Nanoscale CMOS Modeling

Mohan Vamsi Dunga

Electrical Engineering and Computer Sciences
University of California at Berkeley

March 3, 2008

Nanoscale CMOS Modeling

by

Mohan Vamsi Dunga

B. Tech. (Indian Institute of Technology, Bombay) 2001
M. S. (University of California, Berkeley) 2004

A dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Engineering - Electrical Engineering
and Computer Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Ali M. Niknejad, Chair
Professor Chenming Hu
Professor Peter Yu

Spring 2008

The dissertation of Mohan Vamsi Dunga is approved:

Professor Ali M. Niknejad, Chair   Date

Professor Chenming Hu   Date

Professor Peter Yu   Date

University of California, Berkeley

Spring 2008

Nanoscale CMOS Modeling

Copyright © 2008

by

Mohan Vamsi Dunga

Abstract

Nanoscale CMOS Modeling

by

Mohan Vamsi Dunga

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Ali M. Niknejad, Chair

Since its inception almost four decades ago, the conventional planar bulk silicon MOSFET has been scaled relentlessly in accordance with the Moore's Law. However, as the state-of-the-art MOSFET makes inroads into the nanoscale regime, the traditional scaling solutions are being confronted with fundamental physical limitations stunting the rate of CMOS scaling. The insatiable need for scaled CMOS, primarily driven by the economics of computing market, is forcing researchers world-wide to seek scaling solutions in the form of alternative MOSFET structures and new materials for conventional bulk silicon MOSFET. Scaling also introduces new electrical behavior into MOSFETs which had hitherto been unknown or at the least imperceptible. Compact models describing the physics and operation of state-of-the-art bulk planar bulk MOSFETs and new CMOS scaling alternatives are imperative not only for short-term technology design and circuit-level explorations but also for long term product development. The goal of this dissertation is to develop accurate and efficient compact models for emerging nanoscale CMOS devices through a sound un-

1

derstanding of the underlying physics.

New MOSFET architectures - Multi-Gate FETs (MG-FETs) which employ the use of multiple gate electrodes to thwart the deleterious short channel efforts in scaled transistors hold promise to scale CMOS beyond 22nm technology node. The multiple gates in a MG-FET can be electrically interconnected as in FinFETs or can be biased independently as in independent MG-FETs. Surface-potential based compact models describing the electrical characteristics - terminal currents, charges and capacitances - are developed for both categories of MG-FETs. Full scale compact models are developed to describe the start-of-the-art MG-FET technologies.

While new MOSFET architectures are being evaluated, the traditional cost-effective single-gate bulk planar CMOS is still breaking scaling and performance barriers through the use of process-induced strain. A new non-process-specific layout-dependent mobility model for mobility enhancement through process induced strain is developed to improve the modeling of state-of-the-art bulk MOSFETs. Furthermore, the scaled MOSFETs are experiencing increased variability in low-frequency noise characteristics. A thorough understanding of the underlying physics is presented together with a new statistical model for modeling the low frequency noise in scaled MOSFETs.

Professor Ali M. Niknejad, Chair   Date

2

Acknowledgements

I would like to express my sincere gratitude to my advisor Professor Ali M. Niknejad without whose guidance, this work would not have been possible. He always encouraged me to explore new ideas and provided me with a patient ear to listen to even the craziest of them. His in-depth analysis of problems always helped me to transform mere specks of thought into concrete ideas and overcome roadblocks. His vast range of knowledge has inspired me to reach out and seek new fields beyond my research field. I am also highly indebted to Professor Chenming Hu for being my co-advisor and for his constant guidance through his work. The numerous brainstorming sessions with him helped me not only embolden my knowledge but also cultivate new ideas. I am thankful to both Professor Ali M. Niknejad and Professor Chenming Hu for their constant encouragement and endless support over the years which has made my stay both enjoyable and intellectually satisfying.

I would like to thank Professor Peter Yu for teaching Physics 250 in Spring 2004 and for his valuable feedback as a member of my dissertation committee. I owe my "limited" understanding of Solid State Physics to Professor Peter Yu which has proved to be useful time and again during the course of this work.

I am grateful to Dr. Raj Raghuram at Berkeley Design Automation for giving me the opportunity to apply my knowledge as a summer intern with lot of learning about device models and their use in simulators.

I would like to thank industry collaborators for their invaluable support and critical feedback. I would like to thank TI (Keith Green, Claude Cirba, Wade Xiong),

Dedicated to my family

Contents

Chapter 1

Introduction

1.1 CMOS Scaling Challenges

Silicon-based microelectronic devices have revolutionized our world in the past four decades. The need for higher computing power at cheaper cost has fueled incessant CMOS scaling [1]. It all started with the invention of integrated circuit in late 1950's that unveiled the possibility of using transistors in almost all kinds of electronic circuits [2], [3]. The next major breakthrough came with the demonstration of the first metal-oxide semiconductor field-effect transistor (MOSFET) in 1960 by Kahng and Atalla which would enable cost-effective integration of large number of transistors with interconnections on a single silicon chip [4]. Five years later, Gordon Moore made the very important observation that the number of components on minimum-cost integrated circuits had increased roughly by a factor of two per year [5] which then later transformed itself into a law known as the Moore's Law [6]. Defying all
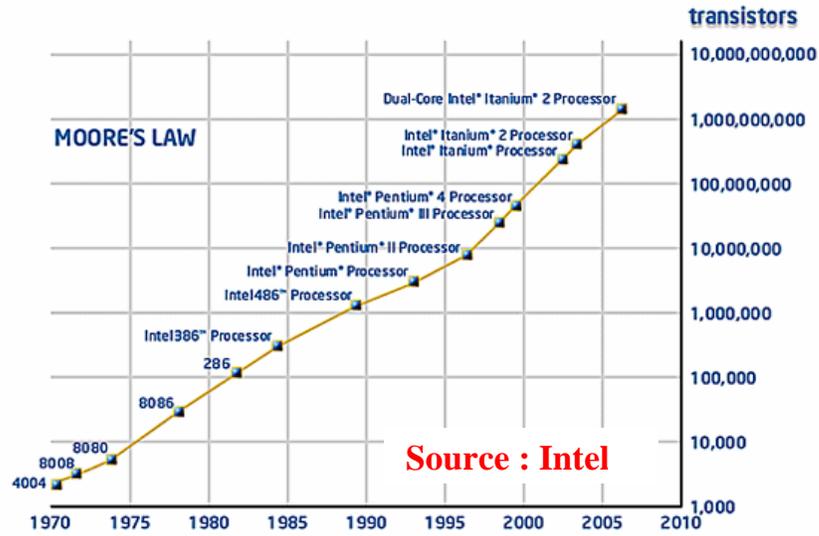
Figure 1.1: Moore's Law in microprocessors.

rumors of its end, the Moore's Law holds even 40 years later and it has become more like a self-fulfilling prophecy. The industry tries to attain the goal set by Moore's Law presuming that the goal will be attained by its competitors at any cost. Fig. 1.1 shows the Moore's Law with regard to the number of transistors in Intel's microprocessors. A modern day microprocessor has about a billion transistors.

Moore's Law is achieved primarily by scaling the transistor dimensions by a factor of 2 every 3 years. The scaling of transistor has several advantages in addition to increasing the on-chip transistor density [7], [8]. Delay of the logic gates decreases and the operating frequency of the transistors increases by a factor of $1/L_g$ (L_g is transistor gate length) enabling faster operation of the circuits. For same amount of functionality, the chip area decreases by a factor of $1/L_g^2$ which allows more dies on a single wafer resulting in large cost savings. Furthermore, since the die size is smaller,

Parameter	Result of Scaling
Dimensions	1/K
V_{DD}	1/K
Fields	1
Current	1/K
Capacitance	1/K
Delay Time	1/K
Power/Circuit	$1/K^2$
Power/Area	1

Figure 1.2: Impact of constant field scaling on various transistor and circuit performance metrics. (K is scaling factor)

the amount of defects/die is also small leading to a higher yield in manufacturing. The active switching power per area remains constant with technology scaling allowing the circuits to run at lower power or allowing the circuits to incorporate more functionality under a fixed power constraint. Planar bulk silicon MOSFET has been the traditional workhorse of the semiconductor industry to achieve constant scaling. However, the bulk planar FET is facing severe challenges for scaling beyond 32nm node as many of the scaling advantages listed in Fig. 1.2 are disappearing, particulary with respect to power density.

Fig. 1.3 shows the schematic of conventional planar bulk transistor. Traditionally, as the channel length L_g is decreased, the oxide thickness T_{ox} is reduced commensurately to maintain strong capacitive coupling between the inversion channel and the
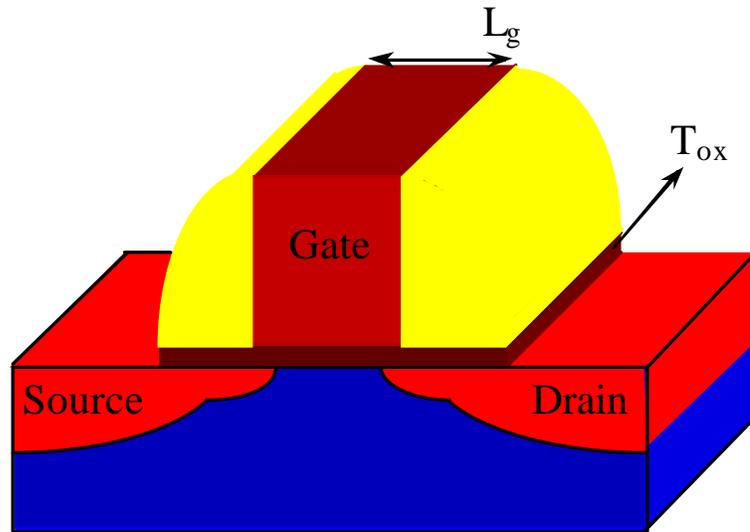
Figure 1.3: Schematic of the conventional planar bulk transistor.

gate terminal compared to other transistor terminals. This helps to control the short

channel effects such as threshold voltage roll-off, sub-threshold slope degradation and

drain induced barrier lowering (DIBL) which act together to increase the off-state

transistor leakage. Scaling of L_g is also accompanied by an corresponding increase

in body doping N_A and decrease in source/drain junction depth X_j to reduce the

leakage path below the inversion channel. This scaling methodology has worked very

well for several decades but in the recent past the conventional CMOS technology is

approaching physical limits whereby further scaling is beginning to have a negative

impact on the transistor performance [9], [10].

The thin gate oxide T_{ox} loses its ideal insulating property and there is significant

increase in gate current through direct quantum tunneling of carriers across the ox-

ide. This results in an increase in the on-state leakage. The increase in body doping

leads to reduction in carrier mobility due to the large vertical electric field and the enhancement in the current drive due to transistor scaling decreases. The large vertical field created at source/drain junction due to the high body doping also leads to large band-to-band tunneling current and increased off-state leakage. Increased body doping also leads to larger body and junction capacitances making the transistor switching slower. In scaled transistors, the random device-to-device variability in the form of random dopant fluctuations is also exacerbated due to the presence of heavy body doping [11]. The reduced junction depths increase the source/drain resistance which reduces the current drive of the transistor.

To solve these scaling issues, two different paths can be chartered. The first path involves introduction of new technologies and new materials into the conventional planar bulk MOSFET to allow further scaling and boost the performance of scaled transistors. The second path involves adoption of new transistor architectures such as ultra-thin body FETs and multi-gate FETs which inherently have superior electrostatic control over the inversion channel. In order to study the circuit level benefits and reliability of these two CMOS scaling solutions, timely understanding and modeling of the associated device physics and device behavior is needed.

1.2   Advanced planar bulk MOSFETs

Even though planar bulk MOSFET is getting constrained by the physical limits of scaling enlisted in Section 1.1, researchers worldwide are trying to extend bulk MOSFET scaling through introduction of new materials and technologies.

In order to combat the problem of increasing gate leakage with decreasing oxide thickness, high-k insulators are being introduced into the gate stack. A high-k gate dielectric with dielectric constant higher than silicon dioxide can generate a larger gate capacitance even with a physical thickness larger than that of silicon dioxide. The larger physical thickness suppresses direct tunneling of carriers through the gate dielectric reducing the gate current by several orders of magnitude [12].

The capacitive coupling of the channel to the gate can be further boosted through the use of metal gates as gate electrode instead of poly-Si. Poly-Si gate electrode suffers from poly-depletion in strong inversion which increases the effective insulator thickness and reduces the gate control over channel. Metal gates do not exhibit any poly-depletion and reduce the equivalent oxide thickness in the total gate capacitance at inversion by ≈ 0.4nm which is significant for scaled transistors. Transistors with poly-Si gate and high-k gate stack show reduced carrier mobility due to additional phonon scattering mechanisms. The use of metal gates for high-k dielectric gate stack improves the carrier mobility close to that of the conventional SiO_2/poly-Si stack by screening the high-k's surface phonons from coupling to the inversion channel [13].

The series resistance of ultra-shallow source/drain (S/D) junctions can degrade the current drive. Use of elevated S/D regions and metal S/D can improve the series resistance. In a metal S/D Schottky-barrier FET, typically a metal silicide replaces the S/D impurity doping to improve the series resistance. These techniques can be used for non-classical CMOS structures also such as the multi-gate FETs.

One of the most important performance boosters for modern day planar bulk MOSFET is process induced strain [14]. New materials are being introduced into the

bulk transistor to create uniaxial and biaxial stress in the inversion channel leading to higher electron and hole mobilities and hence higher current drives. Process-induced strain enhances carrier transport to solve the speed/power trade-off. While scaling improves the transistor speed at the expense of higher off-state leakage, process-induced strain increases CMOS performance without any detrimental effect on transistor performance. However, in extremely scaled CMOS, the performance of a given transistor is influenced by stress from neighboring transistors which introduces systematic variability in CMOS design.

The aggressive scaling of FETs is introducing variability in transistor performance [15]. The variability can be systematic or random. Lithography hotspots, non-uniformity of equipment and layout depending stress effects are some examples of systematic variations. Examples of random variations are random dopant fluctuations, flicker noise and line edge roughness. Even though two transistors may be exactly identical in terms of their layout, there are always minor differences between them. For example, the number of traps and their locations in the gate dielectric or the number of dopant atoms and their locations in the silicon body may differ from transistor to transistor. In large area devices, these differences average out but for small area devices, these differences can cause significant change in performance between transistors with exactly same layout. Modeling and characterization of variability in scaled CMOS poses an interesting challenge.
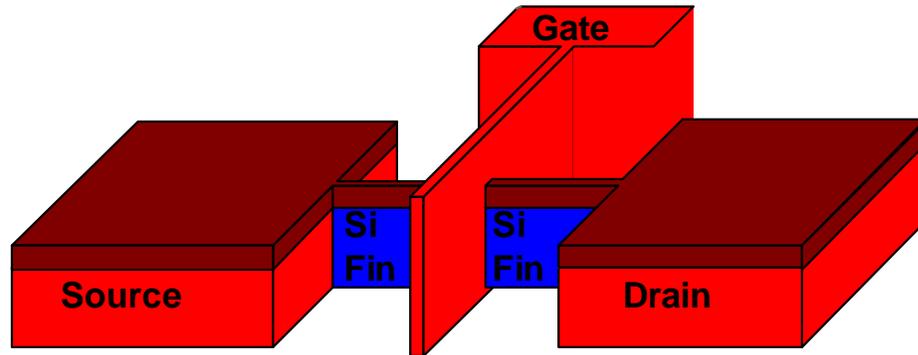
Figure 1.4: Schematic of a FinFET. The gate wraps around the vertical Si fin inducing electrostatic control from the opposite sides of silicon fin.

1.3   Multi-Gate FETs

A promising alternative to extend CMOS scaling beyond 32nm node is the introduction of a new FET architecture - the Multi-gate (MG) FET [16]. MG-FETs can be thought of as an extension to the ultra-thin body FETs or the fully-depleted SOI FETs but with more gates around the thin silicon body. MG-FETs offer stronger electrostatic control of the inversion channel through the use of multiple gates. This reduces the detrimental short channel effects (SCE) and makes the MG-FETs more scalable than the planar bulk CMOS [17], [18]. One of the most feasible multi-gate configurations in terms of being manufacturable is the FinFET. Fig. 1.4 shows the FinFET structure where the gate controls the channel along the silicon sidewalls of the fin.

The good control of SCE in MG-FETs allows the silicon body to be lightly doped compared to single-gate FETs. The reduced body doping results in lower average electric field in the channel which translates to an improvement in carrier mobility,
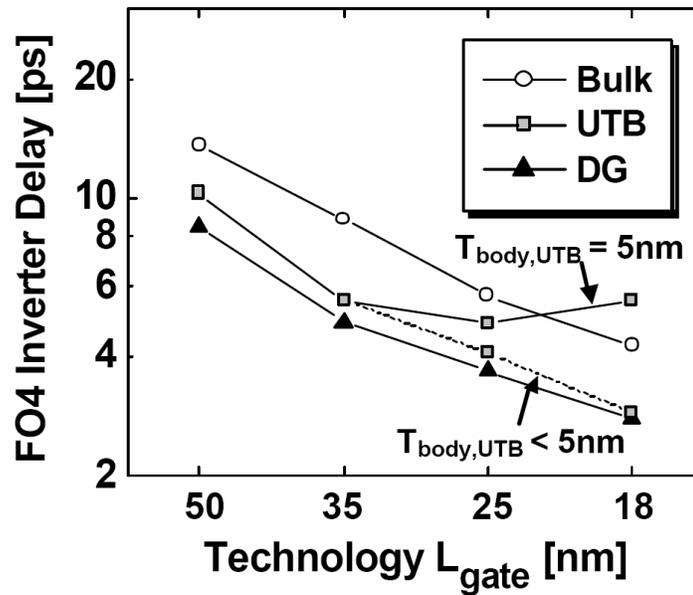
Figure 1.5: DG-FETs show better logic gate delay compared to ultra-thin body (UTB) FETs and planar bulk FETs [19].

gate leakage currents and device reliability. The use of less channel doping also helps to reduce random dopant fluctuations in MG-FETs and hence decreases the variability in device performance which will be critical for some circuit designs such as SRAM arrays. The combination of light body doping and thin body yields steeper sub-threshold swing and lower junction and body capacitance. Because of these additional benefits, MG-FETs show better logic delay than the planar bulk devices (Fig. 1.5).

MG-FETs come in several flavors as shown in Fig. 1.6. The silicon body can be controlled by either two gates or three gates or four gates. The more the number gates the higher the electrostatic control of channel but there is a trade-off with the corresponding process complexity. The gates in a MG-FET can all be electrically
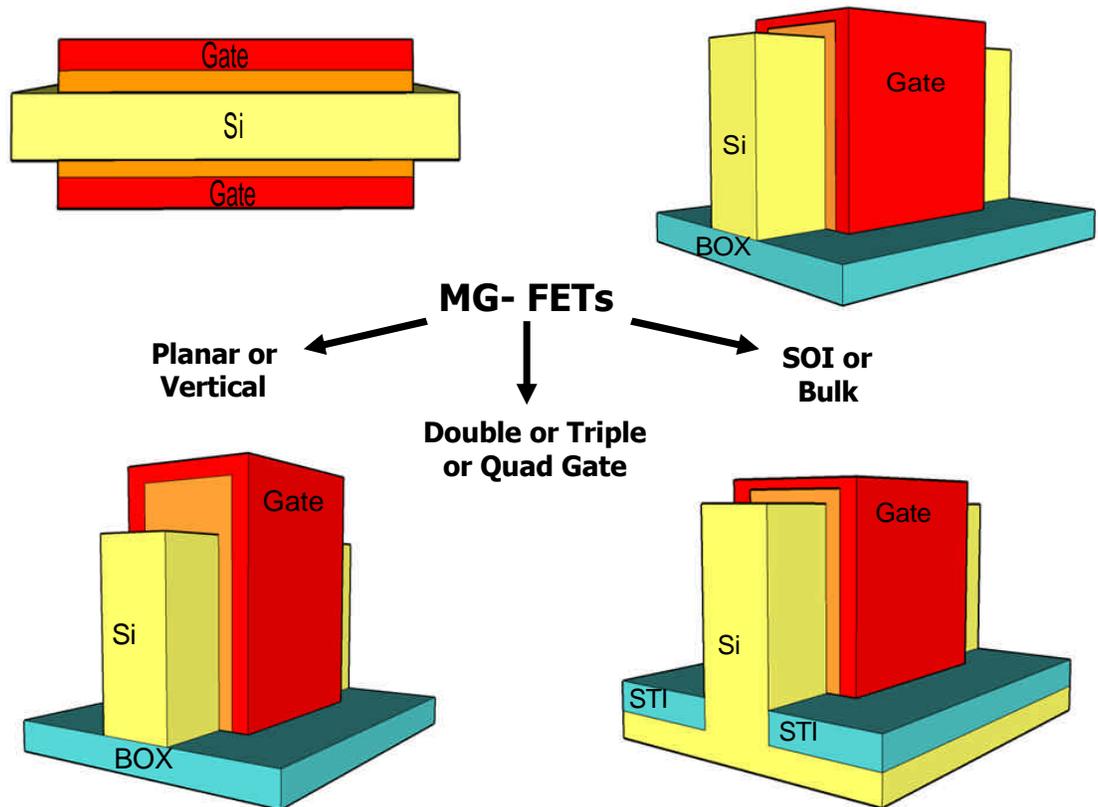
Figure 1.6: Different possible configurations of MG-FETs.

interconnected or they can be biased independently. The independent biasing has been demonstrated with double-gate FETs where the back gate is used to set the threshold voltage of the device while the front gate is used to switch the device. Independent biasing also opens a wide range of applications where both the gates can be switched independently to couple two different time varying signals in silicon body and achieve applications such as frequency mixing or two input logic functions. Furthermore, MG-FETs can be built on SOI or bulk silicon. Bulk MG-FETs are cost

effective compared to SOI MG-FETs and allow the possibility of coexistence with the conventional single-gate bulk FETs on the same wafer. The existence of several MG-FETs configurations together with the control of inversion channel by multiple gates makes modeling of MG-FETs quite challenging.

1.4   Compact MOSFET Models

All integrated circuit designs - digital or analog or mixed-signal, are designed and verified through the use of circuit simulators before being reproduced in real silicon. The origin of circuit simulators can be traced to 1970's when the circuit simulation tool SPICE (Simulation Program with Integrated Circuit Emphasis) was developed at University of California at Berkeley to aid the growing IC industry. Since then, many advanced circuit simulators have been developed such as HSPICE, SPECTRE, etc. with a single goal of reducing the time-to-market for IC designs. By simulating the circuits, the designers can detect errors early in the design process and avoid the costly and time consuming prototyping.

In order for any circuit simulator to predict the performance of a CMOS design, it should have accurate models to describe the behavior of the transistors in the circuit. The device models are innate to all the circuit simulators and they ultimately decide the accuracy of the circuit performance predicted by the circuit simulator. They form a bridge between the design world and the manufacturing world with which the designers can design circuits without worrying about the intricate details of transistor fabrication. The goal of a device model is to accurately predict the

transistor's electrical characteristics - terminal currents, charges and capacitances for a given transistor bias and transistor temperature.

The device models can be classified into three broad categories. The first category comprises of numerical device models. These models predict the device behavior by solving the carrier transport equations numerically. When using such models, the circuit analysis will be extremely slow and hence is not suited for any design with a large number of transistors. The second kind of device models are the table look-up models. As the name suggests, they look-up the value of the transistor characteristics from a table. The table of transistor characteristics is constructed either using measured data or from analytic compact models (which are the third category of device models). To cover a large range of transistor geometry, bias and temperature, extensive measurements and large set of tables are needed. Table look-up models constructed from measured data are limited to the measured devices and are non-predictive. Table look-up models built using analytic compact models find use in fast circuit simulators. The final category of device models are the Compact Models which are the subject of interest here.

A compact model is a concise mathematical description of the complex device physics in the transistor. The compact models are desired to be physical in nature. An accurate model stemming from physics basis allows the process engineer and circuit designer to make projections beyond the available silicon data (scalability) for scaled dimensions and also enables circuit/device co-optimization. A compact model maintains a fine balance between the amount of detailed physics embedded for model accuracy and model compactness (computational efficiency). The simplifications in

the physics enable very fast analysis of device/circuit behavior when compared to the much slower numerical based TCAD simulations. Compact models have been at the heart of CAD tools for circuit design over the past decades, and are playing an ever increasingly important role in the nanometer system-on-chip (SOC) era.

Fig. 1.7 shows the development cycle for a compact model. The first step is to analyze the device behavior by looking at the embedded physics. TCAD simulations can serve as a very useful step in this step. It is also beneficial to examine the measured data from state-of-the-art devices which may reveal new physical phenomenon. The next step is to derive compact mathematical equations to capture the physics of the device. These equations may be verified against TCAD simulations for model accuracy and scalability. In a real device, quantities such as the doping profiles, junction depths, etc. are very complex. In order to precisely model the effect of quantities such as these, physics and technology related model parameters are added to the model. This forms a very important step in model formulation as the ultimate goal of a compact model is to describe any given transistor technology accurately. The model parameters allow to obtain a good description of technology by aiding in data fitting. Equally important is to develop a methodology to extract the value of these model parameters. A model without enough flexibility and without the ease of parameter extraction is virtually useless for real world circuit design. Once the model equations are ready, the next step is to examine the numerical robustness of the model. The model should not have any convergence issues and should be computationally efficient. This may involve modifying some of the physics based equations to accelerate the model computation. The last step in model development cycle is
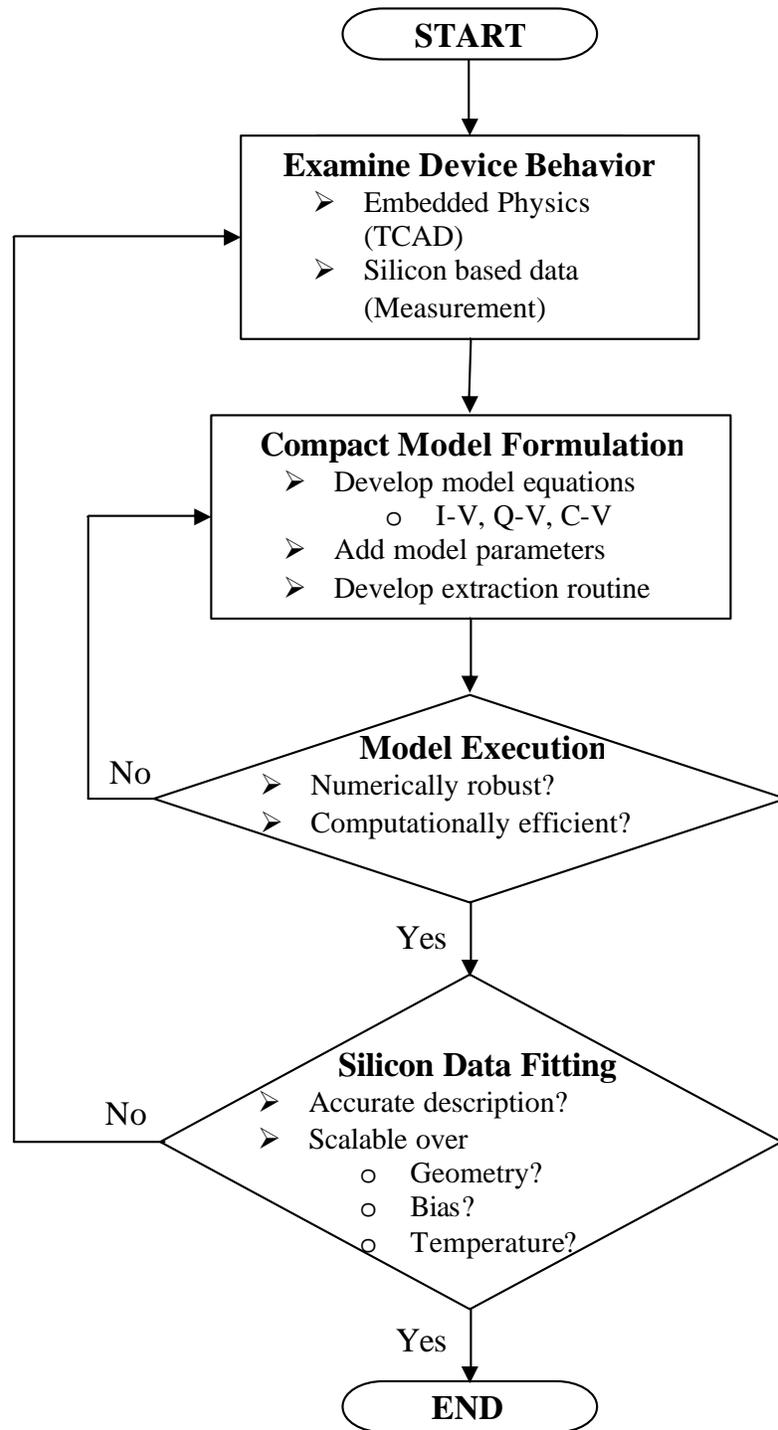
Figure 1.7: Step-by-step guide for developing a Compact Model

14

verification against silicon data. The inability to adequately describe the measured data may require modifying of the physics based equations and/or introduction of new model parameters. Successful description of silicon data over different geometry, bias and temperature marks the completion of model development. It is always preferable to verify the model against more than one technology since the compact model is a universal model which is not tied to any one specific technology.

As the mainstream MOS technology is scaled into the nanometer regime, development of a truly physical and predictive compact model for circuit simulation that covers geometry, bias, temperature, DC, AC, RF, and noise characteristics becomes a major challenge. Some of the popular existing MOS compact models are BSIM3 and BSIM4 [20], PSP [21], HiSIM [22] and EKV [23]. As the bulk planar CMOS technology advances, new physical effects and features need to be incorporated into the existing bulk MOSFET compact models. Modeling of new CMOS architectures such as multi-gate FETs will however require development of completely new compact models. Interestingly, as the CMOS technology is scaled down, the associated compact models become more and more sophisticated reflecting the ever-increasing complexity of the advanced nanoscale CMOS devices.

1.5 Dissertation Goals and Outline

In this dissertation, compact models are developed for nanoscale CMOS technology advancements to aid technology/circuit development in the short term and product design in the longer term. The models are implemented in popular circuit simula-
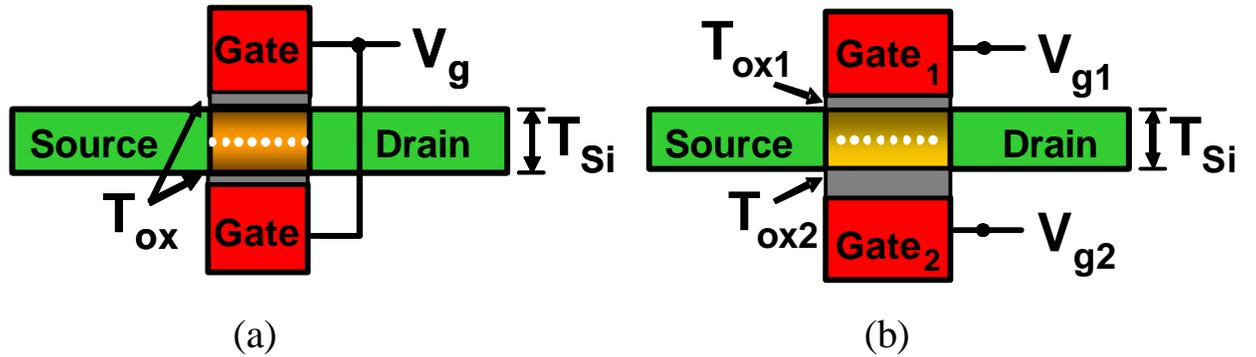
Figure 1.8: The different configurations MG-FETs can be categorized into (a) Common Symmetric MG-FET and (b) Independent MG-FET. Models developed for these two configurations can be extended easily to encompass all the different flavors of MG-FETs.

tors and verified against 2-D TCAD simulations and experimental data whenever available.

The first part of this dissertation models the new FET architecture - Multi-gate FET. Section 1.3 introduced the different flavors of MG-FETs. It is important to obtain a versatile model which can model all the different types of MG-FETs without making the model computationally intensive. One possible technique to handle the different MG-FET architectures is to classify them into two categories as shown in Fig. 1.8 and introduce a separate model for each category: a common symmetric multi-gate model (BSIM - Common Multi-Gate : BSIM-CMG) and an independent multi-gate model (BSIM - Independent Multi-Gate : BSIM-IMG) . The term "common" in common multi-gate model means that all the gates in the MG-FET are electrically interconnected and are biased at the same electrical gate voltage. On the other hand, the independent gate model permits the gates of the MG-FET to be biased independently at different gate voltages.

Chapter 2 describes the surface potential calculation for common symmetric multi-gate FET. The surface potential calculation incorporates the effect of body doping, quantum confinement and poly depletion.

The currents and capacitances for common symmetric MG-FET are formulated in Chapter 3. Chapter 3 also presents the verification of BSIM-CMG model against both 2-D TCAD simulations and experimental data.

In Chapter 4, the BSIM-IMG model is formulated. Comprehensive modeling and verification of the drain current and terminal capacitances for independent multi-gate FET are presented in this chapter.

In the second part of this dissertation, new technologies and new device behavior in advanced planar bulk MOSFET are modeled. The models improve the existing bulk MOSFET model - BSIM4 which is widely used in the industry. The models developed in this part for planar bulk MOSFET are easily extendable to MG-FETs as well.

In Chapter 5, a new non-process specific layout dependent mobility model for mobility enhancement through process induced stress is described. The model is verified against 3-D process simulations for several process induced strain technologies.

Chapter 6 presents a statistical model to capture the device-to-device variations in flicker noise in small area devices. This statistical model is also the first instance of modeling variability in BSIM4 model.

An overall summary of this dissertation is presented in Chapter 7. Chapter 7 highlights the key research contributions and future research directions are suggested.

Chapter 2

Common Symmetric Multi-Gate FET

- Surface Potential Model

2.1 Introduction

As the planar MOSFETs are scaled to sub-45nm channel lengths, achieving a large current drive while maintaining a low off-state leakage becomes challenging [24]. To solve this problem new MOSFET architectures involving the use of multiple gates controlling the transistor have been proposed [25]. Simultaneous efforts are needed in the areas of fabrication and compact modeling to investigate the device and circuit level performances of Multi-Gate FETs (MG-FETs).

A new compact model BSIM-CMG is developed to model the broad category of MG-FETs called the Common Symmetric Multi-Gate FETs (CMG-FET). CMG-FETs are MG-FETs with all the gate electrodes having the same work-function and

all the gate insulators having the same dielectric thickness. Furthermore, the gates on the two, three or four active sides of the silicon fin in the CMG-FETs are all electrically interconnected and are biased at the same electrical gate voltage.

The model BSIM-CMG is a surface potential based model. For any surface potential based model, all the electrical variables, be it the terminal currents (ex. I_{ds}, I_g, etc.) or the terminal charges (ex. Q_g, Q_s, etc.), are either an explicit or implicit function of the surface potentials at the source and drain ends of the transistor. As a result, calculation of the surface potential is the foremost step in formulating a surface potential model and it forms the basis of the entire model. This holds true for BSIM-CMG model as well, where all the electrical variables are an explicit function of the surface potential at the source and drain end.

The BSIM-CMG model starts from a common symmetric DG-FET (CDG-FET) framework and extends it to MG-FETs with more than two gates. An accurate solution to the surface potential for a long channel CDG-FET is core to the model. There have been some efforts in the recent past to calculate the surface potential for the CDG-FET but they are all restricted to transistors with an undoped silicon body [26], [27]. This chapter presents a new surface potential solution for the CDG-FET which takes into account the effect of the body doping on the surface potential.

Section 2.2 and 2.3 describe the surface potential solution for common symmetric DG-FETs with finite body doping. The resulting surface potential equation is numerically complex. Though it can be solved using pure numerical techniques, from the viewpoint of implementing it an a compact model, it is not a desirable approach. For developing a numerically robust compact model, an analytic approximate solu-

tion of the implicit surface potential equation is derived and presented in section 2.4. The elegance of the surface potential solution allows it to incorporate physical effects in MG-FETs such as the poly-depletion effect (PDE) and the quantum mechanical effects (QME). This will be described in Section 2.5.

2.2 Surface Potential Model Formulation

Long channel common symmetric DG-FET forms the core of BSIM-CMG. Fig. 2.1 shows the schematic of the long channel CDG-FET under study. The convention for the axes and the symbols used in this chapter are indicated in Fig. 2.1. The electronic potential in the silicon body is obtained by solving the 2-D Poisson's equation. For a long channel transistor, gradual channel approximation is used which states that the horizontal electric field is much smaller than the vertical electric field. The use of gradual channel approximation simplifies the 2-D Poisson's equation into a 1-D Poisson's equation (in the vertical dimension) in the silicon body.

The 1-D Poisson's equation including both inversion carriers and bulk charge in the body can be written as

$$\frac{\partial^2 \psi(x,y)}{\partial x^2} = \frac{q n_i}{\epsilon_{Si}} \cdot e^{\frac{q(\psi(x,y) - \phi_B - V_{ch}(y))}{kT}} + \frac{q N_A}{\epsilon_{Si}} \tag{2.1}$$

where $\psi(x, y)$ is the electronic potential in the body, $V_{ch}(y)$ is the channel potential ($V_{ch}(0) = 0$ and $V_{ch}(L) = V_{ds}$), N_A is the body doping and

$$\phi_B = \frac{kT}{q} \cdot \ln\left(\frac{N_A}{n_i}\right) \tag{2.2}$$
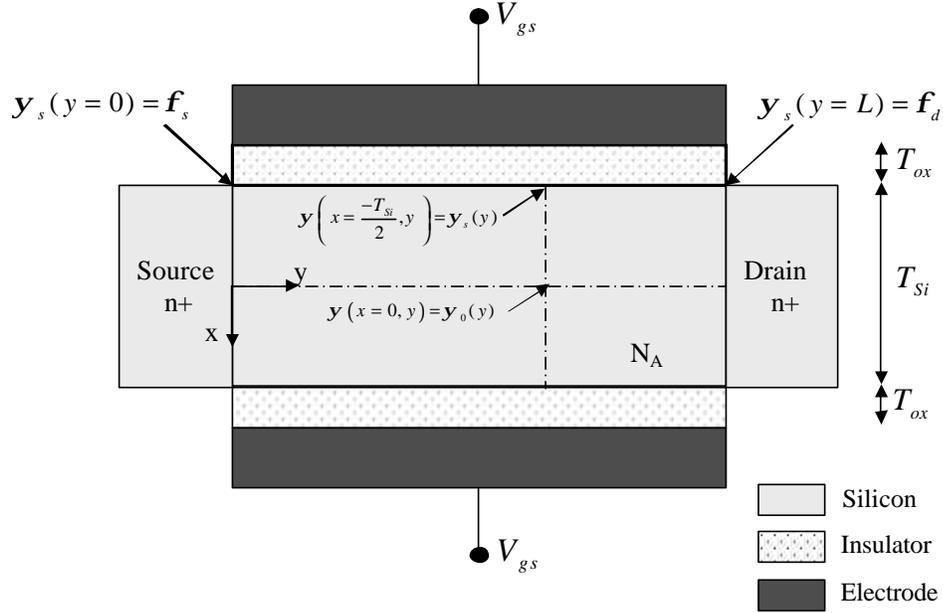
Figure 2.1: Schematic of the common symmetric DG-FET under study.

For a lightly doped body, the body doping can be neglected and Eq. 2.1 can be solved easily by integrating twice and using Gauss law as the boundary condition [26]. However, for moderate to heavy body doping, the doping term in the Poisson's equation cannot be neglected and it complicates the calculation of surface potential as Eq. 2.1 cannot be integrated analytically twice. To overcome this limitation, perturbation approach is used to solve the Poisson's equation in the presence of significant body doping [28]. Under the perturbation approach, the potential in the body can be written as sum of two terms.

$$\psi(x,y) = \psi_1(x,y) + \psi_2(x,y) \tag{2.3}$$

The first term $\psi_1(x,\,y)$ is the potential due to the inversion carriers term in Eq. 2.1. The second term $\psi_2(x,\,y)$ is the perturbation in potential due to the body doping term. The silicon body can be fully depleted or partially depleted depending on applied gate bias (V_{gs}), body doping (N_A) and body thickness (T_{Si}). The perturbation method yields surface potential in both full-depletion and partial-depletion regimes.

In the fully depleted regime, the inversion carriers are spread through the entire body. The contribution of inversion carriers to the potential, $\psi_1(x,\,y)$, is calculated by neglecting the bulk charge term in Eq. 2.1.

$$\frac{\partial^2 \psi_1(x,y)}{\partial x^2} = \frac{q n_i}{\epsilon_{Si}} \cdot e^{\frac{q(\psi_1(x,y) - \phi_B - V_{ch}(y))}{kT}} \tag{2.4}$$

For the symmetric common DG-FET, the vertical electric field at the center plane of the silicon body is zero. As a result, one of the boundary conditions for Eq. 2.4 can be written as

$$\frac{\partial \psi_1(x,y)}{\partial x}\Big|_{x=0} = 0 \tag{2.5}$$

Using the boundary condition Eq. 2.5, $\psi_1(x,\,y)$ can be obtained by integrating Eq. 2.4 twice.

$$\psi_1(x,y) = \psi_0(y) - \frac{2kT}{q}\ln\left(\cos\left(\sqrt{\frac{q^2}{2\epsilon_{Si}kT}\frac{n_i^2}{N_A}e^{\frac{q(\psi_0(y)-V_{ch})}{kT}}} \cdot x\right)\right) \tag{2.6}$$

where $\psi_0(y)$ is the potential at the center of the body as shown in Figure 2.1.

The perturbation in the electronic potential due to the body doping, $\psi_2(x,\,y)$,

will now be determined. Substituting Eq. 2.3 and Eq. 2.4 in Eq. 2.1 yields a second order differential equation in terms of $\psi_2(x, y)$.

$$\frac{\partial^2 \psi_2(x,y)}{\partial x^2} = \frac{qn_i}{\epsilon_{Si}} \cdot e^{\frac{q(\psi(x,y)-\phi_B-V_{ch}(y))}{kT}} \cdot \left(e^{\frac{q\psi_2(x,y)}{kT}} - 1 \right) + \frac{qN_A}{\epsilon_{Si}} \tag{2.7}$$

The mid-plane potential is controlled by the inversion carriers. As a result, the contribution of the bulk charge to potential at the mid-plane is zero. Since the CDG-FET is symmetric, the electric field at the mid-plane due to bulk charge is also zero.

$$\psi_2(x = 0, y) = 0 \quad \text{and} \quad \frac{\partial \psi_2(\mathrm{x, y})}{\partial \mathrm{x}}\big|_{\mathrm{x}=0} = 0 \tag{2.8}$$

Using the boundary conditions on $\psi_2(x, y)$ in Eq. 2.8 and the expression for $\psi_1(x, y)$ in Eq. 2.6, the differential equation Eq. 2.7 can be solved to obtain the perturbation potential $\psi_2(x, y)$.

$$\psi_2(x, y) = \frac{2qn_i}{\epsilon_{Si}} \cdot \frac{e^{\frac{q\phi_B}{kT}}}{a(x, y)} \left(\frac{e^{x\sqrt{a(x,y)}} - 1}{2e^{x\frac{\sqrt{a(x,y)}}{2}}} \right)^2 \tag{2.9}$$

where

$$a(x, y) = \frac{q^2 n_i}{\epsilon_{Si} kT} \cdot e^{\frac{q(\psi_1(x,y)-V_{ch}-\phi_B)}{kT}} \tag{2.10}$$

The surface potential at a point y along the surface is the sum of $\psi_1(x, y)$ and $\psi_2(x,$

y) evaluated at the surface.

$$\psi_s(y) = \psi_1\left(\frac{T_{Si}}{2}, y\right) + \psi_2\left(\frac{T_{Si}}{2}, y\right) \tag{2.11}$$

Note that $\psi_2(x, y)$ is a function of $\psi_1(x, y)$ and $\psi_1(x, y)$ is a function of $\psi_0(y)$. As a result, $\psi_s(y)$ is a function of only one variable : $\psi_0(y)$, the potential at the center of the body. The variable $\psi_0(y)$ can be determined by using the Gauss's Law at the surface as the boundary condition.

The electric field at the surface can be easily obtained by integrating Eq. 2.1 once. The Gauss's Law at the surface can then be expressed as

$$
\begin{aligned}
V_{gs} &= V_{fb} + \psi_s(y) + \\
&\frac{\epsilon_{Si}}{C_{ox}}\sqrt{\frac{2qn_i}{\epsilon_{Si}} \cdot \left(\frac{e^{\frac{q\psi_s(y)}{kT}} - e^{\frac{q\psi_0(y)}{kT}}}{q/kT} \cdot e^{\frac{-q(V_{ch}+\phi_B)}{kT}} + e^{\frac{q\phi_B}{kT}} \cdot (\psi_s(y) - \psi_0(y))\right)}
\end{aligned} \tag{2.12}
$$

For a given gate and drain bias, the only unknown quantity in Eq. 2.12 is $\psi_0(y)$. Solving Eq. 2.12 yields $\psi_0(y)$ and hence $\psi_s(y)$ in the fully depleted regime for the CDG-FET.

In the partially depleted regime, the silicon body is not fully depleted and the depletion width is bias dependent. At the edge of the depletion width, x_{dep}, the electronic potential and the vertical electric field are zero and hence

$$\psi_1(x = x_{dep}, y) = 0 \quad \text{and} \quad \frac{\partial\psi_1(x, y)}{\partial x}\Big|_{(x=x_{dep})} = 0 \tag{2.13}$$

With these new boundary conditions, the contribution to the surface potential from

inversion carriers can be derived for the partially depleted regime similar to the fully depleted regime by integrating Eq. 2.4 twice.

$$\psi_1\left(\frac{T_{Si}}{2}, y\right) = -\frac{2kT}{q}\ln\left(\cos\left(\sqrt{\frac{q^2}{2\epsilon_{Si}kT}\frac{n_i^2}{N_a}e^{\frac{-qV_{ch}(y)}{kT}}} \cdot x_{dep}\right)\right) \tag{2.14}$$

Using the fact that $\psi_2(x = x_{dep}, y) = 0$ for a partially depleted body, the perturbation in the surface potential due to the finite body doping can be expressed as

$$\psi_2\left(\frac{T_{Si}}{2}, y\right) = \frac{2qn_i}{\epsilon_{Si}} \cdot \frac{e^{\frac{q\phi_B}{kT}}}{a(y)}\left(\frac{e^{x_{dep}\sqrt{a(y)}} - 1}{2e^{x_{dep}\frac{\sqrt{a(y)}}{2}}}\right)^2 \tag{2.15}$$

where

$$a(y) = \frac{q^2 n_i}{\epsilon_{Si}kT} \cdot e^{\frac{q\left(\psi_1\left(\frac{T_{Si}}{2}, y\right) - V_{ch}(y) - \phi_B\right)}{kT}} \tag{2.16}$$

The Gauss's Law in the partially depleted regime is obtained by integrating Eq. 2.1 once using the fact that both electronic potential and vertical electric field are zero at the edge of depletion width ($x = x_{dep}$).

$$V_{gs} = V_{fb} + \psi_s(y) + \frac{\epsilon_{Si}}{C_{ox}}\sqrt{\frac{2qn_i}{\epsilon_{Si}} \cdot \left(\frac{e^{\frac{q\psi_s(y)}{kT}} - 1}{q/kT} \cdot e^{\frac{-q(V_{ch} + \phi_B)}{kT}} + e^{\frac{q\phi_B}{kT}} \cdot \psi_s(y)\right)} \tag{2.17}$$

The unknown variable in the surface potential solution of partially depleted regime is x_{dep} which is obtained by solving Eq. 2.17 numerically. Once x_{dep} is determined, the surface potential is calculated using Eq. 2.11 and Eqs. (2.14 - 2.16).

The aforementioned analytical framework yields the surface potential in both
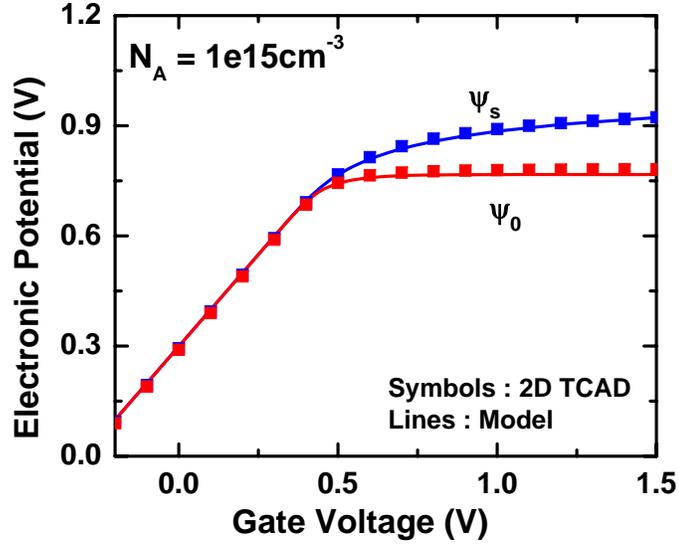
Figure 2.2: Surface potential and center potential for a lightly doped CDG-FET ($N_A =$ 1e15cm^{-3}).

partially depleted and fully depleted regimes. The framework is verified extensively against the 2-D device simulator MEDICI. All the DG-FETs used for verification have metal gates with silicon mid-gap work-function, $L_g = 1\mu$m, $T_{ox} = 2$nm and $T_{Si} = 20$nm unless specified explicitly. The surface potential is verified against 2-D TCAD simulations at the source end ($V_{ch} = 0$V) in all the cases. Surface potential is calculated using the analytical model and compared with the surface potential from MEDICI for both light and heavy body doping.

