS} increases exponentially for more positive V_{GS} with respect to $V_{GS} = -3V$. However, the presence of the normal transistor current is not considered in the extraction of the barrier height reduction in Fig. 3.4 (b). The reduction should gradually saturate when V_{GS} is close to or higher than V_{TH}, where the band bending is for the formation of electron inversion layer. We model V_{GS} effect on barrier height change as an exponential function inside the hyperbolic tangent function

$$\Delta\phi = \mathrm{DSSL} \cdot \tanh\left[e^{\mathrm{ESSL}\cdot(V_{GS}-V_{TH})}\right] \tag{3.3}$$

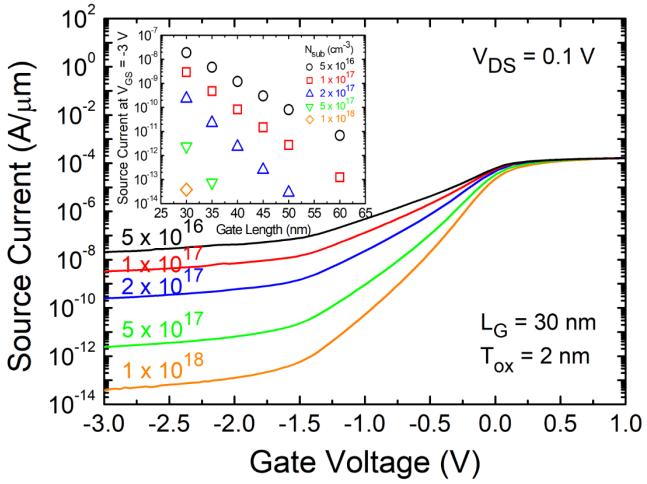


Figure 3.5: $I_{DS} - V_{GS}$ curves for nMOSFET with $L_G = 30$ nm at $V_{DS} = 0.1V$ and $V_S = V_B = 0V$ for different N_{sub} values. Inset: I_{DS} at $V_{GS} = -3V$ and $V_{DS} = 0.1V$ for different values of L_G and N_{sub}.

where DSSL and ESSL in units of V and V^{-1} are the V_{GS}-associated fitting parameters that can be determined from the TCAD data, as shown in Fig. 3.4 (b). The threshold voltage (V_{TH}) will be automatically calculated by BSIM-BULK model [3]. Therefore, from (3.2), the subsurface leakage current can be rewritten as

$$I_{DS,\text{leakage}} = \text{CSSL} \cdot e^{-\text{BSSL} \cdot L_G} \cdot e^{\frac{\Delta\phi}{V_t}} \cdot \left(e^{\frac{\text{ASSL} \cdot V_{DS}}{V_t}} - 1 \right) \cdot W \tag{3.4}$$

Substrate Doping and Temperature Dependence

In the TCAD simulation results of Fig. 3.1 (b), the uniform substrate doping N_{sub} is adopted for simplification. It is evident that in Fig. 3.5, the subsurface leakage current is inversely proportional to N_{sub}, because a heavily doped substrate can reduce the depletion width of source- and drain-to-body and, hence, coupling. However, if N_{sub} is high enough (for example, 10^{18} cm^{-3}), the GIDL current can overwhelm the subsurface leakage current in the drain current due to high electric field [82, 64]. Thus, we only plot the source currents in Fig. 3.5 in order to get rid of the GIDL currents and make the subsurface leakage current clear. If N_{sub} is not fairly high, based on Fig. 3.5 (inset), the leakage currents at $V_{GS} = -3V$ are still linear but with different slopes in a log scale for different N_{sub} values, which implies that the doping effect can be simply modeled by letting the variables BSSL and CSSL be doping concentration-dependent with respect to 10^{23} (m^3)

$$\text{BSSL} = \text{BSSL0} \left(\frac{N_{\text{sub}}}{10^{23}} \right)^{\text{EXP1}}$$

$$\text{CSSL} = \text{CSSL0} \cdot \exp \left[- \left(\frac{N_{\text{sub}}}{10^{23}} \right)^{\text{EXP1}} \right]$$

where N_{sub} is in units of m^{-3}, and BSSL0, CSSL0, and EXP1 are the fitting parameters. If the doping concentration between the source and the drain is high enough

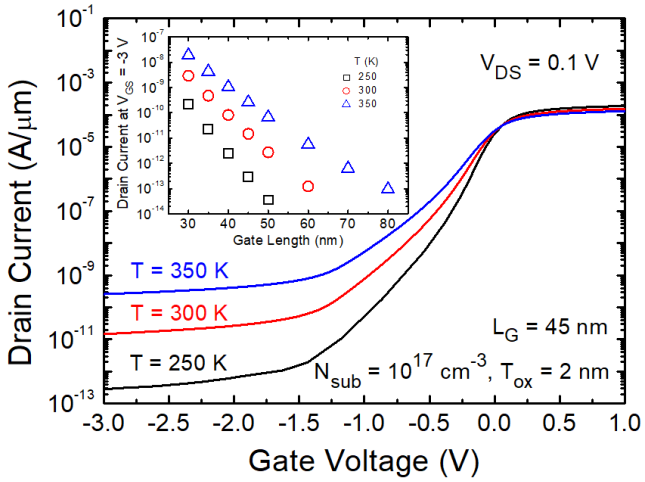


Figure 3.6: $I_{DS} - V_{GS}$ curves for nMOSFET with $L_G = 45$ nm at $V_{DS} = 0.1V$ and $V_S = V_B = 0V$ for different T values. Inset: I_{DS} at $V_{GS} = -3V$ and $V_{DS} = 0.1V$ for different values of L_G and T.

to avoid coupling of the drain-to-body and source-to-body junction depletion regions, the subsurface leakage current can be blocked by the barrier. A complicated doping profile can also suppress the leakage. For example, carrying out the halo implantation and adopting the retrograde doping can suppress the leakage current. It is worth noting that although the substrate doping profile in a real device is sophisticated so that the V_{DS} dependence is not significant, as shown in Fig. 3.3 (b), the barrier-change coefficient ASSL can be set to a small value for matching the experimental data.

In addition to N_{sub}, the temperature dependence should also be included in the model. Due to more energetic electrons at higher temperature passing the barrier, the subsurface leakage would increase as the temperature rises, as shown in Fig. 3.6. Fig. 3.6 (inset) shows the leakage current versus gate length curves for different temperatures, and it can be observed that they are linear functions but with different slopes in a log scale for different temperatures just like N_{sub} effect, so a fitting parameter EXP2 is introduced to capture the temperature effect. As a result, variables BSSL and CSSL can finally be expressed as

$$\text{BSSL} = \text{BSSL0} \left(\frac{N_{\text{sub}}}{10^{23}}\right)^{\text{EXP1}} \left(\frac{300}{T}\right)^{\text{EXP2}} \tag{3.5}$$

$$\text{CSSL} = \text{CSSL0} \cdot \exp\left[-\left(\frac{N_{\text{sub}}}{10^{23}}\right)^{\text{EXP1}} \left(\frac{300}{T}\right)^{\text{EXP2}}\right] \tag{3.6}$$

where T is in units of K. From (3.4), the complete form of subsurface leakage current is

$$I_{DS,\text{leakage}} = \text{CSSL}\left(N_{\text{sub}}, T\right) \cdot e^{-\text{BSSL}(N_{\text{sub}},T)\cdot L_G} \cdot e^{\frac{\Delta\phi}{V_t}} \cdot \left(e^{\frac{\text{ASSL}\cdot V_{DS}}{V_t}} - 1\right) \cdot W \tag{3.7}$$

3.2.2 Results and Discussion

The model of subsurface leakage current (3.7) and related parameters is incorporated in the BSIM-BULK model. Fig. 3.7(a) and (b) show the comparison of the

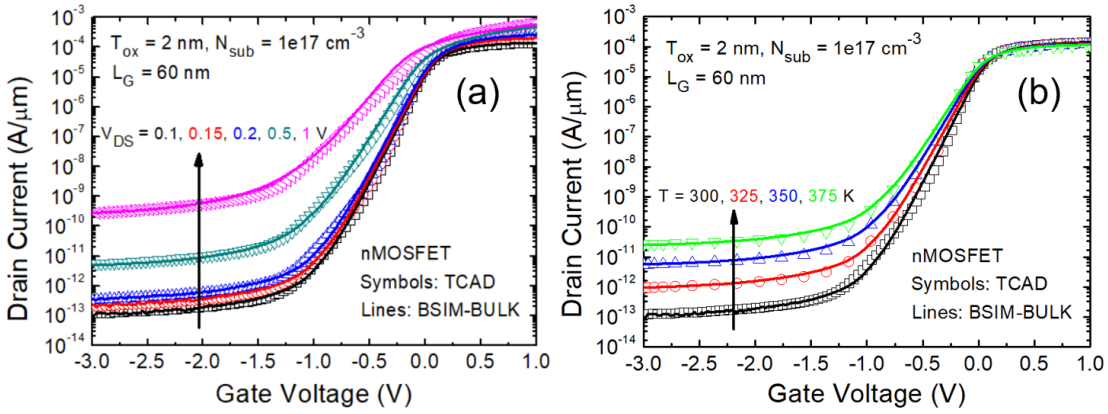


Figure 3.7: Comparison of TCAD and BSIM-BULK with (3.7) for nMOSFET with L_G = 60 nm at $V_S = V_B = 0V$ for (a) different values of V_{DS} and (b) different temperatures.

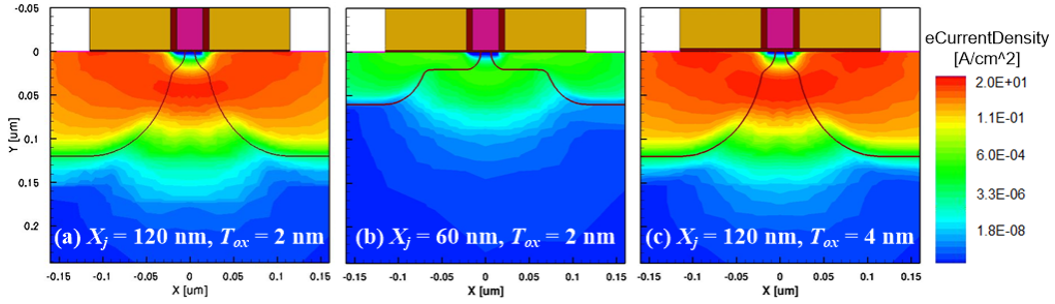


Figure 3.8: Contour plots of the electron current density for L_G = 30 nm nMOSFET with (a) X_j = 120 nm and T_{ox} = 2 nm (for reference), (b) X_j = 60 nm and T_{ox} = 2 nm, and (c) X_j = 120 nm and T_{ox} = 4 nm at $V_{GS} = -3V, V_{DS} = 0.1V$, and $V_S = V_B = 0V$.

TCAD simulation data with the SPICE simulation results for different V_{DS} values (from linear to saturation bias regimes) and temperatures for an nMOSFET with the gate length of 60 nm, the device width of 1 μm, the substrate doping of 10^{17} cm^{-3}, and the gate oxide thickness of 2 nm, respectively. Fig. 3.7 (a) and (b) are with the same set of fitting parameters. The developed model exhibits good match with the TCAD simulation data. The V_{GS} effect appears at around flat-band voltage $V_{FB}(\approx -1V)$ and is successfully included in model. Equation (3.7) captures not only the barrier lowering due to V_{DS} but also the current increment caused by rising temperature. It is worth noting that the influence of the junction depths X_j of source and drain is not explicitly modeled in this paper. As X_j is shallower, the leakage path is much closer to the surface and the leakage is suppressed because of better gate control, as shown in Fig. 3.8. The junction depth effect can be included by adjusting the intrinsic leakage CSSL and barrier-change coefficient ASSL. Nevertheless, a shallow junction could result in higher series resistance and thus degrade the device performance. In addition to X_j, the gate oxide thickness would also affect the subsurface leakage current. If the gate oxide is thicker (less gate control), the subsurface leakage current would be prominent, because drain/source-body capacitance overcomes the gate capacitance, as shown in Fig. 3.8 (c), even though the halo doping or the retrograde doping is adopted (not shown). Similar to X_j, the gate oxide effect can be captured by adjusting CSSL and ASSL to suitable values.

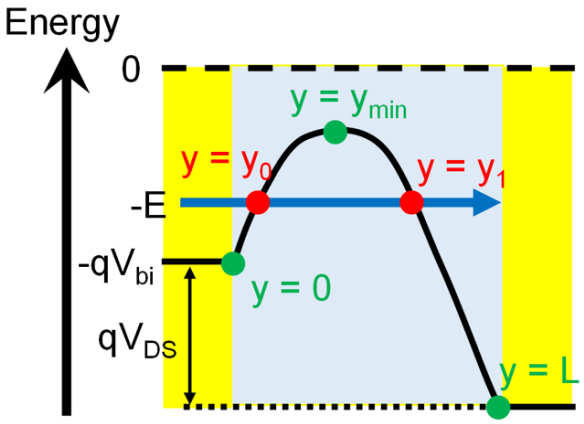


Figure 3.9: Electron potential energy profile of a transistor. The yellow and blue regions represent source/drain and channel, respectively.

3.3 Source-to-Drain Tunneling Leakages in Sub-10 nm Devices

3.3.1 Model Description

Intraband Tunneling With Quadratic Potential Barrier

The source-to-drain tunneling (SDT) appears when the potential barrier height and the width of the channel are small enough, i.e., for very short channel length. The drain bias could affect the electrostatics in the channel against the gate so that the potential as well as tunneling probability are drain- and gate-bias dependent. To evaluate the SDT current, the Landauers equation is adopted [21]

$$I_{\mathrm{SDT}} = \frac{2q}{h} \int M(E)\, T(E)\, [f_S(E) - f_D(E)] \tag{3.8}$$

where q is the charge, h is the Plancks constant, M is the 2-D conduction mode [21], T is the tunneling probability, and $f_{S(D)}$ is the Fermi distribution function at source (drain). The probability of a carrier to tunnel through a potential energy barrier $V(y)$ can be estimated using Wentzel-Kramers-Brillouin (WKB) approximation

$$T(E) = \exp\left[-\frac{2\sqrt{2m^*}}{\hbar} \int_{y_0}^{y_1} \sqrt{V(y) - K}\, dy\right] \tag{3.9}$$

where m^* is the effective mass, $\hbar$ is the reduced Plancks constant, K is the energy of the carrier ($-E$ as defined in Fig. 3.9), and y_0 and y_1 are defined in Fig. 3.9 and are the positions where the potential energy is equal to the kinetic energy. Although the WKB approximation (3.9) may overestimate the tunneling current due to the complex band structure and wave function mismatch at the high field region [95, 96, 97, 98, 99, 100, 101], for compact model purpose the WKB approximation is used to capture the general dependence on biases and gives a simple analytical equation which is good for the simulation speed. The potential profile in a short channel device can be approximated as a quadratic function [102], as shown in Fig. 3.9

$$V(y) = a + by + cy^2 \tag{3.10}$$

with boundary conditions

$$\begin{cases} V\left(y=0\right)=V_{\text{bi}} \\ V\left(y=L\right)=V_{\text{bi}}+V_{DS} \\ V\left(y=y_{\text{min}}\right)=V_{\text{min}} \end{cases} \tag{3.11}$$

Hence, three variables a, b, and c are determined

$$\begin{cases} a=V_{\text{bi}} \\ b=-\frac{y_{\text{min}}V_{DS}}{L(L-y_{\text{min}})}+\frac{L(V_{\text{min}}-V_{\text{bi}})}{y_{\text{min}}(L-y_{\text{min}})} \\ c=\frac{V_{DS}}{L(L-y_{\text{min}})}-\frac{(V_{\text{min}}-V_{\text{bi}})}{y_{\text{min}}(L-y_{\text{min}})} \end{cases} \tag{3.12}$$

At $y=y_{\text{min}}$, the potential has a local minimum, which gives

$$y_{\text{min}}=L\cdot\frac{\sqrt{\frac{V_{\text{bi}}-V_{\text{min}}}{V_{\text{bi}}+V_{DS}-V_{\text{min}}}}}{1+\sqrt{\frac{V_{\text{bi}}-V_{\text{min}}}{V_{\text{bi}}+V_{DS}-V_{\text{min}}}}} \tag{3.13}$$

In (3.11), the minimum potential in the channel is expressed as

$$V_{\text{min}}=V_{GS}-V_{FB}-V_{\text{ox}}-\Delta V_{TH,\text{DIBL}}+\alpha \tag{3.14}$$

where α is the model parameter to capture nonuniform doping at the source (drain)-to-channel junctions because graded doping may effectively affect the barrier, V_{FB} is the flat band voltage, V_{ox} is the oxide voltage calculated by BSIM-CMG core model which automatically includes the quantum effect [14], and $\Delta V_{TH,\text{DIBL}}$ is threshold voltage shift due to drain-induced barrier lowering (DIBL) and is modeled as [14, 15, 103]

$$\Delta V_{TH,\text{DIBL}}=-0.5\cdot\beta\cdot\frac{V_{DS}}{\cosh\left(\gamma\cdot\frac{L}{\lambda}\right)} \tag{3.15}$$

where β and γ are the model parameters to increase the model flexibility for various technologies [14], and λ is the characteristic length [15, 104]. The subthreshold swing degradation due to the interfacial quality and electrostatics control is captured in V_{ox} via the gate transfer factor capacitor divider model [14]. Using the analytical expression of the channel potential energy profile, the integration in (3.9) can be carried out (where the electron is with energy $-E$) and the tunneling probability is

$$T\left(E\right)=\exp\left[-\frac{2\sqrt{2m^{*}}}{\hbar}\cdot\frac{\pi\left(E-qV_{\text{min}}\right)}{2\sqrt{q\cdot c}}\right] \tag{3.16}$$

Fig. 3.10 shows the tunneling probability as a function of energy from the top to the bottom of the channel potential barrier in a device with gate length of 5 nm for different materials: Si, Ge, and InAs. As drain bias increases, the barrier is pulled down and thus the tunneling probability increases. Furthermore, the effective mass also affects the tunneling probability. The carrier with lighter effective mass has more wave nature, which gives higher tunneling probability as predicted in (3.16). The tunneling probability generally increases exponentially with the energy level from the

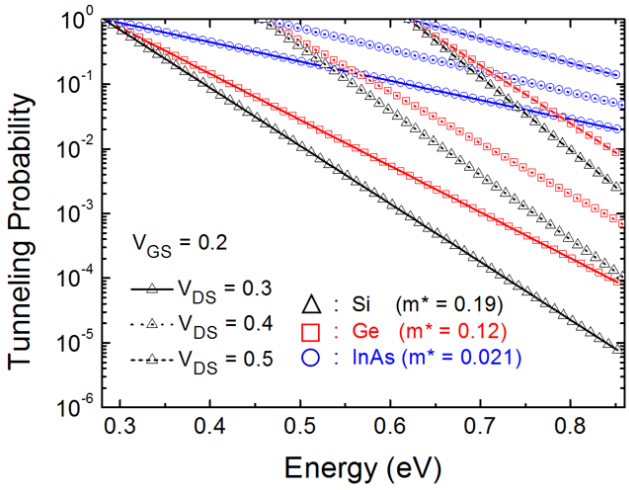


Figure 3.10: Calculated tunneling probability versus energy for the device with gate length of 5 nm. The energy ranges from the top to the bottom of the potential barrier.

bottom of the barrier, which is in agreement with the results considering the complex band structure [89]. In Fig. 3.10, it shows that Si FinFET suffers less SDT and thus has better subthreshold slope [95], although the materials with lighter effective masses and thus higher mobilities are always adopted to boost the current [89, 105]. Note that the same electrostatic potential is assumed for these three materials in Fig. 3.10 for comparison. At high V_{GS}, the quadratic potential approximation will not be accurate, because the potential barrier from the top toward the drain becomes linear [106]. The tunneling probability with the quadratic potential in this scenario could be overestimated. However, at high V_{GS}, the barrier is low and the tunneling window is narrow so that around the top of barrier the linearity will not affect the total integration much if the quadratic function is assumed.

Based on the energy coordinate system shown in Fig. 3.9, the upper and lower limits of the integration in (3.8) are qV_{bi} and qV_{min}. With that, (3.9) becomes

$$I_{\mathrm{SDT}} = \frac{2q}{h}\int_{qV_{\mathrm{min}}}^{qV_{\mathrm{bi}}} M\left(E\right)T\left(E\right)\left[f_S\left(E\right)-f_D\left(E\right)\right] \tag{3.17}$$

and the two-dimensional conduction mode M, which is associated with the average velocity of the carrier and the density of states (DOS) [107, 108], and Fermi distribution functions are

$$\begin{cases} M\left(E\right) = W\cdot g_v\cdot\frac{\sqrt{2m^*\left(qV_{\mathrm{bi}}-E\right)}}{\pi\hbar} \\ f_S\left(E\right) = \left[1+\exp\left(\frac{qV_{\mathrm{bi}}-E}{k_BT}\right)\right]^{-1} \\ f_D\left(E\right) = \left[1+\exp\left(\frac{qV_{\mathrm{bi}}+qV_{DS}-E}{k_BT}\right)\right]^{-1} \end{cases} \tag{3.18}$$

where W is the device width and g_v is the valley degeneracy [107]. Since there is no analytical expression for (3.17), a numerical technique called the Gaussian quadrature method is introduced to complete the integration [7]. This method states that an integral of a well-behaved function can be expressed as a summation by choosing

Table 3.1: Example weight and abscissa for Gaussian quadrature for $N = 6$

i	**Weight** w_i	**Abscissa** x_i
1, 2	0.3607615730481386	±0.6612093864662645
3, 4	0.4679139345726910	±0.2386191860831969
5, 6	0.1713244923791704	±0.9324695142031521

specific weights and abscissa [109]

$$\int_n^m f(x)\,dx \approx \sum_{i=1}^{N} w_i f\left(\frac{(m-n)\,\zeta_i + (m+n)}{2}\right)\frac{(m-n)}{2} \tag{3.19}$$

where m and n are the upper and lower limits, N is the number of Gaussian points, w is the weight, and ζ is the abscissa. An example table of w and ζ for $N = 6$ is shown in Table 3.1. N in this model is 6, which gives accurate results [21]. By the Gaussian quadrature method, (3.17) is carried out easily without making further approximations.

Simplification of Model Equation

Although the Gaussian quadrature method enables complex integration, however, the speed of the simulation would be reduced, which is undesirable in the case of the tunneling model for compact model purpose. The simplification is required based on some approximations. Due to the drain voltage, the energy level corresponding to the tunneling is far away from the drain Fermi level. Thus, the Fermi distribution of the drain side is assumed to be zero, and the Boltzmann approximation is applied to the source side. Hence, the integration of (3.17) becomes

$$\begin{aligned} I_{\mathrm{SDT}} \approx & \frac{4qWg_v\sqrt{2m^*}}{h^2}\exp\left[-\frac{\pi q\sqrt{2m^*}}{\hbar\sqrt{q\cdot c}}\left(V_{\mathrm{bi}}-V_{\mathrm{min}}\right)\right] f_{\mathrm{fermi}} \\ & \times \int_0^{q(V_{\mathrm{bi}}-V_{\mathrm{min}})} \sqrt{x}\cdot\exp\left[-\frac{1}{q}\left(\frac{q}{k_BT}-\frac{q\pi\sqrt{2m^*}}{\hbar\sqrt{q\cdot c}}\right)x\right]dx \end{aligned} \tag{3.20}$$

where f_{fermi} is to ensure zero current at zero drain bias [110]

$$f_{\mathrm{fermi}} = 1 - \frac{2}{1+\exp\left(\frac{qV_{DS}}{k_BT}\right)} \tag{3.21}$$

For Si device with gate length of 5 nm, the coefficient of the exponent in the integration $\frac{q}{k_BT} - \frac{q\pi\sqrt{2m^*}}{\hbar\sqrt{q\cdot c}}$ is much greater than 0, which means that the function in the integration would decrease quickly. Thus, the upper limit of the integration in (3.20) can be

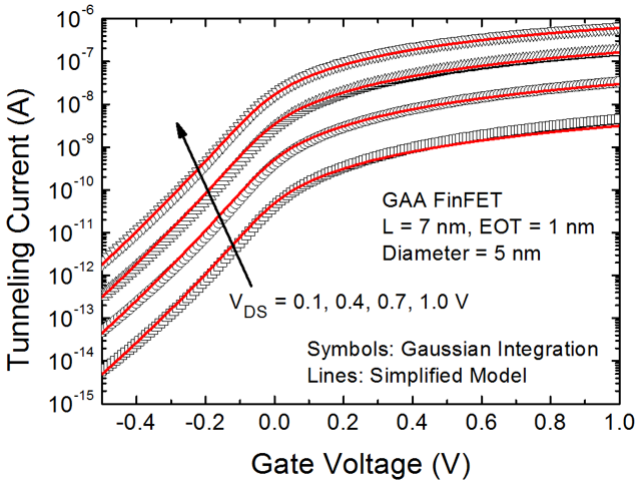


Figure 3.11: Comparison of simplified model and Gaussian integration ($N = 6$).

replaced with infinity, so the integration is carried out analytically

$$\begin{aligned} I_{\text{SDT}} &\approx \frac{4qWg_v\sqrt{2m^*}}{h^2}\exp\left[-\frac{\pi q\sqrt{2m^*}}{\hbar\sqrt{q\cdot c}}\left(V_{\text{bi}}-V_{\text{min}}\right)\right] f_{\text{fermi}} \\ &\quad\times q^{\frac{3}{2}}\cdot\left(\frac{q}{k_BT}-\frac{q\pi\sqrt{2m^*}}{\hbar\sqrt{q\cdot c}}\right)^{-\frac{3}{2}}\cdot\frac{\sqrt{\pi}}{2} \\ &\approx \frac{2qWg_v\sqrt{2m^*\pi}}{h^2}\left(k_BT\right)^{\frac{3}{2}}\exp\left[-\frac{\pi q\sqrt{2m^*}}{\hbar\sqrt{q\cdot c}}\left(V_{\text{bi}}-V_{\text{min}}\right)\right] \\ &\quad\times\left(1+\frac{3\pi\sqrt{2m^*}k_BT}{2\hbar\sqrt{q\cdot c}}\right) f_{\text{fermi}} \end{aligned} \tag{3.22}$$

For the compact model purpose, (3.22) is further rewritten as

$$I_{\text{SDT}} = \text{ASDT}\cdot W\cdot\exp\left[-\frac{\text{BSDT}}{\sqrt{c}}\left(V_{\text{bi}}-V_{\text{min}}\right)^{\text{CSDT}}\right]\cdot\left(1+\frac{3V_t\cdot\text{BSDT}}{2\sqrt{c}}\right)\cdot f_{\text{fermi}} \tag{3.23}$$

where ASDT, BSDT, and CSDT are the model parameters, and V_t is the thermal voltage k_BT/q. Fig. 3.11 shows the comparison of (3.23) with (3.17) using the Gaussian integration. Fine-tuning parameters for (3.23) matches (3.17) well and has more flexibility which is desirable for compact model purpose. Note that both (3.17) and (3.23) are implemented into BSIM-CMG. Fig. 3.12 shows the speed of (3.17) and (3.23) compared with BSIM-CMG without SDT model. With simplification, the speed is improved by 5 times (from +16.4% to +3.3%). Therefore, the simplified equation is adopted.

3.3.2 Results and Discussion

The model is implemented into industry standard model BSIM-CMG where the simulation results are shown in Figs. 5−7. Fig. 3.13 shows the SDT current from the model versus V_{GS} for various gate lengths. The SDT current decreases exponentially

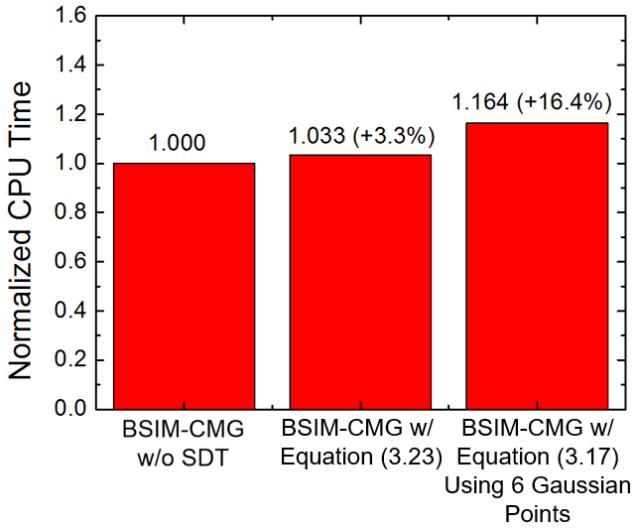


Figure 3.12: Speed comparison of BSIM-CMG without and with SDT models.

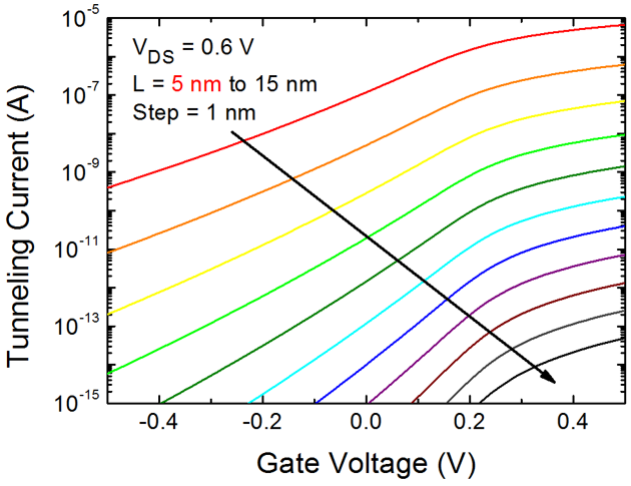


Figure 3.13: SDT current versus V_{GS} for gate lengths from 5 to 15 nm in a Si nanowire GAA transistor with diameter of 4 nm and effective oxide thickness (EOT) of 1 nm. The parameters used in simulation are the same as in Fig. 3.15.

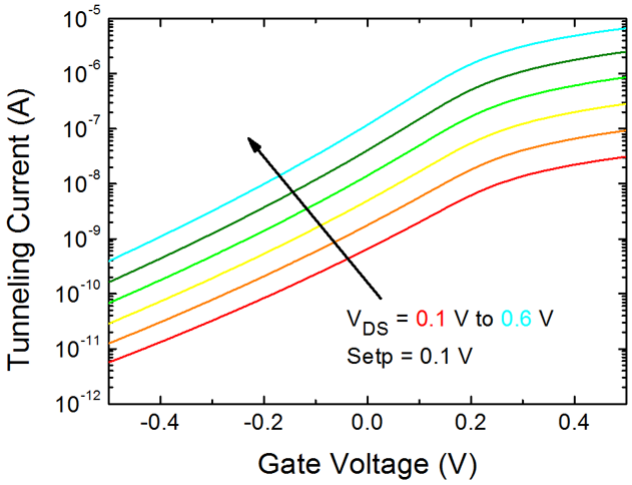


Figure 3.14: SDT current versus V_{GS} for V_{DS} = 0.1, 0.2, 0.3, 0.4, 0.5, and 0.6 V in a Si nanowire GAA transistor with diameter of 4 nm, gate length of 5 nm, and EOT of 1 nm. The parameters used in simulation are the same as in Fig. 3.15.

with the gate length because the bias-dependent function c in (3.12) is inversely proportional to the gate length square. It can be observed that when the gate length is longer than 10 nm, the SDT becomes less important compared to the normal transistor current. Fig. 3.14 shows the tunneling current for different drain biases. As the drain bias increases, the potential barrier is reduced and hence boosts the tunneling probability. Fig. 3.15 shows the model validation with the full-band atomistic quantum transport simulation data of the Si nanowire gate-all-around (GAA) transistors [95]. The extracted model parameters relevant to the SDT current are ASDT = 1.0023×10^4 (A/m), BSDT = 6.15×10^9 $(m^{-1}V^{-0.5})$, CSDT = 1.11, α = 0.3 (V), β = 1, and γ = 0.4. For L = 5 nm, the total current has $\approx$ 96.3% tunneling current at OFF-state ($V_{GS} = 0V, V_{DS} = 0.6V$) and $\approx$ 33.5% at ON-state ($V_{GS} = 0.5V, V_{DS} = 0.6V$), indicating the scaling limit of the transistor. For L = 15 nm, the tunneling current is negligible compared to the drift-diffusion current. The subthreshold slope is degraded from 63 to 90 mV/dec as scaling L from 15 to 5 nm. Although the subthreshold slope can be matched using the interfacial capacitance-related parameters [14], there is no reason assuming an ultra-scaled device has significantly inferior interfacial quality. The short channel effect could come into the picture when the gate length is scaled [14]. In Fig. 3.15, the total current without the SDT current is also shown. The subthreshold slope is 78.6 mV/dec, indicating that the short channel effect is insufficient to explain the subthreshold slope degradation. In addition, when scaling the gate length, the fin thickness or even the device geometry will be changed to suppress the short channel effect and improve the subthreshold slope. For example, a tri-gate transistor of 14-nm technology is reported with subthreshold slope of 61 mV/dec [111]. Therefore, the SDT current is a reasonable cause for the subthreshold slope degradation in an ultra-scaled device. To extract the model parameters of the SDT current, one may start from the shortest length with good electrostatics control among devices with multiple gate lengths to get accurate model parameters for subthreshold slope degradation due to the short channel effect and interfacial quality. Then, for even shorter gate length device, the model parameters of the SDT current are used to fit the data. For instance,

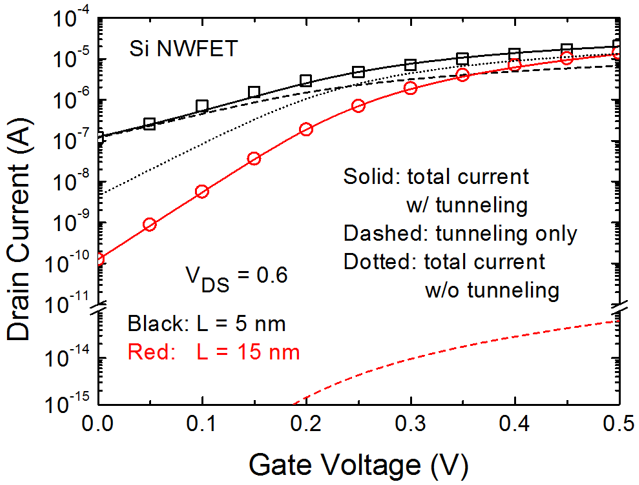


Figure 3.15: Drain current versus V_{GS} for gate lengths of 5 and 15 nm in a Si nanowire GAA transistor with diameter of 4 nm. Symbols: simulation from [95]. Lines: proposed model (Solid: total current with tunneling. Dashed: tunneling only. Dotted: total current without tunneling).

in Fig. 3.15 the classical subthreshold slope degradation parameters are extracted first from $L = 15$ nm device, and then the SDT current parameters are adjusted to capture the subthreshold region for $L = 5$ nm device.

