UNIVERSITY OF CALIFORNIA, SAN DIEGO

Design and Modeling of Non-Classical MOSFETs

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Electrical Engineering (Applied Physics)

by

Bo Yu

Committee in charge:

Professor Yuan Taur, Chair
Professor Peter Asbeck, Co-Chair
Professor Prabhakar Bandaru
Professor Chung-Kuan Cheng
Professor Yuhwa Lo

2009
The dissertation of Bo Yu is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

______________________________________________

______________________________________________

______________________________________________

Co-Chair

______________________________________________

Chair

University of California, San Diego

2009
DEDICATION

This dissertation is dedicated to my wife and my parents.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signature Page</td>
<td>iii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>v</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xiv</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>xv</td>
</tr>
<tr>
<td>Vita</td>
<td>xviii</td>
</tr>
<tr>
<td>Abstract of the Dissertation</td>
<td>xx</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 The Story of CMOS Scaling</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Si-Based Non-Classical MOSFETs</td>
<td>5</td>
</tr>
<tr>
<td>1.3 Non-classical MOSFETs Based on New Materials</td>
<td>11</td>
</tr>
<tr>
<td>1.4 Organization of the Dissertation</td>
<td>14</td>
</tr>
<tr>
<td>2 Compact Model for Long Channel DG and SG MOSFETs</td>
<td>17</td>
</tr>
<tr>
<td>2.1 Drain Current Model for DG MOSFETs</td>
<td>17</td>
</tr>
<tr>
<td>2.2 Charge and Capacitance Model for DG MOSFETs</td>
<td>24</td>
</tr>
<tr>
<td>2.3 Drain Current Model for SG MOSFETs</td>
<td>29</td>
</tr>
<tr>
<td>2.4 Charge and Capacitance Model for SG MOSFETs</td>
<td>32</td>
</tr>
<tr>
<td>2.5 Summary</td>
<td>38</td>
</tr>
<tr>
<td>3 Explicit Solutions for DG and SG MOSFETs</td>
<td>40</td>
</tr>
<tr>
<td>3.1 General Method</td>
<td>40</td>
</tr>
<tr>
<td>3.2 Explicit Solution for DG MOSFETs</td>
<td>46</td>
</tr>
<tr>
<td>3.3 Explicit Solution for SG MOSFETs</td>
<td>50</td>
</tr>
</tbody>
</table>
3.4 Accuracy of the Explicit Solutions ........................................... 53
3.5 Summary ................................................................. 58

4 Unified Analytic Drain Current Model for MG MOSFETs .............. 60
4.1 Basic Idea ............................................................... 60
4.2 Model for QG MOSFETs ............................................... 63
4.3 Model for TG and TPG MOSFETs ..................................... 69
4.4 Unified Drain Current Model .......................................... 73
4.5 Summary ................................................................. 75

5 Modeling of Short-Channel Effects in DG and SG MOSFETs ........ 76
5.1 Scale Length Theory ..................................................... 76
5.2 Analytical Model for SCEs in DG MOSFETs .......................... 86
5.3 Analytical Model for SCEs in SG MOSFETs .......................... 92
5.4 Summary ................................................................. 102

6 Analysis of MOS Capacitance in III-V MOSFETs ....................... 104
6.1 A Review of Inversion Layer Capacitance in Bulk Si MOSFETs: from
  Classical to Quantum-Mechanical ........................................ 104
6.2 Physical Understanding of Inversion Layer Capacitance in III-V MOS-
  FETs ............................................................................. 111
  6.2.1 Density-of-states capacitance ...................................... 111
  6.2.2 Capacitance relation forced by Poisson’s equation ............ 115
  6.2.3 Simplified picture ................................................... 117
  6.2.4 Degradation of gate insulator capacitance ....................... 124
6.3 Degradation of DOS Capacitance in Non-Equilibrium: Ballistic MOS-
  FETs ............................................................................. 126
6.4 Effects of Interface States ................................................ 129
6.5 Summary ................................................................. 131

7 Device Design of III-V MOSFETs for Sub-22 nm Scaling .............. 133
7.1 Selection of Basic Device Structure ..................................... 133
7.2 Short Channel Design .................................................. 139
LIST OF FIGURES

Figure 1.1: 2007 ITRS product technology trends: product functions/chip and industry average “Moore’s Law” trends. Adapted from [5]. ................................................................. 2

Figure 1.2: Scaling of transistor size (physical gate length) with technology node to sustain Moore’s Law. Adapted from [6]. .................. 3

Figure 1.3: (a) Strained-Si p-channel MOSFET process flow for the representative stacked gate transistor and TEM cross-sectional view. (b) dual stress liner process architecture with tensile and compressive silicon nitride capping layers. Adapted from [10]. .................. 4

Figure 1.4: Different topologies of DG MOSFETs: (a) Planar type; (b) Vertical type; (c) Fin type. Adapted from [28]. ................. 6

Figure 1.5: Cross sectional TEM photo of a typical FinFET fabricated by Texas Instruments. Adapted from [30]. .................. 7

Figure 1.6: Schematic diagram of Π-gate and Ω-gate MOSFET cross-sections. Adapted from [17]. .................. 8

Figure 1.7: Three-dimensional schematic diagram of a QG MOSFET. Adapted from [41]. .................. 9

Figure 1.8: Cross sectional TEM photo of a typical SG MOSFET (circular nanowire transistor). Adapted from [42]. ................. 10

Figure 1.9: (a) Schematic diagram of a bulk-like Al$_2$O$_3$/In$_{0.75}$Ga$_{0.25}$As nMOSFET. (b) TEM photo of a similarly fabricated device, with an inserted HRTEM photo showing sharp Al$_2$O$_3$/In$_{0.75}$Ga$_{0.25}$As interface remaining after full device fabrication. Adapted from [49]. .... 12

Figure 1.10: Major parameters for fabricated devices with different gate length. Adapted from [49]. .................. 13

Figure 2.1: (a) Schematic diagram of a DG MOSFET. (b) Band diagram along a vertical cut (AA) in (a). The potential $\psi(x, y)$ is defined as the intrinsic Si (midgap) level referenced to the Fermi level of the n+ source as shown. .................. 19

Figure 2.2: $I_{ds}$-$V_{ds}$ curves calculated from the analytic model (solid curves), compared with the 2-D numerical simulation results (open circles). A constant mobility of 300 cm$^2$/V·s is used in both calculations. .................. 22
Figure 2.3: \( I_{ds} - V_{gs} \) characteristics obtained from the analytic model for two different values of \( t_{si} \) (solid and dashed curves), compared with the 2-D numerical simulation results (symbols). The same currents are plotted on both logarithmic (left) and linear (right) scales.

Figure 2.4: Normalized capacitance coefficients of a DG MOSFET as functions of \( V_{gs} \) obtained from the analytic model (solid lines), in comparison with the 2-D numerical simulation results (open circles). The normalization constant is \( (4\epsilon_{si}L/r)(W/t_{si})(V_{ds} = 2V) \).

Figure 2.5: Normalized capacitance coefficients of a DG MOSFET as functions of \( V_{gs} \) obtained from the analytic model (solid lines), in comparison with the 2-D numerical simulation results (open circles). The normalization constant is \( (4\epsilon_{si}L/r)(W/t_{si})(V_{ds} = 0V) \).

Figure 2.6: Normalized capacitance coefficients of a DG MOSFET as functions of \( V_{ds} \) obtained from the analytic model (solid lines), in comparison with the 2-D numerical simulation results (open circles). The normalization constant is \( (4\epsilon_{si}L/r)(W/t_{si}) \).

Figure 2.7: Schematic diagram of an SG MOSFET (cut through symmetry axis).

Figure 2.8: Normalized terminal charges \( Q_g, Q_s, \) and \( Q_d \) of an SG MOSFET as functions of \( V_{ds} \). \( 8\pi\epsilon_{si}L(kT/q)(1 - \alpha_s)/\alpha_s \) is the normalization constant.

Figure 2.9: Normalized capacitance coefficients of an SG MOSFET as functions of \( V_{gs} \) obtained from the analytic model (solid lines), in comparison with the 2-D numerical simulation results (open circles). The normalization constant is \( 4\pi\epsilon_{si}L/s (V_{ds} = 2V) \).

Figure 2.10: Normalized capacitance coefficients of an SG MOSFET as functions of \( V_{gs} \) obtained from the analytic model (solid lines), in comparison with the 2-D numerical simulation results (open circles). The normalization constant is \( 4\pi\epsilon_{si}L/s (V_{ds} = 1V) \).

Figure 2.11: Normalized capacitance coefficients of an SG MOSFET as functions of \( V_{gs} \) obtained from the analytic model (solid lines), in comparison with the 2-D numerical simulation results (open circles). The normalization constant is \( 4\pi\epsilon_{si}L/s (V_{ds} = 0V) \).

Figure 2.12: Normalized capacitance coefficients of an SG MOSFET as functions of \( V_{gs} \) obtained from the analytic model (solid lines), in comparison with the 2-D numerical simulation results (open circles). The normalization constant is \( 4\pi\epsilon_{si}L/s \).
Figure 3.1: The absolute error $|f|$ as a function of $r$ and $F$ for implicit equation of DG MOSFETs.  

Figure 3.2: The absolute error $|g|$ as a function of $s$ and $G$ for implicit equation of SG MOSFETs.  

Figure 3.3: The absolute error $|g(Q/Q_0; s, G)|$ as a function of $s$ and $G$. The expression for $Q$ is given in [70].  

Figure 3.4: $I_{ds} - V_{gs}$ characteristics obtained from our explicit model (solid curves) compared with those from numerical method (dots) as well as explicit model in [70] (dashed curves). The same currents are plotted on both logarithmic (left) and linear (right) scales.  

Figure 3.5: Transconductances obtained from our explicit model (solid curves) compared with those from numerical method (dots) as well as explicit model in [70] (dashed curves).  

Figure 3.6: Transcapacitances obtained from our explicit model (solid curves) compared with those from numerical method (dots) as well as explicit model in [70] (dashed curves).  

Figure 3.7: $I_{ds} - V_{gs}$ characteristics obtained from our explicit model (solid curves) compared with those from numerical method (dots). The same currents are plotted on both logarithmic (left) and linear (right) scales.  

Figure 3.8: Transconductances obtained from our explicit model (solid curves) compared with those from numerical method (dots).  

Figure 3.9: Transcapacitances obtained from our explicit model (solid curves) compared with those from numerical method (dots).  

Figure 4.1: Schematic diagrams of MG MOSFET cross sections. (a) DG MOSFET. (b) SG MOSFET. (c) QG MOSFET. (d) TG MOSFET. (e) Π-gate MOSFET. (f) Ω-gate MOSFET.  

Figure 4.2: $Q_m(QG)$ and $Q_m(SG)$ as functions of $V_g - V$. $Q_m(SG)$ is calculated from the analytic model, whereas $Q_m(QG)$ is obtained from 2-D TCAD numerical simulations. Radius $R$ of SG is chosen to match subthreshold behavior of QG.  

Figure 4.3: $Q_m(QG)/Q_m(SG)$ as functions of $\alpha$. The results of $Q_m(QG)$ are obtained from 2-D TCAD numerical simulations, whereas the results of $Q_m(SG)$ are simply obtained from the analytic model, which has been validated.  

Figure 4.4: Mobile charge $Q_m$ of QG MOS capacitors as functions of $V_g - V$ obtained from the analytic model (solid lines) in comparison with the 2-D numerical simulation results (open circles).
Figure 4.5: Percentage error of mobile charge $Q_m$ between the analytic model and 2-D numerical simulation results of QG MOS capacitors as functions of $V_g - V$. ........................................... 67

Figure 4.6: Gate capacitance $C_g$ of QG MOS capacitors as functions of $V_g - V$ obtained from the analytic model (solid lines) in comparison with the 2-D numerical simulation results (open circles). .............. 67

Figure 4.7: The first order derivative of $C_g$, i.e., the second order derivative of $Q_m$ of QG MOS capacitors as functions of $V_g - V$ obtained from the analytic model (solid lines) in comparison with the 2-D numerical simulation results (open circles). .................. 68

Figure 4.8: Mobile charge $Q_m$ of TG and TPG MOS capacitors as functions of $V_g - V$ obtained from the analytic model (solid lines) in comparison with the 2-D numerical simulation results (open circles). 70

Figure 4.9: Percentage error of mobile charge $Q_m$ between the analytic model and 2-D numerical simulation results of TG and TPG MOS capacitors as functions of $V_g - V$. ............................... 70

Figure 4.10: Gate capacitance $C_g$ of TG and TPG MOS capacitors as functions of $V_g - V$ obtained from the analytic model (solid lines) in comparison with the 2-D numerical simulation results (open circles). 71

Figure 4.11: The first order derivative of $C_g$, i.e., the second order derivative of $Q_m$ of TG and TPG MOS capacitors as functions of $V_g - V$ obtained from the analytic model (solid lines) in comparison with the 2-D numerical simulation results (open circles). ......... 72

Figure 5.1: Schematic diagram of a DG MOSFET. ....................... 77

Figure 5.2: Schematic diagram of an SG MOSFET. ....................... 78

Figure 5.3: Scale lengths calculated from (5.11) with $\beta = 130.9$ and $\gamma = 27.3$ are compared with the numerical solutions obtained directly from the implicit equation (5.10). The dashed lines show the scale length ($\pi \lambda N$) of Auth and Plummer [76] for comparison. .......... 83

Figure 5.4: Scale lengths calculated from (5.12) with $\beta = 411.4$ and $\gamma = 822.6$ are compared with the numerical solutions obtained directly from the implicit equation (5.6). The dashed lines show the scale length ($\pi \lambda N$) of Suzuki et al. [73] for comparison. .......... 84

Figure 5.5: Normalized gate insulator thickness $t_{ox}/\lambda$ versus normalized silicon cylinder radius $R/\lambda$ for SG MOSFETs with different insulator permittivity. ........................................... 85
Figure 5.6: Normalized gate insulator thickness $t_{ox}/\lambda$ versus normalized silicon cylinder radius $t_{si}/\lambda$ for DG MOSFETs with different insulator permittivity. ........................................ 86

Figure 5.7: $I_{ds}L - V_g$ curves for SG MOSFETs obtained from the analytical model (solid lines) in comparison with the 2-D numerical simulation results (open circles). ........................................ 97

Figure 5.8: Threshold voltage roll-offs for SG MOSFETs as functions of channel length $L$ obtained from the analytical model (lines) in comparison with the 2-D numerical simulation results (symbols). .......... 99

Figure 5.9: DIBL effects for SG MOSFETs as functions of channel length $L$ obtained from the analytical model (lines) in comparison with the 2-D numerical simulation results (symbols). .......... 100

Figure 5.10: Subthreshold slopes for SG MOSFETs as functions of channel length $L$ obtained from the analytical model (lines) in comparison with the 2-D numerical simulation results (symbols). .......... 101

Figure 6.1: Energy-band diagram near the silicon surface of a p-type MOS device. The band bending $\psi$ is defined as positive when the bands bend downward with respect to the bulk. Adapted from [84]. 106

Figure 6.2: MOS capacitance-voltage curves: (a) low frequency, (b) high frequency, (c) deep depletion. Adapted from [88]. ............... 107

Figure 6.3: Equivalent circuit model of an MOS capacitor at low frequency. ................................................................. 108

Figure 6.4: Equivalent circuit model for split $C$-$V$ setup. ............... 108

Figure 6.5: An example of split $C$-$V$ measurement results. ............... 109

Figure 6.6: Comprehensive equivalent circuit model of an MOS capacitor at low frequency. .................................................. 114

Figure 6.7: A simplified equivalent circuit model of an MOS capacitor under the one subband approximation. ............................... 118

Figure 6.8: $x_i$, the effective thickness associated with inversion layer capacitance, as a function of “effective” DOS effective mass. “Effective” DOS effective mass means $m_{d,1}^*/[1 + \exp(E_F - E_i)/kT]$. ................ 124

Figure 6.9: A bar plot indicates different contributions to $x_{inv}$. Besides the important contributions from $x_{ox}$, $x_{av}$, and $x_{DOS}$, we also have a non-negligible term due to $C_{ox}$ degradation. ......................... 125

Figure 6.10: Illustration of how the $k$-states at the virtual source are filled by the two Fermi levels, i.e., Fermi levels from both the source and drain. ....................................................... 126
Figure 6.11: A simplified equivalent circuit model under the one sub-band approximation applicable to the virtual source of a ballistic MOSFET.

Figure 6.12: \(N_s = -(Q_i^+ + Q_i^-)/q\) as a function of \(V_{ds}\) obtained from gate equation (6.63) for (a) InGaAs; (b) Si. In both cases, we use 1.2 nm EOT for \(C_{ox}'\). But \(m_{d,1}^*\) for InGaAs and Si are 0.04\(m_0\) and 0.4\(m_0\), respectively. \(m_0\) is electron mass.

Figure 6.13: Equivalent circuit model of an MOS capacitor with interface states.

Figure 7.1: Field pattern for a short channel thick-BOX-SOI device showing the strong source and drain coupling.

Figure 7.2: Schematic diagram of the baseline device design of thin-BOX-SOI-like III-V MOSFETs for sub-22 nm scaling.

Figure 7.3: High drain \(V_t\) roll-off (\(V_{ds} = 1V\)) is plotted as a function of gate length for baseline design and two types of device variations including pure 5nm InGaAs channel and source/drain doping level (both \(2 \times 10^{19}\)cm\(^{-3}\) and \(8 \times 10^{19}\)cm\(^{-3}\)).

Figure 7.4: High drain \(V_t\) roll-off (\(V_{ds} = 1V\)) is plotted as a function of gate length for baseline design and two types of device variations including 7nm InAlAs bottom barrier and 0.9nm or 1.5nm GaAs cap layer.

Figure 7.5: High drain \(V_t\) roll-off (\(V_{ds} = 1V\)) is plotted as a function of gate length for baseline design and device variations with different dielectric constant or thickness of high-\(\kappa\) gate insulator.

Figure 7.6: Electron density distribution in III-V MOSFETs of 20nm gate length with \(V_{ds} = 1V\): a) baseline design with p+ ground plane only under the gate; b) baseline design but with p+ ground plane under the source/drain as well.

Figure 7.7: Monte-Carlo simulation results of turn-on characteristics at \(V_{ds} = 1V\) for the various device designs (curves shifted to account for the different threshold voltage in the various device designs). Adapted from [102].

Figure 7.8: Calculated transconductance based on Monte-Carlo simulation of the various device designs. Adapted from [102].
LIST OF TABLES

Table 4.1: Structure parameters, $C_1$, and $C_2$ for different devices . . . . 73

Table 7.1: Electron affinity, bandgap, electron effective mass, dielectric constant for InGaAs, InAlAs, InP, and high-$\kappa$ dielectric in ISE-DESSIS simulation. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 143
ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my advisor, Prof. Yuan Taur, for everything he has done for me during the past 4 years. Without his all-around support, I cannot reach this milestone in my life, and all the works in this dissertation would not have been possible. Prof. Taur is a great teacher who has the most significant influence on me. Besides delivering his over 30 years research experience and knowledge in the field of semiconductor device, he also imparts me the way he thinks, which should be even more important the knowledge itself, and I can take advantage of for whole life.

Tons of thanks should also be given to my co-advisor, Prof. Peter Asbeck, for his precious advice and guidance in my graduate study. His vast knowledge and deep insight greatly benefit me throughout our discussions of research. Without his leading the way, I cannot imagine I am able to dig into the project of III-V MOSFETs.

I would also like to thank my current committee members, Prof. Prabhakar Bandaru, Chung-Kuan Cheng, and Yu-Hwa Lo as well as previous committee member Prof. Edward Yu (who is with UT Austin now), for taking their time to serve as committee members, to review my dissertation and give me valuable comments.

Many thanks to my friends in UCSD, in particular, Ming Cai, Xiaoping Liang, Minjian Liu, Huaxin Lu, Wei-Yuan Lu, Jooyoung Song, and Wei Wang in
Prof. Taur’s group, as well as Lingquan Wang in Prof. Asbeck’s group. There are also a lot of friends who didn’t help me on my research work, but helped me a lot in everyday life. I’m not able to cover all of them due to the huge number.

Last, but not least, I want to express sincere gratitude to my family. Thanks to my parents in-law, who have been taking care of me during last 4 months. Finally, thanks to my wife and my parents, for whom I have thousands of words not necessary to be spelt out.

The text of Chapter 2, in part, is a reprint of the material as it appears in “Analytic Charge Model for Surrounding-Gate MOSFETs” by Bo Yu, Wei-Yuan Lu, Huaxin Lu, and Yuan Taur, IEEE Transaction on Electron Devices, Mar 2007. The dissertation author was the primary investigator and author of this paper.

The text of Chapter 2, in part, is a reprint of the material as it appears in “Analytic Charge Model for Double-Gate and Surrounding-Gate MOSFETs” by Bo Yu, Huaxin Lu, Wei-Yuan Lu, and Yuan Taur, Proceedings of NSTI Nanotech, May 2007. The dissertation author was the primary investigator and author of this paper.

The text of Chapter 3, in part, is a reprint of the material as it appears in “Explicit Continuous Models for Double-Gate and Surrounding-Gate MOSFETs” by Bo Yu, Huaxin Lu, Minjian Liu, and Yuan Taur, IEEE Transaction on Electron Devices, Oct 2007. The dissertation author was the primary investigator and
author of this paper.

The text of Chapter 4, in part, is a reprint of the material as it appears in “A Unified Analytic Drain-Current Model for Multiple-Gate MOSFETs” by Bo Yu, Jooyoung Song, Yu Yuan, Wei-Yuan Lu, and Yuan Taur, IEEE Transaction on Electron Devices, Aug 2008. The dissertation author was the primary investigator and author of this paper.

The text of Chapter 5, in part, is a reprint of the material as it appears in “Scaling of nanowire transistors” by Bo Yu, Lingquan Wang, Yu Yuan, Peter Asbeck, and Yuan Taur, IEEE Transaction on Electron Devices, Nov 2008. The dissertation author was the primary investigator and author of this paper.

The text of Chapter 5, in part, is a reprint of the material as it will appear in “A 2-D analytical solution for short-channel effects in nanowire MOSFETs” by Bo Yu, Yu Yuan, Jooyoung Song, and Yuan Taur, IEEE Transaction on Electron Devices, Oct 2009. The dissertation author was the primary investigator and author of this paper.
VITA

2001  B.E., University of Science and Technology of China
2004  M.S., University of Science and Technology of China
2009  Ph.D., University of California, San Diego

PUBLICATIONS


**FIELDS OF STUDY**

Major Field: Electrical Engineering

   Studies in Applied Physics

   Professors Yuan Taur and Peter Asbeck
As bulk CMOS scaling is approaching the limit that is imposed by gate oxide tunneling, body doping, band-to-band tunneling, etc., non-classical MOSFET is becoming an intense subject of very large-scale integration (VLSI) research. Among a variety of non-classical MOSFETs, multiple-gate (MG) MOSFETs which are still based on Si have been proposed to scale down CMOS technology more aggressively because of better control of short-channel effects (SCEs), whereas novel MOSFETs utilizing III-V materials instead of Si are suggested to achieve CMOS performance breakthrough even without scaling down too aggressively due
to the large mobility of mobile carriers. This dissertation focuses on the design and modeling of these two categories of non-classical MOSFETs.

Actually, many different types of Si-based MG MOSFETs have been designed and even fabricated in the last two decades, including double-gate (DG) MOSFETs, surrounding-gate (SG) MOSFETs, quadruple-gate (QG) MOSFETs, triple-gate (TG) MOSFETs, Π-gate MOSFETs, Ω-gate MOSFETs, and so on. Although the design work has been pretty much done, specific compact models for these MG MOSFETs other than BSIM, PSP, and HiSIM are in urgent need, because the charge sheet approximation is no longer appropriate for MG MOSFETs due to the so-called “volume inversion” effect. In this dissertation, we will first introduce the complete non-charge-sheet based analytic models of drain current, terminal charges and capacitance coefficients for long channel symmetric DG and SG MOSFETs. The DG and SG models will be generalized to a unified analytic drain current model for all kinds of MG MOSFETs, with some non-trivial yet reasonable approximations. Efforts will also be focused on making the physics-based model more versatile and computationally efficient.

On the contrary, the research on III-V MOSFETs is still in the primary phase. Compact modeling for III-V MOSFETs is not being considered in the current stage because the device technology itself is far away from maturity, and the interest of this dissertation is in device design and basic physical modeling. With SCEs treated as the top-drawer consideration, a baseline device design of
III-V MOSFET for sub-22nm scaling is proposed based on the thin-BOX-SOI-like structure. Physical modeling of capacitances in III-V MOSFETs has also been carried out to gain a more clear picture of capacitance degradation due to small density-of-states (DOS).
Introduction

1.1 The Story of CMOS Scaling

Although the original idea of field effect transistor (FET) was patented as early as 1930 by Lilenfeld [1], people waited 30 years until Kahng and Atala gave birth to the first metal-oxide-semiconductor field-effect transistor (MOSFET) using SiO$_2$ as the gate insulator [2]. The major breakthrough in the level of integration took place in 1963 thanks to the invention of complimentary MOS (CMOS) [3], in which both n-channel and p-channel MOSFETs are fabricated simultaneously on the same substrate. Ever since then, CMOS gradually became the dominant power in the mushroom semiconductor industry.

In the past few decades, the growth in the silicon very-large-scale-integration (VLSI) technology has been driven by the continued shrinking of transistors to ever smaller physical dimensions. As well known, the CMOS scaling can achieve higher
packing densities, higher circuit speeds, and lower power dissipation, therefore it is considered as the crucial material base in the evolutionary progress leading to the powerful and versatile electronic systems such as computers and communication devices. Actually, the number of transistors on a chip has incessantly doubled about every two years, as shown in Fig. 1.1. This exponential growth in the number of transistors and memory bits per chip is popularly known as Moore’s law [4]. From the transistor size point of view, either technology node or physical gate length has fallen into nanotechnology regime (sub-100nm) for several years, as indicated in Fig. 1.2. Currently, both 45nm bulk and partially-depleted silicon-on-insulator (PDSOI) technology nodes are in volume production, whereas products based on
32nm and beyond will come out not in a few years but in several months. By 2011, the physical gate length is expected to be at or below 10 nm corresponding to the 22nm technology node[6].

At this moment, the strong feeling grows day by day that the scaling limit for conventional bulk and PDSOI CMOS is not too far away from us. This limit is more or less imposed by some fundamental physical factors, e.g., 1) gate oxide leakage current caused by quantum-mechanical direct tunneling; 2) band-to-band tunneling between body and drain; 3) dopant fluctuation resulting in threshold voltage variation from device to device. Strain engineering has been introduced in 90nm technology node and beyond to achieve significant mobility enhancement [7]-[11]. There was a debate on which should be the best path to take, i.e., biaxial substrate stress or uniaxial-process-induced stress. Finally, uniaxial stress became
the choice of industry due to its advantages over biaxial stress, and different types of process-induced uniaxial stress have been developed [10], including 1) stress liners: compressive and tensile Silicon Nitride (SiN) for p- and n-channel devices, respectively, as schematically shown in Fig. 1.3(b); 2) embedded SiGe and SiC source/drain for p- and n-channel devices, respectively, as schematically shown in Fig. 1.3(a); 3) stress memorization technique (SMT) via poly-Si gate only effective for nMOSFETs. Besides the strain engineering, high-\(\kappa\) dielectrics and metal gate (HK/MG) have been adopted in 45nm technology node and beyond [12]–[16]. High-\(\kappa\) dielectrics significantly lower gate leakage and improved manufacturability for future technology nodes, whereas metal gate can largely reduce gate resistance and
eliminate the gate capacitance degradation due to poly-Si depletion effect. Even with all these advanced technologies (strain engineering and HK/MG), people still see the forthcoming scaling limit of conventional bulk and PDSOI CMOS, and thus pin hope on the non-classical MOSFETs to continue CMOS performance improvement beyond the conventional limit. Generally speaking, the non-classical MOSFETs can be divided into two primary categories: one category is still using Si as the channel material, the other is based on Ge or III-V materials. We will introduce them separately in the next two sections.

1.2 Si-Based Non-Classical MOSFETs

Si-based non-classical MOSFETs mainly points to all kinds of multiple-gate (MG) MOSFETs, which have been proposed to scale down CMOS technology more aggressively, because of better control of short-channel effects (SCEs) [17]. MG MOSFETs designed and fabricated in recent years include double-gate (DG) MOSFETs, surrounding-gate (SG) MOSFETs, quadruple-gate (QG) MOSFETs, triple-gate (TG) MOSFETs, Π-gate MOSFETs, Ω-gate MOSFETs, and so on. Besides the benefits on electrostatic integrity, MG MOSFETs also have some other potential promises. First of all, the subthreshold slope is ideally 60mV/Decade by ignoring short-channel degradation. In MG MOSFETs, the SCEs are rather controlled by the dimensions of silicon and oxide films. Because high doping is no longer required to achieve small depletion width, the silicon film is always
undoped or lightly doped, and thus MG MOSFETs are supposed to have better mobility than bulk or PDSOI MOSFETs because of the absence of strong impurity scattering. Due to the low body doping level, the threshold voltage variation caused by the dopant fluctuation will also be suppressed in MG MOSFETs.

