Negative Capacitance Field-Effect Transistor Design and Machine Learning Applications in Compact Models

Ming-Yen Kao

Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2022-34
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2022/EECS-2022-34.html

May 1, 2022

Negative Capacitance Field-Effect Transistor Design and Machine Learning Applications in Compact Models

by

Ming-Yen Kao

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:
Dr. Chenming Hu, Chair
Dr. Sayeef Salahuddin
Dr. Junqiao Wu

Spring 2022

Negative Capacitance Field-Effect Transistor Design and
Machine Learning Applications in Compact Models

Abstract

Negative Capacitance Field-Effect Transistor Design and
Machine Learning Applications in Compact Models
by
Ming-Yen Kao

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley
Dr. Chenming Hu, Chair

With the scaling of the transistor, fabrication of an actual device and modeling of an ultra-short channel transistor becomes more and more challenging. Dr. Salahuddin proposed a negative capacitance field-effect transistor (NCFET) in 2008. By utilizing the ferroelectric negative capacitance region with proper capacitance matching a ferroelectric layer with a dielectric layer, the overall effective oxide thickness could be further thinned down without affecting much of the carrier mobility.

A technique of optimization of an NCFET will be proposed in chapter 2. By utilizing process techniques like mask oxidation, a non-uniform interfacial layer can be formed to create a more uniform metal-oxide-semiconductor capacitance along the channel (C_{mos}). The overall capacitance matching of an NCFET can be improved because of a uniform C_{mos} profile. Chapter 3 will introduce a simulation scheme of NCFETs variation due to dielectric grains within a ferroelectric film. This scheme can be applied to the future estimation of NCFETs variation given the grains' size and the ferroelectric parameters.

The effect and compact modeling of the polarization gradient effect will be demonstrated in Chapter 4. With the feature of polarization gradient effect in an NCFET compact model, the characteristics of an NCFET can be better captured, such as negative drain resistance and negative drain-induce barrier

effect (DIBL). Energy analysis of an NCFET will be presented in Chapter 5. The consistency between TCAD energy calculation by the integral over the grids with the Landau equation and power consumption calculation from the circuit is shown in detail.

Negative capacitance benefits on FinFET and gate-all-around (GAA) FET will be presented in chapters 6 & 7, respectively. Baseline devices of a FinFET and a GAAFET are made in technology computer-aided design (TCAD) and are calibrated to International Roadmap for Devices and Systems (IRDS) tables. NC parameters are also extracted from an experiment on metal-oxide-semiconductor capacitance (MOSCAP). How many nodes NC extends the baseline will be discussed. A compact model of an anti-ferroelectric on an NCFET will be presented in Chapter 8.

Potential applications of machine learning will be illustrated in Chapters 9 & 10. Machine learning-assisted parameter extraction will be presented in Chapter 9. In the long run, using machine learning-assisted models as an alternative to the conventional equation-based compact models will be shown in Chapter 10. In the end, Chapter 11 will conclude chapters and propose some future work.

To my family and friends: Past, Present, and Future

Content

List of Figures

List of Tables

\

Acknowledgement

Firstly, I would like to thank the best advisor, Prof. Chenming Hu. He provided full support throughout my Ph.D. journey. The support is not only the financial support but also all the other aspect of support. He cares so much of students' needs. For example, I remember that he checked every question in detail during the student review to make sure I was satisfied with all the aspects of my life at Cal. Prof. Chenming Hu is also a perfect role model for juniors. He not only works hard but also enjoys his life. During these five years of my Ph.D., I tried to follow his step to increase my working efficiency to have leisure time to enjoy my life here in the US. He is an excellent textbook example of an open-minded person and a life-long learner. For instance, I started machine learning-related projects in my 5th year. He advised me along the way, and we both learned so much together from machine learning theories and applications. He is also an outstanding leader. For instance, he manages the BSIM group very well and maintains a good relationship with industry partners. Last but not least, I learned so much from Prof. Hu's presentation skills. Prof. Hu keeps instructing me on presenting an idea more interestingly. I practice every time at the annual presentation to our industry sponsors. I can even feel the improvements in my presentation, even myself. Under Prof. Hu's supervision, I went through a very fulfilling Ph.D. journey. Suppose anyone asks me if time is back to 5 years again, whether I would choose Ph.D. again. My answer is definitely yes, and I will try my best to join Prof. Hu's group.

I want to show my gratitude toward Prof. Sayeef Salahuddin. Without him, I would not have the NCFET-related research topics throughout my Ph.D. He is very generous in providing his insight on any physics/math-related problems. I feel super confident once he backs up the explanation. He also serves as my preliminary exam, qualification exam, and dissertation committee member. He appeared everywhere and always provided help when I was stuck throughout my Ph.D.

I appreciate Prof. Sayeef Salahuddin, Prof. Ana Claudia Arias, and Prof. Jeffrey Bokor serving on the oral preliminary committee. They challenged us with questions from the very fundamental to the trickiest ones. This process paves a rock-solid foundation for the following years of research. I would like

to also express my gratitude to Prof. Chenming Hu, Prof. Sayeef Salahuddin, Prof. Ali Javey, and Prof. Junqiao Wu for serving as qualification committee members. They examined my research results at the mid-point of my Ph.D. and provided much insight for my following years of research.

Yu-Hung has always been one of my classmates and best co-workers. We studied for four years together at National Taiwan University and did research together during five years of Ph.D. right here at UC Berkeley. We will also join the same group at NVIDIA in the future. Best wishes for our future journey. I would also like to thank the preliminary exam study group: Yu-Hung, Xiaoer, Saavan, Adi, and Jason. Without their help, I could not pass the preliminary exam. I would also like to show my massive gratitude toward the co-authors of my publication, Angada, Huan-Lin, Juan, Avirup, Yen-Kai, Ava, Girish, Hei, Fredo, Harshit, Pragya, Yu-Hung, Suraj, Prof. Hu, Prof. Salahuddin, and Prof. Khandelwal. They are all excellent mentors, and I learn a lot from them, from doing good research to being a good co-worker.

The Taiwanese Community at Cal is closely-connected, and staying at Berkeley is never boring. I appreciate the organization of the Berkeley Association of Taiwanese Students (BATS). While I served as the vice-president and the president of BATS, I met so many new friendly people. Some of them went back to Taiwan, and some stayed in the US for their careers. I will keep all these valuable memories forever.

My family members are always supportive. I cannot express my great appreciation toward my father, mother, and sister here. They provide everything I need to focus on what I would like to pursue. I feel that we are bounded so closely that family is still my priority when I think of anything. I would also like to thank my sister's husband; he takes good care of my sister. I would also like to express my gratitude to my father-in-law and mother-in-law. They treat me like their own son.

Last but not least, my wife, Tien-Ning, is the best person I have ever met in the world. She is kind and enthusiastic about exploring new things. I am always motivated by her to try new things. We plan a long to-do list that we

would like to explore in the future. I am so excited about the future. Tien-Ning is also super supportive of whatever I would like to do. This year is our tenth anniversary. I am so passionate about our next decade's anniversary.

Chapter 1

Introduction

1.1 Challenges of Nano-electronics

Gordon Moore proposed Moore's law in 1975, and after that, the number of components per integrated circuit doubled every 12-24 months [1]. The fast

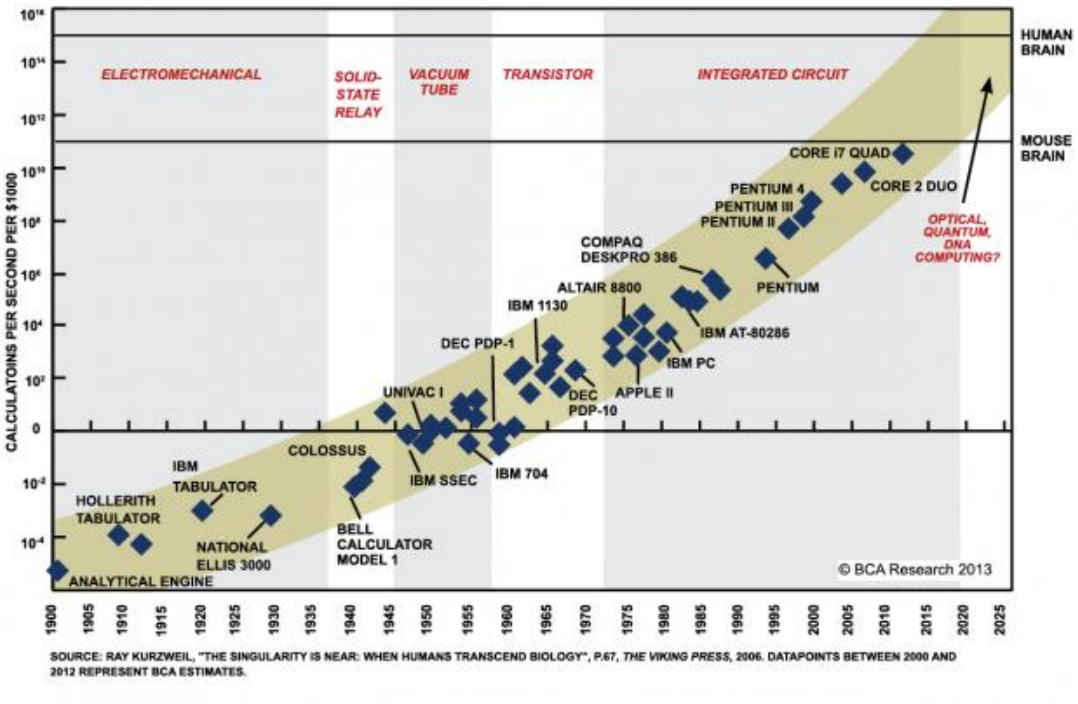


Fig. 1.1. Calculations per second per 1,000 USD versus time. Exponential growth can be observed throughout 100 years. [132]

downscaling of the transistor benefits the whole world in many ways, including communication electronics (like iPhone and Internet routers), personal computers… For example, calculations per second per 1,000 USD of various products throughout 100 years are shown in Fig. 1.1. However, the growth of exponential is difficult to maintain at the same growth rate forever. For example, the exponential growth of the bacteria would be limited by the resources finally. There is no exception for the semiconductor industry. For a long time, the metal-oxide-semiconductor field-effect transistor (MOSFET) benefits from scaling and an increase in clock frequency. By scaling effective oxide thickness (EOT), the same charge density can be achieved at a lower gate voltage. By reducing the gate length, the same current density could be achieved at a lower drain voltage. Several great inventions, including the H-K metal gate, immersion lithography, and FinFET, keep the scaling continuing. Scaling of EOT and increase of clock frequency together increase the performance per area. Recently, scaling of EOT is slowed down because of physical limitations, including direct tunneling, oxide breakdown, and mobility degradation, and the increase of clock frequency is hindered by the difficulty of heat dissipation.

Negative Capacitance Field-Effect Transistor (NCFET) is a promising near-future solution to the slowed-down of EOT scaling. The scaling can be pushed to a few more nodes if NCFET is properly designed. The theory of NCFET will be explained in the next section.

1.2 Theory of Negative Capacitance Field-Effect Transistor

Before explaining NCFET, let's start with where negative capacitance (NC) comes from. Fig. 1.2 (a) shows the energy of energy versus charge in ferroelectric (FE) and dielectric (DE) without an external electric field [2]. The minimum energy of FE is not at zero charge, leading to ferroelectricity (remanent polarization when there is no electric field). In contrast, the minimum energy of DE is at zero charge, and the shape is parabolic. If the FE is stacked on top of the DE, the total energy is plotted as the red curve in Fig. 1 (a). The equilibrium point is the small black ball at zero electric fields, where the individual energy is not minimum for the FE, but the total energy of the system is minimum. Voltage is the derivative of energy with respect to the charge, and the voltage versus charge plot is shown in Fig. 1.2 (b). Capacitance is the derivative of voltage with respect to charge, which is the slope in Fig. (b). The slope of the black curve in Fig. 1 (b) is negative, showing the negative capacitance within the red box.

Fig. 1.3 shows the cartoon graph of a MOSFET. The capacitance from gate to channel can be simplified into two components, insulator capacitance (C_{ins}) and semiconductor capacitance (C_s). The total capacitance can be expressed as Eq. (1.1)

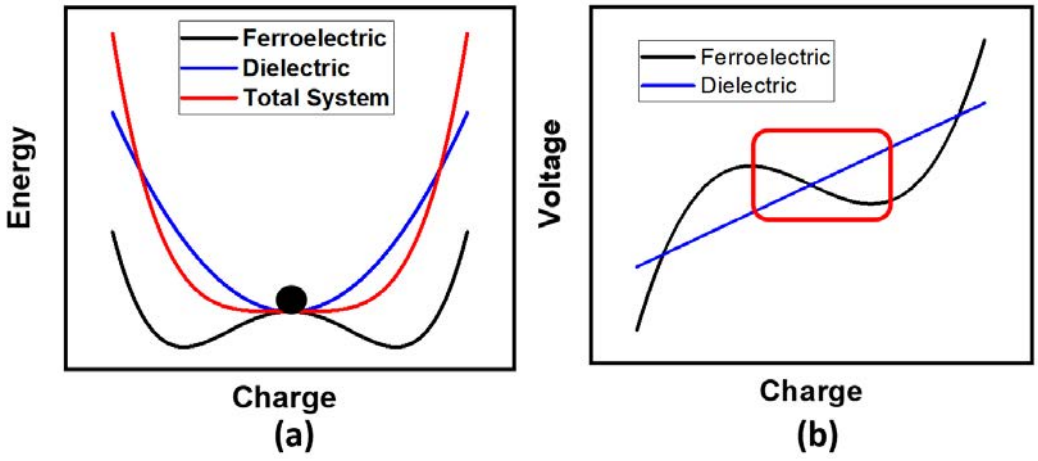


Fig. 1.2. (a) Energy versus charge plot of FE, DE, and the total system of FE and DE. (b) Voltage versus charge plot of FE and DE.

$$C_{total} = \frac{C_s C_{ins}}{C_s + C_{ins}} \quad (1.1)$$

If the value of C_{ins} is negative, C_{total} could be larger than C_s. It is even better than EOT -> 0 scenario. Another way to demonstrate the benefits of NCFET is by looking into the equation of subthreshold slope (SS) [3]:

$$SS = \frac{\partial V_g}{\partial (log_{10} I)} = \frac{\partial V_g}{\partial \varphi s} \frac{\partial \varphi s}{\partial (log_{10} I)} \quad (1.2)$$

The SS of the device is limited by two parts. The first part is $\frac{\partial V_g}{\partial \varphi s}$, which is determined by the capacitance divider as shown in Fig. 2.

$$\frac{\partial V_g}{\partial \varphi s} = 1 + \frac{C_s}{C_{ins}} \quad (1.3)$$

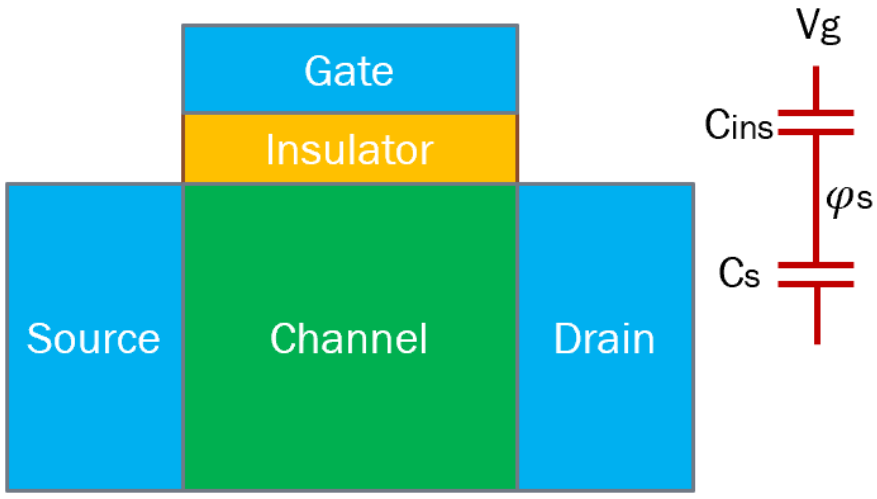


Fig. 1.3. Cartoon graph of a MOSFET.

From Eq. (1.3), $\frac{\partial V_g}{\partial \varphi s}$ is possible to be smaller than one if C_{ins} is negative. The second part is $\frac{\partial \varphi s}{\partial (log_{10} I)}$, which is limited by the current control mechanism of the transistor and the temperature. For the conventional MOSFET, the minimum value of $\frac{\partial \varphi s}{\partial (log_{10} I)}$ is about 60 mV/dec at room temperature, and that

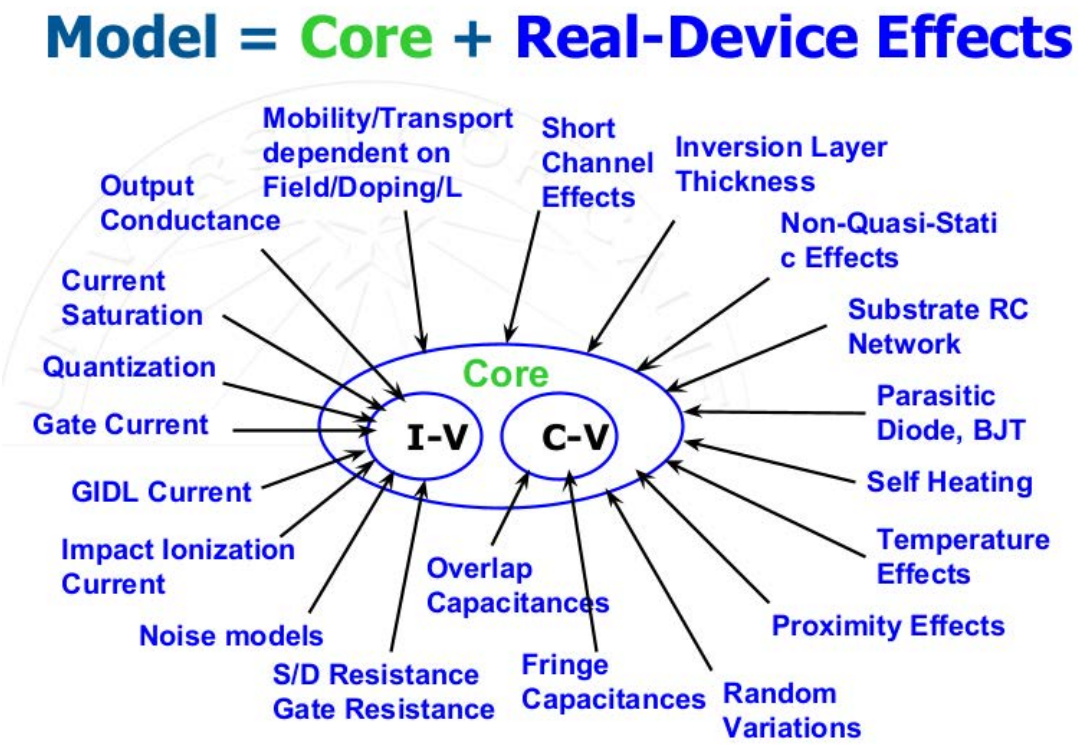


Fig. 1.4. Architecture of BSIM from Prof. Chenming Hu's presentation in Sep. 2013.

is why 60mV/dec is called Boltzmann Tyranny. NC can help to overthrow the Boltzmann Tyranny by providing gain to $\frac{\partial V_g}{\partial \varphi_s}$.

1.3 Berkeley short-channel IGFET Model (BSIM)

People in fabless design houses may not know the knowledge of semiconductor physics; people in foundries may not understand how to design a circuit. That's how a compact model comes in as a bridge to facilitate communication between fabless design houses and foundries. A good compact model has to be both computationally efficient and must predict the characteristics of transistors accurately. Berkeley short-channel insulator-gate field-effect transistors model (BSIM) is the first industry standard compact model for circuit simulation. The architecture of BSIM is demonstrated in Fig. 1.4. With the scaling of the minimum feature size of transistors. More and more non-idea effect (as shown in the peripheral in Fig. 1.4) models have to been added so the maintenance and improvement of BSIM models never stops. To achieve accurate spice/circuit simulation results, both an accurate compact

model and a good model card are needed. A model card is composed of all the necessary model parameters in order to capture the characteristics of a transistor with single or multiple geometries. The procedure of generating a model card is called parameter extraction. The package of compact models and model card are called process design kit (PDK). A PDK will be sent to design houses from foundries to represent their process technology and specifications.

Chapter 2

Optimization of NCFET by Matching Dielectric and Ferroelectric Nonuniformly Along the Channel

A new design to overcome the nonuniformity of capacitance matching along the channel of a negative capacitance field-effect transistor (NCFET) is presented in this section. By introducing non-uniform oxidation, the thickness of SiO_2 at the edge regions of the channel can be increased while maintaining the thickness of SiO_2 at the center region of the channel. As a result, the capacitance along the channel becomes more uniform, and better capacitance matching between the dielectric (DE) and ferroelectric (FE) can be achieved. The Sentaurus TCAD results show matching improvement in the central region and a significant boost of on-current (20% improvement).

2.1 Motivation

The fundamental limit imposed by the Boltzmann distribution (60 mV/decade), which hinders the scaling of CMOS technology, is referred to as the Boltzmann Tyranny [3]-[5]. Tunneling field-effect transistors (TFETs), nano-electromechanical (NEM) switches [6], and negative capacitance field-effect transistors (NCFETs) are promising ways to overcome the Boltzmann Tyranny [7]. However, the NEM switch is subject to reliability and scalability issues [5], while the TFET suffers from low on-current and other non-ideal effects [8]-[9]. NCFETs, on the other hand, can achieve an improved

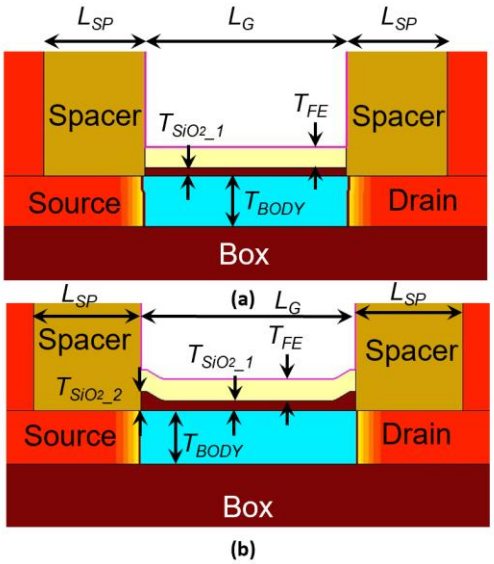
subthreshold slope (SS) while maintaining high on-current (compared to TFETs) and are fabricated through CMOS-compatible processes [10]. Many NCFETs have been made experimentally [10]-[15]. Nevertheless, many demonstrate a subthreshold swing (SS) of only 50-60 mV/decade without hysteresis and an internal metal gate. An inner metal gate breaks the ferroelectric into multiple domains, causing hysteresis [16] and introducing other non-ideal effects like charge trapping, leading to problems during device operation. Despite this, NCFETs without an internal metal gate usually demonstrate worse performance than NCFETs with an inner metal gate [17] because of bad capacitance matching in the center region of the channel. A possible solution to this problem will be proposed in this study. Many papers have discussed the nonuniformity in electric field, nonuniformity in FE [18]-[21], and methods of improving the degree of capacitance matching [22]-[23].

This study will demonstrate the difficulty of further reducing the SS below 60 mV/decade of hysteresis-free NCFETs without an internal metal gate in B (i) and B (ii). The fringing field from the source and drain plays a vital role in the capacitance matching in the subthreshold region, but the fringing field also limits the capacitance matching at the center of the channel. A possible solution will be proposed in B (iii) to improve the performance of NCFETs further.

2.2 Device Characterization and Discussion

Baseline UTB-SOI Device Structure

The baseline structure and design parameters are shown in Fig. 2.1 but without the FE layer. The electric field at $V_g = 0V$ and $V_d = 0.7V$ is plotted in Fig. 2.2 (a). The electric field is stronger at the edges of the gate oxide, as indicated by the red circles. The stronger electric field at the edges is caused



Parameter	Value
Gate Length (L_G)	20 nm
Spacer Length (L_{SP})	10 nm
Body Thickness (T_{BODY})	5 nm
SiO_2 thickness at Center (T_{SiO2_1})	0.9 nm
SiO_2 thickness at Edge (T_{SiO2_2})	1.8 nm
Supply Voltage (V_{DD})	0.7 V
FE Thickness (T_{FE})	2 nm
Baseline ON Current (I_{ON})	429 μA/μm
Baseline OFF Current (I_{OFF})	188 pA/μm
Baseline Subthreshold Slope (SS)	80.3 mV/dec
Basline Threshold Voltage (V_{TH})	327.5 mV

Fig. 2.1. The simulated device structure with (a) uniform thickness of interfacial layer and with (b) nonuniform thickness of interfacial layer. Table on the right-hand side lists important device parameters. The shaded parameters are for the baseline device. The baseline device has the same structure as (a), but without an FE layer.

by both the inner-fringing field (passing through the Si channel) and outer-fringing field (passing through the Si_3N_4 spacer) [24], leading to nonuniform capacitance along the channel as shown in Fig. 2.2 (b) and imposing the limitation on capacitance matching. Note that the capacitance is shown in Fig. 2.2 (b) is from the structures with an FE layer, and the vertical electric field in the SiO_2 (perpendicular to the FE) right below the FE layer is used to extract the nonuniform C_{mos} by using Eq. (2.1)

from the TCAD results.

$$C_{mos} = \frac{dQ}{dV} = \frac{d(E_{\perp FE} \times \epsilon_{SiO2})}{dV(x)} \quad (2.1)$$

where $E_{\perp FE}$ is the electric field perpendicular to and right below the FE, ϵ_{SiO2} is the permittivity of SiO_2, and $dV(x)$ is the step size of electric potential at the interface of SiO_2 and FE as a function of position along the

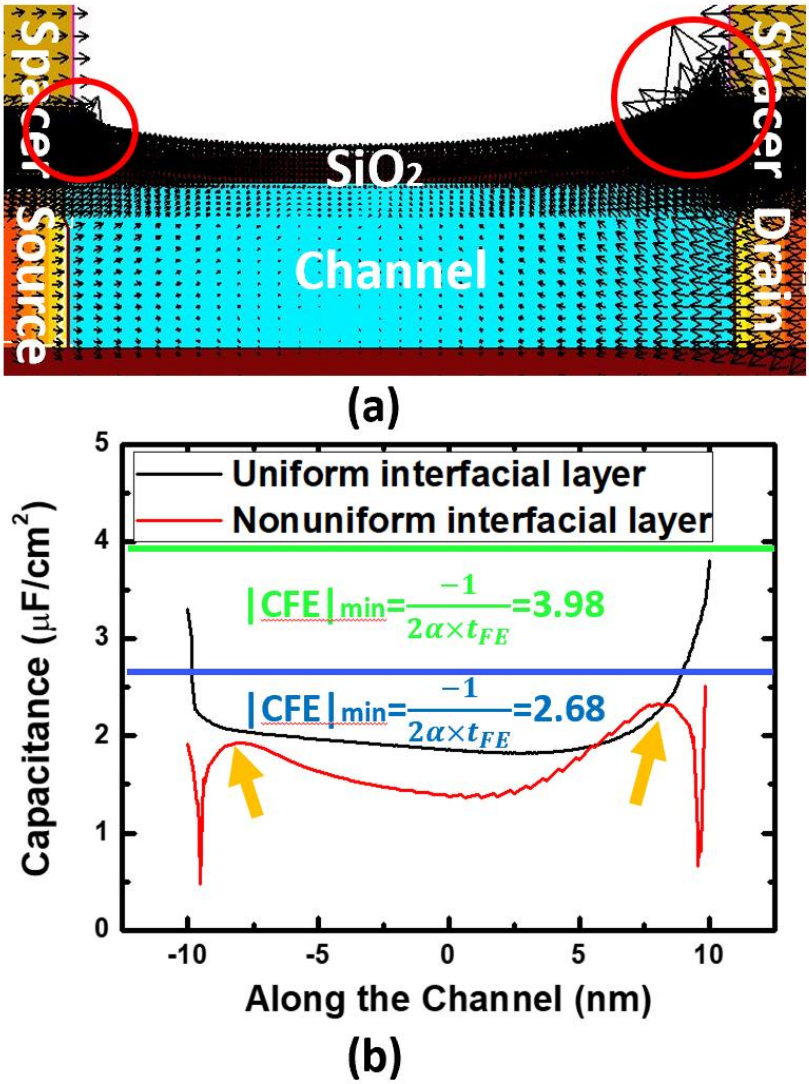


Fig. 2.2. (a) The electric field at V_g = 0V and V_d = 0.7V. The red circles highlight the higher electric field at the edges of the channel. (b) Capacitance versus position along the channel at V_g = 0.7V. The green line and the black curve are the traditional capacitance matching design, and the blue line and the red curve are our proposed capacitance matching design.

channel. To avoid hysteresis, the data in Fig. 2.2 (b) is extracted at V_g = 0.7V and V_d = 0.7V, where the minimum absolute value of FE capacitance is not smaller than any value of the curve.

Uniform interfacial layer NCFET

In traditional NCFET design, a uniform thickness of the interfacial layer

along the channel is assumed, and the corresponding ferroelectric capacitance (C_{FE}) is matched to its limit with C_{mos}. In HfO_2-based NCFETs, the remanent polarization (P_r) is more sensitive to ferroelectric doping concentration than the coercive field (E_c) [25]. Therefore, E_c is fixed to 2MV/cm, and P_r is decreased until C_{FE} touches the largest C_{mos} along the channel. The smallest P_r reached in the TCAD simulation is 11.5 μC/cm^2, which is equivalent to -3.98 μF/cm^2 (dielectric constant = 16 is included from [26]) by using Eq. (2.2). The physics models used in TCAD include the Ginzburg-Landau model for ferroelectric materials, Fermi Statistics, velocity saturation, Philips unified mobility model, Shockley-Read-Hall process, and quantum potential [27]. The Ginzburg-Landau model is shown in Eq. (2.3).

$$C_{FE} = \frac{1}{2\alpha t_{FE}} + \frac{16\varepsilon_0}{t_{FE}} \quad (2.2)$$

$$E = 2\alpha P + 4\beta P^3 + 6\gamma P^5 - 2g\Delta P + \rho\frac{dP}{dt} \quad (2.3)$$

In Eqs. (2.2) and (2.3), |CFE| is the estimation of the minimum absolute value of FE capacitance; t_{FE} is the thickness of the FE; ε_0 is vacuum permittivity; α, β, γ are the parameters for the FE; g is the strength of the polarization gradient (domain coupling) which is set to be 8E-5 cm^3/F in this study (on the same order as [27]); and ρ is the viscosity that represents the finite time required for the polarization to switch. α, β, and γ are related to P_r and E_c by $\alpha = -\frac{3\sqrt{3}}{4} \times \frac{E_c}{P_r}$ and $\beta = \frac{3\sqrt{3}}{8} \times \frac{E_c}{{P_r}^3}$, and $\gamma = 0$ [29]. K. Chatterjee et al. reports that the intrinsic delay of a doped hafnium oxide-based ferroelectric is negligible in digital circuits [30], so ρ is set to be zero in this study.

Proposed nonuniform interfacial layer NCFET

An NCFET with a nonuniform interfacial layer is proposed, as shown in Fig. 2.1 (b). By tuning the thermal gradient or introducing mask oxidation techniques during processing, a thicker SiO_2 can be grown at the edge regions of the channel without affecting the thickness of the SiO_2 at the center region

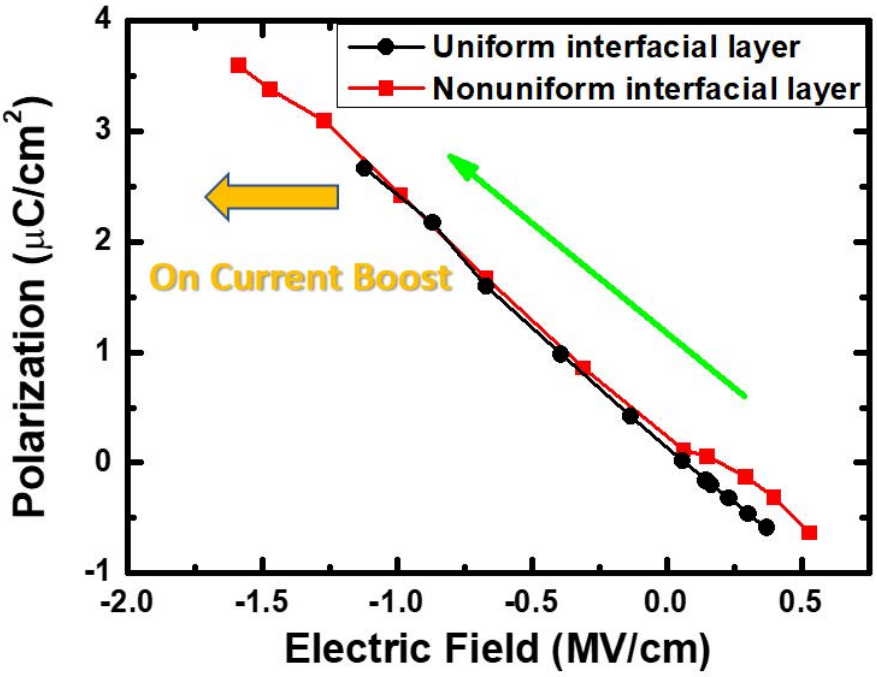


Fig. 2.3. Polarization and electric field of the FE right above the top of barrier (TOB). The transition from bottom right (deamplification) to upper left (amplification) represents the change of the state of the FE from Vg = 0V to Vg = 0.7V.

of the channel. The P_r of the FE is now reduced to 10.1 μC/cm^2, which is equivalent to -2.68 μF/cm^2 (using dielectric constant = 16 from [24]) by using Eq. (2.2). The improvement from the perspective of polarization and electric field in the FE is shown in Fig. 2.3. Note that the metal gate work functions of the two cases are shifted to match the off-current of the baseline, so at V_g = 0 V (bottom right), there is not much difference in the bias points. At V_g = 0.7 V, the bias point of the proposed NCFET demonstrates an improvement via a

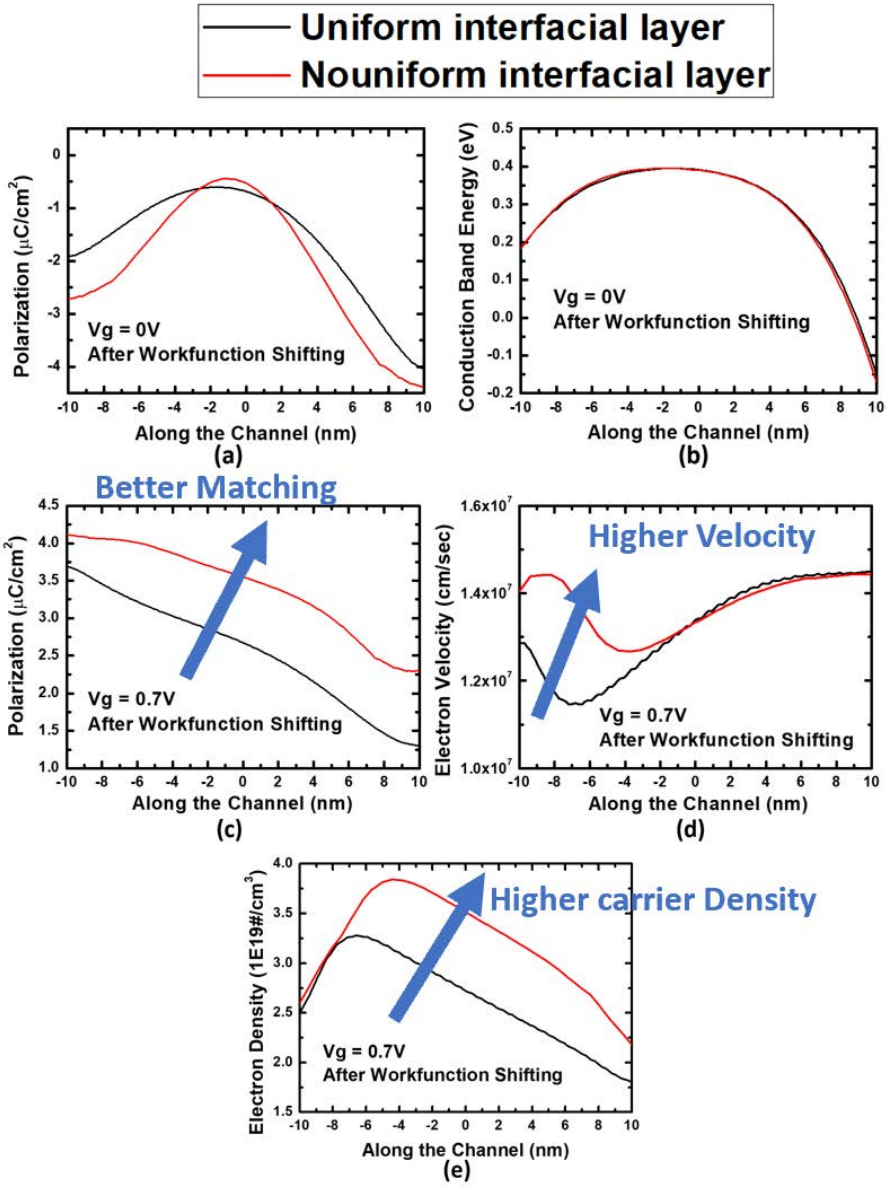


Fig. 2.4. (a) Polarization and (b) conduction band energy versus position along the channel in the off-state. (c) Polarization, (d) electron velocity, and (e) electron density versus position along the channel in the on-state. The black curves represent an NCFET with uniform thickness of the interfacial layer. The red curves represent an NCFET with a thicker interfacial layer at the edges of the channel.

left-shift as indicated by the yellow arrow (more voltage amplification).

Fig. 2.4 (a) plots polarization versus position along the channel at V_g = 0V (off-state). The profile of FE polarization changes because the capacitance matching condition along with the channel changes. Fig. 2.4 (b) shows conduction band energy versus position along the channel. The barrier heights

are the same because the metal gate work function is shifted to match the off-current of the baseline. Note that the position of the TOB is at the middle of the channel, which is why the matching between the FE and DE in the center of the channel is also critical. Fig. 2.4 (c) shows the FE polarization, higher in the NCFET with a nonuniform interfacial layer (more voltage amplification) than the normal NCFET. Fig. 2.4 (d) and 2.4 (e) show the carrier density and carrier velocity at a depth where the carrier concentration is highest (which is not close to the surface due to the quantum confinement effect). A higher current can be supported by higher electron velocity at the source side, as shown in Fig. 2.4 (d). On the other hand, carriers at the drain side have already reached velocity saturation, so the carrier concentration increases more significantly near the drain side, as shown in Fig. 2.4 (e). It is evident from Fig. 2.5 that the proposed NCFET has better ON current (20% improvement) and better SS (minimal SS is 33 mV/decade). In Fig. 2.5, two bias points show a surge in current. The first is around 0.265V, and the second is around 0.615V. Better capacitance matching happens at these two gate voltages. Two gold arrows indicate areas with better capacitance matching in Fig. 2.2 (b). Better capacitance matching over the two areas accounts for the surge in the current of the blue curve in Fig. 2.5 at 0.265V (which corresponds to the left hump of the red curve in Fig. 2.2(b)) and at 0.615V (which corresponds to the right hump of the red curve in Fig. 2.2(b)). Note that these three cases are confirmed to be hysteresis-free by running the transient forward and reverse sweeping test. In comparison, the traditional design has no current surge because the “good matching” parts are not in the channel and therefore contribute little current to the channel.

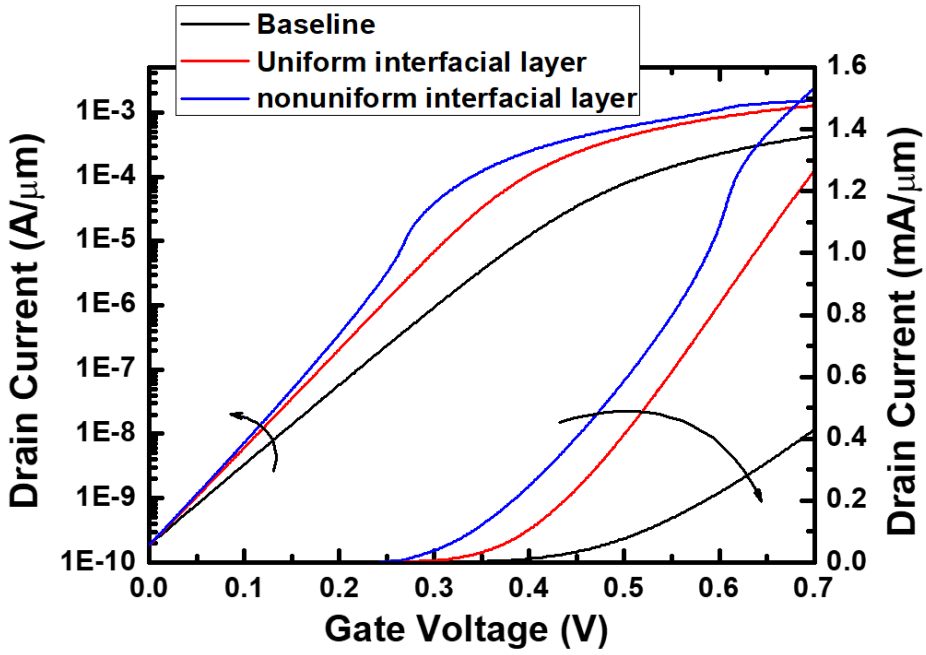


Fig. 2.5. Drain current versus gate voltage of baseline UTBSOI, traditional uniform-interfacial-layer NCFET, and proposed nonuniform-interfacial-layer NCFET.

2.3 Chapter Summary

The difficulty of capacitance matching due to the nonuniformity of capacitance along the channel is pointed out first in Fig. 2.2. To overcome this difficulty, a new design scheme utilizing a thicker interfacial SiO2 at the edges of the channel is proposed. The results show that the performance of the NCFET can be significantly boosted with this scheme. Therefore, the nonuniform capacitance caused by fringing fields should be considered when NCFETs are designed.

Chapter 3

Variation Caused by Spatial Distribution of Dielectric and Ferroelectric Grains in a Negative Capacitance Field-Effect Transistor

A new scheme to consider the dielectric (DE) phases inside polycrystalline ferroelectric (FE) materials will be proposed in this section. The scheme extracts material parameters from experimental Polarization-Electric Field (P-E) measurements from the literature. A Sentaurus TCAD structure is constructed with the extracted parameters, and the simulated P-E curve is in good agreement with experimental data. Furthermore, variation of the device performance in a negative capacitance field-effect transistor (NCFET) due to the spatial distribution of DE and FE phases is studied using Sentaurus TCAD. It is found that the resultant variations of ON and OFF currents can be up to 14.44% and 30.23%, respectively, thus showing the impact of inhomogeneous crystalline phases of the FE material on device performance.

3.1 Motivation

Power consumption becomes the most critical issue as CMOS technology is aggressively scaled. To mitigate this issue, high mobility channel materials and three-dimensional transistors have been explored [31],

[32]. However, the Boltzmann distribution poses a fundamental limit for lowering the energy dissipation in conventional electronics, and this limit is often referred to as the Boltzmann Tyranny [3]-[5], [9]. Negative capacitance FETs (NCFETs) are promising devices to overcome the subthreshold swing limit (60 mV/decade) imposed by the Boltzmann Tyranny and to achieve high Ion [3], [11]-[13], [33]-[39]. Although NCFETs experimentally exhibit sub-60 mV/decade performance [10]-[15], the non-uniformity effects of phases inside the ferroelectric (FE) haven't been investigated. To properly design NCFETs, the non-uniformity effects of phases should be carefully addressed. For example, X-ray diffraction (XRD) experimental data of HfO_2-based ferroelectric thin films [10] shows that, other than the ferroelectric orthorhombic phase, there also exist cubic and monoclinic phases, which are dielectric (DE). Lun Xu et al. also reported the existence of monoclinic phases in ferroelectric HfO_2 films [40]. In previous work [41]-[42], the ferroelectric layer is assumed to be homogeneous. The Landau equation can solely predict its properties in the simulation, which fails to consider the essential physics in real devices.

Symbol	Quantity	Unit
DE %	33.3	%
FE %	66.7	%
α	-5.810E+10	cm/F
β	3.286E+19	$cm^5/(F \cdot C^2)$
γ	2.165E+28	$cm^9/(F \cdot C^4)$
P_0	0.307	$\mu C/cm^2$
E_0	0.185	MV/cm
ε_r	16.38	unit less

Table 3.1. Hysteresis loop fitting results

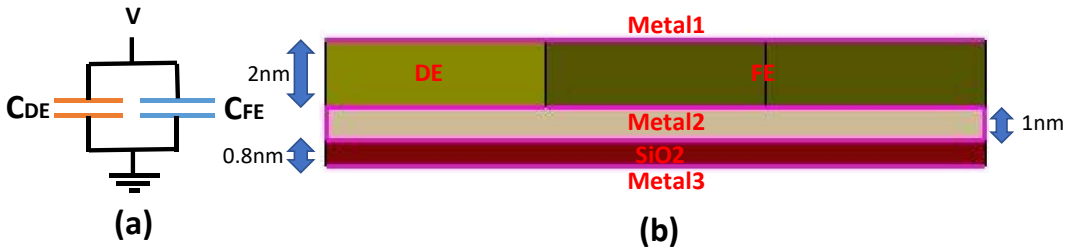


Fig. 3.1. (a) The circuit model used in Eq. (7) to (13) (b) MFMIM structure for TCAD 2D-simulation.