The hole tunneling from the drain conduction band to the source valence band is neglected. If the bandgap is larger than V_{DS} which is low in ultra-scaled FET, this leakage component can be avoided [89]. Although the surface orientation for the transport effective mass would affect the tunneling probability [105], only one effective mass captured by the model parameters is considered to simplify the compact model. Note that in our model the channel length is assumed to be the same as the gate length. To suppress the SDT, one may use the gate underlap structure to effectively increase the channel length [6, 112] but it may degrade the ON current due to worse series resistance. As a result, the compact model of the SDT is needed to evaluate the circuits using ultimately scaled transistors.

3.4 Conclusion

In this Chapter, the leakages in ultra-scaled MOSFETs and FinFETs, including sub-surface leakage and source-to-drain tunneling, are discussed and analyzed. In Section 3.2, a subsurface leakage current model is developed phenomenologically based on the TCAD simulation. The subsurface leakage current is generally caused by the barrier height reduction between the source and the drain influenced by V_{DS} at a distance from the Si/SiO_2 interface. V_{GS} can also affect the barrier height but only significant near V_{FB}. As gate length is scaled down, the current increases due to stronger source to drain coupling. Furthermore, N_{sub} and T are the important factors. When N_{sub} becomes heavier, the depletion widths of junctions reduce and hence lower leakage current. As T rises, there are more electrons with enough energies to surmount the barrier, so that the leakage current increases further. The above effects are successfully included in the developed model. Finally, the model is implemented into the

BSIM-BULK model and is in good agreement with the TCAD simulation for different V_{DS} values and temperatures.

In Section 3.3, a compact model of the SDT current is presented. The tunneling current is evaluated by Landauers equation where the conduction modes, WKB-based tunneling probability to capture the essential bias dependence, and Fermi distribution function are considered. Using the Gaussian quadrature technique, Landauers equation is calculated without approximations. In order to further improve the simulation speed, the formula of the SDT current is simplified. The proposed model is incorporated with the industry standard model BSIM-CMG, and is validated by the GAA FinFET data from the atomistic quantum transport simulations.

Chapter 4

Compact Model of Tunnel FETs

4.1 Introduction

Tunnel FET (TFET) is a promising candidate for low power applications [113, 114] due to its potential to be operated at very low V_{DD} due to its steep subthreshold slope ($<$ 60 mV/dec) and thus very low switching energy [115, 116]. The steep subthreshold slope of TFET comes from the tunneling mechanism happening at source/channel junction, which does not rely on the charge distribution obeying the Boltzmann tail. The experimental results have shown the capability of having sub-60 mV/dec in subthreshold slope in Si TFET [31] and InGaAs TFET [117]. In order to explore the TFET-based circuit, a robust compact model is required. Although compact models of TFETs are available in the literature [110, 118, 119, 120], the electric field of the tunneling junction is always assumed to be constant over the energy range in the junction, resulting in inaccurate tunneling probability. To take nonuniform electric field into account, the tunneling current should be described by Landauer equation [121], which sums up all possible tunneling paths over the tunneling window. However, brute force integration of the Landauer equation for tunneling probability and carrier distribution is numerically inefficient [122, 123]. In this Chapter, to overcome this numerical challenge, we apply the Gaussian quadrature method [109] for the first time to develop compact model of tunnel devices. This enables us to directly integrate Landauer equation with WentzelKramersBrillouin (WKB)-based tunneling, Fermi distribution function, and band edge tailing, resulting in a more accurate and predictive compact model.

4.2 Core Model

4.2.1 Surface Potential and Charge

The TFET in this model is treated as a tunnel diode in series with an MOSFET [110]. Because the tunneling current is much smaller than the drift-diffusion current in the MOSFET, the TFET current is generally limited and determined by the tunneling. Instead of using Kane's model with a constant electric field across the tunnel junction as in literature [110, 118, 119, 120], the TFET band-to-band tunneling current at the

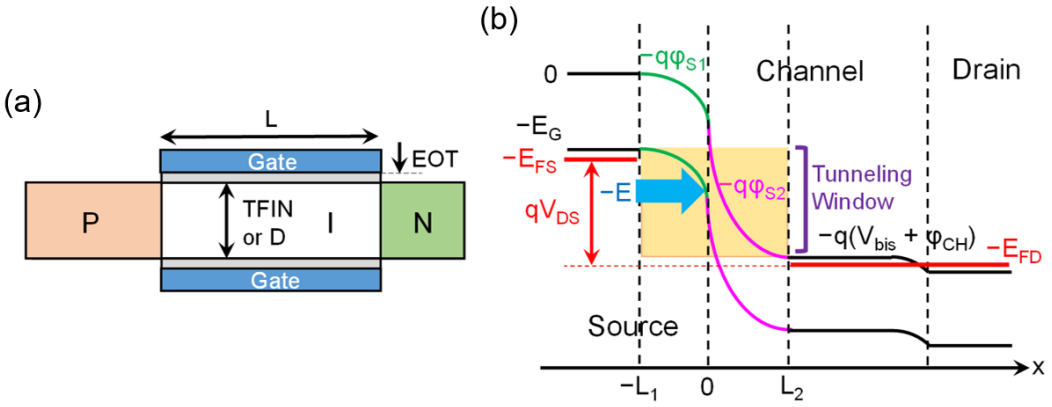


Figure 4.1: (a) Multigate TFET. TFIN is the fin thickness and D is the nanowire diameter. (b) Band diagram when device is ON.

tunnel junction in this paper is described by Landauer equation [121]

$$I = \frac{2q}{h} \int M(E)\, T(E)\, [f_S(E) - f_{CH}(E)]\, dE \tag{4.1}$$

where M is the dimensionless number of the conduction (tunneling) modes, T is the tunneling probability, and f_S and f_{CH} are the Fermi distribution functions of source and channel

$$f_S(E) = \frac{1}{1 + \exp\left[\frac{(E_{FS} - E)}{k_B T}\right]} \tag{4.2}$$

$$f_{CH}(E) = \frac{1}{1 + \exp\left[\frac{(E_{FS} + qV_{DS} - E)}{k_B T}\right]} \tag{4.3}$$

Fig. 4.1 shows the schematic of a multigate TFET and its band diagram in the ON-state. In order to calculate the tunneling probability, the potential profile of sourcechannel junction is required. The electrostatics can be solved from Poisson equation with boundary conditions set in Fig. 1 and continuous electric field at x = 0 [122, 124, 125], resulting in the following potential expression of each region:

$$\psi_{S1}(x) = \frac{qN_S}{2\epsilon_S}(x + L_1)^2 \tag{4.4}$$

$$\psi_{S2}(x) = (V_{GS} - V_{FBS}) - \frac{(V_{GS} - V_{FBB} - \psi_{CH})}{2} \exp\left(\frac{x - L_2}{\lambda}\right) \tag{4.5}$$

where N_S is the source doping concentration, ϵ_S is the source permittivity, V_{FBS} and V_{FBB} $(= V_{FBS} + V_{\text{bis}})$ are the flat-band voltages of the source and channel, V_{bis} is the built-in potential of the source-channel junction, λ is the characteristic length determined by the device geometry [104], ψ_{CH} is the surface potential in the channel, which is determined by both V_{GS} and V_{DS} [122, 126] using the BSIM-CMG model [14], and L_1 and L_2 are the high-field region widths in the source and channel give by

$$L_1 = \sqrt{\frac{2\epsilon_S \psi_S(0)}{qN_S}} \tag{4.6}$$

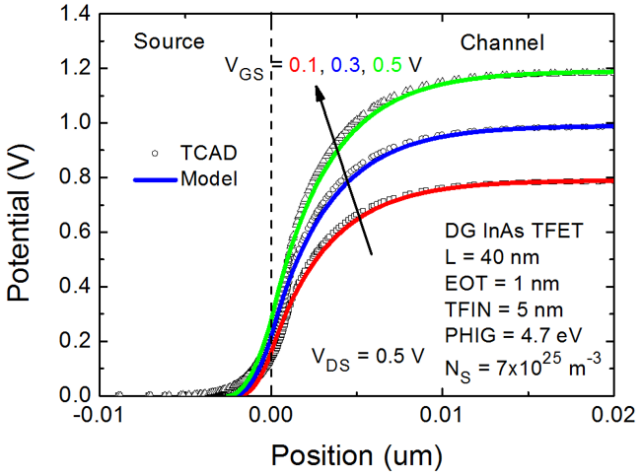


Figure 4.2: Surface potential profiles at various biases in a DG InAs FET. The model results are in good agreement with TCAD simulations.

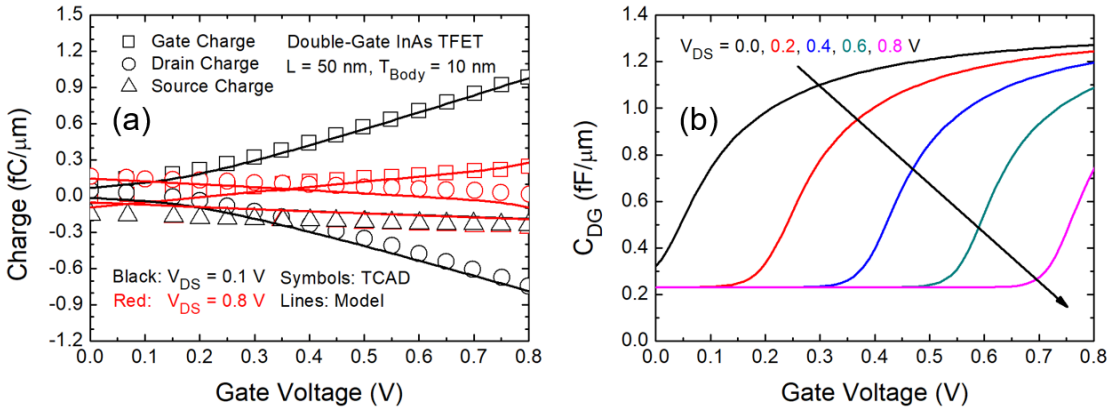


Figure 4.3: (a) Comparison of proposed model and TCAD simulated data [124] for the terminal charges in an DG InAs TFET. (b) C_{DG} of TFET with various V_{DS}.

$$L_2 = \lambda \ln\left[\frac{2\cdot\left(V_{GS}-V_{FBS}-\psi_S\left(0\right)\right)}{V_{GS}-V_{FBS}-V_{\text{bis}}-\psi_{CH}}\right] \tag{4.7}$$

where

$$\psi_S\left(0\right) = -\sqrt{\left(V_{GS}-V_{FBB}-\psi_{CH}\right)^2+2\left(V_{GS}-V_{FBS}\right)\Phi+\Phi^2} + \left(V_{GS}-V_{FBS}+\Phi\right) \tag{4.8}$$

$$\Phi = \frac{qN_S\lambda^2}{\epsilon_S} \tag{4.9}$$

Fig. 4.2 shows the model results of the surface potential profiles at various V_{GS} in a 40-nm double-gate (DG) InAs TFET are in good agreement with 2-D TCAD simulations [2]. The terminal charges and corresponding capacitance of TFET also can be obtained based on derived surface potential as shown in Fig. 4.3 (a) and (b). The TFET has 100/0 (drain/source) charge partition in the channel which is totally different from the MOSFET since the channel charges of TFET are provided from the drain [110, 126]. Therefore, the drain charge (Q_D) is modeled using the channel charge (Q_{CH}) obtained

from BSIM-CMG with additional parasitic charge ($Q_{D,parasitic}$)

$$Q_D = Q_{CH} + Q_{D,parasitic} \tag{4.10}$$

The source charge is coming from the depletion charge and additional parasitic charge ($Q_{S,parasitic}$)

$$Q_S = -qN_S L_1 A_{CH} + Q_{S,parasitic} \tag{4.11}$$

where A_{CH} is the cross-section area of TFET. The gate charge thus is

$$Q_G = -\left(Q_S + Q_D\right) \tag{4.12}$$

Note that as increasing V_{DS}, C_{DG} shifts to the right since the barrier between the channel and drain increases which is resultant from the 100/0 partition.

4.2.2 WKB-Based Tunneling Probability

The WKB approximation is a useful method to evaluate the tunneling probability, even though there are some limitations like wave function mismatch at high electric field region [100], leading to overestimation of the tunneling probability, which will be addressed later by including the electron wave reflectance. Based on WKB approximation, the probability for a carrier to tunnel through a barrier with potential energy $V(x)$ can be expressed as

$$T(E) = \exp\left[-\frac{2\sqrt{2m^*}}{\hbar}\int\sqrt{V(x)-K}dx\right] \tag{4.13}$$

where m^* is the carrier effective mass, $\hbar$ is the reduced Plancks constant, and K is the carrier energy. The potential energy of the tunneling junction can be divided into two regions: source $[V(x) = q\psi_{S1}(x)]$ and channel $[V(x) = q\psi_{S2}(x)]$, as shown in Fig. 4.4. Consider the carrier with energy $-E$ ($E > 0$). In source region, the tunneling probability can be evaluated by using the boundary condition $q\psi_{S1}(x_1) - E + E_G = 0$, and the upper and lower limits are 0 and x_1. The integration result is

$$A(E) = \frac{1}{2}\left[L_1\sqrt{-\alpha+q\psi_S(0)} + \frac{\alpha\ln\left(\sqrt{\beta\alpha}\right)}{\sqrt{\beta}} - \frac{\alpha\ln\left[\beta L_1 + \sqrt{\beta}\sqrt{-\alpha+q\psi_S(0)}\right]}{\sqrt{\beta}}\right] \tag{4.14}$$

where $\alpha = E - E_G$ and $\beta = \frac{q^2 N_S}{2\epsilon_S}$. Note that, α should be smoothed to be $q\psi_S(0)$ to ensure the $T(E)$ of source region at high E is unity, because at high E, the source barrier is infinitesimally thin. In the channel region, the boundary conditions are $q\psi_{S2}(x_2) + E = 0$ for low energy region (i.e., adjacent to source; the upper and lower limits are x_2 and 0) and $q\psi_{S2}(x_0) + E = E_G$ and $q\psi_{S2}(x_2) + E = 0$ for high energy region (i.e., only through channel; the upper and lower limits are x_2 and x_0). These boundary conditions give two results of the integral in (4.13)

$$B(E) = 2\lambda\left[\sqrt{E-q\psi_S(0)} - \sqrt{\gamma}\arctan\left(\frac{\sqrt{E-q\psi_S(0)}}{\sqrt{\gamma}}\right)\right] \tag{4.15}$$

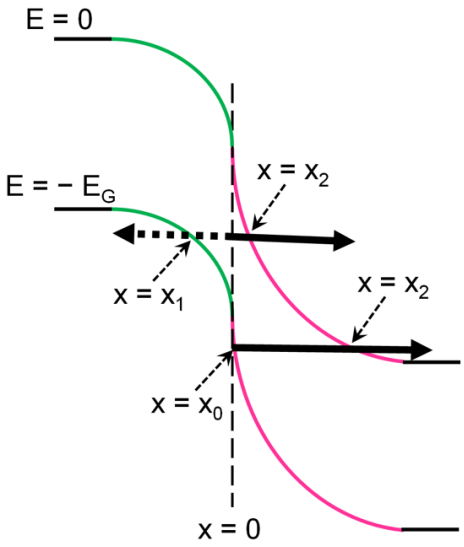


Figure 4.4: Band diagram of the tunneling junction. The green and pink curves represent the bands of source and channel regions, respectively.

and

$$B(E) = 2\lambda \left[\sqrt{E_G} - \sqrt{\gamma} \arctan\left(\frac{\sqrt{E_G}}{\sqrt{\gamma}} \right) \right] \tag{4.16}$$

where $\gamma = qV_{GS} - qV_{FBS} - E$. Equations (4.15) and (4.16) can be combined by smoothing $E - q\psi_S(0)$ to be E_G, when E is large. Therefore, from (4.14)(4.16), the band-to-band tunneling probability can be expressed as

$$T(E) = \exp\left\{ -\frac{2\sqrt{2m^*}}{\hbar} [A(E) + B(E)] \right\} \tag{4.17}$$

Fig. 4.5 shows the tunneling probability as a function of energy using InAs material parameters for different gate biases (V_{GS}), indicating that the tunneling window and probability increase with V_{GS}.

As mentioned earlier, the drawback of WKB approximation is the mismatch of the electron wave function at high electric field region [100, 101], leading to reflection at the boundary. To overcome this issue, a bias-independent electron wave reflectance R is introduced [100] so that (4.1) becomes

$$I = \frac{2q}{h}(1 - R) \int_{E_G}^{q(V_{\text{bis}} + \psi_{CH})} M(E)\, T(E) \left[f_S(E) - f_{CH}(E) \right] dE \tag{4.18}$$

4.2.3 Band Tail Effect

In (4.18), M is the number of the conduction modes, which is associated with the average velocity of the carrier and the density of states (DOS) [107, 108] as mentioned in Chapter 3. Due to the thin body of the device of interest, a 2-D M is adopted and can be written as [107]

$$M(E) = W g_v \frac{\sqrt{2m^* (E - E_G)}}{\pi\hbar} \tag{4.19}$$

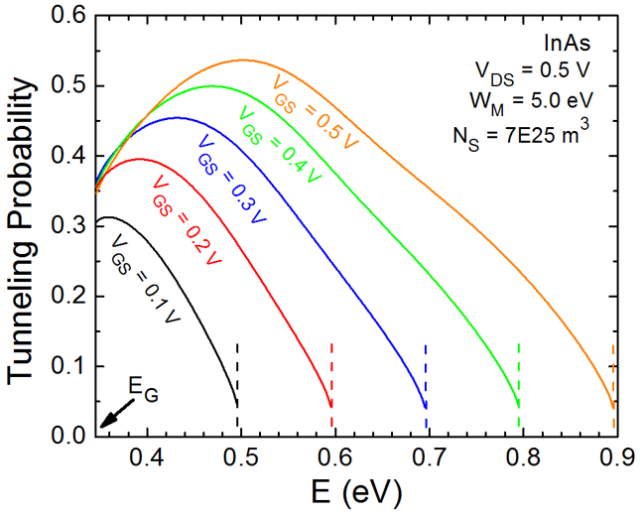


Figure 4.5: Tunneling probability in an InAs TFET. Dashed lines: edges of tunneling windows.

where W is the device width and g_v is the valley degeneracy. From (4.19), it is expected that there is no any conduction mode or DOS in the bandgap so that the lower limit of integration (4.18) is E_G. However, in reality, the band edge is not perfectly sharp [127]. The DOS would extend into the bandgap, called Urbach tail [128], affecting the turn-ON characteristics of TFETs [129]. This effective bandgap states broadening may be caused by the phonon induced tunneling states broadening and doping inhomogeneity. The DOS of the Urbach tail decays exponentially into the bandgap [130]

$$N(E) \propto \exp\left(-\frac{E_G - E}{E_0}\right) \tag{4.20}$$

where E_0 is the Urbach empirical parameter and is comparable to the room temperature thermal energy, which is always determined by optical measurements [129]. Urbach parameter may be doping-dependent [131], but in the model, it is treated as a fitting parameter in order of magnitude of the thermal energy due to nonuniform doping profile in real devices. To incorporate band tail effect into the TFET compact model, (4.19) is multiplied by (4.20). Due to the band states tail, the lower limit of integral in (4.18) becomes $0.5E_G$ (midgap). Although (4.20) exponentially decays in the bandgap, limiting (4.20) not to increase for $E > E_G$ is necessary. This can be numerically solved by replacing the exponential function (4.20) with an exponential function inside a hyperbolic tangent function, ensuring that when E is large (small), the hyperbolic tangent becomes unity (exponential function). Fig. 4.6 shows M as a function of energy. The conduction modes due to Urbach band tail lead to additional current at low gate bias but degrade the subthreshold slope, which will be discussed later.

4.2.4 Gaussian Quadrature Method

Although the Landauer equation captures the physics appropriately, the problem for a compact model is that the integral of such equation is not easy to carry out analytically. Indeed, most of the studies on TFET modeling obtain drain current models

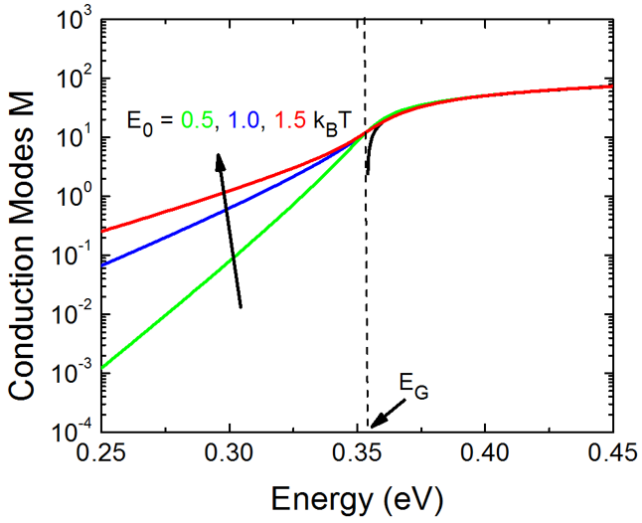


Figure 4.6: Conduction modes as a function of energy using InAs material parameters with $W = 1$ μm. Below E_G, it shows tails for various Urbach parameters E_0.

Table 4.1: Example weight and abscissa for Gaussian quadrature for $N = 8$

i	**Weight** w_i	**Abscissa** x_i
1, 2	0.3626837833783620	±0.1834346424956498
3, 4	0.3137066458778873	±0.5255324099163290
5, 6	0.2223810344533745	±0.7966664774136267
7, 8	0.1012285362903763	±0.9602898564975363

relying on extensive modeling approximations so a close form drain current model can be obtained. Without losing essential physics by making some approximations for the integral, the numerical strategy called Gaussian quadrature method is again introduced here. Gaussian quadrature technique states that an integral of an arbitrary well-behaved function can be simply expressed as a summation by choosing specific weights and abscissa [109]. Mathematically, an integral can be written as

$$\begin{aligned}\int_a^b f(x)\,dx &= \int_{-1}^{1} f\left[\frac{(b-a)\,\zeta + (b+a)}{2}\right]\frac{(b-a)}{2}d\zeta \\ &= \int_{-1}^{1} F(\zeta)\,d\zeta \approx \sum_{i=1}^{N} w_i F(\zeta_i)\end{aligned} \tag{4.21}$$

where a and b are the upper and lower limits of the integral, N is the number of Gaussian points, w_i is the weight, and ζ_i is the abscissa. An example table of w_i and ζ_i for $N = 8$ is shown in Table 4.1. Because Gaussian quadrature involves summation, it is expected that the speed would be slower than a simple analytical equation for the current calculation. However, it is shown that the speed is not a critical issue, since a smaller N gives reasonable accuracy. If function $f(x)$ in (4.21) is continuous, the integration result is also continuous no matter how many N is used. The accuracy, smoothness, and speed of this method will be discussed in the next section.

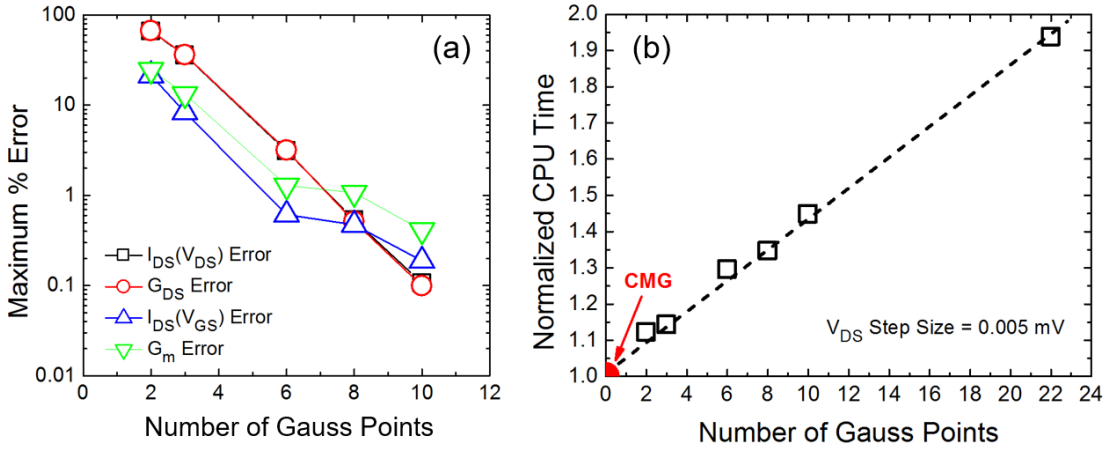


Figure 4.7: (a) Maximum % error of $I_{DS}(V_{DS})$, G_{DS}, $I_{DS}(V_{GS})$, and G_m for different numbers of Gauss points. (b) CPU time versus Gauss points relative to the time used by BSIM-CMG model.

4.3 Validation with Atomistic Simulations

4.3.1 Accuracy, Smoothness, and Speed

For compact model purpose, the accuracy, smoothness, and speed are the key considerations. An optimized N should be obtained since the Gaussian quadrature technique is adopted. Fig. 4.7 (a) shows the maximum % error of $I_{DS}(G_{DS}) - V_{DS}$ and $I_{DS}(G_m) - V_{GS}$ for different N compared to the numerical results, respectively. As expected, more Gauss points reduces error, however, Fig. 4.7 (b) shows that the CPU time increases linearly with N, because the Gaussian quadrature involves summation for N terms. Gaussian quadrature of $N = 8$ keeps the maximum error around 0.5% and only takes 35% more time than the BSIM-CMG model and appears to be satisfactory. The smoothness of the high-order derivative of current is examined due to its importance in the convergence performance of model. Fig. 4.8 (a) and (b) shows that G_{DS3} (the third-order derivative of the drain current) and G_{DS5} (the fifth-order derivative of the drain current) are smooth even if the step size is very small $(= 0.01mV)$ or the number of Gaussian points is only 2. Fig. 4.9 shows the comparison of $I_{DS}(G_{DS}) - V_{DS}$ and $I_{DS}(G_m) - V_{GS}$ curves for $N = 8$ and numerical results, indicating that $N = 8$ is good enough for both speed and accuracy requirements. As a result, $N = 8$ will be taken as the default in this TFET compact model.

4.3.2 Model Validation

Figs. 4.10 (a) and (b) show that the model is in good agreement with the $I_{DS}-V_{GS}$ calculated by atomistic simulation for n-TFETs of several different materials and two geometries [132, 133]. The InAs, $Ge_{0.92}Sn_{0.08}$, and Ge TFETs have cuboid channel with square cross-sectional area of 25 nm^2, while the Si TFET has a cylindrical channel with diameter of 3 nm. The material with small bandgap and low electron effective mass has higher drain current due to higher tunneling probability. Effective bandgap states broadening is included in the model to represent the phonon-induced tunneling states broadening, and doping-induced band tail states. The dashed curves in Fig. 4.10

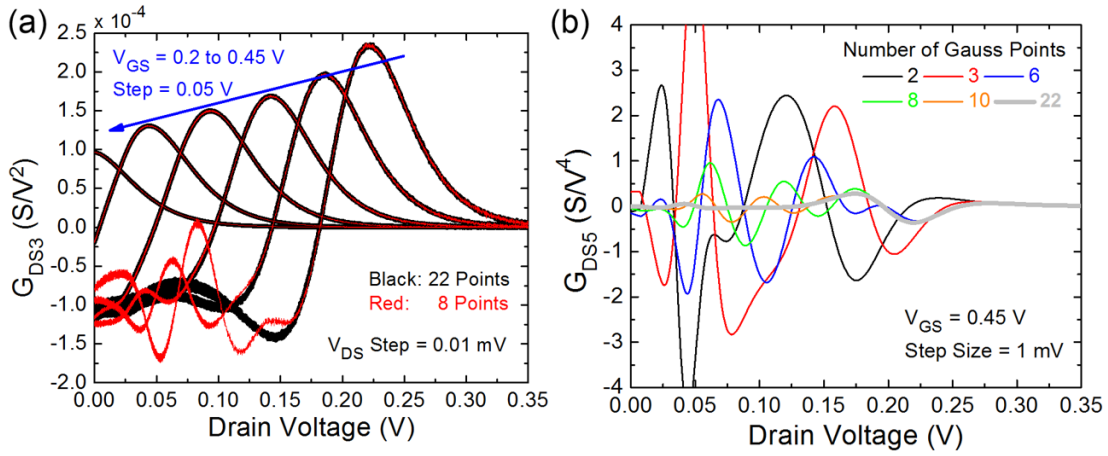


Figure 4.8: (a) G_{DS3} for $N = 8$ and 22 with small V_{DS} Step $= 0.01mV$. (b) G_{DS5} for various N. High-order derivative of current is smooth using Gaussian quadrature.

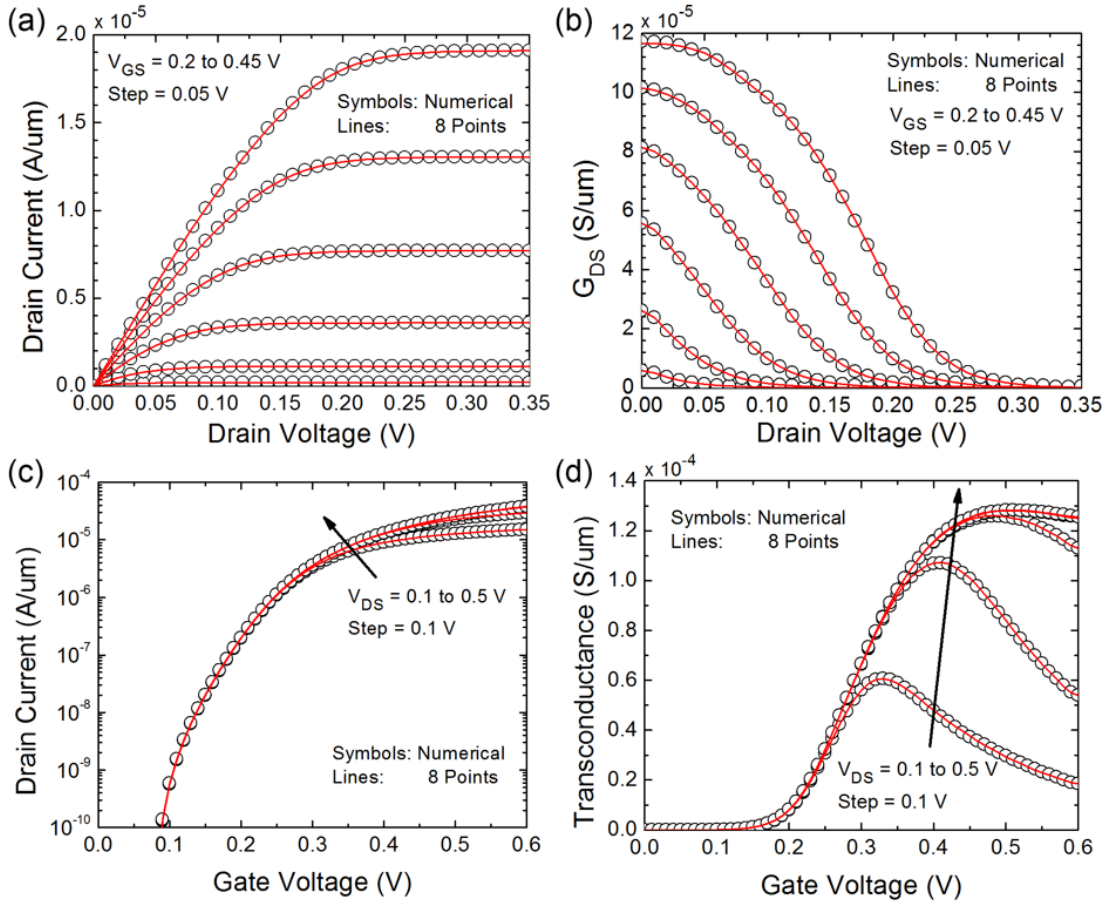


Figure 4.9: (a) $I_{DS} - V_{DS}$, (b) $G_{DS} - V_{DS}$, (c) $I_{DS} - V_{GS}$, and (d) $G_m - V_{GS}$ for $N = 8$ and numerical results. There is little accuracy loss in limiting N to 8.