Fig. 2.2 shows the comparison of electronic potential at the surface and at the center of the body for a light body doping of $N_A = 1$e15cm^{-3}. For light body doping, since the depletion charge is negligible, the surface potential and the center potential are nearly the same in sub-threshold regime as seen in Fig. 2.2. In the strong inversion
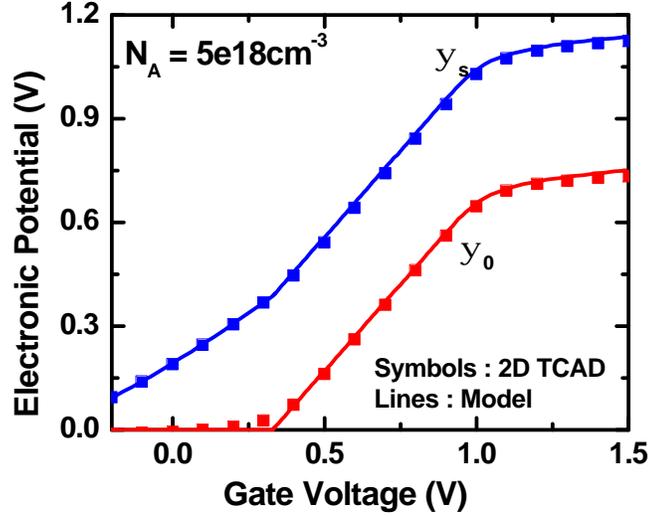
Figure 2.3: Surface potential and center potential for heavily doped CDG-FET ($N_A = $ 5e18cm^{-3}). Model captures surface potential in both partially depleted and fully depleted regimes of CDG-FET operation.

regime, as there is a build up inversion carriers at the surface, the electronic potential at the center starts to differ from the surface potential. Fig. 2.2 shows good agreement between the model and 2-D TCAD for surface potential and the average error is of the order of 0.5%.

Fig. 2.3 compares the potentials obtained from the analytical framework against 2-D TCAD simulations for CDG-FET with a heavy body doping of $N_A = $ 5e18cm^{-3}. At low gate voltages, the transistor is partially depleted and the surface potential is higher than the center potential because of the depletion charge. The transition from partially depleted regime to fully depleted regime is clearly evident at $V_{gs} \approx 0.3$V. Good agreement in the surface potential between the analytical model and 2-D TCAD is clearly evident in both the partially depleted and fully depleted regime with the
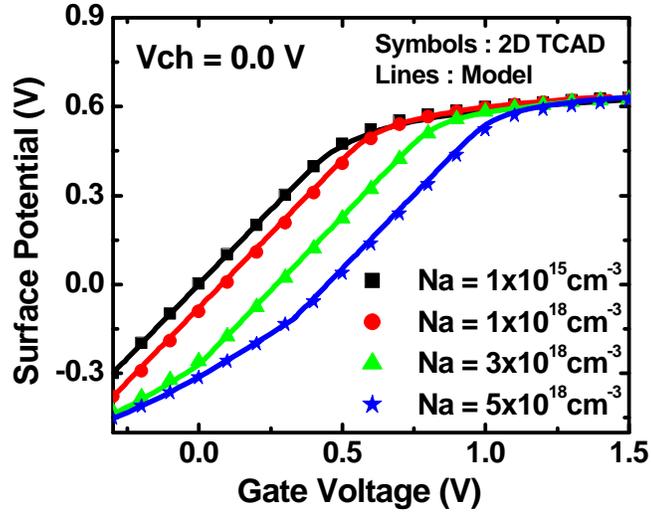
Figure 2.4: Surface potential (referenced to 2-D simulator MEDICI) for DG-FETs with different body doping. Body doping varied from light body doping of $N_A = $ 1e15cm^{-3} to a heavy body doping of $N_A = $ 5e18cm^{-3}.

average error of the order of 1%. The surface potential model is verified against 2-D TCAD for a range of body doping as shown Fig. 2.4 indicating the model's ability to capture the effect of body doping on the electrical characteristics of CDG-FET.

Though the analytical framework captures all the physics of CDG-FET correctly, Fig. 2.3 shows a discontinuity in the transition from partially depleted regime to fully depleted regime which is undesirable for implementation into a compact model. An elegant method to eliminate the discontinuity in the surface potential solution will be developed in the following section.

2.3   Unified Surface Potential Model

In order to obtain continuous expressions for terminal currents and charges, it is necessary to capture the transition between the fully depleted and partially depleted regimes in a smooth manner. Also, the solution of Eq. 2.12 and Eq. 2.17 is computationally expensive due to the complex ψ_2(x,y) perturbation term. To overcome these problems, a simplified expression for ψ_2(x,y) is derived : ψ_{pert}, which is continuous between partially depleted and fully depleted regimes. By using ψ_{pert}, the surface potential in both the regimes is calculated through a single-piece continuous equation.

An optimized high performance and/or low leakage multi-gate transistor design will use a combination of body doping and body thickness such that the level of inversion is very small when the transistor is partially depleted. As a result, in the partially depleted regime, the inversion carriers can be neglected and the 1-D Poisson equation can be written as

$$\frac{\partial^2 \psi(x,y)}{\partial x^2} = \frac{qN_A}{\epsilon_{Si}} \tag{2.18}$$

At the edge of depletion region, x_{dep}, the electronic potential and the vertical electric field are zero.

$$\psi(x = x_{dep}, y) = 0 \text{ and } \frac{\partial \psi(\text{x,y})}{\partial \text{x}}|_{(\text{x=x}_{\text{dep}})} = 0 \tag{2.19}$$

The electric field at the surface can be obtained by integrating Eq. 2.18 once from

29

the edge of depletion width to the surface.

$$\frac{\partial \psi}{\partial x}\Big|_{x=\frac{T_{Si}}{2}} = \sqrt{\frac{2qN_A}{\epsilon_{Si}} \cdot \psi_s} \tag{2.20}$$

where ψ_s is the surface potential. Since the surface potential in Eq. 2.20 is arising solely due to the bulk charge (Eqn. 2.18), it is nothing but the correction in surface potential due to finite body charge in the partially depleted regime (ψ_c). The Gauss's Law at the surface can then be written in partially depleted regime in terms of ψ_c as

$$C_{ox} \cdot (V_{gs} - V_{fb} - \psi_c) = \epsilon_{Si}\sqrt{\frac{2qN_A}{\epsilon_{Si}} \cdot \psi_c} \tag{2.21}$$

This is a simple second order equation in ψ_c which can be solved to obtain the correction in potential due to body doping.

$$\psi_c = \left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{gs} - V_{fb}}\right)^2 \tag{2.22}$$

where

$$\gamma = \sqrt{\frac{2q\,\epsilon_{si}\,N_A}{C_{ox}}} \tag{2.23}$$

Note the similarity of this expression to bulk MOSFET surface potential solution in the sub-threshold regime [29]. Eq. 2.22 indicates that the perturbation potential due to body doping increases indefinitely with increasing V_{gs}. This is not true as Eq. 2.22 is valid only in the partially depleted (PD) regime. When the DG-FET transitions

30

to the fully depleted (FD) regime, the perturbation potential reaches its maximum value and remains constant thereafter. The maximum perturbation potential (ψ_{bulk}) is given by

$$\psi_{bulk} = \frac{1}{2} \frac{qN_A}{\epsilon_{Si}} \left(\frac{T_{Si}}{2}\right)^2 \tag{2.24}$$

It is interesting to note that the perturbation potential in the PD regime (Eq. 2.22) and in the FD regime (Eq. 2.24) can also be obtained as limiting values from the analytical framework presented in section 2.2. The perturbation potential is then given by

$$\psi_{pert} = \text{MIN}\left(\psi_c, \psi_{bulk}\right) \tag{2.25}$$

To implement the perturbation potential as a smooth and continuous function in a compact model, the transition from PD to FD regime is captured through the following smoothing function.

$$\psi_{pert} = \frac{\psi_c \cdot \psi_{bulk}}{\left(\psi_c^N + \psi_{bulk}^N\right)^{1/N}} \tag{2.26}$$

where N is a smoothing parameter.

Using the perturbation approach, the net surface potential (ψ_s) can be written as

$$\psi_s = \psi_1 + \psi_{pert} \tag{2.27}$$

31

where ψ_1 is the electronic potential due to inversion carriers

$$\psi_1 = \psi_0 - \frac{2kT}{q} ln \left(cos \left(\sqrt{\frac{q^2}{2\epsilon_{Si}kT} \frac{n_i^2}{N_A} e^{\frac{q(\psi_0 - V_{ch})}{kT}}} \cdot \frac{T_{Si}}{2} \right) \right) \qquad (2.28)$$

and ψ_0 is the electronic potential at the center of the body. Surface potential is then calculated by solving the Guass's law at the surface.

$$V_{gs} = V_{fb} + \psi_s + \frac{\epsilon_{Si}}{C_{ox}} \sqrt{\frac{2qn_i}{\epsilon_{Si}} \left(\frac{e^{\frac{q\psi_s}{kT}} - e^{\frac{q\psi_0}{kT}}}{q/kT} e^{\frac{-q(V_{ch} + \phi_B)}{kT}} + e^{\frac{q\phi_B}{kT}} (\psi_s - \psi_0) \right)} \qquad (2.29)$$

Eq. 2.27-2.29 together constitute the unified surface potential model. It can be written as a single-piece continuous equation through a simple transformation of variables as will be shown in the section 2.4.

The surface potential and center potential are a function of the smoothing parameter N. The behavior of the transition region between PD and FD regimes depends on the value of N. Optimal value of N can be found by looking at this transition region as illustrated in Fig. 2.5 for the case of DG-FET with a body doping of 5e18cm^{-3} and $T_{Si} = 20$nm. Comparison with 2-D simulator yields an optimal value of $N = 6$ for this case.

It is virtually impossible to extract the value of N from measured data. As a result, the value of N is calculated inside the model. Large number of simulations are performed to find the optimal value of N over a wide range of N_A and T_{Si}. Fig.

(a) (b)
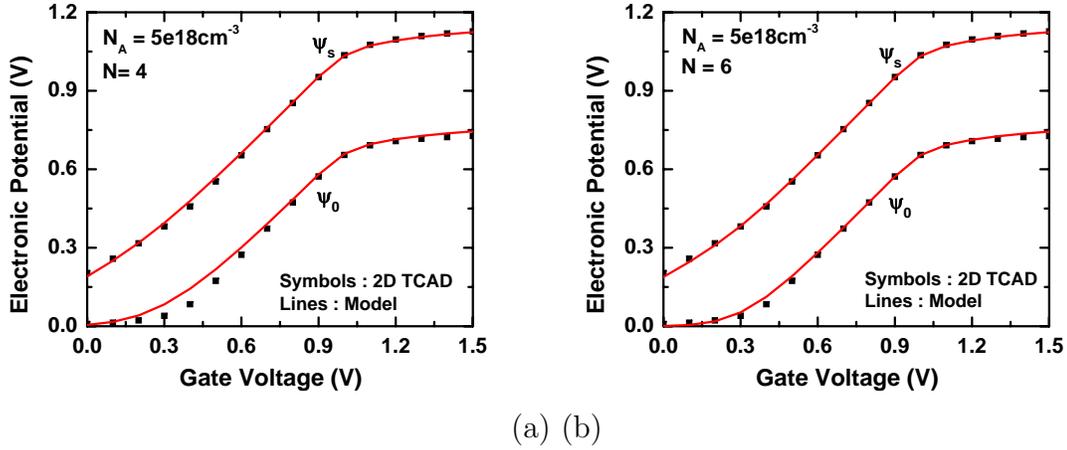
Figure 2.5: Surface potential and center potential for a DG-FET with $N_A = 5e18cm^{-3}$ calculated using the unified surface potential model for (a) $N = 4$ and (b) $N = 6$. Optimal value of N as obtained by comparison with 2-D TCAD in this case is 6.
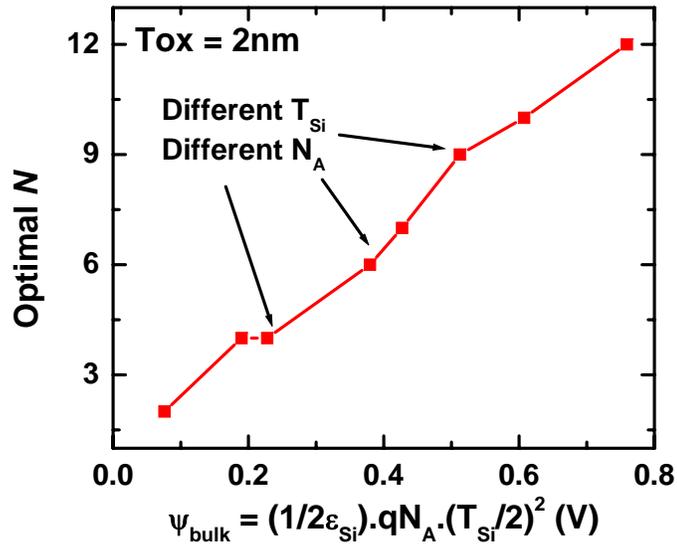
Figure 2.6: Optimal N for the unified surface potential model obtained from 2-D TCAD. A linear fit models the observed N very well.
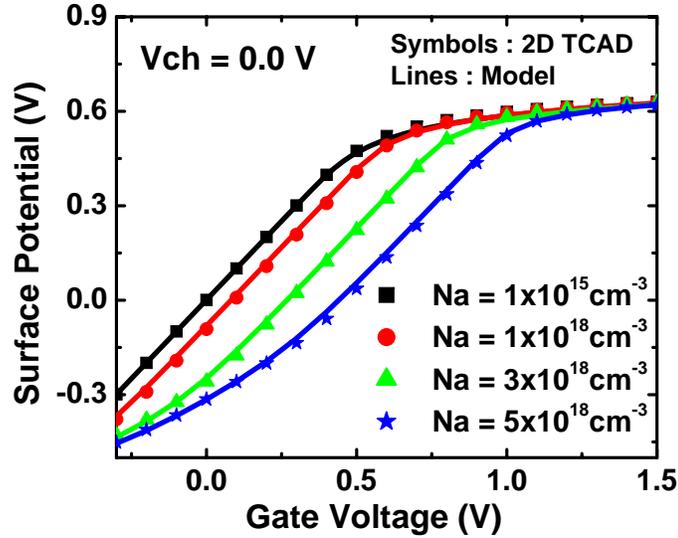
Figure 2.7: Surface potential (referenced to 2-D simulator MEDICI) for DG-FETs with different body doping obtained from the unified surface potential model.

2.6 shows the simulated trend of the optimal value of N and it is modeled through

$$N = MIN\left(2, INT\left(2 \cdot \frac{qN_A}{\epsilon_{Si}} \cdot T_{Si}^2\right)\right) \tag{2.30}$$

where MIN is the minimum function and INT is the integer function which returns the integer portion of the argument number.

The unified surface potential model is verified against 2-D simulator over a wide range of body doping and body thickness as shown in Fig. 2.7. The unified model not only enables a smooth transition between the partially depleted regime and the fully depleted regime but it also yields accurate surface potential. The unified surface potential model formulation is amenable for implementation in any compact model.

It is always important to revisit the assumptions of any model and understand the

resulting limitations. The chief assumption of the unified surface potential model is the presence of negligible inversion carriers when the DG-FET is in PD regime. As a result, the model will be not accurate for DG-FETs with heavy body doping and large body thickness. However, such a MG-FET transistor design will be leakage-prone and hence the underlying assumption of the unified model will never be violated enabling its widespread use in any design environment.

After the unified surface potential model has been developed and verified, the next obvious question is how to implement the model efficiently in a compact modeling framework which is the subject of the next section.

2.4   Analytical Approximation Solution of Unified Surface Potential Model

There are several approaches to solve the unified surface potential model and obtain the numerical value of surface potential. Before divulging into them it is desirable to write the model (Eq. 2.27-2.29) as ONE single-piece continuous equation.

In order to express the unified model in a simpler way, introduce a new variable β which is defined in terms of the center potential, ψ_0, as

$$\beta = \sqrt{\frac{q^2}{2\epsilon_{Si}kT}\frac{n_i^2}{N_A} \cdot e^{\frac{q(\psi_0 - V_{ch})}{kT}}} \cdot \frac{T_{Si}}{2} \tag{2.31}$$

With this transformation, the unknown variable in the earlier formulation ψ_0 is re-

placed with the new variable β.

$$\psi_0 = V_{ch} + 2V_t \left(\ln \beta - \ln \left(\sqrt{\frac{q^2}{2\epsilon_{Si}kT} \frac{n_i^2}{N_A} \cdot \frac{T_{Si}}{2}} \right) \right) \tag{2.32}$$

The contribution to surface potential from inversion carriers, ψ_1 (Eq. 2.28), can be expressed in terms of β using Eq. 2.32

$$\psi_1 = V_{ch} + 2\frac{kT}{q} \left(\ln \beta - \ln \left(\cos \beta \right) - \ln \left(\sqrt{\frac{q^2}{2\epsilon_{Si}kT} \frac{n_i^2}{N_A} \cdot \frac{T_{Si}}{2}} \right) \right) \tag{2.33}$$

Using the definitions of ψ_1 and ψ_0 in terms of β, the Gauss's Law at the surface (Eq. 2.29 can be written as

$$\begin{aligned} f(\beta) \quad \equiv \quad & \ln \beta - \ln \left(\cos \beta \right) + \\ & r \sqrt{\beta^2 \left(\frac{e^{\frac{\psi_{pert}}{V_t}}}{\cos^2 \beta} - 1 \right) + \frac{\psi_{bulk}}{V_t^2} \left(\psi_{pert} - 2V_t \ln \left(\cos \beta \right) \right)} - F = 0 \end{aligned} \tag{2.34}$$

where

$$\begin{aligned} F \quad &= \quad \frac{V_{gs} - V_{fb} - V_{ch} - \psi_{pert}}{2V_t} + \ln \left(\sqrt{\frac{q^2}{2\epsilon_{Si}kT} \frac{n_i^2}{N_A} \cdot \frac{T_{Si}}{2}} \right) \tag{2.35} \\ r \quad &= \quad 2\frac{\epsilon_{Si}}{\epsilon_{ox}} \frac{T_{ox}}{T_{Si}} \\ V_t \quad &= \quad \frac{kT}{q} \end{aligned}$$

Eq. 2.34 is the surface potential equation (SPE) for symmetric common DG-FET and

it forms the core of the BSIM-CMG model. For a given device structure and bias, the only unknown variable in SPE is β. By solving the SPE, β can be determined and hence ψ_1 (Eq. 2.33) and the surface potential ($\psi_s = \psi_1 + \psi_{pert}$) can be obtained.

The SPE can be solved in several ways. One can employ numerical iterative techniques such as Newton-Rhaphson iteration method. The convergence of the solution is not guaranteed in such approaches and it can lead to numerically unstable compact models. Other approach is to use a table look approach. In table look approach, the number of tables can be very large depending on the number of physical variables such as T_{ox}, T_{Si}, etc. Also the size of each table depends on the desired granularity of the bias step in constructing the table. Continuity of terminal currents and charges and the corresponding derivatives is always a question for table look-up models. Another alternative is to solve the SPE using an analytical approximation based on the physics of CDG-FET. An analytical approximate solution of SPE is very desirable as it leads to a numerically robust and efficient compact model.

The analytical approximation solution is based on very good initial approximate solution which is then corrected to obtain the final accurate solution. Let the initial approximate solution of SPE be β_0. The initial solution can then be corrected through a 3rd order correction by the following algorithm

$$\beta = \beta_0 - \frac{f_0}{f_1}\left(1 + \frac{f_0 f_2}{2f_1^2} + \frac{f_0^2\left(3f_2^2 - f_1 f_3\right)}{6f_1^4}\right) \tag{2.36}$$

where $f_n = \frac{\partial^n f}{\partial \beta^n}|_{\beta = \beta_0}$. If the solution is still not accurate, the 3rd order correction can be repeated using the new estimate of β until the desired accuracy is reached. To

keep the time for computing surface potential at a minimum, it is absolutely essential to have a very good initial estimate of β so that only a few stages of correction are required. Hence, the most important step in this technique is obtaining the initial estimate β_0 as close to the final solution β as possible.

To obtain β_0 consider separately the two regimes of operation : sub-threshold regime and strong inversion regime. In sub-threshold regime there are very few carriers in the silicon body. As a result, ψ_1 can be approximated as ψ_0.

$$\psi_1 = \psi_0 - 2\frac{kT}{q}\ln\cos\left(\beta\right) \approx \psi_0 \tag{2.37}$$

Furthermore, since the inversion carriers can be neglected in the sub-threshold regime, the SPE (Eq. 2.34) can be simplified as shown below

$$\ln\beta + r\sqrt{\frac{\psi_{bulk}}{V_t^2}\cdot\psi_{pert}} - F = 0 \tag{2.38}$$

The simplified SPE can be solved to obtain the initial estimate β_0 for the sub-threshold regime of operation.

$$\beta_{0-ST} = e^{\left(F - r\sqrt{\frac{\psi_{bulk}\psi_{pert}}{V_t^2}}\right)} \tag{2.39}$$

β_{0-ST} agrees very well with the exact solution of β in sub-threshold regime. In strong inversion, since F is large, β_{0-ST} becomes very large. To avoid numerical issues while implementing in a compact model, β_{0-ST} needs to be limited for large F. With the

observation that $\beta_{0-ST} << 1$ in the sub-threshold regime, β_{0-ST} can be defined as

$$\beta_{0-ST} = \tan^{-1}\left(e^{\left(F - r\sqrt{\frac{\psi_{bulk}\psi_{pert}}{V_t^2}}\right)}\right) \tag{2.40}$$

without losing any accuracy for low F while obtaining limiting behavior for large F.

Similarly, β_0 for the strong inversion can be estimated. In the strong inversion regime, since the amount of inversion carriers and β are large, the SPE can be expressed as

$$r\sqrt{\beta^2\left(\frac{e^{\frac{\psi_{pert}}{V_t}}}{\cos^2\beta} - 1\right)} - F = 0 \tag{2.41}$$

which can be further simplified and re-written as

$$\beta\tan\beta = \frac{F}{re^{\frac{\psi_{pert}}{2V_t}}} \tag{2.42}$$

For strong inversion, as F increases, β increases and $\tan(\beta)$ will approach ∞. As a result, for large F, β will approach $\frac{\pi}{2}$ and hence the initial estimate for β in strong inversion regime is

$$\beta_{0-SI} = \tan^{-1}\left(\frac{2F}{r\pi} \cdot e^{\frac{-\psi_{pert}}{2V_t}}\right) \tag{2.43}$$

In order to have a well behaved β_{0-SI} in both strong inversion and sub-threshold regimes, a new definition for β_{0-SI} is introduced which retains the correct behavior

at large F but provides the desired limiting at small F.

$$\beta_{0-SI} = \tan^{-1}\left(\frac{2\ln(1+F)}{r\pi} \cdot e^{\frac{-\psi_{pert}}{2V_t}}\right) \qquad (2.44)$$

The estimate of β_0 through all regimes of operation is then simply given by

$$\beta_0 = MIN\left(\beta_{0-ST}, \beta_{0-SI}\right) \qquad (2.45)$$

This exercise of mathematically modifying the physically derived expressions into forms which are better suited for efficient and robust compact models is commonplace in the field of compact modeling and is very crucial. Failing to do so may render excellent physical models unusable for compact models. Mathematical behavior of the equations is as important as the physics buried in them. This is an very important aspect of compact modeling which is often understated and overlooked.

The initial estimate β_0 (Eq. 2.45) with a two stage 3rd order correction (Eq. 2.36) forms the analytical approximation solution of the SPE in BSIM-CMG model. In order to minimize the computational time for the extremely complex derivatives f_1, f_2 and f_3 in Eq. 2.36 it is necessary to share the computation of common variables which is another important facet of developing an efficient compact model.

The accuracy of analytical approximation solution of SPE is verified against exact numerical solution. Fig. 2.8 shows the SPE solution of a DG-FET with light body doping of $N_A = 1e15\text{cm}^{-3}$. In Fig. 2.8(a), the initial approximation β_0 and the approximation after two 3rd order corrections β_2 are compared against numerical solution of the SPE β_n. The initial approximation β_0 is very close to β_n in both
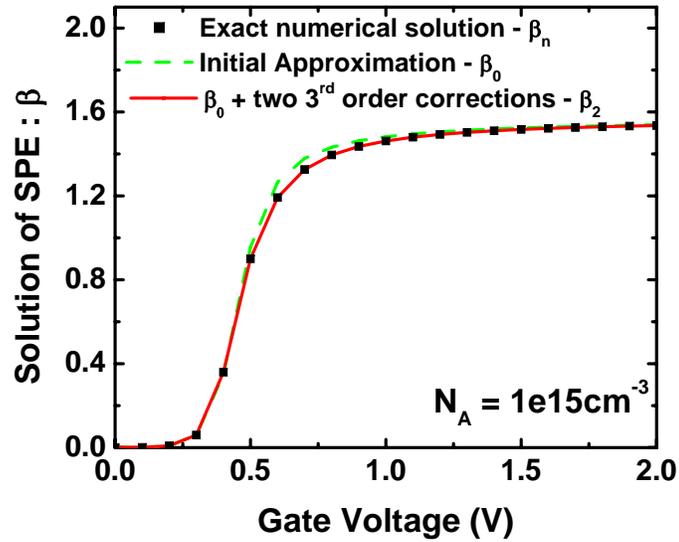
(a)
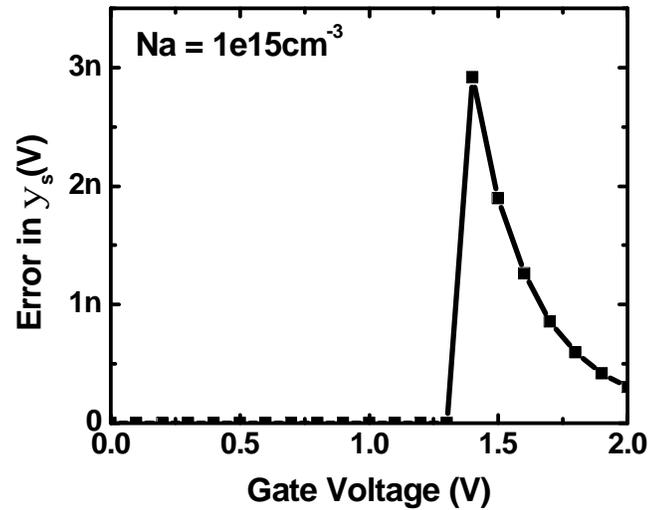
(b)

Figure 2.8: Analytical approximation solution of ψ_s for DG-FET with light body doping of $N_A = 1e15cm^{-3}$. (a) Approximate solutions of SPE : β_0 and β_2 compared against the numerical solution of SPE β_n. (b) Error between ψ_s calculated from analytical approximation and ψ_s calculated using numerical techniques.
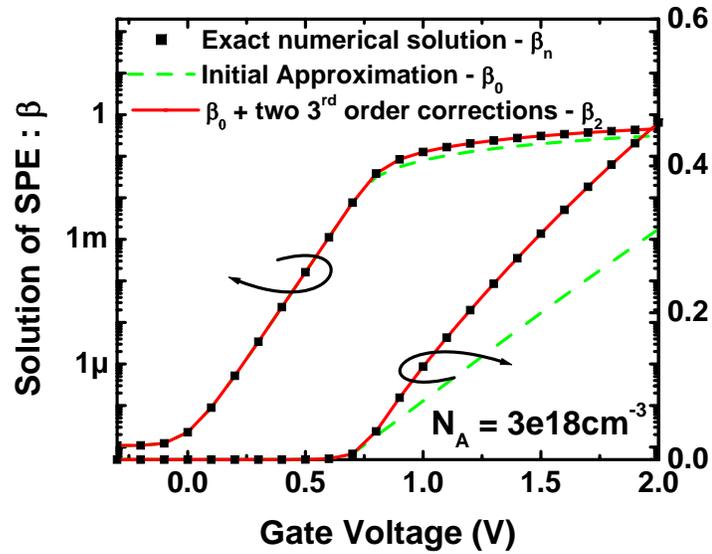
(a)
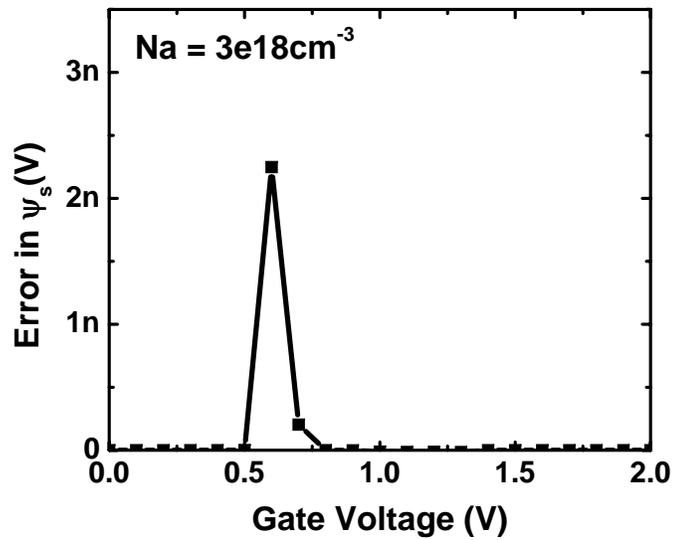
(b)

Figure 2.9: Analytical approximation solution of ψ_s for a DG-FET with heavy body doping of $N_A = 3e18cm^{-3}$. (a) Approximate solutions of SPE : β_0 and β_2 compared against the numerical solution of SPE β_n. (b) Error between ψ_s calculated from analytical approximation and ψ_s calculated using numerical techniques.

weak inversion and strong inversion regimes. After two 3rd order corrections β_2 agrees very well with β_n. Fig. 2.8(b) shows the difference in ψ_s calculated using analytical approximation and numerical solution. The error is limited to a few nano-volts indicating good accuracy of analytical approximation of SPE. The analytical approximation works equally well for DG-FETs with heavy body doping as illustrated in Fig. 2.9 for the case of $N_A = 3e18cm^{-3}$. Fig. 2.9(a) indicates that the β_2 is very close to β_n in both sub-threshold and strong inversion regime. The error in ψ_s calculated using the analytical approximation solution of SPE for a heavily doped body is also limited to a few nano-volts. This has negligible effect on the accuracy of the terminal currents and terminal charges calculated in the model.

2.5   Enhanced Surface Potential Equation

The surface potential equation developed thus far is still rudimentary and devoid of some important physical effects observed in real devices. The surface potential equation will be enhanced in this section through the addition of phenomena such as Poly-Depletion Effects (PDE) and Quantum Mechanical Effects (QME). Upon the addition of these effects into SPE, the analytical approximation solution also needs to be changed to maintain the good accuracy and efficiency of analytical solution.

2.5.1   Poly Depletion Effect

The surface potential model developed so far assumes the use of metal gates. If poly-silicon is used as the gate electrode material, it introduces poly-depletion effect which
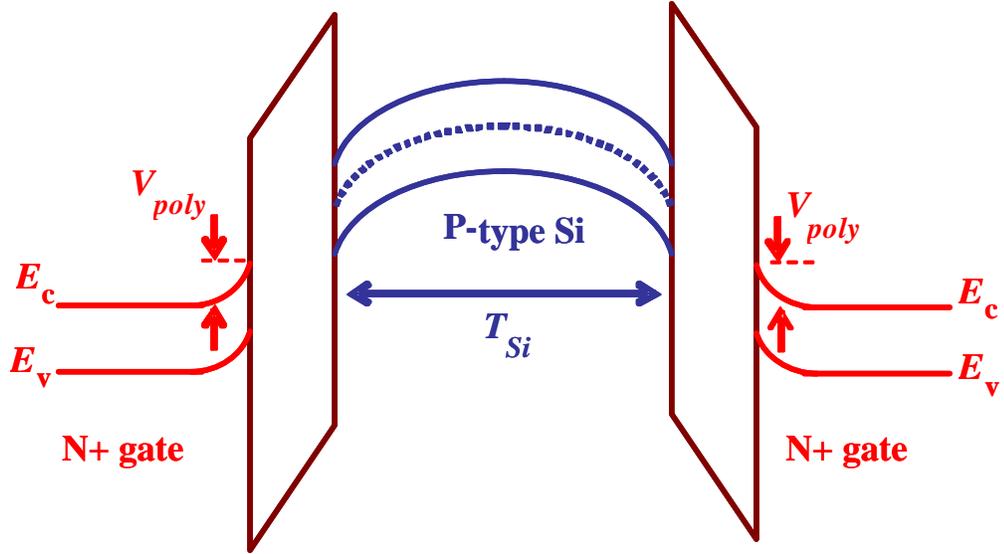
Figure 2.10: Band bending in a n-type DG-FET with N+ poly-Si gates. The use of poly-Si gates causes additional voltage drop V_{poly} in the gate electrode.

leads to a reduction in the inversion charge density compared to the case of metal gates.

Similar to the planar bulk MOSFETs, if N+ poly-Si is used as the gate electrode for a n-type DG-FET, a depletion layer forms in the N+ poly-Si gate electrode. As shown in Fig. 2.10, there is an additional drop in the gate voltage at the interface of poly-silicon gate and the gate insulator. This additional voltage drop in the gate electrode introduces a new term, V_{poly}, in the Gauss's Law which can be written as

$$\underbrace{V_{gs} - V_{fb} = \psi_s + V_{ox}}_{f(\beta)} + V_{poly} \qquad (2.46)$$

where V_{ox} is the voltage drop across the gate insulator and $f(\beta)$ is the basic SPE

introduced in Eq. 2.34. V_{ox} is linearly related to the electric field at the silicon surface (E_{Si}) through

$$V_{ox} = \frac{\epsilon_{Si} E_{Si}}{C_{ox}} \qquad (2.47)$$

V_{poly} can also be expressed as a function of E_{Si} through

$$V_{poly} = \frac{\epsilon_{Si}}{2q \, N_{poly}} E_{Si}^2 \qquad (2.48)$$

where N_{poly} is the doping concentration of the poly-Si gate electrode. E_{Si} and its derivatives need to be calculated for solving the original SPE since V_{ox} is a function of E_{Si}. By expressing V_{poly} in terms of E_{Si}, no additional computational overhead is added when solving for SPE with PDE included. The SPE with PDE included can be finally written as

$$\ln \beta - \ln(\cos \beta) + r_2 \left(\beta^2 \left(\frac{e^{\frac{\psi_{pert}}{V_t}}}{\cos^2 \beta} - 1 \right) + \frac{\psi_{bulk}}{V_t^2} (\psi_{pert} - 2V_t \ln(\cos \beta)) \right) \qquad (2.49)$$

$$+r \sqrt{\beta^2 \left(\frac{e^{\frac{\psi_{pert}}{V_t}}}{\cos^2 \beta} - 1 \right) + \frac{\psi_{bulk}}{V_t^2} (\psi_{pert} - 2V_t \ln(\cos \beta))} - F = 0 \equiv f(\beta)_{SPE+PDE}$$

where the factor r_2 is given by

$$r_2 = \frac{4\epsilon_{Si} V_t}{qT_{Si}^2 N_{poly}} \qquad (2.50)$$

The above equation can be solved to obtain β and hence ψ_s in the presence of

PDE for a CDG-FET. In order to solve the SPE with PDE using the analytical approximation technique, the initial approximate solution β_0 should be changed to incorporate the effect of poly-depletion. In sub-threshold regime, the SPE with PDE can be simplified into a form similar to Eq. 2.38 but with PDE included.

$$\ln \beta + r_2 \frac{\psi_{bulk}}{V_t^2} \cdot \psi_{pert} + r\sqrt{\frac{\psi_{bulk}}{V_t^2} \cdot \psi_{pert}} - F = 0 \tag{2.51}$$

Solving the simplified SPE in Eq. 2.51 yields the initial estimate of β_0 in sub-threshold regime.

$$\beta_{0-ST} = e^{\left(F - r \cdot \sqrt{\frac{\psi_{bulk}\psi_{pert}}{V_t^2}} - r_2 \cdot \frac{\psi_{bulk}\psi_{pert}}{V_t^2} \right)} \tag{2.52}$$

β_{0-ST} needs to be mathematically modified for numerical robustness at large gate voltage bias.

$$\beta_{0-ST} = \tan^{-1}\left(e^{\left(F - r \cdot \sqrt{\frac{\psi_{bulk}\psi_{pert}}{V_t^2}} - r_2 \cdot \frac{\psi_{bulk}\psi_{pert}}{V_t^2} \right)} \right) \tag{2.53}$$

In the limit of strong inversion regime V_{poly} will be negligible compared to V_{ox} and hence β_{0-SI} remains same as Eq. 2.44.

The SPE with PDE has been verified against 2-D TCAD for different poly-Si doping. Fig. 2.11(a) shows the surface potential of a n-type DG-FET with uniform body doping of $N_A = 1e15\text{cm}^{-3}$ for different poly-Si doping. As expected, ψ_s decreases as N_{poly} decreases since there is larger voltage drop in gate electrode and the silicon body sees less effective V_{gs}. Similar results are observed for DG-FET with heavy
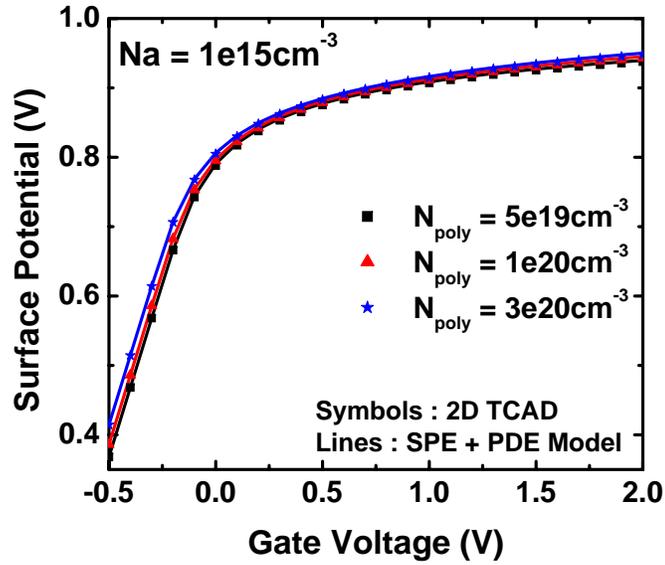
(a)
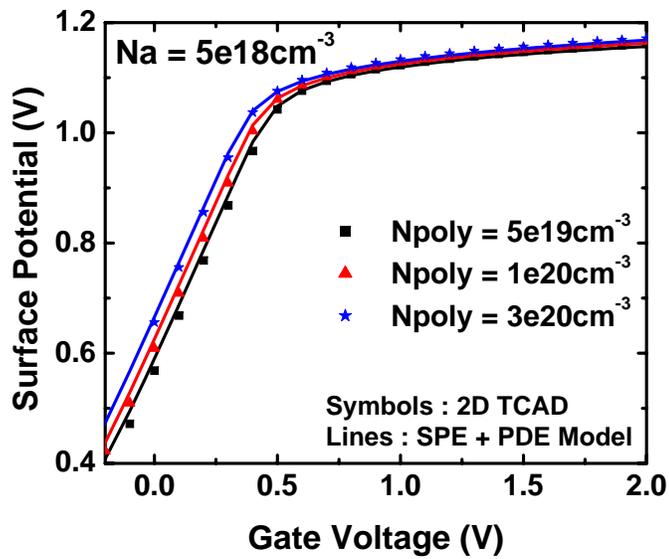
(b)

Figure 2.11: Surface potential for DG-FET with (a) Light body doping and (b) Heavy body doping for different poly-Si doping. The model captures the change in surface potential as a function of poly-Si doping accurately.

body doping of $N_A = $ 5e18cm^{-3} as shown in Fig. 2.11(b). In all the cases, the ψ_s calculated using analytical approximation model shows good agreement with ψ_s obtained through 2-D TCAD.

2.5.2   Quantum Mechanical Effects

Quantum mechanical confinement of inversion carriers is well known in bulk MOS-FETs for a long time [30]. The large vertical electric field leads to strong band bending at the surface and the inversion carriers are confined to dimensions along the length and width of the transistor. This carrier confinement, also known as electrical confinement (EC), leads to splitting of energy bands into discrete subbands which reflects as an increase in the threshold voltage of the transistor and a decrease in the gate capacitance, both of which act to reduce the current drive of the transistor [31].

In the case of DG-FETs, unlike bulk FETs, there is strong carrier confinement even at low electric fields making the QME even more complex. The carriers are bounded by gate insulator on two sides and it is similar to carriers confined in a rectangular well [32]. This is referred to as structural confinement (SC) since it arises from the very physical structure of DG-FET. In order to capture the QME in its entirety it is necessary to model the effect of both EC and SC (Fig. 2.12) on the performance of DG-FETs. Several groups have reported different analytical and numerical approaches to capture the QME in DG-FETs [33], [34].

A new compact approach is developed to model QME which takes advantage of the simple SPE formulation [35]. Carrier confinement causes the splitting of conduction band and the energy of the lowest subband is higher than the conduction band edge.
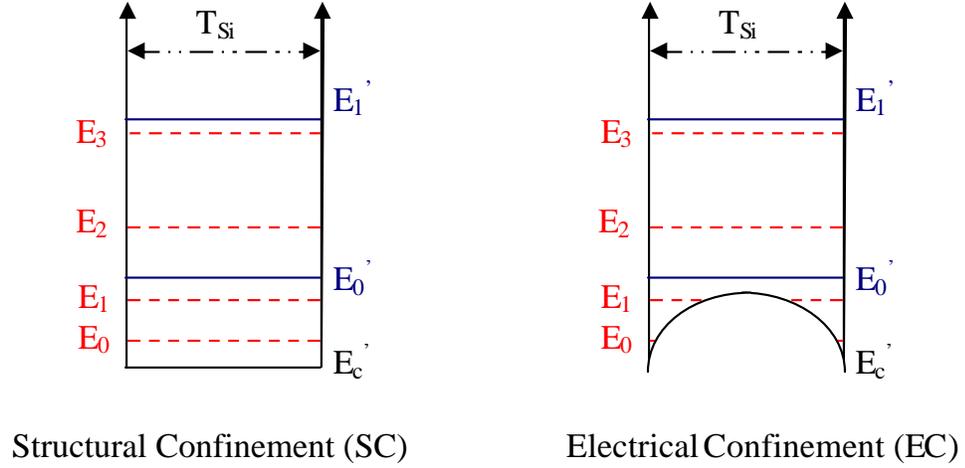
Figure 2.12: Energy-band diagrams showing the carrier confinement and associated quantization of electronic energy levels in DG-FETs due to structural confinement and electrical confinement.

The energy of the lowest subband referenced to the conduction band edge is given by [36]

$$E_0 = \frac{\hbar^2}{2m_x} \left(\left(\frac{\pi}{T_{Si}} \right)^2 + b_0^2 \left(3 - \frac{4}{3 \left((b_0 T_{Si}/\pi)^2 + 1 \right)} \right) \right) \tag{2.54}$$

where m_x is the effective carrier mass perpendicular to the silicon-insulator interface and b_0 is

$$b_0 = \left(\frac{3}{4} \cdot \frac{2m_x q E_{Si}}{\hbar^2} \right)^{\frac{1}{3}} \tag{2.55}$$

and E_{Si} is the vertical electric field at the silicon-insulator surface. The electric field,

E_{Si}, can be written in terms of the surface potential as

$$E_{Si} = \frac{V_{gs} - V_{fb} - \psi_s}{T_{ox}} \cdot \frac{\epsilon_{ox}}{\epsilon_{Si}} \tag{2.56}$$

The first term in Eq. 2.54 represents the subband energy change due to SC and the latter term models the effect of EC at higher gate bias.

Since the lowest subband energy level is higher in presence of QM confinement by E_0, more band bending or equivalently higher surface potential is required to attain same inversion charge density as in the classical case. In other words, for a given gate bias, the QM inversion charge density is smaller than the classical case. This behavior can be captured easily in the SPE formulation by modifying the inversion carrier term to reflect the corresponding decrease in inversion carrier density. The SPE with QME is expressed as

$$
\begin{aligned}
V_{gs} \quad = \quad & V_{fb} + \psi_s + \\
& \frac{\epsilon_{Si}}{C_{ox}} \sqrt{\frac{2qn_i}{\epsilon_{Si}} \left(\frac{e^{\frac{q\psi_s}{kT}} - e^{\frac{q\psi_0}{kT}}}{q/kT} e^{\frac{-q(V_{ch} + \phi_B + QMFIT \cdot E_0/q)}{kT}} + e^{\frac{q\phi_B}{kT}} \left(\psi_s - \psi_0 \right) \right)}
\end{aligned}
\tag{2.57}
$$

where $QMFIT$ is a fitting parameter. As the carrier confinement increases, E_0 increases and the inversion charge decreases as seen clearly in Eq. 2.57. It captures both the structural confinement and electrical confinement. Fig. 2.13 shows the surface potential versus the gate voltage for DG MOSFET with heavy body doping of N_A = 5e18cm^{-3}. The model agrees with 2-D device TCAD very well from subthreshold region to strong inversion region. The increase of the surface potential due to the
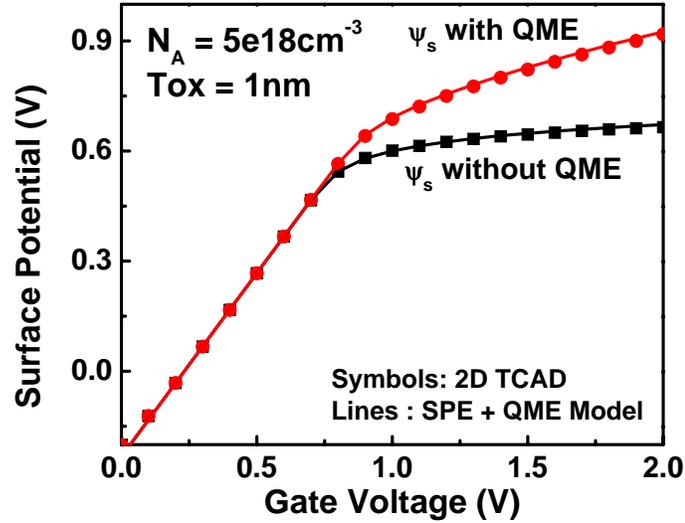
Figure 2.13: Surface potential of a heavily doped DG-FET with and without carrier confinement.

QME is captured by the model.

However, the implementation of Eq. 2.57 into a compact model poses a great challenge. The EC is field dependent and hence gate bias and drain bias dependent. This significantly increases the computational time to solve the SPE. More number of correction stages are required in the analytical approximation framework to achieve surface potential within the desired accuracy of few nano-volts. This is a clear instance where there is a trade-off between the complexity of embedded physics versus the corresponding computational time. As a solution to the trade-off, only SC is calculated in SPE and the EC is captured by explicitly modifying effective insulator thickness (T_{ox}) similar to BSIM4 model [37]. By including only SC in the surface

potential solution, Eq. 2.34 can be modified as

$$
\ln \beta - \ln \left(\cos \beta \right) + \tag{2.58}
$$

$$
r \sqrt{ \underbrace{ e^{-\frac{QMFIT \cdot \left(\frac{\hbar \pi}{T_{Si}} \right)^2}{2V_t \cdot qm^*}} }_{QME} \cdot \beta^2 \left(\frac{e^{\frac{\psi_{pert}}{V_t}}}{\cos^2 \beta} - 1 \right) + \frac{\psi_{bulk}}{V_t^2} \left(\psi_{pert} - 2V_t \ln \left(\cos \beta \right) \right) } - F = 0
$$

When the body thickness T_{Si} decreases, the factor QME decreases and reduces the inversion charge density capturing structural carrier confinement. Eq. 2.58 is the quantum mechanical surface potential equation which captures the effect of structural carrier confinement on surface potential in BSIM-CMG. The SPE with QME is solved using the analytical approximation solution by changing the initial approximate solution β_0 to reflect the impact of QME.

The initial approximate solution β_0 needs to be modified in the presence of QME. In the sub-threshold regime, since β_{0-ST} is estimated by ignoring the inversion carriers, there is no impact of QME and β_{0-ST} remains unaltered from Eq. 2.40. However, in the strong inversion regime, the QME significantly effect the β_{0-SI} through the inversion carrier term. Following a approach similar to the one in Section 2.4, β_{0-SI} in the presence of QME can be calculated as

$$
\beta_{0-SI} = \tan^{-1} \left(\frac{2 \ln(1 + F)}{r\pi} \cdot e^{\frac{-\psi_{pert}}{2V_t}} \cdot e^{\frac{QMFIT \cdot \left(\frac{\hbar \pi}{T_{Si}} \right)^2}{4V_t \cdot qm^*}} \right) \tag{2.59}
$$

2.5.3 SPE with PDE and QME

The surface potential equation formulation enables inclusion of both the poly-depletion effects and quantum mechanical effects at the same time. The complete surface potential equation capturing the effects of both quantum mechanical confinement and poly-depletion can be expressed as

$$f\left(\beta\right) \equiv \ln\beta - \ln\left(\cos\beta\right) + \underbrace{r \cdot g\left(\beta\right)}_{V_{ox}} + \underbrace{r_2 \cdot g\left(\beta\right)^2}_{V_{poly}} - F = 0 \qquad (2.60)$$

where $g(\beta)$ is defined as

$$g(\beta) = \sqrt{QMFIT \cdot \beta^2 \left(\frac{e^{\frac{\psi_{pert}}{V_t}}}{\cos^2\beta} - 1\right) + \frac{\psi_{bulk}}{V_t^2}\left(\psi_{pert} - 2V_t \ln\left(\cos\beta\right)\right)} \qquad (2.61)$$

This equation forms the core of BSIM-CMG model. It is solved through analytical approximation technique where the initial approximation β_0 is given by

$$\beta_0 = MIN\left(\beta_{0-ST}, \beta_{0-SI}\right) \qquad (2.62)$$

where

$$\beta_{0-ST} = \tan^{-1}\left(e^{\left(F - r\sqrt{\frac{\psi_{bulk}\psi_{pert}}{V_t^2}} - r_2\frac{\psi_{bulk}\psi_{pert}}{V_t^2}\right)}\right) \qquad (2.63)$$

$$\beta_{0-SI} = \tan^{-1}\left(\frac{2\ln(1+F)}{r\pi} \cdot e^{\frac{-\psi_{pert}}{2V_t}} \cdot e^{\frac{QMFIT \cdot \left(\frac{\hbar\pi}{T_{Si}}\right)^2}{4V_t \cdot qm^*}}\right)$$

β_0 is corrected through a two stage 3rd order correction defined in Eq. 2.36.

The surface potential at the source end, ϕ_s, is obtained by solving Eq. 2.60 at the source end ($V_{ch} = 0$) and the surface potential the drain end, ϕ_d, is obtained by solving Eq. 2.60 at the drain end ($V_{ch} = V_{ds}$). All the terminal currents and charges for the surface potential model BSIM-CMG are a function of ϕ_s and ϕ_d. In the next chapter, the formulation of drain current and terminal charges as a function of ϕ_s and ϕ_d will be described.