In this section, we investigate how the locations of the dielectric grains affect the behavior of NCFETs using Sentaurus TCAD after extracting the material parameters from experimental data. The percentage of the dielectric phase in the ferroelectric is assumed to be the same under the same process conditions. Still, the position of the dielectric grains varies, which further affects the behavior of NCFETs. This random dielectric distribution imposes an additional variation on top of the variation from fabrication. This variation should be adequately understood to evaluate and minimize the impact on NCFET performance.

3.2 Device Characterization

Ferroelectric parameters extraction

To set up the NCFET simulation using Sentaurus TCAD, the measured hysteresis loop of a 10nm ferroelectric from [14] is used to extract the ferroelectric parameters. As mentioned earlier, there are dielectric grains in the ferroelectric thin film, so only using the Landau equation to fit the hysteresis loop is insufficient. The expression for the Polarization-Electric Field (P-E) relation should include a dielectric component in addition to the Landau equation in the simulation. Therefore, the model should consist of a negative and positive capacitor in parallel (Fig. 3.1 (a)). Note that the dielectric response is incorporated into the model to extend the sixth-order-polynomial approximation of Landau theory which has limited fitting capability in positive capacitance regions (the first and the third quadrants). This dielectric constant value is assumed to be the same as DE grains. Therefore, the polarization for a ferroelectric-dielectric-mixed thin film can be expressed as the following equations:

$$E_{FE} = 2\alpha \times P_{LD} + 4\beta \times {P_{LD}}^3 + 6\gamma \times {P_{LD}}^5 \qquad (3.1)$$

$$\mathrm{P_{FE}} = P_{LD} + \mathrm{E_{FE}} \times \varepsilon_r \times \varepsilon_o \qquad (3.2)$$

$$\mathrm{E_{mix}} = E_{FE} \qquad (3.3)$$

$$\mathrm{P_{mix}} = Area_{DE} \times \mathrm{E_{mix}} \times \varepsilon_r \times \varepsilon_o + Area_{FE} \times P_{FE} \qquad (3.4)$$

where E_{FE} is the electric field across negative capacitance; α, β, and γ are Landau coefficients; P_{LD} is the polarization of ferroelectric given by Landau Equation; $\mathrm{P_{FE}}$ is the polarization of FE part with built-in dielectric constant;

ε_r is the dielectric constant; ε_o is the permittivity of vacuum; E_{mix} is the electric field across the thin film; P_{mix} is the total polarization with units of C/cm^2; $Area_{DE}$ and $Area_{FE}$ are the area percentages of DE and FE in total area.

Two assumptions have to be made in order to extract the material parameters properly. It is assumed that the ferroelectric layer consists of 66.7% ferroelectric grains and 33.3% dielectric grains: a reasonable assumption because it has been reported [40] that the percentage of the monoclinic phase inside HfO2-based ferroelectric layers can range from 10% to 50%, depending on the processing conditions; furthermore, there are cubic, tetragonal, and orthorhombic-dielectric phases possibly coexisting in the ferroelectric thin film. Another assumption is that the dielectric constants of

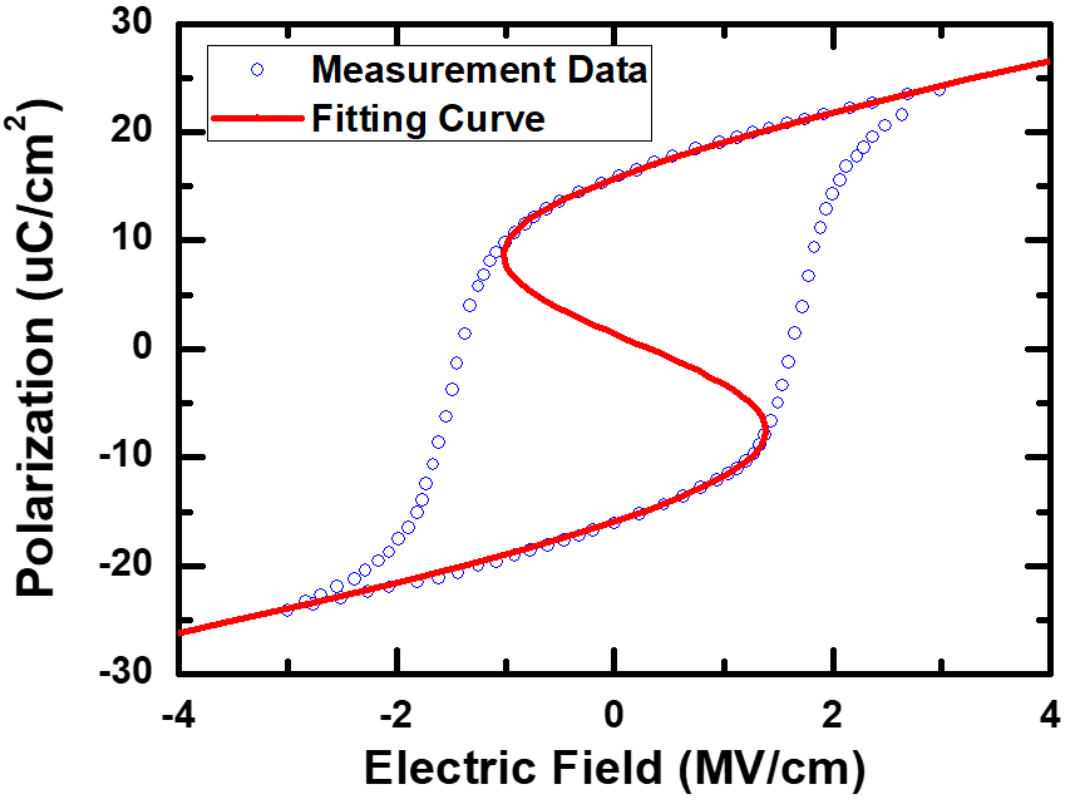


Fig. 3.2. Polarization-Electric Field loop of the proposed model (red line) and measured data (blue dots).

all the grains are the same. Based on these assumptions, we can rewrite the polarization equation from (3.1), (3.2), (3.3), and (3.4):

$$E_{mix} = 2\alpha \times (P_{LD} - P_0) + 4\beta \times (P_{LD} - P_0)^3 + 6\gamma \times$$

$$(P_{LD} - P_0)^5 + E_0 \quad (3.5)$$

$$\mathrm{P_{FE}} = P_{LD} + \mathrm{E_{mix}} \times \varepsilon_r \times \varepsilon_o \quad (3.6)$$

$$P_{mix} = 33.3\% \times E_{mix} \times \varepsilon_r \times \varepsilon_o + 66.7\% \times P_{FE} \quad (3.7)$$

where P_0 and E_0 are the offset polarization and electric field due to the leakage in the thin film [43]. The extracted parameters are listed in Table 3.1, and the fitting results are shown in Fig. 3.2.

Sentaurus TCAD MFMIM structure verification

To study the impact of the dielectric positions on NCFETs, the ferroelectric model (Landau equation) adopted in Sentaurus TCAD should be carefully calibrated. The physical models used in TCAD include the Ginzburg-Landau model for ferroelectric materials, mobility degradation due to carrier-carrier scattering, coulombic scattering, interface scattering, velocity saturation, and the Shockley-Read-Hall process [27]. The structure

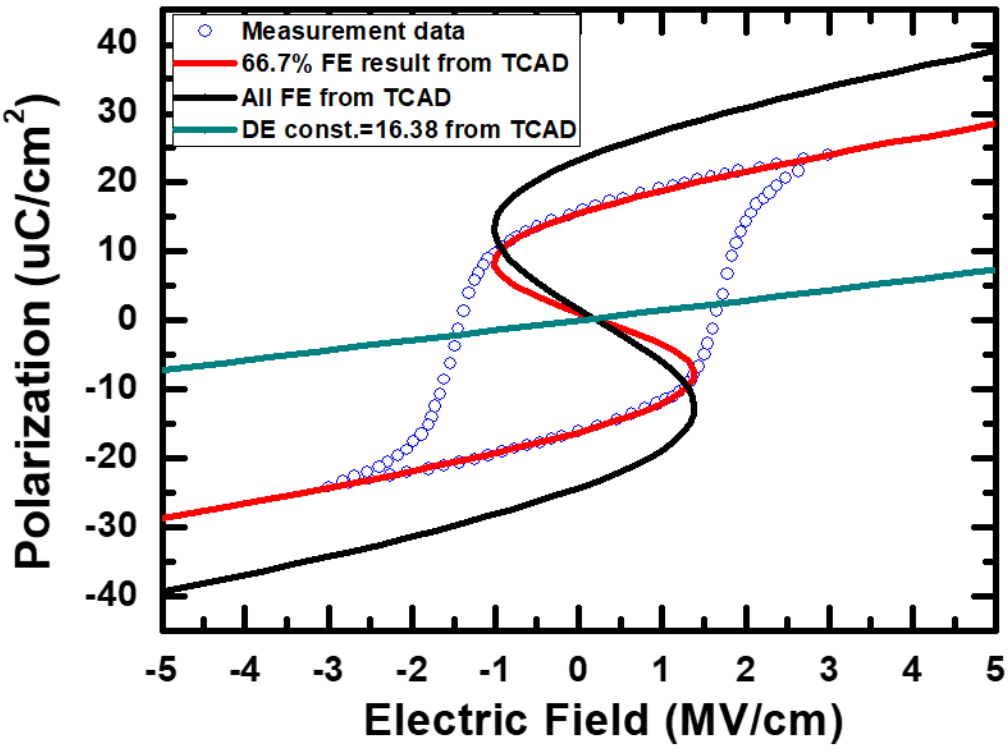


Fig. 3.3. Polarization versus electrical field plot.

for TCAD simulation is shown in Fig. 3.1 (b). In Fig. 3.1 (b), a FE-DE mixed layer consists of 2/3 FE and 1/3 DE sandwiched by metal 1 and metal 2. The SiO_2 layer stabilizes the ferroelectric in the negative capacitance region. The

α, β, and γ values of ferroelectric here are the same as the values in TABLE 3.1. Note that it is assumed no leakage in the FE layer in TCAD, so P_o and E_o are not used here. In Fig. 3.3, the curve generated by TCAD is shifted by P_o and E_o in the y-direction and x-direction, respectively, to align with experimental data.

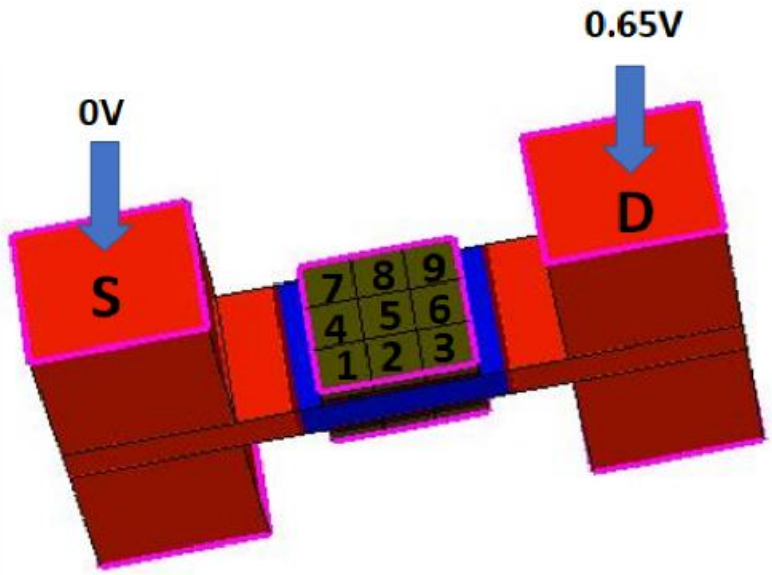


Fig. 3.4. NCFET structure in Sentaurus TCAD simulation. The red regions are source and drain. The blue region is 5nm-thick channel sandwiched by the gate stack, which consists of 0.8nm SiO_2, 2nm segmented FE-DE mixed layer, and metal contact. The n-type source and drain doping are 2E20(#/cm^3), and the p-type channel doping is 1E17(#/cm^3). Gate work function is 4.6eV.

To mimic the P-E loop from experimental measurement, the voltage of metal 1 is swept with metal 2 floating. The P-E loop is measured by sandwiching the FE layer with two metal electrodes, so an internal metal to metal is added in the TCAD simulation. The P-E curve can be extracted by plotting the charge density in metal 1 (P_{mix}) versus the potential difference of metal 1 and metal 2 over the thickness of FE film (E_{mix}). The P-E curve result from TCAD is shown in Fig. 3.3. In Fig. 3.3, the actual remnant polarization is higher because the ferroelectric accounts for only 66.7% of the whole FE-DE mixed layer. At zero electric field, the DE portion contributes zero polarization. The FE part needs 1.5 times of average remnant polarization to build up one time of average remnant polarization, showing that mixed FE-DE phases in the thin

film would degrade the ferroelectricity.

Sentaurus NCFET DE-FE mixed simulation

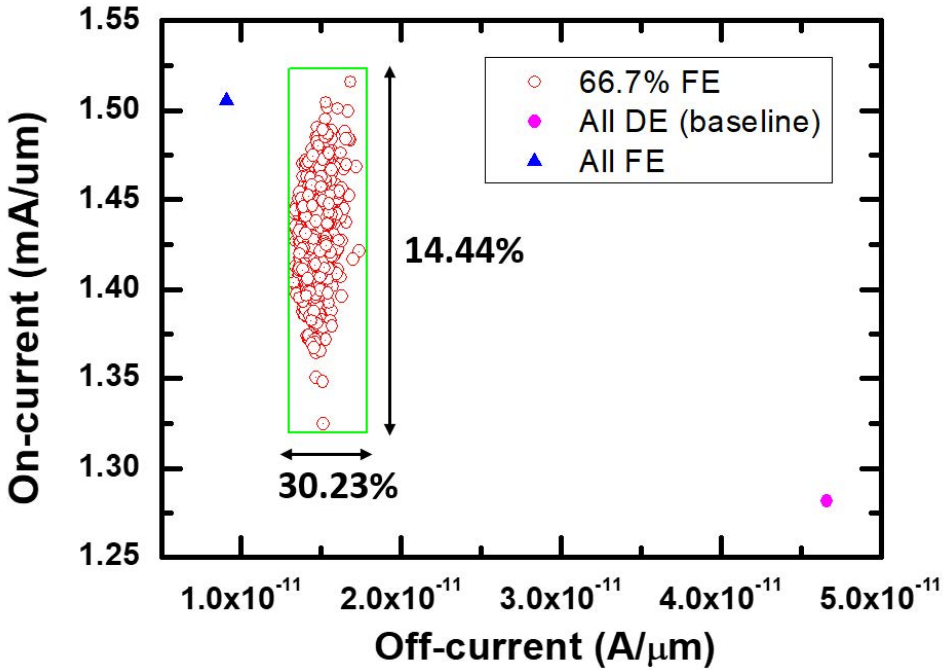


Fig. 3.5. Scatter plot of the random simulation results. The red circles are random simulation results, the pink dot represents that all the segments are DE (baseline), and the blue triangle represents that all the segments are FE.

The ferroelectric layers of the n-channel double-gate NCFET with a gate length of 18 nm and a channel thickness of 5 nm are segmented into 3-by-3 matrix elements in the Sentaurus TCAD simulation, as shown in Fig. 3.4. It is known that the grain size of HZO is in the same order as HZO thin film thickness from the experiment [44], so grain size of 6nm by 6nm by 2nm is a reasonable assumption. Each element is either ferroelectric or dielectric with the same material parameters obtained from section B (ii). The gate stack is the same as the previous MFMIM structure in Fig. 3.2 except the removed intermediate metal layer. Note that without the middle metal layer, the spatial

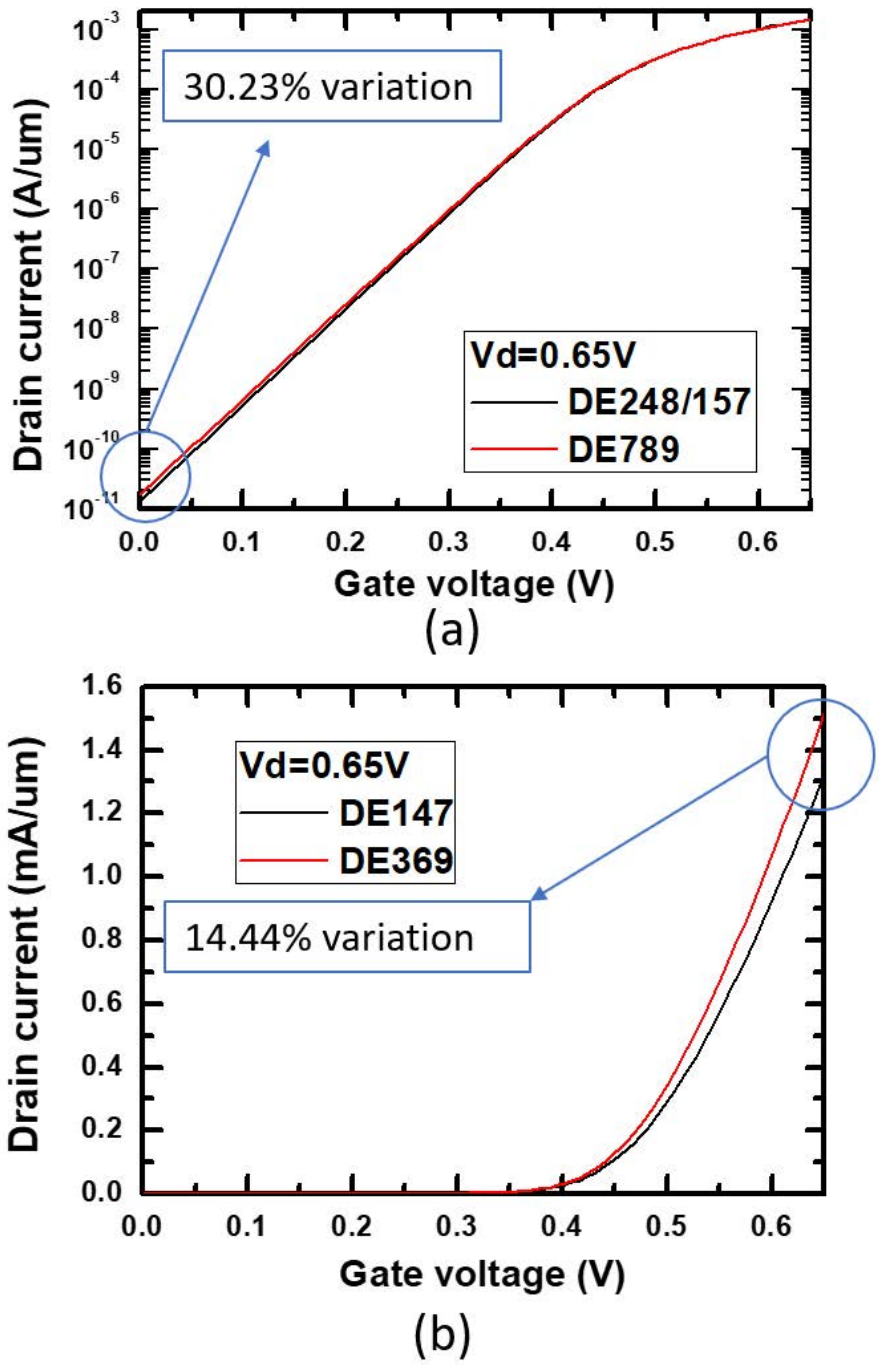


Fig. 3.6. Drain current versus gate voltage characteristics for (a) I_{off} extreme cases with variation of 30.23% and (b) I_{on} extreme cases with variation of 14.44%.

distribution of FE and DE matters. That is why the same characterized FE film can bring out different characteristics in this part.

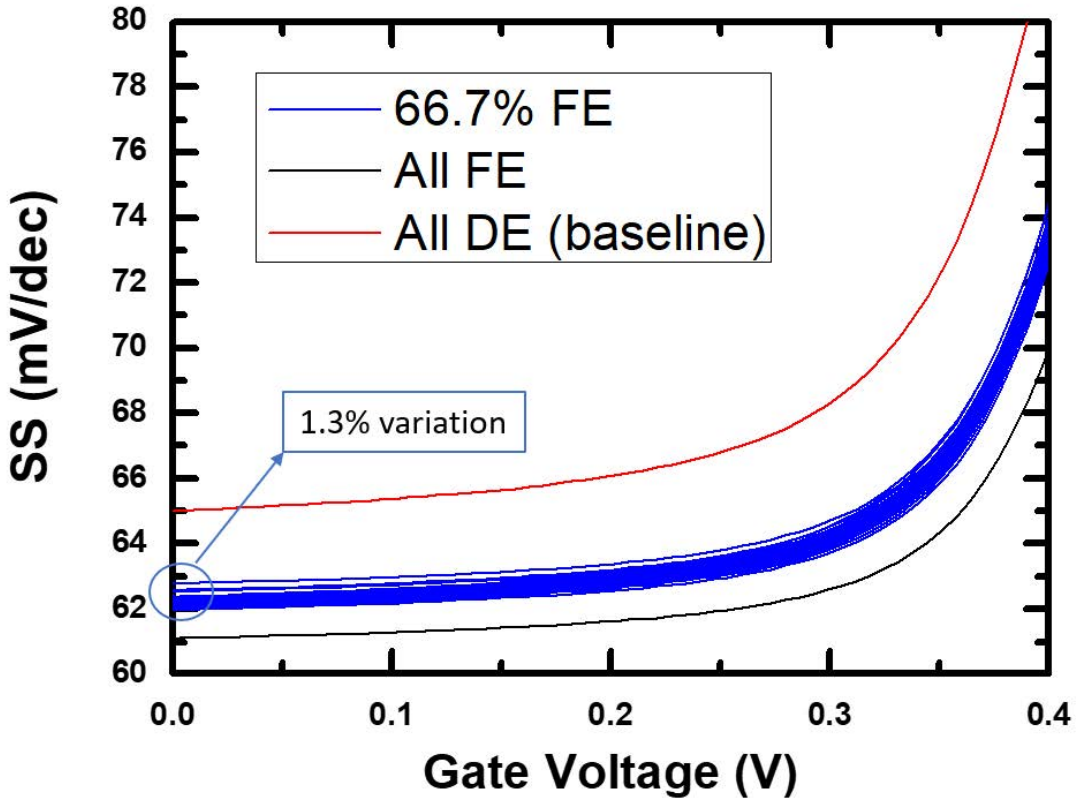


Fig. 3.7. The subthreshold slope versus gate voltage plot of the random DE and FE distribution simulation.

As mentioned in the previous section, 66.7% of the FE-DE layers are ferroelectric, which means two-thirds of segments/grains are ferroelectric. The double-gate NCFET has top and bottom gate stacks with independent DE and FE grains distributions, but 66.7% of each stack consists of FE grains. A random simulation is carried out to examine how the distribution of the DE and FE grains would affect the current of the NCFET. There are 1128 possible combinations by considering two symmetric planes, and 1128 cases are simulated. For each case, V_{DD} is fixed at 0.65V. I_{off} and I_{on} are defined at V_{GS} = 0 and V_{DD} at V_{DS} =V_{DD}, respectively. The results are shown in the scatter plot in Fig. 3.5.

In Fig. 3.5, the green box encloses the boundary of variation due to the different locations of FE and DE grains. Note that the all-FE case means there is no dielectric, but α, β, and γ are modified to fit the experimental P-E loop.

The highest and lowest of I_{off} and I_{on} are shown in Fig. 3.6 (a) and (b), respectively. The highest I_{off} (red curve in Fig. 3.6 (a)) happens when the DE grains on both gates are at the positions of 7, 8, and 9, whereas the lowest I_{off} (black curve) happens when the DE grains are at 2, 4, and 8 on the top gate and at 1, 5, and 7 on the bottom gate. In Fig. 3.6 (b), the highest I_{on} is obtained when the DE grains are at 3, 6, and 9 on both the top and bottom gates,

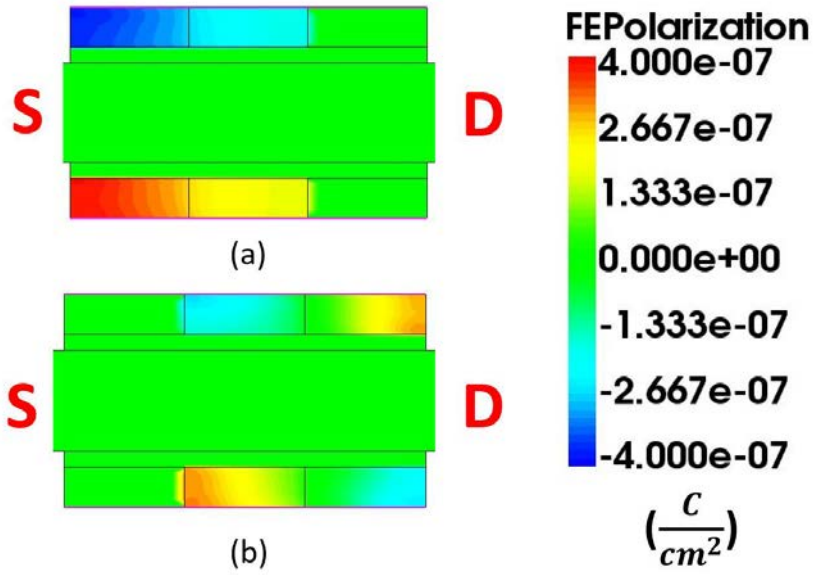


Fig. 3.8. Ferroelectric polarization 2-D plot at $V_{GS} = V_{DD}$. (a) Source-side ferroelectric (location of DE at number 3, 6, and 9 in Fig. 3.4) with the highest Ion, and (b) drain-side ferroelectric (location of DE at number 1, 4, and 7 in Fig. 3.4) with the lowest Ion.

which means all FE grains are on the source side. On the other hand, the lowest Ion appears when DE grains are at 1, 4, and 7 on both top and bottom gates, which means that all FE grains are on the drain side. The subthreshold slope (SS) versus gate voltage is plotted in Fig. 3.7. The SS variation due to the random spatial distribution of DE and FE is approximately 1.3%. Significant SS improvement from the baseline can be seen after adding FE.

3.3 Discussion

In the previous section, the random simulation shows the influence of the DE and FE distributions on the variation in drain current and SS. The physics of the simulation results will be analyzed in detail as follows.

On current extreme cases

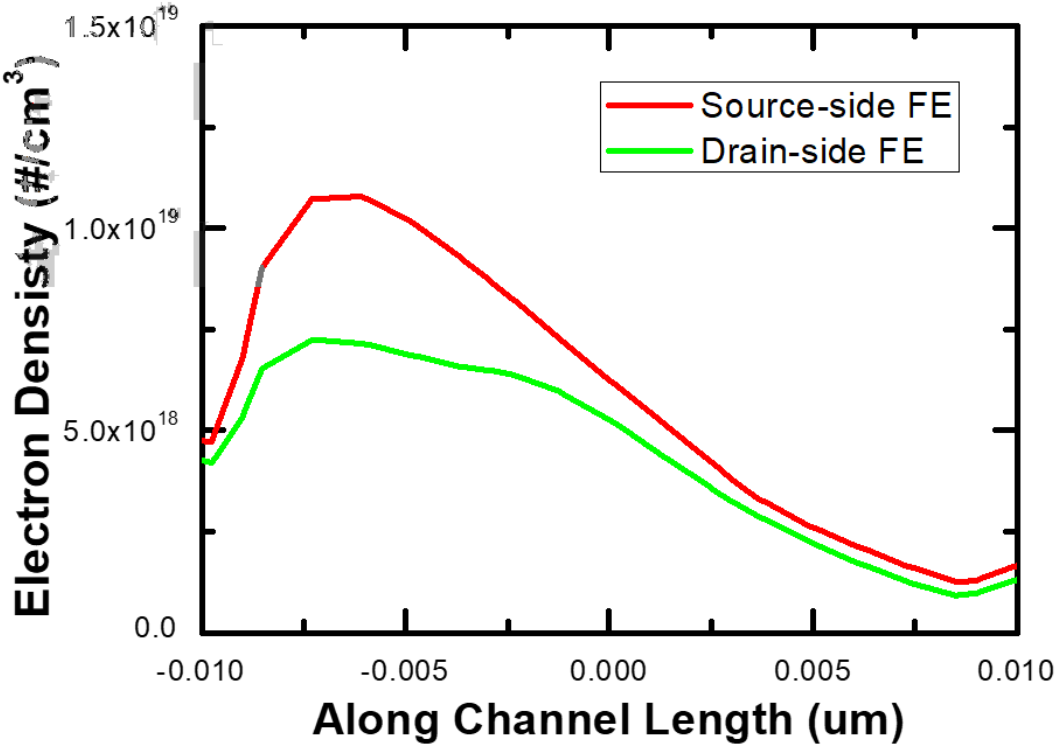


Fig. 3.9. Electron density near the surface along the channel.

As mentioned in the previous section, the highest I_{on} happens when the ferroelectric grains are at the source side, and the lowest I_{on} happens when ferroelectric grains are at the drain side. In Fig. 3.8 (a) and 3.8 (b), the polarization directions of the top and bottom FE are opposite since the electric fields point in opposite directions for the top and bottom gate stacks. To make sure that the FE is in the negative capacitance region, the absolute values of the FE polarization shown in Fig. 3.8 (range from -0.4 μ C/cm^2 to 0.4 μC/cm^2) should be within the range of negative capacitance region (range from -10 μC/cm^2 to 10 μC/cm^2) as determined by the black curve in Fig. 3.3. Therefore, the FE is always in the negative capacitance region in the simulation. For the top gate in Fig. 15 (a), negative P_{FE} (polarization pointing from gate to channel)

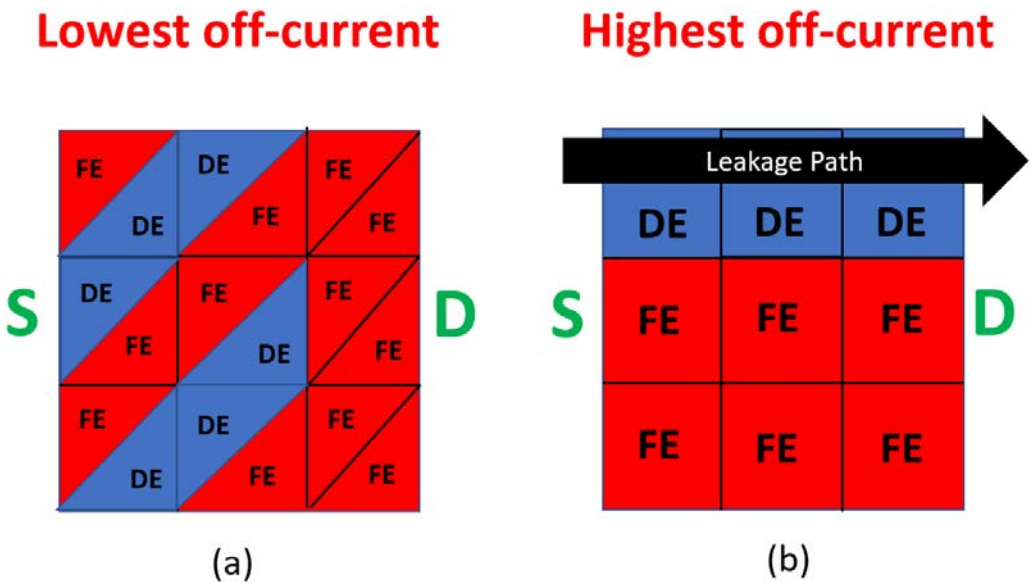


Fig. 3.10. Layout of DE and FE for (a) the lowest I_{off} case (the upper triangles and the lower triangles refer to top gate and bottom gate, respectively) and (b) the highest I_{off} case.

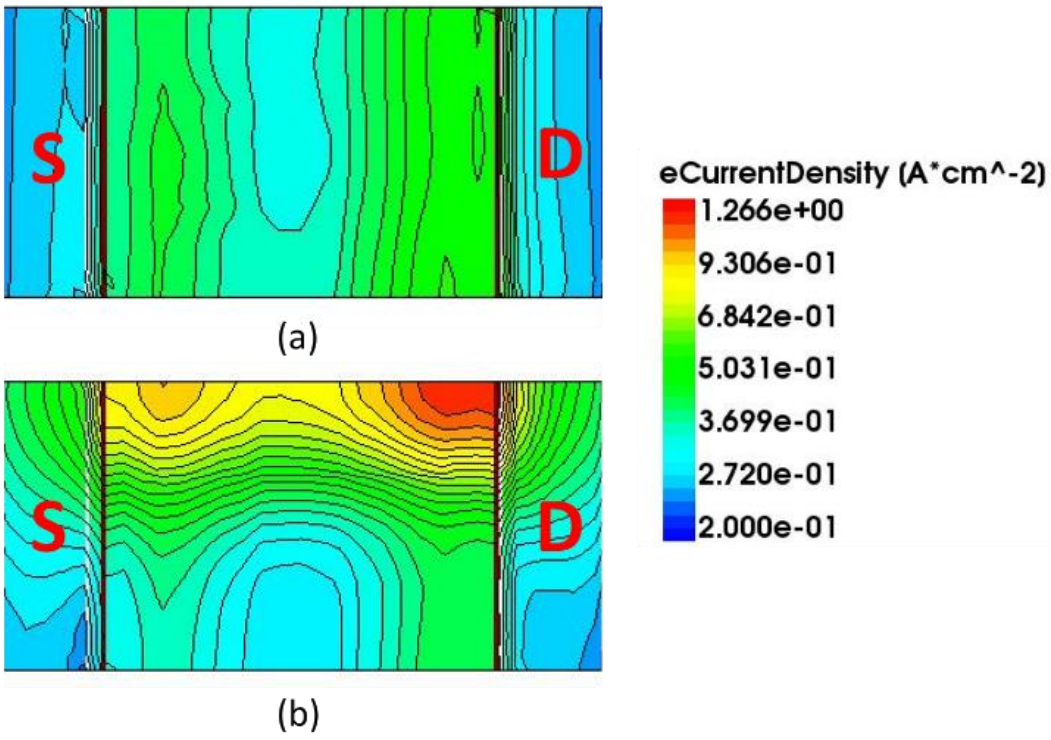


Fig. 3.11. Electron current density in the middle of the channel for (a) the lowest I_{off}

means that the electric field points from channel to gate in the negative

capacitance region. The electric field pointing from channel to gate means that the voltage at the interface of the FE and SiO_2 is higher than the applied gate voltage. The voltage amplification on the bottom gate in Fig. 3.8 (a) can also be explained similarly. However, in Fig. 3.8 (b), the sign of the polarization changes from source to drain because of the fringing field from the drain to the gate [41], which reduces the voltage on the drain side.

To compare the ON current between these two cases, the inversion electron density along the channel length near the surface is plotted in Fig. 3.9. When the FE is at the source side (the carrier injection point of a MOSFET), the source inversion electron density increases due to voltage amplification. In contrast, voltage de-amplification occurs when the FE is at the drain side because the positive drain voltage induces negative gate charges, which reduces the surface potential and inversion electron density. Therefore, the ON current of source-side FE is highest, whereas the drain-side FE is lowest.

Off current extreme cases

Fig. 3.10 (a) shows the lowest I_{off} case where the upper triangles refer to the top gate, and the lower triangles refer to the bottom gate. The highest leakage path is at the center of the channel due to degraded gate control. The layout of DE and FE grains in Fig. 3.10 (a) can control the leakage best among all the cases because the FE grains cover the entire drain side so that the fringing field from the drain can help the ferroelectric suppress the leakage current [41], and also evenly cover over the rest of the channel (see Fig. 3.11 (a)). In contrast, in Fig. 3.10 (b), all the DE grains are in a line along the channel, causing a leakage path along the DE region, clearly seen in Fig. 3.11 (b).

Method of estimating the variation

The proposed method can estimate the additional variation caused by the random spatial distribution of DE and FE grains. First, the percentage of DE and FE grains present should be quantified by XRD or other measurements. The parameters can be extracted by the method described in section B (i). After getting all the parameters required for the simulation, one can get the extrema of I_{on} by putting the FE grains all near the source and the drain, respectively. By putting the DE grains in a line along the channel, the highest I_{off} can be obtained. By putting FE grains in a line on the drain side and

distributing the rest of FE grains evenly but complementary on two sides, the lowest I_{off} can be obtained. Due to the inevitability of dielectric phases inside the ferroelectric [40], this additional variation should be considered. Using this method, device designers can estimate the window of variation by running four cases of TCAD simulations.

The Capacitance matching when FE and DE are mixed

As shown in Fig. 3.3, the actual ferroelectricity is higher than the effectively measured ferroelectricity. As a result, the capacitance matching is not good when the matching is made between an effective negative capacitance and a positive capacitance. By considering the FE-DE mixed model after extracting both DE and FE parameters, device designers can design their NCFETs better.

3.4 Chapter Summary

A dielectric-ferroelectric mixed model is proposed to extract dielectric and ferroelectric material parameters by fitting the experimental hysteresis loop from the literature. Sentaurus TCAD is properly calibrated using an MFMIM capacitor based on these parameters. After that, the impact of dielectric and ferroelectric spatial distribution on NCFET performance is analyzed via Sentaurus TCAD, showing that the ON and OFF current variations can be up to 14.44% and 30.23%, respectively. This dielectric-ferroelectric mixed model is suitable to evaluate the variations in NCFET design.

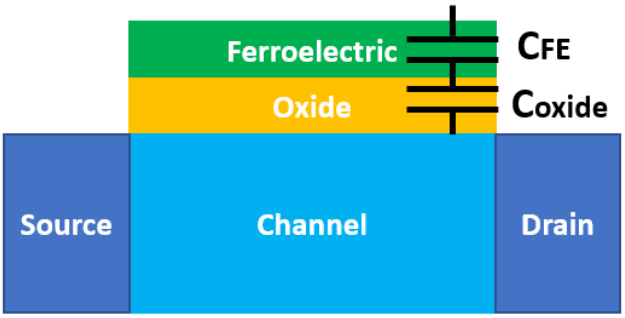
Chapter 4

Analysis and Modeling of Polarization Gradient Effect on Negative Capacitance FET

This section will analyze and provide new insights into the polarization gradient effect of a ferroelectric using TCAD simulation. We demonstrate how to model the polarization gradient effect using the NCFET compact model based on BSIM-framework. A larger value of g (the coefficient of polarization gradient effect) results in improved subthreshold slope, smaller off-current, smaller drain-induced barrier lowering (DIBL), and smaller output conductance. Inclusion of the polarization gradient effect in the NCFET compact model improves usability because it accurately captures the effects of negative DIBL and negative output conductance.

4.1 Motivation

The negative capacitance field-effect transistor (NCFET) is one of the promising technologies for near future logic devices [3]. Fig. 4.1 shows the simplified structure and a representative equation for the NCFET. According to the Landau equation, it is possible to obtain negative differential capacitance from the ferroelectric (FE), and this possibility brings many benefits. The benefits of negative capacitance (NC) include improved subthreshold slope (SS), drain-induced barrier lowering (DIBL), threshold voltage (V_t) roll-off, transconductance (g_m), and output conductance (g_d)



$$C_{total} = \frac{C_{FE}C_{oxide}}{C_{FE} + C_{oxide}}$$

Fig. 4.1. Cartoon structure and equation for capacitance of NCFETs. C_{FE} is possible to become negative according to Landau Equation, and gate control (C_{total}) could be improved.

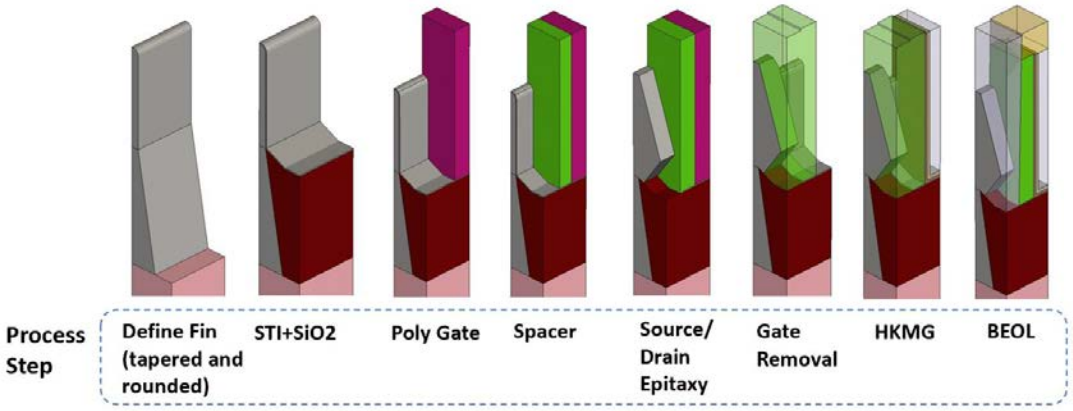


Fig. 4.2. Process simulation flow. Only the source side half of the FinFET is shown.

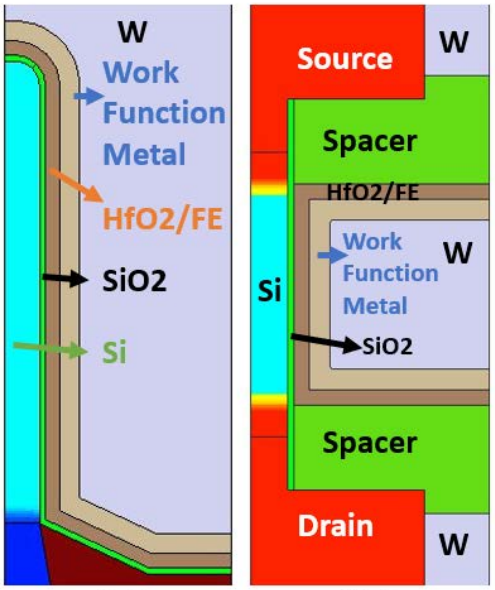
[24,45,46]. Many recent experimental results also demonstrate the benefits of NCFETs [10,36,47,48].

Saha et al. presented an analysis of the polarization gradient effect [19]. By contrast, a different interpretation of the polarization gradient effect will be proposed in this paper. We will also offer a different explanation for why the polarization-electric field points deviate from the S curve when g (the coefficient of polarization gradient effect) increases. The NCFET compact model can help circuit designers to evaluate the performance of NCFETs at the circuit level. Duarte et al. presented both the metal-ferroelectric-insulator-semiconductor (MFIS) and metal-ferroelectric-metal-insulator-semiconductor (MFMIS) compact models [38]. However, the MFMIS structure is not practical because metal between FE and insulator could trap charges and cause V_t shifts. Therefore, in this paper, only the MFIS model

will be discussed. Lin et al. bring the inner-fringing charge model into the MFIS NCFET compact model [49]. Nevertheless, the polarization gradient effect is not considered. The polarization gradient effect captures the polarization interactions between different grid points [19]. This study will demonstrate how the polarization gradient effect can be modeled.

4.2 TCAD SIMULATION SETUP

To build a practical device for TCAD simulations, Sentaurus Process Simulation is first used to construct FinFET structures [50], and the process flow is shown in Fig. 4.2. Note that only half fin and source side is shown for simplicity. The cross-section of the FinFET is demonstrated in Fig. 4.3 (a) and (b), and the geometry and parameters are shown in Fig. 4.3 (c). Gate length and fin widths are 18nm and 7nm, respectively. The gate stack comprises a 0.5nm SiO_2 interfacial later and a 1.5nm high-κ (FinFET baseline) or FE layer (NCFET). Cheema et al. report increasing ferroelectricity with decreasing thickness of HZO [51], and 9nm HZO with P_r = 15μC/cm^2 and E_c = 0.5 MV/cm are reported in [52]. In this paper, P_r and E_c of 1.5nm HZO are assumed to be 20μC/cm^2 and 0.5 MV/cm, respectively. The baseline's intrinsic gate capacitance (Cch) matches the IRDS high performance (HP) requirement, 0.45fF/μm. After the device structure is built, Sentaurus Device Simulation [53] is utilized to simulate the electrical characteristics. The electronic band structure with stress effects includes carrier trajectory, scattering calculations, and the k.p. deformation potential model. Scattering mechanisms such as phonon scattering, impurity scattering, surface roughness scattering, soft optical phonon scattering, remote Coulomb scattering, and impact ionization are included. SRH and Auger recombination is additionally considered. Finally, drift-diffusion with the ballistic mobility model and quantum confinement effect are solved self-consistently with the Sentaurus Device simulator. In Fig. 4.4 (a) and (b), the mobility and the series resistance of the device are calibrated to the Intel 10 nm node [54] by matching the I_d-V_g and I_d-V_d curves. For NCFET simulation, 1-D Landau-Ginzburg Equation is additionally solved, and FE in different regions are assigned with varying polarization directions, and the polarization direction is perpendicular to the channel. Note that instead of the multi-domain state, where sections of the FE material are stable at P_r of different signs, a single domain with



Parameters	Value
Gate Length	18 nm
Spacer Length	7 nm
Fin Height	46 nm
Fin Thickness	7 nm
IL Thickness	0.5 nm
Channel Doping	1E15 cm^{-3}
Source/Drain Doping	2E20 cm^{-3}
HK/FE Thickness	1.5 nm
HK/FE DE constant	21
Coercive Field	0.5 MV/cm
Remnant Polarization	20 $\mu C/cm^2$
Supply Voltage	0.7V

(c)

Fig. 4.3. (a) vertical and (b) horizontal cross sections of the half-FinFET (c) key geometry parameters of the simulated device.