The most intensively studied non-classical MOSFETs are DG MOSFETs [18]-[27]. Fig. 1.4 shows three possible orientations of DG MOSFET on a silicon wafer:

Figure 1.4: Different topologies of DG MOSFETs: (a) Planar type; (b) Vertical type; (c) Fin type. Adapted from [28].
Figure 1.5: Cross sectional TEM photo of a typical FinFET fabricated by Texas Instruments. Adapted from [30].

wafer. The planar type (Fig. 1.4(a)) has the advantage that the channel layer is in the plane of the silicon wafer surface so that the channel thickness is controlled by the thickness of uniform planar layers rather than by lithography [28]. However, it is a big challenge to align the top gate and the bottom gate in this type of topology. The vertical type (Fig. 1.4(b)), which has the channel in the vertical direction, is most compact for DRAM application [29] where low leakage current (hence a long channel) is important and performance is secondary, but it has obvious topological difficulties for a CMOS logic application [28]. At this moment, the Fin type (Fig. 1.4(c)) seems to be the most promising structure. First of all, self-aligned gate
can be easily achieved in this type, with all the terminals accessible from the top. Besides, it has the highest packing density for high-speed logic applications since the channel width, the longest dimension of a logic FET, is perpendicular to the plane of the wafer [28]. A typical cross sectional TEM photo for FinFET is shown in Fig. 1.5. This FinFET device is fabricated by Texas Instruments. The height and width of Si film are about 60nm and 30nm, respectively, and the SiO$_2$ gate insulator is only 2.2nm thick.

Actually, most FinFETs have three gates from geometric point of view, but the gate on top of the Si film is always unimportant because either the top gate insulator is relatively thick or the height of Si film is much larger than the width. However, if the height is comparable to the width and the top gate insulator is as thin as side gate insulator, then the three gates are equally important. Under this
circumstance, the device should be called triple-gate or tri-gate MOSFET [31] (TG MOSFET) rather than FinFET. TG MOSFETs have better characteristics than DG MOSFETs in terms of control of SCEs, but meanwhile they suffer more from quantum-mechanical effects due to 2-D quantum confinement. Π-gate and Ω-gate MOSFETs are simple variations of TG MOSFETs. They can be classified into triple-plus-gate (TPG) MOSFETs because these devices are basically TG devices with an extension of the gate electrode below the active Si island, which increases current drive and improves SCEs. The schematic diagram for Π-gate and Ω-gate MOSFET cross-sections is shown in Fig. 1.6.

In principle, MOSFET with gate all around can provide the best controllability of SCEs. Such a device is usually fabricated using a Si nanowire structure, and thus people sometimes call it nanowire transistor. Nanowire transistors have
generated much research interest lately [32]-[40], with emphasis put on two types of geometry: one has rectangular (or square) cross-section, and is usually named quadruple-gate (QG) MOSFET; the other has circular cross-section, and is always called surrounding-gate (SG) MOSFET. A 3-D schematic diagram of a QG MOSFET is illustrated in Fig. 1.7, whereas a typical cross sectional TEM photo for SG MOSFET is shown in Fig. 1.8.

Although the design work for Si-based MG MOSFETs has been pretty much done, specific models for MG MOSFETs other than BSIM [43], PSP [44], and HiSIM [45] are in urgent need, because the charge sheet approximation [46] is no longer appropriate for MG MOSFETs due to the so-called “volume inversion” effect [47]. In this dissertation, the focus will be put on the compact modeling for the part of Si-based non-classical MOSFETs. Our group has developed an analytic
potential model for DG MOSFETs several years before [48]. The contribution of this dissertation is to extend the DG model to MG MOSFETs. Efforts will also be focused on making the physics-based model more versatile and computationally efficient.

1.3 Non-classical MOSFETs Based on New Materials

Another category of non-classical MOSFETs is based on new materials other than Si, and the candidates of new channel materials for MOSFETs basically include III-V and Ge up to now. III-V materials and Ge have been used in transistors for tens of years, such as SiGe BJTs, heterojunction bipolar transistors (HBTs), and high electron mobility transistors (HEMTs). However, primarily because of the bad quality of dielectric/semiconductor interface, they were not considered as good candidates for CMOS used in logic circuits until the conventional Si bulk CMOS scaling limit is being approached. III-V materials and Ge promise significantly higher mobilities than Si, strained Si, or even SiGe MOSFETs. For example, electron mobility in InGaAs and hole mobility in GaSb are 3-10 times higher than in Si. Theoretically, III-V/Ge MOSFETs can reach the ballistic transport limit at longer gate lengths, e.g., 22nm. Therefore, with the continuing improvement of interface quality recently, people believe that III-V or Ge MOSFET is able to achieve CMOS performance breakthrough even without scaling down too aggressively. In this dissertation, we will only discussion III-V
MOSFETs, because another particular benefit of III-V material is that, by tailoring composition, heterostructures offer additional degrees of freedom to optimize the device.

Gate dielectric is the key factor for III-V MOSFETs. Although insulator with high permittivity and large barrier is much preferred, the quality of insulator/semiconductor interface is always the major concern. Many kinds of high-κ dielectrics have been considered for III-V MOSFETs, including $\text{Al}_2\text{O}_3$ [49]-[54], $\text{HfO}_2$ [54]-[56], $\text{HfAlO}$ [54, 57, 58], $\text{ZrO}_2$ [59], and so on [60, 61]. But till now, it is not clear which one is the best choice for high mobility III-V materials like InGaAs.

III-V MOSFETs based on the bulk-like structure have already been successfully fabricated and excellent device performance has been demonstrated even in relatively long channel MOSFETs [49]-[58]. A typical example is schematically
Figure 1.10: Major parameters for fabricated devices with different gate length. Adapted from [49].

<table>
<thead>
<tr>
<th>$L_g$ (nm)</th>
<th>$I_{on}$ ($\mu A/\mu m$) ($V_{ds}=0.8\text{V}$)</th>
<th>$C_m$ ($\mu F/\mu m$) ($V_{ds}=0.8\text{V}$)</th>
<th>$SS$ ($V_{th}=0.05\text{V}$) (mV/dec)</th>
<th>$SS$ ($V_{th}=0.8\text{V}$) (mV/dec)</th>
<th>$D_{BL}$ (mV/V)</th>
<th>$I_{on}/I_{off}$ ($V_{ds}=0.8\text{V}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>110</td>
<td>685</td>
<td>700</td>
<td>430</td>
<td>1500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>588</td>
<td>704</td>
<td>273</td>
<td>653</td>
<td>956</td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>440</td>
<td>705</td>
<td>168</td>
<td>297</td>
<td>457</td>
<td>60</td>
</tr>
<tr>
<td>140</td>
<td>332</td>
<td>680</td>
<td>165</td>
<td>226</td>
<td>279</td>
<td>605</td>
</tr>
<tr>
<td>150</td>
<td>304</td>
<td>632</td>
<td>160</td>
<td>194</td>
<td>225</td>
<td>$1.2 \times 10^3$</td>
</tr>
<tr>
<td>160</td>
<td>270</td>
<td>580</td>
<td>150</td>
<td>170</td>
<td>187</td>
<td>$1.6 \times 10^3$</td>
</tr>
<tr>
<td>170</td>
<td>250</td>
<td>566</td>
<td>150</td>
<td>165</td>
<td>181</td>
<td>$1.8 \times 10^3$</td>
</tr>
<tr>
<td>180</td>
<td>246</td>
<td>544</td>
<td>149</td>
<td>162</td>
<td>160</td>
<td>$2.2 \times 10^3$</td>
</tr>
<tr>
<td>200</td>
<td>232</td>
<td>538</td>
<td>144</td>
<td>154</td>
<td>137</td>
<td>$2.7 \times 10^3$</td>
</tr>
</tbody>
</table>

illustrated in Fig. 1.9(a), with a TEM photo of a similarly fabricated device shown in Fig. 1.9(b). Such a design has very limited scaling potential. The major parameters for fabricated devices with different gate length are given in Fig. 1.10, from which we can clearly see that gate tends to lose control when gate length shrinks to 130nm and beyond. This observation is totally consistent with the simple estimation based on maximum depletion width which is determined by $1 \times 10^{17} \text{cm}^{-3}$ body doping. People would not be satisfied with 130nm gate length, since Si bulk technology has entered sub-100nm era many years before. Therefore, it is an interesting topic to design III-V MOSFETs that can be extremely scaled down, e.g., 22nm gate length and beyond. In this dissertation, some preliminary work on this topic will be introduced. It will be shown that simply increasing body doping level may not work for III-V MOSFETs.
III-V materials such as InGaAs and GaSb are better than Si in terms of mobility, and this benefit comes from the small effective mass of either electron or hole. However, the small effective brings not only advantages but also disadvantages. For example, the quantum confinement effect in the vertical direction becomes much stronger, and thus results in larger device performance sensitivity with respect to the process fluctuation. Another very significant negative factor caused by the small effective mass is the small density-of-states (DOS), which makes III-V MOSFETs behave differently as compared with Si MOSFETs in many different aspects, especially the gate capacitance degradation due to small DOS. This has to be taken into account in both design and modeling.

1.4 Organization of the Dissertation

This dissertation is devoted to design and modeling of non-classical MOSFETs. In this chapter, we have clarified the motivation of research on non-classical MOSFETs, and we have divided non-classical MOSFETs into two primary categories, namely, Si-based MG MOSFETs and Ge/III-V-based MOSFETs. This dissertation is correspondingly organized into two parts: the first part is on Si MG MOSFESTs and includes Chapter 2, 3, 4 and 5; the second part is on III-V MOSFETs and includes Chapter 6 and 7. For the Si MG part, the emphasis will be put on the compact modeling since the design work has been pretty much done, whereas for the III-V part, the interest is in device design and basic physical
Chapter 2 focuses on the complete analytic models of drain current, terminal charges and capacitance coefficients for long channel symmetric DG and SG MOSFETs. Starting from the exact solution of Poisson’s equation, we derive the models from current continuity equation and Ward-Dutton charge partition without the need for charge sheet approximation.

Similar to the surface-potential-based models (SPBMs) for bulk MOSFETs, we also have implicit equations for intermediate parameters to be solved in the DG and SG models introduced in Chapter 1. Chapter 3 presents explicit solutions for the intermediate parameters used in these models. The explicit solutions are accurate not only in terms of drain current and terminal charges but also their derivatives, including drain conductance, transconductance, and transcapacitances.

Chapter 4 generalizes the DG and SG models to a unified analytic drain current model for MG MOSFETs, covering DG, SG, QG, TG, and TPG MOSFETs. The unified MG drain current model serves as a core long-channel model that can be expanded for compact modeling of various kinds of MG MOSFETs.

In Chapter 5, based on generalized scale length theory, full analytical expressions of potential and subthreshold current have been derived for both short-channel DG and SG MOSFETs by solving 2-D Poisson’s equation in both the semiconductor and insulator regions. The compact models for SCEs including threshold voltage roll-off, DIBL, and subthreshold slope degradation are extracted.
from the subthreshold current expression. It has been verified by 2-D numerical simulation results that the SCE model is predictive with no fitting parameters.

Chapter 6 develops a comprehensive equivalent capacitance circuit model of an MOS capacitor. With the one subband approximation, analytical results for the capacitance components in the equivalent circuit model can be developed in terms of effective thickness, by investigating both Poisson’s equation and Schrödinger’s equation. These analytical results can help us to physically understand the capacitances in III-V MOSFETs, especially the degradation of inversion layer capacitance due to small DOS.

A baseline device design of III-V MOSFETs for sub-22nm scaling is proposed in Chapter 7. The design is based on the thin-BOX-SOI-like structure, therefore it can be scaled down to 19nm roughly without pushing the film thicknesses to the very bottom. Since the film thicknesses are not at physical limits yet, the design considerations are mostly on SCEs.

The last chapter summarizes the whole dissertation.
Compact Model for Long Channel

DG and SG MOSFETs

2.1 Drain Current Model for DG MOSFETs

Compact modeling of DG and SG MOSFETs is quite different from that of conventional bulk MOSFETs. In bulk MOSFETs, Poisson’s equation cannot be solved analytically due to the presence of both depletion charge and mobile charge. To carry out the Pao-Sah integral [62] explicitly, the charge sheet approximation [46] is necessary and thus adopted in all of the compact models for bulk MOSFETs, e.g., BSIM [43], PSP [44], HiSIM [45]. However, for DG and SG MOSFETs, analytical potential function can be obtained without charge sheet approximation. Exact solution to Poisson’s equation is possible due to the absence of depletion charge. Actually, the charge sheet approximation cannot accurately capture the
physics of DG and SG MOSFETs in the subthreshold region due to the volume inversion effect [47], therefore a non charge sheet based model is required.

In this section, we will first review the continuous analytic drain current model developed for DG MOSFETs recently [48]. It is shown that the analytical potential distribution across the silicon film can be obtained by solving the one-dimensional (1-D) Poisson’s equation. The model is derived directly from Pao-Sah integral without charge sheet approximation, and it covers all three regions of MOSFET operations: linear, saturation, and subthreshold, thus maintaining strong continuity between different regions, and yet is completely physics based without the need for ad-hoc fitting parameters.

Consider an undoped (or lightly doped) symmetric DG MOSFET schematically shown in Fig. 2.1(a). Under the gradual channel approximation, Poisson’s equation along a vertical cut perpendicular to the Si film [Fig. 2.1(b)] takes the following form with only the mobile charge (electrons) term:

\[
\frac{d^2 \psi}{dx^2} = \frac{q}{\epsilon_{si}} n_i e^{q(\psi-V)/kT} \tag{2.1}
\]

where \( q \) is the electronic charge, \( \epsilon_{si} \) is the permittivity of silicon, \( n_i \) is the intrinsic carrier density, \( \psi \) is the electrostatic potential (reference shown in Fig. 2.1(b)), and \( V \) is the electron quasi-Fermi potential which is constant in the \( x \)-direction since the current flows predominantly from the source to the drain along the \( y \)-direction. Here we consider an nMOSFET with \( q\psi/kT \gg 1 \) so that the hole density is negligible.
Equation (2.1) can then be integrated twice to yield the solution [63]

\[
\psi(x) = V - \frac{2kT}{q} \ln \left[ \frac{t_{si}}{2L_D} \cos \left( \frac{2\beta x}{t_{si}} \right) \right]
\]  

(2.2)
where \( L_{Di} = \sqrt{2\epsilon_{si}kT/q^2n_i} \) is the intrinsic Debye length, and \( \beta \) is to be determined by the boundary condition

\[
\epsilon_{ox} \frac{V_g - \Delta \phi - \psi(x = \pm \frac{t_{si}}{2})}{t_{ox}} = \pm \epsilon_{si} \frac{d\psi}{dx} \bigg|_{x=\pm \frac{t_{si}}{2}} \tag{2.3}
\]

Here \( \epsilon_{ox} \) is the permittivity of oxide, \( V_g \) is the voltage applied to both gates, \( t_{si} \) and \( t_{ox} \) are the silicon and oxide thicknesses, and \( \Delta \phi \) is the work function of both gate electrodes with respect to the intrinsic silicon. Substituting (2.2) into (2.3) leads to

\[
\ln \beta - \ln (\cos \beta) + r \beta \tan \beta = \frac{q (V_g - \Delta \phi - V)}{2kT} - \ln \left( \frac{2L_{Di}}{t_{si}} \right) \tag{2.4}
\]

where \( r = 2\epsilon_{si}t_{ox}/\epsilon_{ox}t_{si} \) is a structural parameter. The range of \( \beta \) is \( 0 < \beta < \pi/2 \).

Because (2.4) is an implicit equation of \( \beta \), the general approach to solve \( \beta \) is Newton-Ralphson numerical method.

Integrating the current continuity equation

\[
I_{ds} = \mu W Q_i \frac{dV}{dy} = \text{constant} \tag{2.5}
\]

from the source to the drain, the drain current can be written as Pao-Sah’s integral

\[
I_{ds} = \mu \frac{W}{L} \int_{V_s}^{V_d} Q_i(V) dV = \mu \frac{W}{L} \int_{\beta_s}^{\beta_d} Q_i(\beta) \frac{dV}{d\beta} d\beta \tag{2.6}
\]

where \( \mu \) is the effective mobility, \( W \) is the device width, \( L \) is the channel length, \( Q_i \) is the total mobile charge per unit gate area, \( V_s \) and \( V_d \) are the voltages applied to the source and drain, respectively, and \( \beta_s \) and \( \beta_d \) are solutions to (2.4) corresponding to \( V = V_s \) and \( V = V_d \), respectively. From Gauss’s law,

\[
Q_i = 2\epsilon_{si} \frac{d\psi}{dx} \bigg|_{x=\frac{t_{si}}{2}} = 2\epsilon_{si} \frac{2kT}{q} \frac{2\beta \tan \beta}{t_{si}} \tag{2.7}
\]
\[ \frac{dV}{d\beta} = -\frac{2kT}{q} \left( \frac{1}{\beta} + \tan \beta + r \tan \beta + r \beta \sec^2 \beta \right) \] (2.8)

Substitute \( Q_i \) and \( dV/d\beta \) into equation (2.6) and carry out the integration analytically

\[ I_{ds} = \mu \frac{8\varepsilon_{si} W}{L} \left( \frac{kT}{q} \right)^2 [p(\beta_d) - p(\beta_s)] \] (2.9)

with function \( p(\beta) \) defined by

\[ p(\beta) = -2\beta \tan \beta + \beta^2 - r\beta^2 \tan^2 \beta \] (2.10)

All the regions of MOSFET operation are covered under one continuous function (2.9), however, it is not an explicit function of terminal voltages. To understand the how bias conditions affect the drain current, we can simplify equation (2.9) in the linear, saturation, and subthreshold regions with asymptotic behaviors.

In the linear region above threshold, both \( \beta_s \) and \( \beta_d \) are close to \( \pi/2 \), therefore one can obtain

\[ I_{ds} = 2\mu C_{ox} \frac{W}{L} \left( V_{gs} - V_t - \frac{V_{ds}}{2} \right) V_{ds} \] (2.11)

where \( V_{gs} = V_g - V_s \), \( V_{ds} = V_d - V_s \), and

\[ V_t = V_0 + \frac{2kT}{q} \ln \left[ \frac{q(V_{gs} - V_0)}{2r kT} \right] \] (2.12)

with

\[ V_0 = \Delta \phi + \frac{2kT}{q} \ln \left( \frac{2L D_i}{t_{si}} \right) \] (2.13)

In the saturation region, where \( \beta_s \sim \pi/2 \) and \( \beta_d \sim 0 \), one obtains

\[ I_{ds} = \mu C_{ox} \frac{W}{L} \left[ (V_{gs} - V_t)^2 - \frac{4r k^2 T^2}{q^2} e^{-\frac{q(V_{gs} - V_0 - V_{ds})}{kT}} \right] \] (2.14)
Figure 2.2: $I_{ds}$-$V_{ds}$ curves calculated from the analytic model (solid curves), compared with the 2-D numerical simulation results (open circles). A constant mobility of 300 cm$^2$/V·s is used in both calculations.

Note that in this continuous model, the current approaches the saturation value with a difference term exponentially decreasing with $V_{ds}$, in contrast to common piecewise models in which the current is made to be constant in saturation. In the subthreshold region, both $\beta_s, \beta_d \sim 0$, and

$$I_{ds} = \mu \frac{W}{L} kT n_i t_{si} e^{\frac{q(V_{gs} - \Delta \phi)}{kT}} \left( 1 - e^{-\frac{qV_{ds}}{kT}} \right)$$  \hspace{1cm} (2.15)

Note that the subthreshold current is proportional to the silicon thickness $t_{si}$, but independent of $t_{ox}$. In contrast, the above threshold currents, (2.11) and (2.14), are proportional to $C_{ox}$, but independent of silicon film thickness. This is a manifestation of “volume inversion” in the subthreshold region that cannot be reproduced by standard charge-sheet-based drain current models.
Figs. 2.2 and 2.3 show that the analytic $I_{ds}$-$V_{gs}$ curves computed from this model, i.e., equation (2.9), are in complete agreement with 2-D numerical simulations of a long-channel DG MOSFET. “Volume inversion” in which the subthreshold current is proportional to $t_{si}$, is self-evident in Fig. 2.3 without any fitting parameter invoked.

In this section, we have reviewed a continuous I-V model derived from analytic solutions of Poisson’s and current continuity equations for long-channel DG MOSFETs [48]. No charge sheet approximation is invoked—a key to the proper depiction of “volume inversion” in subthreshold. It is shown that the I-V curves constructed by the analytic model are in complete agreement with 2-D numerical
simulation results without fitting terms or parameters. All regions of MOSFET operation, including the linear, saturation, and subthreshold regions, are covered by a continuous function that can be easily calculated. This avoids the pitfalls of the piecewise models, thus making this model ideally suited for compact model applications.

2.2 Charge and Capacitance Model for DG MOSFETs

Drain-current model is not enough for circuit simulation [64]. In order to calculate the dynamic behavior of the device and hence to enable AC and transient circuit simulation, terminal charge model for DG MOSFETs has been developed very recently [65], which will be briefly reviewed in this section.

Symmetric DG MOSFETs only have three terminals, therefore analytic expressions for three terminal charges, i.e., $Q_g$, $Q_s$, and $Q_d$, associated with gate, source, and drain, respectively, are required for circuit simulation of DG MOSFETs. The gate charge can be computed directly by integrating the charge density along the channel. For the source and drain charges, it is physically sound to adopt Ward-Dutton linear charge partition method, which is widely accepted for bulk MOSFETs [66]. Integrating the current continuity equation

$$dy = \mu W Q_i dV/I_{ds}$$

leads to $y = L[p(\beta) - p(\beta_s)]/[p(\beta_d) - p(\beta_s)]$, where $p(\beta)$ is defined by (2.10). With
this result, we can calculate the terminal charges as

\[ Q_g = W \int_0^L Q_i(y) dy = W^2 \frac{\mu}{I_{ds}} \int_{\beta_s}^{\beta_d} Q_i^2(\beta) \frac{dV}{d\beta} d\beta \]

\[ = 8\epsilon_{si} L \frac{W kT}{t_{si} q} \frac{u(\beta_d) - u(\beta_s)}{p(\beta_d) - p(\beta_s)} \]

\[ Q_d = -W \int_0^L \frac{y}{L} Q_i(y) dy = 8\epsilon_{si} L \frac{W kT}{t_{si} q} U_{sd} \]

\[ Q_s = -W \int_0^L \left(1 - \frac{y}{L}\right) Q_i(y) dy = 8\epsilon_{si} L \frac{W kT}{t_{si} q} U_{ds} \quad (2.17) \]

where

\[ U_{ij} = \frac{p(\beta_i) [u(\beta_j) - u(\beta_i)] + [v(\beta_j) - v(\beta_i)]}{[p(\beta_j) - p(\beta_i)]^2} \quad (2.18) \]

and functions \( u(\beta) \) and \( v(\beta) \) are defined as

\[ u(\beta) = -2 \int_0^\beta \beta^2 \tan^2 \beta \left(\frac{1}{\beta} + \tan \beta\right) d\beta - \frac{2}{3} r \beta^3 \tan^3 \beta \]

\[ v(\beta) = 2 \int_0^\beta \beta^2 \tan^2 \beta \left(\frac{1}{\beta} + \tan \beta\right) (\beta^2 - 2\beta \tan \beta) d\beta \]

\[ + 2r \int_0^\beta \beta^2 \tan^2 \beta \left(\frac{\beta}{\cos^2 \beta} + \tan \beta\right) (\beta^2 - 2\beta \tan \beta) d\beta \]

\[ - 2r \int_0^\beta \beta^4 \tan^4 \beta \left(\frac{1}{\beta} + \tan \beta\right) d\beta - \frac{2}{5} r^2 \beta^5 \tan^5 \beta \quad (2.19) \]

The imaginary component of the small-signal behavior of a three-terminal device can be characterized by nine nonreciprocal capacitance coefficients. Here, we shall adopt the convention in which the capacitance coefficient \( C_{mn} \) is defined as \( \delta_{mn} \times \partial Q_m / \partial V_n \), where \( m, n = g, s, d \), and \( \delta_{mn} \) is equal to 1 if \( m = n \), and \(-1\) if otherwise. For simplicity, we will only list the results of four linearly indepen-
dent components, from which all the nine capacitance coefficients can be obtained directly based on the charge conservation law [64]:

\[
C_{gs} = \frac{L^2 (g_{ds} + g_m)^2}{\mu I_{ds}} - \frac{Q_g (g_{ds} + g_m)}{I_{ds}}
\]

\[
= 8\varepsilon_{si} L \frac{W}{t_{si}} \beta_s \tan \beta_s \frac{\beta_s \tan \beta_s + U_{sd} + U_{ds}}{p(\beta_d) - p(\beta_s)}
\]

\[
C_{gd} = -\frac{L^2 g_{ds}^2}{\mu I_{ds}} + \frac{Q_g g_{ds}}{I_{ds}}
\]

\[
= 8\varepsilon_{si} L \frac{W}{t_{si}} \beta_d \tan \beta_d \frac{\beta_d \tan \beta_d + U_{sd} + U_{ds}}{p(\beta_s) - p(\beta_d)}
\]

\[
C_{dg} = -\frac{L^2 g_{ds}^2}{\mu I_{ds}} + \frac{Q_g g_{ds}}{I_{ds}} - \frac{(Q_s - Q_d) g_m}{I_{ds}}
\]

\[
= 8\varepsilon_{si} L \frac{W}{t_{si}} \beta_d \tan \beta_d \frac{\beta_d \tan \beta_d + U_{sd} + U_{ds}}{p(\beta_s) - p(\beta_d)}
\]

\[
+ (U_{ds} - U_{sd}) \frac{\beta_s \tan \beta_s - \beta_d \tan \beta_d}{p(\beta_s) - p(\beta_d)}
\]

\[
C_{sd} = \frac{(Q_s - Q_d) g_{ds}}{I_{ds}} = 8\varepsilon_{si} L \frac{W}{t_{si}} \beta_d \tan \beta_d \frac{(U_{ds} - U_{sd})}{p(\beta_d) - p(\beta_s)}
\]

(2.20)

Here transconductance \( g_m \) and conductance \( g_{ds} \) are given by

\[
g_m = \frac{\partial I_{ds}}{\partial V_{gs}} \Big|_{V_{ds}} = \mu \frac{\varepsilon_{si} W kT}{L \frac{t_{si}}{q}} (\beta_s \tan \beta_s - \beta_d \tan \beta_d)
\]

\[
g_{ds} = \frac{\partial I_{ds}}{\partial V_{ds}} \Big|_{V_{gs}} = \mu \frac{\varepsilon_{si} W kT}{L \frac{t_{si}}{q}} \beta_d \tan \beta_d.
\]

(2.21)

All these analytic expressions have infinite order of continuity, however, they become the form 0/0, i.e., both the numerator and denominator are 0, when \( V_s = V_d \). Inasmuch as the indeterminacies of type 0/0 will cause problems in numerical implementation, expressions for the exact values at \( V_s = V_d \) have to be provided, which are the limits of general expressions as \( V_d \) approaches \( V_s \). These limits can be calculated through L’Hospital’s rule or Taylor expansion. After a
Figure 2.4: Normalized capacitance coefficients of a DG MOSFET as functions of $V_{gs}$ obtained from the analytic model (solid lines), in comparison with the 2-D numerical simulation results (open circles). The normalization constant is $(4\epsilon_{si}L/r)(W/t_{si})$ ($V_{ds} = 2V$).