2.6 Summary

For any surface potential based MOSFET model, accurate calculation of the surface potential is the most crucial part of the model. This chapter presented an analytical framework for calculating the surface potential for a symmetric common double-gate transistor (CDG-FET). The analytical model captures the effect of finite body doping on the surface potential of CDG-FET. A unified surface potential equation was later developed to capture the transition between the partially depleted regime and the fully depleted regime in a smooth and continuous manner without sacrificing any accuracy. The surface potential equation was enhanced to capture real device effects such as poly-depletion and quantum mechanical confinement of the inversion carriers. The numerically complex surface potential equation is solved using an analytical approximation method. The analytical surface potential model has been verified against several 2-D TCAD simulations for a wide range of body doping. The surface potential equation described in this chapter is implemented in the surface potential based model BSIM-CMG and it forms the core of the BSIM-CMG model.

Chapter 3

Common Symmetric Multi-Gate FET - I-V and C-V Model

3.1   Introduction

The goal of a compact model is to provide a concise mathematical model of a transistor to aid circuit design. To enable circuit level simulations, which include DC, AC and transient analysis, a compact model needs to model the current flowing through all the terminals of the transistor and the associated terminal charges and capacitances. The I-V model describes the currents associated with the transistor terminals and the C-V model describes the terminal charges and transcapacitances. One of the most important requirements for both I-V and C-V model formulations is to posses a certain degree of predictivity and scalability without compromising the computational time.

The I-V model for any transistor - single-gate or multi-gate FET, is fairly complex. The drain current of the transistor consists of drift and diffusion components. Depending on the gate bias, one of the two conduction mechanisms contribute significantly to the drain current. It is thus necessary to model both these conduction mechanisms at the same time. The state-of-the-art transistor exhibits several physical mechanisms such as short-channel effects, velocity saturation and velocity overshoot which effect the drain current significantly making the modeling of drain current very challenging. Furthermore, the drain current is just one component of the I-V model. Gate current and body current also need to be accounted due to reduced gate insulator thickness and increased electric fields inside the silicon body in modern transistors. Tunneling through the gate insulator is responsible for the gate current whereas the bulk current is dominated by impact ionization at the drain terminal at large drain bias. Modeling of all these different mechanisms makes the I-V model considerably involved.

The I-V model is usually developed using a bottom-up approach. First the drain current is modeled for a long channel transistor which will be referred to as the core I-V model. Next, short channel effects and other physical mechanisms are incorporated into the core I-V model to enable the prediction and modeling of the behavior of next generation transistors. Other terminal currents are then added into I-V model. This methodology applies to both bulk planar MOSFET and multi-gate FETs. Only the modeling of drain current for MG-FETs will be described in this chapter. Most of the efforts in modeling the drain current of DG-FETs have been limited to DG-FETs with undoped silicon body [27], [38], [39]. Section 3.2 describes a new surface potential

based I-V model for DG-FET which incorporates the effect of finite body doping on the drain current. The phenomenon of volume (bulk) inversion in DG-FETs makes the modeling of drain current even more interesting as will be explained later in this chapter. The drain current of short channel transistors differs significantly from long channel transistors. An elegant way to capture the behavior of short-channel transistors such as threshold voltage roll-off and sub-threshold slope degradation is presented in Section 3.3. By accurate modeling of the short-channel effects, the DG-FET model can be extended to MG-FETs with three or four gates.

The C-V model is equally important as the I-V model. The terminal charges are calculated in accordance with the Ward-Dutton charge partition which dictates the partitioning of the inversion layer charge to the source and drain terminals [40]. The transcapacitances are simply the derivatives of the terminal charges. Some of the important requirements of the C-V model are having non-negative, continuous and symmetric (ex. $C_{gs} = C_{gd}$ when V_{ds}=0) capacitances. Though they may seem very obvious, they are very hard to achieve and care has to be taken at every step in the model development to ensure that the C-V model adheres to these requirements. The same holds true for the currents and its derivatives such as trans-conductance and output conductance for the I-V model also. Following the discussion on I-V model, the formulation of C-V model for DG-FETs is presented in Section 3.4.

The complete compact model for symmetric common-gate multi-gate FETs - BSIM-CMG, consisting of the unified surface potential model, I-V model and C-V model is written in Verilog-A and implemented in popular circuit simulators such as HSPICE and SPECTRE. MG-FETs can be built either on SOI or bulk silicon.
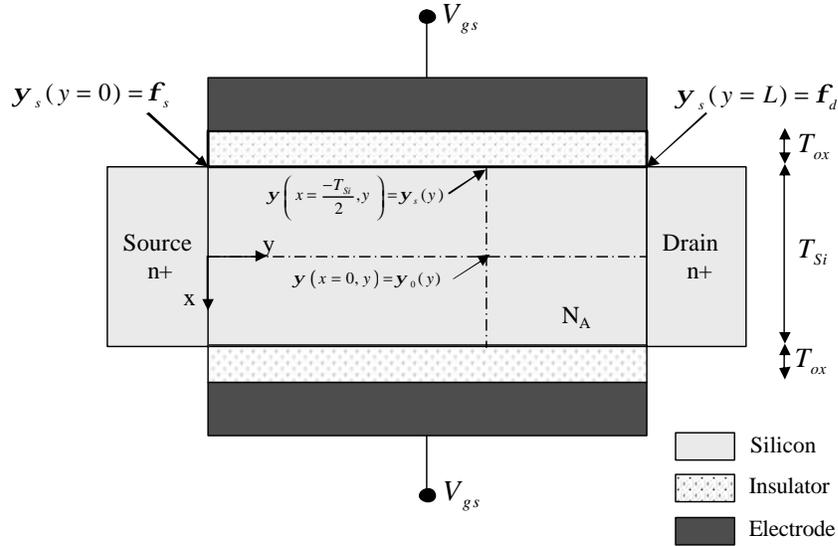
Figure 3.1: Schematic of a symmetric common-gate DG-FET

BSIM-CMG model allows users to model SOI MG-FET and bulk MG-FET through the addition of body node for bulk MG-FET. The BSIM-CMG model has been successfully used to describe the measured electrical characteristics of SOI FinFETs and bulk FinFETs. Section 3.6 in this chapter is devoted to the experimental verification of the complete model.

3.2 Core I-V Model

The drain current consists of both drift and the diffusion components [41]. Fig. 3.1 shows the schematic of the long channel symmetric common DG-FET (CDG-FET) under study. Assuming a constant carrier mobility of μ, the drain current (I_d) at a

plane y along the channel length can be written as

$$I_d(y) \quad = \quad 2 \cdot \mu \cdot W \cdot Q_{inv}(y) \cdot \frac{dV_{ch}(y)}{dy} \tag{3.1}$$

where W is the width of the DG-FET, V_{ch} is the carrier quasi-fermi potential and Q_{inv} is defined as the total inversion charge in upper half (also equal to total inversion charge in the lower half) of the silicon body. The electrical and physical symmetry in CDG-FET forces the charge in the upper half of silicon body to be equal to the charge in the lower half and this explains the factor of 2 in Eq. 3.1. From here on, all the charge terms will refer to charge in one half of the silicon body.

Under quasi-static operation of the transistor, the drain current is same at any point between source and drain. Eq. 3.1 can now be integrated from source to drain yielding

$$I_d \quad = \quad 2 \cdot \mu \cdot \frac{W}{L} \cdot \int_0^{V_{ds}} Q_{inv}(y) \cdot \frac{dV_{ch}(y)}{dy} \tag{3.2}$$

where L is the channel length. The inversion charge is simply the difference of the total charge in silicon body and the bulk charge.

$$Q_{inv}(y) = Q_{Si}(y) - Q_{bulk}(y) \tag{3.3}$$

The total charge in the silicon is given by

$$Q_{Si}(y) = C_{ox} \left(V_{gs} - V_{fb} - \psi_s(y) \right) \tag{3.4}$$

where $\psi_s(y)$ is the surface potential along the channel length. The bulk charge is given by

$$Q_{bulk}(y) = \sqrt{2qn_i\epsilon_{Si}\, e^{\frac{\phi_B}{V_t}}\, \psi_c(y)} \tag{3.5}$$

To evaluate the drain current integral in Eq. 3.2, the rate of change of $V_{ch}(y)$ is also needed. It can be obtained in terms of Q_{inv} by noting that Q_{Si} can be expressed as a function of V_{ch} through

$$Q_{Si}(y) = \sqrt{2qn_i\epsilon_{Si}\left(\frac{e^{\frac{q\psi_s(y)}{kT}} - e^{\frac{q\psi_0(y)}{kT}}}{q/kT}\, e^{\frac{-q(\phi_B+V_{ch}(y))}{kT}} + e^{\frac{q\phi_B}{kT}}\left(\psi_s(y) - \psi_0(y)\right)\right)} \tag{3.6}$$

where all the symbols retain the definitions introduced in Chapter 2. This definition of Q_{Si} is mathematically very complex yielding a complex expression of $\frac{dV_{ch}}{dy}$ and the integral in Eq. 3.2 is not analytically integrable. Analytic expression for drain current valid in all regimes of transistor operation is highly desirable for a numerically robust and an efficient compact model. To obtain an analytical core drain current model, a mathematically simpler approximation of $Q_{inv}(y)$ needs to developed which can be used to evaluate $\frac{dV_{ch}(y)}{dy}$ and hence the drain current. This will be done in a two step approach. First, an analytical approximation of $Q_{inv}(y)$ will be separately derived for lightly doped DG-FETs and highly doped DG-FETs. Then, a unified approximation of $Q_{inv}(y)$ which is valid for both light and heavy body doping is developed to be used for calculating $\frac{dV_{ch}(y)}{dy}$.

3.2.1 Inversion Charge in Lightly Doped DG-FET

For lightly doped DG-FET, the body charge Q_{bulk} is negligible compared to the inversion carriers. Furthermore, since the body is doped lighty, Eq. 3.6 can be further simplified by neglecting the contribution from bulk charge. The inversion charge density for lightly doped DG-FET can then be expressed as

$$Q_{inv}(y) \approx \sqrt{2qn_i\epsilon_{Si} \cdot \left(\frac{e^{\frac{q\psi_s(y)}{kT}} - e^{\frac{q\psi_0(y)}{kT}}}{q/kT} \right) \cdot e^{\frac{-q(\phi_B+V_{ch}(y))}{kT}}} \qquad (3.7)$$

which can be re-written as

$$Q_{inv}(y) = \sqrt{2qn_i\epsilon_{Si}V_t} \cdot e^{\frac{\psi_s(y)-\phi_B+V_{ch}(y)}{2V_t}} \cdot \sqrt{1 - e^{\frac{\psi_0(y)-\psi_s(y)}{V_t}}} \qquad (3.8)$$

The last term with the square root approaches 1 in strong inversion and it decreases in weak inversion regime. In strong inversion regime, the first two terms are a good approximation to the inversion charge. In the weak inversion regime, the last term can be further simplified. The potential profile from the center to the surface of DG-FET depends on the exact value of body doping and amount of inversion. To ease the model derivation, a linear potential profile is assumed for now from center of the body to the surface. This assumption will be revisited towards the end of the section. With a linear potential profile assumption in the weak inversion regime

$$\psi_s(y) - \psi_0(y) = \frac{Q_{inv}(y)}{2C_{Si}} \qquad (3.9)$$
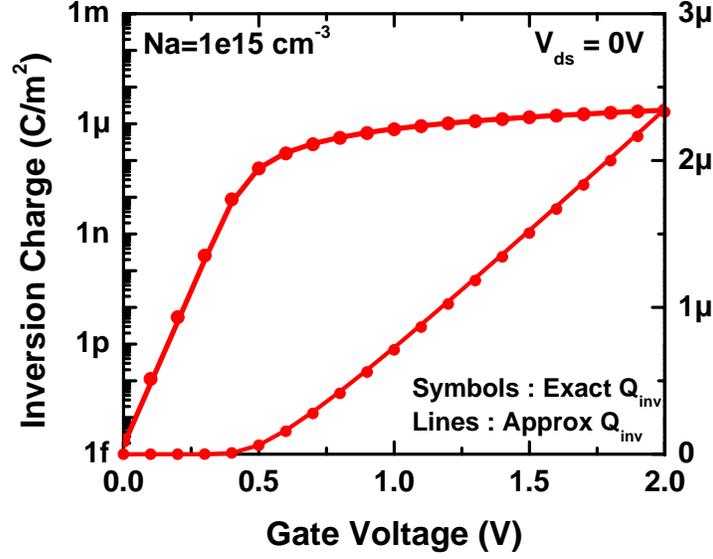
Figure 3.2: Comparison of the approximate Q_{inv} model for lightly doped DG-FET calculated against the exact Q_{inv} in both sub-threshold and strong inversion regime.

where $C_{Si} = \frac{\epsilon_{Si}}{T_{Si}}$. Substituting Eq. 3.9 in Eq. 3.8 and performing a Taylor series expansion, the inversion charge for lightly doped DG-FETs $Q_{inv:LD}$ can be written as

$$Q_{inv:LD}(y) \approx \sqrt{2qn_i\epsilon_{Si}V_t} \cdot e^{\frac{\psi_s(y)-\phi_B-V_{ch}(y)}{2V_t}} \cdot \sqrt{\frac{Q_{inv}(y)}{Q_{inv}(y) + 2C_{Si}V_t}} \qquad (3.10)$$

Fig. 3.2 compares the exact inversion charge from Eq. 3.3 and the approximation Eq. 3.10. The approximation closely matches the exact inversion charge in both weak inversion and strong inversion. Furthermore, $Q_{inv:LD}$ can be used in Eq. 3.2 to obtain an analytic expression for drain current for the lightly doped DG-FETs.

$$I_d = 2 \cdot \mu \cdot \frac{W}{L} \cdot (g(Q_{invs}) - g(Q_{invd})) \qquad (3.11)$$

62
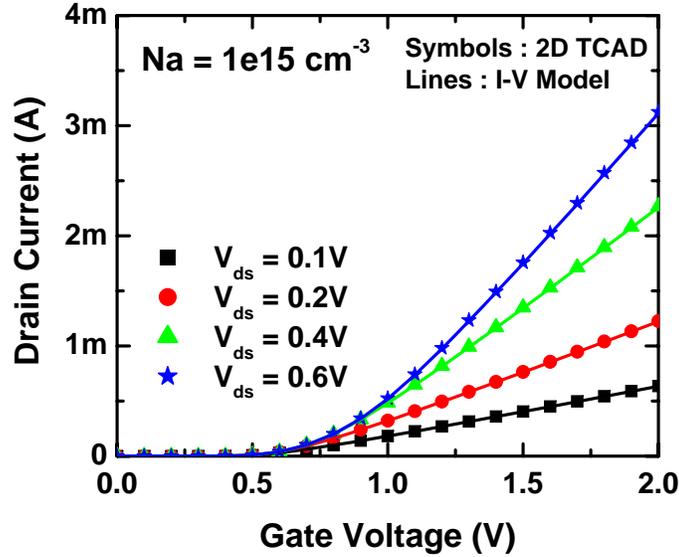
Figure 3.3: Comparison of the drain current of lightly doped DG-FET calculated using the I-V model Eq. 3.11 and Eq. 3.12 against 2-D TCAD simulations.

where $Q_{invs} = Q_{inv}(y = 0)$ and $Q_{invd} = Q_{inv}(y = L)$ (calculated using Eq. 3.3) and the function $g(Q)$ is

$$g(Q) = \frac{Q^2}{2C_{ox}} + 2V_t Q - 2C_{Si}V_t^2 \ln\left(1 + \frac{Q}{2V_t C_{Si}}\right) \qquad (3.12)$$

The drain current calculated from Eq. 3.11 for a DG-FET with $N_A = 1e15$ cm^{-3} is shown in Fig. 3.3. The drain current model prediction agrees very well with 2D TCAD simulations as seen in Fig. 3.3. Note that there is an underlying assumption of linear potential profile. If a parabolic potential profile was chosen instead of linear potential profile, the term $2C_{Si}$ in Eq. 3.10 and in function $g(Q)$ would be replaced

by $4C_{Si}$. With $g(Q)$ defined as

$$g(Q) = \frac{Q^2}{4C_{ox}} + 2V_t Q - 4C_{Si}V_t^2 \ln\left(1 + \frac{Q}{4V_t C_{Si}}\right) \tag{3.13}$$

it can be shown that the drain current model in Eq. 3.11 is exactly identical to the I-V model presented in [39] for undoped DG-FETs.

It is clear that the factor $2C_{Si}$ or $4C_{Si}$ in Eq. 3.10 depends strongly on the potential profile assumed. As a result, the function $g(Q)$ in Eq. 3.13 can be defined in general as

$$g(Q) = \frac{Q^2}{2C_{ox}} + 2V_t Q - xC_{Si}V_t^2 \ln\left(1 + \frac{Q}{xV_t C_{Si}}\right) \tag{3.14}$$

where the optimal x can be found by comparing the drain current using the analytical model against 2-D TCAD simulations.

Fig. 3.4 plots the error in drain current calculated using the analytical model for different x. From Fig. 3.4, it can be concluded that $x = 5$ gives the most accurate drain current and hence $Q_{inv:LD}(y)$ will be re-defined as

$$Q_{inv:LD}(y) \approx \sqrt{2qn_i\epsilon_{Si}V_t} \cdot e^{\frac{\psi_s(y) - \phi_B - V_{ch}(y)}{2V_t}} \cdot \sqrt{\frac{Q_{inv}(y)}{Q_{inv}(y) + 5C_{Si}V_t}} \tag{3.15}$$
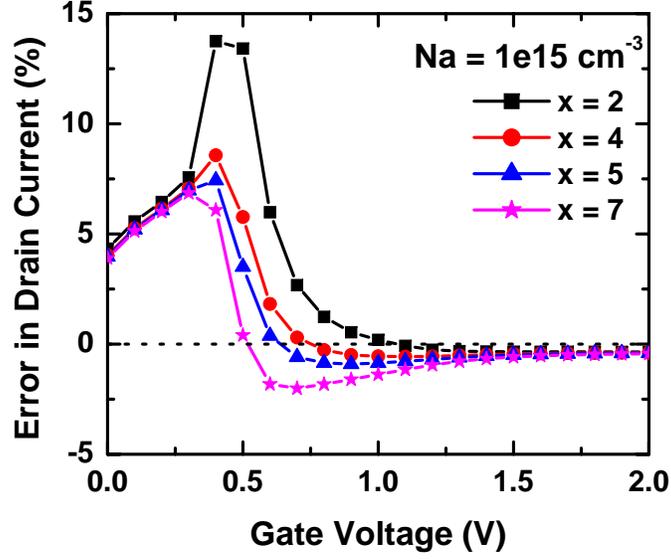
Figure 3.4: Error in the model predicted drain current of lightly doped DG-FET for different x when compared with 2-D TCAD simulations. Least error is obtained for an optimum $x = 5$.

3.2.2 Inversion Charge in Heavily Doped DG-FET

Recall that the total charge in the silicon body is defined as

$$Q_{Si}(y) = \sqrt{2qn_i\epsilon_{Si}\left(\frac{e^{\frac{q\psi_s(y)}{kT}} - e^{\frac{q\psi_o(y)}{kT}}}{q/kT} \cdot e^{\frac{-q(\phi_B+V_{ch}(y))}{kT}} + e^{\frac{q\phi_B}{kT}} \cdot (\psi_s(y) - \psi_o(y))\right)} \quad (3.16)$$

In the strong inversion regime, the exponential $\psi_s(y)$ term is large and the bulk charge is less than the inversion charge. For a heavily doped DG-FET, in the weak inversion regime the inversion carrier density is much smaller than the bulk charge and hence

$$\psi_s(y) - \psi_0(y) \approx \psi_{pert} \quad (3.17)$$

Using these assumptions and the definition of Q_{bulk}, the total charge in the silicon body can be approximated as

$$Q_{Si}(y) \approx \sqrt{2qn_i\epsilon_{Si}V_t \cdot e^{\frac{q\psi_s(y)}{kT}} \cdot e^{\frac{-q(\phi_B+V_{ch}(y))}{kT}} + Q_{bulk}^2} \qquad (3.18)$$

Utilizing the fact that $Q_{Si} = Q_{inv} + Q_{bulk}$, one can rearrange the above expression and express the inversion charge density of a heavily doped DG-FET, $Q_{inv:HD}$, as

$$Q_{inv:HD}(y) \approx \sqrt{2qn_i\epsilon_{Si}V_t} \cdot e^{\frac{\psi_s(y)-\phi_B-V_{ch}(y)}{2V_t}} \cdot \sqrt{\frac{Q_{inv}(y)}{Q_{inv}(y)+2Q_{bulk}}} \qquad (3.19)$$

Notice the similarity in the expressions for the approximate inversion charge density of lightly doped DG-FET (Eq. 3.15) and heavily doped DG-FET (Eq. 3.19). This striking similarity will be put to use in the next section in developing a unified approximate inversion charge model over all doping ranges.

3.2.3 Unified Charge Density Model

To have one analytical drain current model which is accurate over all doping ranges, it is necessary to have a unified charge density model. Looking at the expressions of approximate Q_{inv} for lightly doped DG-FETs and heavily doped DG-FETs, the following model is proposed to predict the inversion charge density accurately for wide range of body doping

$$Q_{inv}(y) \approx \sqrt{2qn_i\epsilon_{Si}V_t} \cdot e^{\frac{\psi_s(y)-\phi_B-V_{ch}(y)}{2V_t}} \cdot \sqrt{\frac{Q_{inv}(y)}{Q_{inv}(y)+2Q_{bulk}+5C_{Si}V_t}} \qquad (3.20)$$
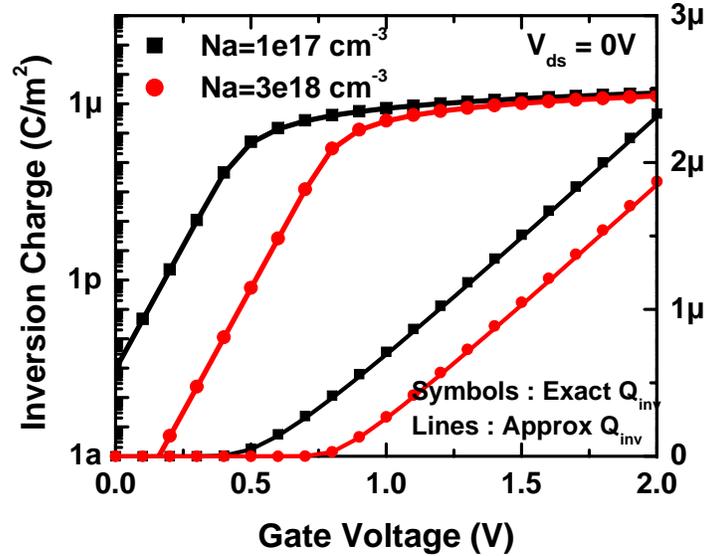
Figure 3.5: Inversion charge calculated using the unified charge density model for lightly doped DG-FET and heavily doped DG-FET. The unified charge density model agrees well with the inversion charge density calculated using exact charge equation.

When the body doping is low, the term Q_{bulk} can be ignored and the model is similar to Eq. 3.15 for lightly doped DG-FETs. Similarly, when the body doping is high the term $5C_{Si}V_t$ can be neglected compared to Q_{bulk} and the model is similar to Eq. 3.19 for heavily doped DG-FETs.

The unified charge model is verified against the exact charge equation in Eq. 3.3. Fig. 3.5 shows the inversion charge density calculated for both $N_A = 1\text{e}17\text{cm}^{-3}$ and $N_A = 3\text{e}18\text{cm}^{-3}$ using the unified inversion charge density model. The model predicts the inversion charge density very well when compared with charge density from the exact charge equation. The unified charge density model is used to calculate the drain current in the next section.

3.2.4   Drain Current Model

An analytical model for the drain current for DG-FETs which captures the effect of finite body doping can be developed using the unified charge density expression. To evaluate the integral in the drift-diffusion formulation for the drain current (Eq. 3.2) the gradient in the quasi-fermi potential can be expressed in terms of Q_{inv}. From Eq. 3.20, the gradient in V_{ch} can be written as

$$\frac{dV_{ch}}{dy} = \frac{d\psi_s}{dy} + V_t \frac{dQ_{inv}}{dy} \left(\frac{2Q_{bulk} + 5C_{Si}V_t}{Q_{inv} + 2Q_{bulk} + 5C_{Si}V_t} - \frac{2}{Q_{inv}} \right) \tag{3.21}$$

The drain current is then simply found by integrating Eq. 3.2.

$$I_d = 2 \cdot \mu \cdot \frac{W}{L} \cdot \left(h\left(Q_{invs}\right) - h\left(Q_{invd}\right) \right) \tag{3.22}$$

where Q_{invs} and Q_{invd} are inversion charge density at the source and drain end and the function $h(Q)$ is

$$h(Q) = \frac{Q^2}{2C_{ox}} + 2V_t Q - V_t \cdot (2Q_{bulk} + 5C_{Si}V_t) \cdot \ln\left(1 + \frac{Q}{2Q_{bulk} + 5V_t C_{Si}}\right) \tag{3.23}$$

Eq. 3.22 and Eq. 3.23 together constitute the core I-V model for DG-FETs and have been implemented in BSIM-CMG model. Note that the approximate charge density model was developed only to enable the formulation of an analytical drain current model. To maintain high accuracy of the drain current model, Q_{invs} and Q_{invd} are

still evaluated using Gauss's Law instead of the approximate charge density model.

$$Q_{invs} = C_{ox} (V_{gs} - V_{fb} - \phi_s) - Q_{bulk} \tag{3.24}$$

$$Q_{invd} = C_{ox} (V_{gs} - V_{fb} - \phi_d) - Q_{bulk}$$

where ϕ_s and ϕ_d are the surface potentials at the source and drain end respectively.

The I-V model has been verified against 2-D TCAD simulations without using any fitting parameters. Fig. 3.6(a) shows the I_d-V_{ds} characteristics of a long channel DG-FET with body doping of $N_A = 3e18$cm^{-3}. Fig. 3.6(b) shows the I_d-V_{gs} characteristics for the same DG-FET. The model shows good agreement with 2-D TCAD simulations in all the regimes of operation : sub-threshold, linear and saturation regimes.

One of the main features of this I-V model is the ability to predict the drain current for DG-FETs over a wide range of body doping. The model is used to calculate the drain current for DG-FETs starting from low body doping of $N_A = 1e15$cm^{-3} to a heavy body doping of $N_A = 1e19$cm^{-3}. As shown in the Fig. 3.7, the model prediction is very accurate over the wide range of body doping in both sub-threshold and linear region. Furthermore, the smooth transition from partial depletion to full depletion regimes can be seen clearly for DG-FETs with heavy body doping. Note that the excellent agreement to the 2-D TCAD simulations in all these cases has been obtained without the use of any fitting parameters.

A unique property of the lightly doped thin body DG-FETs is volume inversion (a.k.a bulk inversion) [42] which states that the drain current is linearly proportional to the body thickness in the sub-threshold regime. This phenomenon can be easily
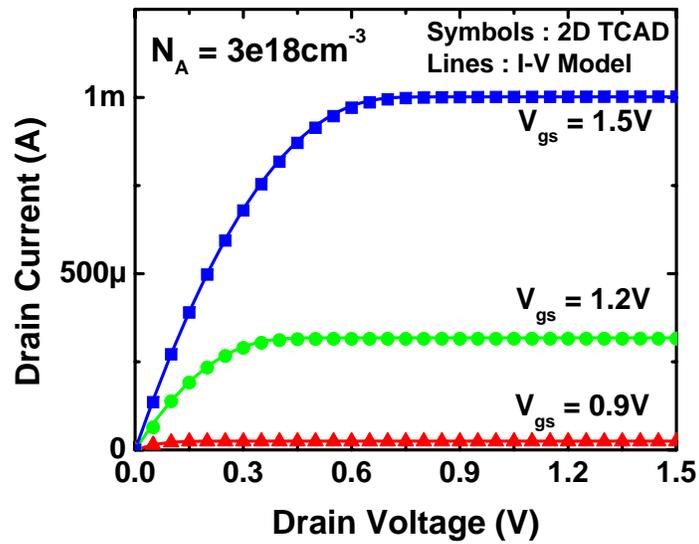
(a)
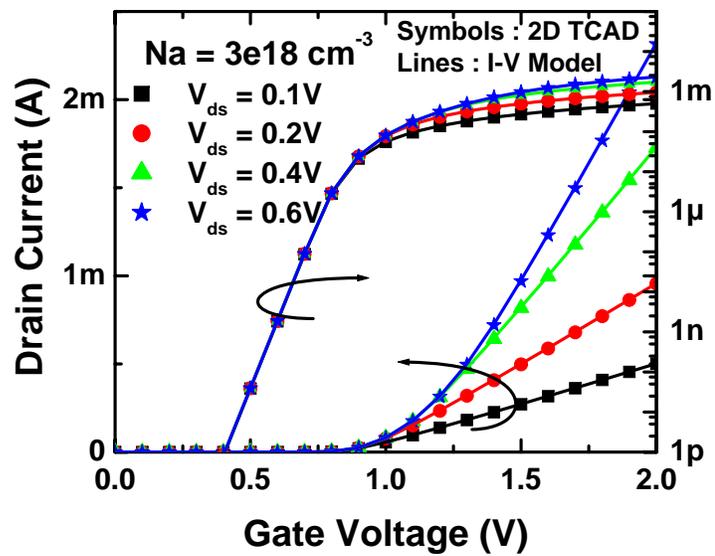
(b)

Figure 3.6: (a) I_d-V_{ds} and (b) I_d-V_{gs} of a DG-FET with heavy body doping of $N_A = $ 3e18cm^{-3}. Model predicted I_d agrees very well with 2-D TCAD simulations in all regimes of transistor operation.
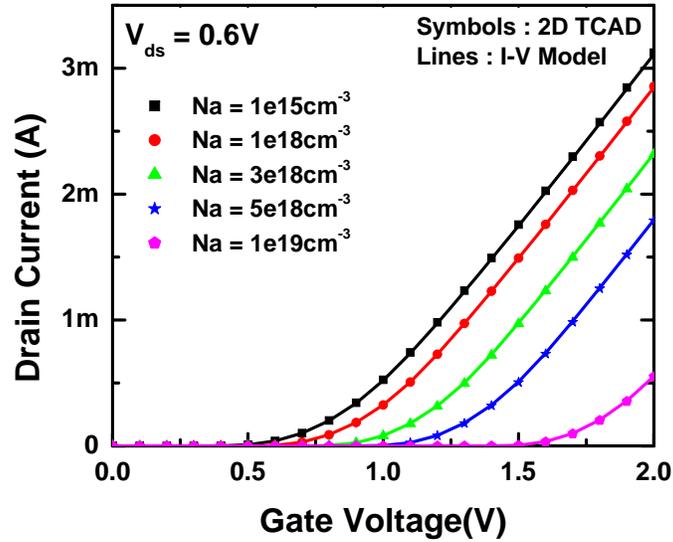
(a)
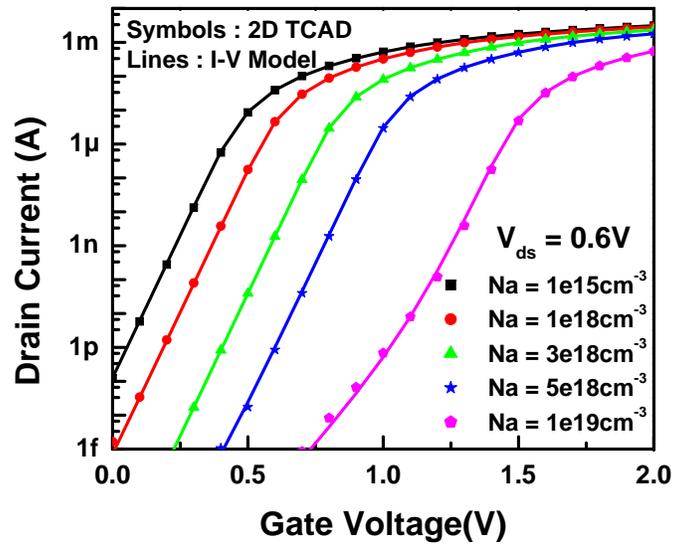
(b)

Figure 3.7: Verification of the I-V model against 2-D TCAD for both lightly doped DG-FETs and heavily doped DG-FETs. I-V model agrees very well with 2D TCAD for large range of body doping without using any fitting parameters. The transition from partial depletion to full depletion is very smooth as for the case of $N_A = $ 1e19cm^{-3}.
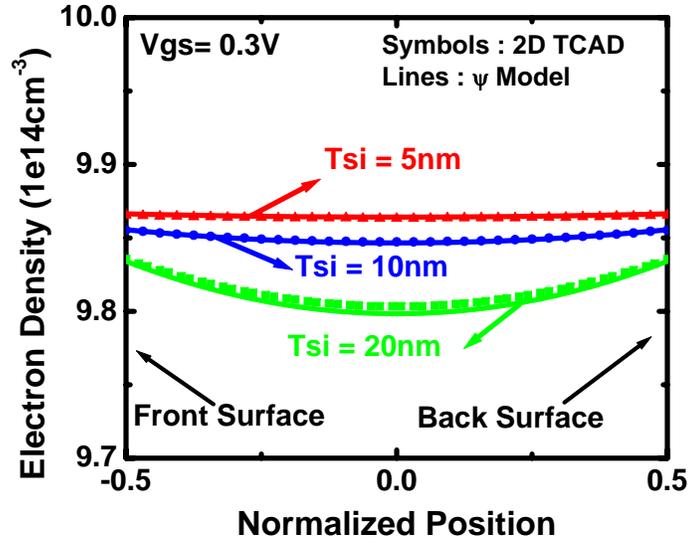
Figure 3.8: Carrier density profile along the depth of silicon body for a n-type DG-FET with $N_A = $ 1e15cm^{-3} in the sub-threshold regime for three different T_{Si}. The nearly constant electron density marks the origin of volume inversion.

explained using the surface potential model framework developed in Chapter 2.

For lightly doped DG-FET, in the sub-threshold regime, the surface potential is almost same as the center potential. As a result, the inversion charge density is nearly constant throughout the body thickness. Fig. 3.8 shows the electron density profile in the sub-threshold regime from the front to the back surface for a n-type DG-FET with $N_A = $ 1e15cm^{-3} for three different body thicknesses. Firstly, Fig. 3.8 shows that electron density is nearly independent of the position inside the body. Secondly, it is also independent of the body thickness. It can also be seen that the surface potential model predicts very good carrier profiles when compared against 2-D TCAD simulations. Finally, the total integrated charge inside the body is proportional to the body thickness. As a result, in the sub-threshold regime the drain current is also propor-
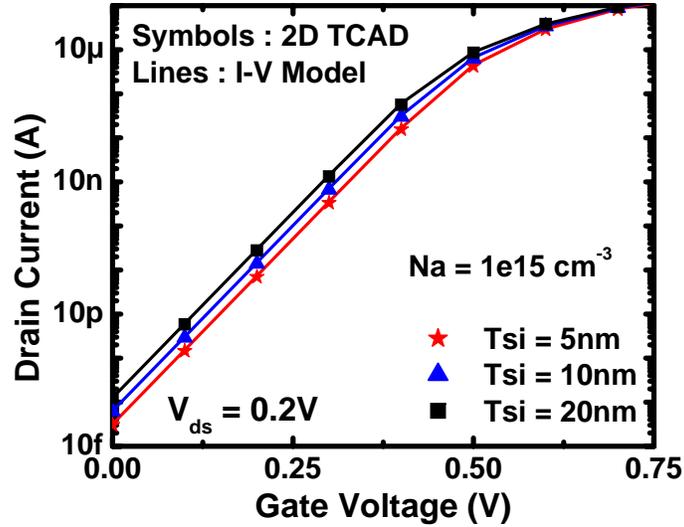
Figure 3.9: I_d-V_{gs} in the sub-threshold regime for n-type DG-FET with $N_A = 1e15cm^{-3}$. Linear relationship of I_d with respect to T_{Si} in sub-threshold regime is the signature of volume inversion. The I-V model captures volume inversion very well.

tional to T_{Si}. Since the I-V model does not make any charge sheet approximation, it should be able to capture volume inversion. Fig. 3.9 shows that the drain current predicted by the I-V model is linearly related to T_{Si} in the sub-threshold regime for DG-FET with $N_A = 1e15cm^{-3}$ which further attests to the accuracy of the model.

The core drain current model predicts the drain current only for long channel transistors. A start-of-the-art transistor with $L < 45nm$ will exhibit numerous other physical effects which will alter the drain characteristics significantly. In the next section, modeling of some of the important short channel effects will be described.

3.3   Modeling Short Channel Effects (SCE)

Short channel MG-FETs are plagued by the same SCE as the bulk transistors. How-
ever, because of the stronger electrostatic control from use of multiple gates, the SCE
are not as severe as in the bulk MOSFETs. Nevertheless, the SCE in MG-FETs are
still strong enough to warrant attention to modeling them [43]. Some of the important
SCE are

- Threshold voltage (V_{th}) roll-off

- Sub-threshold slope (n) degradation

- Drain Induced Barrier Lowering (DIBL)

- Channel Length Modulation (CLM)

- Carrier velocity saturation/overshoot

In this section only the former three which are related to lowering of potential barrier
at the source end are discussed. A SCE model must be scalable over a wide range of
device parameters such as gate length L, gate insulator thickness T_{ox}, body thickness
T_{Si}, fin height H_{fin} and channel doping N_A.

SCE are essentially 2-D effects where the drain significantly effects the potential
barrier at the source due to its close proximity to source region in a short channel
transistor. There have been numerous efforts in the past to model the SCE in DG-
FETs. Invariably all the methods to model SCE solve the 2-D Poisson's equation
in sub-threshold region inside the silicon body with varying degree of simplifying

assumptions. Some of the methods are better suited to implementation in a compact model environment while some are not. In [44] the authors solve the complete 2-D boundary value problem by expressing the potential through a infinite sum of sin functions. The solution is numerically complex and will be hard to extend to the case of MG-FETs with 3 or 4 gates. In [45] the authors first solve a 1-D Poisson equation in the channel length direction and then solve the 2-D Poisson equation. The solution in this case is independent of V_{ds} and hence modeling DIBL is a challenge in this approach. Another approach assumes a parabolic potential function perpendicular to the silicon-insulator interface and solves the 2-D Poisson's equation [46], [47]. This approach maintains a balance between the model accuracy and model computation time and hence is used to develop a SCE model for DG-FETs.

Start with the 2-D Poisson's equation inside the body in the sub-threshold regime where the inversion carriers can be neglected in comparison to the bulk charge.

$$\frac{d^2\psi(x,y)}{dx^2} + \frac{d^2\psi(x,y)}{dy^2} = \frac{qN_A}{\epsilon_{Si}} \tag{3.25}$$

Assume a simple parabolic function in the x direction (Fig. 3.1) for the potential distribution [48].

$$\psi(x,y) = C_0(y) + C_1(y) \cdot x + C_2(y) \cdot x^2 \tag{3.26}$$

The three quantities C_0, C_1 and C_2 can be found by applying three boundary conditions.

$$
\begin{aligned}
\psi\left(x = \frac{T_{Si}}{2}, y\right) &= \psi_s(y) \\
\frac{d\psi(x,y)}{dx}\Big|_{x=0} &= 0 \\
\frac{d\psi(x,y)}{dx}\Big|_{x=\frac{T_{Si}}{2}} &= \frac{V_{gs} - V_{fb} - \psi_s}{T_{ox}} \cdot \frac{\epsilon_{ox}}{\epsilon_{Si}}
\end{aligned}
\tag{3.27}
$$

where $\psi_s(y)$ is the surface potential. Solving the above boundary conditions, the potential profile in the body can be expressed as

$$
\psi(x,y) = \psi_s(y) - \frac{V_{gs} - V_{fb} - \psi_s(y)}{T_{ox}} \cdot \frac{\epsilon_{ox}}{\epsilon_{Si}} \cdot \frac{T_{Si}}{4} + \frac{(V_{gs} - V_{fb} - \psi_s(y))\,\epsilon_{ox}}{T_{ox}T_{Si}\epsilon_{Si}} \cdot x^2 \tag{3.28}
$$

The SCE are determined by the minimum potential barrier $(\psi_{c(min)})$ seen by the carriers entering at the source end. In the DG-FETs, the minimum barrier height exists at the center plane of the body. To determine $\psi_{c(min)}$ it is necessary to transform the 2-D Poisson's equation into a 1-D problem of potential distribution along the center plane. This can be achieved by expressing the potential profile in terms of center plane potential $\psi_c(y)$. The center plane potential is obtained by evaluating Eq. 3.28 at $x = 0$.

$$
\psi_c(y) = \psi_s(y) - \frac{V_{gs} - V_{fb} - \psi_s(y)}{T_{ox}} \cdot \frac{\epsilon_{ox}}{\epsilon_{Si}} \cdot \frac{T_{Si}}{4} \tag{3.29}
$$

Using this definition of $\psi_c(y)$, one can express the potential profile $\psi(x, y)$ in terms of $\psi_c(y)$ instead of $\psi_s(y)$ and the 2-D Poisson's equation 3.25 simply becomes

$$\frac{d^2\psi_c(y)}{dy^2} + \frac{V_{gs} - V_{fb} - \psi_c(y)}{\lambda^2} = \frac{qN_A}{\epsilon_{Si}} \tag{3.30}$$

where λ is defined as

$$\lambda = \sqrt{\frac{\epsilon_{Si}}{2\epsilon_{ox}} \left(1 + \frac{\epsilon_{ox}T_{Si}}{4\epsilon_{Si}T_{ox}}\right) T_{Si}T_{ox}} \tag{3.31}$$

λ is a characteristic field penetration length, also known as scale length, which defines the amount of SCE in the transistor. It captures the variation in the amount of drain field penetrating into the silicon body as a function of physical parameters such as T_{Si} and T_{ox}. Eq. 3.30 describes the variation of the potential at the center of the body along the channel length from source to drain. The center potential $\psi_c(y)$ can be determined by solving Eq. 3.30 under the following constraints.

$$\psi_c(y = 0) = V_{bi} \tag{3.32}$$

$$\psi_c(y = L) = V_{bi} + V_{ds}$$

where V_{bi} is the built-in potential at the source end. Upon solving Eq. 3.30 the center potential is given by

$$\psi_c(y) = V_{SL} + (V_{bi} - V_{SL})\frac{\sinh\left(\frac{L-y}{\lambda}\right)}{\sinh\left(\frac{L}{\lambda}\right)} + (V_{bi} + V_{ds} - V_{SL})\frac{\sinh\left(\frac{y}{\lambda}\right)}{\sinh\left(\frac{L}{\lambda}\right)} \tag{3.33}$$

where V_{SL} is equivalent to the center potential for a long channel transistor

$$V_{SL} = V_{gs} - V_{fb} - \frac{qN_A}{\epsilon_{Si}}\lambda^2 \tag{3.34}$$

The minimum potential along the center of body determines the short channel effects. This potential $\psi_{c(min)}$ can be easily calculated by finding the minimum of Eq. 3.33.

$$\psi_{c(min)} = V_{SL} - \frac{V_{ds}^2 \cdot e^{-L/2\lambda}}{\left(e^{L/\lambda} - e^{-L/\lambda}\right)\sqrt{Z_0 Z_L}} + 2\sqrt{Z_0 Z_L} \cdot \frac{\sinh(L/2\lambda)}{\sinh(L/\lambda)} \tag{3.35}$$

where $Z_0 = V_{bi} - V_{SL}$ and $Z_L = V_{bi} + V_{ds} - V_{SL}$. For a long channel transistor, $\psi_{c(min)} \approx V_{SL} \propto V_{gs}$. The surface potential model developed in previous chapter is a long channel framework. To obtain the correct minimum barrier height for a short channel DG-FET at a certain V_{gs} using the long channel surface potential model, one can use the long channel surface potential model with an effective V_{gs}' where

$$V_{gs}' = V_{gs} + \underbrace{\left(\psi_{c(min)} - V_{SL}\right)}_{\Delta V_{gs}} \tag{3.36}$$

This method can be very easily implemented into any compact model by simply changing the gate voltage bias according to Eq. 3.36. The new gate voltage V_{gs}' is used to calculate the surface potential for the short channel device and the new surface potential is used in the I-V model to calculate the drain current. The SCE model in Eq. 3.36 not only captures the V_{th} roll-off and DIBL, but it also captures the degradation of sub-threshold slope simultaneously since the correction term ΔV_{gs} is a function of gate voltage.
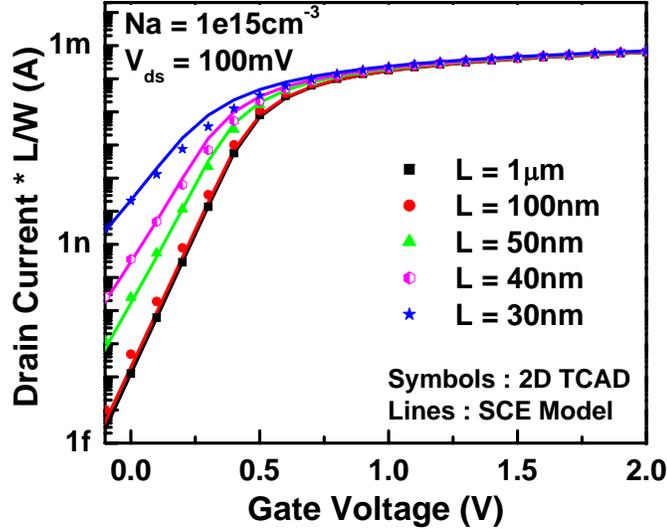
Figure 3.10: Normalized drain current for DG-FETs with different channel lengths, from $L = 1\mu$m down to $L = 30$nm. All the important SCE are captured by the model and agree well with 2-D TCAD simulations.

The SCE model is verified against 2-D TCAD simulations. Drain current is calculated using the SCE and I-V model for a wide range of channel lengths ranging from a long channel $L = 1\mu$m DG-FET to a short channel DG-FET of $L = 30$nm. Fig. 3.10 shows the $I_{ds} * L/W$ calculated from the model against the current obtained from 2-D TCAD simulations. Good agreement in the SCE behavior such as V_{th} roll-off and n degradation is observed between the model and 2-D TCAD. Closer match can be easily obtained by introducing fitting parameters into the SCE model.

Good scalability of the SCE model is highly desirable. A scalable model allows to stretch the limited available silicon data to perform technology projections. The scalability of the SCE model is examined extensively with respect to physical parameters such as L, T_{Si} and T_{ox}. In Fig. 3.11 the scalability of the SCE model with respect to
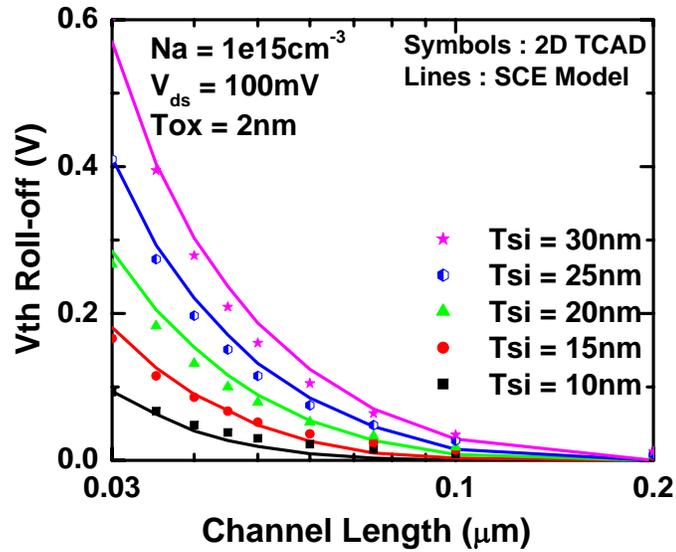
Figure 3.11: Threshold voltage roll-off of DG-FETs for different T_{Si}. Model demonstrates excellent scalability when compared against 2-D TCAD simulations.
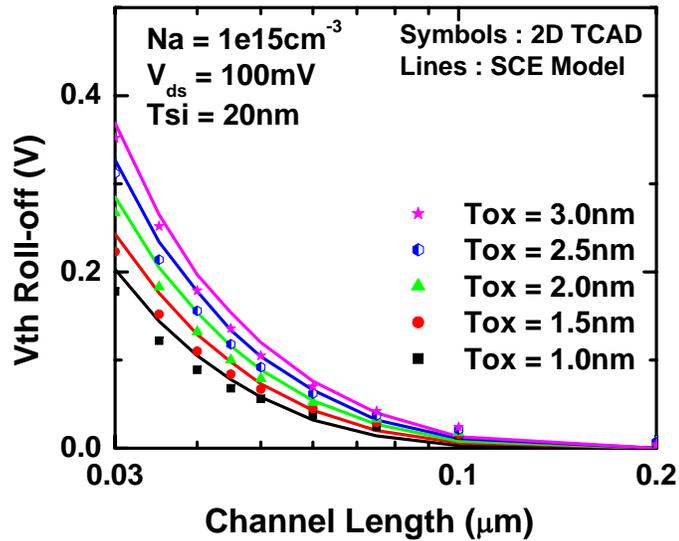
Figure 3.12: Threshold voltage roll-off of DG-FETs for different T_{ox}. Model demonstrates excellent scalability when compared against 2-D TCAD simulations.

T_{Si} is examined. Fig. 3.11 shows the V_{th} roll-off extracted from the model and 2-D TCAD for different T_{Si}. Here V_{th} is defined using a constant current definition. As expected, DG-FETs with smaller T_{Si} have smaller drain field penetration and hence the model predicts smaller V_{th} roll-off for thinner T_{Si}. Fig. 3.12 shows the roll-off in V_{th} extracted from the model for different T_{ox} and compares it against the V_{th} extracted from 2-D TCAD simulations . The smaller the T_{ox}, the closer the gate electrode is to the inversion layer enforcing a stronger electrostatic gate control and hence less V_{th} roll-off. The V_{th} roll-off predicted by the analytical model for different T_{Si} and T_{ox} agree very closely with those obtained from 2-D TCAD simulations confirming the scalability of the SCE model.

Thus far, only DG-FETs, i.e. MG-FETs with gate electrodes only on two opposite sides of slicon body, have been analyzed. MG-FETs can have three or four gate electrodes to further improve the electrostatic control. In MG-FETs with three or more gates, the penetration of the drain field is minimized and so the characteristic drain field penetration length λ is different. The current SCE modeling framework for DG-FETs can be easily extended to the model MG-FETs by simply changing λ. Based on [49], a new λ_{eff} is defined to include the effect of fin height H_{fin}.

$$\lambda_{eff} = \frac{1}{\sqrt{\frac{1}{\lambda}^2 + \frac{A}{\lambda_{Hfin}}^2}} \tag{3.37}$$

where

$$\lambda_{Hfin} = \sqrt{\frac{\epsilon_{Si}}{4\epsilon_{ox}}\left(1 + \frac{\epsilon_{ox}T_{Si}}{2\epsilon_{Si}T_{ox}}\right)H_{fin}T_{ox}} \tag{3.38}$$

and

$$
\begin{aligned}
A &= 0 && \text{for DG} - \text{FET} && (3.39) \\
&= \frac{1}{2} && \text{for Triple} - \text{gate FET} \\
&= 1 && \text{for Surround} - \text{gate FET}
\end{aligned}
$$

The new field penetration field λ_{eff} has been extensively verified against 3-D TCAD simulations for different MG-FET configurations in [50].