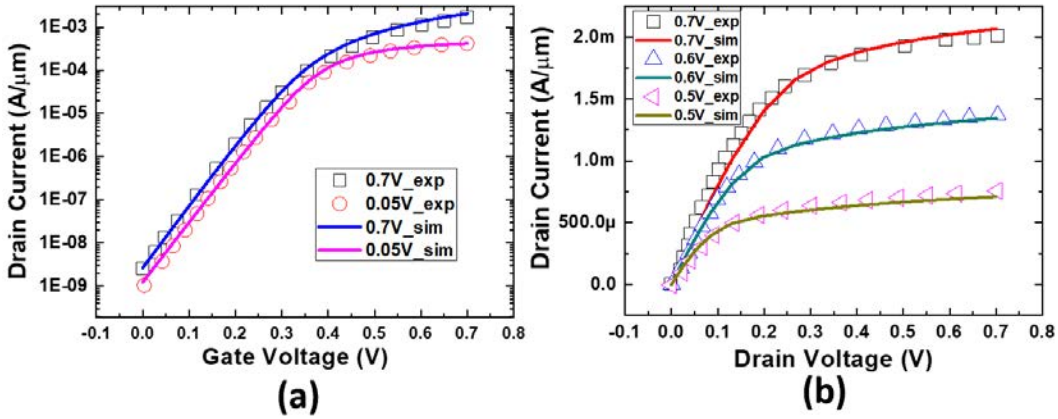


Fig. 4.4 (a) I_d-V_g curve at V_d=0.05 and 0.7V. (b) I_d-V_d curve at V_g=0.5, 0.6, and 0.7V. The data in this figure is normalized to fin pitch according to Intel's convention.

inhomogeneous polarization (polarization gradient) is considered in this simulation [55].

4.3 TCAD ANALYSIS OF POLARIZATION GRADIENT EFFECT

Fig. 4.5 (a) shows polarization-electric field data extracted from the FE layer at the middle of the device along the channel length. At $g = 5 \times 10^{-5}$cm^3/F, which is the minimum value of g that this simulation can converge, the data point (red) follows the S curve. As g becomes larger, e.g., 5×10^{-3}cm^3/F, the data point (green) begins to deviate away from the S curve. Eq. 4.1 shows the relationship between total energy and polarization of the FE [56]

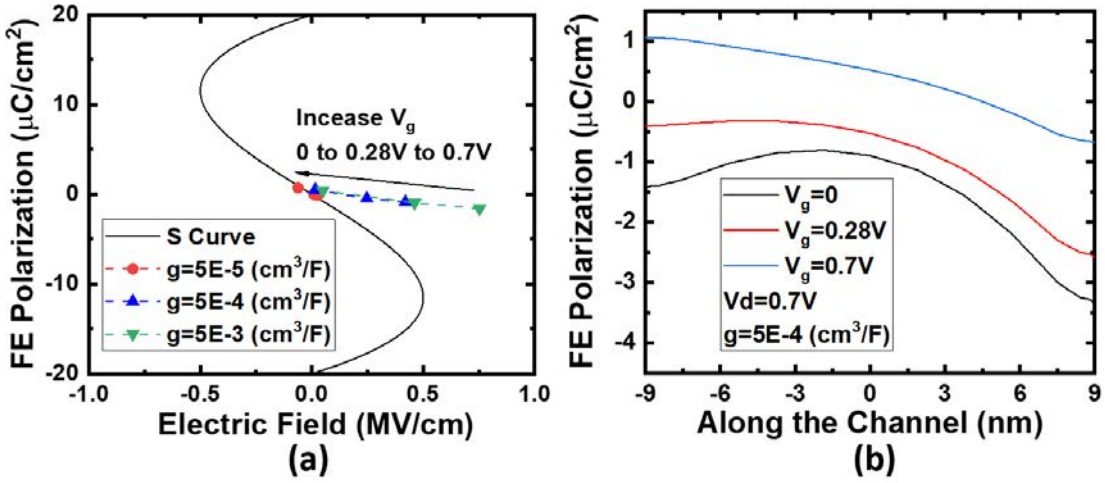


Fig. 4.5 (a) Polarization-Electric Field data points extracted from the middle of the channel length at V_g=0, 0.28, and 0.7V. (b) FE polarization along the channel at V_g=0, 0.28, and 0.7V.

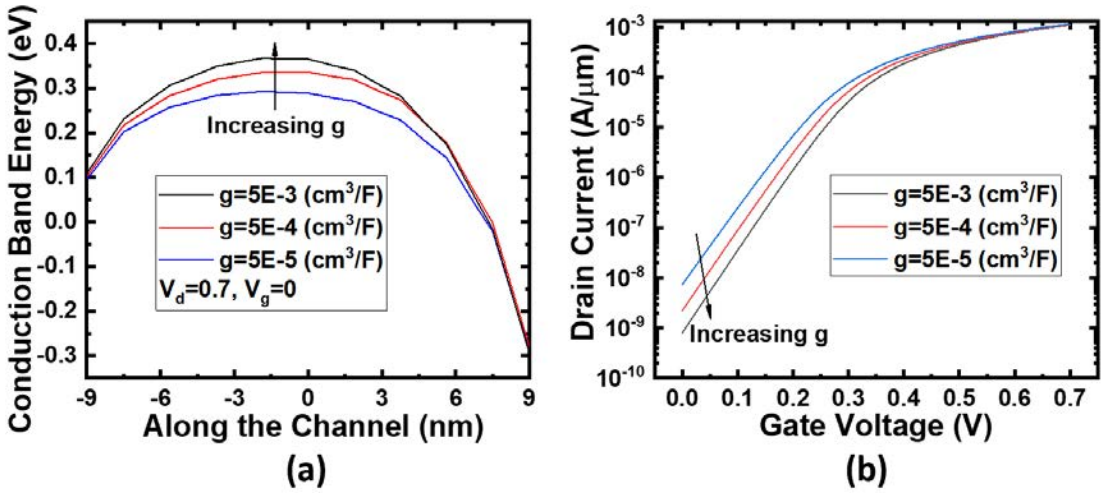


Fig. 4.6. (a) Conduction band energy at off-state and V_d=0.7V along the channel and (b) I_d-V_g curve at g=5×10^{-3}, 5×10^{-4}, 5×10^{-5} (cm^3/F).

$$\rho \frac{dP}{dt} + \nabla_P U_{total\_FE} = 0 \quad (4.1)$$

where ρ is the viscosity coefficient of FE, and U_{total_FE} is the total energy of the FE system. Eq. 4.2 describes the relationship between energy density and polarization of the FE

$$U = \alpha P^2 + \beta P^4 + \gamma P^6 + g|\nabla P|^2 - E \cdot P \quad (4.2)$$

where U is energy density at a position in FE, α, β, and γ are FE material parameters, and g is the coefficient for a polarization gradient effect.

By combining Eq. 4.1 and 4.2 with the functional derivative rule, Eq. 4.3 illustrates the relation between the electric field in the FE and the FE polarization.

$$E = 2\alpha P + 4\beta P^3 + 6\gamma P^5 - 2g\Delta P \quad (4.3)$$

where Δ is Laplace operator. Eq. 4.3 and Fig. 4.5 (b) explain why the FE polarization-electric field data points deviate from the S curve in Fig. 4.5 (a). As shown in Fig. 4.5(b), the curvature of FE polarization is more negative at low V_g than at high V_g, and this effect contributes a positive electric field when g is reasonably large. Therefore, the green data point at low V_g is shifted to the right of the S curve in Fig. 4.5 (a).

Fig. 4.6 (a) shows the conduction band energy along the channel. As g increases, the product of g and ΔP decreases (becomes more negative), leading to positive change in electric field in the FE from Eq. (4.3). From Fig. 4.5 (a), the FE polarization-electric field data point at V_g=0 moves to the right and contributes a more positive electric field (the direction of electric field points down), thus lowering the potential of the silicon channel. The conduction band energy is the negative of potential, so the conduction band energy barrier height increases with g. Hence, the drain current in Fig. 4.6 (b) decreases when g increases due to the increase in barrier height. Fig. 4.7 (a) illustrates the relation between SS and g. SS improves when g increases. This effect is also explained by Fig. 4.5 (a). For an equivalent increase in V_g, the green data (corresponding to a larger g) traverses a bigger change in the electric field compared to the red data (smaller g), thus leading to a larger differential voltage gain contributed by NC. Fig. 4.7 (b) shows DIBL and output conductance versus different g values. DIBL and drain conductance both decrease when g increases, and both phenomena can be explained by the same reason. When drain voltage increases, the curvature of the FE polarization becomes larger, and the data points on the FE polarization-

electric field plot will be shifted to the right (larger electric field), leading to lower channel potential. When g is reasonably large, this effect could overwhelm the short channel effect and lead to negative DIBL and negative output conductance. Fig. 4.8 (a) explains the negative DIBL at $g = 5 \times 10^{-3}$cm^3/F by showing energy barrier increases when V_d increases. Fig. 4.8 (b) shows carrier density in the channel reduces more at a larger g value when V_d increases, leading to a negative output conductance effect when g is significantly large.

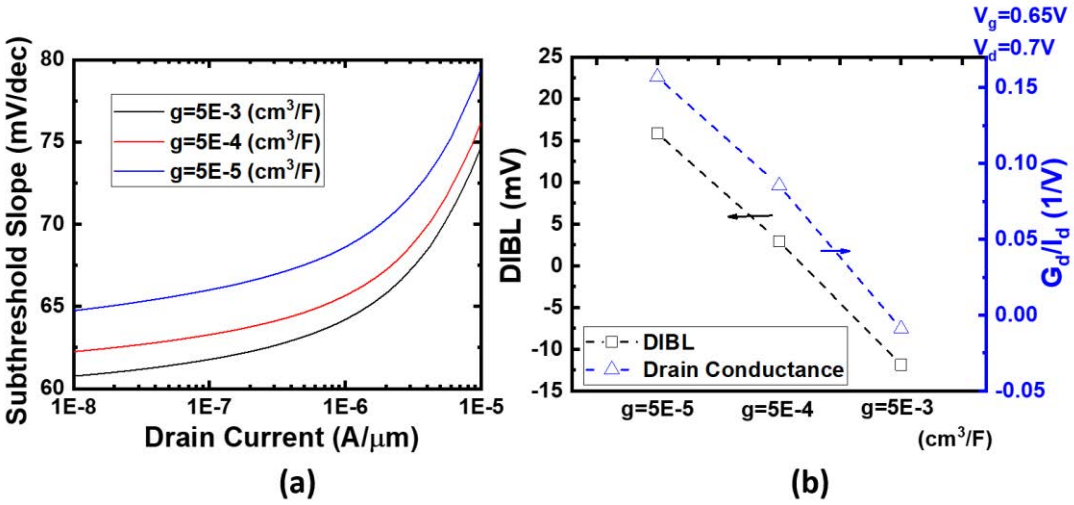


Fig. 4.7. (a) SS versus drain current at g=5×10^{-3}, 5×10^{-4}, 5×10^{-5} (cm^3/F). (b) DIBL and drain conductance versus different g values.

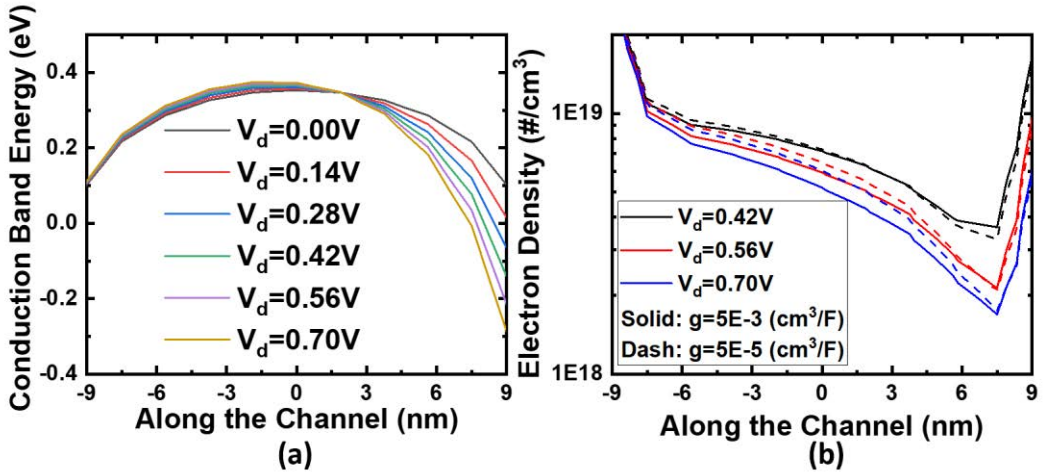


Fig. 4.8. (a) Conduction band energy along the channel at V_g=0 and various V_d. (b) Electron Density along the channel at V_g=0.65V and various V_d.

4.4 Compact model of Polarization Gradient Effect

For the NCFET compact model, the 1-D Landau-Ginzburg equation is considered, and $P \equiv Q_g$ is assumed, so, considering a 1-D scenario ΔP can be simplified as $\frac{d^2 Q_g}{dx^2}$, where x is along the channel direction.

$$Q_g = Q_{inv} + Q_{dep} + Q_{fring} \quad (4.4)$$

Here, Q_{inv} is inversion charge density, Q_{dep} is depletion charge density, and Q_{fring} is fringing charge density [11]. Q_{dep} is assumed to be constant, so it disappears from the second derivative of Q_g.

$$\frac{d^2 Q_g}{dx^2} = \frac{d^2 Q_{inv}}{dx^2} + \frac{d^2 Q_{fring}}{dx^2} \quad (4.5)$$

The second derivative of inversion charge density is negligible in subthreshold, so only strong inversion is considered. The distribution of inversion charge density along the channel can be described empirically [17] as

$$Q_{Inv} = -(B - A\xi)^{\frac{1}{n}} \quad (4.6)$$

where n is a parameter, and ξ is normalized distance. A and B can be solved by using the boundary conditions:

$$At\ \xi = 0,\ q_{Inv} = Q_{is} \quad (4.7)$$

$$At\ \xi = 1,\ q_{Inv} = Q_{idsat} \quad (4.8)$$

where Q_{is} stands for the inversion charge at the source side, and Q_{idsat} stands for the inversion charge density at the drain side. By solving the equation above,

$$A = (-Q_{is})^n - (-Q_{idsat})^n \quad (4.9)$$

$$B = (-Q_{is})^n \quad (4.10)$$

The 2nd derivative of inversion charge density at the source side is assumed to be zero. By taking 2nd derivative of Eq. (4.6) and replacing A and B with Eq. (4.9) and (4.10) respectively, the second derivative of inversion charge density at the drain side can be represented as:

$$\frac{d^2 Q_{Invd}}{d\xi^2} \approx \frac{(n-1)[(-Q_{Invs})^n - (-Q_{Invd})^n]^2}{n^2} (-Q_{Invd})^{1-2n} \quad (4.11)$$

The fringing charge can be represented by Eq. 4.12:

$$Q_{fring} \equiv -Cox\left\{(V_{bi}-V_{SL})\frac{\sinh\left[\frac{Lg-x}{\lambda}\right]}{\sinh\left[\frac{Lg}{\lambda}\right]}+(V_{bi}+V_{DS}-V_{SL})\frac{\sinh\left[\frac{x}{\lambda}\right]}{\sinh\left[\frac{Lg}{\lambda}\right]}\right\} \quad (4.12)$$

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 that depends on the device's geometry [57]. V_{SL} is

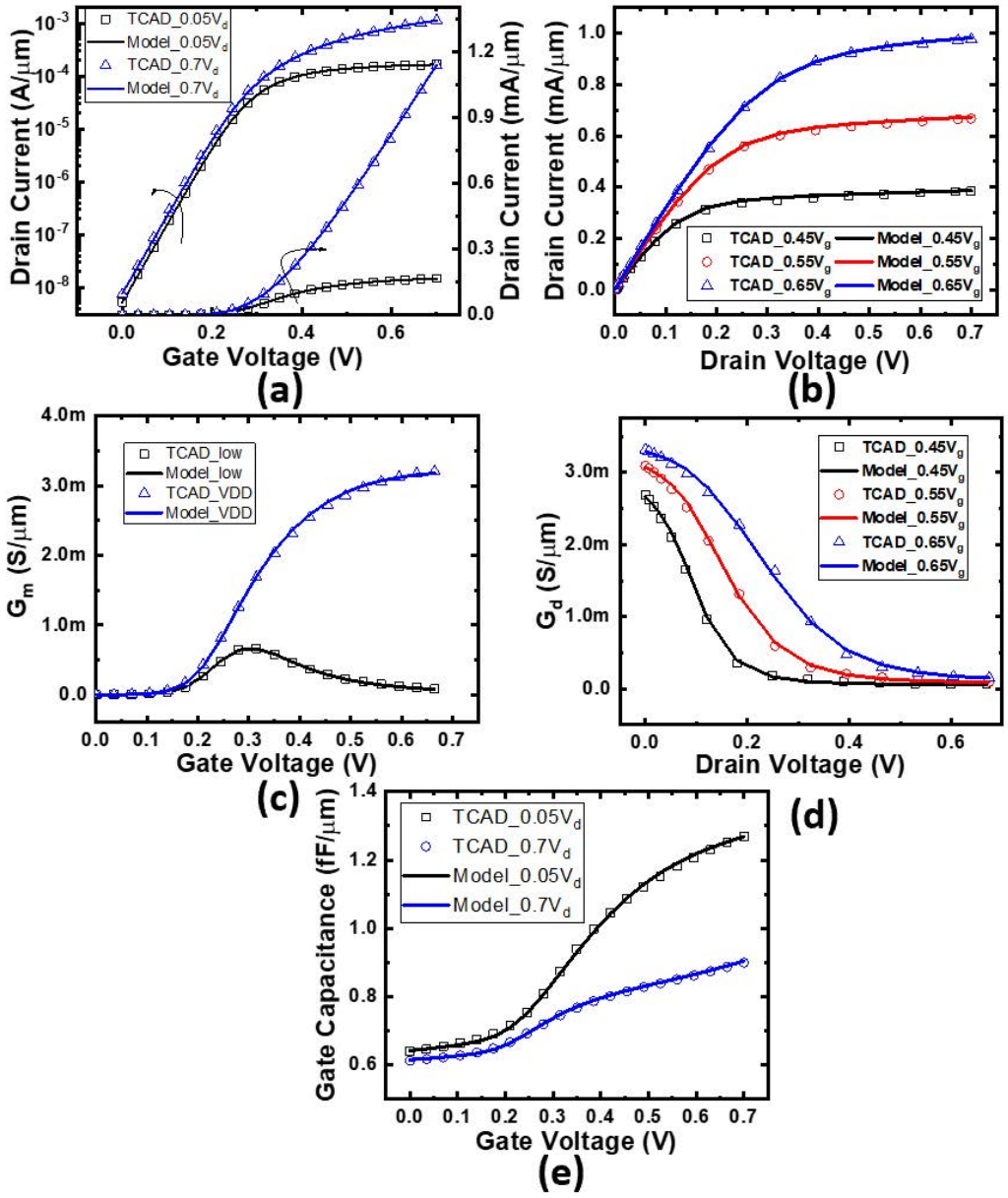


Fig. 4.9. Model validation with TCAD simulation data at $g = 5 \times 10^{-5}$ cm^3/F. (a) I_d-V_g curve, (b) I_d-V_d curve, (c) transconductance, (d) output conductance, and (e) gate capacitance.

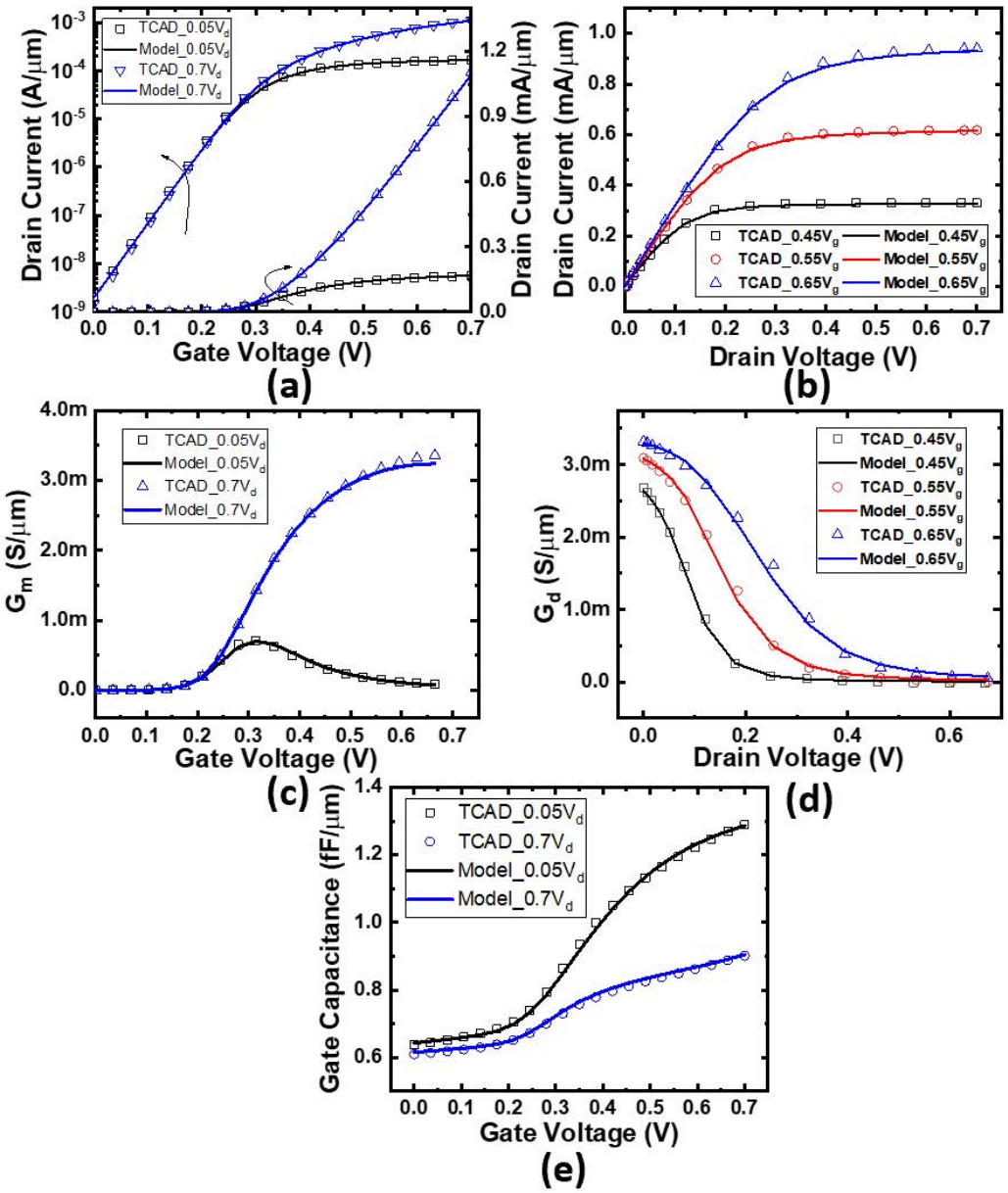


Fig. 4.10. Model validation with TCAD simulation data at $g = 1 \times 10^{-3}$ cm^3/F. (a) I_d-V_g curve, (b) I_d-V_d curve, (c) transconductance, (d) output conductance, and (e) gate capacitance.

defined as

$$V_{SL} = V_{GS} - V_{FB} - \frac{qN_{ch}}{\varepsilon_{Si}}\lambda^2 \quad (4.13)$$

where V_{FB} is the flat band voltage and N_{ch} is the doping concentration in the channel. The 2nd derivative of Q_{fring} can be expressed as:

$$\frac{d^2Q_{fring}}{dx^2} \equiv -\frac{Cox}{\lambda^2}\left\{(V_{bi}-V_{SL})\frac{\sinh\left[\frac{Lg-x}{\lambda}\right]}{\sinh\left[\frac{Lg}{\lambda}\right]}+(V_{bi}+V_{DS}-V_{SL})\frac{\sinh\left[\frac{x}{\lambda}\right]}{\sinh\left[\frac{Lg}{\lambda}\right]}\right\} \quad (4.14)$$

The effect of polarization gradient can be modeled by coupling Eq. 4.11 and Eq. 4.14 with Eq. 4.3 and the unified charge model [10]. Fig. 4.9 shows I_d-V_g, I_d-V_d, transconductance, output conductance, and gate capacitance TCAD

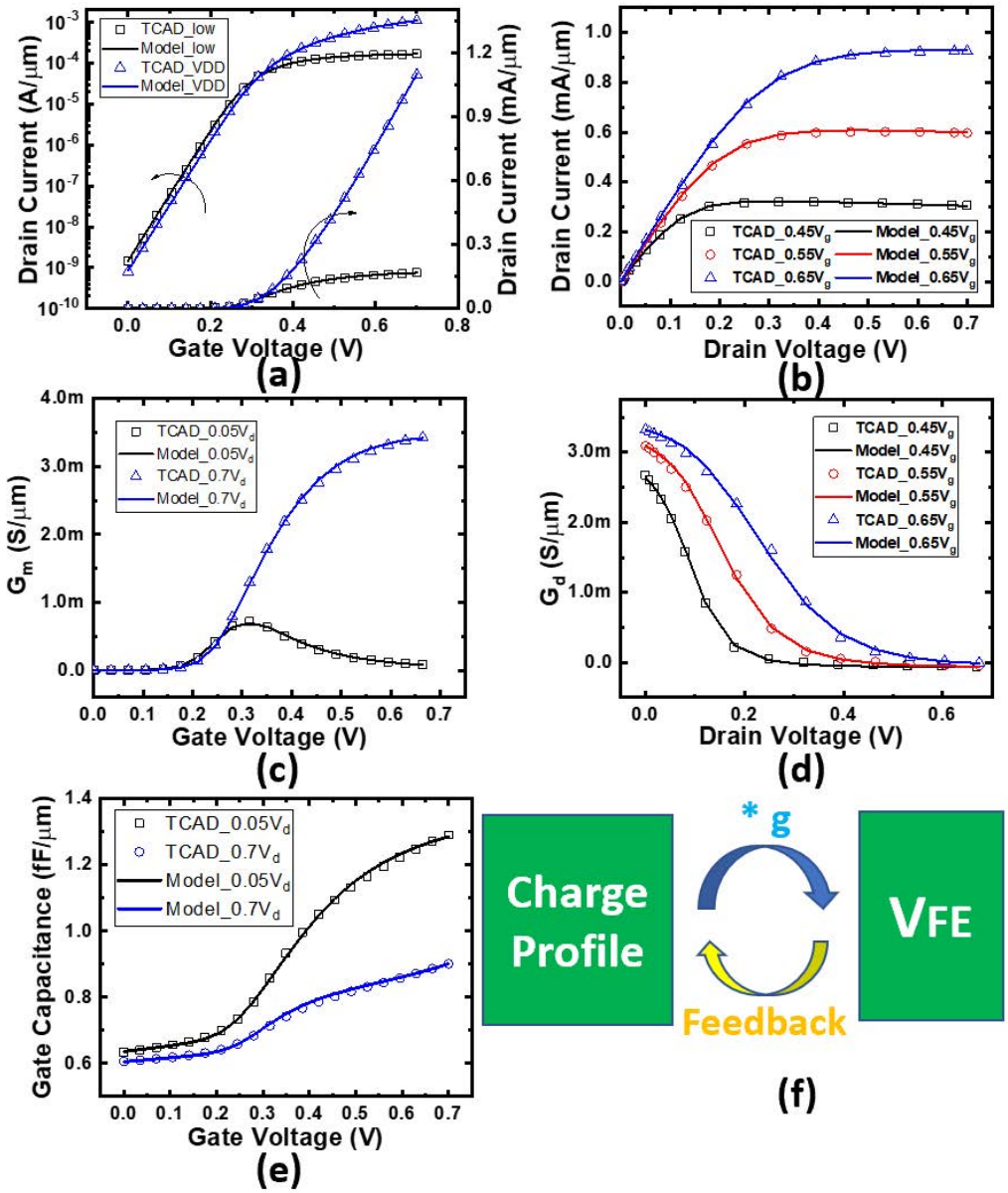


Fig. 4.11. Model validation with TCAD simulation data at $g = 5 \times 10^{-3}$ cm^3/F. (a) I_d-V_g curve, (b) I_d-V_d curve, (c) transconductance, (d) output conductance, and (e) gate capacitance. (f) Diagram of how charge profile and potential of FE affect each other.

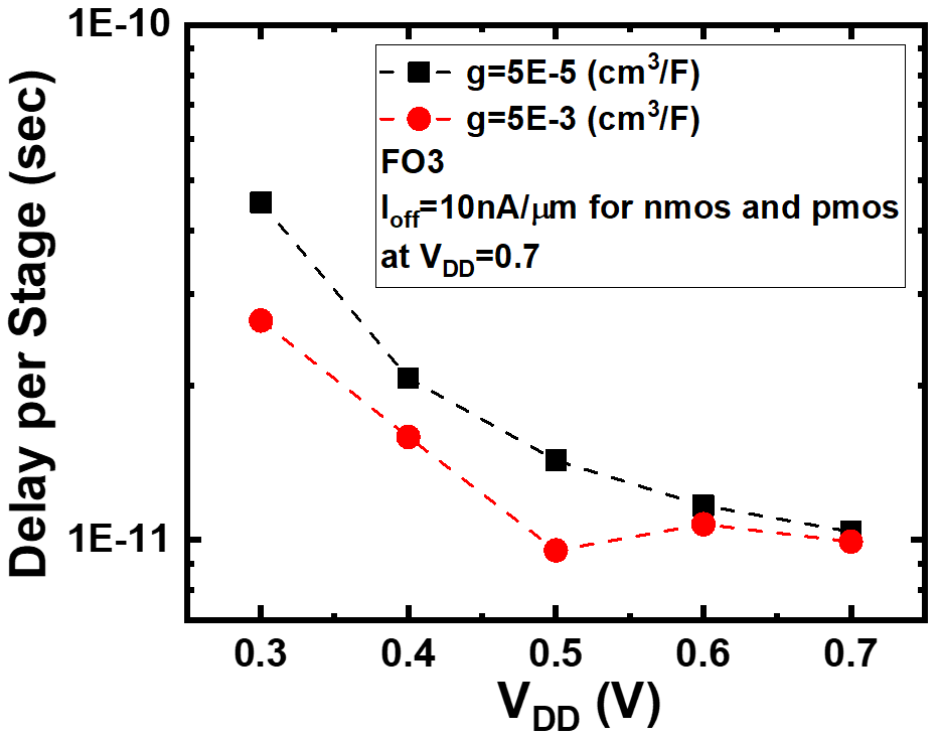


Fig. 4.12. 17-stage ring oscillator simulation with FO3 under different V_{DD}. The delay per state becomes lower at larger g and V_{DD}=0.5V.

data fitted with NCFET compact model at $g = 5 \times 10^{-5}$cm^3/F. The effect of g is negligible at $g = 5 \times 10^{-5}$cm^3/F which can be confirmed by Fig. 4.5 (a). The compact model is represented by lines, and TCAD data is represented by symbols in Fig. 4.9-4.11. The fitting results of I_d-V_g, I_d-V_d, transconductance, output conductance, and gate capacitance at $g = 1 \times 10^{-3}$cm^3/F are shown in Fig. 4.10. From Fig. 4.9 to Fig. 4.10, only the g parameter is changed, demonstrating the model's predictability. DIBL in Fig. 4.10 (a) becomes close to 0, and the slope of the I_d-V_d curve in the saturation region (output conductance) in Fig. 4.10 (b) becomes flatter as compared to Fig. 4.9 (b). Both the change of DIBL and output conductance are consistent with the results in Fig. 4.7 (b). Fig. 4.11 shows the fitting at $g = 5 \times 10^{-3}$cm^3/F where g is large enough to see the negative DIBL and negative output conductance effect. For $g = 5 \times 10^{-3}$cm^3/F, the g parameter along with other parameters like work function, mobility, and parasitic capacitance are tuned to obtain a good match between TCAD and this model as shown in Fig. 4.11 (a)-(e). The reason why the fitting cannot be achieved by tuning only the g parameter is shown in Fig.

4.11 (f). The model captures the blue arrow in Fig. 4.11 (f) by evaluating the second derivative of the charge profile along the channel at g=0. Still, the feedback of the potential to the charge profile is not considered. Even though the model cannot be used as a predictive model at $g = 5 \times 10^{-3}$cm^3/F, it is still an excellent fitting model which can capture both negative DIBL and negative drain conductance as shown in Fig. 4.11 (a) and (d) respectively.

Fig. 4.12 shows abnormal behavior of NCFET with a larger g value. The delay per stage of a ring oscillator decreases when V_{DD} increases normally. However, the delay for $g = 5 \times 10^{-3}$cm^3/F increases when V_{DD} rises from 0.5V to 0.6V. This is attributed to negative output conductance and stronger drain potential coupling to the channel due to the g effect. When the drain bias increases, the current further decreases, leading to an increase in delay. You et al. also report similar behavior in an NCFET circuit [58]. Overall, the delay is less for larger g, especially at lower $V_{DD,}$ because of better SS.

4.5 Chapter Summary

The effect of polarization gradient (g term) is analyzed using TCAD simulation. Off-current reduction, SS improvement, a decrease of DIBL, and a decrease of output conductance are explained by the plot of FE polarization versus electric field in Fig. 4.5 (a). The polarization gradient effect has been implemented in the NCFET compact model based on BSIM-framework. For small to moderate g values, this predictive model can help circuit designers to evaluate potential NCFET circuits. This model also captures and explains the effects of negative DIBL and output conductance.

Chapter 5

Energy Storage and Reuse in Negative Capacitance

This paper analyzes how a ferroelectric (FE) acts as a rechargeable energy storage medium that stores, releases, and retrieves energy and helps the gate achieve the desired charge density with reduced energy (voltage) from the external gate drive. During transistor turn-on, the FE releases energy while the whole system is absorbing energy, and during turn-off, the FE retrieves energy while the entire system is releasing energy. Capacitor energy is analyzed using two different approaches: static material free energy integrals and transient circuit power integrals. The two results agree within 1%. Energy analysis is also performed for a metal-oxide-semiconductor field-effect transistor structure for two gate lengths, 20nm, and 2μm, in an inverter circuit. At 2μm gate length, the values of energy match under these two different approaches with less than a 6% difference. The difference is more significant in the 20nm gate length case due to larger parasitic capacitances, such as gate-to-drain and gate-to-source capacitance, affecting the transient circuit analysis. Even so, most of the energy storage and retrieval benefit is retained even in small-size negative capacitance transistors.

5.1 Motivation

Moore's law has gone a long way in scaling from feature sizes of 10μm to below 30 nm in the past 40 years [59]. One of the obstacles facing scaling is the difficulty in reducing effective oxide thickness (EOT) [60]. Reduction of EOT by simply thinning down the thickness of SiO_2 increases the direct tunneling leakage current exponentially [61,62]. In 2008, Intel adopted the high-κ (HK) metal gate process and improved the direct tunneling gate leakage problem [63]. Even with the help of HK, EOT reduction is still limited because there is a tradeoff between EOT and mobility [64]. In addition, reducing EOT cannot potentially provide less-than-60mV/decade steep turn-on. Salahuddin et al. [3] proposed using a ferroelectric (FE) in the gate stack as a voltage amplifier. As shown in Fig. 5.1, when an FE material stands alone at zero electric field, it is polarized with remnant polarization (P_r) to stay at an energy minimum. When the FE is placed in series with a dielectric (DE),

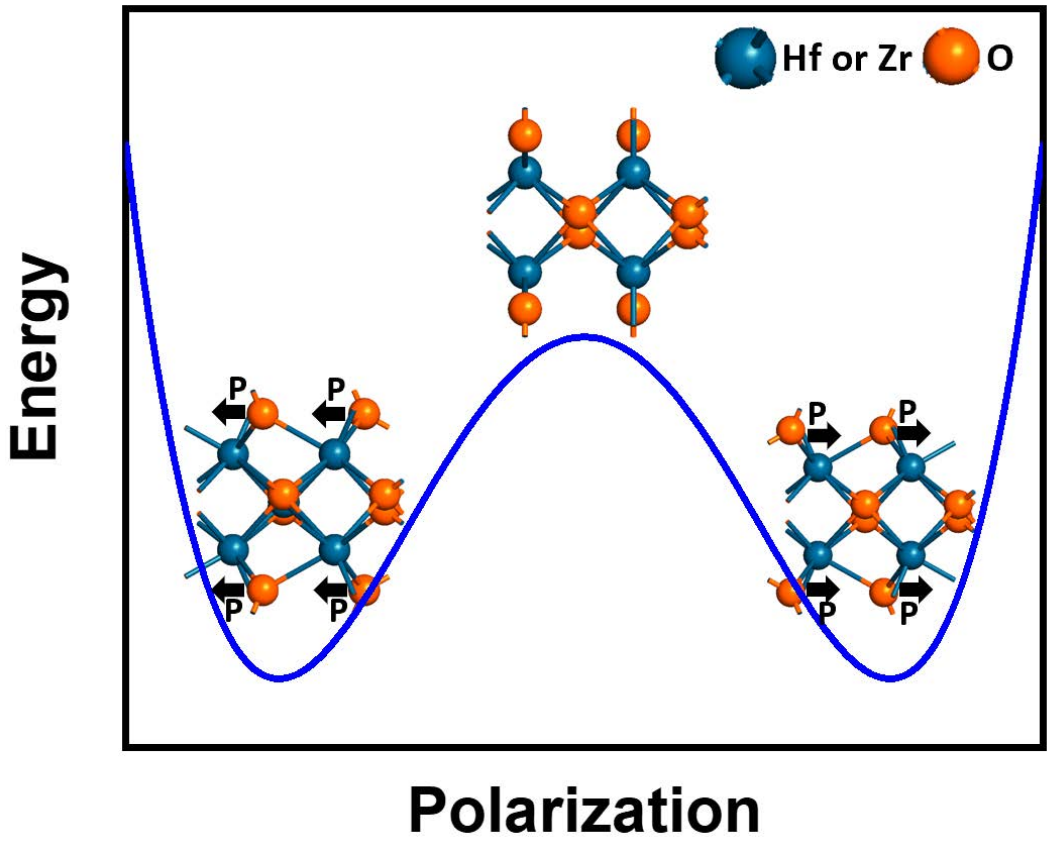
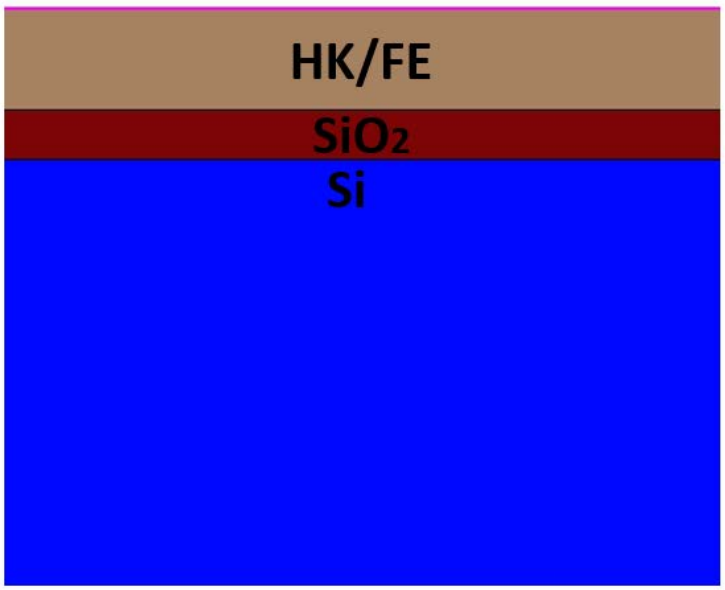


Fig. 5.1. Energy polarization plot at different HZO states without external electric field. HZO with remanent polarization has lowest energy; whereas, the energy is higher at zero polarization because lattice is distorted.



Property	Quantity
Thickness of HK/FE	2nm
Thickness of SiO_2	1nm
Length	1µm
Width	1µm

Fig. 5.2. The MOSCAP Structure used in TCAD simulation and its key parameters. FE parameters will be described in the text.

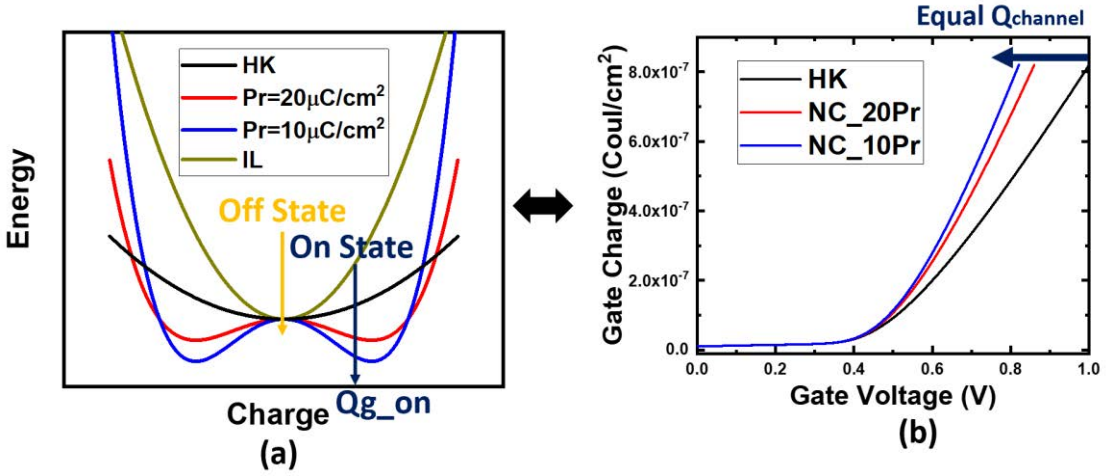


Fig. 5.3. (a) Energy versus charge plot of four materials. (b) Gate charge versus gate voltage of three cases.

whose energy minimum happens at zero charge (polarization), and the capacitance of the DE is small enough so that the energy minimum of the combined system of DE and FE together happens at charge (polarization) equals to zero, this phenomenon is called depolarization of the FE. The curvature of the energy-polarization plot of FE at zero polarization (see Fig. 5.1) is negative, so the capacitance here is also negative. This negative

capacitance (NC) can reduce the EOT of the gate stack. It has been pointed out that NCFET is a quasi-adiabatic device in that part of the energy required for the gate to create the inversion channel is retrieved by the FE material [37,65].

Many NC experiment studies have been published [34,66-71]. The FE Landau model has recently been incorporated into the Sentaurus technology computer-aided design (TCAD) tool [53]. With the help of a commercial TCAD tool, studying what is happening inside NC devices has become possible. In [72,73], Hoffman et al. and Wong analyze the energy within NC-metal-oxide-semiconductor capacitor (MOSCAP) devices. In this paper, energy analysis of NCFETs will be done with the help of TCAD. This paper is composed of two parts. The first part will be energy analysis on simple NC-MOSCAPs, and the second part will be energy analysis on NC- metal-oxide-semiconductor field-effect transistors (MOSFET). Note that the static energy is integrated by the values of meshed points along the channel for both parts.

5.2 ENERGY ANALYSIS OF NC-MOSCAPS

We begin with the simpler discussion of the metal-oxide-semiconductor capacitor (MOSCAP) structure modeled with a 1-D electric field. The MOSCAP structure used in the TCAD simulation is shown in Fig. 5.2. It comprises a Si substrate of 500 nm thickness, a SiO_2 interfacial layer (IL) of 1 nm thickness, and a HK or FE layer of 2 nm thickness. There are three cases used in simulation: 1) HK with dielectric constant 16; 2) FE with remanent polarization (P_r) = 20μC/cm^2 and coercive field (E_c) =1MV/cm (NC_20Pr); and 3) FE with P_r=10μC/cm^2 and E_c=1MV/cm (NC_10Pr). As a single domain study of FE, the viscosity term (ρ) is set to be 0.18 Ω·cm [30] in cases 2 and 3. The minimum $|C_{FE}|$ (P_{FE} = 0) of NC_10Pr is $\sim 1.2 \times 10^{-5}\ \frac{F}{cm^2}$, and the C_{IL} is $\sim 3.5 \times 10^{-6}$. *Hence,* NC condition is satisfied ($|C_{FE}|>C_{IL}$) [3] for both NC_10Pr and NC_20Pr (NC_20Pr has larger $|C_{FE}|$). All the simulations are done by both forward and reverse sweep to make sure hysteresis does not happen. It is easier to visualize the energy benefits of NC if the FE layer is treated as a rechargeable energy storage medium. In Fig. 5.3 (a), the red and the blue curves are the energy-charge profile of FE with P_r= 20μC/cm^2 and P_r= 10μC/cm^2, respectively. The FEs represented by red and the blue curves

have E_c = 1MV/cm. The energy of the red and the blue is higher at the off-state (yellow arrow in Fig. 5.3 (a)) as compared with on-state (blue arrow in Fig. 5.3 (a)), and the FE acts as a charged energy storage medium. When the gate voltage of MOSCAP is swept from V_g=0 to V_g=V_{DD}, energy flows out of the FE, leading to lower energy in the FE at the on-state (blue arrow in Fig. 5.3 (a)). The simulation in this section tries to determine where the FE energy goes. The V_{DD}s of NC_20Pr and NC_10Pr are reduced to achieve the same gate charge density (Q_{gate}) at V_g = V_{DD}. Under the same Q_{gate} among different cases, the channel charge ($Q_{channel}$) will be the same, so the energy stored in IL and Si substrate will be the same among the three cases. Alignment of Q_{gate} at V_g=V_{DD} can exclude factors other than the energy stored in HK/FE, like the energy stored in the IL and the Si substrate.