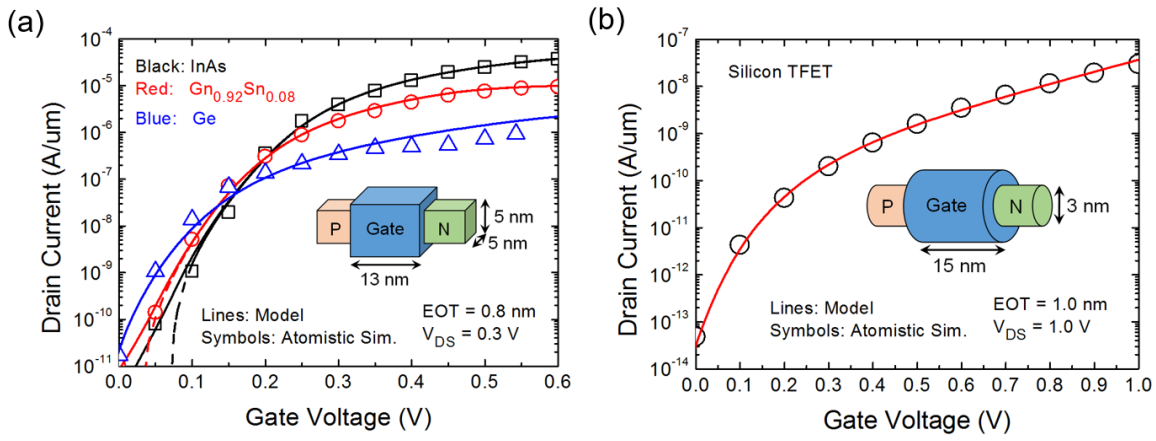


Figure 4.10: (a) $I_{DS} - V_{GS}$ of proposed model and atomistic simulations from [132] using various materials. The fin height (HFIN) and thickness (TFIN) are 5 nm. (Dashed lines: model without band tail.) (b) Silicon TFET $I_{DS} - V_{GS}$ of proposed model and atomistic simulation from [133]. The diameter of the channel is 3 nm.

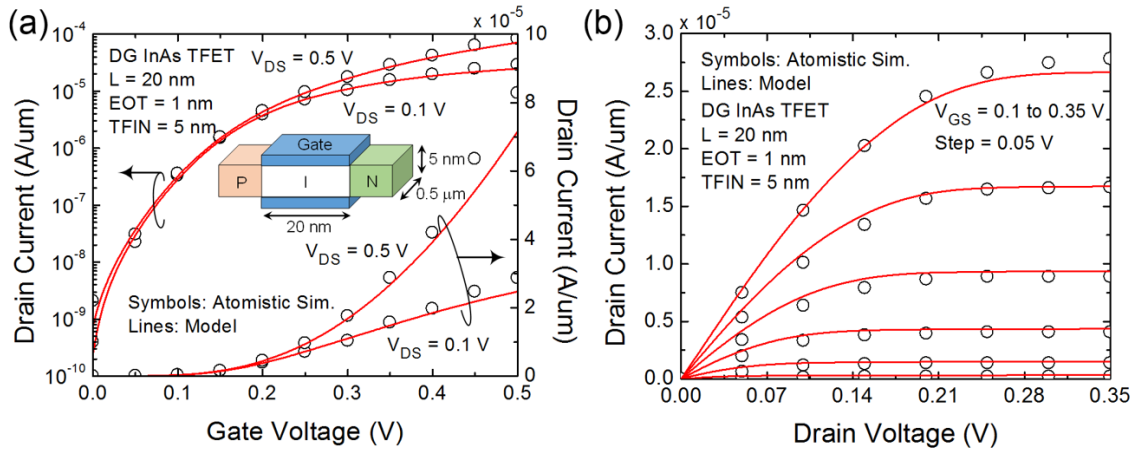


Figure 4.11: (a) $I_{DS} - V_{GS}$ and (b) $I_{DS} - V_{DS}$ of an $L = 20$ nm DG InAs TFET with TFIN = 5 nm and EOT = 1 nm. The model exhibits good agreement with the atomistic simulations [115].

(a) shows that the turn-OFF characteristic will be sharper than the simulation result if this effect is not included. If the bandgap is wider, the current coming from the band tail becomes smaller due to lower tunneling probability so that in Ge and Si TFETs the band tail effect is not observed from the atomistic simulation data. The parameters used in the model are listed in Table 4.2. The material parameters are taken from the literatures [132, 133, 134, 135, 136]. The gate work function W_M , source doping concentration N_S , reflectance R, Urbach parameter E_0, and R_{SD} are the fitting parameters. Fig. 4.11 shows that the model can not only capture $I_{DS} - V_{GS}$ but also $I_{DS} - V_{DS}$ characteristics of a DG InAs TFET with $L = 20$ nm and TFIN = 5nm [115]. In Fig. 4.12, the superlinear output characteristic at low V_{DS} is also demonstrated in $I_{DS} - V_{DS}$ of an InAs TFET using the proposed model by varying the source doping concentration N_S, showing the model capability of capturing the carrier occupancy function and tunneling probability over the tunnel window [137].

Table 4.2: Major parameters used in Fig. 4.10 (a) and (b). The shadowed rows represent the parameters tuned for fitting.

Parameters	**InAs**	**$Ge_{0.92}Sn_{0.08}$**	**Ge**	**Si**
E_G [eV]	0.354	0.52	0.661	1.619
m^*/m_0	0.021	0.12	0.12	0.32
N_V [m^{-3}]	6.66×10^{24}	5.00×10^{24}	5.00×10^{24}	1.80×10^{25}
χ [eV]	4.90	4.00	4.00	4.05
ϵ_r	15.15	16.2	16.2	11.9
$1-R$	0.0444	0.04663	0.03964	0.01114
R_{SD} [$\Omega\mu$m]	500	400	400	100
W_M [eV]	5.13	4.156	4.06	4.08
E_0 [meV]	39.0	44.2	26.0	26.0
N_S [m^{-3}]	7.00×10^{25}	4.00×10^{25}	2.00×10^{25}	9.00×10^{25}

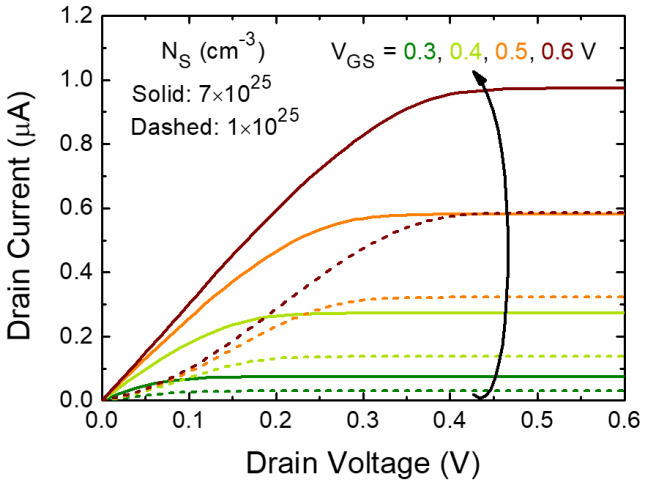


Figure 4.12: The superlinear output characteristic of InAs TFET with the same model and structure parameters as used in Fig. 4.10 (a) except the source doping concentration N_S.

Therefore, the proposed model can accurately describe the current behaviors with different materials, device geometries, and biases.

4.4 Conclusion

A predictive TFET compact model is presented. It can capture the WKB-based band-to-band tunneling probability, the electron wave reflectance, and the band tail effect using Landauer equation. Because there is no close form results of integration of Landauer equation, we adopt the Gaussian quadrature method and show that this numerical technique as an accurate, computationally efficient, and smooth technique eminently suitable for compact model. The predictive nature model is validated by atomistic simulation using different materials and device structures.

Chapter 5

Negative Capacitance FETs: Analysis and Modeling

5.1 Introduction

In Chapter 4, one of the promising transistors for low power applications, tunnel FET, is discussed. In this Chapter, another device, negative capacitance FET (NCFET), will be analyzed, simulated, and modeled. NCFET research is an increasingly growing research topic in device community. NCFET was invented by Salahuddin and Datta in 2008 [138]. NCFET has a structure like traditional MOSFET but with additional ferroelectric (FE) layer sandwiched by gate metal and interfacial oxide, which amplifies the surface potential relative to the gate voltage and bring the subthreshold slope (SS) below 60 mV/dec to reduce the supply voltage. To evaluate NCFET-based circuits, the compact model of NCFET is urgently needed and will be explored in Section 5.2. The detailed mechanism and operation of NCFET will be elaborated subsequently.

5.1.1 NCFET Operation

To explain the subthrehsold characteristics of MOSFET, the capacitor divider is useful. Let's consider a traditional MOSFET with the capacitor network shown in Fig. 5.1. The change in surface potential (ψ_S) in semiconductor (channel) can be expressed as

$$\partial\psi_S = \frac{C_{\mathrm{OX}}}{C_{\mathrm{OX}} + C_S}\partial V_G \tag{5.1}$$

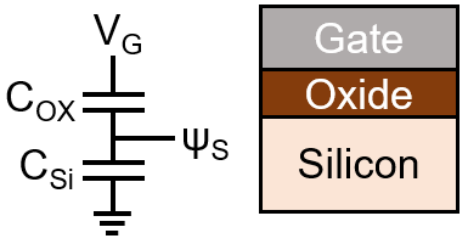


Figure 5.1: Capacitor divider model of MOSFET. C_{OX}: oxide (insulator) capacitance; C_{Si}: semiconductor capacitance; ψ_S: surface potential in semiconductor.

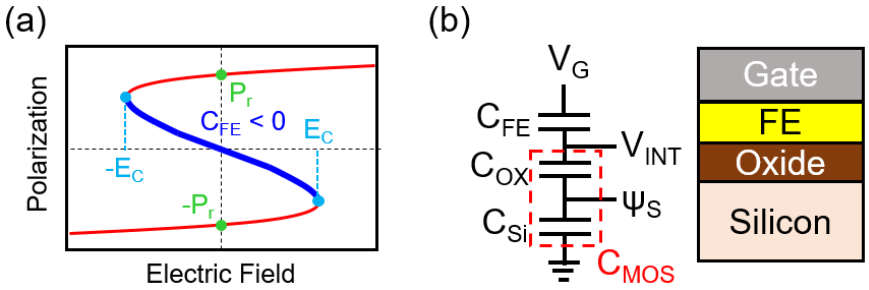


Figure 5.2: (a) Polarization versus electric field in ferroelectric. Blue region has a negative capacitance. (P_r: remnant polarization; E_C: coercive field) (b) capacitor divider model of NCFET.

(5.1) yields the subthreshold current characteristics since

$$I_{DS} \propto \exp\left(\frac{q\psi_S}{k_B T}\right) \tag{5.2}$$

and

$$\begin{aligned} SS &= \left(\frac{\partial \log I_{DS}}{\partial V_G}\right)^{-1} = \left(\frac{\partial \log I_{DS}}{\partial \psi_S}\frac{\partial \psi_S}{\partial V_G}\right)^{-1} = \ln 10 \cdot k_B T \left(1 + \frac{C_S}{C_{\text{OX}}}\right) \\ &= 60 \cdot \left(1 + \frac{C_S}{C_{\text{OX}}}\right) \text{ [mV/dec]} = 60 \cdot m \text{ [mV/dec]} \quad \text{at T} = 300 \text{ K} \end{aligned} \tag{5.3}$$

where

$$m \equiv 1 + \frac{C_S}{C_{\text{OX}}} \tag{5.4}$$

is called the body factor. It can be seen from above equations that in traditional MOSFET the SS cannot be reduced below 60 mV/dec at room temperature because the body factor is always postive, which is called the Boltzmann tyranny. To overcome this fundamental limit, the body factor should be smaller than 1. The ferroelectric layer between the gate and interfacial oxide serves as a voltage booster by providing negative capacitance. The negative capacitance effect in ferroelectric material is coming from the depolarization field from the semiconductor capacitance, which drives the FE into a locally energitically unstable state where $\partial Q/\partial V$ is negative [139] as shown in Fig. 5.2 (a). The FE can be biased at this negative capacitance region if the system is stabilized by adding a dielectric capacitance in series [139, 140]. The Landau-Khalatnikov equation

$$E_{FE} = 2\alpha P + 4\beta P^3 + 6\gamma P^5 - 2g\Delta P + \rho\frac{dP}{dt} \tag{5.5}$$

phenomenologically describes the FE polarization (P) behavior with ferroelectric parameters α, β, γ, g, and ρ, which recently has been experimentally observed in pulsed charge-voltage measurements without hysteresis [141, 142]. The fourth and fifth terms in (5.5) are called the domain interaction[1] coefficient [143] and the viscosity associated

[1] Δ here is the Laplacian.

with polarization-switching dynamics. By applying similar voltage divider model to NCFET [Fig. 5.2 (b)], the internal voltage (V_{INT}) can be written as

$$\partial V_{INT} = \frac{|C_{FE}|}{|C_{FE}| - C_{MOS}} \partial V_G \equiv A_V \cdot \partial V_G \tag{5.6}$$

and thus

$$SS = \frac{\partial V_G}{\partial V_{INT}} \frac{\partial V_{INT}}{\partial \log I_{DS}} = \frac{1}{A_V} \cdot SS_{MOSFET} \tag{5.7}$$

From (5.6) and (5.7), one can immediately see that to achieve sub-60 mV/dec sub-threshold slope, the magnitude of "negative" capacitance from ferroelectric should be as close as possible to C_{MOS} and be greater than C_{MOS} which includes all parasitic and intrinsic capacitances of MOSFET systems. This is called the capacitance matching of NCFET, and this determines how good improvement in SS and $_{DS}$ a NCFET is able to obtain. Several ways to improve the capacitance matching have been proposed [144, 145, 146]. To further boost NCFET-based circuit performances, designing the structure like spacer engineering is crucial and will be discussed in Section 5.4.

The commonly used FE materials are HfO_2-based materials with various dopant species [147] since they are fully compatible to standard CMOS process, where HfO_2 was introduced for high-κ metal gate in 45 nm process [20], without changing device structures. However, the polyscrystallinity nature of HfO_2-based ferroelectric may lead to device performance variations imposed on traditional MOSFET/FinFET variations. The impact of ferroelectric material property variations will be discussed in Section 5.3.

5.2 Compact Model

5.2.1 Review

As mentioned in Chapter 1, the BSIM-CMG is a unified compact model for an arbitrary channel cross-section. The relationship between the mobile charge (Q_m) and the applied terminal voltages (V_D, V_G, V_S, V_B) in a normalized form is [7]

$$v_G - v_O - v_{ch} = -q_m + \ln(-q_m) + \ln\left(\frac{q_t^2}{e^{q_t} - q_t - 1}\right) \tag{5.8}$$

where v_G and v_{ch} are the normalized gate and channel potential, and the definition of quantities in (5.8) and the FE variables are summarized in Table 5.1. Note that the normalized quantities are usually written in lower cases. In the unified model, there are only four model parameters required for the charge and surface potential calculation [15]: gate oxide capacitance (C_{OX}), channel area (A_{CH}), channel doping (N_{CH}), and effective channel width (W_{eff}). From (5.5), the normalized ferroelectric voltage (v_{FE}) is

$$v_{FE} = a_0 p_{FE} + b_0 p_{FE}^3 + c_0 p_{FE}^5 \tag{5.9}$$

where p_{FE} is the normalized polarization (P) in FE obtained by iteratively solving the displacement field $D = \epsilon_{FE} E_{FE} + P = Q_G = -Q_{ch}$ with the channel charge (Q_{ch}) including the mobile charge, depletion charge, and parasitic charge and with the background permittivity of FE which is assumed to be 16 for Hf-based FE in

Table 5.1: BSIM-CMG unified model and FE model variables

Variable	Definition
v_G, v_{ch}	$\frac{V_G}{V_t}$, $\frac{V_{ch}}{V_t}$
q_m, q_{dep}	$\frac{Q_m}{V_t C_{OX}}$, $-\frac{qN_{CH}A_{CH}}{V_t C_{OX}}$
v_O	$v_{FB} - q_{dep} - \ln\left(\frac{2qn_i^2 A_{CH}}{V_t C_{OX} N_{CH}}\right)$
q_t	$(q_m + q_{dep})\, r_N$
r_N	$\frac{A_{CH}C_{OX}}{\epsilon_{CH}W_{eff}^2}$
a_0	$\frac{2\alpha t_{FE}C_{OX}}{W_{eff}}$
b_0	$\frac{4\beta t_{FE}\left(V_t C_{OX}/W_{eff}\right)^3}{V_t}$
c_0	$\frac{6\gamma t_{FE}\left(V_t C_{OX}/W_{eff}\right)^5}{V_t}$

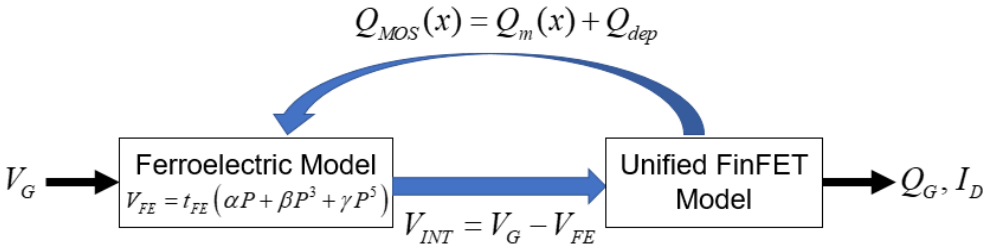


Figure 5.3: Compact modeling flow of NCFET.

simulations later on. Note that the fourth and fifth terms in (5.5) are ignored now since g is relatively small reported in some FE materials [140] and a small ρ is measured [148]. The compact modeling flow of NCFET is illustrated in Fig. 5.3. Since the polarization is non-uniform along the channel due to the complicated electric fields, the drain current involves integration from the source to drain to capture the distributed nature. This integration can be achieved by adopting the previously mentioned numerical technique - Gaussian quadrature method and by replacing v_G with $v_G - v_{FE}$ at each channel position

$$i_{DS} = \int_0^{v_{DS}} q_m dv_{ch} \approx \sum_{i=1}^{N} q_m\left(v_{ch,i}\right) w_i \tag{5.10}$$

where $v_{ch,i} = \left(v_D - v_S\right)\left(\zeta_i + 1\right)/2 + v_S$, N is the number of Gauss points, ζ_i and w_i are the abscissas and weights as mentioned in previous Chapters. The proposed NCFET compact model has been validated with the experimental data as shwon in Fig. 5.4.

5.2.2 Short Channel Effects

So far, the core model of NCFET has been discussed. The assumption of the unified FinFET core model is that the device has a long channel with validation of gradual channel approximation. Normally, the short channel effect is modeled by obtaining the effective gate voltage. However, since the NCFET operation involves capacitance matching, the charge calculation with polarization becomes crucial. The inclusion of

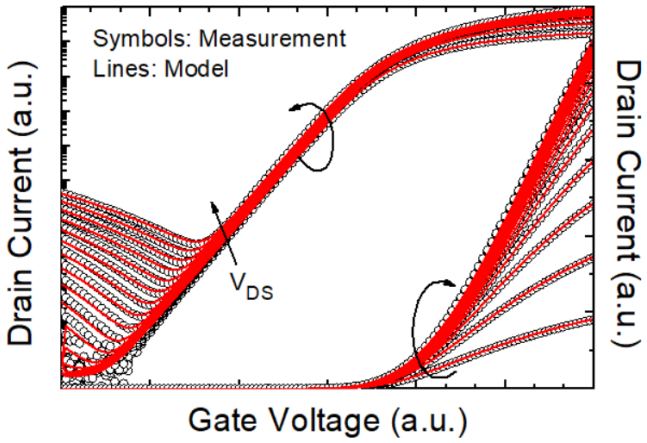


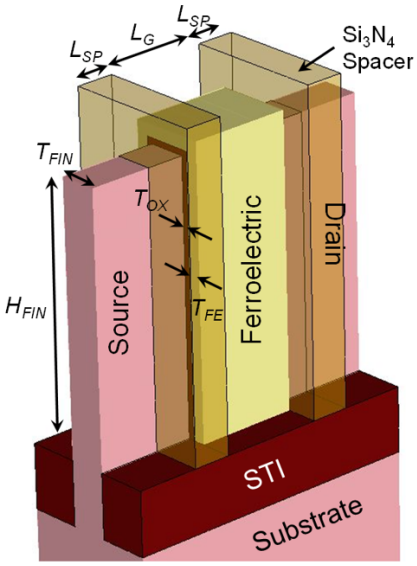
Figure 5.4: Model validation with an experimental NC-FinFET at 20 different V_{DS}. (Data from [149])

inner fringing field appearing in short channel device is essential. The effects of the fringing field and parasitic capacitances are discussed in [145, 150]; however, a floating metal gate between the FE layer and the interfacial oxide is used in these analyses, which changes the electrical boundary conditions [140] and is impractical in commercial FETs. Moreover, although Pahwa *et al.* [151] and Duarte *et al.* [152] recently reported that the fringing fields impact the subthreshold characteristics of NCFETs, the detailed analysis of the FE properties across the channel and its impact on the ON current remain absent. Therefore, the impact of inner fringing field on NCFET and the corresponding compact model will be explored in this section.

Simulation and Analysis

Sentaurus TCAD is used to simulate the p-type silicon body NC-FinFET [2], which solves Poissons equation, the continuity equation, and the Landau equation simultaneously. The FE polarization is assumed to be perpendicular to the fin surface. The domain interaction term [143] in the Landau equation is included with a coefficient of 10^{-8} m^3/F. This does not affect the S-shaped curve predicted by the Landau equation. The NC-FinFET is designed based on the low-power 8-/7-nm technology node of the IEEE International Roadmap for Devices and Systems [153]. The doping profile in the S/D extension region (with ≈ 0.5 nm underlap) and the metal gate work function are tuned to match targeted currents (I_{OFF} = 100 pA/μm and I_{ON} = 637 μA/μm) for the baseline FinFET. We vary the remnant polarization (P_r) of the FE layer, which is achieved in practice by varying doping conditions in the hafnia layer [147, 154], without changing the FinFET structure. The FE layer is kept thin (= 2 nm) to ensure the compatibility with device scaling and hysteresis-free operation for P_r used in this paper. The simulated device structure and dimensions are shown and listed in Fig. 5.5. Note that there is no internal metal gate between the FE and interlayer dielectric, which is commonly used in many experimental and simulated devices and is prone to hysteresis [140]. To examine the impact of the channel length scaling, several gate lengths (L_G = 16, 30, 80, and 200 nm) are explored based on the structure shown in Fig. 5.5.

Fig. 5.6 (a) shows the transfer characteristics of the NC-FinFET at L_G = 16,



Parameter	Value
Gate Length (L_G)	16, 30, 80, 200 nm
Spacer Length (L_{SP})	6 nm
Fin Height (H_{FIN})	42 nm
Fin Thickness (T_{FIN})	6 nm
Interfacial EOT (T_{OX})	0.9 nm
Channel Doping (N_{CH})	10^{18} cm^{-3}
Source/Drain Doping (N_{SD})	2×10^{20} cm^{-3}
FE Thickness (T_{FE})	2 nm
Coercive Field (E_C)	1 MV/cm
Remnant Polarization (P_r)	12, 15, 18, 20 μC/cm^2
Supply Voltage (V_{DD})	0.7 V

Figure 5.5: Simulated device structure and the important device parameters.

30, 80, and 200 nm for various values of remnant polarization (P_r). Fig. 5.6 (b)−(d) summarizes the subthreshold slope (SS), OFF current (I_{OFF} at $V_{GS} = 0V$ and $V_{DS} = 0.7V$), and ON current (I_{ON} at $V_{GS} = 0.7V$ and $V_{DS} = 0.7V$). It is found that the OFF current decreases with decreasing P_r at short-channel lengths, while it increases with decreasing P_r at long-channel lengths [see Fig. 5.6 (c)]. In contrast, SS and ON current are improved by decreasing P_r at all lengths [see Fig. 5.6 (b) and (d)]. The increase in the ON current can be attributed to amplification from the FE layer due to the support of depletion and strong inversion charges [7]. Note that due to the short-channel effect (SCE), the NC-FinFETs with $L_G < 30$ nm do not show sub-60 mV/dec SS, as can be seen in Fig. 5.6 (b). Fig. 5.7 (a) shows the polarization versus FE voltage (V_{FE}) curve (S curve). In the strong inversion, the device is biased in the second quadrant due to negative channel inversion charges (positive gate charges Q_G) throughout the channel, leading to an amplified gate voltage ($V_{GS,\text{amp}} = V_{GS} - V_{FE}$). In the subthreshold regime, Q_G is induced from (1), the depletion charges (positive Q_G), and (2), fringing field-induced charges (negative Q_G). There are outer and inner fringing fields in the MOSFET structure. The outer fringing fields through the spacer can only change the gate charges at the channel edges, unlike the inner fields which are able to penetrate into the channel [Fig. 5.7 (b)] [155]. The device is now biased in either the fourth or the second quadrants, depending on the polarity of gate charges and position along the channel. Fig. 5.7 (c) shows the electric fields along the gate-stack at the middle of the channel with $L_G = 16$ and 200 nm and $P_r = 12$ μC/cm^2 at both ON and OFF states. Different signs of the electric field in the OFF state in short- and long-channel devices can be seen, indicating different operation points on the S curve. Fig. 5.8 shows the spatial distribution of the polarization in the FE (OFF state) along the channel direction for different gate lengths. It can be seen that at very short-channel lengths (for example, $L_G = 16$ nm), the polarization is negative throughout the FE due to the inner fringing field-induced Q_G, which means that the device is biased in the fourth quadrant [point C shown in Fig 5.7 (a)]. The channel potential is, thus, reduced, and the OFF current is suppressed [see Fig. 5.6 (c)]. Note

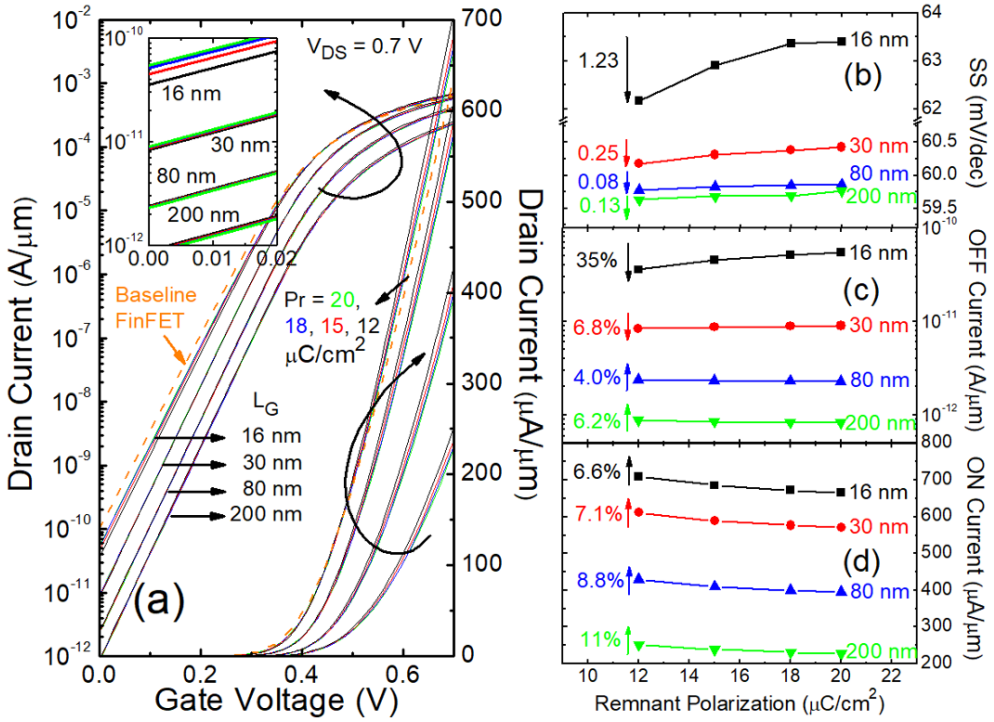


Figure 5.6: (a) Drain current versus gate voltage at L_G = 16, 30, 80, and 200 nm with P_r = 12, 15, 18, and 20 μC/cm^2 at $V_{DS} = 0.7V$. The impact of L_G and P_r on (b) SS, (c) OFF current, and (d) ON current. Note that the SS, I_{OFF}, and I_{ON} of the baseline FinFET at L_G = 16 nm [dashed lines in (a)] are 65.8 mV/dec, 100 pA/μm, and 637 μA/μm, respectively.

that a smaller P_r [156] improves capacitance matching ($C_{FE} \propto P_r$). However, as the gate length is increased, the polarization at the middle of the channel gradually becomes positive. These positive charges are induced from the negative depletion charges, as mentioned earlier. This can be confirmed by the fact that the polarization at the center of the channel is the same for L_G = 80 and 200 nm, where the inner fringing fields cannot have influence. If the channel is lightly doped or undoped, the polarization at the center of the channel would become negligible, and the subthreshold characteristics would be insensitive to P_r in long-channel devices. Furthermore, due to better capacitance matching with smaller P_r, the polarization at the center of the channel is boosted to be more positive. Therefore, the channel potential is raised [biased at point B shown in Fig. 5.7 (a)], and the OFF current gets worse with smaller P_r, as shown in Fig. 5.6 (c). Note that at the edges (i.e., the source and drain sides), the polarizations for L_G = 80 and 200 nm are within an order of magnitude, implying that the outer fringing fields remain the same but cannot affect the potential barrier of the channel. The difference between the edge polarizations of short- and long-channel devices results from the inner fringing field, where more field lines terminate at the gate (silicon channel) in short (long)-channel devices.

If there is an internal metal gate (between the FE and interfacial oxide layers) which creates an equipotential surface, the outer fringing field-induced gate charges could be coupled into the central part of the channel through the internal metal [150], and one will overestimate the effect of drain coupling [157] compared to the case without the internal gate. Furthermore, it has been reported that an internal metal

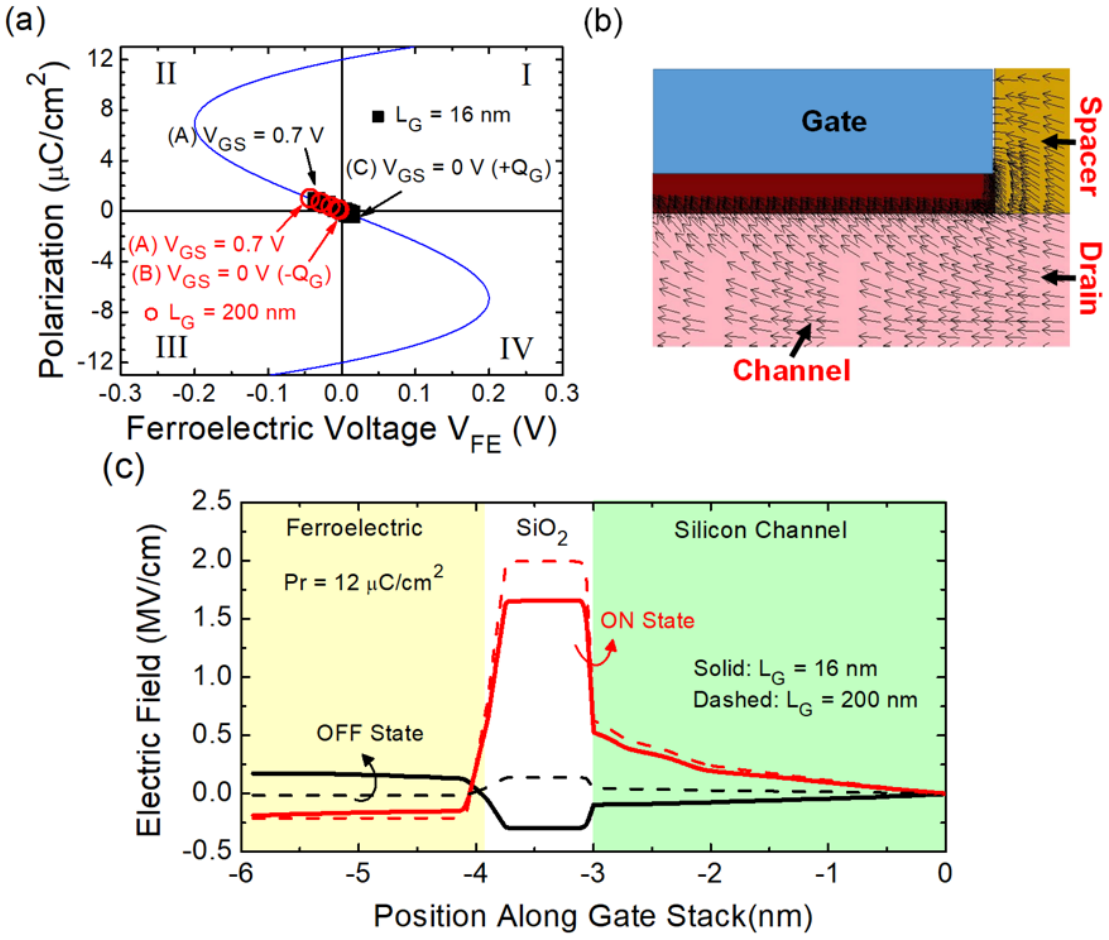


Figure 5.7: (a) Polarization ($P_r = 12\ \mu$C/cm^2) versus FE voltage of $L_G = 16$ and 200 nm at $V_{DS} = 0.7V$. (Dots are extracted from the simulations). (b) Simulated electric field lines near drain in MOSFET at $V_{GS} = 0V$ and $V_{DS} = 0.7V$. The spatial distribution of the inner fringing field (penetration into the channel) is evident. (c) Electric field at the middle of the channel along the gate-stack of the OFF and ON states in $L_G = 16$ and 200 nm NC-FinFET.