3.4 C-V Model

A compact model is virtually unusable for circuit simulation without a capacitance model. The I-V model is useful only for DC analysis and operating point analysis where the voltages in the circuit are time independent. In real world, almost all the circuits run with either periodic or aperiodic time varying voltages, be it large signal or small signal. The functionality of such circuits is determined through AC and transient analysis. AC analysis needs the capacitances associated with the terminals of the transistor. To analyze the transient behavior it is essential to model the charges stored in the transistor. The C-V model defines both the charges and the associated capacitances for the transistor. The capacitance model consists of capacitances internal to the transistor and capacitances which are extrinsic such as overlap capacitances and fringe capacitances. Several works can be found for modeling extrinsic capacitances for MG-FETs [51], [52]. This section describes the formulation of only

the intrinsic capacitances.

The intrinsic capacitances of the DG-FET are derived from the terminal charges. As a result, the first step in formulating the C-V model is to find the charges stored in the transistor. The charge on the top and bottom gate electrodes is equal to the total integrated charge in the body. The charge located in a vertical plane at a distance y from the source end is given by

$$Q_{Si}(y) = 2WC_{ox}\left(V_{gs} - V_{fb} - \psi_s(y)\right) \tag{3.40}$$

The total integrated charge in the body is computed by integrating the charge along the channel from the source end to the drain end. Since the two gates are electrically connected,

$$Q_g = 2WC_{ox}\int_0^L \left(V_{gs} - V_{fb} - \psi_s(y)\right) dy \tag{3.41}$$

where Q_g denotes the charge on the electrically interconnected gate. The inversion charge in the body is divided between the source and drain terminals using the Ward-Dutton charge partition approach [40]. The charge on the source terminal Q_s is given by

$$Q_s = -2WC_{ox}\int_0^L \left(1 - \frac{y}{L}\right) \cdot \left(V_{gs} - V_{fb} - \psi_s(y) - \frac{Q_{bulk}}{C_{ox}}\right) dy \tag{3.42}$$

Since the total charge has to be conserved, the charge on drain terminal Q_d can be

expressed as

$$Q_d = -2WC_{ox} \int_0^L \frac{y}{L} \cdot \left(V_{gs} - V_{fb} - \psi_s(y) - \frac{Q_{bulk}}{C_{ox}} \right) dy \qquad (3.43)$$

To evaluate the above charge integrals, the surface potential along the channel length $\psi_s(y)$ is needed. $\psi_s(y)$ is obtained using current continuity. Current continuity states that the current is conserved all along the length of the transistor under quasi-static operation of the transistor.

$$I_d(y) = I_d(L) \quad \forall \ \ 0 \le y \le L \qquad (3.44)$$

Recall that the drain current is given by Eq. 3.22-3.23. The expression for drain current is not practical for applying current continuity. For the purpose of determining $\psi_s(y)$, a simplified expression of drain current as shown below is used.

$$I_d \ \ = \ \ 2 \cdot \mu \cdot \frac{W}{L} \cdot (h_{cv}(Q_{invs}) - h_{cv}(Q_{invd})) \qquad (3.45)$$

where the function $h_{cv}(Q)$ is defined as

$$h_{cv}(Q) = \frac{Q^2}{2C_{ox}} + 2V_t Q \qquad (3.46)$$

The function $h_{cv}(Q)$ is a simplified version of $h(Q)$ in Eq. 3.23. Fig. 3.13 shows the drain current as predicted by this simplified I-V model. It can be seen that $h_{cv}(Q)$ retains the drain current accuracy in strong inversion regime but overestimates the
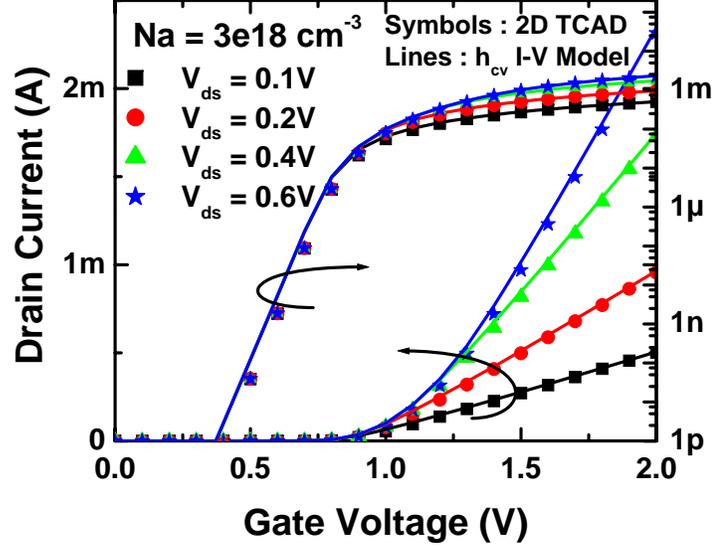
Figure 3.13: I_d-V_{ds} of a DG-FET with heavy body doping of $N_A = $ 3e18cm^{-3} using the approximate I-V model (Eq. 3.45) for determining $\psi_s(y)$. Model predicted I_d agrees very well with 2-D TCAD simulations in the strong inversion regime.

drain current in sub-threshold regime. Note that the approximate I-V model is used only for applying current continuity to determine terminal charges but not for calculating the drain current. The advantage of being able to obtain mathematically simple analytical expressions for terminal charges and capacitances due to this approach outweighs any resulting loss in accuracy which may occur in the C-V model in sub-threshold regime. Applying the current continuity to Eq. 3.45-3.46, $\psi_s(y)$ can be related to ϕ_s and ϕ_d through

$$\frac{y}{L} \cdot (B - \phi_s - \phi_d)(\phi_d - \phi_s) = (B - \phi_s - \psi_s(y))(\psi_s(y) - \phi_s) \qquad (3.47)$$

where

$$B = 2\left(V_{gs} - V_{fb} - \frac{Q_{bulk}}{C_{ox}} + 2V_t\right) \tag{3.48}$$

The terminal charges can now be obtained by substituting $\psi_s(y)$ in Eqs. 3.41-3.43 and evaluating the integrals.

$$
\begin{aligned}
\frac{Q_g}{2C_{ox}WL} &= \left(V_{gs} - V_{fb} - \frac{\phi_s + \phi_d}{2} + \frac{(\phi_d - \phi_s)^2}{6(B - \phi_d - \phi_s)}\right) \\
\frac{-Q_d}{2C_{ox}WL} &= \frac{V_{gs} - V_{fb} - \frac{Q_{bulk}}{C_{ox}}}{2} - \frac{\phi_s + \phi_d}{4} + \frac{(\phi_d - \phi_s)^2}{60(B - \phi_d - \phi_s)} \\
&\quad + \frac{(5B - 6\phi_s - 4\phi_d)(B - 2\phi_d)(\phi_s - \phi_d)}{60(B - \phi_d - \phi_s)^2} \\
Q_s &= -(Q_g + Q_{bulk} + Q_d)
\end{aligned}
\tag{3.49}
$$

The expressions for the terminal charges are continuous and are valid over sub-threshold, linear and saturation regimes of operation. Fig. 3.14 shows the terminal charges calculated using Eq. 3.49 as a function of V_{ds} for a DG-FET with heavy body doping. The ratio of the drain charge to source charge is 40/60 in the saturation region as seen in Fig. 3.14. This is due to the Ward-Dutton charge partition which is physically correct under quasi-static operation of the transistor. It can also be observed that $Q_d = Q_s$ when $V_{ds} = 0$ showing the symmetry property in the charge equations. Fig. 3.15 shows the variation of the terminal charges for the same DG-FET as a function of gate voltage.

Eq. 3.49 forms the C-V model for BSIM-CMG. The terminal charges are used as state variables in the circuit simulation. All transcapacitances are derived from
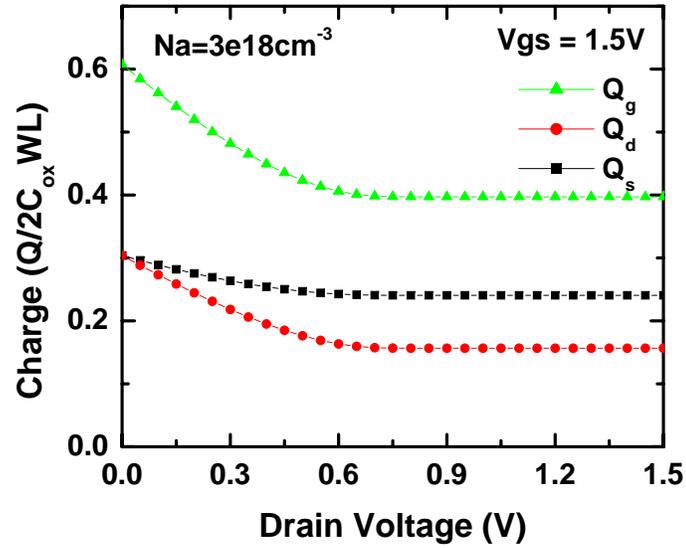
Figure 3.14: Terminal charges of a DG-FET with $N_A = 3e18cm^{-3}$ calculated using Eq. 3.49 as a function of V_{ds}. In the saturation region, Q_s and Q_d are in the ratio of 60/40 in accordance with Ward-Dutton charge partition.
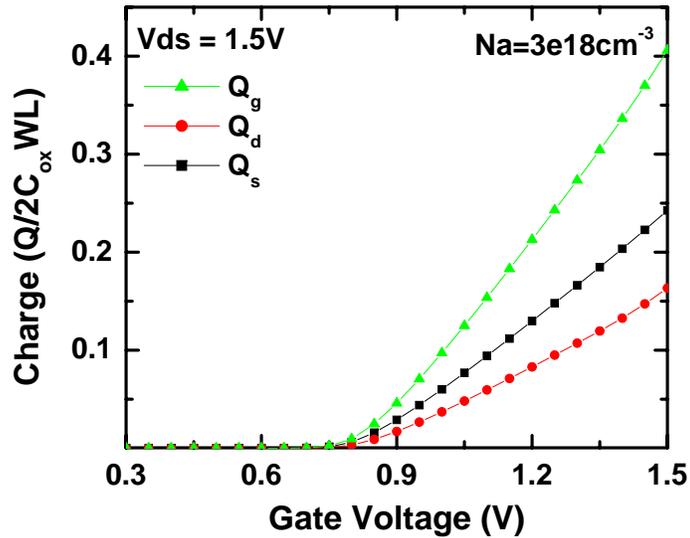
Figure 3.15: Terminal charges of a DG-FET with $N_A = 3e18cm^{-3}$ calculated using Eq. 3.49 as a function of V_{gs}.
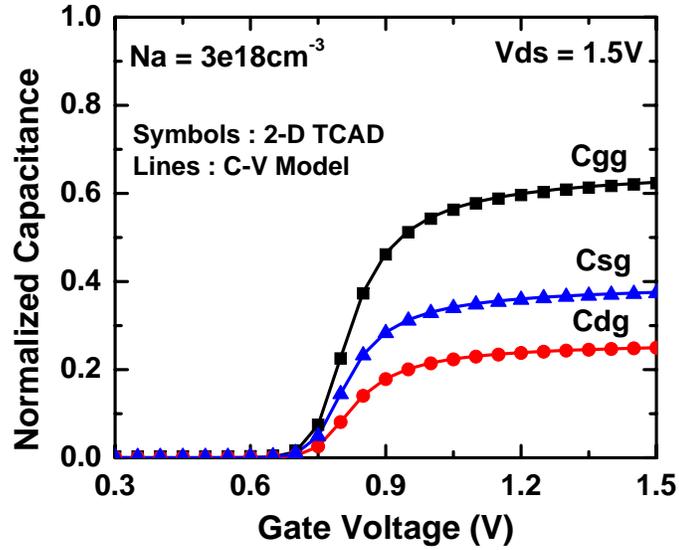
(a)
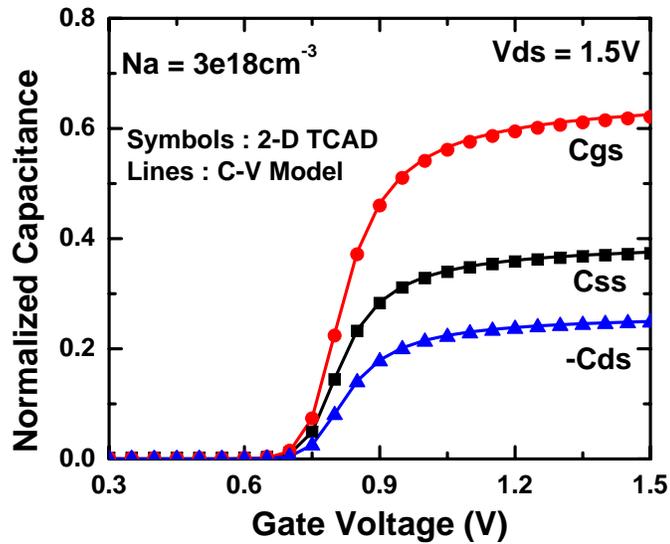
(b)

Figure 3.16: Capacitances (normalized to $2WLC_{ox}$ calculated from the C-V model and 2D TCAD as a function of V_{gs} for DG-FET with heavy body doping of $N_A = $ 3e18cm^{-3}. Good agreement between the model and TCAD is seen for all the capacitances without the use of any fitting parameters.
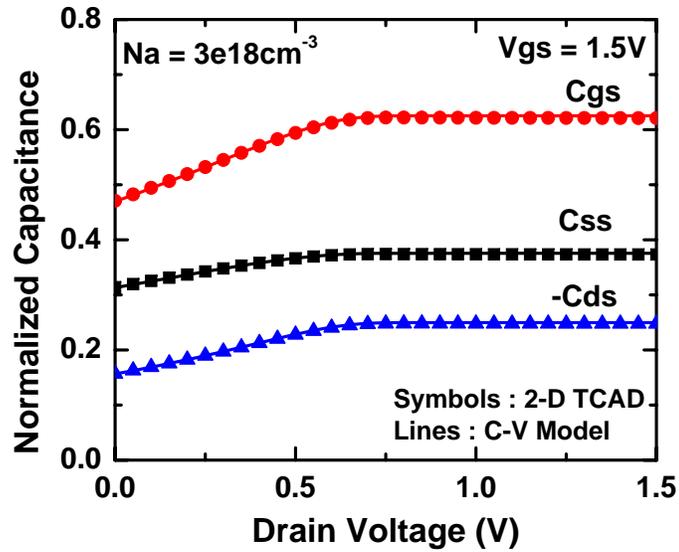
(a)
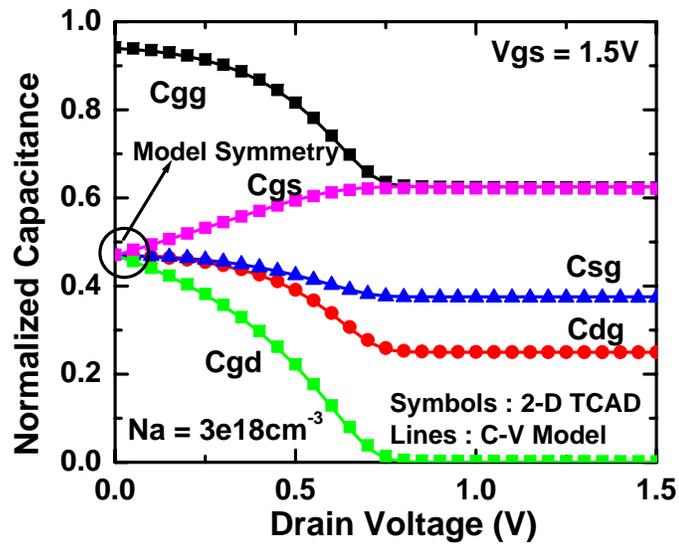
(b)

Figure 3.17: Capacitances (normalized to $2WLC_{ox}$ calculated from the C-V model and 2D TCAD as a function of V_{ds} for DG-FET with heavy body doping of $N_A = $ 3e18cm^{-3}. Good agreement between the model and TCAD is seen for all the capacitances without the use of any fitting parameters. The symmetry in the model is also clearly visible at V_{ds} = 0V.

the terminal charges to ensure charge conservation. The intrinsic capacitances are defined simply as

$$
\begin{aligned}
C_{ij} &= \frac{\partial Q_i}{\partial V_j} \qquad i \neq j \\
&= -\frac{\partial Q_i}{\partial V_j} \qquad i = j
\end{aligned}
\tag{3.50}
$$

where i and j denote the transistor terminals. Note that the capacitances C_{ij} satisfy

$$
\sum_i C_{ij} = \sum_j C_{ij} = 0
\tag{3.51}
$$

due to charge conservation [37].

The C-V model is verified against 2-D TCAD simulations without the use of any fitting parameters. The capacitances from the C-V model are plotted as a function of gate voltage and drain voltage in Fig. 3.16 and Fig. 3.17 respectively. The capacitance values from the model are in excellent agreement with TCAD simulations in all regimes of transistor operation. At $V_{ds} = 0$V, Fig. 3.17(b) shows that $C_{sg} = C_{dg}$ and $C_{gs} = C_{gd}$. This equality in capacitances at $V_{ds} = 0$V demonstrates the symmetry of the C-V model. Model symmetry is important for predicting correct distortion metrics for circuits switching about $V_{ds} = 0$V especially in the analog and RF world [53].

1) Quantum Mechanical Effects
2) Short Channel Effects
a) Vth roll-off
b) DIBL
c) Sub-threshold slope
d) Channel length modulation
3) Polysilicon-gate Depletion Effects (PDE)
4) Series resistance
5) Mobility degradation
6) Velocity Saturation
7) Velocity Overshoot/Source-End Velocity Limit
8) Gate Induced Drain (Source) Leakage (GIDL, GISL)
9) Impact Ionization
10) S/D Junction leakage
11) Gate tunneling
12) Parasitic capacitance

Table 3.1: List of physical effects modeled in BSIM-CMG

3.5   BSIM-CMG Model

The unified surface potential model together with the I-V model and C-V model for

DG-FET form the core of BSIM-CMG. The core model is highly predictive and has

a high degree of accuracy. However, it is only the beginning of any compact model.

BSIM-CMG, in the tradition of BSIM3 and BSIM4, models numerous physical phe-

nomena that are expected to be important in advanced multi-gate FET technolo-

gies. These additional phenomenon are added to the framework presented by the

core model. Table 3.1 presents a comprehensive list of real device effects which are

modeled in BSIM-CMG. Modeling of physical effects such as quantum mechanical

carrier confinement and short channel effects have already been described. Other

physical effects are modeled in BSIM-CMG using approaches similar to the ones used

in BSIM4 [20] and PSP [21] bulk transistor compact models but with appropriate

changes for MG-FETs. Leakage current mechanisms such as gate tunneling and gate-induced drain leakage are also modeled. The substrate current model comprising of impact ionization current and diode current extends the 3-terminal SOI MG-FET model to a 4-terminal bulk MG-FET. The model has been written in Verilog-A and implemented in popular circuit simulators such as SPECTRE and HSPICE.

3.6   Experimental Verification

The final step in testing a compact model is against real silicon data. The model BSIM-CMG has been verified against two different FinFET technologies - SOI Fin-FETs and bulk FinFETs [54]. The Verilog-A model was implemented in ICCAP environment and used to fit to the measured data. The model successfully described the measured drain current and its derivatives, transconductance (g_m) and output conductance (g_{ds}), for both long channel and short channel MG-FETs.

The SOI FinFETs were fabricated on a lightly doped 60nm thick film with 2nm SiO_2 dielectric and a strained $TiSiN$ gate [55]. The strained gate in turn strains the channel to enhance the electron mobility, hence increasing the current drive. Measured devices had 20 parallel fins, where each fin is 22nm thick. Fig. 3.18 shows the model fitting to $I_d - V_{gs}$ and its derivatives for a short channel $L = 90$nm SOI FinFET as a function of gate voltage in the linear and saturation regime. In Fig. 3.18 (a) precise modeling of physical phenomenon such as DIBL and GIDL is clearly visible. Fig. 3.18 (b) shows the g_m of the device and it highlights the model's ability to capture mobility degradation at large vertical electric field in both
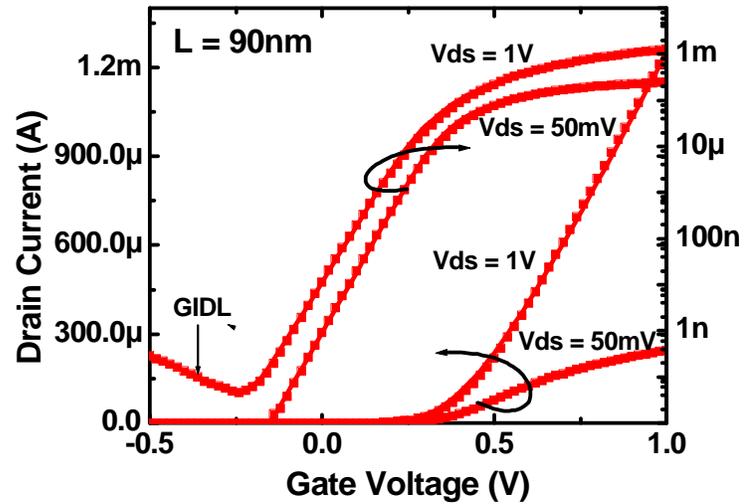
(a)
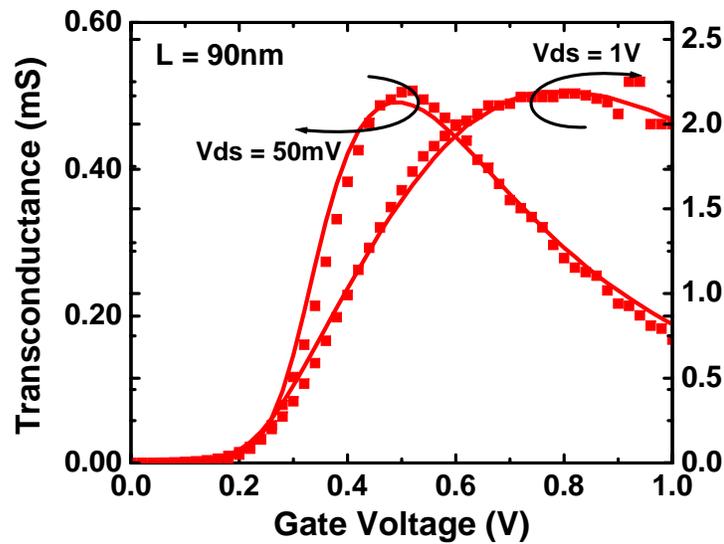
(b)

Figure 3.18: Model fitting to short channel $L = 90$nm SOI FinFET measurements (a) Drain Current (I_d) and (b) Transconductace (g_m) as a function of V_{gs} (Symbols: measured data, Lines : BSIM-CMG model). The model describes short channel SOI FinFETs very well.
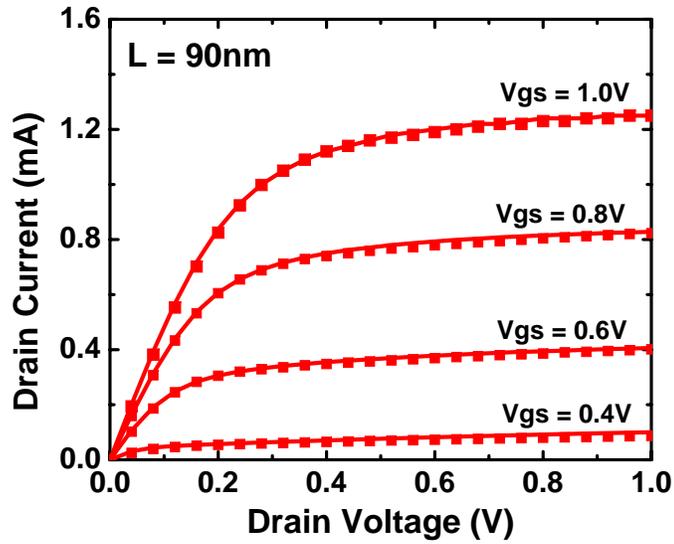
(a)
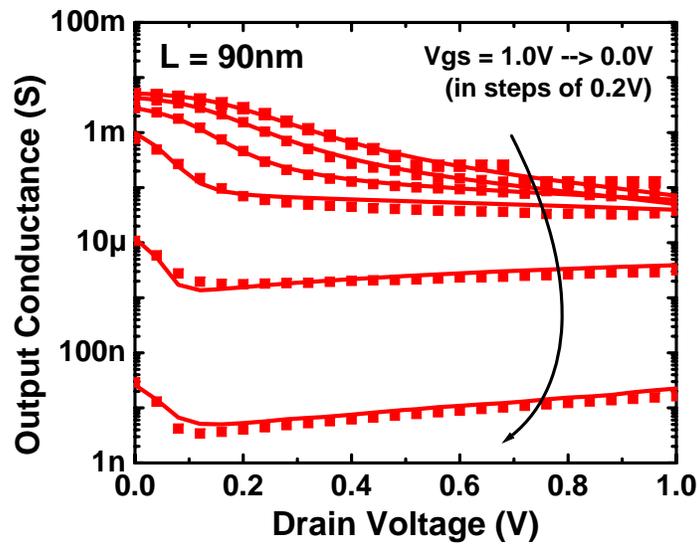
(b)

Figure 3.19: Model fitting to short channel $L = 90$nm SOI FinFET measurements (a) Drain Current (I_d) and (b) Transconductace (g_{ds}) as a function of V_{ds} (Symbols: measured data, Lines : BSIM-CMG model). The model describes short channel SOI FinFETs very well.
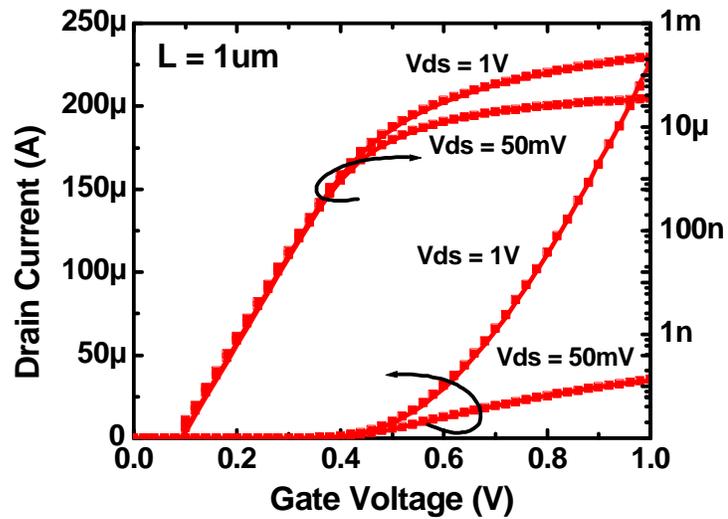
(a)
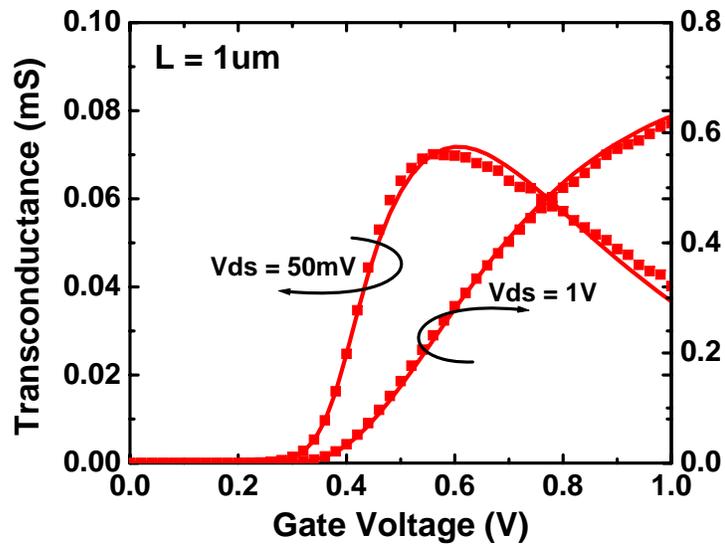
(b)

Figure 3.20: Model fitting to long channel $L = 1\mu m$ SOI FinFET measurements (a) Drain Current (I_d) and (b) Transconductace (g_m) as a function of V_{gs} (Symbols: measured data, Lines : BSIM-CMG model). The model describes long channel SOI Finfets very well.
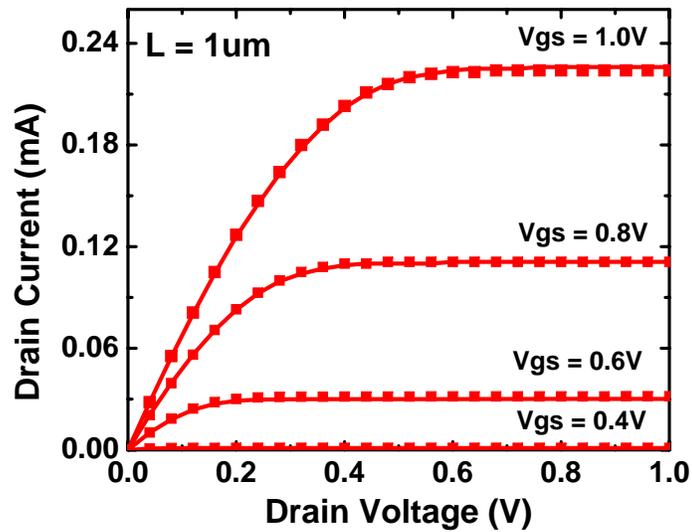
(a)
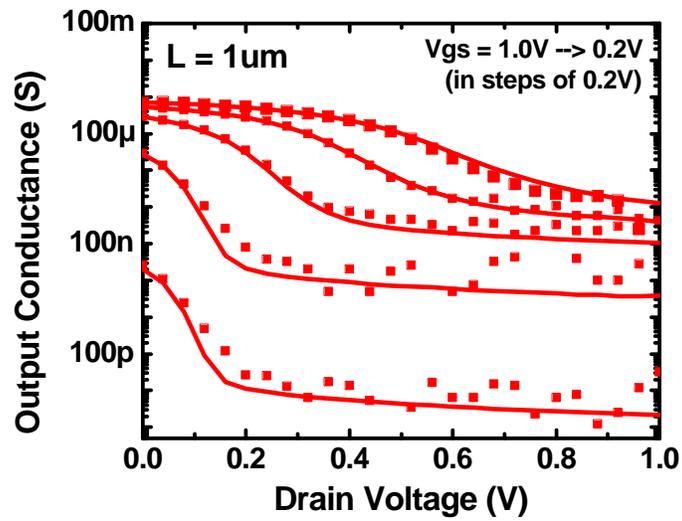
(b)

Figure 3.21: Model fitting to long channel $L = 1\mu m$ SOI FinFET measurements (a) Drain Current (I_d) and (b) Transconductace (g_{ds}) as a function of V_{ds} (Symbols: measured data, Lines : BSIM-CMG model). The model describes long channel SOI FinFETs very well.
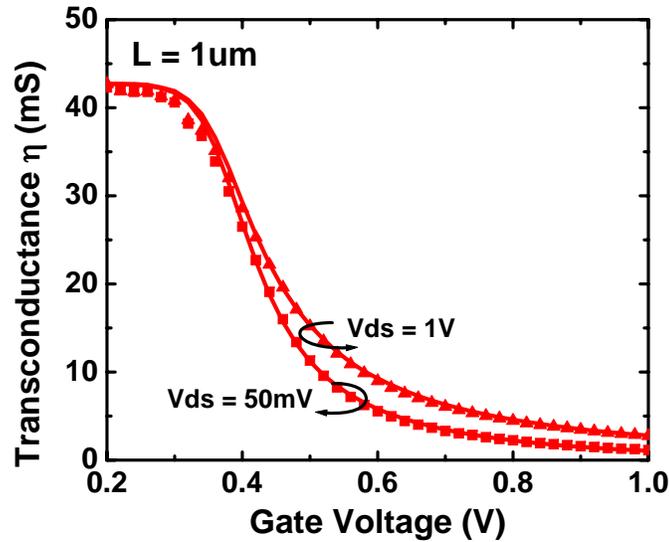
Figure 3.22: Model description of the g_m efficiency of the long channel $L = 1\mu$m SOI FinFET. (Symbols: measured data, Lines : BSIM-CMG model). The model is apt for use in both analog and digital design worlds.

linear and saturation regime. Fig. 3.19 illustrates the $I_d - V_{ds}$ and its derivative g_{ds} model fitting. Model captures the short channel phenomenon such as channel length modulation very well. BSIM-CMG model was also used to describe the $L = 1\mu$m long channel devices. The model fitting results to the long channel SOI FinFET device are shown in Fig. 3.20 and Fig. 3.21. To illustrate the strength of BSIM-CMG model for analog design, Fig. 3.22 shows the model description of transconductance efficiency (g_m/I_d) for the long channel SOI FinFET.

BSIM-CMG has also been verified against bulk FinFET measurements. Measured devices have 25nm thick fins and an EOT of 1.95nm. The model was again able to describe both the long and short channel transistors. Fig. 3.23 shows the measured short channel $(L = 50$nm$)$ characteristics and the corresponding BSIM-CMG fitting
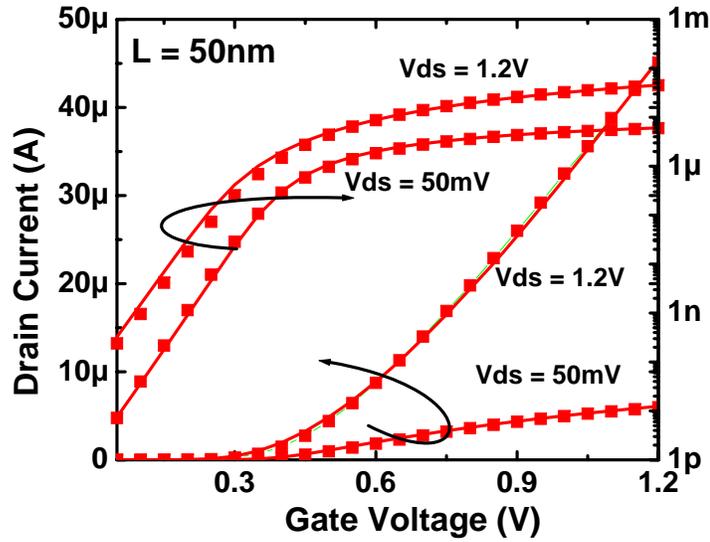
(a)
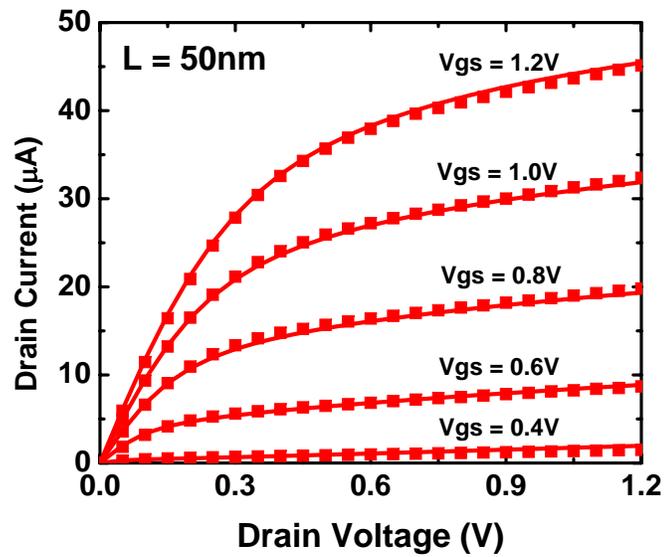
(b)

Figure 3.23: Model fitting to short channel $L = 50$nm bulk FinFET measurements (a) I_d-V_{gs} and (b) I_d-V_{ds} (Symbols: measured data, Lines : BSIM-CMG model). The model describes short channel bulk FinFETs very well.

results.

The experimental verification shows that BSIM-CMG model successfully captures the characteristics of advanced multi-gate FETs. Triple-gate multi-gate FETs were used for model verification demonstrating the ability of the model to capture phenomena such as corner effect which are unique to tri-gate and quadruple-gate FETs. The model is able to describe both SOI and bulk silicon based multi-gate FET technologies. Accurate description of the drain current and its derivatives warrants the use of BSIM-CMG for both digital and analog design.

3.7 Summary

A full scale compact model for common-gate symmetric Multi-gate FETs is developed in this chapter. An analytical drain current model valid over all regimes of transistor operation for long channel DG-FET is first developed without making any charge-sheet approximation. The I-V model exhibits excellent accuracy over a wide range of body doping. The I-V model is extended to short channel transistors and MG-FETs with more than two gates by careful modeling of short channel effects. Short channel effects such as threshold voltage roll-off and sub-threshold slope degradation are captured simultaneously. The I-V model shows excellent agreement with TCAD simulations without any fitting parameters. C-V model comprising of terminal charges and capacitances is also developed for MG-FETs with finite body doping. BSIM-CMG, a complete model for MG-FET based circuit simulation is then built using these I-V and C-V models. BSIM-CMG model is experimentally verified against

both SOI FinFET and bulk FinFET technologies for both long and short channel transistors. The model describes both analog and digital design metrics very well making it suitable for mixed-signal design applications.

Chapter 4

Independent Multi-Gate FET Model

4.1   Introduction

MG-FETs provide an elegant solution to the problem of planar bulk CMOS scaling by reducing the sub-threshold leakage due to the stronger electrostatic control provided by the multiple gate electrodes. One class of MG-FETs is the independent DG-FETs where the two gate electrodes are electrically independent and can be biased separately at two different gate voltages. Different front and back gate work-functions and gate dielectric thicknesses may be employed potentially making the DG-FETs asymmetric. The independent DG-FETs (from now on referred to as IDG-FETs) can be likened to ultra-thin body FETs or the fully depleted SOI-FETs with a thin back gate dielectric.

The flexibility of independent biasing of second gate electrode gives rise to several interesting applications and design benefits. The back gate can be biased to achieve

the desired threshold voltage without the need of significant body doping. This helps avoid the statistical fluctuations in threshold voltage due to random dopant fluctuations [56]. Threshold voltage tuning by back gate biasing can improve the circuit performance and provide power savings compared to common-gate DG-FETs [57]. The back gate biasing also opens a plethora of possible RF design applications such as RF mixers where a large LO signal can be applied on one gate and a small RF signal can be applied on the other gate [58]. In order to explore the IDG-FET based circuits, it is imperative to have a compact model for IDG-FETs. Some efforts have been made in developing such a compact model [59]. This chapter presents a new framework for modeling IDG-FETs which allows to borrow many features from BSIM-CMG model.

The model for IDG-FETs is also surface potential based, i.e. all the electrical variables inside the model such as currents, charges and capacitances are all functions of the surface potential at the source and drain end. A separate core IDG-FET model is developed for the independent double-gate FETs with a lightly-doped body. Determination of the surface potential for IDG-FETs is described in Section 4.2. Using the surface potential solution, an analytical I-V model and C-V model for IDG-FETs are derived in Section 4.3 and Section 4.4 respectively. The core model comprising of the surface potential model, I-V model and C-V model is extensively verified using 2-D TCAD simulations. Following the core model development, several physical effects are added to obtain a full scale compact model for IDG-FETs called BSIM-IMG (BSIM-Independent Multiple Gate FET model) [60].
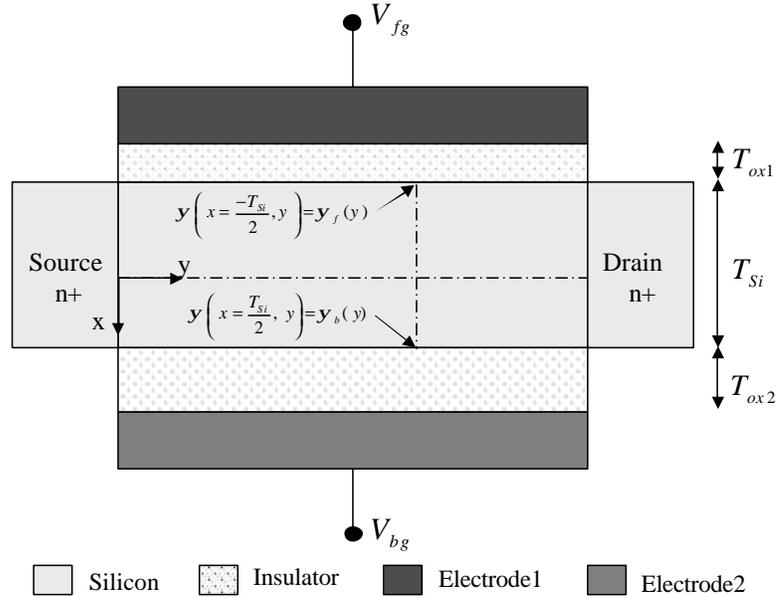
Figure 4.1: Schematic of the asymmetric independent DG-FET under study showing the asymmetry in the dielectric thickness and gate work function.

4.2 Surface Potential Model

Fig. 4.1 shows the schematic of IDG-FET under study. The asymmetric IDG-FET can have different front and back gate dielectric thickness (T_{ox1} and T_{ox2}) and different gate work-functions (ϕ_{m1} and ϕ_{m2}). Since the threshold voltage of an independent IDG-FET can be tuned by adjusting the back gate bias (V_{bg}), there is no need for significant doping in the body. The surface potential model for IDG-FET is developed assuming a lightly doped body.

The electronic potential in the body of a long channel IDG-FET is obtained by solving the 1-D Poisson's equation together with the Gauss's Law at the front and back surfaces as the boundary conditions. The 1-D Poisson's equation for the lightly

doped IDG-FET is

$$\frac{\partial^2 \psi(x,y)}{\partial x^2} = \frac{qn_i}{\epsilon_{Si}} \cdot e^{\frac{q(\psi(x,y)-V_{ch}(y))}{kT}} \tag{4.1}$$

where $\psi(x,y)$ is the electronic potential in the body and $V_{ch}(y)$ is the channel potential ($V_{ch}(0) = 0$ and $V_{ch}(L = V_{ds})$). The Gauss's Law at the front and the back silicon surfaces can be written as

$$
\begin{aligned}
V_{fg} - V_{fb1} - \psi_f(y) &= -\frac{\epsilon_{Si}}{C_{ox1}} \frac{\partial \psi(x,y)}{\partial x}\Big|_{x=-T_{Si}/2} \\
V_{bg} - V_{fb2} - \psi_b(y) &= +\frac{\epsilon_{Si}}{C_{ox2}} \frac{\partial \psi(x,y)}{\partial x}\Big|_{x=+T_{Si}/2}
\end{aligned}
\tag{4.2}
$$

where $\psi_f(y)$ and $\psi_b(y)$ are the front surface and back surface potentials respectively. The solution to Eq. 4.1 depends on the existence of the zero electric field plane, $\frac{\partial \psi(x,y)}{\partial x}\big|_{x=x_0} = 0$.

If the zero electric field plane (x_0) exists, i.e $-\infty < x_0 < \infty$, the potential in the body can be written as [59]

$$\psi(x,y) = V_{ch}(y) - \frac{2kT}{q} \ln\left[\frac{T_{Si}}{2\beta(y)} \sqrt{\frac{q^2 n_i}{2\epsilon_{Si}kT}} \sin\left(\frac{2\beta(y)x}{T_{Si}} + \alpha(y) \right) \right] \tag{4.3}$$

where both $\alpha(y)$ and $\beta(y)$ are a function of the applied bias. The electronic potential in Eq. 4.3 is valid for lightly doped IDG-FETs as well as for lightly doped symmetric DG-FETs. For a symmetric DG-FET, the electric field at the center plane ($x_0 = 0$) is zero. Using this fact, it can be concluded from Eq. 4.3 that $\alpha = \pi/2$ for a symmetric DG-FET. Interestingly, the expression of surface potential using Eq. 4.3 with $\alpha =$

$\pi/2$ matches with the expression of surface potential for a lightly doped symmetric DG-FET. The boundary conditions for determining $\alpha(y)$ and $\beta(y)$ can be obtained by substituting the definition of electronic potential Eq. 4.3 in Eq. 4.2.

$$
\begin{aligned}
V_{fg} - V_{fb1} - V_{ch}(y) &= \frac{kT}{q} \ln\left(\frac{2\epsilon_{Si}kT}{q^2 n_i T_{Si}^2}\right) + \frac{2kT}{q} ln\left(\frac{2\beta(y)}{\sin(\alpha(y) - \beta(y))}\right) \quad (4.4) \\
&\quad + r_1 \frac{4kT}{q} \beta(y) \cot(\alpha(y) - \beta(y)) \\
V_{bgs} - V_{fb2} - V_{ch}(y) &= \frac{kT}{q} \ln\left(\frac{2\epsilon_{Si}kT}{q^2 n_i T_{Si}^2}\right) + \frac{2kT}{q} ln\left(\frac{2\beta(y)}{\sin(\alpha(y) + \beta(y))}\right) \\
&\quad - r_2 \frac{4kT}{q} \beta(y) \cot\left((\alpha(y) + \beta(y))\right)
\end{aligned}
$$

where

$$
r_1 = \frac{4kT\epsilon_{Si}}{qC_{ox1}T_{Si}} \quad \text{and} \quad r_2 = \frac{4kT\epsilon_{Si}}{qC_{ox2}T_{Si}} \tag{4.5}
$$

The unknown parameters $\alpha(y)$ and $\beta(y)$ are calculated by solving the system of equations in Eq. 4.4.

If the IDG-FET is heavily asymmetric due to the difference in work-functions between the two gate electrodes and different bias applied to the two gates, the zero electric field plane may not exist anywhere. In this case the potential profile in the body is given by

$$
\psi(x,y) = V_{ch}(y) - \frac{2kT}{q} \ln\left[\frac{T_{Si}}{2\beta(y)} \sqrt{\frac{q^2 n_i}{2\epsilon_{Si}kT}} \sinh\left(\frac{2\beta(y)x}{T_{Si}} + \alpha(y)\right)\right] \tag{4.6}
$$

105
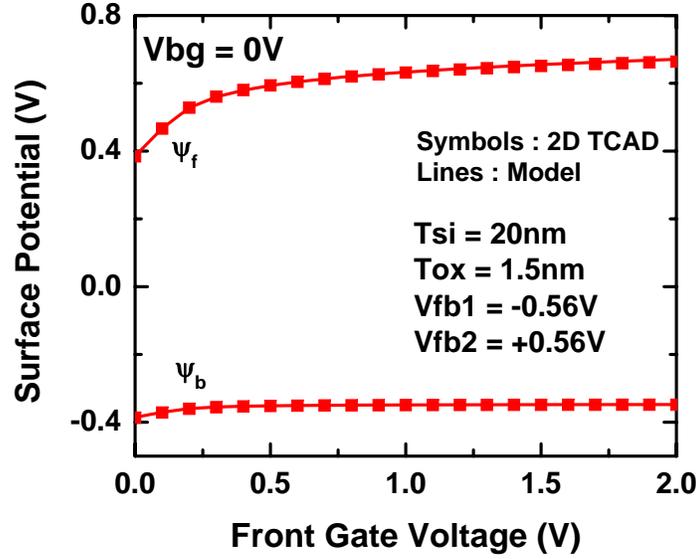
Figure 4.2: Front and back surface potentials for an IDG-FET.

and the boundary conditions for evaluating $\alpha(y)$ and $\beta(y)$ are

$$
\begin{aligned}
V_{fg} - V_{fb1} - V_{ch}(y) &= \frac{kT}{q} \ln \left(\frac{2\epsilon_{Si} kT}{q^2 n_i T_{Si}^2} \right) + \frac{2kT}{q} ln \left(\frac{2\beta(y)}{\sinh(\alpha(y) - \beta(y))} \right) \quad (4.7)\\
&\quad + r_1 \frac{4kT}{q} \beta(y) \coth(\alpha(y) - \beta(y))\\
V_{bg} - V_{fb2} - V_{ch}(y) &= \frac{kT}{q} \ln \left(\frac{2\epsilon_{Si} kT}{q^2 n_i T_{Si}^2} \right) + \frac{2kT}{q} ln \left(\frac{2\beta(y)}{\sinh(\alpha(y) + \beta(y))} \right)\\
&\quad - r_2 \frac{4kT}{q} \beta(y) \coth \left((\alpha(y) + \beta(y)) \right)
\end{aligned}
$$

Fig. 4.2 shows the potentials at the front surface (ψ_f) and the back surface (ψ_b) for an IDG-FET biased at $V_{ds} = 0$V. The IDG-FET has a n^+ front gate and p^+ back gate. The back gate is biased at $V_{bg} = 0$V and the front gate voltage is varied.

The biggest challenge in implementing the surface potential model for IDG-FET

in a compact model is solving the system of equations in Eq. 4.4 and Eq. 4.7 efficiently. Several clever approaches such as 2-D Newton-Rhapson method, Shooting method and analytical approximation technique were attempted in [61]. The analytical approximation technique stands out in all the approaches as it avoids the convergence issues which are faced by the other iterative approaches and it is also computationally efficient. The analytical approximation approach however assumes that the inversion charge density at the back surface of the IDG-FET is negligibly small. As a result, it is valid only for IDG-FETs where the back gate voltage is used primarily as a control for adjusting threshold voltage without forcing the back surface into inversion. Attempts need to be made to expand the analytical approximation technique to model even back surface inversion.

4.3   I-V Model

To have a computationally efficient analytical model for drain current, charge sheet approximation is used for the asymmetric IDG-FET. This is based on the observation that the use of charge sheet approximation for a symmetric common DG-FET introduces errors on the order of only a few percent while yielding a very simple expression for drain current. Charge sheet approximation has also been traditionally used in almost all the bulk planar MOSFET models [62]. First a charge sheet based I-V model for CDG-FET is derived and its accuracy is studied. Based on this model, a charge sheet based I-V model for IDG-FET is developed.