Two approaches will be used to check the energy. The first approach involves static material free energy integrals throughout the material of HK or FE using Eq. (5.1) and (5.2) [30]

$$U_{HK} = \int \left[\frac{\epsilon {E_z}^2}{2} + \frac{\epsilon {E_x}^2}{2}\right] dxdydz \quad (5.1)$$

$$U_{FE} = \int \begin{bmatrix} \alpha P^2 + \beta P^4 + \gamma P^6 + \\ g|\nabla P|^2 + \frac{\epsilon {E_z}^2}{2} + \frac{\epsilon {E_x}^2}{2} \end{bmatrix} dxdydz \quad (5.2)$$

where U_{HK} and U_{FE} are free energy of HK and FE respectively; E_z and E_x are the electric fields along the vertical gate stack direction and along the gate length direction respectively; α, β, and γ are the FE parameters, and g is the polarization gradient coefficient. 1-D polarization is assumed here, and the integral is performed over the volume of HK or FE. The second approach involves integrating the gate voltage with respect to the gate charge to determine energy inflow to the gate:

$$U_{in} = \int V_g dQ_g \quad (5.3)$$

where U_{in} is the energy inflow, V_g is the gate voltage, and Q_g is the gate charge.

Fig. 5.4 (a) shows the static free energy integration of the HK and FEs. Off-state energy is minimal and negligible. The energy-saving in comparing the HK to NC_20Pr is 581 aJ, and the energy-saving in comparing the NC_20Pr to NC_10Pr is 161 aJ. Fig. 5.3 (a) also helps visualize the differences in energy. At the off-state, every case is essentially at zero energy. At the on-state, the

free energy of HK is positive, but the energies of NC_20Pr (red) and NC_10Pr (blue) are negative, leading to a static free energy saving going from HK to NC. Fig. 5.4 (b) shows the energy inflow to the gate. Energy inflow to the gate saves 582 aJ going from HK to NC_20Pr and 160 aJ going from NC_20Pr to NC_10Pr. Comparing Fig. 5.4 (a) and (b), the energy-saving calculated from the two different approaches is consistent. FE acts as a rechargeable energy storage medium, which can recycle energy from the on-state to the off-state and releases the energy back to the system from the off-state to the on-state, making the energy efficiency of the whole system better. Moreover, the consistency between Fig. 5.4 (a) and (b) indicates that the FE's energy storage

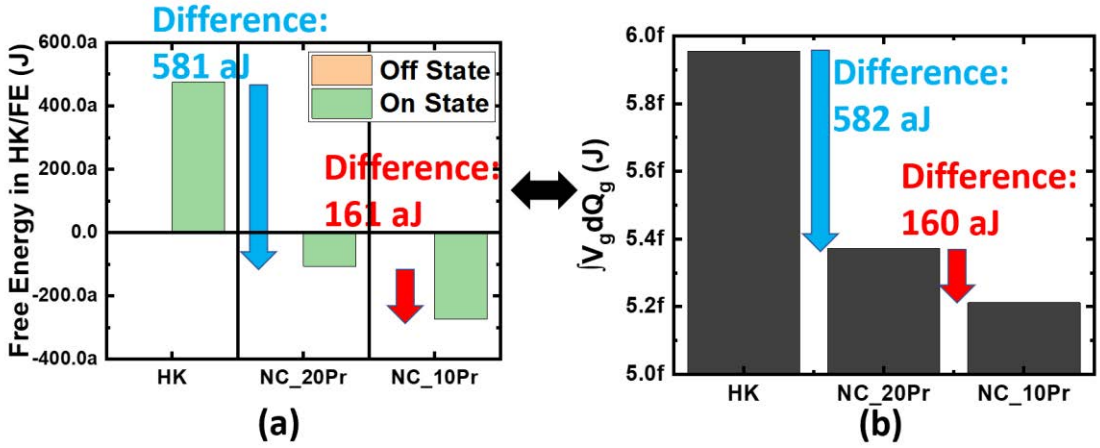
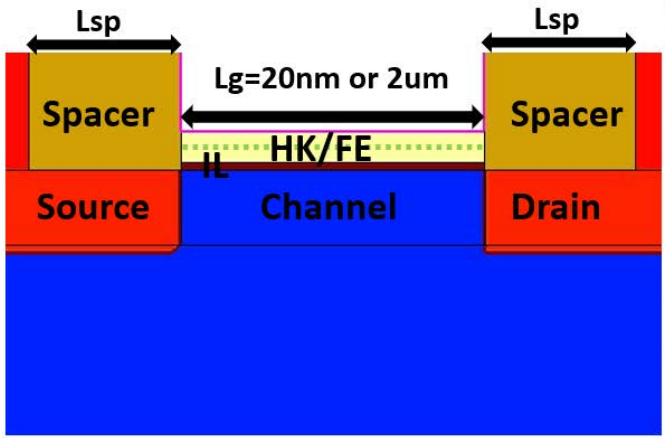


Fig. 5.4. (a) Static material free energy integration throughout either HK or FE. (b) Energy inflow to the gate from the off-state to the on-state.



Property	Quantity
L_G	20nm/2μm
L_{SP}	10nm
ε_{SP}	2
T_{IL}	1nm
$T_{HK/FE}$	2nm
$\varepsilon_{HK/FE}$	16
S/D Doping	$2 \times 10^{20} cm^{-3}$
Channel Doping	$5 \times 10^{18} cm^{-3}$

Fig. 5.5. The MOSFET structure used in TCAD simulation, and its key parameters. FE parameters are the same as previous section.

and release process is nearly lossless and can be almost entirely accounted for in the energy inflow calculation results.

5.3 ENERGY ANALYSIS OF NC-MOSFETS

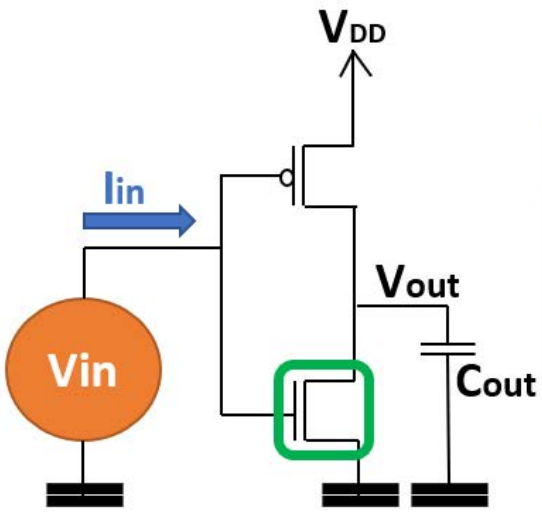
The second portion of this paper covers the simulation of the MOSFET structure. Fig. 5.5 shows the MOSFET structure used in the TCAD simulation. Some important device parameters are listed in the table in Fig. 5.5. Fig. 5.6 shows the inverter circuit, which will be later used for the transient simulation in TCAD. The table in Fig. 5.6 indicates 6 cases: three for L_g = 20nm, and three for L_g = 2μm. For each gate length, there are three cases: HK, NC_20Pr, and NC_10Pr, where the definitions are the same as described in section II.

The simulation flow should be carefully designed to exclude the effect from other factors. For each gate length, the n-type MOSFET (nmos) work function with HK (case 1 & 4) is adjusted to match I_{off} = 10nA/μm. Work function and V_{DD} of NC_20Pr and NC_10Pr are adjusted to match the gate charges at the off-state (V_g = 0 & V_d = V_{DD}) and the gate charges at the on-state (V_g = V_{DD} & V_d = 0) with HK. The solid line in Fig. 5.7 (a) shows that gate charges of HK, NC_20Pr, and NC_10Pr are aligned at the off-state (from 0 to 1ns & from 4ns to 5ns) and on-state (from 2ns to 3ns) at L_g = 20nm. The solid line in Fig. 5.7 (b) shows the gate charges of HK, NC_20Pr, and NC_10Pr are aligned at the off-state and on-state at L_g = 2μm. P-type MOSFET (pmos) is required for an inverter simulation. To achieve symmetric I_d-V_g at V_d=V_{DD} between nmos and pmos, the width of pmos at L_g = 20nm is designed to be 1.81μm, and the width of pmos at L_g = 2μm is designed to be 3μm. The widths of nmos at L_g=20nm and L_g = 2μm are both designed to be 1μm.

Fig. 5.8 (a) shows the I_d-V_g plot at L_g = 20nm, and Fig. 5.8 (b) shows the I_d-V_g plot at L_g = 2μm. Note that the on-current and off-current of pmos are matched with the on-current and off-current of nmos for every case. The on-currents in Fig. 5.8 (a) at L_g = 20nm are very similar among the different cases under the same gate charge at the on-state because the drain current in the saturation region for a short channel MOSFET can be approximated as:

$$I_{ds\_sat\_short} \approx Q_{ch} \times V_{sat} \quad (5.4)$$

where $I_{ds\_sat\_short}$ is the drain current in saturation for short channel. Q_{ch} is the channel charge density, and V_{sat} is saturation velocity, which is not a strong function of V_{DD} in short channel MOSFETs. Under the same Q_g, Q_{ch} will be



	Lg=20nm	Lg=2um
HK	1	4
NC_20Pr	2	5
NC_10Pr	3	6

Fig. 5.6. The inverter circuit which will be used in TCAD simulation. The table on the right-hand side indicates there are 6 cases in this simulation.

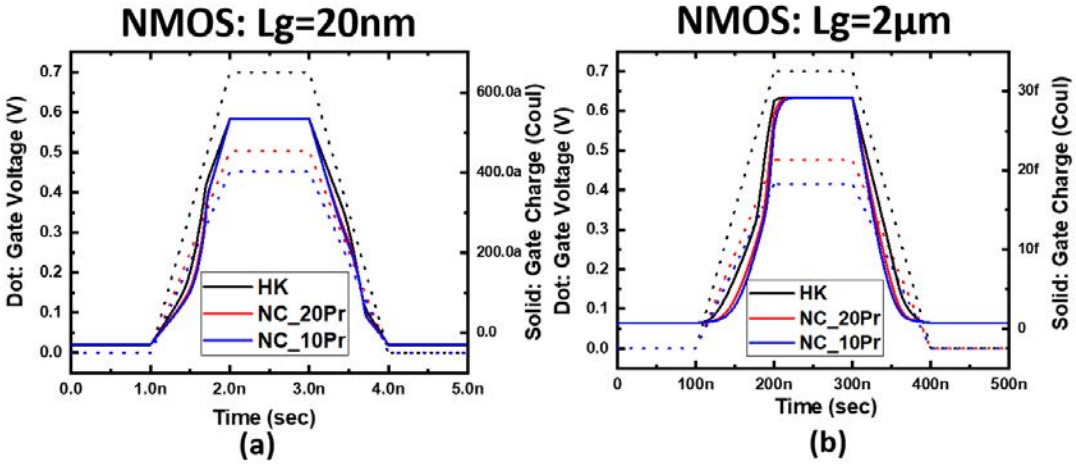


Fig. 5.7. Gate voltage (dash & left axis) and gate charge (solid and right axis) versus time at (a) L_g = 20nm and (b) L_g = 2µm.

very close, so $I_{ds\_sat\_short}$ will also be very similar. In contrast, the drain current in saturation at L_g = 2µm can be approximated by the long channel equation:

$$I_{ds\_sat\_long} \approx Q_{ch} \times \mu \times \frac{(V_{gs}-V_t)}{L_g} \quad (5.5)$$

where V_{gs} is the gate to source voltage, V_t is the threshold voltage, and L_g is gate length. $I_{ds\_sat\_long}$ depends on V_{gs}, which will be affected by V_{DD}, under the same Q_{ch}. Hence, the on-current of the NC case is smaller than the on-current of the HK case at L_g = 2µm, even under the same Q_g in Fig. 5.8 (b).

The output capacitor (C_{out}) in the inverter circuit in Fig. 5.6 is set to be 50

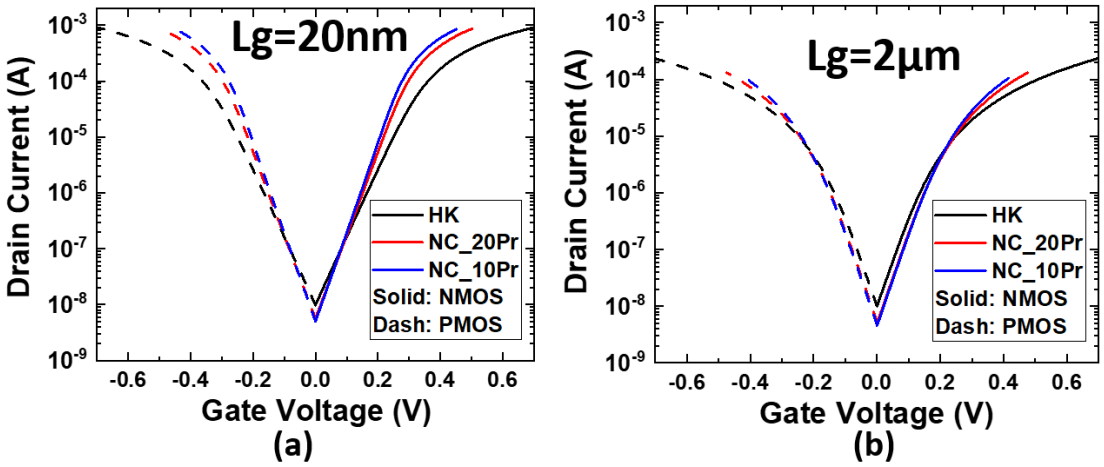


Fig. 5.8. I_d-V_g plot at (a) L_g=20nm and at (b) L_g=2µm.

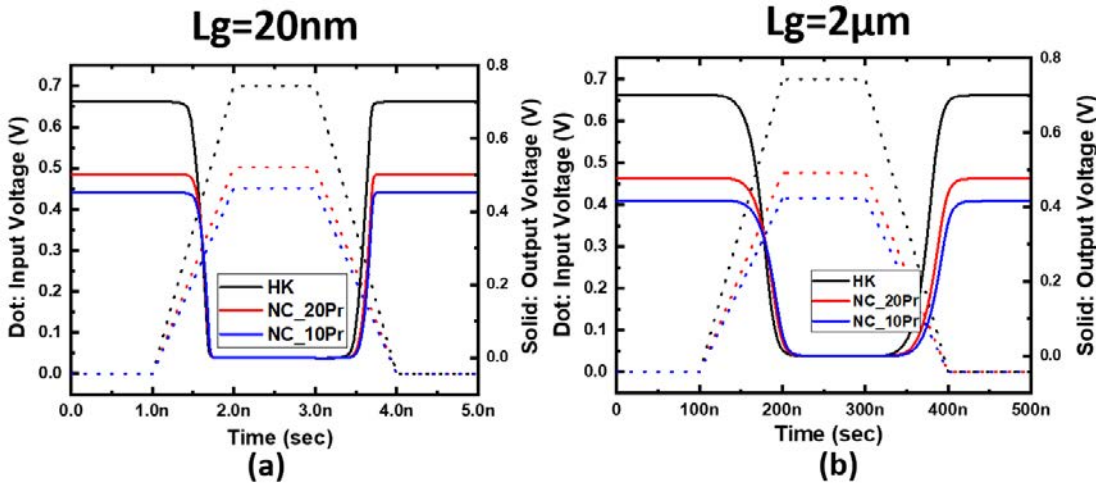


Fig. 5.9. Input and output voltage from the TCAD transient simulation of the circuit as shown in Fig. 6 at (a) L_g = 20nm and at (b) L_g = 2µm. Input voltage is dash line and corresponds to left axis, and output voltage is solid line and corresponds to right axis.

fF at L_g = 20nm and 5pF at L_g = 2µm. The ramp-up time and ramp-down time of gate input voltage are set to be 1 ps at L_g = 20nm and 100ps at L_g = 2µm. The gate-to-drain coupling transient behavior, an undesirable factor in this study, can be minimized by increasing C_{out} and increasing ramping time with increasing L_g. Input and output voltage simulation results of the inverter are shown in Fig. 5.9. Note that there is no rise in the output voltage when the input voltage ramps up, meaning that transient gate-to-drain coupling behavior is minimized. Note that there is an additional delay for the NC compared to HK due to smaller I_{ds} under reduced V_{DD} at L_g = 2µm, as explained earlier. Two approaches for energy calculation are adopted in this

energy analysis. The first is the static free energy integration throughout the entire volume of the HK or FE, as explained in section II. The second is transient input power integration using the equation:

$$U_{in} = \int V_{gs} I_g dt \quad (5.6)$$

where V_{gs} is the input voltage, and I_g is the input current. Note that only gate energy input is considered here, and drain displacement current is ignored so that the gate-to-drain parasitic capacitance may cause a mismatch between static material free energy integrals and transient circuit power integrals at

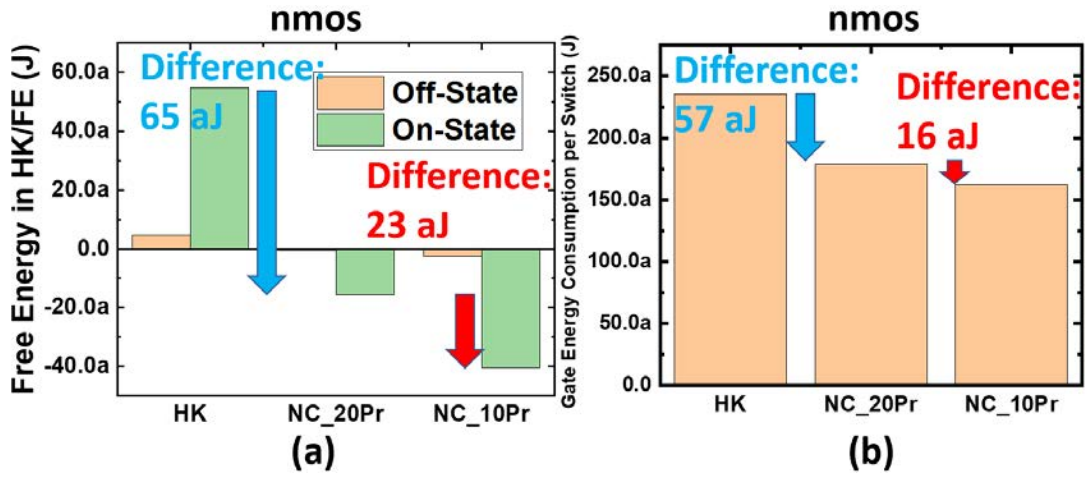


Fig. 5.10. (a) Static free energy of HK/FE in different cases at L_g = 20nm. (b) Gate energy consumption of nmos in the inverter transient simulation per switch at L_g = 20nm.

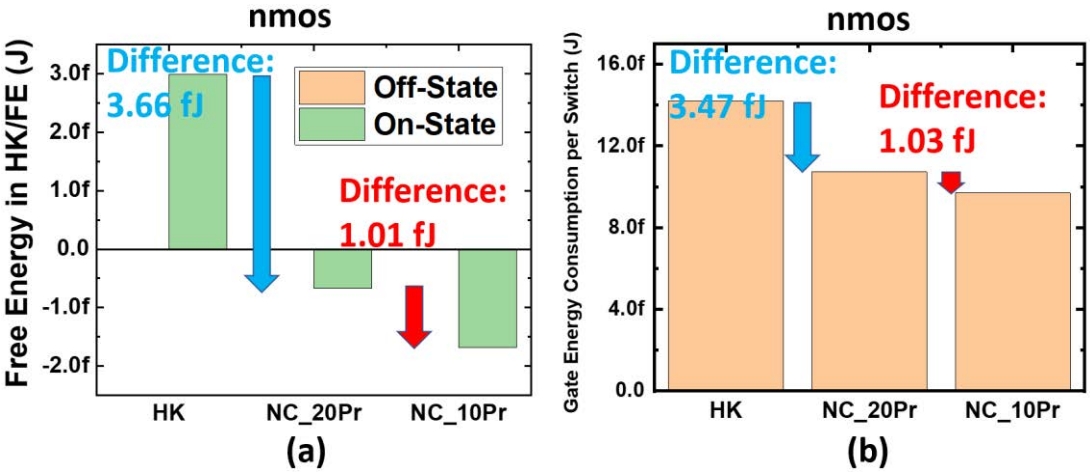


Fig. 5.11. (a) Static free energy of HK/FE in different cases at L_g=2μm. (b) Gate energy consumption of nmos in the inverter transient simulation per switch at L_g=2μm.

short L_g. Also, note that only the nmos (indicated by the green box in Fig. 5.6) energy is calculated because only nmos devices have aligned on-state and off-state gate charges. Remember that matching on-state and off-state gate charges are critical to maintaining similar energies stored in the IL and Si channel.

In a MOSFET structure, the polarization and the electric field change with positions. The middle along the thickness direction (green dot line in Fig. 5.5) is taken to do the static free energy calculations (Eq. 5.2), and the total energy is integrated by the values of meshed points along the L_g. Fig. 5.10 (a) shows the static free energy of HK/FE for the cases of HK, NC_20Pr, and NC_10Pr. The energy difference between off-state and on-state is the energy absorbed or emitted by the HK/FE when the MOSFET is swept from off-state to on-state. The free energy difference in the nmos HK/FE between the on and off state is 50aJ, -15aJ, and -38aJ for HK, NC_20Pr, and NC_10Pr, respectively. There is a 65aJ free energy saving going from HK to NC_20Pr and a 23aJ energy saving going from NC_20Pr to NC_10Pr. Fig. 5.10 (b) shows the input power to the gate of the nmos using Eq. (5.6). The energy saving going from HK to NC_20Pr is 57aJ, and the energy saving going from NC_20Pr to NC_10Pr is 16aJ. Although the value in Fig. 5.10 (a) and (b) are close, there is still around a 30% mismatch. The major factor which causes the mismatch is the gate-to-source and gate-to-drain parasitic capacitance. Longer gate length can reduce the parasitic effect since the percentage of parasitic capacitance in the total gate capacitance ($\frac{C_{GS}+C_{GD}}{CG}$) reduces with increasing L_g.

To illustrate this point, the same calculation is done at L_g = 2μm. Fig. 5.11 (a) shows the free energy of HK/FE for the cases of HK, NC_20Pr, and NC_10Pr. The difference of free energy stored in the nmos HK/FE between on-state and off-state is 2.99fJ, -0.67fJ, and -1.68fJ for HK, NC_20Pr, and NC_10Pr, respectively. There is a 3.66fJ free energy saving going from HK to NC_20Pr and 1.01fJ energy saving going from NC_20Pr to NC_10Pr. Compared with the input power integration results in Fig. 5.11(b), there is only a ~5% mismatch when comparing the energies from the two approaches: static free energy integration of the materials and transient circuit power integration. The results in Fig. 5.11 clearly show how the energy-charge profile of the FE benefits energy efficiency in the inverter circuit.

Fig. 5.12 shows energy per switch versus delay at L_g=20nm. Solid symbols

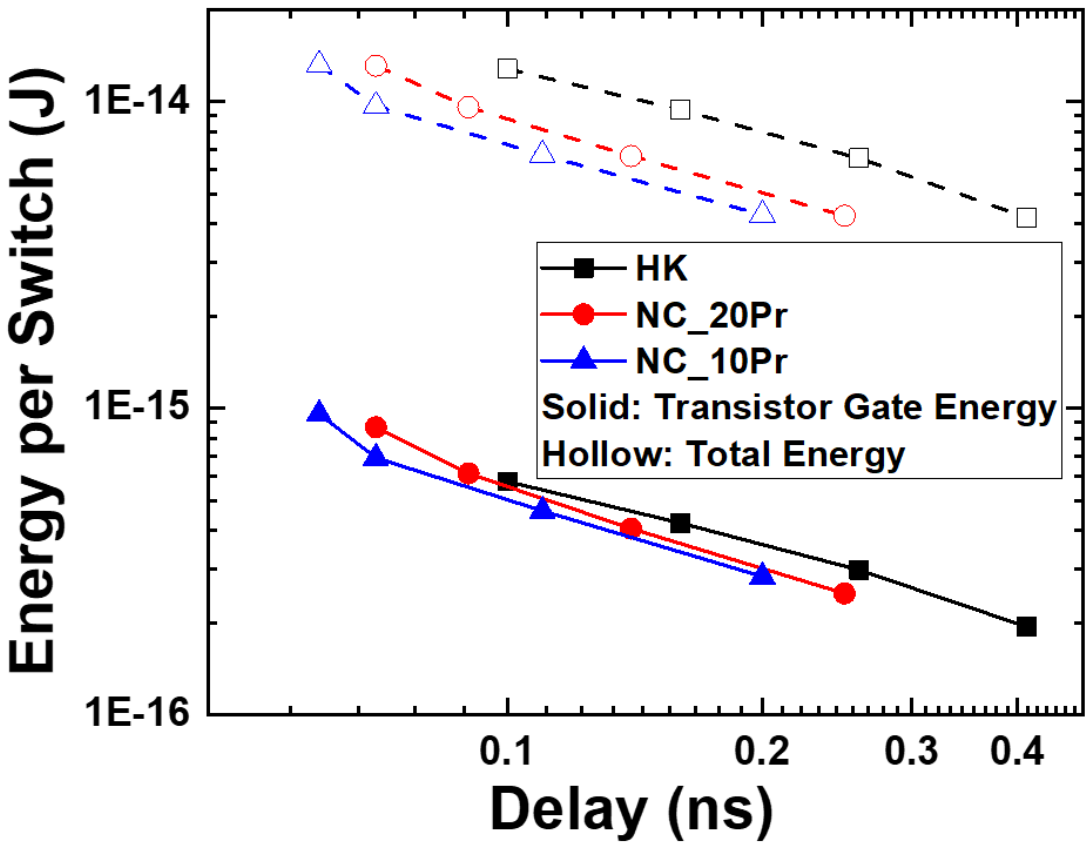


Fig. 5.12. Energy per switch versus delay at L_g=20nm. Solid symbols only consider gate energy consumption of nmos and pmos, and hollow symbols consider the additional energy consumption due to C_{out}. C_{out} is set to be 50fF.

indicate the energy consumed only by the gates of the nmos and pmos. These calculations show that NC benefits energy consumption through energy storage separately from V_{DD} reduction on C_{load}. The unique energy-charge profile of FE reduces the input energy required to achieve the same speed (average current). The hollow symbols in Fig. 5.12 indicate the energy consumed by the gate and C_{out}. The energy-saving provided by NC is even larger here because of the impact of V_{DD} reduction on C_{load}.

5.4 Chapter Summary

FE acts as a rechargeable energy storage medium, which recycles energy when the energy of the whole system decreases, and releases the energy again when the energy of the system increases. This quasi-adiabatic behavior [65]

improves the energy efficiency of the system. Energy analysis has been conducted in both MOSCAP and MOSFET structures. Two energy calculation approaches were investigated: the static materials free energy integral and the transient power integral in the circuit. The numbers calculated by both approaches match in both MOSCAP and MOSFET structures with less than a 1% and 6% error, respectively, showing how the energy saved by a rechargeable energy storage medium can be transferred almost entirely to the system, thereby improving the energy efficiency of the system.

Chapter 6

Negative Capacitance Enables FinFET Scaling Beyond 3nm Node

A comprehensive study of the scaling of negative capacitance FinFET (NC-FinFET) is conducted with TCAD. We show that the NC-FinFET can be scaled to "2.1nm node" and almost "1.5nm node" that comes two nodes after the industry "3nm node," which has 16nm L_g and is the last FinFET node according to the International Roadmap for Devices and Systems (IRDS). In addition, for the intervening nodes, NC-FinFET can meet IRDS I_{on} and I_{off} target at target-beating V_{DD}. The benefits of negative capacitance (NC) include improved subthreshold slope (SS), drain-induced barrier lowering (DIBL), V_t roll-off, transconductance over I_d (G_m/I_d), output conductance over I_d (G_d/I_d), and lower V_{DD}. Further scaling may be achieved by improving capacitance matching between ferroelectric (FE) and dielectric (DE).

6.1 Motivation

The negative capacitance field-effect transistor (NCFET) is a promising technology for near future logic devices [3]. SS, DIBL, V_t roll-off, G_m/I_d, and G_d/I_d of the FinFET can be improved by doping Zr into the HfO_2 high-κ gate dielectric [3,74]. The show stopper to scaling of FinFET, according to IRDS, is the difficulty in reducing fin-thickness (T_{fin}) and reduction of effective oxide thickness (EOT). Our study shows that NC enables FinFET scaling beyond the "3nm node" without requiring further thinning of T_{fin} and gate

stack. The scalability of the metal-ferroelectric-metal-insulator-semiconductor (MFMIS) NCFET (with internal metal) has been discussed in [75]. Nevertheless, the electrical characteristics of the metal-ferroelectric-insulator-semiconductor (MFIS) NCFET (without internal metal) is different from the MFMIS NCFET [38], and an internal metal is not desirable in practical logic devices. In this paper, the scalability of MFIS NC-FinFET will be discussed. Moreover, the parameters of FE used in this paper are extracted from the MFIS capacitor, not from the polarization-electric field (PE) loop of the metal-ferroelectric-metal (MFM) structure. The extracted FE parameters are experimentally available and ready for MFIS NCFET.

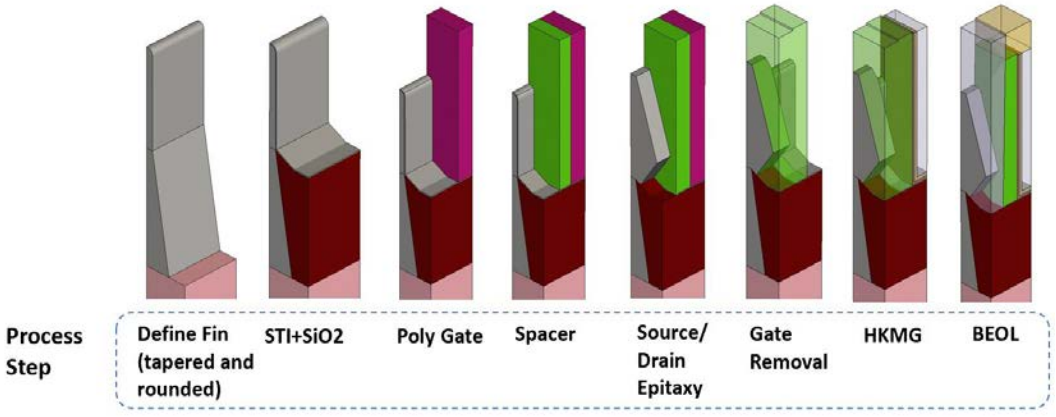
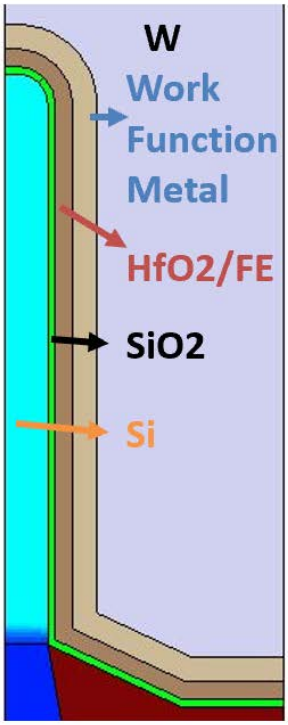


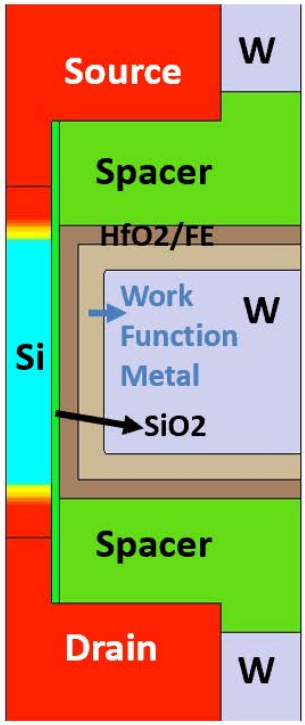
Fig. 6.1. Process simulation flow. Only the source side half of the FinFET is shown.

6.2 TCAD Simulation

Sentaurus Process Simulation [50] is used to build a realistic device for device simulations. The flow of the process simulation is shown in Fig. 6.1. A gate-last process is adopted using a high-κ metal gate. Only the source side of the FinFET is shown for simplicity. Fig. 6.2 (a) and (b) show the cross-section of the fin. After the device structure is built, Sentaurus Device Simulation [53] is utilized to simulate the electrical characteristics. The electronic band structure with stress effects is included with carrier trajectory and scattering calculations and the k.p. deformation potential model. Scattering mechanisms such as phonon scattering, impurity scattering, surface



(a)



(b)

Node Labeling (nm)	"5"
Year of Production	2020
Gate Length (nm)	18
Fin Height (nm)	46
Fin Thickness (nm)	7
IL Thickness (nm)	0.5
HK/FE Thickness (nm)	1.5
Cch (fF/um) - HP	0.45

(c)

Fig. 6.2. (a) vertical and (b) horizontal cross sections of the half-FinFET (c) key geometry parameters of the simulated device. Note that the capacitance of the gate stack (C_{ch}) matches with IRDS high performance (HP) requirement, 0.45fF/μm.

roughness scattering, remote Coulomb scattering, and impact ionization are also included. SRH and Auger recombination are additionally considered. Finally, drift-diffusion with the quantum confinement effect is solved self-consistently with the Sentaurus Device simulator.

An n-type FinFET with fin-height (H_{fin}) of 46nm, T_{fin} of 7nm, and gate length (L_g) of 18nm is calibrated to the "Intel 10nm node" (equivalent to "5nm node" using the IRDS node definition) experimental data [54,76] as shown in Fig. 6.3. Note that the definition of *per-foot-print* (drain current (I_d) normalized to the fin pitch) is used only in Fig. 6.3 to match the Intel presentation [54]. In contrast, all other figures and tables on I_d in this paper present I_d *per-channel-width* (I_d normalized to the sum of 2 times H_{fin} plus T_{fin}) because this work is not concerned with fin pitch. After this calibration, all TCAD parameters, including contact resistivity, are fixed, except the change of fin width from 7nm at "5nm node" to 6nm at "3nm node" according to the IRDS definition [76]. The gate stack is composed of a 0.5nm SiO_2 interfacial

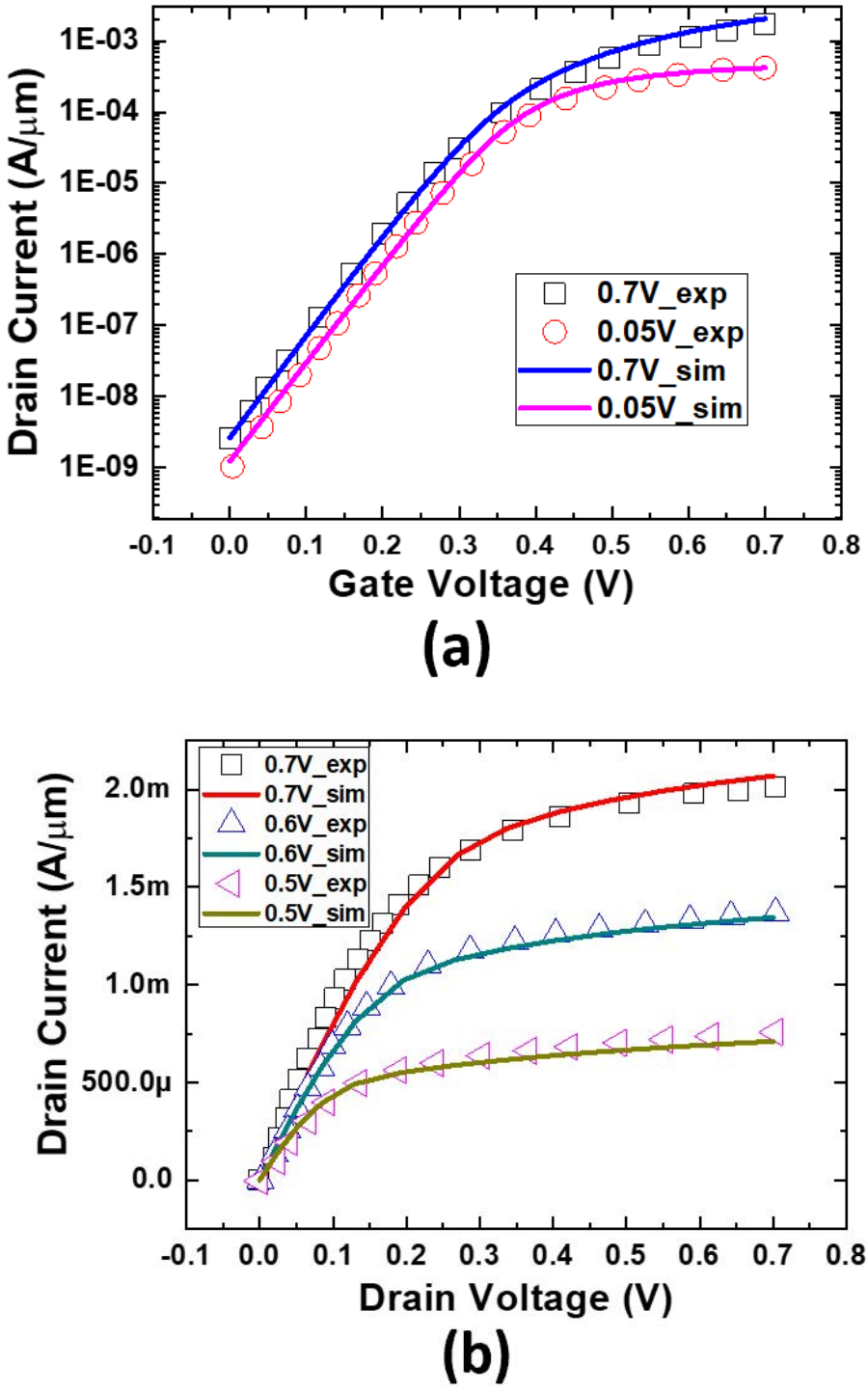


Fig. 6.3. TCAD FinFET calibration to 18nm L_g Intel experimental data. *Per-foot-print* normalization of I_d is done in this figure to match the Intel presentation. (a) I_d-V_g fitting plot at high and low drain voltage (b) I_d-V_d fitting plot at three different gate voltages, 0.7V, 0.6V, and 0.5V.

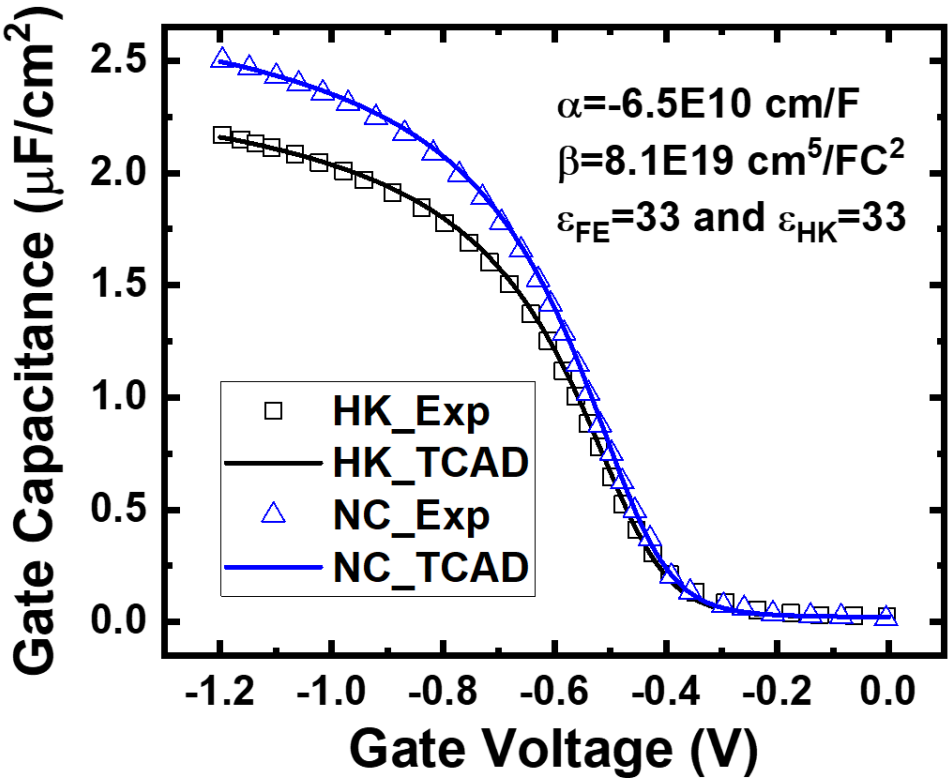


Fig. 6.4. C-V fitting of the experiment data of NC-MOSCAP. The extracted α and βare equivalent to P_r = 20μC/cm^2 and E_c = 1MV/cm.

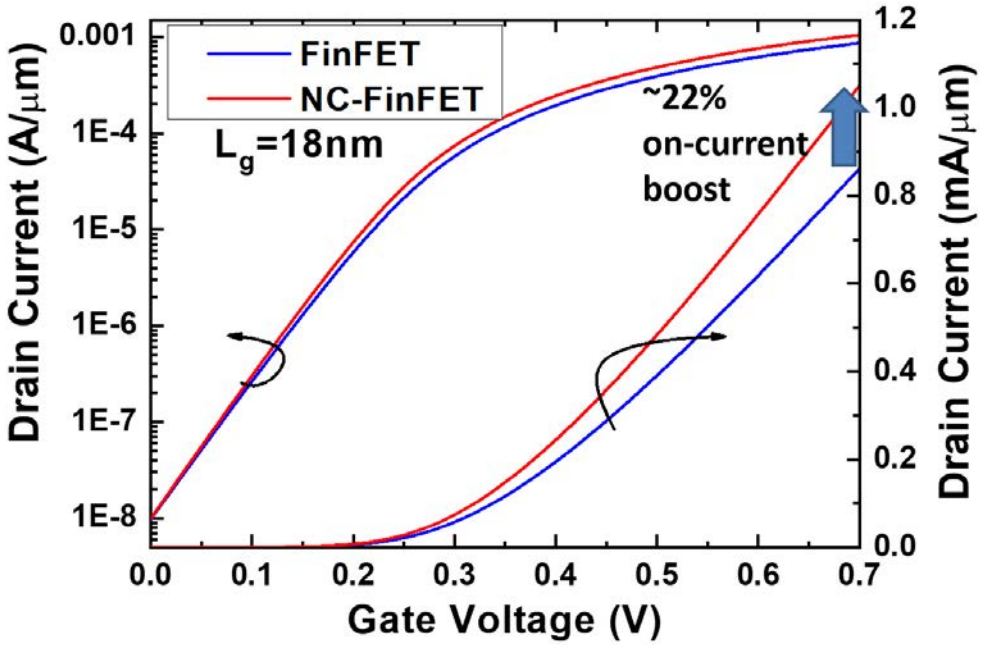


Fig. 6.5. I_d-V_g plot of the FinFET and the NC-FinFET at L_g=18nm.

layer (IL) and either 1.5nm HfO_2 (FinFET) or 1.5nm HZO (NC-FinFET), the latter meaning that HfO_2 is doped with Zr and becomes FE in the NC-FinFET simulation. In NC-FinFET simulation, FE parameters, including α and β from Landau's Equation and background dielectric of FE (ϵ_{FE}), are extracted from experimental C-V of NC MFIS structure [77] (see Fig. 6.4), and strength of the polarization gradient (domain coupling) is set to be 5×10^{-5} cm^3/F (on the same order as [16]). In Fig. 6.4, the gate insulators for HK and NC constitute a chemical oxide (8Å) and 2.8nm layer of HfO_2 or HZO. The extracted dielectric constant of HfO_2 (HK) is 33, which corresponds to 1.1 nm EOT gate stack. On the other hand, a very high dielectric constant (>100) exceeding theoretical predictions for Hf and Zr-based dielectrics [78-82] is required to fit the HZO C-V, and the anomalous I-V behavior in [77] must be explained by the non-linear response of the gate insulator [83]. Therefore, a model with a partially active FE layer in HZO is presented in [83] to explain both the C-V and I-V results in [77]. This work adopts the same methodology to use the Landau-Khalatnikov (L-K) model to extract HZO parameters and fit to Fig. 6.4, with $\alpha = -6.5 \times 10^{10} \frac{cm}{F}$ and $\beta = 8.1 \times 10^{19} \frac{cm^5}{FC^2}$.

6.3 Results and Discussion

Fig. 6.5 demonstrates the improvement from FinFET to NC-FinFET at L_g=18nm, and the boost of I_{on} is about 22% when the I_{off} is aligned at 10nA/μm. Table 6.1 shows the IRDS scaling targets from "5nm" to "1.5 nm". The second row shows the year of production [76]. The third row shows the physical gate length. Physical gate length is predicted to reach the scaling limit of 12nm according to IRDS. T_{fin} (the fourth row of Table 6.1) is set to be 6nm from "3nm" to "1.5nm" according to the IRDS definition. The 5th row in Table 6.1 shows IRDS target V_{DD}, and the 6th row displays the IRDS I_{on} targets at I_{off} = 10nA/μm. The 7th row shows that simulated FinFETs cannot meet the targets after "3nm node," hence the red color. The 8th row shows that NC-FinFETs with extracted FE parameters can meet IRDS requirement one more node beyond "3nm node" and almost meet "1.5nm node" with I_{on} only 3% less than the target I_{on}.