In a lengthy derivation, one can finally obtain the results as

$$Q_g = -2Q_s = -2Q_d = 8\epsilon_{si}L \frac{WkT}{t_{si}} \beta_s \tan \beta_s$$

$$C_{gg} = 2C_{gs} = 2C_{gd} = 2C_{sg} = 2C_{dg}$$

$$= -6C_{sd} = -6C_{ds} = 3C_{ss} = 3C_{dd}$$

$$= 8\epsilon_{si}L \frac{W}{t_{si}} \left( \frac{1}{\beta_s} + \tan \beta_s \right) + 2r \left( \frac{\beta_s}{\cos^2 \beta_s} + \tan \beta_s \right)$$

In Fig. 2.4, 2.5, and 2.6, we have compared the capacitance coefficient model with numerical simulation results of a long channel DG MOSFET. Fig. 2.4, where $V_{ds} = 2V$, shows that the general expressions of the model are in complete agreement with numerical simulations. In this case, the device is always in satura-
tion region, and it is reasonable to assume $\beta_d = 0$. Under this approximation, we can easily obtain the simple relation between capacitance coefficients from (2.20):

$$ C_{gg} = C_{gs}, \quad C_{sg} = C_{ss}, \quad C_{dg} = -C_{ds}, \quad \text{and} \quad C_{gd} = C_{sd} = C_{dd} = 0. $$

If we further consider the well-above threshold region, where $\beta_s \approx \pi/2$, we can even derive the simplified approximate expressions for these capacitance coefficients as follow

$$ C_{gg} \simeq \frac{10}{6} C_{sg} \simeq \frac{10}{4} C_{dg} \simeq \frac{10}{15} \frac{4\epsilon_{si}LW}{r} \left(1 - \frac{1}{r\beta_s \tan \beta_s}\right) \quad (2.24) $$

Fig. 2.5 indicates the validity of (2.23) for the special case of $V_s = V_d$. In Fig. 2.6, the capacitance coefficients from both analytic model and numerical simulation are compared as functions of $V_{ds}$. We can clearly observe the transition between two degenerate cases, which have been introduced above: one degeneracy is from

![Normalized capacitance coefficients of a DG MOSFET as functions of $V_{gs}$ obtained from the analytic model (solid lines), in comparison with the 2-D numerical simulation results (open circles). The normalization constant is $(4\epsilon_{si}L/r)(W/t_{si}) \ (V_{ds} = 0V)$.](image-url)
source-drain-interchange symmetry at $V_{ds} = 0$; another is induced by saturation at the drain end.

### 2.3 Drain Current Model for SG MOSFETs

A similar drain current model has been developed for SG MOSFETs [67] by following the approach we introduced in Section 2.1 [48].

Consider an undoped (or lightly doped) cylindrical SG MOSFET schematically shown in Fig. 2.7. Under the gradual channel approximation, Poisson’s
Figure 2.7: Schematic diagram of an SG MOSFET (cut through symmetry axis).

equation takes the following form with only the mobile charge (electrons) term:

\[
\frac{d^2 \psi}{d \rho^2} + \frac{1}{\rho} \frac{d \psi}{d \rho} = \frac{q}{\epsilon_{si}} \frac{n_i e^{(\psi-V)/kT}}{\rho}
\]  

(2.25)

where \( \psi(\rho) \) is the electrostatic potential, and \( V \) is the electron quasi-Fermi potential which is constant in the \( \rho \)-direction. Here we consider an nMOSFET with \( q\psi/kT \gg 1 \) so that the hole density is negligible.

The solution of equation (2.25) has been presented in [67]. We reorganize it as

\[
\psi(\rho) = V - \frac{2kT}{q} \ln \left[ \frac{R}{2} \sqrt{\frac{q^2 n_i}{2\epsilon_{si} kT (1-\alpha)}} \left( 1 - \frac{(1-\alpha) \rho^2}{R^2} \right) \right]
\]

(2.26)

where \( \alpha \) is to be determined by the boundary condition which has final form as

\[
\frac{1}{2} \ln (1-\alpha) - \ln \alpha + s \frac{1-\alpha}{\alpha} = \frac{q(V_g - \Delta \phi - V)}{2kT} - \ln \left( \frac{2}{R} \sqrt[2]{\frac{2\epsilon_{si} kT}{q^2 n_i}} \right)
\]

(2.27)
Here \( s = 2\varepsilon_{si}\ln(1 + \frac{t_{ox}}{R})/\epsilon_{ox} \) is a structural parameter, \( R \) is the radius of silicon cylinder. The range of \( \alpha \) is \( 0 < \alpha < 1 \). Because (2.27) is an implicit equation of \( \alpha \), the general approach to solve \( \alpha \) is Newton-Ralphson numerical method.

From Gauss’s law, the total mobile charge per unit gate area

\[
Q_i = 2\varepsilon_{si}\frac{2kT}{q} \frac{1 - \alpha}{\alpha R}
\]  

(2.28)

\( dV/d\alpha \) can also be expressed as a function of \( \alpha \) by differentiating (2.27):

\[
\frac{dV}{d\alpha} = \frac{kT}{q} \left( \frac{1}{1 - \alpha} + \frac{2}{\alpha} + \frac{2s}{\alpha^2} \right)
\]  

(2.29)

Therefore, the drain current can be calculated as

\[
I_{ds} = \mu \frac{2\pi R}{L} \int_{V_s}^{V_d} Q_i(V) dV = \mu \frac{2\pi R}{L} \int_{\alpha_s}^{\alpha_d} Q_i(\alpha) \frac{dV}{d\alpha} d\alpha
\]  

\[
= \mu \frac{8\pi\epsilon_{si}}{L} \left( \frac{kT}{q} \right)^2 [f(\alpha_d) - f(\alpha_s)]
\]  

(2.30)

where \( \alpha_s, \alpha_d \) are solutions to (2.27) corresponding to \( V = V_s \) and \( V = V_d \), respectively, and function \( f(\alpha) \) is defined by

\[
f(\alpha) = \left( -\frac{2}{\alpha} - \ln \alpha \right) + s \left( -\frac{1}{\alpha^2} + \frac{2}{\alpha} \right)
\]  

(2.31)

It has been validated by numerical simulations that this drain current model for SG MOSFETs is accurate in all the three operation regions and properly describe the transition behaviors [67].
2.4 Charge and Capacitance Model for SG MOSFETs

In the last section, we have shown the analytic drain current model for SG MOSFETs. However, as have been mentioned previously, drain current model is not enough for circuit simulation. For AC and transient simulation, analytic and continuous expressions for three terminal charges, i.e., $Q_g$, $Q_s$, and $Q_d$, associated with gate, source, and drain, respectively, are needed to be developed.

For SG MOSFETs, we will also adopt Ward-Dutton linear charge partition method, which has been proven to be reasonable for DG MOSFETs [65]. Integrating the current continuity equation

$$dy = -\mu (2\pi R)Q_i dV/I_{ds}$$

leads to $y = L[f(\alpha) - f(\alpha_s)]/[f(\alpha_d) - f(\alpha_s)]$, where $f(\alpha)$ is defined by (2.31). With this result, we can calculate the terminal charges as

$$Q_g = 2\pi R \int_0^L Q_i(y) dy = (2\pi R)^2 \frac{\mu I_{ds}}{T_{ds}} \int_{\alpha_s}^{\alpha_d} Q_i^2(\alpha) \frac{dV}{d\alpha} d\alpha$$

$$= 8\pi \epsilon_s L kT q \frac{g(\alpha_d) - g(\alpha_s)}{f(\alpha_d) - f(\alpha_s)}$$

$$Q_d = -2\pi R \int_0^L y Q_i(y) dy = 8\pi \epsilon_s L kT q W_{sd}$$

$$Q_s = -2\pi R \int_0^L \left(1 - \frac{y}{L} \right) Q_i(y) dy = 8\pi \epsilon_s L kT q W_{ds} \quad (2.32)$$

where

$$W_{ij} = \frac{f(\alpha_i) [g(\alpha_j) - g(\alpha_i)] + [h(\alpha_j) - h(\alpha_i)]}{[f(\alpha_j) - f(\alpha_i)]^2} \quad (2.33)$$
and functions \( g(\alpha) \) and \( h(\alpha) \) are defined as

\[
g(\alpha) = \left( -\frac{1}{\alpha^2} + \frac{3}{\alpha} + \ln \alpha \right) + s \left( -\frac{2}{3\alpha^3} + \frac{2}{\alpha^2} - \frac{2}{\alpha} \right)
\]

\[
h(\alpha) = \left( -\frac{4}{3\alpha^3} + \frac{5}{2\alpha^2} + \frac{1}{\alpha} - \frac{1 - 3\alpha}{\alpha^2} \ln \alpha + \frac{(\ln \alpha)^2}{2} \right)
- s \left( \frac{3}{2\alpha^4} - \frac{43}{9\alpha^3} + \frac{9}{2\alpha^2} + \frac{2 - 6\alpha + 6\alpha^2}{3\alpha^3} \ln \alpha \right)
- s^2 \left( \frac{2}{5\alpha^5} - \frac{2}{\alpha^4} + \frac{10}{3\alpha^3} - \frac{2}{\alpha^2} \right)
\]

(2.34)

It is straightforward to obtain expressions for conductance and transconductance

\[
g_m = \frac{\partial I_{ds}}{\partial V_{gs}} \bigg|_{V_{ds}} = \mu \frac{8\pi \epsilon_{si} kT}{L q} \left( \frac{1}{\alpha_s} - \frac{1}{\alpha_d} \right)
\]

\[
g_{ds} = \frac{\partial I_{ds}}{\partial V_{ds}} \bigg|_{V_{gs}} = \mu \frac{8\pi \epsilon_{si} kT}{L q} \frac{1 - \alpha_d}{\alpha_d}
\]

(2.35)

Combining above results, we can finally derive the capacitance coefficients for SG MOSFETs. We only show the results of four linearly independent components, from which all the nine capacitance coefficients can be obtained directly based on the charge conservation law:

\[
C_{gs} = 8\pi \epsilon_{si} L \frac{1 - \alpha_s}{\alpha_s} \left\{ \frac{(1 - \alpha_s)}{\alpha_s} \left[ W_{sd} + W_{ds} \right] \right\}
\]

\[
C_{gd} = 8\pi \epsilon_{si} L \frac{1 - \alpha_d}{\alpha_d} \left\{ \frac{(1 - \alpha_d)}{\alpha_d} \left[ W_{sd} + W_{ds} \right] \right\}
\]

\[
C_{dg} = 8\pi \epsilon_{si} L \left\{ \frac{1 - \alpha_d}{\alpha_d} \left[ \frac{(1 - \alpha_d)}{\alpha_d} \left( W_{sd} + W_{ds} \right) \right] \right\}
\]

\[
+ \left( W_{ds} - W_{sd} \right) \frac{1}{\alpha_s} \left[ \frac{1}{\alpha_s} - \frac{1}{\alpha_d} \right]
\]

\[
C_{sd} = 8\pi \epsilon_{si} L \frac{1 - \alpha_d}{\alpha_d} \left( \frac{W_{ds} - W_{sd}}{f(\alpha_d) - f(\alpha_s)} \right)
\]

(2.36)
Till now, the analytic charge and capacitance coefficient model for SG MOSFETs has been derived. Compared with the model for DG MOSFETs [65], this model is more suitable for implementation in circuit simulation programs. In the model for DG, the resulting expressions for the terminal charges and capacitance coefficients are not composed of fundamental functions but numerical integrations, which are time-consuming to be carried out. But our model for SG has no numerical integrations involved in the final expressions any more and the most time-consuming parts are just logarithmic functions. It is worthy to mention that this model is non-source-referenced and automatically preserves the symmetry with respect to source-drain interchange. This property can be obviously observed in the final expressions of the terminal charges as well as the capacitance coefficients.

When $V_s = V_d$, specific expressions are needed to avoid numerical instability:

$$Q_g = -2Q_s = -2Q_d = 8\pi\epsilon_{si}L\frac{kT}{q} \frac{1 - \alpha_s}{\alpha_s} \quad (2.37)$$

$$C_{gg} = 2C_{gs} = 2C_{gd} = 2C_{sq} = 2C_{dq} = -6C_{sd} = -6C_{ds}$$

$$= 3C_{ss} = 3C_{dd} = 8\pi\epsilon_{si}L \frac{1 - \alpha_s}{2\alpha_s - \alpha_s^2 + 2s (1 - \alpha_s)} \quad (2.38)$$

Fig. 2.8 plots terminal charges as functions of drain voltage. In this case, $\alpha_s$ is small. When $V_{ds} = 0$, $Q_g = -2Q_s = -2Q_d \simeq 8\pi\epsilon_{si}LkT/q\alpha_s$. When the device enters saturation region, $Q_g \simeq -10Q_s/6 \simeq -10Q_d/4 \simeq 10 (8\pi\epsilon_{si}LkT/q\alpha_s)/15$ according to (2.32). These approximate relations are exactly the same with those for bulk MOSFETs [64], and are clearly reflected in Fig. 2.8.
Figure 2.8: Normalized terminal charges $Q_g$, $Q_s$, and $Q_d$ of an SG MOSFET as functions of $V_{ds}$. $8\pi \epsilon_{si} L (kT/q)(1 - \alpha_s)/\alpha_s$ is the normalization constant.

Figure 2.9: Normalized capacitance coefficients of an SG MOSFET as functions of $V_{gs}$ obtained from the analytic model (solid lines), in comparison with the 2-D numerical simulation results (open circles). The normalization constant is $4\pi \epsilon_{si} L/s$ ($V_{ds} = 2V$).
Figure 2.10: Normalized capacitance coefficients of an SG MOSFET as functions of $V_{gs}$ obtained from the analytic model (solid lines), in comparison with the 2-D numerical simulation results (open circles). The normalization constant is $4\pi \epsilon_{si} L / s$ ($V_{ds} = 1V$).

In Fig. 2.9, 2.10, 2.11, and 2.12, we have compared our capacitance coefficient model with numerical simulation results of a long channel SG MOSFET. Because we use cylindrical coordinate instead of Cartesian coordinate, the numerical simulation is essentially two-dimensional. Fig. 2.9, where $V_{ds} = 2V$, shows that the general expressions of the model are in complete agreement with numerical simulations. In this case, the device is always in saturation region, and it is reasonable to assume $\alpha_d = 1$. Under this approximation, we can easily obtain the simple relation between capacitance coefficients from (2.36): $C_{gg} = C_{gs}$, $C_{sg} = C_{ss}$, $C_{dg} = -C_{ds}$, and $C_{gd} = C_{sd} = C_{dd} = 0$. If we further consider the well-above threshold region, where $\alpha_s$ is small, we can even derive the simplified
Figure 2.11: Normalized capacitance coefficients of an SG MOSFET as functions of $V_{gs}$ obtained from the analytic model (solid lines), in comparison with the 2-D numerical simulation results (open circles). The normalization constant is $4\pi \epsilon_{si} L/s$ ($V_{ds} = 0V$).

approximate expressions for these capacitance coefficients as follow

$$C_{gg} \approx \frac{10}{6} C_{sg} \approx \frac{10}{4} C_{dg} \approx \frac{10}{15} \frac{4\pi \epsilon_{si} L}{s} \left(1 - \frac{\alpha_s}{s}\right)$$

(2.39)

Fig. 2.10 has proven our model to be accurate for the transition from saturation to non-saturation, whereas Fig. 2.11 indicates the validity of (2.38) for the special case of $V_s = V_d$. In Fig. 2.12, the capacitance coefficients from both analytic model and numerical simulation are compared as functions of $V_{ds}$. We can clearly observe the transition between two degenerate cases, which have been introduced above: one degeneracy is from source-drain-interchange symmetry at $V_{ds} = 0$; another is induced by saturation at the drain end.
Figure 2.12: Normalized capacitance coefficients of an SG MOSFET as functions of \( V_{ds} \) obtained from the analytic model (solid lines), in comparison with the 2-D numerical simulation results (open circles). The normalization constant is \( 4\pi \epsilon_{si} L/s \).

2.5 Summary

In this chapter, we have shown complete analytic models of drain current, terminal charges and capacitance coefficients for long channel symmetric DG and SG MOSFETs. Starting from the exact solution of Poisson’s equation, we derive the models from current continuity equation and Ward-Dutton charge partition without the need for charge sheet approximation. Inasmuch as the essential physics has been preserved, it has been verified by numerical simulations that all the operation and transition regions are properly described by continuous functions in the absence of non-physical fitting parameters. Furthermore, the models are non-source-referenced and preserves the source-drain symmetry inherently. With
all these favorable features, the models are ideal candidates of compact model for DC, AC, and transient circuit simulation of DG and SG MOSFETs.

The text of Chapter 2, in part, is a reprint of the material as it appears in “Analytic Charge Model for Surrounding-Gate MOSFETs” by Bo Yu, Wei-Yuan Lu, Huaxin Lu, and Yuan Taur, IEEE Transaction on Electron Devices, Mar 2007. The dissertation author was the primary investigator and author of this paper.

The text of Chapter 2, in part, is a reprint of the material as it appears in “Analytic Charge Model for Double-Gate and Surrounding-Gate MOSFETs” by Bo Yu, Huaxin Lu, Wei-Yuan Lu, and Yuan Taur, Proceedings of NSTI Nanotech, May 2007. The dissertation author was the primary investigator and author of this paper.
Explicit Solutions for DG and SG MOSFETs

3.1 General Method

In the previous chapter, we have presented continuous analytic models for DG and SG MOSFETs. Without charge sheet approximation, these two models are derived directly from the Pao-Sah integral for DG and SG MOSFETs with undoped (or lightly doped) silicon body. It has been validated by numerical simulations that these models can continuously cover all the three operation regions without the need for non-physical fitting parameters. However, a common problem of these models and the surface-potential-based models (SPBMs) for bulk MOSFETs [68] is to solve an implicit transcendental equation.

Iterative method is the most universal approach to deal with implicit equa-
tions because of high accuracy it can achieve, but it has some concomitant unfavor- 
able features such as computational inefficiency and sporadic exceptions (eg. 
divergence). Therefore, an explicit approximation with enough accuracy is always 
preferred, if it exists. Chen and Gildenblat have proposed an accurate explicit 
solution for the bulk MOSFET surface potential [69]. This result was then applied 
in the well-known PSP model, which is selected by the Compact Model Council 
as the new industry standard [44].

Recently, based on a reformulation of the previous model [67], Iniguez et al. 
have developed an explicit continuous drain current model for SG MOSFETs [70]. 
Although the model can catch the physical essentials and match two asymptotic 
behaviors, it is not mathematically accurate, especially in the transition region 
around the threshold voltage. Actually, as the power supply voltage is scaled 
down, this transition region becomes more and more important. Therefore, a 
highly accurate explicit model is in urgent need.

In this chapter, we will introduce explicit solutions for the implicit tran-
scendental equations of both DG and SG MOSFETs. Being quite different from 
[70], our explicit models are extremely accurate because high order mathematical 
corrections are used. They are accurate not only in terms of drain current but 
also in terms of the derivatives such as transconductance and transcapacitance. 
First of all, the general method to develop the explicit solutions is presented in 
this section.
As mentioned previously, an accurate explicit approximation for the bulk MOSFET surface potential was proposed in [69], and then was applied in SPBM s such as PSP model [44]. Although this approximation is considered as a superb result, it seems to be an ad hoc method which can only be used in this special case. However, we have found the underlying general method of this approximation after a careful study, and it will play the role of guideline in developing explicit models for DG and SG MOSFETs.

To solve the famous implicit equation of surface potential, a three-step method was adopted in [69].

1) Compose a continuous starting function as the initial approximation

The first step gives a rough estimation of surface potential as an explicit continuous function of gate voltage, quasi-Fermi potential, etc. It is the most important step and the determinant factor of the method. The function should not be too complicated, but must be close enough to the exact implicit solution. The feasibility of this general method lies on whether we can find a proper starting function. If the implicit equation can be largely simplified in the region well above or below threshold, the asymptotic limits can be easily obtained. Using appropriate smoothing function to joint two asymptotic behaviors, we may achieve an ideal initial approximation.
2) Modify the starting function with high order correction

The starting function is just a crude estimation and far from accurate. To achieve accuracy, a third order correction is used in [69] to modify the starting function. Although the third order correction in [69] is described in a concrete manner and applicable to the specific equation, it is easy to reorganize it into a more universal fashion. Assume we are going to solve the implicit equation \( f(x; a, b, c) = 0 \), where \( x \) is to be solved, and \( a, b, \) and \( c \) are independent variables. \( x \) is initially approximated by \( g(a, b, c) \). Then \( f \) can be expanded into Taylor series near \( g \) as

\[
\sum_{n=0}^{\infty} \frac{1}{n!} \frac{\partial^n f(x; a, b, c)}{\partial x^n} \bigg|_{x=g(a, b, c)} [x - g(a, b, c)]^n
\]  

(3.1)

if \( g(a, b, c) \) is close enough to the exact solution. Keeping the expansion to the third order, we will have

\[
f_{g0} + f_{g1}h + \frac{f_{g2}}{2} h^2 + \frac{f_{g3}}{6} h^3 + O(h^4) = 0
\]  

(3.2)

with \( h = x - g(a, b, c) \) and

\[
f_{gn} = \frac{\partial^n f(x; a, b, c)}{\partial x^n} \bigg|_{x=g(a, b, c)}
\]  

(3.3)

where \( n = 0, 1, 2, 3 \). Cubic equation has exact analytical solutions. However, considering \( h \) is a small quantity, the solution should be approximately simplified into a rational form for the purpose of efficiency. It is well known that arithmetic

\[\text{In [69], equation } (x_g - x)^2 - G^2[x - 1 + exp(x - x_n)] = 0 \text{ is transformed to } ln[(x_g - x)^2/G^2 - x + 1] - x + x_n = 0 \text{ to make the third order correction in the second step.}\]
operations are much faster than square root and cubic root. Following the steps in [69], we can obtain a possible rational solution for equation (3.2). The general form of the third order correction used in the second step of [69] is given by

$$h = -f_{g0} \frac{f_{g1}^3 - f_{g0}f_{g1}f_{g2}/2}{f_{g1}^4 - f_{g0}f_{g1}^2f_{g2} + f_{g0}f_{g2}^2/4 + f_{g0}f_{g1}f_{g3}/6}$$  \hspace{1cm} (3.4)$$

Actually, there are many choices of approximate rational solutions for (3.2). We compared several typical ones and chose the optimal one for our explicit models:

$$h = -\frac{f_{g0}}{f_{g1}} \left( 1 + \frac{f_{g0}f_{g2}}{2f_{g1}^2} + \frac{f_{g0}(3f_{g2}^2 - f_{g1}f_{g3})}{6f_{g1}^4} \right)$$  \hspace{1cm} (3.5)$$

Because the third order is enough for high accuracy, we do not involve higher orders for simplicity, but the second order correction may be an alternative. The optimal second order correction for our explicit models is found to be

$$h = -\frac{f_{g0}}{f_{g1}} \left( 1 + \frac{f_{g0}f_{g2}}{2f_{g1}^2} \right)$$  \hspace{1cm} (3.6)$$

which is an approximate solution of

$$f_{g0} + f_{g1}h + \frac{f_{g2}}{2} h^2 + O(h^3) = 0$$  \hspace{1cm} (3.7)$$

Here we would like to show why (3.5) is an approximate solution of equation (3.2) as well as why (3.6) is an approximate solution of equation (3.7). To obtain the solution of (3.2), one feasible idea is to assume

$$h = d + p_2d^2 + p_3d^3$$  \hspace{1cm} (3.8)$$

where $d = -f_{g0}/f_{g1}$ is exactly Newton’s first-order correction term. Substituting (3.8) into (3.2), we can rewrite it as

$$\left( f_{g1}p_2 + \frac{f_{g2}}{2} \right) d^2 + \left( f_{g1}p_3 + f_{g2}p_2 + \frac{f_{g3}}{6} \right) d^3 + O(d^4) = 0$$  \hspace{1cm} (3.9)$$
It is very obvious that the coefficients $p_2$ and $p_3$ must be chosen to cancel $d^2$ and $d^3$ terms:

\[
p_2 = -\frac{f_{g2}}{2f_{g1}} \\
p_3 = \frac{3f_{g2}^2 - f_{g1}f_{g3}}{6f_{g1}^2}
\]  

(3.10)

Therefore, (3.5) is an approximate rational solution of (3.2), and certainly (3.6) is an approximate rational solution of (3.7).

Now, $x$ has a new approximate expression

\[
u(a, b, c) = g(a, b, c) + h(a, b, c)
\]

(3.11)

which is much more accurate than the original $g(a, b, c)$. Fortunately, $f_{g1}$ will never be zero in the case of our explicit solutions for DG and SG MOSFETs.

3) Make another correction to improve accuracy

Although a rough estimation evolves into an accurate one through last step, another correction may be needed to further improve accuracy to reach a required level. To realize this objective, we apply one more step which is similar to the last one. For example, [69] employed a second order correction, which is exactly equivalent to the Halley’s irrational formula.

Finally, we have an ultra accurate approximation for $x$:

\[
v(a, b, c) = g(a, b, c) + h(a, b, c) + w(a, b, c)
\]

(3.12)
where $w(a, b, c)$ is the new correction term in this step. If $f(x; a, b, c)$ has infinite order of continuity as expected in the cases we concern, $v(a, b, c)$ will share this favorable feature which is desirable in circuit CAD tools.

### 3.2 Explicit Solution for DG MOSFETs

The implicit equation of DG MOSFET analytic model is given by equation (2.4), which is repeated here

$$\ln \beta - \ln (\cos \beta) + r \beta \tan \beta - F = 0 \quad (3.13)$$

where $r = \frac{2 \epsilon_{si} t_{ox}}{\epsilon_{ox} t_{si}}$, and

$$F = \frac{q (V_g - \Delta \phi - V)}{2kT} - \ln \left( \frac{2L_{Di}}{t_{si}} \right) \quad (3.14)$$

The above implicit equation must be solved to complete the model, however, it has no exact analytical solution and has to be solved approximately. There are two independent variables in the equation ($r$ and $F$). It is much more complicated than the implicit equation with single variable. Here, we apply the general method in Section 3.1 to this equation. Substituting $\beta = \arctan z$ into (3.13), we obtain

$$f (z; r, F) = \ln \left( \arctan z \sqrt{1 + z^2} \right) + rz \arctan z - F = 0 \quad (3.15)$$

Then $z$ can be computed in three steps.

1) The asymptotic behaviors of $z$ are

$$z = \exp F \quad (as \ z \to 0); \quad z = \frac{2F}{\pi r} \quad (as \ z \to \infty) \quad (3.16)$$
Use smoothing function to connect them:

\[ z_1 = \sqrt{\left(\frac{8}{\pi^2 r^2}\right)^2 + \left(\frac{4}{\pi r}\right)^2 \ln^2 (1 + e^{F/2}) - \frac{8}{\pi^2 r^2}} \]  

(3.17)

2) Compute

\[ \gamma_2 = \arctan z_1 \]

\[ \delta_2 = 1 / (1 + z_1^2) \]

\[ \eta_0 = \frac{1}{2} \ln \left(\frac{\gamma_2^2}{\delta_2}\right) + rz_1 \gamma_2 - F \]

\[ \eta_1 = \frac{\delta_2}{\gamma_2} + (1 + r) z_1 \delta_2 + r \gamma_2 \]

\[ \eta_2 = -\frac{\delta_2^2}{\gamma_2^2} - \frac{2z_1 \delta_2}{\gamma_2} + \left(1 + 2r - z_1^2\right) \delta_2^2 \]

\[ \eta_3 = \frac{2\delta_3^3}{\gamma_3^3} + \frac{6z_2 \delta_3^3}{\gamma_3^3} + \left(6z_2^2 - 2\right) \frac{\delta_3^3}{\gamma_3} + \left(2z_2^2 - 6 - 8r\right) z_1 \delta_3^3 \]

\[ z_2 = z_1 - \frac{\eta_0}{\eta_1} \left(1 + \frac{\eta_0 \eta_2}{2\eta_1^2} + \frac{\eta_0^2 (3\eta_2^2 - \eta_1 \eta_3)}{6\eta_1^4}\right) \]  

(3.18)

3) Compute

\[ \gamma_3 = \arctan z_2 \]

\[ \delta_3 = 1 / (1 + z_2^2) \]

\[ \lambda_0 = \frac{1}{2} \ln \left(\frac{\gamma_3^2}{\delta_3}\right) + rz_2 \gamma_3 - F \]

\[ \lambda_1 = \frac{\delta_3}{\gamma_3} + (1 + r) z_2 \delta_3 + r \gamma_3 \]

\[ \lambda_2 = -\frac{\delta_3^2}{\gamma_3^2} - \frac{2z_2 \delta_3^2}{\gamma_3} + \left(1 + 2r - z_2^2\right) \delta_3^2 \]

\[ \lambda_3 = \frac{2\delta_3^3}{\gamma_3^3} + \frac{6z_2 \delta_3^3}{\gamma_3^3} + \left(6z_2^2 - 2\right) \frac{\delta_3^3}{\gamma_3} + \left(2z_2^2 - 6 - 8r\right) z_2 \delta_3^3 \]

\[ z_3 = z_2 - \frac{\lambda_0}{\lambda_1} \left(1 + \frac{\lambda_0 \lambda_2}{2\lambda_1^2} + \frac{\lambda_0^2 (3\lambda_2^2 - \lambda_1 \lambda_3)}{6\lambda_1^4}\right) \]  

(3.19)
To achieve high accuracy, we choose the optimal third order correction term which is given by (3.5) in both the second and third steps. The above algorithm includes computation of one exponent, one square root, three logarithms, and two inverse tangents. It is worthy to mention that the minimum of $\partial f(z; r, F)/\partial z$ is greater than 0 in the range of $z > 0$ and $r > 0$, so this approximation will not be ruined by the possibility of $\eta_1$ or $\lambda_1$ approaching 0.