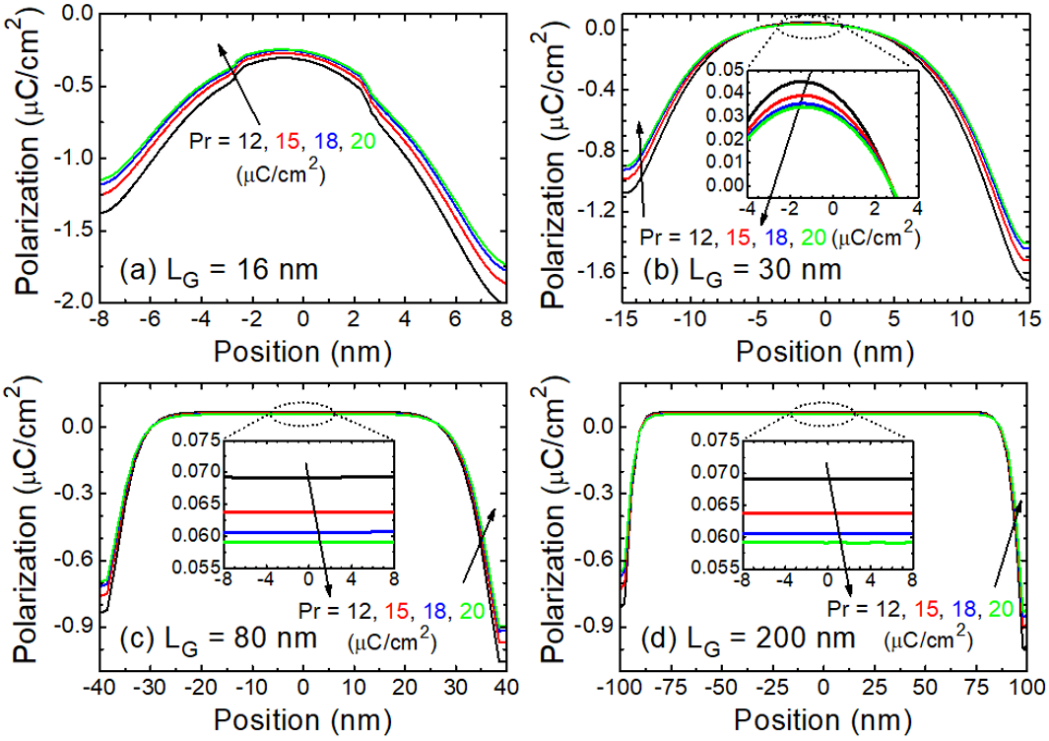


Figure 5.8: Polarization ($V_{GS} = 0V$ and $V_{DS} = 0.7V$) inside the FE along the channel [source (left) and drain (right)] at (a) $L_G = 16$ nm, (b) $L_G = 30$ nm, (c) $L_G = 80$ nm, and (d) $L_G = 200$ nm. Insets: zoomed-in at the central channel

gate leads to hysteresis [140], so we do not include it here.

The ON current of long-channel devices improves with decreasing P_r. Due to the effect of depletion charges, the long-channel device is operated in the second quadrant of the S curve [point B shown in Fig. 5.7 (a) and see Fig. 5.7 (c)] in the subthreshold region. As V_GS is ramped up, the bias point moves deeper (more positive polarization and more negative V_{FE}) into the second quadrant, and thus V_{GS} (and thus I_{ON}) is amplified [see Figs. 5.7 (c) and 5.9 (a)]. In contrast, the bias point of the short-channel device starts from the fourth quadrant due to the inner fringing field-induced charges. Even in the strong inversion, the short-channel device is not biased in the second quadrant as deeply as the long-channel device, so the increase in I_{ON} by reducing P_r is not as much as in the long-channel case.

Fig. 5.10 summarizes the threshold voltages (V_{TH}) and drain-induced barrier lowering (DIBL) for various P_r and L_G. The threshold voltage is defined as the voltage when the drain current is equal to 100nA$\times W/L$. As shown in Fig. 5.9 (b), due to the inner fringing field induced by V_{DS}, the change of V_{TH} at $V_{DS} = 0.7V$ ($V_{TH,sat}$) with P_r of $L_G = 16$ nm is more than that at $V_{DS} = 0.05V$ ($V_{TH,lin}$), so DIBL is suppressed more with smaller P_r. However, at $L_G = 200$ nm, both $V_{TH,sat}$ and $V_{TH,lin}$ decrease with decreasing P_r, which can be attributed to the effect of depletion charges. Furthermore, because of the reduction in the impact of inner fringing fields in the long-channel device, DIBL changes only slightly with P_r [see Fig. 5.9 (b)]. Therefore, the DIBL of NC-FinFETs is improved compared with the baseline FinFETs, especially in short-channel devices [Fig. 5.10 (b)]. NC-FinFETs, thus, show a superior output resistance, with an improvement of 57.8% (Fig. 5.11), indicating the potential for

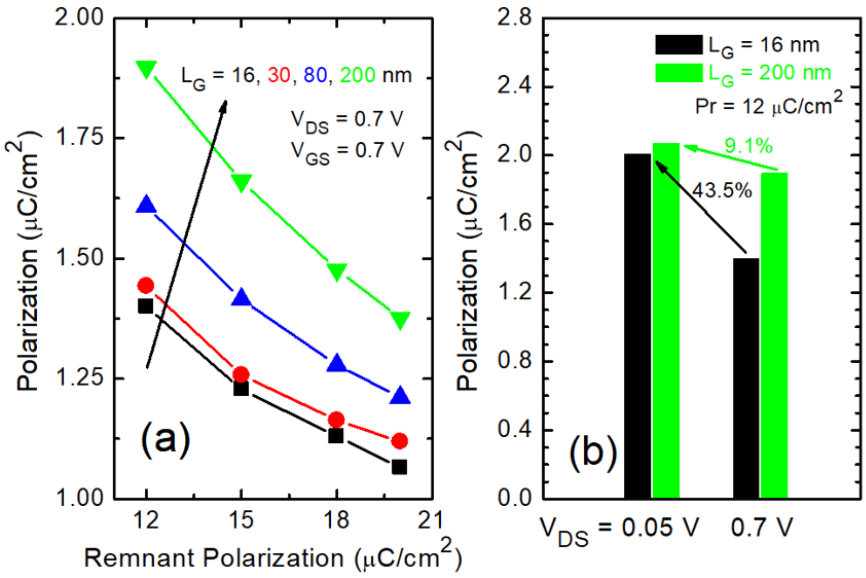


Figure 5.9: (a) Polarization versus P_r at the ON state. The polarization is measured at the carrier injection point (the highest electron density). (b) Comparison of polarization at different V_{DS} and at $V_{GS} = 0.7V$ for $L_G = 16$ and 200 nm.

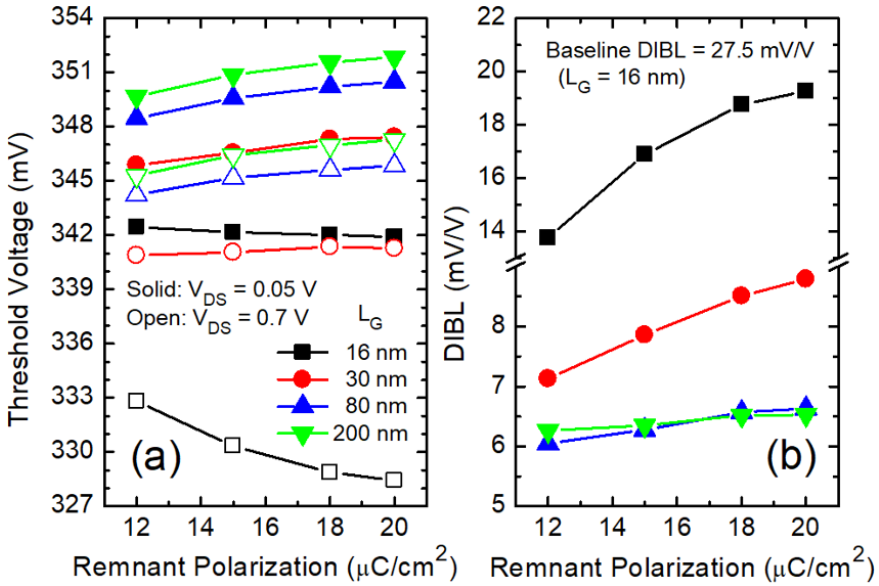


Figure 5.10: (a) Threshold voltage at $V_{DS} = 0.05V$ and $0.7V$ versus P_r at various L_G. (b) DIBL versus P_r at various L_G. In comparison with baseline FinFET, NC-FinFET ($E_C = 1$ MV/cm and $T_{FE} = 2$ nm) exhibits better DIBL.

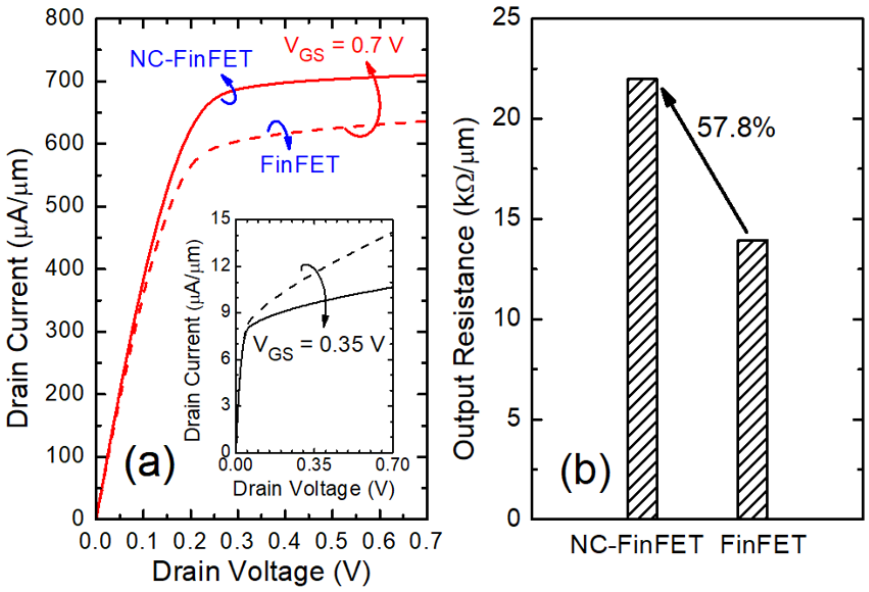


Figure 5.11: (a) $I_{DS} - V_{DS}$ characteristics of NC-FinFET with $P_r = 12\ \mu\text{C/cm}^2$ (solid lines) and baseline FinFET (dashed lines) at $V_{GS} = 0.35$ and $0.7V$. (b) Output resistance of NC-FinFET and FinFET.

better analog circuit performance [158].

Compact Modeling

The BSIM-compatible compact model for NC-FinFETs was proposed in [7]. However, the long-channel potential approximation ignores the effect of inner fringing fields on NC-FinFETs, and the P_r scaling mentioned in previous section is not included. To properly take the inner fringing field effect into account, the amount of change in the gate charges in the subthreshold due to the SCE is modeled using the quasi-2-D model [103]

$$Q_G(x) = -C_{\text{OX}}\left[V_{\text{SL}} + (V_{\text{bi}} - V_{\text{SL}})\frac{\sinh\frac{L_G - x}{\lambda}}{\sinh\frac{L_G}{\lambda}} + (V_{\text{bi}} + V_{DS} - V_{\text{SL}})\frac{\sinh\frac{x}{\lambda}}{\sinh\frac{L_G}{\lambda}}\right] \tag{5.11}$$

where C_{OX} is the oxide capacitance, V_{bi} is the built-in potential between the source- and drain-body junctions, x is the position along the channel, and λ is the characteristic length which depends on the geometry of device [104], and V_{SL} is defined

$$V_{\text{SL}} = V_{GS} - V_{FB} - \frac{qN_{CH}}{\epsilon_{Si}}\lambda^2 \tag{5.12}$$

where V_{FB} is the flat band voltage and N_{CH} is the doping concentration in the channel. The second and third terms in the right-hand side of (5.11) represent the short-channel charges (position dependent), while the first term represents the long-channel charges (position independent). Fig. 5.12 shows the model validation with the TCAD simulation of double-gate MOSFETs with $L_G = 16$ and 30 nm. Since in (5.10), the Gaussian integration method is adopted to overcome the difficulty of obtaining the analytical current expression, the proposed position-dependent short-channel charges

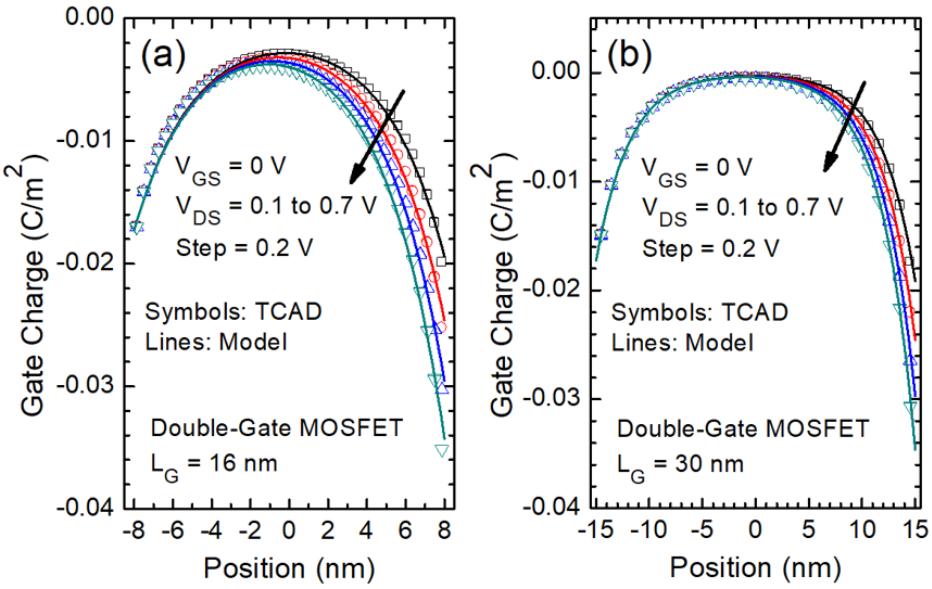


Figure 5.12: Gate charge of double-gate MOSFETs with (a) $L_G = 16$ nm and (b) $L_G = 30$ nm. The effective oxide thickness is 0.9 nm, and the silicon body thickness is 6 nm.

can be integrated into the current calculation through (5.9). (5.10) would yield the drain current with the spatial dependence of the contribution of the inner fringing fields. Fig. 13 (a) shows the channel length dependence of the NC-FinFETs with P_r = 20 μC/cm^2 at $V_{DS} = 0.7$ and $0.05V$, which validates the proposed model with the TCAD simulations. Note that only one parameter set is needed for multiple channel lengths. After extracting the parameters for multiple lengths, the model is further validated for P_r scaling, which involves holding all parameters constant except P_r. Fig. 13 (b) shows the distinct characteristics of scaling P_r in NC-FinFETs at $L_G = 16$ and 200 nm compared to the baseline FinFET at $L_G = 16$ nm, indicating the predictive nature of the compact model.

5.3 Polycrystallinity Variability

Doped HfO_2 [147] is the most promising ferroelectric material for this application due to its compatibility with modern CMOS process. However, the polycrystalline material may contain multiple phases including dielectric (DE) phases [147, 159]. Even the ferroelectric phase grains may have variance in ferroelectric characteristics such as the remnant polarization, P_r, due to local strain variance. In this section, the influence of polycrystallinity variations on NCFET will be comprehensively discussed based on Monte Carlo TCAD simulations.

5.3.1 Simulation Methodology

The NC-FinFETs are simulated using Sentaurus TCAD tool, which simultaneously solves Poisson's equation with Landua-Khalatnikov (LK) equation. The baseline FinFET is designed based on the low power (LP) 8/7 nm technology node of the International Roadmap for Devices and Systems (IRDS) [153]. The remnant polarization

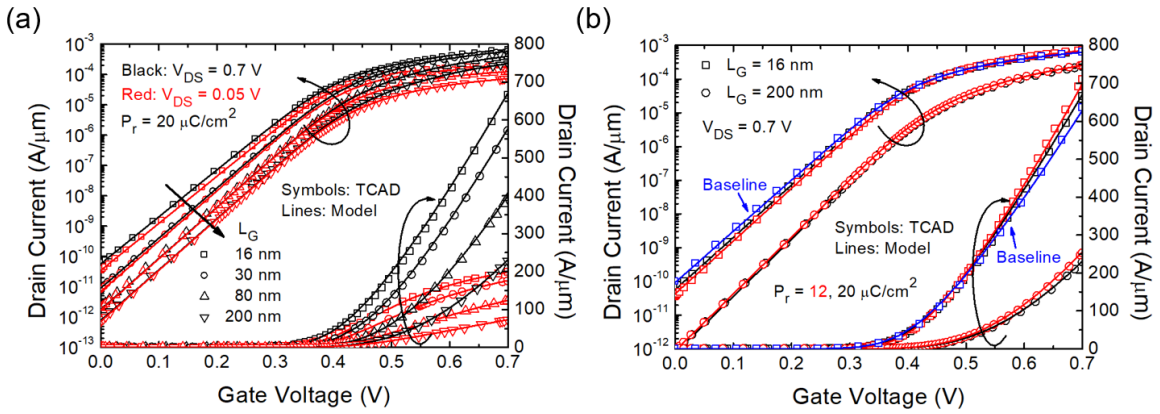


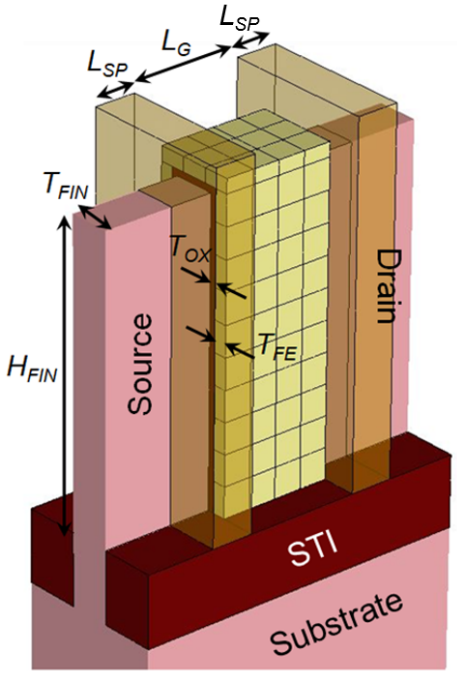
Figure 5.13: (a) $I_{DS} - V_{GS}$ characteristics of NC-FinFETs with various gate lengths at $V_{DS} = 0.05$ and $0.7V$. The proposed model accurately captures the length scaling. (b) $I_{DS} - V_{GS}$ characteristics of NC-FinFETs at $L_G = 16$ and 200 nm with $P_r = 12$ and 20 μC/cm^2. The P_r scaling is successfully predicted since the inner fringing field-induced charges are included. The model also captures the baseline FinFET if setting T_{FE} to be zero.

(P_r) of the doped HfO_2 is an assignable parameter [147]. To support fin spacing scaling, the FE thickness is kept as thin as 2 nm. We assume columnar grains and consider three variation sources: (1) P_r variation among the FE grains, (2) DE to FE grain ratio variation, and (3) grain size. Monte Carlo simulations are carried out to randomly assign the P_r (E_C variation is not considered for simplicity) and whether a grain is DE or FE within the confines of chosen μ and σ (means and sigmas). The grain size is kept constant for simplicity, but the effect of grain size on FinFET sensitivity to the material variations is separately studied. Fig. 1 shows the simulated device structure and parameters. The channel area is divided into 45 to 92 tiles (grains) depending on the assumed grain size. Note that there is no internal gate in NC-FinFET in this work [140].

5.3.2 Results and Discussion

Ferroelectric P_r Variations

Due to local material or strain variations, different FE grains may have varying properties. Fig. 5.15 (a) shows $I_{DS}-V_{GS}$ of the baseline FinFET and NC-FinFET with $\mu P_r = 20$ and $\sigma P_r = 5$ μC/cm^2 (grain size = 6.2 nm $\times$ 6.2 nm). The FE variation leads to spatially non-uniform FE amplification and thus nonuniform current flow [Fig. 5.15 (b)]. The channel under the grain with smaller P_r (with better capacitance matching to the channel capacitance [156]) carries more current [Fig. 5.15 (b)]. Fig. 5.16 (a) and (b) exhibit the distributions of I_{ON}, I_{OFF}, V_{TH}, and SS. V_{TH} is defined as the V_{GS} where the current (I_{TH}) is equal to 100nA$\times W/L_G$ (W = 2HFIN + TFIN), and SS is the average subthreshold slope from I_{OFF} to $I_{TH}/10$. As expected, larger P_r variation results in larger device variation. Interestingly, increasing P_r variation also causes the mean values of I_{ON}, I_{OFF}, V_{TH}, and SS to shift in Fig. 5.16 (a) although the mean P_r is kept constant. This can be understood from the fact that the voltage amplification is a nonlinear function of Pr [156] [Fig. 5.16 (c)]. Since the grains with smaller P_r provides more voltage amplification advantage than the disadvantage due



Parameter	Value
Gate Length (L_G)	16 nm
Spacer Length (L_{SP})	6 nm
Fin Height (H_{FIN})	42 nm
Fin Thickness (T_{FIN})	6 nm
Interfacial EOT (T_{OX})	0.9 nm
FE Thickness (T_{FE})	2 nm
Remnant Polarization (P_r)	15, 20 μC/cm^2
Coercive Field (E_C)	1 MV/cm
Grain Size	4^2, 5.3^2, 6.2^2 nm^2
Supply Voltage (V_{DD})	0.7 V
ON current (I_{ON})	637 μA/μm
OFF Current (I_{OFF})	100 pA/μm
Subthreshold Slope (SS)	65.8 mV/dec
Threshold Voltage (V_{TH})	321.1 mV

Figure 5.14: Simulated 8/7 nm NC-FinFET structure and device parameters. The shadowed rows present the baseline FinFET performance.

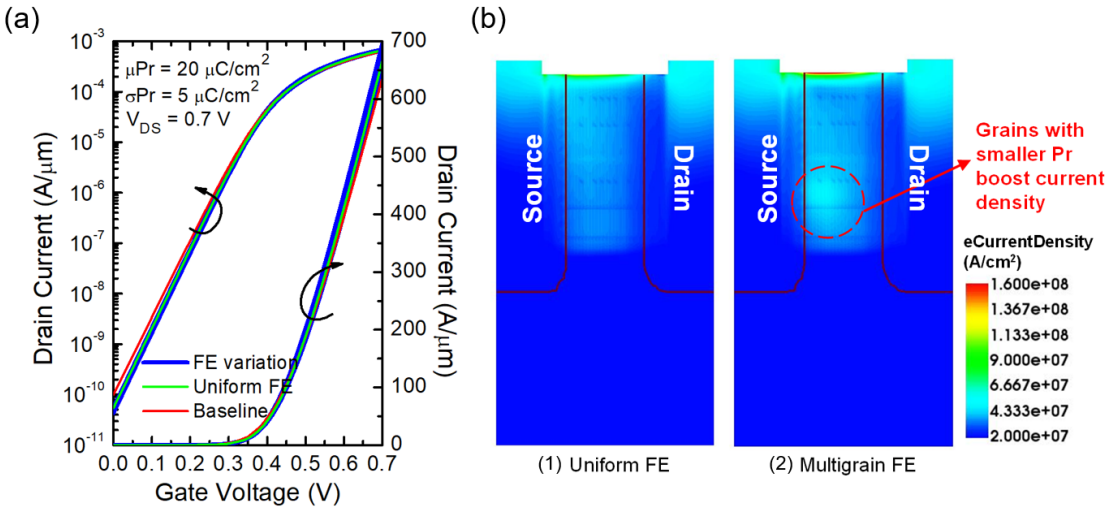


Figure 5.15: (a) $I_{DS} - V_{GS}$ of baseline FinFET and NC-FinFET with P_r variation only. (b) Current density in NC-FinFETs with (1) uniform FE with 20 μC/cm^2 P_r and (2) FE variation with $\sigma P_r = 5$ μC/cm^2 (maximum I_{ON} case shown).

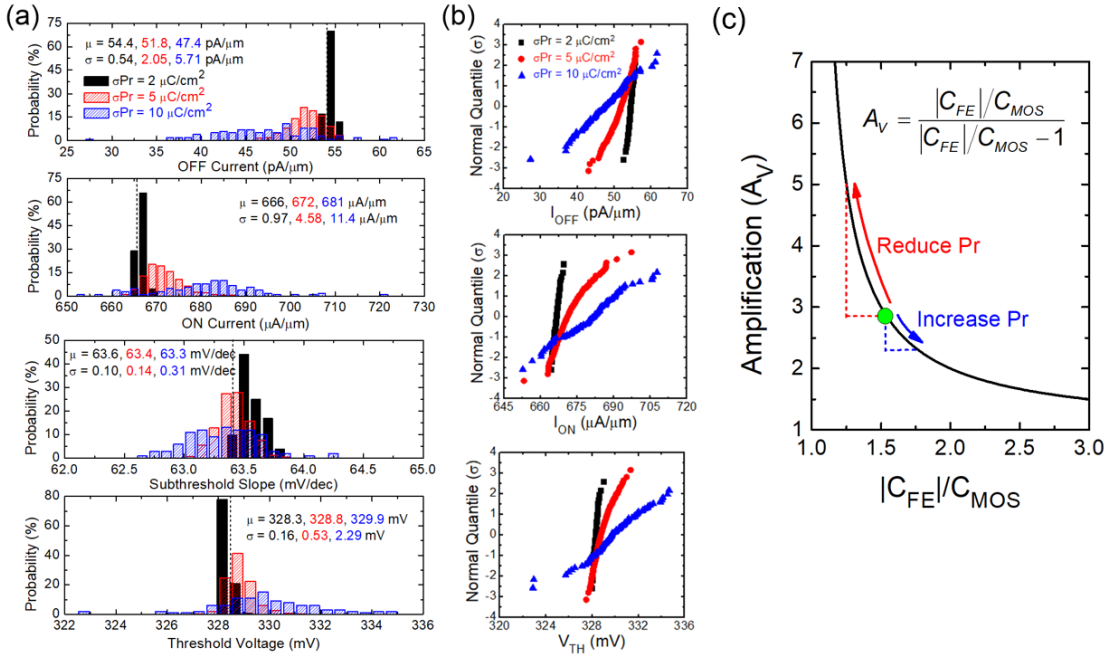


Figure 5.16: (a) Distributions of I_{OFF}, I_{ON}, V_{TH}, and SS in NC-FinFET with $\mu P_r = 20$ μC/cm^2, $\sigma P_r = 2$, 5, and 10 μC/cm^2. The dashed lines represent NCFinFET with uniform FE ($\mu P_r = 20$ μC/cm^2). (b) Normal quantile plots. (c) Voltage amplification (A_V) as a function of $|C_{FE}|/C_{MOS}$ (P_r), showing nonlinearity of A_V when changing P_r.

to the grains with larger P_r. P_r variations cause variations in I_{ON} and other device properties and shifts of their mean values. Based on Fig. 5.16, we suggest that P_r variation in the FE should be kept below 25% (e.g. $\mu P_r = 20$ and $\sigma P_r = 5$ μC/cm^2) to ensure that the device performance variations are negligible compared with other variation sources, e.g., random dopant fluctuation (RDF), in FinFETs [160, 161].

Fig. 5.17 shows the scatter plot of I_{ON} and I_{OFF} versus V_{TH}. It is found that the I_{OFF} and V_{TH} are strongly negatively correlated, while I_{ON} and V_{TH} are weakly positively correlated. The sign change of correlation can be attributed to the amplification from FE due to the inner fringing field at OFF state and inversion charges at ON state. However, to explain the strength of correlation of I_{ON} and I_{OFF} with V_{TH}, the spatial distribution of P_r in different grains should be carefully considered. Fig. 5.18 shows the $I_{DS} - V_{GS}$ and the corresponding conduction band profiles and mobile charge (electron) density along the channel for three different P_r distributions with the same mean P_r. At OFF state, the grain with smaller P_r located at either source or drain side will give suppressed OFF current due to the inner fringing field (but the drain side has more impact because of V_{DS}). In contrast, at ON state, only the source grain controls the electron density at the carrier injection point and thus I_{ON} is mainly related to the source grain FE property. However, the threshold voltage is more related to the OFF state because the FE is still biased near the origin of S-shaped curve of Landau-Khalatnikov equation. Fig. 5.19 further illustrates the importance of drain and source grains at OFF [Fig. 5.19 (a)] and ON [Fig. 5.19 (b)] states. After strong inversion layer appears, the source grain dominates and thus the impact of V_{DS} (output conductance) becomes similar among three P_r spatial configurations. Therefore, the ON current is weakly correlated to the threshold voltage, which is analogous

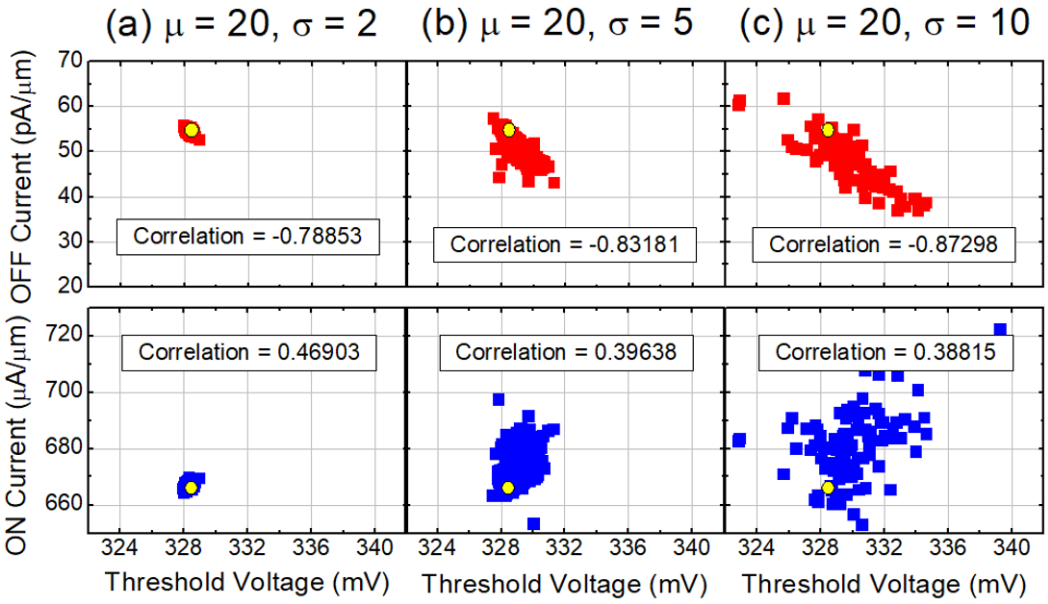


Figure 5.17: Scatter plot of I_{OFF} and I_{ON} versus V_{TH} with various μP_r and σP_r. The yellow dots represent NC-FinFET with uniform FE.