Due to the symmetry in CDG-FET, the current flowing in the upper half of the

silicon body is equal to the current flowing in the lower half of the body. Within the charge sheet approximation (CSA), all the inversion charge in the body is assumed to flow in an infinitesimally thin layer at the surface [29]. Using the CSA, the drain current in CDG-FET comprising of both drift and diffusion components can be expressed as [63]

$$I_d(y) \;=\; 2 \cdot \mu \cdot W \cdot \left(Q_i(y) \frac{d\psi_s(y)}{dy} - \frac{kT}{q} \frac{dQ_i(y)}{dy} \right) \tag{4.8}$$

where $Q_i(y)$ is the inversion charge in the upper half of silicon body for CDG-FET and $\psi_s(y)$ is the surface potential. For lightly doped CDG-FET, the inversion charge $Q_i(y)$ can be expressed as

$$Q_i(y) = C_{ox} \left(V_{gs} - V_{fb} - \psi_s(y) \right) \tag{4.9}$$

Substituting $Q_i(y)$ in Eq. 4.8 and integrating from source to drain, the drain current for CDG-FET with CSA is obtained.

$$I_d \;=\; 2 \cdot \frac{\mu W}{L} \cdot \left(\frac{Q_s + Q_d}{2} + C_{ox} \frac{kT}{q} \right) \cdot (\phi_s - \phi_d) \tag{4.10}$$

where Q_s and Q_d are the inversion charges and ϕ_s and ϕ_d are the surface potentials at the source and drain end respectively. The model is mathematically very simple and has good accuracy when compared with 2-D TCAD simulations as show in Fig. 4.3 in both strong inversion and sub-threshold regimes.

To model the drain current for IDG-FETs, follow the same framework as the CSA
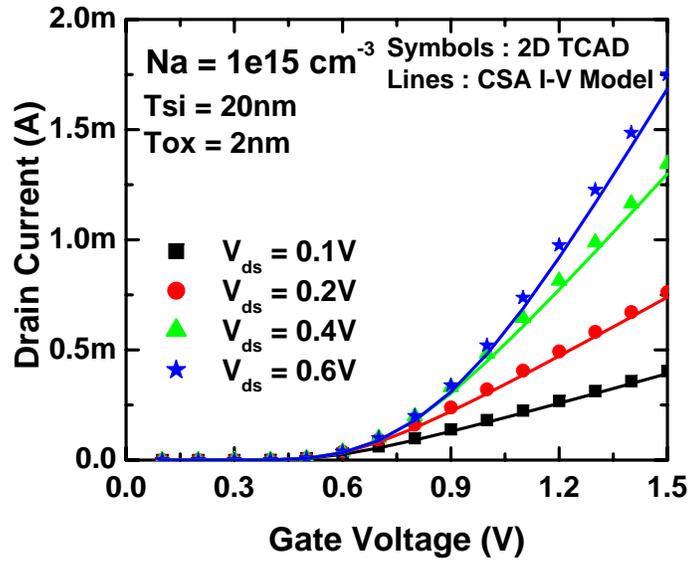
(a)
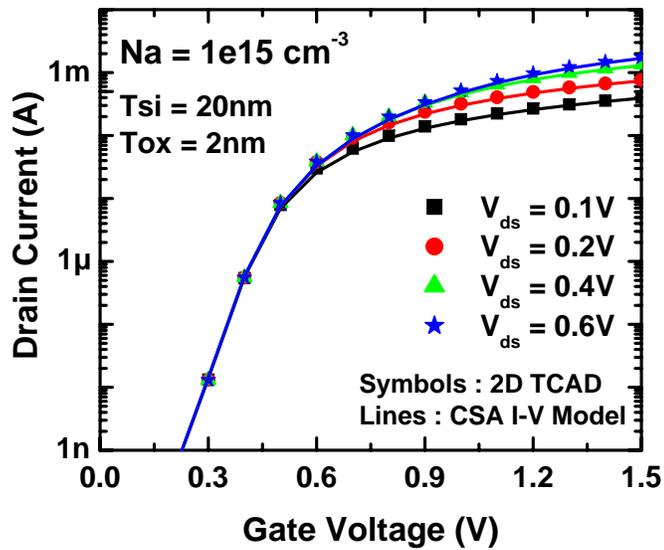
(b)

Figure 4.3: Charge sheet approximation based drain current model for symmetric common DG-FET. Model is close to 2-D TCAD simulations.
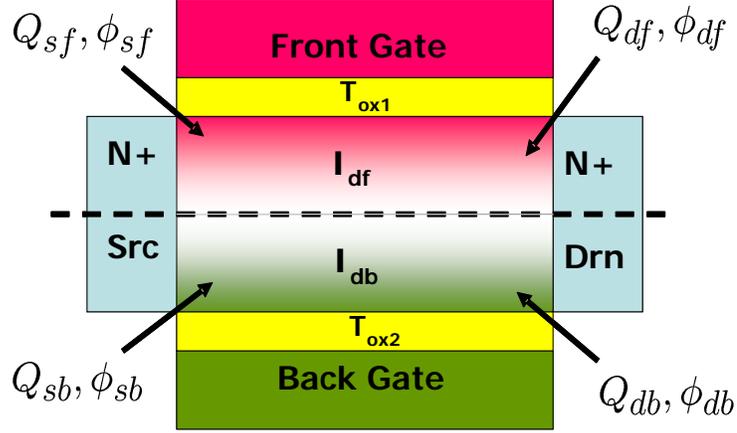
Figure 4.4: Schematic of the asymmetric IDG-FET showing the definitions of charges and surface potentials at the source and drain end at front and back surface of the silicon body.

based I-V model for CDG-FET in Eq. 4.10 since it is simple and accurate. Fig. 4.4 shows the schematic of the IDG-FET under study. The definitions of the symbols used in this section are indicated in Fig. 4.4. The total drain current in IDG-FET can be decomposed into two different channel currents : I_{df} and I_{db}. I_{df} is the current flowing in the upper half of the body while I_{db} is the current flowing in the lower half of the body as show in Fig. 4.4.

Each of the channel currents can be formulated using CSA similar to Eq. 4.10. The front channel current, I_{df}, can be expressed as

$$I_{df} \quad = \quad \frac{\mu_1 \cdot W}{L} \cdot \left(\frac{Q_{sf} + Q_{df}}{2} + C_{ox1} \frac{kT}{q} \right) \cdot (\phi_{sf} - \phi_{df}) \qquad (4.11)$$

It can be interpreted as the charge in the upper half of the body being driven by front surface electric field $(\phi_{sf} - \phi_{df})/L$ and front surface mobility μ_1. The inversion charge density at any point in the silicon body is given by

$$Q_i(x, y) = qn_i e^{\frac{q(\psi(x,y) - V_{ch}(y))}{kT}} \tag{4.12}$$

The charges in the upper half of the body at the source and drain end can be easily computed by integrating Eq. 4.12 from front surface to the center of the body. When the zero electric field plane exists, $\psi(x, y)$ is defined by Eq. 4.3 and the charges are given by

$$Q_{sf} = \frac{4\beta_s \epsilon_{Si} kT}{qT_{Si}} \left(\cot \left(\alpha_s - \beta_s \right) - \cot \left(\alpha_s \right) \right) \tag{4.13}$$
$$Q_{df} = \frac{4\beta_d \epsilon_{Si} kT}{qT_{Si}} \left(\cot \left(\alpha_d - \beta_d \right) - \cot \left(\alpha_d \right) \right)$$

where $\alpha_s = \alpha(y = 0)$, $\alpha_d = \alpha(y = L)$, $\beta_s = \beta(y = 0)$ and $\beta_d = \beta(y = L)$ When the zero electric field plane does not exist, $\psi(x, y)$ is defined by Eq. 4.6 and the charges are given by

$$Q_{sf} = \frac{4\beta_s \epsilon_{Si} kT}{qT_{Si}} \left(\coth \left(\alpha_s - \beta_s \right) - \coth \left(\alpha_s \right) \right) \tag{4.14}$$
$$Q_{df} = \frac{4\beta_d \epsilon_{Si} kT}{qT_{Si}} \left(\coth \left(\alpha_d - \beta_d \right) - \coth \left(\alpha_d \right) \right)$$

On similar lines as front channel current, the back channel current I_{db} can be expressed

111

as

$$I_{db} = \frac{\mu_2 \cdot W}{L} \cdot \left(\frac{Q_{sb} + Q_{db}}{2} + C_{ox2} \frac{kT}{q} \right) \cdot (\phi_{sb} - \phi_{db}) \tag{4.15}$$

I_{db} can be interpreted as the charge in the lower half of the body being driven by back surface electric field $(\phi_{sb} - \phi_{db})/L$ and back surface mobility μ_2. The charges in the lower half of the body at the source and drain end are given by

$$Q_{sb} = \frac{4\beta_s \epsilon_{Si} kT}{qT_{Si}} \left(\cot(\alpha_s) - \cot(\alpha_s + \beta_s) \right) \tag{4.16}$$
$$Q_{db} = \frac{4\beta_d \epsilon_{Si} kT}{qT_{Si}} \left(\cot(\alpha_d) - \cot(\alpha_d + \beta_d) \right)$$

where zero electric field plane exists and

$$Q_{sb} = \frac{4\beta_s \epsilon_{Si} kT}{qT_{Si}} \left(\coth(\alpha_s) - \coth(\alpha_s + \beta_s) \right) \tag{4.17}$$
$$Q_{db} = \frac{4\beta_d \epsilon_{Si} kT}{qT_{Si}} \left(\coth(\alpha_d) - \coth(\alpha_d + \beta_d) \right)$$

when the zero electric field does not exist.

The net drain current for IDG-FET is simply the sum of the two channel currents.

$$I_d = I_{df} + I_{db} \tag{4.18}$$

There is no rigorous proof for this analytical drain current model but it can be justified looking at the two asymptotic cases. In case the IDG-FET is symmetric, then Eq. 4.18 is identical to Eq. 4.10. When the device is highly asymmetric, all the channel

current is expected to flow at one surface only and again the model is accurate. For all the intermediate cases, verification against 2-D TCAD simulations will be taken as a measure of model accuracy. Fig. 4.5 shows the model predicted drain current for an IDG-FET with asymmetric front and back oxide and asymmetric work functions. The comparison with 2-D TCAD simulations shows that the model yields accurate drain current in sub-threshold, linear and saturation regimes of operation.

One of the main advantages of independent gate operation in IDG-FETs is the ability to tune the threshold voltage of the transistor by varying back gate voltage. Any I-V model for IDG-FETs should be able to capture this behavior very well. In order to examine the ability of the model to predict threshold voltage (V_{th}) tuning, two different IDG-FETs were studied. Device A had a thick back oxide $T_{ox2} = 50$nm while the Device B had a thinner back oxide $T_{ox2} = 20$nm. The back gate bias was changed from -1V to +1V and the resulting drain currents as predicted by model and by 2-D TCAD simulations are shown in Fig. 4.6. As expected, the I-V model shows a greater V_{th} tuning range for same V_{bg} range in Device B which has thinner T_{ox2}. The model predicted values of drain current agree very well with 2-D TCAD simulations. To further verify the scalability of the I-V model, the body thickness of the IDG-FET was varied from $T_{Si} = 5$nm to $T_{Si} = 20$nm. Fig. 4.7 shows that the I-V model predicts correct scaling of drain current as a function of T_{Si}.

The I-V model discussed thus far can handle inversion at both surfaces. However, if the operation of IDG-FET is restricted to single conduction channel (ex. only front channel inversion), then the corresponding I-V model can be simplified further. For only front channel inversion, the charge in the upper half of silicon body is much
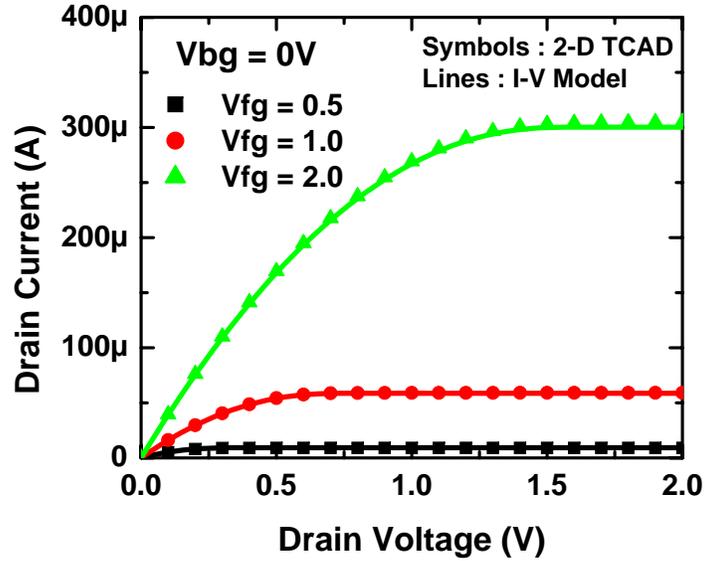
(a)
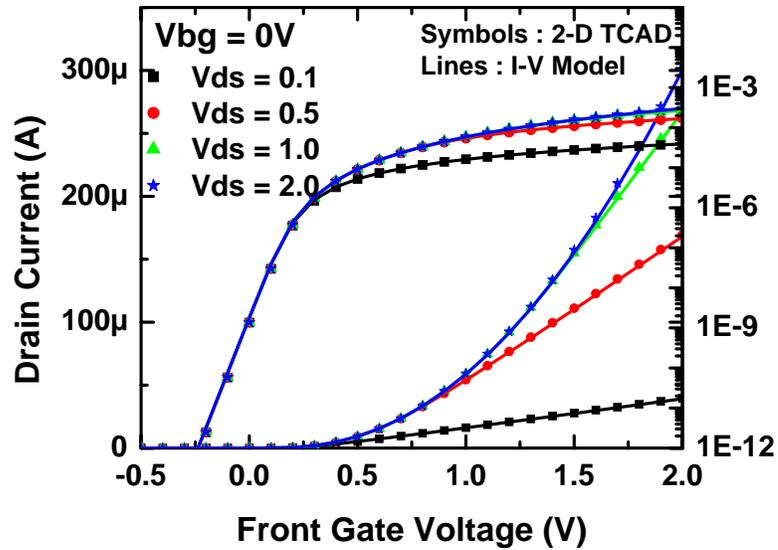
(b)

Figure 4.5: Drain current characteristics of an asymmetric IDG-FET (a) I_d-V_{fg} (b) I_d-V_{bg}. Device data : $T_{ox1} = 2$nm, $T_{ox2} = 4$nm, $T_{Si} = 20$nm, $N_A = 1e16 \mathrm{cm}^{-3}$, n^+ front gate and p^+ back gate. Model agrees very well with TCAD in all regimes of operation.
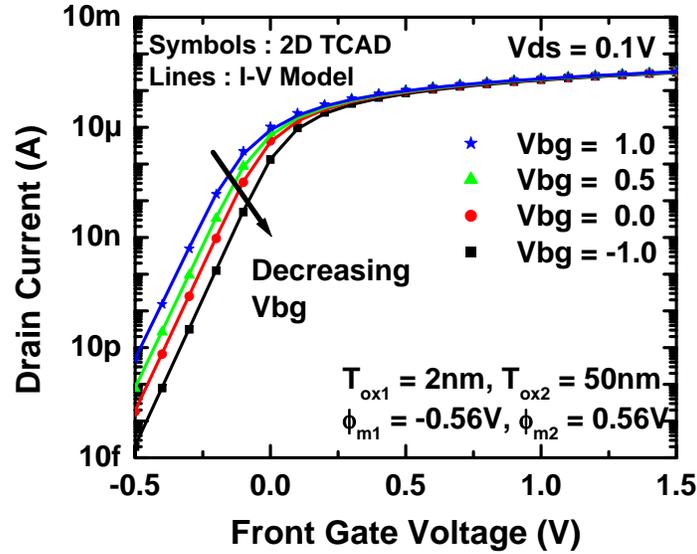
(a)
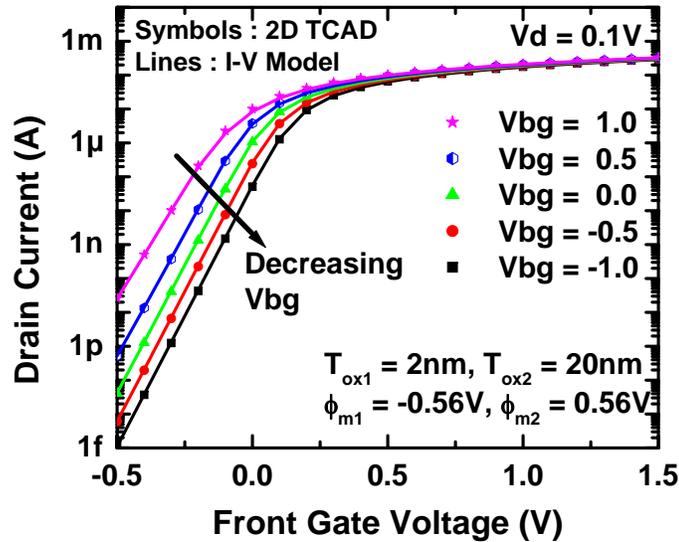
(b)

Figure 4.6: Drain current characteristics of two different asymmetric IDG-FETs illustrating the threshold voltage tuning by changing the back gate bias for devices with (a) Thick $T_{ox2} = 50$nm (Device A) and (b) Thin $T_{ox2} = 20$nm (Device B). Model correctly predicts larger V_{th} shift for Device B with the thinner T_{ox2}.
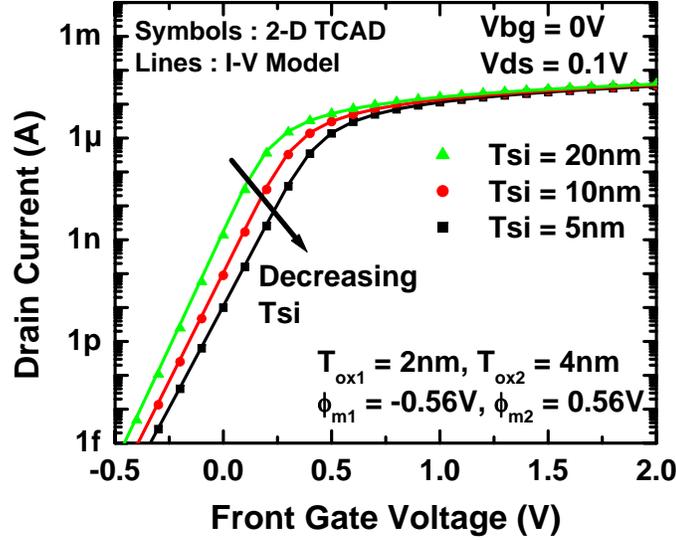
Figure 4.7: Drain current for IDG-FET with different body thickness. The good agreement between the model and 2-D TCAD simulations demonstrates the scalability of the I-V model.

larger than charge in the lower half.

$$Q_{sf} \gg Q_{sb} \quad \text{and} \quad Q_{df} \gg Q_{db} \tag{4.19}$$

The electric field at the front surface is also larger than the electric field at back surface.

$$\phi_{sf} - \phi_{df} \gg \phi_{sb} - \phi_{db} \tag{4.20}$$

As a result, one can evaluate drain current in this case by allowing the front surface electric field to drive the entire charge in the body from source to drain. In other

words, the drain current for only front surface conduction can be formulated as

$$I_d \;=\; \frac{\mu \cdot W}{L} \cdot \left(\frac{Q_s + Q_d}{2} + C_{ox1} \frac{kT}{q} \right) \cdot (\phi_{sf} - \phi_{df}) \qquad (4.21)$$

where Q_s and Q_d are the total integrated charge in the body.

$$Q_s = C_{ox1} \left(V_{fg} - V_{fb1} - \phi_{sf} \right) + C_{ox2} \left(V_{bg} - V_{fb2} - \phi_{sb} \right) \qquad (4.22)$$

$$Q_d = C_{ox2} \left(V_{fg} - V_{fb2} - \phi_{df} \right) + C_{ox2} \left(V_{bg} - V_{fb2} - \phi_{db} \right)$$

This drain current model with single channel conduction assumption Eq. 4.21 is computationally faster than the drain current model for both channel conduction in Eq. 4.18. Its simplicity also makes it useful for deriving the charge model for IDG-FET under single channel conduction assumption as will be described in the following section.

4.4   C-V Model

The C-V model is indispensable for any compact model and it complements the I-V model in making the compact model suitable for all kinds of circuits analysis including AC and transient analysis. The C-V model starts with calculation of the terminal charges.

For the lightly doped IDG-FET, the total integrated charge on the front and back

gate terminals can be written as

$$
\begin{aligned}
Q_{fg} &= \int_0^L C_{ox1}\left(V_{fg} - V_{fb1} - \psi_f(y)\right) dy \\
Q_{bg} &= \int_0^L C_{ox2}\left(V_{bg} - V_{fb2} - \psi_b(y)\right) dy
\end{aligned}
\tag{4.23}
$$

In accordance with the Ward-Dutton charge partition [40], the charge on the drain terminal can be written as

$$
Q_d = -\int_0^L \left(C_{ox1}\left(V_{fg} - V_{fb1} - \psi_f(y)\right) + C_{ox2}\left(V_{bg} - V_{fb2} - \psi_b(y)\right)\right) \frac{y}{L} dy \tag{4.24}
$$

The charge on the source terminal is found through simple charge conservation

$$
Q_s = -(Q_{fg} + Q_{bg} + Q_d) \tag{4.25}
$$

To obtain the terminal charges, both $\psi_f(y)$ and $\psi_b(y)$ should be determined. Current continuity gives only one equation and there are two variables to be solved. To circumvent this obstacle, single surface conduction is assumed. Analytic expressions for the terminal charges and capacitances are obtained where there is strong inversion only at one surface of the IDG-FET.

For the purpose of illustrating the model formulation, front channel inversion is assumed. The assumption of single channel inversion allows to express the charge at

the back gate in terms of front surface potential through a capacitive divider.

$$Q_{bg}(y) = C_{ox2} \underbrace{\frac{C_{Si}}{C_{ox2} + C_{Si}}}_{F} (V_{gb} - V_{fb2} - \psi_f(y)) \tag{4.26}$$

Using Eq. 4.26, the drain current for IDG-FET (Eq. 4.21) under single channel conduction can be re-written as

$$I_d = \mu \cdot C_{ox1} \cdot \frac{W}{L} \cdot (A - B(\phi_{df} + \phi_{sf})) \cdot (\phi_{sf} - \phi_{df}) \tag{4.27}$$

where

$$B = \frac{1 + F \cdot \frac{C_{ox2}}{C_{ox1}}}{2} \tag{4.28}$$

$$A = V_{fg} - V_{fb1} - F \cdot \frac{C_{ox2}}{C_{ox1}} (V_{bg} - V_{fb2}) + V_t \left(1 + \frac{C_{ox2}}{C_{ox1}}\right)$$

When current continuity is applied with I_d as defined in Eq. 4.27, there is only one variable $\psi_f(y)$ which can be determined deterministically. Applying current continuity under quasi-static operation of the transistor, $\psi_f(y)$ can be related to ϕ_{sf} and ϕ_{df} through

$$y \cdot (A - B(\phi_{df} + \phi_{sf})) (\phi_{df} - \phi_{sf}) = L \cdot (A - B(\psi_f(y) + \phi_{sf})) (\psi_f(y) - \phi_{sf}) \tag{4.29}$$

The front gate terminal charge can now be obtained by substituting the expression

for $\psi_f(y)$ in Eq. 4.23

$$\frac{Q_{fg}}{C_{ox1}\,W\,L} = V_{fg} - V_{fb1} - \frac{\phi_{sf} + \phi_{df}}{2} + \frac{B\left(\phi_{df} - \phi_{sf}\right)^2}{6\left(A - B\left(\phi_{df} + \phi_{sf}\right)\right)} \tag{4.30}$$

The charge on the drain terminal Q_d consists of contribution from the front gate terminal Q_{df} and the back gate terminal Q_{db} as seen from Eq. 4.24. The drain charge component arising from the front gate terminal is

$$\begin{aligned}
\frac{Q_{df}}{C_{ox}\,L} = \frac{V_{fg} - V_{fb1}}{2} \;&-\; \frac{\phi_{sf} + \phi_{df}}{4} + \frac{B\left(\phi_{df} - \phi_{sf}\right)^2}{60\left(A - B\left(\phi_{df} + \phi_{sf}\right)\right)} \\
&-\; \frac{\left(5A - 4B\phi_{df} - 6B\phi_{sf}\right)\left(A - 2B\phi_{df}\right)\left(\phi_{df} - \phi_{sf}\right)}{60\left(A - B\left(\phi_{df} + \phi_{sf}\right)\right)^2}
\end{aligned} \tag{4.31}$$

Similarly, the charge on the back gate terminal Q_{bg} and its contribution to drain charge Q_{db} can be obtained by replacing the replacing all the front surface quantities with the back surface quantities and vice versa.

The expressions for the terminal charges are continuous and valid over all regimes of transistor operation. Terminal capacitances are obtained by differentiating the terminal charges with respect to the terminal voltages. The model is verified against 2-D TCAD simulations without any fitting parameters. Fig. 4.8 shows the variation of capacitances associated with the front gate terminal and drain terminal as a function of drain voltage for an IDG-FET with asymmetric front and back gates and $V_{bg} = 0\text{V}$. The desired symmetry property of the C-V model is evident in Fig. 4.8(a) which shows that $C_{fgs} = C_{fgd}$ at $V_{ds} = 0\text{V}$. Fig. 4.9 plots the capacitances associated with the front gate and back gate terminal as a function of front gate voltage for

the same IDG-FET. There is small deviation in the capacitances associated with the back gate terminal (C_{bgx} where x represents any of the four transistor terminals). However, since the magnitude of the capacitances associated with the back gate is small, the discrepancy will not effect the accuracy of circuit simulation. If the back oxide thickness is larger than the front oxide thickness, then the magnitude of C_{gbx} decreases even further and the error is even less of a concern. This is illustrated in Fig. 4.10. The capacitances of an IDG-FET with $N+$ front gate, $P+$ back gate, $T_{ox1} = 1.2$nm, $T_{ox2} = 40$nm and $T_{Si} = 15$nm are calculated using the C-V model and compared against 2-D TCAD simulations. It can be seen from Fig. 4.10 (a) that the magnitude of capacitance associated with the back gate terminal is negligible compared to the other capacitances and hence any small error associated with it will have negligible effect on the circuit simulation. Agreement between C-V model and 2-D TCAD simulations is again achieved without the use of any fitting parameters highlighting the accuracy and predictive nature of the C-V model.

4.5   BSIM-IMG Model

The I-V model and C-V model presented in Section 4.3 and Section 4.4 form the core of BSIM-IMG model - the compact model for IDG-FETs. The core model is highly predictive and has a high degree of accuracy as demonstrated in the earlier sections by comparing against 2-D TCAD simulations without using any fitting parameters. Similar to BSIM-CMG, modeling of several real device effects that will determine the electrical behavior of advanced independent MG-FET technologies are added to the
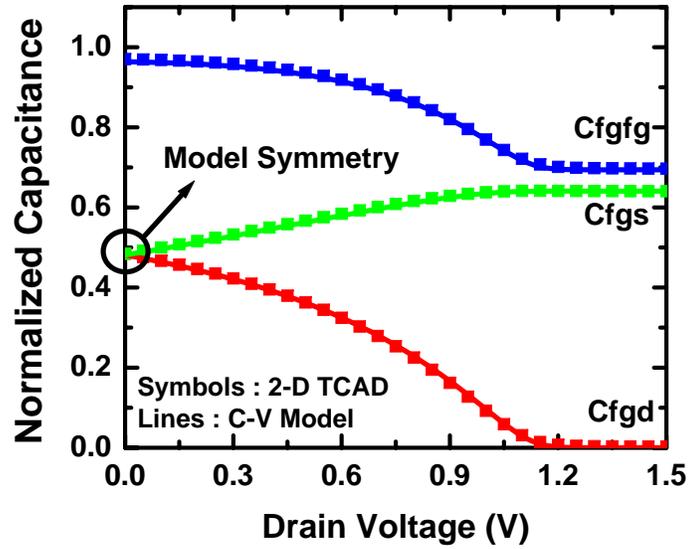
(a)
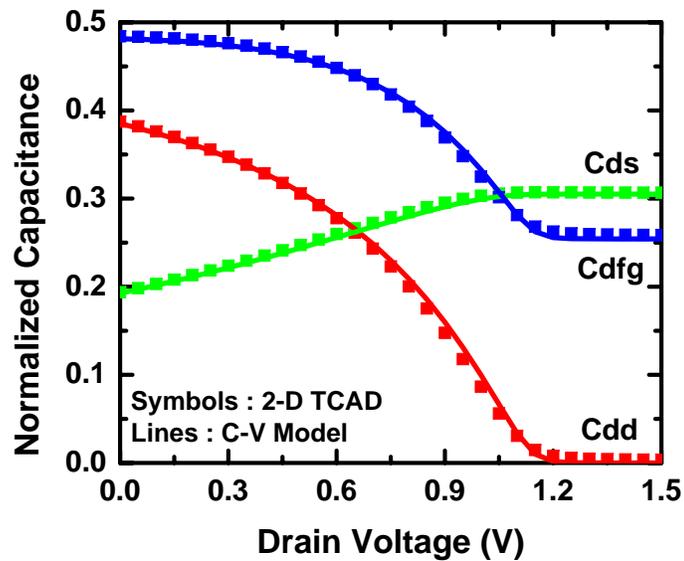
(b)

Figure 4.8: Model predicted capacitances associated with (a) front gate terminal and (b) drain terminal for a lightly doped IDG-FET with $N+$ front gate, $P+$ back gate, $T_{Si} = 20$nm, $T_{ox1} = T_{ox2} = 1.5$nm, V_{fg}=1.5V and $V_{bg} = 0$V. Model agrees very will 2-D TCAD simulation without any fitting parameters.
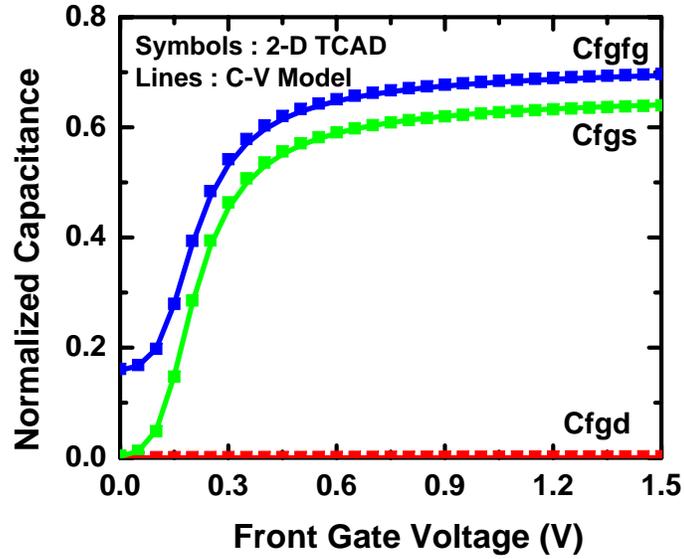
(a)
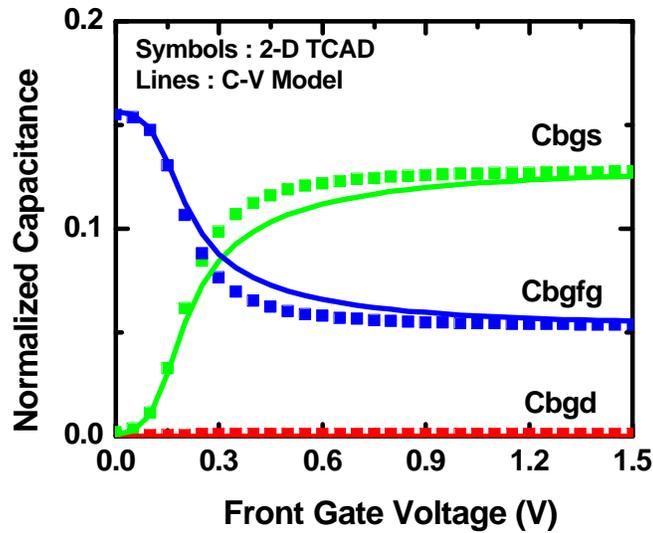
(b)

Figure 4.9: Model predicted capacitances associated with (a) Front gate terminal and (b) Back gate terminal for a lightly IDG-FET with $N+$ front gate, $P+$ back gate, T_{Si} = 20nm and $T_{ox1} = T_{ox2}$ = 1.5nm biased at V_d=1.5V and $V_{bg} = 0$V. Small difference observed between the capacitances associated with the back gate terminal.
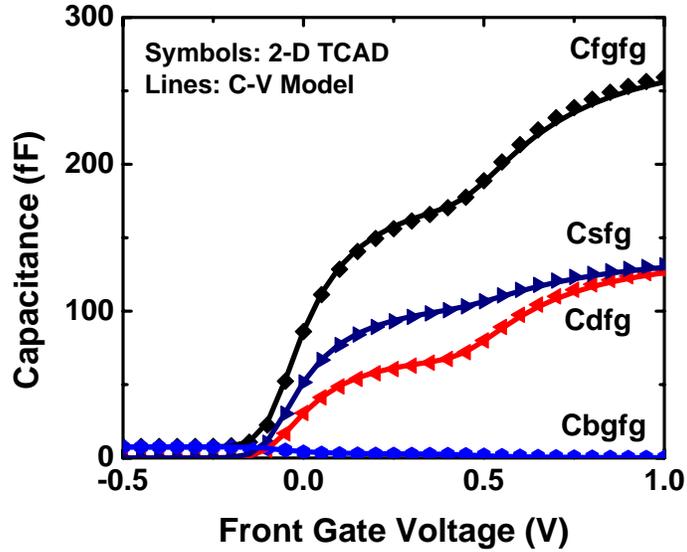
(a)
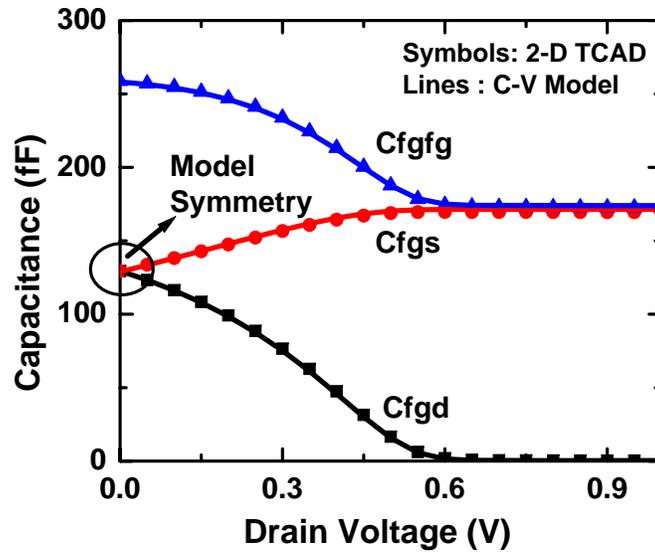
(b)

Figure 4.10: Model predicted capacitances as a function of (a) Front gate voltage at V_{ds} = 0.5V and (b) Drain voltage at V_{fg}=0.5V for a lightly IDG-FET with $N+$ front gate, $P+$ back gate, T_{Si} = 15nm, T_{ox1} = 1.2nm and T_{ox2} = 40nm.

1) Quantum Mechanical Effects
2) Short Channel Effects
a) Vth roll-off
b) DIBL
c) Sub-threshold slope
d) Channel length modulation
3) Series resistance
4) Mobility degradation
5) Velocity Saturation
6) Velocity Overshoot/Source-End Velocity Limit
7) Gate Induced Drain (Source) Leakage (GIDL, GISL)
8) Impact Ionization
9) S/D Junction leakage
10) Gate tunneling
11) Parasitic capacitance

Table 4.1: List of physical effects modeled in BSIM-IMG

core model of BSIM-IMG. Table 4.1 enlists the real device effects modeled in BSIM-IMG. The model has been written in Verilog-A and implemented in popular circuit simulators such as SPECTRE and HSPICE.

4.6 Summary

A surface potential based core model for modeling independent MG-FETs is developed in this chapter. An analytical I-V model for IDG-FETs is derived using charge sheet approximation to model the drain current under any operation regime. Surface potential calculation when there is strong inversion at both front and back surface of IDG-FET is a significant challenge. As a result, a simplified and computationally faster I-V model is also presented for the case when there is significant current conduction at only one surface of IDG-FET. An analytical C-V model is also presented

for IDG-FETs with single surface conduction. The I-V model and C-V model show excellent agreement with 2-D TCAD simulations without the use of any fitting parameters demonstrating the inherent scalability and predictive nature of the core model of BSIM-IMG.

Chapter 5

Layout Dependent Mobility Model for Process Induced-Stress

5.1 Introduction

Continued scaling of the conventional bulk planar MOSFET is becoming more and more challenging in the sub-100nm regime. New materials innovations in the state-of-the-art planar bulk silicon transistors that can boost the transistor performance hold the key to scaling. One such innovation that provides a dramatic performance boost with little added process complexity is process-induced strain. Strain can be applied to a transistor such that it enhances the carrier mobility and increases the current drive making the transistor faster. Since the introduction of strain in mainstream CMOS technology for the first time in 90nm technology node, it has become a mainstay in all the future CMOS generations [14].
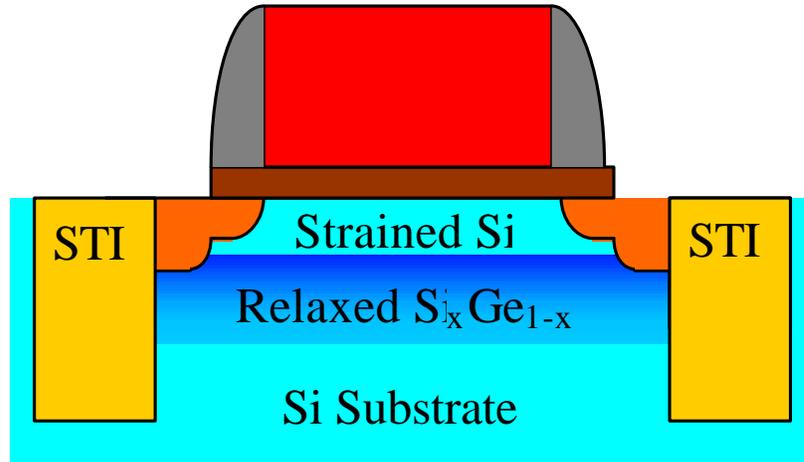
Figure 5.1: Schematic of strained Si MOSFET which uses a relaxed Si_xGe_{1-x} layer to generate biaxial strain in the inversion channel.

Strain can be introduced into the inversion channel of a transistor primarily through two different ways. One approach to strain the inversion channel is through the use of global strain or wafer-level strain where strain is created through the entire wafer. Strained silicon wafers can be created by a epitaxial deposition of thin film silicon on a substrate with different lattice constant such as relaxed Si_xGe_{1-x} (Fig. 5.1) to generate biaxial stain [64]. Due to the high cost of the wafers with Si_xGe_{1-x} films and associated integration challenges, this technique is not yet popular in production. Incidentally, most of the early theoretical and experimental studies on strained silicon were based on global strain. The second approach to strain the inversion channel is through the application of local strain. Stressed films can be deposited on the transistor and stressed structures can be created adjacent to the transistor to generate strain locally. Since this generally involves only a slight alteration in the conventional MOSFET process flow, local strain is commonly referred to as process-induced strain
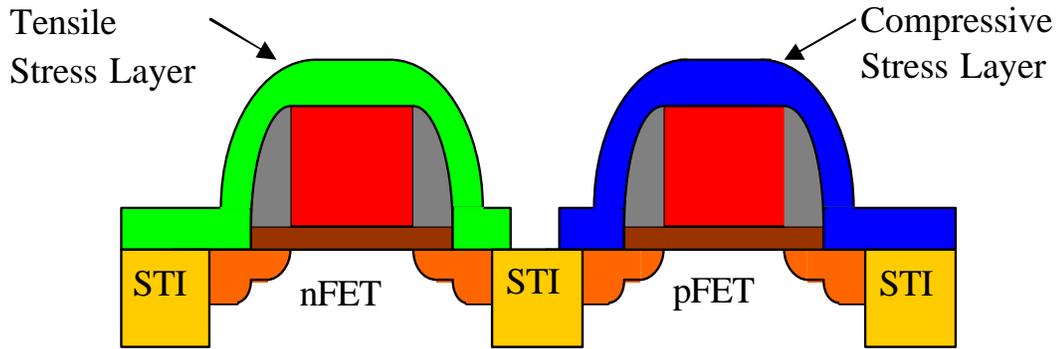
Figure 5.2: Generation of local strain by using the Dual Etch Stop Layer approach with tensile and compressive nitride layers over nFET and pFET respectively.

and has been employed in mainstream production by almost all the companies. Generally, the wafer-level strain creates a biaxial stress while the process-induced strain is optimized for a strong uniaxial stress.

The most common form of introducing process induced strain into CMOS is through depositing stressed nitride layers on top of transistors [65]. A tensile nitride capping layer is deposited over nFETs and a compressive nitride capping layer is deposited over pFETs to generate strong longitudinal uniaxial tensile and uniaxial compressive stress respectively as shown in Fig. 5.2. Since these stress layers also serve as contact etch stops, this technique is also known as Dual Etch Stop Layer process (dESL). In a different approach to create uniaxial strain, local epitaxial films are grown in the source/drain regions of the transistor [66]. SiGe and SiC can be epitaxially grown in S/D regions for pFETs and nFETs to boost hole and electron mobilities respectively. Another approach to create local strain is the stress memorization technique [67]. A temporary capping layer with certain internal stress is deposited on an amorphized gate electrode which when recrystallized memorizes
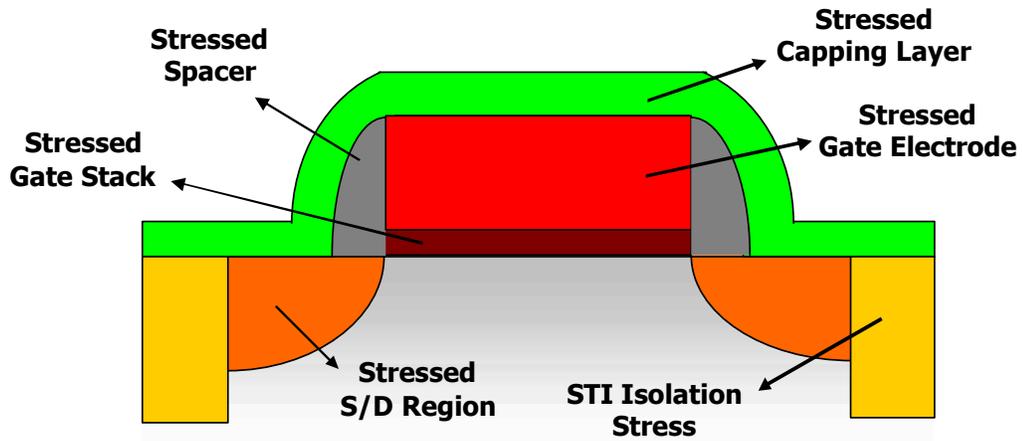
Figure 5.3: Different ways in which stress can be introduced into the inversion channel in the bulk MOSFET.

some of the stress from the temporary capping layer and transfers it to the silicon channel. The temporary capping layer may then be removed and followed by a dESL process to induce more local strain. Different stress inducing techniques can be simultaneously incorporated into a transistor to generate additive stresses resulting in large gains in transistor performance as shown in Fig. 5.3.

The transistor performance gain due to process-induced strain is a strong function of the transistor layout [68]. The improvement in carrier mobility depends on transistor layout variables such as channel length, channel width and distance to its neighbors. To evaluate the circuits designed using strained silicon technology, modeling of layout dependent carrier mobility is absolutely essential. It is not practical to model the layout dependent mobility change for each of the stress inducing technology individually. An ideal model would be one that is non-process specific and which can capture the effect of any given number of stressors in a transistor. With
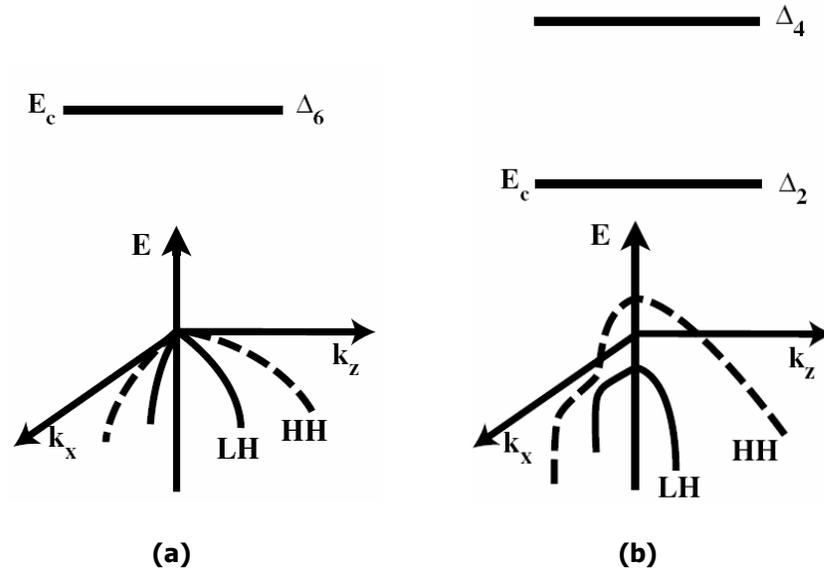
Figure 5.4: Band structure of (a) relaxed silicon and (b) silicon under uniaxial tensile strain [69]. Lowering of conductivity effective mass and scattering rates improves electron mobility.

this goal in mind, this chapter presents a new holistic, layout dependent model for mobility enhancement through process-induced stress.

5.2 Stress Modeling : Theory and Methodology

When strain is applied to silicon, it leads to a change in the band-structure. The valence band and the conduction band behave differently under tensile and compressive strain.

Under the influence of uniaxial tensile strain, the 6-fold degeneracy in the conduction band between the four in-plane valleys (Δ_4) and the two out-of-plane valleys (Δ_2) is lost as shown in the Fig 5.4. The electrons now preferentially occupy the

lower energy Δ_2 valleys which have lower in-plane conductivity mass (m^*). Furthermore, the splitting of the energy levels also reduces the amount of intervalley phonon scattering between Δ_4 and Δ_2 valleys. The degeneracy in the valence band between the heavy hole band (HH) and light hole band (LH) is also lost and the holes now preferentially occupy the HH band. Note that the carrier mobility is given by

$$\mu = \frac{q\tau}{m^*} \tag{5.1}$$

where τ is the average time between carrier scattering and m^* is the conductivity effective mass of the carriers. As a result, it can be seen that uniaxial tensile stress improves the electron mobility due to improvements in both conductivity mass and scattering rates but it degrades the hole mobility due to increase in hole conductivity mass [70].

Fig. 5.5 illustrates the impact of uniaxial compressive strain on silicon band structure. The 6-fold degeneracy in conduction band is lost and Δ_4 valley, which has higher conductivity mass, is lowered in energy degrading the electron mobility. The valence band experiences band splitting and band warping and the holes preferentially occupy the top band which has a lower in-plane conductivity effective mass. Unlike the conduction band splitting, the valence band splitting is small and so the intervalley scattering among holes does not decrease significantly. However, the band warping is significant enough to result in a large hole mobility enhancement. The large increase in hole mobility due to uniaxial compressive stress is another reason for using process-induced strain instead of whole-wafer strain which generates biaxial strain [71].
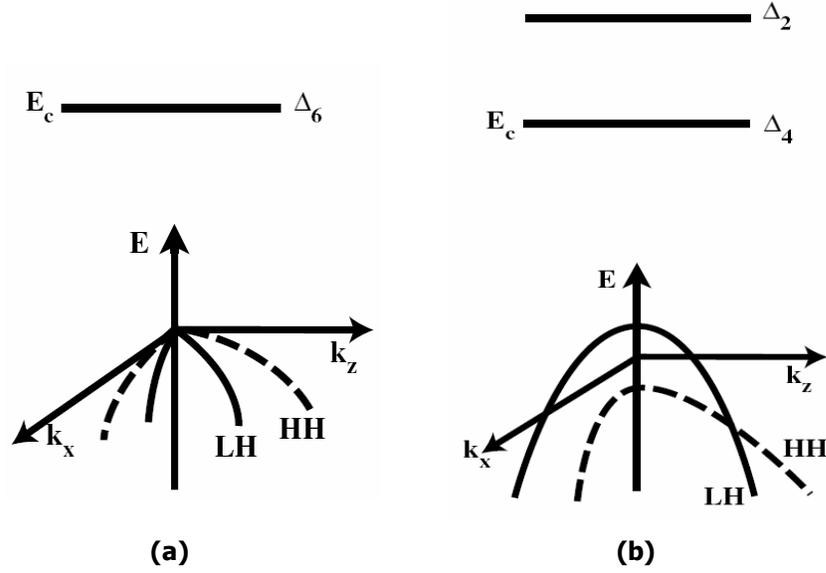
Figure 5.5: Band structure of (a) relaxed silicon and (b) silicon under uniaxial compressive strain [69]. Lowering of hole conductivity effective mass increases hole mobility.