If we substitute $z_3$ back into equation (3.13), $f(z_3(r, F); r, F)$ is a function of $r$ and $F$, which is not equal to 0 any more, because $z_3$ is only an approximation but not exact solution. The final error $|f(z_3(r, F); r, F)|$ is a good criteria for the accuracy of our explicit solution. In Fig. 3.1, $|f|$ is plotted as a function of both $r$ and $F$. The ranges of $r$ and $F$ are carefully chosen to cover all the reasonable
cases. It is easy to observe that, for any specified \( r \), the largest error appears in the transition region around the threshold. This phenomenon is consistent with how we construct the approximate solution. The worst case in Fig. 3.1 has been found to be \(|f| \simeq 1.1 \times 10^{-11}\), which means our explicit solution is extremely accurate. Through the definition of \( F \), we can convert above error into error of gate voltage by multiplying \( 2kT/q \), which becomes \( 6 \times 10^{-13} \) V roughly, i.e., 0.6 pV. In other words, the error of \( z \) caused by our algorithm is less than the difference of \( z \) caused by a gate voltage shift of 0.6 pV.

Using \( z_3 \) as an approximation for \( z \), we can obtain the explicit solution of \( \beta \), which is given by \( \arctan(z_3) \). In Chapter 2, it has been shown that drain current and terminal charges as well as transconductances and transcapacitances of DG MOSFETs can be expressed as explicit functions of \( \beta_s \) and \( \beta_d \), which are \( \beta \) at the source and drain sides:

\[
\beta_s = \arctan z_3|_{V = V_s} \\
\beta_d = \arctan z_3|_{V = V_d}
\]  

(3.20)

Therefore, we have finally obtained an explicit continuous model for DG MOSFETs.
3.3 Explicit Solution for SG MOSFETs

The implicit equation of SG MOSFET analytic model is given by equation (2.27), which is repeated here

\[
\frac{1}{2} \ln (1 - \alpha) - \ln \alpha + s \frac{1 - \alpha}{\alpha} = G
\]  

(3.21)

where \( s = 2\epsilon_{si} \ln(1 + \frac{t_{ox}}{R})/\epsilon_{ox} \) and

\[
G = \frac{q (V_g - \Delta \phi - V)}{2kT} - \ln \left( \frac{2}{R} \sqrt{\frac{2\epsilon_{si}kT}{q^2n_i}} \right)
\]  

(3.22)

Again, we apply the general method in Section 3.1 to this equation. Substituting \( \alpha = 1/(z + 1) \) into (3.21), we obtain

\[
g(z; s, G) = \ln \left( \sqrt{z + z^2} \right) + sz - G = 0
\]  

(3.23)

Then \( z \) can be computed in three steps.

1) The asymptotic behaviors of \( z \) are

\[
z = \exp(2G) \text{ (as } z \to 0); \quad z = \frac{G}{s} \text{ (as } z \to \infty)
\]  

(3.24)

Use smoothing function to connect them:

\[
z_1 = \sqrt{\left( \frac{1}{2s^2} \right)^2 + \left( \frac{1}{s} \right)^2 \ln^2 (1 + e^G) - \frac{1}{2s^2}}
\]  

(3.25)

2) Compute

\[
\eta_0 = \frac{1}{2} \ln \left( z_1 + z_1^2 \right) + sz_1 - G
\]

\[
\eta_1 = \frac{1}{2z_1} + \frac{1}{2(1 + z_1)} + s
\]
\[ \eta_2 = -\frac{1}{2z_1^2} - \frac{1}{2(1 + z_1)^2} \]

\[ \eta_3 = \frac{1}{z_1^3} + \frac{1}{(1 + z_1)^3} \]

\[ z_2 = z_1 - \frac{\eta_0}{\eta_1} \left( 1 + \frac{\eta_0 \eta_2}{2\eta_1^2} + \frac{\eta_0^2 (3\eta_2^2 - \eta_1 \eta_3)}{6\eta_1^4} \right) \] (3.26)

3) Compute

\[ \lambda_0 = \frac{1}{2} \ln \left( z_2 + z_2^2 \right) + sz_2 - G \]

\[ \lambda_1 = \frac{1}{2z_2} + \frac{1}{2(1 + z_2)} + s \]

\[ \lambda_2 = -\frac{1}{2z_2^2} - \frac{1}{2(1 + z_2)^2} \]

\[ z_3 = z_2 - \frac{\lambda_0}{\lambda_1} \left( 1 + \frac{\lambda_0 \lambda_2}{2\lambda_1^2} \right) \] (3.27)

In the second step, we choose (3.5) which has the optimal performance among the third order correction terms. In the third step, even the optimal second order correction, i.e., (3.6) can provide extremely high accuracy. The above algorithm includes computation of one exponent, one square root, and three logarithms. It is worthy to mention that the minimum of \( \partial g(z; s, G) / \partial z \) is greater than 0 in the range of \( z > 0 \) and \( s > 0 \), so this approximation will not be ruined by the possibility of \( \eta_1 \) or \( \lambda_1 \) approaching 0.

If we substitute \( z_3 \) back into equation (3.23), \( g(z_3(s, G); s, G) \) is a function of \( s \) and \( G \), which is not exactly equal to 0 any more. The final error \( |g(z_3(s, G); s, G)| \) is a good criteria for the accuracy of our explicit solution. In Fig. 3.2, \( |g| \) is plotted as a function of both \( s \) and \( G \). The ranges of \( s \) and \( G \) are carefully chosen to cover all the reasonable cases. Similarly, it is easy to observe that, for any specified
Figure 3.2: The absolute error $|g|$ as a function of $s$ and $G$ for implicit equation of SG MOSFETs.

$s$, the largest error appears in the transition region around the threshold. The worst case in Fig. 3.2 has been found to be $|g| \simeq 3.2 \times 10^{-13}$, which means our explicit solution for SG MOSFETs is extremely accurate, and even more accurate than that for DG MOSFETs. Through the definition of $G$, we can convert above error into error of gate voltage by multiplying $2kT/q$, which becomes $1.7 \times 10^{-14}$ V roughly, i.e., 17 fV. In other words, the error of $z$ caused by our algorithm is less then the difference of $z$ caused by a gate voltage shift of 17 fV.

Using $z_3$ as an approximation for $z$, we can obtain the explicit solution of $\alpha$, which is given by $1/(z_3+1)$. In Chapter 2, it has been shown that drain current and terminal charges as well as transconductances and transcapacitances of SG MOSFETs can be expressed as explicit functions of $\alpha_s$ and $\alpha_d$, which are $\alpha$ at the
source and drain sides:

\[
\alpha_s = \left. \frac{1}{z_3 + 1} \right|_{V=V_s}
\]
\[
\alpha_d = \left. \frac{1}{z_3 + 1} \right|_{V=V_d}
\]

(3.28)

Therefore, we have finally obtained an explicit continuous model for SG MOSFETs.

### 3.4 Accuracy of the Explicit Solutions

In this section, we will discuss the accuracy of our explicit solutions in detail. We will first consider SG MOSFETs model, because we can compare our model with the previously proposed explicit model for SG MOSFETs [70]. In [70], Iniguez et al. have put forward an explicit approximate solution for \(Q\), which is

![Figure 3.3: The absolute error \(|g(Q/Q_0; s, G)|\) as a function of \(s\) and \(G\). The expression for \(Q\) is given in [70].](image)
defined as the mobile charge per unit gate area. It is straightforward to find the relation between $Q$ and $z$: $z = Q/Q_0$, where $Q_0 = (4\epsilon_{si}/R)(kT/q)$. Therefore we can similarly plot $|g(Q/Q_0; s, G)|$ as a function of $s$ and $G$, as shown in Fig. 3.3. We can see that the approximation for $Q$ is not accurate, especially in the transition region. The worst case in Fig. 3.3 has been found to be $|g| \approx 0.52$, which corresponds to 27 mV in terms of gate voltage. Compared with 17 fV for our model, 27 mV is excessively large. Furthermore, lower accuracy does not mean higher efficiency. The explicit expression for $Q$ includes two exponents, two square roots, and three logarithms. Although it has more compact form, our algorithm is not expected to be computationally more complex for the lack of one exponent and one square root.

![Figure 3.4: $I_{ds}-V_{gs}$ characteristics obtained from our explicit model (solid curves) compared with those from numerical method (dots) as well as explicit model in [70] (dashed curves). The same currents are plotted on both logarithmic (left) and linear (right) scales.](image)
Figure 3.5: Transconductances obtained from our explicit model (solid curves) compared with those from numerical method (dots) as well as explicit model in [70] (dashed curves).

Till now, we only show the accuracy from viewpoint of $|g|$, but it is not intuitive enough. A more direct way to understand the accuracy of our explicit solution is to compare the results of drain current or terminal charges calculated by numerical method and our explicit solution. In Fig. 3.4, we show one example of $I_{ds} - V_{gs}$ characteristics. The numerical results are obtained by solving equation (3.23) with second-order Newton iteration method. Actually, it is shown that our model provides excellent accuracy everywhere. The explicit model in [70] also has a good performance well above or below the threshold, however in the transition region around the threshold, it does not agree with the numerical results very well.

The accuracy of drain current or terminal charges is always easy to achieve,
Figure 3.6: Transcapacitances obtained from our explicit model (solid curves) compared with those from numerical method (dots) as well as explicit model in [70] (dashed curves).

but it is not sufficient for circuit simulation. Partial derivatives such as drain conductance, transconductance, and transcapacitance are required for AC and transient circuit simulation, and these derivatives are more sensitive to the errors in the intermediate parameters. Therefore, although in the drain current plot we cannot observe the prominent advantage of our explicit model over that in [70], but it is expected that our model can exceed quite a bit in derivative plots, which has been demonstrated in Fig. 3.5 and 3.6. Although [70] did not provide transconductance and transcapacitance models, we can straightforwardly transform $Q$ in [70] into $\alpha$, and use expressions in Chapter 2 to calculate transconductance and transcapacitance. It is indicated that the explicit model in [70] has obvious deviation from numerical results, whereas our model is accurate not only in terms of drain current
Figure 3.7: \( I_{ds} - V_{gs} \) characteristics obtained from our explicit model (solid curves) compared with those from numerical method (dots). The same currents are plotted on both logarithmic (left) and linear (right) scales.

and terminal charges but also their derivatives.
Figure 3.9: Transcapacitances obtained from our explicit model (solid curves) compared with those from numerical method (dots).

We can obtain similar results for DG MOSFETs. In Fig. 3.7, 3.8, and 3.9, our explicit solutions of drain current, transconductance, and transcapacitance, are compared with those calculated from numerical method, respectively. Excellent agreement has been achieved again.

3.5 Summary

In this chapter, we have presented explicit models for both DG and SG MOSFETs, or equivalently, we have provided the explicit solutions for the intermediate parameters used in DG and SG MOSFETs models introduced in the previous chapter. In addition to the favorable features inherited from those models,
our explicit models manifest their advantages in some other aspects. In the explicit models, drain current, terminal charges, drain conductance, transconductance, and transcapacitances can be expressed as explicit functions of applied voltages and structural parameters, and even these expressions have infinite order of continuity. Because high order mathematical corrections have been used, our models can achieve extreme accuracy over a wide range of device parameters and also in all the operation regions, especially including the transition region around the threshold. They are accurate not only in terms of drain current and terminal charges but also their derivatives, which involve drain conductance, transconductance, and transcapacitances.

The text of Chapter 3, in part, is a reprint of the material as it appears in “Explicit Continuous Models for Double-Gate and Surrounding-Gate MOSFETs” by Bo Yu, Huaxin Lu, Minjian Liu, and Yuan Taur, IEEE Transaction on Electron Devices, Oct 2007. The dissertation author was the primary investigator and author of this paper.
Unified Analytic Drain Current Model for MG MOSFETs

4.1 Basic Idea

Analytic compact models for DG and SG MOSFETs have been discussed in Chapter 2. Due to the inherent symmetry of DG and SG MOSFETs, Poisson’s equation can be reduced to ordinary differential equations (one-dimensional) in Cartesian and cylindrical coordinates, respectively, based on the gradual channel approximation. These ordinary differential equations have analytical solutions, therefore analytic models can be obtained for these highly symmetric structures.

However, other MG MOSFETs have no analytical results even with gradual channel approximation. In triple-gate (TG), triple-plus-gate (TPG), and quadruple-gate (QG) MOSFETs [17], Poisson’s equation cannot be reduced to 1D equation
for lack of symmetry, therefore modeling these MG structures is challenging, and more approximations are required besides gradual channel approximation. In this
chapter, we will show a unified MG MOSFET model developed from those for DG and SG. The basic idea is as follows: QG and SG MOSFETs are both devices with gate all around, therefore it is possible to modify SG model to work for QG structures; TG device can be approximated by the average of QG and DG devices; from the TG MOSFET model, we are able to further develop the model for TPG devices, such as Π-gate and Ω-gate MOSFETs. By following the above idea, we first propose analytic models for inversion charge density as a function of gate voltage in MG MOS capacitors, which have been verified to be accurate by numerical simulations. Then the Pao-Sah integral [62] is used to obtain a unified analytic drain-current model for MG MOSFETs. All of the devices shown in Fig. 4.1 are covered by our unified MG MOSFET model.

To implement the basic idea, we need to use reasonable approximations consistent with the numerical simulation results from TCAD tools. MOS capacitor behaviors can be captured through 2D simulation of the cross section perpendicular to channel length. Pao-Sah integral can then be carried out to obtain the long-channel drain current as

\[ I_{ds} = \frac{\mu}{L} \int_{-V_s}^{V_d} Q_m(V_g, V) dV \]  

which has been used to derive the analytic models for DG and SG MOSFETs. Here, \( Q_m \) is defined as the total mobile charge per unit gate length along the channel direction. For DG and SG, we have \( Q_m = WQ_i \) and \( Q_m = 2\pi RQ_i \), respectively. We consider nMOSFETs, and \( V \) is the electron quasi-Fermi potential. For MG
MOSFETs, the mobile charge per unit gate length $Q_m(V_g, V)$ is essentially a single-variable function of $V_g - V$, and can be written as $Q_m(V_g - V)$.

### 4.2 Model for QG MOSFETs

We will first study QG MOS capacitors which are two terminal devices, so their characteristics can be easily described: the mobile charge $Q_m$ (or capacitance) is only a single-variable function of voltage $V_g - V$, where $V$ is the n+ source (and drain) potential (later on $V$ is allowed to vary between $V_s$ and $V_d$ for MOSFETs).

The function $Q_m$ has two asymptotic behaviors. In the subthreshold region, volume inversion takes place for undoped (or lightly doped) silicon body [63], which is the default assumption throughout this paper. Therefore, $Q_m$ is proportional to the area of the cross section of silicon body, and the dependence on $V_g - V$ is exponential with an inverse subthreshold slope of 60 mV/decade [63]. Whereas well above threshold, $Q_m$ is proportional to the perimeter of the cross section of silicon body in the limit of $t_{ox} \to 0$, and the dependence on $V_g - V$ is linear.

Because of the above asymptotic behaviors, we are able to model a QG MOS capacitor from that of SG. For a QG MOS capacitor with structure parameters $t_{si}$, $H$, and $t_{ox}$, as indicated in Fig. 4.1 (c), we consider a SG MOS capacitor with a radius $R = \sqrt{t_{si}H/\pi}$ so that they are in possession of the same area of silicon body. The subthreshold behaviors of these two capacitors are then identical, as shown in Fig. 4.2. Above threshold, $Q_m(QG)/Q_m(SG)$ is given by $(t_{si} + H)/(\pi R)$.
Figure 4.2: $Q_m(QG)$ and $Q_m(SG)$ as functions of $V_g - V$. $Q_m(SG)$ is calculated from the analytic model, whereas $Q_m(QG)$ is obtained from 2-D TCAD numerical simulations. Radius $R$ of SG is chosen to match subthreshold behavior of QG.

in the limit of $t_{ox} \to 0$, which is exactly the ratio of perimeters of silicon body. Because $R = \sqrt{t_{si}H/\pi}$, the above ratio $(t_{si} + H)/(\pi R) \geq 2/\sqrt{\pi} (\approx 1.13)$, and equality holds in the special case of $t_{si} = H$. In Fig. 4.3, $Q_m$ ratios are plotted as functions of $\alpha$, which has one-to-one correspondence with $V_g - V$ through equation (2.27). The curves are more or less linear from 1 at $\alpha \to 1$ (subthreshold region) to $1 + C_1$ at $\alpha \to 0$ (well-above threshold). Therefore, for the entire range, we can approximate $Q_m$ ratio by

$$Q_m(QG)/Q_m(SG) \simeq 1 + C_1(1 - \alpha) \quad (4.2)$$

Here $C_1 > 0$ is a parameter that depends on the device structure. In the limiting case of $t_{ox} \to 0$, $C_1$ is given by $(t_{si} + H)/\sqrt{\pi t_{si}H} - 1$. For finite $t_{ox}$, $C_1$ can be extracted from the $Q_m$ ratio versus $\alpha$ curve generated by 2-D numerical simulation.
For the case of square cross section \((H = t_{si})\), considered in Fig. 4.3 (a), \(C_1\) is mainly a single-variable function of \(t_{ox}/t_{si}\). We have generalized an empirical
formula for \( C_1 \) based on extensive simulation results:

\[
C_1 = 0.130 - 0.220 \frac{t_{ox}}{t_{si}}, \quad \text{for} \quad \frac{t_{ox}}{t_{si}} < 1/3 \quad (4.3)
\]

For cases where \( H \neq t_{si} \) shown in Fig. 4.3 (b), \( C_1 \) can be treated as a function of dimension ratios \( t_{ox}/t_{si} \) and \( t_{ox}/H \). The linear approximation equation (4.2) works well (less than 3% error) for \( 1/2 < \frac{t_{si}}{H} < 2 \). For QG MOSFETs of aspect ratios much different from unity, a quadratic approximation \( 1 + A(1 - \alpha) + B(1 - \alpha)^2 \) gives better agreement. The focus of this chapter is on the linear approximation which applies for the technically relevant cases of near square cross sections.

Multiplying \( Q_m(SG) \) by \( Q_m \) ratio, we immediately obtain the analytical expression for \( Q_m(QG) \) as

\[
Q_m(QG) = 8\pi \varepsilon_{si} \frac{kT}{q} \frac{1 - \alpha}{\alpha} [1 + C_1 (1 - \alpha)] \quad (4.4)
\]

Figure 4.4: Mobile charge \( Q_m \) of QG MOS capacitors as functions of \( V_g - V \) obtained from the analytic model (solid lines) in comparison with the 2-D numerical simulation results (open circles).
Figure 4.5: Percentage error of mobile charge $Q_m$ between the analytic model and 2-D numerical simulation results of QG MOS capacitors as functions of $V_g - V$.

Figure 4.6: Gate capacitance $C_g$ of QG MOS capacitors as functions of $V_g - V$ obtained from the analytic model (solid lines) in comparison with the 2-D numerical simulation results (open circles).
where \(\alpha\) is related to \(V_g - V\) by equation (2.27) with \(R = \sqrt{t_{si}H/\pi}\). In Fig. 4.4, we compare this analytic model with numerical simulation results. Two examples are used to illustrate the validity of the model: one with \(H = 25\) nm, \(t_{si} = 20\) nm, \(t_{ox} = 1\) nm, and \(C_1\) is extracted to be 0.129; another with \(H = t_{si} = 10\) nm, \(t_{ox} = 1\) nm, and \(C_1\) is extracted, or directly calculated from the empirical formula (4.3), to be 0.108. The percentage error between the analytic model and simulation results is shown in Fig. 4.5, which demonstrates the accuracy of the model. To further examine its accuracy, we show the first-order and second-order derivatives of \(Q_m\) with respect to \(V_g - V\) in Fig. 4.6 and 4.7, respectively. The first-order derivative is the gate capacitance \(C_g\), and the second-order derivative is \(dC_g/d(V_g - V)\). In both figures, the analytic model agrees very well with the simulation results. The
humps of $dC_g/d(V_g - V)$ curves correspond to the onset of strong inversion. Each curve in Fig. 4.7 exhibits a single hump, indicating that the corner effect is smeared out [71, 72], i.e., the early inversion of corners does not generate another distinctive threshold.

### 4.3 Model for TG and TPG MOSFETs

For TG MOS capacitors, we need to approximate

$$Q_m(TG) \approx \frac{Q_m(DG) + Q_m(QG)}{2} \tag{4.5}$$

Here DG and QG devices have the same structure parameters as the TG device. In other words, if the silicon dimensions of the TG device to be modeled are $t_{si}$ and $H$, then the QG on the right side of equation (4.5) uses the same $t_{si}$ and $H$, whereas the DG uses the same $t_{si}$ and $W = H$. This approximation makes sense from three different aspects. First of all, both sides of (4.5) have identical subthreshold behavior because of the same cross-sectional area of silicon body $t_{si} \times H$. Second, they have the same gated width $t_{si} + 2H$ for inversion. Finally, they have the same number of corners. The last two factors ensure that both sides of equation (4.5) have identical above-threshold behavior. The validity of the approximation in the transition region needs to be checked by 2-D numerical simulations. Based on the above approximation, we obtain the analytic model for TG MOS devices as follows

$$Q_m(TG) = 4\pi \varepsilon_{si} \frac{kT}{q} \frac{1 - \alpha}{\alpha} [1 + C_1 (1 - \alpha)] + 4\varepsilon_{si} \frac{H kT}{t_{si}^2} \frac{\beta}{q} \tan \beta \tag{4.6}$$
where \( \alpha \) is related to \( V_g - V \) by equation (2.27), and \( \beta \) is related to \( V_g - V \) by equation (2.4).

A TPG device is between TG and QG devices. Therefore, to model TPG MOS capacitors, we can use this relation:

\[
Q_m(TPG) \approx \frac{1 - C_2}{2} Q_m(DG) + \frac{1 + C_2}{2} Q_m(QG) \tag{4.7}
\]

where \( C_2 \) is a parameter to be extracted from 2-D numerical simulation results.

Finally, the analytical expression of \( Q_m(TPG) \) is given by

\[
Q_m(TPG) = (1 + C_2) 4\pi\epsilon_{si} \frac{kT}{q} \frac{1 - \alpha}{\alpha} [1 + C_1 (1 - \alpha)]
+ (1 - C_2) 4\epsilon_{si} \frac{H kT}{q} \beta \tan \beta \tag{4.8}
\]

where \( \alpha \) is related to \( V_g - V \) by equation (2.27), and \( \beta \) is related to \( V_g - V \) by equation (2.4).

Figure 4.8: Mobile charge \( Q_m \) of TG and TPG MOS capacitors as functions of \( V_g - V \) obtained from the analytic model (solid lines) in comparison with the 2-D numerical simulation results (open circles).
Figure 4.9: Percentage error of mobile charge $Q_m$ between the analytic model and 2-D numerical simulation results of TG and TPG MOS capacitors as functions of $V_g - V$.

Figure 4.10: Gate capacitance $C_g$ of TG and TPG MOS capacitors as functions of $V_g - V$ obtained from the analytic model (solid lines) in comparison with the 2-D numerical simulation results (open circles).
Figure 4.11: The first order derivative of $C_g$, i.e., the second order derivative of $Q_m$ of TG and TPG MOS capacitors as functions of $V_g - V$ obtained from the analytic model (solid lines) in comparison with the 2-D numerical simulation results (open circles).

In Fig. 4.8, we compare the analytic model for TG and TPG MOS capacitors with numerical simulation results. Three examples are used to illustrate the validity of the model. The first is an TG MOS capacitor with $H = t_{si} = 10nm$, $t_{ox} = 1nm$, and $C_1 = 0.108$. The second is an Π-gate MOS capacitor with $H = t_{si} = 20nm$, $t_{ex} = 10nm$, $t_{ox} = 1nm$, and $C_1 = 0.119$ according to (4.3), and $C_2$ is extracted to be 0.218. The third is an Ω-gate MOS capacitor with $H = 25nm$, $t_{si} = 20nm$, $t_{ex} = 10nm$, $t_{es} = 16nm$, $t_{ox} = 1nm$, and $C_1 = 0.129$, and $C_2$ is extracted to be 0.440. The percentage error between the analytic model and simulation results is shown in Fig. 4.9, which demonstrates the accuracy of the model. To further examine its accuracy, we show the first-order and second-order derivatives of $Q_m$ with respect to $V_g - V$ in Fig. 4.10 and 4.11, respectively. In
Table 4.1: Structure parameters, $C_1$, and $C_2$ for different devices

<table>
<thead>
<tr>
<th>Device</th>
<th>$H$</th>
<th>$t_{si}$</th>
<th>$t_{ox}$</th>
<th>$C_1$</th>
<th>$C_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DG</td>
<td>$W$</td>
<td>$t_{si}$</td>
<td>$t_{ox}$</td>
<td>0</td>
<td>−1</td>
</tr>
<tr>
<td>SG</td>
<td>$\sqrt{\pi}R$</td>
<td>$\sqrt{\pi}R$</td>
<td>$t_{ox}$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>QG</td>
<td>$H$</td>
<td>$t_{si}$</td>
<td>$t_{ox}$</td>
<td>$C_1$</td>
<td>1</td>
</tr>
<tr>
<td>TG</td>
<td>$H$</td>
<td>$t_{si}$</td>
<td>$t_{ox}$</td>
<td>$C_1$</td>
<td>0</td>
</tr>
<tr>
<td>TPG</td>
<td>$H$</td>
<td>$t_{si}$</td>
<td>$t_{ox}$</td>
<td>$C_1$</td>
<td>$C_2$</td>
</tr>
</tbody>
</table>

both figures, the analytic model agrees very well with the simulation results. Each curve in Fig. 4.11 exhibits a single hump, indicating that there are no multiple threshold voltages.

From the analytic models for the mobile charge density of MG MOS capacitors, we can carry out the Pao-Sah integral to obtain the drain current model for MG MOSFETs.

4.4 Unified Drain Current Model

The expression of $Q_m(TPG)$ is most general because it can cover all kinds of different MG MOS capacitors by choosing different values for $C_1$ and $C_2$. We use it as an example to show how we can derive the drain current model. There are two terms in $Q_m(TPG)$: one is a function of the intermediate parameter $\alpha$,
the other is a function of $\beta$. They can be separated as follows.

$$I_{ds} = \frac{\mu}{L} \int_{V_s}^{V_d} Q_m(TPG) dV$$

$$= 4\pi \varepsilon_{si} \frac{kT}{q} \frac{\mu}{L} (1 + C_2) \int_{\alpha_s}^{\alpha_d} [1 + C_1 (1 - \alpha)] \frac{1 - \alpha}{\alpha} \frac{dV}{d\alpha} d\alpha$$

$$+ 4\varepsilon_{si} \frac{H}{t_{si}} \frac{kT}{q} \frac{\mu}{L} (1 - C_2) \int_{\beta_s}^{\beta_d} \beta \tan \frac{\beta}{\beta} \frac{dV}{d\beta} d\beta$$

(4.9)

where $\alpha_s$ and $\alpha_d$ are solutions to equation (2.27) corresponding to $V = V_s$ and $V = V_d$, respectively, with $R = \sqrt{t_{si}H/\pi}$; $\beta_s$ and $\beta_d$ are solutions to equation (2.4) corresponding to $V = V_s$ and $V = V_d$, respectively. The integrals in equation (4.9) can be carried out in closed forms to yield the final expression of drain current

$$I_{ds} = \mu \frac{4\pi \varepsilon_{si}}{L} \left( \frac{kT}{q} \right)^2 (1 + C_2) [f(\alpha) + C_1 d(\alpha)]_{\alpha_s}^{\alpha_d}$$

$$+ \mu \frac{4\varepsilon_{si} H}{L t_{si}} \left( \frac{kT}{q} \right)^2 (1 - C_2) [p(\beta)]_{\beta_s}^{\beta_d}$$

(4.10)

where functions $p$ and $f$ have been defined previously with equations (2.10) and (2.31), and $d$ is defined as

$$d(\alpha) = \left( -\frac{2}{\alpha} - 3 \ln \alpha + \alpha \right) + s \left( -\frac{1}{\alpha^2} + \frac{4}{\alpha} + 2 \ln \alpha \right)$$

(4.11)

The unified analytic drain current model can be applied to different MG MOSFETs by substituting different $C_1$, $C_2$, and structure parameters into the expression, as listed in Table I.
4.5 Summary

In conclusion, we have presented a unified analytic drain current model for MG MOSFETs, including DG, SG, QG, TG, and TPG MOSFETs. The model is based on previous analytic DG and SG MOSFET models [48, 67], and thus inherits most of the favorable features from those models, such as source-drain-interchange symmetry, absence of charge sheet approximation, and coverage of all the operation regions by one continuous function. The accuracy of the model is shown by comparing the inversion charge density as a function of gate voltage to 2-D numerical simulation results. The drain current is then obtained by Pao-Sah integral of the inversion charge density. The analytic drain current model developed serves as a core long-channel model that can be expanded for compact modeling of various kinds of MG MOSFETs.