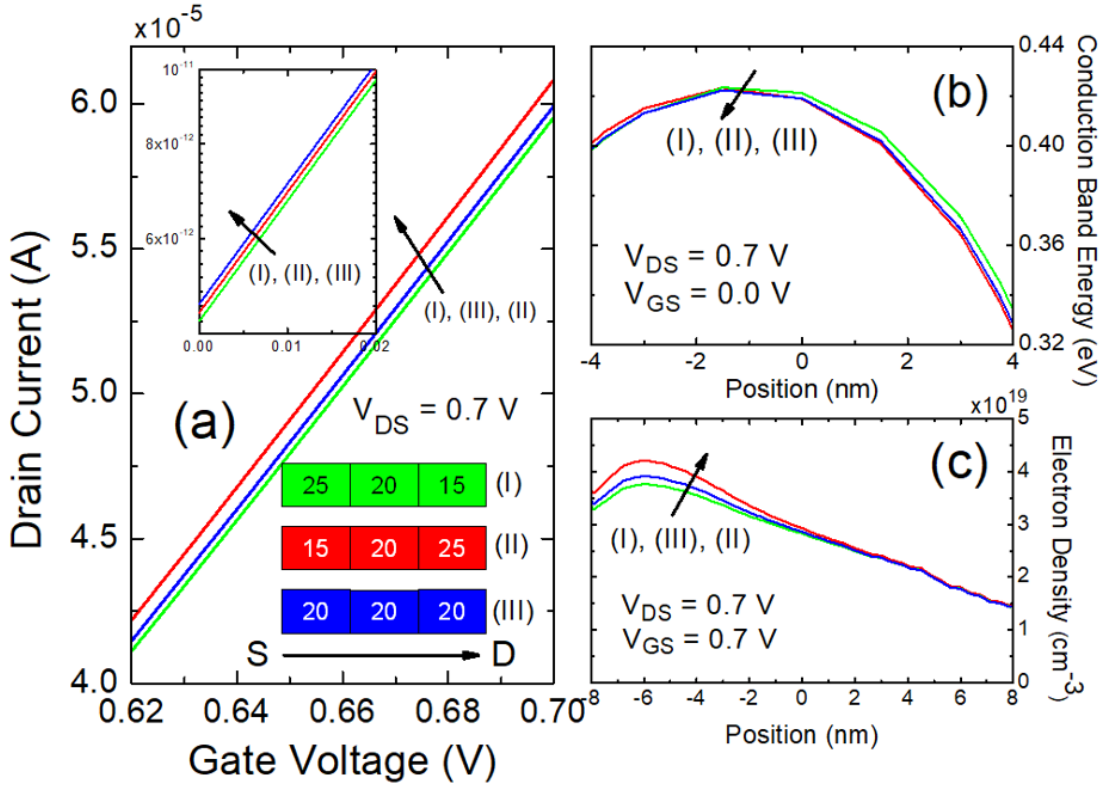


Figure 5.18: (a) $I_{DS} - V_{GS}$ of three different P_r configurations along the channel. The inset shows OFF state characteristics. (b) Conduction band profile at OFF state along the channel. (c) Electron density at ON state along the channel.

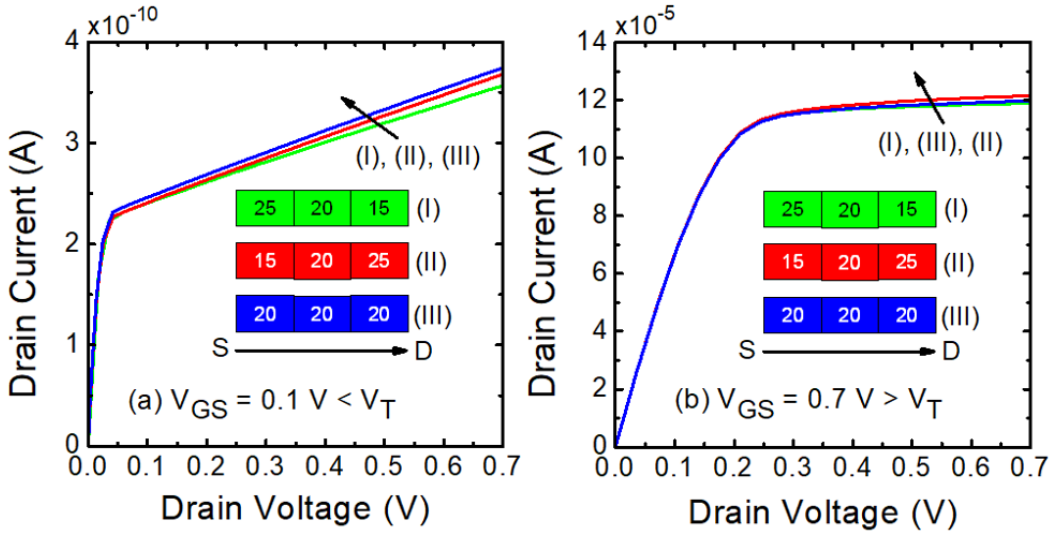


Figure 5.19: $I_{DS} - V_{DS}$ at V_{GS} = (a) 0.1 and (b) 0.7V for different P_r spatial distribution.

to the phenomenon reported in halo transistors [57].

DE and FE Grain Ratio

Ignoring P_r variation for now, three DE (assumed permittivity is 16) tile probabilities are considered in Fig. 5.20 (a) and (b): 33%, 50%, and 67%. These values are used because the literature suggested values as high as 70% [162] and as low as 0% [159]. The grain size is assumed to be 6.2 nm × 6.2 nm. DE content broadens the device parameter distributions. Of course, one expects 100% DE to cause no variation as shown in Fig. 5.21 (a)−(d). As the DE content increases in Fig. 5.21 (a)−(d), the mean I_{ON} and the mean V_{TH} decrease, while the mean I_{OFF} and the mean SS increase. All four trends support the mental model that the presence of DE effectively increases the EOT of the NC-FinFET. Fig. 5.21 (e) shows the current flow contour of the maximum ION case for 33% DE at $V_{GS} = V_{DS} = V_{DD}$. The low current density regions are over DE grains as expected, showing the effect of effective larger local EOT and the absence of the voltage amplification effect provided by FE. Note that the device variations due to the assumed DE content are much more harmful than that due to P_r variation (Fig. 5.16). Fortunately, a theoretical study [159] has shown that 0% DE (pure FE) is possible. The material should be carefully engineered to get rid of the DE grains. In general, the performances of NC-FinFET with DE grains can be understood as a baseline FinFET with thicker interfacial EOT (100% DE would mean the thickest EOT) and partial voltage amplification benefit of a 0% DE NC-FinFET. Empirically the relation between the DE content and device performance parameters can be described by the bowing equation

$$F\left(\mathrm{FE}_{1-x}\mathrm{DE}_x\right) = F\left(\mathrm{FE}\right)\cdot(1-x) + F\left(\mathrm{DE}\right)\cdot x - b\cdot x\cdot(1-x) \tag{5.13}$$

where F can be I_{ON}, I_{OFF}, V_{TH}, and SS, x is the content of DE ranging from 0 to 1, and b is the bowing parameter. Fig. 5.21 (a)−(d) show the empirical bowing parameters. Note that the device with 100% DE has a thicker EOT (= 0.9 nm + 2 × 3.9/16 nm ≈ 1.39 nm) than the baseline FinFET (EOT = 0.9 nm).

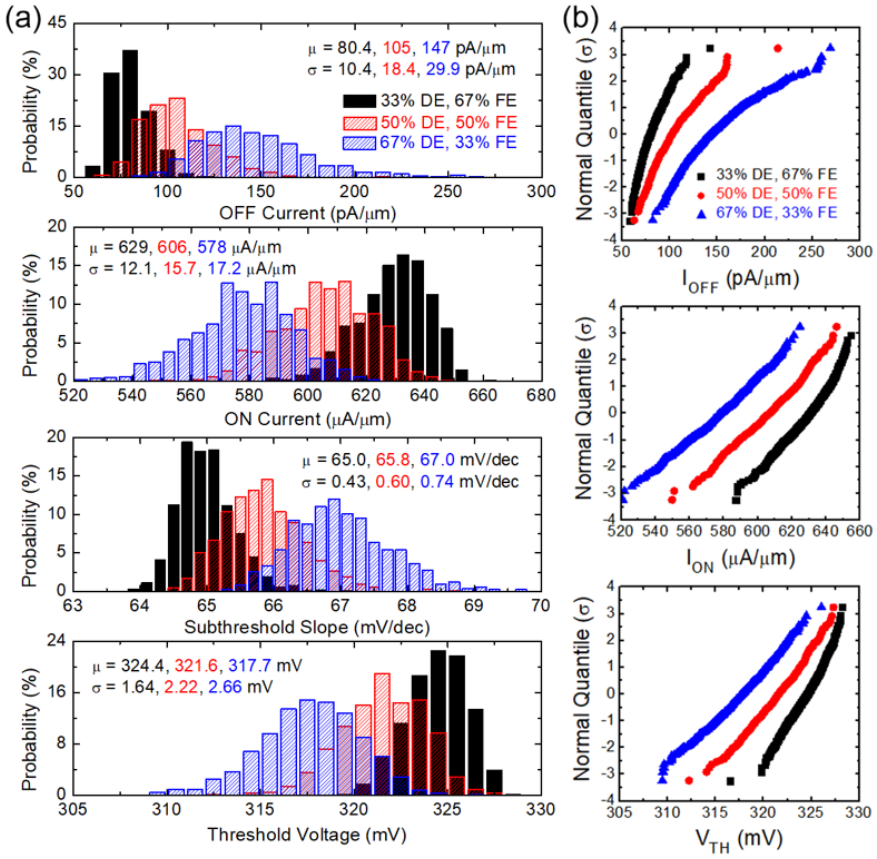


Figure 5.20: (a) Distributions of I_{OFF}, I_{ON}, V_{TH}, and SS in NC-FinFET with DE variation (33%, 50%, and 67% DE). (b) Normal quantile plots.

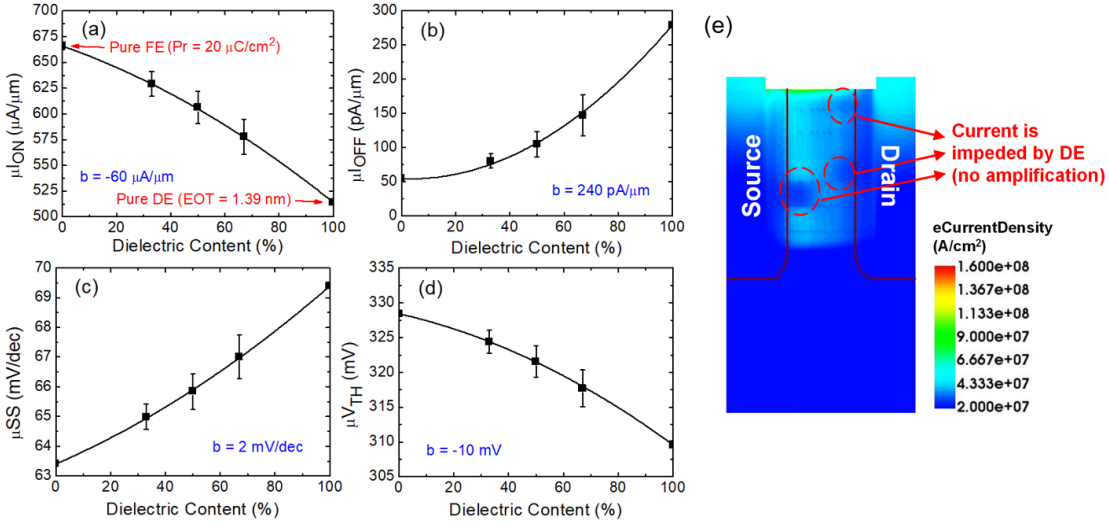


Figure 5.21: (a) μI_{ON}, (b) μI_{OFF}, (c) μSS, and (d) μV_{TH} of NC-FinFET with various DE content. The bars show the $\pm\sigma$. The solid lines represent the bowing equations with bowing parameter b. Note that 0% DE and 100% DE are NC-FinFET with uniform ferroelectric film and FinFET with thicker EOT (= 1.39 nm) than the baseline EOT (= 0.9 nm). (e) Current density in NCFinFET with 33% DE (maximum I_{ON} case shown). The dark blue regions are covered by DE.

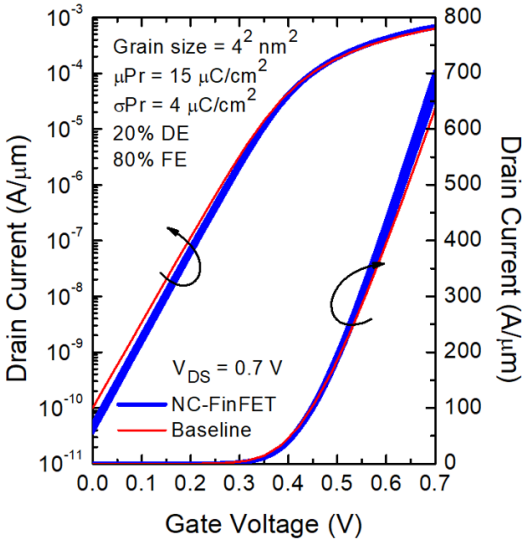


Figure 5.22: $I_{DS} - V_{GS}$ of baseline FinFET and NCFinFET with 20% DE content and P_r variations.

Combined Effects of DE and Pr Variation

The combined effect of DE phase and P_r variation is shown in Fig. 5.22 and Fig. 5.23. The DE variation is the dominant source of device parameter variations (Fig. 5.23) as mentioned earlier. In order to keep FE induced device variations below those due to other FinFET variation sources [e.g. gate edge roughness (GER), fin edge roughness (FER), and metal gate granularity (MGG)] [27, 160, 161], the DE content must be less than 20% and σP_r less than 27% of the mean. Fig. 5.24 further shows the coefficient of variation at different DE content. The normal $Q-Q$ test on the I_{ON}, I_{OFF}, and V_{TH} distributions due to the DE content variation shows that they closely follow the Gaussian distribution (Fig. 5.25). Some outliers of I_{ON}, I_{OFF}, and V_{TH} are observed in 60% DE case due to stronger short channel effect.

Grain Size Effect on Device Variations

The effect of grain size (comparing 4 nm × 4 nm and 5.3 nm × 5.3 nm) are also investigated (Fig. 5.26 and Fig. 5.27). Grain size can be engineered through the stress and doping [159] and perhaps deposition method. Small grain reduces the device variations because the larger number of grains in each device reduces the number of outliers. Indeed, the smaller grain cases in Fig. 5.27 follows the Gaussian distribution well. The small grain can average out the variation from the adjacent grains. Similar results have been observed in metal gate granularity (MGG) effect [161]. In addition to I_{ON}, I_{OFF}, V_{TH}, and SS, the effective drive current (I_{EFF}) and output resistance (R_{out}) are also important from circuit perspective. Fig. 5.28 shows the quantile plot of I_{EFF} and R_{out}. FE-induced I_{EFF} and R_{out} variations of NC-FinFET is less than the variations induced by other sourced in the baseline FinFET if the DE content is kept below 20% and P_r variation below 27%.

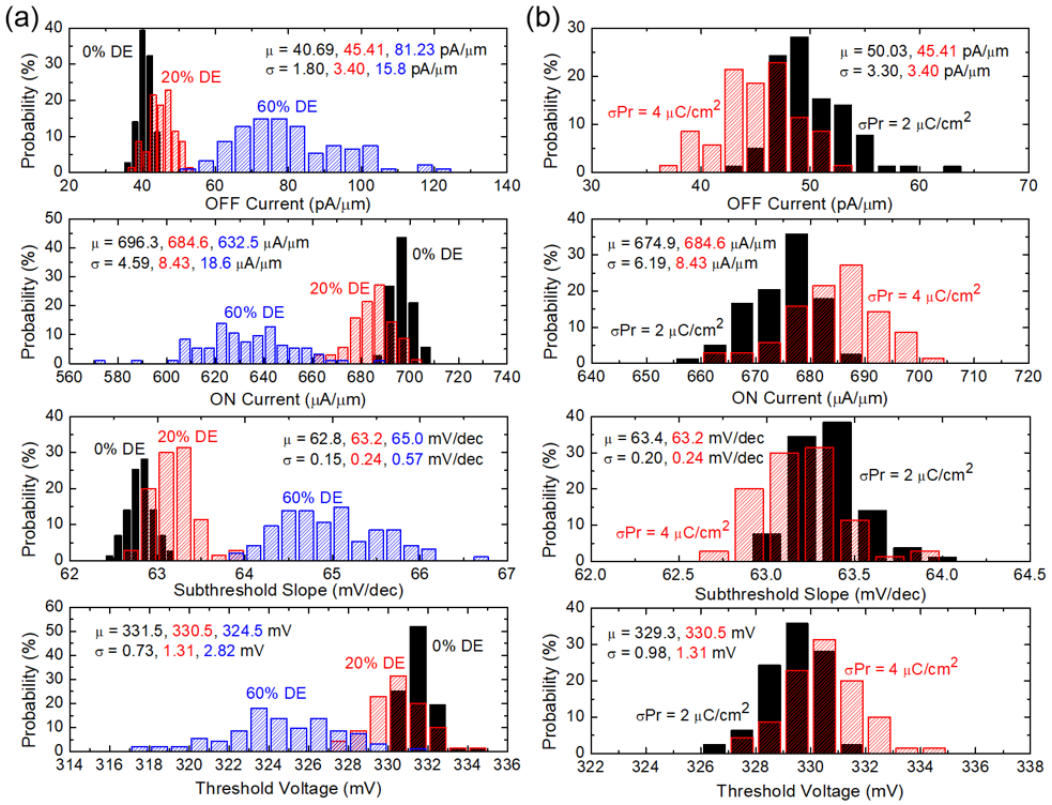


Figure 5.23: Statistical distributions of NC-FinFET with (a) different DE content variations (μP_r = 15 μC/cm^2 and σP_r = 4 μC/cm^2) and (b) different P_r variations (20% DE and μP_r = 15 μC/cm^2). The DE content variation has more impact on NC-FinFET.

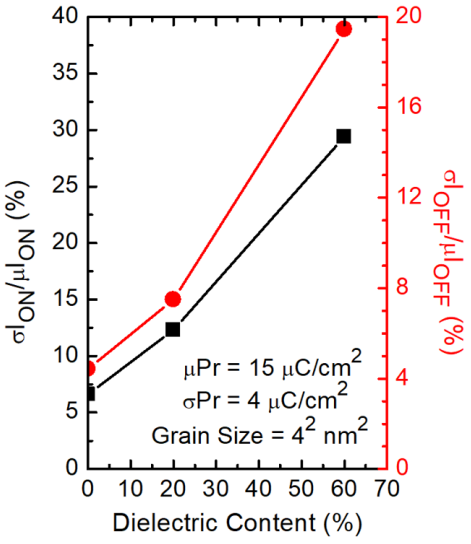


Figure 5.24: Coefficient of variation of I_{ON} and I_{OFF} for different DE content.

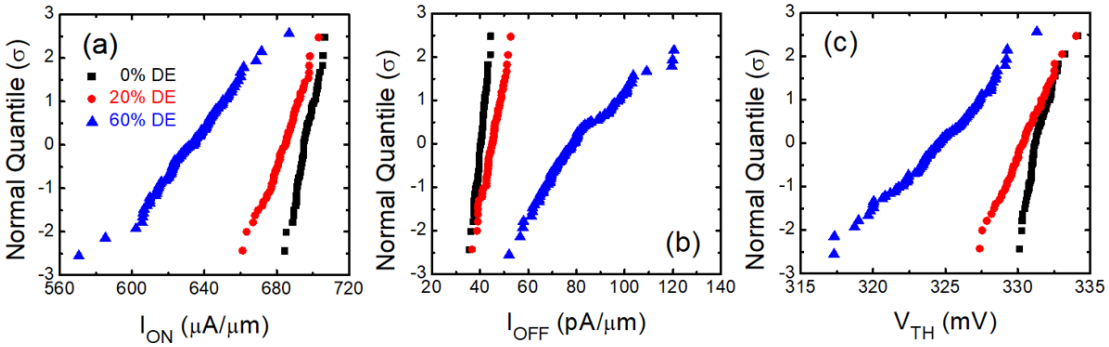


Figure 5.25: Normal quantile plot for (a) I_{ON}, (b) I_{OFF}, and (c) V_{TH} of NC-FinFET with various DE content (grain size = 4^2 nm^2, $\mu P_r = 15$ μC/cm^2, and $\sigma P_r = 4$ μC/cm^2).

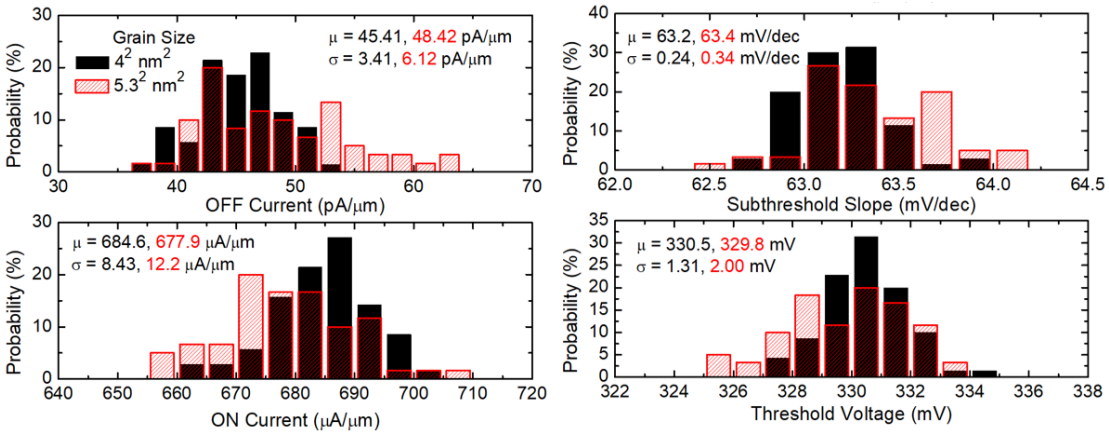


Figure 5.26: Grain size effect on the distributions of device performances (20% DE, $\mu P_r = 15$ μC/cm^2 and $\sigma P_r = 4$ μC/cm^2).

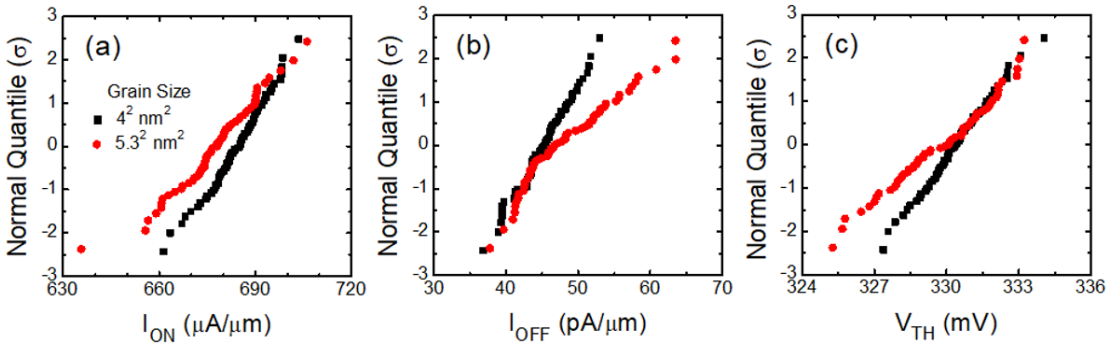


Figure 5.27: Normal quantile plot for (a) I_{ON}, (b) I_{OFF}, and (c) V_{TH} of NC-FinFET with various grain sizes (20% DE, $\mu P_r = 15$ μC/cm^2 and $\sigma P_r = 4$ μC/cm^2).

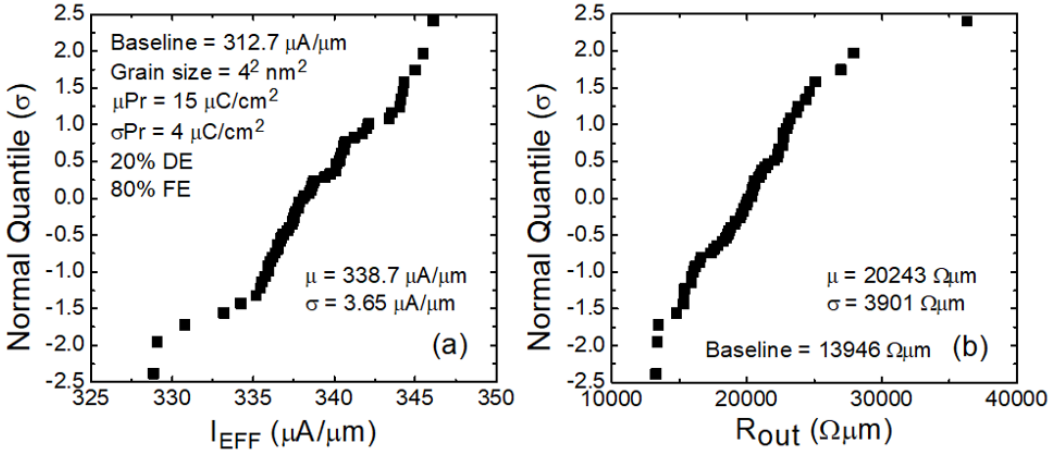


Figure 5.28: Normal quantile plot for (a) I_{EFF} and (b) R_{out} of NC-FinFET considering both P_r and DE variations. The coefficient of variation of I_{EFF} ($\approx 10.77\%$) is less than the I_{ON} variation in Fig. 2.26 and is a better indicator of circuit speed variation.

5.4 Spacer Optimization

In Section 5.2, the impact of inner fringing field has been discussed, indicating that the usage of fringing fields at the edge of NCFET would be an efficient way to boost the NCFET performances. To further utilize the negative capacitance (NC) effect, the device structure design and optimization of NCFET become more and more important. For example, the impact of spacer permittivity on NCFET has been discussed in [150, 151] since the outer fringing field affects the capacitance matching of the NC effect. Furthermore, it has been reported that the parasitic capacitances, such as gate-to-contact and gate-to-sidewall capacitances, play important roles in the advanced technology [163], the spacer engineering to reduce the parasitic capacitances is urgently required. Although the spacer optimization is proposed in FinFET [164] and nanowire FET [165], it has not yet been investigated in NCFET. In this section, the spacer design of NC-FinFET is studied using Sentaurus TCAD. The device and circuit performances with various spacer configurations are analyzed, deriving the insight into NC-FinFET optimization.

5.4.1 Simulation Methodology

The NC-FinFET, designed based on low power (LP) 8/7 nm technology node FinFET (baseline) reported in IEEE International Roadmap for Devices and Systems (IRDS) [153], is simulated by using Sentaurus Technology computer-aided design (TCAD). The device performances of baseline FinFET are calibrated with the targets in IRDS by tuning the mobility parameters, doping profile, and metal gate workfunction. The FE layer is deposited on the gate dielectric of the baseline FinFET, and the direction of polarization of FE is perpendicular to the fin surface. Sentaurus TCAD is capable of solving Landau-Khalatnikov equation of electric field in FE as a function of polarization (P): $E_{FE} = 2\alpha P + 4\beta P^3 + 6\gamma P^5 - 2g\Delta P + \rho\,(dP/dt)$ with Poissons equation and other physical models, including mobility degradation due to surface roughness scattering and impurity scattering. The domain interaction coefficient g

[143] and the viscosity coefficient ρ in Landau-Khalatnikov equation are 10^{-10} m^3/F [140] and 0.18 Ωcm [148], respectively. The remnant polarization (P_r) and coercive field (E_C) are 6 $\mu C/cm^2$ and 1 MV/cm corresponding to $\alpha = 2.165 \times 10^{11}$ cm/F, $\beta = 3.007 \times 10^{21}$ $cm^5/(FC^2)$, and $\gamma = 0.000$ $cm^9/(FC^4)$, which are achievable in HfO_2-based FE by employing doping technique [147]. The NC-FinFET is simulated with the same model parameters except the FE parameters. The simulated device structure and dimensions are summarized in Fig. 5.26 (a). There are five spacer configurations: (1) full (nitride); (2) dual (air + nitride); (3) fin corner; (4) fin selective; and (5) air, which are defined in Fig. 5.29 (b). The fabrication processes of these spacer configurations will be discussed in next section. In addition to NC-FinFET, the spacer design of baseline FinFET is also identical for simplified comparison. Note that the OFF currents of all devices are set to be 100 pA/μm by only tuning the metal gate workfunction since the doping profile may change the electric field distribution in the spacer region [165].

5.4.2 Analysis and Discussion

Proposal of Spacer Fabrication Process

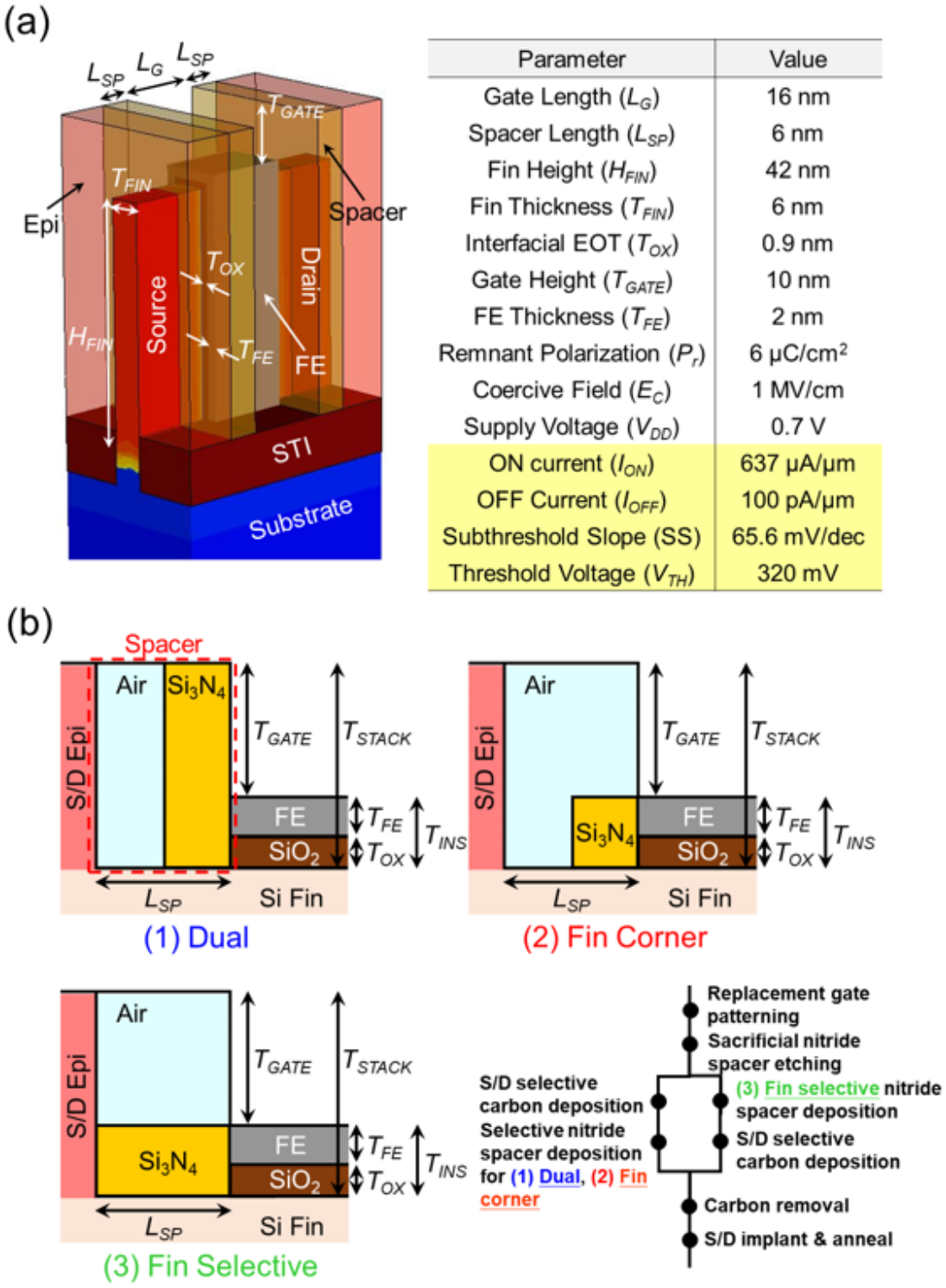
The process flows of the vacuum spacer for FinFET [164] and corner spacer for planar MOSFET [166] have been developed. To form the fin selective and dual/fin corner spacers, the process is proposed as shown in Fig. 5.29 (b). After the removal of the sacrificial spacer from the gate-last process, the selective deposition is used to define the air and nitride spacer region. The air spacer region is formed by filling in the carbon. The carbon layer is then removed by being exposed to a mild oxygen plasma [167], forming the byproducts escaping through the opening in the spacer region [164]. Finally, the ion implantation is carried out to form the highly-doped source/drain region, and the dopants are activated by the annealing.