Fig. 6.6 shows I_d-V_g simulation results at V_d = IRDS V_{DD} from "5nm node"

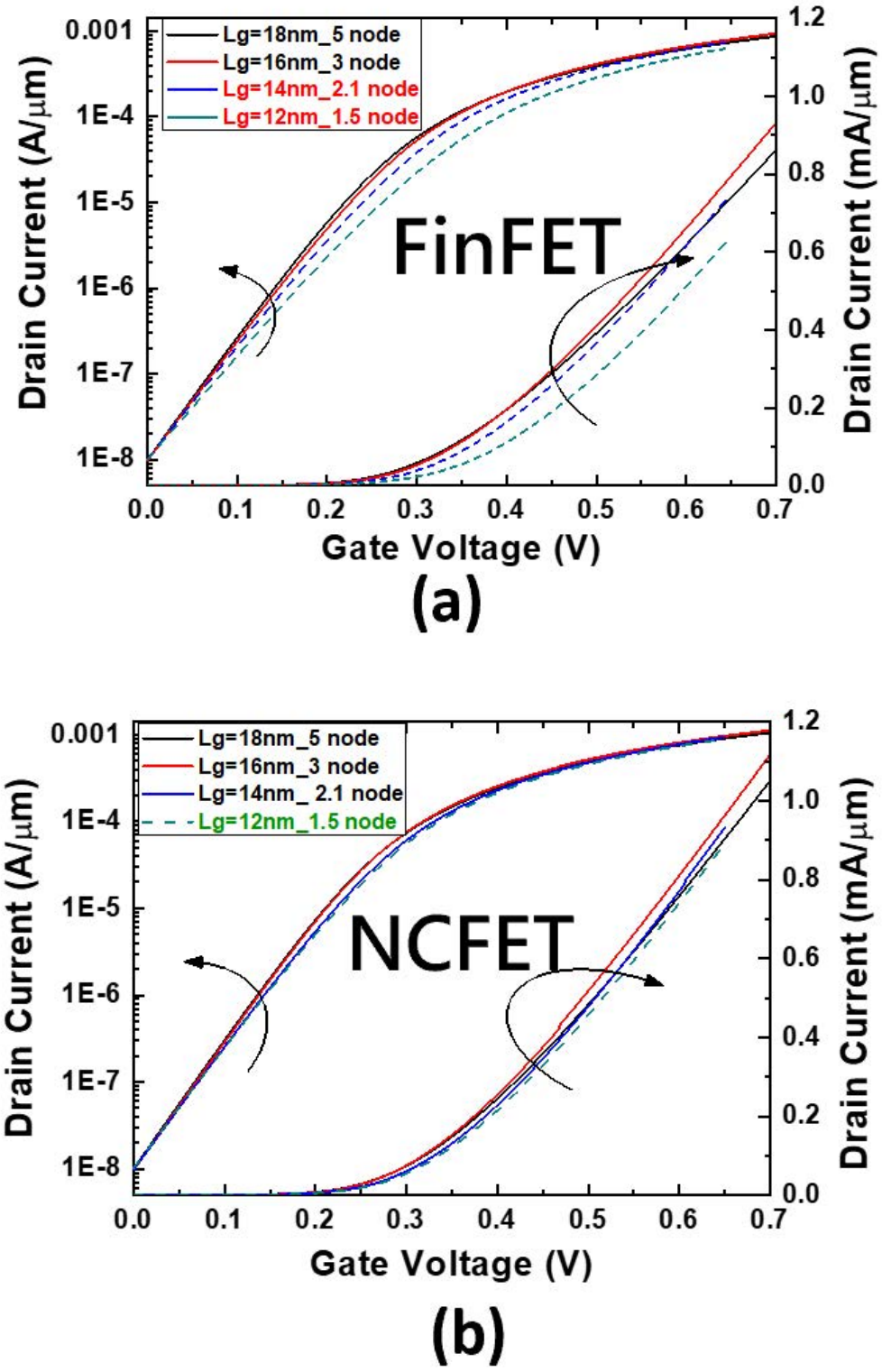


Fig. 6.6. I_d-V_g of (a) FinFETs and (b) NC-FinFETs with work functions shifted to align the off-current with the IRDS high-performance requirement.

1	"2018 IRDS" Node	"5"	"3"	"2.1"	"1.5"
2	Year of Production	2020	2022	2025	2028
3	Physical Gate Length (nm)	18	16	14	12
4	Fin Width (nm)	7	6		
5	IRDS Target VDD (V)	0.70		0.65	
6	Ion Target (mA/μm)	0.85	0.91	0.82	0.92
7	Ion of FinFET(mA/μm)	0.86	0.93	0.75	0.64
8	Ion of NC-FinFET(mA/μm)	1.05	1.12	0.93	0.89

Table 6.1. Simulation plan follows the IRDS 2018 roadmap. Red highlighting of the TCAD results indicate failure to meet the on-current targets at all future nodes.

to "1.5nm node" with work function shifted to align the I_{off} at 10nA/μm. One can see that FinFETs beyond the "3nm node" cannot meet the IRDS targets in Fig. 6.6 (a). The nodes that fail to reach the IRDS targets are labeled red and plotted in a dashed line. NC-FinFET, on the other hand, can meet the IRDS targets at "2.1nm node" and almost complete the IRDS target at "1.5nm node" respectively- two more nodes than FinFET in our simulations. The NC-FinFET simulation results are summarized in the 8th row of Table 6.1. For several nodes, I_{on} is significantly larger than the targets.

SS versus L_g is shown in Fig. 6.7. SS degrades for both the NC-FinFET and FinFET when L_g decreases. Still, the SS degrades at a lower rate for the NC-FinFET because the inner-fringing field, which becomes stronger at shorter L_g, helps capacitance matching and enhances V_g-amplification. Note that even if the SS of the NCFET is not below 60mV/dec in the "weak NC" FinFET studied here, the NC effect improves the on/off ratio improvement is large enough to enable two more nodes of scaling than simple FinFETs. Fig. 6.8 shows the DIBL versus different gate lengths of FinFETs and NC-FinFETs. NC helps relieve the degradation of DIBL, as it can be seen that the slower rate of increase in DIBL when gate length scales. Note that negative DIBL does not appear in this extracted parameter set because the NC effect is "weak"

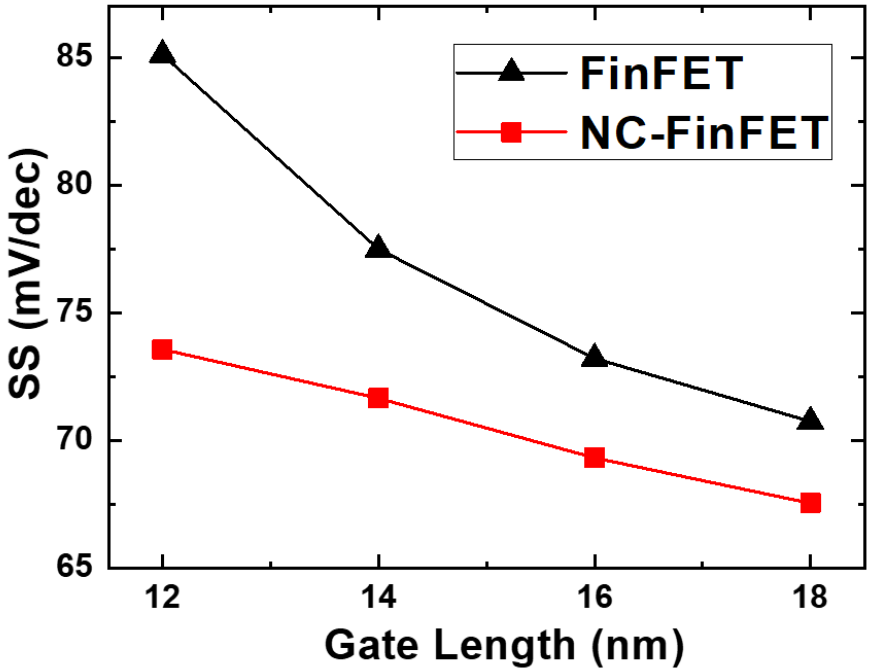


Fig. 6.7. SS of NC-FinFET is smaller than FinFET at 18nm L_g and rises at lower rate with decreasing L_g.

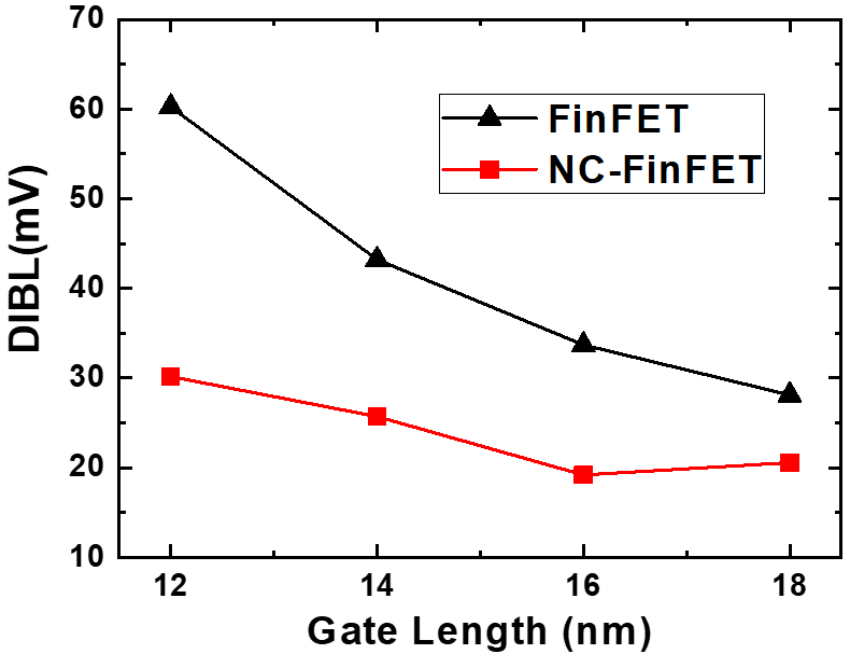


Fig. 6.8. DIBL versus gate lengths. DIBL is smaller in NC-FinFET and rises at lower rate as L_g shrinks.

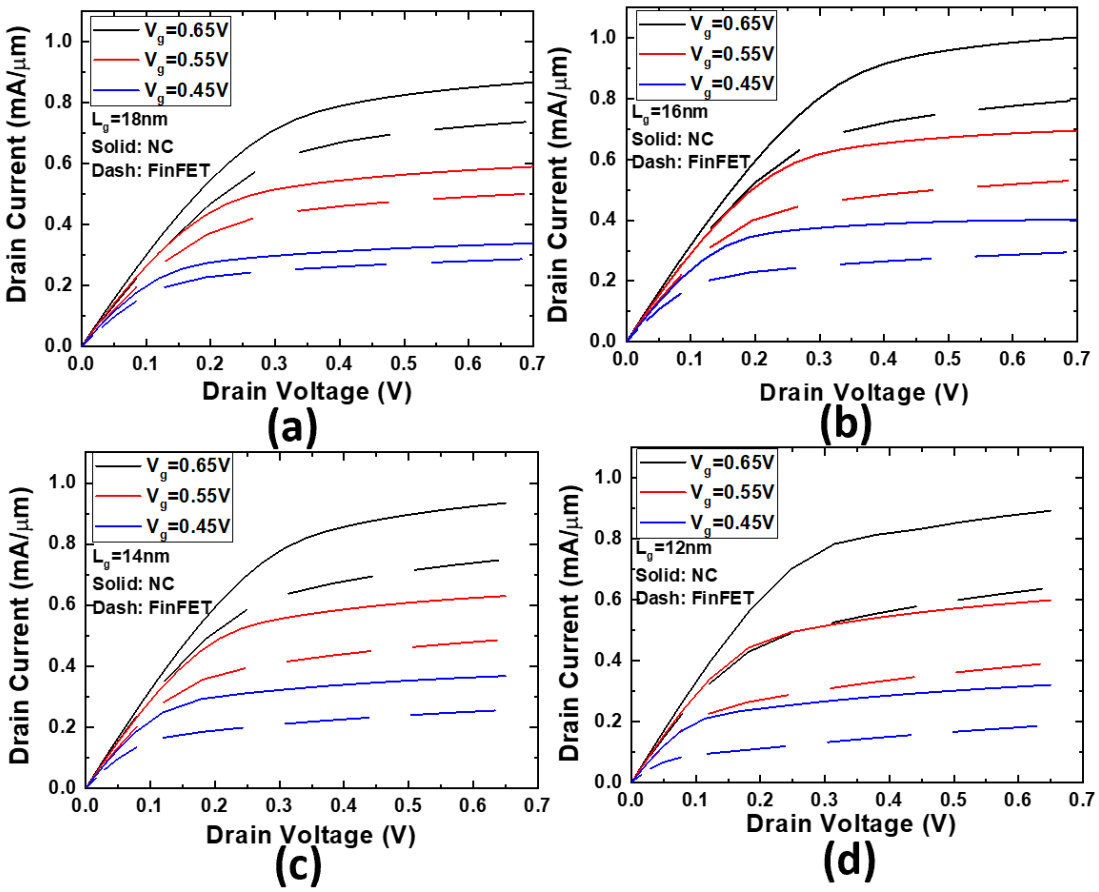


Fig. 6.9. I_d is larger and g_d is smaller in NC-FinFET than in FinFET at (a). L_g=18nm, at (b) L_g=16nm, at (c) L_g=14nm, and at (d) L_g=12nm.

(only 36% of the HZO is active FE layer). The SS_{NC}/SS_{FinFET} and $DIBL_{NC}/DIBL_{FinFET}$ trends are consistent with Y. Liao et al.'s results [84]. Fig. 6.9 compares the I_d-V_d characteristics of the FinFETs and the NC-FinFETs from L_g=18nm to L_g=12nm. In Fig. 6.9, NC-FinFETs are plotted in solid lines, and FinFETs are plotted in dash lines. The current of FinFETs decreases a lot from L_g=18nm to L_g=12nm because of the short channel effect, whereas the current of NC-FinFETs barely decreases from L_g=18nm to L_g=12nm.

The scaled NC-FinFETs are also suitable for analog applications. The g_m/I_d versus I_d of FinFET and NC-FinFET at L_g=18nm, 16nm, 14nm, and 12nm are presented in Fig. 6.10. The g_m/I_d of the NC-FinFET is better than the conventional FinFET overall, and the g_m/I_d of NC-FinFET at L_g=18nm and 16nm nearly hit the theoretical limits of 40 1/V at room temperature [85]. The g_d to drain current ratio is shown in Fig. 6.11. While g_d/I_d increases with shorter channel length for the FinFET due to the short channel effects, the

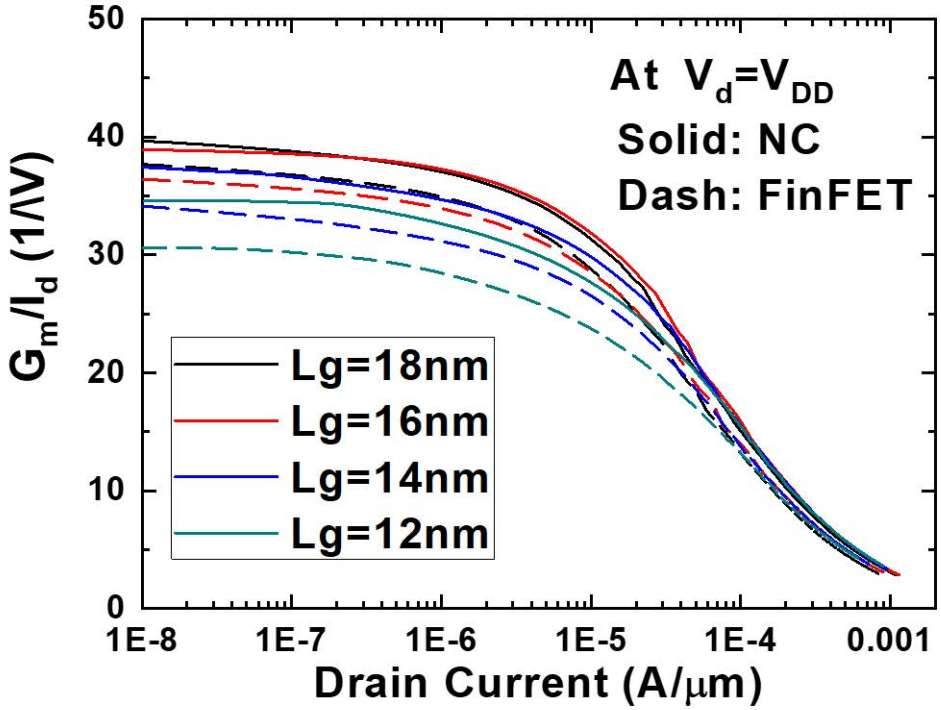


Fig. 6.10. G_m/I_d versus drain current. NC-FinFETs have better G_m/I_d performance than FinFETs overall.

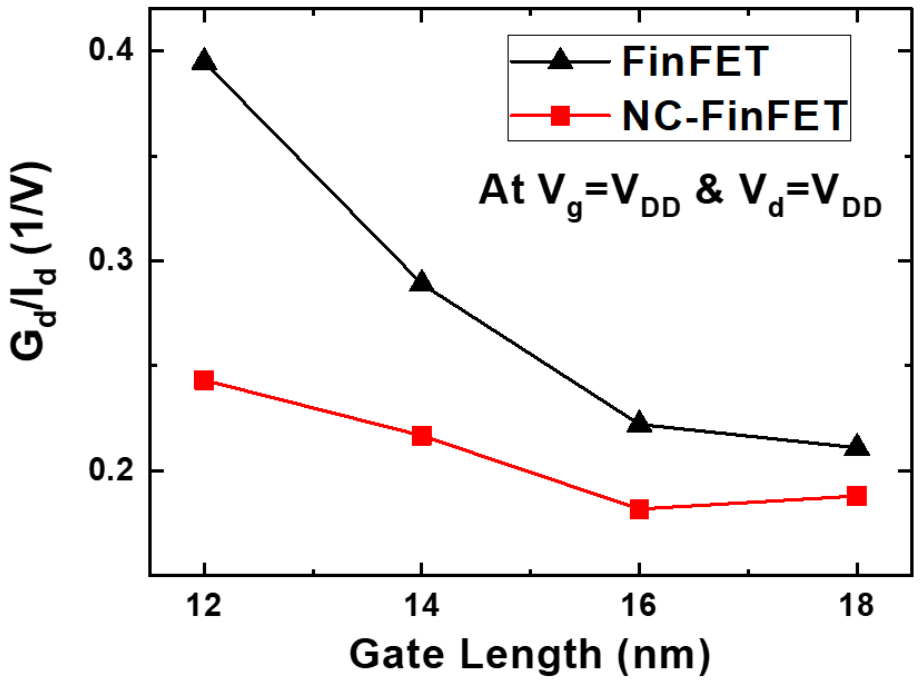


Fig. 6.11. G_d/I_d versus drain current. NC-FinFET has higher Gm/I_d and lower G_d/I_d leading to better analog performance.

"2018 IRDS" Node	"5"	"3"	"2.1"	"1.5"
Year of Production	2020	2022	2025	2028
IRDS Target VDD (V)	0.70		0.65	
FinFET V_{DD} needed to meet IRDS Target Ion	0.70	0.69	X	X
NC –FinFET V_{DD} needed to meet IRDS Target Ion	0.63	0.63	0.61	X

Table 6.2. V_{DD} needed to reach IRDS target on-current (I_{off} is fixed at 10nA/μm) for different IRDS nodes.

trend of g_d/I_d is the opposite for the NC-FinFET from L_g=18nm to L_g=16nm since the FE polarization induced by the inner-fringing field at short L_g, and high V_d negates and overwhelms the short channel effect. For NC-FinFET at L_g=14nm and L_g=12nm, g_d/I_d increases at a much slower rate compared with FinFET. This benefits the intrinsic voltage gain, speed of both the static and pass-transistor logic, and noise margin of logic gates [86].

Table 6.2 shows another way to utilize NC-FinFETs' potential "excess horsepower" in the intervening nodes. The V_{DD} of the NC-FinFET is reduced for each L_g by trial and error until I_{on} and I_{off} at V_d=reduced V_{DD} match the IRDS I_{on} and I_{off} targets, respectively. The reduced (needed) V_{DD} for the scaled NC-FinFET is shown in the last row of Table 6.2. Some IRDS I_{off} and I_{on} targets may be reachable below the target V_{DD} by 70mV at significant power reduction.

Fig. 6.12 (a) shows the conduction band energy along the channel at V_g=0 and V_d=V_{DD}. The black arrow in Fig. 6.12 (a) demonstrates the reduction of the top-of-barrier (TOB) of FinFETs due to gate length scaling. In comparison, the red arrow in Fig. 6.12 (a), which indicates the reduction of the TOB in NC-FinFETs due to gate length scaling, is much shorter than the black arrow. Fig. 6.12 (a) shows that NC-FinFETs have better immunity toward gate length scaling. Fig. 6.12 (b) illustrates the reduced drain-induced barrier lowering effects at L_g=12nm. When the inner-fringing field, more significant at high

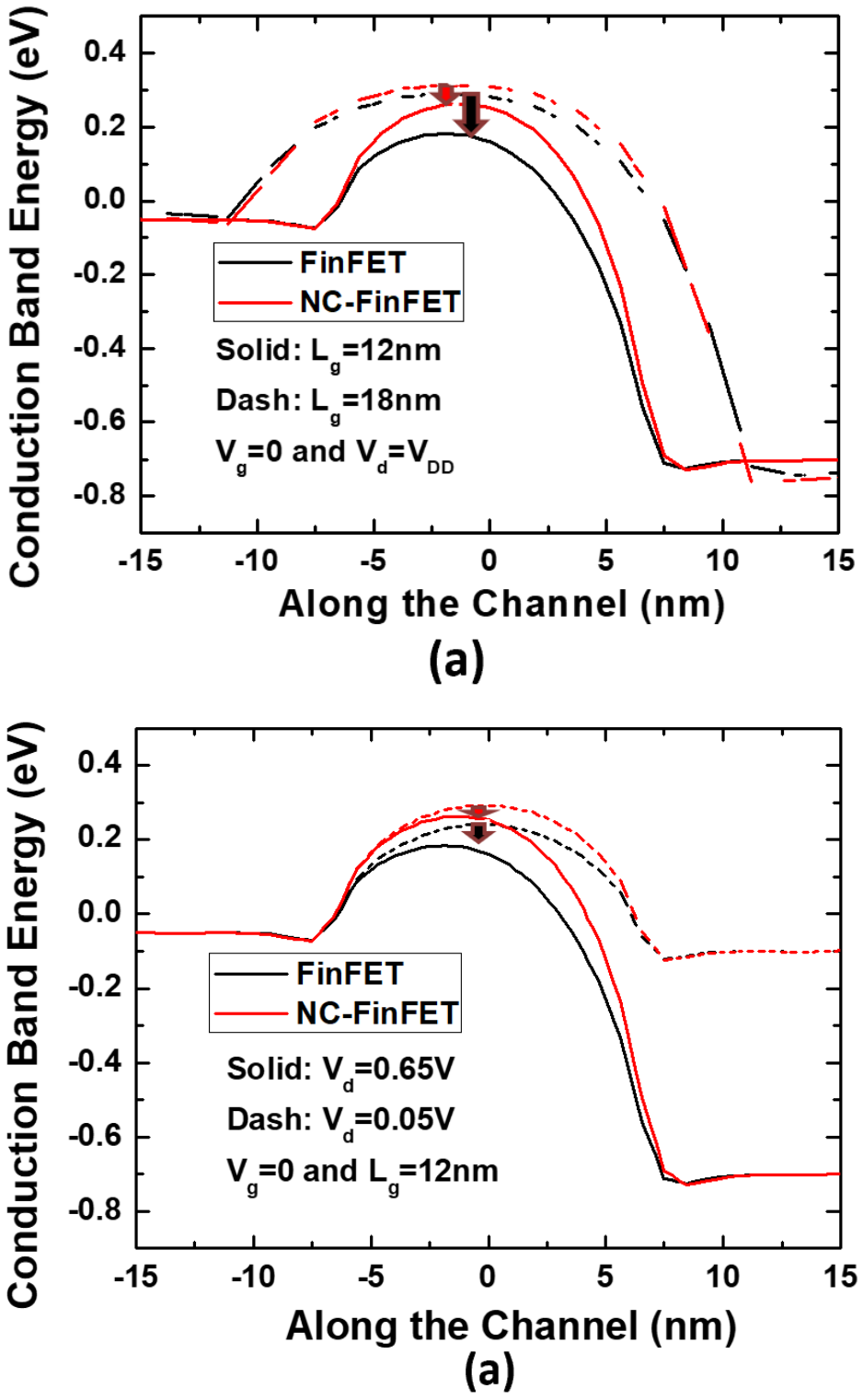


Fig. 6.12. (a) The potential barrier is higher in NC-FinFET (red) than in FinFET (black), and the difference is larger at Lg=12nm (solid) than at Lg=18nm (dash). (b) DIBL of NC-FinFET is smaller than FinFET. Note that the work function is not shifted in this figure.

V_d, goes through the channel to the FE film, it induces polarization in the FE such that the NC-FinFETs' channel potential barrier (the red line in Fig. 6.12 (b)) becomes higher compared with FinFETs' channel potential barrier (the black line in Fig. 6.12 (b)). That is why the reduction of the TOB due to the increase of drain bias in NC-FinFETs (the red arrow in Fig. 6.12 (b)) is smaller than the reduction of the TOB of FinFETs (the black arrow in Fig. 6.12 (b)).

6.4 Chapter Summary

NC may enable FinFET scaling 2 nodes beyond the "3nm node" without requiring thinner W_{fin} or high-k film. We note that this is a TCAD simulation study that assumes the uniform FE film, which can be scaled from large-area NC-MOSCAP to small NCFET device without changing the properties of the FE film and can be put into production. On the other hand, future HZO optimization with a larger portion of active FE layer, multi-layer FEs [45], or varying FE along the channel [87], may lead to even much better NC performance in the future. NC may delay the need for nano-sheet FET in the near term and extend the nanosheet scalability in the long term.

Chapter 7

Negative Capacitance Enables GAA Scaling Beyond 0.5nm Node

We present a TCAD simulation of the negative capacitance gate-all-around (NCGAA) field-effect transistor with the 3-D Ginzburg-Landau-Khalatnikov Model. The baseline device is based on the 2020 IRDS Table, and the mobility model is calibrated to account for ballistic transport and to match the “1.5nm node” IRDS on-current requirement. The NC parameters are extracted from experimental C-V data. The NC-GAA shows a reduction in the off current by one order of magnitude and a 40% on-current boost. If the gate work function is shifted to align the NC-GAA’s off-current with the IRDS high-performance requirement, it is shown that NC-GAA can achieve every node's on-current and VDD requirement through the “0.7eq node,” which is the last node predicted in the 2020 IRDS Table. Furthermore, NC-GAA can even achieve a “0.5eq node,” which is three additional nodes beyond the baseline “1.5nm node.” We also show that these benefits are retained over a varying set of ferroelectric parameters.

7.1 Motivation

The gate length (L_g) of transistors has witnessed continuous scaling [59]. At the nanometer scale, non-ideal effects such as drain-induced barrier lowering (DIBL), subthreshold slope (SS) degradation, random dopant fluctuation, etc., deteriorate device performance [88]. The high-κ (HK) metal gate process [89], a key enabler for the extension of Moore’s Law, has helped reduce effective

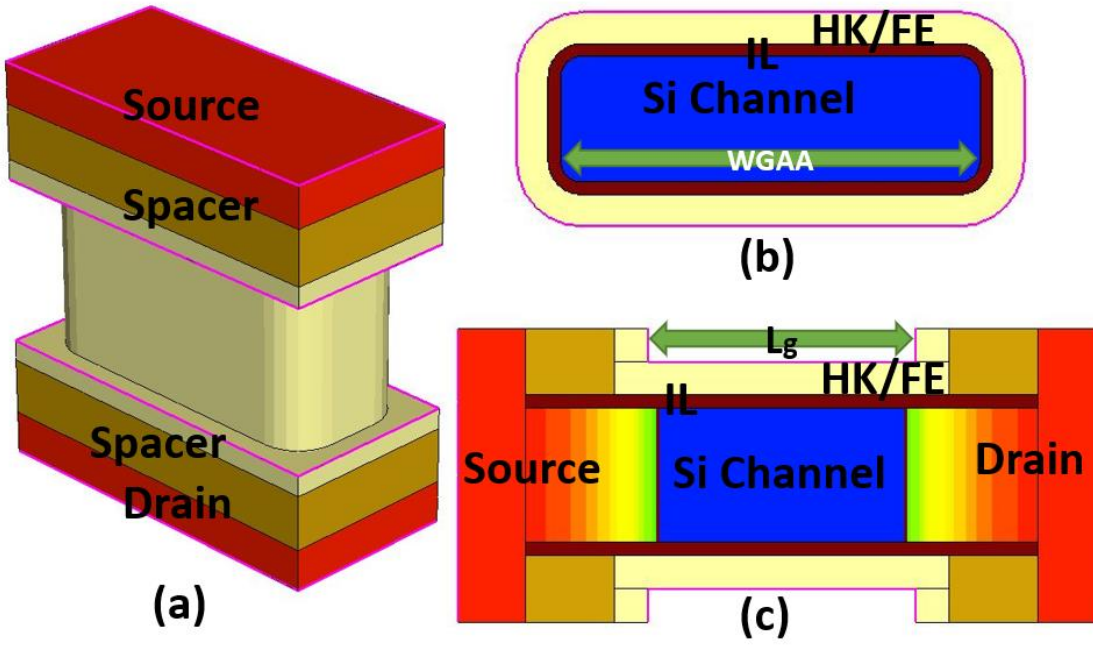


Fig. 7.1. (a) 3-D view of the GAA structure. The metal layer is transparent in these plots to better view the inside structure. (b) Cross section at mid-channel. The corners are rounded to make it closer to real-world devices. (c) Cross section along the channel. The diffusion of dopant from source/drain can be clearly seen.

oxide thickness (EOT) for DIBL and SS improvement, and has also allowed for flexibility of metal work function tuning to resolve the issue of random dopant fluctuation. The FinFET was later proposed to improve the gate-control and transistor density per-foot-print [90]. However, challenges remain when L_g is more aggressively scaled. HK metal gate technologies have reached the end of their scalability limit as further reduction of SiO_2 thickness leads to severe mobility degradation due to remote Coulomb scattering [91].
Negative capacitance (NC) as a solution to extend Moore's Law was proposed for use in the metal-oxide-semiconductor field-effect transistor (MOSFET) by Salahuddin et al. in 2008 [3]. Subsequent experiments on perovskite ferroelectrics (FE) confirmed the existence of NC [4,92]. However, perovskite FEs have scalability and reliability issues when integrating with complementary metal-oxide-semiconductor (CMOS) technologies. Fortunately, many experiments [51-52,66,68-69] reported ferroelectricity in thin films of hafnium oxide (HfO_2) and zirconium oxide (ZrO_2), materials that are already used in the production of advanced CMOS technologies. Thus, the application of NC for state-of-the-art CMOS technologies has become more

Node Labeling (nm)	“1.5”
Year of Production	2028
Gate Length (nm)	12
GAA Width (nm)	20
GAA Thickness (nm)	6
IL Thickness (nm)	0.5
HK/FE Thickness (nm)	1.5

Table 7.1. Key parameters of the simulated GAA device.

practical and feasible.

Two studies on NC’s benefits for FinFETs based on the International Roadmap for Devices and Systems (IRDS) have been accomplished via technology computer-aided design (TCAD) [75,93]. Nevertheless, state-of-the-art technology shifts from FinFETs to gate-all-around (GAA) FETs [94]. Sakib et al. studied metal-ferroelectric-metal-insulator-semiconductor (MFMIS) NC-GAAs via TCAD [95]. However, the simulated structure in the paper has an internal metal gate which is not desirable in practical logic devices, and the electrical characteristics of MFMIS are different from the metal-ferroelectric-insulator-semiconductor (MFIS) NCFET [38]. This paper will present the MFIS NC-GAA using Sentaurus TCAD and the 3-D Ginzburg-Landau-Khalatnikov Equation. The FE parameters are extracted

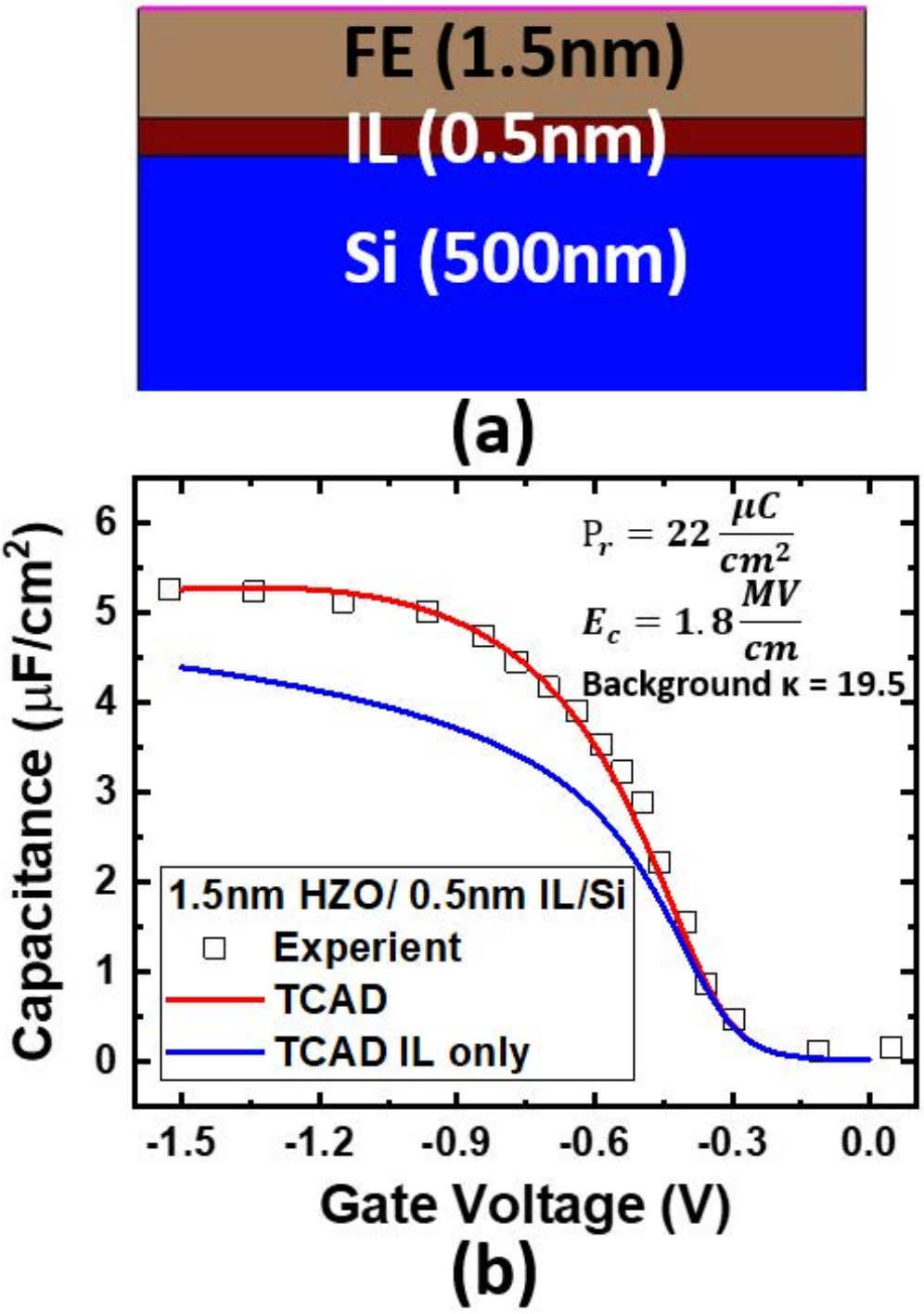


Fig. 7.2. (a) NC gate stack that mimics the device structure in [19]. (b) Capacitance versus gate voltage. The black square symbols are experiment data from [19], the blue curve is with only 0.5 nm IL layer, and the red curve is with 0.5 nm IL layer and 1.5 nm FE layer. The equivalent extracted P_r and E_c are shown on the top-right corner.

from the C-V of a MOS capacitor in published experimental data [96].

7.2 TCAD Simulation Setup

In this section, the assumption of this study, device parameters, and physics models will be explained in detail. A GAA geometry is created using Sentaurus Structure Editor [97]. In Fig. 7.1 (b), the width of the GAA (W_{GAA}) is 20 nm, the thickness of the GAA (T_{GAA}) is 6 nm, the thickness of the interfacial layer (IL) is 0.5 nm, of the HK/FE is 1.5 nm, and the corners are rounded. In Fig. 7.1 (c), the gate length (L_g) is 12 nm, the spacer length (L_{SP}) is 4 nm with κ=3, the doping concentration in the channel is 5×10^{15}, the doping concertation in the source/drain is 2×10^{20} with decay as gaussian distribution into the channel. L_g, W_{GAA}, T_{GAA}, and L_{SP} follow the specifications of the 2020 IRDS "1.5 nm node" [98]. The gate stack of the baseline in the simulation is calibrated to an EOT of 0.9 nm, which is specified in the IRDS table [98].

The physics models used in the simulation include fermi statistics, phonon scattering model, Coulomb scattering, velocity saturation, thin-layer mobility, Shockley-Read-Hall (SRH) recombination, and quantum potential model [53]. The 3-D Ginzburg-Landau-Khalatnikov Equation is additionally solved for the NC-GAA simulations. The on/off current specified by the IRDS "1.5 nm node" is used to calibrate the mobility parameters. The saturation velocity parameter is increased by around 28% to account for ballistic transport and to reach the IRDS "1.5nm node" on-current requirement while keeping all other mobility parameters the same as the Sentaurus default Si-calibrated ones. FE parameters are extracted from a C-V of an NC-MOSCAP with a 1.5 nm thick FE [96]. The 1.5 nm thickness was confirmed by high-resolution transmission electron microscopy (HE-TEM) in [96]. Direct measurement of the P-E loop of the FE is too leaky, and the characteristics of the FE when sandwiched by metals are not necessarily the same as the characteristics of an FE incorporated in a MOSCAP structure, hence why the FE parameters are extracted from the NC-MOSCAP C-V. The gate stack structure is rebuilt in Sentaurus TCAD according to the HE-TEM image [96] (Fig. 7.2 (a)). In Fig. 7.2 (b), the blue curve is the simulated C-V when there is no 1.5 nm FE layer with only 0.5 nm IL remaining. Still, the capacitance value is smaller than the measurement data

from [96]. After bringing in FE parameters of $P_r = 22\,\frac{\mu C}{cm^2}, E_c = 1.8\,\frac{MV}{cm}$, and background dielectric constant of 19.5, the red curve, which is the simulated C-V of the structure in Fig. 7.2 (a), fits the measurement data well. Equivalently, $\alpha \approx -1.1 \times 10^{11}\,\frac{cm}{F}$ and $\beta \approx 1.1 \times 10^{20}\,\frac{cm^5}{FC^2}$, the strength of the polarization gradient (domain coupling) is set to be $5 \times 10^{-5}\,\frac{cm^3}{F}$ (on the same order as [16]), and the ρviscosity is set to be 0 since there is no transient simulation in this study.

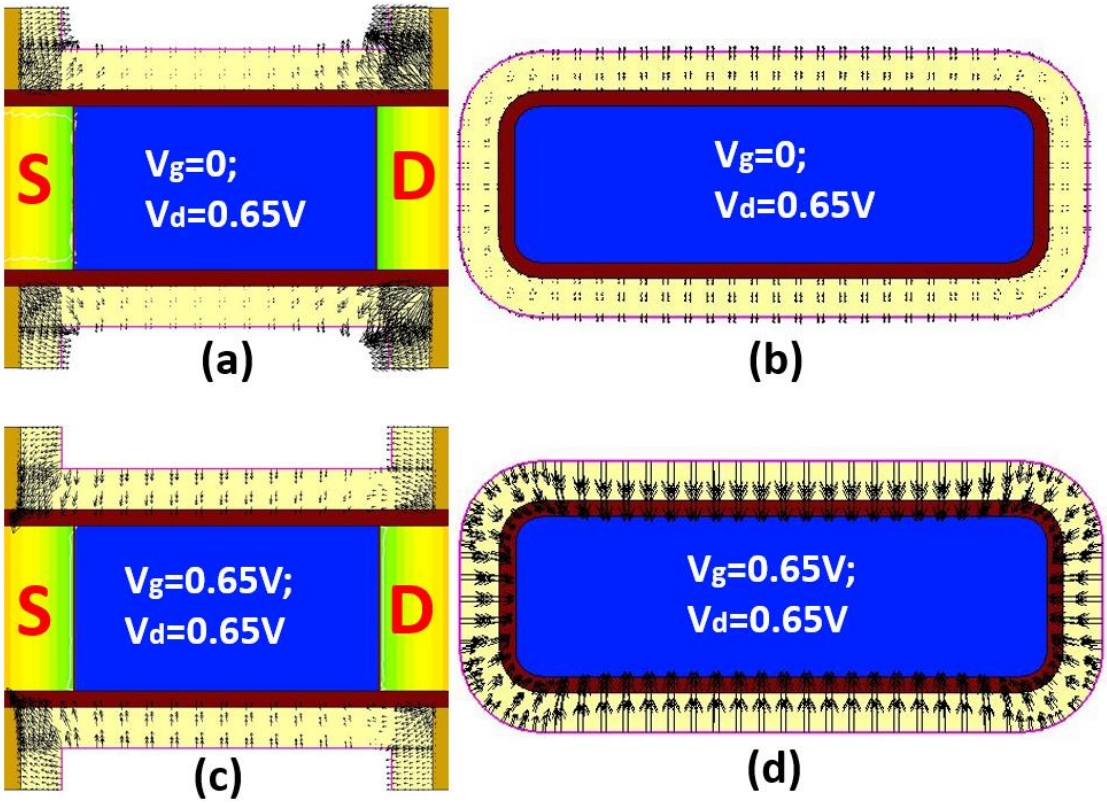


Fig. 7.3. 2-D FE polarization vector plot. (a) Cross section along the channel at $V_g = 0$ and $V_d = 0.65V$. (b) Cross section at middle of the channel at $V_g = 0$ and $V_d = 0.65V$. (c) Cross section along the channel at $V_g = 0.65V$ and $V_d = 0.65V$. (d) Cross section at middle of the channel at $V_g = 0.65V$ and $V_d = 0.65V$.

7.3 Results and Discussion

For the following section, a GAA with an HfO_2 high-κlayer will be called HK-GAA, and a GAA with an HZO ferroelectric layer will be called NC-GAA. Once the HK-GAA baseline is set up and NC parameters extracted, the following analysis and comparison of NC-GAA and HK-GAA can be performed. Fig. 7.3 shows the 2-D FE polarization vector plot at different cross-sections. Fringing field-induced polarization can be seen in Fig. 7.3 (a), and the direction of the polarization vector is from source/drain to gate. In the on-state, where V_g = 0.65V and V_d = 0.65V, the direction of the FE polarization is from gate to source at the source side, while the direction of the FE polarization is from drain to gate at the drain side, as shown in Fig. 7.3 (c). Fig. 7.3 demonstrates that the 3-D Ginzburg-Landau-Khalatnikov Equation used in this simulation work more accurately reflects the behavior of the actual device in comparison to an FET simulation only coupled with

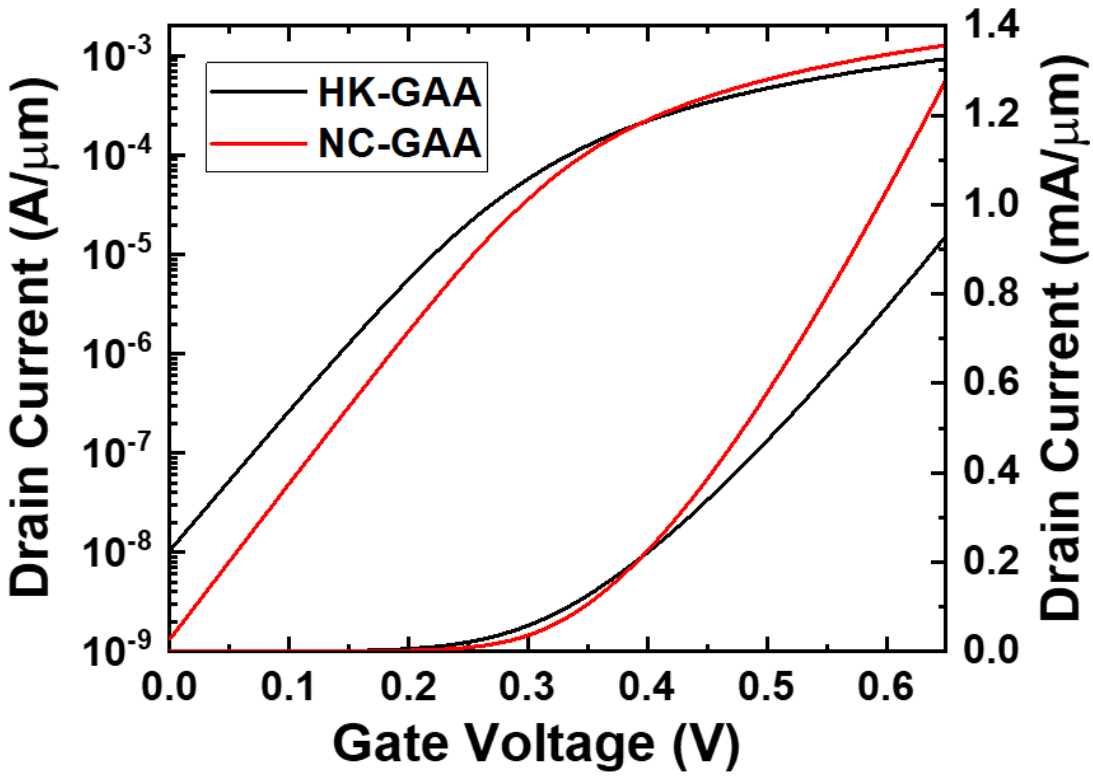


Fig. 7.4 I_d-V_g plot of HK-GAA and NC-GAA before work function shift. NC-GAA has around one order off-current reduction and around 40% on-current boost.