The text of Chapter 4, in part, is a reprint of the material as it appears in “A Unified Analytic Drain-Current Model for Multiple-Gate MOSFETs” by Bo Yu, Jooyoung Song, Yu Yuan, Wei-Yuan Lu, and Yuan Taur, IEEE Transaction on Electron Devices, Aug 2008. The dissertation author was the primary investigator and author of this paper.
Modeling of Short-Channel Effects in DG and SG MOSFETs

5.1 Scale Length Theory

We only consider long-channel MG MOSFETs in Chapter 2–4, while in this chapter, we will put emphasis on the modeling of short-channel effects (SCEs) in DG and SG MOSFETs. Based on various assumptions, two-dimensional (2-D) analysis of Poisson’s equation for short-channel DG and SG MOSFETs has been previously investigated. However, most of these analytical models for SCEs assume one-dimensional (1-D) Poisson’s equation in the oxide region, i.e., ignore any lateral variation of the oxide field in the source/drain direction [73]-[79].

Recently, a two-region analytical model for SCEs in DG MOSFETs [80, 81] has been developed by following the generalized scale length theory [82]. The
model solves 2-D Poisson’s equation in both the silicon and insulator regions consistent with the 2-D boundary conditions at the silicon/insulator interface. Therefore it can be applied to arbitrary silicon and insulator thicknesses and dielectric constants, including high-κ gate dielectrics. The generalized scale length for SG MOSFETs was also derived several years before [83]. We have developed a two-region analytical model for SCEs in SG MOSFETs based on the generalized scale length. In this section, we will first give a thorough discussion on the scale length theory. The analytical models for SCEs in DG and SG MOSFETs will be shown in detail in the next two sections.

Suzuki et al. [73] have proposed a scaling theory for DG MOSFETs. The natural length for fully depleted DG MOSFETs (shown in Fig. 5.1) has been
Figure 5.2: Schematic diagram of an SG MOSFET.

derived to be

$$\lambda_N = \frac{\sqrt{\epsilon_{si}}}{2\epsilon_{ox}} \left( 1 + \frac{\epsilon_{ox} t_{si}}{4\epsilon_{si} t_{ox}} \right) t_{si} t_{ox}$$  \hspace{1cm} (5.1)$$

Following the same methodology, the natural length for fully depleted SG MOSFETs (shown in Fig. 5.2) can also be obtained, which is given by [76]

$$\lambda_N = \sqrt{\frac{2\epsilon_{si} R^2 \ln \left( 1 + \frac{t_{ox}}{R} \right) + \epsilon_{ox} R^2}{4\epsilon_{ox}}}$$ \hspace{1cm} (5.2)$$

Analysis of numerical simulation results indicates that the effective channel length of a transistor should be larger than $\sim 5$ times the natural length to maintain electrostatic integrity and to produce a reasonable subthreshold behavior [73, 76].

The scaling theory [73, 76] and the analytical SCE model [74, 77] for DG and SG MOSFETs based on the concept of natural length are successful to a certain extent. However, in this approach, the electrostatic potential in the channel is assumed to be a parabolic function along the transverse direction without a firm
physical basis. Besides, this approach applies the 1-D boundary condition at the silicon/oxide interface, which ignores any lateral variation of the oxide field in the source/drain direction. Significant error can occur in devices with a relatively thick insulator, especially those with high-\(\kappa\) dielectrics. Other approaches \([75, 78, 79]\) do not assume the parabolic potential profile, but still use 1-D Poisson’s equation in the insulator region.

To address the issue of 2-D effects in the gate insulator, a more generalized concept of scale length has been proposed recently \([82]\). The scale length theory is based on a two-region (silicon and insulator) model by matching the 2-D boundary conditions at the silicon/insulator interface. It can apply to arbitrary semiconductor and insulator thicknesses and dielectric constants.

We consider an undoped (or lightly-doped) symmetric DG MOSFET (Fig. 5.1) in the subthreshold region, where both the fixed and mobile charge densities are negligible. Therefore Poisson’s equation takes the following form in both the silicon and insulator regions:

\[
\frac{\partial}{\partial x} \left( \epsilon(x) \frac{\partial \psi}{\partial x} \right) + \frac{\partial}{\partial y} \left( \epsilon(x) \frac{\partial \psi}{\partial y} \right) = 0 \quad (5.3)
\]

where the electrostatic potential \(\psi(x, y)\) is defined as the intrinsic potential at a point \((x, y)\) with respect to the Fermi level in the n+ source, and \(\epsilon(x)\) is the dielectric constant of silicon or the insulator. By applying the superposition principle, the electrostatic potential can be written as

\[
\psi(x, y) = v(x) + u_L(x, y) + u_R(x, y) \quad (5.4)
\]
where \( v(x) \) is the 1-D solution to Poisson’s equation satisfying gate boundary condition. \( u_L \) and \( u_R \) are solutions to Poisson’s equation and satisfy the source and drain boundary conditions, respectively. For example, \( u_L \) is zero on the gate and drain boundaries, but \( v + u_L \) satisfies the source boundary condition. Similarly, \( u_R \) is zero on the gate and source boundaries, but \( v + u_R \) satisfies the drain boundary condition [82, 84]. \( u_L \) can be written as series in the form

\[
\begin{align*}
&u_L(x, y) = \left\{ \begin{array}{l}
\sum_{n=1}^{\infty} b_{2n} \sin \left[ k_n \left( x + \frac{t_{si}}{2} + t_{ox} \right) \right] \frac{\sinh[k_n(L-y)]}{\sinh(k_nL)}, &-\frac{t_{si}}{2} - t_{ox} \leq x < -\frac{t_{si}}{2} \\
\sum_{n=1}^{\infty} b_{1n} \cos \left[ k_n x \right] \frac{\sinh[k_n(L-y)]}{\sinh(k_nL)}, &-\frac{t_{si}}{2} \leq x \leq \frac{t_{si}}{2} \\
\sum_{n=1}^{\infty} b_{2n} \sin \left[ k_n \left( \frac{t_{si}}{2} + t_{ox} - x \right) \right] \frac{\sinh[k_n(L-y)]}{\sinh(k_nL)}, &\frac{t_{si}}{2} < x \leq \frac{t_{si}}{2} + t_{ox}
\end{array} \right.
\end{align*}
\]

(5.5)

With different coefficients, we can write similar series expression for \( u_R \) with \( L - y \) replaced by \( y \). The dielectric boundary conditions at the silicon/insulator interface require \( u_{L,R} \) and \( \epsilon(\partial u_{L,R}/\partial x) \) to be continuous at \( x = \pm t_{si}/2 \). Taking the ratio of these two boundary conditions finally yields an implicit equation for the eigenvalues \( k_n \):

\[
\tan (k_n t_{ox}) \tan (k_n t_{si}/2) = \frac{\epsilon_{ox}}{\epsilon_{si}}
\]

(5.6)

Except in extremely short MOSFETs, the \( u_{L,R} \) series decay much too fast to have a significant effect on the potential, therefore the most important terms are the lowest order ones. The generalized scale length \( \lambda \) is defined by \( \lambda = \pi/k_1 \), where \( k_1 \) is the smallest solution of the eigenvalue equation.

The scale length for SG MOSFETs can also be derived [83] with Bessel functions for the case of cylindrical coordinates [85]. Fig. 5.2 shows the schematic
diagram of an undoped (or lightly doped) SG MOSFET. In the subthreshold region, both the fixed and mobile charge densities are negligible, therefore Poisson’s equation takes the following form in both the silicon and insulator regions:

\[
\frac{1}{\rho} \frac{\partial}{\partial \rho} \rho \left( \epsilon (\rho) \frac{\partial \psi}{\partial \rho} \right) + \frac{\partial}{\partial y} \left( \epsilon (\rho) \frac{\partial \psi}{\partial y} \right) = 0
\] (5.7)

where the electrostatic potential \( \psi(\rho, y) \) is defined as the intrinsic potential at a point \((\rho, y)\) with respect to the Fermi level in the n+ source, and \( \epsilon (\rho) \) is the dielectric constant of silicon or the insulator. By applying the superposition principle, the electrostatic potential can be written as

\[
\psi (\rho, y) = v(\rho) + u_L (\rho, y) + u_R (\rho, y)
\] (5.8)

where \( v(\rho) \) is the 1-D solution to Poisson’s equation satisfying gate boundary condition. \( u_L \) and \( u_R \) are solutions to Poisson’s equation and satisfy the source and drain boundary conditions, respectively. \( u_L \) can be written as series in the form

\[
u_L(\rho, y) = \begin{cases} 
\sum_{n=1}^{\infty} b_{1n} J_0 (k_n \rho) \frac{\sinh[k_n(L-y)]}{\sinh(k_n L)}, & 0 \leq \rho \leq R \\
\sum_{n=1}^{\infty} b_{2n} \left[ J_0 (k_n \rho) - \frac{J_0(k_n R_i)}{Y_0(k_n R_i)} Y_0 (k_n \rho) \right] \frac{\sinh[k_n(L-y)]}{\sinh(k_n L)}, & R < \rho \leq R_i
\end{cases}
\] (5.9)

where \( R_i = R + t_{ox} \), \( J_0 (r) \) denotes the 0-th order Bessel function of the first kind (Bessel function), and \( Y_0 (r) \) denotes the 0-th order Bessel function of the second kind (Neumann function) [85]. With different coefficients, we can write similar series expression for \( u_R \) with \( L - y \) replaced by \( y \). The dielectric boundary conditions at the silicon/insulator interface require \( u_{L,R} \) and \( \epsilon (\partial u_{L,R}/\partial \rho) \) to be
continuous at $\rho = R$. Taking the ratio of these two boundary conditions finally yields an implicit equation for the eigenvalues $k_n$:

$$
\frac{\epsilon_{si} J_1 (k_n R)}{J_0 (k_n R)} = \frac{\epsilon_{ox} J_1 (k_n R) Y_0 (k_n R_i) - J_0 (k_n R_i) Y_1 (k_n R)}{J_0 (k_n R) Y_0 (k_n R_i) - J_0 (k_n R_i) Y_0 (k_n R)}
$$

(5.10)

where $J_1 (r)$ and $Y_1 (r)$ are the 1-st order Bessel function and Neumann function, respectively. This eigenvalue equation is equivalent to that in [83]. By solving this implicit equation, we can obtain the eigenvalues of $k_n$, which form a sequence $k_1 < k_2 < k_3 < \cdots$ with $k_n/k_1 \sim n$. Again, the generalized scale length $\lambda$ is defined by $\lambda = \pi/k_1$. The scale length theory shows that the minimum potential has an effective channel length dependence of $\sim \exp(-\pi L/2\lambda)$ to the first order, so the ratio $L/\lambda$ is a fundamental indicator of SCEs in MOSFETs. Numerical simulations reveal that the minimum channel length for transistors with acceptable SCEs should be larger than $\sim 1.5\lambda$ [84]. Note that $\lambda$ is analogous to $\pi$ times the natural length previously discussed [73, 76].

The implicit eigenvalue equation (5.10) has been solved numerically for SiO$_2$ gate insulator, and $\lambda/R$ as a function of $R/t_{ox}$ is plotted in Fig. 5.3. $\pi\lambda_N/R$ versus $R/t_{ox}$ obtained from equation (5.2) is also shown for comparison. It is clear that the natural length $\lambda_N$ leads to large errors when one of the thicknesses is much larger than the other. In the limit of $t_{ox} \ll R$, (5.10) can be approximately solved as $\lambda = \pi[R + (\epsilon_{si}/\epsilon_{ox}) t_{ox}]/\alpha_1$, where $\alpha_1 \simeq 2.405$ is the first zero of $J_0 (r)$. Equation (5.10) can also be approximated in the opposite extreme, i.e., $t_{ox} \gg R$, in which case $\lambda = \pi(t_{ox} + R)/\alpha_1$. To avoid the complexity of numerically solving the
Figure 5.3: Scale lengths calculated from (5.11) with $\beta = 130.9$ and $\gamma = 27.3$ are compared with the numerical solutions obtained directly from the implicit equation (5.10). The dashed lines show the scale length ($\pi \lambda N / R$) of Auth and Plummer [76] for comparison.

Implicit equation (5.10), we can construct an approximate explicit expression for $\lambda$ according to the asymptotic behaviors: $\lambda \simeq \frac{\pi}{\alpha_1} \left( \frac{R^3}{\epsilon_{ox}} + \frac{3 \epsilon_{si}}{\epsilon_{ox}} R^2 t_{ox} + 3 R t_{ox}^2 + t_{ox}^3 \right)^{1/3}$. This result applies to different $\epsilon_{ox}$, including high-$\kappa$ gate dielectrics, but the relative error exceeds 10% in some cases. However, we can add in fitting parameters to achieve high accuracy, e.g.,

$$
\lambda \simeq \frac{\pi}{\alpha_1} \left( R^5 + \frac{5 \epsilon_{si}}{\epsilon_{ox}} R^4 t_{ox} + \beta R^3 t_{ox}^2 + \gamma R^2 t_{ox}^3 + 5 R t_{ox}^4 + t_{ox}^5 \right)^{1/5}
$$

(5.11)

The fitting parameters $\beta$ and $\gamma$ have to be determined for different gate insulator permittivity. For SiO$_2$, the numerically solved $\lambda$ can be fitted quite well with $\beta = 130.9$ and $\gamma = 27.3$ for a wide range. This has been demonstrated in Fig. 5.3.

Similarly, we can also obtain accurate explicit solutions for $\lambda$ of DG MOS-
The fitting parameters $\beta$ and $\gamma$ have to be determined for different gate insulator permittivity. For SiO$_2$, the numerically solved $\lambda$ can be fitted quite well with $\beta = 411.4$ and $\gamma = 822.6$ for a wide range. This has been demonstrated in Fig. 5.4. $\pi \lambda_N/t_{si}$ versus $t_{si}/t_{ox}$ obtained from equation (5.1) is also shown for comparison. It is clear that the natural length $\lambda_N$ leads to large errors when one
Figure 5.5: Normalized gate insulator thickness \( t_{ox}/\lambda \) versus normalized silicon cylinder radius \( R/\lambda \) for SG MOSFETs with different insulator permittivity.

of the thicknesses is much larger than the other.

To provide some insight into the effects of varying the permittivity of the insulator, we also solve the implicit eigenvalue equations (5.6) and (5.10) for high-\( \kappa \) dielectrics. Typical results are shown in Fig. 5.5 and 5.6, where \( t_{ox}/\lambda \) is plotted as a function of \( R/\lambda \) or \( t_{si}/\lambda \) for \( \epsilon_{ox}/\epsilon_{si} = 1/3, 1, 3, 10 \). In Fig. 5.5, we note that \( \lambda > \pi R/\alpha_1 \) and \( \lambda > \pi t_{ox}/\alpha_1 \) in any case. The straight line result \( \lambda = \pi(t_{ox}+R)/\alpha_1 \), is expected for \( \epsilon_{ox} = \epsilon_{si} \). It is indicated in Fig. 5.5 that, as \( t_{ox}/R \) increases over 1, high-\( \kappa \) gate insulators gradually become less beneficial assuming the dielectric extends over the drain region. In Fig. 5.6, we note that \( \lambda > t_{si} \) and \( \lambda > 2t_{ox} \) in any case. The straight line result \( \lambda = t_{si} + 2t_{ox} \), is expected for \( \epsilon_{ox} = \epsilon_{si} \). Similarly, it is indicated in Fig. 5.6 that, as \( t_{ox}/t_{si} \) increases over 0.5, high-\( \kappa \) gate insulators
Figure 5.6: Normalized gate insulator thickness $t_{ox}/\lambda$ versus normalized silicon cylinder radius $t_{si}/\lambda$ for DG MOSFETs with different insulator permittivity.

gradually become less beneficial.

## 5.2 Analytical Model for SCEs in DG MOSFETs

In this section, we will review the two-region analytical model for SCEs in DG MOSFETs developed in [80, 81], which is based on the generalized scale length theory introduced in the last section.

For an undoped (or lightly-doped) symmetric DG MOSFET (Fig. 5.1) in the subthreshold region, Poisson’s equation is written as equation (5.3). If we assume the source and drain junctions are abrupt, the boundary conditions are as
follows

\[
\text{Gate} : \quad \psi \left( \pm \left( \frac{t_{si}}{2} + t_{ox} \right), y \right) = V_g - \Delta \phi, \quad 0 \leq y \leq L
\]

\[
\text{Source} : \quad \psi (x, 0) = V_s + V_{bi}, -\frac{t_{si}}{2} \leq x \leq \frac{t_{si}}{2}
\]

\[
\text{Drain} : \quad \psi (x, L) = V_d + V_{bi}, -\frac{t_{si}}{2} \leq x \leq \frac{t_{si}}{2}
\]

(5.13)

where \(\Delta \phi\) is the work function of gate electrode with respect to the intrinsic silicon, and \(V_{bi}\) is the build-in voltage roughly given by \(V_{bi} = E_g/2q\). It can be noticed that we have no explicit boundary conditions in the insulator gap regions between the source/drain and gate. Numerical simulation results show that the potential in this region is approximately a linear function of \(x\). The potential in the insulator gap region can then be expressed by

\[
\psi (x, 0) = (V_s + V_{bi}) + \frac{V_g - \Delta \phi - V_s - V_{bi}}{t_{ox}} \left| x - \frac{t_{si}}{2} \right|, \quad -\frac{t_{si}}{2} < x < \frac{t_{si}}{2} + t_{ox}
\]

\[
\psi (x, L) = (V_d + V_{bi}) + \frac{V_g - \Delta \phi - V_d - V_{bi}}{t_{ox}} \left| x - \frac{t_{si}}{2} \right|, \quad -\frac{t_{si}}{2} < x < \frac{t_{si}}{2} + t_{ox}
\]

(5.14)

By applying the superposition principle, the electrostatic potential is given by equation (5.4). Owing to the volume inversion effect in the subthreshold region, we can choose \(v (x) = constant = V_g - \Delta \phi\). Meanwhile, \(u_L\) and \(u_R\) can be written as

\[
u_L(x, y) = \sum_{n=1}^{\infty} b_n B_n (x, y)\]

\[
u_R(x, y) = \sum_{n=1}^{\infty} c_n C_n (x, y),\]

respectively, where

\[
B_n (x, y) = \begin{cases} 
\cos \left( k_n \frac{t_{si}}{2} \right) \frac{\sin \left[ k_n \left( \frac{t_{si}}{2} + t_{ox} + x \right) \right]}{\sin \left( k_n t_{ox} \right)} \frac{\sinh \left[ k_n (L - y) \right]}{\sinh \left( k_n L \right)}, & -\frac{t_{si}}{2} - t_{ox} \leq x < -\frac{t_{si}}{2} \\
\cos \left( k_n \frac{t_{si}}{2} \right) \frac{\sin \left[ k_n \left( \frac{t_{si}}{2} + t_{ox} + x \right) \right]}{\sin \left( k_n t_{ox} \right)}, & -\frac{t_{si}}{2} \leq x < \frac{t_{si}}{2} \\
\cos \left( k_n \frac{t_{si}}{2} \right) \frac{\sin \left[ k_n \left( \frac{t_{si}}{2} + t_{ox} - x \right) \right]}{\sin \left( k_n t_{ox} \right)} \frac{\sinh \left[ k_n (L - y) \right]}{\sinh \left( k_n L \right)}, & \frac{t_{si}}{2} < x \leq \frac{t_{si}}{2} + t_{ox}
\end{cases}
\]

(5.15)
and
\[
C_n(x, y) = \begin{cases} 
\cos \left( k_n \frac{t_{si}}{2} \right) \frac{\sin \left[ k_n \left( \frac{t_{si}}{2} + t_{ox} + x \right) \right]}{\sin (k_n t_{ox})}, & -\frac{t_{si}}{2} - t_{ox} \leq x < -\frac{t_{si}}{2} \\
\cos \left( k_n x \right) \frac{\sin (k_n y)}{\sin (k_n L)}, & -\frac{t_{si}}{2} \leq x \leq \frac{t_{si}}{2} \\
\cos \left( k_n \frac{t_{si}}{2} \right) \frac{\sin \left[ k_n \left( \frac{t_{si}}{2} + t_{ox} - x \right) \right]}{\sin (k_n t_{ox})}, & -\frac{t_{si}}{2} < x \leq \frac{t_{si}}{2} + t_{ox}
\end{cases}
\] (5.16)

Here \( k_n \) is the solution of the eigenvalue equation (5.6).

The eigenfunctions \( B_n(x, 0) (= C_n(x, L)) \) are not orthogonal to each other. In order to evaluate the coefficients \( b_n \) and \( c_n \), we need to find the orthogonality relationship in this case. It is easy to be verified that the orthogonality relationship takes the form
\[
\int_{-\frac{t_{si}}{2} - t_{ox}}^{\frac{t_{si}}{2} + t_{ox}} B_n(x, 0) B_m(x, 0) \epsilon(x) \, dx = 0, \quad n \neq m \tag{5.17}
\]

By multiplying \( u_{LR} \) with \( B_n(x, 0) \epsilon(x) \) and integrating, we will obtain the coefficients \( b_n \) and \( c_n \). \( b_n \) can be expressed as
\[
b_n = \frac{\int_{-\frac{t_{si}}{2} - t_{ox}}^{\frac{t_{si}}{2} + t_{ox}} \left[ \psi(x, 0) - v(x) \right] B_n(x, 0) \epsilon(x) \, dx}{\int_{-\frac{t_{si}}{2} - t_{ox}}^{\frac{t_{si}}{2} + t_{ox}} B_n^2(x, 0) \epsilon(x) \, dx} \tag{5.18}
\]
and \( c_n \) can be expressed as
\[
c_n = \frac{\int_{-\frac{t_{si}}{2} - t_{ox}}^{\frac{t_{si}}{2} + t_{ox}} \left[ \psi(x, L) - v(x) \right] C_n(x, L) \epsilon(x) \, dx}{\int_{-\frac{t_{si}}{2} - t_{ox}}^{\frac{t_{si}}{2} + t_{ox}} C_n^2(x, L) \epsilon(x) \, dx} \tag{5.19}
\]

The above integrals can be carried out to obtain explicit expressions for both \( b_n \) and \( c_n \). The final results are
\[
b_n = \frac{2 \tan (k_n t_{ox}) \sin \left( k_n \frac{t_{si}}{2} \right)}{k_n^2 t_{ox} \left[ \frac{t_{si}}{2} + \frac{\sin (k_n t_{si})}{\sin 2(k_n t_{ox})} \right]} (V_s + V_{bi} + \Delta \phi - V_g) \tag{5.20}
\]
and
\[ c_n = \frac{2 \tan \left( k_n t_{ox} \right) \sin \left( \frac{k_n t_{ox}}{2} \right)}{k_n^2 t_{ox} \left[ \frac{t_{ox}}{2} + \frac{\sin (k_n t_{ox})}{\sin 2(k_n t_{ox})} \frac{t_{ox}}{2} \right]} (V_d + V_{bi} + \Delta \phi - V_g) \] (5.21)

For \( L > 1.5 \lambda \), the \( u_{L,R} \) series decay rapidly since \( k_n / k_1 \sim n \). Therefore only the lowest order terms in the \( u_{L,R} \) series need to be kept.\(^1\) Then the expression for the full 2-D potential in the subthreshold region is \( \psi (x, y) = V_g - \Delta \phi + b_1 B_1 (x, y) + c_1 C_1 (x, y) \), which in the silicon region is explicitly given by
\[ \psi (x, y) = V_g - \Delta \phi + \frac{b_1 \sinh [k_1 (L - y)] + c_1 \sinh (k_1 y)}{\sinh (k_1 L)} \cos (k_1 x) \] (5.22)

Once the analytical potential is obtained, the subthreshold current can be derived based on the current continuity equation
\[ I_{ds} (y) = \mu dV (y) Q_m (y) \] (5.23)

where
\[ Q_m (y) = q n_i W \int_{t_{si}^{1/2}}^{t_{si}^{1/2}} e^{q[\psi(x,y) - V(y)]/kT} dx \] (5.24)

is the inversion charge per unit gate length. Here, \( \mu \) is the effective mobility, \( n_i \) is the intrinsic carrier density, and \( V(y) \) is the electron quasi-Fermi potential, which is a constant in the \( x \)-direction. Current continuity requires \( I_{ds} \) to be independent of \( y \). Therefore, integration of (5.23) with respect to \( y \) from 0 to \( L \) yields
\[ I_{ds} = \mu q n_i W \int_{t_{si}^{1/2}}^{t_{si}^{1/2}} e^{-qV(y)/kT} dV (y) \]
\[ = \mu kT n_i W \int_{t_{si}^{1/2}}^{t_{si}^{1/2}} e^{-qV_{ds}/kT} dy \]
\[ = \mu kT n_i W \int_{t_{si}^{1/2}}^{t_{si}^{1/2}} e^{-qV_{ds}/kT} dy \] (5.25)

\(^1\)For extremely scaled MOSFETs, e.g., \( L < 1.5 \lambda \), some higher order terms of \( u_{L,R} \) are not negligible. But these cases are not the focus of our discussion because of the severe SCEs.
Threshold voltage roll-off $\Delta V_t$ can be obtained from the parallel shift of $I_{ds}L - V_g$ curves (in log-scale) of a short-channel device with respect to the long-channel device, i.e.,

$$I_{ds}L\text{(short channel)} = I_{ds}L\text{(long channel)} \times e^{-\frac{q\Delta V_t}{kT}} \quad (5.26)$$

The subthreshold drain current expression (5.25) contains a double integral in the denominator that cannot be carried out analytically. To derive an explicit expression for compact model purposes, we need to simplify the 2-D potential function $\psi(x, y)$.

The drain current is predominantly controlled by the point of maximum electron energy barrier (minimum electrostatic potential) in the channel direction located at $(0, y_c)$, where $y_c$ is obtained by solving $\partial \psi(x, y) / \partial y|_{y=y_c} = 0$,

$$y_c = \frac{L}{2} - \frac{1}{2k_1} \ln \frac{c_1 e^{k_1 L/2} - b_1 e^{-k_1 L/2}}{b_1 e^{k_1 L/2} - c_1 e^{-k_1 L/2}} \simeq \frac{L}{2} - \frac{1}{2k_1} \ln \frac{c_1}{b_1} \quad (5.27)$$

The minimum potential can then be expressed as

$$\psi_{\text{min}} = \psi(0, y_c) = V_g - \Delta \phi + 2\sqrt{b_1 c_1 e^{-k_1 L/2}} \quad (5.28)$$

We expand $\psi(x, y)$ at $(0, y_c)$ and keep to the second order term,

$$\psi(x, y) \simeq \psi_{\text{min}} + k_1^2 \sqrt{b_1 c_1 e^{-k_1 L/2}} (y - y_c)^2 \quad (5.29)$$

Substituting (5.29) into the extraction equation (5.26), one can derive

$$\Delta V_t = -2\sqrt{b_1 c_1 e^{-k_1 L/2}} + \frac{kT}{q} \ln \left\{ \frac{\sqrt{\pi}}{2D_1L} [\text{erf}(D_1(L - y_c)) + \text{erf}(D_1y_c)] \right\} \quad (5.30)$$
where erf(z) is the error function defined as

\[ \text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-u^2} \, du \] (5.31)

and

\[ D_1 = \left( \frac{q k_1^2 \sqrt{b_1} c_1 e^{-k_1 L}}{k T} \right)^{1/2} \] (5.32)

As can be observed from (5.21) and (5.30), DIBL effect is incorporated in the threshold voltage roll-off model through the \( V_d \)-dependent parameter \( c_1 \), and can be estimated by

\[ DIBL = \Delta V_t(V_{ds} = 0.05V) - \Delta V_t(V_{ds} = 1V) \] (5.33)

In order to extract an explicit expression for the subthreshold slope from (5.25), the 2-D potential function is further simplified by \( \psi(x, y) \simeq \psi_{\text{min}} \). We can approximate the subthreshold current as \( I_{ds} \propto e^{q \psi_{\text{min}}/kT} \). Then the inverse subthreshold slope is given by

\[ S = \left[ \frac{\partial (\log_{10} I_{ds})}{\partial V_g} \right]^{-1} \simeq \left( \frac{\partial \psi_{\text{min}}}{\partial V_g} \right)^{-1} \times 60 \text{ mV/Dec} \] (5.34)

With the definition of \( \psi_{\text{min}} \) in (5.28), one obtains the expression for \( \partial \psi_{\text{min}}/\partial V_g \), which is smaller than 1 and thus accounts for the subthreshold slope degradation:

\[ \frac{\partial \psi_{\text{min}}}{\partial V_g} = 1 - 2B \left[ 1 + \frac{1}{8} \left( \frac{V_d - V_s}{V_{bi} + \frac{V_d - V_s}{2} - V_g} \right)^2 \right] e^{-k_1 L} \] (5.35)

where

\[ B = \frac{2 \tan \left( k_1 t_{ox} \right) \sin \left( k_1 \frac{t_{bi}}{2} \right)}{k_1^2 t_{ox} \left[ \frac{t_{bi}}{2} + \frac{\sin(k_1 t_{ox})}{2 \sin(2k_1 t_{ox})} t_{ox} \right]} \] (5.36)
Explicit expressions for three typical parameters of SCEs, i.e., threshold voltage roll-off, DIBL, and subthreshold slope degradation, are given by equations (5.30), (5.33), and (5.35), respectively. These equations are insensitive to $V_g$ as long as the device is biased in the subthreshold region. This analytical model for SCEs in DG MOSFETs has been validated by TCAD numerical simulation results, and model-to-simulation comparison is extensively shown in [80, 81].