Parasitic Capacitance in FinFET Structure

Since the source/drain epitaxy region overlaps with the gate electrode for the contacts, there is a parasitic capacitance (C_{EPI}) which increases the total gate capacitance (C_{GG}) [see Fig. 5.30 (a)] [165]. To mitigate C_{EPI}, the air spacer has been proposed in FinFET technology [164, 166]. However, such low-κ spacer potentially leads to degradation of the ON current due to lack of control to the S/D extension region through the outer fringing field [capacitance $C_{OF,G}$ defined in Fig. 5.30 (a)]. Thus, the corner spacer is adopted [165] to further boost the current without much penalty of parasitic capacitance [23.4% reduction, see Fig. 5.31 (b)]. However, with the concern of capacitance matching in NC-FinFET [152], the spacer design has the impact on not only the current but also the total gate capacitance.

Spacer Engineering of FinFET

Fig. 5.31 (a) and (b) show the drain current and total gate capacitance of FinFET with various spacer designs shown in Fig. 5.29 (b). The increment in the length and height of the Si_3N_4 region improves the drain current but degrades the total gate capacitance wiht the increased switching charge [Fig. 5.31 and 5.32]. Although the



Parameter	Value
Gate Length (L_G)	16 nm
Spacer Length (L_{SP})	6 nm
Fin Height (H_{FIN})	42 nm
Fin Thickness (T_{FIN})	6 nm
Interfacial EOT (T_{OX})	0.9 nm
Gate Height (T_{GATE})	10 nm
FE Thickness (T_{FE})	2 nm
Remnant Polarization (P_r)	6 μC/cm^2
Coercive Field (E_C)	1 MV/cm
Supply Voltage (V_{DD})	0.7 V
ON current (I_{ON})	637 μA/μm
OFF Current (I_{OFF})	100 pA/μm
Subthreshold Slope (SS)	65.6 mV/dec
Threshold Voltage (V_{TH})	320 mV

Figure 5.29: (a) The simulated device structure and the device parameters. The rows in yellow represent the electrical performance of the baseline FinFET. (b) The cross-section and definition of the spacer configurations with their corresponding process flows.

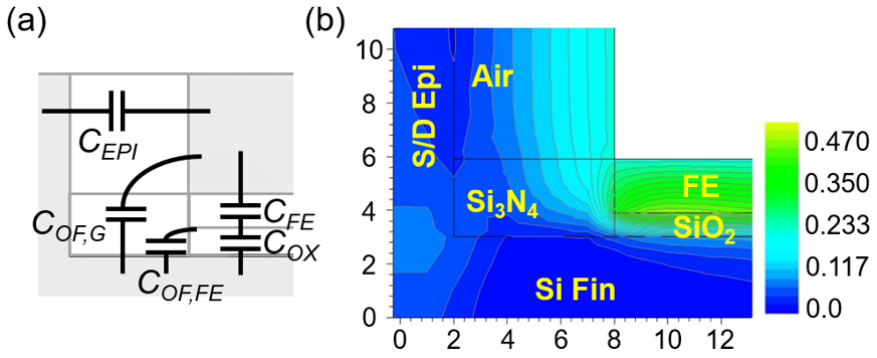


Figure 5.30: (a) Capacitance network of an NC-FinFET. (b) Simulated electrostatic potential contours through fin selective spacer region near the source at $V_{GS} = 0.7V$ (the scales are in unit of nm).

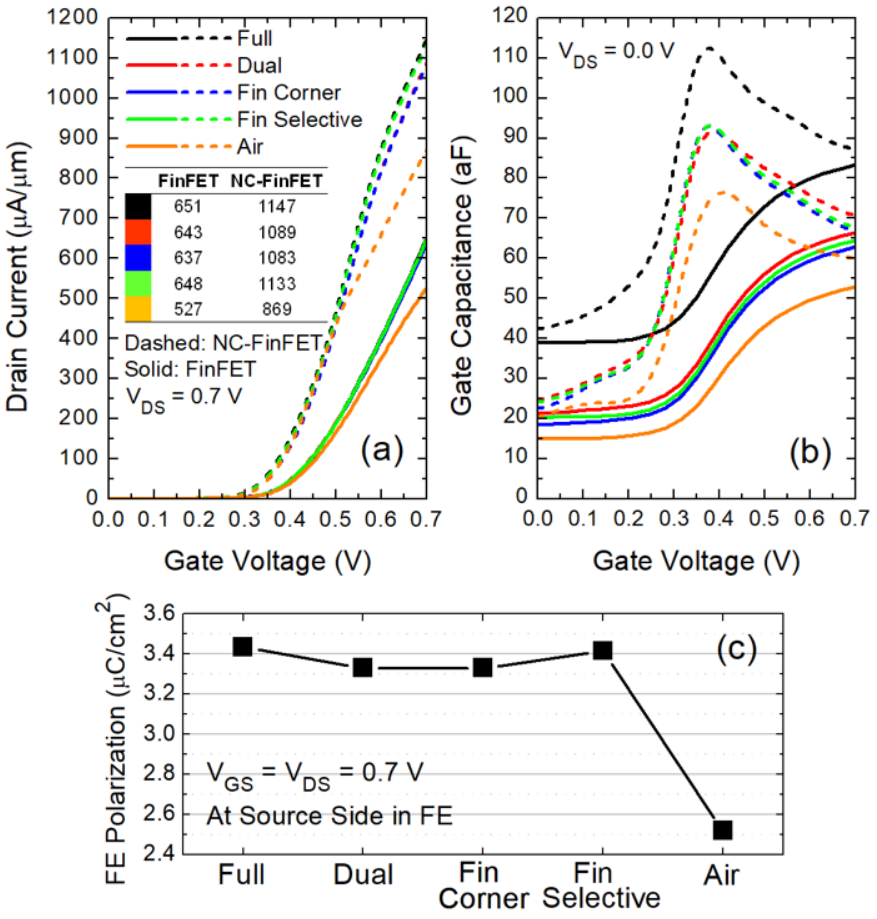


Figure 5.31: (a) Drain current (the table summarizes I_{ON}) and (b) gate capacitance with the spacer engineering in the baseline FinFET (solid) and NC-FinFET (dashed). (c) FE polarizations at source side in NC-FinFET.

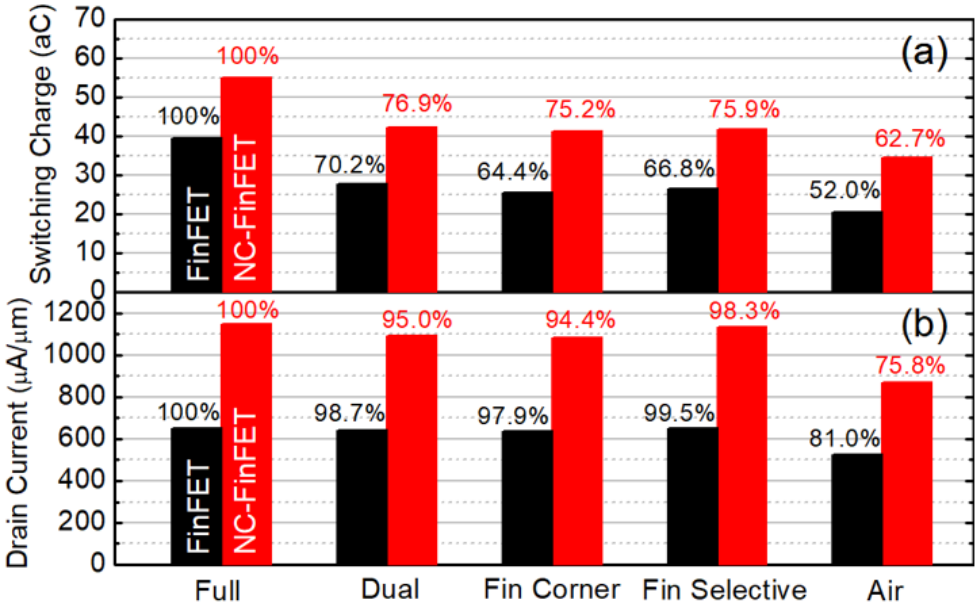


Figure 5.32: Impact of spacer designs on (a) switching charge (integrated gate capacitance from $V_{GS} = 0$ to $0.7V$ at $V_{DS} = 0V$) and (b) drain current (ON-state: $V_{GS} = V_{DS} = 0.7V$).

total gate capacitance can be effectively reduced by mitigating C_{EPI} through the air spacer, the drain current gets worse, leading to more inverter propagation delay especially when the wire or load capacitance overwhelms the gate capacitance, for example, in the state-of-the art technology [153], as shown in Fig. 5.33. Thus, the fin corner spacer would be the better choice for the circuits consisting of FinFETs as reported in [165] where the drain current is only degraded by 2.1% (compared to the full spacer design) but the switching charge is improved by 35.6% (Fig. 5.32) and thus 7.9% improvement (at $C_{\text{Load}} = 200$ aF $\approx$ 3x gate capacitance) in the inverter propagation delay (Fig. 5.33). The benefit of fin corner spacer is even more visible when the circuit is heavily loaded (Fig. 5.33).

Spacer Engineering of NC-FinFET

As mentioned earlier, the capacitance matching in NC-FinFETs is critical for the design. To gain the better capacitance matching, the outer fringing field should be properly engineered. Fig. 5.31 and Fig. 5.32 show the resultant drain current, total gate capacitance, and switching charge characteristics with various spacer designs shown in Fig. 5.29 (b). Although the gate capacitance and switching charge are significantly suppressed, the drain current is also degraded even more than the baseline FinFET compared to their full nitride spacer configurations due to lack of capacitance matching. This results that although the fin corner spacer is of interest for the baseline FinFET, the inverter propagation delay can be further reduced if the $C_{OF,FE}$ [Fig. 5.30 (a) and (b)] is engineered by elongating the fin corner spacer for the NC-FinFET which is called the fin selective spacer (with thickness of Si_3N_4 up to $T_{FE} + T_{OX}$). The field between the gate and the S/D epitaxy/extension layer leads to more ineffective capacitances (C_{EPI} and $C_{OF,G}$) for capacitance matching, which cannot boost the current but degrade the total capacitance. Fig. 5.31 (c) shows the FE polarizations

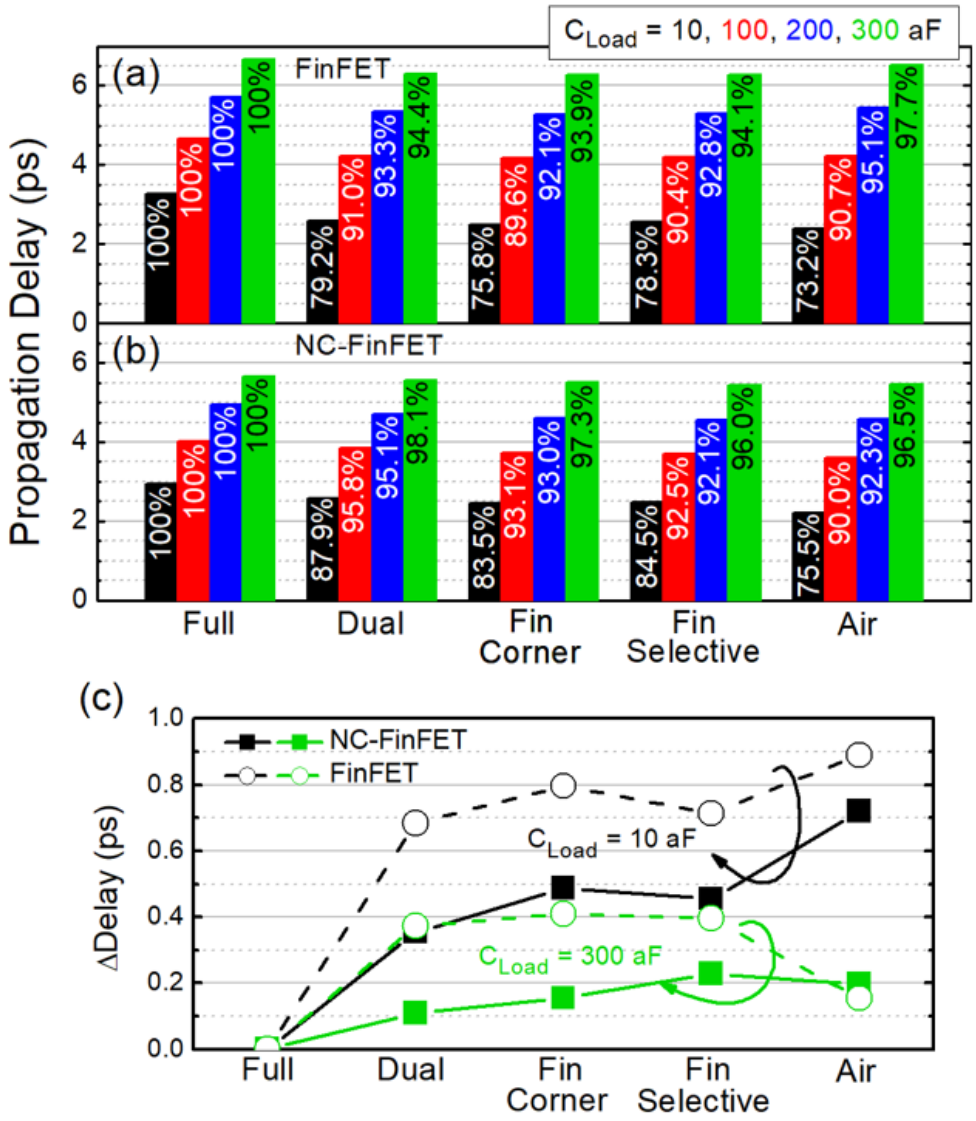


Figure 5.33: Inverter propagation delay of (a) FinFET and (b) NC-FinFET with various spacer configurations at load capacitances C_{Load} = 10 aF, 100 aF, 200 aF, and 300 aF. The relative percentage changes are calculated by comparing to their full spacer designs. (c) Delay reduction (ΔDelay) with respect to the full spacer designs.

for various spacer designs at source side, indicating better capacitance matching and more voltage amplification in the full and fin selective spacers than in the dual and fin corner because of $C_{OF,FE}$. If without much aid of $C_{OF,FE}$, the capacitance matching of NC-FinFET with the dual or fin corner spacer gets more reduction in current compared with FinFET [see Fig. 5.31 (c) and Fig. 5.32 (b)]. The drain current of fin selective spacer is only degraded by 1.7% (compared to the full spacer design) but the switching charge is reduced by 24.1%, giving rise to 7.9% improvement in the inverter delay when the load capacitance is 200 aF ($\approx$ 3x gate capacitance) as shown in Fig. 5.33. The benefit provided by the fin selective spacer is more obvious when the load capacitance gets worse. The improvements of $\approx$ 30 fs and $\approx$ 230 fs against the air and full spacers are obtained at C_{Load} = 300 aF as shown in Fig. 5.33 (c). However, the air spacer beats other designs when the circuit is lightly loaded, and it only loses a little benefit when the circuit is heavily loaded due to removal of amplified capacitance in NC-FinFET. From the fabrication complexity perspective, the practicality of the air spacer is more applicable than other designs since it only needs carbon deposition and removal. Finally, we note the superiority of the NC-FinFET that the inverter delay is only about 80% of that of the baseline FinFET (Fig. 5.33) due to the significant improvement of the drive current.

5.5 Conclusion

In this Chapter, the physics, compact model, variability, and design of NCFET have been comprehensively explored and analyzed with emphasis on the future technology node.

In Section 5.2, the compact model of NCFET is first reviewed, and then the role of inner fringing fields in NC-FinFETs is discussed. The TCAD simulated results show that in short channel devices, the inner fringing fields induce negative charges and thus suppress the OFF current. In contrast, in long channel devices, the inner fringing fields cannot affect the central portion of the channel and the depletion charges dominate the subthreshold characteristics. The ON current is boosted with decreasing P_r due to FE voltage amplification resulting from the bias point determined by the inversion charge. Due to the NC effect, the DIBL of NC-FinFETs is reduced as compared with baseline FinFETs, which is promising for the scaling capability of NC-FinFETs. A compact model to capture the inner fringing field effect in NC-FinFETs is also proposed. Thanks to the Gaussian quadrature method, the spatial distribution of the inner fringing field-induced gate charges can be properly modeled with an analytical short channel charge expression. The predictive nature of the developed compact model regarding channel length and P_r scaling is validated by TCAD simulation.

In Section 5.3, Monte Carlo simulation of NC-FinFET variations induced by remnant polarization variation and the presence of dielectric phases in the ferroelectric film is presented. It is found that the DE variation is a more serious source of NC-FinFET variation than the P_r variaion. To keep the FE variation effects below those induced in FinFET by other sources (RDF, GER, FER, and MGG), the DE content must be less than 20% which is theoretically possible. The P_r variation must be less than 27%. Our simulation shows that 4 nm $\times$ 4 nm grains (92 grains in the channel) leads to less device variations than 5.3 nm $\times$ 5.3 nm grains (54 grains in the channel).

While uniform single crystalline ferroelectric film would minimize the device variation, nano-crystalline ferroelectric film may be preferrable to poly-crystalline film.

In Section 5.4, the spacer designs for FinFET and NC-FinFET are analyzed using Sentaurus TCAD. It is found that for a heavily loaded circuit ($\approx$ 3 times gate capacitance), the fin corner and fin selective spacers are helpful for the circuit speed in FinFET ($\approx$ 7.9% improvement) and NC-FinFET ($\approx$ 7.9% improvement), respectively. However, if taking into account the process complexity, the air spacer ($\approx$ 7.7% improvement in speed) would be economical in NC-FinFET but the fin selective spacer may still be attractive.

Chapter 6

Magnetic Memory Compact Model: STT-MTJ MRAM

6.1 Introduction

In previous few Chapters, the transistor compact models are emphasized and have been implemented into the industry standard BSIM models. The philosophy of transistor compact model is still applicable to the memory model: analytical equations with flexible model parameters. Different flavors of memories offer the needs for various applications but they have their own pros and cons as summerized in Table 1.1. Commonly used volatile memories include SRAM and DRAM. SRAM has a fast operation with the usage of positive feedback between two inverters, but it requires more cell area and suffers from leakages with aggressively scaled transistors [33]. DRAM has a high density cell but relies on frequently refreshing [34]. Non-volatile flash memory requires high electric field to generate the tunneling which leads to reliability issues and low speeds [34, 168]. STT-MTJ MRAM serves as an universal memory with the speed of SRAM, the density of DRAM, and non-volatility of flash memories but simultaneously with power efficiency and higher endurance. In this Chapter, we will develop the compact model of STT-MTJ MRAM.

6.1.1 STT-MTJ Operation

MTJ is a two-terminal device with an oxide tunnel barrier (for example, MgO) sandwiched by two ferromagnetic (FM) layers (for example, CoFeB) [36]. One of the FM layers is with fixed (permanent) magnetization (called the fixed or reference or pinned layer), while the other can be flipped by the input (free layer) as shown in Fig. 6.1 (a). The tunneling magnetoresistance (TMR) effect leads to bi-state characteristics of MTJ. There are two states in STTMTJ: (1) parallel and (2) anti-parallel states, depending on the magnetization configurations. In parallel (P) state the resistance is low (denoted as 0) because of good band matching [36, 169], while in anti-parallel (AP) state the resistance is high (denoted as 1) as shown in Fig. 6.1 (b). To change from AP to P, the electrons should be injected from pinned layer to free layer. The electrons would be spin-polarized in pinned layer so that only the electrons with same polarization direction as the magnetization direction of pinned layer are able to tunnel through the barrier. These electrons now are with opposite direction

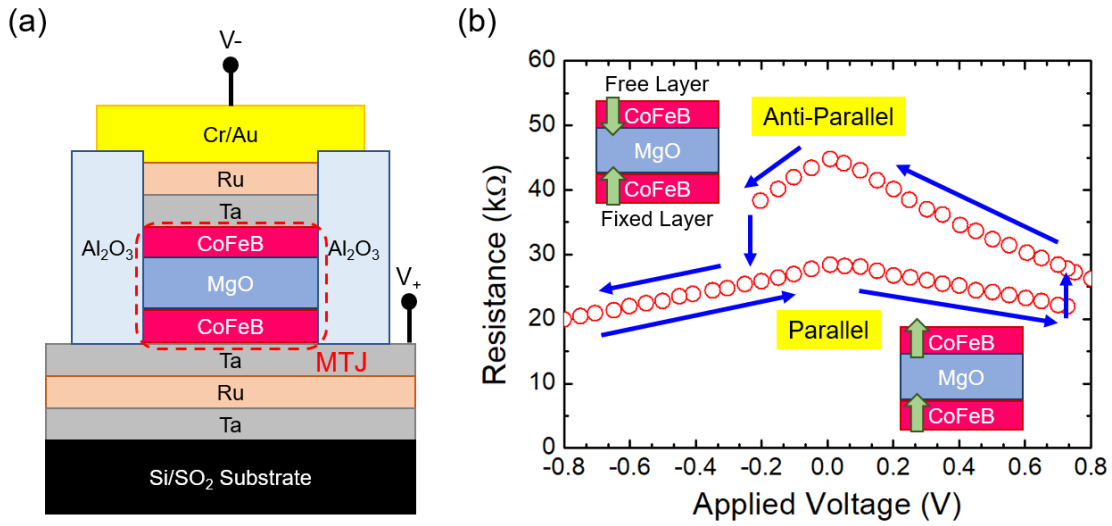


Figure 6.1: (a) A typical MTJ structure (reproduce from [170]). (b) Illustration of parallel and anti-parallel states in the perpendicular MTJ. The green arrows in FM layers represent magnetization. (Applied voltage = $V_+ - V_-$)

of magnetization against that of free layer. Thus, they give a torque to free layer and flip the magnetization. This mechanism is known as spin-transfer-torque (STT) [171], which needs smaller current than that of traditional MRAM generating magnetic field to flip the magnetization [171]. For changing P to AP, the electrons are injected from free layer to pinned layer. Only the reflected electrons from barrier with opposite polarization to the magnetization direction of free layer give a torque and flip the magnetization [171]. Because the reflected electrons are minority, the critical current to change from P to AP is higher than that from AP to P, resulting in asymmetric writing of STT-MTJ.

There are two flavors of MTJs depending on the magnetization directions: in-plane and perpendicular MTJs. The perpendicular MTJ (PMTJ) has the magnetization perpendicular to the surface of the tunnel oxide [see Fig. 6.1 (b)], while the in-plane MTJ (IMTJ) has that parallel to the surface of the tunnel oxide. Since the IMTJ relies on shape magnetic anisotropy field to maintain the thermal stability, the scalability of IMTJ is not suitable [172]. In contrast, the PMTJ has good interfacial magnetic anisotropy field with better thermal stability so that scaling down to sub-20 nm is possible [173]. Therefore, the PMTJ will be emphasized and modeled in this Chapter.

6.2 Core Model

6.2.1 Landau-Lifshitz-Gilbert (LLG) Equation

The magnetization switching behavior in the free layer of MTJ is described by the Landau-Lifshitz-Gilbert (LLG) equation[1] [36]

$$\frac{1+\alpha^2}{\gamma}\frac{d\hat{\boldsymbol{m}}}{dt} = \hat{\boldsymbol{m}} \times [(\boldsymbol{H_{eff}} - \boldsymbol{H_s}) - \hat{\boldsymbol{m}} \times (\alpha \boldsymbol{H_{eff}} - \boldsymbol{H_s})] \tag{6.1}$$

[1]The bold symbols represent the vectors.

where $\hat{\boldsymbol{m}}$ is the unit vector of magnetization in the free layer, α is the damping parameter, γ is the gyromagnetic ratio, $\boldsymbol{H_{eff}}$ is the effective magnetic field, and $\boldsymbol{H_s}$ is the field characterizing the STT effect [36, 174, 175]

$$\boldsymbol{H_s} = \eta \frac{\hbar}{2} \frac{J}{q} \frac{1}{M_S t} \hat{\boldsymbol{m}}_{\boldsymbol{p}} \tag{6.2}$$

since the spin-transfer torque is

$$\boldsymbol{\Gamma}_{STT} = -\gamma\eta \frac{\hbar}{2} \frac{J}{q} \frac{1}{M_S t} \hat{\boldsymbol{m}} \times \hat{\boldsymbol{m}} \times \hat{\boldsymbol{m}}_{\boldsymbol{p}} \tag{6.3}$$

where η is the STT-efficiency factor related to the polarization of the injected current, J is the injected current density, M_S is the saturation magnetization of the free layer, t is the thickness of the free layer, and $\hat{\boldsymbol{m}}_{\boldsymbol{p}}$ is the spin-polarized direction of current [174, 175]. If the STT overcomes the damping (i.e., $H_s \geq \alpha H_{eff}$), the magnetization motion is excited and the critical current density (J_{c0}) for PMTJ to excite such motion can be derived from (6.1)−(6.3) with uniform switching approximation [174, 175]

$$J_{c0} = \frac{1}{\eta} \frac{2\alpha q}{\hbar} M_S t H_K \tag{6.4}$$

where H_K is the anisotropy field. If the magnetization is uniformly switched with rotation, the energy barrier to overcome is governed by the anisotropy energy $E_b = K_u V$, where $K_u = \frac{M_S}{2H_K}$ is the anisotropy energy density, and V is the volume [36]. The thermal stability Δ is then given by

$$\Delta = \frac{E_b}{k_B T} = \frac{K_u V}{k_B T} = \frac{H_K M_S V}{k_B T} \tag{6.5}$$

Thus, (6.3) can be re-writen as

$$J_{c0} = \frac{1}{\eta} \frac{2\alpha q}{\hbar} t \frac{2k_B T}{V} \Delta = \left(\frac{4qk_B T}{\hbar} \right) \frac{\alpha}{\eta} \Delta \frac{1}{A} \tag{6.6}$$

where A is the area. Hence, the critical current I_{c0} is

$$I_{c0} = \left(\frac{4qk_B T}{\hbar} \right) \frac{\alpha}{\eta} \Delta \tag{6.7}$$

The STT-efficiency factor η consists of the components from the spin valve (η_{sv}) and tunneling η_{tunnel} [172, 176]

$$\begin{aligned} \eta &= \eta_{sv} + \eta_{tunnel} \\ &= \left[-4 + \left(P^{-\frac{1}{2}} + P^{\frac{1}{2}} \right)^3 \frac{(3 + \cos\theta)}{4} \right]^{-1} + \frac{P}{2\left(1 + P^2 \cos\theta\right)} \end{aligned} \tag{6.8}$$

where θ is the angle between the magnetization of fixed and free layer (P state: $\theta = 0$; AP state: $\theta = \pi$), and P is the spin polarization efficiency of the injected current. Fig. 6.2 (a) shows η as a function of P. The difference in η between P and AP states results in asymmetric forward and reverse switching in PMTJ. Note that the thermal stability (thus the critical current) is a function of temperature since the saturation magnetization and anisotropy field are affected by temperature which can be modeled using a linear function [177, 178, 179] with a slope parameter DELTAS, $\Delta(T) = [\Delta_{300\mathrm{K}} + \mathrm{DELTAS}\,(T - 300)]\,\frac{300}{T}$, as shown in Fig. 6.2 (b).

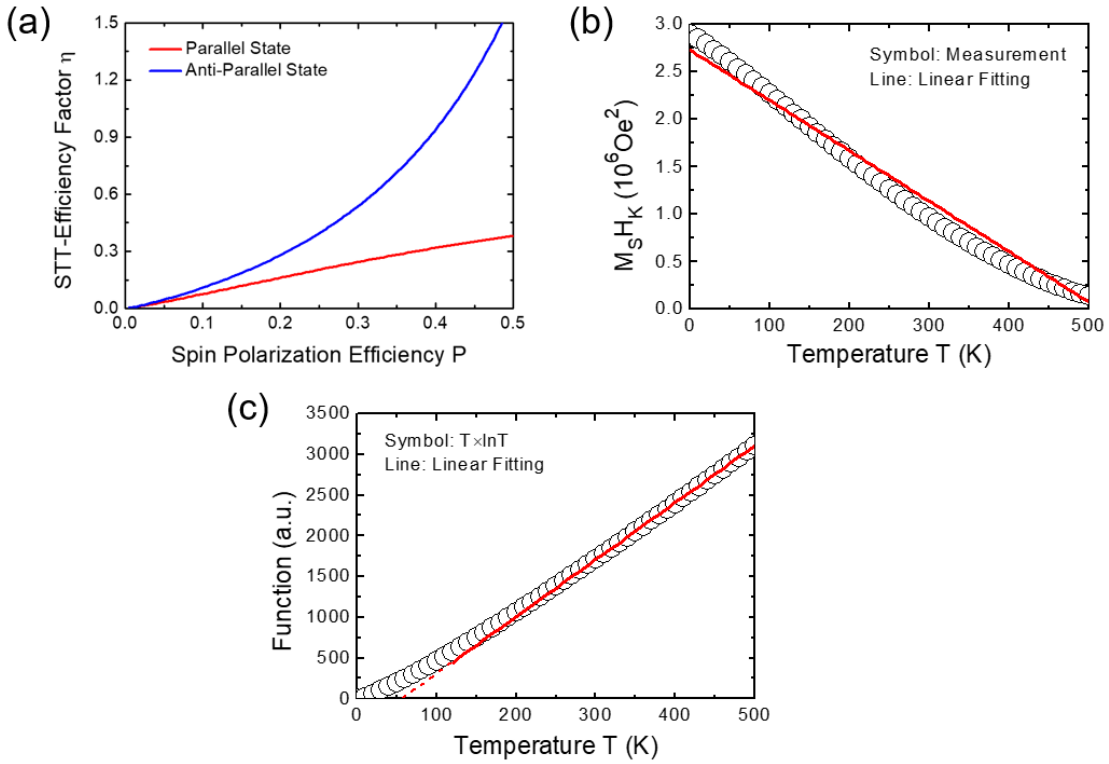


Figure 6.2: (a) STT-efficiency factor η versus spin polarization efficiency P, showing asymmetric switching in PMTJ. (b) Comparison of $M_S H_K$ data (from [177, 178]) and linear fitting. (c) Comparison of $T \ln T$ and T in TMR model.

6.2.2 Tunneling Magnetoresistance

The tunneling magnetoresistance is the key feature of MTJ, which determines the reliability of sensing states in MRAM. The general definition of tunneling magnetoresistance ratio (TMR) is

$$\text{TMR} = \frac{R_{AP} - R_P}{R_P} \tag{6.9}$$

where R_P and R_{AP} are the resistances in P and AP states, respectively. R_P generally depends on the tunnel oxide thickness and will be an model parameter of MTJ compact model. Once R_P is known, R_{AP} is determined by TMR. Theoretically and experimentally TMR shows bias-dependence, also known as the zero bias anomaly, which can be explained by the hot carrier-induced spin excitations (magnons) localized at the interfaces between the ferromagnetic layer and the tunnel oxide (inelastic tunneling) [41, 169] and the Brinkman-Dynes-Rowell model [180].

$$\text{TMR} = \frac{\text{TMR}_0}{1 + \frac{V_{\text{apply}}^2}{V_{\text{H}}^2}} \tag{6.10}$$

where V_{apply} is the applied voltage to the fixed (reference) layer with respect to the free layer (see Fig. 6.1), TMR_0 is the TMR at $V_{\text{apply}} = 0$, V_{H} is a model parameter associated with the material properties. Note that TMR_0 is a function of temperature and it decreases with increasing temperature, which can be explained by also the hot

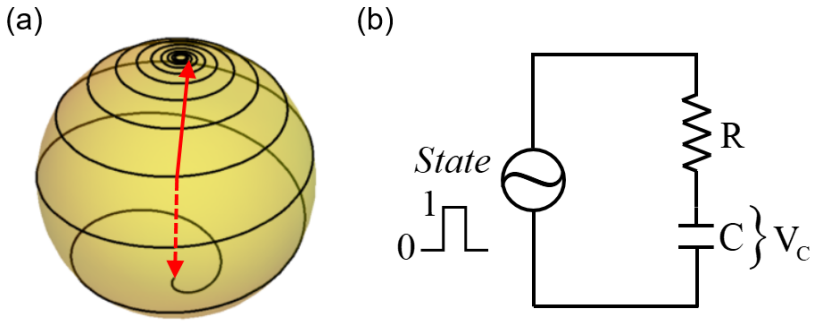


Figure 6.3: (a) Illustration of dynamic switching of magnetization simulated based LLG equation [174]. (b) RC sub-circuit for magnetization dynamic switching.

carrier-induced magnons (obey Bose-Einstein statistics and increase the number as increasing temperature) and minor additional spin-independent tunneling from hopping through the localized states in oxides [181]. The temperature dependence in TMR_0 shown in [169] is $T \ln T$ at high temperature which can be approximated by a linear function of T with a slope parameter TMRS [see Fig. 6.2 (c)] to simplify the equation: $\mathrm{TMR}_0(T) = \mathrm{TMR}_{300\mathrm{K}} + \mathrm{TMRS}(T - 300)$.