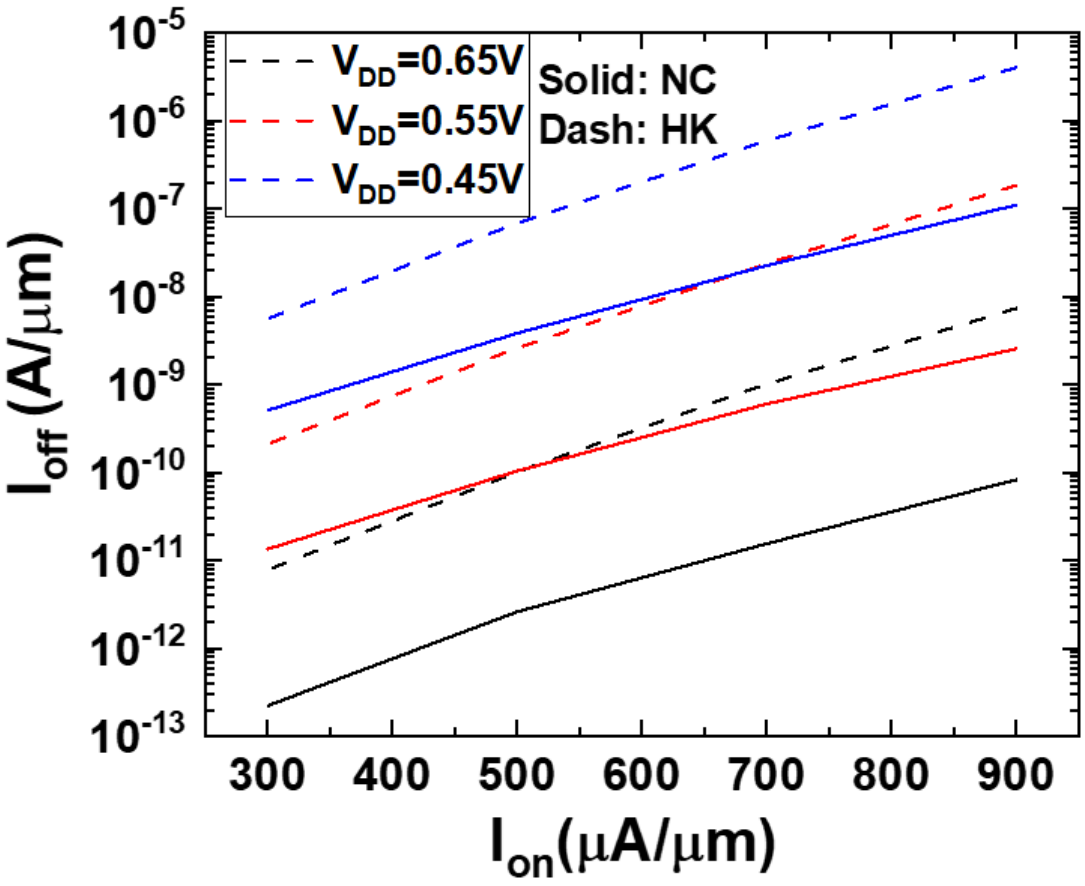


Fig. 7.5. I_{off}-I_{on} plot of HK-GAA (dash lines) and NC-GAA (solid lines) under different V_{DD} conditions. NC-GAA with V_{DD} of 0.55V (solid red line) performs even better than HK-GAA with V_{DD} of 0.65V (dashed black line) in the high-performance region (I_{on} > 550μA/μm).

the simplified 1-D Landau Equation. Fig. 7.4 shows the I_d-V_g plot of the HK-GAA and NC-GAA. With the help of the fringing-field-induced potential, the NC-GAA shows ~1 order off-current reduction, so the barrier height of conduction band energy is higher. NC-GAA also shows about a 40% on-current boost thanks to voltage amplification in the on-state. To compare the resulting on/off-current more comprehensively, I_{on}-I_{off} curves for V_{DD} values of 0.65V, 0.55V, and 0.45V are plotted in Fig. 7.5. V_{DD} can be reduced by more than 0.1V while maintaining the same on-off ratio in the high-performance region. For example, the dashed black curve (HK-GAA at V_{DD} = 0.65V) lies above the solid red curve (NC-GAA at V_{DD}=0.55V) when I_{on} > 550 μA/μm, meaning that the off-current of the NC-GAA is smaller than HK-GAA for the same on-current and at a reduced V_{DD}. For the low-power case

Year of Production	2028	2031	2034	2037
Node Labeling (nm)	"1.5"	"1.0 eq"	"0.7eq"	"0.5 eq"
Target VDD (V)	0.65	0.60	0.55	0.50
Ion (μA/μm) at Ioff=10nmA/μm	924	849	760	760
Required V_{DD} (V) for HK-GAA	0.65	X	X	X
Required V_{DD} (V) for NC-GAA	0.52	0.51	0.49	0.49

Table 7.2 Simulation results compared with IRDS specifications. The row with yellow background is NC-GAA's required V_{DD}, and the column with green background is the additional predicted node which is not on the IRDS Table.

(I_{on} = 521 μA/μm and I_{off} = 100pA/μm), the benefit of V_{DD} reduction is smaller in comparison to the high-performance case, but the benefit is still a 0.1V V_{DD} reduction. The reason why the V_{DD} benefit is larger at higher on-current is due to the fact that NC's benefits are the most substantial in weak to strong inversion, where the capacitance of the IL and Si channel (C_{mos}) is large enough for good capacitance matching. This result differs from the sub-60 mV/dec benefit, most often cited for NCFET devices.

Table 7.2 lists some key numbers from IRDS and this simulation work. The first row is the year of production; the second row is the IRDS node label; the third row is the target V_{DD} given by IRDS; the fourth row is the on-current requirement specified by IRDS; and the fifth and sixth rows are V_{DD} simulation results of HK-GAA and NC-GAA, respectively. The last column shaded in green is our projection, which may differ from the future IRDS specification. HK-GAA can only reach the target V_{DD} at the "1.5nm node" in the fifth row. In contrast, the NC-GAA can beat the target V_{DD}s at every node until our predicted "0.5nm node." The target VDD can be reduced to below 0.5 under the on-current requirement of the IRDS "0.7eq node."

Fig. 7.6 shows the electron density of HK-GAA and NC-GAA, where their on-current is matched with each other by reducing the V_{DD} of the NC-GAA from 0.65V to 0.52V. The channel is plotted from x = -6 nm to x = 6 nm, with x = -5 nm near the source side, x = 0 at middle of the channel, and x = 5 nm

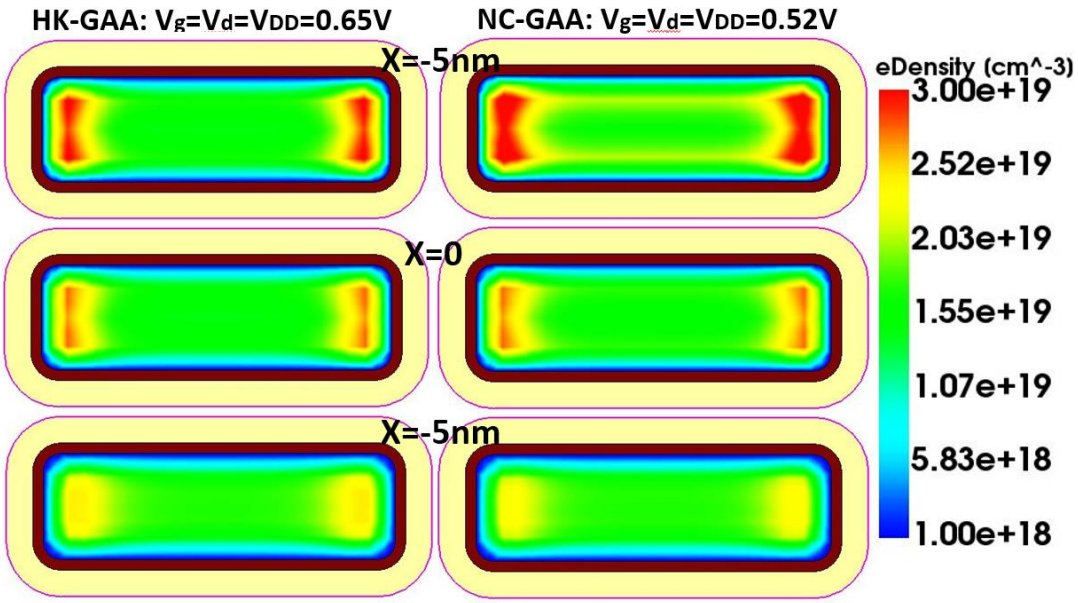


Fig. 7.6. Electron density plot at different cross section planes of HK-GAA and NC-GAA. The V_{DD} of NC-GAA is reduced to match NC-GAA's on-current with HK-GAA's on-current.

near the drain side. At the source side (x = -5 nm), NC-GAA has a higher electron density to compensate for the effect of reduced V_d in order to achieve the same on-current level. However, the electron density is not higher near the drain side (x = 5 nm) due to the effect of voltage reduction by the FE, as explained in Fig. 7.3(c). Fig. 7.6 also shows the corner effect, demonstrating that electron density is higher at the corners. SS versus I_d is shown in Fig. 7.7. The minimum SS of HK-GAA is around 70 mV/dec, the minimum SS of NC-GAA is around 63 mV/dec, and a 7 mV/dec improvement is achieved. NC's benefits are larger in the weak inversion region when C_{mos} is larger to achieve a better capacitance matching condition. Fig. 7.7 shows that a large SS improvement (13 mV/dec) is achieved at the threshold current. Note that although sub-60 mV/dec swing is not achieved in this simulation result, the NC benefits on the GAA structure are still substantial enough to extend the IRDS roadmap by three nodes.

Fig. 7.8 shows the potential at the metal gate, the FE/IL interface, and the Si surface at V_g=0.25V, 0.35V, and 0.45V. Near the source side (x = -6 nm), FE provides voltage reduction (since the dashed black line sits below the solid

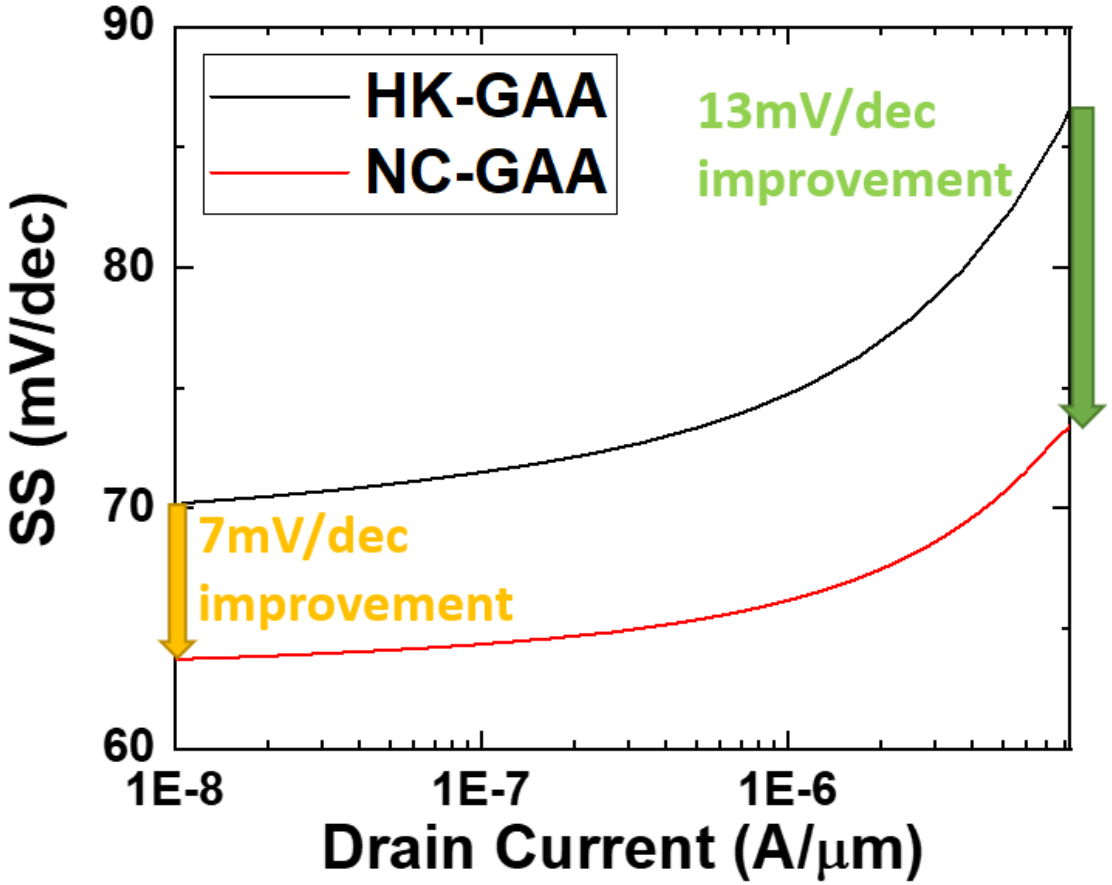


Fig. 7.7. Subthreshold slope versus drain current of HK-GAA (black) and NC-GAA (red).

black line) at V_g = 0.25V, whereas the FE provides voltage amplification (since the dashed blue line sits above the solid blue line) at V_g = 0.45V. This transition happens around the threshold voltage. At V_g = 0.45V, there is voltage reduction at the drain side (x = 6 nm), indicated by the green arrow. When interpreting Fig. 7.8, there are two important values to look at: the voltage provided by the FE and the electric field in the IL. The difference between the solid line and the dashed line is the voltage provided by FE. When the solid line is higher than the dashed line, FE provides voltage reduction and vice versa. The difference between the dashed line and the dotted line is the voltage consumed by IL. When the dashed line is higher than the dotted line, the direction of the electric field is pointing from the gate to the channel and vice versa. Two aspects of the short channel effect are demonstrated in Fig. 7.9. Fig. 7.9 (a) shows a comparison of DIBL between the HK-GAA and NC-GAA. The NC-GAA has a significantly reduced barrier drop (indicated by the

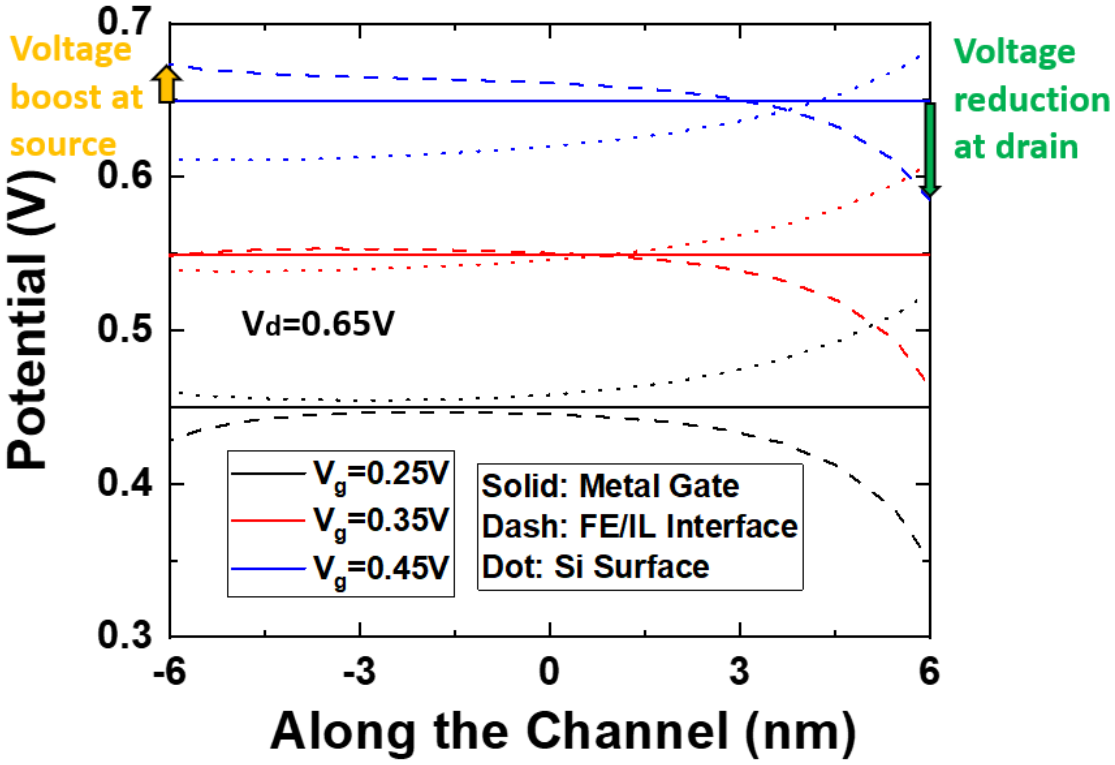


Fig. 7.8. Potential at metal gate, FE/IL interface, and Si surface under different V_g condition along the channel. Noted that V_d is set to be 0.65V in this plot.

red arrow) when V_d increases from 0.05V to 0.65V compared to the HK-GAA (indicated by the black arrow). Fig. 7.9 (b) compares barrier lowering due to L_g reduction between the HK-GAA and the NC-GAA. The NC-GAA also has a much smaller barrier lowering compared with the HK-GAA. Fig. 7.9 demonstrates that NC has better gate control and weaker DIBL. FE polarization points more from the drain to the gate when the drain bias increases. The potential drops more from the gate to the channel (since the polarization and electric field are in the opposite directions in the NC region) such that the barrier height increases. When this effect is stronger than the capacitance coupling from the drain to the top-of-the-barrier, negative DIBL occurs.

Fig. 7.10 shows additional scenarios in which either P_r is increased from 22 to 30 μC/cm^2 or E_c is reduced from 1.8 to 1.0 MV/cm. The on-current boost and the off-current reduction benefits are still substantial, showing the robustness of NC-related benefits against variation in FE parameters.

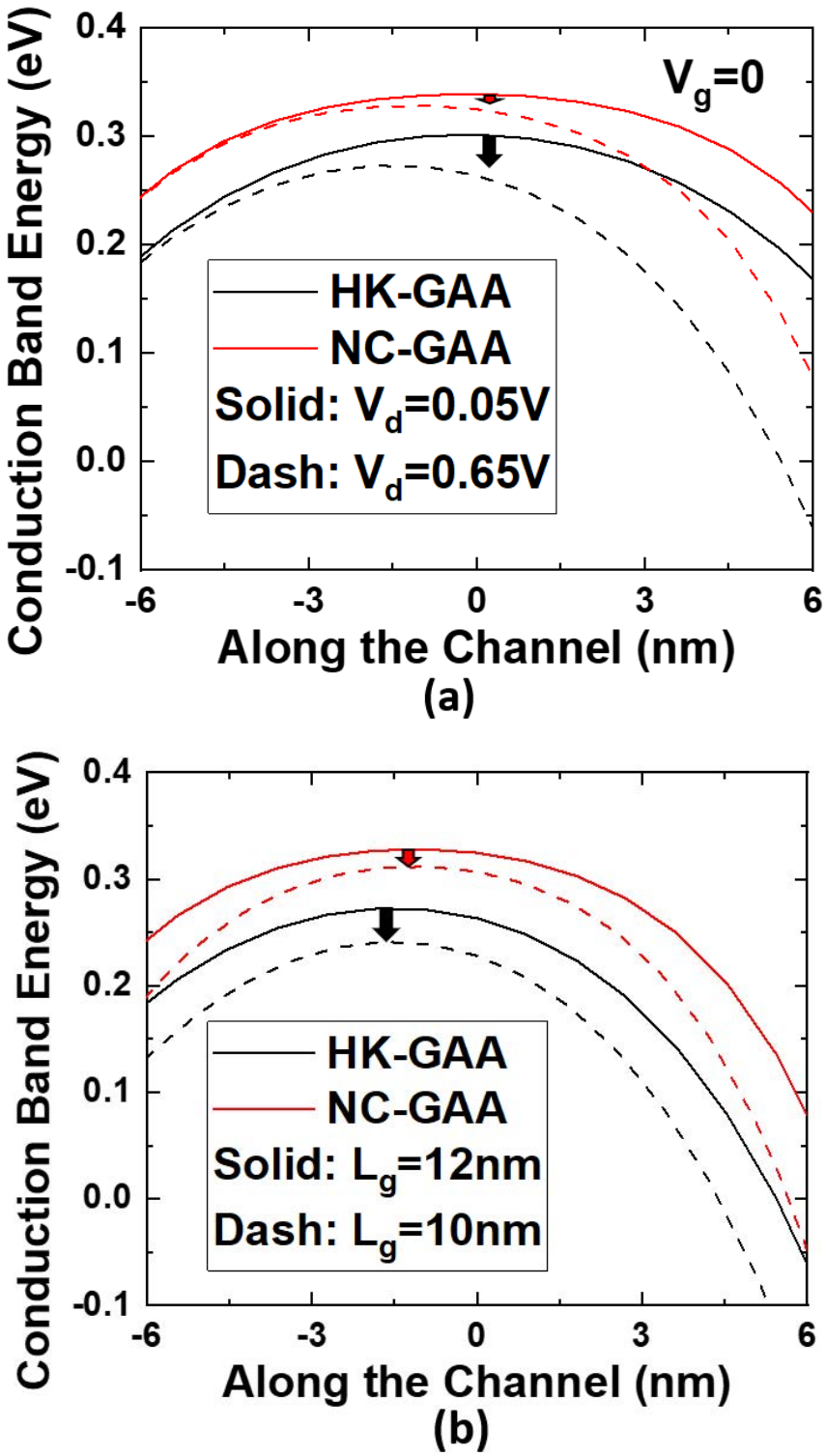


Fig. 7.9. Conduction band energy at V_g=0 along the channel. Comparison of (a) drain-induced barrier lowering and (b) barrier lowering due to L_g reduction between HK-GAA and NC-GAA.

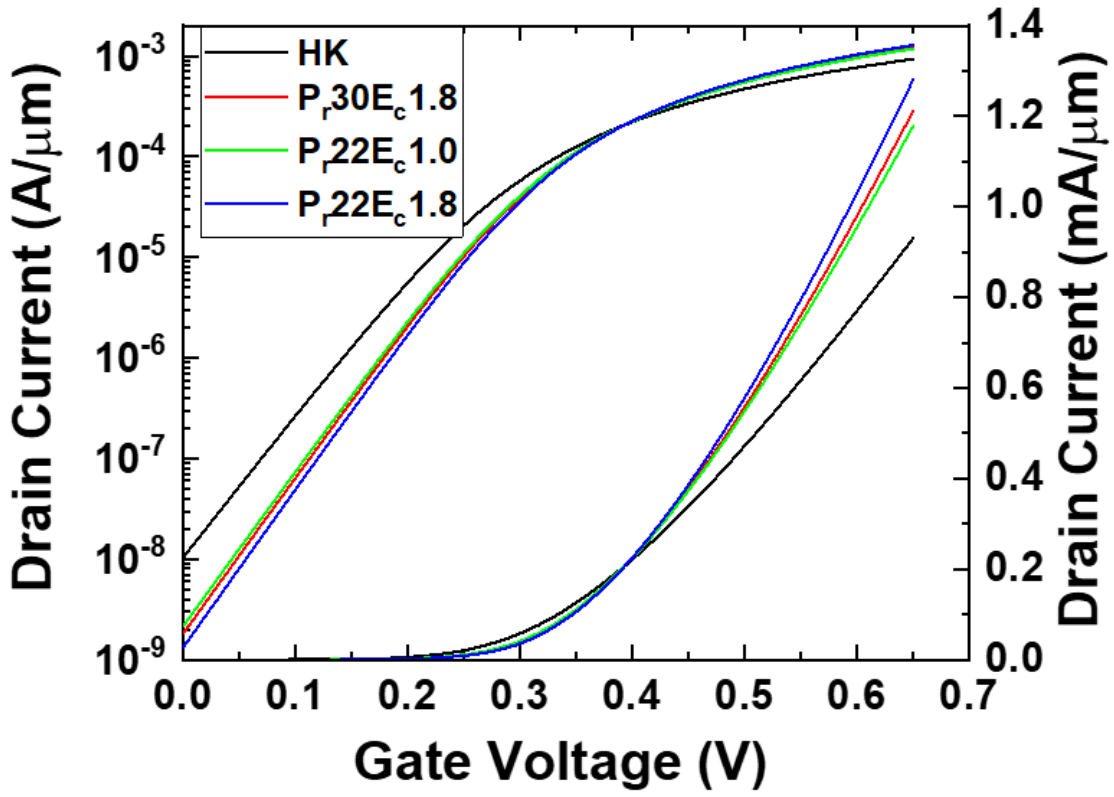


Fig. 7.10. I_d-V_g plot at V_d=0.65V under different NC parameters. The black is HK-GAA baseline, the red is NC of P_r = 30 μC/cm^2 and E_c = 1.8MV/cm, the green is NC of P_r = 22 μC/cm^2 and E_c = 1.0MV/cm, and the blue is NC of P_r = 22 μC/cm^2 and E_c = 1.8MV/cm.

7.4 Chapter Summary

The GAA structure is built based on the 2020 IRDS Table. The NC parameters are extracted from published 1.5 nm HZO experimental MOSCAP C-V data. The direct comparison of NC-GAA and HK-GAA shows one order off-current reduction and 40% on-current boost. If the work function is shifted to align the off-current at 10 nA/μm (IRDS HP standard), the NC-GAA can beat the on-current and VDD requirement of the "0.7eq node" is the last node predicted by the 2020 IRDS Table. Our predicted node, the "0.5eq node," is also achieved by the NC-GAA. Detailed analyses, including electron density, potentials, and short channel effects, are also presented. It is shown that NC has more benefits at weak inversion when C_{mos} is closer to the absolute value

of C_{FE}, which differs from the conventional understanding that a sub-60 mV/dec swing is the signature of the NC effect. Last but not least, the variation sensitivity test is performed, which shows that NC's benefits are not sensitive to variations of NC parameters.

Chapter 8

Compact Model and Benefits of Antiferroelectric in NCFETs

8.1 Motivation

Ferroelectric (FE) and anti-ferroelectric (AFE) have various applications in electron devices, including negative capacitance field-effect transistor (NCFET) [3], ferroelectric random-access memory (FeRAM) [99], ferroelectric FET (FeFET) [100], and ferroelectric tunnel junctions (FTJ) [101-102]. NCFETs can achieve better gate control by stacking an FE layer on top of a dielectric (DE) gate oxide layer. An NCFET is called stabilized when the DE capacitance is larger than the absolute value of the FE capacitance, leading to larger gate capacitance (compared with the gate capacitance without an FE layer) [3]. NCFETs show steeper subthreshold slope (SS), improved drain-induced barrier lowering, larger on-off ratio under the same supply voltage. There are a lot of NCFET re-search going on [20,29,103,104], and many researchers dedicate themself to making this technology realized in the real world.

Many NCFET papers focus on FE-based NCFETs, but few of them discuss the role of AFE in NCFETs [105]. In this paper, a compact model of multilayers of FE and AFE will be proposed first. The model will be later used for proof of concept of benefits of AFE in NCFET gate stack. Finally, the model will be tested under 17-stage ring oscillator to demonstrate its robustness.

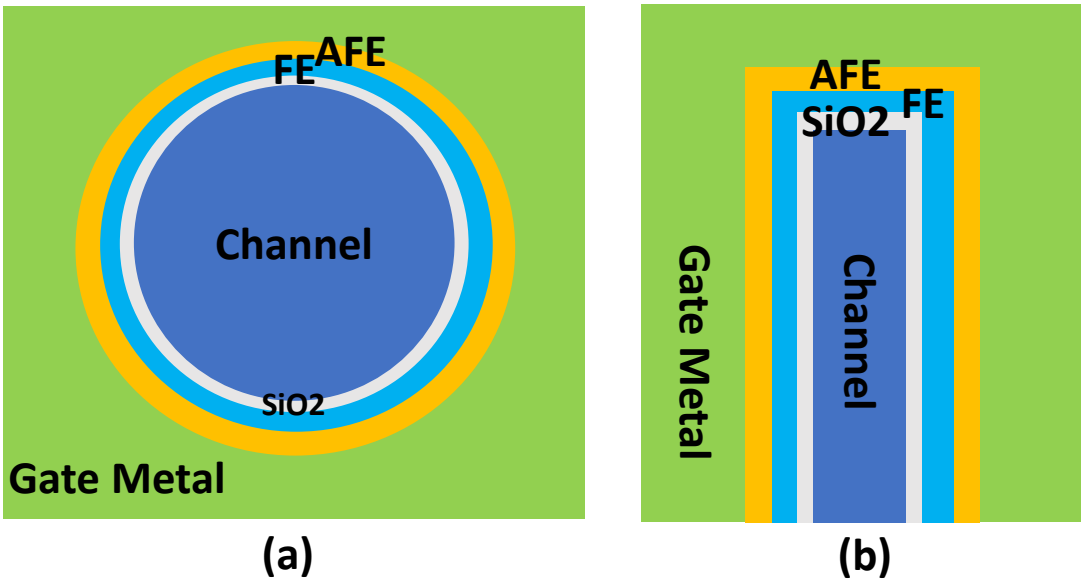


Fig. 8.1. Schematic of the proposed model. The NC model is implemented in BSIM-CMG. BSIM-CMG has a few geometry options, including but not limited to (a) cylindrical gate and (b) FinFET.

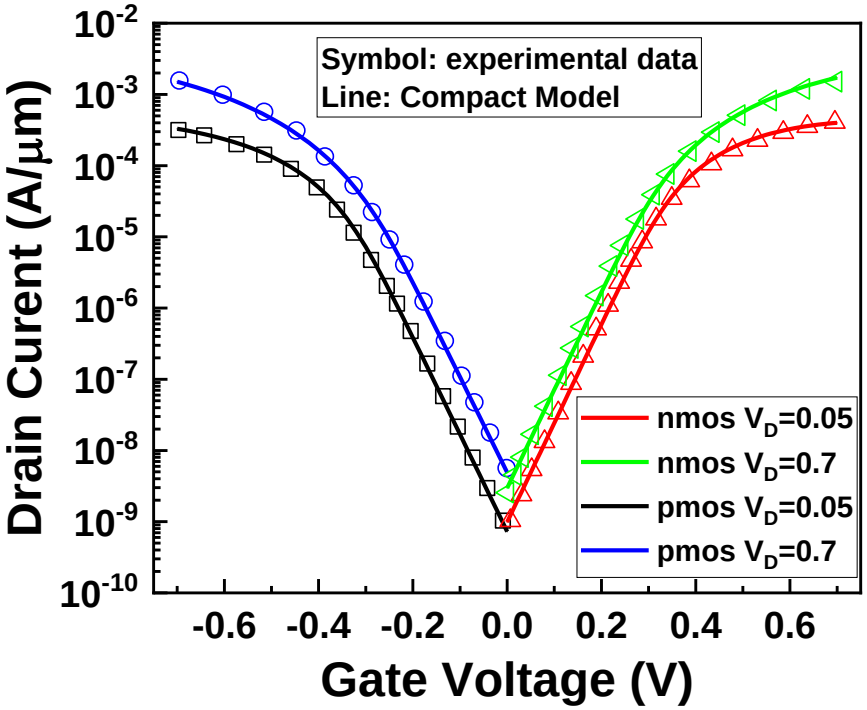


Fig. 8.2. Baseline model fitting with Intel 10nm high performance and low-power measurement data.

8.2 Methods

The NC model is built on Berkeley short-channel IGFET model-common multi-gate (BSIM-CMG), and it has a few options for geometry, including double gate, triple gate, nanosheet, cylindrical gate, and FinFET (schematics of cylindrical gate and FinFET are shown in Fig. 8.1). BSIM-CMG predicts charges and current-voltage characteristics accurately [106]. The model can be divided into two major steps. Firstly, a single unified charge model (UCM), which is a closed form equation between mobile charges and four terminal voltages (VG, VD, VS, VB) as follows [38]:

$$v_G - v_o - v_{ch} = -q_m + \ln(-q_m) + \ln\left(\frac{{q_t}^2}{e^{q_t} - q_t - 1}\right) \quad (8.1)$$

where vG is VG normalized by thermal voltage (kT/q) ,vo and qt are defined in Table 8.1, and vch is channel potential normalized by thermal voltage. In Table 8.1, there are four model parameters, insulator capacitance (Cins), channel area (Ach), channel doping (Nch), and effective channel width (Weff), respectively. This baseline model is tested for 10nm node high performance and low-power FinFET [54], and it models the experimental data accurately as shown in Fig. 8.2.

elationship between electric-field (E) and polarization (P) in FE and AFE are described by Landau-Khalatnikov (L-K) Equation (Eq. 8.2) [107] and Kittel Model (Eq. 8.3) [108-110], respectively:

$$E_{FE} = 2\alpha P + 4\beta P^3 + 6\gamma P^5 \quad (8.2)$$

$$E_{AFE} = \left(2\alpha_p + \alpha_n + 12\beta_{AFE}{P_a}^2\right)P - P_a(\alpha_n + 18\beta_{AFE}P^2) - \beta_{AFE}(2P_a - P)^3 + 7\beta_{AFE}P^3 \quad (8.3)$$

where $\alpha, \beta, and\ \gamma$ in Eq. 8.2 are material parameters of the FE and AFE, $\alpha_p, \alpha_n, \beta_{AFE}, and\ P_a$ are AFE parameters. To solve these equations seamlessly, Eq. (8.2) and (8.3) are further normalized to be Eq. (8.3) and (8.4).

$$v_{FE} = -(a_0 q_{ch} + b_0 {q_{ch}}^3 + c_0 {q_{ch}}^5) \quad (8.4)$$

$$v_{AFE} = -\left(2a_{1p} + \alpha_{1n} + 12b_1{p_{a1}}^2\right)q_{ch} - p_{a1}(a_{1n} + 18b_1{q_{ch}}^2) - b_1(2p_{a1} - q_{ch})^3 + 7b_1{q_{ch}}^3 \quad (8.5)$$

where $a_0, b_0, c_0, a_{1p}, a_{1n}, b_1$, and p_{a1} are listed in Table 8.1. With additional

v_T	$\frac{kT}{q}$ (Thermal voltage)
q_m, q_{dep}	$\frac{Q_m}{v_T C_{ins}}$, $\frac{-qN_{ch}A_{ch}}{v_T C_{ins}}$
v_o	$v_{FB} - q_{dep} - \ln\left(\frac{2q{n_i}^2 A_{ch}}{v_T C_{ins} N_{ch}}\right)$
γ_n	$\frac{A_{ch}C_{ins}}{\epsilon_{ch}{W_{eff}}^2}$
q_t	$(q_m + q_{dep})\gamma_n$
a_0	$\frac{2\alpha t_{FE}(C_{ins})}{W_{eff}}$
b_0	$\frac{4\beta t_{FE}}{v_T}\left(\frac{v_T C_{ins}}{W_{eff}}\right)^3$
c_0	$\frac{6\gamma t_{FE}}{v_T}\left(\frac{v_T C_{ins}}{W_{eff}}\right)^5$
χ_{AFE}	$2\alpha_{AFE} + 16\beta_{AFE}{P_a}^2$
a_{1p}	$\frac{(\alpha_{AFE} + \chi_{AFE}/2)t_{AFE}(C_{ins})}{W_{eff}}$
a_{1n}	$\frac{(\alpha_{AFE} - \chi_{AFE}/2)t_{AFE}(C_{ins})}{W_{eff}}$
b_1	$\frac{4\beta_{AFE} t_{FE}}{v_T}\left(\frac{v_T C_{ins}}{W_{eff}}\right)^3$
p_{a1}	$\frac{-P_a W_{eff}}{v_T C_{ins}}$

Table 8.1: Unified compact model, FE model, and AFE model parameters and variables.

FE and AFE layers Eq. 8.1 should be modified as follow:

$$v_G - v_{FE} - v_{AFE} - v_o - v_{ch} = -q_m + \ln(-q_m) + \ln\left(\frac{{q_t}^2}{e^{q_t - q_t - 1}}\right) \quad (8.6)$$

By solving Eq. 8.4-8.6 iteratively, source side charges and drain side charges can be calculated by plugging in channel voltage = 0 or V_D. Noted that the model can be only applied when the capacitance matching condition (the total differential capacitance is positive) is satisfied [3].

Once the source and drain charges are computed, the current is calculated by integrated inversion charge in the channel in BSIM-CMG [106]:

$$i_{DS} = \int_0^{v_{DS}} q_m dv_{ch} \quad (8.7)$$

where i_{DS} is normalized drain current, q_m is inversion charge, and v_{ch} is

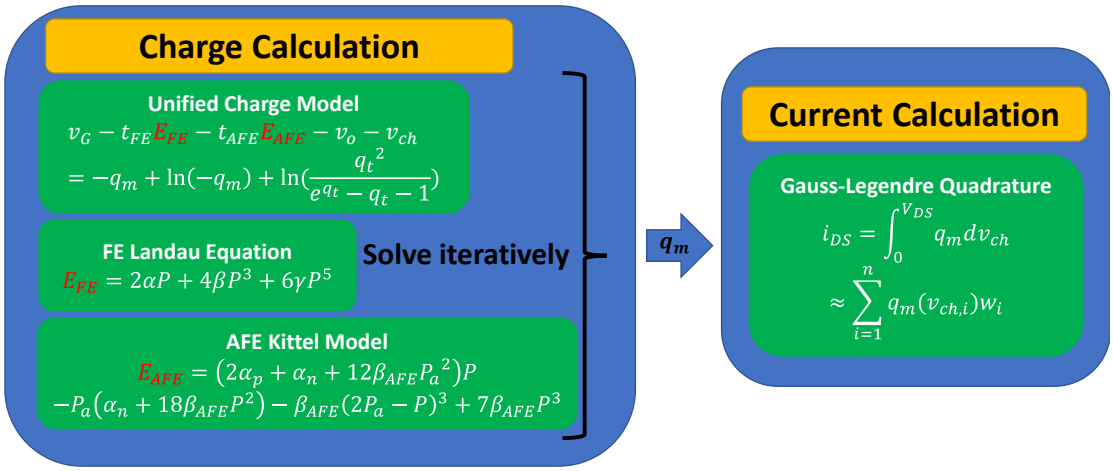


Fig. 8.3. Architecture of the proposed model. Source and drain side channel inversion charges are calcualted first. Drain current are calculated by Gauss-Legendre Quadrature.

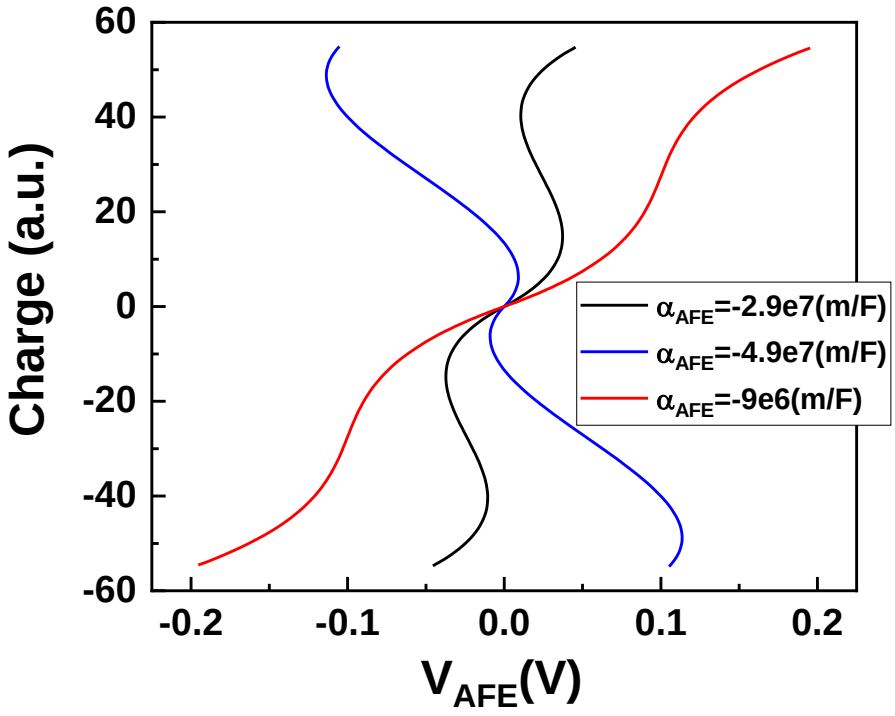


Fig. 8.4. Charge verus voltage relation in AFE. Different α values are simulated, and beta value is fixed.

potential in the channel. Eq. 8.7 can be implemented using Gauss-Legendre quadrature:

$$i_{DS} \approx \sum_{i=1}^{n} q_m(v_{ch,i}) w_i \quad (8.8)$$

where $v_{ch,i}$ is given by $v_{ch,i} = \frac{(v_D - v_S)(x_i + 1)}{2 + v_S}$, n is the number of Gauss sample points in the integration, and x_i *and* w_i are the abscissas and weights of the Gauss-Legendre quadrature [38]. The architecture of the model is illustrated in Fig. 8.3.

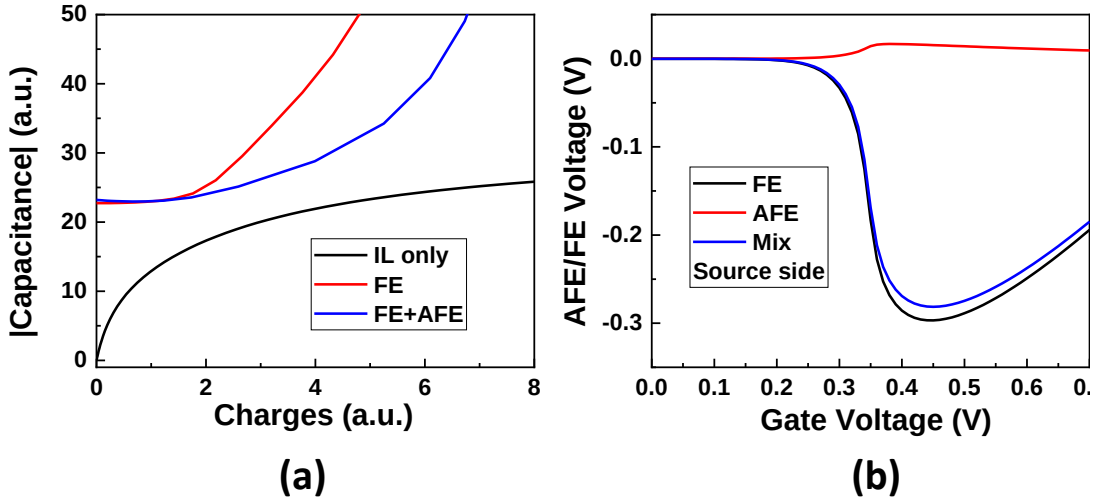


Fig. 8.5. (a) Absolute value of capacitance versus charge. The black curve is gate capacitance with only a 1nm SiO2 layer. The red curve is the capacitance absolute value of a 2nm FE layer with $\alpha_{FE} = -4.33 \times 10^9 \left(\frac{m}{F}\right)$ *and* $\beta_{FE} = 2.4 \times 10^{12} \left(\frac{Coul^2 m^5}{F}\right)$. The blue curve is the capacitance absolute value of a 2nm FE layer and a 0.5nm AFE layer with $\alpha_{FE} = -5.63 \times 10^9 \left(\frac{m}{F}\right)$, $\beta_{FE} = 2.4 \times 10^{12} \left(\frac{Coul^2 m^5}{F}\right)$, $\alpha_{AFE} = -3.1 \times 10^7 \left(\frac{m}{F}\right)$, $\beta_{AFE} = 9.8 \times 10^9 \left(\frac{Coul^2 m^5}{F}\right)$, *and* $P_a = 3 \times 10^{-2} \left(\frac{Coul}{m^2}\right)$. (b) Voltage of AFE or FE versus gate charge. The blue curve is the sum of the black and the red curves.

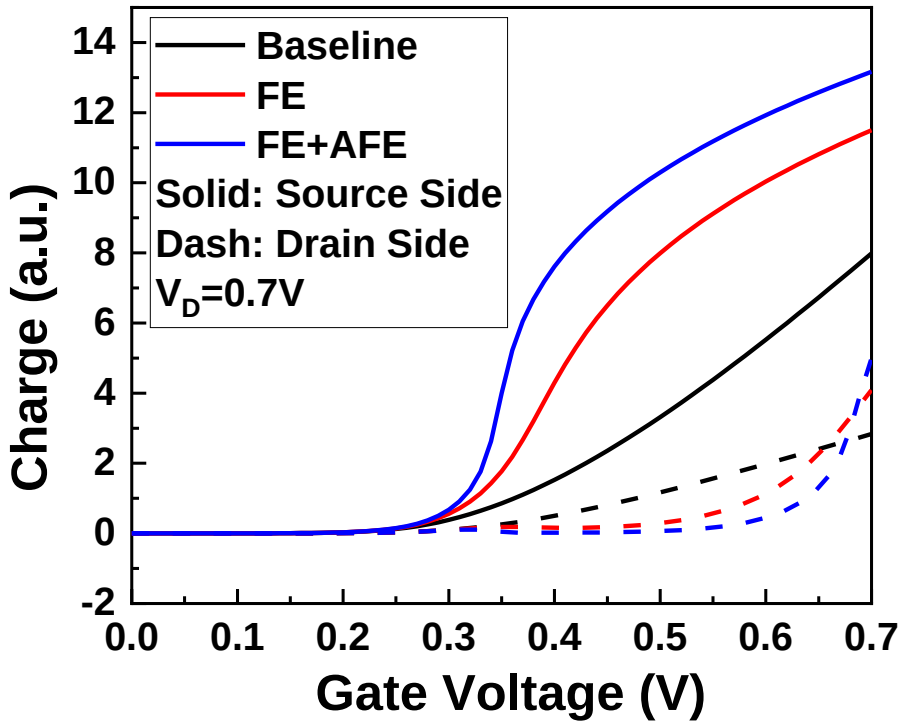


Fig. 8.6. Charge versus gate voltage for baseline, FE, and FE+AFE cases.