The change in carrier mobilities due to strain is captured by piezoresistive coefficients. The piezoresistive coefficients (π) linearly relate the change in local resistivity (ρ) to the local stress (σ).

$$\frac{\Delta\rho}{\rho} = \pi\sigma \tag{5.2}$$

When a transistor is stressed by a process-induced strain techniques, the stress in the inversion channel is invariably non-uniform between source and drain. Using Eq. 5.2, the change in channel mobility (μ) taking into account the non-uniformity of stress can be written as

$$\frac{\Delta\mu}{\mu} = -\pi\sigma_{AVG} \tag{5.3}$$
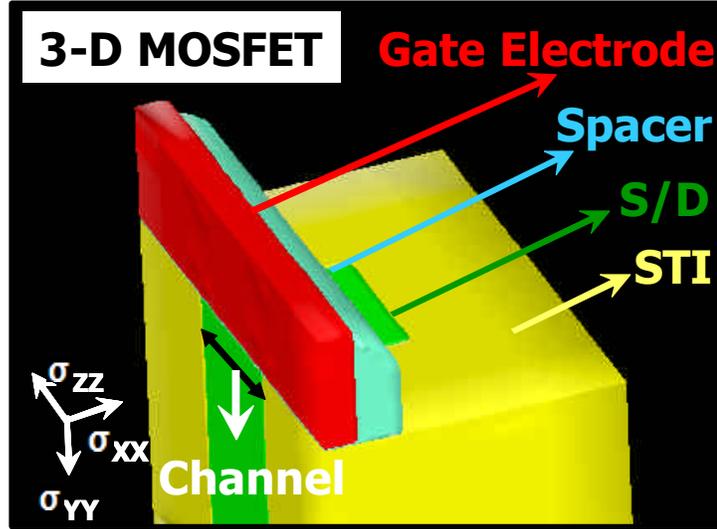
Figure 5.6: 3-D cross section of a bulk transistor showing the conventions for stress vectors.

where σ_{AVG} represents the average stress in the inversion channel. Eq. 5.3 shows that the mobility enhancement is proportional to the average stress in the channel. Note that stress is a vectorial quantity. Fig. 5.6 shows the 3-D cross-section of bulk transistor and the three stress vectors. σ_{XX} is the average in-plane longitudinal channel stress, σ_{ZZ} is the average in-plane transverse channel stress and σ_{YY} is the average out-of-plane channel stress. Each stress vector contributes to net mobility change differently and is dictated by the piezoresistive coefficient along the direction of the stress vector. The net carrier mobility change incorporating the effect of all three stress vectors can be written as

$$\frac{\Delta \mu}{\mu_0} = -(\pi_{XX} \cdot \sigma_{XX} + \pi_{YY} \cdot \sigma_{YY} + \pi_{ZZ} \cdot \sigma_{ZZ}) \qquad (5.4)$$

where π_{XX}, π_{YY} and π_{ZZ} are the piezoresistive coefficients along XX, YY and ZZ
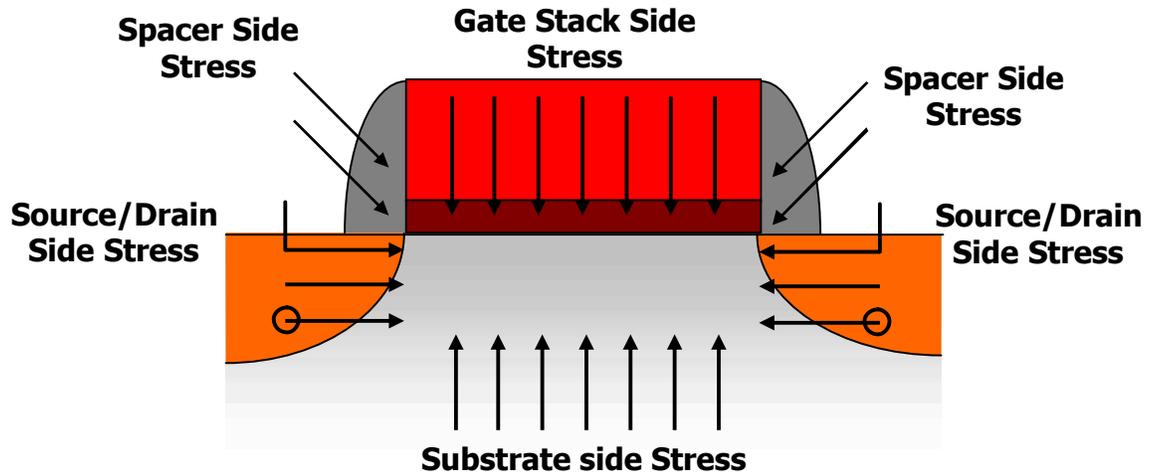
Figure 5.7: Possible directions in which stress gets transferred into the inversion channel.

directions respectively. The average channel stress vectors are a function of transistor layout variables such as channel length (L) and channel width (W). Looking at Eq. 5.4, modeling of the layout dependent mobility change for any stress-inducing process boils down to evaluating the layout dependency of the three average channel stress vectors.

Irrespective of the type and number of stressors present around the transistor, the stress from transistor's surroundings reaches the inversion channel through four distinct directions as shown in Fig. 5.7. The stress can reach the inversion channel through the gate dielectric (gate-stack side stress), through the spacer (spacer side stress), through source/drain or transverse to source/drain parallel to Si/SiO_2 plane (source-drain side stress) or through the substrate (substrate-side stress). When stress reaches the inversion channel from any of these four directions, it would generate all

the three stress vectors (σ_{XX}, σ_{YY} and σ_{ZZ}). However, some of the stress vectors maybe insignificant in comparison with the others. As a result, the first step in simplifying the model would be to identify the significant stress vectors for each of the stress transfer direction. Next, these significant channel stress vectors due to each stress transfer direction will be modeled as a function of layout variables of interest. Finally, the individual stress models would be combined to obtain the final layout dependent mobility model for process-induced stress. In this methodology no assumption has been made about the kind and number of stressors and hence the model is truly non-process specific and holistic in nature.

5.3 Modeling of Stress Transfer Components

Stress which gets transferred from substrate side (usually wafer-level strain) is mostly uniform along the inversion channel. It is also nearly independent of transistor layout and hence can be modeled simply through one layout independent parameter. Stress transferred from the other three stress transfer directions are strongly layout dependent and need to be modeled individually. In order to model the layout dependent stress, stressors are added to the transistor such that stress is transferred from only one transfer direction at a time and 3-D process simulations are run using the process simulator TAURUS. In this exercise, only the layout variables L and W are examined. Dependence of stress on other layout variables can be studied in a similar fashion.
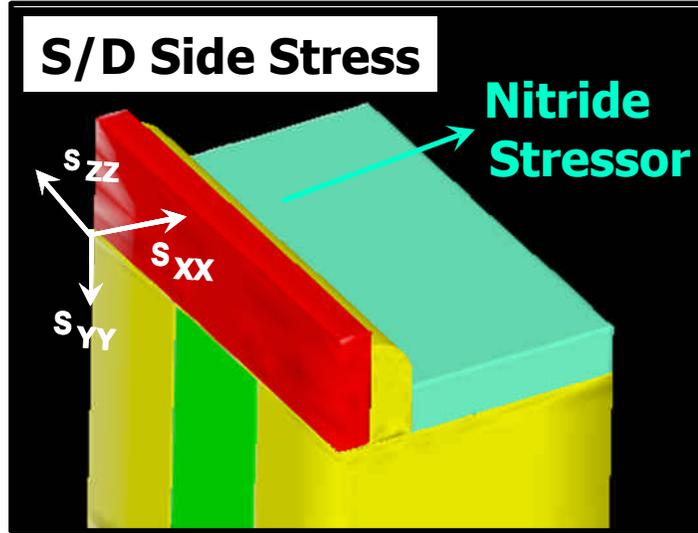
Figure 5.8: The MOSFET used to analyze S/D side stress transfer has a tensile nitride layer on the S/D area as the stressor.

5.3.1 Source-Drain (S/D) Side Stress

S/D side stress refers to the stress transferred from S/D region or transverse to S/D region, i.e along the width edge. The stressor in this case can be on top of S/D (ex. capping layer), inside the S/D (ex. SiGe S/D) or adjacent to the source drain region (ex. STI). To analyze this component of stress transfer, a 75nm thick stressed nitride layer with an internal tensile stress of 1800MPa is deposited on the S/D area of nFETs as shown in Fig. 5.8. The tensile nitride layer on S/D generates significant tensile stress in XX direction and compressive stress in YY direction both of which enhance electron mobility in nFETs. Transistors with different L and W are simulated and the resulting average channel stresses σ_{XX} and σ_{YY} are shown in Fig. 5.9. As L increases, the stressor which is on top of S/D region moves away from the center of the channel and it results in a decrease of the average stress in the channel. As W
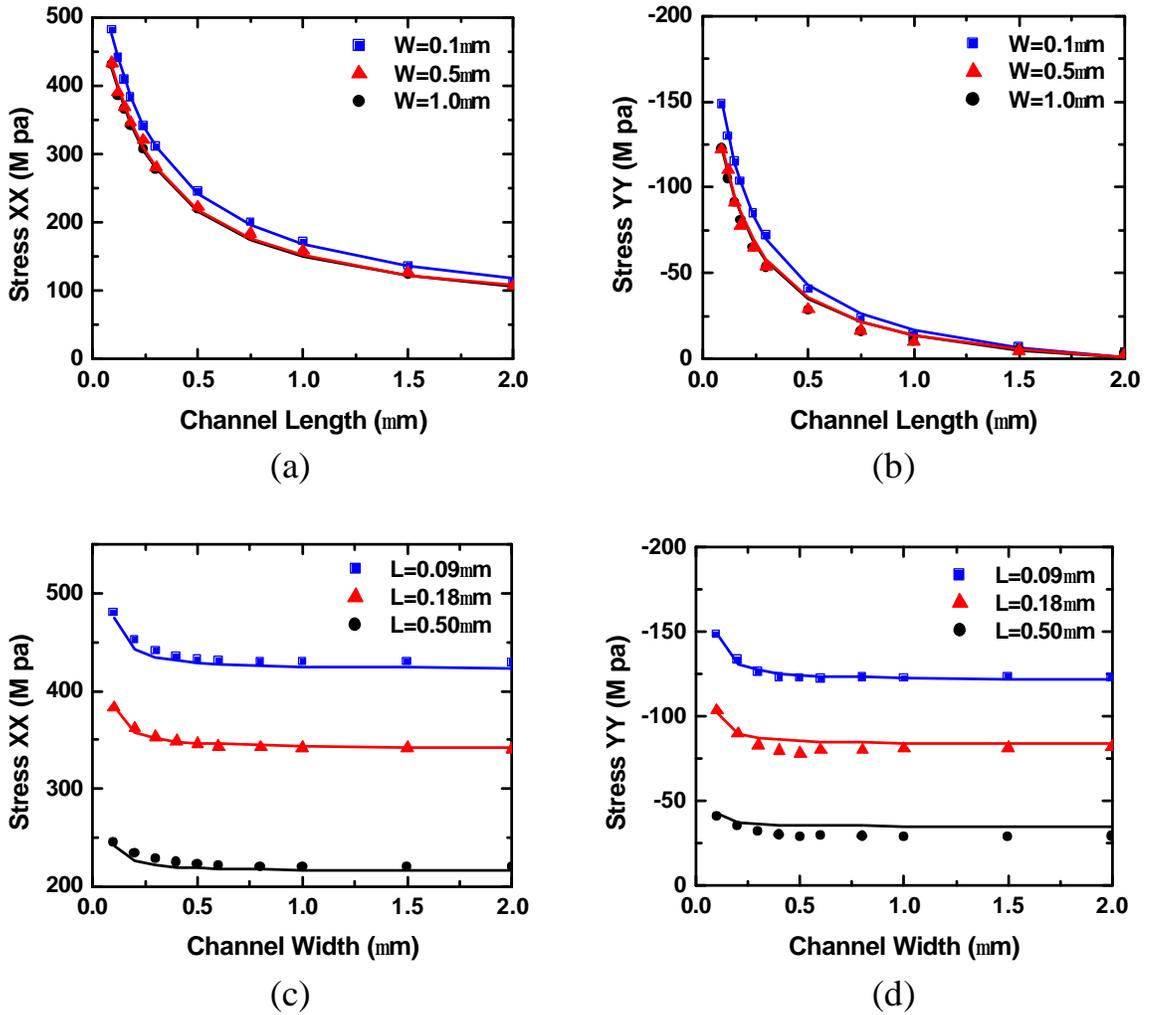
Figure 5.9: Average channel stress due to S/D side stress transfer obtained from 3-D TCAD (symbols) and model fitting (lines) (a) σ_{XX} vs. L (b) σ_{YY} vs. L (c) σ_{XX} vs. W (d) σ_{YY} vs. W

increases, there is more area for the stress to relax and so the average stress in the channel decreases. The channel length and channel width dependence of σ_{XX} and σ_{YY} can be empirically modeled through

$$
\begin{aligned}
\sigma_{XX} &= \left(A1 + \frac{A2}{A3 + L}\right) \cdot \left(B1 + \frac{B2}{B3 + W}\right) \\
\sigma_{YY} &= \left(A4 + \frac{A5}{A6 + L}\right) \cdot \left(B4 + \frac{B5}{B6 + W}\right)
\end{aligned}
\tag{5.5}
$$

where $A1 - A6$ and $B1 - B6$ are fitting parameters. It is interesting to observe that the layout dependence of both stress vectors σ_{XX} and σ_{YY} are identical.

5.3.2 Gate-stack Side Stress

Stress can also get transferred to the channel from gate-stack side (Fig. 5.7). The stressor can be a layer on top of gate electrode (ex. capping layer), the gate electrode itself (ex. stress memorization technique) or the gate-stack. In order to study this stress transfer component, transistors with 0.18μm thick gate electrode and 1800Mpa tensile intrinsic stress are simulated as shown in Fig. 5.10. Process simulations revealed significant σ_{XX} in comparison to other two stress vectors. The tensile gate electrode resulted in a compressive stress in the inversion channel which is beneficial for holes in pFETs. Fig. 5.11 shows the simulated channel length and channel width dependence of gate-stack side stress transfer which can be captured through the following empirical model

$$
\sigma_{XX} = \left(C1 + \frac{C2}{C3 + L} - \frac{C4}{C5 + L}\right) \cdot \left(D1 + \frac{D2}{D3 + W}\right)
\tag{5.6}
$$
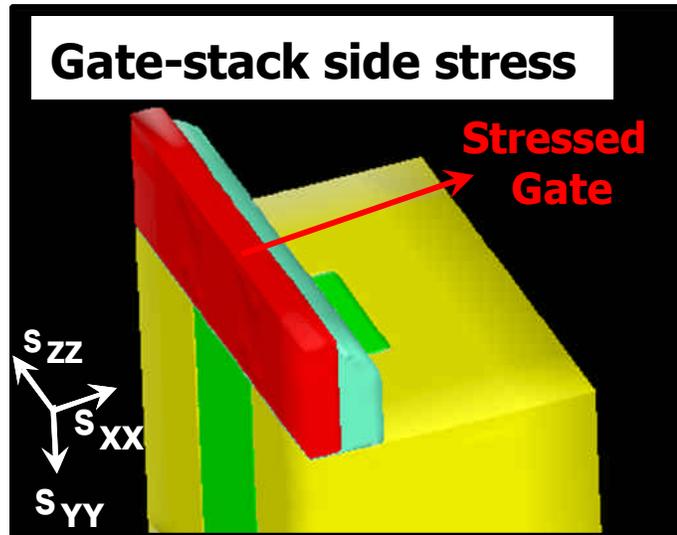
Figure 5.10: The MOSFET structure used to analyze gate stack side stress transfer. The stressed gate electrode serves as the stressor.
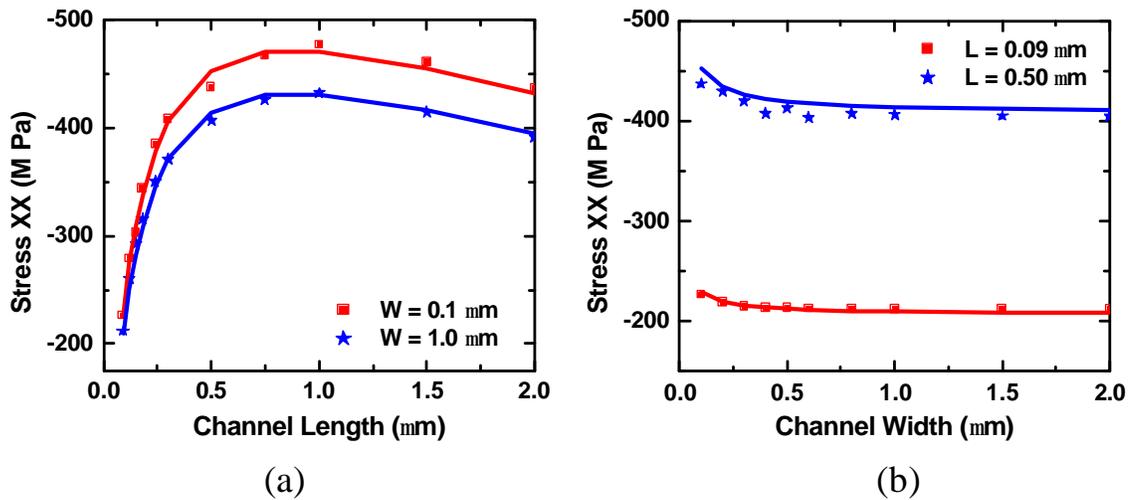
Figure 5.11: Average channel stress due to gate-stack side transfer obtained from 3-D TCAD (symbols) and model fitting (lines) (a) σ_{XX} vs. L (b) σ_{XX} vs. W. Only one stress vector - σ_{XX} is dominant.

where $C1 - C5$ and $D1 - D3$ are fitting parameters. The model fitting to the 3-D simulations is also shown in Fig. 5.11.

The non-monotonic behavior for the L dependence in Fig. 5.11(a) can be explained as follows. The volume of the stressor in Fig. 5.10 is proportional to gate length. For very short L, stressor volume is small making the average channel stress small. As L increases, the increasing stressor volume enhances the channel stress. At longer L, the stress in the center of gate is relaxed. The maximum σ_{XX} is observed at an intermediate L where the relaxation effect begins to dominate the stressor volume increase.

5.3.3 Spacer Side Stress

A stressed sidewall spacer or any stressed film outlining the spacer (ex. capping layer) can transfer stress to the channel from the spacer side. Maximum stress transfer will occur if the spacer itself is stressed. To study spacer-side stress transfer, a tensile spacer with 1800MPa intrinsic stress is used as shown in Fig. 5.12. Process simulations yielded significant compressive stress vectors in both XX and YY direction as shown in Fig. 5.13. The L and W trends for the spacer side stress are identical to those for S/D side stress and can be modeled through

$$
\begin{aligned}
\sigma_{XX} &= \left(\frac{E1}{E2+L} + E3\right) \cdot \left(F1 + \frac{F2}{F3+W}\right) \\
\sigma_{YY} &= \left(\frac{E4}{E5+L} + E6\right) \cdot \left(F4 + \frac{F5}{F6+W}\right)
\end{aligned}
\tag{5.7}
$$
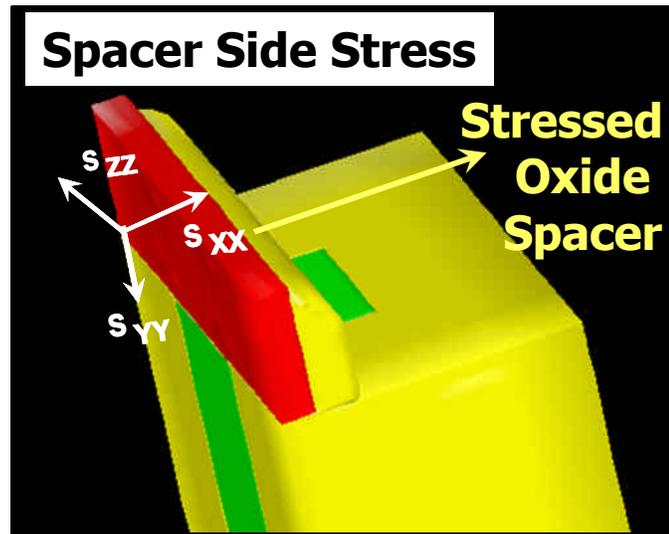
Figure 5.12: MOSFET structure used to analyze spacer-side stress transfer. The stressed spacer serves as the stressor.

where $E1 - E6$ and $F1 - F6$ are fitting parameters. Note the similarity in the layout dependencies for all the stress-transfer directions in Eq. 5.5-5.7. This observation is very useful for developing the holistic model in the next section.

5.4   Holistic Model

A complete model would be a generalized sum of the significant layout dependent stress vectors from the three stress transfer directions studied in Section 5.3. The number of model parameters in such a generalized model would be very large and impossible to extract. The efficacy of any model strongly depends not only on the flexibility provided by the fitting parameters but also the ease of model parameter extraction. In order to make parameter extraction easier, simplifying assumptions need to be made. Assuming that multiple stressors will result in two dominant stress
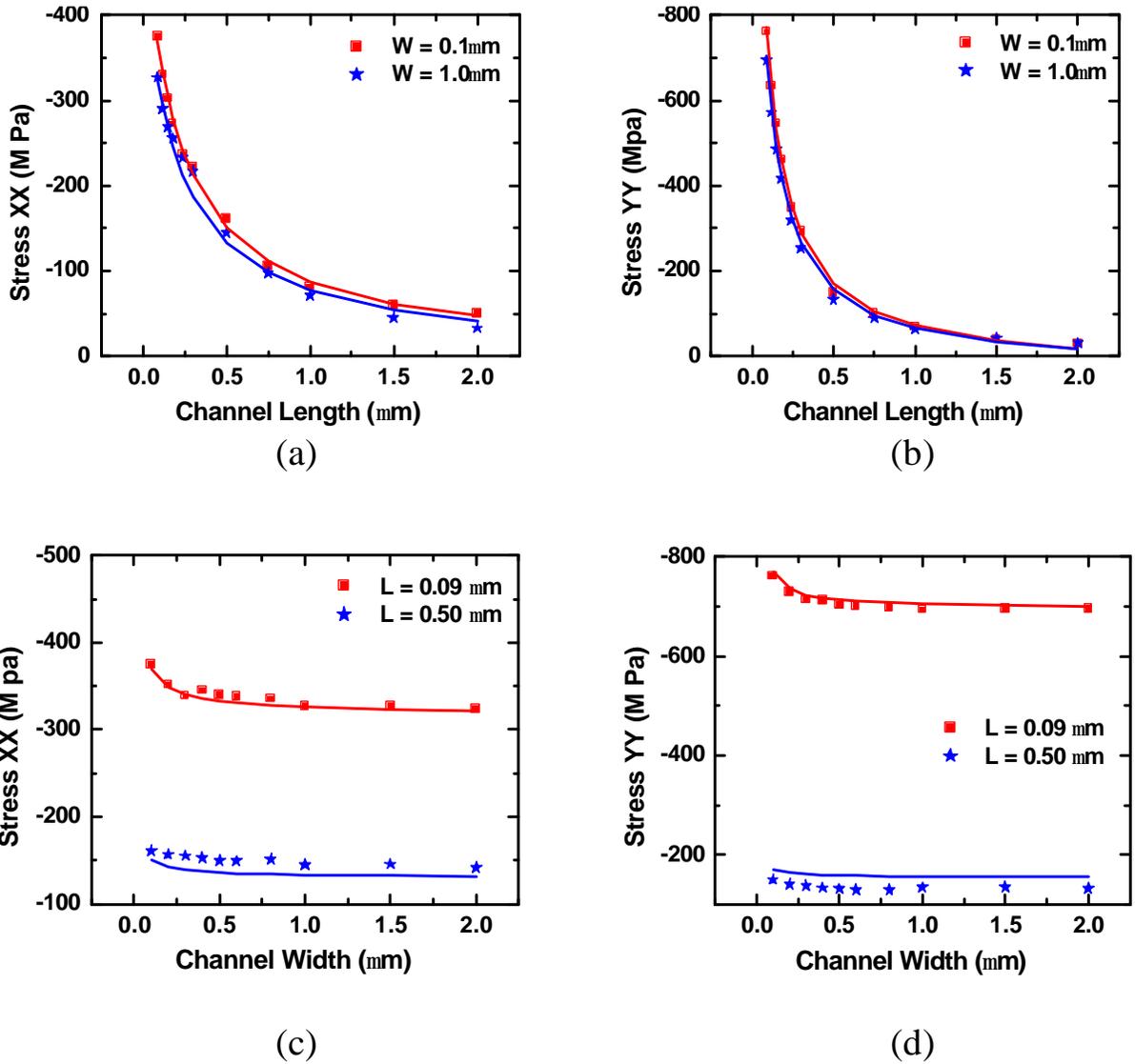
Figure 5.13: Average channel stress due to spacer side transfer obtained from 3-D TCAD (symbols) and model fitting (lines) (a) σ_{XX} vs. L (b) σ_{YY} vs. L (c) σ_{XX} vs. W (d) σ_{YY} vs. W

vetors, a simplified layout dependent mobility model can be written as

$$
\begin{aligned}
\frac{\mu}{\mu_0} \;=\;& 1 + \left(A1 + \frac{A2}{A3 + L} - \frac{A4}{A5 + L} \right) \cdot \left(B1 + \frac{B2}{B3 + W} - \frac{B4}{B5 + W} \right) \\
& + \left(A6 + \frac{A7}{A8 + L} - \frac{A9}{A10 + L} \right) \cdot \left(B6 + \frac{B7}{B8 + W} - \frac{B9}{B10 + W} \right)
\end{aligned}
\tag{5.8}
$$

where $A1 - A10$ and $B1 - B10$ are fitting parameters. This simplification relies on the fact that the L and W dependence of average channel stress in all stress transfer directions follows the same $1/L$ and $1/W$ scaling trends. No assumptions have been made on the specifics of the stressor. The model is applicable to any number and any kind of stressors around the transistor which makes it holistic in nature. The fitting parameters can be extracted through mobility measurements on different L and W devices.

5.5   Model Verification

The holistic model is verified against several process-induced strain technologies through full scale 3-D process simulations using the device simulator TAURUS. The verification is done as follows. For any given strain technology, the average channel stress is first calculated for all the three stress vectors. Then, the average stress vectors are multiplied with the corresponding piezoresistive coefficients and added together to yield the net simulated change in carrier mobility.

$$
\frac{\Delta \mu}{\mu_0} \;=\; - \left(\pi_{XX} \cdot \sigma_{XX} + \pi_{YY} \cdot \sigma_{YY} + \pi_{ZZ} \cdot \sigma_{ZZ} \right)
\tag{5.9}
$$

Carrier / Piezo	Holes (1e-11 Pa$^{-1)}$)	Electrons (1e-11 Pa$^{-1)}$)
p_{xx}	71.8	-31.6
p_{yy}	-1.1	53.4
p_{zz}	-66.3	-17.6

Figure 5.14: Piezoresistive coefficients for electrons and holes on (001) silicon surface.

Fig. 5.14 shows the values of piezoresistive coefficients for a (100) silicon surface for holes and electrons used in this section. The holistic model in Eq. 5.8 is then used to describe the simulated change in carrier mobility as a function of the layout variables L and W.

The holistic model is first verified against the capping layer process. The capping layer process transfers stress to the channel through all the three layout dependent stress transfer directions. A 75nm thick nitride layer with an intrinsic tensile stress of 1800MPa is deposited on a nFET as shown in Fig. 5.15. It results in a strong tensile stress along the channel length and a compressive stress into the substrate both of which benefit the electron mobility. Fig. 5.16 shows the simulated change in electron mobility as a function of L and W. The holistic model is able to capture the simulated electron mobility change very well. In this case only 6 model parameters were sufficient to capture both the L and W variation.

$$\frac{\mu}{\mu_0} = 1 + \left(A1 + \frac{A2}{A3 + L}\right) \cdot \left(B1 + \frac{B2}{B3 + W}\right) \quad (5.10)$$
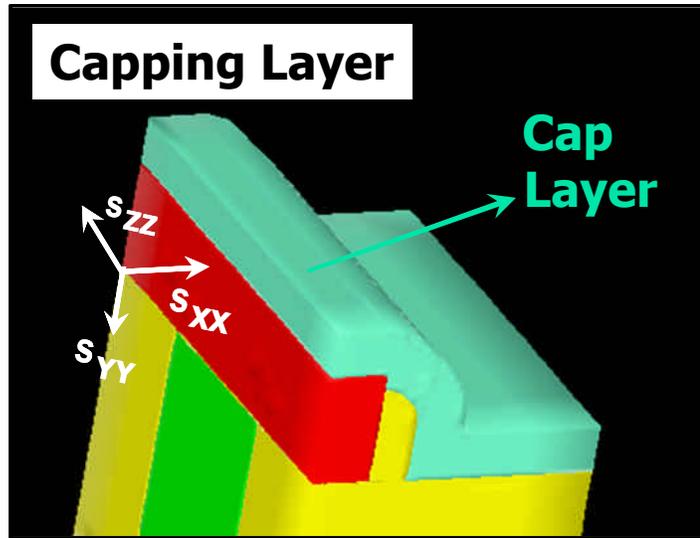
Figure 5.15: MOSFET with a tensile capping layer.
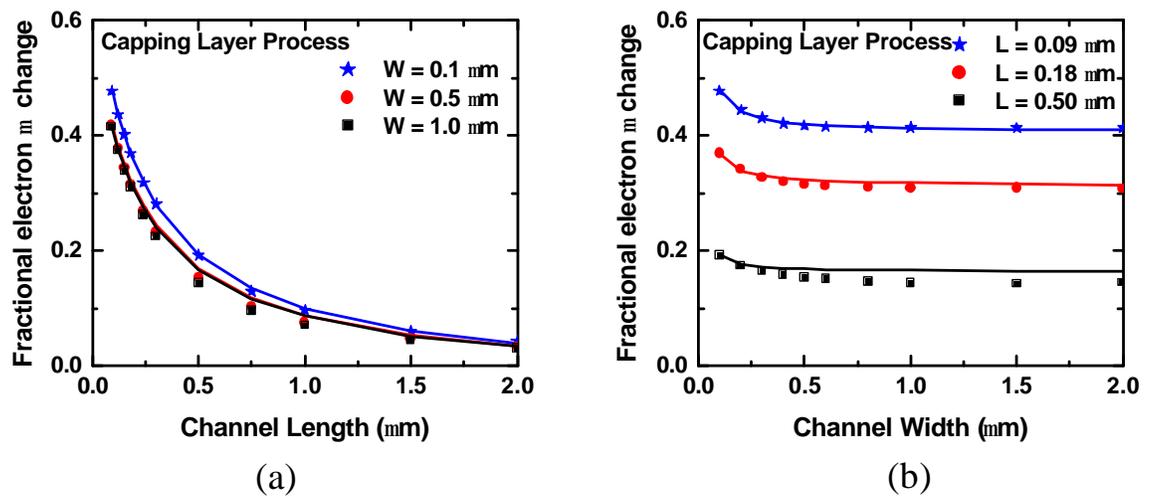
Figure 5.16: Fractional change in electron mobility for nFETs with a tensile capping layer as a function of (a) channel length (L) and (b) channel width (W). (Symbols : 3-D TCAD, Lines : Holistic Model Fitting)
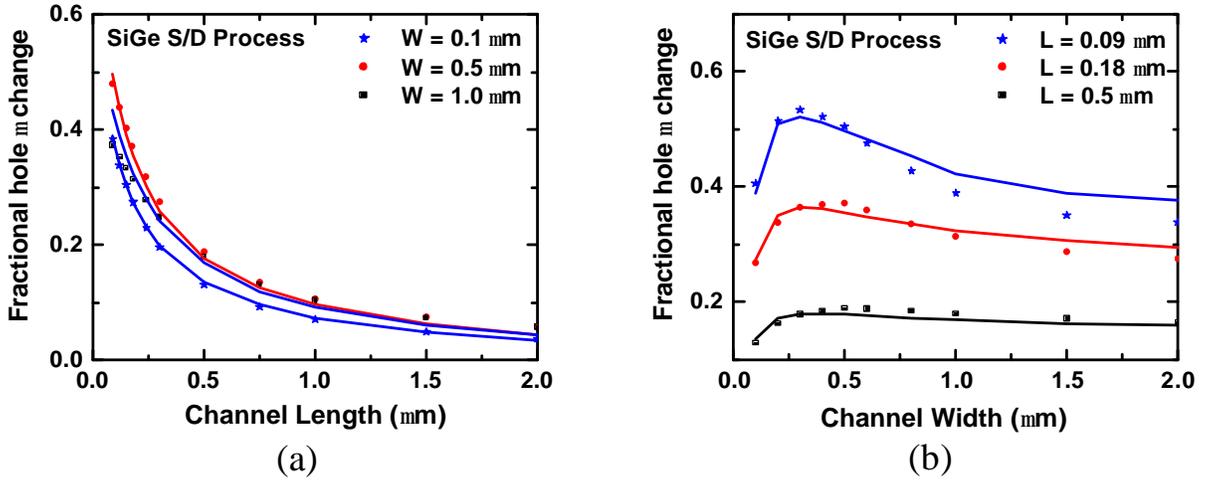
Figure 5.17: Fractional change in hole mobility for pFETs with SiGe S/D as a function of (a) Channel length and (b) Channel width. (Symbols : 3-D TCAD, Lines : Holistic Model Fitting)

The holistic model is next verified against SiGe S/D technology which is used in state-of-the-art pFETs. The mismatch in the lattice constant between SiGe S/D and silicon channel induces a strong compressive stress along the channel length which enhances the hole mobility in pFETs. The SiGe S/D also creates tensile stress along ZZ direction which further boosts the hole mobility. The interplay of these two stress vectors determines the net gain in hole mobility for pFETs. p-FETs with 100nm deep SiGe S/D and 20% Ge content are simulated. Fig. 5.17 shows the simulated hole mobility change for the SiGe S/D process as a function of L and W. As expected, when L increases, both σ_{XX} and σ_{ZZ} decrease resulting in net decrease in hole mobility. However, when W increases, σ_{XX} and σ_{ZZ} show different behavior. σ_{XX} increases monotonically but σ_{ZZ} shows non-monotonic behavior which explains the non-monotonicity in Fig. 5.17(b). The model fitting to the simulated mobility
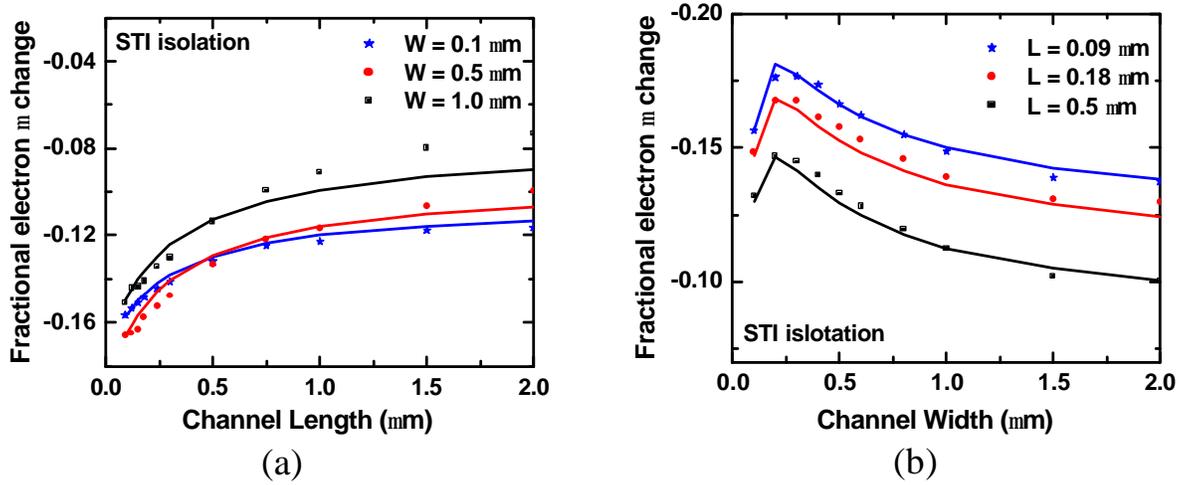
Figure 5.18: Fractional change in electron mobility in nFETs due to STI isolation as a function of (a) Channel length and (b) Channel width. (Symbols : 3-D TCAD, Lines : Holistic Model Fitting)

change is also shown in Fig. 5.17.

Modern CMOS technologies use STI for transistor isolation. The oxidation step involved in STI process together with the thermal expansion coefficient mismatch and internal stress of the deposited material in the trench induces stress in neighboring transistors. To simulate the STI stress, 500nm deep trench is used. The STI generates strong compressive stress along XX and ZZ directions. The coefficients π_{XX} and π_{ZZ} for electrons are of the same order in magnitude but opposite in sign (Fig. 5.14). This results in interesting trends for electron mobility in nFETs as shown in Fig. 5.18. The crossover in L dependence of mobility change in Fig. 5.18(a) demonstrates the necessity of modeling two dominant stress vectors. The crossover cannot be captured by a single term in Eq. 5.8. Fig. 5.18 demonstrates that the holistic model is able to capture the simulated layout dependent mobility changes for STI process as well.
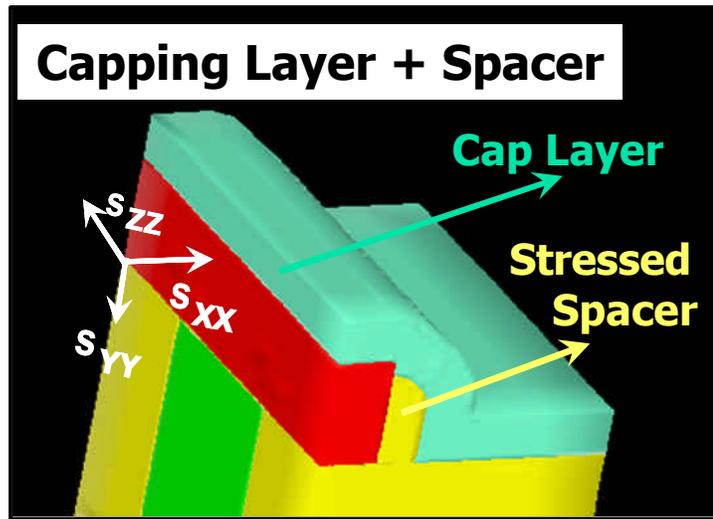
Figure 5.19: MOSFET with two stressors - stressed capping layer and stress spacer.

The holistic nature of the model can be further tested against a multiple stressor technology. Fig. 5.19 shows a n-channel transistor with two stressors - a tensile capping layer and a tensile stressed spacer. The tensile capping layer is 75nm thick with an intrinsic stress of 1800MPa. The spacer also has a tensile intrinsic stress of 1800MPa. The tensile capping layer produces a strong tensile σ_{XX} and a compressive σ_{YY} and the tensile spacer produces strong compressive σ_{XX} and σ_{YY}. As a result, the transistor exhibits a strong compressive σ_{YY}. This simulation exercise also demonstrates the additive nature of stresses. Fig. 5.20 shows that the sum of σ_{YY} of a transistor with only tensile spacer and σ_{YY} of a transistor with only capping layer is nearly equal to the σ_{YY} of a transistor with both the stressors. Fig. 5.21 shows the simulated electron mobility change for transistors with both stressors and the corresponding holistic model fitting. The good fitting results affirm the belief in the holistic nature of the model.
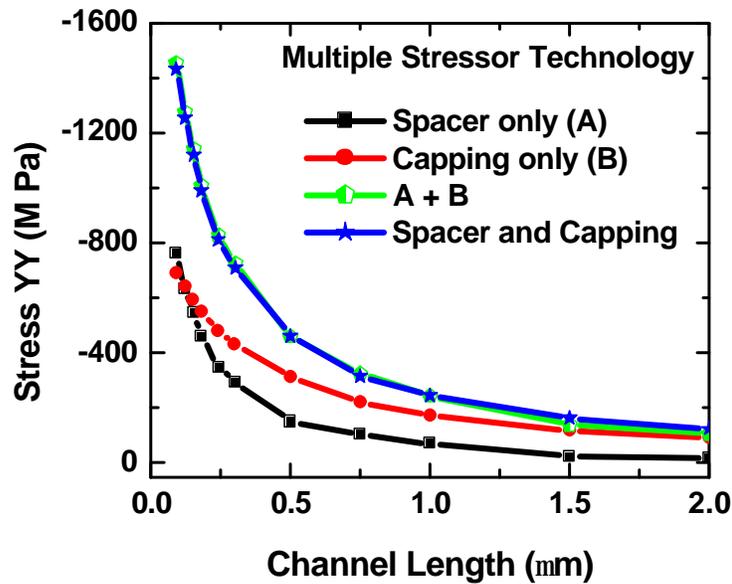
Figure 5.20: Total stress in a multiple stressor technology is equal to the sum of stress due to the individual stressors one at a time.

(a)

(b)
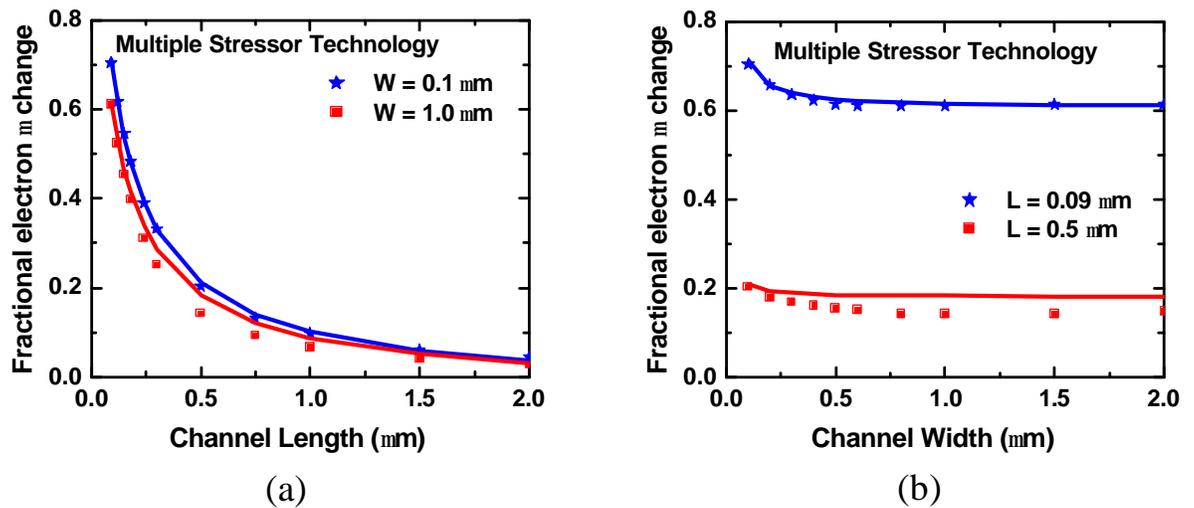
Figure 5.21: Fractional change in electron mobility in nFETs due the multiple stressor technology in Fig. 5.19 (tensile capping and tensile spacer) as a function of (a) Channel length and (b) Channel width. (Symbols : 3-D TCAD, Lines : Holistic Model Fitting)
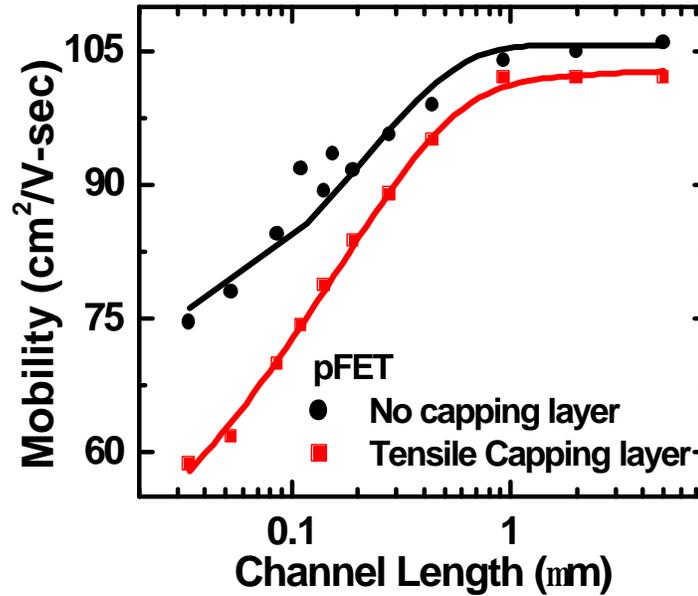
Figure 5.22: Holistic model fitting to pFET mobility with a tensile capping layer. (Symbols : Experimental data, Lines : Holistic Model Fitting)

The model is also verified against experimental data from literature. Experimental data for a capping layer process is obtained from [72] which reports mobility reduction for PMOS devices with a tensile capping layer. BSIM mobility model is first matched to the control wafer. Holistic mobility model (Eq. 5.8) is then added to BSIM4 mobility model to capture the mobility change due to capping layer. A good fit is obtained to the measured mobility of transistors with capping layer using the holistic mobility model as seen in Fig 5.22. The model is also verified against experimental data for a SiGe S/D process. Gain in linear drain current after correcting for $V_g - V_{th}$ was reported for pFETs with SiGe S/D in [66] . The holistic model is able to match the measured gain in linear drain current for different channel lengths through a change in hole mobility as illustrated in Fig. 5.23.
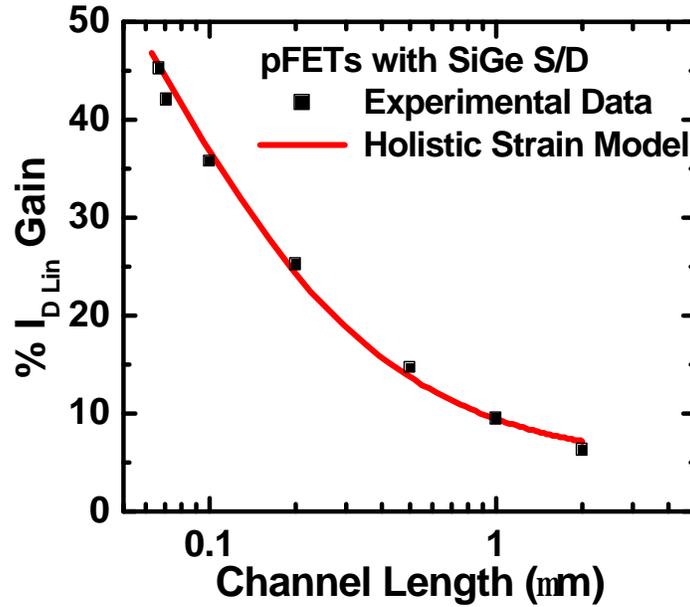
Figure 5.23: Holistic model fitting to gain in linear regime drain current for pFETs with SiGe S/D. (Symbols : Experimental data, Lines : Holistic Model Fitting)

5.6 Conclusions

A non-process specific holistic model for layout dependent mobility enhancement through process-induced stress is developed. Individual stress transfer directions are first identified and then analyzed to model the corresponding layout dependent average channel stress vectors. A holistic model is then developed by observing the layout dependencies of the individual stress transfer directions. The developed holistic model can be easily incorporated into any compact model through a simple modification of the mobility term.

Channel length and channel width dependencies of mobility enhancement have been modeled. The approach used in this work can be extended to add other layout variables when need arises. The model has been verified against both 3-D process

simulations and experimental data for a wide range of stress-inducing processes including capping layer, SiGe S/D and STI process. The model has also been verified against the interesting case of multiple stressors surrounding a single transistor.

Chapter 6

A Statistical Model for Flicker Noise

in Small Area MOSFETs

6.1   Introduction

Recent years have seen an enormous growth in RF circuit applications. The continuous downscaling of bulk planar CMOS into deep-submicron dimensions has led to MOSFETs with very high unity-gain frequencies making CMOS attractive for RF applications [73], [74]. To investigate the potential applications of advanced CMOS in low-noise RF circuits, accurate modeling of noise in transistors is vital. The major sources of noise in a transistor are the thermal noise and the low frequency $1/f$ noise.

The low frequency $1/f$ noise in CMOS impacts the performance of RF CMOS circuits due to noise up-conversion such as in VCOs and low IF mixers where it causes a significant increase in the noise figure [75]. The low frequency noise also

154
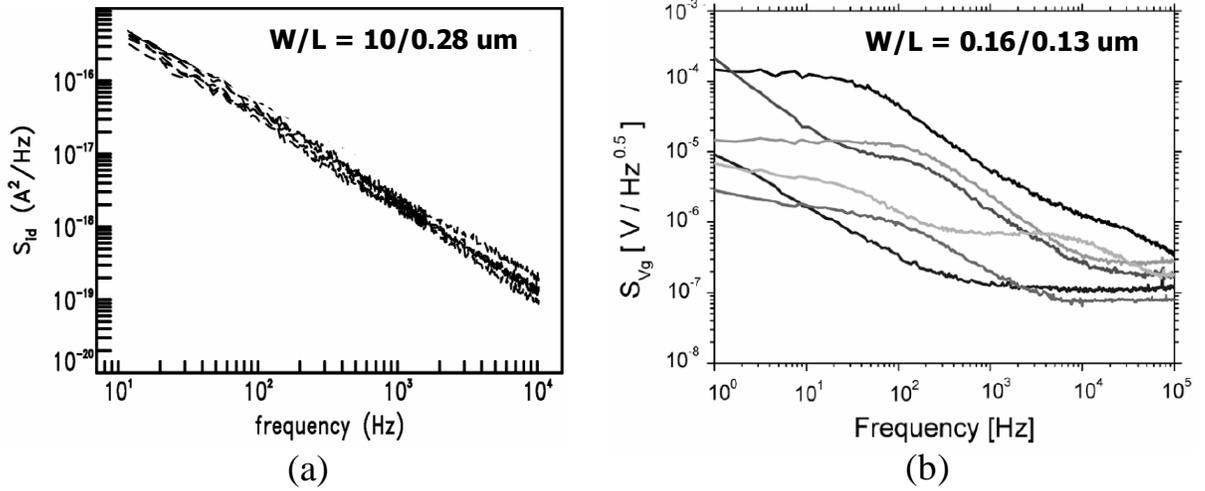
Figure 6.1: Measured FN characteristics of (a) large area device [76] and (b) small area device [77]. In small area devices, the flicker noise spectrum differs greatly from $1/f$ and exhibits very large variation.

affects analog circuits such as high performance operational amplifiers and precision ADC/DACs. In scaled CMOS, the reduction in the maximum signal swing lowers the Signal-to-Noise Ratio (SNR) making the low frequency $1/f$ noise even more important. Compact models describing the low frequency $1/f$ noise (a.k.a Flicker Noise - FN) characteristics of MOSFETs are needed to predict the noise performance of both RF and analog circuits.

The FN characteristics of small area MOSFETs are very different from those of large area MOSFETs. Fig. 6.1(a) shows FN of large area FETs with $W/L = 10/0.28$ μm [76] and Fig. 6.1(b) shows FN of small area FETs with $W/L = 0.16/0.13$ μm [77]. The large area devices show neat $1/f$ behavior and the variation in noise among different devices with same area is very small. A unified flicker noise model was developed in [78] to model the FN for large area devices. As the transistor area is scaled down,

the low frequency noise appears more Lorenztian-like instead of $1/f$ shape and it exhibits more device-to-device variations. The variability in FN is becoming increasingly important for circuits using scaled CMOS and needs to be captured in MOSFET noise models. It is thus necessary to develop a statistical compact model for modeling flicker noise to enable a robust design of RF and analog circuits using advanced CMOS technologies. In this chapter, such a model is developed which captures the noise characteristics as a function of device size and voltage bias.

Section 6.2 briefly describes the origin of FN in CMOS and explains the cause of differences in FN among large and small area devices. The statistical modeling framework for FN is introduced in Section 6.3 and the complete statistical model is developed in Section 6.4 and Section 6.5. The implementation of the model in BSIM4 compact model is discussed in Section 6.6. Finally, the verification of the model and model results are presented in Section 6.7.