### 5.3 Analytical Model for SCEs in SG MOSFETs

Following the methodology in the last section, we, for the first time, develop a two-region analytical model for SCEs in SG MOSFETs based on the generalized scale length theory.

For an undoped (or lightly-doped) SG MOSFET (Fig. 5.2) in the sub-threshold region, Poisson’s equation is written as equation (5.7). If we assume the source and drain junctions are abrupt, the boundary conditions are as follows

\[
\begin{align*}
Gate & : \psi(R_i, y) = V_g - \Delta \phi, 0 \leq y \leq L \\
Source & : \psi(\rho, 0) = V_s + V_{bi}, 0 \leq \rho \leq R \\
Drain & : \psi(\rho, L) = V_d + V_{bi}, 0 \leq \rho \leq R 
\end{align*}
\]  

(5.37)

where $\Delta \phi$ is the work function of gate electrode with respect to the intrinsic silicon, and $V_{bi}$ is the build-in voltage roughly given by $V_{bi} = E_g/2q$. It can be noticed that we have no explicit boundary conditions in the insulator gap regions between
the source/drain and gate. Numerical simulation results show that the potential in this region is approximately a logarithm function of $\rho$. The potential in the insulator gap region can then be expressed by

$$
\psi (\rho, 0) = (V_s + V_{bb}) + \frac{V_g - \Delta \phi - V_s - V_{bi}}{\ln (R_i/R)} \ln \frac{\rho}{R}, \quad R < \rho \leqslant R_i
$$

$$
\psi (\rho, L) = (V_d + V_{bb}) + \frac{V_g - \Delta \phi - V_d - V_{bi}}{\ln (R_i/R)} \ln \frac{\rho}{R}, \quad R < \rho \leqslant R_i
$$

(5.38)

By applying the superposition principle, the electrostatic potential is given by equation (5.8), where $v (\rho)$ is the 1-D solution of

$$
\frac{1}{\rho} \frac{\partial}{\partial \rho} \rho \left( \epsilon (\rho) \frac{\partial v (\rho)}{\partial \rho} \right) = 0
$$

(5.39)

Owing to the volume inversion effect in the subthreshold region, we can choose $v (\rho) = \text{constant} = V_g - \Delta \phi$. Meanwhile, $u_L$ and $u_R$ can be written as $u_L (\rho, y) = \sum_{n=1}^{\infty} b_n B_n (\rho, y)$ and $u_R (\rho, y) = \sum_{n=1}^{\infty} c_n C_n (\rho, y)$, respectively, where

$$
B_n (\rho, y) = \begin{cases} 
J_0 (k_n \rho) \frac{\sinh[k_n (L-y)]}{\sinh[k_n L]}, & 0 \leqslant \rho \leqslant R \\
J_0 (k_n R) \frac{J_0 (k_n R) Y_0 (k_n R_i) - J_0 (k_n R_i) Y_0 (k_n \rho)}{J_0 (k_n R) Y_0 (k_n R_i) - J_0 (k_n R_i) Y_0 (k_n R)} \frac{\sinh[k_n (L-y)]}{\sinh[k_n L]}, & R < \rho \leqslant R_i
\end{cases}
$$

(5.40)

and

$$
C_n (\rho, y) = \begin{cases} 
J_0 (k_n \rho) \frac{\sinh(k_n y)}{\sinh(k_n L)}, & 0 \leqslant \rho \leqslant R \\
J_0 (k_n R) \frac{J_0 (k_n R) Y_0 (k_n R_i) - J_0 (k_n R_i) Y_0 (k_n \rho)}{J_0 (k_n R) Y_0 (k_n R_i) - J_0 (k_n R_i) Y_0 (k_n R)} \frac{\sinh(k_n y)}{\sinh(k_n L)}, & R < \rho \leqslant R_i
\end{cases}
$$

(5.41)

Here $k_n$ is the solution of the eigenvalue equation (5.10).

The eigenfunctions $B_n (\rho, 0)$ ($= C_n (\rho, L)$) are not orthogonal to each other. In order to evaluate the coefficients $b_n$ and $c_n$, we need to find the orthogonality
relationship in this case. It can be approved that the orthogonality relationship takes the form

\[ \int_0^{R_i} B_n(\rho, 0) B_m(\rho, 0) \epsilon(\rho) \rho d\rho = 0, n \neq m \]  

(5.42)

The key point to prove equation (5.42) is to analytically carry out the integrals in the form of \( \int_{t_1}^{t_2} t f_1(at) f_2(bt) dt \) \((a \neq b)\), where the functions \( f_1 \) and \( f_2 \) can be either Bessel function \( J_0 \) or Neumann function \( Y_0 \). Here, we will take \( \int_{t_1}^{t_2} t J_0(at) Y_0(bt) dt \) as an example to show how to calculate this type of integral.

According to the definitions, one has

\[ \frac{1}{t} \frac{d}{dt} \left[ t \frac{dJ_0(at)}{dt} \right] + a^2 J_0(at) = 0 \]  

(5.43)

\[ \frac{1}{t} \frac{d}{dt} \left[ t \frac{dY_0(bt)}{dt} \right] + b^2 Y_0(bt) = 0 \]  

(5.44)

(5.43) \times (tY_0(bt)) - (5.44) \times (tJ_0(at)) leads to

\[ (a^2 - b^2) t J_0(at) Y_0(bt) = -Y_0(bt) \frac{d}{dt} \left[ t \frac{dJ_0(at)}{dt} \right] + J_0(at) \frac{d}{dt} \left[ t \frac{dY_0(bt)}{dt} \right] \]  

(5.45)

Integrating the above equation, we finally obtain

\[ (a^2 - b^2) \int_{t_1}^{t_2} t J_0(at) Y_0(bt) dt = [at J_1(at) Y_0(bt) - bt J_0(at) Y_1(bt)]_{t=t_1}^{t=t_2} \]  

(5.46)

By multiplying \( u_{L,R} \) with \( B_n(\rho, 0) \epsilon(\rho) \rho \) and integrating, we will obtain the coefficients \( b_n \) and \( c_n \). \( b_n \) can be expressed as

\[ b_n = \frac{\int_0^{R_i} \left[ \psi(\rho, 0) - v(\rho) \right] B_n(\rho, 0) \epsilon(\rho) \rho d\rho}{\int_0^{R_i} B_n^2(\rho, 0) \epsilon(\rho) \rho d\rho} \]  

(5.47)

and \( c_n \) can be expressed as

\[ c_n = \frac{\int_0^{R_i} \left[ \psi(\rho, L) - v(\rho) \right] C_n(\rho, L) \epsilon(\rho) \rho d\rho}{\int_0^{R_i} C_n^2(\rho, L) \epsilon(\rho) \rho d\rho} \]  

(5.48)
The above tedious integrals can be carried out by using the basic characteristics of Bessel and Neumann functions [85]. To derive explicit expressions for $b_n$ and $c_n$, we have to analytically calculate following types of integral: $\int_{t_1}^{t_2} tf_1 (at) \, dt$, $\int_{t_1}^{t_2} t \ln tf_1 (at) \, dt$, and $\int_{t_1}^{t_2} tf_1 (at) f_2 (at) \, dt$. Here, we will list the result of an example for each type, and the others can be obtained by replacing $J_{0,1}$ with $Y_{0,1}$, or reversely:

$$\int_{t_1}^{t_2} tJ_0 (at) \, dt = \left[ \frac{tJ_1 (at)}{a} \right]_{t=t_1}^{t=t_2} \quad (5.49)$$
$$\int_{t_1}^{t_2} t \ln tJ_0 (at) \, dt = \left[ \frac{t \ln tJ_1 (at)}{a} + \frac{J_0 (at)}{a^2} \right]_{t=t_1}^{t=t_2} \quad (5.50)$$
$$\int_{t_1}^{t_2} tJ_0 (at) Y_0 (at) \, dt = \left[ \frac{t^2 J_1 (at) Y_1 (at)}{2} + \frac{t^2 J_0 (at) Y_0 (at)}{2} \right]_{t=t_1}^{t=t_2} \quad (5.51)$$

According to these intermediate results, we can obtain the final expressions of $b_n$ and $c_n$ as

$$b_n = E (V_s + V_{bi} + \Delta \phi - V_g) \quad (5.52)$$

and

$$c_n = E (V_d + V_{bi} + \Delta \phi - V_g) \quad (5.53)$$

where

$$E = \frac{\pi^2}{2} \frac{1}{\ln (R_i / R)} \frac{\epsilon_i}{\epsilon_{si}} \frac{J_0 (k_n R)}{A \cdot B + \frac{\pi^2 k_n^2 R^2}{4} C} \quad (5.54)$$

with

$$A = \frac{J_0 (k_n R)}{J_0 (k_n R) Y_0 (k_n R_i) - J_0 (k_n R_i) Y_0 (k_n R)}$$

$$B = \frac{J_1 (k_n R)}{J_1 (k_n R) Y_0 (k_n R_i) - J_0 (k_n R_i) Y_1 (k_n R)}$$

$$C = \left( 1 - \frac{\epsilon_i}{\epsilon_{si}} \right) J_0^2 (k_n R) + \left( 1 - \frac{\epsilon_{si}}{\epsilon_i} \right) J_1^2 (k_n R) \quad (5.55)$$
Note that $b_n$ and $c_n$ are bias dependent. The $V_d$ dependence in $c_n$ is responsible for drain induced barrier lowering (DIBL), and the $V_g$ dependence in both $b_n$ and $c_n$ is responsible for subthreshold slope degradation, which will be discussed later.

For $L > 1.5\lambda$, the $u_{L,R}$ series decay rapidly since $k_n/k_1 \sim n$. Therefore only the lowest order terms in the $u_{L,R}$ series need to be kept. Then the expression for the full 2-D potential in the subthreshold region is

$$\psi (\rho, y) = V_g - \Delta \phi + b_1 B_1 (\rho, y) + c_1 C_1 (\rho, y)$$

which in the silicon region is explicitly given by

$$\psi (\rho, y) = V_g - \Delta \phi + \frac{b_1 \sinh [k_1 (L - y)] + c_1 \sinh (k_1 y)}{\sinh (k_1 L)} J_0 (k_1 \rho)$$

(5.56)

Once the analytical potential is obtained, the subthreshold current can be derived based on the current continuity equation

$$I_{ds} (y) = \mu \frac{dV (y)}{dy} Q_m (y)$$

(5.57)

where

$$Q_m (y) = 2\pi q n_i \int_0^R e^{q[\psi (\rho, y) - V (y)]/kT} \rho d\rho$$

(5.58)

is the inversion charge per unit gate length. Current continuity requires $I_{ds}$ to be independent of $y$. Therefore, integration of (5.57) with respect to $y$ from 0 to $L$ yields

$$I_{ds} = 2\pi \mu q n_i \left[ \int_{V_d}^{V_s} e^{-qV (y)/kT} dy \right] \frac{\int_0^L \frac{dy}{\int_0^R e^{q\psi (\rho, y)/kT} \rho d\rho}}{\int_0^L \frac{dy}{\int_0^R e^{q\psi (\rho, y)/kT} \rho d\rho}}$$

(5.59)
Figure 5.7: $I_{ds}L - V_g$ curves for SG MOSFETs obtained from the analytical model (solid lines) in comparison with the 2-D numerical simulation results (open circles).

The subthreshold current can be calculated analytically as a function of $V_g$, $V_s$, and $V_d$ according to (5.59). In Fig. 5.7, we compare the model-predicted $I_{ds}L - V_g$ curves with the numerical simulation results. Because we use cylindrical coordinates instead of Cartesian coordinates in ISE-DESSIS, the numerical simulation is essentially 2-D. We observe excellent agreement in the subthreshold region where the analytical solution is valid.

Threshold voltage roll-off $\Delta V_t$ can be obtained from the parallel shift of $I_{ds}L - V_g$ curves (in log-scale) of a short-channel device with respect to the long-channel device, i.e.,

$$I_{ds}L_{(short\ channel)} = I_{ds}L_{(long\ channel)} \times e^{-q\Delta V_t / kT} \quad (5.60)$$
The subthreshold drain current expression (5.59) contains a double integral in the denominator that cannot be carried out analytically. To derive an explicit expression for compact model purposes, we need to simplify the 2-D potential function $\psi(\rho, y)$.

The drain current is predominantly controlled by the point of maximum electron energy barrier (minimum electrostatic potential) in the channel direction located at $(0, y_c)$, where $y_c$ is obtained by solving $\partial \psi (\rho, y) / \partial y |_{y=y_c} = 0$,

$$
y_c = \frac{L}{2} - \frac{1}{2k_1} \ln \frac{c_1 e^{k_1 L/2} - b_1 e^{-k_1 L/2}}{b_1 e^{k_1 L/2} - c_1 e^{-k_1 L/2}} \simeq \frac{L}{2} - \frac{1}{2k_1} \ln \frac{c_1}{b_1} \tag{5.61}
$$

The minimum potential can then be expressed as

$$
\psi_{\text{min}} = \psi(0, y_c) = V_g - \Delta \phi + 2\sqrt{b_1 c_1 e^{-k_1 L/2}} \tag{5.62}
$$

We expand $\psi(\rho, y)$ at $(0, y_c)$ and keep to the second order term,

$$
\psi(\rho, y) \simeq \psi_{\text{min}} + k_1^2 \sqrt{b_1 c_1 e^{-k_1 L/2}} (y - y_c)^2 - \frac{1}{2} k_1^2 \sqrt{b_1 c_1 e^{-k_1 L/2}} \rho^2 \tag{5.63}
$$

Substituting (5.63) into the extraction equation (5.60), one can derive

$$
\Delta V_t = \Delta V_{t1} + \Delta V_{t2} + \Delta V_{t3} \tag{5.64}
$$

where $\Delta V_{t1}$ corresponds to the minimum potential term given by

$$
\Delta V_{t1} = -2\sqrt{b_1 c_1 e^{-k_1 L/2}} \tag{5.65}
$$

and $\Delta V_{t2}$ and $\Delta V_{t3}$ correspond to the modifications induced by the second order terms in the channel direction and radial direction, respectively. They are

$$
\Delta V_{t2} = \frac{kT}{q} \ln \left\{ \frac{\sqrt{\pi}}{2D_1 L} \left[ \text{erf} (D_1 (L - y_c)) + \text{erf} (D_1 y_c) \right] \right\}
$$
Figure 5.8: Threshold voltage roll-offs for SG MOSFETs as functions of channel length $L$ obtained from the analytical model (lines) in comparison with the 2-D numerical simulation results (symbols).

$$
\Delta V_{t3} = \frac{kT}{q} \ln \left[ \frac{D_1^2 R^2}{2 \left( 1 - e^{-D_1^2 R^2 / 2} \right)} \right] 
$$

(5.66)

where $\text{erf}(z)$ is the error function defined as

$$
\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-u^2} du 
$$

(5.67)

and

$$
D_1 = \left( \frac{qkT}{k} \sqrt{b_1 c_1 e^{-k_1 L / 2}} \right)^{1/2}
$$

(5.68)

Fig. 5.8 shows the model-predicted threshold voltage roll-off $\Delta V_t$ as a function of channel length, compared with 2-D numerical simulation results. The shift is evaluated at a constant $I_{dsL}$ level corresponding to the current of the long channel device at $V_g - \Delta \phi = 0$ V, deeply into the subthreshold region. Inasmuch as the
subthreshold slope changes slightly when \( L > 1.5\lambda \), the choice of constant current level is non-critical. Good agreement has been achieved for channel length down to \( \sim 1.5\lambda \).

As can be observed from (5.53) and (5.64), DIBL effect is incorporated in the threshold voltage roll-off model through the \( V_d \)-dependent parameter \( c_1 \), and can be estimated by

\[
DIBL = \Delta V_t(V_{ds} = 0.05V) - \Delta V_t(V_{ds} = 1V)
\]

(5.69)

Fig. 5.9 indicates the validity of our DIBL model by comparison with simulation results.
Figure 5.10: Subthreshold slopes for SG MOSFETs as functions of channel length $L$ obtained from the analytical model (lines) in comparison with the 2-D numerical simulation results (symbols).

In order to extract an explicit expression for the subthreshold slope from (5.59), the 2-D potential function is further simplified by $\psi(\rho, y) \simeq \psi_{\text{min}}$. We can approximate the subthreshold current as $I_{ds} \propto e^{q\psi_{\text{min}}/kT}$. Then the inverse subthreshold slope is given by

$$S = \left[ \frac{\partial \left( \log_{10} I_{ds} \right)}{\partial V_g} \right]^{-1} \simeq \left( \frac{\partial \psi_{\text{min}}}{\partial V_g} \right)^{-1} \times 60 \text{ mV/Dec} \quad (5.70)$$

With the definition of $\psi_{\text{min}}$ in (5.62), one obtains the expression for $\partial \psi_{\text{min}}/\partial V_g$, which is smaller than 1 and thus accounts for the subthreshold slope degradation:

$$\frac{\partial \psi_{\text{min}}}{\partial V_g} = 1 - 2E \frac{(V_{bi} + \Delta \phi + \frac{V_d + V_s}{2} - V_g) e^{-\frac{k_1 L}{2}}}{\sqrt{(V_s + V_{bi} + \Delta \phi - V_g)(V_d + V_{bi} + \Delta \phi - V_g)}} \quad (5.71)$$

where $E$ is given by (5.54) ($n = 1$).
Fig. 5.10 shows that the analytical subthreshold slope for SG MOSFETs is in good agreement with 2-D numerical simulation results for $L > 1.5\lambda$. Again, the gate voltage is chosen to satisfy $V_g - \Delta\phi = 0$ V so that the device is biased in the deep subthreshold region. Actually, the subthreshold slope is almost independent of bias condition as long as it is in the subthreshold region.

5.4 Summary

In conclusion, based on generalized scale length theory, full analytical expressions of potential and subthreshold current have been derived for both short-channel DG and SG MOSFETs by solving 2-D Poisson’s equation in both the semiconductor and insulator regions. The compact models for SCEs including threshold voltage roll-off, DIBL, and subthreshold slope degradation are extracted from the subthreshold current expression. It has been verified by 2-D numerical simulation results that the SCE model is predictive with no fitting parameters.

The text of Chapter 5, in part, is a reprint of the material as it appears in “Scaling of nanowire transistors” by Bo Yu, Lingquan Wang, Yu Yuan, Peter Asbeck, and Yuan Taur, IEEE Transaction on Electron Devices, Nov 2008. The dissertation author was the primary investigator and author of this paper.

The text of Chapter 5, in part, is a reprint of the material as it will appear in “A 2-D analytical solution for short-channel effects in nanowire MOSFETs” by Bo Yu, Yu Yuan, Jooyoung Song, and Yuan Taur, IEEE Transaction on Electron
Devices, Oct 2009. The dissertation author was the primary investigator and author of this paper.
Analysis of MOS Capacitance in III-V MOSFETs

6.1 A Review of Inversion Layer Capacitance in Bulk Si MOSFETs: from Classical to Quantum-Mechanical

The focus of this dissertation was heretofore on the Si-based MG MOSFETs, but we will change the focus to III-V MOSFETs from now on. In this chapter, the emphasis is put on the 1-D structure, i.e., the III-V MOS capacitor, by neglecting any transport or 2-D electrostatics related issues. Under this circumstance, we only need to solve 1-D Poisson’s equation. However, to get the reasonable solution for III-V MOS capacitor, 1-D Poisson’s equation has to be solved with 1-D Schrödinger’s equation self-consistently, and Maxwell-Boltzmann (MB) distribution, which is widely used in Si-based devices, has to be replaced by
Fermi-Dirac (FD) distribution. Before we get into analysis of MOS capacitance in III-V MOSFETs, it is necessary to take a quick review of MOS capacitance in bulk Si MOSFETs.

The total MOS capacitance per unit area is defined as

$$C = \frac{dQ_g}{dV_g}$$  

(6.1)

where $Q_g$ is the charge per unit area on the gate. With oxide and interface trapped charges ignored, simple analysis indicates that the total MOS capacitance in general consists of three components which are connected in series [84, 86, 87]: gate insulator capacitance $C_{ox}$ given by $C_{ox} = \epsilon_{ox}/t_{ox}$; polysilicon-gate depletion capacitance $C_p$; semiconductor capacitance $C_{sc}$. Therefore, $C$ is given by

$$C = \left( \frac{1}{C_{ox}'} + \frac{1}{C_p'} + \frac{1}{C_{sc}'} \right)^{-1}$$  

(6.2)

In the following discussion, we will neglect the contribution by polysilicon-gate depletion effect, because even in Si-based MOSFETs polysilicon-gate tends to be replaced by metal gate [12]–[16]. Thus we have

$$C = \left( \frac{1}{C_{ox}'} + \frac{1}{C_{sc}'} \right)^{-1}$$  

(6.3)

Gate insulator capacitance $C_{ox}$ has very clear physical meaning, and actually is constant when device parameters ($\epsilon_{ox}$ and $t_{ox}$) are determined. However, semiconductor capacitance $C_{sc}$ is much more complicated. It is defined as

$$C_{sc} = \frac{dQ_g}{d\psi_s}$$  

(6.4)
Figure 6.1: Energy-band diagram near the silicon surface of a p-type MOS device. The band bending $\psi$ is defined as positive when the bands bend downward with respect to the bulk. Adapted from [84].

where $\psi_s$ is the surface potential (reference shown in Fig. 6.1). As known, $C_{sc}$ is a function of $\psi_s$, and depends on the frequency. Therefore, the total capacitance $C$ can be plotted as a function of gate voltage $V_g$, and the $C$-$V$ characteristics are frequency-dependent, as indicated in Fig. 6.2. However, we only concentrate on the low-frequency $C$-$V$ here.

The total charge per unit area induced in the silicon $Q_s$ is equal and opposite to the charge on the metal gate, and $Q_s$ is a combination of both inversion charge $Q_i$ and depletion charge $Q_d$. Therefore, $C_{sc}$ can be broken up into two components in parallel:

$$C_{sc} = C_i + C_d$$  \hspace{1cm} (6.5)

where $C_i$ and $C_d$ are usually called inversion layer capacitance and depletion charge
Figure 6.2: MOS capacitance-voltage curves: (a) low frequency, (b) high frequency, (c) deep depletion. Adapted from [88].

According to the equations (6.3) and (6.5), an equivalent circuit for MOS capacitance $C$ is shown in Fig. 6.3. MOS capacitances are defined as small-signal capacitances and thus can be easily measured by applying a small AC voltage on top of a DC bias across the structure and sensing the out-of-phase AC current at the same frequency. Actually, it is possible to measure $C_i$ and $C_d$ separately in a split $C$-V measurement on MOSFET [89]. The equivalent circuit model for split

$$
C_i = \frac{d (-Q_i)}{d\psi_s}
$$

$$
C_d = \frac{d (-Q_d)}{d\psi_s}
$$

In the accumulation region, $C_d$ is also used to denote the capacitance associated with majority carriers, but it cannot be called depletion charge capacitance anymore.
Figure 6.3: Equivalent circuit model of an MOS capacitor at low frequency.

C-V setup (Fig. 6.4) is slightly different from that of an MOS capacitor. In the split C-V measurement, the n+ Source/Drain part of the total MOS capacitance is $C_{inv}$

$$C_{inv} = \frac{d(-Q_i)}{dV_g} = \frac{C_{ox}C_i}{C_{ox} + C_i + C_d}$$  \hspace{1cm} (6.7)

and the p-type Substrate part of the total MOS capacitance is $C_D$

$$C_D = \frac{d(-Q_d)}{dV_g} = \frac{C_{ox}C_d}{C_{ox} + C_i + C_d}$$  \hspace{1cm} (6.8)

The importance of $C_{inv}$ is manifest. Being the slope of DC $Q_i-V_g$ curves, $C_{inv}$ can be integrated to regenerate $Q_i-V_g$ curves. Such a technique is used, for example, in channel mobility measurements where the inversion charge density must be

Figure 6.4: Equivalent circuit model for split C-V setup.
An example of split C-V measurement results is demonstrated in Fig. 6.5. It can be observed that before inversion layer forms only $C_D$ contributes. But after inversion layer forms, $C_{\text{inv}}$ becomes dominant quickly, and $C_D$ drops significantly due to the strong screening by inversion layer. Therefore, it is a good approximation in the strong inversion region that

$$C_{\text{inv}} = \frac{C_{\text{ox}} C_i}{C_{\text{ox}} + C_i}$$

(6.9)

i.e., $C_{\text{inv}}$ equals $C_{\text{ox}}$ and $C_i$ connected in series.

Using the classical theory (MB distribution), $Q_i$ in the strong inversion
region can be approximated by [84]

\[ Q_i = -\sqrt{\frac{2\epsilon_s k T n_i^2}{N_a}} e^{q\psi_s/2kT} \] (6.10)

Thus the inversion layer capacitance can be expressed as

\[ C_i = \frac{|Q_i|}{2kT/q} \] (6.11)

In this case, \( C_i \) is always much larger than \( C_{ox} \), and thus both \( C_{inv} \) and \( C \) are very close to \( C_{ox} \), as indicated in Fig. 6.2. However, in general, electrons in the inversion layer should be treated quantum-mechanically as a 2-D electron gas [90, 91]. The classical theory is a good compromise for small \( C_{ox} \). But for large \( C_{ox} \), quantum-mechanical theory is necessary to predict the degradation of \( C_{inv} \) from \( C_{ox} \). Due to quantum-mechanical effect, the inversion charge distribution is pushed further into Si, i.e., away from the oxide interface. This finite inversion layer thickness effectively increases the oxide thickness in terms of capacitance \( C_{inv} \). Inversion layer capacitance \( C_i \) due to the quantum-mechanical inversion layer thickness is reasonably given by

\[ C_i = \frac{\epsilon_{si}}{x_c} \] (6.12)

where \( x_c \) is the centroid of inversion layer, and defined as

\[ x_c = \frac{\int_0^{W_d} x \cdot n(x) \, dx}{\int_0^{W_d} n(x) \, dx} \] (6.13)

Here, \( W_d \) is the depletion width. With the triangular potential well approximation and assuming that only one subband is occupied, an analytical expression for the centroid position \( x_c \) has been derived by Stern [91] based on the variational
\[
x_c = \left[ \frac{4q m^*}{9 \epsilon_{si} \hbar^2} \left| Q_d + \frac{11}{32} Q_i \right| \right]^{-1/3} \approx \left[ \frac{11 q m^*}{72 \epsilon_{si} \hbar^2} |Q_i| \right]^{-1/3}
\] (6.14)

where $\hbar$ is the reduced Plank constant and $m^*$ is the effective mass of Si perpendicular to the Si/SiO$_2$ interface.

### 6.2 Physical Understanding of Inversion Layer Capacitance in III-V MOSFETs

#### 6.2.1 Density-of-states capacitance

For Si bulk MOSFETs, the inversion layer capacitance $C_i$ can be approximated by $\epsilon_{si}/x_c$, as given in equation (6.12). However, this approximation is not valid anymore for III-V MOSFETs. It has been pointed out that there are two physical origins of $C_i$ [92]-[95]: One origin is the finite band bending associated with increasing $Q_i$, which is due to the finite density-of-states (DOS) in the band; the other origin is the finite inversion layer thickness, i.e., expanding inversion charge distribution, which is due to the quantum confinement effect. Si has relatively large DOS effective mass with degeneracy taken into account, therefore the contribution to inversion layer capacitance $C_i$ from finite DOS is negligible to a certain extent. In other words, the finite inversion layer thickness dominates $C_i$ in Si bulk MOSFETs. As we have discussed in Chapter 1, III-V MOSFETs are proposed to achieve CMOS performance breakthrough because III-V materials
promise significantly higher mobilities than Si, strained Si, or even SiGe MOS-FETs. For example, electron mobility in InGaAs and hole mobility in GaSb are 3-10 times higher than in Si. This benefit of mobility comes from the small effective mass of electron or hole. Due to the relatively small DOS effective mass, the contribution to inversion layer capacitance $C_i$ from finite DOS might be even greater than that from finite inversion layer thickness in III-V MOSFETs.