8.3 Results

In this section, a few output plots from the proposed model will be demonstrated. In Fig. 8.4, effect of changing α_AFE value is demonstrated. When α_AFE=-9×10^6 m/F (the red curve), there is no negative capacitance region. When α_AFE =-2.9×10^7 m/F (the black curve), negative capacitance region starts to appear. With α_AFE =-4.9×10^7 m/F (the blue curve), negative capacitance region further enlarges. In general, negative capacitance region is larger, and the absolute value of the negative capacitance becomes smaller with more negative the α_AFE.

To better demonstrate the proposed model, an optimized NCFET with an FE layer and an AFE layer is compared with an optimized NCFET with only an FE layer. Capacitance matching condition is shown in Fig. 8.5 (a). Combination of FE and AFE (the blue curve) has better overall capacitance

matching. Fig. 8.5 (b) shows the AFE layer has non-linear capacitance at different charge, so the total capacitance, the blue curve in Fig. 8.5 (a) can be modulated and maintain the same negative capacitance value at higher gate charges. Fig 8.6. shows that the FE+AFE case (the blue) has sharper transition than the FE only case (the red), leading to the improved on/off ratio or lower required V_{DD}.

8.4 Chapter summary

This session demonstrates how AFE Kittel's equation could be implemented as an additional layer in NCFET compact model, including the charge calculation and current integral approximated by Gauss-Legendre Quadrature. In addition, the benefits of additional AFE layers are demonstrated. Additional AFE layer can improve capacitance matching from subthreshold region to strong inversion region.

Chapter 9

Deep Learning-Based BSIM-CMG Parameter Extraction for 10nm FinFET

A new deep learning-based parameter extraction method is presented in this session. 50K training cases are generated by Monte Carlo simulations of these pre-selected parameters in BSIM-CMG. Deep learning models are trained using backward propagation with C_{gg}-V_g and I_d-V_g as the input and selected BSIM-CMG parameters as the output. A TCAD simulated FinFET device, calibrated to Intel 10nm node, is used to test the deep learning models. The deep learning-based parameters extraction results show an excellent fit to capacitance and drain current data, with 0.16% RMS error in C_{gg}-V_g and 6.1% RMS error in I_d-V_g (0.69% RMS error in above-threshold-voltage I_d-V_g), respectively. In addition, devices with a 10% variation in gate length and oxide-thickness are successfully modeled with the trained deep learning model. The results show tremendous promise in using the deep learning-based models for parameters extraction.

9.1 Motivation

Compact models are the fundamental block for integrated circuit simulations. Circuit simulation results would not be accurate and efficient without accurate and fast compact models. A device compact model models the charges and current behavior of a transistor/device given the voltages at terminals. An excellent compact model models the measurement data well and has a short computation time. A compact model cannot be used without the

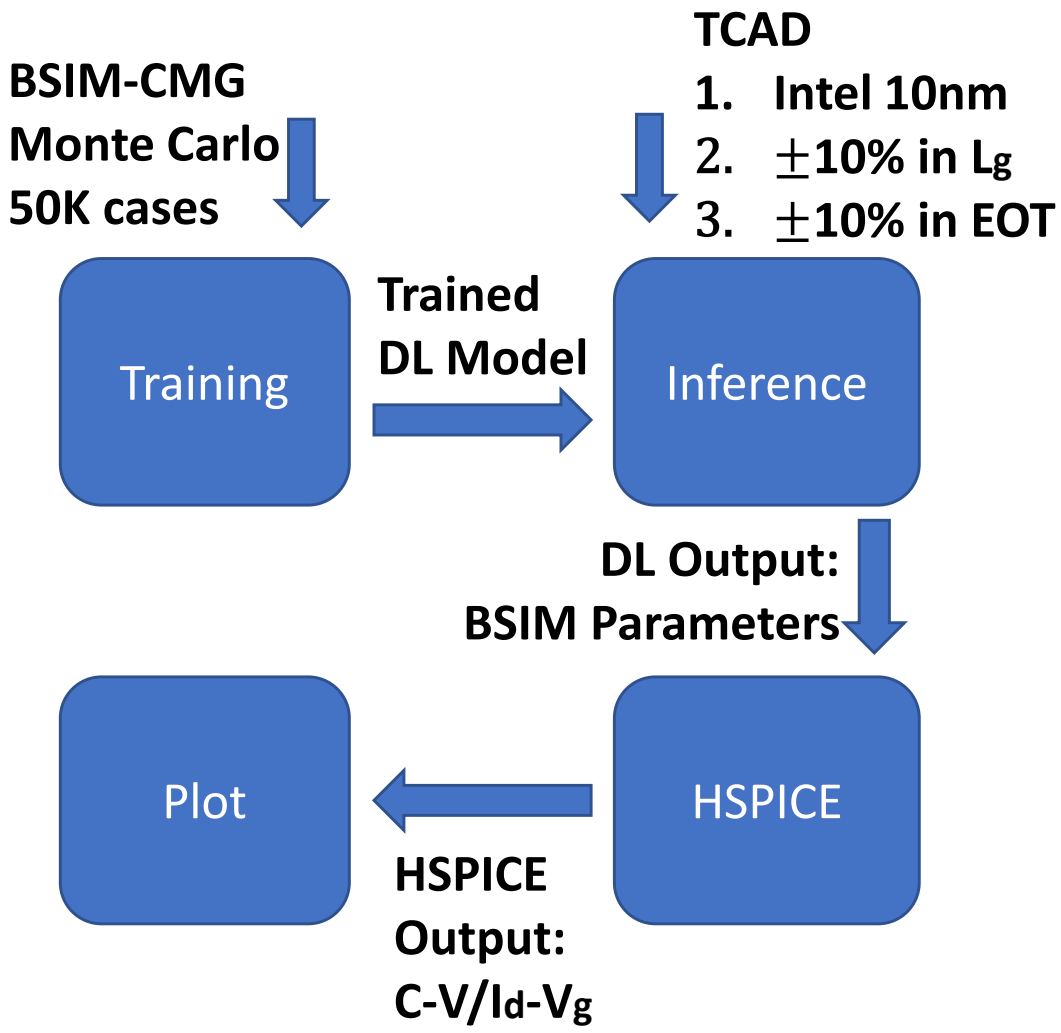


Fig 9.1. The framework of the deep learning-based parameter extraction.

crucial step of parameter extraction. During parameter extraction, the parameters in the compact model will be determined by fitting the compact model simulations to the measurement data of specific device technology.

With the scaling of complementary metal-oxide-semiconductor (CMOS) technology, more and more non-ideal effects such as the short channel effect start to appear [111]-[112]. Nowadays, equations and parameters in compact models have become more complicated to capture the non-ideal effects in state-of-the-art devices. For example, there are more than a thousand parameters in the latest version (111.1.0) of Berkeley Short-channel IGFET (BSIM) Common Multi-Gate (CMG) [113]. Traditionally, device model

engineers extract compact model parameters by tuning parameters manually and checking the output capacitance and current fitting by humans and/or semi-automated extraction programs. The extraction process takes a lot of effort and requires experience.

There are a few machine learning applications in compact models. A few papers use the artificial neural network as an alternative to conventional compact models [114-117]. Alia et al. propose using machine learning to accelerate the genetic algorithm-based parameter extraction [118]. Nevertheless, the total replacement of conventional compact models by machine learning has a long way to go (methods proposed in [114-117]), and using genetic algorithm bases parameter extraction is computationally too expensive (the technique proposed in [118]).

This session will present a deep-learning (DL)-based parameter extraction method that can generate a set of BSIM-CMG parameters within a second. Parameters extraction for C_{gg}-V_g and I_d-V_g will be demonstrated in this paper. The rest of the article follows the flow in Fig. 9.1. It is organized as follows: In section II, the Proposed Method, parameters extraction flow, DL architecture, and how the training data are generated will be explained. In section III, the DL-based parameters extraction results from TCAD data will be shown. Finally, the conclusion will be drawn in section IV.

9.2 Proposed Methods

BSIM-CMG is the industry-standard compact model for FinFET and beyond technologies. Key BSIM-CMG parameters are selected as DL model outputs, listed in Fig. 9.2. In the C-V fit (Fig. 9.2 (a)), PHIG represents work function, CFS represents outer fringing capacitance, TOXP represents physical oxide thickness in the capacitance model, and CGSL represents the overlap capacitance of the gate and source/drain extension area. In the Id-Vg fit (Fig. 9.2 (b)), CIT represents subthreshold slope (SS) degradation, U0 means mobility, UA represents mobility degradation in the strong inversion region, EU can change the transconductance (gm) shape in the strong inversion region, ETA0 models the drain-induced barrier lowering (DIBL) effect, CDSCD models drain bias-dependent sub-threshold slope degradation, VSAT1 models the saturation velocity, and KSATIV is related to drain bias-dependency in current.

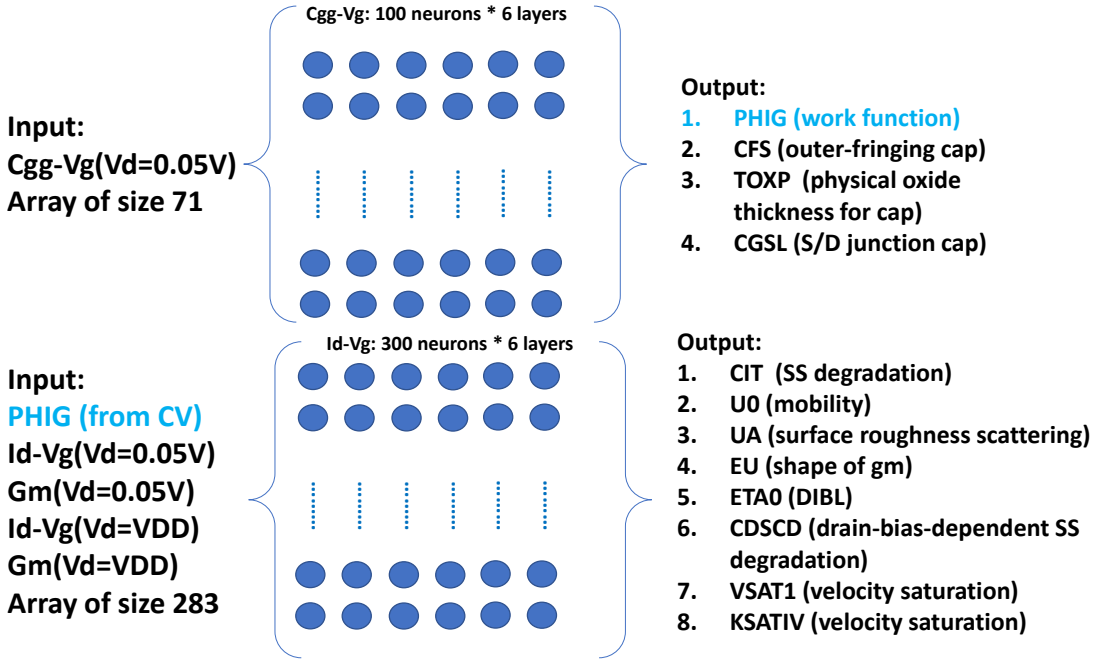


Fig 9.2. The architecture of the deep learning models. (a) C_{gg}-V_g parameters extraction model (b) Id-Vg parameters extraction model.

Fig. 9.2 also shows the architecture of the DL model. Scikit-learn Multi-layer Perceptron package [119] is used. In Fig. 9.2 (a), the DL model for C_{gg}-V_g related parameters extraction consists of six hidden layers with each size of 100 neurons (total parameters number for the C-V DL model is ~57,500). The input is Cgg-Vg data range from 0V to 0.7V with a step size of 0.01V. In Fig. 9.2 (b), the DL model for Id-Vg related parameters extraction comprises six hidden layers with each size of 300 neurons (total parameters number for the I-V DL model is ~537,600). The input is Id-Vg at Vd=0.05V and 0.7V, gm at Vd=0.05V and 0.7V with Vg step size of 0.01V, and the PHIG value from Fig. 9.2 (a) output. ReLU [120] is used as the activation function for the hidden layers in the DL models, and there is no activation function at the output layer. "Adam" solver is used for optimization; batch size, learning rate, and regularization parameters are set to be default values.

The baseline TCAD data is calibrated to Intel 10nm node [54]. A baseline BSIM-CMG model is extracted from the TCAD data manually. From this BSIM-CMG model, 50k of Monte-Carlo simulations are performed to generate training data sets. The Monte Carlo sampling method is chosen because it has better coverage for the nonlinear dataset than pre-assigned-

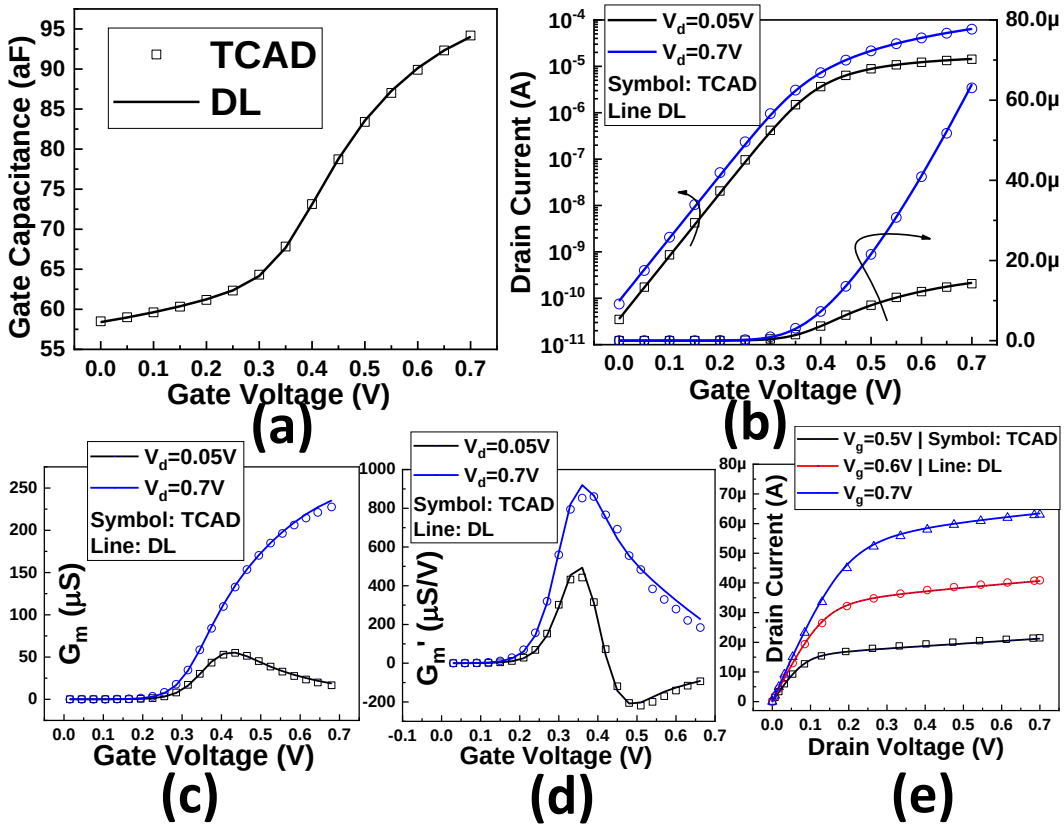


Fig 9.3. The DL-based paramters extraction model testing results by the TCAD data. (a) C_{gg}-V_g. (b) I_d-V_g at V_d=0.05V and V_d=0.7V. (c) gm at V_d=0.05V and V_d=0.7V. (d) gm' at V_d=0.05V and V_d=0.7V. (e) I_d-V_d.

value sampling. A total 45k number of data points are used to train the DL model. The rest 5k number of data points are used for validation. According to the specification, geometry-related parameters defined by technology are pre-assigned values, such as gate length (L), effective oxide thickness (EOT), fin height (HFIN), fin thickness (TFIN), and the number of fingers (NFIN). Selected BSIM-CMG parameters listed in Fig. 9.2, such as PHIG, TOXP, CIT, U0, and KSATIV, are varied randomly as uniform distribution during Monte Carlo simulations. The variation ranges for these parameters have been selected based on our experience of modeling various FinFET technologies (±0.1 for PHIG and ±50% for the others). All the other BSIM-CMG parameters are set to their default values. 10% of Monte Carlo simulation data is assigned as testing data, and 10% of the training data is assigned as

validation data in the model training stage. The training stops after no improvement in loss function over the validation data for ten consecutive epochs to avoid overfitting. The score difference between the training and testing sets is within 1% to ensure no overfitting. All the input and output are scaled between 0 and 1. Note that the scalars for different bias points will be different. For example, Vg=0.5V & Vd=0.5V and Vg=0.2V & Vd=0.5V used different scalar because there are at the different elements in the input array.

9.3 Results and Discussion

After training, in the inference stage, the baseline TCAD data is used as the input to test if the trained DL model can extract BSIM-CMG parameters for this data. DL-based engine outputs BSIM-CMG parameters, then put in the HSPICE simulations. The results are plotted in Fig. 9.3. The RMS error of the Cgg-Vg fit in (a) is 0.16%, and the RMS error of the Id-Vg fit in (b) is 6.1% (the subthreshold region dominates the overall RMS error). The RMS error of the above-threshold-voltage Id-Vg fit in (b) is 0.69%. The first and second derivatives of the drain current with respect to the gate voltage are shown in (c) and (d), respectively. Id-Vd results are shown in (e). These results show an excellent model accuracy in all the characteristics.

There is no guarantee that the actual devices always comply with technology specifications because of the process variations [121] and other uncertainties, such as slight deviation from the roadmap designs. To further test the DL-based extraction process, we generated additional TCAD data for four devices with a ±10% variation in Lg and EOT compared to the baseline TCAD device. Before feeding these various TCAD cases into the DL model, we checked their C-V and I-V and ensured they were bounded by the training data set. DL-based model extraction showed good accuracy for these devices as well as shown in Fig. 9.4. Note that Lg is not changed in BSIM-CMG for TCAD gate length variation, and the deep learning model can still fit the TCAD data. In Fig. 9.4, the RMS error of the Cgg-Vg fit is 0.24%, and the RMS error of the Id-Vg fit is 13.47% (the RMS error of the above-threshold-voltage Id-Vg fit is 1.6%). For example, the DL model captures the increase of capacitance due to the rise of Lg (magenta in Fig. 9.4 (a)). Furthermore, the SS degradation due to a decrease in Lg (blue in Fig. 9.4 (b)) is also captured.

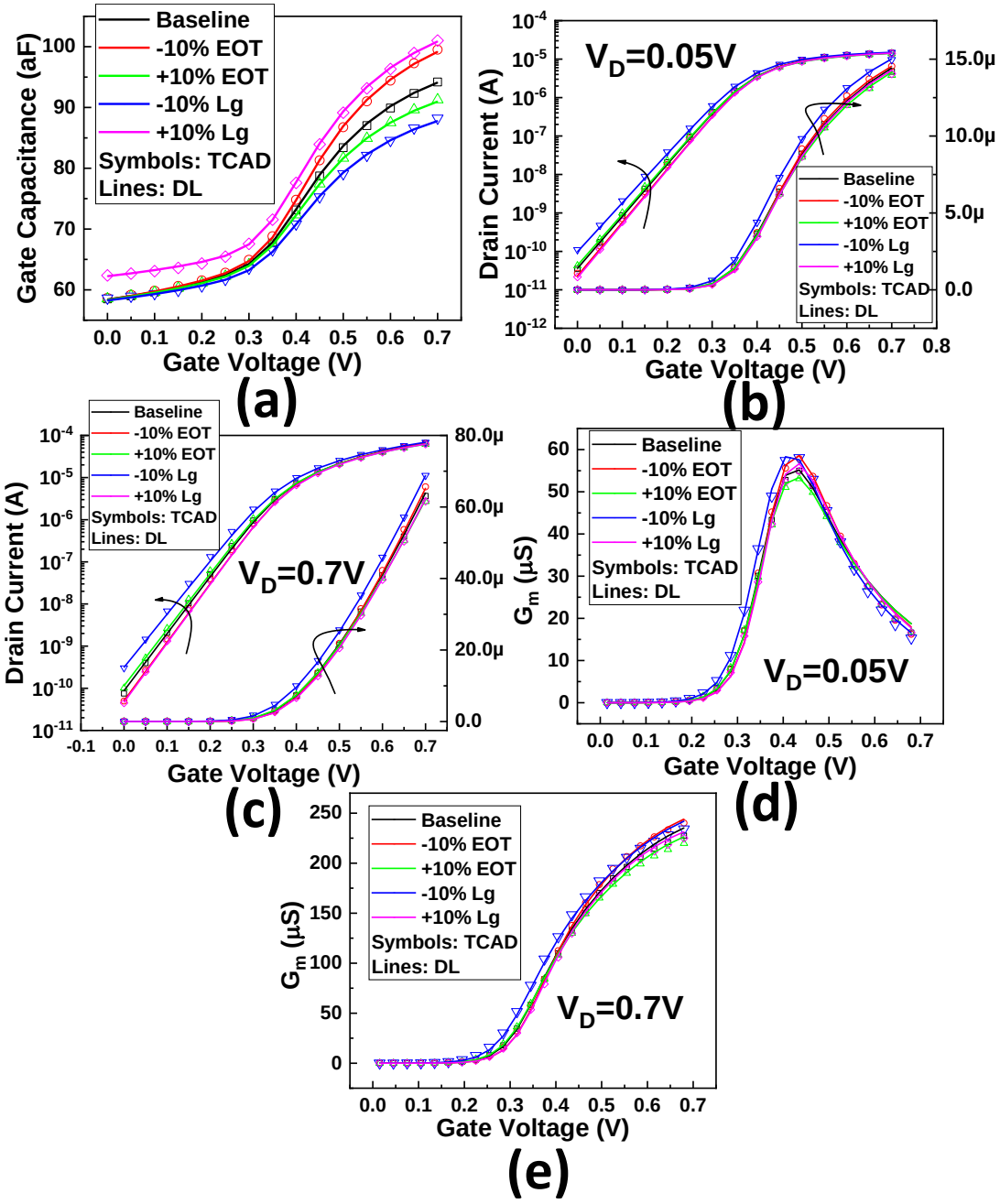


Fig. 9.4. DL-based model testing results by TCAD data with 10% variation in EOT and L_g. (a) C_{gg}-V_g. (b) I_d-V_g at V_d=0.05V. (c) I_d-V_g at V_d=0.7V. (d) gm at V_d=0.05V. (e) gm at V_d=0.7V.

Table 9.1 shows the output of the DL parameters extraction models for five testing cases. In the column of TOXP, a physical oxides thickness parameter,

the value in +10% EOT is 1.97n, which is 11.3% higher than the value of 1.77n in the baseline. The value of TOXP in -10% EOT is 1.51n, which is 14.7% lower than the value in the baseline. In the column of CIT (an SS degradation parameter), values in +10% EOT and -10% Lg are more significant than the value of the baseline, which means SS degrades. On the other hand, the CIT values in +10% Lg and -10% EOT are smaller than the baseline, which means better SS. In the column of U0 and UA, U0 values are both larger in -10% EOT and -10% Lg compared with the baseline, representing higher on-current. The UA (a surface roughness scattering related

case\parameter	PHIG (V)	CFS (F/m)	TOXP (m)	CGSL (F/m)
baseline	4.519	154p	1.77n	137p
+10% EOT	4.516	157p	1.97n	131p
+10% Lg	4.516	189p	1.45n	110p
-10% EOT	4.523	146p	1.51n	152p
-10% Lg	4.517	164p	2.15n	116p

case\parameter	CIT (F/m^2)	U0 (m^2/(V*s))	UA	EU
baseline	3.97	54.8m	4.38	0.935
+10% EOT	4.33	48.8m	3.97	0.944
+10% Lg	3.22	48.6m	3.97	0.912
-10% EOT	3.43	64.4m	5.21	0.954
-10% Lg	5.39	64.4m	4.88	1.017

case\parameter	ETA0	CDSCD (F/m^2)	VSAT1 (m/s)	KSATIV
baseline	4.49	0.731m	116k	1.88
+10% EOT	5.30	0.814m	109k	1.79
+10% Lg	3.70	1.035m	110k	1.85
-10% EOT	3.63	0.491m	128k	2.00
-10% Lg	6.32	6.158m	121k	1.91

Table 9.1. DL model output values under different testing cases. The colored values will be discussed.

parameter) value in -10% EOT is larger than the UA value in -10% Lg, consistent with the physics that reducing EOT will have a more significant effect on the surface roughness scattering than reducing Lg. In the column of ETA0 and CDSCD, the values in -10% Lg are larger, representing larger DIBL and drain-bias dependent SS degradation.

The baseline model card can be determined by the IRDS specification for future technology. The variation of the selected BSIM-CMG parameters in the Montel Carlo training data generation should be designed well according to the potential deviation of silicon data from the IRDS table.

9.4 Chapter Summary

We demonstrate a new approach for parameters extraction for BSIM-CMG, a deep learning-based model. The paper covers the flow and the architecture of the DL models, how training data are generated, and the results from TCAD data. The DL shows an excellent fit with the baseline TCAD data with 0.16% RMS error in C_{gg}-V_g and 0.69% RMS error in above-threshold-voltage I_d-V_g. This DL approach can extract parameters within a second (compared to the conventional manual fitting of hours to days) with pre-trained models and be applied to general compact models.

Chapter 10

Deep-Learning-Assisted Physics-Driven GAAFET I-V Modeling

In this work, we propose using deep learning to improve the accuracy of the partially-physics-based conventional MOSFET current-voltage model. The benefits of having some physics-driven features in the model are discussed. Using a portion of the Berkeley Short-channel IGFET Common-Multi-Gate (BSIM-CMG), the industry-standard FinFET and GAAFET compact model, as the physics model and a 3-layer neural network with 6 neurons per layer, the resultant model can well predict *IV*, output conductance, and transconductance of a TCAD-simulated gate-all-around transistor (GAAFET) with outstanding 3-sigma errors of 1.3%, 4.1%, and 2.9%, respectively. Implications for circuit simulation are also discussed.

10.1 Motivation

Accurate and fast current-voltage (I–V) models are critical for integrated circuit simulation. Compact models are traditionally derived from physics [106,122,123]; they involve solving nonlinear differential equations approximately where closed-form solutions may not exist. Furthermore, as transistors are scaled, equations that account for nanoscale effects and non-idealities become more mathematically complicated. By sacrificing accuracy, semi-empirical models with fitting parameters are commonly used. To improve model accuracy, pure look-up-table [124] or deep-learning-based

transistor models [114-117,125-127] have recently been proposed in which the measured IV data were tabulated or used for model training. However, some of these approaches result in non-physical behavior such as non-zero current at VDS = 0V and/or asymmetric IV model. Furthermore, the modeling of variation is complex in the pure deep-learning-based model.

To alleviate these, a deep-learning assisted, partially-physics-based, IV model that combines the advantages of both strategies is presented in this paper. The rest of this article is organized as follows. Section II sets up the model framework and discusses the physics-driven requirements. Then, sections III and IV present the neural network design and dataset, respectively. Using Berkeley Short-channel IGFET Common-Multi-Gate (BSIM-CMG) and the simulated IV characteristics of 12nm gate-length gate-all-around (GAAFET) technology as the physics model and the dataset, respectively, section V presents the modeling results. Implication on circuit simulations is also discussed. Finally, we conclude this paper in Section VI.

10.2 Deep-Learning-Assisted IV Model

We select the core BSIM-CMG [106] (IDS,BSIM) as the starting physics-based model. The charge density model, the transport model, and the terminal charge model are included in the core model. Then, a bias-dependent correction function ε(VGS,VGD) is introduced to account for the aforementioned non-idealities not included in the core model:

$$I_{DS}(V_{GD},V_{GD}) = I_{DS,BSIM}(V_{GS},V_{GD}) \times \varepsilon(V_{GS},V_{GD}) \quad (10.1)$$

in which ε is to be trained by deep learning. Before doing so, we first discuss the physics-driven requirements on IDS and ε.

Physics requires IDS to be zero when VDS=0. Physics-based $I_{DS,BSIM}$ and therefore (1) always satisfy this requirement.

Furthermore, since transistors are symmetric devices, the direction of the current flow change if we swap the source and drain voltage, i.e. IDS(VGS,VGD) = - IDS(VGD,VGS), substituting this condition into (1), we get:

$$I_{DS,BSIM}(V_{GS},V_{GD}) \times \varepsilon(V_{GS},V_{GD}) = -I_{DS,BSIM}(V_{GD},V_{GS}) \times \varepsilon(V_{GD},V_{GS})$$

Since BSIM-CMG is symmetric, i.e. IDS,BSIM(VGS,VGD) = - IDS,BSIM(VGD,VGS), the correction function must be symmetric:

$$\varepsilon(V_{GS},V_{DS}) = \varepsilon(V_{GD},V_{GS})$$

This condition can be satisfied if we transform input (VGS,VGD) by T before feeding into the neural network:

$$(u_1, u_2) = T(V_{GS}, V_{GD})$$

where $u_1 = V_{GS} + V_{GD}$ and $u_2 = (V_{GS} - V_{GD})^2$ (10.2)

To prove that T is symmetric, we simply swap VD and VS:

$$(u_1^*, u_2^*) = T(V_{GD}, V_{GS})$$

where $u_1^* = V_{GD} + V_{GS}$ and $u_2^* = (V_{GD} - V_{GS})^2$

Therefore $u_1^* = u_1$, $u_2^* = u_2$ and $T(V_{GD}, V_{GS}) = T(V_{GS}, V_{GD})$.

Finally, IDS must be infinitely differentiable with respect to VGS and VDS. The core model is physics-based and already meets this requirement. As a result, only the correction function, i.e., the activation function, must be infinitely differentiable. In addition, analog circuit applications also require accurate prediction for the output conductance gds and transconductance gm. Another advantage of having a physics-based core model is that it is capable of predicting the effects of changing gate-length, channel-width, channel (fin or nano-sheet) thickness and doping concentration, gate work-function, and the gate effective-oxide-thickness-at least over a particular range such as +/-

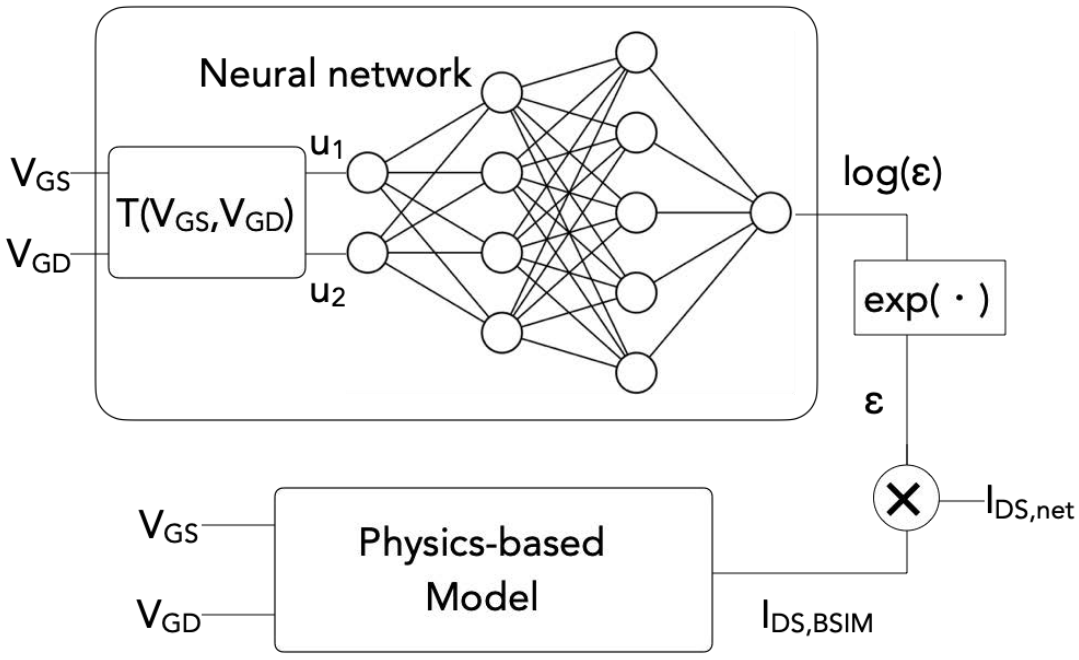


Fig 10.1. Deep-learning-assisted *IV* model architecture.

20% for the training target. Since the machine-learning-based correction function is independent of these physical variables and is the only function of VGS and VGD, the final model I-V should have a similar predictive capability as the physical core model. This ability is valuable for model manufacturing process variations. For a wide range of L and W, the core model's parameters can be binned [128] or made functions of L and W. We will discuss the neural network design in the next section.

10.3 Neural Network Design

Fig. 10.1 shows the neural network design used in this work. Input (VGS,VGD) is first transformed with T before feeding into the neural network. After comparing several neural network designs, a 3-hidden-layer network with 6 neurons per layer is found to provide good model accuracy and fast training and is selected. Hyperbolic tangent function tanh is used as the activation function for the input and hidden layers due to its infinite differentiability. Since ε is always positive, the network is designed to train log(ε) for improved convergence. The resulting ε is multiplied by the physics model IDS,BSIM to obtain the final IDS,net, where the subscript net denotes the network output.

The cost function J is defined as the average root mean square relative errors in IDS, gds, and gm:

$$J = \frac{1}{3}\left[\sqrt{\frac{1}{m}\sum_{i=1}^{m} RE^2\left(I_{DS,net}^{(i)}, I_{DS,data}^{(i)}\right)}\right.$$

$$+\sqrt{\frac{1}{m}\sum_{i=1}^{m} RE^2\left(gds_{net}^{(i)}, gds_{data}^{(i)}\right)} \quad \left. +\sqrt{\frac{1}{m}\sum_{i=1}^{m} RE^2\left(gm_{net}^{(i)}, gm_{data}^{(i)}\right)}\right] \tag{10.3}$$

where m is the training set sample size, i is the i-th sample, and subscript data denotes measured data. RE is the relative error:

$$RE(x_{net}, x_{data}) = (x_{net} - x_{data})/(x_{data} + \delta)$$

where δ is a user-defined infinitesimal parameter that prevents numerical overflow. δ = 1E-10 is used in this study.

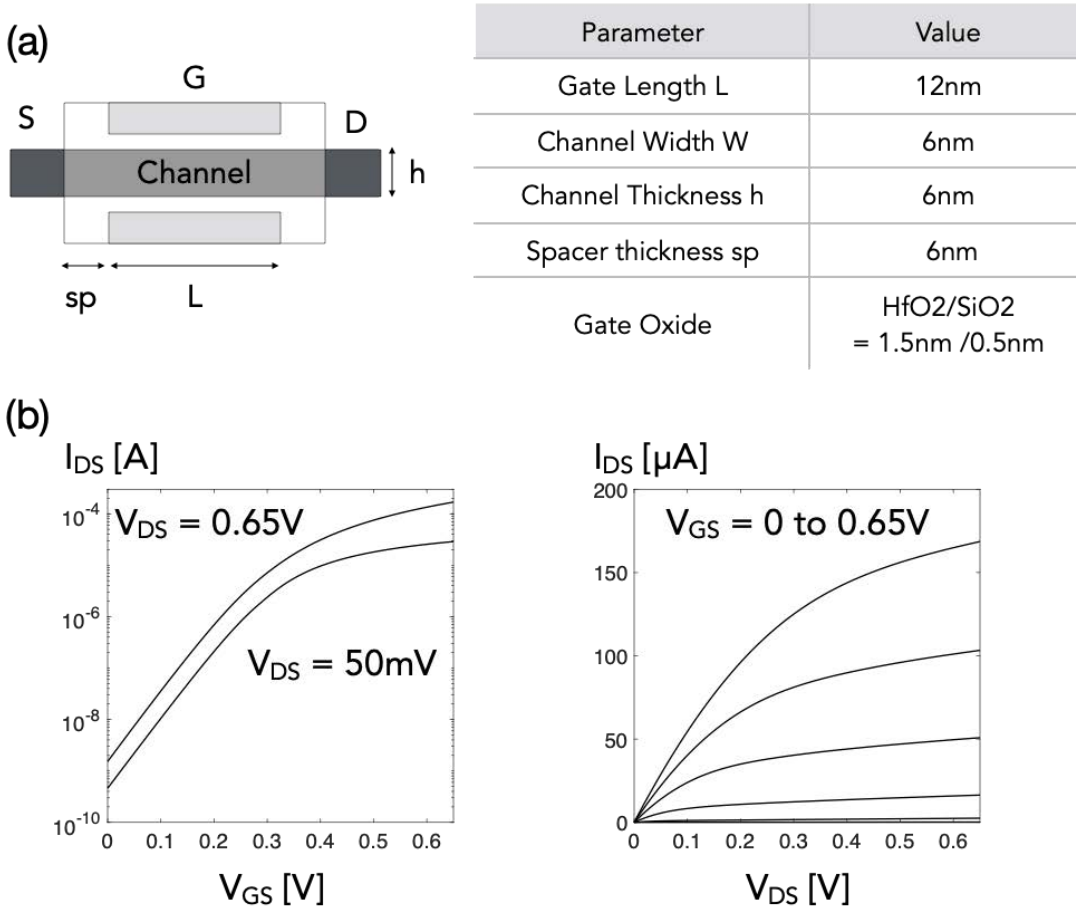
The central difference method is used to estimate the derivatives. We first

define a perturbation voltage dv. During training, we feed (VGS,VGD) together with (VGS+dv,VGD+dv), (VGS-dv,VGD-dv), (VGS,VGD-dv), (VGS,VGD+dv) into the neural network to compute IDS,net(VGS,VGD), IDS,net(VGS+), IDS,net (VGS-), IDS,net(VDS+), IDS,net(VDS-) respectively. We then approximate the partial derivatives by:

$$gm_{nn} \approx \left[I_{DS,nnet}(V_{GS}+) - I_{DS,nnet}(V_{GS}-)\right]/(2dv)$$

$$gds_{nn} \approx \left[I_{DS,nnet}(V_{DS}+) - I_{DS,nnet}(V_{DS}-)\right]/(2dv)$$

With the network design established, we will discuss data preparation in the



Parameter	Value
Gate Length L	12nm
Channel Width W	6nm
Channel Thickness h	6nm
Spacer thickness sp	6nm
Gate Oxide	HfO2/SiO2 = 1.5nm /0.5nm

Fig 10.2. (a) Cross-sectional schematic of the simulated GAAFET structure and (b) simulated transfer and output *IV* characteristics.

next section.

10.4 Dataset

Ideally, we would use the measured device data from the silicon foundry as the dataset. In this work, it suffices to use the finite element TCAD tool, SENTARUSTM, to generate the training I-V dataset (IDS,data). Fig. 10.2 shows the GAAFET structure and the simulated IV characteristics calibrated to the IRDS "1.5nm node" technology [98]. Physics models such as the Philips unified mobility model [129], thin layer mobility, Fermi Dirac Distribution, and correction due to Fermi statistics and quantization effect are included in the simulation.

The same device geometry, including the channel length, width, and thickness, together with the equivalent oxide thickness and gate work function, all of which would be known to the silicon foundry, are also used in the core BSIM-CMG model. Since default parameters are used for all other model parameters without calibration, it results in a large mean and standard deviation of IDS,data/IDS,BSIM 6.2 and 5.1, respectively.

The neural network is trained with a dataset of (VGS,data, VDS,data, IDS,data) with VGS,data and VDS,data range from 0V to 0.65V with 50mV increment. Then, we test the trained model using a dataset with the same voltage ranges but a smaller voltage increment of 5mV. Training and testing dataset size are therefore $14\times14=196$ and $131\times131=17161$, respectively, resulting in a training-to-test-set ratio of $\sim1/100$. Partial derivatives gdsdata, gmdata are computed using the central difference formula with $dv = 10mV$.

TensorFlow with adaptive moment estimation ("Adam") optimization with a learning rate of 1E-4 is used for training. Adam optimization is a gradient-descent-based optimization algorithm in which the adaptive learning rate is estimated based upon the first and second moments of the gradients [130]. Given the small training size of < 200, a single batch is fed into the neural network.

10.5 Results and Discussion

Fig. 10.3 shows the cost vs. epoch and the testing results for the 3-layer

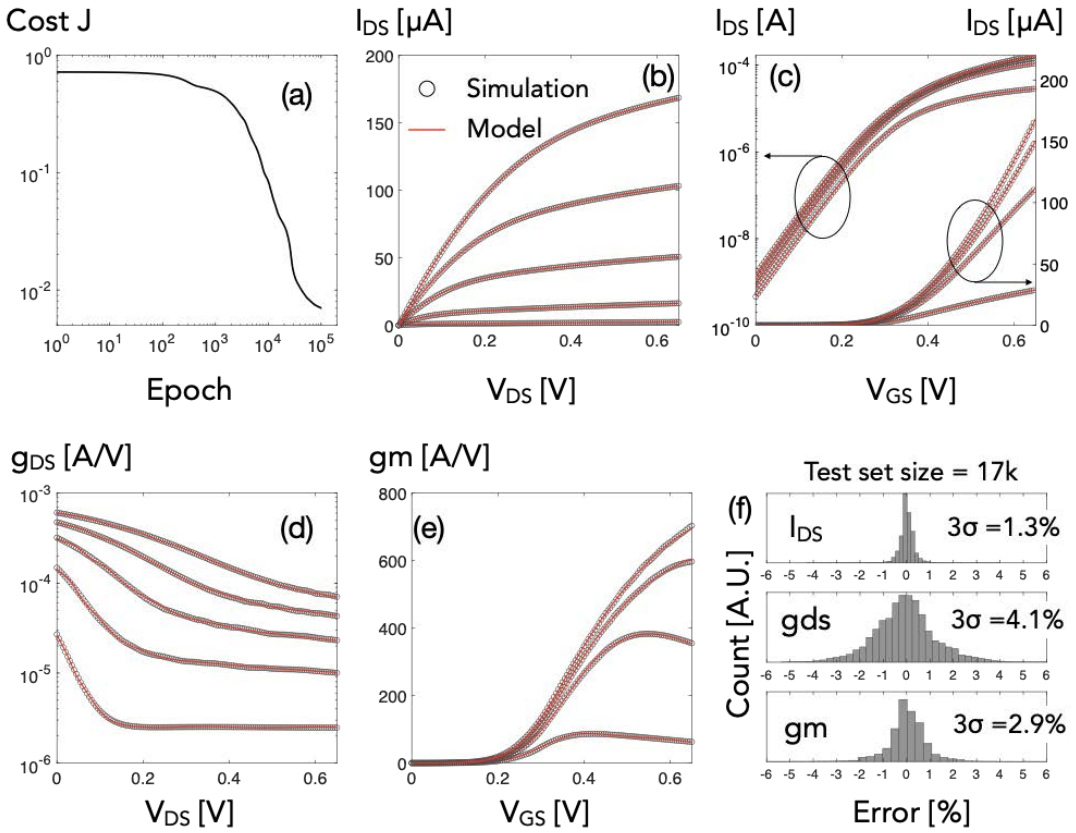


Fig 10.3. (a) Training error vs epoch. Model well-matches to the test dataset in (b) output and (c) transfer characteristic (d) output conductance, and (e) transconductance. (e) Percentage error distribution in I_{DS}, *gds* and *gm*.

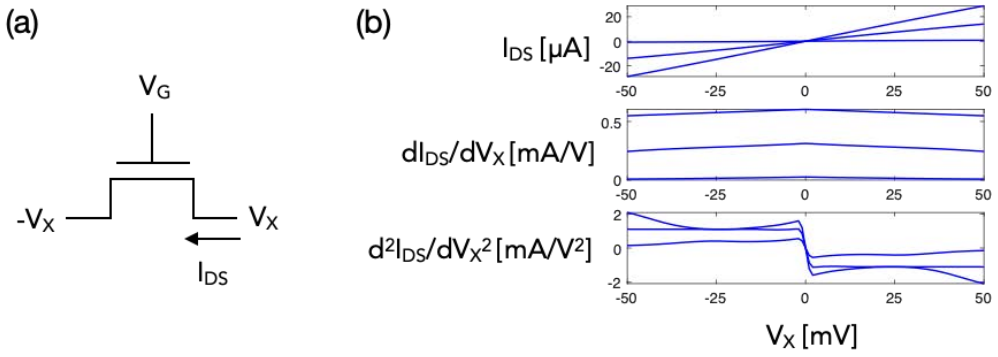


Fig. 10.4. (a) Gummel symmetry test setup. (b) I_{DS}, first and second derivative of I_{DS} with respect to V_X.

neural network. The trained model well predicts the test IV, gds, and gm to 3σ error of 1.3%, 4.1%, and 2.9%, respectively. Furthermore, as shown in Fig. 10.4, the trained model satisfies the Gummel symmetry [131]. This is unsurprising because ε is symmetric by design, as earlier discussed.

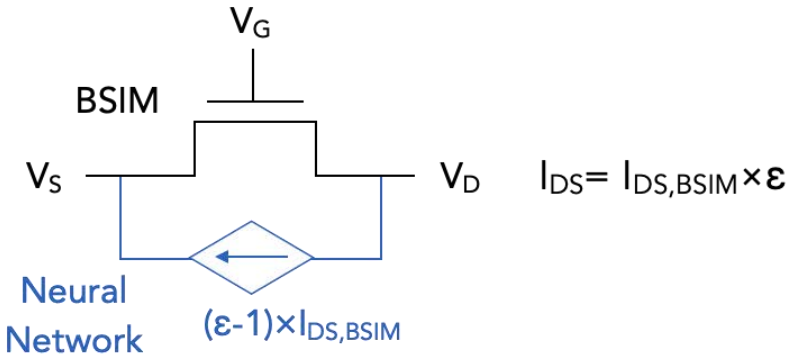


Fig. 10.5. (a) Model implementation for circuit simulations

Fig 10.5 shows the model implementation for circuit simulation. By rewriting (1) as IDS= IDS,BSIM +IDS,BSIM×(ε-1), one can model the neural network as an add-on current source.