6.2   Flicker Noise : Large area vs. Small area FETs

The origin of FN lies in the traps present in the gate dielectric. When the inversion channel carriers are trapped and de-trapped in the dielectric traps, the drain current fluctuates with time and is commonly referred to as Random Telegraphic Noise (RTN) [79]. The fluctuations in the drain current are dictated by the changes in number of carriers in inversion channel and by the changes in surface carrier mobility associated with carrier scattering by the trapped charges in gate dielectric [78]. Flicker noise is simply the frequency domain representation of RTN and it represents the total drain
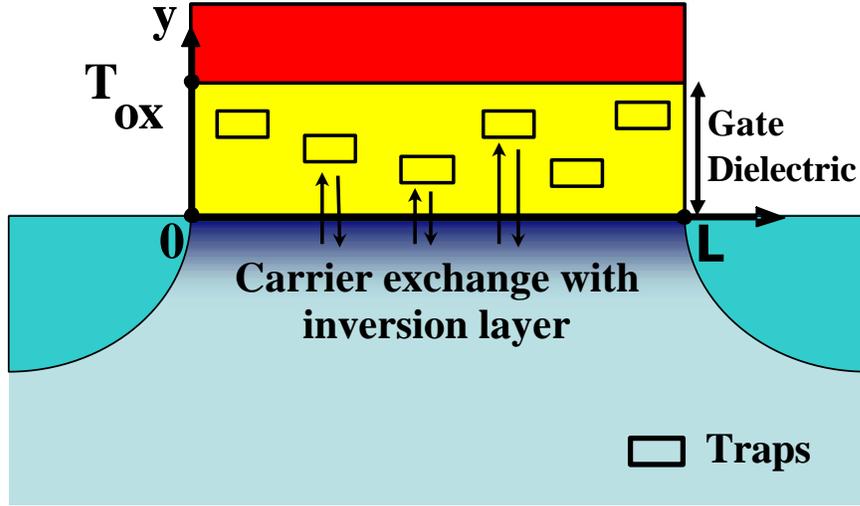
Figure 6.2: Schematic of the transistor under study. FN is caused by carrier exchange between dielectric traps and inversion layer. y axis is along the depth of gate dielectric.

current noise power. It is often denoted by $S_{I_d}(f)$ or by $S_{V_g}(f)$ when referred to the gate terminal $(S_{V_g}(f) = S_{I_d}(f)/g_m^2$ where g_m is the MOSFET trans-conductance).

In a large area device, there are several traps in the gate dielectric as shown in Fig. 6.2. Each trap is characterized by a unique time constant τ representing the rate of carrier capture and carrier emission. Since the trap capture and emission happens primarily through a tunneling process, the trapping rate decreases exponentially with the depth of the trap in the dielectric, y, and hence the trapping time constant τ is given by

$$\tau = \tau_0 \cdot \exp(\gamma y) \tag{6.1}$$

where τ_0 is the time constant for traps at silicon interface and γ is the tunneling coefficient. Typically, τ_0 is 10^{-10}s and γ is 10^8cm^{-1} for SiO$_2$ dielectric. The RTN

associated with an individual trap with time constant τ can be expressed as [80]

$$RTN(y,f) \propto \frac{\tau}{1+(\omega\tau)^2} \tag{6.2}$$

where $\omega = 2\pi f$. The FN in a transistor is simply the sum of all the RTNs. Assuming a spatially uniform trap density inside the dielectric, the FN for a large area device can be written as

$$S_{I_d}(f) \propto \int_0^{T_{ox}} RTN(y,f)dy \tag{6.3}$$

Carrying out the above integration, the FN for large area device can be expressed as

$$S_{I_d}(f) \propto \frac{\tan^{-1}(e^{\gamma T_{ox}}\omega\tau_0) - \tan^{-1}(\omega\tau_0)}{\omega\gamma} \tag{6.4}$$

For large γT_{ox}, in the frequencies of interest, $\tan^{-1}(e^{\gamma T_{ox}}\omega\tau_0) \approx \pi/2$ and $\tan^{-1}(\omega\tau_0) \approx 0$ yielding the well known "$1/f$" expression for FN in large area devices.

$$S_{I_d}(f) \propto \frac{1}{4\gamma f} \tag{6.5}$$

The summation of large number of RTNs leading to $1/f$ frequency dependence of FN is pictorially depicted in Fig. 6.3 (a). Each trap contributes a Lorenztian with a certain corner frequency. The sum of many Lorenztian spectra with the corner frequencies distributed exponentially yields the experimentally observed $1/f$ shape for FN in large area devices as shown in Fig. 6.1(a). Since the number of traps in
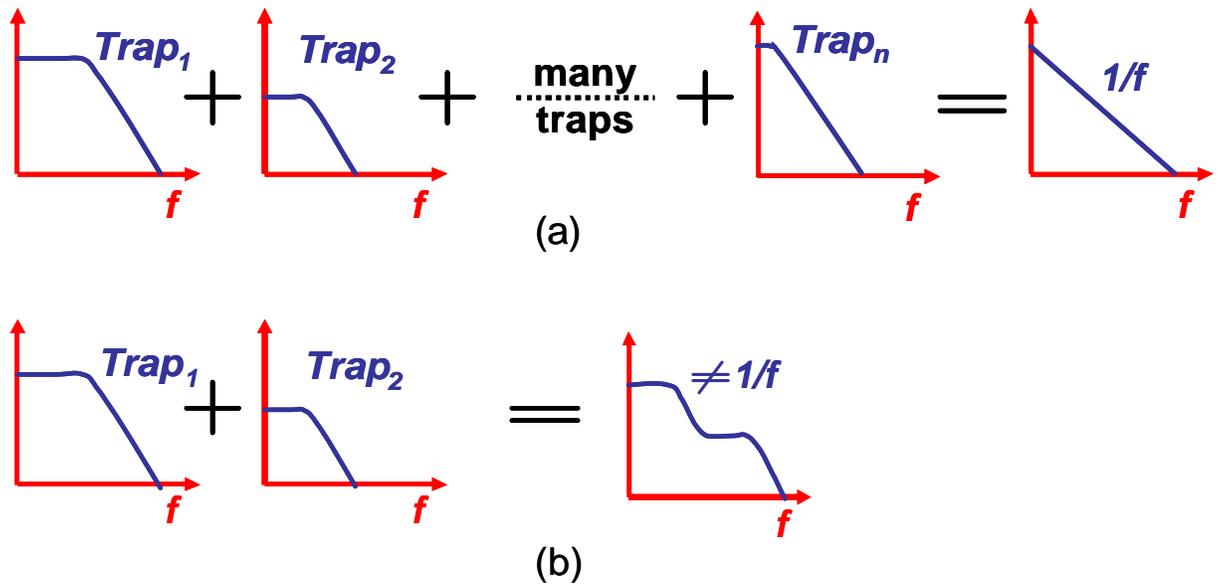
Figure 6.3: Graphical illustration of the origin of (a) 1/f noise in large devices (b) Lorentzian-like spectrum in small devices. The individual spectra in the summation in (a) and (b) represent the RTN from a single trap in the dielectric.

large area device is large, any small device-to-device variation in the number of traps is averaged out and FN remains essentially the same for different devices with same geometries as seen in Fig. 6.1(a).

The FN characteristics are very different in small area devices compared to large devices even though the physical origin of FN is the same - carrier trapping and de-trapping. In a small area device, there are only a few traps in the gate dielectric. As a result, only few time constants exist and the sum of the few RTN yields Lorenztian-like spectrum as illustrated in Fig. 6.3(b) and is observed experimentally as shown in Fig. 6.1(b). The FN in small area devices can vary a lot depending on the time constants associated with the few traps in the device. For example, if one of the two traps in Fig. 6.3(b) is absent, the FN for the small area device will be very different.

Due to the inherent random nature of trap distribution inside the gate dielectric, the trap profile in any given transistor is not known beforehand. Since the trap profiles of different small area devices with same geometries can be very different, a statistical model is the only way to capture the FN in small area devices. The first step in formulating the statistical FN model is identifying the necessary statistical variables which is the subject of the next section.

6.3   Statistical Model Variables

In a small area device, there are only few traps present in the gate dielectric and the FN ($S_{Id}(f)$) is obtained by a discrete sum of the RTNs of the individual traps.

$$S_{I_d}(f) = \sum_{0}^{N_{tr}} RTN(y, f)dy \tag{6.6}$$

N_{tr} is the number of traps present in the small area device. In order to evaluate Eq. 6.6, the number of traps in the device (N_{tr}) and $RTN(y, f)$ for each trap should be known. The RTN is a function of several trap characteristics.

A trap can be physically located anywhere inside the insulator as shown Fig. 6.4(a). The depth of the trap inside the insulator (y location) determines the trap time constant (τ) as shown in Eq. 6.1. Traps which are closer to the silicon interface are fast traps and they have smaller τ contributing to noise at high frequencies. Traps located at the far end of dielectric at gate interface are slow traps since the carriers have to tunnel though entire gate dielectric and they contribute to the noise at lower frequencies.
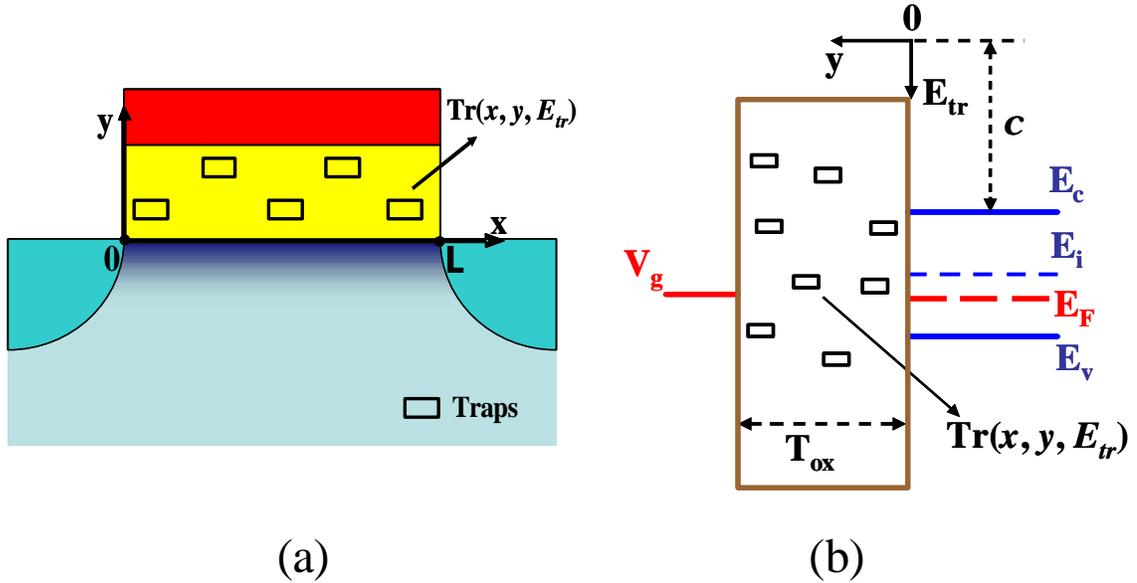
Figure 6.4: The RTN generated by a trap depends on (a) Spatial location of the trap inside the gate dielectric (x and y location) and (b) Energy of the trap relative to carrier fermi-level. (χ is the electron affinity of silicon). The number of traps together with the individual trap characteristics constitute the statistical model variables.

The RTN generated by the trap is also a strong function of the location of the trap along the channel length (x location). Traps which are located close to source and drain have almost no impact on the FN while the traps located at the minimum potential barrier location between source and drain have the largest impact on FN [81]. This result assumes a silicon substrate with uniformly distributed dopant atoms. In extremely small devices, the dopants are scattered in the channel and the current is no longer a streamline flow. The random dopant atoms in small area devices create surface potential fluctuations. The potential fluctuations result in strong current flow (current percolation) along the potential valleys. The location of the trap with respect to the percolation paths decides the RTN generated by the trap. This is referred to as the percolation effect. The effect of uncertainty in x location and percolation

phenomenon on FN is nearly indistinguishable. They both increase the amount of device-to-device variation in FN. As a result, they will be quantified together by a parameter PF (percolation factor) which changes the magnitude of RTN generated by an individual trap.

The energy level of the trap (E_{tr}) also determines the RTN generated by the trap. The traps are distributed in the energy space as shown in Fig. 6.4(b). Traps which are located close to the fermi-level of the device (E_F) exchange carriers easily and contribute maximally to FN. Traps which are away from E_F are relatively inactive and contribute less to FN.

For a given trap density and given device area, the number of traps in different devices is variable. Furthermore, the trap profile for each trap in terms of its location, percolation factor and trap energy is also variable. The number of traps in the device, N_{tr}, together with the parameters - y, E_{tr} and PF for each trap in the device constitute the statistical model variables. Monte Carlo estimation of these statistical model variables will be presented in the next section.

6.4   Statistical Trap Profiling

In order to determine the RTNs contributing to the total FN, the number of traps in the device and trap profiles for each trap need to be determined. This process will be referred to as trap profiling. Monte-Carlo estimation of the statistical model variables will be presented one by one.
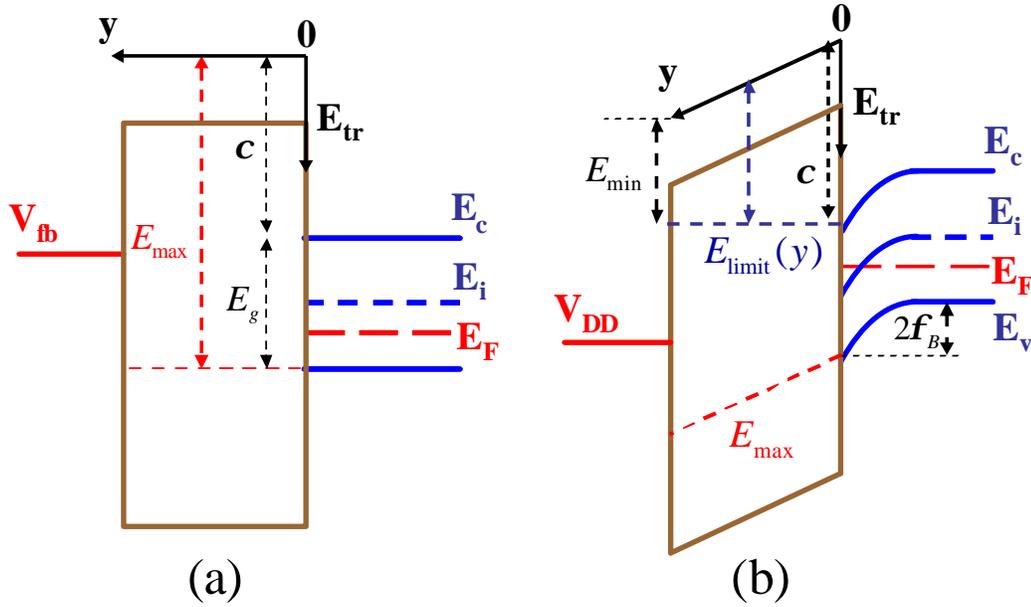
Figure 6.5: (a) Maximum trap energy (E_{max}) and (b) Minimum trap energy (E_{min}) which contributes to the FN. Using this band of trap energies to estimate N_{avg}.

6.4.1   Number of traps : N_{tr}

Though the number of traps varies from device to device, the average number of traps for an ensemble of devices with a given device area is fixed in a given technology. Assume the average trap density is N_T cm^{-3}eV^{-1} for a certain technology which is fixed for all devices in that technology. The average number of traps generating FN depends on the areal volume $W \cdot L \cdot T_{ox}$ and the range of trap energies which will influence FN. It is reasonable to assume that only those traps which lie between the silicon conduction band (E_c) and valence band (E_v) contribute to FN. Traps outside the silicon bandgap will have negligible effect on FN. Fig. 6.5(a) shows the band diagram of a nFET biased in flatband ($V_g = V_{fb}$). The traps which are aligned with

the E_v represent the maximum energy (E_{max}) traps that will ever contribute to FN.

$$E_{max} = \chi + E_g \tag{6.7}$$

where χ and E_g are the electron affinity and band-gap of silicon substrate respectively. Similarly, when the transistor is biased in strong inversion ($V_g = V_{DD}$), the traps aligned with the silicon E_c represent the minimum energy traps which will ever contribute to FN as shown in Fig. 6.5(b). The smallest trap energies contributing to FN exist at gate interface and the minimum trap energy E_{min} can be calculated as

$$E_{min} = max(\chi - (V_{DD} - V_{FB} - 2\phi_B), \chi - \phi_{barrier}) \tag{6.8}$$

where the surface potential in strong inversion is assumed to be $2\phi_B$ and $\phi_B = \frac{kT}{q} \ln\left(\frac{N_A}{n_i}\right)$ where N_A is body doping. $\phi_{barrier}$ is the barrier height between silicon channel and gate dielectric. With the range of trap energies defined, the average number of traps in the device (N_{avg}) is given by

$$N_{avg} = N_T \cdot W \cdot L \cdot T_{ox} \cdot (E_{max} - E_{min}) \tag{6.9}$$

The number of traps in an arbitrary device sample is dictated by Poisson's statistics. The probability of finding exactly N_{tr} traps in one device among a group of similar sized devices with average trap number N_{avg} is

$$P(N_{tr}) = \frac{N_{avg}^{N_{tr}} \cdot e^{-N_{avg}}}{N_{tr}!} \tag{6.10}$$
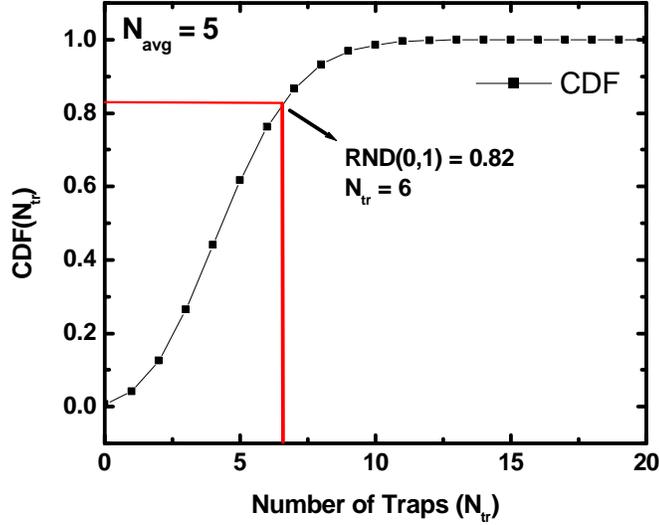
Figure 6.6: Graphical illustration of the Monte-Carlo selection process of N_{tr} by using $CDF(N_{tr})$. The random number generator produces $RND(0,1) = 0.82$ which corresponds to a random device sample with $N_{avg} = 5$ having 6 traps.

The number of traps N_{tr} in the device sample is then estimated through a Monte Carlo process. A cumulative distribution function $CDF(N_{tr})$ defined as

$$CDF(N_{tr}) = \sum_{0}^{N_{tr}} P(N_{tr}) \qquad (6.11)$$

is first constructed. Next, a random number is picked between 0 and 1 and the number of traps N_{tr} in the device is chosen such that

$$CDF(N_{tr} - 1) < RND(0,1) < CDF(N_{tr}) \qquad (6.12)$$

The function $RND(0,1)$ generates a random real number between 0 and 1. This process of Monte Carlo estimation of Poisson's distributed N_{tr} is graphically illus-

trated in Fig. 6.6 for a case of $N_{avg} = 5$.

6.4.2 Trap location inside insulator : y

A uniform spatial density in the insulator yields $1/f$ FN for large device (Eq. 6.5). However, measurements on large area devices show $1/f^\alpha$ FN where $0.8 < \alpha < 1.2$ suggesting non-uniform trap density in the dielectric [82]. $\alpha > 1$ implies existence of more low frequency traps, i.e more traps away from the silicon interface. Similarly, $\alpha < 1$ can be interpreted as more high frequency traps i.e higher trap density at the interface. The $\alpha \neq 1$ behavior in large area devices can be modeled through exponentially distributed traps along the depth of the insulator with a distribution $N_T(y)$ defined as $N_T(y) = N_{T0} \exp(ay)$ as shown in Fig. 6.7 [83]. FN of a large area device with an exponential trap distribution can be shown to be

$$S_{I_d}(f) \propto \frac{1}{\gamma f^\alpha} \quad \text{where} \quad \alpha = 1 + a/\gamma \tag{6.13}$$

When the trap distribution is uniform ($a = 0$) $\alpha = 1$ as expected from Eq. 6.5. The trap distribution coefficient a can be easily extracted from the slope of large area FN.

In the statistical FN model, both exponential and uniform trap distributions $N_T(y)$ are allowed. The probability distribution function ($PDF(y)$) for an exponentially distributed trap density is

$$PDF(y) = \frac{a \cdot e^{ay}}{e^{aT_{ox}} - 1} \tag{6.14}$$
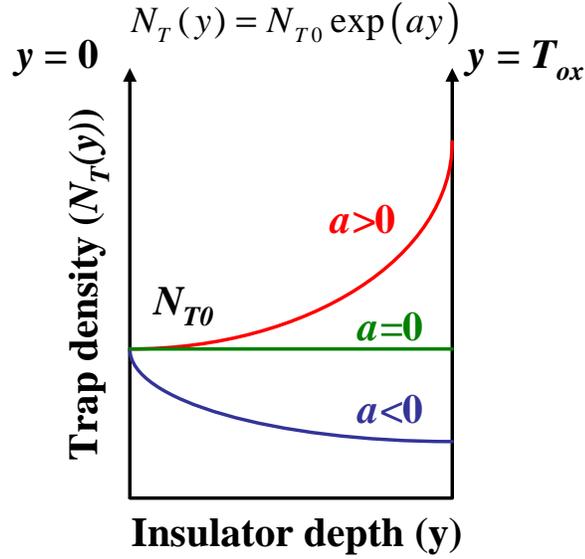
Figure 6.7: Allowed trap density distributions as a function of y in the FN statistical model. The exponential trap density distribution enables modeling of experimentally observed $1/f^\alpha$ noise spectrum where $\alpha \neq 1$.

For Monte-Carlo estimation of y location, the corresponding cumulative distribution function $CDF(y)$ is obtained.

$$CDF(y) \;\; = \;\; \frac{e^{ay} - 1}{e^{aT_{ox}} - 1} \tag{6.15}$$

Since y is a continuous variable and $CDF(y)$ is invertible, Monte Carlo generated y location for any given trap is simply

$$
\begin{aligned}
y \;\; &= \;\; \frac{1}{a} \ln \left[RND(0,1) \cdot \left(e^{aT_{ox}} - 1 \right) + 1 \right] && \text{if} \;\; a \neq 0 \\
&= \;\; RND(0,1) \cdot T_{ox} && \text{if} \;\; a = 0
\end{aligned} \tag{6.16}
$$

6.4.3 Trap Energy : E_{tr}

The energy of the traps can vary between E_{min} and E_{max}. The trap density as a function of energy can be uniform or non-uniform depending on the technology. A non-uniform trap distribution such as an exponential distribution will lead to large area devices exhibiting gate voltage (V_g) dependent slope for FN, i.e $1/f^{\alpha(V_g)}$ [83], [84]. Due to the lack of strong experimental evidence supporting α being a function of V_g, only a uniform trap density is assumed. If future technologies generate non-uniform trap density as a function of energy, it can be incorporated into the statistical model in a fashion similar to inclusion of exponential trap density as a function of y.

Under the assumption of constant trap density, the energy of any given trap can be Monte Carlo selected through

$$E_{tr} = E_{min} + RND(0,1) \cdot (E_{max} - E_{min}) \tag{6.17}$$

The shaded region in Fig. 6.8 shows the energy bounds within which trap energies are chosen from Monte-Carlo process. However, from Fig. 6.5(b), the range of traps contributing to FN is a function of insulator depth y and is represented by the region between the blue dotted line $(E_{limit}(y))$ and the red dotted line (E_{max}). For example, at the silicon interface, traps between E_{max} and χ contribute to FN and at the gate electrode interface, traps between E_{max} and E_{min} generate FN. The trapezoid in Fig. 6.8 represents this area which corresponds to the set of traps that lie within the silicon band-gap at some operation regime of the transistor and contribute to FN. As a result any trap which has energy outside the trapezoid in Fig. 6.8 is discarded and this helps
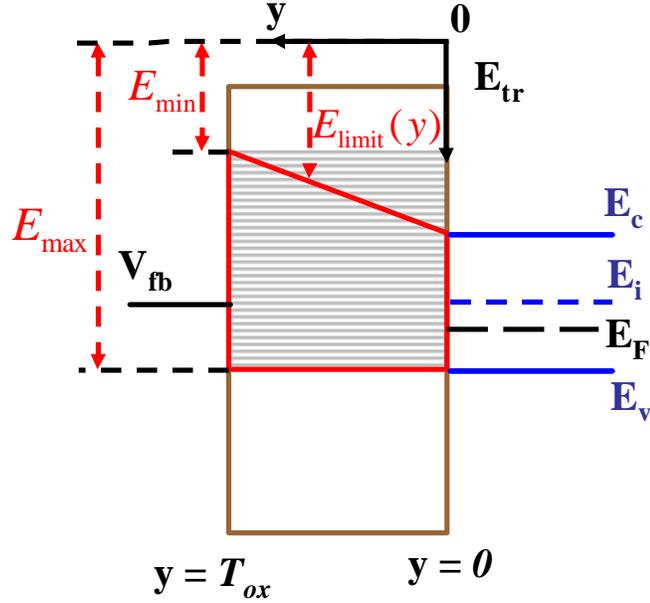
Figure 6.8: The trapezoid encloses all the traps that lie within the silicon band-gap at some operation regime of the transistor and contribute to FN. Only the traps enclosed within the trapezoid are using for calculated FN in the statistical model.

speed up the statistical model computation. This is achieved by discarding all the traps located at y which have $E_{tr} < E_{limit}(y)$ where the limit $E_{limit}(y)$ is defined as

$$E_{limit}(y) = \chi - \frac{V_{DD} - V_{fb} - 2\phi_B}{T_{ox}} \cdot y \qquad (6.18)$$

6.4.4 Percolation Factor

The percolation effect adds another level of randomness into the determination of flicker noise. Fig. 6.9 shows the simulated distribution of RTN amplitudes in W/L = 50nm/50nm FETs due to random discrete dopants [81]. The corresponding prob-
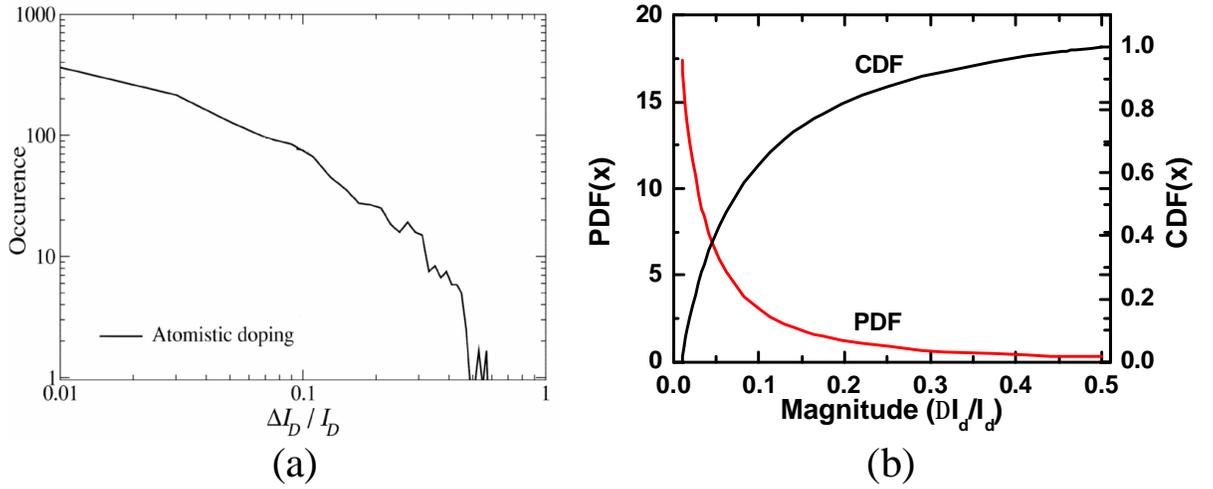
Figure 6.9: Distribution of RTN amplitudes in small area FET ($W/L = 50$nm/50nm) due to percolation effect. (a) Simulation result [81] (b) Analytical model for PDF and CDF of RTN magnitudes.

ability distribution of different RTN amplitudes can be modeled through

$$PDF(x) = \frac{0.102657}{x^2 + 0.2x + 0.0038} \tag{6.19}$$

where x here represents the RTN amplitude. The PF in the statistical flicker noise model represents the uncertainty in ΔI_d due to a single trap. As a result, the mean of the statistical variable PF should be 1. Keeping this in mind, Eq. 6.19 can be used to generate Monte Carlo estimate of PF through

$$PF = \frac{1.625 * e^{1.534RND(0,1)} - 1.17}{6.05 - e^{1.534RND(0,1)}} \tag{6.20}$$

However, in the absence of strong experimental evidence for percolation effect through the associated large variation in noise, PF is set to unity in this work ($PF = 1$).

6.4.5 Trap Profiling Algorithm

For a given device instance, the number of traps N_{tr} present in the device is first estimated using Eq. 6.12. Each trap in the device is then assigned a y location based on Eq. 6.16. Next, the trap energy E_{tr} for each trap is calculated using Eq. 6.17. Depending on the trap's y location, the trap is filtered out if E_{tr} is smaller than the limit $E_{limit}(y)$ defined in Eq. 6.18. Finally, each trap is assigned a percolation factor PF, which in this work is set to 1, completing the trap profiling process. With all the traps well defined in the device, the next step is to calculate the RTN due to the individual traps and combine them to obtain the FN for the device. The calculation of RTN and FN will be presented in the next section.

6.5 Statistical Flicker Noise Model

A trap generates RTN by trapping/de-trapping charge carriers from the inversion channel. The energy needed for carrier emission is usually higher than that needed for capture resulting in different rates for carrier capture (τ_c) and carrier emission (τ_e) [85]. The power spectral density (PSD) of the noise generated by a trap with different capture and emission times is given by [86]

$$S_{tr}(f) = \Delta I_d^2 \cdot \frac{\beta}{(1 + \beta)^2} \cdot \frac{4\tau}{(1 + 2\pi f \tau)^2} \tag{6.21}$$

where $\beta = \tau_c/\tau_e$, ΔI_d is the difference in current levels between the trapped and de-trapped states and τ is the time constant of the trap which depends on the trap's

y location through

$$\tau = \tau_0 \cdot \exp(\gamma y) \tag{6.22}$$

β, the ratio of time constants, is defined as

$$\beta = g \exp\left(\frac{E_T - E_F}{kT}\right) \tag{6.23}$$

where g is the trap degeneracy factor and $E_T - E_F$ is the trap energy level relative to the Fermi level. $E_T - E_F$ can be expressed in terms of the statistical model variables through

$$E_T - E_F = q\left(\chi - \phi_s + E_g/2 + \phi_B - \left(E_{tr} + \frac{V_g - V_{fb} - \phi_s}{T_{ox}} \cdot y\right)\right) \tag{6.24}$$

where ϕ_s is the surface potential. Eq. 6.24 tracks the trap energy relative to E_F capturing the bias-dependent noise contribution of the trap. The trap contributes maximum noise when its energy is aligned with E_F as seen from Eq. 6.21 and Eq. 6.23. When the trap moves away from E_F, it is almost always filled or empty and the noise generated by the trap decreases exponentially.

In strong inversion, the change in V_g in Eq. 6.24 will mask the negligible change in surface potential. As a result, an approximate solution of surface potential accurate only in the sub-threshold regime will suffice. A simplified expression for ϕ_s is [29]

$$\phi_s = MIN\left[2\phi_B, \left(-\frac{\eta}{2} + \sqrt{\frac{\eta^2}{4} + V_g - V_{fb}}\right)\right] \tag{6.25}$$

where η is the body factor $\frac{\sqrt{2q\epsilon_{Si}N_A}}{C_{ox}}$.

The only quantity remaining to be determined for calculating RTN in Eq. 6.21 is ΔI_d. When the trap is empty, assume that the drain current is given by

$$I_{d0} = q\mu_0 W N_{inv}(x)\frac{dV_{ch}}{dx} \tag{6.26}$$

where μ_0 is the carrier mobility, N_{inv} is the inversion charge/unit area and V_{ch} is the channel potential. When the trap is filled the number of carrier in the inversion channel decreases by one. Also, the carrier mobility changes due to scattering from the charge carrier trapped in the gate dielectric. As a result, the drain current when the trap is filled can be approximated as

$$I_d = q \cdot \left(\mu_0 \pm \frac{\delta}{WL}\mu_0^2\right) \cdot W \cdot \left(N_{inv}(x) - \frac{1}{WL}\right) \cdot \frac{dV_{ch}}{dx} \tag{6.27}$$

where δ is mobility scattering coefficient [78]. ΔI_d can be easily calculated from the difference of Eq. 6.26 and Eq. 6.27.

$$\Delta I_d = \frac{I_d}{WL}\left(\frac{1}{N_{inv}} \pm \mu_0\alpha\right) \tag{6.28}$$

The RTN of a single trap i can finally be written as

$$RTN(f) = \frac{I_d^2}{W^2L^2}\left(\frac{1}{N_{inv}} \pm \mu_0\alpha\right)^2 \cdot PF_i\frac{\beta_i}{(1+\beta_i)^2} \cdot \frac{4\tau_i}{(1+2\pi f\tau_i)^2} \tag{6.29}$$

where PF_i, τ_i and β_i are the percolation factor, time constant and time constant ratio

for the trap i respectively. The total flicker noise of the device is simply the sum of all the RTNs and can be written as

$$S_{I_d}(f) = \frac{I_d^2}{W^2 L^2} \left(\frac{1}{N_{inv}} \pm \mu_0 \alpha \right)^2 \cdot \sum_{i=1}^{N_{tr}} PF_i \frac{\beta_i}{(1 + \beta_i)^2} \cdot \frac{4\tau_i}{(1 + 2\pi f \tau_i)^2} \tag{6.30}$$

It is interesting to note that the statistical model when applied to a large area device with uniform trap density (a=1) yields the expected $1/f$ shape for large area FN. This can be clearly seen from the following important equality when N_{tr} is large.

$$\sum_{i=1}^{N_{tr}} \frac{\beta_i}{(1 + \beta_i)^2} \cdot \frac{4\tau_i}{(1 + 2\pi f \tau_i)^2} = N_T \cdot W \cdot L \cdot \frac{kT}{q} \frac{1}{\gamma f} \tag{6.31}$$

6.6 Statistical Flicker Noise Model Implementation

The statistical flicker noise model (SFN) is implemented in BSIM4 with the goal of capturing the FN in small area devices. Currently, FN is described in BSIM4 by the unified flicker noise model (UFN) which has been successfully used by the industry for several years. The UFN captures the FN in large area devices in both strong inversion and weak inversion regime accurately [78]. The UFN parameters $NOIA$, $NOIB$ and $NOIC$ capture the bias dependence while the parameter EF captures the frequency dependence of FN ($FN \propto 1/f^{EF}$). The unified flicker noise model is implemented in BSIM4 assuming a fixed γ of 1e10s^{-1}. The parameter EF is equivalent to α in Eq. 6.13 and it determines the spatial trap density distribution coefficient a in the statistical flicker noise model. The UFN parameter $NOIA$ is related to the average trap density through N_T/q.

The statistical model SFN is implemented in such a way that the mean of SFN ($\langle SFN \rangle$) is equal to UFN for any transistor under any operating condition. In this way, the bias dependence of FN in large area devices modeled by UFN is retained in SFN. Furthermore, the model implementation also allows the parameters extracted for UFN to be used for evaluating SFN. To satisfy these two goals of model implementation, define SFN as

$$SFN = S \cdot \frac{UFN}{\frac{kT}{1e10 \cdot f^{EF}} \cdot NOIA \cdot WL} \cdot \sum_{i=1}^{N_{tr}} \frac{\beta_i}{(1+\beta_i)^2} \cdot \frac{4\tau_i}{(1+2\pi f \tau_i)^2} \tag{6.32}$$

where S is a scaling factor. Clearly, all the model parameters of UFN are used in SFN. To attain the goal of $\langle SFN \rangle = UFN$, S can be calculated through

$$S = \frac{\frac{kT/q}{1e10 \cdot f^{EF}} \cdot N_T \cdot WL}{\sum_{i=1}^{N_{tr}} \frac{\beta_i}{(1+\beta_i)^2} \cdot \frac{4\tau_i}{(1+2\pi f \tau_i)^2}} \tag{6.33}$$

for a large area device.

When $EF = 1$ (i.e $a = 0$), using Eq. 6.31 S is simply given by

$$S = \frac{\gamma}{1e10} \tag{6.34}$$

When $EF \neq 1$ an exponential trap density exists in the dielectric $N(y)$. Define $N(y) = N_{T0}e^{ay}$ where N_{T0} represents surface trap density and $a = \gamma(EF - 1)$. N_{T0} is related to the average trap density N_T through

$$N_{T0} = N_T \cdot \frac{aT_{ox}}{\exp(aT_{ox}) - 1} \tag{6.35}$$

175
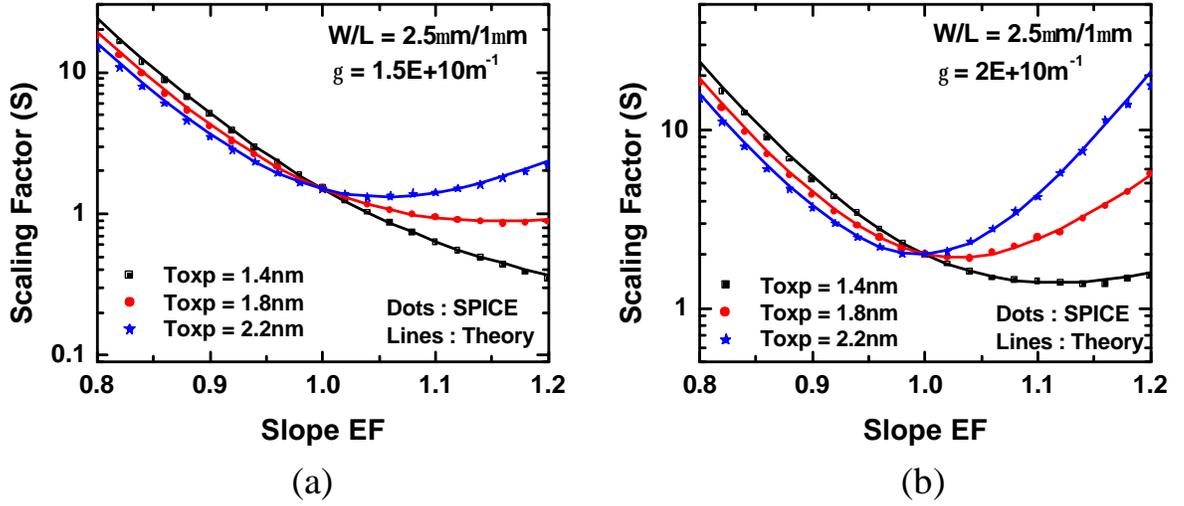
Figure 6.10: Scaling factor S for SFN as a function of EF and T_{ox} for (a) $\gamma = 1.5e10\text{m}^{-1}$ (b) $\gamma = 2.0e10\text{m}^{-1}$. The analytical model agrees very well with SPICE values.

In the limit of a large area device, Eq. 6.33 can be written as

$$S = \frac{N_T}{N_{T0}} \cdot \frac{1}{1e10 \cdot f^{EF}} \cdot \frac{1}{\int_0^{T_{ox}} \frac{4e^{ay} \cdot \tau_0 \, e^{\gamma y}}{1 + (2\pi f \, \tau_0 e^{\gamma y})^2} \, dy} \tag{6.36}$$

For the case of $\tau_{max} = \tau_0 e^{\gamma T_{ox}} > 1$, Eq. 6.36 can be approximated by

$$S \approx \frac{\gamma}{1e10} \cdot \frac{e^{aT_{ox}} - 1}{aT_{ox}} \cdot e^{-21.21 \cdot (EF-1)} \tag{6.37}$$

Fig. 6.10 shows the S calculated from the analytical model in Eq. 6.37 for different T_{ox} and different γ as a function of EF. The analytical values of S agree very well with the values obtained from SPICE. S is obtained from SPICE by calculating $UFN/\langle SFN(S=1)\rangle$.

Eq. 6.32 constitutes the statistical flicker noise model implemented in BSIM4
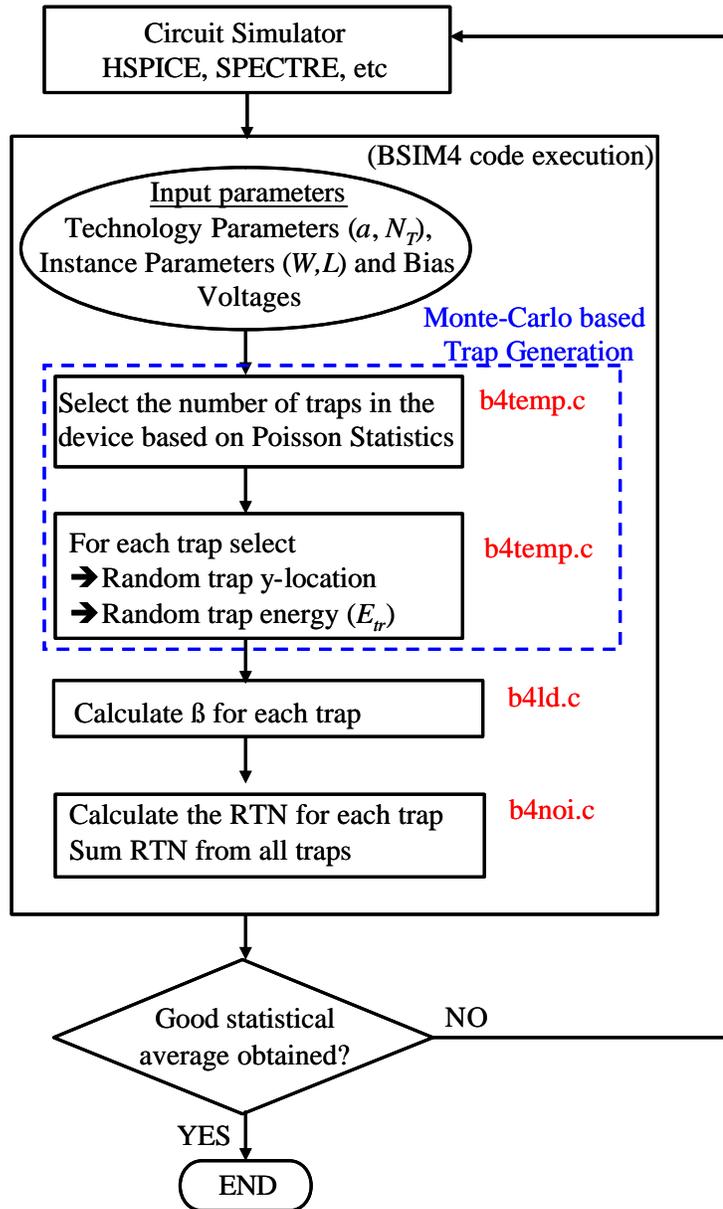
Figure 6.11: Execution flow for the statistical flicker noise model in BSIM4 framework.

model. Fig. 6.11 summarizes the algorithm for calculating flicker noise using the statistical noise model. A designer simulates the desired circuit using any popular circuit simulator such as HSPICE, SPECTRE, etc. which internally calls the BSIM4 model. The SFN is activated by setting the BSIM4 model parameter $FNOIMOD$ = 2. The technology parameters for calculating SFN such as a and N_T are obtained from BSIM4 model parameters EF and $NOIA$ which are extracted from the noise spectrum of large area device. Based on the technology parameters and a user-defined random number generator seed, Monte-Carlo based trap profiling is performed which comprises of estimating the number of traps in the device and assigning a y-location and an energy level E_{tr} for each trap. Next, the time constant ratio β is calculated for each trap as a function of bias voltages. Finally, the RTN for each trap is calculated and added together to give the final bias-dependent FN in the device. Based on the FN of the individual devices in the circuit, the circuit simulator calculates the noise characteristics of the circuit. The circuit can be simulated several times using different random number generator seeds to obtain a good statistical average.

The behavior of SFN model is verified against experimental data in Fig. 6.12. Using the large area $W/L = 10/0.28$ μm device noise spectrum, technology parameters are extracted and used to simulate FN characteristics. Fig. 6.12(a) shows the measured and simulated FN for 10 different device samples. As expected, the model and measurements show small device-to-device variation in large area devices. In Fig. 6.12(b), using the same technology parameters, smaller area devices with W/L = $0.5/0.28\mu$m are simulated. When the area is scaled down, the average number of traps in the dielectric decreases and the variation in FN increases. This is seen from
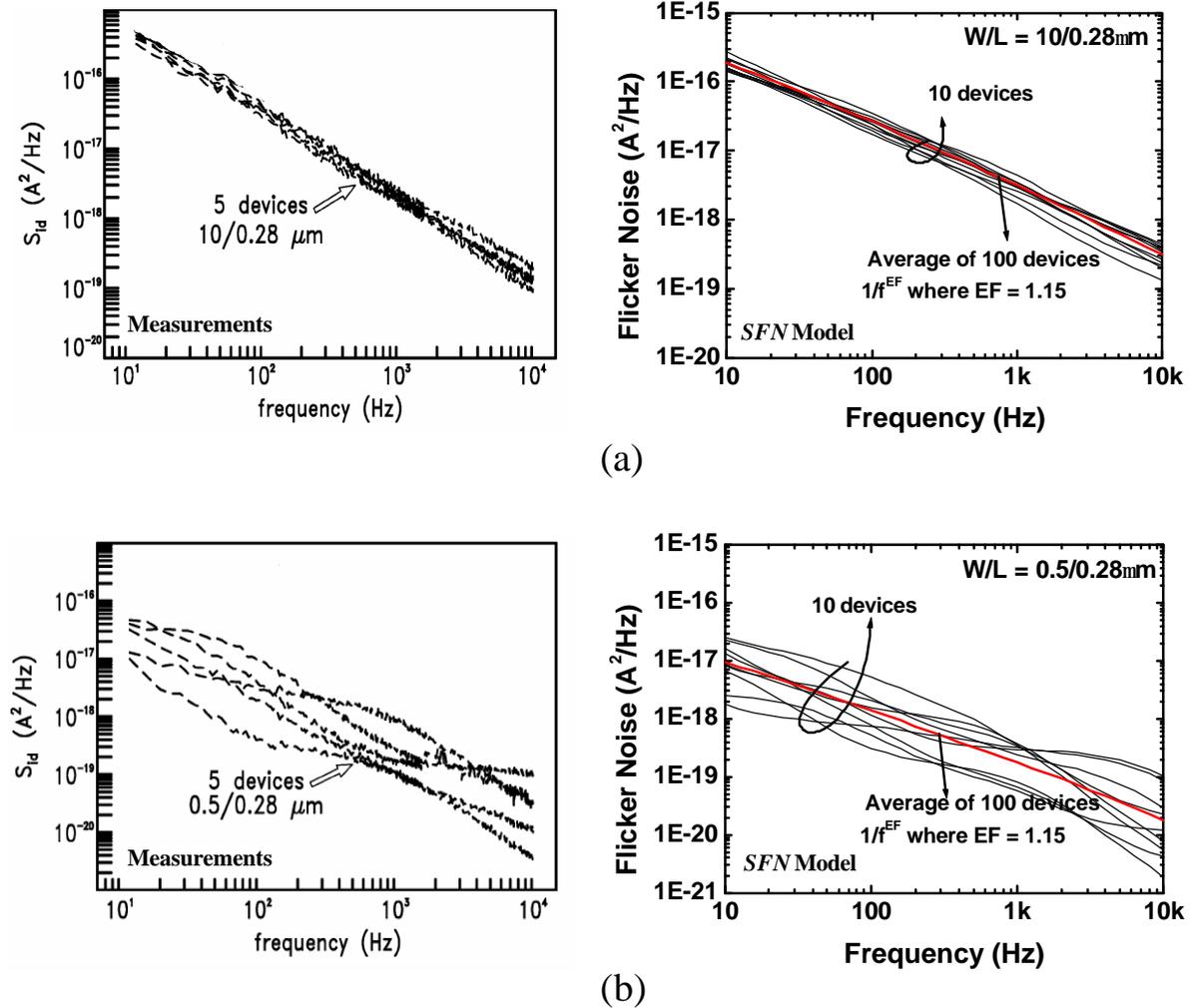
Figure 6.12: Flicker noise in (a) Large area devices and (b) Small area devices. SFN model parameters are extracted from the large area noise measurements and used to simulate FN for both large and small area devices. The model correctly predicts the increased variation in FN in small area devices.

both the measurements and the statistical model. Going to even smaller devices gives rise to the interesting possibility of finding some devices with zero traps and hence exhibiting no flicker noise.

6.7 Model Statistics

Circuit designers can garner useful statistical noise information by running multiple Monte Carlo simulations. However, as the area of the transistor decreases, to have a good estimate of mean $\langle SFN \rangle$ and standard deviation σ_{FN}, the number of Monte Carlo simulations increases dramatically. Analytical models for the statistics will prove to be very useful in getting quick estimates of noise variation. In this section, analytical models for the mean and standard deviation of FN will be developed using SFN formulation.