Roughly speaking, $C_i$ has two physical origins, but to get a deeper physical understanding we need to dig into the details from the quantitative point of view. Another advantage of III-V materials is that, by tailoring composition, heterostructures offer additional degrees of freedom to optimize the device. Therefore the triangular potential well approximation does not apply to III-V MOSFETs. Even so, we can still derive some general analytical results directly from Schrödinger’s equation and Poisson’s equation.

According to the quantum-mechanical theory, one is able to expand $Q_i$ into $\sum_n Q_{i,n}$, where $Q_{i,n}$ is the inversion charge associated with the $n$-th subband of the quantum well formed by both band offset (structural confinement) and transverse electrical field (electrical confinement). If we define

$$C_{i,n} = \frac{d (-Q_{i,n})}{d\psi_s} \quad (6.15)$$

it is straightforward to get

$$C_i = \sum_n C_{i,n} \quad (6.16)$$

Since the surface potential $\psi_s$ is determined by the bending of conduction band
edge at the semiconductor/insulator interface, it is obvious that

\[ C_{i,n} = \frac{d(-Q_{i,n})}{d\psi_s} = \frac{d(-Q_{i,n})}{d(E_f-E_n(0)/q)} \]  

(6.17)

where \( E_f \) is the Fermi level in the semiconductor, and \( E_c(0) \) is the surface conduction band edge. We can further decompose \( C_{i,n} \) like this:

\[
C_{i,n} = \frac{d(-Q_{i,n})}{d\left(\frac{E_f-E_n}{q}\right)} + \frac{d(-Q_{i,n})}{d\left(\frac{E_n-E_c(0)}{q}\right)}
\]

(6.18)

where \( E_n \) is the eigen-energy of the \( n \)-th subband. We define

\[ C_{DOS,n} = \frac{d(-Q_{i,n})}{d\left(\frac{E_f-E_n}{q}\right)} \]  

(6.19)

since it is the capacitance associated with DOS for the \( n \)-th subband. The other term is much more complicated, which contains the information of the solutions of Schrödinger’s equation and Poisson’s equation. To get its exact value, the two master equations must be solved self-consistently. But anyway it is the capacitance associated with quantum well confinement effect for the \( n \)-th subband, so we define

\[ C_{qw,n} = \frac{d(-Q_{i,n})}{d\left(\frac{E_n-E_c(0)}{q}\right)} \]  

(6.20)

Finally, equation (6.16) is explicitly written as

\[ C_i = \sum_n \frac{1}{C_{DOS,n}} + \frac{1}{C_{qw,n}} \]  

(6.21)

Combining equations (6.3), (6.5), and (6.21), a more comprehensive equivalent circuit model of an MOS capacitor can be obtained, as illustrated in Fig. 6.6.
As we mentioned, $C_{DOS,n}$ is the capacitance associated with DOS for the $n$-th subband. Here, we will show in details why it is the DOS capacitance. To unveil the physical meaning and derive the analytical expression, we need to make a couple of necessary approximations: 1) No electrons penetrate into the insulator, i.e., the insulator has infinite barrier; 2) the whole device has parabolic dispersion relation. Under these approximations, it is well known based on the quantum-mechanical theory that

$$Q_{i,n} = -\frac{q m^*_{d,n}}{\pi \hbar^2} kT \ln \left[ 1 + \exp \left( \frac{E_f - E_n}{kT} \right) \right]$$  \hspace{1cm} (6.22)

where $m^*_{d,n}$ is the DOS effective mass for the $n$-th subband. For materials with subband degeneracy (e.g., Si), degeneracy $g$ is incorporated into $m^*_{d,n}$.
Substituting equation (6.22) into equation (6.19) yields

\[ C_{DOS,n} = q^2 \frac{m_{d,n}^*}{\pi \hbar^2} \frac{1}{1 + \exp \left( \frac{E_n - E_f}{kT} \right)} \] (6.23)

The last term \(1/[1 + \exp((E_n - E_f)/kT)]\) is in the form of Fermi-Dirac distribution function, which approaches unity when \(E_f\) is more than a few \(kT/q\) above \(E_n\). We know that \(m_{d,n}^*/\pi \hbar^2\) is usually recognized as the 2-D DOS, therefore it is clear now why (6.19) is called the DOS capacitance.

### 6.2.2 Capacitance relation forced by Poisson’s equation

We will start from the generalized 1-D Poisson’s equation given by [84]

\[ \frac{d^2 \psi}{dx^2} = \frac{q}{\epsilon(x)} \left[ n(x) + N_a(x) - p(x) - N_d(x) \right] \] (6.24)

where fully donor/acceptor ionization is assumed. Poisson’s equation always holds whether quantum-mechanical theory is turned on or off. For simplicity, we assume donor-type doping concentration \(N_d(x) = 0\). We consider the region between gate-insulator/channel interface \((x = 0)\) and substrate metal contact \((x = W_m)\).

In between, we can have doped or undoped semiconductor layers and possibly insulator layers. However, to have a neat final result of capacitance relation, we assume all these materials have the same dielectric constant \(\epsilon_s\). This is a reasonable approximation for III-V MOSFETs. Then the 1-D Poisson’s equation is reduced to

\[ \frac{d^2 \psi}{dx^2} = \frac{q}{\epsilon_s} [n(x) + N(x)], \quad 0 < x < W_m \] (6.25)
where depletion charge density $N(x)$ is defined as

$$N(x) = N_a(x) - p(x)$$

(6.26)

Integrating (6.25) once, we obtain

$$\frac{d\psi(x)}{dx} = \frac{d\psi}{dx} \bigg|_{x=W_m} - \int_x^{W_m} \frac{q}{\varepsilon_s} \left[ n(x') + N(x') \right] dx'$$

(6.27)

According to Gauss’s Law, one know $\frac{d\psi}{dx} \bigg|_{x=W_m} = Q_m/\varepsilon_s$, where $Q_m$ is the charge on the substrate metal contact. Since $Q_m$ equals 0 in general, we can further integrate (6.27) and get

$$\psi(W_m) - \psi(0) = \int_0^{W_m} \left\{ - \int_x^{W_m} \frac{q}{\varepsilon_s} \left[ n(x') + N(x') \right] dx' \right\} dx$$

(6.28)

Using integration by parts, one have

$$\int_0^{W_m} \left[ \int_x^{W_m} \frac{q}{\varepsilon_s} n(x') dx' \right] dx = \frac{q}{\varepsilon_s} \int_0^{W_m} x \cdot n(x) dx$$

(6.29)

$$\int_0^{W_m} \left[ \int_x^{W_m} \frac{q}{\varepsilon_s} N(x') dx' \right] dx = \frac{q}{\varepsilon_s} \int_0^{W_m} x \cdot N(x) dx$$

(6.30)

Finally,

$$\psi(W_m) - \psi_s = -\frac{q}{\varepsilon_s} \int_0^{W_m} x \cdot n(x) dx - \frac{q}{\varepsilon_s} \int_0^{W_m} x \cdot N(x) dx$$

(6.31)

where $\psi_s = \psi(0)$ is the surface potential, and $\psi(W_m)$ is the potential at substrate metal contact interface, which should be a constant.

Now differentiating (6.31) with respect to $\psi_s$, one obtain

$$1 = \frac{x_{av}}{\varepsilon_s} \frac{d}{d\psi_s} \int_0^{W_m} q \cdot n(x) dx + \frac{W_{av}}{\varepsilon_s} \frac{d}{d\psi_s} \int_0^{W_m} q \cdot N(x) dx$$

(6.32)
where \( x_{av} \) is defined as

\[
x_{av} = \frac{\int_0^{W_m} x \cdot \frac{dn(x)}{d\psi_s} dx}{\int_0^{W_m} \frac{dn(x)}{d\psi_s} dx}
\]

and \( W_{av} \) is defined as

\[
W_{av} = \frac{\int_0^{W_m} x \cdot \frac{dN(x)}{d\psi_s} dx}{\int_0^{W_m} \frac{dN(x)}{d\psi_s} dx}
\]

Based on the above definitions, \( x_{av} \) and \( W_{av} \) can be physically understood as the centroid of the differential inversion and depletion charge, respectively. Because

\[
Q_i = -\int_0^{W_m} q \cdot n(x) dx \quad \text{and} \quad Q_d = -\int_0^{W_m} q \cdot N(x) dx
\]

(6.32) becomes

\[
\frac{x_{av}}{\epsilon_s} C_i + \frac{W_{av}}{\epsilon_s} C_d = 1
\]

(6.35)

This is a capacitance relation forced by Poisson’s equation only, i.e., independent of Schrödinger’s equation.

### 6.2.3 Simplified picture

As we have shown previously, \( C_{DOS,n} \) has very clear physical meaning, but not \( C_{qw,n} \). The complexity of the self-consistent solutions of Schrödinger’s equation and Poisson’s equation has been fully lumped into \( C_{qw,n} \). The coupling between populations in different subbands is reflected in \( C_{qw,n} \), and this coupling prevents us from developing a compact analytical result for \( C_{qw,n} \). Fortunately, in modern ultra scaled MOSFETs, it is reasonable to make the assumption that only the lowest subband is occupied. Under this approximation, the comprehensive equivalent circuit model of an MOS capacitor shown in Fig. 6.6 is reduced to a simplified one, as illustrated in Fig. 6.7.
Figure 6.7: A simplified equivalent circuit model of an MOS capacitor under the one subband approximation.

According to the one subband approximation, one can obtain

\[
C_{qw,1} = \frac{d(-Q_{t,1})}{d\psi_s} = \frac{C_i}{d\left(\frac{E_1-E_c(0)}{q}\right)/d\psi_s} = \frac{C_i}{d\psi_s} (6.36)
\]

Following the quantum-mechanical theory, we get

\[
d\left(\frac{E_1-E_c(0)}{q}\right)/d\psi_s = d\left(\langle \phi_1| \frac{H-E_c(0)}{q} |\phi_1\rangle \right)/d\psi_s = \langle \phi_1| \frac{H-E_c(0)}{q} |\phi_1\rangle (6.37)
\]

where \(H\) represents the Hamiltonian of the quantum well, and \(|\phi_1\rangle\) is the eigen wavefunction corresponding to the lowest subband. It can be approved that

\[
d\langle \phi_1| \left(\frac{H-E_c(0)}{q}\right) |\phi_1\rangle = \langle \phi_1| \left(\frac{H-E_c(0)}{q}\right) d|\phi_1\rangle/d\psi_s = 0 (6.38)
\]

Since the kinetic operator in Hamiltonian \(H\) is independent of \(\psi_s\) and the band-offsets between different materials are invariables, it is straightforward to
have
\[ \langle \phi_1 | d \left( \frac{H - \mathcal{E}_c(0)}{q} \right) | \phi_1 \rangle = \langle \phi_1 | d(\psi_s - \psi(x)) | \phi_1 \rangle \] (6.39)

Integrating equation (6.27) once yields
\[ \psi_s - \psi(x) = \frac{q}{\epsilon_s} \left[ \int_0^x x' n(x') dx' + \int_0^x x' N(x') dx' + x \int_x^{W_m} n(x') dx' + x \int_x^{W_m} N(x') dx' \right] \] (6.40)

Substituting (6.40) into (6.39), we will have four terms, which are calculated separately as follows:
\[ \langle \phi_1 | \frac{d}{d\psi_s} \int_0^x x' n(x') dx' | \phi_1 \rangle = \int_0^{W_m} \left[ \int_0^x x' \frac{dn(x')}{d\psi_s} dx' \right] dx \]
\[ = \int_0^{W_m} \frac{n(x)}{\int_0^{W_m} n(x) dx} \left( \int_0^x x' \frac{dn(x')}{d\psi_s} dx' \right) dx \]
\[ = \int_0^{W_m} \frac{dn(x)}{x \frac{d}{d\psi_s} dx} \]
\[ - \int_0^{W_m} \frac{dn(x)}{x \frac{d}{d\psi_s} dx} \left( \int_0^x \frac{n(x')}{\int_0^{W_m} n(x) dx} dx' \right) dx \] (6.41)

The second step is based on the one subband assumption and infinite insulator barrier approximation. The third step is simply integration by parts.
\[ \langle \phi_1 | \frac{d}{d\psi_s} \int_0^x x' N(x') dx' | \phi_1 \rangle = \int_0^{W_m} \frac{n(x)}{\int_0^{W_m} n(x) dx} \left( \int_0^x x' \frac{dN(x')}{d\psi_s} dx' \right) dx \]
\[ = \int_0^{W_m} \frac{x dN(x)}{d\psi_s} dx \]
\[ - \int_0^{W_m} \frac{x dN(x)}{d\psi_s} \left( \int_0^x \frac{n(x')}{\int_0^{W_m} n(x) dx} dx' \right) dx \]
\[ = 0 \] (6.42)

Here the third step is an approximation which requires inversion charge \( n(x) \) and differential depletion charge \( dN(x)/d\psi_s \) are spatially separated. It is a reason-
able approximation because inversion charge is always close to the gate insulator whereas the differential depletion charge is located around the end of depletion region which is farther away from the gate insulator.

\[
\langle \phi_1 | d \left[ x \int_x^{W_m} n(x') dx' \right] \frac{d}{d\psi_s} | \phi_1 \rangle = \int_0^{W_m} \frac{n(x)}{\int_0^{W_m} n(x) dx} \left( \int_x^{W_m} \frac{dN(x')}{d\psi_s} dx' \right) dx
\]

(6.43)

\[
\langle \phi_1 | d \left[ x \int_x^{W_m} N(x') dx' \right] \frac{d}{d\psi_s} | \phi_1 \rangle = \int_0^{W_m} \frac{n(x)}{\int_0^{W_m} n(x) dx} \left( \int_x^{W_m} \frac{dN(x')}{d\psi_s} dx' \right) dx
\]

(6.44)

Again, in the third step of equation (6.44), we adopt the approximation that \( n(x) \) and \( dN(x)/d\psi_s \) are spatially separated.

Combining equations (6.37)-(6.44), one obtain

\[
\frac{d \left( \frac{E_1 - E_c}{q} \right)}{d\psi_s} = \frac{q}{\epsilon_s} (x_{av} - x_{sh}) \int_0^{W_m} \frac{dn(x)}{d\psi_s} dx + \frac{q}{\epsilon_s} x_c \int_0^{W_m} \frac{dN(x)}{d\psi_s} dx
\]

\[
= \frac{x_{av} - x_{sh}}{\epsilon_s} C_i + \frac{x_c}{\epsilon_s} C_d
\]

(6.45)

where \( x_{av} \) is defined by equation (6.33), and \( x_c \) and \( x_{sh} \) are defined as follows:

\[
x_c = \int_0^{W_m} \frac{x \cdot n(x) dx}{\int_0^{W_m} n(x) dx}
\]

(6.46)

\[
x_{sh} = \frac{\int_0^{W_m} \frac{dn(x)}{d\psi_s} \left( \int_0^x \frac{(x-x') n(x')}{\int_0^{W_m} n(x) dx} dx' \right) dx}{\int_0^{W_m} \frac{dn(x)}{d\psi_s} dx}
\]

(6.47)
$x_c$ is the centroid of the inversion charge, and is different from $x_{av}$, which should be understood as the centroid of the differential inversion charge. $x_{sh}$ is very complicated, but we know that it is like the square root of a second order central moment. In other words, $x_{sh}$ is related to the spreading of the inversion charge. To show this point manifestly, we can assume Gaussian distribution approximately, i.e., assume

$$
\frac{n(x)}{\int_0^W n(x)dx} \approx \frac{1}{\sqrt{2\pi}\sigma_c} \exp \left[ -\frac{(x-x_c)^2}{2\sigma_c^2} \right]
$$

$$
\frac{d\nu(x)}{d\psi_s} \approx \frac{1}{\sqrt{2\pi}\sigma_{av}} \exp \left[ -\frac{(x-x_{av})^2}{2\sigma_{av}^2} \right]
$$

Then we can calculated $x_{sh}$ as

$$
x_{sh} = \frac{\int_0^W \frac{d\nu(x)}{d\psi_s} \left( \int_0^x \frac{dx'}{\int_0^W n(x')dx'} \right) dx}{\int_0^W \frac{d\nu(x)}{d\psi_s} dx}
$$

$$
\approx \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_{av}} \exp \left[ -\frac{(x-x_{av})^2}{2\sigma_{av}^2} \right] \int_{-\infty}^x \frac{dy}{\sqrt{2\pi}\sigma_c} \exp \left[ -\frac{(y-x_c)^2}{2\sigma_c^2} \right] dy
$$

$$
= \sqrt{\frac{\sigma_c^2 + \sigma_{av}^2}{2\pi}} \exp \left[ -\frac{(x_c-x_{av})^2}{2(\sigma_c^2 + \sigma_{av}^2)} \right]
$$

$$
- \frac{x_c - x_{av}}{2\pi\sigma_c\sigma_{av}} \int_{-\infty}^{\infty} \int_{-\infty}^x \exp \left[ -\frac{(y-y_{av})^2}{2\sigma_{av}^2} \right] \exp \left[ -\frac{(y-x_c)^2}{2\sigma_c^2} \right] dy dx
$$

$$
= \sqrt{\frac{\sigma_c^2 + \sigma_{av}^2}{2\pi}} - \frac{x_c - x_{av}}{2} + O(\delta^2)
$$

(6.49)

$x_c$ is very close to $x_{av}$, and $\sigma_c$ is very close to $\sigma_{av}$. So $(x_c - x_{av}), (\sigma_c - \sigma_{av}) \sim O(\delta)$. In the last step of (6.49), we only keep to the first order of $\delta$. Now, it is clear that the primary part of $x_{sh}$ is $\sqrt{\sigma_c^2 + \sigma_{av}^2}/2\pi$, which reinforces our previous understanding of $x_{sh}$ as a second order moment. The difference between $x_c$ and $x_{av}$ will also make a small part of contribution.
Substituting equation (6.45) into (6.36) yields

\[ C_{qw,1} = \frac{\epsilon_s}{x_{av} - x_{sh} + x_c \frac{C_d}{C_i}} \] (6.50)

Because the capacitances are connected in series except \( C_d \), it is more convenient to use the concept of effective thickness, which is defined by the reciprocal of capacitance normalized to the dielectric constant \( \epsilon_s \). For example, we define

\[ C_{DOS,1} = q^2 \frac{m_{d,1}^*}{\pi \hbar^2} \frac{1}{1 + \exp \left( \frac{E_1 - E_f}{kT} \right)} = \frac{\epsilon_s}{x_{DOS}} \] (6.51)

According to the one subband approximation, we have

\[ C_i = \frac{1}{C_{DOS,1}^{-1} + C_{qw,1}^{-1}} \] (6.52)

Combining equations (6.50), (6.51), (6.52), and the capacitance relation forced by Poisson’s equation (6.35), one finally obtain

\[ C_i = \frac{\epsilon_s \left( 1 - \frac{x_{av}}{W_{av}} \right)}{x_{DOS} + x_{av} - x_{sh} - x_c \frac{x_{av}}{W_{av}}} \triangleq \frac{\epsilon_s}{x_i} \] (6.53)

i.e., \( x_i \) is the effective thickness associated with the inversion layer capacitance \( C_i \) and given by

\[ x_i = x_{av} + \frac{W_{av}}{W_{av} - x_c} \left( x_{DOS} - x_{sh} \right) \] (6.54)

Given following effective thickness definition

\[ C_{ox} \triangleq \frac{\epsilon_s}{x_{ox}} \] (6.55)

\[ C_{inv} \triangleq \frac{\epsilon_s}{x_{inv}} \] (6.56)

where \( C_{inv} \) is defined by equation (6.7), one also have

\[ x_{inv} = x_{av} + \frac{W_{av}}{W_{av} - x_c} \left( x_{DOS} - x_{sh} \right) + x_{ox} + \frac{x_{ox}}{W_{av} - x_c} \left( x_{DOS} - x_{sh} \right) \] (6.57)
Equations (6.54) and (6.57) represent our final analytical results for capacitances in terms of effective thicknesses, which help us to physically understand the capacitances in III-V MOSFETs, especially the inversion layer capacitance. Since these analytical expressions are somewhat complicated, the following simplified approximations have been proposed.

Approximation version A, which assumes $C_{DOS,1} = \infty$ and $x_{sh} = 0$:

\[
\begin{align*}
    x_{inv} &= x_{ox} + x_{av} \\
    x_i &= x_{av}
\end{align*}
\]  

(6.58)

Approximation version B, which assumes $C_{qw,1} = \frac{\epsilon_s}{x_{av}}$:  

\[
\begin{align*}
    x_{inv} &= x_{av} + x_{DOS} + x_{ox} + \frac{x_{DOS}}{W_{av}} x_{ox} \\
    x_i &= x_{av} + x_{DOS}
\end{align*}
\]  

(6.59)

Approximation version C, which assumes $x_c = x_{av}$ and $x_{sh} = 0$:

\[
\begin{align*}
    x_{inv} &= x_{av} + \frac{W_{av}}{W_{av} - x_{av}} x_{DOS} + x_{ox} + \frac{x_{ox}}{W_{av} - x_{av}} x_{DOS} \\
    x_i &= x_{av} + \frac{W_{av}}{W_{av} - x_{av}} x_{DOS}
\end{align*}
\]  

(6.60)

We have compared all the three versions of approximations with the accurate result in Fig. 6.8. Here, the layer structure for the simulated device is Metal Gate/2nm Insulator/5nm Semiconductor/7nm Insulator/100nm p+ Semiconductor/Back Contact. First of all, the accurate result ((6.54) and (6.57)) has been verified by the numerical simulations, which is obtained by ISE-DESSIS. Obviously, approximation A is only good for very large $C_{DOS,1}$ (e.g., silicon). It is definitely...
Figure 6.8: $x_i$, the effective thickness associated with inversion layer capacitance, as a function of “effective” DOS effective mass. “Effective” DOS effective mass means $m_{d,1}^*/[1 + \exp(E_1-E_F/kT)]$.

unsuitable for the other cases. The accuracy of approximation B depends on the difference between $W_{av}/x_c$ and $x_{DOS}/x_{sh}$. Approximation C is always larger than numerical results, and it works very well for small $C_{DOS,1}$. According to Fig. 6.8, we know that $x_{sh}$ is not totally negligible, but approximation B and C are good approximations for III-V MOSFETs.

6.2.4 Degradation of gate insulator capacitance

For Si, $C_{DOS,1}$ is very large above threshold. Therefore $C_i$ is limited by the quantum confinement effect, and $C_d$ is negligible due to the screening of inversion charge. But for III-V, the picture is quite different because of the small $C_{DOS,1}$.
III-V MOSFETs, due to the non-negligibility of $C_d$, $C_{ox}$ is shared by $C_i$ and $C_d$. By comparing the expressions of $x_i$ and $x_{inv}$ in the equations (6.54) and (6.57), we can find that the contribution by gate insulator is not $x_{ox}$.

To show the importance of $C_{ox}$ degradation in terms of inversion charge gate capacitance $C_{inv}$, we will use the approximation version B, which is a good approximation for III-V MOSFETs. The typical numbers for the baseline design of III-V MOSFETs, which will be discussed in detail in the next chapter, are given in effective oxide thickness (EOT), i.e., normalized to $3.9\varepsilon_0$ (dielectric constant of SiO$_2$) instead of semiconductor dielectric constant: $x_{ox} \approx 0.6 \text{ nm EOT}$, $x_{DOS} \approx 1.3 \text{ nm EOT}$, $x_{av} \approx 0.4 \text{ nm EOT}$, and $W_{av} \approx 3.2 \text{ nm EOT}$. All the contributions to $C_{inv}$ are illustrated by the bar plot in Fig. 6.9. We can see that $C_{ox}$ has a 30% degradation, which results in 10% loss of the total inversion charge gate capacitance $C_{inv}$. From this bar plot, we can also imagine that the thin gate insulator has marginal benefit for $C_{inv}$ because the large $x_{DOS}$ is not scalable, but small gate insulator EOT is needed to scale down gate length from SCEs point of
6.3 Degradation of DOS Capacitance in Non-Equilibrium: Ballistic MOSFETs

Till now, we only consider the MOS capacitors or the MOSFETs in equilibrium. For a ballistic MOSFET, when the drain is applied a positive voltage (assume nFET), we can expect the degradation of DOS capacitance in non-equilibrium. According to the theory of ballistic MOSFETs [96, 97, 98], the equivalent circuit model can be used at the virtual source because of the quasi-equilibrium condition. Virtual source is the point with the maximum electron energy barrier, as indicated in Fig. 6.10. However, in this case, $C_{DOS,1}$ is divided into two parts, one
Figure 6.11: A simplified equivalent circuit model under the one subband approximation applicable to the virtual source of a ballistic MOSFET.

is connected to the source, the other is connected to the drain. They are different from each other since the Fermi levels in the source and drain are different. The simplified equivalent circuit model for this case is shown in Fig. 6.11, where

\[
C_{DOS,1}^+ = \frac{q^2 m_{d,1}^*}{2\pi \hbar^2} \frac{1}{1 + \exp\left(\frac{E_1 - E_{FS}}{kT}\right)}
\]

(6.61)

\[
C_{DOS,1}^- = \frac{q^2 m_{d,1}^*}{2\pi \hbar^2} \frac{1}{1 + \exp\left(\frac{E_1 - E_{FD}}{kT}\right)}
\]

(6.62)

where \(E_{FS}\) and \(E_{FD}\) are Fermi levels of the source and drain, respectively.

DOS capacitance is the primary component in III-V MOSFETs, which has been indicated in Fig. 6.9, so the degradation of DOS capacitance in non-equilibrium is significant. In the limit of high drain bias, DOS capacitance is reduced by a factor of 2, which results in the inversion charge degradation at high drain. This phenomenon can be predicted by the gate equation at the virtual source, which is given by

\[
V_g - V_t = \frac{E_{FS} - E_1}{q} - \frac{Q_i^+ + Q_i^-}{C_{ox}'}
\]

(6.63)
Figure 6.12: \( N_s = - (Q_i^+ + Q_i^-)/q \) as a function of \( V_{ds} \) obtained from gate equation (6.63) for (a) InGaAs; (b) Si. In both cases, we use 1.2 nm EOT for \( C_{ox}' \). But \( m_{d,1}^* \) for InGaAs and Si are 0.04\( m_0 \) and 0.4\( m_0 \), respectively. \( m_0 \) is electron mass.

where

\[
Q_i^+ = \frac{qkTm_{d,1}^*}{2\pi\hbar^2} \ln \left[ 1 + \exp \left( \frac{E_{FS} - E_1}{kT} \right) \right]
\]  

(6.64)
\[ Q_i^- = -\frac{qkTm_{d1}^2}{2\pi\hbar^2} \ln \left[ 1 + \exp \left( \frac{E_{FS} - E_1 - qV_{ds}}{kT} \right) \right] \]  

(6.65)

Here, \( V_t \) is the threshold voltage. \( C'_{ox} \) is different from \( C_{ox} \). \( C_{qw,1} \) and the degradation of \( C_{ox} \) have been lumped into \( C'_{ox} \). According to the gate equation, the inversion charge at virtual source is dependent on \( V_{ds} \), especially for small DOS. This has been demonstrated in Fig. 6.12, where \( N_s = -(Q_i^+ + Q_i^-)/q \) is plotted as a function of \( V_{ds} \) for both In\textsubscript{0.53}Ga\textsubscript{0.47}As and Si. We can observe \( N_s \) decreases as \( V_{ds} \) increases in both cases. But inversion charge degradation in In\textsubscript{0.53}Ga\textsubscript{0.47}As ballistic MOSFET is much more significant due to small DOS.

### 6.4 Effects of Interface States

As mentioned in Chapter 1, primarily because of the bad quality of dielectric/semiconductor interface, III-V materials were not considered as good candidates for CMOS used in logic circuits until the conventional Si bulk CMOS scaling limit is being approached. Although the interface quality keeps being improved, interface states still significantly degrade device performance in current stage. In this section, we will discuss two major effects of interface states with the help of equivalent capacitance circuit model: 1) subthreshold slope degradation in the subthreshold region; 2) \( Q_i \) vs \( V_g \) slope degradation, i.e., inversion charge gate capacitance \( C_{inv} \) degradation above threshold.