Since the core BSIM-CMG (~100 lines of codes) is 10× shorter than the full version counterparts (~ 2000 lines), the deep-learning approach discussed herein reduces the BSIM model calculation time. The sum of that and the added neural network operations, which include matrix multiplications and hyperbolic tangent functions, is expected to be substantially faster than calculating the full BSIM-CMG I-V model if the neural network operations can be parallelized using specialized hardware such as the graphical processing units (GPU). The proposed model also accelerates the parameter extraction and process design kit (PDK) development. All these make our approach very attractive for I-V modelling.

10.6 Chapter Summary

In this work, hybrid analytical and deep-learning-assisted MOSFET IV modeling is proposed. The reasons for using a physics-based core are discussed. Using a 12nm gate length GAAFET technology as an example, our model utilizing a 3-layer neural network with 18 neurons can predict the simulated IV to 1.3% in 3-sigma error. Our model is Gummel symmetric and can readily be extended to model a wide range of L and W for circuit simulation.

Chapter 11

Summary

11.1 Chapter Summary

Challenges to the advancement of nanoelectronics have been brought out at first. The concept of NCFET and the BSIM model have also been explained in Chapter 1. Chapter 2 demonstrates how an NCFET's capacitance matching could be further optimized by introducing a non-uniform interfacial layer. A non-uniform interfacial layer makes the MOSFET capacitance more uniform along the channel, and the overall capacitance matching can be enhanced. A new scheme of variation TCAD simulation of dielectric and ferroelectric grains has been proposed in Chapter 3. The method proposed in Chapter 3 can be applied to NCFETs with impurity DE-phase grains in the FE film.

In Chapter 4, the effect of polarization gradient effect has been illustrated. The polarization gradient effect is the interaction between neighbors of the FE dipole that force the neighbors to be continuous. The effect can lead to negative DIBL and negative drain resistor phenomena. Compact modeling of the polarization gradient effect has also been explained in Chapter 4. Energy flow in the system of NCFETs has been shown in Chapter 5. The energy calculation shows consistency by two

different approaches: energy integration across the materials in TCAD and energy flow calculation from a circuit point of view.

NC's benefits on FinFET and GAAFET according to the IRDS roadmap have been illustrated in Chapters 6 and 7. FinFET structure can be extended by two more nodes with the help of NC, and the GAA structure can be developed by three more nodes with the support of NC. Compact modeling of anti-FE has been demonstrated in Chapter 8. With proper capacitance matching, an additional layer of anti-FE could potentially enhance the performance of NCFET.

Potential applications of machine learning in the compact model have been shown in Chapters 9 and 10. Chapter 9 offers a near-term ML application – a deep learning application on compact models' parameter extraction. With the help of ML, the process design kit developer could save tons of time on the manual parameter extraction process. Chapter 10 shows a longer-term goal of ML application as an alternative to an equation-based compact model. This could potentially speed up spice simulation by ten times.

11.1 Future Work

There are many exciting future research topics in nanoelectronics, including neuromorphic in-memory computing, production of NCFETs, machine learning applications in compact modeling, etc. The evolution of electronics can and will be speeded up with the help of software-hardware co-optimization.

Bibliography

[1] G. E. Moore, "Cramming More Components Onto Integrated Circuits," in Proceedings of the IEEE, vol. 86, no. 1, pp. 82-85, Jan. 1998. doi: 10.1109/JPROC.1998.658762

[2] S. Salahuddin and S. Datta, "Can the subthreshold swing in a classical FET be lowered below 60 mV/decade?," 2008 IEEE International Electron Devices Meeting, San Francisco, CA, 2008, pp. 1-4. doi: 10.1109/IEDM.2008.4796789

[3] S. Salahuddin and S. Datta, “Use of Negative Capacitance to Provide Voltage Amplification for Low Power Nanoscale Devices,” Nano Letters, vol 8, no. 2, pp. 405-410. doi: 10.1021/nl071804g

[4] A. I. Khan, K. Chatterjee, B. Wang, S. Drapcho, L. You, C. Serrao, S. R. Bakaul, R. Ramesh, and S. Salahuddin, “Negative capacitance in a ferroelectric capacitor,” Nature Materials, vol. 14, pp. 182–186, Dec. 2014, doi:10.1038/nmat4148.

[5] V. V. Zhirnov, and R. K. Cavin, “Negative capacitance to the rescue,” Nature Nanotech, vol. 3, pp. 77–78, Feb. 2008, doi: 10.1038/nnano.2008.18.

[6] O. Y. Loh and H. D. Espinosa, “Nanoelectromechanical contact switches,” Nature Nanotech., vol. 7, pp. 283–295, May 2012. doi: 10.1038/NNANO.2012.40.

[7] W. Y. Choi, B.-G. Park, J. D. Lee, and T.-J. K. Liu, “Tunneling field-effect transistors (TFETs) with subthreshold swing (SS) less than 60 mV/dec,” IEEE Electron Device Lett., vol. 28, no. 8, pp. 743–745, Aug. 2007. doi:

10.1109/LED.2007.901273.

[8] U. E. Avci, B. Chu-Kung, A. Agrawal, G. Dewey, V. Le, R. Rios, D. H. Morris, S. Hasan, R. Kotlyar, J. Kavalieros, and I. A. Young, “Study of TFET non-ideality effects for determination of geometry and defect density requirements for sub-60mV/dec Ge TFET,” in IEDM Tech. Dig., Dec. 2015, pp. 891–894. doi: 10.1109/IEDM.2015.7409828.

[9] T. N. Theis, P. M. Solomon, “It’s Time to Reinvent the Transistor,” Science, vol. 327, no. 5973, pp. 1600-1601, Mar. 2010, doi: 10.1126/science.1187597

[10] Z. Krivokapic, U. Rana, R. Galatage, A. Razavieh, A. Aziz, J.Liu, J.Shi, H. J. Kim, R. Sporer, C. Serrao, A. Busquet, P. Polakowski, J. Müller, W. Kleemeier, A. Jacob, D. Brown, A. Knorr, R. Carter, and S. Banna, “14nm Ferroelectric FinFET technology with steep subthreshold slope for ultra low power applications,” in IEDM, San Francisco, CA, USA, 2017, doi: 10.1109/IEDM.2017.8268393.

[11] M. H. Lee, P.-G. Chen, C. Liu, K. Chu, C.-C. Cheng, M.-J. Xie, S.-N. Liu, J.-W. Lee, S.-J. Huang, M.-H. Liao, M. Tang, K.-S. Li, and M.-C. Chen, “Prospects for ferroelectric HfZrOx FETs with experimentally CET=0.98 nm, SSfor=42 mV/dec, SSrev=28 mV/dec, switch-off <0.2 V, and hysteresis-free strategies,” in IEDM Tech. Dig., Dec. 2015, pp. 22.5.1–22.5.4, doi: 10.1109/IEDM.2015.7409759.

[12] K.-S. Li, P.-G. Chen, T.-Y. Lai, C.-H. Lin, C.-C. Cheng, C.-C. Chen, Y.-J. Wei, Y.-F. Hou, M.-H. Liao, M.-H. Lee, M.-C. Chen, J.-M. Sheih, W.-K. Yeh, F.-L. Yang, S. Salahuddin, and C. Hu, “Sub-60 mV-swing negative-capacitance FinFET without hysteresis,” in IEDM Tech. Dig., Dec. 2015, pp. 22.6.1–22.6.4, doi: 10.1109/IEDM.2015.7409760

[13] A. Rusu, G. A. Salvatore, D. Jimenéz, and A. M. Ionescu, “Metal-ferroelectric-meta-oxide-semiconductor field effect transistor with sub-60 mV/decade subthreshold swing and internal voltage amplification,” in IEDM Tech. Dig., Dec. 2010, pp. 16.3.1–16.3.4, doi: 10.1109/IEDM.2010.5703374.

[14] W. Chung, M. Si, and P. D. Ye, “Hysteresis-free negative capacitance germanium CMOS FinFETs with Bi-directional Sub-60 mV/dec,” in IEDM,

San Francisco, CA, USA, 2017, doi: 10.1109/IEDM.2017.8268395.

[15] C.-C. Fan, C.-H. Cheng, Y.-R. Chen, C. Liu, and C.-Y. Chang, “Energy-efficient HfAlOx NCFET: Using gate strain and defect passivation to realize nearly hysteresis-free sub-25mV/dec switch with ultralow leakage,” in IEDM, San Francisco, CA, USA, 2017, doi: 10.1109/IEDM.2017.8268444.

[16] M. Hoffmann, M. Pesic, S. Slesazeck, U. Schroeder, and T. Mikolajick, “On the stabilization of ferroelectric negative capacitance in nanoscale devices,” Nanoscale, vol. 10, pp. 10891–10899, May 2018. doi: 10.1039/C8NR02752H.

[17] G. Pahwa, T. Dutta, A. Agarwal, and Y. S. Chauhan, “Physical Insights on Negative Capacitance Transistors in Nonhysteresis and Hysteresis Regimes: MFMIS Versus MFIS Structures,” IEEE Trans. Electron Devices, vol. 65, no. 3, pp. 867-873, Mar. 2018, doi: 10.1109/TED.2018.2794499.

[18] V. Garcia, and M. Bibes, “Ferroelectric tunnel junctions for information storage and processing,” Nature Communications, no. 5, pp 4289, Jul. 2014, doi: 10.1038/ncomms5289.

[19] 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 IEDM, San Francisco, CA, USA, 2017, doi: 10.1109/IEDM.2017.8268385.

[20] Z. C. Yuan, S. Rizwan, M. Wong, K. Holland, S. Anderson, T. B. Hook, S. Kienle, S. Gadelrab, P. S. Gudem, M. Vaidyanathan, “Switching-Speed Limitations of Ferroelectric Negative-Capacitance FETs,” IEEE Trans. Electron Devices, vol. 63, no. 10, pp. 4046-4052, Oct. 2016, doi: 10.1109/TED.2016.2602209.

[21] A. Cano, and D. Jimenez, “Multidomain ferroelectricity as a limiting factor for voltage amplification in ferroelectric field-effect transistors,” Appl. Phys. Lett. Vol. 97, no. 13, Sep. 2010, doi: 10.1063/1.3494533.

[22] A. I. Khan, U. Radhakrishna, S. Salahuddin, and D. Antoniadis, “Work

Function Engineering for Performance Improvement in Leaky Negative Capacitance FETs," IEEE Electron Device Letters, vol. 38, no. 9, pp. 1335-1338, Sept. 2017, doi: 10.1109/LED.2017.2733382.

[23] J. Zhou, Y. Peng, G. Han, Q. Li, Y. Liu, J. Zhang, M. Liao, Q.-Q. Sun, D. W. Zhang, Y. Zhou, Y. Hao, "Hysteresis Reduction in Negative Capacitance Ge PFETs Enabled by Modulating Ferroelectric Properties in HfZrOx," IEEE Journal of the Electron Devices Society, vol. 6, pp. 41-48, Oct. 2017, doi: 10.1109/JEDS.2017.2764678.

[24] G. Pahwa, A. Agarwal, and Y. S. Chauhan, "Numerical Investigation of Short-Channel Effects in Negative Capacitance MFIS and MFMIS Transistors: Subthreshold Behavior," IEEE Trans. Electron Devices, Sep. 2018, doi: 10.1109/TED.2018.2870519.

[25] M. H. Park, Y. H. Lee, H. J. Kim, Y. J. Kim, T. Moon, K. D. Kim, J. Muller, A. Kersch, U. Schroeder, T. Mikolajick, and C. S. Hwang, "Ferroelectricity and antiferroelectricity of doped thin HfO2-based films," Adv. Mater., vol. 27, no. 11, pp. 1811–1831, Mar. 2015. doi: 10.1002/adma.201404531.

[26] M.-Y. Kao, A. B. Sachid, Y.-K. Lin, Y.-H. Liao, H.Agarwal, P. Kushwaha, J. P. Duarte, H.-L. Chang, S. Salahuddin, and C. Hu, "Variation Caused by Spatial Distribution of Dielectric and Ferroelectric Grains in a Negative Capacitance Field-Effect Transistor," IEEE Trans. Electron Devices, vol. 65, no. 10, pp. 4652-4658, Oct. 2018, doi: 10.1109/TED.2018.2864971.

[27] Sentaurus Device, Synopsys, Inc., CA, USA, 2017.

[28] M. Hoffmann, M. Pesic, S. Slesazeck, U. Schroeder, and T. Mikolajick, "On the stabilization of ferroelectric negative capacitance in nanoscale devices," Nanoscale, no. 23, pp. 10891-10899, May. 2018, doi: 10.1039/C8NR02752H.

[29] C.-I. Lin, A. I. khan, S. Salahuddin, and C. Hu, "Effects of the Variation of Ferroelectric Properties on Negative Capacitance FET Characteristics," IEEE Trans. Electron Devices, vol. 63, no. 5, pp. 2197-2199, Jan. 2016, doi: 10.1109/TED.2016.2514783

[30] K. Chatterjee, A. J. Rosner, and S. Salahuddin, “Intrinsic speed limit of negative capacitance transistors,” IEEE Electron Device Letters, vol. 38, no. 9, pp. 1328-1330, Sept. 2017, doi: 10.1109/LED.2017.2731343.

[31] 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 Trans. Electron Devices, vol. 47, no. 12, pp. 2320–2325, Dec. 2000, doi: 10.1109/16.887014.

[32] S. Joglekar, U. Radhakrishna, D. Piedra, D. Antoniadis, and T. Palacios, “Large Signal Linearity Enhancement of AlGaN/GaN High Electron Mobility Transistors by Device-level VT Engineering for Transconductance Compensation,” in IEDM, San Francisco, CA, USA, 2017, doi: 10.1109/IEDM.2017.8268457.

[33] S. Dasgupta, A. Rajashekhar, K. Majumdar, N. Agrawal, A. Razavieh, S. Troiier-McKinstry, and S. Datta, “Sub-kT/q switching in strong inversion in PbZr0.52Ti0.48O3 gated negative capacitance FETs,” IEEE J. Exploratory Solid-State Comput. Device Circuits, vol. 1, pp. 43–48, Dec. 2015, doi: 10.1109/JXCDC.2015.2448414.

[34] A. I. Khan, K. Chatterjee, J. P. Duarte, Z. Lu, A. Sachid, S. Khandelwal, R. Ramesh, C. Hu, and S. Salahuddin, “Negative capacitance in short-channel FinFETs externally connected to an epitaxial ferroelectric capacitor,” IEEE Electron Device Lett., vol. 37, no. 1, pp. 111–114, Jan. 2016, doi: 10.1109/LED.2015.2501319

[35] A. Nourbakhsh, A. Zubair, S. Joglekar, M. Dresselhaus, and T. Palacios, “Subthreshold swing improvement in MoS2 transistors by the negative-capacitance effect in a ferroelectric Al-doped-HfO2/HfO2 gate dielectric stack,” Nanoscale, vol. 9, pp. 6122–6127, Apr. 2017, doi: 10.1039/C7NR00088J.

[36] J. Zhou, G. Han, Q. Li, Y. Peng, X. Lu, C. Zhang, J. Zhang, Q.-Q. Sun, D. W. Zhang, and Y. Hao, “Ferroelectric HfZrOx Ge and GeSn PMOSFETs with sub-60 mV/decade subthreshold swing, negligible hysteresis, and improved Ids,” in IEDM Tech. Dig., Dec. 2016, pp. 12.2.1–12.2.4, doi: 10.1109/LED.2015.2501319.

[37] C. Hu, S. Salahuddin, C. I. Lin, and A. Khan, “0.2V adiabatic NC-FinFET with 0.6mA/ um ION and 0.1nA/um IOFF,” in IEEE 73rd Annual Device Research Conference (DRC), Columbus, OH, USA, 2015, doi: 10.1109/DRC.2015.7175542.

[38] J. P. Duarte, S. Khandelwal, A. I. Khan, A. Sachid, Y.-K. Lin, H.-L. Chang, S. Salahuddin, and C. Hu, “Compact models of negative-capacitance FinFETs: Lumped and distributed charge models,” in IEDM, San Francisco, CA, USA, 2016, doi: 10.1109/IEDM.2016.7838514.

[39] H. Agarwal et al., "Engineering Negative Differential Resistance in NCFETs for Analog Applications," in IEEE Transactions on Electron Devices, vol. 65, no. 5, pp. 2033-2039, May 2018. doi: 10.1109/TED.2018.2817238

[40] L. Xu, T. Nishimura1, S. Shibayama1, T. Yajima1, S. Migita, and A. Toriumi, “Kinetic pathway of the ferroelectric phase formation in doped HfO2 films,” Journal of Applied Physics, vol. 122, no. 12, pp. 124104, Sep. 2017, doi: 10.1063/1.5003918.

[41] A. K. Saha, P. Sharma, I. Dabo1, 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 IEDM, San Francisco, CA, USA, 2017, doi: 10.1109/IEDM.2017.8268385.

[42] H. Ota, K. Fukuda, T. Ikegami, J. Hattori, H. Asai, S. Migita, and A. Toriumi, “Perspective of Negative Capacitance FinFETs Investigated by Transient TCAD Simulation,” in IEDM, San Francisco, CA, USA, 2017, doi: 10.1109/IEDM.2017.8268394.

[43] L. D. Filip, L. Pintilie, W.-S. Tam, C.-W. Kok, “Leakage current for thin film metal-ferroelectric-metal device,” in Next-Generation Electronics, Hsinchu, Taiwan, 2016, doi: 10.1109/ISNE.2016.7543292.

[44] M. H. Park, H. J. Kim, Y. J. Kim, T. Moon, and C. S. Hwanga, “The effects of crystallographic orientation and strain of thin Hf0.5Zr0.5O2 film on its ferroelectricity,” Appl. Phys. Lett., vol. 104, no. 7, Feb. 2014, doi: 10.1063/1.4866008.

[45] 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," in IEEE Electron Device Letters, vol. 40, no. 3, pp. 463-466, March 2019. doi: 10.1109/LED.2019.2891540.

[46] G. Pahwa, A. Agarwal and Y. S. Chauhan, "Numerical Investigation of Short-Channel Effects in Negative Capacitance MFIS and MFMIS Transistors: Above-Threshold Behavior," in IEEE Transactions on Electron Devices, vol. 66, no. 3, pp. 1591-1598, March 2019. doi: 10.1109/TED.2019.2892186.

[47] C. Liu, H. Chen, C. Hsu, C. Fan, H. Hsu and C. Cheng, "Negative Capacitance CMOS Field-Effect Transistors with Non-Hysteretic Steep Sub-60mV/dec Swing and Defect-Passivated Multidomain Switching," 2019 Symposium on VLSI Technology, Kyoto, Japan, 2019, pp. T224-T225. doi: 10.23919/VLSIT.2019.8776482

[48] M. H. Lee, K. Chen, C. Liao, G. Siang, C. Lo, H. Chen, Y. Tseng, C. Chueh, C. Chang, Y. Lin, Y. Yang, F. Hsieh, S. Chang, M. Liao, K. Li, and C. Liu, "Bi-directional Sub-60mV/dec, Hysteresis-Free, Reducing Onset Voltage and High Speed Response of Ferroelectric-AntiFerroelectric Hf0.25Zr0.75O2 Negative Capacitance FETs," 2019 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2019, pp. 23.6.1-23.6.4. doi: 10.1109/IEDM19573.2019.8993581

[49] Y.-K. Lin, H. Agarwal, P. Kushwaha, M. Kao, Y. Liao, K. Chatterjee, S. Salahuddin, and C. Hu, "Analysis and Modeling of Inner Fringing Field Effect on Negative Capacitance FinFETs," in IEEE Transactions on Electron Devices, vol. 66, no. 4, pp. 2023-2027, April 2019. doi: 10.1109/TED.2019.2899810.

[50] Sentaurus Process User Guide, Version O-2018.06, Synopsys, Mountain View, CA, USA, Sep. 2018.

[51] S. S. Cheema, D. Kwon, N. Shanker, R. d. Reis, S.-L. Hsu, J. Xiao, H. Zhang, R. Wagner, A. Datar, M. R. McCarter, C. R. Serrao, A. K. Yadav, G. Karbasian, C.-H. Hsu, A. J. Tan, L.-C. Wang, V. Thakare, X. Zhang, A. Mehta, E. Karapetrova, R. V. Chopdekar, P. Shafer, E. Arenholz, C. H, R. Proksch, R. Ramesh, J. Ciston, and S. Salahuddin, “Enhanced ferroelectricity in ultrathin

films grown directly on silicon,” in Nature, vol. 580, pp. 478-482, Apr. 2020, doi: 10.1038/s41586-020-2208-x

[52] J. Muller, T. S. Boscke, U. Schroder, S. Mueller, D. Brauhaus, U. Bottger, L. Frey, and T. Mikolajick, “Ferroelectricity in Simple Binary ZrO2 and HfO2,” in Nano Letters, vol. 12, no. 8, pp. 4318-4323, Aug. 2012, doi: 10.1021/nl302049k

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

[54] C. Auth, A. Aluyarukunju, M. Asoro, D. Bergstrom, V. Bhagwat, J. Birdsall, N. Bisnik, M. Buehler, V. Chikarmane, G. Ding, Q. Fu, H. Gomez, W. Han, D. Hanken, M. Haran, M. Hattendorf, R. Heussner, H. Hiramatsu, B. Ho, S. Jaloviar, I. Jin, S. Joshi, S. Kirby, S. Kosaraju, H. Kothari, G. Leatherman, K. Lee, J. Leib, A. Madhavan, and K. Maria, “A 10nm high performance and low-power CMOS technology featuring 3rd generation FinFET transistors, Self-Aligned Quad Patterning, contact over active gate and cobalt local interconnects,” in IEDM Tech Dig., Dec 2017, pp. 29.1.1-29.1.4, doi: 10.1109/IEDM.2017.8268472.

[55] J. C. Wong and S. Salahuddin, "Negative Capacitance Transistors," in Proceedings of the IEEE, vol. 107, no. 1, pp. 49-62, Jan. 2019, doi: 10.1109/JPROC.2018.2884518.

[56] N. Ng, R. Ahluwalia, and D. Srolovitz, “Depletion-layer-induced size effects in ferroelectric thin films: A Ginzburg-Landau model study,” Phys. Rev. B, vol. 86, no. 9, pp. 094104-094110, Sep. 2012. doi: 10.1103/PhysRevB.86.094104.

[57] J.-P. Colinge, “Multiple-gate SOI MOSFETs,” Solid-State Electron., vol. 48, no. 6, pp. 897–905, Jun. 2004. doi: 10.1016/j.sse.2003.12.020.

[58] W. You, P. Su and C. Hu, "Evaluation of NC-FinFET Based Subsystem-Level Logic Circuits," in IEEE Transactions on Electron Devices, vol. 66, no. 4, pp. 2004-2009, April 2019. doi: 10.1109/TED.2019.2898445

[59] S. E. Thompson, and S. Parthasarathy, “Moore's law: the future of Si microelectronics,” in *Materials Today*, vol. 9, no. 6, pp. 20-25, Jun. 2006, doi: 10.1016/S1369-7021(06)71539-5

[60] H. Wong, and H. Iwai, “On the scaling of subnanometer EOT gate dielectrics for ultimate nano CMOS technology,” in *Microelectronic Engineering*, vol. 138, pp. 57-76, Apr. 2015, doi: 10.1016/j.mee.2015.02.023

[61] Y. C. Yeo, Q. Lu, W. C. Lee, T.-J. King, C. Hu, X. Wang, X. Guo, and T.P. Ma, "Direct tunneling gate leakage current in transistors with ultrathin silicon nitride gate dielectric," in *IEEE Electron Device Letters*, vol. 21, no. 11, pp. 540-542, Nov. 2000, doi: 10.1109/55.877204

[62] Y. C. Yeo, T.-J. King, C. Hu, “Direct tunneling leakage current and

scalability of alternative gate dielectrics,” in *Applied Physics Letters*, vol. 81, no. 11, pp. 2091-2093, Sep. 2002, doi: 10.1063/1.1506941

[63] C. Auth, A. Cappellani, J.-. Chun, A. Dalis, A. Davis, T. Ghani, G. Glass, T. Glassman, M. Harper, M. Hattendorf, P. Hentges, S. Jaloviar, S. Joshi, J. Klaus, K. Kuhn, D. Lavric, M. Lu, H. Mariappan, K. Mistry, B. Norris, N. Rahhal-orabi, P. Ranade, J. Sandford, L. Shifren, V. Souw, K. Tone, F. Tambwe, A. Thompson, D. Towner, T. Troeger, P. Vandervoorn, C. Wallace, J. Wiedemer, and C. Wiegand, "45nm High-k + metal gate strain-enhanced transistors," *2008 Symposium on VLSI Technology*, Honolulu, HI, 2008, pp. 128-129, doi: 10.1109/VLSIT.2008.4588589.

[64] S. Barraud, L. Thevenod, M. Casse, O. Bonno, and M. Mouis, “Modeling of remote Coulomb scattering limited mobility in MOSFET with HfO2/SiO2 gate stacks,” in *Microelectronic Engineering*, vol. 84, no. 9-10, pp. 2404-2407, Sep. 2007, doi: 10.1016/j.mee.2007.04.032

[65] J. P. Duarte, Y.-K. Lin, Y.-H. Liao, A. Sachid, M.-Y. Kao, H. Agarwal, P. Kushwaha, K. Chatterjee, D. Kwon, H. Chang, S. Salahuddin, and C. Hu, "Negative-Capacitance FinFETs: Numerical Simulation, Compact Modeling and Circuit Evaluation," *2018 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Austin, TX, 2018, pp. 123-128, doi: 10.1109/SISPAD.2018.8551641.

[66] D. Kwon, K. Chatterjee, A. J. Tan, A. K. Yadav, H. Zhou, A. B. Sachid, R. D. Reis, C. Hu, and S. Salahuddin, "Improved Subthreshold Swing and Short Channel Effect in FDSOI n-Channel Negative Capacitance Field Effect Transistors," in *IEEE Electron Device Letters*, vol. 39, no. 2, pp. 300-303, Feb. 2018, doi: 10.1109/LED.2017.2787063.

[67] J. Jo and C. Shin, "Negative Capacitance Field Effect Transistor With Hysteresis-Free Sub-60-mV/Decade Switching," in *IEEE Electron Device Letters*, vol. 37, no. 3, pp. 245-248, March 2016, doi: 10.1109/LED.2016.2523681.

[68] M. H. Lee, Y. -. Wei, K. -. Chu, J. -. Huang, C. -. Chen, C. -. Cheng, M. -. Chen, H. -. Lee, Y. -. Chen, L. -. Lee, and M. -. Tsai, "Steep Slope and Near Non-Hysteresis of FETs With Antiferroelectric-Like HfZrO for Low-Power Electronics," in *IEEE Electron Device Letters*, vol. 36, no. 4, pp. 294-296, April 2015, doi: 10.1109/LED.2015.2402517.

[69] K. Li, P. Chen, T. Lai, C. Lin, C. Cheng, C. Chen, Y. Wei, Y. Hou, M. Liao, M. Lee, M. Chen, J. Sheih, W. Yeh, F. Yang, S. Salahuddin, and C. Hu, "Sub-60mV-swing negative-capacitance FinFET without hysteresis," *2015 IEEE International Electron Devices Meeting (IEDM)*, Washington, DC, 2015, pp. 22.6.1-22.6.4, doi: 10.1109/IEDM.2015.7409760.

[70] Hyeon Woo Park, Jangho Roh, Yong Bin Lee, and Cheol Seong Hwang, “Modeling of Negative Capacitance in Ferroelectric Thin Films,” in Advanced Materials, vol. 31, no. 32, pp. 1805266, Jun. 2019, doi: 10.1002/adma.201805266

[71] S. Chang, U. E. Avci, D. E. Nikonov and I. A. Young, "A Thermodynamic Perspective of Negative-Capacitance Field-Effect Transistors," in IEEE Journal on Exploratory Solid-State Computational Devices and Circuits, vol. 3, pp. 56-64, Dec. 2017, doi: 10.1109/JXCDC.2017.2750108.

[72] M. Hoffmann, F. P. G. Fengler, B. Max, U. Schroeder, S. Slesazeck, and T. Mikolajick, “Negative Capacitance for Electrostatic Supercapacitors,” in Adv. Energy Mater, vol. 9, no. 40, pp. 1901154, Sep. 2019, doi: 10.1002/aenm.201901154

[73] J. Wong, “Negative capacitance and hyperdimensional computing for unconventional low-power computing,” Ph.D. dissertation, EECS Department, University of California, Berkeley, Berkeley, 2018. Accessed on: July 14, 2020. [Online]. Available: http://www2.eecs.berkeley.edu/Pubs/TechRpts/2018/EECS-2018-187.html

[74] K.-S. Li, Y.-J. Wei, Y.-J Chen, W.-C. Chiu, H.-C. Chen, and M.-H Lee, Y.-F Chiu, F.-K. Hsueh, B.-W. Wu, P.-G Chen, T.-Y. Lai, C.-C. Chen, J.-M. Shieh, W.-K. Yeh, S. Salahuddin, and C. Hu, “Negative-Capacitance FinFET Inverter, Ring Oscillator, SRAM Cell, and Ft,” in *IEDM Tech Dig*., Dec. 2018, pp. 31.7.1-31.7.4, doi: 10.1109/IEDM.2018.8614521.

[75] V. P.-H. Hu, P.-C. Chiu, A. B Sachid, C. Hu, “Negative capacitance enables FinFET and FDSOI scaling to 2 nm node,” in *IEDM Tech Dig*., Dec. 2017, pp. 23.1.1-23.1.4, doi: 10.1109/IEDM.2017.8268443.

[76] IRDS 2018 Edition, 2018. [https://irds.ieee.org/editions/2018]

[77] D. Kwon, S. Cheema, Y. Lin, Y. Liao, K. Chatterjee, A. Tan, C. Hu, S. Salahuddin, "Near Threshold Capacitance Matching in a Negative Capacitance FET With 1 nm Effective Oxide Thickness Gate Stack," in *IEEE Electron Device Letters*, vol. 41, no. 1, pp. 179-182, Jan. 2020, doi: 10.1109/LED.2019.2951705.

[78] G.-M. Rignanese, X. Gonze, G. Jun, K. Cho, and A. Pasquarello,“First-principles investigation of high-K dielectrics: Comparison between the silicates and oxides of hafnium and zirconium,”*Phys. Rev. B*, vol. 69, p. 184301, May 2004. doi: 10.1103/PhysRevB.69.184301

[79] X. Zhao, D. Ceresoli, and D. Vanderbilt, “Structural, electronic, and dielectric properties of amorphous Zro2 from ab initio molecular dynamics,” Phys. Rev. B, vol. 71, p. 085107, Feb 2005. doi: 10.1103/PhysRevB.71.085107

[80] T.-J. Chen and C.-L. Kuo, “First principles study of the structural, electronic, and dielectric properties of amorphous hfo2,” *Journal of Applied Physics*, vol. 110, no. 6, p. 064105, 2011. doi: 10.1063/1.3636362

[81] D. Fischer and A. Kersch, “The effect of dopants on the dielectric constant of hfo2 and zro2 from first principles,” *Applied Physics Letters*, vol. 92, no. 1, p. 012908, 2008. doi: 10.1063/1.2828696

[82] G. Dutta, K. P. S. S. Hembram, G. M. Rao, and U. V. Waghmare,“Effects of o vacancies and c doping on dielectric properties of zro2: A first-principles study,” *Applied Physics Letters*, vol. 89, no. 20, p. 202904, 2006. doi: 10.1063/1.2388146

[83] Y. Liao, D. Kwon, S. Cheema, A. Tan, M. Kao, L. Wang, C. Hu, and S. Salahuddin, “Anomalous Subthreshold Behaviors in Negative Capacitance Transistors,” Jun. 2020. [Online]. Available: https://arxiv.org/abs/2006.02594v1

[84] Y.-H. Liao, D. Kwon, Y.-K. Lin, A. J. Tan, C. Hu, S. Salahuddin, “Anomalously Beneficial Gate-Length Scaling Trend of Negative Capacitance Transistors,” *IEEE Electron Device Letters*, vol. 40, no. 11, pp. 1860-1863, Nov. 2019, doi: 10.1109/LED.2019.2940715.

[85] B. K. Kaushik, “Chapter 1 Tunnel FET: Devices and Circuits,” in Nanoelectronics Device, Circuits and Systems 1st ed., Dutch: *Elsevier*, 2018, pp. 18-20.

[86] M.H. Na, E. J. Nowak, W. Haensch, and J. Cai, “The effective drive current in CMOS inverters,” in *IEDM Tech Dig*., Dec. 2012, pp. 121-124, doi: 10.1109/IEDM.2002.1175793.

[87] M.-Y. Kao, Y-K. Lin, H. Agarwal, Y.-H. Liao, P. Kushwha, A. Dasgupta, S. Salahuddin, and C. Hu, “Optimization of NCFET by Matching Dielectric and Ferroelectric Nonuniformly Along the Channel,” *IEEE Electron Device Letters*, vol. 40, no. 5, pp. 822-825, May 2019, doi: 10.1109/LED.2019.2906314.

[88] K. Roy, S. Mukhopadhyay and H. Mahmoodi-Meimand, "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits," in Proceedings of the IEEE, vol. 91, no. 2, pp. 305-327, Feb. 2003, doi: 10.1109/JPROC.2002.808156.

[89] K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellani, R. Chau, C.-H. Choi, G. Ding, K. Fischer, T. Ghani, R. Grover, W. Han, D. Hanken, M. Hattendorf, J. He, H. Hicks, R. Huessner, D. Ingerly, P. Jain, R. James, L. Jong, S. Joshi, C. Kenyon, K. Kuhn, K. Lee, H. Liu, J. Maiz, B. Mclntyre, P. Moon, J. Neirnck, S. Pae, C. Parker, D. Parsons, C. Prasad, L. Pipes, M. Prince, P. Ranade, T. Reynolds, J. Sandford, L. Shifren, J. Sebastian, J. Seiple, S. Simon, S. Sivakumar, P. Smith, C. Thomos, T. Troeger, P. Vandervoom, S. Williams, and K. Zawadzki, "A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging," 2007 IEEE International Electron Devices Meeting, Washington, DC, 2007, pp. 247-250, doi: 10.1109/IEDM.2007.4418914.

[90] 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," in IEEE Transactions on Electron Devices, vol. 47, no. 12, pp. 2320-2325, Dec. 2000, doi: 10.1109/16.887014.

[91] D. Esseni and A. Abramo, "Modeling of electron mobility degradation by remote Coulomb scattering in ultrathin oxide MOSFETs," in IEEE Transactions on Electron Devices, vol. 50, no. 7, pp. 1665-1674, July 2003, doi: 10.1109/TED.2003.814973.

[92] A. I. Khan, D. Bhowmik, P. Yu, S. Kim, X. Pan, R. Ramesh, S. Salahuddin, “Experimental evidence of ferroelectric negative capacitance in nanoscale heterostructures,” in Applied Physics Letters, vol. 99, no. 11, pp. 113501, Sep. 2011, doi: 10.1063/1.3634072

[93] M.-Y. Kao, H. Agarwal, Y.-H. Liao, S. Cheema, A. Dasgupta, P. Kushwaha, A. Tan, S. Salahuddin, and C. Hu, “Negative Capacitance Enables FinFET Scaling Beyond 3nm Node,” Jul. 2020. [Online]. Available: https://arxiv.org/abs/2007.14448.

[94] G. Bae, D.-I. Bae, M. Kang, S.M. Hwang, S.S Kim, B. Seo, T.Y. Kwon, T.J. Lee, C. Moon, Y.M. Chio, K. Oikawa, S. Masuoka, K.Y. Chun, S.H. Park, H.J. Shin, J.C. Kim, K.K. Bhuwalka, D.H. Kim, J. Yoo, H.Y. Jeon, M.S. Yang, S.-J. Chung, D. Kim, B.H. Ham, K.J. Park, G. Song, Y.H. Ki, M.S. Kang, K.H.

Hwang, C.-H. Park, J.-H. Lee, D.-W. Kim, S.-M. Jung, and H.K. Kang, "3nm GAA Technology featuring Multi-Bridge-Channel FET for Low Power and High Performance Applications," 2018 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, 2018, pp. 28.7.1-28.7.4, doi: 10.1109/IEDM.2018.8614629.

[95] F. I. Sakib, M. A. Hasan and M. Hossain, "Exploration of Negative Capacitance in Gate-All-Around Si Nanosheet Transistors," in IEEE Transactions on Electron Devices, doi: 10.1109/TED.2020.3025524.

[96] M. H. Lee, S.-T. Fan, C.-H. Tang, P.-G. Chen, Y.-C. Chou, H.-H. Chen, J.-Y. Kuo, M.-J. Xie, S.-N. Liu, M.-H. Liao, C.-A. Jong, K.-S. Li, M.-C. Chen, and C. W. Liu, "Physical thickness 1.x nm ferroelectric HfZrOx negative capacitance FETs," 2016 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, 2016, pp. 12.1.1-12.1.4, doi: 10.1109/IEDM.2016.7838400.

[97] Sentaurus Structure Editor User Guide, Version O-2018.06, Synopsys, Mountain View, CA, USA, Sep. 2018.

[98] IRDS 2020 Edition, 2020. [https://irds.ieee.org/editions/2020]

[99] D. Takashima, "Overview of FeRAMs: Trends and perspectives," 2011 11th Annual Non-Volatile Memory Technology Symposium Proceeding, 2011, pp. 1-6, doi: 10.1109/NVMTS.2011.6137107.

[100] S. Dünkel et al., "A FeFET based super-low-power ultra-fast embedded NVM technology for 22nm FDSOI and beyond," 2017 IEEE International Electron Devices Meeting (IEDM), 2017, pp. 19.7.1-19.7.4, doi: 10.1109/IEDM.2017.8268425.

[101] F. Ambriz-Vargas et al., “A Complementary Metal Oxide Semiconductor Process-Compatible Ferroelectric Tunnel Junction,” ACS Appl. Mater. Interfaces, Sept. 15, 2017, vol. 9, no. 15, pp. 13262-13268, doi: 10.1021/acsami.6b16173

[102] T. Mikolajick et al., “Next generation ferroelectric materials for semiconductor process integration and their applications,” Journal of Applied

Physics, Mar. 10, 2021, vol. 129, pp. 100901, doi: 10.1063/5.0037617

[103] J. Seo, J. Lee and M. Shin, "Analysis of Drain-Induced Barrier Rising in Short-Channel Negative-Capacitance FETs and Its Applications," in IEEE Transactions on Electron Devices, vol. 64, no. 4, pp. 1793-1798, April 2017, doi: 10.1109/TED.2017.2658673.

[104] A. I. Khan, C. W. Yeung, Chenming Hu and S. Salahuddin, "Ferroelectric negative capacitance MOSFET: Capacitance tuning & antiferroelectric operation," 2011 International Electron Devices Meeting, 2011, pp. 11.3.1-11.3.4, doi: 10.1109/IEDM.2011.6131532.\

[105] S. -E. Huang, P. Su and C. Hu, "S-Curve Engineering for ON-State Performance Using Anti-Ferroelectric/Ferroelectric Stack Negative-Capacitance FinFET," in IEEE Transactions on Electron Devices, vol. 68, no. 9, pp. 4787-4792, Sept. 2021, doi: 10.1109/TED.2021.3099090.

[106] J. P. Duarte, “BSIM-CMG: Standard FinFET compact model for advanced circuit design,” ESSCIRC, 2015, doi: 10.1109/ESSCIRC.2015.7313862

[107] D. T. Haar, “Collected Papers of L.D. Landau,” Elsevier, Jan. 1, 1965, eBook ISBN: 9781483152707

[108] K. Karda, “An anti-ferroelectric gated Landau transistor to achieve sub-60 mV/dec switching at low voltage and high speed,” Appl. Phys. Lett., Apr. 21, 2015, vol. 106, no. 16, pp. 163501, doi: 10.1063/1.4918649

[109] L. E. Cross, “Antiferroelectric-Ferroelectric Switching in a Simple “Kittel” Antiferroelectric,” Nov. 7, 1966, vol. 23, no. 1, pp. 77-82, doi: 10.1143/JPSJ.23.77

[110] K. Okada, “Phenomenological Theory of Antiferroelectric Transition. I. Second-Order Transition,” Apr. 12, 1969, vol. 27, no. 2, pp. 420-428, doi: 10.1143/JPSJ.27.420

[111] Sang-Hyun Oh, D. Monroe and J. M. Hergenrother, "Analytic description of short-channel effects in fully-depleted double-gate and cylindrical, surrounding-gate MOSFETs," in IEEE Electron Device Letters,

vol. 21, no. 9, pp. 445-447, Sept. 2000, doi: 10.1109/55.863106.

[112] K. K. Young, "Short-channel effect in fully depleted SOI MOSFETs," in IEEE Transactions on Electron Devices, vol. 36, no. 2, pp. 399-402, Feb. 1989, doi: 10.1109/16.19942.

[113] BSIM-CMG 111.1.0, 2020. [https://bsim.berkeley.edu/models/bsimcmg/]

[114] J. Wang, Y. -H. Kim, J. Ryu, C. Jeong, W. Choi and D. Kim, "Artificial Neural Network-Based Compact Modeling Methodology for Advanced Transistors," in IEEE Transactions on Electron Devices, vol. 68, no. 3, pp. 1318-1325, March 2021, doi: 10.1109/TED.2020.3048918.

[115] M. Li, O. İrsoy, C. Cardie and H. G. Xing, "Physics-Inspired Neural Networks for Efficient Device Compact Modeling," in IEEE Journal on Exploratory Solid-State Computational Devices and Circuits, vol. 2, pp. 44-49, Dec. 2016, doi: 10.1109/JXCDC.2016.2636161

[116] K. Mehta and H. -Y. Wong, "Prediction of FinFET Current-Voltage and Capacitance-Voltage Curves Using Machine Learning With Autoencoder," in IEEE Electron Device Letters, vol. 42, no. 2, pp. 136-139, Feb. 2021, doi: 10.1109/LED.2020.3045064.

[117] J. Xu and D. E. Root, "Advances in artificial neural network models of active devices," 2015 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO), 2015, pp. 1-3, doi: 10.1109/NEMO.2015.7415102.

[118] G. Alia, A. Buzo, H. Maier-Flaig, K. -W. Pieper, L. Maurer and G. Pelz, "Machine learning-based acceleration of Genetic Algorithms for Parameter Extraction of highly dimensional MOSFET Compact Models," 2021 28th IEEE International Conference on Electronics, Circuits, and Systems (ICECS), 2021, pp. 1-4, doi: 10.1109/ICECS53924.2021.9665517.

[119] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, “Scikit-

learn: Machine Learning in Python,” in JMLR, vol. 12, no. 85, pp. 2825-2830, 2011, doi: 12(85):2825−2830, 2011.

[120] A. F. Agarap, “Deep learning using rectified linear units (relu),” arXiv:1803.08375, 2018.

[121] J. Lorenz et al., "Simultaneous simulation of systematic and stochastic process variations," 2014 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), 2014, pp. 289-292, doi: 10.1109/SISPAD.2014.6931620.

[122] C. C. Enz and E. A. Vittoz, “Charge-Based MOS Transistor Modeling: The EKV Model for Low-Power and RF IC Design." John Willey, New York (2006).

[123] G. Gildenblat et al. "PSP Model." Department of Electrical Engineering, The Pennsylvania State University and Philips Research, Aug. 2005

[124] P. G. A. Jespers and B. Murmann, “Systematic Design of Analog CMOS Circuits,” Cambridge University Press, 2017.

[125] L. Zhang and M. Chan. "Artificial neural network design for compact modeling of generic transistors," in Journal of Computational Electronics, vol. 16, no. 3, pp. 825-832, Sep. 2017, doi: 10.1007/s10825-017-0984-9

[126] H. B. Hammouda, M. Mhiri, Z. Gafsi, and K. Besbes, ‘‘Neural-based models of semiconductor devices for SPICE simulator,’’ in American Journal of Applied Sciences, vol. 5, no. 4, pp. 785–791, Apr. 2008, doi: 10.3844/ajassp.2008.385.391

[127] Fang Wang and Qi-Jun Zhang, "Knowledge-based neural models for microwave design," in IEEE Transactions on Microwave Theory and Techniques, vol. 45, no. 12, pp. 2333-2343, Dec. 1997, doi: 10.1109/22.643839.

[128] C. Hu, “BSIM-making the first international standard MOSFET model,” in Science in China Series F: information Sciences, vol: 51, pp 765-773, May 2008, doi: 10.1007/s11432-008-0053-x

[129] D. B. M. Klaassen, "A unified mobility model for device simulation," International Technical Digest on Electron Devices, 1990, pp. 357-360, doi: 10.1109/IEDM.1990.237157.

[130] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).

[131] C. C. Mcandrew, "Validation of MOSFET model Source–Drain Symmetry," in IEEE Transactions on Electron Devices, vol. 53, no. 9, pp. 2202-2206, Sept. 2006, doi: 10.1109/TED.2006.881005.

[132] G. Templeton, “What is Moore’s Law?,” EXTREMETECH , July 29, 2015, [https://www.extremetech.com/extreme/210872-extremetech-explains-what-is-moores-law]