6.7.1 Mean : $\langle SFN \rangle$

SFN was developed in Section 6.6 with the goal $\langle SFN \rangle = UFN$. An underlying assumption in this equality is the requirement of $\tau_{max} = \tau_0 e^{\gamma T_{ox}} > 1s$. If the dielectric thickness T_{ox} or the tunneling coefficient γ is small, the largest time constant may become smaller than 1s i.e. $\tau_{max} < 1s$. Under such conditions, the scaling factors in Eq. 6.34 and Eq. 6.37 are no longer accurate and $\langle SFN \rangle \neq UFN$. This is illustrated in Fig. 6.13 which shows the $\langle SFN \rangle$ obtained from Monte-Carlo simulations for different γ. Due to the absence of traps with time constants larger than τ_{max}, $\langle SFN \rangle$ becomes nearly constant for frequencies less than $(2\pi\tau_{max})^{-1}$. This is also evident
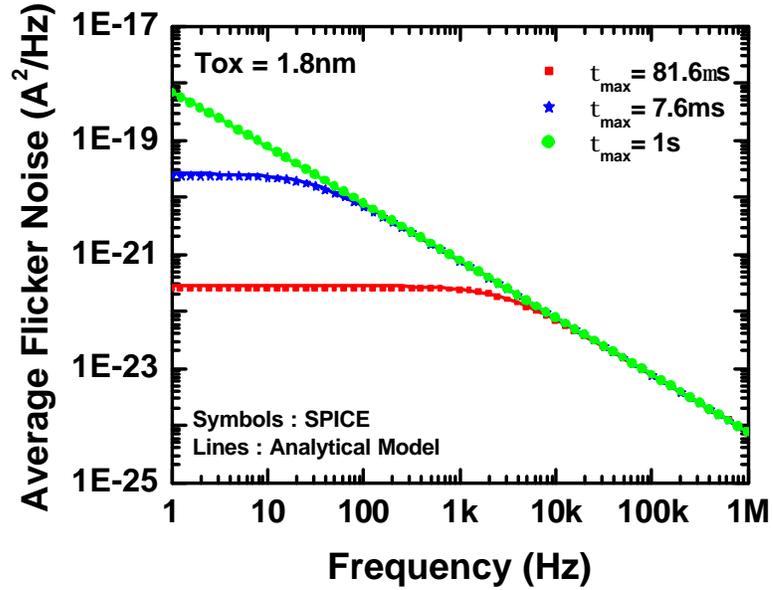
Figure 6.13: Average flicker noise for devices with different γ i.e. different τ_{max}. $\langle SFN \rangle$ remains constant for $f < 1/(2\pi\tau_{max})$. $\langle SFN \rangle$ from SPICE is generated by averaging 2000 Monte-Carlo runs. Good agreement is seen between the analytical model and SPICE simulated values.

from Eq. 6.32. $\langle SFN \rangle$ can be analytically modeled by modifying UFN as shown below.

$$\langle SFN \rangle = UFN \cdot \left(\frac{f}{f_{\langle SFN \rangle}} \right)^{EF} \quad \text{where} \quad f_{\langle SFN \rangle} = \left(f^{3/2} + \left(\frac{\sqrt{2}}{2\pi\tau_{\max}} \right)^{3/2} \right)^{2/3} \quad (6.38)$$

$\langle SFN \rangle$ calculated using the analytical model in Eq. 6.38 is shown in Fig. 6.13. Good agreement is observed between the analytical model and the mean obtained from Monte-Carlo simulations in all the cases.
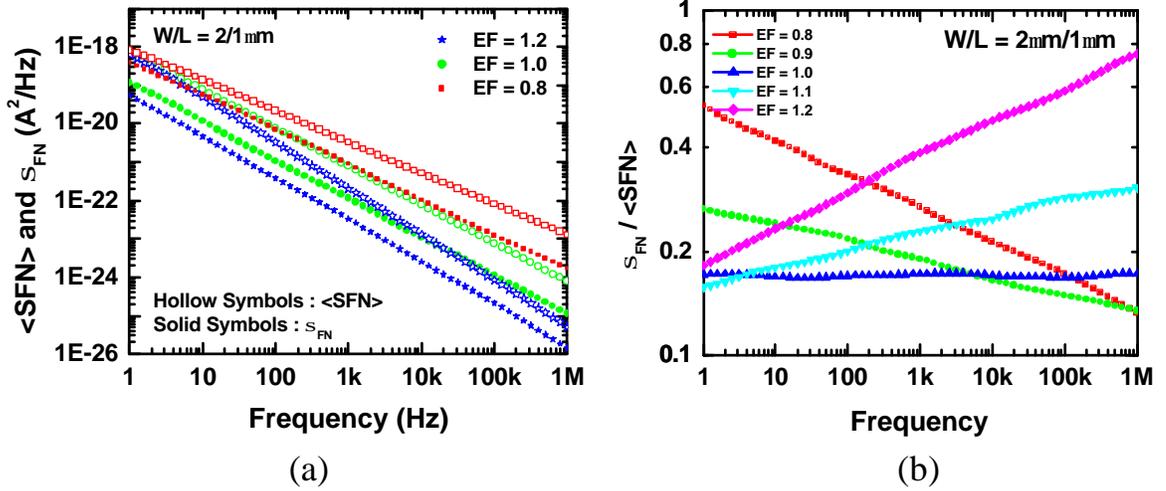
Figure 6.14: (a)$\langle SFN \rangle$ and σ_{FN} for a large area device for different EF. (b) Normalized standard deviation $\sigma_n(f)$

6.7.2 Standard Deviation : σ_{FN}

A detailed analytical study of σ_{FN} is nearly impossible due to the presence of several statistical variables. Furthermore, since the SFN is not a linear function of the random variables, propagation of variances is very hard to use. As a result, physics-based arguments together with Monte-Carlo simulations are used to determine σ_{FN}.

Fig. 6.14(a) shows the simulated σ_{FN} for a large area device for different EF. It is interesting to observe that the ratio $\sigma_{FN}/\langle SFN \rangle$ ($= \sigma_n(f)$) changes with frequency for non-unity EF. Fig. 6.14(b) plots explicitly the $\sigma_n(f)$ for different EF as a function of frequency. The magnitude of $\sigma_n(f)$ at $f = 1$Hz is also a function of EF. When the area of the device decreases, the average number of traps in the device decreases and σ_{FN} increases. This is illustrated in Fig. 6.15 which shows the $\sigma_n(f)$ for different area devices ranging from a large area device of $WL = 2\mu$m^2 to a small area device of WL
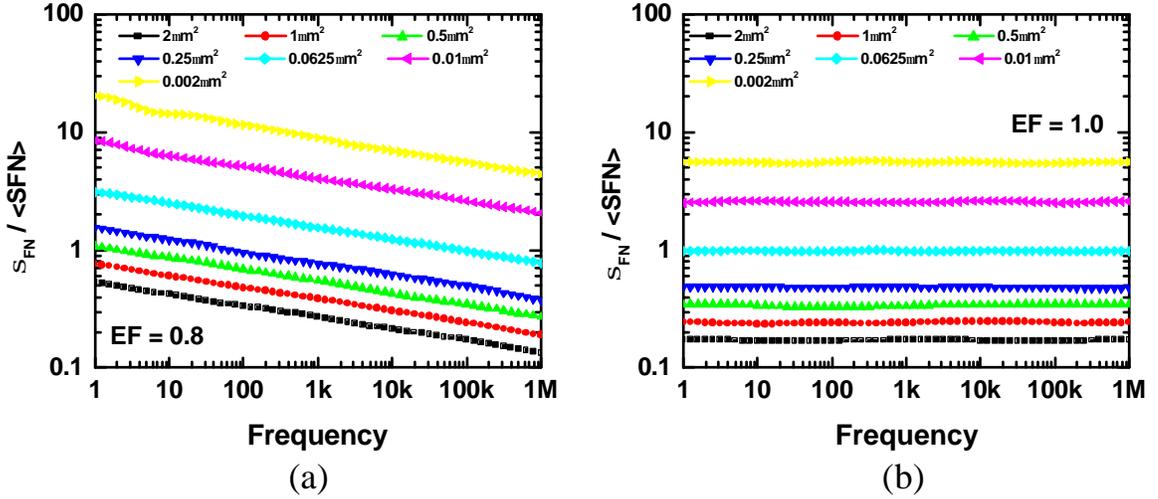
Figure 6.15: Normalized standard deviation $\sigma_n(f)$ for different area devices (a) $EF = 0.8$ (b) $EF = 1.0$. As the area decreases, the variation in flicker noise increases.

$= 0.002\mu\mathrm{m}^2$. Fig. 6.15 shows that the $\sigma_n(f)$ increases with decreasing area and the increase is independent of EF. This is because of the fact that the average number of traps in a device simply scales by $1/WL$ independent of EF. This indicates that the modeling of WL and EF is orthogonal. With these insights, the normalized σ_{FN} can be modeled through

$$\sigma_n(f) = K(EF) \cdot \sigma_{n0} \cdot f^\eta \tag{6.39}$$

where the factor σ_{n0} represents $\sigma_n(f)$ for a uniform trap density distribution ($EF = 1$) and the factors η and $K(EF)$ model the impact of exponential trap density distribution ($EF \neq 1$) on σ_{FN}.

The standard deviation will be first determined for the case of uniform trap density distribution. σ_{n0} is expected to depend on the average number of traps in the device
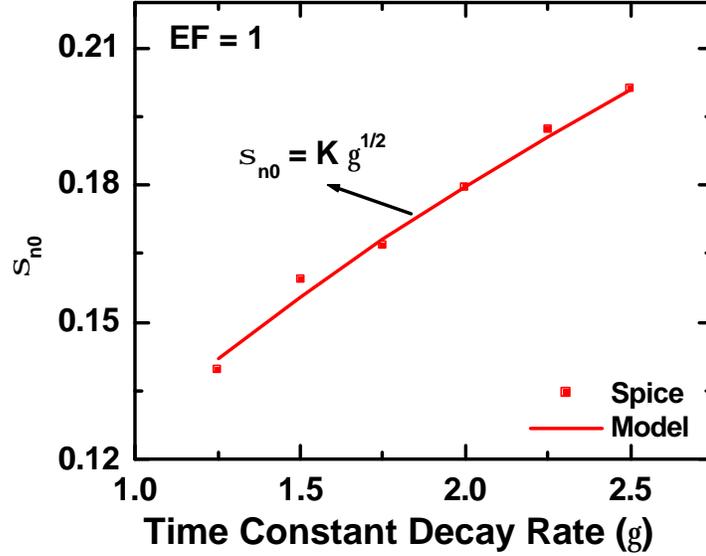
Figure 6.16: Normalized standard deviation $\sigma_n(f)$ as a function of γ. $\sigma_n(f)$ increases with increasing γ.

and the time constant spread in each frequency decade. In the case of uniform trap density distribution (a=0), the time constants are spread exponentially at a rate of γ and hence the σ_{n0} is expected to increase with increasing γ. Fig. 6.16 shows that the γ dependence can be captured through

$$\sigma_{n0} \propto \sqrt{(\gamma)} \tag{6.40}$$

The average number of traps in a given device area is proportional to $\langle N \rangle > N_T \cdot WL$. Since the variation in noise is inversely proportional to $\sqrt{\langle N \rangle}$,

$$\sigma_{n0} \propto \sqrt{\frac{\gamma}{N_T WL}} \tag{6.41}$$

184
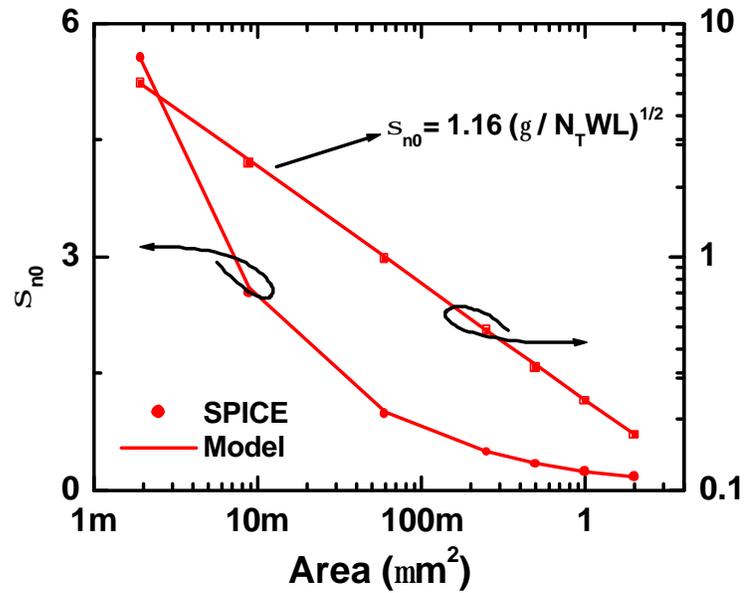
Figure 6.17: Normalized standard deviation $\sigma_n(f)$ as a function of device area WL. $\sigma_n(f)$ decreases with increasing WL.

The proportionality constant is determined to be 1.16 upon comparing the model against simulations. Fig. 6.17 shows the variation of σ_{n0} as a function of device area illustrating the $1/\sqrt{WL}$ dependence. Fig. 6.18 verifies the dependence of σ_{n0} on the trap density N_T.

For an exponential trap density distribution, the number of traps giving rise to low frequency and high frequency noise are different. Consider a technology exhibiting $EF > 1$ which corresponds to more traps away from the interface $(a = \gamma \cdot (EF - 1))$. As a result the number of traps contributing to low frequency noise is large and hence the variation in noise at low frequencies is small compared to the variation at high frequencies. Since the σ_{FN} is inversely proportional to $\sqrt{\langle N \rangle}$, for an exponential trap
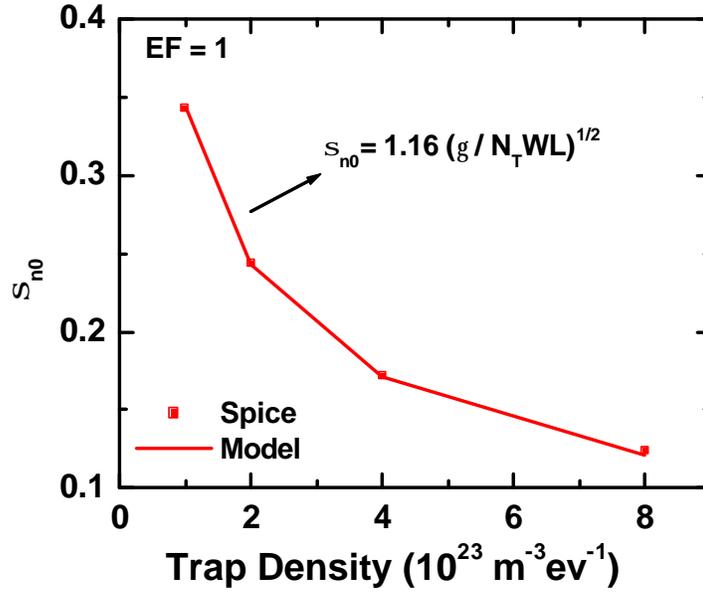
Figure 6.18: Normalized standard deviation $\sigma_n(f)$ as a function of trap density N_T. $\sigma_n(f)$ decreases with increasing N_T.

density distribution, σ_{FN} is expected to demonstrate

$$\sigma_{FN}(f) \propto \frac{1}{f^{\frac{1+EF}{2}}} \quad (6.42)$$

Since $\langle SFN \rangle \propto 1/f^{EF}$, the frequency dependence of $\sigma_n(f)$ can be captured by

$$\sigma_n(f) \propto f^{\frac{EF-1}{2}} \quad (6.43)$$

Fig. 6.19 shows that the frequency dependence model agrees very well with the simulated statistical trends for different EF. This however implicitly assumes that the maximum time constant $\tau_{max} > 1s$. This assumption will be revisited towards
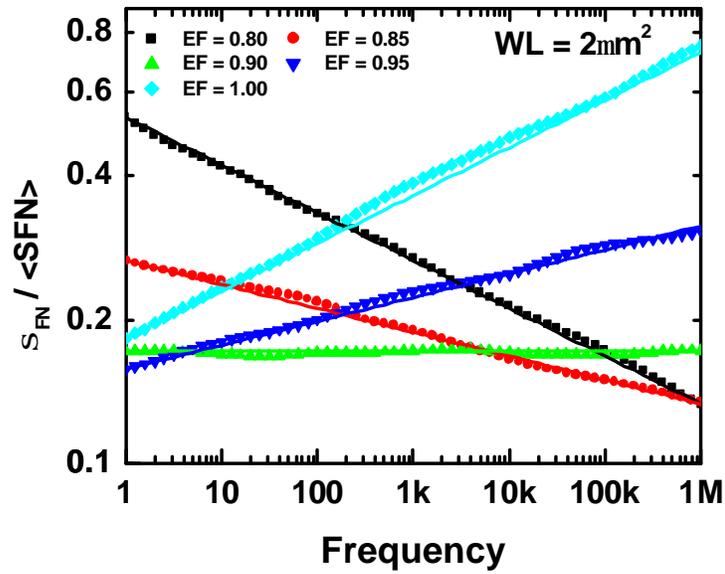
Figure 6.19: Normalized standard deviation $\sigma_n(f)$ as a function EF.
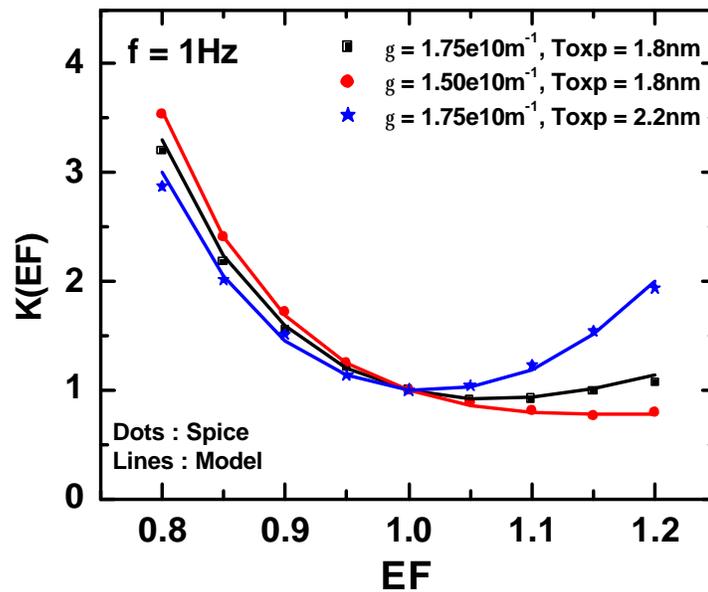
Figure 6.20: K(EF) for different γ and T_{ox}.

the end of this section. Furthermore, the change in magnitude of $\sigma_n(f)$ at 1Hz for an exponential trap density distribution can be modeled through

$$K(EF) = \sqrt{\frac{e^{aT_{ox}} - 1}{aT_{ox}}} \cdot e^{-21.21*(EF-1)} \qquad (6.44)$$

Fig. 6.20 shows the values of $K(EF)$ obtained from SPICE simulations for different combinations of γ and T_{ox}. In all the cases, the model prediction agrees very will the simulated trends. Using the analytical models for $K(EF)$, σ_{n0} and η, the standard deviation in SFN can be written as

$$\sigma_{FN}(f) = UFN \cdot \sqrt{\frac{e^{aT_{ox}} - 1}{aT_{ox}}} \cdot e^{-21.21*(EF-1)} \cdot 1.16 \cdot \sqrt{\frac{\gamma}{N_T}} \cdot \frac{1}{\sqrt{WL}} \cdot f^{\frac{EF-1}{2}} \quad (6.45)$$

As mentioned earlier, the frequency dependence modeling assumed $\tau_{max} > 1$s. When $\tau_{max} < 1$s, there are no traps contributing generating noise at frequencies below $(2\pi\tau_{max})^{-1}$Hz and hence there would be no change in standard deviation of noise in frequencies below $(2\pi\tau_{max})^{-1}$Hz. In order to capture this effect, UFN in Eq. 6.45 is replaced with $\langle SFN \rangle$ and the frequency dependence is limited to the maximum trap time constant. The final standard deviation model for SFN can be then written as

$$\sigma_{FN}(f) = \langle SFN \rangle \cdot \sqrt{\frac{e^{aT_{ox}} - 1}{aT_{ox}}} \cdot e^{-21.21*(EF-1)} \cdot 1.16 \cdot \sqrt{\frac{\gamma}{N_T}} \cdot \frac{1}{\sqrt{WL}} \cdot f_{eff}^{\frac{EF-1}{2}} \quad (6.46)$$

where f_{eff} is defined as

$$f_{eff} = \left(f^3 + \left(\frac{1}{2\pi\tau_{max}} \right)^3 \right)^{1/3} \qquad (6.47)$$
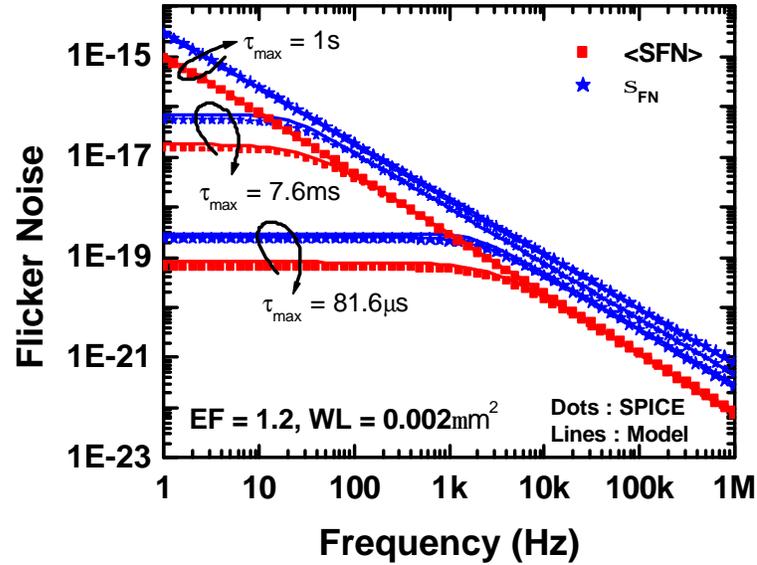
Figure 6.21: $\langle SFN \rangle$ and σ_{FN} for small area device with $WL = 0.002 \mu m^2$ and $EF = 1.2$ but varying γ. Analytical models for the statistics of FN agree very well with SPICE simulated values.

Fig. 6.21 shows the standard deviation and the mean of FN predicted by the analytical model for a small area device with $EF = 1.2$ for two different γ. The analytical model results are compared against the statistics obtained from 15000 Monte Carlo SPICE simulations. When γ is small, τ_{max} is small and FN at low frequencies becomes independent of frequency. Good agreement is seen between the analytical model and SPICE simulations suggesting that the analytical expressions for $\langle SFN \rangle$ and σ_{FN} can be used to quickly gauge the noise variation in small area devices without resorting to tedious Monte Carlo simulation process.

6.8 Conclusion

A novel statistical compact model for flicker noise in scaled transistor is developed. The model captures a large number of physical effects to accurately model both the geometry and voltage bias dependence of flicker noise. The model is yet easy to use with all the technology parameters extractable from the large device flicker noise spectrum. The model has been implemented in BSIM4 and is the first instance of modeling any kind of variability within BSIM4 framework. In addition, analytical models for statistics of FN in small area devices - mean and standard deviation - have been developed. The statistical flicker noise model can be used for Monte Carlo noise simulation as well as a guide for analog designers.

Chapter 7

Conclusions

7.1 Summary of Contributions

The single-gate planar bulk silicon MOSFET has served the industry very well for the last several decades. With the conventional scaling techniques approaching fundamental physical limits, new CMOS architectures are being investigated. At the same time, new materials are being incorporated into the conventional planar MOSFET to extend it beyond 45nm technology node. Scaled MOSFETs are exhibiting new device behaviors which were previously insignificant or absent.

Compact MOSFET models describe the physics and operation of a MOSFET in a compact mathematical way to enable fast circuit simulation. A compact model describing a new technology is almost always developed prior to the adoption of that technology by the semiconductor industry. The compact models are useful not only for long term product designs but also for early evaluation of a technology for circuit

applications.

In this dissertation, new compact models were developed for modeling alternative CMOS device structures and for modeling new materials and new behavior observed in state-of-the-art bulk planar MOSFETs with a goal to enable timely evaluation of the new nanoscale CMOS technologies and their circuit-level benefits and enable efficient product designs in the future.

7.1.1 Multi-Gate FET Modeling

Multi-Gate FETs (MG-FETs) are the most promising candidates for CMOS scaling beyond 22nm technology node. The multiple gate electrodes in a MG-FET can be electrically interconnected as in Common Multi-Gate FETs or they can be biased independently as in Independent Multi-Gate FETs. Compact models were developed separately for both CMG-FETs and IMG-FETs in this work.

A surface-potential based compact model for CMG-FETs was developed. A new unified surface-potential model was first developed which calculates the surface potential in a CMG-FET in the presence of finite body doping unlike some of the other CMG-FET compact models. The surface potential model captures the effect of structural carrier confinement and poly-depletion. A new drain current model was formulated for the long channel DG-FET and extended to short channel MG-FETs through electrostatic modeling of short channel effects such as threshold voltage roll-off and DIBL. The capacitances for the CMG-FET were also derived as a function of surface-potentials at the source and drain end. The unified surface-potential model together with the I-V and C-V model developed in this dissertation forms the core

of BSIM-CMG model. Through incorporation of additional real device effects into the core model, BSIM-CMG can describe any modern CMG-FET technology. The core model's accuracy and scalability was verified extensively using TCAD simulations. BSIM-CMG was also shown to successfully describe two different experimental FinFET technologies : SOI FinFETs and bulk FinFETs.

Surface-potential based compact model for IMG-FETs was also developed. The framework for calculating the surface potential in a IMG-FET was first outlined. Next, a new charge-sheet based I-V model was developed for IMG-FET which can accurately model the drain current under all bias conditions. A C-V model describing the terminal charges and capacitances of IMG-FET was also presented. The C-V model however is valid when only one surface is allowed to conduct current. This is useful in applications where the back gate is biased to tune to front channel threshold voltage without forcing any inversion at the back surface. The I-V and C-V model developed in this work constitute the core of BSIM-IMG model.

7.1.2 Advanced planar bulk Si MOSFET Modeling

This dissertations also improves the existing compact models for the planar bulk silicon MOSFETs. State-of-the-art MOSFETs use process-induced strain to boost the carrier mobilities and extend the scaling limits. A new non-process-specific layout-dependent mobility model for mobility enhancement through process induced strain was developed. The model captures the mobility change due to process-induced strain as a function of channel length and width. The model was verified extensively using 3-D process simulations and experimental data from published literature. The model

can be easily incorporated into any compact MOSFET model.

In small area devices, random variability is exacerbated with transistors showing large variability in different electrical parameters such as drain current and low frequency noise. A new statistical model was developed for modeling the low frequency noise in small area FETs due to charge trapping/de-trapping from the traps in the gate dielectric. The statistical model predicts same behavior as the unified flicker noise model for large area devices. The model has been incorporated into BSIM4 and is the first ever instance of incorporating variability into the BSIM framework. This approach can also be extended to model other sources of variability in transistor's performance.

7.2 Future Research Directions

Advances in nanoscale CMOS technology to solve scaling issues opens up a plethora of venues for exciting research in field of compact modeling. The compact models developed in this dissertation are simply the start of an exciting era in the field of modeling.

7.2.1 Multi-Gate FET Models

The MG-FET models developed in this dissertation can be further enhanced in several ways. Resistive and capacitive parasitics play an increasingly important role in determining the performance of scaled MG-FETs. Inclusion of layout-dependent models of device parasitics will significantly strengthen the accuracy of the model.

Real device effects such as mobility degradation, gate tunneling current and GIDL may be different between conventional planar bulk MOSFET and MG-FETs. Currently, many of these real device effects are modeled along the lines of bulk MOSFET model. It is judicious to perform a detailed study of the real device effects with respect to MG-FETs and identify any differences from bulk MOSFET.

The surface potential model and the C-V model for IMG-FETs assume single surface conduction. This restricts the versatility of the BSIM-IMG model to a large degree. The development of a surface potential solution technique and a C-V model for IMG-FETs which are valid even under both channel conduction is very important and perhaps one of the most important challenges in MG-FET modeling in the future.

The use of process-induced strain to enhance the carrier mobilities in MG-FETs is inevitable. This calls for a layout-dependent mobility enhancement model for MG-FETs, similar to the one developed for bulk FETs in this dissertation.

7.2.2   Advanced planar bulk Si MOSFET Models

The continuing advancements in bulk CMOS technology demand constant improvements in bulk FET compact models. This dissertation addressed modeling of mobility change due to process-induced strain. Modeling of high-k/metal gate is also needed to describe advanced CMOS technologies. Accordingly, the new BSIM4 model - BSIM4.6.1, captures many of the effects of high-k/metal gate on CMOS performance. Further improvements may still be needed in terms of gate current modeling and flicker noise modeling in presence of high-k gate dielectric.

The variability in low frequency noise is just one of the several manifestations of

random variability in scaled CMOS. Another important random variability is random dopant fluctuations (RDF) which can cause large deviations in threshold voltages in scaled transistors critically effecting the functionality of circuits. Modeling of RDF is another venue of exciting research for bulk FET models.

Variability in CMOS performance can also occur in the form of systematic variability. As the transistors come closer in scaled CMOS, layout-dependent effects become important. Modeling of layout-dependent effects is very critical though it may be prove to be overbearing in an academic environment. A collective effort between academia and industry is needed in this area.

Development of a framework bringing both random and systematic under the same umbrella is one of the most exciting research opportunities in bulk FET models. This also entails redefining the existing corner simulation approach to capture the variability in a process more efficiently. Corner simulations usually give overly pessimistic or optimistic performance prediction due to insufficient attention to the correlations between variations and to electrical test variation data and yield conservative margins in many design instances.

The future research directions mentioned for bulk FETs are valid for MG-FETs also. For example, even the MG-FETs are expected to be plagued by variability. The lower body doping in MG-FETs reduces RDF but the variation in body thickness now adds a new source of variability.

7.3 Conclusions

The economics behind the CMOS scaling is strong enough to keep it alive for several more generations to come. Multi-Gate transistors present an interesting alternative to extend the CMOS scaling. But at the same time, the conventional planar bulk silicon MOSFET continues to shrink aided by incorporation of new materials into the conventional MOSFET structure, be it the gate-dielectric or the gate electrode or the source/drain region. New compact models will continue to emerge alongside new FET architectures to enable evaluation of these new architectures. At the same time, the existing models for bulk FETs will continue to incorporate new physics to describe advanced bulk FETs. Nanoscale CMOS is an endless exciting road for both technology and model developers.

Bibliography

[1] D. L. Critchlow, "Mosfet scaling - the Driver of VLSI Technology," *Proc. IEEE*, vol. 87, pp. 659–667, 1999.

[2] J. S. Kilby, "Invention of the Integrated Circuit," *IEEE Trans. Electron Devices*, vol. 23, pp. 648–654, July 1976.

[3] ——, "Turning Potential into Realities: The Invention of the Integrated Circuit (Noble Lecture)," *ChemPhysChem*, vol. 2, pp. 482–489, 2001.

[4] D. Kahng and M. M. Atalla, "Silicon-silicon dioxide field induced surface devices," in *IEE-AIEE Solid-State Device Res. Conf.*, 1960.

[5] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, pp. 114–117, 1965.

[6] ——, "Progress in digital integrated electronics," in *IEDM Tech. Dig.*, 1975, pp. 11–13.

[7] R. H. Dennard, F. H. Gaensslen, L. Kuhn, and H. N. Yu, "Design of micron MOS switching devices," in *IEDM Tech. Dig.*, 1972.

[8] R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, and A. L. Blanc, "Design of ion-implanted MOSFET's with very small physical dimenstions," *IEEE J. Solid-State Circuits*, vol. SC-9, pp. 256–268, 1974.

[9] Y. Taur, D. A. Buchanan, W. Chen, D. J. Frank, K. E. Ismail, S. H. Lo, G. A. Sai-Halasz, R. G. Viswanathan, H. J. C. Wann, S. J. Wind, and H.-S. Wong, "CMOS Scaling into the Nanometer Regime," *Proc. IEEE*, vol. 85, pp. 486–504, 1997.

[10] D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, , and H. S. P. Wong, "Device Scaling Limits of Si MOSFETs and Their Application Dependencies," *Proc. IEEE*, vol. 89, pp. 259–288, 2001.

[11] A. Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 μm MOSFET's: A 3-D "atomistic" simulation study," *IEEE Trans. Electron Devices*, vol. 45, pp. 2505–2513, 1998.

[12] J. C. Lee, H. J. Cho, C. S. Kang, S. Rhee, Y. H. Kim, R. Choi, C. Y. Yang, C. Choi, and M. Abkar, "High-K dielectrics and MOSFET characteristics," in *IEDM Tech. Dig.*, 2003, pp. 95–98.

[13] R. Chau, S. Datta, M. Doczy, B. Doyle, J. Kavalieros, and M. Metz, "High-k/Metal-gate stack and its MOSFET characteristics," *IEEE Electron Device Lett.*, vol. 25, no. 6, pp. 408–410, 2004.

[14] T. Ghani, M. Armstrong, C. Auth, M. Bost, P. Charvat, G. Glass, T. Hoffmann, K. Johnson, C. Kenyon, J. Klaus, B. McIntyre, K. Mistry, A. Murthy, J. Sandford, M. Silberstein, S. Sivakumar, P. Smith, K. Zawadzki, S. Thompson, , and M. Bohr, "A 90 nm high volume manufacturing logic technology featuring novel 45 nm gate length strained silicon CMOS transistors," in *IEDM Tech. Dig.*, 2003, pp. 978–980.

[15] by K. Bernstein, D. J. Frank, A. E. Gattiker, W. Haensch, B. L. Ji, S. R. Nassif, E. J. Nowak, D. J. Pearson, and N. J. Rohrer, "High-performance CMOS variability in the 65-nm regime and beyond," *IBM J. Res. and Dev.*, vol. 50, no. 4/5, pp. 433–450, 2006.

[16] D. Hisamoto, W.-C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kua, E. Anderson, T.-J. King, J. Bokor, and C. Hu, "FinFET - A Self-Aligned Double Gate MOSFET scalable to 20nm," *IEEE Trans. Electron Devices*, vol. 47, no. 12, pp. 2320–2325, 2000.

[17] H.-S. P. Wong, D. J. Frank, and P. M. Solomon, "Device Design Considerations for Double-Gate, Ground Plane and Single-Gate Ultra-Thin SOI MOSFETs at the 25nm Channel Length Consideration," in *IEDM Tech. Dig.*, 1998, pp. 407–410.

[18] K. Suzuki, T. Tanaka, Y. Tosaka, H. Horie, and Y. Arimoto, "Scaling theory for double-gate SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 40, no. 12, pp. 2326–2329, 1993.

[19] L. Chang, "Nanoscale Thin-Body CMOS Devices," Ph.D. dissertation, University of California, Berkeley, Department of Elec. Engg. and Computer Sciences, 2003.

[20] (2007, May) BSIM4.6.1 Manual. [Online]. Available: http://www-device.eecs.berkeley.edu/~bsim3/bsim4.html

[21] (2007, April) PSP 102.1 Manual. [Online]. Available: http://pspmodel.asu.edu/documentation.htm

[22] (2006) HiSIM. [Online]. Available: http://home.hiroshima-u.ac.jp/usdl/HiSIM

[23] (2006, August) EKV. [Online]. Available: http://legwww.epfl.ch/ekv/model.html

[24] S. Borkar, "Design challenges of technology scaling," *IEEE Micro*, vol. 19, no. 4, pp. 23–29, 1999.

[25] B. Yu, L. Chang, S. Ahmed, H. Wang, S. Bell, C.-Y. Yang, C. Tabery, C. Ho, Q. Xiang, T.-J. King, J. Bokor, C. Hu, M.-R. Lin, and D. Kyser, "FinFET scaling to 10 nm gate length," in *IEDM Tech. Dig.*, 2002, pp. 251–254.

[26] Y. Taur, "Analytic solutions of charge and capacitance in symmetric and asymmetric double-gate MOSFETs," *IEEE Trans. Electron Devices*, vol. 48, no. 12, pp. 2861–2869, 2001.

[27] G. D. J. Smit, A. J. Scholten, N. Serra, R. M. T. Pijper, R. van Langevelde, A. Mercha, G. Gildenblat, and D. B. M. Klaassen, "PSP-based compact FinFET model describing dc and RF measurements," in *IEDM Tech. Dig.*, 2006, pp. 1–4.

[28] M. V. Dunga, C.-H. Lin, X. Xi, D. D. Lu, A. M. Niknejad, and C. Hu, "Modeling Advanced FET Technology in a Compact Model," *IEEE Trans. Electron Devices*, vol. 53, pp. 1971–1978, Sept. 2006.

[29] Y. Tsividis, *Operation and Modeling of the MOS Transistor*. New York: McGraw-Hill, 1987.

[30] F. Stern, "Electronic properties of two-dimensional systems," *Reviews of Modern Physics*, vol. 54, pp. 437–672, April 1982.

[31] R. Rios, N. D. Arora, C.-L. Huang, N. Khabil, J. Faricelli, and L. Gruber, "A physical compact MOSFET model, including quantum mechanicals effects, for statistical circuit design applications," in *IEDM Tech. Dig.*, 1995, pp. 947–950.

[32] E. Merzbacher, *Quantum Mechanics*. Wiley, 1997.

[33] G. Baccarani and S. Reggiani, "A compact double-gate MOSFET model comprising quantum-mechanical and nonstatic effects," *IEEE Trans. Electron Devices*, vol. 46, no. 8, pp. 1656–1666, 1999.

[34] L. Ge and J. G. Fossum, "Analytical modeling of quantization and volume inversion in thin Si-film double gate MOSFETs," *IEEE Trans. Electron Devices*, vol. 49, no. 2, pp. 287–294, 2002.

[35] C.-H. Lin, M. V. Dunga, A. M. Niknejad, and C. Hu, "A compact quantum-mechanical model for Double-Gate MOSFET," in *Solid State and Integrated Circuit Technology, ICSICT*, 2006, pp. 1272–1274.

[36] V. P. Trivedi and J. G. Fossum, "Quantum-Mechanical effects on the Threshold Voltage of Undoped Double-Gate MOSFETs," *IEEE Electron Device Lett.*, vol. 26, no. 8, pp. 579–582, 2005.

[37] Y. Cheng and C. Hu, *MOSFET Modeling and BSIM3 User's Guide*. Kluwer Academic, 1999.

[38] Y. Taur, X. Liang, W. Wang, and H. Lu, "A Continuous, Analytic Drain-Current Model for DG MOSFETs," *IEEE Electron Device Lett.*, vol. 25, no. 2, pp. 107–109, 2004.

[39] J.-M. Sallese, F. Krummenacher, F. Pregaldiny, C. Lallement, A. Roy, and C. Enz, "A design oriented charge-based current model for symmetric DG MOSFET and its correlation with the EKV formalism," *Solid-State Electronics*, vol. 49, pp. 485–489, 2005.

[40] D. Ward and R. Dutton, "A charge-oriented model for MOS transistor capacitances," *IEEE J. Solid-State Circuits*, vol. 13, no. 5, pp. 703–708, 1978.

[41] H. C. Pao and C. T. Sah, "Effects of diffusion current on charectistics of metal-oxide(insulator)-semiconductor transistors," *Solid-State Electronics*, vol. 9, no. 10, pp. 927–937, 1966.

[42] F. Balestra, S. Cristoloveanu, M. Benachir, J. Brini, and T. Elewa, "Double-gate silicon-on-insulator transistor with volume inversion: A new device with greatly enhanced performance," *IEEE Electron Device Lett.*, vol. 8, no. 9, pp. 410–412, 1987.

[43] J. Kedzierski, D. M. Fried, E. J. Nowak, T. Kanarsky, J. H. Rankin, H. Hanafi, W. Natzle, D. Boyd, Z. Ying, R. A. Roy, J. Newbury, Y. Chienfan, Y. Qingyun, P. Saunders, C. P. Willets, A. Johnson, S. P. Cole, H. E. Young, N. Carpenter, and Rakows, "High-performance symmetric-gate and CMOS-compatible Vt asymmetric-gate FinFET devices," in *IEDM Tech. Dig.*, 2001, pp. 437–440.

[44] X. Liang and Y. Taur, "A 2-D analytical solution for SCEs in DG MOSFETs," *IEEE Trans. Electron Devices*, vol. 51, no. 9, pp. 1385–1391, 2004.

[45] Q. Chen, E. M. Harrell, and J. D. Meindl, "A physical short-channel threshold voltage model for undoped symmetric double-gate MOSFETs," *IEEE Trans. Electron Devices*, vol. 50, no. 7, pp. 1631–1637, 2003.

[46] K. Suzuki, T. Tanaka, Y. Tosaka, H. Horie, and Y. Arimoto, "Scaling theory for double-gate SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 40, no. 12, pp. 2326–2329, 1993.

[47] K. Suzuki, Y. Tosaka, and T. Sugii, "Analytical threshold voltage model for short channel n+/p+ double-gate SOI MOSFETs,," *IEEE Trans. Electron Devices*, vol. 43, no. 5, pp. 732–738, 1996.

[48] K. K. Young, "Short-channel effect in fully depleted SOI MOSFET's," *IEEE Trans. Electron Devices*, vol. 36, no. 2, pp. 399–402, 1989.

[49] G. Pei, J. Kedzierski, P. Oldiges, M. Ieong, and E. Kan, "FinFET design considerations based on 3-D simulation and analytical modeling," *IEEE Trans. Electron Devices*, vol. 48, no. 8, pp. 1441–1419, 2002.

[50] C.-H. Lin, "Modeling," Ph.D. dissertation, University of California, Berkeley, 373 Cory Hall, Berkeley CA 94720, July 2007.

[51] A. Bansal, B. C. Paul, , and K. Roy, "Modeling and optimization of fringe capacitance of nanoscale DGMOS devices," *IEEE Trans. Electron Devices*, vol. 52, no. 2, pp. 256–262, 2005.

[52] W. Wu and M. Chan, "Analysis of Geometry-Dependent Parasitics in Multifin Double-Gate FinFETs," *IEEE Trans. Electron Devices*, vol. 54, no. 4, pp. 692–698, 2007.

[53] P. Bendix, P. Rakers, P. Wagh, L. Lemaitre, W. Grabinski, C. C. McAndrew, X. Gu, and G. Gildenblat, "RF distortion analysis with compact MOSFET models," in *Proc. of the IEEE 2004 Custom Integrated Circuits Conference*, Oct. 2004, pp. 9–12.

[54] M. V. Dunga, C.-H. Lin, D. D. Lu, W. Xiong, C. R. Cleavelin, P. Patruno, J.-R. Hwang, F.-L. Yang, A. M. Niknejad, and C. Hu, "BSIM-MG : A Versatile Multi-Gate FET Model for Mixed-Signal Design," in *Proc. Symp. VLSI Technol.*, 2007, pp. 60–61.

[55] W. Xiong, K. Shin, C. R. Cleavelin, T. Schulz, K. Schruefer, I. Cayrefourcg, M. Kennard, C. Mazure, P. Patruno, and T.-J. K. Liu, "FinFET Performance Enhancement with Tensile Metal Gates and Strained Silicon on Insulator (sSOI)

Substrate," in *Proc. of the IEEE 2006 Device Research Conference*, Jun. 2006, pp. 39–40.

[56] D. Frank, Y. Taur, M. Ieong, and H.-S. Wong, "Monte Carlo modeling of threshold variation due to dopant fluctuations," in *Proc. Symp. VLSI Technol.*, 1999, pp. 169–170.

[57] S. Balasubramanian, J. L. Garrett, V. Vidya, B. Nikolic, and T. J. King, "Energy-delay optimization of thin-body MOSFETs for the sub-15 nm regime," in *Proc. of the IEEE 2004 International SOI Conference*, Oct. 2004, pp. 27–29.

[58] L. Mathew, Y. Du, A.-Y. Thean, M. Sadd, A. Vandooren, C. Parker, T. Stephens, R. Mora, R. Rai, M. Zavala, D. Sing, S. Kalpat, J. Hughes, R. Shimer, S. Jallepalli, G. Workman, W. Zhang, J. Fossum, B. White, B.-Y. Nguyen, and J. Mogab, "CMOS Vertical Multiple Independent Gate Field Effect Transistor (MIGFET)," in *Proc. of the IEEE 2004 International SOI Conference*, Oct. 2004, pp. 187–180.

[59] H. Lu and Y. Taur, "Analytic Potential Model for Symmetric and Asymmetric DG MOSFETs," *IEEE Trans. Electron Devices*, vol. 53, no. 5, pp. 1161–1168, 2006.

[60] D. Lu, M. Dunga, C.-H. Lin, A. Niknejad, and C. Hu, "A Multi-Gate MOSFET Compact Model Featuring Independent-Gate Operation," in *IEDM Tech. Dig.*, 2007.

[61] D. Lu, "Efficient Surface Potential Calculation for the Asymmetric Independent Double-Gate MOSFET," Master's thesis, University of California, Berkeley, 373 Cory Hall, Berkeley CA 94720, 2007.

[62] G. Gildenblat, X. Li, W. Wu, H. Wang, A. Jha, R. van Langevelde, G. Smit, A. Scholten, and D. Klaassen, "PSP: an advanced surface-potential-based MOSFET model for circuit simulation," *IEEE Trans. Electron Devices*, vol. 53, no. 9, pp. 1979–1993, 2006.

[63] J. R. Brews, "A charge sheet model for the MOSFET," *Solid-State Electronics*, vol. 21, pp. 345–355, 1978.

[64] K. Rim, S. Koester, M. Hargrove, J. Chu, P. Mooney, J. Ott, T. Kanarsky, P. Ronsheim, M. Ieong, A. Grill, and H.-S. P. Wong, "Strained Si NMOSFETs for high performance CMOS technology," in *VLSI Tech. Dig.*, 2001, pp. 59–60.

[65] A. Shimizu, K. Hachimine, N. Ohki, H. Ohta, M. Koguchi, Y. Nonaka, H. Sato, and F. Ootsuka, "Local mechanical-stress control (LMC): a new technique for CMOS-performance enhancement," in *IEDM Tech. Dig.*, 2001, pp. 433–436.

[66] K. Mistry, M. Armstrong, C. Auth, S. Cea, T. Coan, T. Ghani, T. Hoffmann, A. Murthy, J. Sandford, R. Shaheed, K. Zawadzki, K. Zhang, S. Thompson, and M. Bohr, "Delaying forever: Uniaxial strained silicon transistors in a 90 nm CMOS technology," in *VLSI Tech. Dig.*, 2004, pp. 50–51.

[67] C. H. Chen, T. L. Lee, T. H. Hou, C. L. Chen, C. C. Chen, J. W. Hsu, K. L. Cheng, Y. H. Chiu, H. J. Tao, Y. Jin, C. H. Diaz, S. C. Chen, and M. S. Liiang, "Stress memorization technique (SMT) by selectively strained-nitride capping for sub-65 nm high-performance strained-Si device application," in *VLSI Tech. Dig.*, 2004, pp. 56–57.

[68] V. Moroz, L. Smith, X.-W. Lin, D. Pramanik, and G. Rollins, "Stress-aware design methodology," in *Proc. of the 7th ISQED*, 2006.

[69] D. J. Paul, "Si/SiGe heterostructures: from material and physics to devices and circuits," *Semicond. Sci. Technol.*, vol. 19, pp. 75–108, 2004.

[70] S. E. Thompson, G. Sun, Y. S. Choi, and T. Nishida, "Uniaxial-Process-Induced Strained-Si: Extending the CMOS Roadmap," *IEEE Trans. Electron Devices*, vol. 53, no. 5, pp. 1010–1020, 2006.

[71] S. E. Thompson, G. Sun, K. Wu, J. Lim, and T. Nishida, "Key differences for process-induced uniaxial vs. substrate-induced biaxial stressed Si and Ge channel MOSFETs," in *IEDM Tech. Dig.*, 2004, pp. 221–224.

[72] F. Andrieu, T. Ernst, F. Lime, F. Rochette, K. Romanjek, S. Barraud, C. Ravit, F. Boeuf, M. Jurczak, M. Casse, O. Weber, L. Brevard, G. Reimbold, G. Ghibaudo, and S. Deleonibus, "Experimental and Comparative Investigation of Low and High Field Transport in Substrate- and Process-Induced Strained Nanoscale MOSFETs," in *VLSI Tech. Dig.*, 2005, pp. 176–177.

[73] A. A. Abidi, "RF CMOS comes of age," *IEEE J. Solid-State Circuits*, vol. 39, no. 4, pp. 549–561, 2004.

[74] C. H. Doan, S. Emami, A. M. Niknejad, , and R. W. Brodersen, "Millimeter-wave CMOS design," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 144–155, 2005.

[75] A.Hajimiri and T. H. Lee, "A general theory of phase noise in electrical oscillators," *IEEE J. Solid-State Circuits*, vol. 33, no. 2, pp. 179–194, 1998.

[76] A. J. Scholten, L. F. Tiemeijer, R. Langevelde, R. J. Havens, A. T. A. Z. van Duijnhoven, and V. C. Venezia, "Noise modeling for RF CMOS circuit simulations," *IEEE Trans. Electron Devices*, vol. 50, no. 3, pp. 618–632, 2003.

[77] G. I. Wirth, J. Koh, R. da Silva, R. Thewes, and R. Brederlow, "Modelling of Statistical Low-Frequency Noise of Deep-Submicron MOSFETs," *IEEE Trans. Electron Devices*, vol. 52, no. 7, pp. 1576–1588, 2005.

[78] K. K. Hung, P. K. Ko, C. Hu, and Y. C. Cheng, "A unified model for the flicker noise in metal-oxide-semiconductor field effect transistors," *IEEE Trans. Electron Devices*, vol. 37, no. 3, pp. 654–665, 1990.

[79] M. J. Uren, D. J. Day, and M. J. Kirton, "$1/f$ and random telegraph noise in silicon metal-oxide-semiconductor field-effect transistors," *Appl. Phys. Lett.*, vol. 47, no. 11, pp. 1195–1197, 1985.

[80] S. Christensson, I. Lundstrom, and C. Svensson, "Low frequency noise in MOS transistorsI Theory," *Solid-State Electron.*, vol. 11, pp. 797–812, 1968.

[81] A. Asenov, R. Balasubramaniam, A. R. Brown, and J. H. Davies, "RTS amplitudes in decananometer MOSFETs: a 3D simulation study," *IEEE Trans. Electron Devices*, vol. 50, no. 3, pp. 839–845, 2003.

[82] M. T. Yang, C. W. Kuo, A. K. L. Chang, Y. J. Wang, and S. Liu, "Statistical characterization and Monte-Carlo simulation of low-frequency noise variations in foundry AMS/RF CMOS technology," in *Silicon Monolithic Integrated Circuits in RF Systems Dig.*, 2006, pp. 18–20.

[83] C. Surya and T. Y. Hsiang, "Theory and experiment on the $1/f^\gamma$ noise in p-channel metal-oxide-semiconductor field-effect transistors at low drain bias," *Phy. Rev. B*, vol. 33, pp. 4898–4905, Apr. 1986.

[84] Z. Celik-Butler and T. Y. Hsiang, "Spectral Dependence of $1/f^\gamma$ Noise on Gate Bias in N-MOSFETs," *Solid-State Electron.*, vol. 30, no. 4, pp. 419–423, 1987.

[85] Z. Shi, J.-P. Mieville, and M. Dutoit, "Random Telegraph Signals in Deep Submicron n-MOSFET's," *IEEE Trans. Electron Devices*, vol. 41, no. 7, pp. 1161–1168, 1994.

[86] M. J. Kirton and M. J. Uren, "Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency $(1/f)$ noise," *Adv. Phys.*, vol. 38, no. 4, pp. 367–468, 1989.