However, it is observed that such voltage- and temperature-dependence are more visible in the AP state than in the P state [41], which has been successfully explained by (only) the theory of hot carrier-induced magnons since the prefactors of these dependencies in the AP and P states are $(\rho^m)^2 + (\rho^M)^2$ and $2\rho^m\rho^M$ respectively[2], where $\rho^{m(M)}$ is density of states of itinerant minority (majority) electrons [169]. This can be attributed to the angular momentum conversation so that the spin of these electrons needs to be flipped and it contributes to another spin-channel (band) [169, 182]. In the compact model, the similar equation form of voltage-dependence in R_P is adopted but the model parameter V_{HP} will be much greater than V_{H}

$$R_P = \frac{R_{P0}}{1 + \frac{V_{\mathrm{apply}}^2}{V_{\mathrm{HP}}^2}} \tag{6.11}$$

and the temperature-dependence in R_P is ignored. Note that R_P is scaled with cell area, i.e., $R_{P0} \propto 1/A$. The experimental results have been shown that the TMR of PMTJ can be higher than 100% in L10-ordered FePt/MgO/Fe/ L10-ordered FePt and CoFeB/MgO/CoFeB systems, which trigger the shift from IMTJ to PMTJ although IMTJ has even higher TMR [171].

6.2.3 Dynamic Switching Model

From (6.1), one can see that the magnetization switching is a time-dependent behavior, indicating that the magnetization cannot be immediately flipped by the spin-transfer torque due to precession and thermal fluctuation [183] [see Fig. 6.3 (a)]. This dynamic switching behavior is modeled by an RC sub-circuit since the switching gradually takes place and charging C can mimic such behavior. The RC network is

[2]According to the inequality of arithmetic and geometric means, $(\rho^m)^2 + (\rho^M)^2 \geq 2\rho^m\rho^M$.

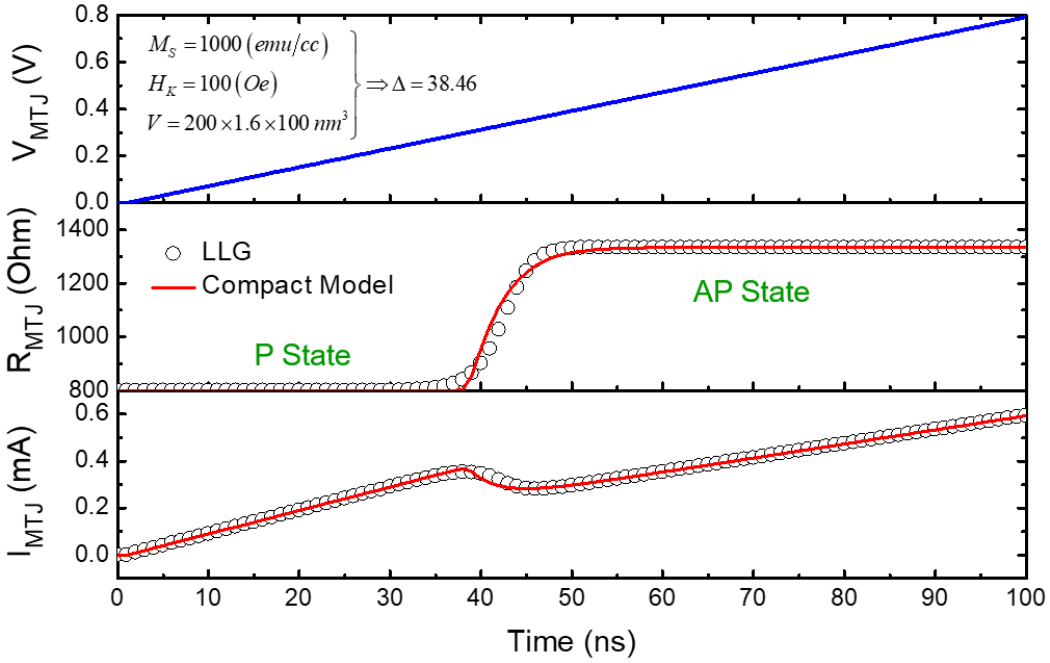


Figure 6.4: Comparison of transient simultions of the LLG solution and proposed compact model. The gradual transition from P to AP state is evident. ($R = 0.001$ and $C = 10^{-7}$)

shown in Fig. 6.3 (b), where the parameter STATE represent P (= 0) and AP (= 1) states. The MTJ resistance (R_{MTJ}) thus is

$$R_{\mathrm{MTJ}} = (1 - V_C)\, R_P + V_C R_{AP} \tag{6.12}$$

where V_C is the voltage across the capacitor. The capacitance C is determined by the ferroelectric material. (6.1) can be re-writen by normalizing to H_K since we consider STATE (from 0 to 1) in normalization manner [184]

$$\frac{1+\alpha^2}{\gamma H_K}\frac{d\hat{\boldsymbol{m}}}{dt} = \hat{\boldsymbol{m}} \times [(\boldsymbol{h_{eff}} - \boldsymbol{h_s}) - \hat{\boldsymbol{m}} \times (\alpha \boldsymbol{h_{eff}} - \boldsymbol{h_s})] \tag{6.13}$$

where $\boldsymbol{h_{eff}} = \boldsymbol{H_{eff}}/H_K$ and $\boldsymbol{h_s} = \boldsymbol{H_s}/H_K$. (6.13) is similar to the voltage across a capacitor since $C\frac{dV}{dt} = I$. Thus, the order of magnitude of C can be estimated by $\frac{1+\alpha^2}{\gamma H_K} \approx 10^{-7}$ if $\alpha = 0.01$, $\gamma = 1.76 \times 10^{11}$ $\frac{\mathrm{rad}}{\mathrm{sT}}$, and $H_K = 100$ Oe $= \frac{10^5}{4\pi}$ A/m, which is confirmed by the transient simulation of the LLG equation from [185] as shown in Fig. 6.4.

6.2.4 Model Structure and Parameter Extraction

The resistances of P and AP states and critical currents are calculated analytically in previous section. The structure of STT-MTJ compact model is shown in Fig. 6.5. To properly extract the model parameters, it is recommended to follow the procedure below step-by-step.

Step 0: R_P (or RA and AREA) and Δ are the instance parameters that are determined by the technology.

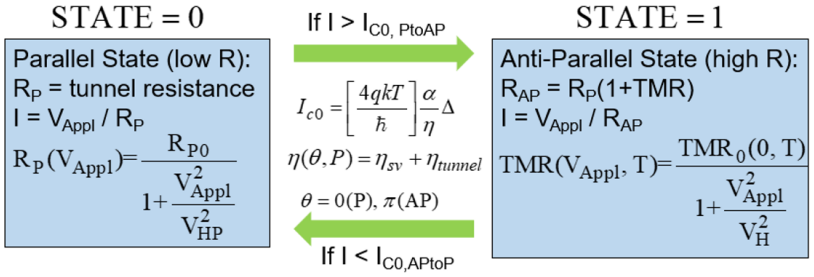


Figure 6.5: Compact model flow of STT-MTJ model.

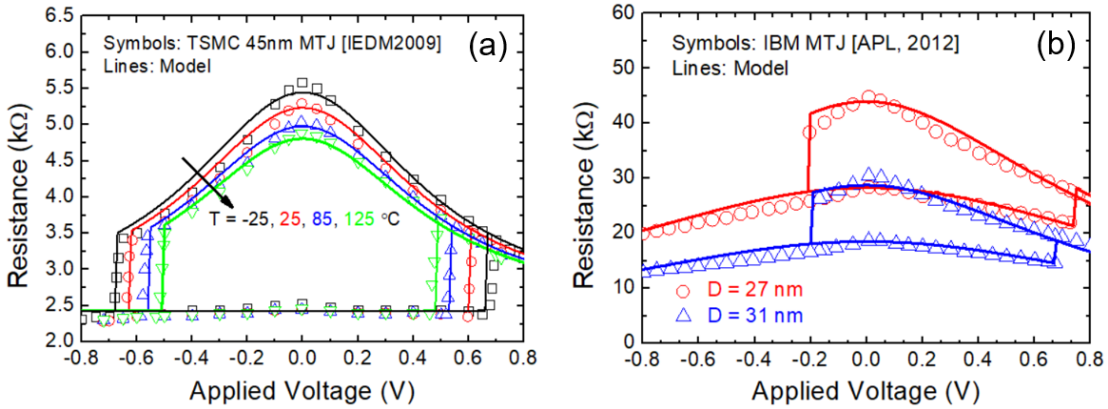


Figure 6.6: Comparison of resistance hysteresis loops of model and silicon data of (a) TSMC [41] and (b) IBM [35] MTJs with various temperatures and diameters (D).

Step 1: At T = 300 K, observe resistance ratio of two states at $V_{\text{apply}} = 0$ to extract TMR0, and then observe V_{apply} at which resistances are half of that at $V_{\text{apply}} = 0$ for AP (V_{H}) and P (V_{HP}) states.

Step 2: At T = 300 K, observe V_{apply} for AP to P and P to AP to get α and P. Parameter P changes the ratio of critical current for these two states.

Step 3: Observe resistances at $V_{\text{apply}} = 0$ for different temperatures to obtain TMRS, and then observe V_{apply} for AP to P and P to AP at different temperatures to extract DELTAS.

Step 4: Start with $C = \frac{1+\alpha^2}{\gamma H_K}$ (the default is 10^{-7}) using the material property parameters, and observe transient switching of MTJ either from P to AP or AP to P by fine tuning C to extract R.

6.2.5 Model Validation

By follwoing the parameter extraction procedure in the previous section, the proposed model has been validated with the experimental data of TSMC and IBM MTJs [41, 35] as shown in Fig. 6.6. The temperature effect in resistance is successfully captured since the TMR and critical current are temperature-dependent as discussed in Section 6.2.1 and 6.2.2. The model also shows scalability nature because the resistance is scaled with the cell area. The asymmetric switching is evident in Fig. 6.6 (b), which is modeled by the STT-efficiency factor η (6.8). To exhibit the capability of simulating circuit for compact purpose, an 1Transistor-1MTJ MRAM cell is simulated using the proposed MTJ model and BSIM-CMG. The transistor model is calibrated with 14 nm

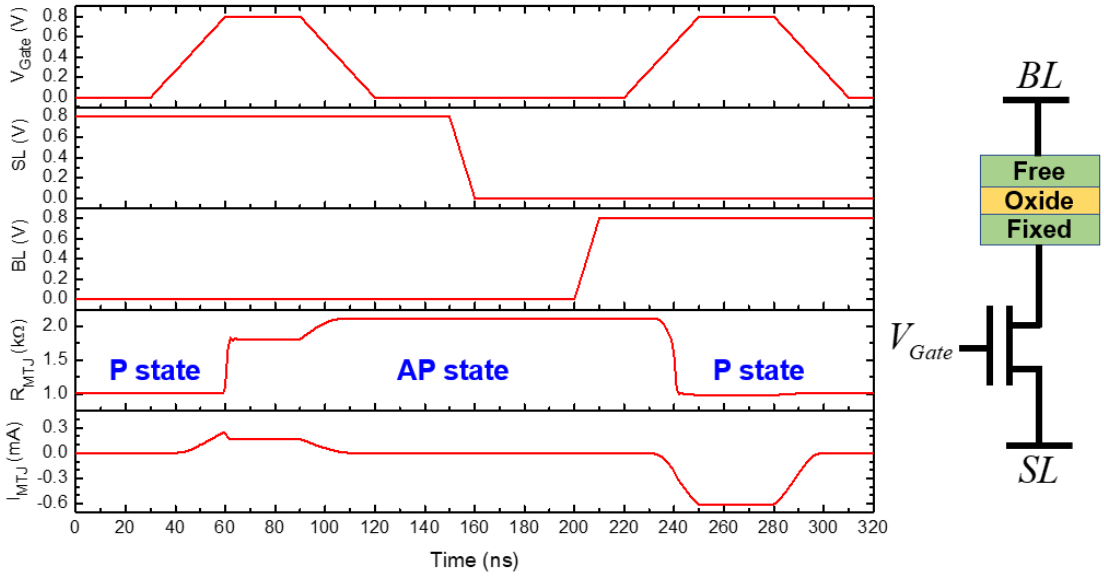


Figure 6.7: Trasient bahavior of MTJ using an 1Transistor-1MTJ cell. The key model parameters of MTJ are TMR = 115%, $\Delta = 51$, and $R_P = 1000\ \Omega$. 10 fins are used in the FinFET model.

FinFET silicon data [111]. Fig. 6.7 shows the transient switching behavior of MTJ. The MTJ state is flipped successfully from P to AP states and from AP to P states if reversing the bias in source line (SL) and bit line (BL) with turning ON the transistor. The memory characteristics are clearly obtained.

6.3 Conclusion

The compact model of MTJ is proposed in this Chapter. The analytical critical current is simplified from the LLG equation, and the zero bias anomaly and temperature effect in TMR is also modeled based on the hot electron-induced magnons model. To capture the dynamic switching due to the precession and thermal fluctuation, the RC sub-circuit is used by deriving from the LLG equation, which has been confirmed directly with the LLG solution. Finally, the compact model is validated by the silicon data from the industry, showing the capability of capturing the temperature dependence and MTJ size scaling.

Chapter 7

Conclusions

7.1 Chapters Summary

Chapter 2 presented the back-gate bias-dependent effects in FDSOI devices for different flavors of circuit applications, which results from the fact that the electrostatic in channel is affected by the back-gate bias. The GIDL and gate leakages are modeled using an analytical equation with back-gate bias. The proposed compact models have been validated with the experimental data from an industry lab. In addition to the leakages, the carrier transport property, mobility, is also exhibiting the back-gate bias-dependence. The back side inversion effect accounts for this dependence since it creates a second channel. This model is crucial to capture the transconductance behaviors with a wide range of back-gate biases, especially in FDSOIs with thick front-gate oxide where the back channel is comparable to the front channel. The non-monotonic trend of DIBL with back-gate biases is also found in such thick front-gate oxide devices. The channel length modulation is found to be the root cause. Both the fringing fields through the front-gate oxide and buried oxide are responsible for the DIBL degradation with back-gate biases, which is verified by TCAD simulations. Finally, the back-gate bias-dependent overlap capacitance is modeled by introducing the back-gate bias into the voltage dropped in the overlap region and is validated with TCAD simulation and experimental data.

The leakage models in extremely scaled MOSFET and FinFET are the focus in Chapter 3. In the zero threshold voltage devices such as I/O devices, the subsurface leakage may be important since the substrate doping is not high enough to avoid the electrostatic control from the drain bias. This subsurface leakage is modeled using the concept of barrier lowering and it is a function of drain bias, substrate doping, and temperature, which has been validated with the TCAD simulated data. To broaden the horizon of the understanding in ultra-scaled FinFET-based circuits, the source-to-drain tunneling current model, which plays an important role in sub-10 nm devices, is investigated. The potential barrier in channel is assumed to be quadratic, leading to a close form result of WKB-based tunneling probability. By adopting the numerical technique of Gaussian quadrature method, the integration result of the Landauer's equation is obtained. However, for compact model purpose of leakage models, the equation is further simplified assuming Boltzmann statistics, showing a good computational efficiency and an excellent agreement with the atomistic simulation data.

To achieve the urgent needs of low power chips in future technology nodes, a

new paradigm of steep subthreshold slope devices - Tunnel FET (TFET) is proposed. Chapter 4 presented the compact model of TFETs. The surface potentials of source and channel regions are solved analytically. The close form result of WKN-based tunneling probability using the derived surface potentials is obtained. To overcome the brute force of integration in Landauer's equation, the Gaussian quadrature method is again adopted. It is shown that the computational efficiency, smoothness, and accuracy are satisfactory. Finally, the developed model is validated with the atomistic simulation data with various materials and channel cross-sections.

In Chapter 5, another steep subthreshold slope device - negative capacitance FET (NCFET) is discussed in terms of compact model, ferroelectrics variability, and device design. The enhancement of existent compact model in short channel effect is proposed. The TCAD simulations suggest that the change in OFF current when varying P_r in short channel NCFET is opposite to that in long channel NCFET because of the inner fringing fields-induced gate charges, which can be taken into account by applying a quasi-2D position-dependent gate charge model to the Gaussian quadrature-based current calculation. The updated NCFET compact models are validated by the TCAD simulation date with wide range of P_r and L_G. In addition to the compact model, the variability in ferroelectric materials, such FE properties variation and existent dielectric phases, is also explored using Monte Carlo TCAD simulations. In comparison with FinFET variation sources, the ferroelectric variations should be well controlled that $\sigma P_r/\mu P_r < 27\%$ and DE percentage $< 20\%$. Finally, the spacer design of NC-FinFET is discussed. It is found that the fin selective spacer can help NC-FinFET via capacitance matching in propagation delay, while the fin corner is better in FinFET due to the reduction of parasitic capacitance. However, if taking into account the process complexity, the air spacer would be the best choice where the delay is not degraded much compared with the fin selective spacer design.

In addition to the transistor models, the memory compact model for spin-transfer-torque magnetic tunnel junction (STT-MTJ) MRAM is the emphasis in Chapter 6 to fulfill the need in cirecuit designs. The core model consists of resistances of parallel and anti-parallel stats and critical current for switching using the analytical equations derived from the Landau-Lifshitz-Gilbert (LLG) equation. To capture the dynamic switching behavior due to the precession and thermal fluctuation, for compact model purpose a RC sub-circuit based on the LLG equation is utilized and is verified with the LLG solution. The parameter extraction procedure is also suggested, showing that the compact model is accurate to describe the silicon data from the industrial devices.

7.2 Future Works

Since a portion of this dissertation focuses on enhancing the existent BSIM compact model, more real device effects are expected to appear and be modeled in future technology nodes. In FDSOIs, the back-gate bias effect is still an important aspect. For example, the parasitic resistance and capacitance could show back-gate bias-dependence and need to be modeled predicatively. In extremely scaled multi-gate devices such as naonsheet devices which show more visible quantum confinement, the source-to-drain tunneling equation might be modified since the sub-band effect changes the tunneling scheme. However, for compact model purpose, considering the complex-

ity of band structure will make the model significantly computationally inefficient and be unsuitable for SPICE simulation.

In Chapter 4, the homojunction TFET is emphasized. However, the heterojunction TFET has been shown that the tunneling is more efficient and thus higher ON current. Therefore, the heterojunction TFET compact model is required. Basically, the developed homojunction TFET model can be extended to heterojunction TFET since they have similar band structure except the band offset. Nevertheless, this idea needs to be further checked and tested since it might cause the discontinuity in the integration and diverge. Furthermore, the carrier transport effect in MOSFET part of TFET should be also modeled if the short channel effect is significant and thus large lateral electric field. Additional internal node is needed for the calculation, and it is prone to degrade the speed.

In Chapter 5, since only the polynomial terms are considered in the Landau-Khalatnikov equation. The domain interaction and viscosity terms are ignored. To best the predictive nature of compact model in transient simulation, the viscosity effect in FE is important and should be properly modeled. In variability study, the separated FE variation sources are explored. Indeed, the correlation between the grain size and FE properties still remains unknown and is required further investigation. In NC-FinFET spacer design, the doping profile in the source/drain extension region affects the distribution of electric fields, and thus the co-optimization of doping profile and size of Si_3N_4 region might be a good future work.

The STT-MTJ operation and compact model are discussed in Chapter 6. However, to continue scaling the MTJ MRAM, the spin-orbit-torque MTJ (SOT-MTJ) is paid much attention in spintronics community. The structure of compact model of SOT-MTJ might be the same as that of STT-MTJ. The critical current can be accordingly modified based on the LLG equation. Nevertheless, nowaday there is no abundant experimental data of SOT-MTJ. Once the data is reliable, the SOT-MTJ compact model thus can be validated and other real device effects are able to be included properly.

Bibliography

[1] BSIM Group. `http://bsim.berkeley.edu/`. Accessed: 2019-02-19.

[2] Synopsys, Mountain View, CA, USA. *Sentaurus Device User Guide, Version O-2018.06*, Sep. 2018.

[3] Y. S. Chauhan, S. Venugopalan, M. Chalkiadaki, M. A. U. Karim, H. Agarwal, S. Khandelwal, N. Paydavosi, J. P. Duarte, C. C. Enz, A. M. Niknejad, and C. Hu. BSIM6: Analog and RF Compact Model for Bulk MOSFET. *IEEE Transactions on Electron Devices*, 61(2):234–244, Feb. 2014.

[4] Sriramkumar Venugopalan. *From Poisson to Silicon - Advancing Compact SPICE Models for IC Design*. PhD thesis, University of California, Berkeley, 2013.

[5] D. Hisamoto, W.-C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, E. Anderson, T.-J. King, J. Bokor, and C. Hu. FinFET-a self-aligned double-gate MOSFET scalable to 20 nm. *IEEE Transactions on Electron Devices*, 47(12):2320–2325, Dec. 2000.

[6] F. L. Yang et al. 5nm-gate nanowire FinFET. In *Symposium on VLSI Technology,*, pages 196–197, Jun. 2004.

[7] J. P. Duarte, S. Khandelwal, A. I. Khan, A. Sachid, Y. Lin, H. Chang, S. Salahuddin, and C. Hu. Compact models of negative-capacitance FinFETs: Lumped and distributed charge models. In *2016 IEEE International Electron Devices Meeting*, pages 30.5.1–30.5.4, Dec. 2016.

[8] A. Vladimirescu and S. Liu. The Simulation of MOS Integrated Circuits Using SPICE2. Technical report, EECS Department, University of California, Berkeley, 1980.

[9] B. Sheu, D. Scharfetter, C. Hu, and D. Pederson. A compact IGFET charge model. *IEEE Transactions on Circuits and Systems*, 31(8):745–748, Aug. 1984.

[10] N. Paydavosi, S. Venugopalan, Y. S. Chauhan, J. P. Duarte, S. Jandhyala, A. M. Niknejad, and C. C. Hu. BSIMSPICE Models Enable FinFET and UTB IC Designs. *IEEE Access*, 1:201–215, 2013.

[11] Department of Electrical Engineering and Computer Science, UC Berkeley, Berkeley, CA, USA. *BSIM4v4.8.0 Technical Manual*, 2013.

[12] S. Khandelwal, Y. S. Chauhan, D. D. Lu, S. Venugopalan, M. Ahosan Ul Karim, A. B. Sachid, B. Nguyen, O. Rozeau, O. Faynot, A. M. Niknejad, and C. C. Hu. BSIM-IMG: A Compact Model for Ultrathin-Body SOI MOSFETs With Back-Gate Control. *IEEE Transactions on Electron Devices*, 59(8):2019–2026, Aug. 2012.

[13] J. P. Duarte, S. Khandelwal, H. L. Chang, Y. K. Lin, P. Kushwaha, Y. S. Chauhan, and C. Hu. Modeling independent multi-gate MOSFETs. In *Workshop of Compact Modeling (WCM)*, May 2016.

[14] Y. S. Chauhan, D. D. Lu, S. Vanugopalan, S. Khandelwal, J. P. Duarte, N. Paydavosi, A. Niknejad, and C. Hu. *FinFET Modeling for IC Simulation and Design*. Academic Press, 1st edition, 2015.

[15] J. P. Duarte, N. Paydavosi, S. Venugopalan, A. Sachid, and C. Hu. Unified FinFET compact model: Modelling Trapezoidal Triple-Gate FinFETs. In *2013 International Conference on Simulation of Semiconductor Processes and Devices*, pages 135–138, Sep. 2013.

[16] C. Hu. *Modern Semiconductor Devices for Integrated Circuits*. Pearson, 1st edition, 2009.

[17] J. Wang and M. Lundstrom. Ballistic transport in high electron mobility transistors. *IEEE Transactions on Electron Devices*, 50(7):1604–1609, Jul. 2003.

[18] W. C. Lee and C. Hu. Modeling CMOS tunneling currents through ultrathin gate oxide due to conduction- and valence-band electron and hole tunneling. *IEEE Transactions on Electron Devices*, 48(7):1366–1373, Jul. 2001.

[19] G. Ribes, J. Mitard, M. Denais, S. Bruyere, F. Monsieur, C. Parthasarathy, E. Vincent, and G. Ghibaudo. Review on high-k dielectrics reliability issues. *IEEE Transactions on Device and Materials Reliability*, 5(1):5–19, Mar. 2005.

[20] K. Mistry et al. A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging. In *2007 IEEE International Electron Devices Meeting*, pages 247–250, Dec. 2007.

[21] Y. K. Lin, S. Khandelwal, J. P. Duarte, H. L. Chang, S. Salahuddin, and C. Hu. A Predictive Tunnel FET Compact Model With Atomistic Simulation Validation. *IEEE Transactions on Electron Devices*, 64(2):599–605, Feb. 2017.

[22] N. Loubet et al. Stacked nanosheet gate-all-around transistor to enable scaling beyond FinFET. In *2017 Symposium on VLSI Technology*, pages T230–T231, Jun. 2017.

[23] Y. K. Choi, K. Asano, N. Lindert, V. Subramanian, T. J. King, J. Bokor, and C. Hu. Ultra-thin body SOI MOSFET for deep-sub-tenth micron era. In *1999 International Electron Devices Meeting*, pages 919–921, Dec. 1999.

[24] Y. K. Lin, P. Kushwaha, H. Agarwal, H. L. Chang, J. P. Duarte, A. B. Sachid, S. Khandelwal, S. Salahuddin, and C. Hu. Modeling of Back-Gate Effects on Gate-Induced Drain Leakage and Gate Currents in UTB SOI MOSFETs. *IEEE Transactions on Electron Devices*, 64(10):3986–3990, Oct. 2017.

[25] Darsen Lu. *Compact Models for Future Generation CMOS*. PhD thesis, University of California, Berkeley, 2011.

[26] D. Jin, D. Kim, T. Kim, and J. A. del Alamo. Quantum capacitance in scaled down IIIV FETs. In *2009 IEEE International Electron Devices Meeting*, pages 1–4, Dec. 2009.

[27] S. Natarajan et al. A 14nm logic technology featuring 2nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm^2SRAM cell size. In *2014 IEEE International Electron Devices Meeting*, pages 3.7.1–3.7.3, Dec. 2014.

[28] T. Masuhara. Quest for Low-Voltage and Low-Power Integrated Circuits: Towards a Sustainable Future. *IEEE Solid-State Circuits Magazine*, 5(1):8–26, winter 2013.

[29] O. Y. Loh and H. D. Espinosa. Nanoelectromechanical contact switches. *Nature Nanotechnology*, 7:283–295, May 2012.

[30] K. Gopalakrishnan, P. B. Griffin, and J. D. Plummer. I-MOS: a novel semiconductor device with a subthreshold slope lower than kT/q. In *2002 International Electron Devices Meeting,*, pages 289–292, Dec. 2002.

[31] W. Y. Choi, B. Park, J. D. Lee, and T. K. Liu. Tunneling Field-Effect Transistors (TFETs) With Subthreshold Swing (SS) Less Than 60 mV/dec. *IEEE Electron Device Letters*, 28(8):743–745, Aug. 2007.

[32] S. Salahuddin and S. Datta. Use of Negative Capacitance to Provide Voltage Amplification for Low Power Nanoscale Devices. *Nano Letters*, 8(2):405–410, 2008.

[33] Avirup Dasgupta. *Compact Modeling of Electrostatics, Quasi-ballistic Transport and Noise in MOSFETs and Magnetic Tunnel Junction*. PhD thesis, Indian Institute of Technology Kanpur, 2018.

[34] K. L. Wang, J. G. Alzate, and P. K. Amiri. Low-power non-volatile spintronic memory: STT-RAM and beyond. *Journal of Physics D: Applied Physics*, 46(7):074003, Feb. 2013.

[35] M. Gajek et al. Spin torque switching of 20nm magnetic tunnel junctions with perpendicular anisotropy. *Applied Physics Letters*, 100(13):132408, 2012.

[36] A. V. Khvalkovskiy et al. Basic principles of STT-MRAM cell operation in memory arrays. *Journal of Physics D: Applied Physics*, 46(7):074001, Feb. 2013.

[37] J. J. Nowak et al. Dependence of Voltage and Size on Write Error Rates in Spin-Transfer Torque Magnetic Random-Access Memory. *IEEE Magnetics Letters*, 7:1–4, 2016.

[38] Y. J. Song et al. Highly functional and reliable 8Mb STT-MRAM embedded in 28nm logic. In *2016 IEEE International Electron Devices Meeting*, pages 27.2.1–27.2.4, Dec. 2016.

[39] D. Shum et al. CMOS-embedded STT-MRAM arrays in 2x nm nodes for GP-MCU applications. In *2017 Symposium on VLSI Technology*, pages T208–T209, Jun. 2017.

[40] J. J. Kan et al. Systematic validation of 2x nm diameter perpendicular MTJ arrays and MgO barrier for sub-10 nm embedded STT-MRAM with practically unlimited endurance. In *2016 IEEE International Electron Devices Meeting*, pages 27.4.1–27.4.4, Dec. 2016.

[41] C. J. Lin et al. 45nm low power CMOS logic compatible embedded STT MRAM utilizing a reverse-connection 1T/1MTJ cell. In *2009 IEEE International Electron Devices Meeting*, pages 1–4, Dec. 2009.

[42] Y. K. Lin, P. Kushwaha, J. P. Duarte, H. Chang, H. Agarwal, S. Khandelwal, A. B. Sachid, M. Harter, J. Watts, Y. S. Chauhan, S. Salahuddin, and C. Hu. New Mobility Model for Accurate Modeling of Transconductance in FD-SOI MOSFETs. *IEEE Transactions on Electron Devices*, 65(2):463–469, Feb. 2018.

[43] V. Suntharalingam et al. Megapixel CMOS image sensor fabricated in three-dimensional integrated circuit technology. In *2005 IEEE International Solid-State Circuits Conference*, pages 356–357, Feb. 2005.

[44] P. Coudrain, P. Magnan, P. Batude, X. Gagnard, C. Leyris, M. Vinet, A. Castex, C. Lagahe-Blanchard, A. Pouydebasque, Y. Cazaux, B. Giffard, and P. Ancey. Investigation of a Sequential Three-Dimensional Process for Back-Illuminated CMOS Image Sensors With Miniaturized Pixels. *IEEE Transactions on Electron Devices*, 56(11):2403–2413, Nov. 2009.

[45] M. Goto, K. Hagiwara, Y. Iguchi, H. Ohtake, T. Saraya, M. Kobayashi, E. Higurashi, H. Toshiyoshi, and T. Hiramoto. Pixel-Parallel 3-D Integrated CMOS Image Sensors With Pulse Frequency Modulation A/D Converters Developed by Direct Bonding of SOI Layers. *IEEE Transactions on Electron Devices*, 62(11):3530–3535, Nov. 2015.

[46] A. Litty, S. Ortolland, D. Golanski, and S. Cristoloveanu. Dual Ground Plane EDMOS in ultrathin FDSOI for 5V energy management applications. In *2014 44th European Solid State Device Research Conference*, pages 134–137, Sep. 2014.

[47] S. Cristoloveanu, N. Rodriguez, and F. Gamiz. Why the Universal Mobility Is Not. *IEEE Transactions on Electron Devices*, 57(6):1327–1333, Jun. 2010.

[48] C. Navarro, N. Rodriguez, A. Ohata, F. Gamiz, F. Andrieu, C. Fenouillet-Beranger, O. Faynot, and S. Cristoloveanu. Multibranch Mobility Analysis for the Characterization of FDSOI Transistors. *IEEE Electron Device Letters*, 33(8):1102–1104, Aug. 2012.

[49] E. R. Fossum. CMOS image sensors: electronic camera-on-a-chip. *IEEE Transactions on Electron Devices*, 44(10):1689–1698, Oct. 1997.

[50] J. Chen, T. Y. Chan, I. C. Chen, P. K. Ko, and C. Hu. Subbreakdown drain leakage current in MOSFET. *IEEE Electron Device Letters*, 8(11):515–517, Nov. 1987.

[51] P. Kushwaha, N. Paydavosi, S. Khandelwal, C. Yadav, H. Agarwal, J. P. Duarte, C. Hu, and Y. S. Chauhan. Modeling the impact of substrate depletion in FDSOI MOSFETs. *Solid-State Electronics*, 104:6–11, 2015.

[52] K. M. Cao et al. BSIM4 gate leakage model including source-drain partition. In *2000 International Electron Devices Meeting*, pages 815–818, Dec. 2000.

[53] A. Nazarov, J. P. Colinge, F. Balestra, J. P. Raskin, F. Gamiz, and V. S. Lysenko. *Semiconductor-On-Insulator Materials for Nanoelectronics Applications*. Springer, 1st edition, 2011.

[54] Y. T. Hou, M. F. Li, T. Low, and D. L. Kwong. Metal gate work function engineering on gate leakage of MOSFETs. *IEEE Transactions on Electron Devices*, 51(11):1783–1789, Nov. 2004.

[55] H. Agarwal, P. Kushwaha, Y. S. Chauhan, S. Khandelwal, J. P. Duarte, Y. K. Lin, H. L. Chang, C. Hu, H. Wu, and P. D. Ye. Modeling of GeOI and validation with Ge-CMOS inverter circuit using BSIM-IMG industry standard model. In *2016 IEEE International Conference on Electron Devices and Solid-State Circuits*, pages 444–447, Aug. 2016.