The equivalent circuit model of an MOS capacitor with interface states is shown in Fig. 6.13. Compared with Fig. 6.3, an extra capacitance associated with
interface traps $C_{it}$ has been added. The definition of $C_{it}$ is

$$
C_{it}(\psi_s) = \frac{d(-Q_{it})}{d\psi_s} \tag{6.66}
$$

where $Q_{it}$ is the total interface charge. Assuming FD distribution is a step function, it is straightforward to get

$$
C_{it}(\psi_s) = q^2 D_{it}(\psi_s) \tag{6.67}
$$

where $D_{it}$ is the so-called interface state density, and usually given in the unit of cm$^{-2}$eV$^{-1}$. Let us first examine $C_{inv}$ in the presence of interface traps. According to the definition in (6.7), one obtain

$$
C_{inv} = C_{ox} C_i \frac{C_{ox} C_i}{C_{ox} + C_i + C_d + C_{it}} \tag{6.68}
$$

This result manifests the degradation of $C_{inv}$ due to interface traps. We can also conclude that, even given same gate insulator capacitance and interface state density, III-V material, e.g., In$_{0.53}$Ga$_{0.47}$As suffers more from interface traps than
Si in terms of $C_{inv}$ degradation, because In$_{0.53}$Ga$_{0.47}$As has smaller $C_i$ resulting from smaller DOS effective mass. Another important effect is the subthreshold slope degradation. Subthreshold slope in the long channel limit can be calculated by

$$S = \frac{dV_g}{d\psi_s} \times 60 \text{ mV/Dec} = \frac{C_i}{C_{inv}} \times 60 \text{ mV/Dec}$$

$$= \left(1 + \frac{C_i + C_d + C_{it}}{C_{ox}} \right) \times 60 \text{ mV/Dec}$$

$$\approx \left(1 + \frac{C_d + C_{it}}{C_{ox}} \right) \times 60 \text{ mV/Dec} \quad (6.69)$$

It is obvious that the subthreshold slope degradation only depends on $C_{it}/C_{ox}$ and has nothing to do with the channel material. The good news is that reducing gate insulator EOT can help to suppress effects of interface states on the device performance.

### 6.5 Summary

In this chapter, we have proposed a comprehensive equivalent capacitance circuit model of an MOS capacitor. With the one subband approximation, analytical results for the capacitance components in the equivalent circuit model can be developed in terms of effective thickness, by investigating both Poisson’s equation and Schrödinger’s equation. These analytical results can help us to physically understand the capacitances in III-V MOSFETs, especially the inversion layer capacitance. Besides, these analytical results can predict the degradation of gate insulator capacitance in inversion charge gate capacitance $C_{inv}$, and degradation
of DOS capacitance in non-equilibrium in ballistic MOSFETs. The effects of interface traps have also been examined by analyzing the equivalent circuit model with interface states capacitance incorporated.
Device Design of III-V MOSFETs for Sub-22 nm Scaling

7.1 Selection of Basic Device Structure

In this chapter, we will present our results on the device design of III-V MOSFETs for sub-22 nm scaling. First of all, the selection of basic device structure needs to be made. For Si MOSFETs, multiple-gate structures have been proposed to improve control of SCEs, but for III-V MOSFETs, we are not at that stage yet. Here, we will only consider planar structures for III-V MOSFETs, including bulk-like structure, thick-BOX-SOI-like structure, and thin-BOX-SOI-like structure. Actually, planar structures can take advantage of one important benefit of III-V materials: heterostructures. By tailoring composition, heterostructures offer additional degrees of freedom to optimize the device. We would like to com-
pare the scaling potential of different planar structures with specific considerations of III-V materials.

Without doubt, we should start with the bulk-like structure. Actually, III-V MOSFETs based on the bulk-like structure have already been successfully fabricated and excellent device performance has been demonstrated even in relatively long channel MOSFETs [49]-[58]. SCEs are the most important effects limit MOSFET scalability, and can be commendably indicated by a parameter so-called generalized scale length [82]. The eigenvalue equation for the bulk-like structure is repeated here [82]:

$$\epsilon_s \tan(\frac{\pi t_{ox}}{\lambda}) + \epsilon_{ox} \tan(\frac{\pi W_d}{\lambda}) = 0$$

(7.1)

where generalized scale length \(\lambda\) is the largest solution of the above equation, \(\epsilon_s\) is the dielectric constant of semiconductor, and \(W_d\) represents the maximum depletion width. The scale length theory shows that the minimum potential has an effective channel length dependence of \(\sim \exp(-\pi L/2\lambda)\) to the first order, so the ratio \(L/\lambda\) is a fundamental indicator of SCEs in MOSFETs. Numerical simulations reveal that the minimum channel length for transistors with acceptable SCEs should be larger than \(\sim 1.5\lambda\) [84]. Therefore, to achieve very short gate length, e.g., sub-22 nm, very high body doping is required to determine a small \(W_d\), at least in the level of \(1 \times 10^{19} \text{ cm}^{-3}\). However, it is well known that high doping in the channel region is harmful to the device performance. Besides the significant mobility degradation, the high body doping leads to the concern of band-to-band
tunneling [84, 99]. The band-to-band tunneling current density for a p-n junction is given by [99]

$$J_{b-b} = \frac{\sqrt{2m^* q^3 \varepsilon V_{app}}}{4\pi^3 \hbar^2 E_g^{1/2}} \exp \left( -\frac{4\sqrt{2m^* E_g^{3/2}}}{3q\varepsilon \hbar} \right)$$

where $\varepsilon$ is the electric field, and $V_{app}$ is the applied reverse voltage across the junction. Since equation (7.2) indicates that $J_{b-b}$ increases exponentially with decreasing bandgap $E_g$, III-V MOSFETs usually suffer more from band-to-band tunneling leakage, because narrow band III-V material, e.g., In$_{0.53}$Ga$_{0.47}$As, is used for channel region to achieve high transport mobility. Under this circumstance, due to the limitation imposed by band-to-band tunneling, the body doping cannot be too high and thus the scalability of bulk-like structure based III-V MOSFETs is severely affected. Presently, the body doping of III-V MOSFETs is in the level of $1 \times 10^{17}$ cm$^{-3}$ [49]-[58], therefore the gate lengths are usually at least a few hundred nanometers.

Since bulk-like structure is not a good candidate for extremely scaled III-V MOSFETs due to the band-to-band tunneling leakage current limitation, we should consider SOI-like structure. Of course, the partially-depleted SOI structure will not be considered where the SCEs are also controlled by body doping. In other words, only fully-depleted SOI-like structure will be considered. As known, SOI is the abbreviation for “silicon-on-insulator”. In III-V MOSFETs, “silicon” will be replaced by high-mobility III-V materials, but how about “insulator”? To take the advantage of III-V material integration, it seems like a formidable mission to put
an insulator layer at the backside with high interface quality, considering that the
top gate insulator is not perfect yet. But fortunately, we have wide bandgap III-V
materials which can serve as semi-insulator, e.g., InAlAs, and these wide bandgap
materials can be easily integrated with substrate and narrow bandgap channel
materials, even lattice matched sometimes. This is the magic of III-V! In real
“silicon-on-insulator” devices, the bottom insulator region is always called buried-
oxide (BOX). The BOX can be either thick or thin with a ground plane underneath.
The ground plane is a back conductor consists of highly doped semiconductor
region connected to 0 voltage supply, i.e., ground. We will consider these two
types of SOI-like structures for III-V MOSFETs: thick-BOX-SOI-like and thin-
BOX-SOI-like.

The short-channel scaling of thick-BOX-SOI-like structure depends on chan-
nel film thickness, gate insulator thickness and all the dielectric constants, but not the BOX thickness. The generalized scale length model [82] is inapplicable here due to the source/drain-to-channel lateral-field coupling through the BOX as shown in Fig. 7.1. The lateral-field penetration into the BOX depends on the channel length, i.e., the source-to-drain distance, which implies that the “effective” BOX thickness for short-channel device is much thinner than the physical thickness [100, 101]. Although the general scale length theory does not work here, an empirical scaling rule has been obtained for real “silicon-on-insulator” devices based extensive simulation results [100, 101]:

$$L_{\text{min}} \approx 4.5 \left( t_{\text{si}} + \frac{\epsilon_{\text{si}}}{\epsilon_{\text{ox}} t_{\text{ox}}} \right)$$

(7.3)

where $L_{\text{min}}$ is defined as the channel length with the $V_t$ roll-off of 100 mV, which is relatively a loose criterion. Equation (7.3) assumes $\epsilon_{\text{BOX}} = 3.9\epsilon_0$, where $\epsilon_{\text{BOX}}$ represents the dielectric constant of BOX. Actually, SCEs are significantly affected by $\epsilon_{\text{BOX}}$. If we assume $\epsilon_{\text{BOX}} = 11.7\epsilon_0$, which is similar to the scenario in III-V MOSFETs, another quite different empirical scaling rule can be obtained [101]:

$$L_{\text{min}} \approx 7.9 \left( t_{\text{si}} + \frac{\epsilon_{\text{si}}}{\epsilon_{\text{ox}} t_{\text{ox}}} \right)$$

(7.4)

which indicates that SCEs are aggravated high-$\kappa$ BOX. Although equation (7.4) might be not able to be applied to III-V directly, we can still expect similar trends for thick-BOX-SOI-like III-V MOSFETs.

However, the generalized scale length theory is applicable to thin-BOX-
SOI-like structure, with the eigenvalue equation given by [82]

$$\frac{1}{\epsilon_{ox}} \tan \left( \frac{\pi t_{ox}}{\lambda} \right) + \frac{1}{\epsilon_s} \tan \left( \frac{\pi t_s}{\lambda} \right) + \frac{1}{\epsilon_{BOX}} \tan \left( \frac{\pi t_{BOX}}{\lambda} \right) = \frac{\epsilon_s}{\epsilon_{ox} \epsilon_{BOX}} \tan \left( \frac{\pi t_{ox}}{\lambda} \right) \tan \left( \frac{\pi t_s}{\lambda} \right) \tan \left( \frac{\pi t_{BOX}}{\lambda} \right)$$  \hspace{1cm} (7.5)

where \( t_s \) is the thickness of channel semiconductor film, and \( t_{BOX} \) is the thickness of BOX (or wide bandgap III-V material). This is a complicated transcendental equation. But if we assume \( \epsilon_{ox} = \epsilon_s = \epsilon_{BOX} \), a simple relation can be obtained as

$$\lambda = t_s + t_{ox} + t_{BOX}$$  \hspace{1cm} (7.6)

which indicates that, from SCEs point of view, \( t_{BOX} \) is as important as the thickness of front gate insulator \( t_{ox} \). In other words, SCEs can be effectively improved by scaling down \( t_{BOX} \). If \( t_{ox}, t_{BOX} \ll t_s \) is assumed, one have

$$\lambda = t_s + \frac{\epsilon_s}{\epsilon_{ox}} t_{ox} + \frac{\epsilon_s}{\epsilon_{BOX}} t_{BOX}$$  \hspace{1cm} (7.7)

In contrast to a thick-BOX-SOI-like device, a thin-BOX-SOI-like device still needs high-\( \kappa \) material to reduce the effective oxide thickness (EOT) of the BOX, and thus improve SCEs. Based on this point, thin-BOX-SOI-like structure is more suitable for III-V MOSFETs than thick-BOX-SOI-like structure since wide bandgap III-V “BOX” has relatively high dielectric constant. Moreover, for Si-based device, it is not easy to thin down the BOX, but the thickness of wide bandgap III-V “BOX” can be well controlled in III-V MOSFETs.

The scaling limit of SOI-like structures depends on the limit of channel film thickness and gate insulator EOT. Even with ultra high-\( \kappa \) dielectrics, there
is still a physical limit on gate insulator EOT due to the quantum-mechanical gate tunneling leakage. Actually, high-κ dielectrics always have smaller barrier to resist tunneling. On the other hand, channel film thickness is limited by quantum confinement effect, which results in that the ground state energy and thus the threshold voltage is very sensitive to channel film thickness when it is small. To avoid a large threshold voltage fluctuation, the minimum channel film thickness is determined by how accurate the film thickness can be controlled, i.e., the device-to-device, die-to-die, and wafer-to-wafer variations in film thickness. Besides, the transport mobility decreases as the channel film thickness is reduced. Given the same limit on channel film thickness and gate insulator EOT, it is obvious that thin-BOX-SOI-like structure is much more scalable than thick-BOX-SOI-like structure in terms of gate length.

In conclusion, thin-BOX-SOI-like structure is the best candidate for extremely scaled III-V MOSFETs. It will be shown in the next section that thin-BOX-SOI-like III-V MOSFETs can achieve sub-22 nm scaling without pushing the film thicknesses to the very bottom.

### 7.2 Short Channel Design

We will first introduce our baseline design of III-V MOSFETs for sub-22 nm scaling, as illustrated in Fig. 7.2. This is a thin-BOX-SOI-like structure, and later we will discuss in detail how short the gate length can be scaled down to. Many
kinds of high-\(\kappa\) dielectrics have been considered for III-V MOSFETs, including Al\(_2\)O\(_3\) [49]-[54], HfO\(_2\) [54]-[56], HfAlO [54, 57, 58], ZrO\(_2\) [59], and so on [60, 61]. But till now, it is not clear which one is the best choice for high mobility III-V materials like InGaAs. Therefore, in our baseline design, we simply assume the dielectric constant of gate insulator to be 13\(\varepsilon_0\), which is close to the semiconductor permittivities. The thickness of high-\(\kappa\) gate insulator is 2 nm, and it is corresponding to 0.6 nm EOT. This design consists of a 2.5nm-thick In\(_{0.53}\)Ga\(_{0.47}\)As/2.5nm-thick InP composite channel. Compared with 5nm pure In\(_{0.53}\)Ga\(_{0.47}\)As channel, the composite channel is supposed to provide a few extra benefits. For example, due to wider bandgap of InP, the band-to-band tunneling leakage current can be largely suppressed. Under the channel, there is a 3.5nm-thick lattice-matched InAlAs layer, which serves as the insulating bottom barrier, i.e., the counterpart of BOX in Si SOI devices. From SCEs point of view, this barrier layer is expected
to be very thin. However, although InAlAs has relatively wide bandgap compared with InGaAs, it is still semiconductor rather than insulator, therefore too thin bottom barrier layer results in large leakage current through the ground plane. For the sub-22 nm scaling, all the thicknesses, including high-\(\kappa\) dielectric EOT, channel thickness, and bottom barrier thickness, do not have to be pushed to the physical limits. The p-type InGaAs ground plane beneath the InAlAs barrier is highly doped to \(5 \times 10^{19}\) cm\(^{-3}\). Because the depletion width in ground plane also tends to degrade SCEs on top of barrier thickness, i.e., the effective “BOX” thickness is increased by the depletion region in ground plane, we need very high doping level to obtain a small depletion width. Fortunately, the band-to-band tunneling leakage current between n+ drain and p+ ground plane can be well controlled by the InAlAs barrier thickness. It is important to have a self-aligned ground plane to avoid excessive parasitic capacitance and resistance, so Professor Mark Rodwell of UCSB suggested to apply a shallow Ti implant after gate patterning to damage the p+ regions under the source/drain. Moreover, the baseline design has a high source/drain doping of \(5 \times 10^{19}\) cm\(^{-3}\), which is required to promise a high drain current. This will be discussed in the next section. Finally, metals for both gate and source/drain must be carefully chosen. Especial attention should be paid to the gate metal, because it is a very important way to adjust threshold voltage in this design. Of course, we also have some other ways to adjust threshold voltage, e.g., apply a non-zero voltage to the ground plane, or add a \(\delta\)-doping layer in InAlAs barrier.
In this section, we only put emphasis on the SCEs, i.e., we will focus on the device electrostatic integrity in the subthreshold region. ISE-DESSIS is used for 2-D numerical device simulation. It is not easy to tell directly from 2-D potential distribution whether electrostatic integrity is well maintained or not, so we need to establish reasonable and clear criteria. A regular method is to map electrostatics into drain current, and then extract typical parameters for SCEs from drain current, including threshold voltage ($V_t$) roll-off, DIBL, and subthreshold slope degradation. The criteria should be based on either these parameters or their sensitivity with respect to the gate length. We use drift-diffusion transport model with constant mobility in simulation to calculate current. This model is not suitable for short channel devices, however the task is not to give an accurate estimation of drain current in the subthreshold region but to simply quantify 2-D electrostatics by this mapping technique. Therefore, it is rational to adopt this simplest transport model. Actually, the 2-D electrostatics in the subthreshold region is governed by Poisson’s equation dominantly. Quantum-mechanical effect would make a second-order contribution to SCEs. We tends to incorporate quantum-mechanical effect in our simulation, due to small electron effective mass of In$_{0.53}$Ga$_{0.47}$As and very thin channel film. ISE-DESSIS is able to solve 2-D Poisson’s equation and 1-D Schrödinger’s equation in the direction perpendicular to the insulator/semiconductor interface self-consistently. It must be pointed out that MB distribution is no long a good approximation for In$_{0.53}$Ga$_{0.47}$As, so FD distribution has to be turned on in the simulation. For the quantum-mechanical
Table 7.1: Electron affinity, bandgap, electron effective mass, dielectric constant for InGaAs, InAlAs, InP, and high-\(\kappa\) dielectric in ISE-DESSIS simulation.

<table>
<thead>
<tr>
<th>Material</th>
<th>Electron Affinity</th>
<th>Bandgap</th>
<th>Effective Mass</th>
<th>Dielectric Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>InGaAs</td>
<td>4.70 eV</td>
<td>0.78 eV</td>
<td>0.041 (m_0)</td>
<td>14.00(\epsilon_0)</td>
</tr>
<tr>
<td>InAlAs</td>
<td>4.20 eV</td>
<td>1.46 eV</td>
<td>0.080 (m_0)</td>
<td>12.45(\epsilon_0)</td>
</tr>
<tr>
<td>InP</td>
<td>4.45 eV</td>
<td>1.42 eV</td>
<td>0.079 (m_0)</td>
<td>12.40(\epsilon_0)</td>
</tr>
<tr>
<td>High-(\kappa)</td>
<td>3 eV</td>
<td>6 eV</td>
<td>0.5 (m_0)</td>
<td>13(\epsilon_0)</td>
</tr>
</tbody>
</table>

Based on the simulation results, we can use high drain \(V_t\) roll-off as the simplest criterion to determine the minimum gate length: for example, the magnitude of \(V_t\) roll-off is required to be below 200mV for \(V_{ds} = 1V\). In Fig. 7.3, 7.4, and 7.5, high drain \(V_t\) roll-off \((V_{ds} = 1V)\) is plotted as a function of gate length for baseline design and different device variations. Here, high drain \(V_t\) roll-off is extracted from simulation results with quantum-mechanical effect taken into account. According to the above criterion, the minimum gate length for baseline design is expected to be 19nm roughly. The first type of device variation under consideration is the pure 5nm InGaAs channel. We can see from Fig. 7.3 that SCEs are not sensitive to channel composition. The second type of device variation is the source/drain
Figure 7.3: High drain $V_t$ roll-off ($V_{ds} = 1$V) is plotted as a function of gate length for baseline design and two types of device variations including pure 5nm InGaAs channel and source/drain doping level (both $2 \times 10^{19}$ cm$^{-3}$ and $8 \times 10^{19}$ cm$^{-3}$).

doping level. It is shown in Fig. 7.3 that higher doping ($8 \times 10^{19}$ cm$^{-3}$) tends to aggravate SCEs, and SCEs are relatively sensitive to the source/drain doping level. This fact can be explained by the small DOS of InGaAs: Due to small DOS, the build-in potential between source/drain and channel keeps increasing as doping level increases$^1$, and thus the induced larger lateral electric field can penetrate more into channel. In Fig. 7.4, another two types of variations have been compared with the baseline design. If the thickness of InAlAs bottom barrier is 7nm, then the corresponding minimum gate length is about 2nm larger than that of baseline structure, i.e., 21nm. GaAs cap layer is the layer between high-κ gate

$^1$On the contrary, the build-in potential between source/drain and channel can be treated as a constant in Si MOSFETs
Figure 7.4: High drain $V_t$ roll-off ($V_{ds} = 1\text{V}$) is plotted as a function of gate length for baseline design and two types of device variations including 7nm InAlAs bottom barrier and 0.9nm or 1.5nm GaAs cap layer.

insulator and InGaAs channel. The primary purpose of this layer is to achieve better insulator/semiconductor interface quality. However, adding this cap layer will significantly aggravate SCEs. With a 0.9nm GaAs cap layer, the minimum gate length under the same criterion is roughly 22nm, while a 1.5nm GaAs cap layer corresponds to 24nm minimum gate length. In other words, for every nanometer increase of GaAs cap layer, the minimum gate length increases by $\sim 3\text{nm}$. Fig. 7.5 focuses on the device variations with different dielectric constant or thickness of high-$\kappa$ gate insulator. It is very clear that the minimum gate length is not sensitive to dielectric constant if high-$\kappa$ gate insulator EOT is kept same (e.g., the case labeled “HK=26 T=4nm” or “HK=19.5 T=3nm” in Fig. 7.5). On the other
Figure 7.5: High drain $V_t$ roll-off ($V_{ds} = 1V$) is plotted as a function of gate length for baseline design and device variations with different dielectric constant or thickness of high-$\kappa$ gate insulator.

On the other hand, the minimum gate length is very sensitive to the high-$\kappa$ gate insulator EOT, which has been verified by the two cases in Fig. 7.5 with only dielectric constant increased, i.e., physical thickness of high-$\kappa$ is kept same.

### 7.3 Some Design Considerations Beyond SCEs

In our baseline device design of III-V MOSFETs for sub-22nm scaling shown in Fig. 7.2, a self-aligned shallow Ti implant (or some other similar technique that can achieve the same goal) is suggested to be applied after gate patterning to damage the p+ regions under the source/drain to avoid excessive parasitic capacitance and resistance. This is very important because InAlAs bottom barrier
Figure 7.6: Electron density distribution in III-V MOSFETs of 20nm gate length with $V_{ds} = 1V$: a) baseline design with p+ ground plane only under the gate; b) baseline design but with p+ ground plane under the source/drain as well.

layer is only 3.5nm thick in our baseline structure. If there is p+ ground plane under the source/drain, we can imagine that the parasitic capacitance between
source/drain and ground plane would kill device performance. Besides, the p+ ground plane under the source/drain can severely deplete source/drain, and the severe depletion effect is illustrated in Fig. 7.6. Compared with baseline design which only has p+ ground plane under the gate (Fig. 7.6(a)), the device variation with p+ ground plane also under the source/drain (Fig. 7.6(b)) has much lower electron density in the source/drain region next to the channel. Here, in both devices, source/drain are doped to $2 \times 10^{19}$ cm$^{-3}$ instead of nominal $5 \times 10^{19}$ cm$^{-3}$ in baseline design. Due to the high drain bias, the depletion effect is even stronger at the drain side. A straightforward deleterious result caused by the depletion effect is the terribly large source/drain series resistance.

We have shown that the SCEs are very sensitive to the source/drain doping level in the baseline device design, and lower doping can achieve better electrostatic integrity. However, heavily doped source/drain are necessary to sustain a large drain current. To get a reasonable prediction of drain current of short channel III-V MOSFETs, a state-of-art Monte-Carlo simulator is required to simulate electronic transport in the devices. Using DAMOCLES, a self-consistent Poisson/-Monte Carlo computer program by IBM corporation, Fischetti et al. have studied various device designs related to the baseline structure given in Fig. 7.2 [102]. The simulation results of $I_d$-$V_g$ characteristics with $V_{ds} = 1$V have been recurred in Fig. 7.7 in both linear and semi-log scales, while Fig. 7.8 summarizes the corresponding maximum transconductances. From the subthreshold swing shown, the good short-channel behavior exhibited by most of the device designs considered
Figure 7.7: Monte-Carlo simulation results of turn-on characteristics at $V_{ds} = 1V$ for the various device designs (curves shifted to account for the different threshold voltage in the various device designs). Adapted from [102].

It is obvious that designs with higher source/drain doping (numbered 5), raised source/drain (numbered 6), and thick channel (numbered 7) dramatically improves the drive current as well as transconductance. As transport approaches the quasi-ballistic limit, the strong flux of carriers exiting the source may cause
a “starvation” of carriers occupying longitudinal k-states (i.e., oriented along the channel direction), thus rendering the source unable to supply enough carriers and sustain a high on-current. However, all of these design variations (#5, 6, and 7) allow electrons in the source to scatter and effectively replenish longitudinal $k$-states which are injected into the channel in ballistic-transport conditions [102].

Because raised and heavily-doped source/drain can help to alleviate “source starvation” effect, our baseline device design is finally decided to be that in Fig. 7.2, which is basically the device design #6 in [102] with raised source/drain doped to $5 \times 10^{19}$ cm$^{-3}$.

As shown in Fig. 7.7 and Fig. 7.8, by using 5nm pure In$_{0.53}$Ga$_{0.47}$As channel
instead of 2.5nm $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$/2.5nm InP composite channel, device design #7
is expected to have even large on-current and maximum transconductance than baseline structure. Meanwhile, according to Fig. 7.3, using pure InGaAs channel only has a mild degradation on SCEs. Therefore, it might be a good idea to upgrade the composite channel in the baseline device design to a 5nm pure InGaAs channel. But actually we are trading off low ground plane leakage for high drain current. With the substitution done, it is possible that we have large ground plane leakage current, but we are able to suppress it by slightly increasing the thickness of InAlAs bottom barrier. This is a trade-off again.

7.4 Summary

In conclusion, we have proposed a baseline device design of III-V MOSFETs for sub-22nm scaling. The design is based on the thin-BOX-SOI-like structure, therefore it can be scaled down to 19nm roughly without pushing the film thicknesses to the very bottom. Since the film thicknesses are not at physical limits yet, the design considerations are mostly on SCEs. The sensitivity of SCEs to different device variations have been studied, and these variations include pure InGaAs channel, source/drain doping, InAlAs barrier thickness, GaAs cap layer thickness, and high-$\kappa$ EOT. We have also clarified some design considerations beyond SCEs in the baseline structure, e.g., achieving high on-current by reducing source/drain series resistance and alleviate “source starvation” effect.
Conclusion

This dissertation presents a comprehensive study on the design and modeling of non-classical MOSFETs, including MG MOSFETS based on Si as well as novel III-V MOSFETs.

First of all, we have introduced the complete analytic models of drain current, terminal charges and capacitance coefficients for long channel symmetric DG and SG MOSFETs. Starting from the exact solution of Poisson’s equation, we derive the models from current continuity equation and Ward-Dutton charge partition without the need for charge sheet approximation. Inasmuch as the essential physics has been preserved, it has been verified by numerical simulations that all the operation and transition regions are properly described by continuous functions in the absence of non-physical fitting parameters. There are implicit equations for the intermediate parameters to be solved in the DG and SG models, but explicit
solutions for the intermediate parameters have been developed. The explicit solutions are accurate not only in terms of drain current and terminal charges but also their derivatives, including drain conductance, transconductance, and transcapacitances. With these explicit solutions, the models are ideal candidates of compact model for DC, AC, and transient circuit simulation of DG and SG MOSFETs.

With some non-trivial yet reasonable approximations, the DG and SG models have also been generalized to a unified analytic drain current model for all kinds of MG MOSFETs, including DG, SG, QG, TG, and TPG MOSFETs. The accuracy of the unified model is shown by comparing the inversion charge density as a function of gate voltage to 2-D numerical simulation results. The drain current is then obtained by Pao-Sah integral of the inversion charge density. The analytic drain current model serves as a core long-channel model that can be expanded for compact modeling of various kinds of MG MOSFETs.

One of the most important add-ons to the core long-channel model is the model for SCEs. Based on generalized scale length theory, full analytical expressions of potential and subthreshold current have been derived for both short-channel DG and SG MOSFETs by solving 2-D Poisson’s equation in both the semiconductor and insulator regions. The compact models for SCEs including threshold voltage roll-off, DIBL, and subthreshold slope degradation are extracted from the subthreshold current expression. It has been verified by 2-D numerical simulation results that the SCE model is predictive with no fitting parameters.
Apart from compact modeling of Si-based MG MOSFETs, we have also developed a comprehensive equivalent capacitance circuit model of an MOS capacitor. With the one subband approximation, analytical results for the capacitance components in the equivalent circuit model can be developed in terms of effective thickness, by investigating both Poisson’s equation and Schrödinger’s equation. These analytical results can help us to physically understand the capacitances in III-V MOSFETs, especially the inversion layer capacitance. Besides, these analytical results can predict the degradation of gate insulator capacitance in inversion charge gate capacitance $C_{inv}$, and degradation of DOS capacitance in non-equilibrium in ballistic MOSFETs. The effects of interface traps have also been examined by analyzing the equivalent circuit model with interface states capacitance incorporated.

Finally, we have proposed a baseline device design of III-V MOSFETs for sub-22nm scaling. The design is based on the thin-BOX-SOI-like structure, therefore it can be scaled down to 19nm roughly without pushing the film thicknesses to the very bottom. Since the film thicknesses are not at physical limits yet, the design considerations are mostly on SCEs. The sensitivity of SCEs to different device variations have been studied, and these variations include pure InGaAs channel, source/drain doping, InAlAs barrier thickness, GaAs cap layer thickness, and high-$\kappa$ EOT. We have also clarified some design considerations beyond SCEs in the baseline structure, e.g., achieving high on-current by reducing source/drain series resistance and alleviate “source starvation” effect.
Bibliography


BSIM model is available at http://www-device.eecs.berkeley.edu/bsim3/.

PSP model is available at http://pspmodel.asu.edu/.

HiSIM model is available at http://home.hiroshima-u.ac.jp/usdl/.