[56] A. B. Sachid, N. Paydovosi, S. Khandelwal, and C. Hu. Multi-gate MOSFET with electrically tunable VT for power management. In *2013 International Semiconductor Device Research Symposium*, pages 1–2, Dec. 2013.

[57] H. Agarwal, C. Gupta, S. Dey, S. Khandelwal, C. Hu, and Y. S. Chauhan. Anomalous Transconductance in Long Channel Halo Implanted MOSFETs: Analysis and Modeling. *IEEE Transactions on Electron Devices*, 64(2):376–383, Feb. 2017.

[58] T. Rudenko, V. Kilchytska, S. Burignat, J.-P. Raskin, F. Andrieu, O. Faynot, Y. Le Tiec, K. Landry, A. Nazarov, V. S. Lysenko, and D. Flandre. Experimental study of transconductance and mobility behaviors in ultra-thin SOI MOSFETs with standard and thin buried oxides. *Solid-State Electronics*, 54(2):164–170, 2010.

[59] P. Kushwaha, H. Agarwal, Y. S. Chauhan, M. Bhoir, N. R. Mohapatra, S. Khandelwal, J. P. Duarte, Y. K. Lin, H. L. Chang, and C. Hu. Predictive effective mobility model for FDSOI transistors using technology parameters. In *2016*

IEEE International Conference on Electron Devices and Solid-State Circuits, pages 448–451, Aug. 2016.

[60] S. Khandelwal, J. P. Duarte, Y. S. Chauhan, and C. Hu. Modeling 20-nm Germanium FinFET With the Industry Standard FinFET Model. *IEEE Electron Device Letters*, 35(7):711–713, Jul. 2014.

[61] Y. Sahu, P. Kushwaha, A. Dasgupta, C. Hu, and Y. S. Chauhan. Compact Modeling of Drain Current Thermal Noise in FDSOI MOSFETs Including Back-Bias Effect. *IEEE Transactions on Microwave Theory and Techniques*, 65(7):2261–2270, Jul. 2017.

[62] T. Poiroux, O. Rozeau, P. Scheer, S. Martinie, M. A. Jaud, M. Minondo, A. Juge, J. C. Barb, and M. Vinet. Leti-UTSOI2.1: A Compact Model for UTBB-FDSOI TechnologiesPart I: Interface Potentials Analytical Model. *IEEE Transactions on Electron Devices*, 62(9):2751–2759, Sep. 2015.

[63] T. Poiroux, O. Rozeau, P. Scheer, S. Martinie, M. Jaud, M. Minondo, A. Juge, J. C. Barb, and M. Vinet. Leti-UTSOI2.1: A Compact Model for UTBB-FDSOI TechnologiesPart II: DC and AC Model Description. *IEEE Transactions on Electron Devices*, 62(9):2760–2768, Sep. 2015.

[64] Y. Tsividis and C. McAndrew. *Operation and Modeling of the MOS Transitor*. Oxford University Press, 3rd edition, 2011.

[65] A. Ortiz-Conde and F. J. G. Snchez. Unification of asymmetric DG, symmetric DG and bulk undoped-body MOSFET drain current. *Solid-State Electronics*, 50(11):1796–1800, 2006.

[66] T. Poiroux, O. Rozeau, S. Martinie, P. Scheer, S. Puget, M. A. Jaud, S. E. Ghouli, J. C. Barb, A. Juge, and O. Faynot. UTSOI2: A complete physical compact model for UTBB and independent double gate MOSFETs. In *2013 IEEE International Electron Devices Meeting*, pages 12.4.1–12.4.4, Dec. 2013.

[67] S. Takagi, M. Takayanagi, and A. Toriumi. Characterization of inversion-layer capacitance of holes in Si MOSFET's. *IEEE Transactions on Electron Devices*, 46(7):1446–1450, Jul. 1999.

[68] A. Majumdar, C. Ouyang, S. J. Koester, and W. Haensch. Effects of Substrate Orientation and Channel Stress on Short-Channel Thin SOI MOSFETs. *IEEE Transactions on Electron Devices*, 57(9):2067–2072, Sep. 2010.

[69] Q. Liu et al. Impact of back bias on ultra-thin body and BOX (UTBB) devices. In *2011 Symposium on VLSI Technology*, pages 160–161, Jun. 2011.

[70] L. Grenouillet et al. UTBB FDSOI transistors with dual STI for a multi-Vtstrategy at 20nm node and below. In *2012 International Electron Devices Meeting*, pages 3.6.1–3.6.4, Dec. 2012.

[71] C. Yu and P. Su. Built-in Effective Body-Bias Effect in Ultra-Thin-Body Hetero-Channel IIIV-on-Insulator n-MOSFETs. *IEEE Electron Device Letters*, 35(8):823–825, Aug. 2014.

[72] V. P. Trivedi and J. G. Fossum. Nanoscale FD/SOI CMOS: thick or thin BOX? *IEEE Electron Device Letters*, 26(1):26–28, Jan. 2005.

[73] T. Ernst, R. Ritzenthaler, O. Faynot, and S. Cristoloveanu. A Model of Fringing Fields in Short-Channel Planar and Triple-Gate SOI MOSFETs. *IEEE Transactions on Electron Devices*, 54(6):1366–1375, Jun. 2007.

[74] Q. Xie, C. Lee, J. Xu, C. Wann, J. Y. C. Sun, and Y. Taur. Comprehensive Analysis of Short-Channel Effects in Ultrathin SOI MOSFETs. *IEEE Transactions on Electron Devices*, 60(6):1814–1819, Jun. 2013.

[75] B. H. Lee et al. Ultrathin hafnium oxide with low leakage and excellent reliability for alternative gate dielectric application. In *1999 International Electron Devices Meeting*, pages 133–136, Dec. 1999.

[76] M. K. M. Arshad, J. Raskin, V. Kilchytska, F. Andrieu, P. Scheiblin, O. Faynot, and D. Flandre. Extended MASTAR Modeling of DIBL in UTB and UTBB SOI MOSFETs. *IEEE Transactions on Electron Devices*, 59(1):247–251, Jan. 2012.

[77] S. A. Parke, J. E. Moon, H. C. Wann, P. K. Ko, and C. Hu. Design for suppression of gate-induced drain leakage in LDD MOSFETs using a quasi-two-dimensional analytical model. *IEEE Transactions on Electron Devices*, 39(7):1694–1703, Jul. 1992.

[78] T. Wang, T. E. Chang, C. M. Huang, J. Y. Yang, K. M. Chang, and L. P. Chiang. Structural effect on band-trap-band tunneling induced drain leakage in n-MOSFET's. *IEEE Electron Device Letters*, 16(12):566–568, Dec. 1995.

[79] T. Wang, T. E. Chang, C. M. Huang, J. Y. Yang, K. M. Chang, and L. P. Chiang. A comprehensive study of hot carrier stress-induced drain leakage current degradation in thin-oxide n-MOSFETs. *IEEE Transactions on Electron Devices*, 46(9):1877–1882, Sep. 1999.

[80] L. Huang, P. T. Lai, J. P. Xu, and Y. C. Cheng. Mechanism analysis of gate-induced drain leakage in off-state n-MOSFET. *Microelectronics Reliability*, 38(9):1425–1431, 1998.

[81] J. Zhu, R. A. Martin, and J. Y. Chen. Punchthrough current for submicrometer MOSFETs in CMOS VLSI. *IEEE Transactions on Electron Devices*, 35(2):145–151, Feb. 1988.

[82] V. P. Hu, M. Fan, P. Su, and C. Chuang. Band-to-Band-Tunneling Leakage Suppression for Ultra-Thin-Body GeOI MOSFETs Using Transistor Stacking. *IEEE Electron Device Letters*, 33(2):197–199, Feb. 2012.

[83] V. Pako, S. Ban, D. Prejda, and J. Dobe. MOSFET gate dimension dependent drain and source leakage modeling by standard SPICE models. *Solid-State Electronics*, 81:144–150, 2013.

[84] G. Eneman, G. Hellings, A. De Keersgieter, N. Collaert, and A. Thean. Quantum-barriers and ground-plane isolation: A path for scaling bulk-FinFET technologies to the 7 nm-node and beyond. In *2013 IEEE International Electron Devices Meeting*, pages 12.3.1–12.3.4, Dec. 2013.

[85] S. Wu et al. A 7nm CMOS platform technology featuring 4th generation FinFET transistors with a 0.027um2 high density 6-T SRAM cell for mobile SoC applications. In *2016 IEEE International Electron Devices Meeting*, pages 2.6.1–2.6.4, Dec. 2016.

[86] D. S. Huang et al. Comprehensive device and product level reliability studies on advanced CMOS technologies featuring 7nm high-k metal gate FinFET transistors. In *2018 IEEE International Reliability Physics Symposium*, pages 6F.7–1–6F.7–5, Mar. 2018.

[87] J. Wang and M. Lundstrom. Does source-to-drain tunneling limit the ultimate scaling of MOSFETs? In *2002 International Electron Devices Meeting*, pages 707–710, Dec. 2002.

[88] N. Seoane and A. Martinez. A detailed coupled-mode-space non-equilibrium Green's function simulation study of source-to-drain tunnelling in gate-all-around Si nanowire metal oxide semiconductor field effect transistors. *Journal of Applied Physics*, 114(10):104307, 2013.

[89] M. Luisier, M. Lundstrom, D. A. Antoniadis, and J. Bokor. Ultimate device scaling: Intrinsic performance comparisons of carbon-based, InGaAs, and Si field-effect transistors for 5 nm gate length. In *2011 International Electron Devices Meeting*, pages 11.2.1–11.2.4, Dec. 2011.

[90] Q. Rafhay, R. Clerc, G. Ghibaudo, and G. Pananakakis. Impact of source-to-drain tunnelling on the scalability of arbitrary oriented alternative channel material nMOSFETs. *Solid-State Electronics*, 52(10):1474–1481, 2008.

[91] T. Ohguro et al. 0.18 μm low voltage/low power RF CMOS with zero Vth analog MOSFETs made by undoped epitaxial channel technique. In *1997 International Electron Devices Meeting*, pages 837–840, Dec. 1997.

[92] Steven H. Voldman. *ESD Circuits and Devices*. Wiley, 2nd edition, 2015.

[93] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand. Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits. *Proceedings of the IEEE*, 91(2):305–327, Feb. 2003.

[94] Simon M. Sze and Kwok K. Ng. *Physics of Semiconductor Devices*. Wiley, 3rd edition, 2007.

[95] A. Szabo and M. Luisier. Under-the-Barrier Model: An Extension of the Top-of-the-Barrier Model to Efficiently and Accurately Simulate Ultrascaled Nanowire Transistors. *IEEE Transactions on Electron Devices*, 60(7):2353–2360, Jul. 2013.

[96] M. Luisier and G. Klimeck. Simulation of nanowire tunneling transistors: From the WentzelKramersBrillouin approximation to full-band phonon-assisted tunneling. *Journal of Applied Physics*, 107(8):084507, 2010.

[97] R. B. Salazar, H. Ilatikhameneh, R. Rahman, G. Klimeck, and J. Appenzeller. A predictive analytic model for high-performance tunneling field-effect transistors approaching non-equilibrium Green's function simulations. *Journal of Applied Physics*, 118(16):164305, 2015.

[98] D. Esseni, M. G. Pala, and T. Rollo. Essential Physics of the OFF-State Current in Nanoscale MOSFETs and Tunnel FETs. *IEEE Transactions on Electron Devices*, 62(9):3084–3091, Sep. 2015.

[99] H. Ilatikhameneh, R. B. Salazar, G. Klimeck, R. Rahman, and J. Appenzeller. From FowlerNordheim to Nonequilibrium Greens Function Modeling of Tunneling. *IEEE Transactions on Electron Devices*, 63(7):2871–2878, Jul. 2016.

[100] Y. Gao, T. Low, and M. Lundstrom. Possibilities for VDD = 0.1V logic using carbon-based tunneling field effect transistors. In *2009 Symposium on VLSI Technology*, pages 180–181, Jun. 2006.

[101] S. Heinze, J. Tersoff, R. Martel, V. Derycke, J. Appenzeller, and Ph. Avouris. Carbon Nanotubes as Schottky Barrier Transistors. *Physical Review Letters*, 89:106801, Aug. 2002.

[102] G. Hiblot, Q. Rafhay, F. Boeuf, and G. Ghibaudo. Analytical relationship between subthreshold swing of thermionic and tunnelling currents. *Electronics Letters*, 50(23):1745–1747, 2014.

[103] Z. H. Liu, C. Hu, J. H. Huang, T. Y. Chan, M. C. Jeng, P. K. Ko, and Y. C. Cheng. Threshold voltage model for deep-submicrometer MOSFETs. *IEEE Transactions on Electron Devices*, 40(1):86–95, Jan. 1993.

[104] J. P. Colinge. Multiple-gate SOI MOSFETs. *Solid-State Electronics*, 48(6):897–905, 2004.

[105] Z. Jiang, J. Wang, H. Park, A. Pham, N. Xu, Y. Lu, S. Jin, W. Choi, M. A. Pourghaderi, J. Kim, and K. Lee. Comprehensive Simulation Study of Direct Source-to-Drain Tunneling in Ultra-Scaled Si, Ge, and III-V DG-FETs. *IEEE Transactions on Electron Devices*, 64(3):945–952, Mar. 2017.

[106] A. Rahman, J. Guo, S. Datta, and M. S. Lundstrom. Theory of ballistic nanotransistors. *IEEE Transactions on Electron Devices*, 50(9):1853–1864, Sep. 2003.

[107] Y. A. Kruglyak. A generalized Landauer-Datta-Lundstrom electron transport model. *Russian Journal of Physical Chemistry A*, 88(11):1826–1836, Nov. 2014.

[108] B. Bhushan, K. Nayak, and V. R. Rao. DC Compact Model for SOI Tunnel Field-Effect Transistors. *IEEE Transactions on Electron Devices*, 59(10):2635–2642, Oct. 2012.

[109] Tarek I. Zohdi. *A Finite Element Primer for Beginners The Basics*. Springer, 1st edition, 2015.

[110] L. Zhang and M. Chan. SPICE Modeling of Double-Gate Tunnel-FETs Including Channel Transports. *IEEE Transactions on Electron Devices*, 61(2):300–307, Feb. 2014.

[111] C. H. Jan et al. A 14 nm SoC platform technology featuring 2nd generation Tri-Gate transistors, 70 nm gate pitch, 52 nm metal pitch, and 0.0499 μm^2 SRAM cells, optimized for low power, high performance and high density SoC products. In *2015 Symposium on VLSI Technology*, pages T12–T13, Jun. 2015.

[112] H. Lee et al. Sub-5nm All-Around Gate FinFET for Ultimate Scaling. In *2006 Symposium on VLSI Technology*, pages 58–59, Jun. 2006.

[113] H. Lu and A. Seabaugh. Tunnel Field-Effect Transistors: State-of-the-Art. *IEEE Journal of the Electron Devices Society*, 2(4):44–49, Jul. 2014.

[114] U. E. Avci, D. H. Morris, and I. A. Young. Tunnel Field-Effect Transistors: Prospects and Challenges. *IEEE Journal of the Electron Devices Society*, 3(3):88–95, May 2015.

[115] U. E. Avci, R. Rios, K. Kuhn, and I. A. Young. Comparison of performance, switching energy and process variations for the TFET and MOSFET in logic. In *2011 Symposium on VLSI Technology*, pages 124–125, Jun. 2011.

[116] C. Hu, D. Chou, P. Patel, and A. Bowonder. Green Transistor - A VDD Scaling Path for Future Low Power ICs. In *2008 International Symposium on VLSI Technology, Systems and Applications*, pages 14–15, Apr. 2008.

[117] G. Dewey et al. Fabrication, characterization, and physics of IIIV heterojunction tunneling Field Effect Transistors (H-TFET) for steep sub-threshold swing. In *International Electron Devices Meeting*, pages 33.6.1–33.6.4, Dec. 2011.

[118] R. Vishnoi and M. J. Kumar. Compact Analytical Drain Current Model of Gate-All-Around Nanowire Tunneling FET. *IEEE Transactions on Electron Devices*, 61(7):2599–2603, Jul. 2014.

[119] R. Vishnoi and M. J. Kumar. An Accurate Compact Analytical Model for the Drain Current of a TFET From Subthreshold to Strong Inversion. *IEEE Transactions on Electron Devices*, 62(2):478–484, Feb. 2015.

[120] H. Lu, D. Esseni, and A. Seabaugh. Universal analytic model for tunnel FET circuit simulation. *Solid-State Electronics*, 108:110–117, 2015.

[121] Supriyo Datta. *Quantum Transport: Atom to Transistor*. Cambridge University Press, 1st edition, 2005.

[122] J. Wu, J. Min, and Y. Taur. Short-Channel Effects in Tunnel FETs. *IEEE Transactions on Electron Devices*, 62(9):3019–3024, Sep. 2015.

[123] Y. Taur, J. Wu, and J. Min. Dimensionality Dependence of TFET Performance Down to 0.1 V Supply Voltage. *IEEE Transactions on Electron Devices*, 63(2):877–880, Feb. 2016.

[124] L. Zhang, X. Lin, J. He, and M. Chan. An Analytical Charge Model for Double-Gate Tunnel FETs. *IEEE Transactions on Electron Devices*, 59(12):3217–3223, Dec. 2012.

[125] Y. Taur, J. Wu, and J. Min. An Analytic Model for Heterojunction Tunnel FETs With Exponential Barrier. *IEEE Transactions on Electron Devices*, 62(5):1399–1404, May 2015.

[126] B. Rajamohanan, D. Mohata, A. Ali, and S. Datta. Insight into the output characteristics of III-V tunneling field effect transistors. *Applied Physics Letters*, 102(9):092105, 2013.

[127] S. Agarwal and E. Yablonovitch. Band-Edge Steepness Obtained From Esaki/Backward Diode CurrentVoltage Characteristics. *IEEE Transactions on Electron Devices*, 61(5):1488–1493, May 2014.

[128] S. John, C. Soukoulis, M. H. Cohen, and E. N. Economou. Theory of Electron Band Tails and the Urbach Optical-Absorption Edge. *Physical Review Letters*, 57:1777–1780, Oct. 1986.

[129] M. Abul Khayer and Roger K. Lake. Effects of band-tails on the subthreshold characteristics of nanowire band-to-band tunneling transistors. *Journal of Applied Physics*, 110(7):074508, 2011.

[130] M. H. Cohen, M. Y. Chou, E. N. Economou, S. John, and C. M. Soukoulis. Band tails, path integrals, instantons, polarons, and all that. *IBM Journal of Research and Development*, 32(1):82–92, Jan. 1988.

[131] A. Iribarren, R. Castro-Rodríguez, V. Sosa, and J. L. Peña. Band-tail parameter modeling in semiconductor materials. *Physical Review B*, 58:1907–1911, Jul. 1998.

[132] U. E. Avci et al. Energy efficiency comparison of nanowire heterojunction TFET and Si MOSFET at Lg=13nm, including P-TFET and variation considerations. In *2013 IEEE International Electron Devices Meeting*, pages 33.4.1–33.4.4, Dec. 2013.

[133] M. Luisier and G. Klimeck. Simulation of nanowire tunneling transistors: From the WentzelKramersBrillouin approximation to full-band phonon-assisted tunneling. *Journal of Applied Physics*, 107(8):084507, 2010.

[134] M. Levinshtein, S. Rumyantsev, and M. Shur. *Handbook Series on Semiconductor Parameters,*. World Scientific, 1st edition, 2000.

[135] S. Gupta et al. GeSn technology: Extending the Ge electronics roadmap. In *2011 International Electron Devices Meeting*, pages 16.6.1–16.6.4, Dec. 2011.

[136] J. Wang, A. Rahman, A. Ghosh, G. Klimeck, and M. Lundstrom. On the validity of the parabolic effective-mass approximation for the I-V calculation of silicon nanowire transistors. *IEEE Transactions on Electron Devices*, 52(7):1589–1595, Jul. 2005.

[137] L. De Michielis, L. Lattanzio, and A. M. Ionescu. Understanding the Superlinear Onset of Tunnel-FET Output Characteristic. *IEEE Electron Device Letters*, 33(11):1523–1525, Nov. 2012.

[138] S. Salahuddin and S. Datta. Use of Negative Capacitance to Provide Voltage Amplification for Low Power Nanoscale Devices. *Nano Letters*, 8(2):405–410, 2008.

[139] Korok Chatterjee. *Design and Characterization of Ferroelectric Negative Capacitance*. PhD thesis, University of California, Berkeley, 2018.

[140] M. Hoffmann, M. Pei, S. Slesazeck, U. Schroeder, and T. Mikolajick. On the stabilization of ferroelectric negative capacitance in nanoscale devices. *Nanoscale*, 10:10891–10899, 2018.

[141] M. Hoffmann, B. Max, T. Mittmann, U. Schroeder, S. Slesazeck, and T. Mikolajick. Demonstration of High-speed Hysteresis-free Negative Capacitance in Ferroelectric $Hf_{0.5}Zr_{0.5}O_2$. In *2018 IEEE International Electron Devices Meeting*, pages 31.6.1–31.6.4, Dec. 2018.

[142] M. Hoffmann, F. P. G. Fengler, M. Herzig, T. Mittmann, B. Max, U. Schroeder, R. Negrea, P. Lucian, S. Slesazeck, and T. Mikolajick. Unveiling the double-well energy landscape in a ferroelectric layer. *Nature*, 565:464–467, 2019.

[143] A. K. Saha, P. Sharma, I. Dabo, S. Datta, and S. K. Gupta. Ferroelectric transistor model based on self-consistent solution of 2D Poisson's, non-equilibrium Green's function and multi-domain Landau Khalatnikov equations. In *2017 IEEE International Electron Devices Meeting*, pages 13.5.1–13.5.4, Dec. 2017.

[144] H. Agarwal, P. Kushwaha, Y. Lin, M. Kao, Y. Liao, A. Dasgupta, S. Salahuddin, and C. Hu. Proposal for Capacitance Matching in Negative Capacitance Field-Effect Transistors. *IEEE Electron Device Letters*, 40(3):463–466, Mar. 2019.

[145] S. Khandelwal, J. P. Duarte, A. I. Khan, S. Salahuddin, and C. Hu. Impact of Parasitic Capacitance and Ferroelectric Parameters on Negative Capacitance FinFET Characteristics. *IEEE Electron Device Letters*, 38(1):142–144, Jan. 2017.

[146] A. I. Khan, C. W. Yeung, C. Hu, and S. Salahuddin. Ferroelectric negative capacitance MOSFET: Capacitance tuning amp; antiferroelectric operation. In *2011 International Electron Devices Meeting*, pages 11.3.1–11.3.4, Dec. 2011.

[147] M. H. Park, Y. H. Lee, H. J. Kim, Y. J. Kim, T. Moon, K. D. Kim, J. Mller, A. Kersch, U. Schroeder, T. Mikolajick, and C. S. Hwang. Ferroelectricity and Antiferroelectricity of Doped Thin HfO_2-Based Films. *Advanced Materials*, 27(11):1811–1831, 2015.

[148] K. Chatterjee, A. J. Rosner, and S. Salahuddin. Intrinsic speed limit of negative capacitance transistors. *IEEE Electron Device Letters*, 38(9):1328–1330, Sep. 2017.

[149] D. Kwon, Y. Liao, Y. K. Lin, J. P. Duarte, K. Chatterjee, A. J. Tan, A. K. Yadav, C. Hu, Z. Krivokapic, and S. Salahuddin. Response speed of negative capacitance finfets. In *2018 IEEE Symposium on VLSI Technology*, pages 49–50, Jun. 2018.

[150] J. Hattori, K. Fukuda, T. Ikegami, H. Ota, S. Migita, H. Asai, and A. Toriumi. Fringing field effects in negative capacitance field-effect transistors with a ferroelectric gate insulator. *Japanese Journal of Applied Physics*, 57(4S):04FD07, Mar. 2018.

[151] G. Pahwa, A. Agarwal, and Y. S. Chauhan. Numerical Investigation of Short-Channel Effects in Negative Capacitance MFIS and MFMIS Transistors: Subthreshold Behavior. *IEEE Transactions on Electron Devices*, 65(11):5130–5136, Nov. 2018.

[152] J. P. Duarte, Y. K. Lin, Y. H. Liao, A. Sachid, M. Y. Kao, H. Agarwal, P. Kushwaha, K. Chatterjee, D. Kwon, H. L. Chang, S. Salahuddin, and C. Hu. Negative-Capacitance FinFETs: Numerical Simulation, Compact Modeling and Circuit Evaluation. In *2018 International Conference on Simulation of Semiconductor Processes and Devices*, pages 123–128, Sep. 2018.

[153] More Moore White Paper-International Roadmap for Devices and Systems. `https://irds.ieee.org/reports`. Accessed: 2019-03-12.

[154] Z. Krivokapic et al. 14nm Ferroelectric FinFET technology with steep subthreshold slope for ultra low power applications. In *2017 IEEE International Electron Devices Meeting*, pages 15.1.1–15.1.4, Dec. 2017.

[155] S. Agrawal and J. G. Fossum. A Physical Model for Fringe Capacitance in Double-Gate MOSFETs With Non-Abrupt Source/Drain Junctions and Gate Underlap. *IEEE Transactions on Electron Devices*, 57(5):1069–1075, May 2010.

[156] H. Agarwal, P. Kushwaha, Y. K. Lin, M. Kao, Y. Liao, J. Duarte, S. Salahuddin, and C. Hu. NCFET Design Considering Maximum Interface Electric Field. *IEEE Electron Device Letters*, 39(8):1254–1257, Aug. 2018.

[157] J. Seo, J. Lee, and M. Shin. Analysis of Drain-Induced Barrier Rising in Short-Channel Negative-Capacitance FETs and Its Applications. *IEEE Transactions on Electron Devices*, 64(4):1793–1798, Apr. 2017.

[158] H. Agarwal, P. Kushwaha, J. P. Duarte, Y. K. Lin, A. B. Sachid, M. Kao, H. Chang, S. Salahuddin, and C. Hu. Engineering Negative Differential Resistance in NCFETs for Analog Applications. *IEEE Transactions on Electron Devices*, 65(5):2033–2039, May 2018.

[159] Y. T. Tang et al. A Comprehensive Study of Polymorphic Phase Distribution of Ferroelectric-Dielectrics and Interfacial Layer Effects on Negative Capacitance FETs for Sub-5 nm Node. In *2018 IEEE Symposium on VLSI Technology*, pages 45–46, Jun. 2018.

[160] X. Wang, B. Cheng, A. R. Brown, C. Millar, J. B. Kuang, S. Nassif, and A. Asenov. Interplay Between Process-Induced and Statistical Variability in 14-nm CMOS Technology Double-Gate SOI FinFETs. *IEEE Transactions on Electron Devices*, 60(8):2485–2492, Aug. 2013.

[161] X. Wang, A. R. Brown, , and A. Asenov. Statistical variability and reliability in nanoscale FinFETs. In *2011 International Electron Devices Meeting*, pages 5.4.1–5.4.4, Dec. 2011.

[162] L. Xu, T. Nishimura, S. Shibayama, T. Yajima, S. Migita, and A. Toriumi. Kinetic pathway of the ferroelectric phase formation in doped HfO_2 films. *Journal of Applied Physics*, 122(12):124104, 2017.

[163] S. Y. Mun et al. Quantitative model of CMOS inverter chain ring oscillator's effective capacitance and its improvements in 14nm FinFET technology. In *2018 IEEE International Conference on Microelectronic Test Structures*, pages 153–156, Mar. 2018.

[164] A. B. Sachid, Y. Huang, Y. Chen, C. Chen, D. D. Lu, M. Chen, and C. Hu. FinFET With Encased Air-Gap Spacers for High-Performance and Low-Energy Circuits. *IEEE Electron Device Letters*, 38(1):16–19, Jan. 2017.

[165] A. B. Sachid, H. Lin, and C. Hu. Nanowire FET With Corner Spacer for High-Performance, Energy-Efficient Applications. *IEEE Transactions on Electron Devices*, 64(12):5181–5187, Dec. 2017.

[166] Jemin Park. *Improving CMOS Speed and Switching Power with Air-gap Structures*. PhD thesis, University of California, Berkeley, 2011.

[167] S. Pang, H. N. Tsao, X. Feng, and K. Mllen. Patterned Graphene Electrodes from Solution-Processed Graphite Oxide Films for Organic Field-Effect Transistors. *Advanced Materials*, 21(34):3488–3491, 2009.

[168] S. Aritome, R. Shirota, G. Hemink, T. Endoh, and F. Masuoka. Reliability issues of flash memory cells. *Proceedings of the IEEE*, 81(5):776–788, May 1993.

[169] S. Zhang, P. M. Levy, A. C. Marley, and S. S. P. Parkin. Quenching of Magnetoresistance by Hot Electrons in Magnetic Tunnel Junctions. *Phys. Rev. Lett.*, 79:3744–3747, Nov. 1997.

[170] S. Ikeda, K. Miura, H. Yamamoto, K. Mizunuma, H. D. Gan, M. Endo, S. Kanai, J. Hayakawa, F. Matsukura, and H. Ohno. A perpendicular-anisotropy CoFeB-MgO magnetic tunnel junction. *Nature Materials*, 9:721724, Sep. 2010.

[171] T. Kawahara, K. Ito, R. Takemura, and H. Ohno. Spin-transfer torque RAM technology: Review and prospect. *Microelectronics Reliability*, 52(4):613 – 627, 2012.

[172] Y. Zhang, W. Zhao, Y. Lakys, J. Klein, J. Kim, D. Ravelosona, and C. Chappert. Compact Modeling of Perpendicular-Anisotropy CoFeB/MgO Magnetic Tunnel Junctions. *IEEE Transactions on Electron Devices*, 59(3):819–826, Mar. 2012.

[173] W. Kim et al. Extended scalability of perpendicular STT-MRAM towards sub-20nm MTJ node. In *2011 International Electron Devices Meeting*, pages 24.1.1–24.1.4, Dec. 2011.

[174] J. Z. Sun. Spin-current interaction with a monodomain magnetic body: A model study. *Phys. Rev. B*, 62:570–578, Jul. 2000.

[175] R. H. Koch, J. A. Katine, and J. Z. Sun. Time-Resolved Reversal of Spin-Transfer Switching in a Nanomagnet. *Phys. Rev. Lett.*, 92:088302, Feb. 2004.

[176] G. D. Fuchs, I. N. Krivorotov, P. M. Braganca, N. C. Emley, A. G. F. Garcia, D. C. Ralph, and R. A. Buhrman. Adjustable spin torque in magnetic tunnel junctions with two fixed layers. *Applied Physics Letters*, 86(15):152509, 2005.

[177] J. G. Alzate, P. Khalili Amiri, G. Yu, P. Upadhyaya, J. A. Katine, J. Langer, B. Ocker, I. N. Krivorotov, and K. L. Wang. Temperature dependence of the voltage-controlled perpendicular anisotropy in nanoscale MgO—CoFeB—Ta magnetic tunnel junctions. *Applied Physics Letters*, 104(11):112410, 2014.

[178] Y. Wang, H. Cai, L.A.B. Naviner, Y. Zhang, J.O. Klein, and W.S. Zhao. Compact thermal modeling of spin transfer torque magnetic tunnel junction. *Microelectronics Reliability*, 55(9):1649–1653, 2015.

[179] K. Lee and S. H. Kang. Design Consideration of Magnetic Tunnel Junctions for Reliable High-Temperature Operation of STT-MRAM. *IEEE Transactions on Magnetics*, 46(6):1537–1540, Jun. 2010.

[180] L. Yuan, S. H. Liou, and Dexin Wang. Temperature dependence of magnetoresistance in magnetic tunnel junctions with different free layer structures. *Phys. Rev. B*, 73:134403, Apr. 2006.

[181] C. H. Shang, J. Nowak, R. Jansen, and J. S. Moodera. Temperature dependence of magnetoresistance and surface magnetization in ferromagnetic tunnel junctions. *Phys. Rev. B*, 58:R2917–R2920, Aug. 1998.

[182] V. Drewello, J. Schmalhorst, A. Thomas, and G. Reiss. Evidence for strong magnon contribution to the TMR temperature dependence in MgO based tunnel junctions. *Phys. Rev. B*, 77:014440, Jan. 2008.

[183] L. Faber, W. Zhao, J. Klein, T. Devolder, and C. Chappert. Dynamic compact model of Spin-Transfer Torque based Magnetic Tunnel Junction (MTJ). In *2009 4th International Conference on Design Technology of Integrated Systems in Nanoscal Era*, pages 130–135, Apr. 2009.

[184] G. D. Panagopoulos, C. Augustine, and K. Roy. Physics-Based SPICE-Compatible Compact Model for Simulating Hybrid MTJ/CMOS Circuits. *IEEE Transactions on Electron Devices*, 60(9):2808–2814, Sep. 2013.

[185] Kerem Yunus Camsari, Samiran Ganguly, Shehrin Sayed, and Supriyo Datta. Modular Spintronics Library. https://nanohub.org/resources/17831, May 2013.